

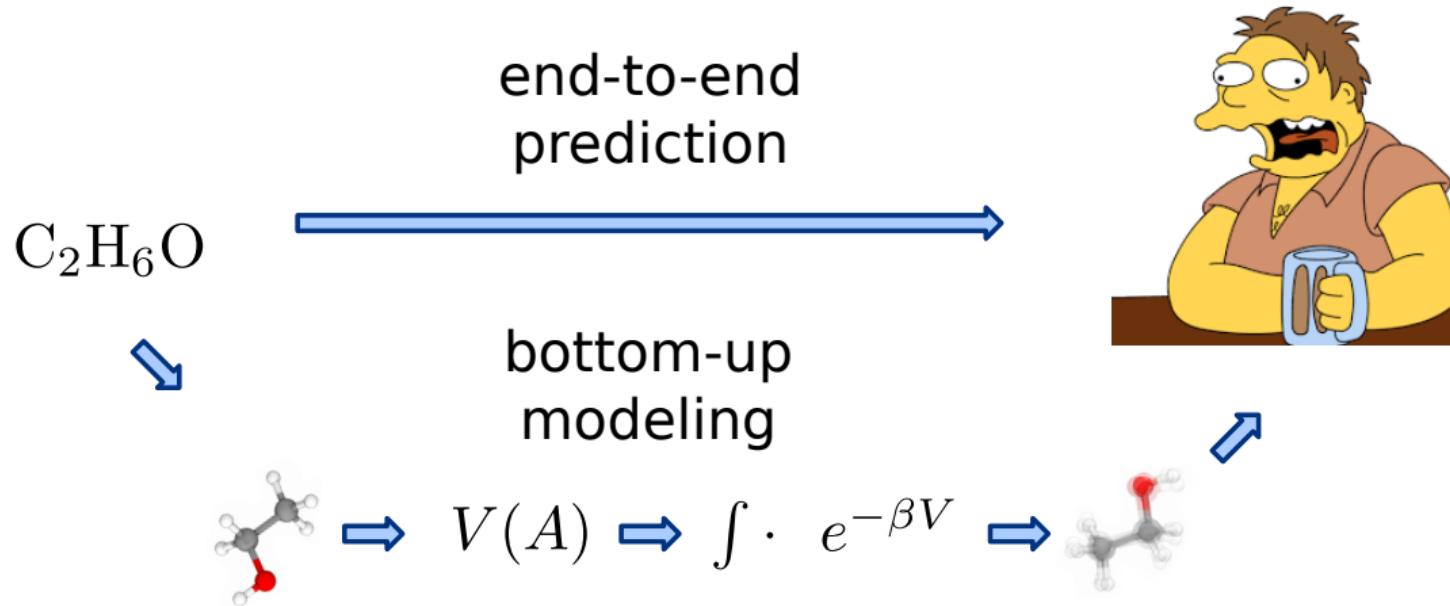
Introduction to atomic-scale machine learning

Michele Ceriotti
cosmo.epfl.ch

- Supervised and unsupervised learning with linear methods
 - An overview of ML for materials and molecules
 - Linear methods: PCA, ridge regression, PCovR
 - The kernel trick
- Atom-centered representations for atomistic machine learning
 - Mathematical requirements for a well-behaved featurization
 - Internal coordinates and symmetry functions
 - Construction of structural representations as correlations of the atom-density
- Equivariant models for covariant properties
 - Rotations and the rotation group
 - Equivariant representations
 - Equivariant models of tensorial properties

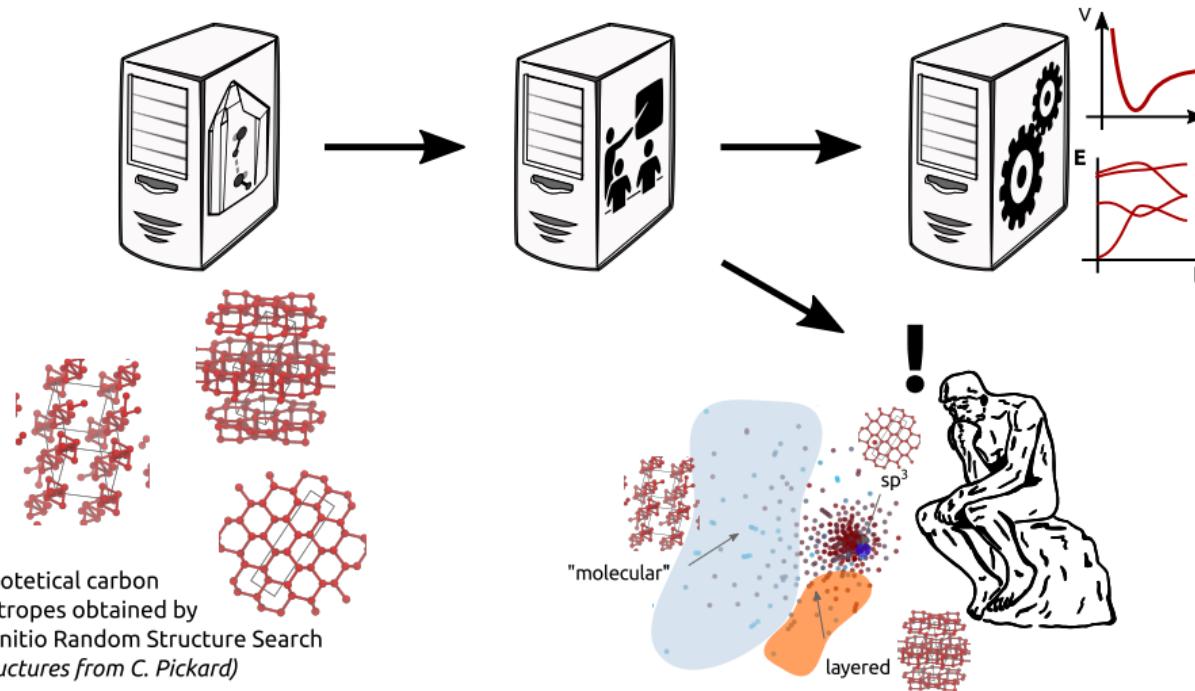
Bottom-up modeling for chemistry

- Most widespread application of ML targets the prediction of the end property
- These lectures focus on the application as part of “bottom-up”, physics-based modeling



Black-box predictions and intuitive understanding

- Machine learning can be used to predict molecules or materials' properties or to process large amounts of complex simulation data to produce a human-understandable form
- Focus on interpretability helps achieve *both* accurate predictions and intuitive understanding



Hands-on examples

- Will be showing examples of how to compute many of these models.
- Jupyter notebooks available for download on <https://github.com/ceriottm/ml-intro>
- These are *not* practical examples: over-simplified models and over-verbose implementations, purpose is to clarify concepts

The screenshot shows a Jupyter notebook interface. On the left, there is a sidebar with a tree view of the notebook's structure, including sections like "1.3. Visualize the environments ...", "2. A structure-property map for az...", and "2.4. Ridge regression". The main area contains a code cell with the following Python code:

```
if chemiscope.jupyter._is_running_in_notebook():
    from IPython.display import display
    display(cs)
```

Below the code cell is a PCA plot titled "PCA[1]" with a color scale from -0.05 (dark blue) to 0.05 (red). The plot shows a dense cloud of points with two distinct clusters. A red dot is highlighted on the right side of the plot. At the bottom of the plot are two buttons: "structure 1" (blue) and "atom 2" (green). Below the plot are two sets of navigation controls: a top row with arrows and a bottom row with arrows, each accompanied by a colored dot (blue for top, green for bottom).

2. A structure-property map for azaphenacene organic conductors

This example uses a small dataset of 156 molecular crystals (Yang et al., Chem. Mater. (2018)), assembled from 28 different structural isomers of azaphenacene derivatives. Lattice energy and electronic mobility have been computed for each structure, as well as a favorable classification of the resulting materials as conductors or insulators based on the mean number of hydrogen bonds per molecule.