Supplementary Information

Quantum Active Learning for Structural Determination of Doped Nanoparticles - A Case Study of 4Al@Si_{11}

Maicon Pierre Lourenço, *[*,a] **Mosayeb Naseri,**[b,c] **Lizandra Barrios Herrera,**[c] **Hadi Zadeh-Haghighi,**[d] **Daya Gaur,**[e] **Christoph Simon**[d] **and Dennis R. Salahub**[c]*

[a]Departamento de Química e Física, Centro de Ciências Exatas, Naturais e da Saúde (CCENS), Universidade Federal do Espírito Santo, 29500-000 Alegre-ES, Brazil

[b]Digital Technologies Research Centre, National Research Council Canada, 1200 Montreal Road, Ottawa, K1A0R6, Ontario, Canada

[c]Department of Chemistry, Department of Physics and Astronomy, CMS Centre for Molecular Simulation, IQST Institute for Quantum Science and Technology, Quantum Alberta, QHA Quantum Horizons Alberta, University of Calgary, 2500 University Drive NW, Calgary, AB, T2N 1N4, Canada

[d]Department of Physics and Astronomy, Institute for Quantum Science and Technology, Quantum Alberta, and Hotchkiss Brain Institute, University of Calgary, Calgary, AB T2N 1N4, Canada

[e]Department of Math and Computer Science, University of Lethbridge, 4401 University Dr. West Lethbridge, AB T1K 3M4, Canada

*e-mail: maiconpl01@gmail.com

Root mean square error of optimum hyperparameters used during the QAL (with QGPR) and classical AL (with GPR) cycles

Table S1. Hyperparameters used in this work[1] as defined in the scikit-learn and sQUlearn libraries for the 4Al@Si_{11} nanoparticle. 95% of the data were randomly chosen for the training set and 5% for testing for 4Al@Si_{11}, as defined during the QAL cycles using exploitation for decision making. The cost function used to evaluate the quality of the regression was the root mean square error (RMSE) and it is presented for different data sizes obtained by QAL

Hyper.	20 data		ca. 100 data		ca. 200 data	
	RMSE train / Hatree × 10^{-2}	RMSE test / Hatree × 10^{-2}	RMSE train / Hatree × 10^{-2}	RMSE test / Hatree × 10^{-2}	RMSE train / Hatree × 10^{-2}	RMSE test / Hatree × 10^{-2}
featureMap = HighDim; qkernel1 = PQK; σ = 0.0001	0.0012	0.0200	0.0067	0.0497	0.0061	0.0797
featureMap = YZ_CX; qkernel2 = PQK; σ = 0.0001	0.0129	0.0165	0.0066	0.0805	0.0065	0.1230
kernel1 = DotProduct + White, σ_ bounds[1.0, (10^{-3}- 10^{3})], noise_level_bounds[10.0, (10^{-3}- 10^{3})]; α = 1.0	0.1691	0.1407	0.4766	0.6039	0.4207	0.4601
Kernel2 = Constant*RBF, constant_value_bounds[1.0, (10^{-3}- 10^{3})], length_scale_bounds[10.0, (10^{-2}- 10^{2})]; α = 1.0	0.0340	67.9609	0.128	0.2349	0.0898	0.1485

Hyper.: hyperparameters used in the QGPR (using projected quantum kernel, PQK, or fidelity quantum kernel, FQK) or classical GPR (radial basis function, RBF, kernel) regression. The RMSE is in Hatree × 10^{-2}. The MBTR descriptor was used and PCA was applied, allowing a dimensionality reduction of 4.

Average total energy of $4Al@Si_{11}$ with standard deviation with 4 qubits in the quantum circuits

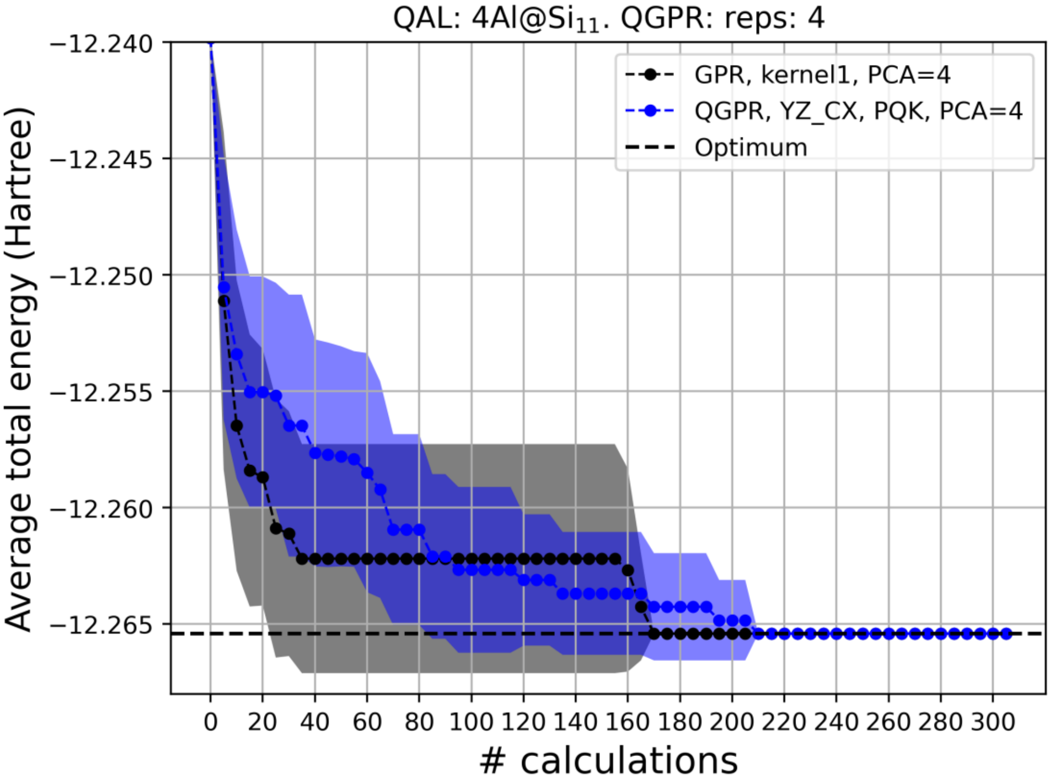

Figure S1. Average total energy (in Hartree) and the standard deviation obtained by DFTB for $4Al@Si_{11}$ obtained by 10 QAL and AL independent runs as a function of new calculations (with $N_{cycles} = 60$ and $N_{selected} = 5$). GPR: Classical Gaussian process (GPR) regression with DotProduct and White kernel combination - kernel1. QGPR: Quantum Gaussian process regression with XFeatureMap (X = YZ_CX) and Projected Quantum Kernel (PQK). PCA = X: principal component analysis with X main components chosen from MBTR descriptor. "reps": number of times the quantum circuit is decomposed. The black and blue shadows are the standard deviation obtained by 10 independent AL and QAL runs, respectively.

Average total energy of $4Al@Si_{11}$ with standard deviation with 8 qubits in the quantum circuits

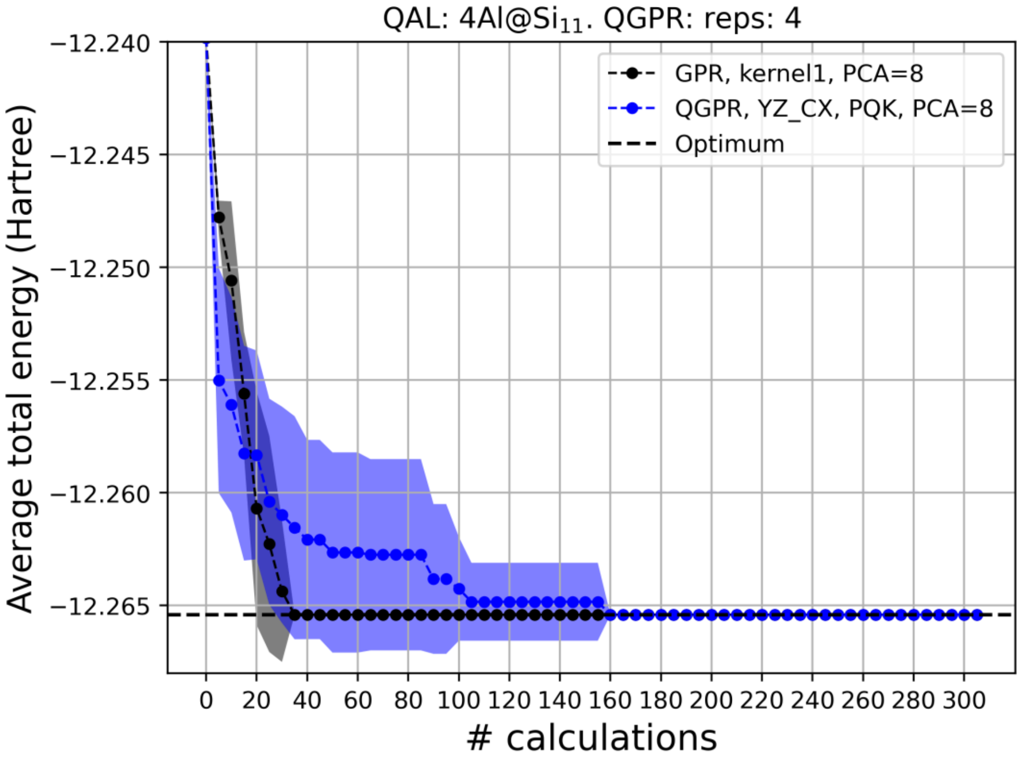

Figure S2. Average total energy (in Hartree) obtained by DFTB for $4Al@Si_{11}$ obtained by 10 QAL and AL independent runs as a function of new calculations (with $N_{cycles} = 60$ and $N_{selected} = 5$). GPR: Classical Gaussian process (GPR) regression with DotProduct and White kernel combination – kernel1. QGPR: Quantum Gaussian process regression with XFeatureMap (X = YZ_CX) and Projected Quantum Kernel (PQK). PCA = X: principal component analysis with X main components chosen from MBTR descriptor. "reps": number of times the quantum circuit is decomposed. The black and blue shadows are the standard deviation obtained by 10 independent AL and QAL runs, respectively.

Grid search hyperparameterization for classical (GPR) and quantum (QGPR) machine learning models

The hyperparameters for GPR (using "DotProduct + white" kernel) and QGPR (using HighDim encoding and projected quantum kernel) were obtained using a gridsearch. MBTR descriptor was used with dimensionality reduction using PCA. The first 4 PCA principal components were used as input features (descriptors) for both the GPR and the QGPR algorithms. Subsequently, the optimal hyperparameters were selected and used in the QAL and the AL methods for 60 $4Al@Si_{11}$ isomers or homotops. This grid search is in line with the supporting information of reference[2] and scikit-learn[3] for classical machine learning (ML) and sQULearn[4] for quantum ML were used.

Structural of 4Al@Si₁₁:

Data size: 60 isomers or homotops.

alpha hyperparameters grid for classical GPR:

'gpr__alpha': [1e15, 1e10, 1e5, 1e3, 10**, 1.0**, 1e-1, 1e-3, 1e-5, 1e-10, 1e-15]

sigma hyperparameters grid for quantum QGPR:

‘qgpr__sigma': [1e-6, 1e-5,**1e-4**, 1e-3, 1e-2, 1e-1, 1e0, 1e1, 1e2, 1e3]

In bold is presented the optimum hyperparameters found.

References

1. Lourenço, M. P.; Naseri, M.; Barrios Herrera, L.; Zadeh-Haghighi, H.; Gaur, D.; Simon, C.; Salahub, D.; *arXiv* **2024**. [Crossref]

2. Huang, H.-Y.; Broughton, M.; Mohseni, M.; Babbush, R.; Boixo, S.; Neven, H.; McClean, J. R.; *Nat. Commun.* **2021**, *12*, 2631. [Crossref]

3. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É.; *J. Mach. Learn. Res.* **2011**, *12*, 2825. [Crossref]

4. Kreplin, D.; Willmann, M.; Schnabel, J.; Rapp, F.; Roth, M.; *arXiv* **2024**. [Crossref]