# LMG 1 - General Methods for Electron Density Prediction in the Condensed Phase and Calculation of Vibrational Sum-Frequency Spectroscopy

<u>Alan Lewis<sup>1</sup></u>, Paolo Lazarroni<sup>1</sup>, Karen Fidanyan<sup>1</sup> and Mariana Rossi<sup>1</sup>

1) Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany

## Machine Learning Electron Densities in the Condensed Phase

#### Introduction

The electron density  $\rho$  is a fundamental quantity of quantum chemistry and physics, which in principle can determine all of the ground state properties of a system. Using density-functional theory (DFT), a wide variety of these properties can be derived directly from the electron density. As a result, obtaining accurate electron densities is central to many applications within computational chemistry, physics and material science. The machine learning method we developed to predicted these densities works for isolated and periodic systems on the same footing, and is able to treat molecular solids, semiconductors and metals to very high accuracy.



# Simulating SFG Spectra without **Molecular Decompositions**

### Introduction

Vibrational sum-frequency-generation (SFG) spectroscopy is a powerful tool used to selectively study nuclear structure and dynamics at interfaces. The signal arising from this non-linear optical process vanishes in centrosymmetric environments, making the technique inherently surface sensitive. The SFG signal is proportional to the resonant part of the second-order susceptibility  $\chi^{(2), \text{res}}_{par}$ 

## A General Method to Predict $\rho$

We introduced the symmetry-adapted learning of threedimensional electron densities (SALTED) method [1]. The electron density  $\rho(\mathbf{r})$  is expanded as a linear combination of numeric atom-centred basis functions  $\phi_{i,\sigma}$ 

$$\rho(\mathbf{r}) \approx \tilde{
ho}(\mathbf{r}) = \sum_{i,\sigma,\mathbf{U}} c_{i,\sigma} \phi_{i,\sigma}(\mathbf{r} - \mathbf{R}_i + \mathbf{T}(\mathbf{U}))$$
(1)

where  $R_i$  are the nuclear positions and T(U) is a unit-cell translation vector. This expansion uses the auxiliary basis functions of the FHI-aims [2] code.

SALTED predicts the coefficients  $c_{i,\sigma}$  ( $\sigma = n\lambda\mu$ ) of this expansion. These are approximated as

$$c_{n\lambda\mu}(A_i) \approx \sum_{j\in M} \sum_{|\mu'|\leq \lambda} b_{n\lambda\mu'}(M_j) \, \mathbf{k}^{\lambda}_{\mu\mu'}(A_i, M_j) \delta_{a_i a_j},$$
 (2)

where  $b_{n\lambda\mu'}(M_j)$  are regression weights and  $k^{\lambda}_{\mu\mu'}(A_i, M_j)$  are covariant kernel functions which measure similarities between

We show the ability to accurately predict *ab initio* quality densities of large systems using inexpensive training data, avoiding the need for expensive self-consistent calculations. We predicted densities on ice cells containing up to 512 molecules using a SALTED model trained on cells containing just 4 molecules. The errors introduced to the electron density are a very small fraction of the total electron density, as illustrated in Figure 4.

 $\Delta 
ho(oldsymbol{r}) \left[ e/{
m \AA}^3 
ight]$ 

$$-10^{-3} - 10^{-4} - 10^{-5} - 10^{-6} 0 10^{-6} 10^{-5} 10^{-4} 10^{-3}$$



Figure 4: A twodimensional cut of the predicted electron density of a 64-molecule ice supercell (lower *slice and colorbar*) and of the error in the density with respect to the reference DFT calculation (upper slice and colorbar.)

$$\chi_{pqr}^{(2),\text{res}}(\omega) = i \int_{0}^{+\infty} dt \ e^{i\omega t} \ \langle \alpha_{pq}(t) \mu_{r}(0) \rangle_{T},$$

where p, q, r are Cartesian directions,  $\alpha_{pq}$  is a component of the polarizability tensor and  $\mu_r$  a component of the dipole.

Current simulation protocols rely on a molecular decomposition of lpha and  $\mu$  to exclude noise arising from the bulk, and to separate the signals arising from different interfaces inside the periodic simulation box.[5] This molecular decomposition is arbitrary, and not easily transferable to solids or systems with delocalised protons. We propose a method which overcomes this practical challenge.

#### **Real-Space Method and Results**



Figure 6: Schematic illustration of the real-space domain of integration.

The real-space approach developed here limits the domain of the real-space integrals used to calculate the dipole moment and polarizability in the FHI-aims code [2] to the interfacial region, shown schematically in Figure 6. The surface which defines this region,  $Z_{CN}(x, y; t)$ , is obtained by finding a constrained instantaneous atomic density surface close to a chosen plane perpendicular to the interface, [6] which is then adjusted such that the domain of integration is charge neutral. This surface (illustrated in Figure 7) fluctuates over the course of a molecular dynamics simulation without significant

atomic environments  $A_i$  and  $M_i$ . The regression weights are given by

> $\boldsymbol{b}_{M} = (\boldsymbol{K}_{NM}^{T} \boldsymbol{S}_{NN} \boldsymbol{K}_{NM} + \eta \boldsymbol{K}_{MM})^{-1} \boldsymbol{K}_{NM}^{T} \boldsymbol{w}_{N},$ (3)

where  $S_{NN}$  is the overlap matrix of the expansion basis functions, and  $w_N$  contains the projections of the training densities on these basis functions.  $K_{NM}$  contains the kernels, which we evaluate using the  $\lambda$ -SOAP formalism.[3] We collaborated with A. Grisafi and M. Ceriotti (EPFL) in this work.

#### Validation

- Validation on three prototypical systems: a metal (AI), a semiconductor (Si), and a molecular solid (ice).
- Accuracy was assessed by:
  - Direct comparison between predicted electron densities and those obtained self-consistently from DFT: RMSE of less than 2% for each system (Figure 1).
  - Indirect evaluation of energetic properties derived from the predicted electron density: Errors of tens of meV (Figure 2).



Figure 1: Learning curves for each of the test datasets. For each point, the percentroot mean age square error is averaged across 10 randomly selected validation sets, each containing 20

errors

of 20

 $ho(oldsymbol{r}) \left[ e/{
m \AA}^3 
ight]$ 

 $10^{-1}$ 

The exchange-correlation and electrostatic energies indirectly calculated from these predicted electron densities are found to be *more* accurate than those obtained from comparable direct machine learning models (Table 1).

Molecules	$\overline{\epsilon}_{xc}^{ML}(I)$	$\overline{\epsilon}_{xc}^{ML}(D)$	$\overline{\epsilon}_{el}^{ML}(I)$	$\bar{\epsilon}_{el}^{ML}(D)$
64	1.57	2.25	2.90	8.19
128	1.29	3.21	1.80	8.82
256	1.66	3.67	1.41	9.63
512	1.82	3.60	1.09	9.51

Table 1: The mean absolute errors in the exchange-correlation and electrostatic energies ( $\bar{\epsilon}_{xc}^{\text{ML}}$  and  $\bar{\epsilon}_{el}^{\text{ML}}$ ) from the predicted electron densities (the indirect errors, I) and from a direct prediction using Gaussian process regression (D) for each size of ice supercell These errors are relative to the DFT reference values. All energies are in meV per atom

## Water at Charged Interfaces

#### **Outlook: Dissociation with Applied Fields**

We study the process of water dissociation on a Pd (111) slab under applied electric fields. We investigate the interplay between electric-field induced surface charge and nuclear quantum effects on the dissociation process. Calculations based on a nudged-elastic-band implementation in the i-PI code [4] and harmonic phonon calculations.

eV) 1.4 no ZPE with ZPE 1.3

Figure 7: An illustra-

of the

surface

tive snapshot

instantaneous

 $Z_{CN}(x, y; t).$ 

#### Preliminary Results

spectral artefacts.

As a test, we calculated the IR absorption spectrum of the watervacuum interface (Figure 8). As expected, the intensity of the main O-H stretching peak is diminished when the dipoles are only evaluated in the interfacial region, but the intensity of the high frequency shoulder ( $\sim 3800 \text{ cm}^{-1}$ ) corresponding to O-H stretching modes into the vacuum is unchanged.



Figure 8: The absorption spectra of a 24molecule water slab, using the dipole moment calculated from the full slab and the surface regions alone. The high-frequency shoulder present at equal intensity in both spectra indicates O-H stretches into the vacuum.



of water monomer dissociation on Pd(111)  $4 \times 4$ surface (PBE+vdW<sup>surf</sup>, 7 layers) under different applied voltages, inferred from the electric field strength. Black circles show the energy without the ZPE contribution, red squares include

ZPE.

Figure 5: Energy barrier

#### Voltage on Pd (111) surface (V)

The monomer adsorption site switches from a planar "atop" site to a "bridge" site with hydrogen atoms pointing towards the surface between -2 and -3 V. Zero-point energy (ZPE) decreases the dissociation barrier; this effect is more pronounced for the planar atop configuration.

Calculations of polarizabilities in an analogous manner in order to simulate SFG spectra are under way.

#### References

[1] A. M. Lewis *et al.*, J. Chem. Theory Comput. **17**, 7203 (2021). [2] V. Blum et al., Comp. Phys. Comm. 180, 2175 (2009). [3] A. Grisafi et al., Phys. Rev. Lett. 120, 36002 (2018). [4] V. Kapil et al., Comp. Phys. Comm. 236, 214 (2019). [5] F. Tang et al., Chem. Rev. 120, 3633 (2020). [6] A. P. Willard and D. Chandler, J. of Phys. Chem. B **114**, 1954 (2010).







