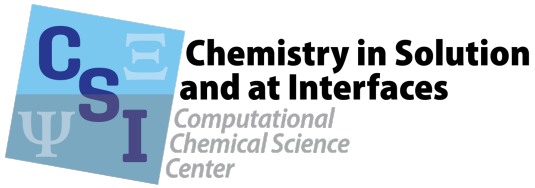
Bottom-up Ab-initio Multiscale Modeling of Materials with Machine Learning

Roberto Car

Princeton University

6th AI for Materials Science Workshop, July 9-10, 2025, NIST, MD (USA)



Suspended since 3/31/25

DPMD: potential energy, polarization, and polarizability surfaces are represented by DNNs trained on DFT data → AIMD on steroids

$$\Phi(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3, \dots, \boldsymbol{R}_N) = \sum_i \Phi_i, \qquad \Phi_i = \mathcal{F}(\{\boldsymbol{R}_k \in \mathcal{N}_i\})$$

$$V\boldsymbol{\mathcal{P}}(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3, \dots, \boldsymbol{R}_N) = \sum_i z_i \boldsymbol{R}_i - 2\sum_w \boldsymbol{R}_w + C = \sum_c \boldsymbol{d}_c + C, \qquad \boldsymbol{R}_w = \boldsymbol{\mathcal{F}}(\{\boldsymbol{R}_k \in \mathcal{N}_w\})$$

$$\boldsymbol{\alpha}(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3, \dots, \boldsymbol{R}_N) = V\left(\frac{\partial \boldsymbol{\mathcal{P}}}{\partial \boldsymbol{\mathcal{E}}}\right)_{\boldsymbol{R}^N} = -2\sum_{\mathcal{W}} \frac{\partial \boldsymbol{R}_{\mathcal{W}}}{\partial \boldsymbol{\mathcal{E}}} = \sum_{\mathcal{W}} \boldsymbol{\alpha}_{\mathcal{W}}(\{\boldsymbol{R}_k \in \mathcal{N}_{\mathcal{W}}\})$$

DPLR includes long-range electrostatics with Ewald sum on top of DP

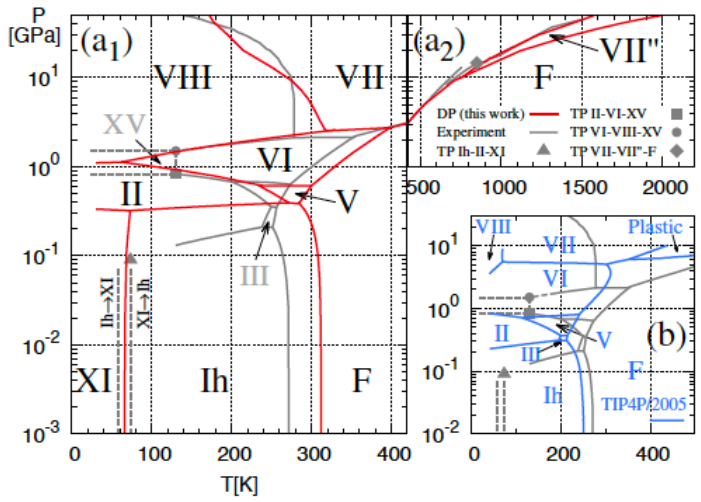
Coupling with an external electric field (static or dynamic)

$$\Phi_{\boldsymbol{\mathcal{E}}(t)} = \Phi_{\boldsymbol{\mathcal{E}}=\boldsymbol{0}} - V\boldsymbol{\mathcal{P}} \cdot \boldsymbol{\mathcal{E}}(t)$$

L. Zhang *et al*, *Phys Rev Lett* **120**, 143001 (2018); L. Zhang et *al*, in *Advances in Neural Information Processing Systems* **31**, 4441 (2018); L. Zhang et *al*, *Phys Rev Materials* **3**, 023804 (2019); L. Zhang et *al*, Phys *Rev B* **102**, 041121(R) (2020); G. Sommers et *al*, PCCP **22**, 10592 (2020); L. Zhang et *al*, *J Chem Phys* **156**, 124107 (2022); J. Zheng *et al, "DeePMD kit v2…", J Chem Phys* **159**, 054801 (2023)

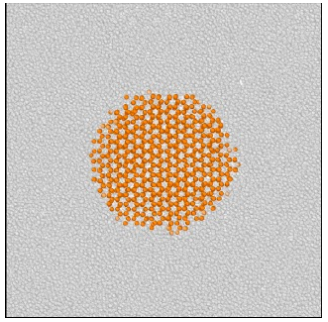
DPMD made possible AIMD studies with unprecedented size and time scales

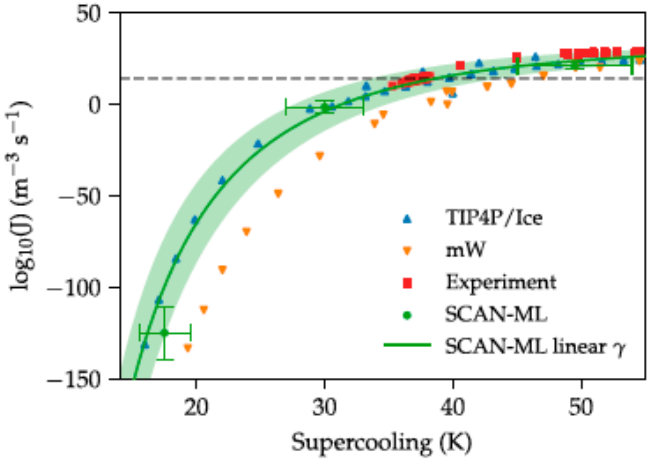
Water phase diagram



L. Zhang, H. Wang, R.C., and W.E,
Phys Rev Lett **126**, 236001 (2021)

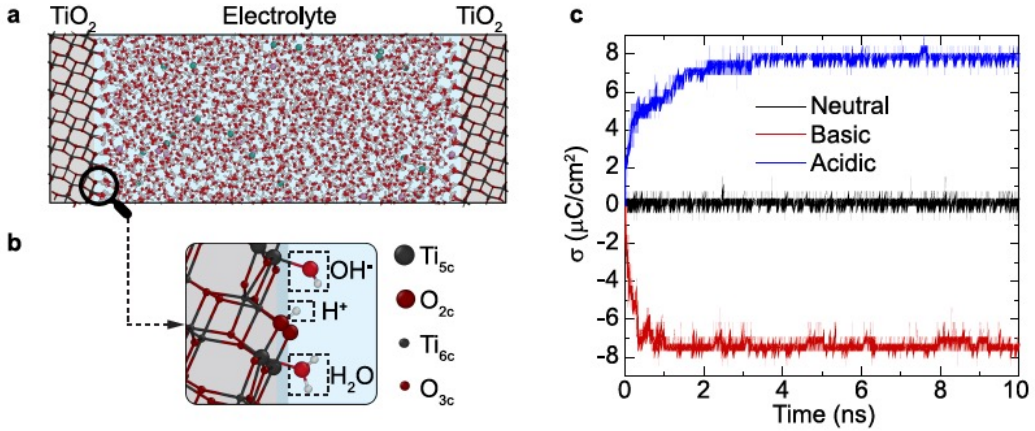
Homogeneous nucleation of ice





P. Piaggi, J. Weis, A. Panagiotopoulos, P. Debenedetti, RC, *PNAS* (2022)

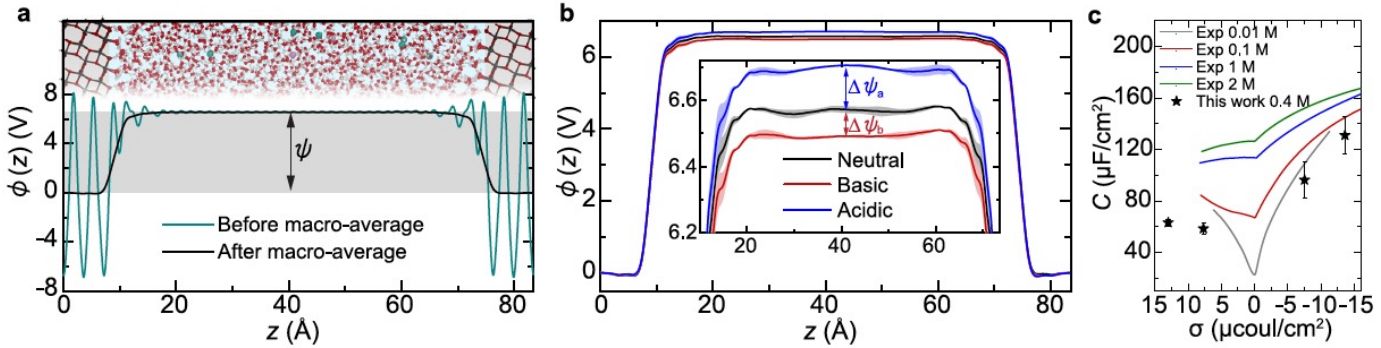
Electrolyte solutions in contact with TiO_2



0.4 M NaCl + 0.2 M HCl

O.4 M NaCl

0.4 M NaCl + 0.2 M NaOH

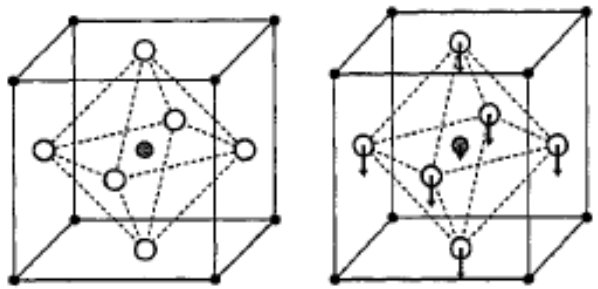


Differential Capacitance

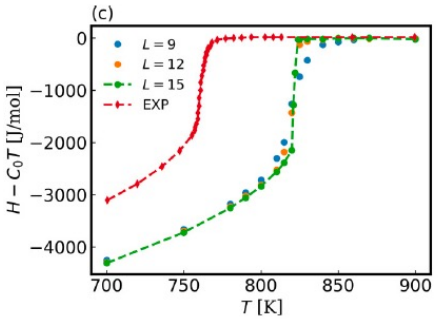
$$C = \frac{d\sigma}{d\psi} \approx \frac{\Delta\sigma}{\Delta\psi}$$

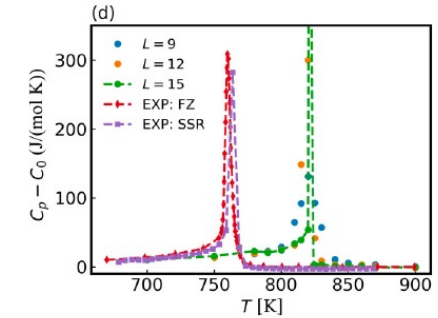
C Zhang, MC Andrade, ZK Goldsmith, AS Raman, Y Li, P Piaggi, X Wu, RC, A Selloni, *Nature Comm* (2024)

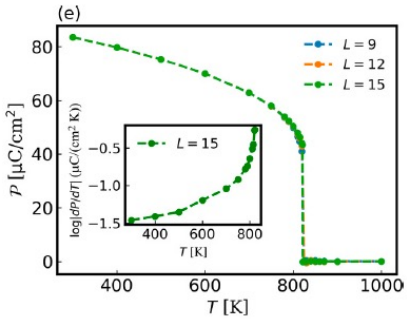
The ferroelectric phase transition in $PbTiO_3$ (PTO)

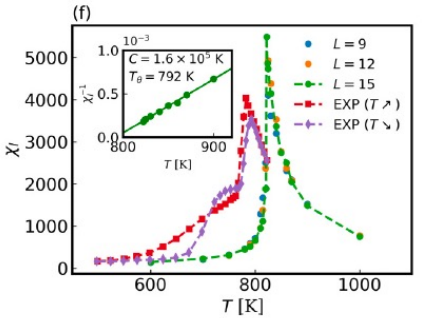


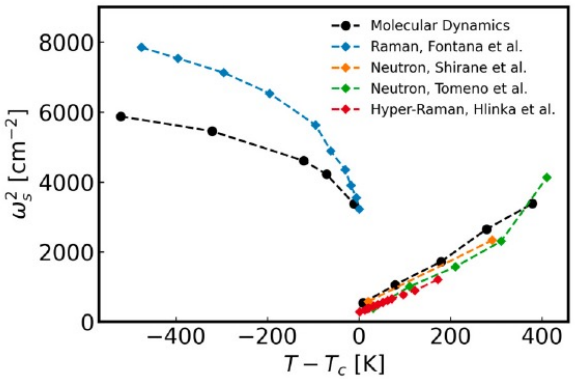
$T>T_c$ $T<T_c$

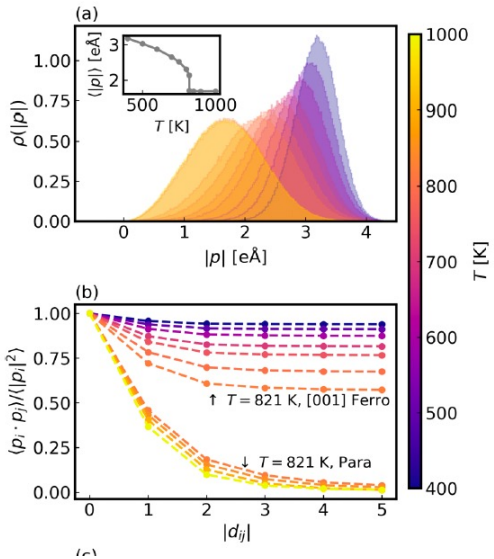








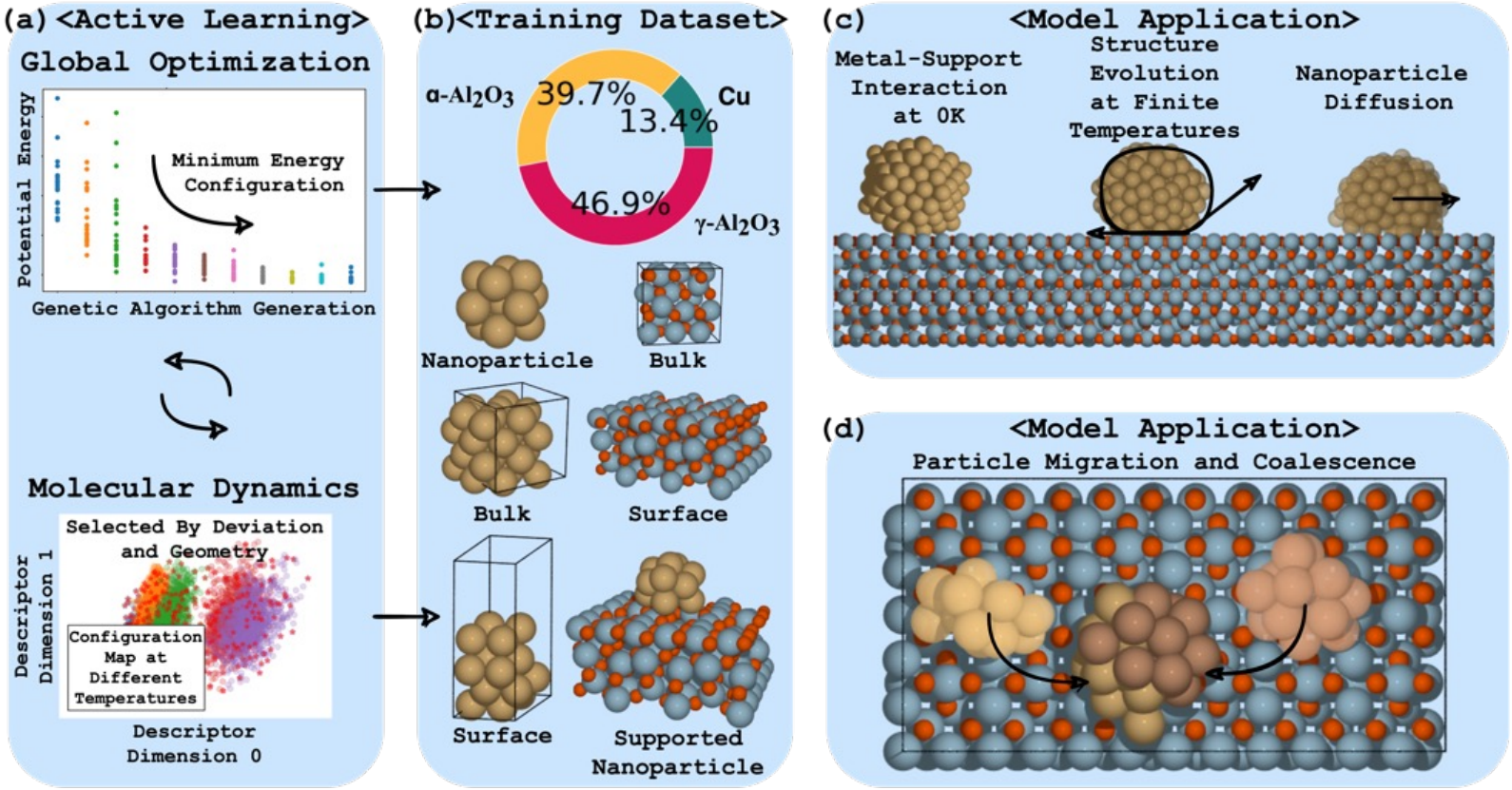




The dominant order-disorder character of the transition

P. Xie, Y. Chen, W. E, R. C., PRB **111**, 094113 (2025)

Cu nanoparticles on Al_2O_3 surfaces

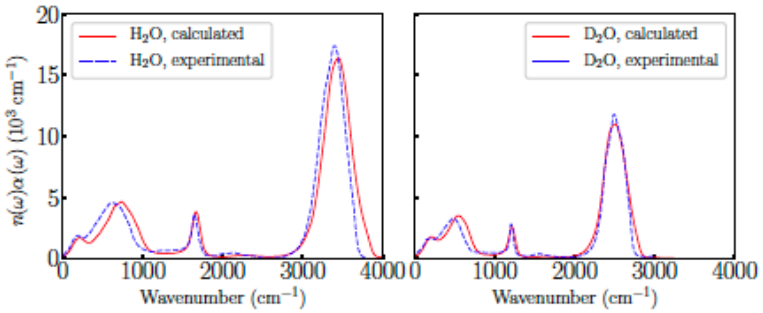


The DP model accuracy: E_{rmse} of 0.004 eV/atom, F_{rmse} of 0.057 eV/Å upon training on ~150k structures representing Cu_n on γ-Al_2O_3(100), γ-Al_2O_3(110), and α-Al_2O_3(0001).

J. Xu, S. Das, A.D. Pathak, A. Patra, S. Shetty, D. Hohl, R.C., arXiv (2025)

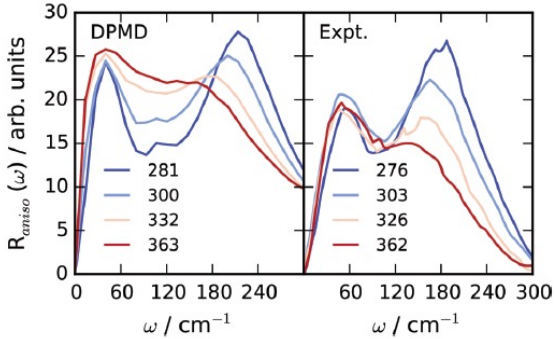
Vibrational spectroscopies (Infrared, Raman, SFG)

Infrared absorption of liquid water

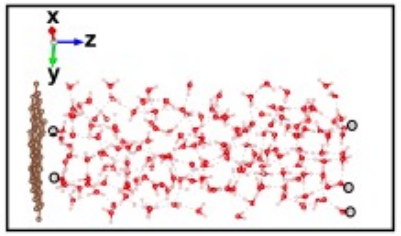


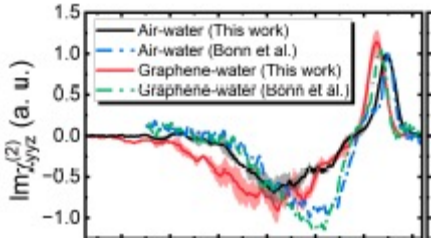
L. Zhang et al., PRB (2020)

Raman spectra of H-bonding modes in liquid water



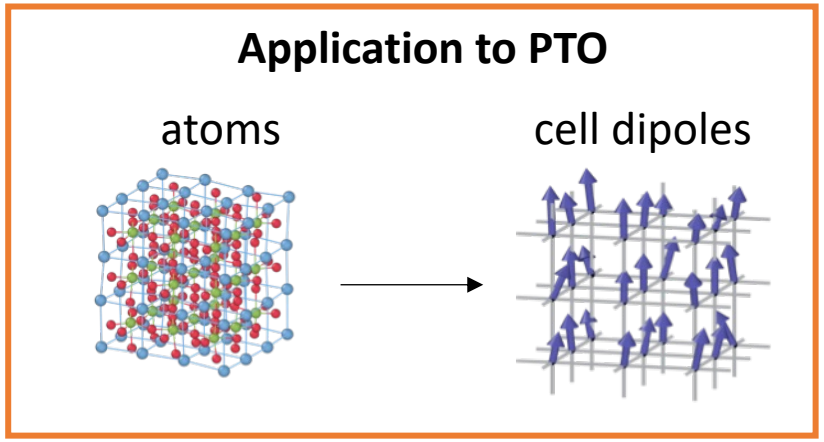
G. Sommers et al., PCCP (2020)





Y. Wang et al., arXiv (2025)

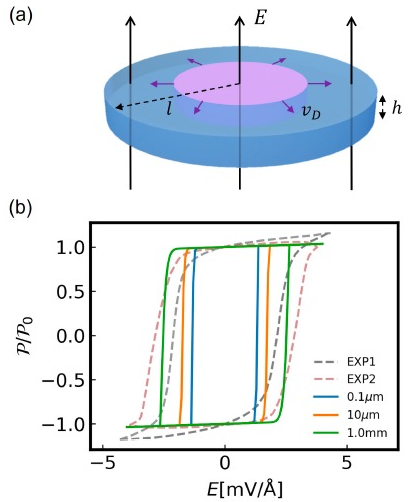
From all-atom to collective variable dynamics: Ab-initio Generalized Langevin Equation (AIGLE)



$$M\ddot{X}(t) = -\nabla G(X) + F(t) + M\int_0^t dsK(s)\dot{X}(t-s) + R(t)$$

Effective Hamiltonian model or free energy model learned from DPMD

ML is used to learn colored noise; 2FDT is assumed



Crude 1D model of hysteresis loop

Curvature of domain wall as well as edge, thickness (*h*), and quenched disorder (O vacancies) effects are neglected.

Loops are calculated with AIGLE for different radii *l*, with switching field period $t_P = 100\ \mu s$, and remnant polarization $P_0 = 72\ \mu C/cm^2$

EXP1: $t_P = 100\ \mu s$, unknown *l*, $P_0 = 94\ \mu C/cm^2$

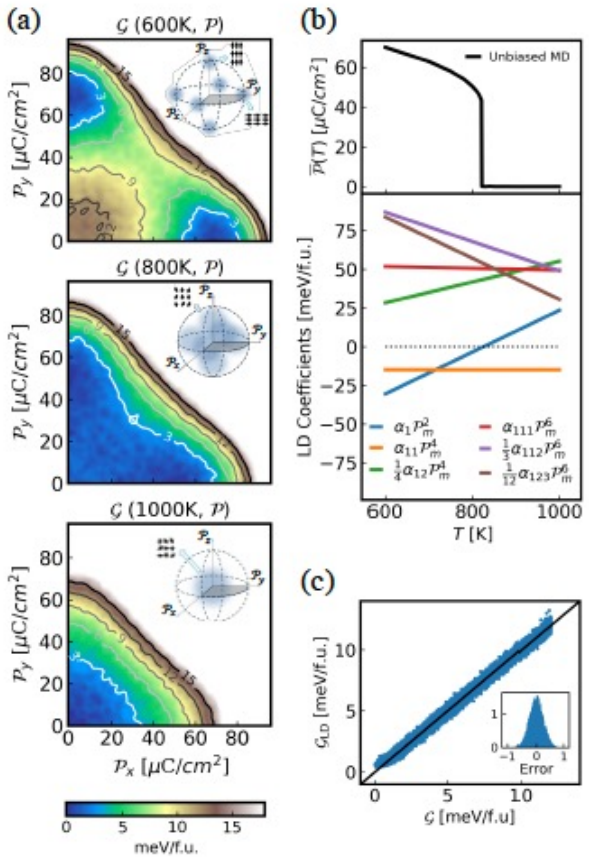
EXP2: unknown t_P , $l \approx 144\ \mu m$, $P_0 = 96\ \mu C/cm^2$

From P. Xie, RC, W. E, *PNAS* **121**, e23086668121 (2024)

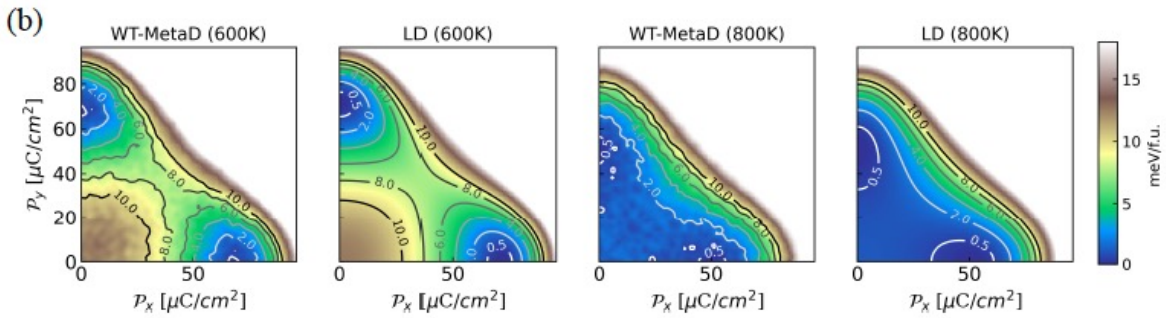
From atoms to fields (Ginzburg-Landau models)

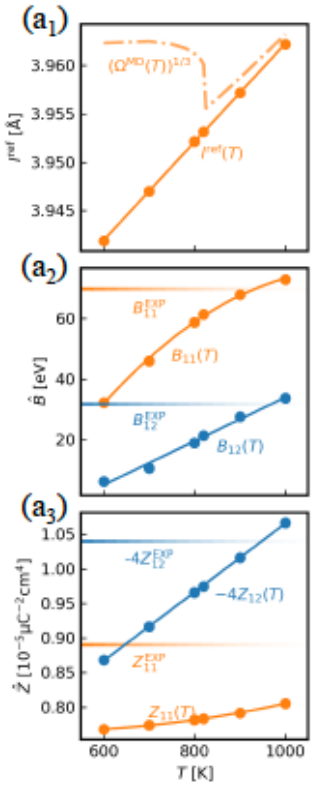
$$\mathcal{F}(T,\boldsymbol{\mathcal{P}},\boldsymbol{\eta}) = \mathcal{G}(T,\boldsymbol{\mathcal{P}}) + \frac{N}{2}\left(\boldsymbol{\eta}^T - \boldsymbol{Q}^T\hat{Z}^T\right)\hat{B}\left(\boldsymbol{\eta} - \hat{Z}\boldsymbol{Q}\right) \qquad \boldsymbol{Q}^T \equiv \left(\mathcal{P}_x^2, \mathcal{P}_y^2, \mathcal{P}_z^2\right)$$

$\mathcal{G}(T,\boldsymbol{\mathcal{P}})$ is obtained with WT-MetaD and the coupling matrices $\hat{B}$ and $\hat{Z}$ are obtained with linear regression



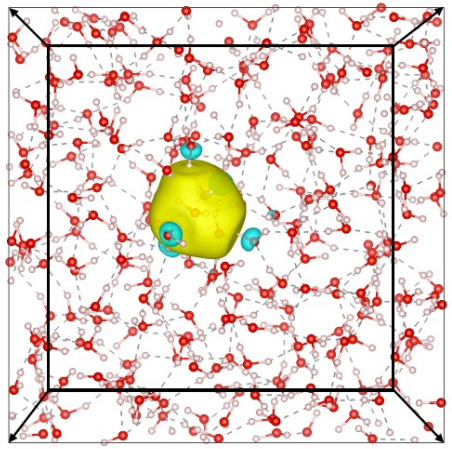
$\mathcal{G}(T,\boldsymbol{\mathcal{P}})$ is fitted (denoised) extremely well by a 6th order Landau-Devonshire phenomenological model





(P. Xie, Y. Chen, X. Xu, W. E, RC, submitted (2025))

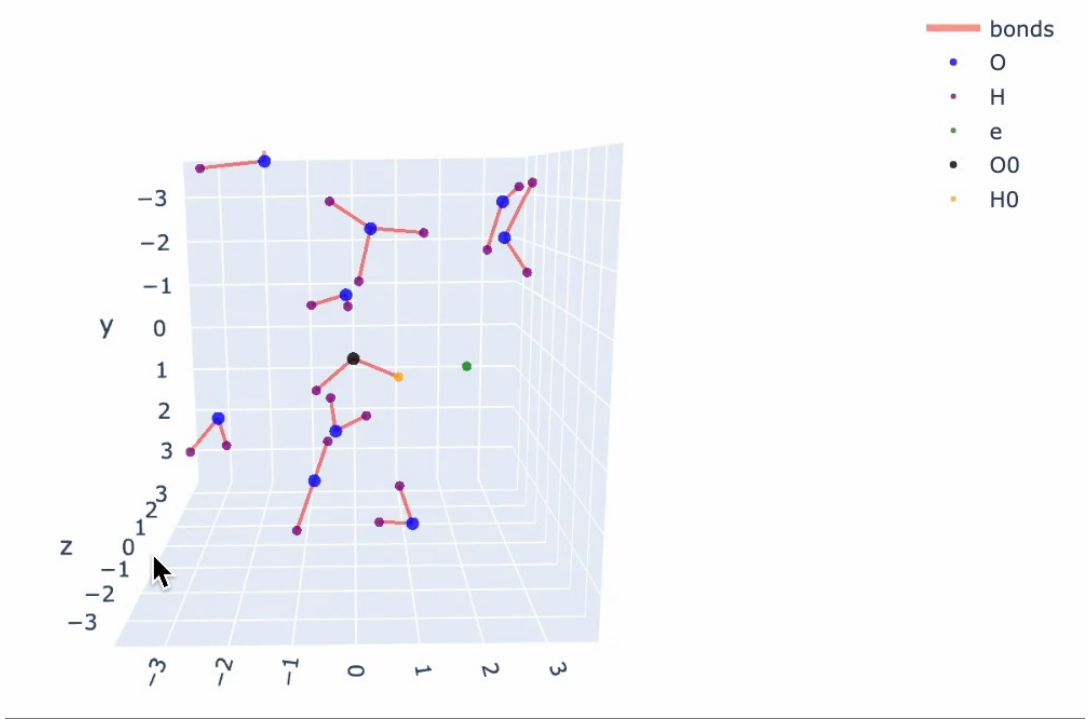
The solvated electron: a challenge for ML coarse-graining



R. Gao, Y. Li, R.C., *PCCP* **26**, 23080 (2024)

Learning the reaction $e^- + H_3O^+ \rightarrow H_2O + H^{\cdot}$ is even more challenging!

R. Gao et al., in preparation (2025)



Concluding Remarks

- ML and DNN representations boost the power of AIMD, opening the way for ab-initio multiscale models in which the parameters are derived non-empirically from quantum mechanics
- It is essential to construct ML models that are solidly based on physics
- Multiscale ab-initio models will help to understand complex processes and to design materials and devices
- New physics may emerge from the simulations but that requires human intelligence as opposed to AI
- Accuracy of the models will improve with more accurate quantum mechanical reference models and transfer learning techniques