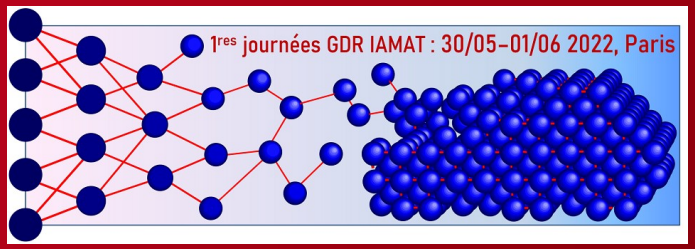
DE LA RECHERCHE À L'INDUSTRIE

cea

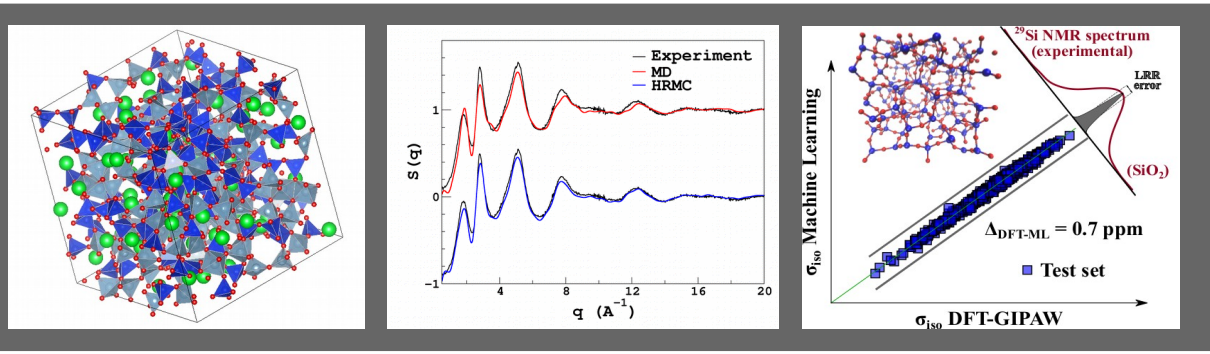
1ères journées
GDR IAMAT

30 mai - 1 juin 2022
Sorbonne Université
Paris



www.cea.fr

Modelling NMR Spectroscopy of Oxide Glasses with Machine Learning



Thibault Charpentier
Université Paris-Saclay, CEA, CNRS,
NIMBE, CEA Saclay France
thibault.charpentier@cea.fr

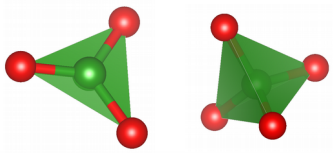
IRAMIS cnrs dépasser les frontières université PARIS-SACLAY

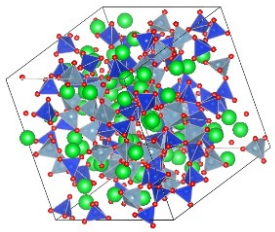
CONTEXT: MAS NMR IN OXIDE GLASSES

MAS NMR of Quadrupolar Nuclei in Glasses

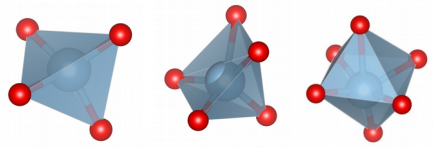
Deciphering the **Short Range Order** (SRO) in Glasses

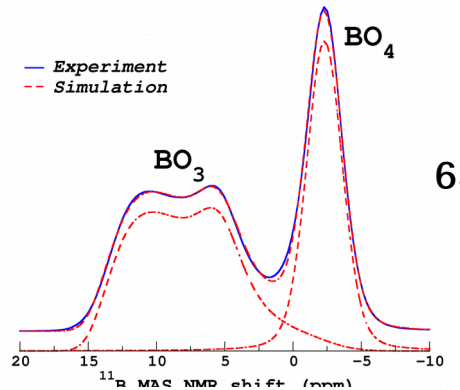
^{11}B $I=3/2$





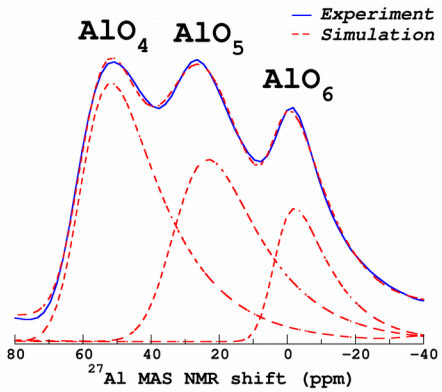
^{27}Al $I=5/2$





Aluminoborate glass

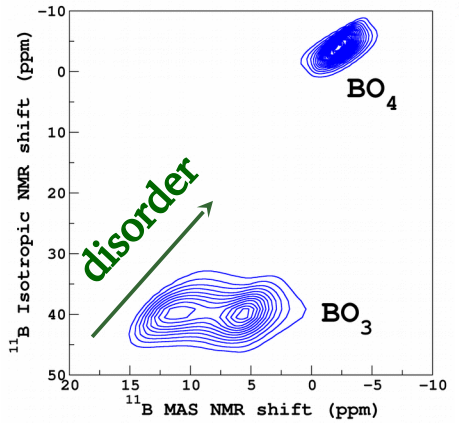
63 B_2O_3-21 La_2O_3 -16 Al_2O_3

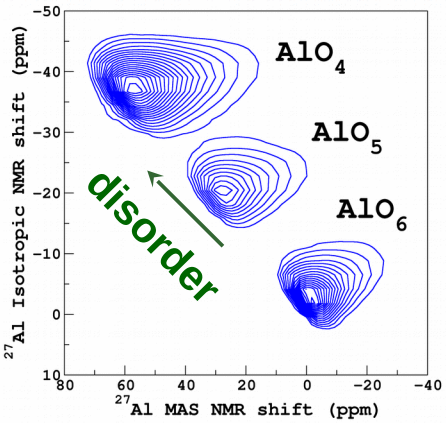


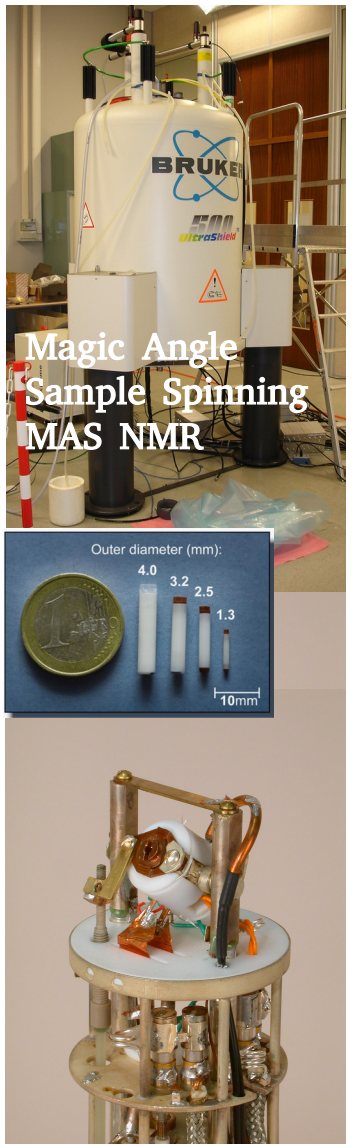
1D MAS NMR

Gain in resolution with 2D NMR

2D MQMAS NMR (Triple Quantum MAS)





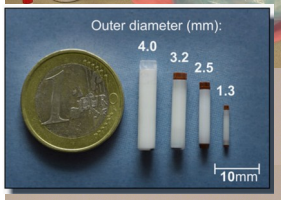


High-resolution MAS NMR reveals **the structural motifs** (network formers) building the glass network. NMR lineshapes are reflective of the local disorder and reveals the distribution of NMR parameters.

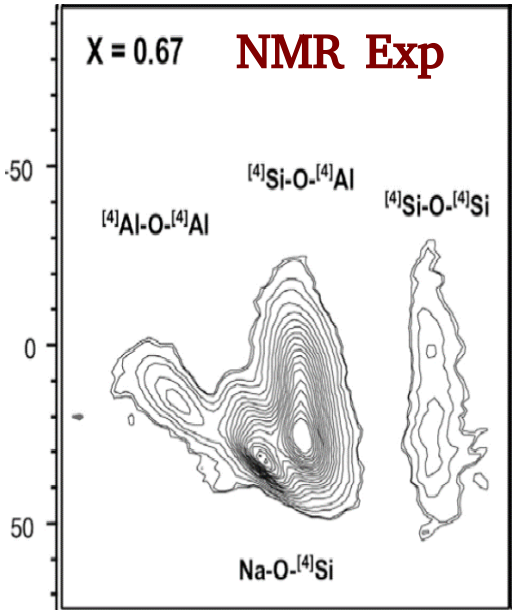
E. Chesneau, D. Caurant, T. Charpentier, in preparation.

MOTIVATIONS: NMR from first-principles

Coupling MD simulations with NMR Experiments through DFT/NMR computations (GIPAW)

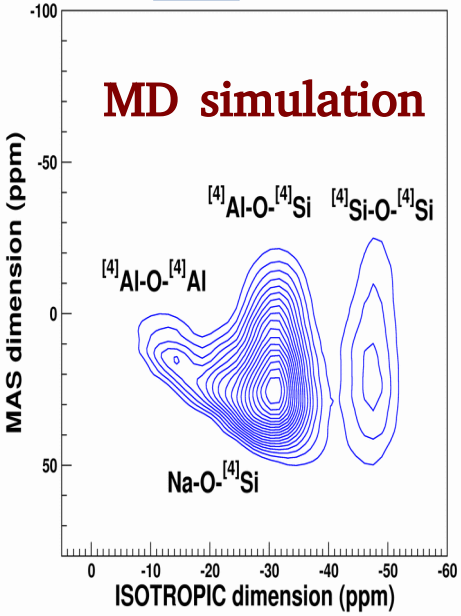


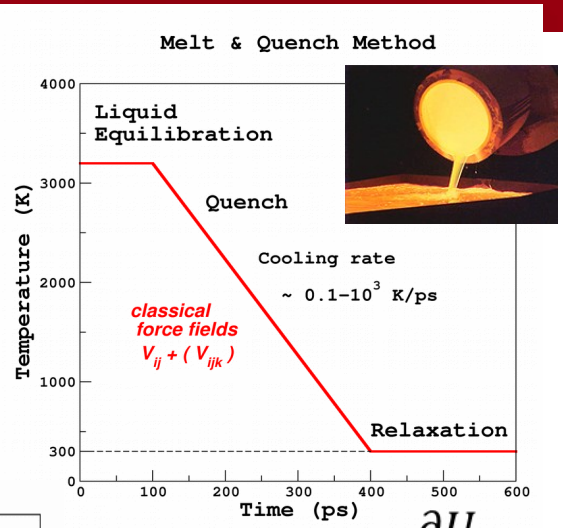
^{17}O MQMAS (11.7 T)
~1-4 days



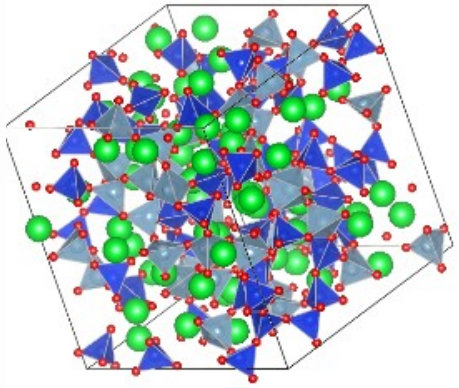
MD assessment

Experiment Interpretation





$$m_i \ddot{\boldsymbol{r}}_i = \boldsymbol{F}_i = -\frac{\partial U}{\partial \boldsymbol{r}_i}$$

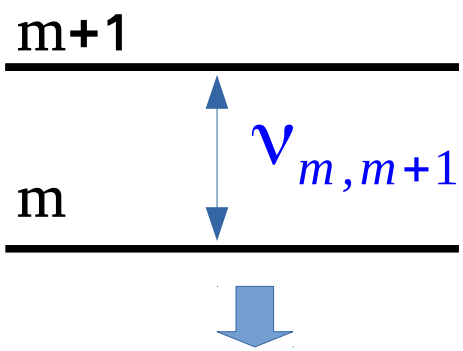


VASP-GIPAW, 800at.
~ 5h, 96 procs

T. Charpentier et al., *Computational simulations of solid state NMR spectra: a new era in structure determination of oxide glasses* ; RSC Adv. 3, 10550–10578 (2013).

NMR Parameters : The Effective Spin Hamiltonian
What do we need to compute for MAS NMR simulations ?

NMR transitions



NMR

$$H_{NMR} = -\hbar \gamma_i \vec{I} \cdot (1-\sigma) \cdot \vec{B}_0 + \vec{I} \cdot V \cdot \vec{I}$$

DFT

MAS NMR simulation : $\nu_{m,m+1} = \delta_{iso} + \nu^{(2)}_{Q,(m,m+1)}$

Magnetic Shielding Tensor

$$\sigma = \frac{\delta^2 E_0}{\delta \vec{\mu} \delta \vec{B}_0}$$

Isotropic magnetic shielding : $\sigma_{iso} = Tr\{\sigma\}$ (isotropic chemical shift)

scalar

today

Electric Field Gradient Tensor (Traceless Symmetric Matrix)

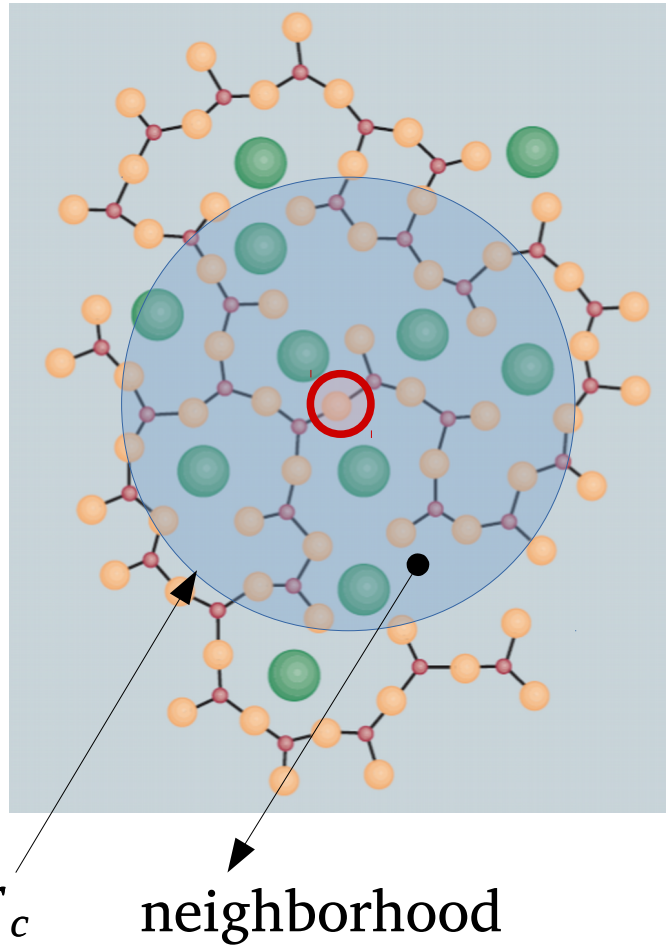
$$V_{\alpha,\beta}(r) = \int du \frac{n_0(r)}{|r-u|} \left\{ \delta_{\alpha,\beta} - 3 \frac{(r_\alpha - u_\alpha)(r_\beta - u_\beta)}{|r-u|^2} \right\}$$

matrix

Atomic Centered Descriptors
Chemical and Geometrical Properties

It assumed that the quantity of interest is local (e.g., NMR shift). The Cartesian coordinates of each atom cannot be used as such in ML. They must be transformed into a set of descriptors that fulfill some invariance/symmetry properties.

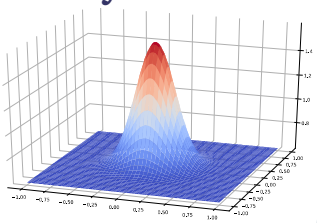
- Structural description (faithfullness)
- Translational invariance
- **Rotational invariance for a scalar quantity**
  (Energy, NMR isotropic chemical shift, …)
- Atomic permutation independent
- Zero value and derivative at cutoff radius
  (locality, short range)
- Continuous in value and derivative (smoothness)
- Of course, the design of descriptors must be driven by the properties of interest…



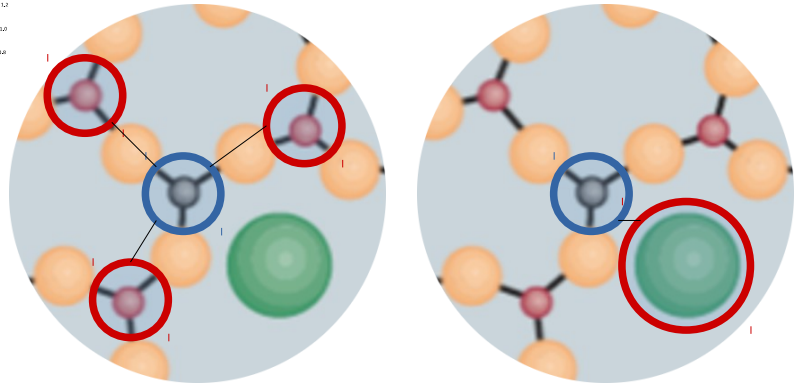
Atom-Centered Descriptors
Smooth Overlap of Atomic Positions (SOAP)

Smooth expansion of the atomic density with a radial cutoff function

$$\rho_i^{\nu}(\vec{r}) = \sum_{j \in N_i^{\nu}} G_{\sigma}(\vec{r} - \vec{r}_{ij}) f_c(r_{ij})$$

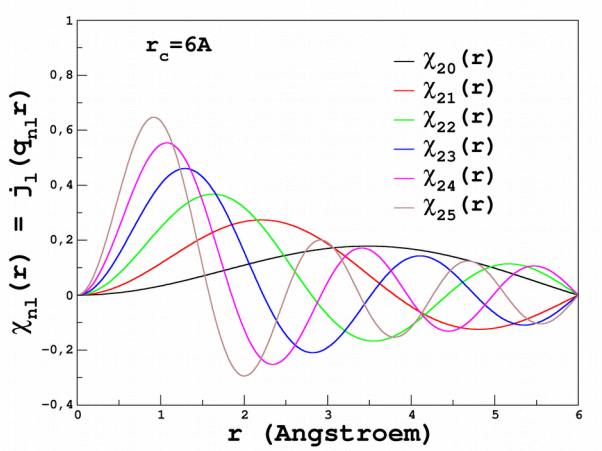


$\rho_i^{\nu}(\vec{r})$ $\rho_i^{\mu}(\vec{r})$

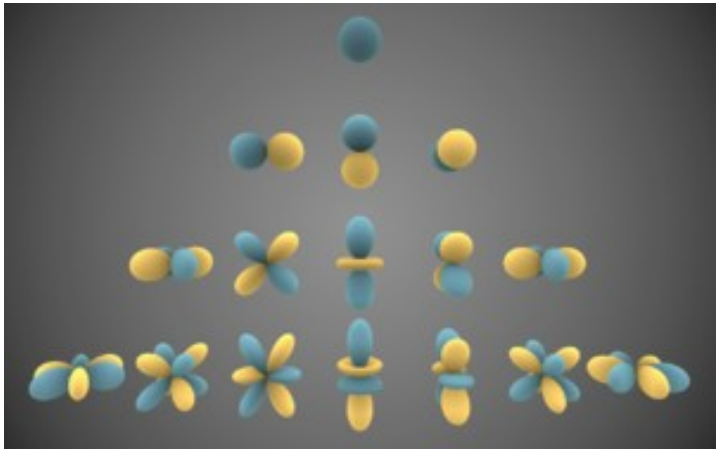


Spherical Harmonics & radial basis expansion

$$\rho_i^{\nu}(\vec{r}) = \sum_{nlm} c_{nlm}^{\nu} \chi_{nl}(r) Y_{lm}(\hat{r})$$



We found the spherical Bessel functions to be superior to other functions for the radial Basis Expansion. They form a natural orthonormal basis on the segment [0,r_c].



A.P. Bartok et al. (2013) On representing chemical environments, Phys. Rev. B

R. Jinnouchi et al. (2019) On-the-fly machine learning force field generation: Application to melting points, Phys. Rev. B

E. Kocer et al. (2020) Continuous and optimally complete description of chemical environments using Spherical Bessel descriptors, AIP Advances 10, 015021

Smooth Overlap of Atomic Positions (SOAP)

$$\rho_i^{\nu}(\vec{r}) = \sum_{nlm} c_{nlm}^{\nu} \chi_{nl}(r) Y_{lm}(\theta, \phi)$$

The problem is that the only **rotational invariant** descriptors are obtained for l=m=0, which limits severely the number of descriptors (typically l_{max}=4, n_{max}=8)

$$c_{n00}^{\nu}$$ **2-body terms**

One option is to couple channels in such a way that they form rotational invariants:

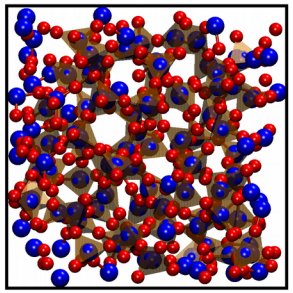
$$c_{nl+m}^{\nu} \times c_{n'l-m}^{\mu}$$

This is the so-called **SOAP** power spectrum

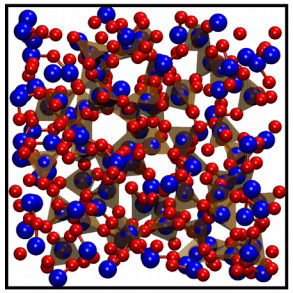
$$p_{nn'l}^{\nu\mu} = \sum_{m} c_{nl+m}^{\nu} \times c_{n'l-m}^{\mu}$$ **3-body terms**

A.P. Bartok et al. (2013), « On representing chemical environments », Physical Review B

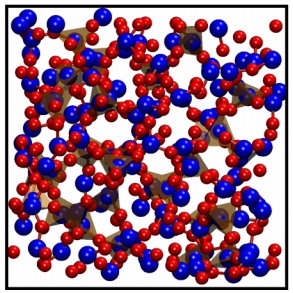
Methodology : DATABASE Generation



SiO_2 300K



SiO_2 1000K



SiO_2 2000K

10-20 MD models (400 at.) for each T

DFT-GIPAW **QM/MM**

ab initio **MD**

Scalar (Nmr shifts, charges)
Vector (Forces, dipole moments)
Tensors (Polarizability, Stress tensor)

Database for Machine Learning (ML)

OUTPUT (DFT-NMR)
10^3-10^4 NMR shift values

Machine learning engine

Invariant descriptors construction
(Atom-centered functions)

ML algorithms
(Regression, classification)

Cross-Validation
(Optimize hyperparameters)

ML Box

Kernel Ridge Regression (Gaussian Processes)
Non-Linear Regression

Kernel Ridge Regression builds a linear model with a **non-linear** similarity measurement (Kernel) :

$$\delta(\chi) = \sum_{(j \in train)} \alpha_j K_G(\chi, \chi_j)$$

$$\boldsymbol{\alpha} = \left(\boldsymbol{K}_{\boldsymbol{G}}^{\boldsymbol{train}} + \lambda \boldsymbol{I} \right)^{-1} \boldsymbol{\delta}^{\boldsymbol{train}}$$

$\chi : local\ environment\ (descriptors)$

$\chi_j, \delta_j : database\ (train)$

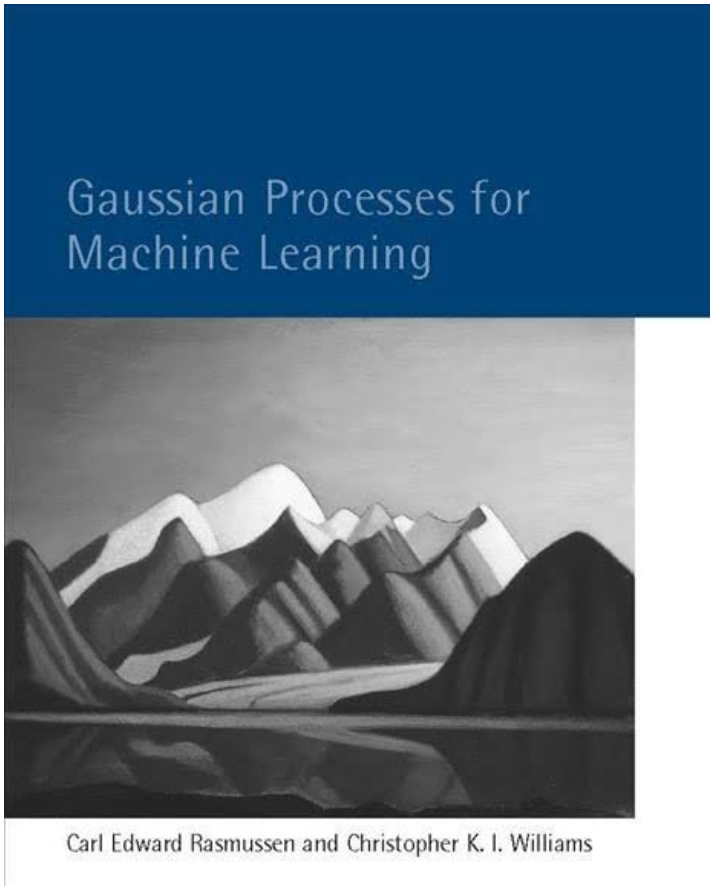
$\alpha_j : regression\ parameters$

$$Kernel : K_G(\chi, \chi_j) = e^{-\kappa \|\chi - \chi_j\|^2}$$

λ : *ridge parameter*

κ : *Kernel width*

hyperparameters

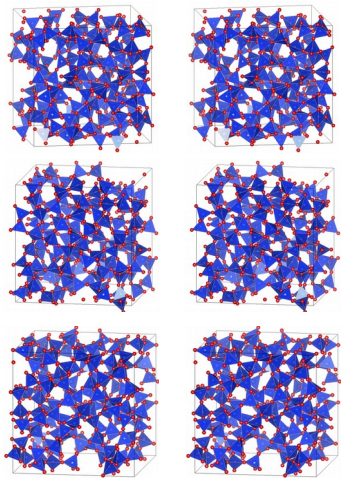


The ridge parameter prevents from overfitting and ensures a good transferability.

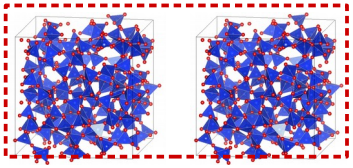
Kernel Ridge Regression : Hyperparameters optimization

DATABASE

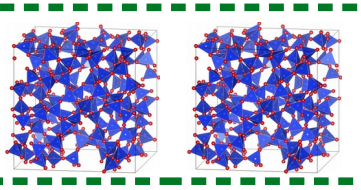
TRAINING



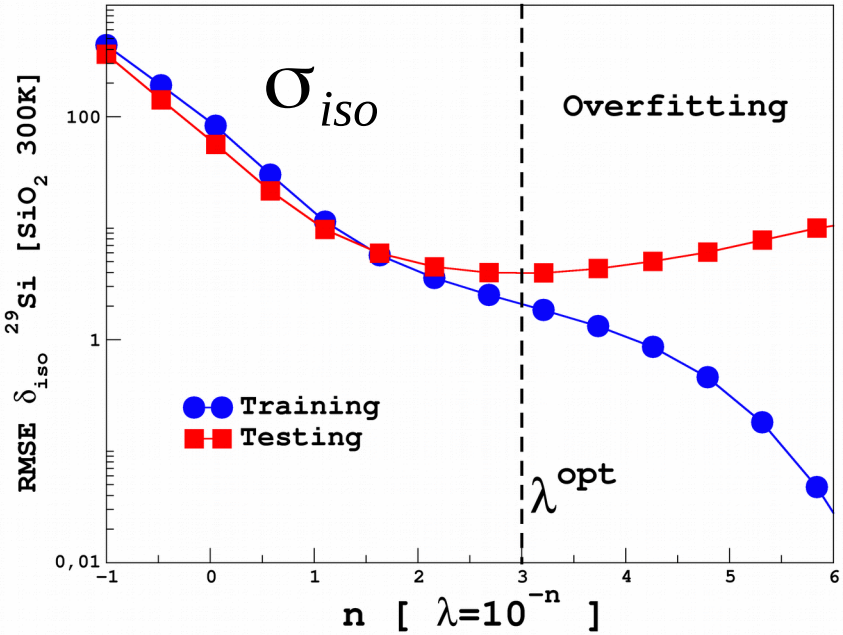
VALIDATION



TESTING

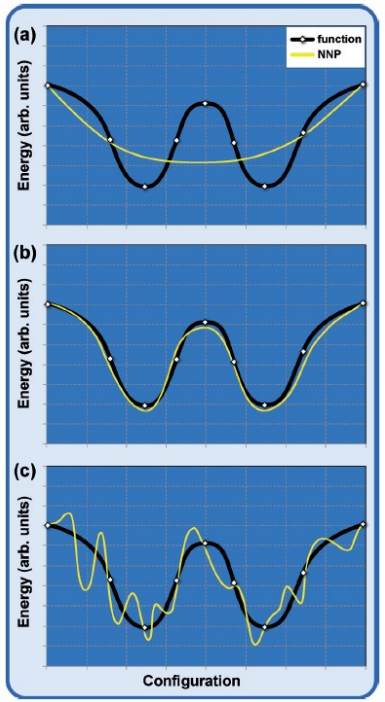


λ and κ must be optimized on a validation set before testing.

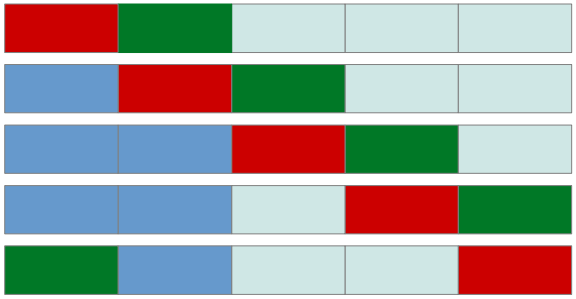


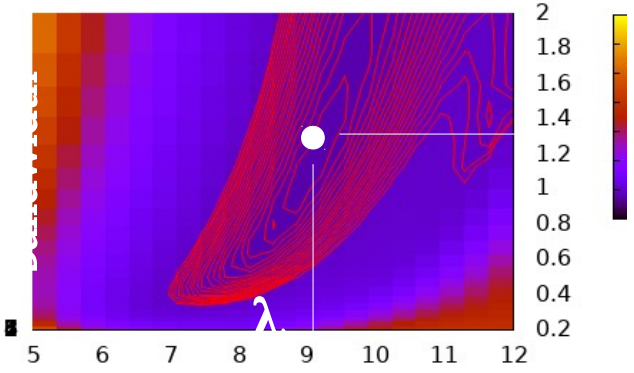
$$\alpha = \left(K_G^{train} + \lambda I \right)^{-1} \delta^{train}$$

J. Behler et al. Angewandte Chemie 2017

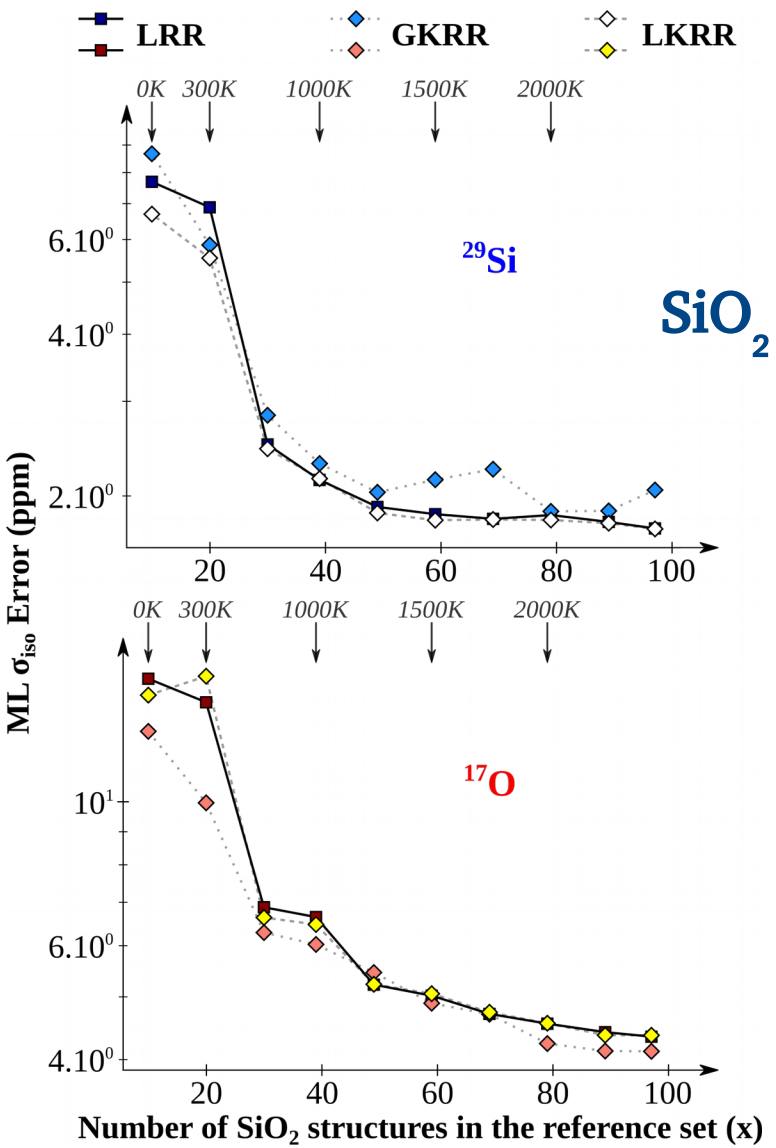


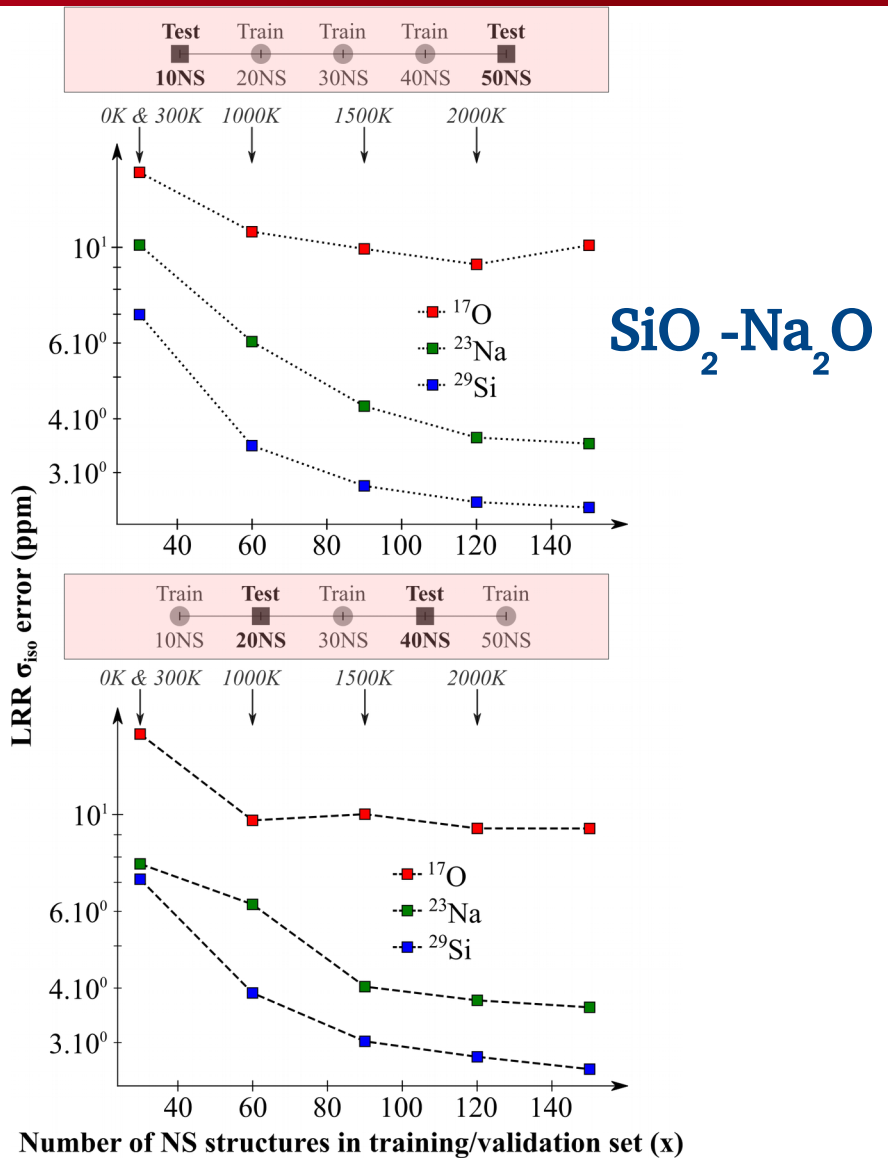
k-fold Cross Validation





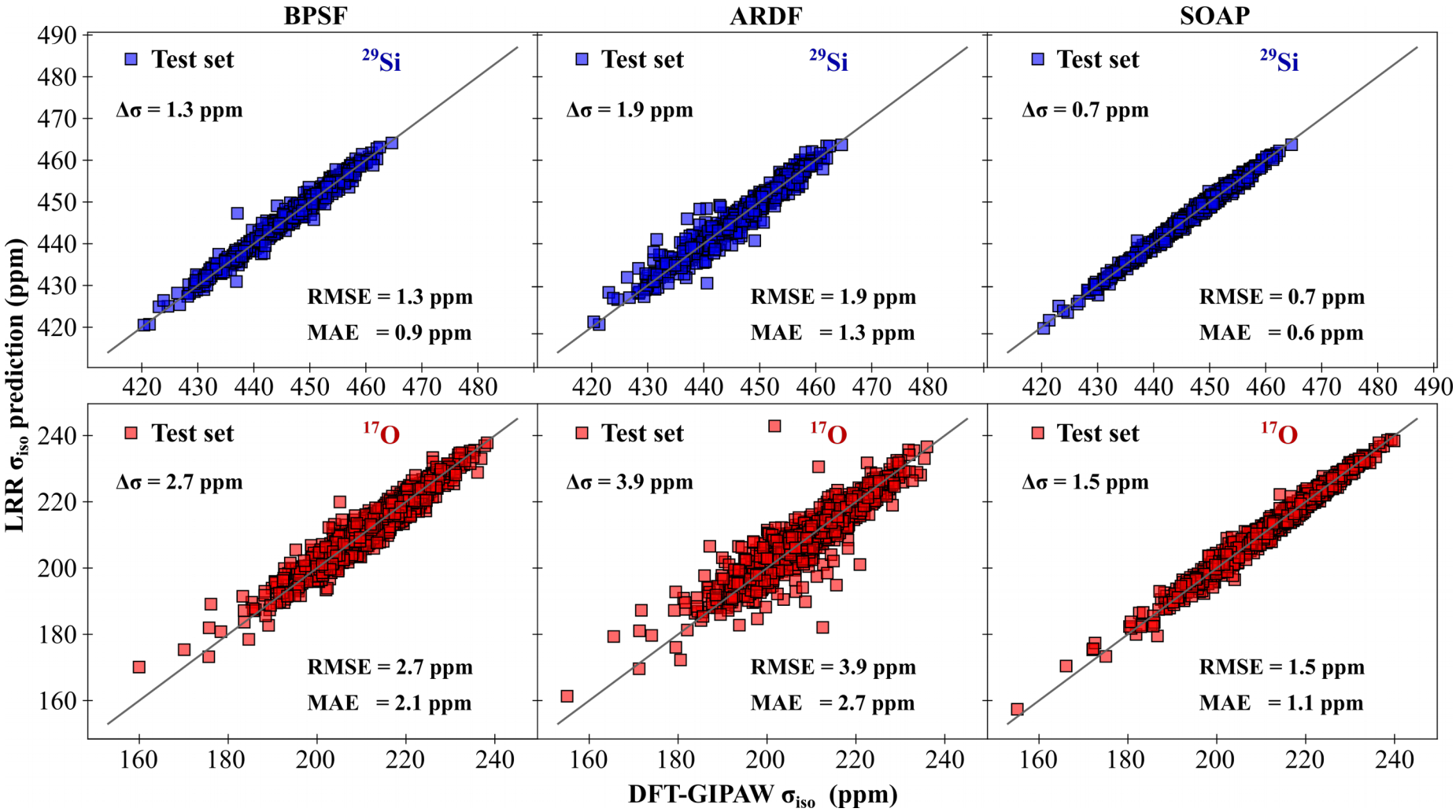
Machine Learning NMR shifts : Database Generation
MD-generated structures at various Temperatures





Z. Chaker et al. Phys. Chem. Chem. Phys. **21**, 21667 (2019)

Machine Learning NMR shifts : Descriptor's Performances

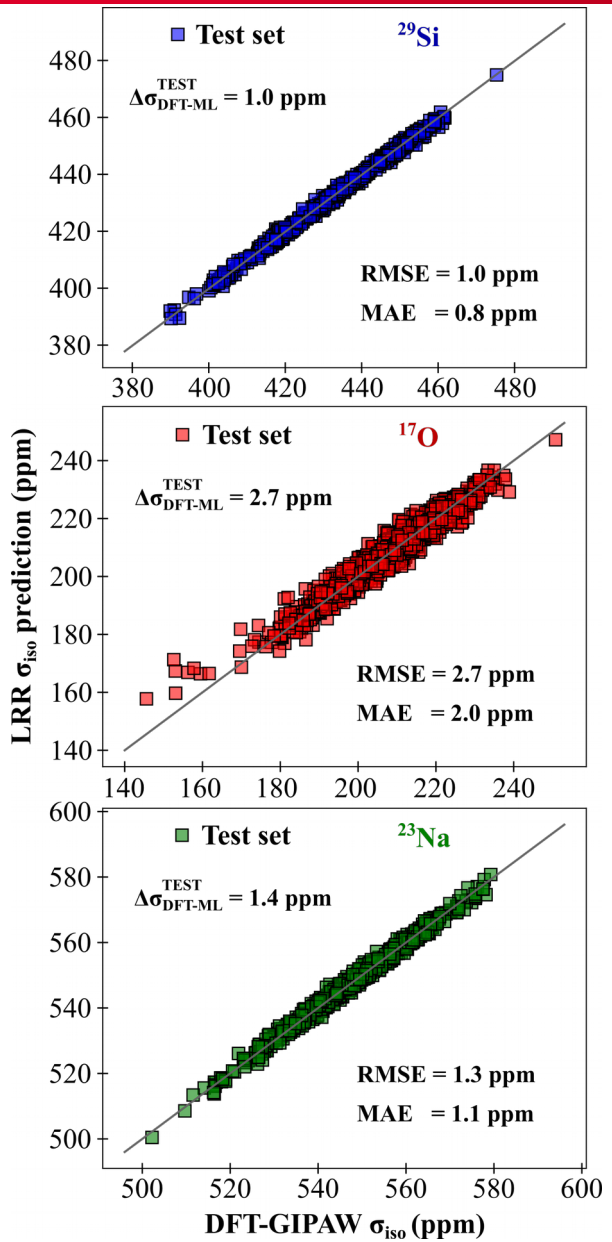


Symmetry Functions (**BPSF**) : J. Behler and M. Parrinello, Phys. Rev. Lett. 98, 146401 (2007)

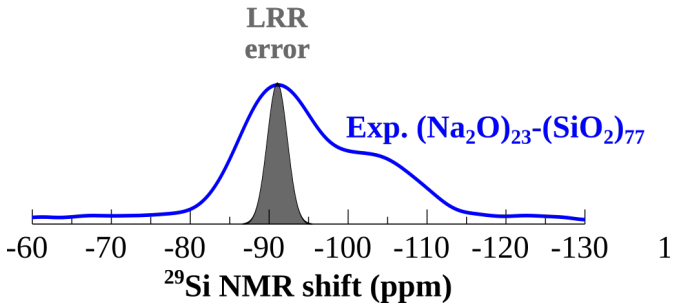
Angular-Radial Distribution Functions (**ARDF**) : N. Artrith, A. Urban, and G. Ceder, PHYSICAL REVIEW B 96, 014112 (2017)

Smooth Overlap of Atomic Positions (**SOAP**) : A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B 87, 184115 (2013)

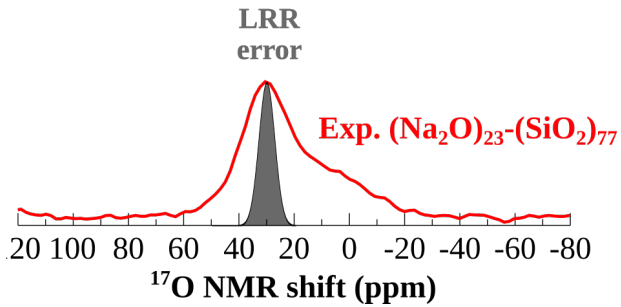
Machine Learning NMR shifts
Accuracy versus experimental data



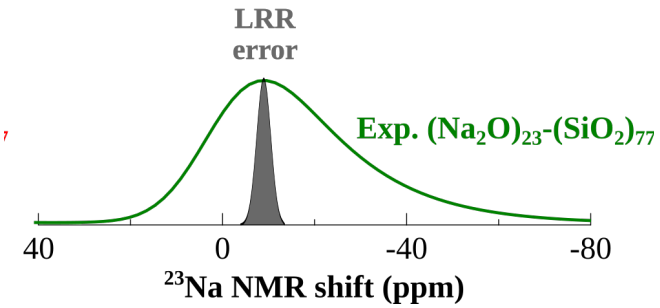
$\Delta\sigma$ = 1.0 ppm



$\Delta\sigma$ = 2.7 ppm



$\Delta\sigma$ = 1.4 ppm



SiO_2-Na_2O

Experiments : F. Angeli et al. GCA 2011

NMR shifts prediction : sparcification
Least-Square Support Vector Regression

The **Nyström** approximation of the Kernel uses a small set of **landmark/inducing points** (support vectors) :

$$K_{(\chi,\chi)} \approx K_{(\chi,\xi)} K^{-1}_{(\xi,\xi)} K_{(\xi,\chi)}$$

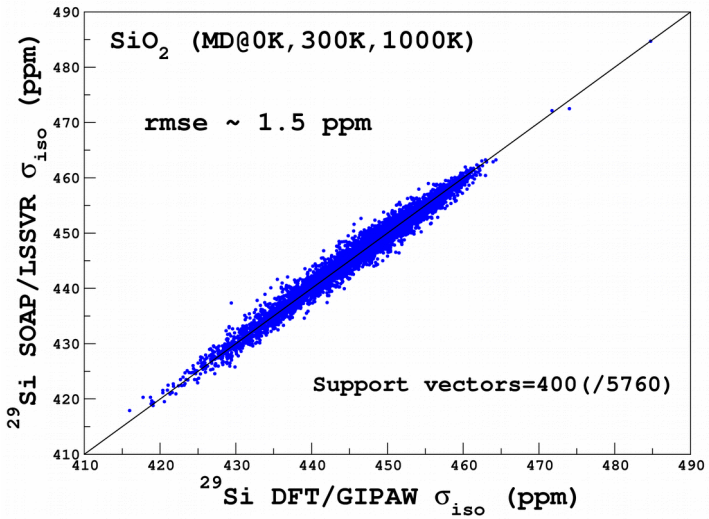
In the **RKHS**, the Kernel can be defined as a scalar product of **non-linear feature map** (Mercer Th.) :

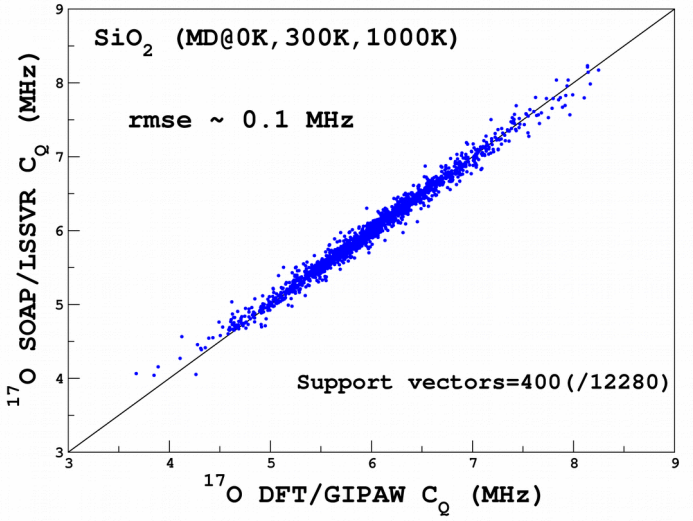
$$K(\chi_i,\chi_j) = \langle \phi(\chi_i)\phi(\chi_j) \rangle$$

Then, according to the Reproducing Kernel Hilbert Space (**RKHS**) formalism, the KRR prediction can be expressed as linear combination of the **non-linear feature maps** approximated on the support vectors :

$$\delta(\chi) = \sum_{(\xi)} w_\xi \phi_\xi(\chi)$$

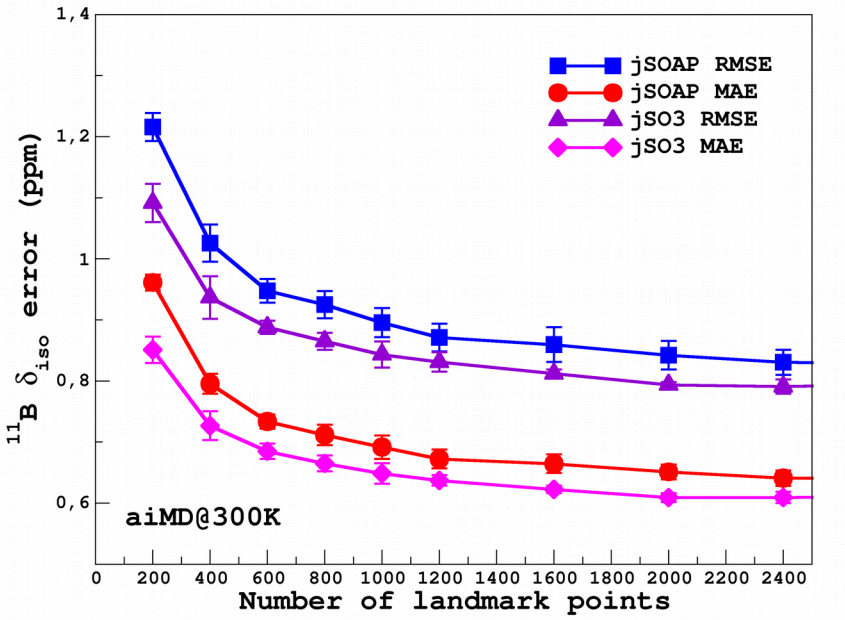
The dimensionality of the problem is reduced (sparcification). Several options exist for the choice of the support vectors (Kmeans, Max. Entropy). We found the **Incomplete Cholesky Decomposition** to be the most efficient.

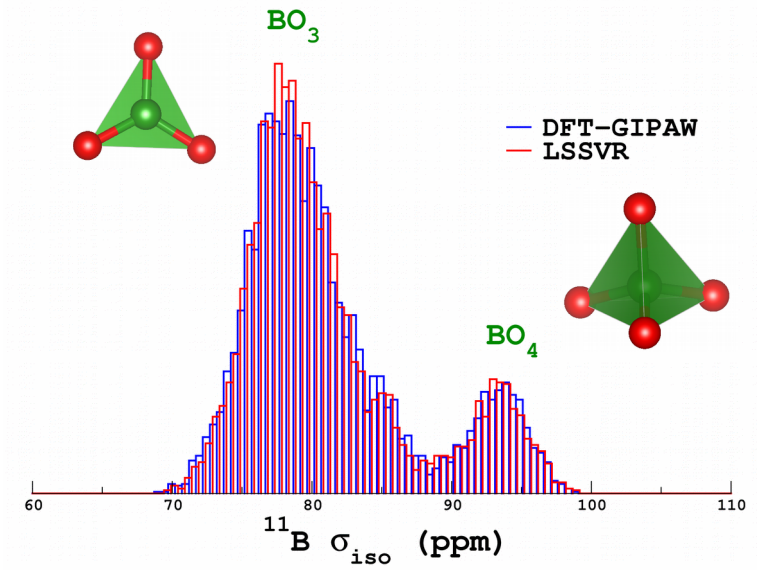




De Brabanter K, et al., (2010) Optimized fixed-size kernel models for large data sets, Computational Statistics & Data Analysis
Oglic D. et al. (2017) Nyström method with Kernel K-means++ samples as landmarks, Journal of Machine Learning Research
T. Charpentier et al., NMR interactions in sodium borate glasses by machine learning, In preparation

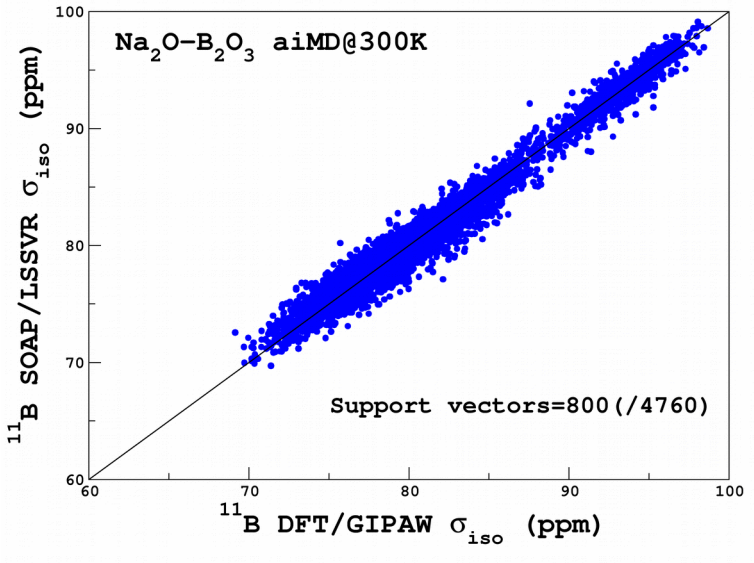
NMR Shifts in sodium borate glasses
Least Square Support Vector Regression (LSSVR)





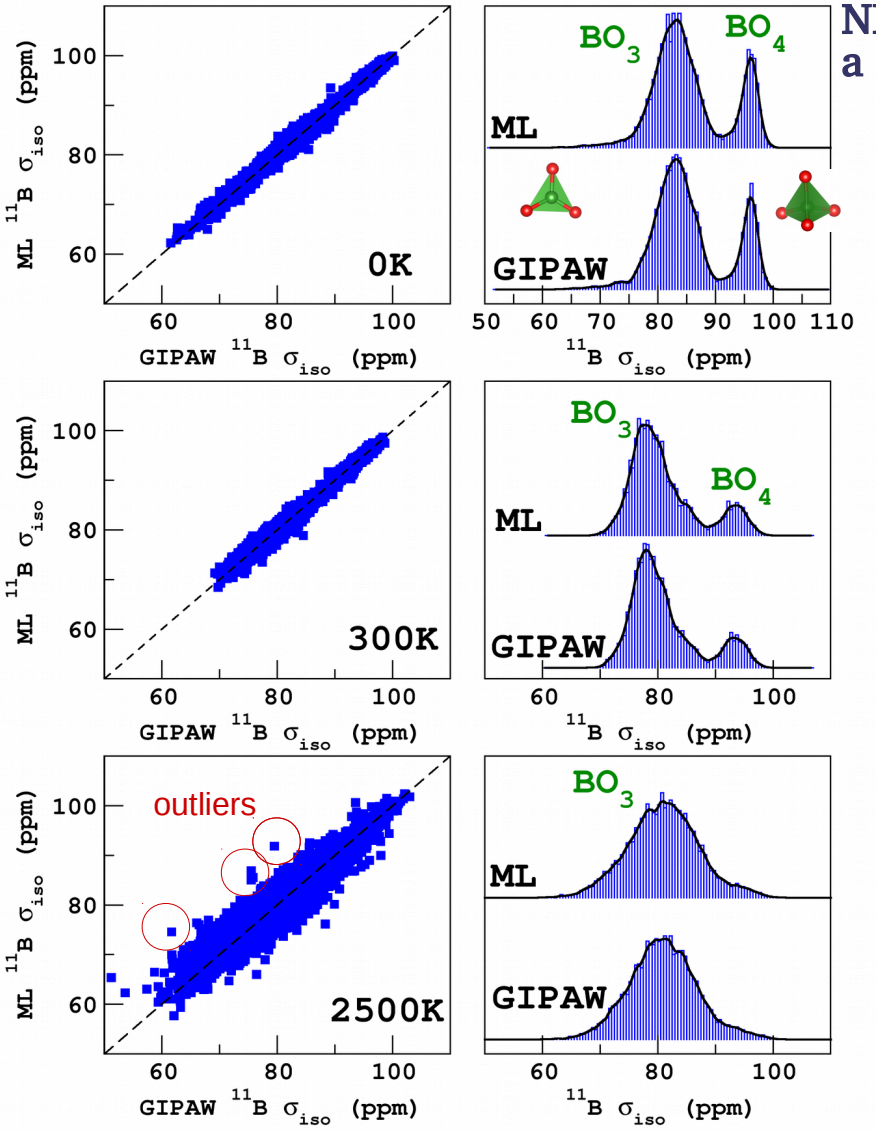
(Top) Comparison of convergence of the ^{11}B NMR shifts error in sodium borate glasses (aiMD at 300K) for two descriptors.

Right panel shows the comparison between DFT calculated and ML predicted distribution of isotropic ^{11}B NMR shifts.

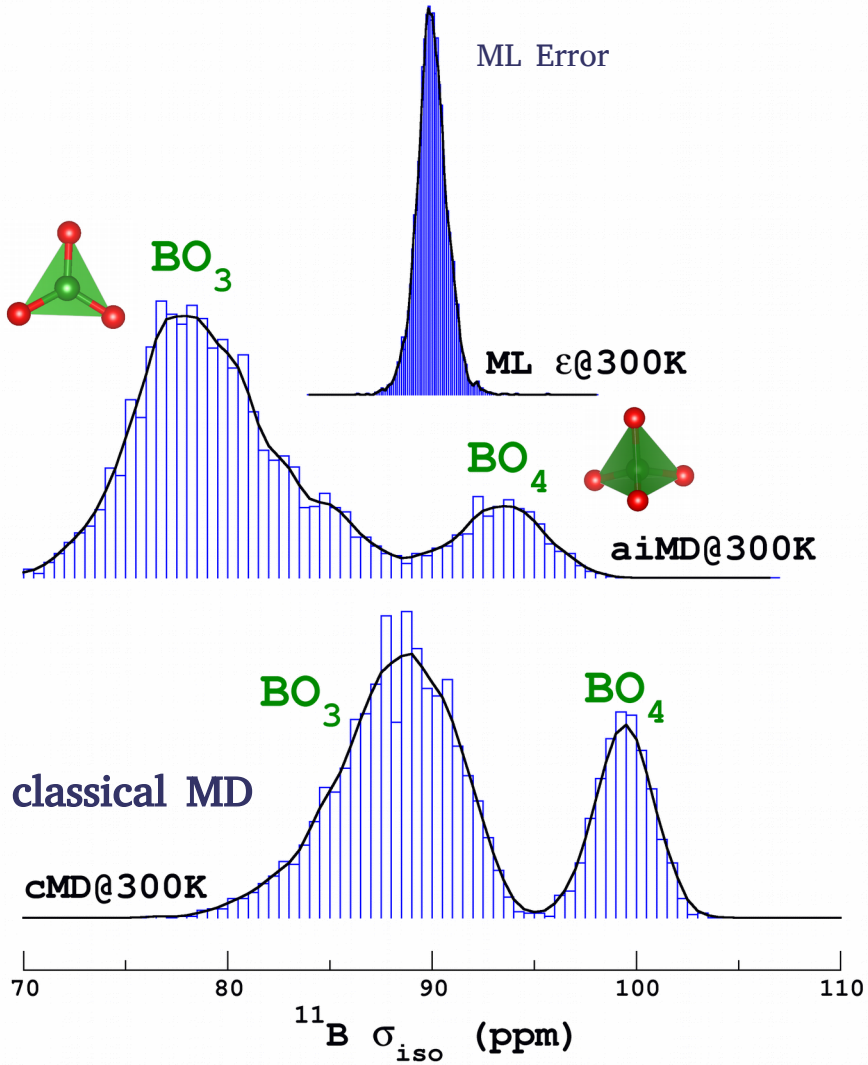


T. Charpentier et al., NMR interactions in sodium borate glasses by machine learning, In preparation

NMR Shifts in sodium borate glasses
Support Vector Regression (LSSVR)



NMR shift : a driving force for data driven modeling ?



T. Charpentier, E. Chesneau, Z. Chaker, NMR interactions in sodium borate glasses by machine learning, in preparation

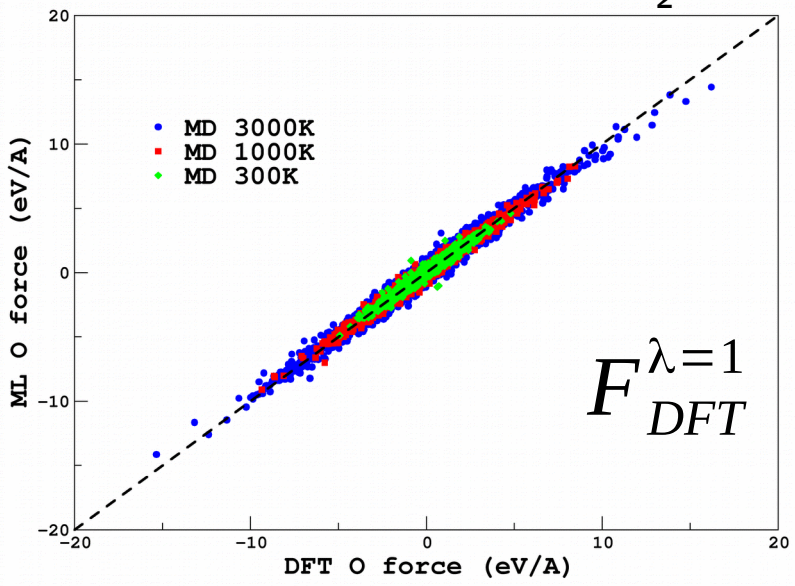
Perspectives: vector-valued Kernels Machine Learning NMR Tensors (λ – SOAP)

Each descriptor is a **Tensor** . With SOAP :

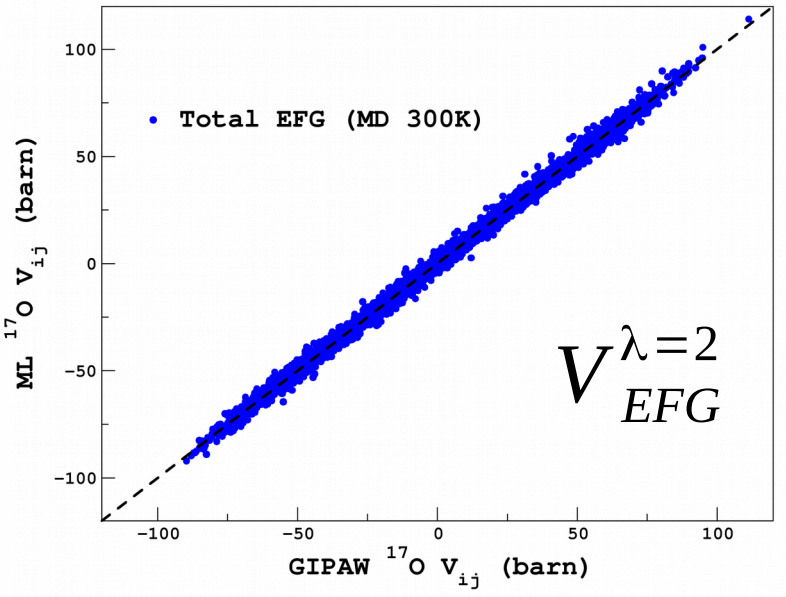
$$T^{\lambda}_{m}(\chi)=\sum_{n,\nu}\alpha_{n,\nu}\,c^{\nu}_{n\lambda m}(\chi)$$

2-body terms

Force prediction (SiO_2)



Matrix prediction (SiO_2)



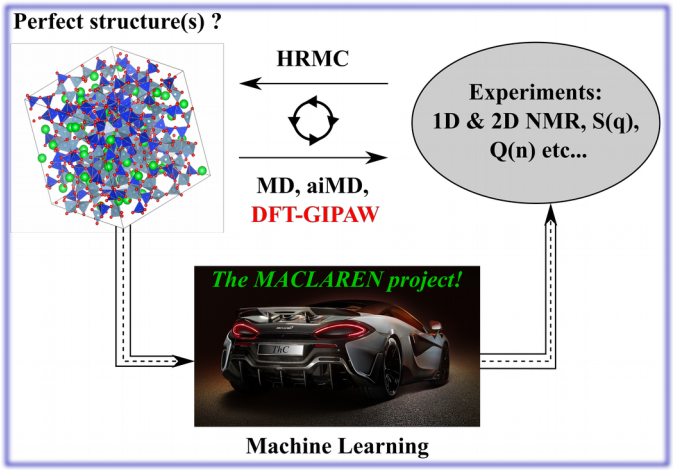
$$P^{\lambda,m}_{n_1 n_2 l_1 l_2}(\chi)=\sum_{m_1 m_2} c_{n_1 l_1 m_1}(\chi)\,c_{n_2 l_2 m_2}(\chi)\,C^{\lambda,m}_{l_1,m_1,l_2,m_2}$$

3-body terms

→ E(3) Equivariant Graph Neural Network (NequIP) arXiv:2101.03164

RAMAN example : Using Gaussian process regression to simulate the vibrational Raman spectra of molecular crystals, N. Raimbault, A. Grifasi, M. Ceriotti, M. Rossi September 2019 New Journal of Physics 21(10)

- **Z. CHAKER** (Machine Learning NMR)
- CEA Paris-Saclay, France, Post-Doc 2018-2020



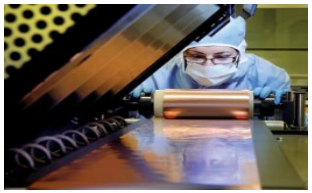
Z. Chaker et al. Phys. Chem. Chem. Phys. **21**, 21667 (2019)

We acknowledge the financial support of the Cross-Disciplinary Program on Numerical Simulation of CEA, the French Alternative Energies and Atomic Energy Commission

DE LA RECHERCHE À L'INDUSTRIE

cea

Thank you for your attention


FROM RESEARCH TO INDUSTRY



French Alternative Energies and Atomic Energy Commission

Alternative Energies and Atomic Energy Commission
Centre de Saclay | 91191 Gif-sur-Yvette Cedex
T. +33 (0)1 64 50 10 00| F. +33 (0)1 64 50 11 86
Etablissement public à caractère industriel et commercial | RCS Paris B 775 685 019

DRF DEN
IRAMIS DE2D
NIMBE SEVT