



Local Chemical Environments in Machine Learning

M. El Khatib, A. A. Peterson

Catalyst Design Laboratory, School of Engineering, Brown University,
Providence, RI, United States.

Introduction

- Recent advances on electronic structure theory have been devoted to decrease the computational time needed to explicitly solve the Schrödinger equation without the loss of accuracy. Machine-learning (ML) has emerged as a new and powerful tool to decrease the computational time to predict electronic properties by orders of magnitude.
- An important aspect in ML workflows lies in the so-called feature space (FS). In the context of atom-centered models, the FS exploits the concept of locality using local chemical environments (LCE) which are a vectorial representation of atoms in molecules¹. These LCEs can be seen as a functional form of atoms that is extracted from their atom positions, atomic element type, and neighboring atoms (input variables). They fulfill a series of requirements such as rotational and translational invariance, uniqueness, and size-extensivity. These properties make them more adequate to be plugged into learning algorithms and allow the creation of universal ML models.
- In this contribution, we explore different compositions of Behler-Parrinello symmetry functions and Zernike LCEs and evaluate these feature vectors and how they can play an important role to improve accuracy and model generalization towards the prediction of electronic properties.

Computational Details

Molecular Systems

- FCC copper bulk data set obtained from molecular dynamics (Velocity Verlet; 1600K) computed at density functional theory level (DFT) using the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional in plane-wave mode with the GPAW package.
- ISO17 MD Trajectories of $\text{C}_7\text{O}_2\text{H}_{10}$ isomers obtained with the Fritz-Haber Institute package (FHI-aims), using PBE and van der Waals corrections².

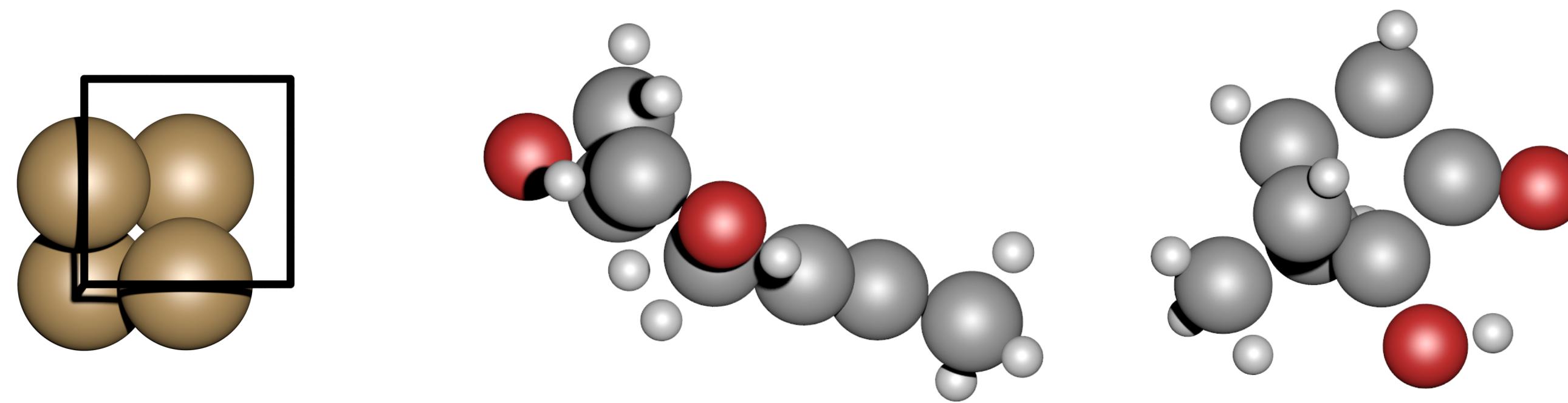


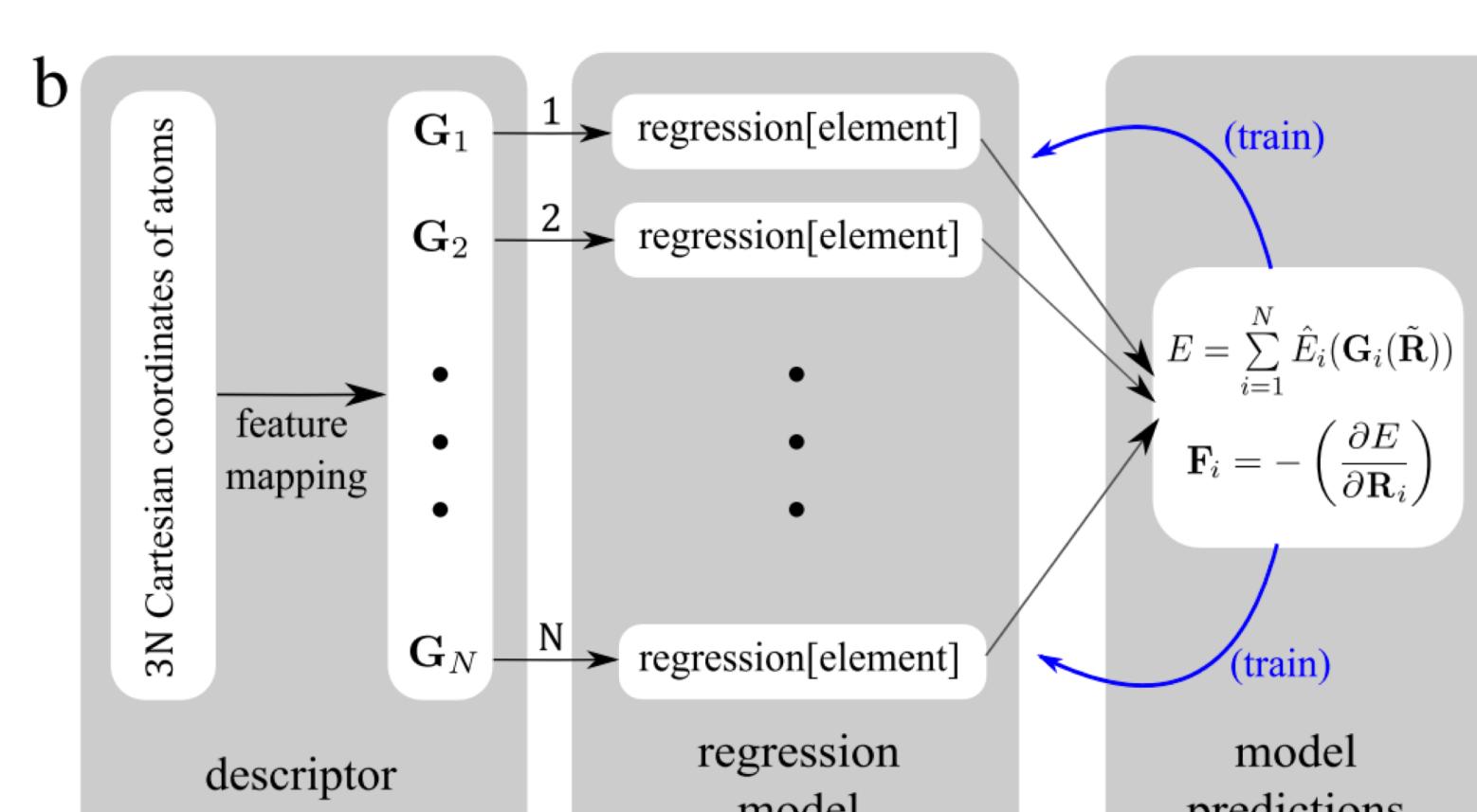
Figure 1. Schematic representation of FCC copper bulk, and two examples of $\text{C}_7\text{O}_2\text{H}_{10}$ isomers.

Machine-Learning Models

- Atom-centered calculations were carried out using the Atomistic Machine-Learning Package³ (Amp).
- Kernel Ridge Regression (KRR) applied using Cholesky decomposition⁴, and two stationary kernels:

Isotropic Kernel $k(\mathbf{x}, \mathbf{y})_{rbf} = \exp\left(-\frac{\ \mathbf{x} - \mathbf{y}\ _2^2}{2\sigma^2}\right)$	$\text{Anisotropic Kernel}$ $k(\mathbf{x}, \mathbf{y})_{rbf} = \exp\left(-\sum_{i=1}^D \frac{(x_i - y_i)^2}{2\sigma_i^2}\right)$
---	--

- Artificial Neural Nets (ANN) with topology (FS, HL, HL, output).



Descriptors

Behler-Parrinello Descriptor

$$f_i^I = \sum_{j \neq i}^{\text{atoms } j \text{ within } R_c \text{ distance of atom } i} e^{-\eta(R_{ij}-R_s)^2/R_c^2} f_c(R_{ij})$$

$$f_c(r) = \begin{cases} 0.5(1 + \cos(\pi r/R_c)) & \text{if } r \leq R_c \\ 0 & \text{if } r > R_c \end{cases}$$

$$f_i^{II} = 2^{1-\zeta} \sum_{\substack{j, k \neq i \\ (j \neq k)}}^{\text{atoms } j, k \text{ within } R_c \text{ distance of atom } i} (1 + \lambda \cos \theta_{ijk})^\zeta e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)/R_c^2} \times f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}).$$

Zernike Descriptors

$$c_{nl}^m = \langle Z_{nl}^m(\tilde{r}, \theta, \phi), f(\tilde{r}, \theta, \phi) \rangle \quad f(\tilde{r}, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{l=-l}^l c_{nl}^m Z_{nl}^m(\tilde{r}, \theta, \phi)$$

Experiments

FCC Copper Bulk

Neural Networks

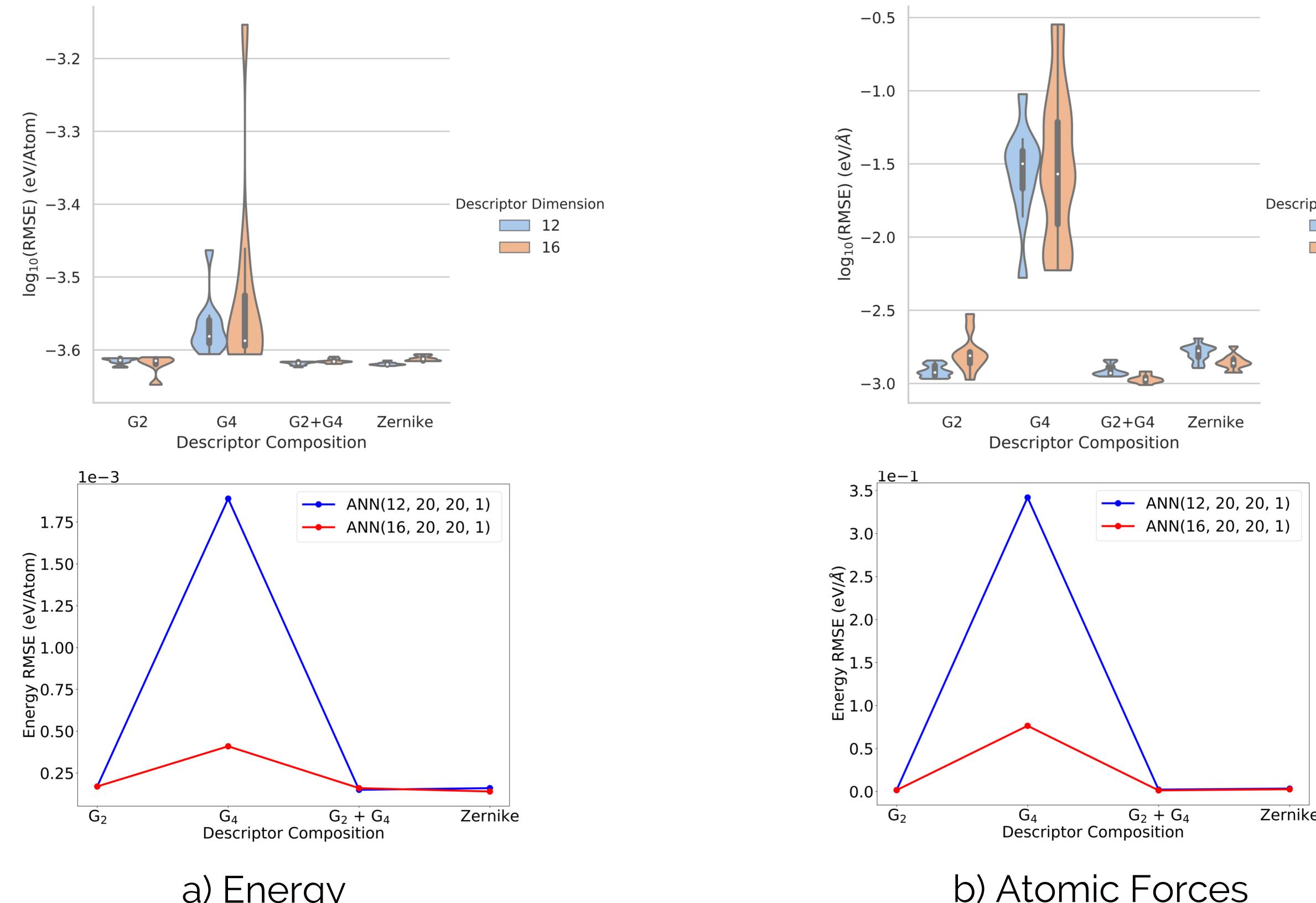


Figure 2: Violin plots of predictions' RMSE over the test set for 10 ANN models, and RMSE as a function of descriptor composition for the best model.

Kernel Ridge Regression

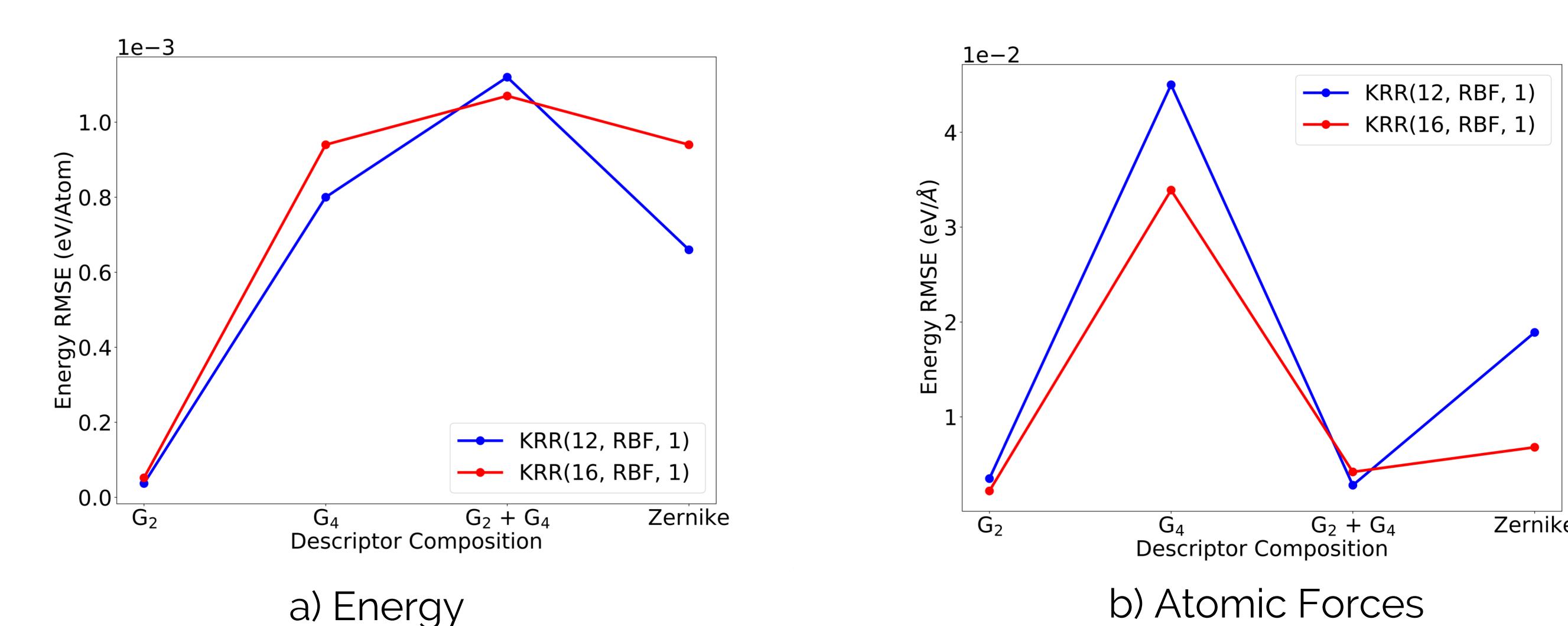


Figure 3: RMSE as a function of local chemical environment composition.

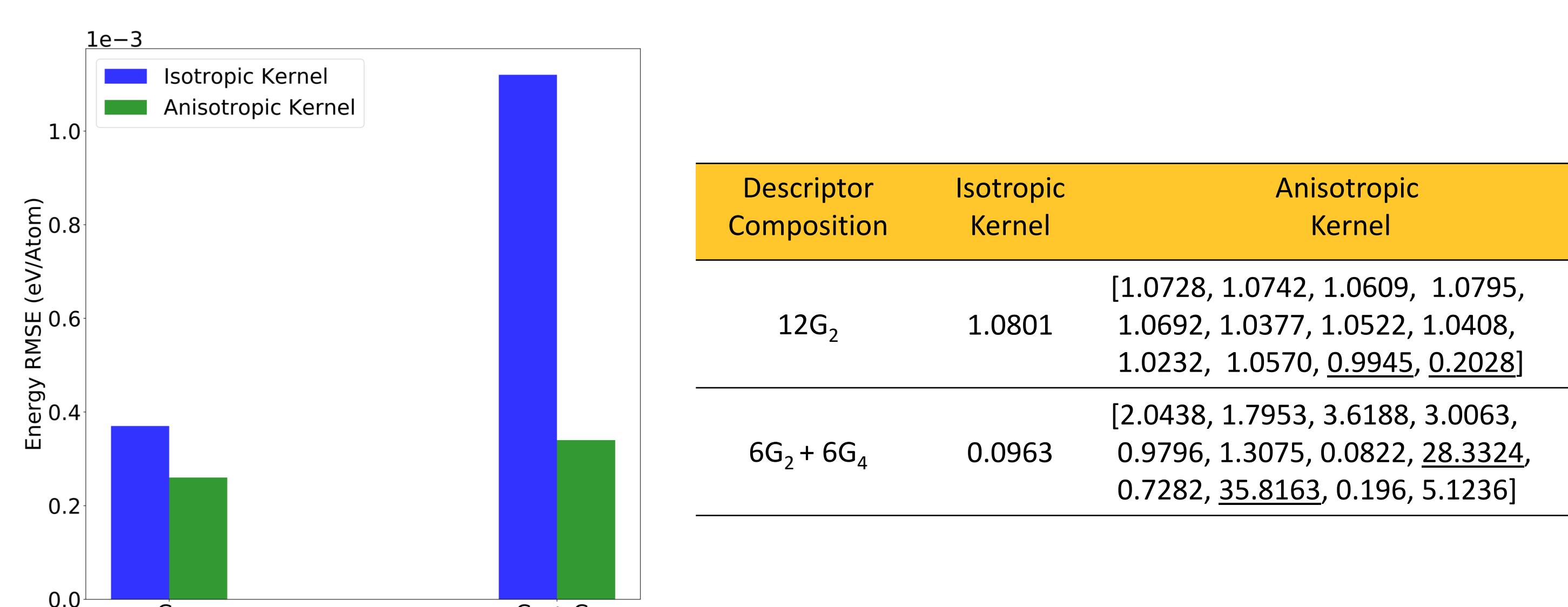


Figure 4: Effect of anisotropy of the feature vectors on the prediction of total energy.

ISO17 MD Trajectories

Kernel Ridge Regression

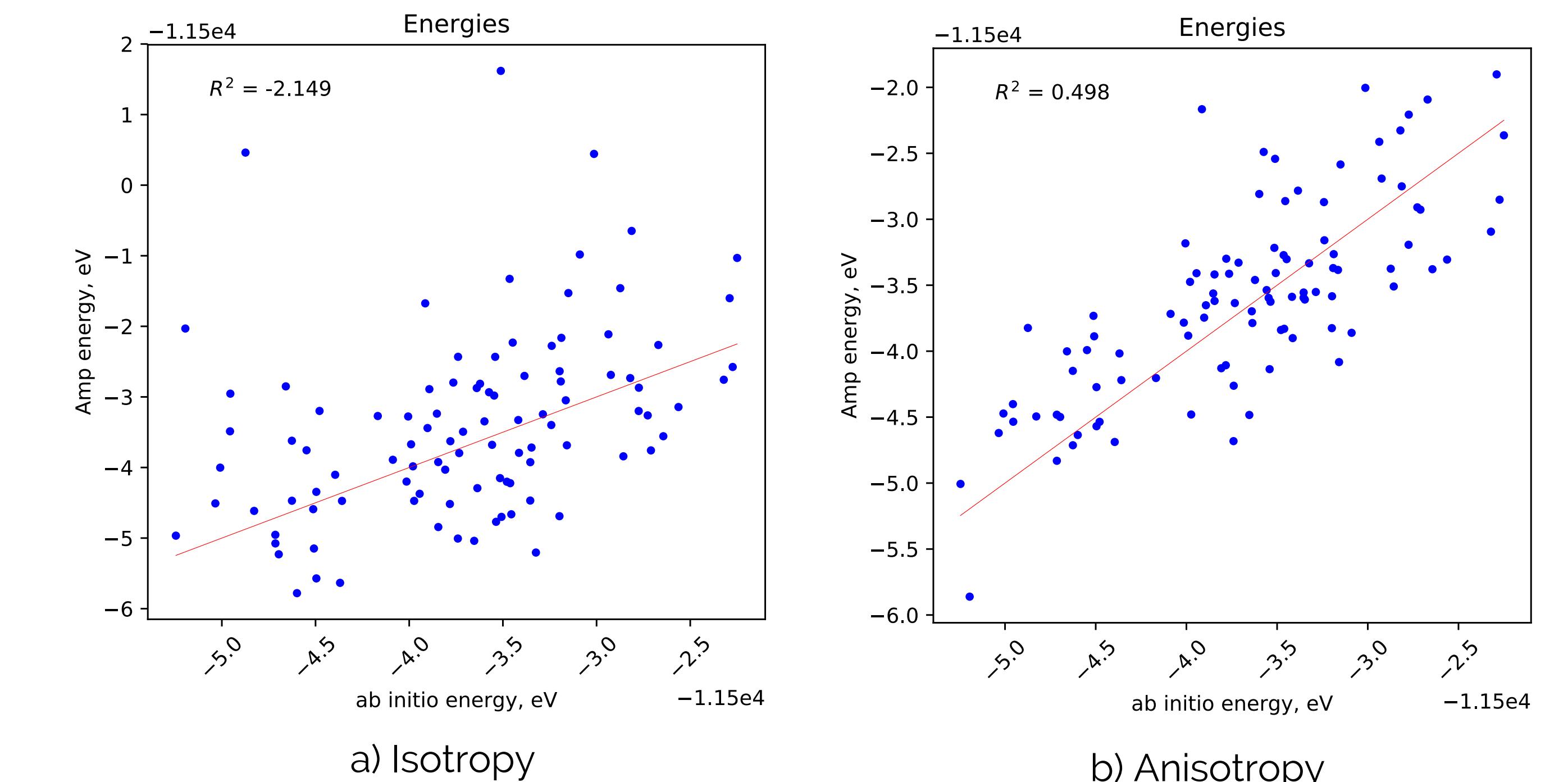


Figure 5: Parity plots for energy predictions over ISO17 MD test set using G2 + G4 symmetry functions with 12 dimensions.

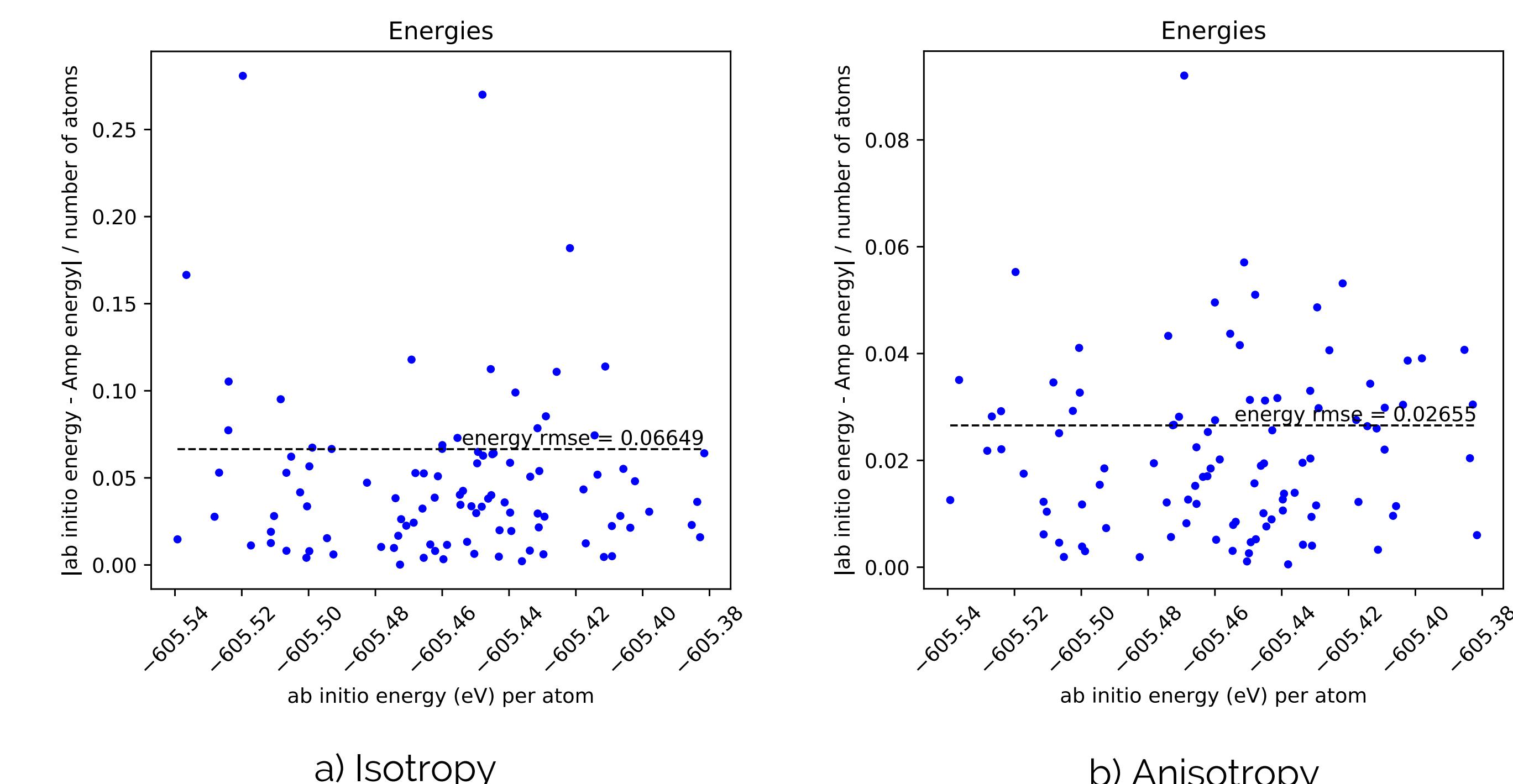


Figure 6: Error plots for energy predictions over ISO17 MD test set using G2 + G4 symmetry functions with 12-D.

Conclusions

- The dimension of the LCE affects model's RMSE over the test set. Larger dimensionalities tend to improve predictions.
- Feature vectors accounting only for diatomic symmetry functions are more important than triatomic ones.
- The anisotropy of the feature vector improves accuracy and generalization of KRR models as a consequence of a weighted "flexibility" given to each dimension of the feature vector through the hyper parameter σ of the kernel function.
- For the case of the ISO17 MD structures, the addition of anisotropy per atom type makes KRR to show chemical accuracy (1kcal/mol = 0.043eV) for atomic energy predictions with a rather small feature vector (12-D).

References

- J. Behler and M. Parrinello, Phys. Rev. Lett., vol. 98, no. 14, p. 146401, 2007.
- R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. Von Lilienfeld, Sci. Data, vol. 1, pp. 1–7, 2014.
- A. Khorshidi and A. A. Peterson, Comput. Phys. Commun., vol. 207, pp. 310–324, 2016. <https://bitbucket.org/andrewpeterson/amp>
- M. Rupp, Int. J. Quantum Chem., vol. 115, no. 16, pp. 1058–1073, 2015.