Analysis of high-resolution AFM images with deep learning

Rubén Pérez[1,2]

[1] *Departamento de Física Teórica de la Materia Condensada and*
[2] *Condensed Matter Physics Center (IFIMAC)*
Universidad Autónoma de Madrid, E-28049 Madrid, Spain
ruben.perez@uam.es

Non-contact atomic force microscopy with CO-functionalized metal tips (HR-AFM) reveals the internal structure of adsorbed organic molecules with unprecedented resolution, resolving intermolecular features, determining bond orders, and characterizing intermediates and final products generated in on-surface reactions [1]. Recent advances in the interpretation of the AFM contrast using DFT-based AFM simulation methods [2,3] show that there are clear connections between fundamental chemical properties of molecules and key features imprinted in HR-AFM images.

Deep learning (DL) models are nowadays routinely used to classify, interpret, describe and analyze images, providing machines with capabilities that surpass human beings. The DL ability to recognize patterns has been exploited, in the context of HR-AFM, to extract information about the structure and chemical composition [4], and the electrostatic properties [5] of molecular systems using convolutional neural networks (CNN). The training of these DL models requires very large data sets, that mostly comprised theoretically simulated HR-AFM images.

Using one of these data sets, QUAM-AFM [6], different deep learning models have been designed and trained to achieve a complete identification of an arbitrarily complex, unknown molecule. Multimodal recurrent networks (M-RNNs) [7], that combine CNNs for image analysis and RNNs for natural language processing can determine the chemical groups and their connection within the molecule to predict the IUPAC name from HR-AFM images with an accuracy better than the best language translation models. Some of the limitations of M-RNNs can be overcome using Conditional Generative Adversarial Networks (CGANs) [8] or exploiting other descriptors as the molecular fingerprints [9].

[1] L. Gross, et al., *Angew. Chem*.Int. Ed. **57**, 3888 (2018)
[2] M. Ellner, P. Pou, R. Perez, ACS Nano 13, 786 (2019)
[3] A. Liebig, P. Hapala, A. J. Weymouth, & F. J. Giessibl, Sci. Rep. 10, 14104 (2020)
[4] B. Alldritt, et al., Sci. Adv. 6, eaay6913 (2020).
[5] N. Oinonen, et al., ACS Nano 16, 89 (2022).
[6] J. Carracedo-Cosme, et al., *J. Chem. Inf. Model.* 62, 1214 (2022)
[7] J. Carracedo-Cosme, et al., ACS Appl. Mater. Interfaces 15, 22692 (2023)
[8] J. Carracedo-Cosme and R. Perez, npj Comp. Mater. 10:19 (2024)
[9] M. Gonzalez-Lastre, et al., arXiv:2405.04321v1 (2024)