



Machine Learning for Molecular Physics

WiSe 2021/22

L06 - GAP and long-range interactions



1. Gaussian Approximation Potential
2. Long-range interactions
3. “Nitty gritties”

Gaussian Approximation Potential

Gaussian approximation potential

GAP

- proposed in 2010
- smooth overlap of atomic positions (SOAP) + Gaussian process regression

PRL 104, 136403 (2010)

PHYSICAL REVIEW LETTERS

week ending
2 APRIL 2010

Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons

Albert P. Bartók and Mike C. Payne

Cavendish Laboratory, University of Cambridge, J J Thomson Avenue, Cambridge, CB3 0HE, United Kingdom

Risi Kondor

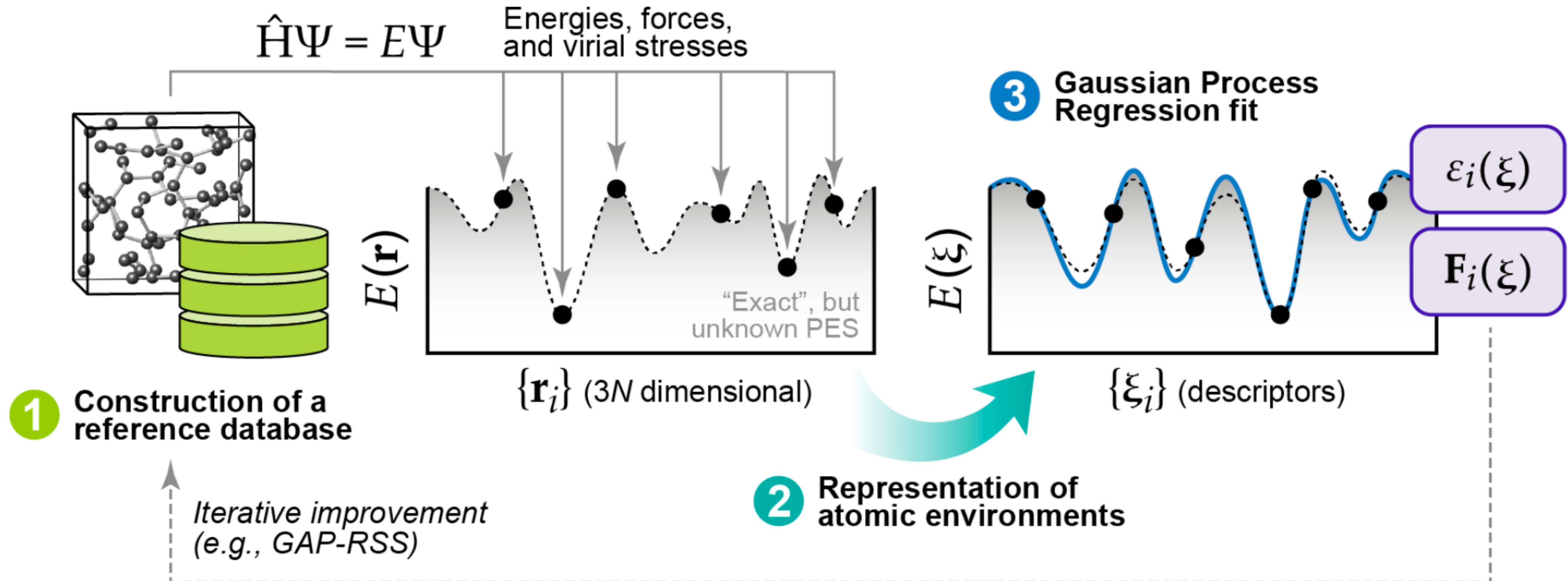
Center for the Mathematics of Information, California Institute of Technology, MC 305-16, Pasadena, California 91125, USA

Gábor Csányi

Engineering Laboratory, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom

(Received 1 October 2009; published 1 April 2010)

Gaussian approximation potential



Deringer et al., Chem. Rev 121, 10073 (2021)

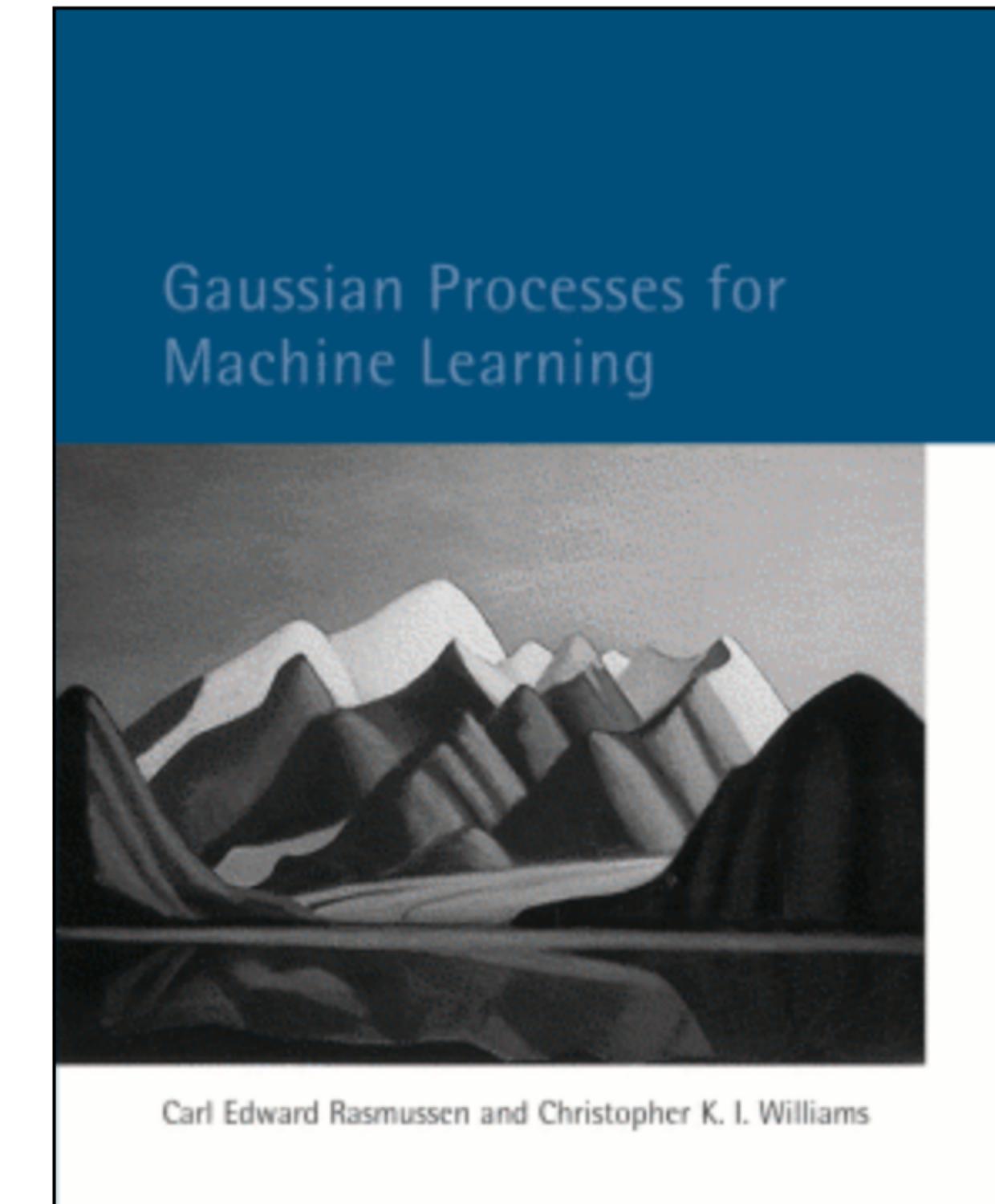
Gaussian process regression



- general introduction:

Gaussian Processes for Machine Learning

Carl Edward Rasmussen and Christopher K. I. Williams
The MIT Press, 2006. ISBN 0-262-18253-X.



Gaussian process regression

- application to materials:

CHEMICAL REVIEWS

pubs.acs.org/CR



Review

Gaussian Process Regression for Materials and Molecules

Volker L. Deringer,* Albert P. Bartók,* Noam Bernstein, David M. Wilkins, Michele Ceriotti, and Gábor Csányi*



Cite This: *Chem. Rev.* 2021, 121, 10073–10141



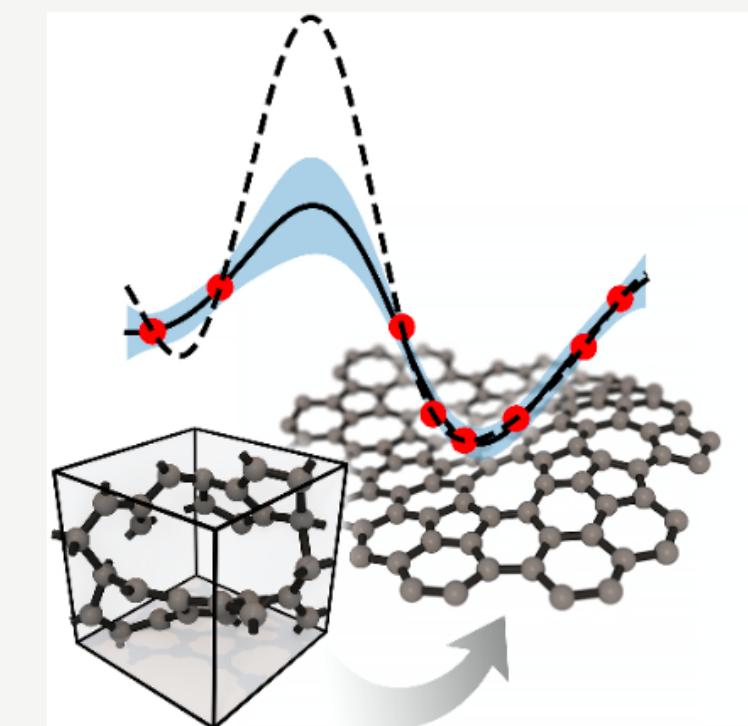
Read Online

ACCESS |

Metrics & More

Article Recommendations

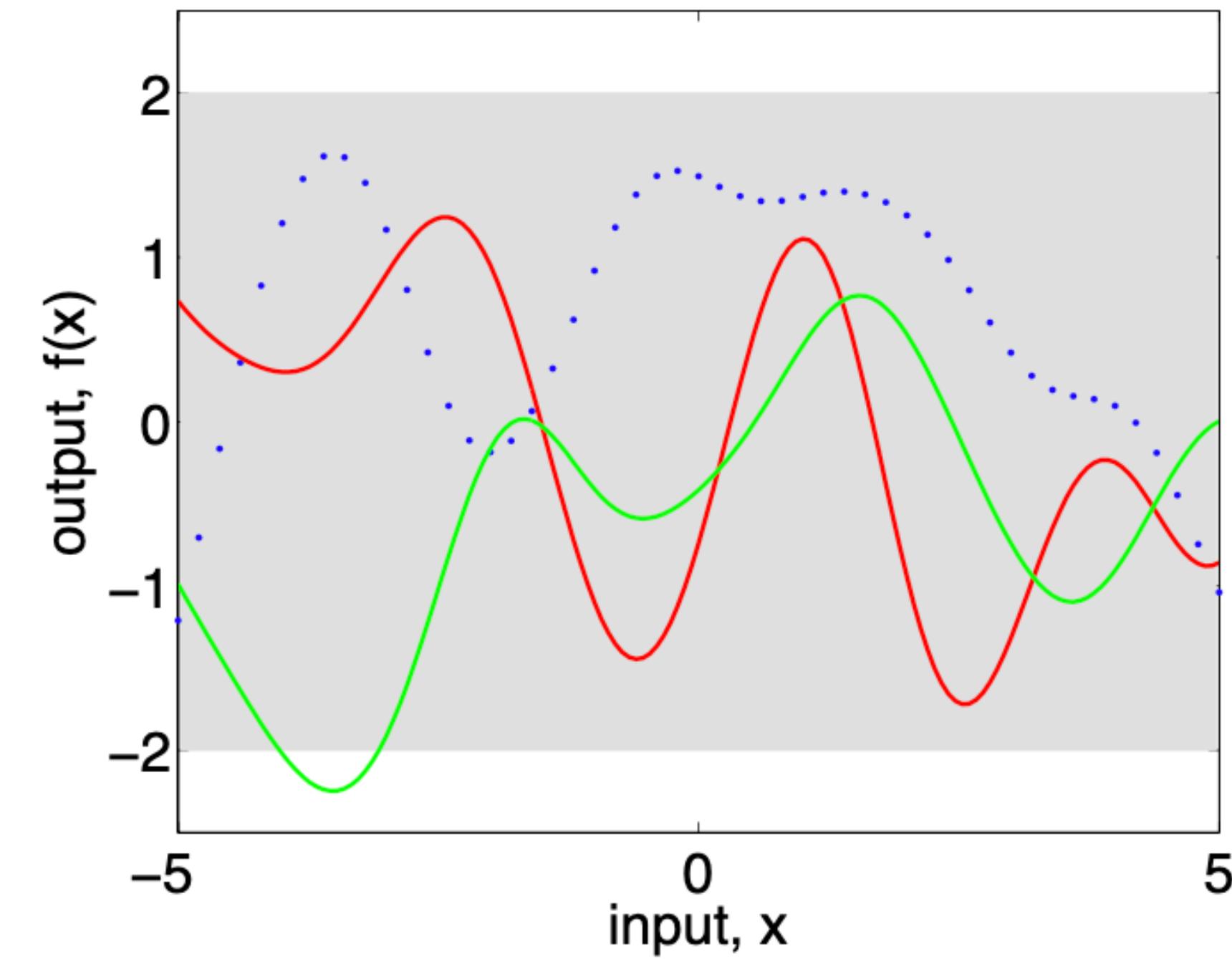
ABSTRACT: We provide an introduction to Gaussian process regression (GPR) machine-learning methods in computational materials science and chemistry. The focus of the present review is on the regression of atomistic properties: in particular, on the construction of interatomic potentials, or force fields, in the Gaussian Approximation Potential (GAP) framework; beyond this, we also discuss the fitting of arbitrary scalar, vectorial, and tensorial quantities. Methodological aspects of reference data generation, representation, and regression, as well as the question of how a data-driven model may be validated, are reviewed and critically discussed. A survey of applications to a variety of research questions in chemistry and materials science illustrates the rapid growth in the field. A vision is outlined for the development of the methodology in the years to come.



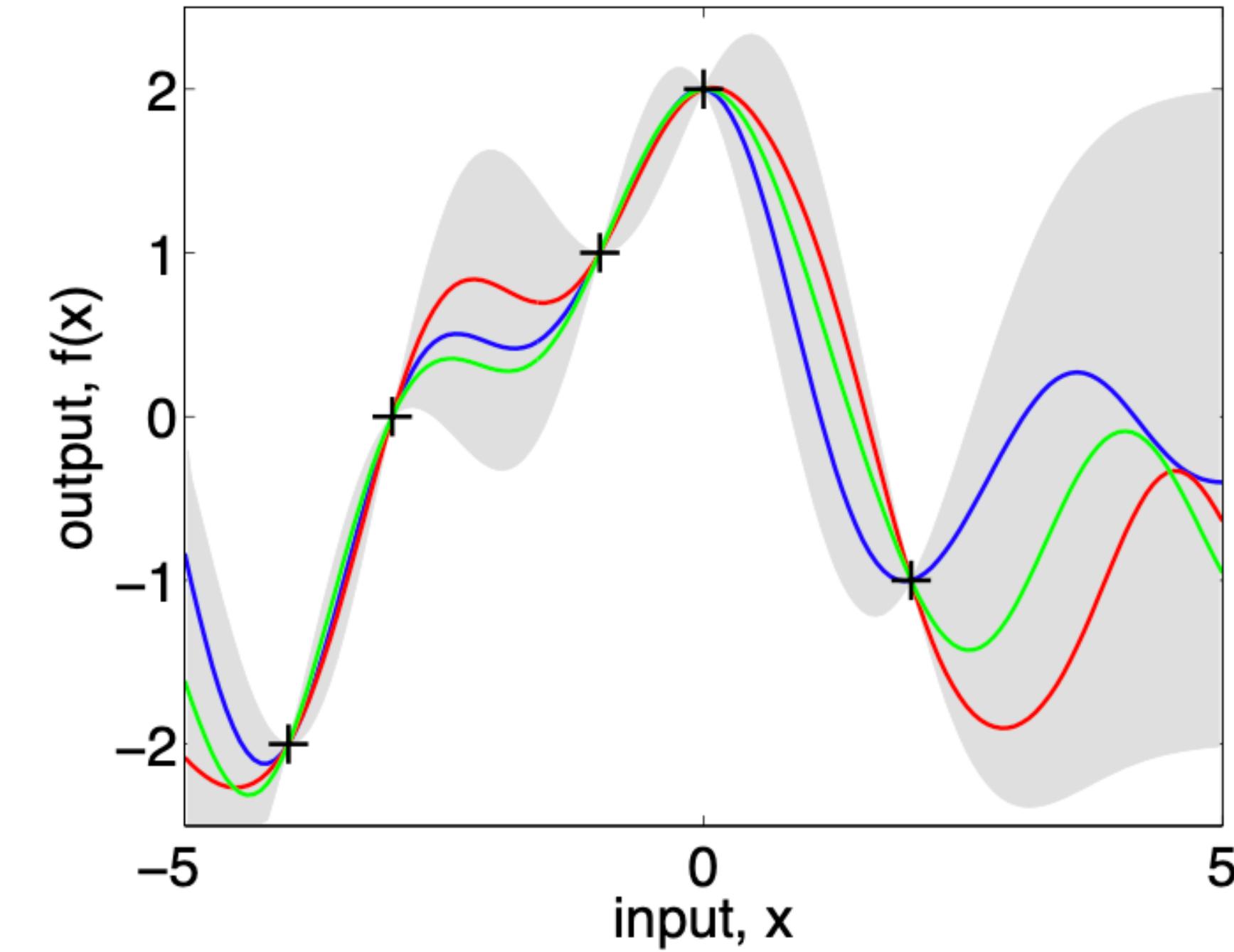
Gaussian process



- **Definition:** “A Gaussian process *is a collection of random variables, any finite number of which have a joint Gaussian distribution.*”
- “Bayesian version of support vector machines”
- non-parametric model



(a), prior



(b), posterior

Weight-space view of GPR

- dataset

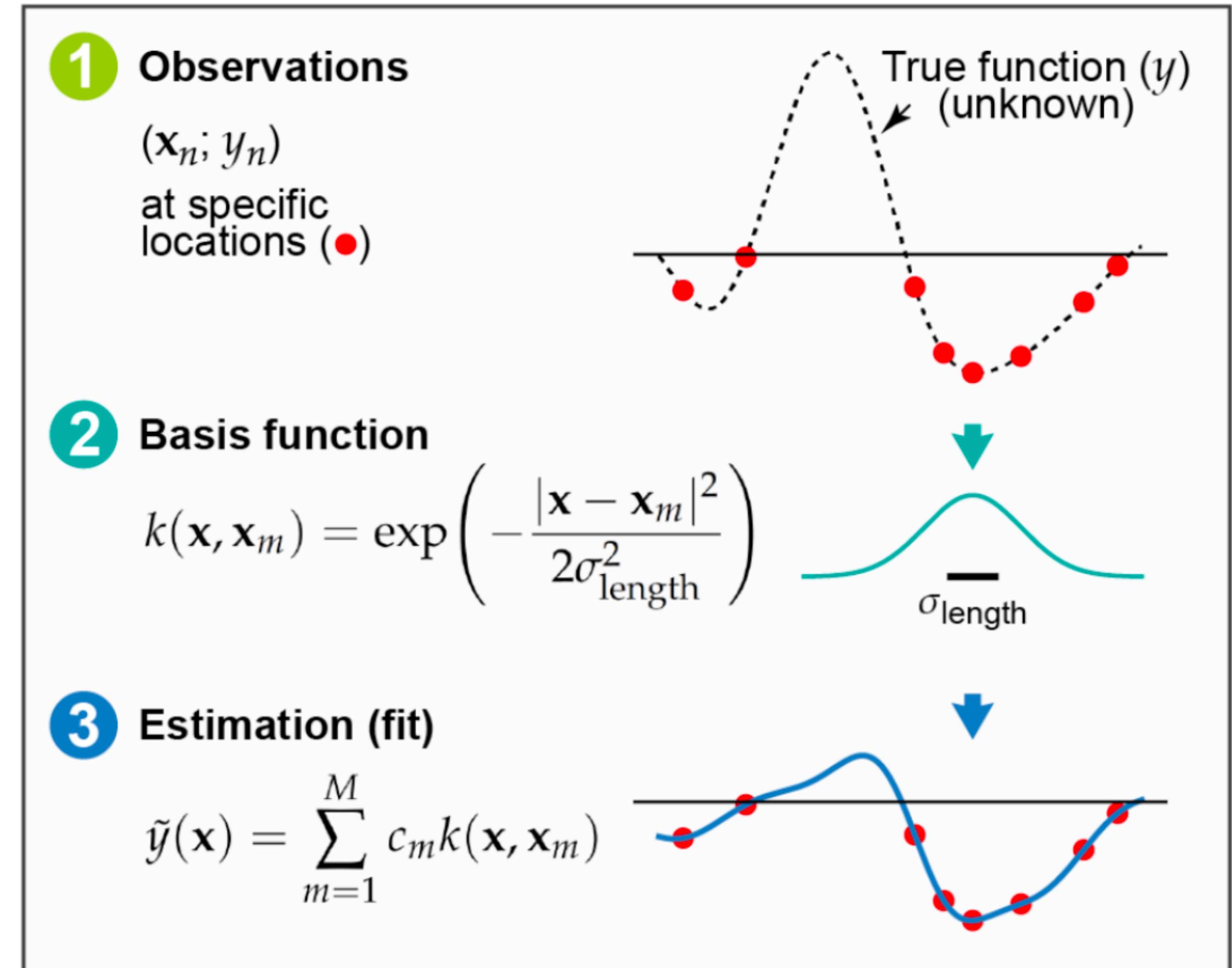
$$\mathcal{D} = \{\mathbf{x}_n; y_n\}_{n=1}^N$$

- similar to kernel ridge regression

$$\tilde{y}(\mathbf{x}) = \sum_{m=1}^M c_m k(\mathbf{x}, \mathbf{x}_m)$$

- loss function

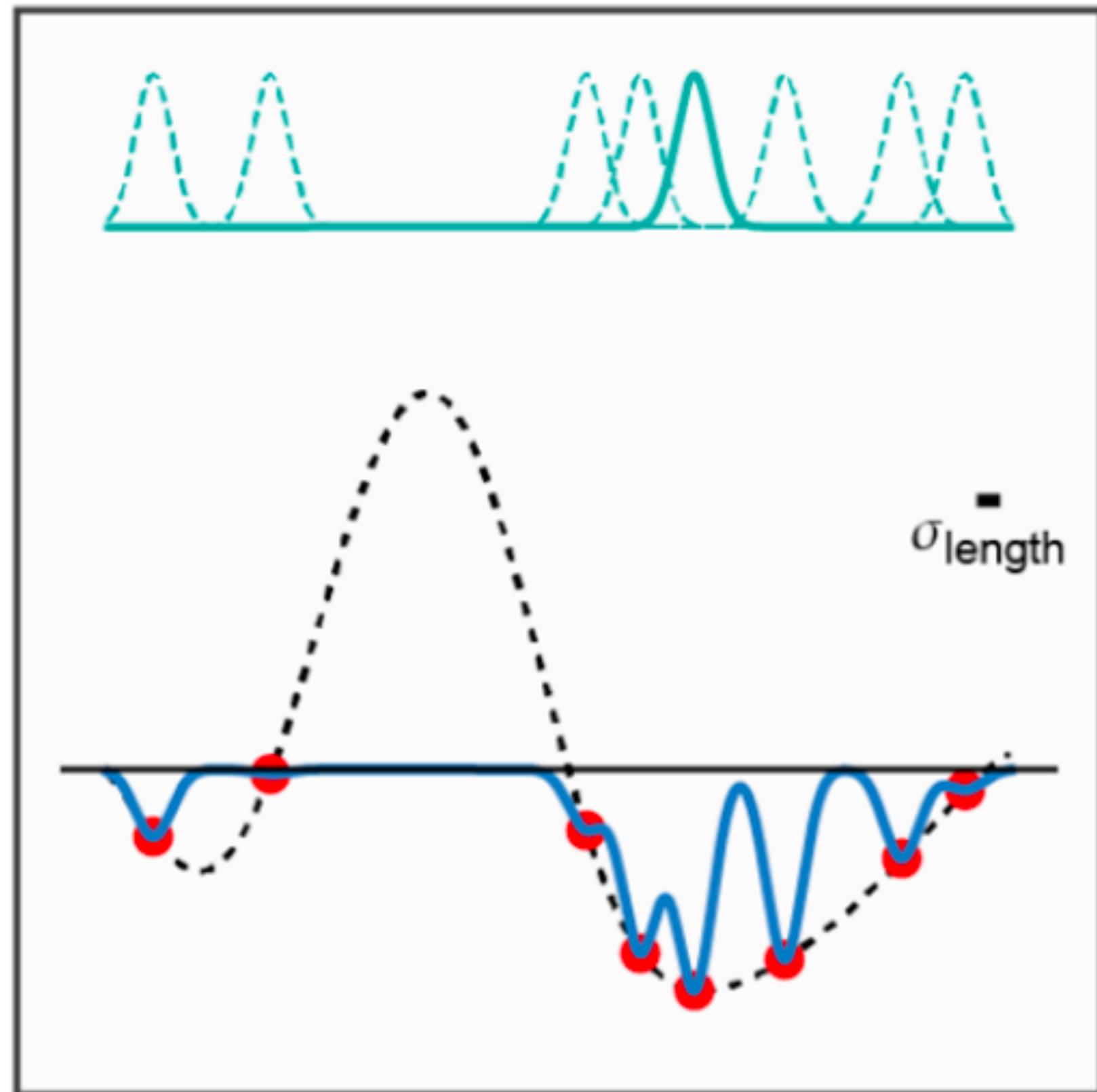
$$\mathcal{L} = \sum_{n=1}^N [y_n - \tilde{y}(\mathbf{x}_n)]^2 + \sigma^2 \sum_{m,m'=1}^M c_m k(\mathbf{x}_m, \mathbf{x}_{m'}) c_{m'}$$



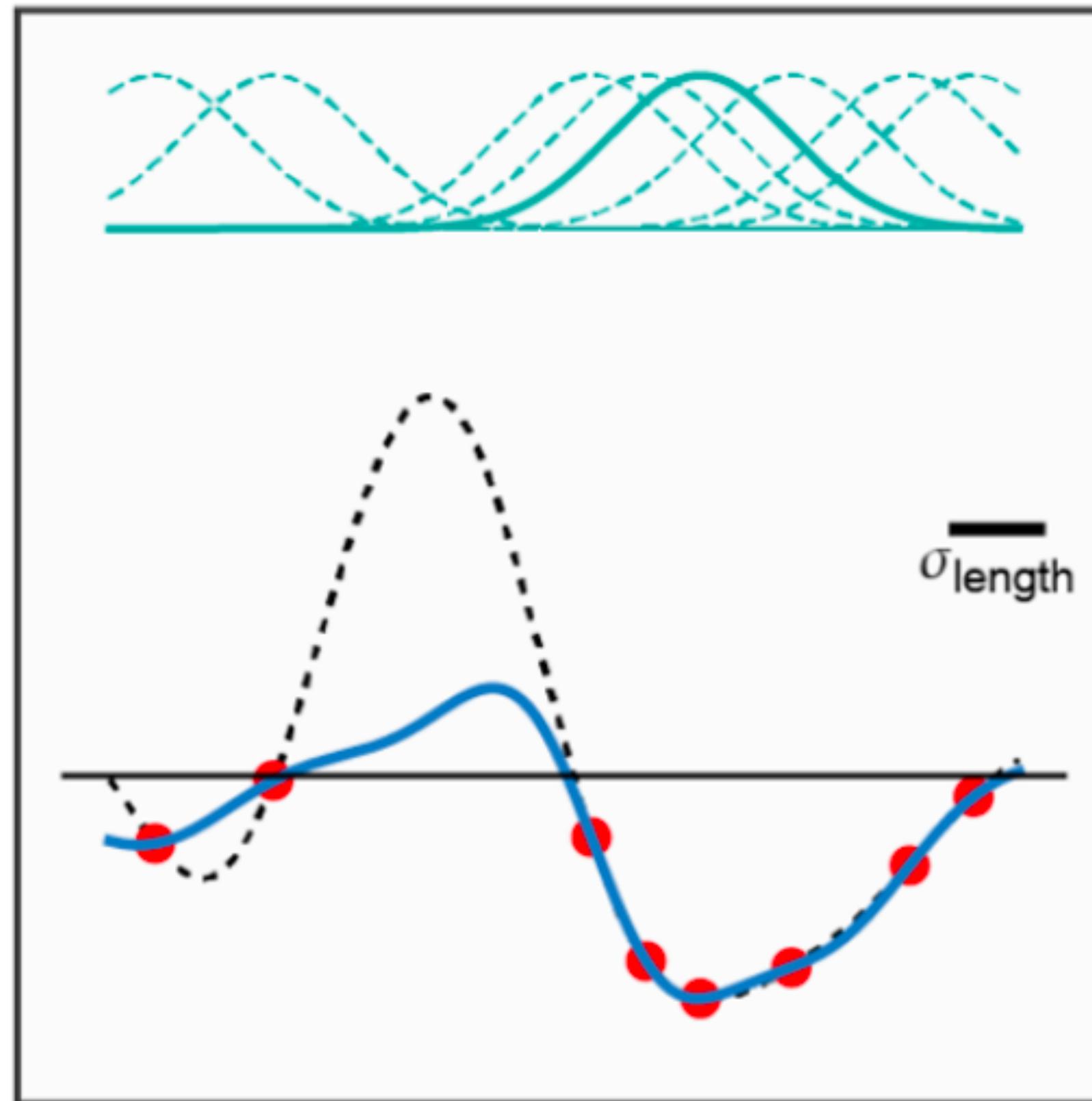
Deringer et al., Chem. Rev 121, 10073 (2021)

Kernel function

Too small σ_{length} (overfitting)



Appropriate σ_{length}

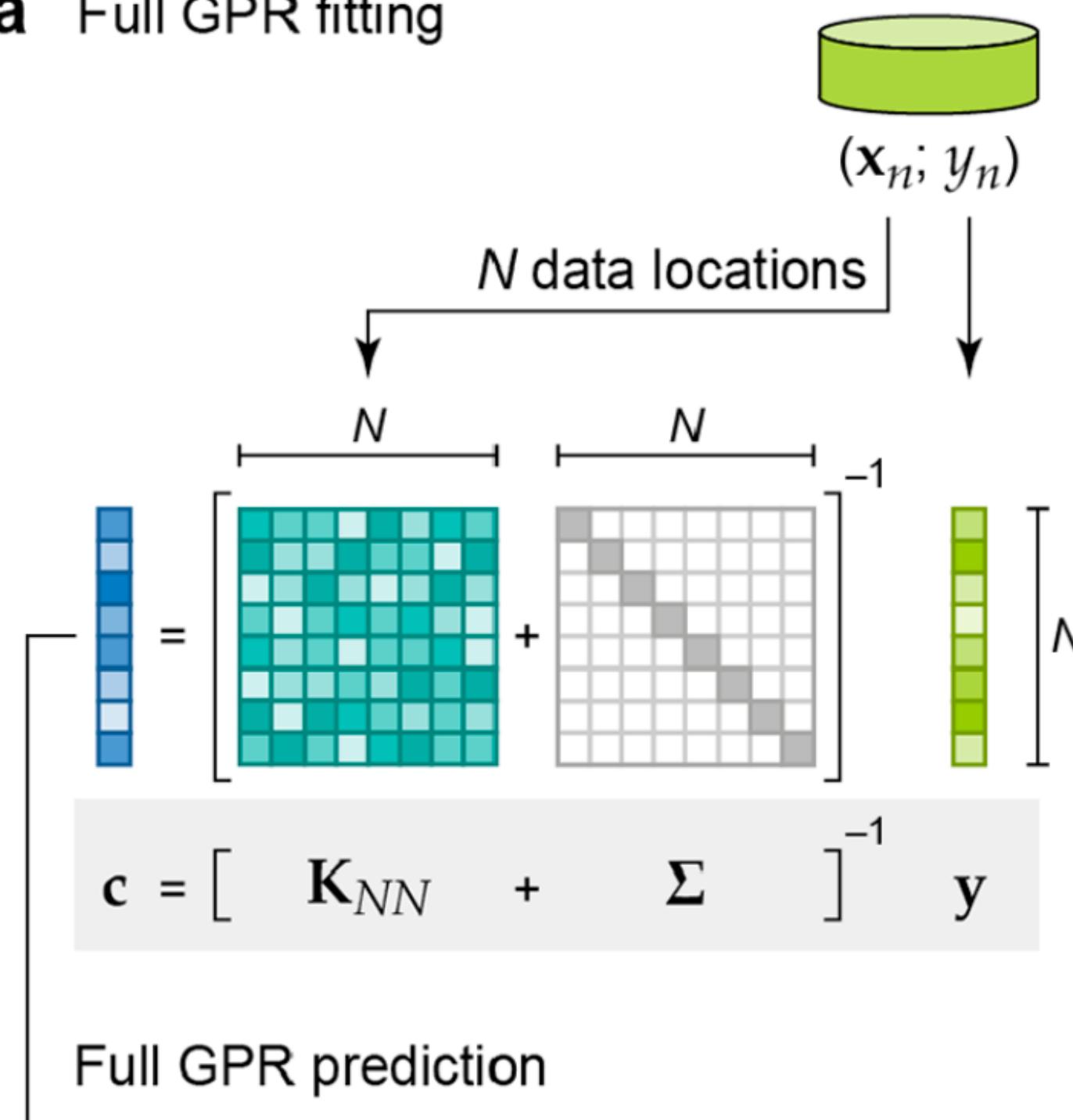


$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\sigma_{\text{length}}^2}\right)$$

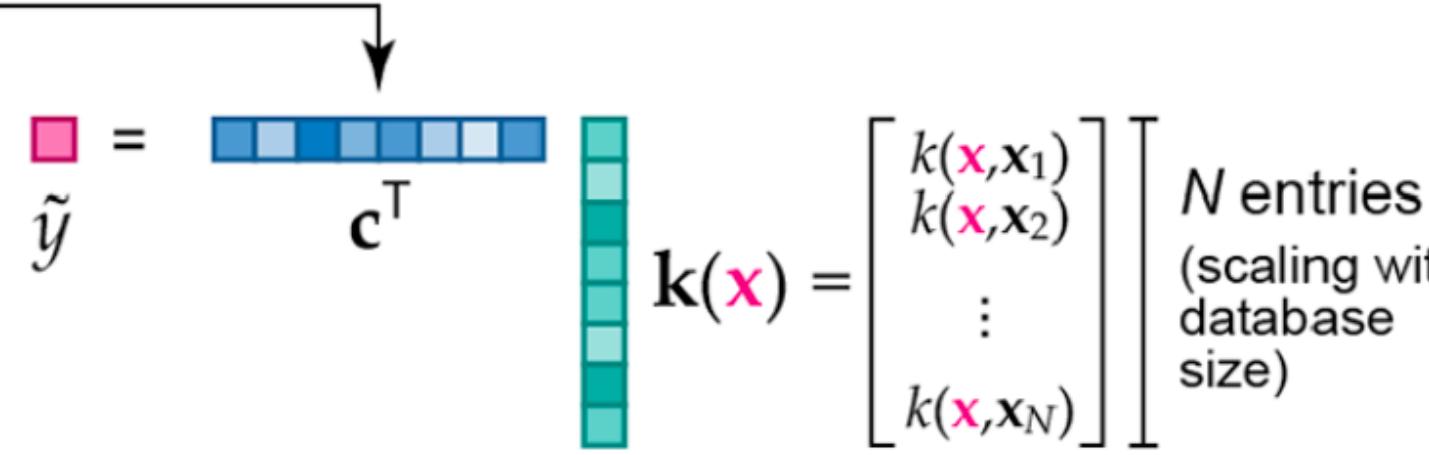
Deringer et al., Chem. Rev 121, 10073 (2021)

Sparse GPR prediction

a Full GPR fitting



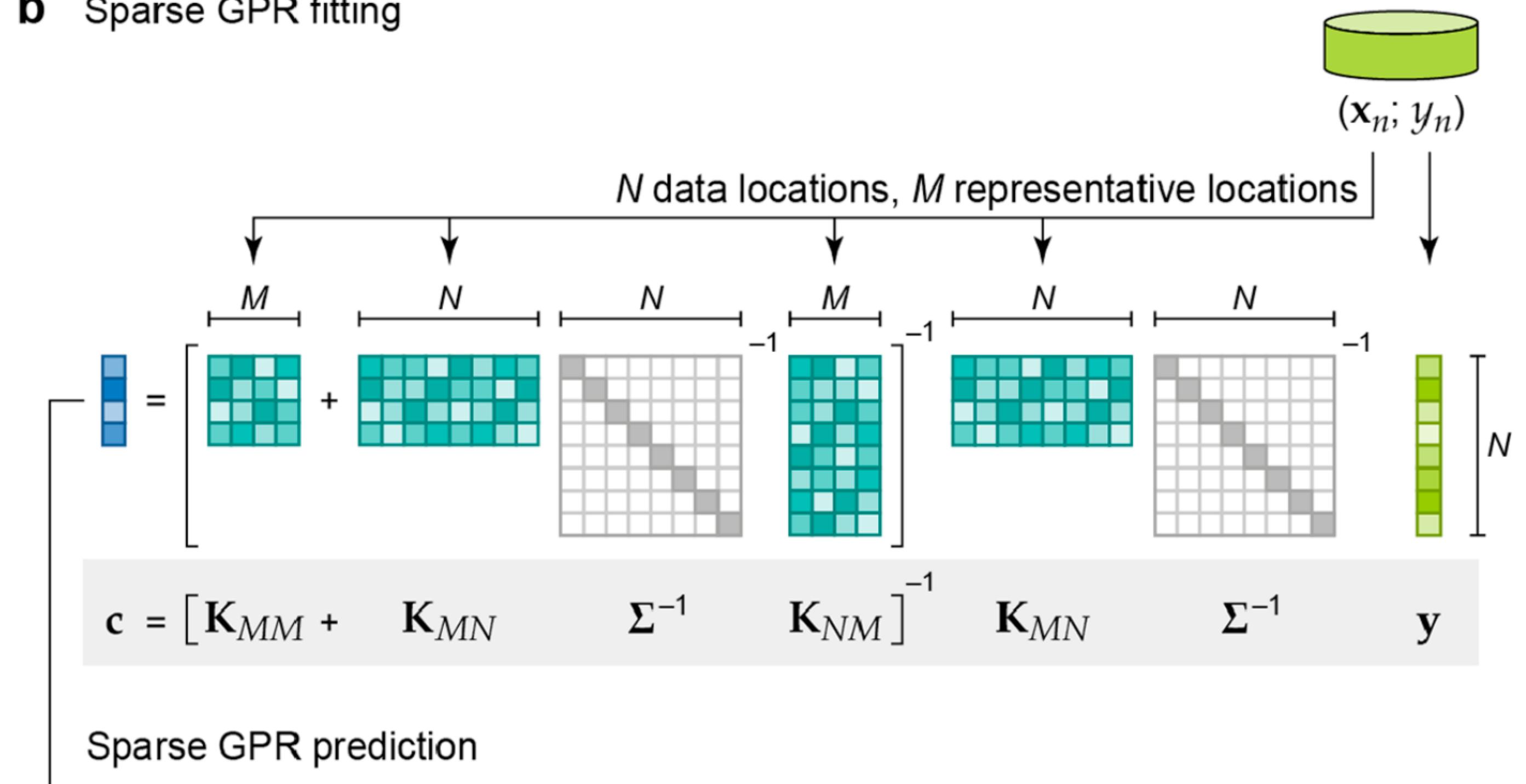
Full GPR prediction



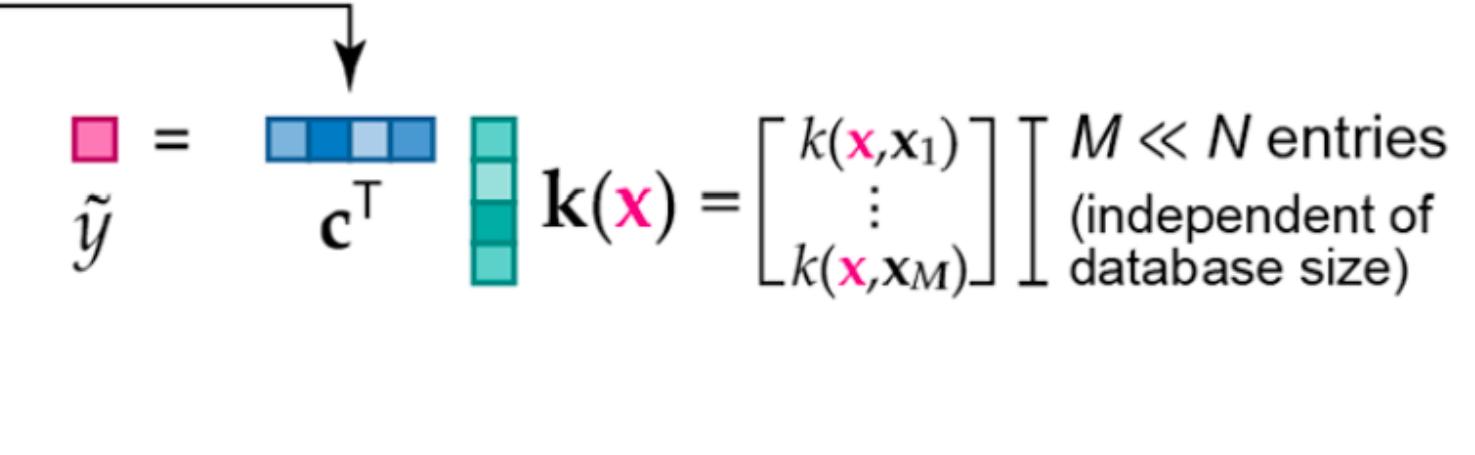
The prediction \tilde{y} is calculated as the product of the transpose of the coefficient vector \mathbf{c}^T and the kernel matrix $\mathbf{k}(\mathbf{x})$. The kernel matrix $\mathbf{k}(\mathbf{x})$ is defined as:

$$\mathbf{k}(\mathbf{x}) = \begin{bmatrix} k(\mathbf{x}, \mathbf{x}_1) \\ k(\mathbf{x}, \mathbf{x}_2) \\ \vdots \\ k(\mathbf{x}, \mathbf{x}_N) \end{bmatrix} \quad \text{N entries (scaling with database size)}$$

b Sparse GPR fitting



Sparse GPR prediction



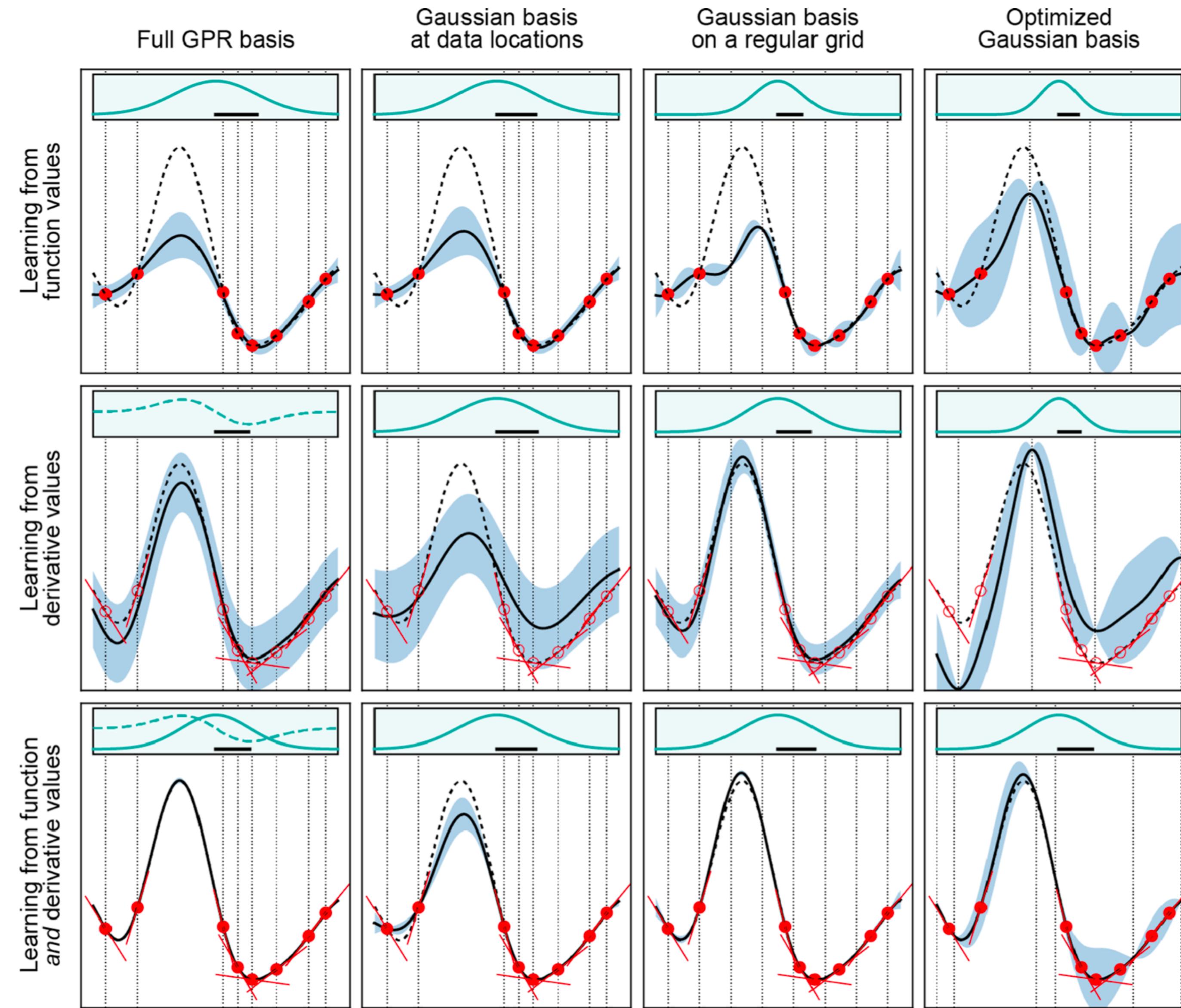
The prediction \tilde{y} is calculated as the product of the transpose of the coefficient vector \mathbf{c}^T and the kernel matrix $\mathbf{k}(\mathbf{x})$. The kernel matrix $\mathbf{k}(\mathbf{x})$ is defined as:

$$\mathbf{k}(\mathbf{x}) = \begin{bmatrix} k(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ k(\mathbf{x}, \mathbf{x}_M) \end{bmatrix} \quad M \ll N \text{ entries (independent of database size)}$$

■ Reference data (y)
■ Kernel values (k)
■ Coefficients (c)
■ Prediction ($\mathbf{x} \rightarrow \tilde{y}$)

Deringer et al., Chem. Rev 121, 10073 (2021)

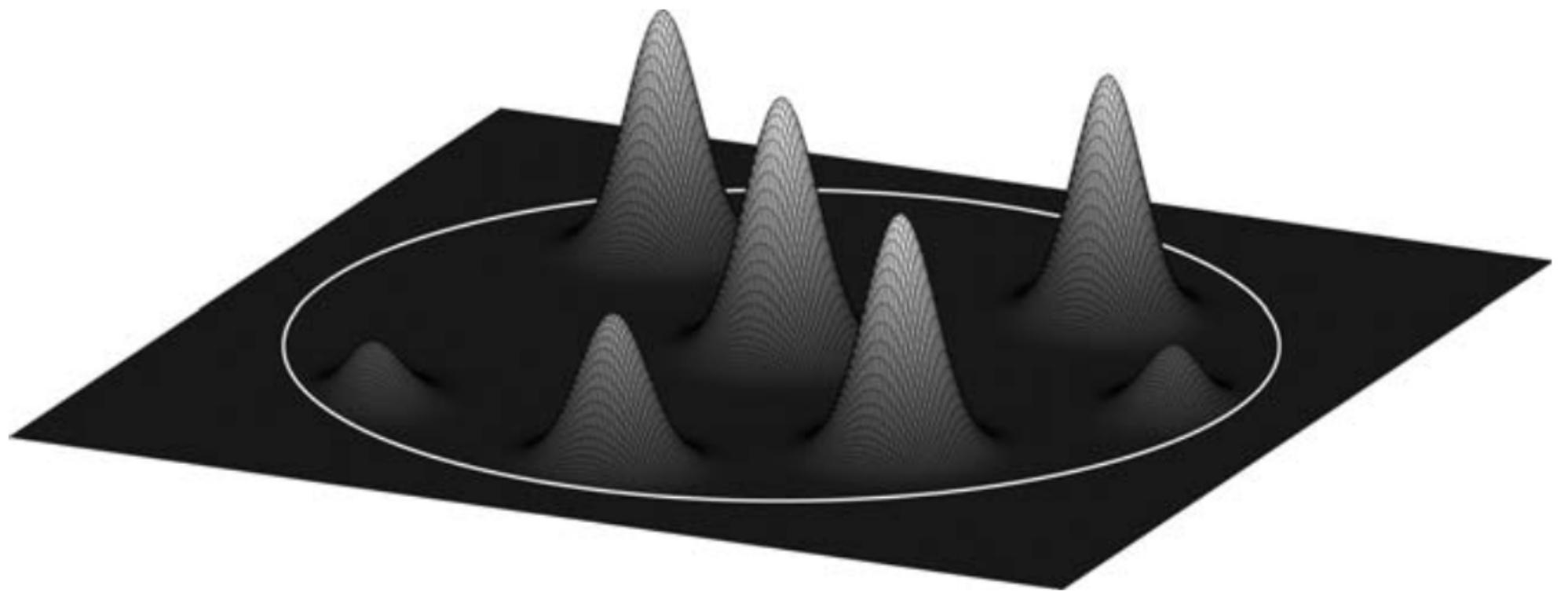
Fit quality



Smooth overlap of atomic positions

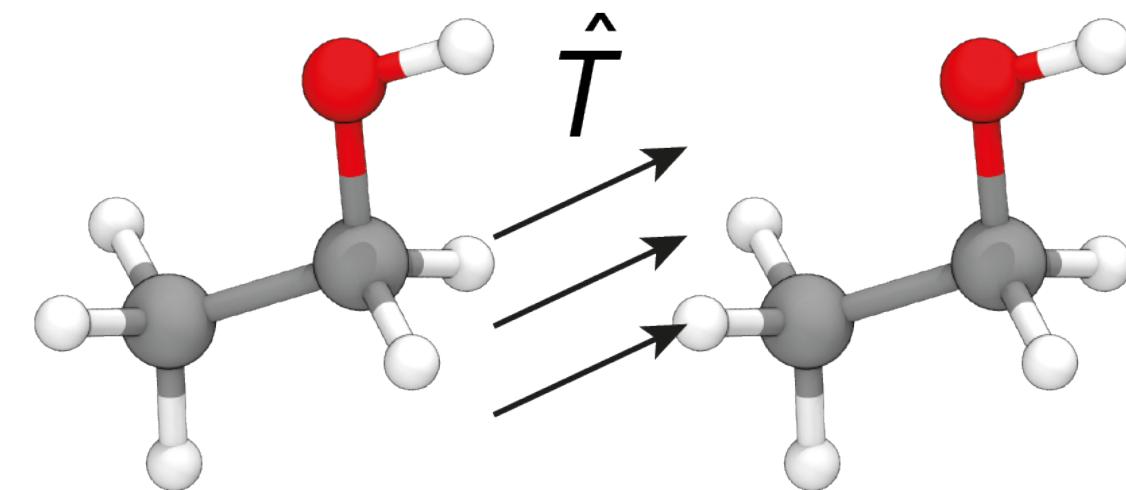
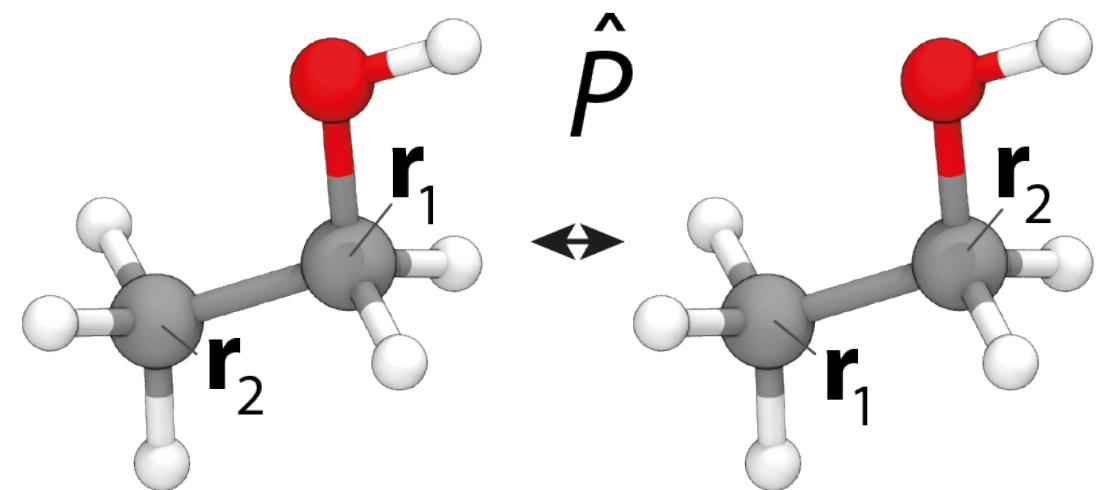
- SOAP
- set of neighbour densities

$$\rho^{i,a}(\mathbf{r}) = \sum_j \delta_{aa_j} \exp\left[\frac{-|\mathbf{r} - \mathbf{r}_{ij}|^2}{2\sigma_a^2}\right] f_{\text{cut}}(r_{ij})$$



Bartók, Csányi, Int. J. Quant. Chem. 115, 1051 (2015)

- invariant wrt. to permutation and translation



SOAP - rotational invariance

- expand in terms of radial functions and spherical harmonics

$$\rho^{i,a}(\mathbf{r}) = \sum_{nlm} c_{nlm}^{i,a} R_n(r) Y_l^m(\hat{\mathbf{r}})$$

$$c_{nlm}^{i,a} = \int d\mathbf{r} R_n(r)^* Y_l^m(\hat{\mathbf{r}})^* \rho^{i,a}(\mathbf{r})$$

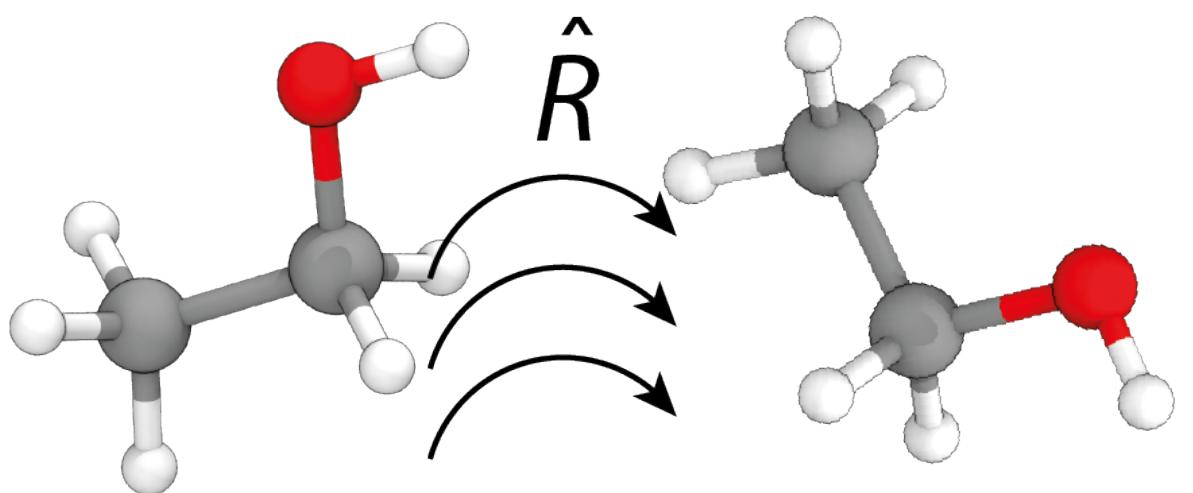
- change of basis, not rotationally invariant

SOAP - rotational invariance

- power spectrum

$$p_{nn'l}^{i,aa'} = \frac{1}{\sqrt{2l+1}} \sum_m (c_{nlm}^{i,a})^* c_{n'lm}^{i,a'}$$

- rotational invariant



- SOAP descriptor is vector hundreds/thousands of entries

- similarity between two environments

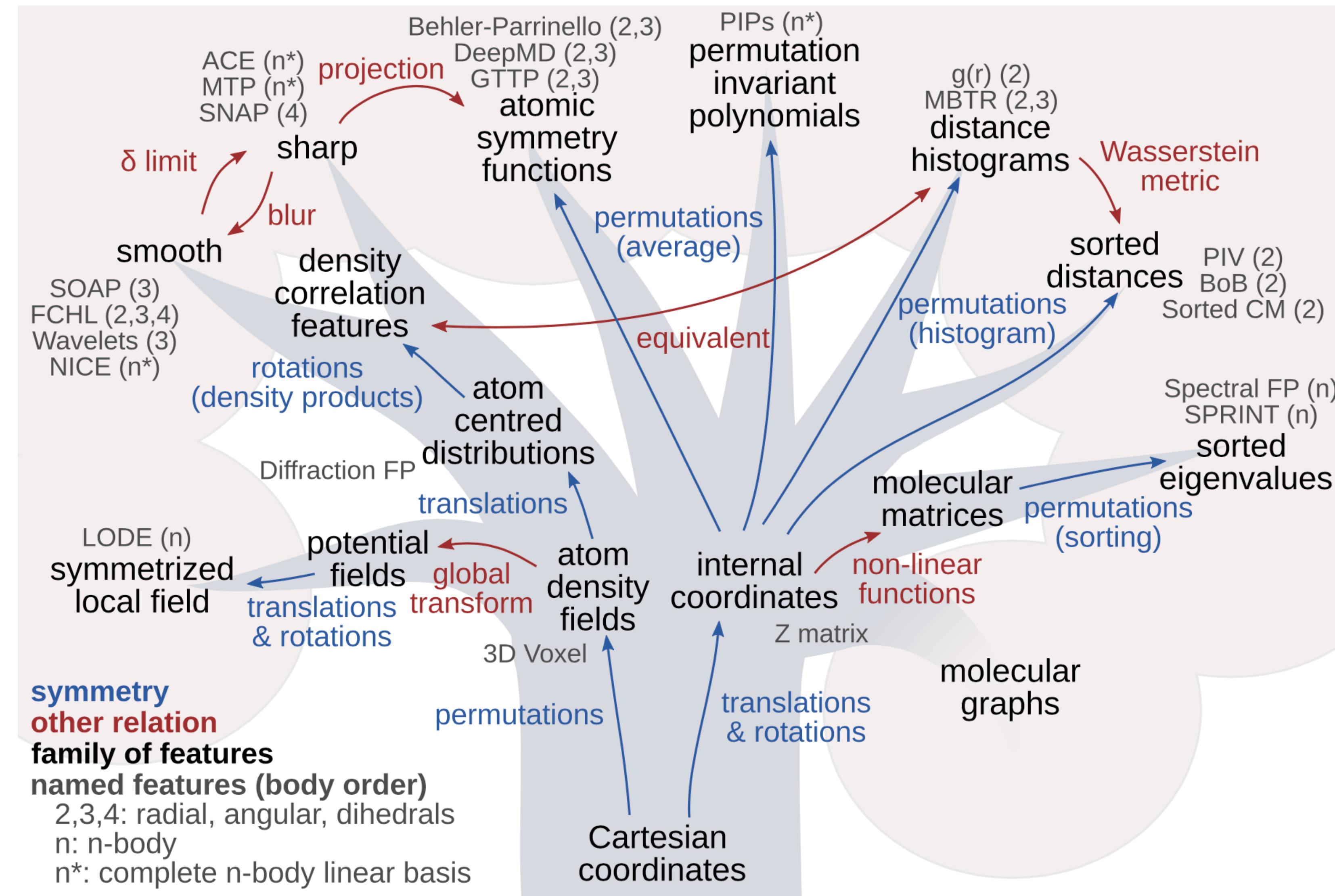
$$k(A, A') = (\xi \cdot \xi')^\zeta$$

$$\xi = \mathbf{p}/|\mathbf{p}|$$

$$\mathbf{p}_i = \{p_{nn'l}^{i,aa'}\}$$

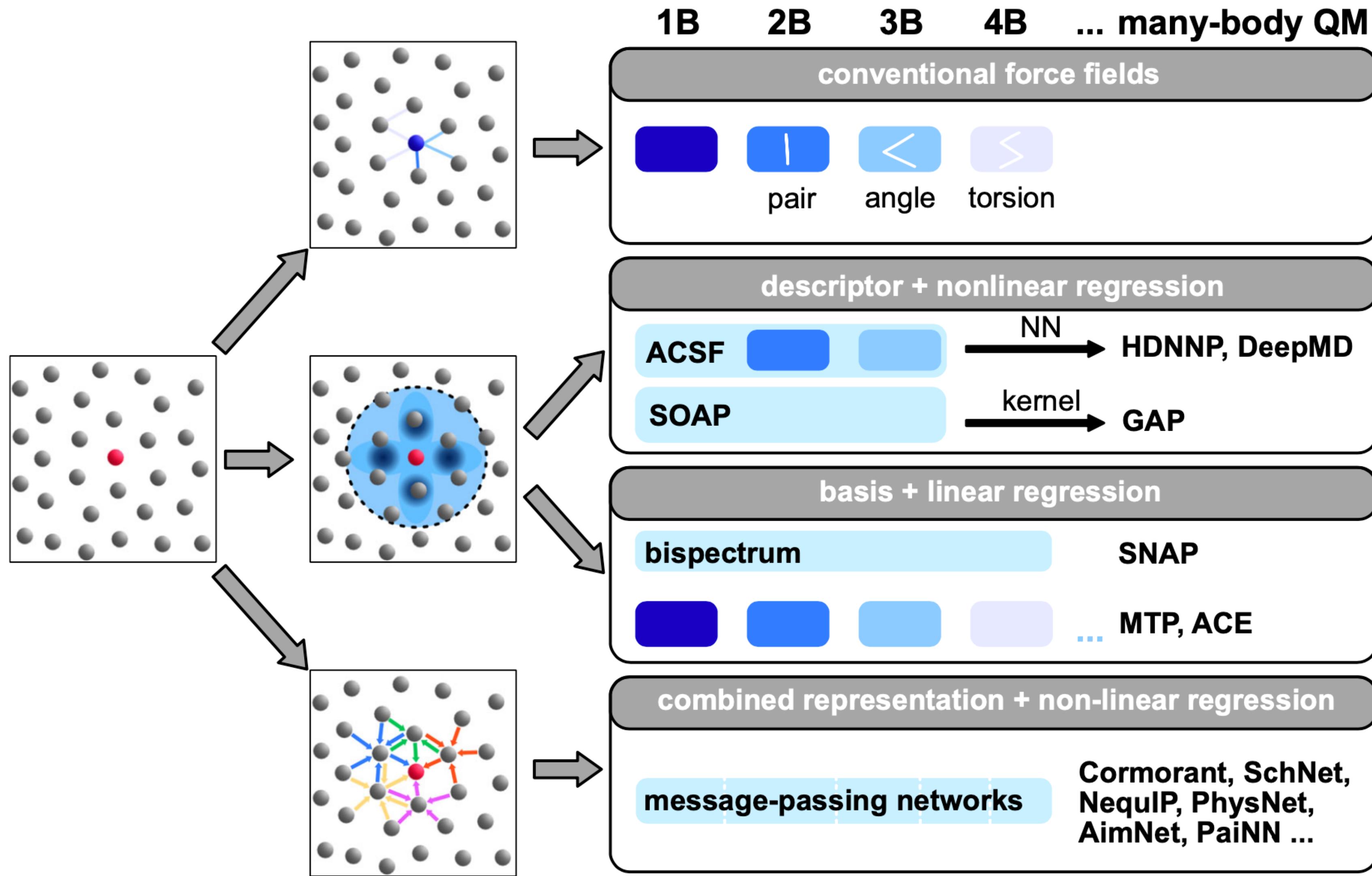
Structural representation

Structural representation



Musil et al., Chem. Rev. 121, 9759 (2021)

ML potential with local environments



Behler, Csányi, Eur. Phys. J. B 94, 142 (2021)

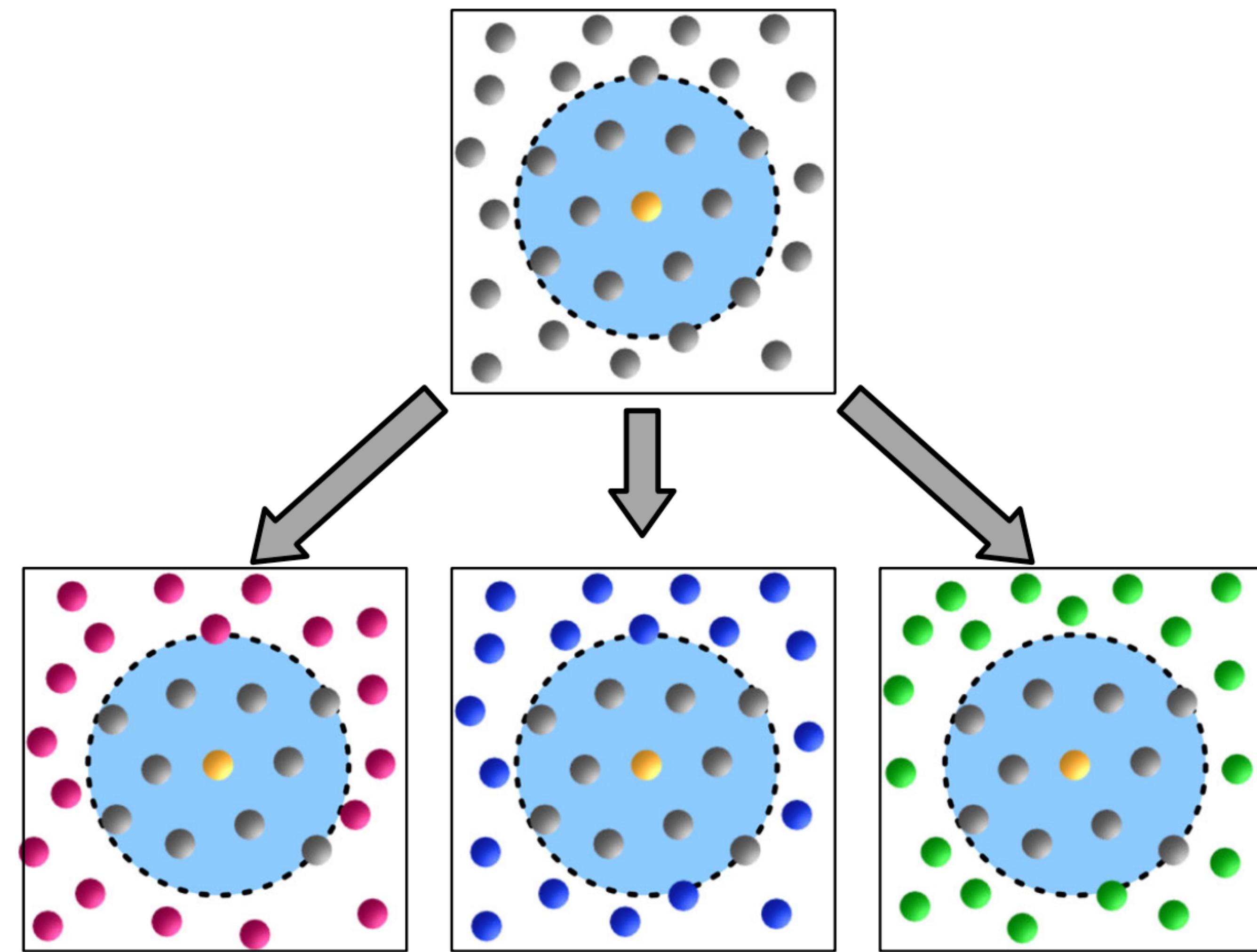
Long-range interactions

Long-range interactions

- *local* atomic environment (locality approximation)
- interactions beyond cutoff result in inconsistent data ('noise')
- long-range interactions:
 - electrostatics
 - dispersion interactions (Van der Waals)
 - delocalised electrons in aromatic systems
 - conjugated π -systems

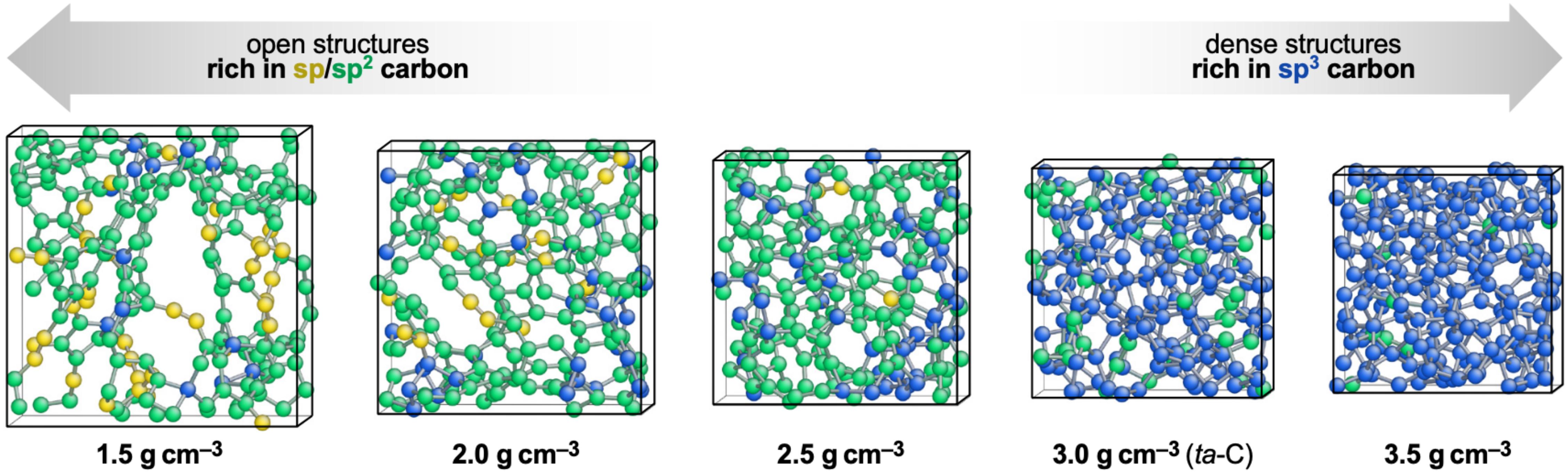
Quantifying long-range interactions

- locality test



Behler, Csányi, Eur. Phys. J. B 94, 142 (2021)

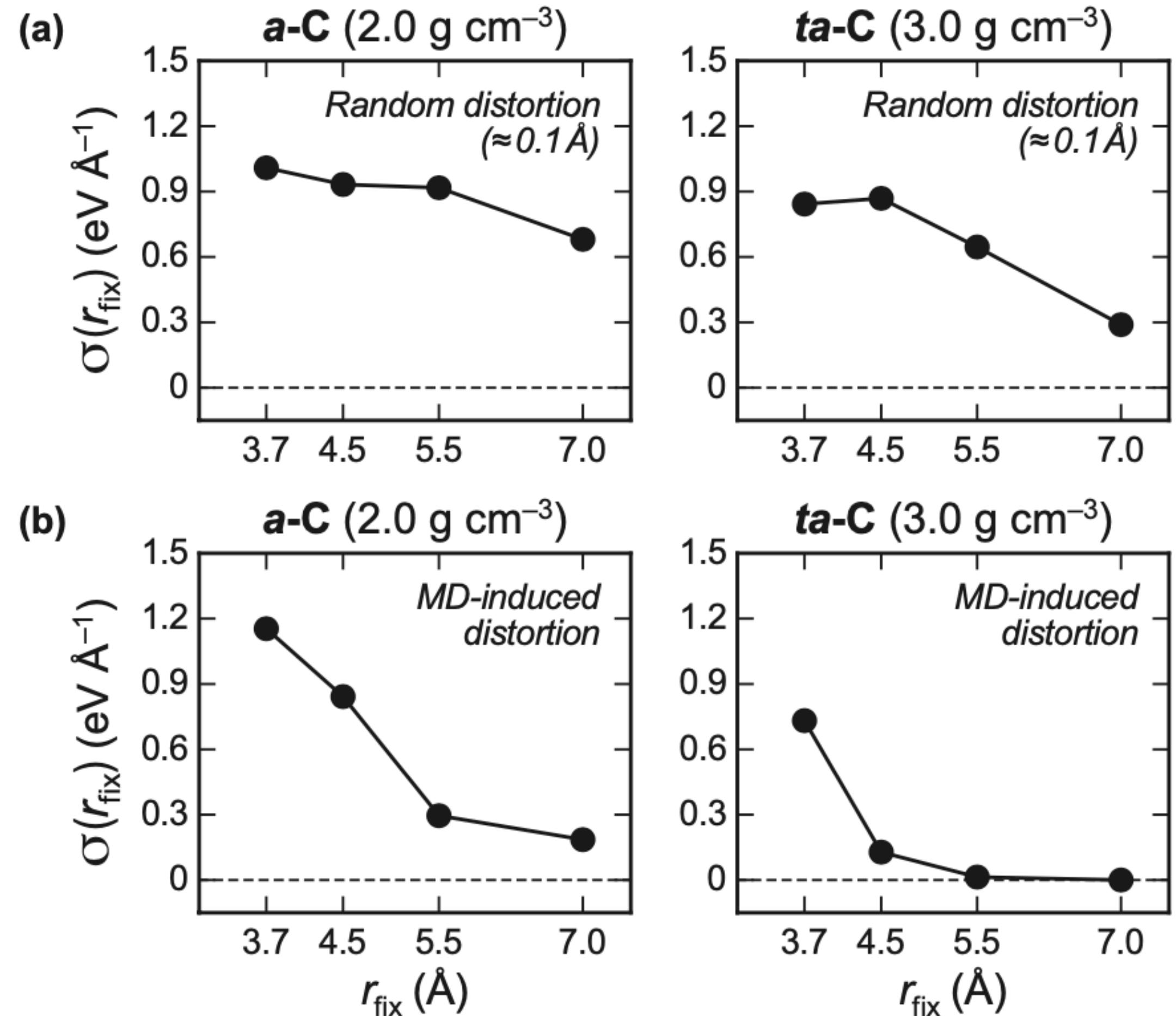
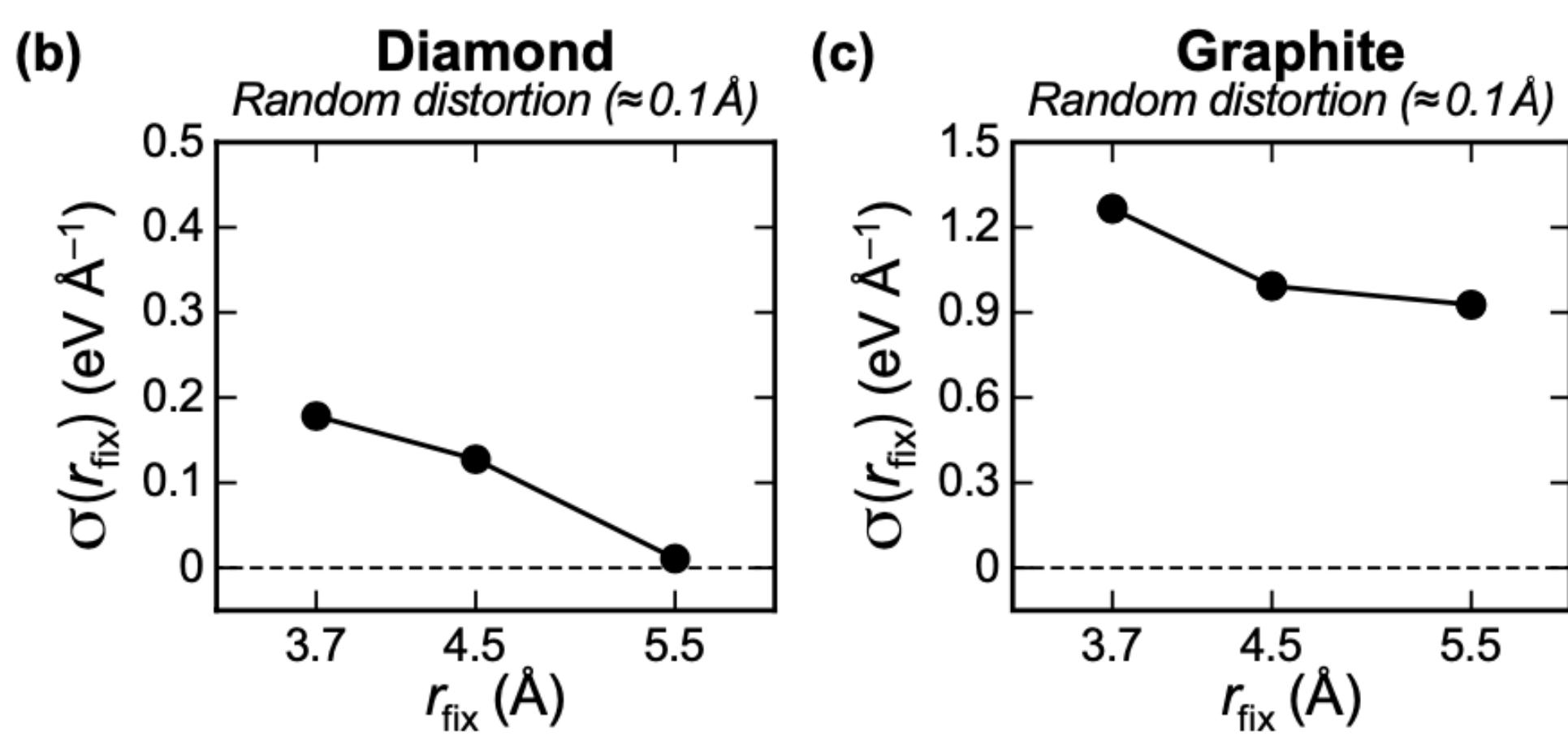
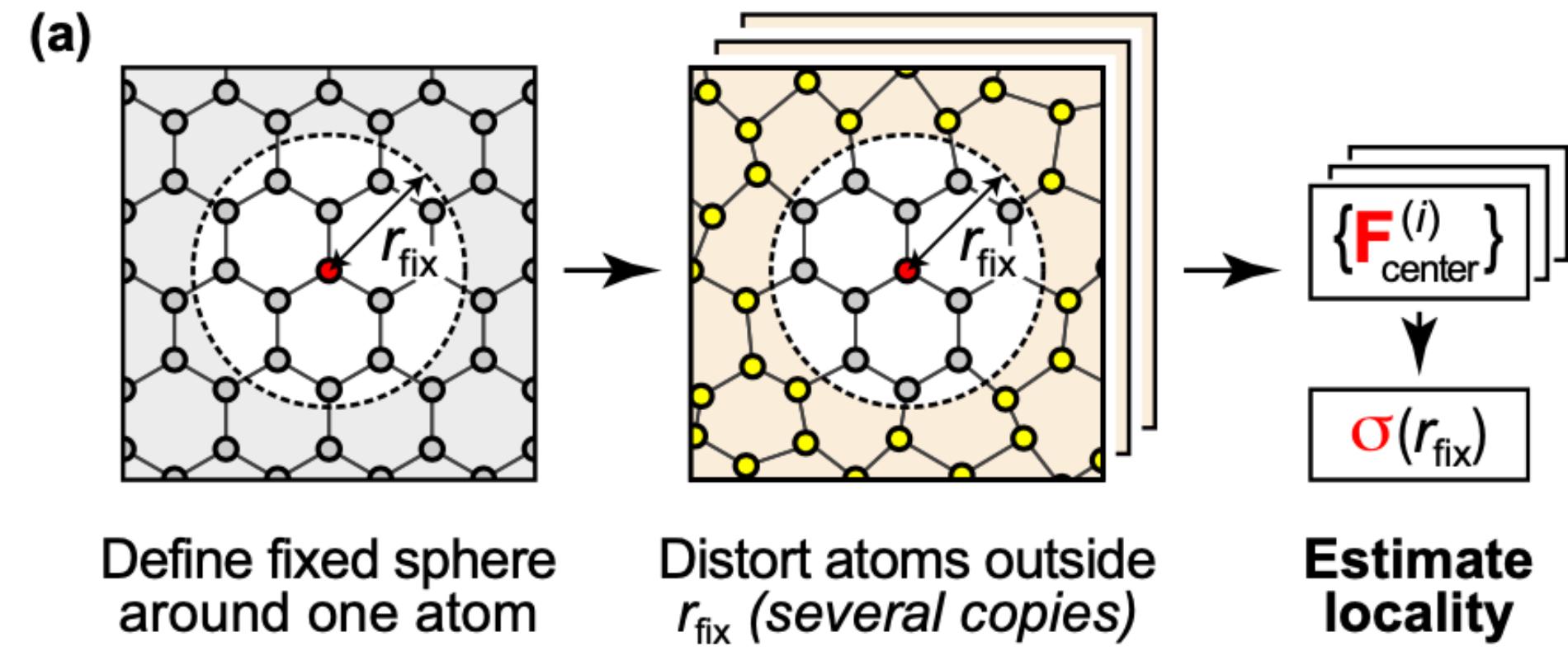
Example: amorphous carbon



- GAP development
- amorphous (a-C) and tetrahedral amorphous (ta-C)

Deringer, Csányi, PRB 95, 094203 (2017)

Example: amorphous carbon



Deringer, Csányi, PRB 95, 094203 (2017)

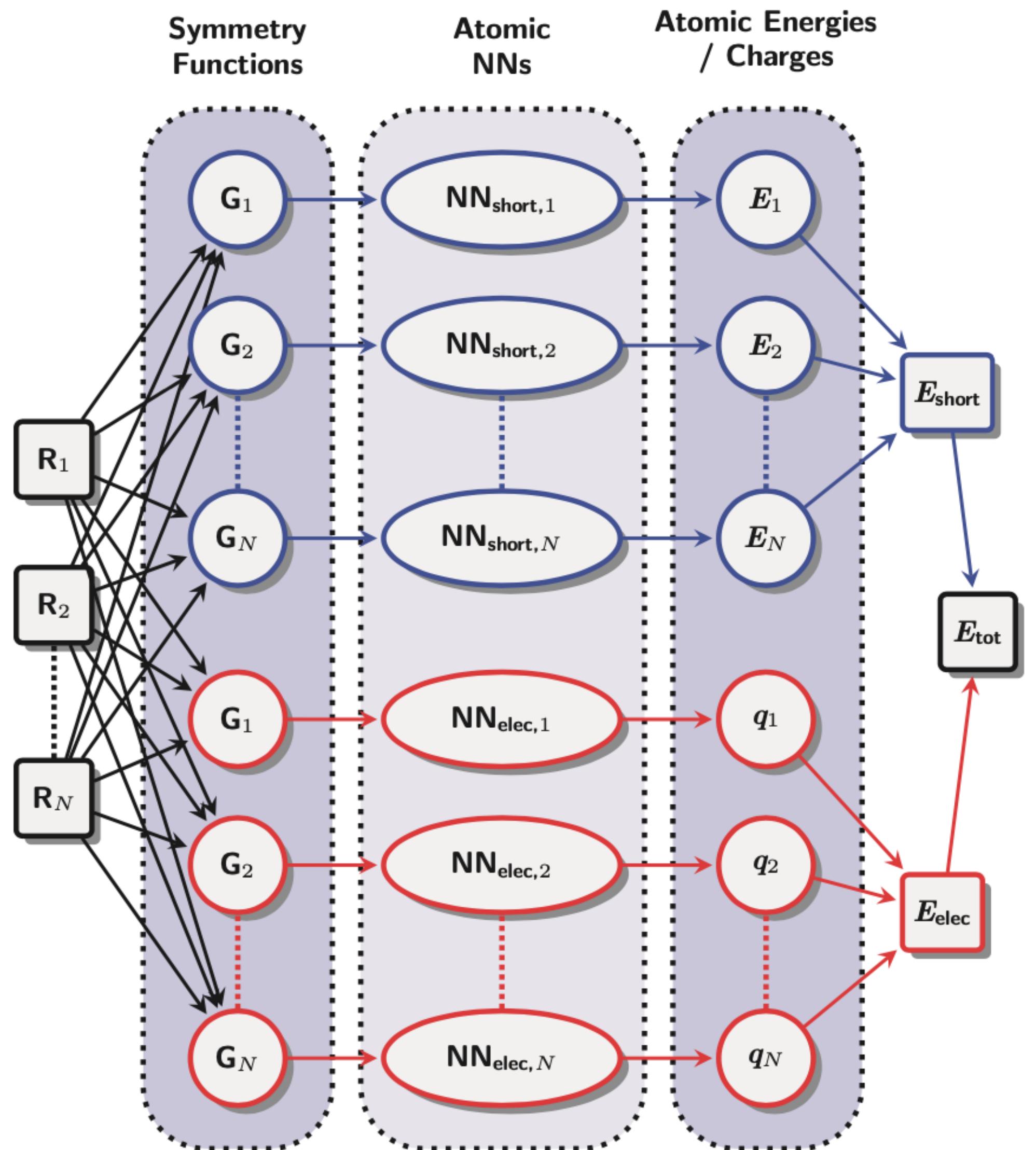
Electrostatic interactions

- decompose energy into short range and electrostatic part

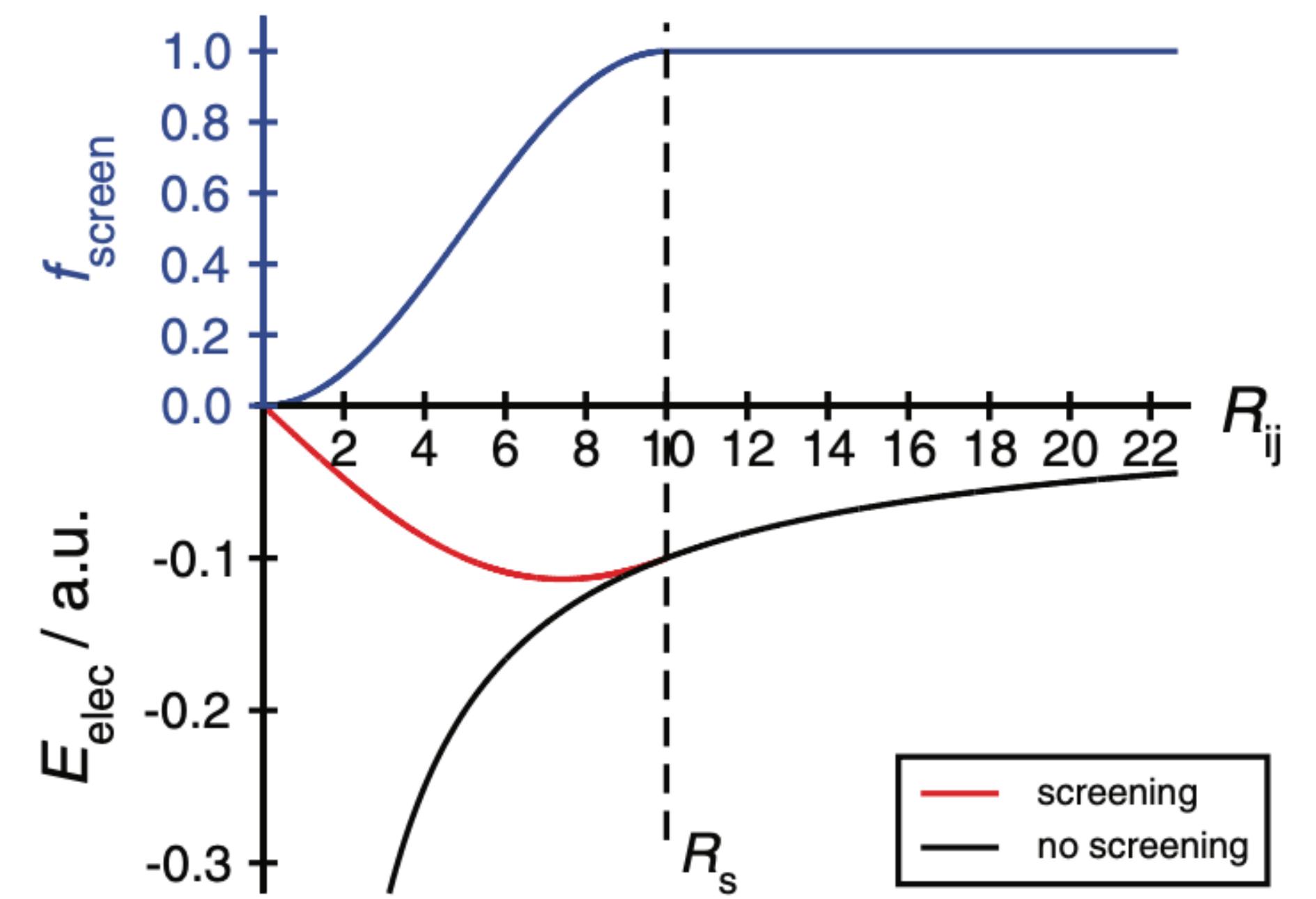
$$E_{\text{tot}} = E_{\text{short}} + E_{\text{elec}}$$

- second NN to construct *environment dependent* charges
- fit charges and short-range interactions separately
- screening function for short distances
- rescale charges to maintain overall charge neutrality

Electrostatic interactions



- 3rd generation NN potentials
- screening:



Morawietz et al., JCP 136, 064103 (2012)

Electrostatic interactions

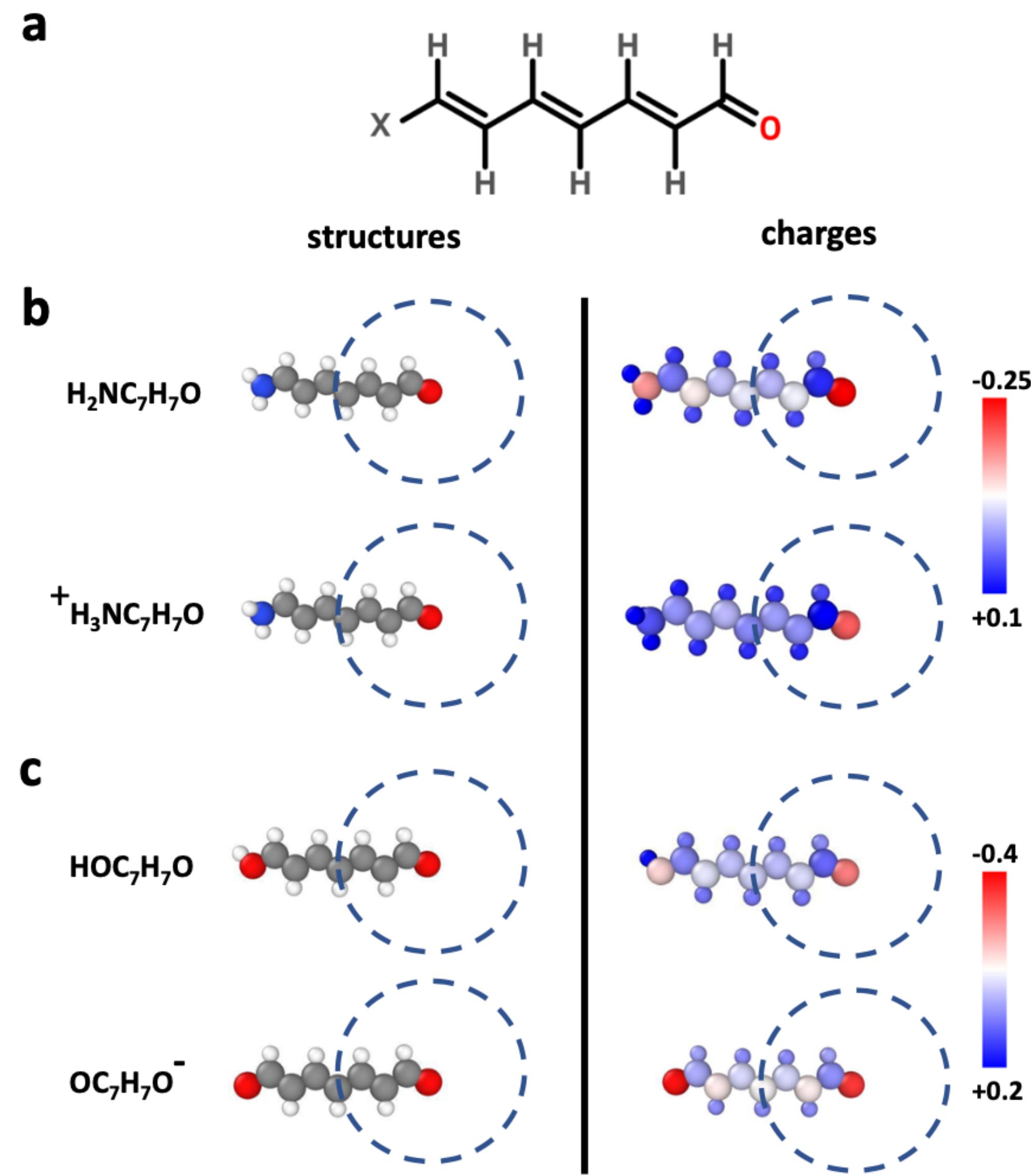
- 3rd generation NN potentials
- long-range interactions vs. non-local interactions
- restricted to single charge state (cannot describe electron removal/attachment or (de)protonation)
- → 4th generation machine learning potentials

4th generation machine learning potentials

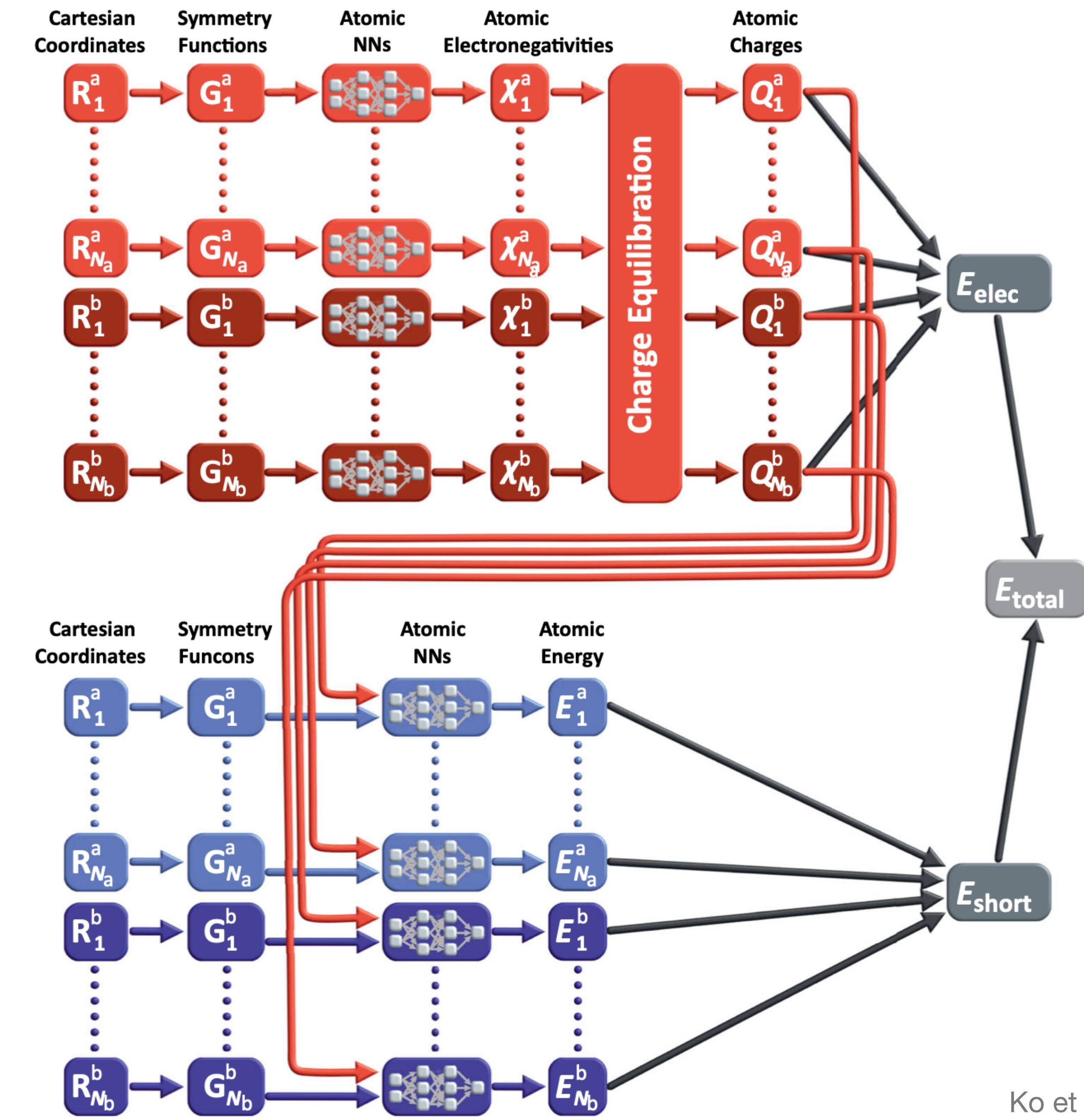


- charge equilibration neural network technique (CENT)
- 4th generation high-dimensional neural network potentials (4G-HDNNP)
- Becke population neural network (BpopNN)

Long-range charge transfer

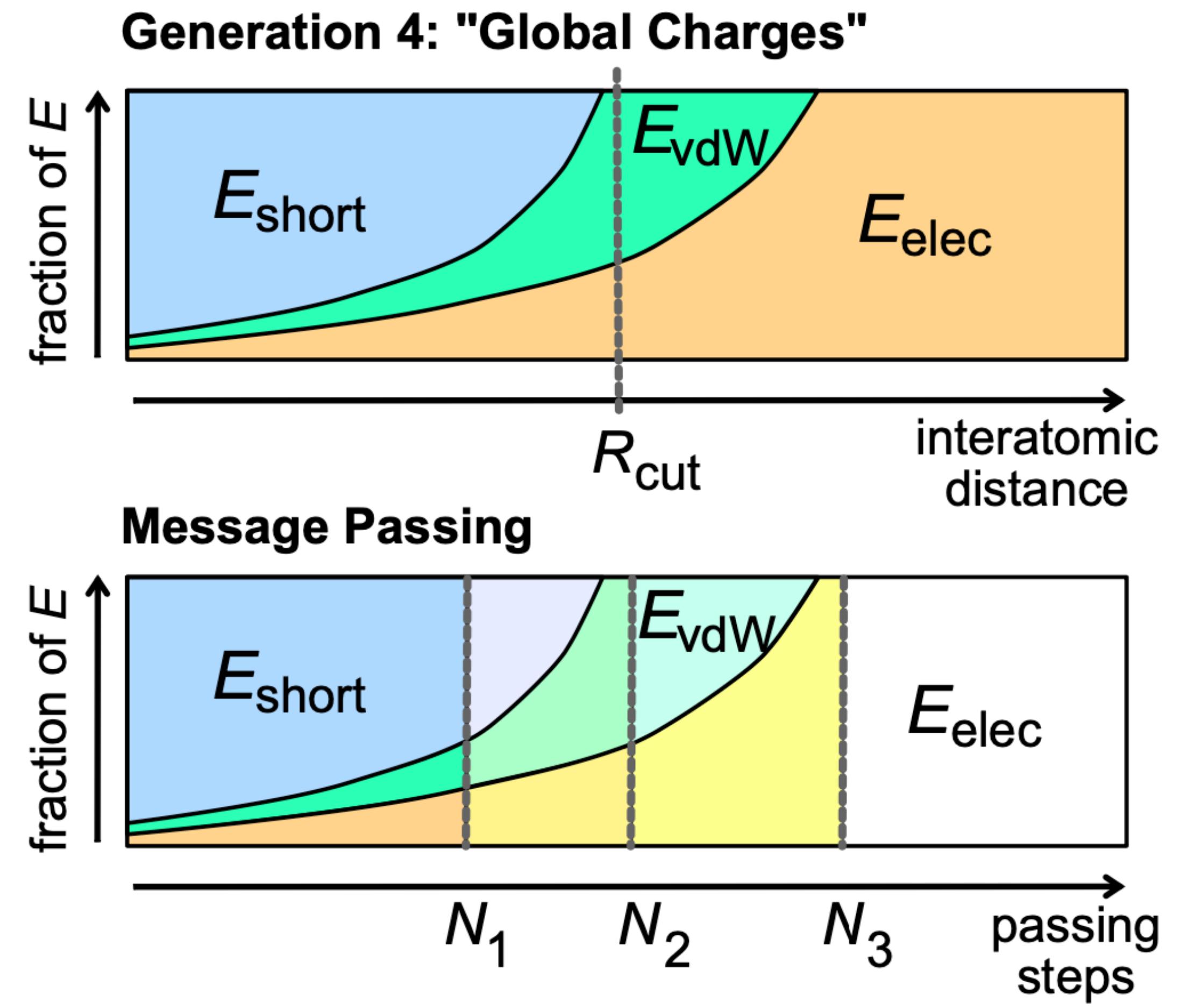
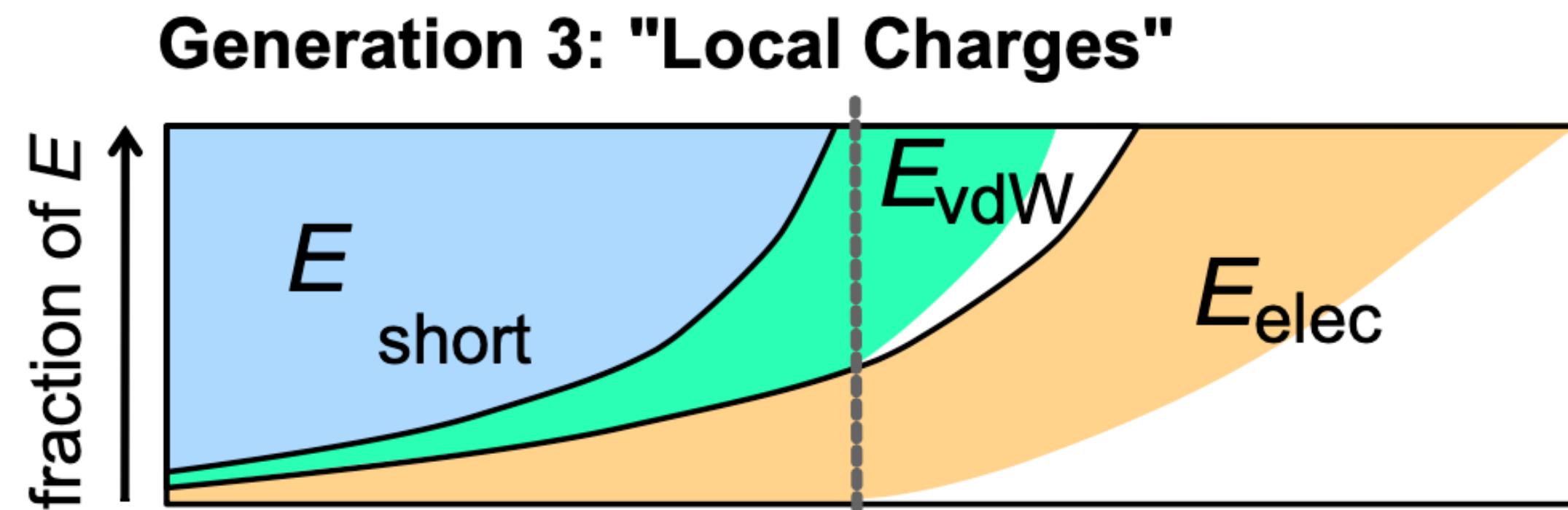
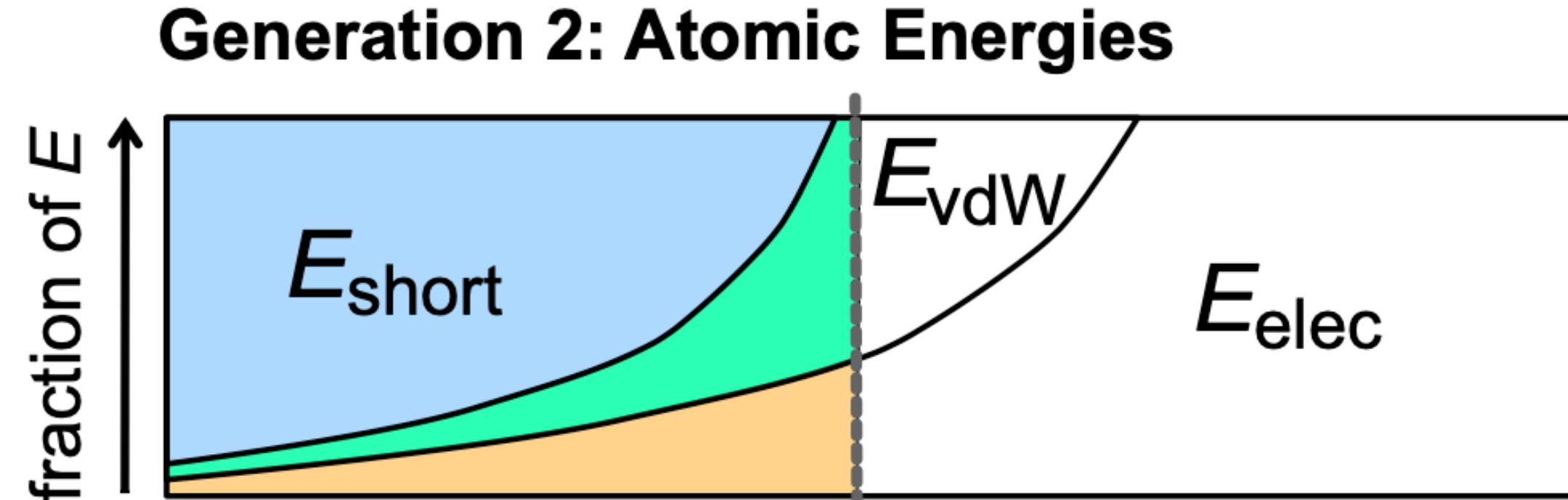


Ko et al., Nat. Commun. 12, 398 (2021)



Ko et al., Nat. Commun. 12, 398 (2021)

Relative range and coverage



Behler, Csányi, Eur. Phys. J. B 94, 142 (2021)

Nitty gritties

Nitty gritties

- transferability vs. special purpose potentials
- interpolation vs. extrapolation, range of validity
- datasets, accurate training data
- multicomponent systems (currently 3-4 elements)
- molecules, atomic bulk systems, (molecular crystals)
- active learning, Δ ML approaches

