Computational Chemistry and Machine Learning

PFAS identification and transition metal nitride oxidation

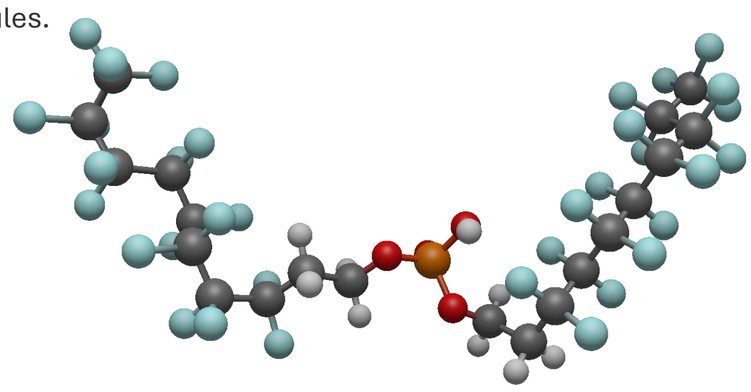
Daniel Barron | Supervised by Dr. Michael Nolan
Materials Modelling for Devices | MicroNano Systems

Tyndall

PFAS Spectroscopy and Machine Learning

Background

- PFAS molecules or "forever chemicals" are persistent environmental contaminants with links to cancer, immune disruption, and ecosystem damage.
- 68 PFAS molecules were computed at DFT level to generate IR and ^{19}F NMR spectra.
- Goal: use ML to identify PFAS from spectra and predict spectra of unseen molecules.



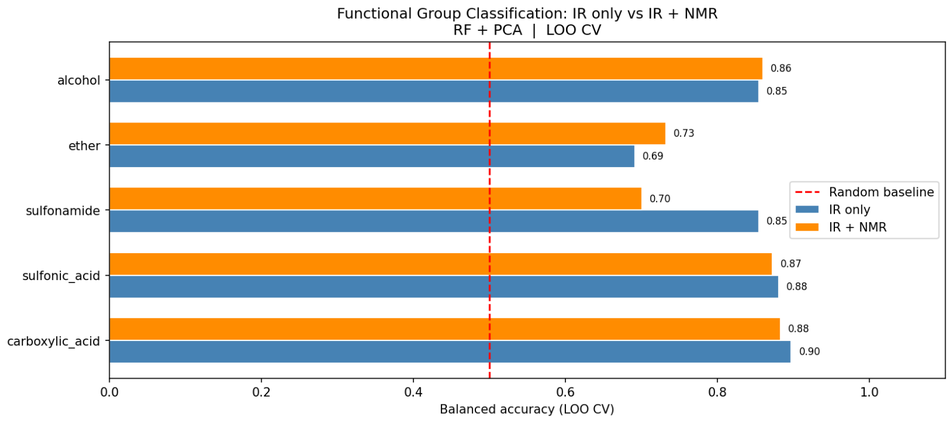
8:2 Fluorotelomer phosphate diester

ML Pipeline (Leave-one-out CV, n = 68)

- Functional group classification: Random Forest + Principal Component Analysis (PCA).
- Spectrum prediction: Ridge Regression + PCA from RDKit descriptors.
- Cross-modal: ^{19}F NMR → IR (r = 0.837) stronger than IR → ^{19}F NMR (r = 0.645).

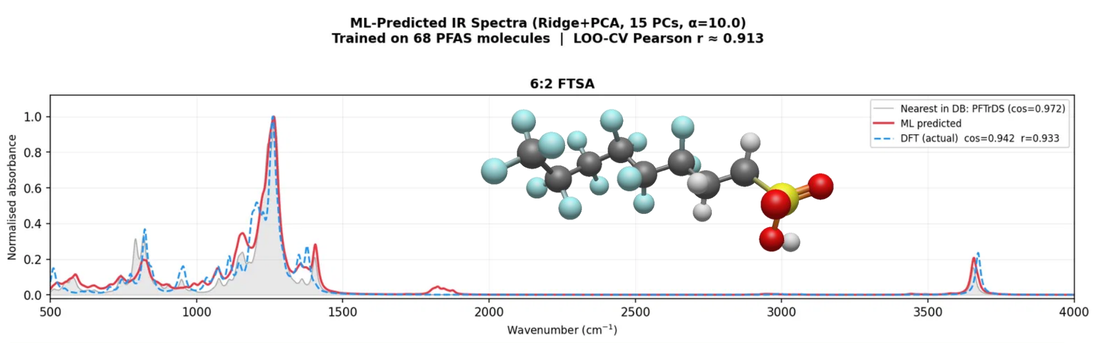
Functional Group Classification

- IR-only Random Forest classifier achieves F1 ≥ 0.85 for carboxylic acid, sulfonic acid, sulfonamide, and alcohol.
- Ether linkages are hardest to classify (F1 = 0.69), weakest IR signature.
- Adding ^{19}F NMR does not consistently improve classification, confirming IR alone is highly discriminative.



Spectrum Prediction

- Full IR spectra predicted from 18 RDKit molecular descriptors.
- Best case: PFTrDS r = 0.980 | Hardest: 2,2,3,3-TFPA r = 0.589.



3 unseen molecules validated: 6:2 FTSA (r = 0.933), DONA 3H interior (r = 0.950), 6:2 FTAB protonated (r = 0.798).

Future Work

- Expand dataset beyond 68 molecules and improve model generalisability.
- Apply model to experimental IR spectra for real-world PFAS identification.
- Combine IR + NMR in a unified system to resolve IR-confusable PFAS pairs.

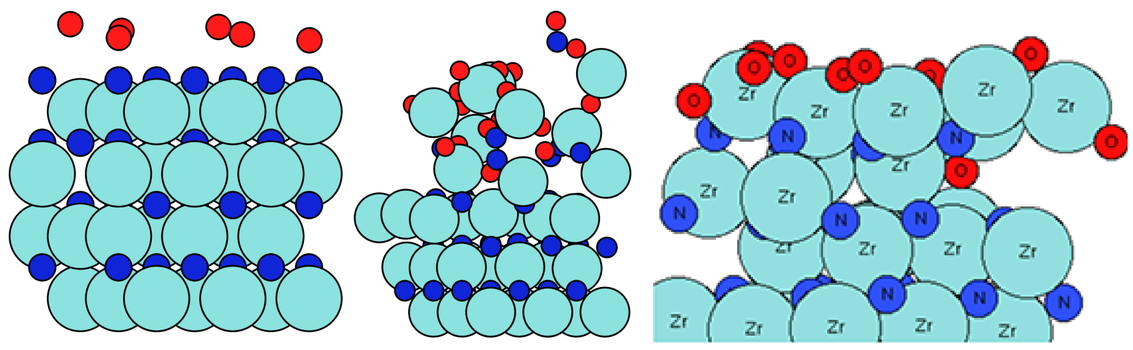
Transition Metal Nitride Oxidation using Matlantis

Background

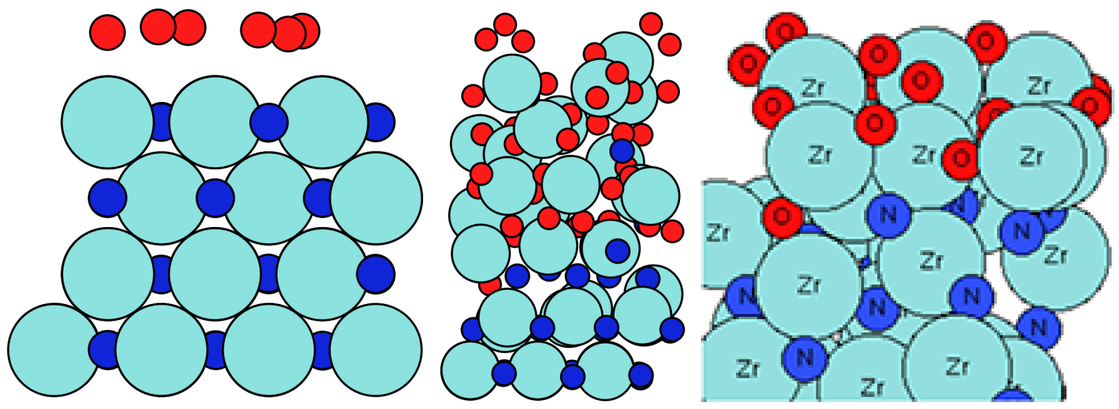
- Liu et al. (J. Phys. Chem. C 2025, 129, 11173) performed ab initio MD (VASP/PBE) on ZrN and VN surface oxidation, limited to ~50 ps by DFT cost.
- Key finding: ZrN oxidises more readily than VN; ZrN forms stable ZrO_2 layer, VN forms only thin VO_x clusters.
- Goal: extend to longer timescales using Matlantis' universal ML potential.

Oxidation Structures

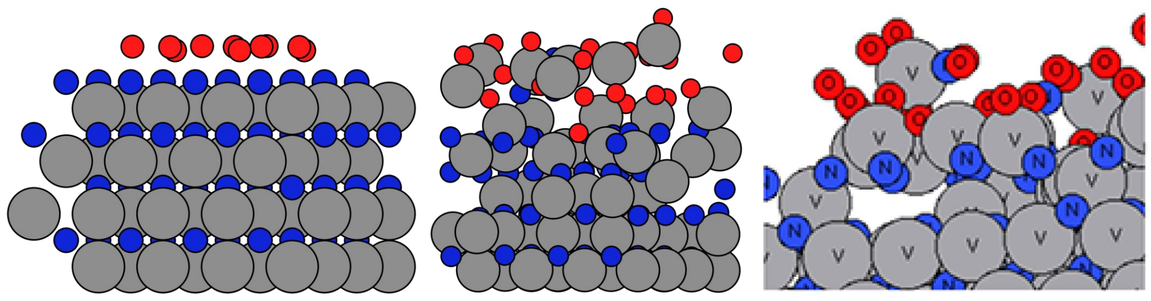
ZrN (100):



ZrN (110):



VN (111):



*Before (left) and after (middle) **800 ps** oxidation simulation at 363 K. Shorter timescale DFT simulations for same system are shown on the right for comparison. Simulations were also done at 295 K and higher temperatures.*

Future Work

- Finetune system-specific foundational potential on ZrN/VN DFT data and compare against Matlantis potential results.
- Extend to longer timescales to study more oxidation dynamics.

Shared Conclusions

- Both projects demonstrate that machine learning extends the reach of DFT, in timescale and in prediction power.
- Neither approach replaces DFT, each uses it as high-quality training data for faster, scalable simulation.
- PFAS result: DFT IR spectra were computed for over 70 PFAS molecules; ^{19}F NMR and IR probe orthogonal chemical information.
- TMN result: 800 ps Matlantis simulations reveal oxidation dynamics missed at 50 ps DFT timescales.

References:
- *Ji Liu et al., J. Phys. Chem. C 2025, 129, 11173–11182.*
- *ORCA 5 (Neese, WIREs Comput. Mol. Sci. 2022) .*
- *Matlantis (Takamoto et al., Nat. Comput. Sci. 2022).*

Rialtas na hÉireann
Government of Ireland

UCC
Coláiste na hOllscoile Corcaigh
University College Cork, Ireland