#

Introducing Espeem Web-App: A Convenient Simulation Service for Organic Synthesis Researchers

# Main Features and Benefits

As an experimental researcher, identifying molecules on a crystal surface under UHV is a vital part of organic on-surface synthesis research. However, comparing experimental STM images to simulated ones is often challenging due to poor quality simulations found in scientific literature. The problem is not the lack of good methods for STM simulation, but rather the inconvenience of using them. Espeem is a simulation method designed specifically for regular use by experimental researchers, developed by a collaboration between a theorist and two organic synthesis researchers.

One of the primary issues with simulation methods is managing hardware and software resources, as advanced simulation methods can be computationally costly. The installation of (often pre-commercial) software to run on such clusters can also be challenging. To solve this issue, Espeem offers a single-page Software-as-a-Service (Saas) solution, allowing experimentalists to input only a few relevant parameters and receive simulations without maintaining a cluster or installing programs.

Espeem also addresses three common problems in STM simulations found in scientific literature. Firstly, the unrealistic assumption of tip distance to the surface can lead to inaccuracies. In Espeem, we solve this problem by using the Paz/Soler method, allowing us to evaluate arbitrary distances from the surface.

Secondly, one common problem with STM simulations in scientific literature is the lack of consideration for spatial noise, which can result in overly sharp features in theoretical predictions compared to experimental images, leading to inaccurate comparisons. In some cases, when spatial noise is modeled, it is done incorrectly, resulting in smoothed features that do not accurately represent the actual experimental data. Espeem addresses this issue by applying a Gaussian lateral filter to the current before generating current iso-surface images, providing more accurate simulations that can reproduce sharp features in experimental images where the height of the topography drops suddenly.

Finally, when molecules become non-planar, the choice of the assumed angle compared to the surface can be difficult. In Espeem, we introduce a potential that simulates the effects of a surface potential, making comparisons between molecules possible.

Espeem presents a convenient simulation service designed for experimental researchers that need to identify molecules on a crystal surface under UHV. With our single-page Software-as-a-Service solution and addressing common STM simulation issues, we offer a convenient and accurate simulation method for organic on-surface synthesis research.

# Examples

Small nanoflake – as we can see it is not always obvious to relate low spatial noise images and high spatial noise images.



Figure Left: Atomic model of small nanoflake, Center/Right: Constant-current image simulation of the LUMO frontier orbital with a low/high spatial broadening.

Non-planar molecule – molecules that are not flat can have relatively sharp features. A critical feature in our web-app is a consistent way to orient a non-planar molecule.



Figure Left: Atomic model of compact nonplanar molecule, Right: Constant-current image simulation of the HOMO frontier orbital with a medium spatial broadening.

# STM Method: Design and Details

## Design Principles

Our STM simulation method is designed to produce a prediction solely from the molecule ignoring the effects of both substrate and tip. The idea is that we assume both substrate and tip to be as generic and feature-less as possible and thereby generate a type of reference simulation. If the assumptions are correct for a specific substrate and tip the experimental image should match the simulated one. If the assumptions are not correct and experiment and theory diverges, our method will at least show that the substrate and tip diverge from the assumptions.

We also aim to eliminate unnecessary parameters for the end user. For instance, precise positions of HOMO/LUMO levels are difficult to reproduce, but experimentalists typically increase the voltage until a strong response is obtained. Therefore, we let users choose the desired frontier orbital rather than an absolute voltage.

## Step-by-Step Method

1. Geometric relaxations are performed using spin-polarized density-functional theory (DFT) with the SIESTA code [1]. A double-zeta polarized (DZP) basis set is employed, with orbital radii defined by a 100 meV energy shift. The Perdew-Burke-Ernzerhof (PBE) exchange-correlation potential [2] was utilized, along with a real-space grid equivalent to a 200 Ry plane-wave cutoff. Forces were relaxed until they reached values below 0.020 eV/Å. 6 Å of vacuum is used in all directions. The molecule was relaxed with a potential emulating a generic surface: P[z] = 1.00 \* (2\*(z+2)\*\*(-3) - (z+2)\*\*(-2)) (in units of [eV, Ang], minimum at z=0).
2. A final evaluation of the electronic structure is conducted with 12 Å of vacuum.
3. For STM images, we adopted the surface integration technique proposed by Paz and Soler [3]. The first step in Paz/Soler is to store the value of wavefunctions within the bias window on the isosurface of the density – here we chose 10^(-3) Ang^(-3) as this iso-value.
4. As the second step in Paz/Soler we find the overlap with an assumed spherical tip frontier orbital – also known as the Tersoff-Hamann approximation [4], with a Fast-Fourier transform convolution technique. The wavefunctions overlaps are evaluated assuming a flat electrostatic potential beyond the iso-density surface. All wavefunctions are calculated as if they are at the Fermi Level [5]. The connection between overlap and current is arbitrarily chosen as a proportionality factor of 1.00 nA·Å^(-3).
5. States are included in the image evaluation using a bias window step function broadened by a Gaussian kernel (Voltage Broadening),
6. To simulate the effect of spatial uncertainty, which reduces the resolution of STM images, we convolute the currents with a 2-D Gaussian kernel (Spatial Broadening).
7. To emulate constant-current imaging the chosen iso-current value (Setpoint Current) is used to generate the image.

[1] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter, 2002, 14, 2745.

[2] J. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett., 1996, 77, 3865.

[3] O. Paz and J. M. Soler, Phys. Status Solidi B, 2006, 243, 1080-1094.

[4] J. Tersoff and D. R. Hamann, Phys. Rev. Lett., 1983, 50, 1998-2001.

[5] This choice represents the average electrostatic potential between substrate and tip when the voltage is such that it just includes the electronic state. This is a minor improvement over standard Tersoff-Hamann, which evaluates projected DOS as if the bias is 0.

# Usage Guide

## Choosing a molecule input

The first step is to select a molecule to simulate. There are two ways to do this:

* Choose a molecule from the Chemspider database[1], which holds over 100 million structures. This can be done using the Espeem feature in the app.
* Provide a SMILES code if you are working with a unique molecule not available in the Chemspider database. RD-Kit's pre-optimizer[2] will be used to generate the molecule. Please note that this feature is limited to organic molecules.

After importing the molecule, the app will automatically pre-process it. This includes rotating the molecule to lie as flat as possible on the surface with the principal axis along the x-axis. The app will also resolve any ambiguity due to spatial symmetry using the Pymatgen[3]/Spglib[4] symmetry package.

[1] ChemSpider, Royal Society of Chemistry. <https://chemspider.com>.

[2] RDKit, open-source chem-informatics package, <http://www.rdkit.org>.

[3] Pymatgen, structure manipulation package, <https://pymatgen.org>

[4] Spglib, Atsushi Togo, [https://spglib.readthedocs.io](https://spglib.readthedocs.io/en/latest/)

## Choosing the Simulation Parameters

The app has simplified the complexity of the input, resulting in an interface with only four parameters to set:

1. **Setpoint Current [nA]:** This parameter allows you to explore the effect of lower/higher setpoint current. However, you cannot expect quantitative agreement with your experiment due to limitations in both the algorithm and our knowledge of the precise tip.
2. **Bias level from MidGapLevel to:** This parameter allows you to choose what molecular states to include in the calculation. For example, if you pick LUMO+2, the calculation of the STM image will include both LUMO+0 and LUMO+1 in addition to half the contribution of the LUMO+2 level. Please note that the last level is included only with half the contribution since the voltage is assumed to correspond to the peak value you would observe in STS measurement.
3. **Spatial Broadening [Ang]:** This parameter represents the assumed uncertainty in the tip position. After the current is generated as a function of the tip position, a Gaussian kernel is applied in the lateral directions with this standard deviation. This is a kind of fitting parameter that should be adjusted to fit your experimental setup. Try different values and pick the one that best matches your experimental image/setup. Please note that this method of applying lateral uncertainty is more correct than simply applying a Gaussian to the final image.
4. **Voltage Broadening [V]:** This parameter is the assumed uncertainty in the voltage. A Gaussian kernel is applied to the bias window to soften the transition between the molecular states included and not included in the calculation. When molecular states are close in energy, increasing this parameter reduces the difference in the predicted image. We suggest picking a value close to the minimum width of peaks in your STS spectra.

## Using the results

Once you have set the simulation parameters, you can browse the results in the app. If you have several results on the same molecule, you can browse them in the app using a system with "pins" on the already calculated values. This allows you to switch seamlessly between browsing calculations and requesting new ones. Additionally, you have the option to download the data from the interface. The downloaded zip-file will include:

* The topographic image generated as a '.png'-file.
* A '.png'-image of the molecule with the atoms shown as balls. Note that the scale of this model and that of the topographic image is exactly the same and can easily be used side-by-side or overlaid.
* The raw simulated data as a '.gsf'-file (Gwyddion Simple Field). This format can be opened by the free Gwyddion software along with many data formats for experimental data, allowing for direct comparison with exactly the same color scale.
* A molecule DFT-relaxed version of the molecule as an'.xyz'-file.
* An 'explanation.txt' file which shows an explanation of the theory used along with suggested citations.
* A '.py' which shows the script that was used internally in CalcTroll. This file is mostly provided for transparency into the workings of the interface.
* (Only for ChemSpider) The original input file downloaded from the ChemSpider database as a '.mol'-file.

Overall, using the web-app for simulating STM images of molecules involves selecting a molecule input, pre-processing the molecule, choosing simulation parameters, and then browsing or downloading the results. Give it a try and explore the different features of the app to better understand how it can be used for your research needs.