Material Discovery With Quantum-Enhanced Machine Learning Algorithms

Ignacio Fernández Graña[1], Savvas Varsamopoulos[1], Tatsuhito Ando[2], Hiroyuki Maeshima[2], and Nobuyuki N. Matsuzawa[2]

[1]*Pasqal, 7 rue Léonard de Vinci, 91300, Massy, France*
[2]*Panasonic Industry Co., Ltd., 1006, Oaza-Kadoma, Kadoma, Osaka, Japan*
Contact email: *ignacio.fernandez-grana@pasqal.com*

Developing computational methods for material discovery is a famously complex challenge due to the large size of the chemical design space. An important family of chemical structures are *organic semiconductors*, due to their versatile material properties and ease of manufacturing. However, the applications of such compounds are currently limited by their low carrier mobilities, which is a key property that is hard to access experimentally. Figures of merit such as the Hole Reorganization Energy (HRE) have been identified as good estimators of the charge mobility while remaining computationally tractable. Virtual screenings calculating these figures of merit for a large body of candidates are commonly used to identify new promising compounds, despite being computationally intensive due to the large solution space.

Recently, Machine Learning (ML) models have been proposed as reliable predictors of target molecular properties such as the HRE [1], potentially reducing the computational requirements of data-hungry virtual screenings. In parallel, Quantum Machine Learning (QML) has emerged as a new paradigm that may provide certain advantages compared to conventional ML algorithms, such as a greater expressibility and generalization capabilities. The main goal of this work is to study the potential of *quantum-enhanced* ML algorithms in the field of material discovery by focusing on the specific problem of identifying low-HRE organic semiconductors, classically emulating the quantum algorithms in a supercomputer cluster.

We focus on algorithms based on the *extremal learning* framework, whose aim is to find the inputs that *extremize* the figure(s) of merit of a given dataset. They do so by first building a model of the relation between the inputs and the figure(s) of merit (*learning phase*), potentially from very few data. Then, a procedure is followed to identify the optimal (or extremal) input given the learned model (*selection phase*). The outcome of the algorithm is a list of candidate inputs with a high probability of being close to optimal solutions. Depending on the particular algorithm, both phases can be done via either classical or quantum computation. We focus on two quantum-enhanced algorithms, namely *Quantum Extremal Learning* (QEL) [2] and the *Factorization Machine + Quantum Approximate Optimization Algorithm* (FM+QAOA) [3]. QEL makes use of a QML model (Quantum Neural Networks) in the learning phase and a brute-force classical method in the selection phase. On the other hand, FM+QAOA follows the opposite approach, using a classical ML model (FM) in the learning phase and a QML model in the selection phase (QAOA).

To construct the training dataset, we leverage a dataset previously generated in a large-scale virtual screening of heteroacene compounds [4] from which only the the lowest HRE compounds were used. Each compound is then mapped into a binary representation using N bits to encode information about its molecular structure comprising their rings and bonds. Three different datasets were constructed with $N = 7, 16$ and 22 bits. Using more bits allows the access to a larger solution space as a wider variety of molecules can be encoded with increasing number of bits. The QEL algorithm could not be applied to the $N = 22$ dataset due to the large computational requirements of simulating the algorithm in classical hardware. For the $N = 16, 22$ cases, an iterative process (Figure 1) was followed to progressively increase the number of low-HRE compounds in the training dataset. At every iteration, each algorithm was trained to predict a list of candidates with the lowest predicted HRE. The HREs of these candidates were subsequently calculated through Density Functional Theory (DFT) methods to assess their validity, and then incorporated into the existing training dataset for the next learning iteration.

The main results are condensed in Figure 2. The quantum-enhanced algorithms consistently predicted low-HRE compounds, and both were able to extrapolate to molecules with a lower HRE than any of the molecules in the training dataset (0.0724eV), showcasing good generalization capabilities. As iterations progress and the algorithms are trained with more quality data, the predicted candidates are clustered more towards the low-HRE regime. The lowest HRE achieved in each iteration (bottom whiskers in the plot) also decreases steadily for all algorithms. In the 16 bits dataset case, both algorithms extrapolated to lower HRE molecules while training with roughly 3% of the complete solution space, predicting the same molecule with a HRE=0.0705eV. The FM+QAOA required as little as 0.47% of the solution space for training when applied to the 22 bits dataset, predicting a molecule with a HRE=0.0615eV already at iteration 1.

These results, although still limited, offer an exciting prospect regarding the use of QML algorithms for material discovery. However, future work to improve the scalability of the algorithms will be required in order to scale up the size of the datasets. In QEL, the brute-force selection must be substituted with a more efficient method, for example a QNN adapted to extremize over discrete inputs, while in FM+QAOA improvements to the training process of the QAOA algorithm would benefit convergence with a larger number of qubits.

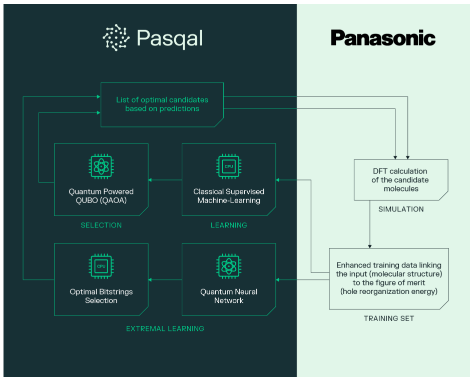

Figure 1: Schematic diagram of the iterative process.

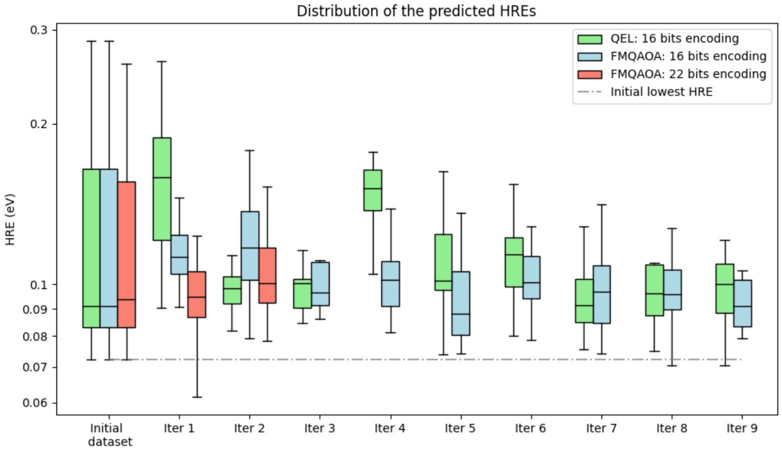

Figure 2: Box plot showing the distribution of the Hole Reorganization Energies of the candidates predicted by each algorithm in each iteration, calculated through DFT methods.

References

[1] G. Marques *et al.*, en, *J. Phys. Chem. A*, vol. 125, no. 33, pp. 7331–7343, 2021.

[2] S. Varsamopoulos *et al.*, en, *Quantum Mach. Intell.*, vol. 6, no. 2, pp. 1–27, 2024.

[3] K. Kitai *et al.*, *Phys. Rev. Res.*, vol. 2, p. 013 319, 1 2020.

[4] N. N. Matsuzawa *et al.*, en, *J. Phys. Chem. A*, vol. 124, no. 10, pp. 1981–1992, 2020.