

Transferable Machine-Learning Model of the Electron Density

Supporting Information

Andrea Grisafi,[†] Alberto Fabrizio,[‡] Benjamin Meyer,[‡] David M. Wilkins,[†]
Clemence Corminboeuf,[‡] and Michele Ceriotti^{*,†}

[†]*Laboratory of Computational Science and Modeling, IMX, École Polytechnique Fédérale de
Lausanne, 1015 Lausanne, Switzerland*

[‡]*Laboratory for Computational Molecular Design, Institute of Chemical Sciences and
Engineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

E-mail: michele.ceriotti@epfl.ch

Symmetry-adapted Gaussian Process Regression

Symmetry-adapted Gaussian process regression (SA-GPR) is a machine-learning framework which makes it possible to include in the regression algorithm the geometric covariances in three dimensions of spherical tensors of arbitrary order λ . This is done by designing a tensorial kernel function $\mathbf{k}^\lambda(\mathcal{X}, \mathcal{X}')$ which transforms like a Wigner-D matrix of the proper order λ , as follows:

$$\mathbf{k}^\lambda(\mathcal{X}, \mathcal{X}') = \int d\hat{R} \mathbf{D}^\lambda(\hat{R}) \left| \langle \mathcal{X} | \hat{R} \mathcal{X}' \rangle \right|^2 \quad (\text{S1})$$

On the top of this environmental similarity measure, a normalization can follow by considering

$$\tilde{\mathbf{k}}^\lambda(\mathcal{X}, \mathcal{X}') = \frac{\mathbf{k}^\lambda(\mathcal{X}, \mathcal{X}')}{\sqrt{\|\mathbf{k}^\lambda(\mathcal{X}, \mathcal{X}')\|_F \|\mathbf{k}^\lambda(\mathcal{X}, \mathcal{X}')\|_F}} \quad (\text{S2})$$

Finally, the normalized kernel can be elevated to a positive integer power ζ to enhance its non-linear character and, as a consequence, its ability to describe many-body structural correlations of higher order. Given that the proper geometric covariances of the kernel function are strictly related to its linear relationship with the Wigner-D matrix, such exponentiation actually consists in multiplying the primitive tensorial kernel by its scalar, and therefore rotationally invariant, counterpart:

$$k_{\mu\mu'}^{\lambda,\zeta}(\mathcal{X}, \mathcal{X}') = \tilde{k}_{\mu\mu'}^\lambda(\mathcal{X}, \mathcal{X}') \tilde{k}_{00}^0(\mathcal{X}, \mathcal{X}')^{\zeta-1} \quad (\text{S3})$$

As a practical implementation of Eq. (S2), we followed the choice made in Ref. S1 where such kernel functions have been implemented by choosing as the primitive similarity measure $\langle \mathcal{X} | \mathcal{X}' \rangle$ the smooth overlap of atomic positions (SOAP) of Ref. S2, consisting of a real space overlap of Gaussian smoothed atomic densities between atomistic environments. From this hierarchy of λ -SOAP kernels, the prediction of a spherical tensor component \mathbf{T}^λ associated

with a given atomic environment \mathcal{X} is given by

$$T_{\mu}^{\lambda}(\mathcal{X}) = \sum_{j \in M} \sum_{|\mu'| < \lambda} k_{\mu\mu'}^{\lambda}(\mathcal{X}, \mathcal{X}_j) x_{\mu'}^{\lambda}(\mathcal{X}_j) \tag{S4}$$

with $x_{\nu}^{\lambda}(\mathcal{X}_j)$ the regression weights associated with a set M of training environments $\{\mathcal{X}_j\}$, being in turn spherical tensors of dimension $2\lambda + 1$.

Derivation of the regression formula

We report here a detailed mathematical derivation of the regression formula for the charge density.

1. We start by making the ansatz of expanding the charge density field in an atom-centred basis set:

$$\rho(\mathbf{r}) = \sum_{i \in \mathcal{A}} \rho_i(\mathbf{r}) = \sum_{i \in \mathcal{A}} \sum_{nlm} c_{nlm}^i \phi_{nlm}(\mathbf{r} - \mathbf{r}_i) \delta_{\alpha_i \alpha_n} \tag{S5}$$

with i atomic labels for molecule \mathcal{A} . Given $\mathbf{S} = \mathbf{r} - \mathbf{r}_i$ the atom centred position vector, basis functions are factorized into spherical harmonics $Y_m^l(\hat{\mathbf{S}})$ and radial functions $R_n(S)$. Letting n run over any arbitrary set of radial functions, the Kronecker delta $\delta_{\alpha_i \alpha_n}$ has the role of associating different radial functions with different atomic species α .

2. We continue by making the assumption that the expansion coefficients arise from a symmetry-adapted Gaussian process regression framework, where different atomic species, radial channels and spherical harmonic components are predicted independently of each other. In practice, once the full series of spherical-tensor kernels $\{\mathbf{k}^l\}$ is computed up to the maximum angular momentum value l_{\max} included in the density expansion, the local prediction of a generic density component associated with a given local environment \mathcal{X}_i is

carried out by the following linear combination:

$$c_{nlm}^i(\mathbf{x}) = \sum_{j \in M} \sum_{|m'| < l} k_{mm'}^l(\mathcal{X}_i, \mathcal{X}_j) x_{nlm'}^j \delta_{\alpha_i \alpha_j}, \quad (\text{S6})$$

where j runs over a given reference set M of representative atom-centred environments \mathcal{X}_j , while the sum over m' provides the correct geometric covariance of the predicted density component. \mathbf{x} refers to the vector of regression weights $x_{nlm'}^j$ to be determined.

3. The regression weights $x_{nlm'}^j$ come from the minimization of a loss function describing the collective error in representing the density of N training molecules. Formally we can write:

$$\ell(\mathbf{x}) = \sum_{\mathcal{A} \in N} \int d\mathbf{r} \left| \rho_{\mathcal{A}}(\mathbf{r}) - \sum_{i \in \mathcal{A}} \sum_{nlm} c_{nlm}^i(\mathbf{x}) \phi_{nlm}(\mathbf{r} - \mathbf{r}_i) \delta_{\alpha_i \alpha_n} \right|^2 + \eta |\mathbf{x}|^2 \quad (\text{S7})$$

with $c_{nlm}^i(\mathbf{x})$ depending parametrically on the set of regression weights \mathbf{x} to be determined while η accounting for the intrinsic noise of the training densities. Differentiating the loss function of Eq. (S7) with respect to the regression weights $x_{\nu\lambda\mu}^\gamma$, we find:

$$\begin{aligned} \frac{\partial \ell(\mathbf{x})}{\partial x_{\nu\lambda\mu}^\gamma} &= 2\eta x_{\nu\lambda\mu}^\gamma - 2 \sum_{\mathcal{A} \in N} \int d\mathbf{r} \left[\rho_{\mathcal{A}}(\mathbf{r}) - \sum_{i \in \mathcal{A}} \sum_{nlm} c_{nlm}^i(\mathbf{x}) \phi_{nlm}(\mathbf{r} - \mathbf{r}_i) \delta_{\alpha_i \alpha_n} \right] \times \\ &\times \left[\sum_{i' \in \mathcal{A}} \sum_{n'l'm'} \frac{\partial c_{n'l'm'}^{i'}(\mathbf{x})}{\partial x_{\nu\lambda\mu}^\gamma} \phi_{n'l'm'}(\mathbf{r} - \mathbf{r}_{i'}) \delta_{\alpha_{i'} \alpha_{n'}} \right] = \\ &= 2\eta x_{\nu\lambda\mu}^\gamma - 2 \sum_{\mathcal{A} \in N} \sum_{i' \in \mathcal{A}} \sum_{n'l'm'} \left[w_{n'l'm'}^{i'} - \sum_{i \in \mathcal{A}} \sum_{nlm} c_{nlm}^i(\mathbf{x}) S_{nlm, n'l'm'}^{i, i'} \right] \frac{\partial c_{n'l'm'}^{i'}(\mathbf{x})}{\partial x_{\nu\lambda\mu}^\gamma} \end{aligned} \quad (\text{S8})$$

where we have defined the projections of the density on basis functions as

$$w_{n'l'm'}^{i'} = \int d\mathbf{r} \rho_{\mathcal{A}}(\mathbf{r}) \phi_{n'l'm'}(\mathbf{r} - \mathbf{r}_{i'}) \delta_{\alpha_{i'} \alpha_{n'}} \quad (\text{S9})$$

and the overlap between basis functions as

$$S_{nlm,n'l'm'}^{i,i'} = \int d\mathbf{r} \phi_{nlm}(\mathbf{r} - \mathbf{r}_i) \phi_{n'l'm'}(\mathbf{r} - \mathbf{r}_{i'}) \delta_{\alpha_i \alpha_n} \delta_{\alpha_{i'} \alpha_{n'}}. \quad (\text{S10})$$

Note that the overlap matrix \mathbf{S} is block diagonal in the space of dataset configurations since it only couples basis functions relative to the atoms of the same molecule \mathcal{A} . Upon plugging Eq. (S6) in place of the expansion coefficients, derivatives with respect to the regression weights are given by:

$$\frac{\partial c_{n'l'm'}^{i,i'}(\mathbf{x})}{\partial x_{\nu\lambda\mu}^\gamma} = k_{m'\mu}^\lambda(\mathcal{X}_{i'}, \mathcal{X}_\gamma) \delta_{\lambda i'} \delta_{\nu n'} \delta_{\alpha_{i'} \alpha_\gamma} \quad (\text{S11})$$

which leaves us with

$$\begin{aligned} \frac{\partial \ell(\mathbf{x})}{\partial x_{\nu\lambda\mu}^\gamma} &= 2\eta x_{\nu\lambda\mu}^\gamma - 2 \sum_{\mathcal{A} \in \mathcal{N}} \sum_{i' \in \mathcal{A}} \sum_{|m'| < \lambda} w_{\nu\lambda m'}^{i'} k_{m'\mu}^\lambda(\mathcal{X}_{i'}, \mathcal{X}_\gamma) \delta_{\alpha_{i'} \alpha_\gamma} \\ &+ 2 \sum_{j \in \mathcal{M}} \sum_{nlm''} x_{nlm''}^j \left\{ \sum_{\mathcal{A} \in \mathcal{N}} \sum_{i \in \mathcal{A}} \sum_{|m| < l} [k_{m''m}^l(\mathcal{X}_j, \mathcal{X}_i) \delta_{\alpha_j \alpha_i}]^T \sum_{i' \in \mathcal{A}} \sum_{|m'| < \lambda} S_{nlm, \nu\lambda m'}^{i,i'} k_{m'\mu}^\lambda(\mathcal{X}_{i'}, \mathcal{X}_\gamma) \delta_{\alpha_{i'} \alpha_\gamma} \right\} \end{aligned} \quad (\text{S12})$$

Setting to zero the previous equations, we get a system of coupled equations for the regression coefficients \mathbf{x} which can be linearly solved by a simple matrix inversion. Defining the problem dimensionality via the collective index $\{\gamma, \nu\lambda\mu\}$ this system can be written in matrix notation as follows:

$$0 = \eta \mathbf{x} - \mathbf{w} \mathbf{K} + \mathbf{x} \mathbf{K}^T \mathbf{S} \mathbf{K} \quad (\text{S13})$$

which corresponds to the formal solution:

$$\mathbf{x} = (\mathbf{K}^T \mathbf{S} \mathbf{K} + \eta \mathbf{1})^{-1} \mathbf{K}^T \mathbf{w} \quad (\text{S14})$$

Basis functions and overlap matrix

The atom-centered basis functions chosen are Gaussian Type Orbitals (GTOs), commonly used in quantum chemistry codes as primitive functions to construct optimized atomic functions. They are defined as:

$$G_{nlm}(\mathbf{r} - \mathbf{r}_i) = Y_{lm}(\theta, \phi) \times R_n^l(r) = Y_{lm}(\theta, \phi) \times \mathcal{N}_n^l |\mathbf{r} - \mathbf{r}_i|^l \exp \left\{ -\frac{1}{2} \left(\frac{|\mathbf{r} - \mathbf{r}_i|}{\sigma(n)} \right)^2 \right\} \quad (\text{S15})$$

where the normalization factor \mathcal{N}_n^l is defined such that:

$$\int_0^\infty dr r^2 R_n^l(r) R_n^l(r) = 1 \quad (\text{S16})$$

This choice allow us to get an analytic implementation of the overlap matrix. In particular, the overlap matrix can be computed analytically if working with the Cartesian counterpart of GTOs. For each l value of angular momentum, there are $(l+1)(l+2)/2$ Cartesian GTOs, namely:

$$\mathcal{G}_{\lambda_x \lambda_y \lambda_z}(\mathbf{r} - \mathbf{r}_i; n) = (x - x_i)^{\lambda_x} (y - y_i)^{\lambda_y} (z - z_i)^{\lambda_z} \exp \left\{ -\frac{1}{2} \frac{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}{\sigma^2(n)} \right\}$$

such that $l = \lambda_x + \lambda_y + \lambda_z$.

In particular, since each Cartesian direction is independent from each other, we have $\mathcal{S}_{\{\lambda_x \lambda_y \lambda_z\} \{\lambda_x \lambda_y \lambda_z\}'} = \mathcal{S}_{\lambda_x \lambda_x'} \mathcal{S}_{\lambda_y \lambda_y'} \mathcal{S}_{\lambda_z \lambda_z'}$, where, for instance,

$$\begin{aligned} \mathcal{S}_{\lambda_x \lambda_x'} &= \sqrt{\frac{\pi}{\alpha(n) + \alpha(n')}} \exp \{ -E (A_x - A_x')^2 \} \times \\ &\times \sum_{i_x=0}^{\lambda_x} \sum_{i_x'=0}^{\lambda_x'} \binom{\lambda_x}{i_x} \binom{\lambda_x'}{i_x'} \frac{(i_x + i_x' - 1)!!}{[2\alpha(n) + 2\alpha(n')]^{(i_x + i_x')/2}} (P_x - A_x)^{\lambda_x - i_x} (P_x - A_x')^{\lambda_x' - i_x'} \end{aligned} \quad (\text{S17})$$

with $(i_x + i'_x)$ assumed even, and

$$\alpha(n) = \frac{1}{2\sigma(n)}, \quad E = \frac{\alpha(n)\alpha(n')}{\alpha(n) + \alpha(n')}, \quad P_x = \frac{\alpha(n)A_x + \alpha(n')A'_x}{\alpha(n) + \alpha(n')}.$$

Going back to the spherical representation requires the proper transformation matrix \mathbf{C}_l which can be found, for instance, in Ref. S3. Finally a transformation matrix \mathbf{R}_l is also used to go from the complex to real space. With these choices, the final real and spherical overlap matrix $\mathbf{S}_{ll'}$ for a given couple of atoms i, i' and radial functions n, n' reads:

$$\mathbf{S}_{ll'} = \int d\mathbf{r} \mathbf{G}_l(\mathbf{r} - \mathbf{r}_i; n) \mathbf{G}_{l'}^T(\mathbf{r} - \mathbf{r}_{i'}; n') = \mathbf{R}_l^* \mathbf{C}_l^* \left(\int d\mathbf{r} \mathcal{G}_l(\mathbf{r} - \mathbf{r}_i; n) \mathcal{G}_{l'}^T(\mathbf{r} - \mathbf{r}_{i'}; n') \right) \mathbf{C}_{l'}^T \mathbf{R}_{l'}^T \quad (\text{S18})$$

which requires to pick the proper Cartesian overlap matrix of dimension $\frac{(l+1)(l+2)}{2} \times \frac{(l'+1)(l'+2)}{2}$ included in the parentheses (\cdot) .

Density projections over basis functions

The projection of the molecular density onto the atom-centered basis functions reads

$$w_{nlm}^i = \langle \phi_{nlm}^i | \rho \rangle = \int d\mathbf{S} S^2 R_n^l(S) Y_{lm}^*(\hat{\mathbf{S}}) \rho(\mathbf{r}_i + \mathbf{S}) \quad (\text{S19})$$

with $\mathbf{S} = \mathbf{r} - \mathbf{r}_i$. To accurately integrate spherical harmonic components over the unit sphere we use the Lebedev quadrature with 2030 points.^{S4-S6} The radial integration is performed with an equispaced radial mesh of 200 points spanning a cutoff distance of $r_{\text{cut}} = 6 \text{ \AA}$.

Error estimation on uniform grids

The comparison between the predicted machine-learning densities and the reference quantum-mechanical densities are carried out on Cartesian uniform grids with 0.1 Bohr of spacing and

with a number of points per side defined in such a way that the grid extends for at least 5 Å from any atom of the molecule. The density printing for quantum mechanical densities is carried out by using the Dgrid software.^{S7}

Basis set optimization

When using a multi-centred expansion one can easily encounter ill-conditioning issues consisting in a lowering of the overlap matrix rank. The standard strategy to avoid these problems involves the contraction of the radial basis set in a smaller set of optimized functions $R'_n(r)$ expressed as a linear combination of the starting primitive functions $R_{n'}(r)$. Formally we can write,

$$R'_n(r) = \sum_{n'} a_{nn'} R_{n'}(r) \tag{S20}$$

with $a_{nn'}$ the rectangular contraction matrix to be optimized. In our case, 12 primitive radial GTOs are contracted down to a total of 4 optimized functions in such a way that both the root mean square density error ε_ρ coming from the linear problem solution $\mathbf{c} = \mathbf{S}^{-1}\mathbf{w}$ and the condition-number ω of the overlap matrix \mathbf{S} are simultaneously minimized.^{S8} The two quantities are for this purpose suitably combined to modulate their relative importance in the final minimization target. In particular, we found that a good compromise consists in combining the root mean square density error with $10^{-3} \log \omega$. The primitive set of GTOs is defined in order to uniformly span the radial interval via the Gaussian width σ_n . The minimization is carried out by using the L-BFGS-B method of the *Scipy* optimization libraries. The systems which have been used for the optimization are given by 3 molecules of butane and 3 molecules of butadiene. This choice has demonstrated to encode enough chemical variability to lead to a basis set that can suitably represent all the carbon molecules considered in this work with comparable accuracy.

Dataset generation

Molecular structures included in the dataset were generated by sampling the trajectory of the lowest temperature (300 K) replica of a replica exchange molecular dynamics (REMD) simulation.^{S9,S10} Forces were computed at the DFTB3-UFF/3OB^{S11-S13} level as implemented in the DFTB+ software,^{S14} while the REMD was driven by the i-PI software.^{S15} The REMD simulation was performed in the NVT ensemble with a generalised Langevin equation (GLE) thermostat.^{S16} The time step of the dynamics was set at 0.25 fs for a total simulation time of at least 130 ps. 16 replica at 300 K, 364.3 K, 430.6 K, 498.9 K, 569.2 K, 641.7 K, 716.4 K, 793.4 K, 872.7 K, 954.5 K, 1038.7 K, 1125.5 K, 1215.0 K, 1307.2 K, 1402.1 K, 1500.0 K, were found sufficient to ensure enough exchanges within the simulation time. Upon removing the first picosecond of dynamics to get rid of the equilibration steps of the canonical ensemble at 300K, the resulting REMD trajectory consists of at least 10'000 independent configurations. On the top of this selection, farthest point sampling^{S17,S18} coupled with scalar 0-SOAP average kernel similarity measure, with 4.5Å cutoff radius and 0.2Å of Gaussian smearing, is finally used to extract the 1000 most diverse structures that we used to provide a diverse, challenging benchmark for the machine-learning model for the charge density.

Ab-initio computations

Ab-initio computations of the electron density (ρ) have been performed within the framework of density functional theory, using the PBE functional.^{S19,S20} The SBKJC-LFK basis set^{S21,S22} has been used for all atoms, combined with SBKJC effective core potentials^{S23} for the core electrons of second row elements. All DFT computations have been performed in a locally modified version of the GAMESS-US software.^{S24,S25}

Machine-learning parameters

All the tensorial λ -SOAP kernels for all the molecules of this work have been computed with the following parameters:

- environment cutoff: $r_{\text{cut}} = 4.5\text{\AA}$
- Gaussian smearing: $\sigma = 0.2\text{\AA}$
- angular cutoff: $l_{\text{cut}} = 6$
- radial cutoff: $n_{\text{cut}} = 8$
- environmental kernel exponent: $\zeta = 2$

It should be noted that the angular expansion of the SOAP descriptor has nothing to do with the expansion of the charge density field, which instead refers to the tensorial order λ of the SOAP kernel of Ref. S1 used to encode geometric covariances of density components. Since the density is expanded in real spherical harmonics, these tensorial kernel components are also expressed as real quantities.

On the top of the SOAP descriptor defined above, the feature space is reduced to the 500 principal components obtained by diagonalizing the covariance matrix defined in space of data points. The parameter η regularizing the regression has been set to 10^{-6} .

Selection of reference environments

The 1500 reference environments defining the dimensionality of the regression problem are defined via farthest point sampling (FPS) coupled with environmental 0-SOAP kernel computed using the same parameters described above. Fig. S1 reports a comparison between the charge density learning curves of butane computed with a different number of FPS reference environments.

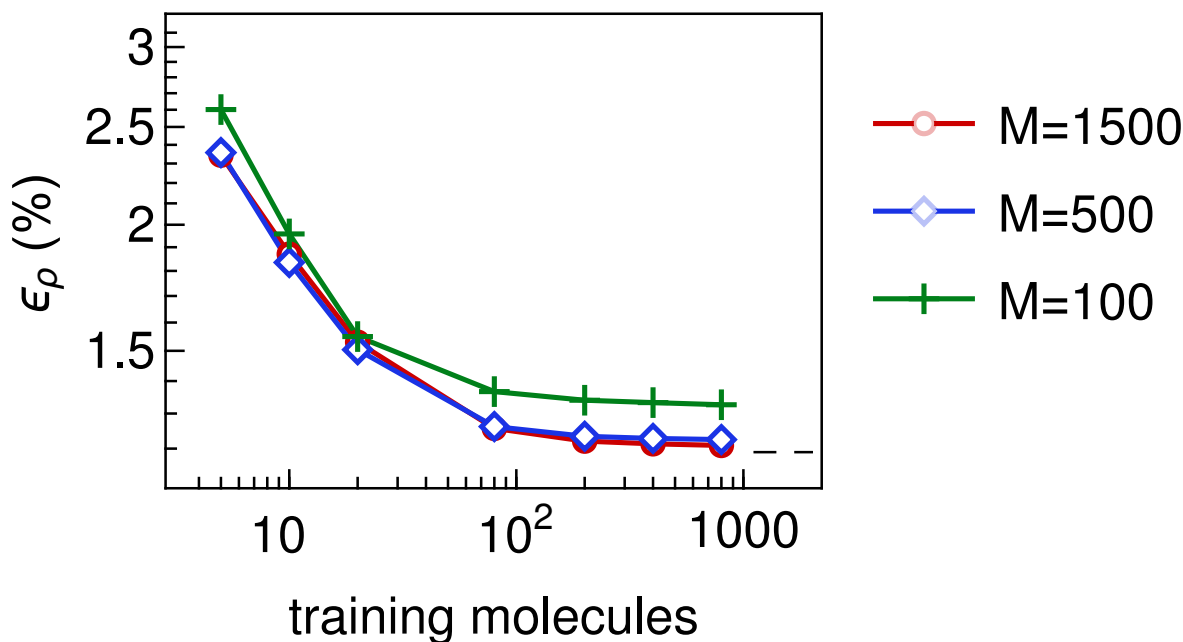


Figure S1: Learning curves comparison for the charge density prediction of butane at different numbers of reference environments M . Black dashed line represents the error brought by the basis set representation.

As shown, all models are limited by the basis set accuracy, and not by the sparsification of the reference environments. The case of $M = 1500$, that we used in the main text, is overconverged: 500 environments leads to almost indistinguishable accuracy, and even just 100 environments lead to an error increase of less than 20%.

Comparison between datasets

In Figure. S2 we compare the learning curves for density of ethane, as obtained for the same dataset we used in the main text, and from the dataset of Ref. S26. It can be seen that – although comparable – errors are slightly larger for our dataset, because structures were selected by FPS rather than randomly, and are therefore more diverse.

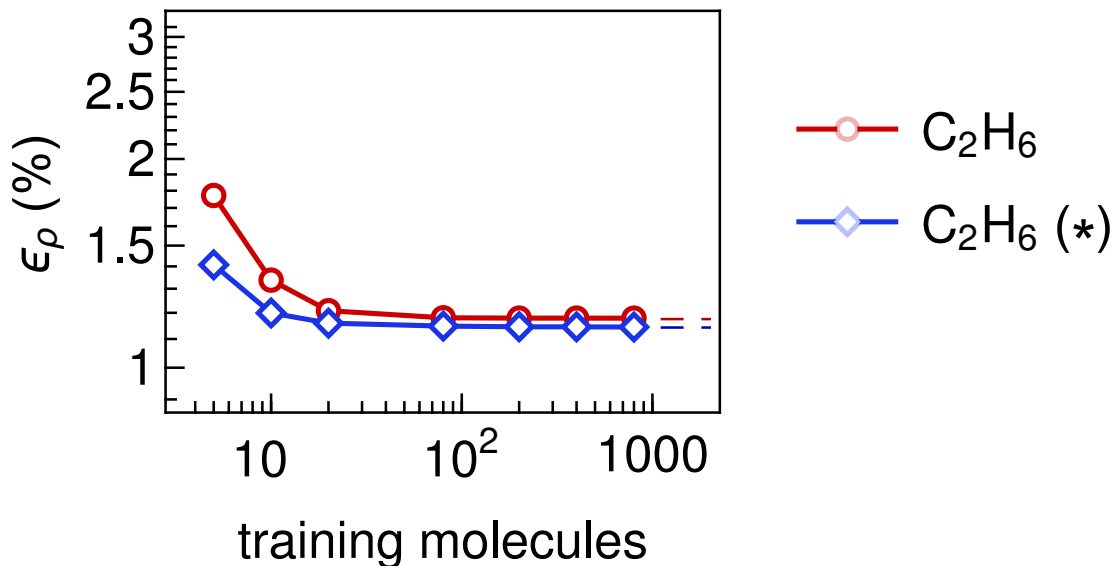


Figure S2: Comparison between the density learning curves of ethane (C_2H_6) for our 300K FPS selection (red) and the 300K selection used in Ref. ^{S26} (blue, \star).

Summary of regression performance

We report here a Table that summarize the SA-GPR performances both in terms of density and PBE exchange correlation energies at different training set sizes.

Table 1: Comparison among machine-learning (ML) performances of the four carbon dataset at different training size N , to be compared with the corresponding basis set (BS) error. Errors are averaged over three independent random selections of 200 test molecules.

N	ϵ_ρ (%)				ϵ_{XC} [kcal/mol]			
	BS	ML			BS	ML		
		5	80	800		5	80	800
C_2H_4	1.039	1.443	1.044	1.042	0.784	18.700	1.212	0.926
C_2H_6	1.142	1.771	1.180	1.145	1.633	26.853	2.442	1.700
C_4H_6	0.975	2.442	1.025	0.990	1.503	56.286	4.004	1.912
C_4H_{10}	1.190	2.344	1.257	1.209	2.721	49.544	7.557	3.367

References

- [S1] Grisafi, A.; Wilkins, D. D. M.; Csányi, G.; Ceriotti, M. Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems. *Phys. Rev. Lett.* **2018**, *120*, 036002.
- [S2] Bartók, A. P.; Gillan, M. J.; Manby, F. R.; Csányi, G. Machine-Learning Approach for One- and Two-Body Corrections to Density Functional Theory: Applications to Molecular and Condensed Water. *Phys. Rev. B* **2013**, *88*, 054104.
- [S3] Schlegel, H. B.; Frisch, M. J. Transformation Between Cartesian and Pure Spherical Harmonic Gaussians. *Int. J. Quantum Chem.* **1995**, *54*, 83–87, DOI: 10.1002/qua.560540202.
- [S4] Lebedev, V. I. Quadratures on a Sphere. *USSR Comput. Maths Math. Phys.* **1976**, *16*, 10–24.
- [S5] Murray, C. W.; Handy, N. C.; Laming, G. J. Quadrature Schemes for Integrals of Density Functional Theory. *Mol. Phys.* **1993**, *78*, 997–1014.
- [S6] Gill, P. M.; Johnson, B. G.; Pople, J. A. A Standard Grid for Density Functional Calculations. *Chem. Phys. Lett.* **1993**, *209*, 506–512.
- [S7] Kohout, M. DGrid, Version 5.0. *Dresden, Germany* **2017**,
- [S8] VandeVondele, J.; Hutter, J. Gaussian Basis Sets for Accurate Calculations on Molecular Systems in Gas and Condensed Phases. *J. Chem. Phys.* **2007**, *127*, 114105.
- [S9] Earl, D. J.; Deem, M. W. Parallel Tempering: Theory, Applications, and New Perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910.
- [S10] Petraglia, R.; Nicolai, A.; Wodrich, M. M. D.; Ceriotti, M.; Corminboeuf, C. Beyond Static Structures: Putting Forth REMD As a Tool to Solve Problems in Computational Organic Chemistry. *J. Comp. Chem.* **2016**, *37*, 83–92.

- [S11] Zhechkov, L.; Heine, T.; Patchkovskii, S.; Seifert, G.; Duarte, H. A. An Efficient a Posteriori Treatment for Dispersion Interaction in Density-Functional-Based Tight Binding. *J. Chem. Theory Comput.* **2005**, *1*, 841–847.
- [S12] Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem Theory Comput.* **2011**, *7*, 931–948.
- [S13] Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem Theory Comput.* **2013**, *9*, 338–354.
- [S14] Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.
- [S15] Ceriotti, M.; More, J.; Manolopoulos, D. E. i-PI: A Python Interface for Ab Initio Path Integral Molecular Dynamics Simulations. *Comp. Phys. Comm.* **2014**, *185*, 1019–1026.
- [S16] Ceriotti, M.; Bussi, G.; Parrinello, M. Colored-Noise Thermostats À La Carte. *J. Chem. Theory Comput.* **2010**, *6*, 1170–1180.
- [S17] Ceriotti, M.; Tribello, G. A.; Parrinello, M. Demonstrating the Transferability and the Descriptive Power of Sketch-Map. *J. Chem. Theory Comput.* **2013**, *9*, 1521–1532.
- [S18] De, S.; Bartók, A. A. P.; Csányi, G.; Ceriotti, M. Comparing Molecules and Solids Across Structural and Alchemical Space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.
- [S19] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [S20] Perdew, J. P.; Burke, K.; Ernzerhof, M. Erratum: Generalized Gradient Approximation Made Simple (Physical Review Letters (1996) 77 (3865)). *Phys. Rev. Lett.* **1997**, *78*, 1396.

- [S21] Labello, N. P.; Ferreira, A. M.; Kurtz, H. A. An Augmented Effective Core Potential Basis Set for the Calculation of Molecular Polarizabilities. *J. Comput. Chem.* **2005**, *26*, 1464–1471.
- [S22] Labello, N. P.; Ferreira, A. M.; Kurtz, H. A. Correlated, Relativistic, and Basis Set Limit Molecular Polarizability Calculations to Evaluate an Augmented Effective Core Potential Basis Set. *Int. J. Quantum Chem.* **2006**, *106*, 3140–3148.
- [S23] Stevens, W. J.; Basch, H.; Krauss, M. Compact Effective Potentials and Efficient Shared-Exponent Basis Sets for the First- and Second-Row Atoms. *J. Chem. Phys.* **1984**, *81*, 6026–6033.
- [S24] Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- [S25] Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry*; Dykstra, C., Frenking, G., Kim, K., Scuseria, G., Eds.; Elsevier: Amsterdam, 2005.
- [S26] Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham Equations with Machine Learning. *Nat. Comm.* **2017**, *8*, 872.