XACS
Xiamen Atomistic Computing Suite
XACScloud.com

MLatom.com

Accelerating and improving quantum chemistry and dynamics with artificial intelligence

Pavlo O. Dral

dral@xmu.edu.cn

Xiamen University, P. R. China

28th June 2023

Group & Acknowledgements

Hiring post-docs, PhD & MSc students!

Funding

PCOSS

国家自然科学基金委员会
National Natural Science Foundation of China

嘉庚创新实验室
TAN KAH KEE INNOVATION LABORATORY

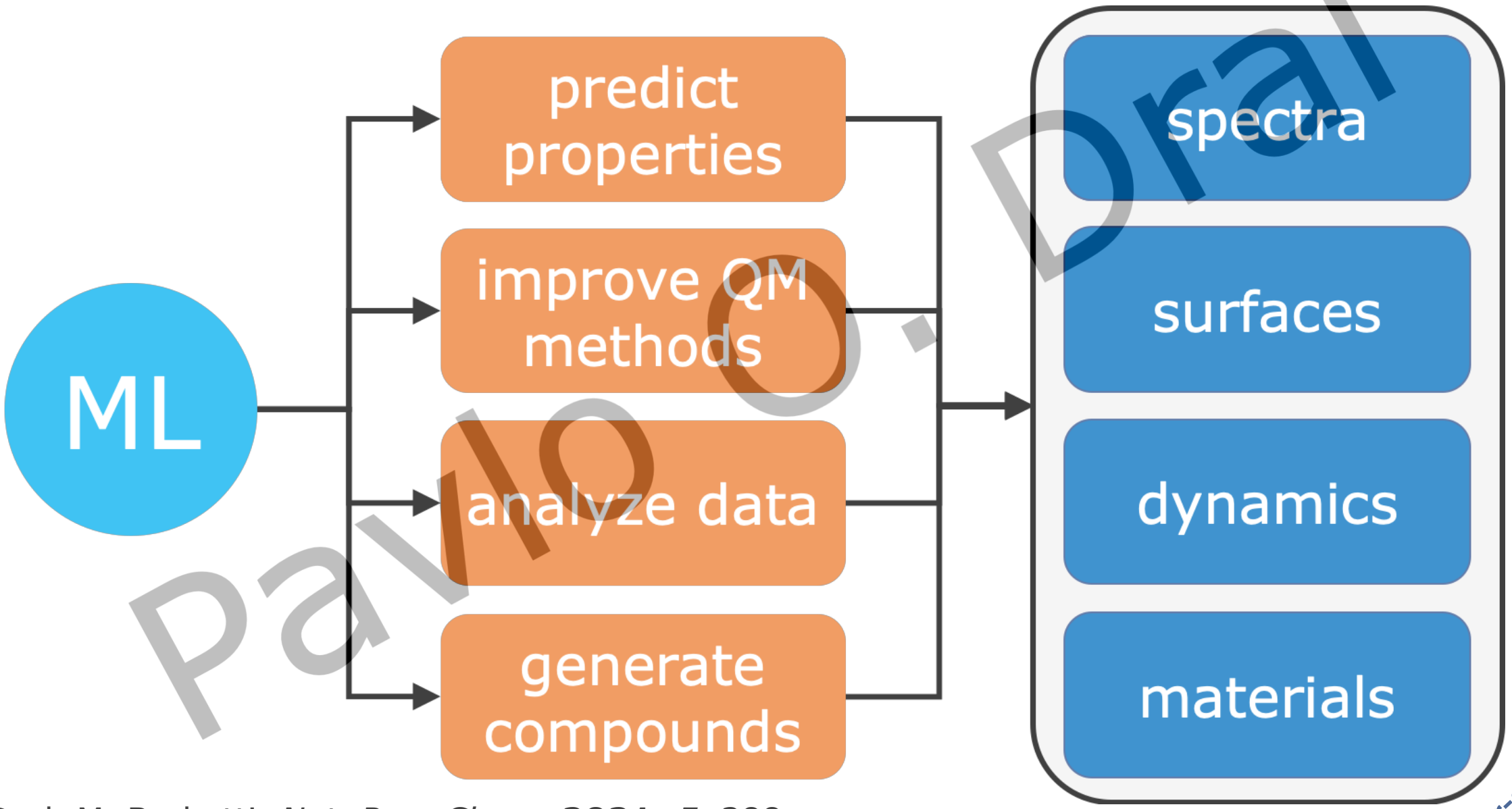
Xiamen University, Siming campus

Air Photo: CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=27584741

16 subjects rank top 1% globally. 11th in Mainland China.

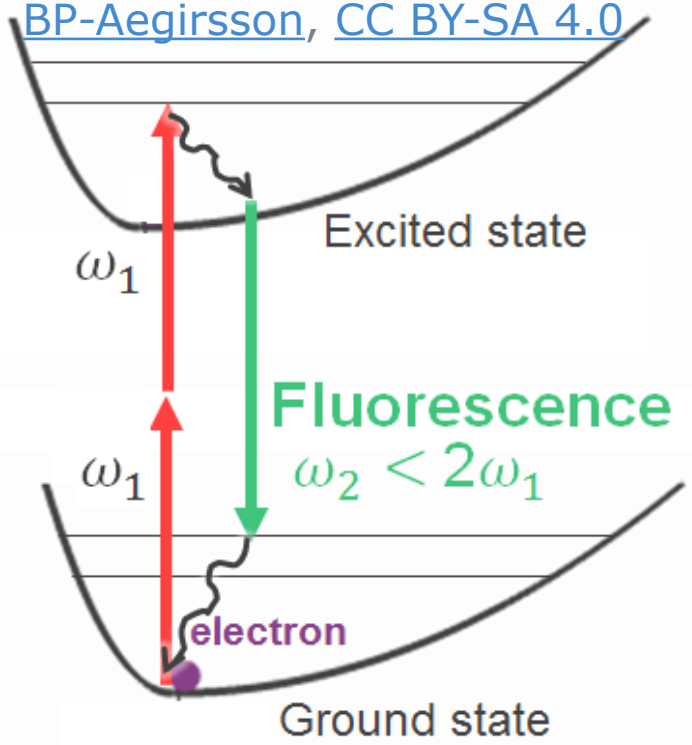
Chemistry ranks top 1‰ globally (*ESI as of March 2019).*

Machine learning in chemistry



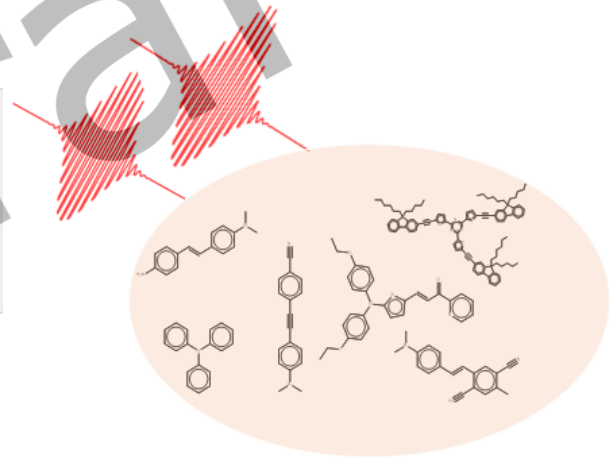
P. O. Dral, M. Barbatti, *Nat. Rev. Chem.* **2021,** *5*, 388

Two-photon absorption



Here we show how to calculate TPA cross section for RHODAMINE 6G and RHODAMINE 123 molecules with MLatom input file `mltpa.inp`:

```
MLTPA
SMILESfile=Smiles.csv
auxfile=_aux.txt
```



This input requires `Smiles.csv` file with SMILES of molecules:

```
CCNC1=CC2=C(C=C1C)C(=C3C=C(C(=[NH+]CC)C=C3O2)C)C4=CC=CC=C4C(=O)OCC.[Cl-]
COC(=O)C1=CC=CC=C1C2=C3C=CC(=N)C=C3OC4=C2C=CC(=C4)N.Cl
```

Two-photon absorption applications:

- two-photon lithography
- Photodynamic therapy
- Bioimaging
- 3D printing
- Upconverted laser

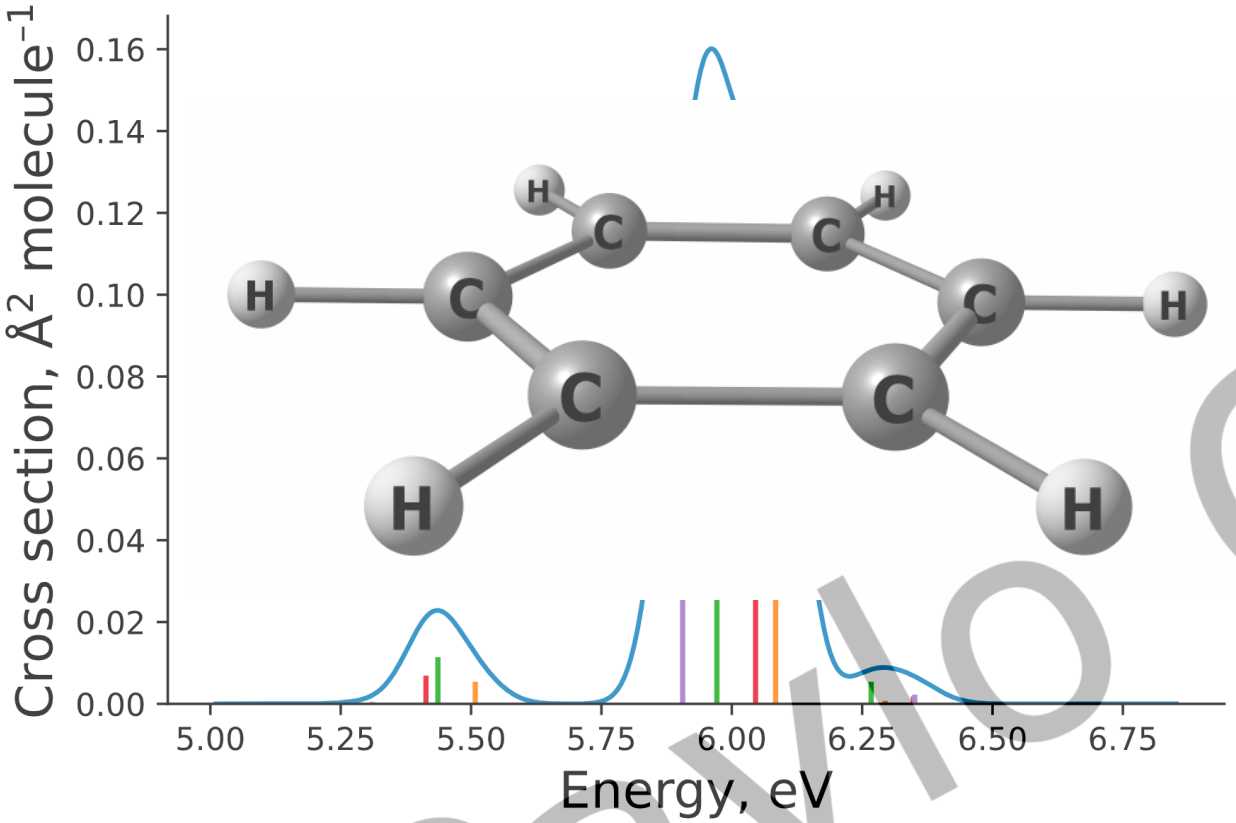
Yuming Su

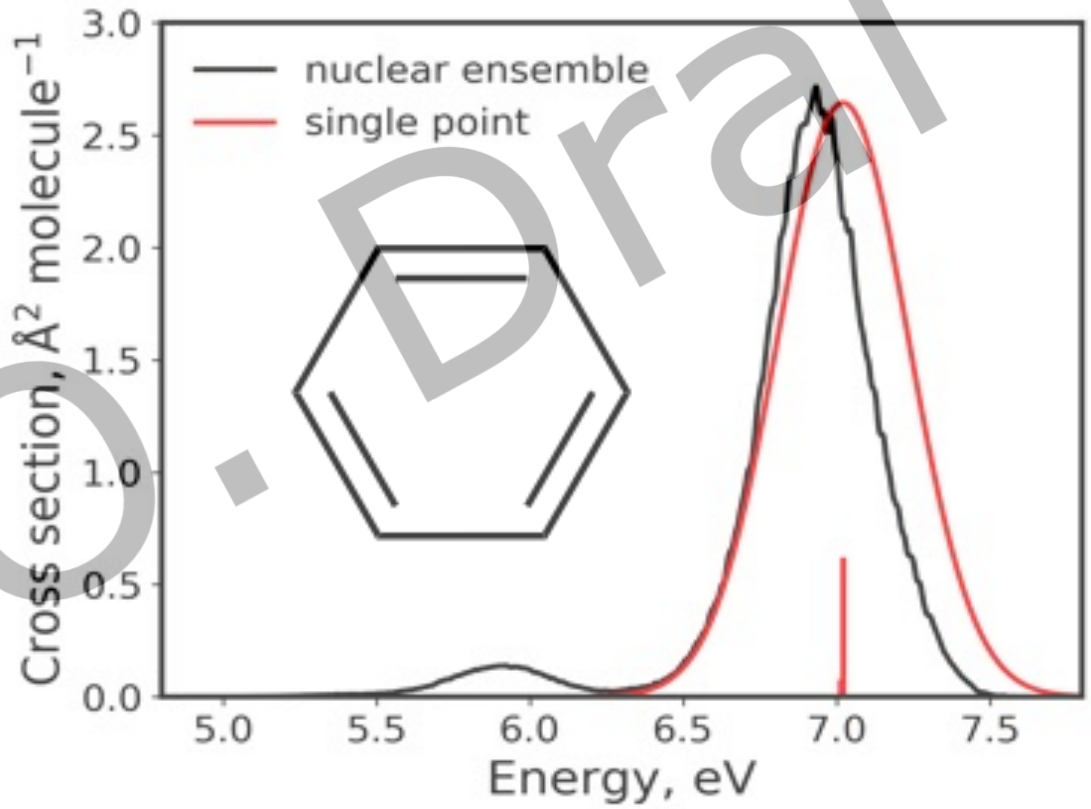
Zhou Da

Cheng Wang

Y. Su, Y. Dai, Y. Zeng, …, P. Zheng, D. Zhou, P. O. Dral, C. Wang. *Adv. Sci.* **2023**, 2204902

Machine learning UV/vis absorption spectra





Nuclear Ensemble Approach (NEA) calculates cross section by averaging over multiple normalized broadening functions at different conformations (marked by different colors).

Compared with the single point convolution, NEA correctly predicts absorption intensity for forbidden transitions (example: benzene)

$$\sigma(E) = \frac{\pi e^2 h}{2mc\varepsilon_0 E} \sum_n^{N_{fs}} \frac{1}{N_p} \sum_i^{N_p} \Delta E_{0n}(\mathbf{x}_i) f_{0n}(\mathbf{x}_i) \frac{1}{\sqrt{2\pi(\delta/2)^2}} \exp\left(-\frac{(E-\Delta E_{0n})^2}{2(\delta/2)^2}\right)$$

R. Crespo-Otero, M. Barbatti, *Theor. Chem. Acc.* **2012,** *131*, 1237

Quantum chemistry vs Machine learning

Conventional programming in quantum chemistry:

→ Code for molecular orbitals

→ Code for excitation energies

→ Code for oscillator strengths

→ …

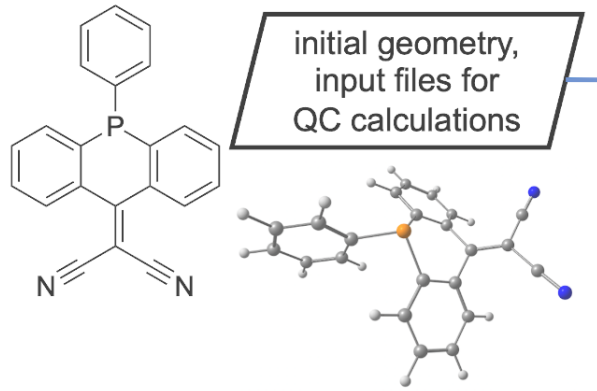
Machine learning (in principle – adaptations required!):

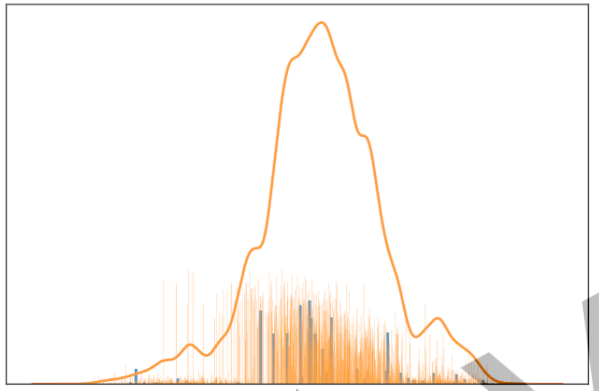
→ The same code for all above

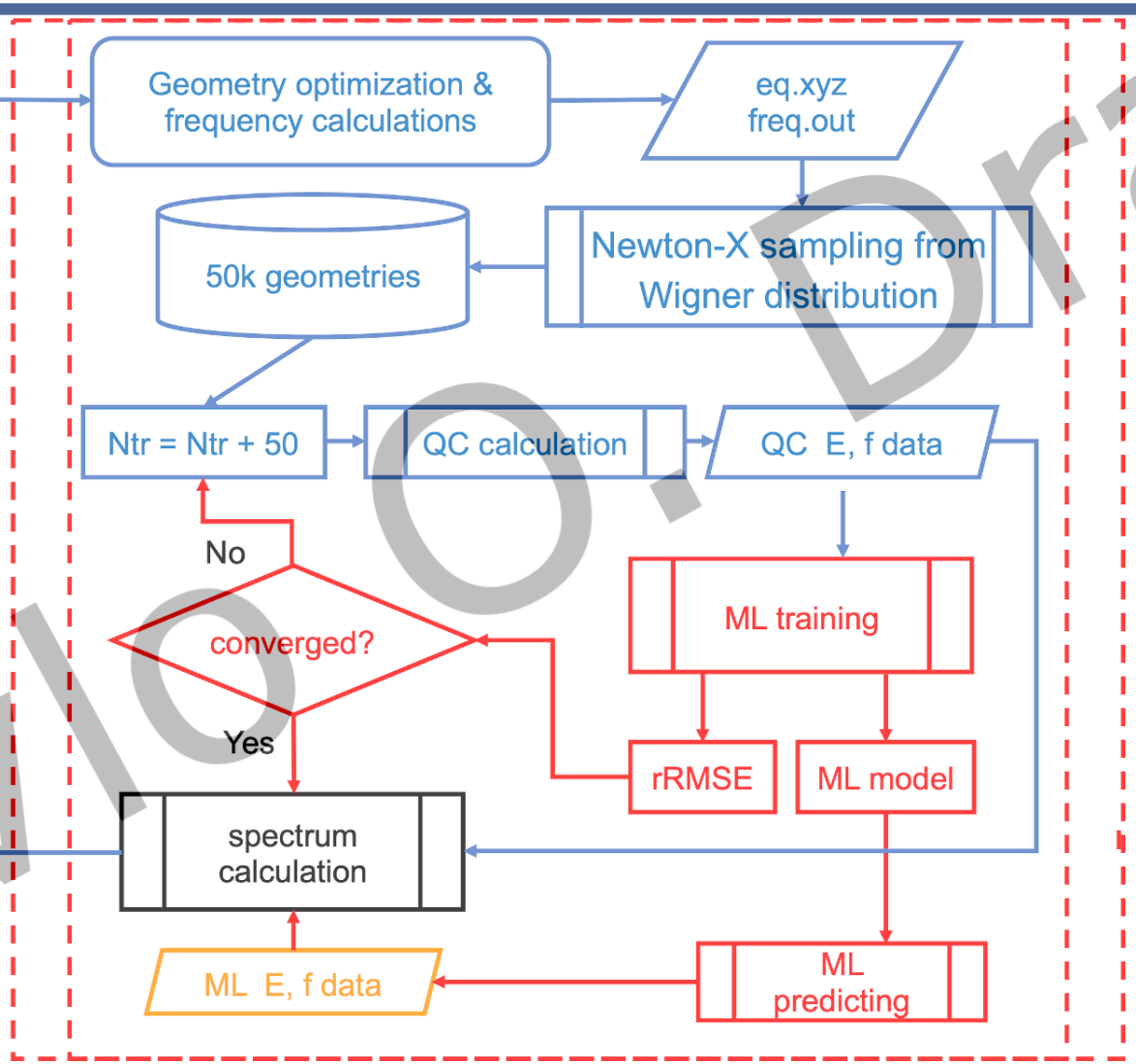
↑ MOs ↑ excitation energies ↑ Oscillator strength

Training data

Machine learning single-photon absorption spectra







Mario Barbatti

Bao-Xin Xue

ML-NEA method: B.-X. Xue, P. O. Dral, M. Barbatti, *J. Phys. Chem. A* **2020**, *124*, 7199–7210
Implementation in MLatom: P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou, M. Pinheiro Jr, J. Huang, M. Barbatti, *Top. Curr. Chem.*, ***2021***, *379*, 27

MLatom: Platform for atomistic machine learning

Simulations

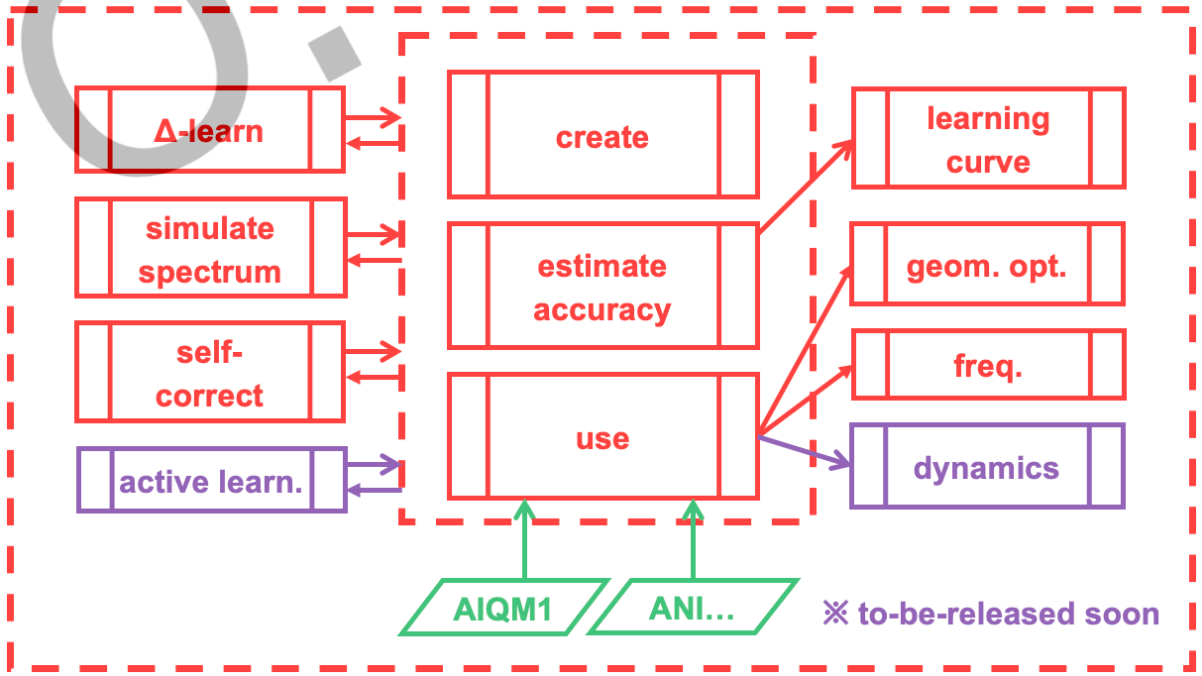
- single-point calculations
- geometry optimizations (minima and transition states, IRC)
- frequencies & thermochemistry
- UV/vis spectra (ML-NEA)
- simulations with pre-trained models (AIQM1, ANI-1ccx, etc.)
- simulations with user-trained models
- NEW! two-photon absorption cross sections (ML-TPA)
- Only available on MLatom@XACS cloud:
  - quantum dynamics with machine learning
- Coming soon:
  - molecular dynamics
  - IR spectra

Learning

- training popular ML models (KREG, ANI, sGDML, PhysNet, DPMD, GAP-SOAP, KRR-CM)
- training generic ML models (kernel ridge regression with many kernels)
- optimizing hyperparameters
- evaluating ML models (also with learning curves)
- Δ-learning
- self-correction

Data

- converting XYZ coordinates to molecular descriptor (RE, Coulomb matrix, ...)
- analyzing data sets
- sampling (random, structure-based, farthest-point) and splitting datasets



P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou, M. Pinheiro Jr, J. Huang, M. Barbatti, *Top. Curr. Chem.* **2021**, *379*, 27

XACS

Xiamen Atomistic Computing Suite

XMVB **XEDA** **MLatom**

Cloud computing (free!)

XACScloud.com

- *Ab initio* valence bond calculations (**VBSCF**, **VBCI**, **BOVB**, …)
- Generalized Kohn–Sham energy decomposition analysis (**GKS-EDA**)
- Artificial intelligence-enhanced quantum mechanical method 1 (**AIQM1**, faster and more accurate than B3LYP)
- Fast geometry optimization, MD, thermochemistry
- … and much more

Principal investigators (Xiamen University)

- Wei Wu
- Peifeng Su
- Pavlo O. Dral

Partners

- Mario Barbatti, Aix Marseille University
- Benoît Braïda, Sorbonne Université
- Philippe Hiberty, University of Paris-Saclay
- Olexandr Isayev, Carnegie Mellon University
- Yirong Mo, UNC Greensboro
- Sason Shaik, Hebrew University
- Avital Shurki, Hebrew University
- Cheng Wang, Xiamen University

Interfaces

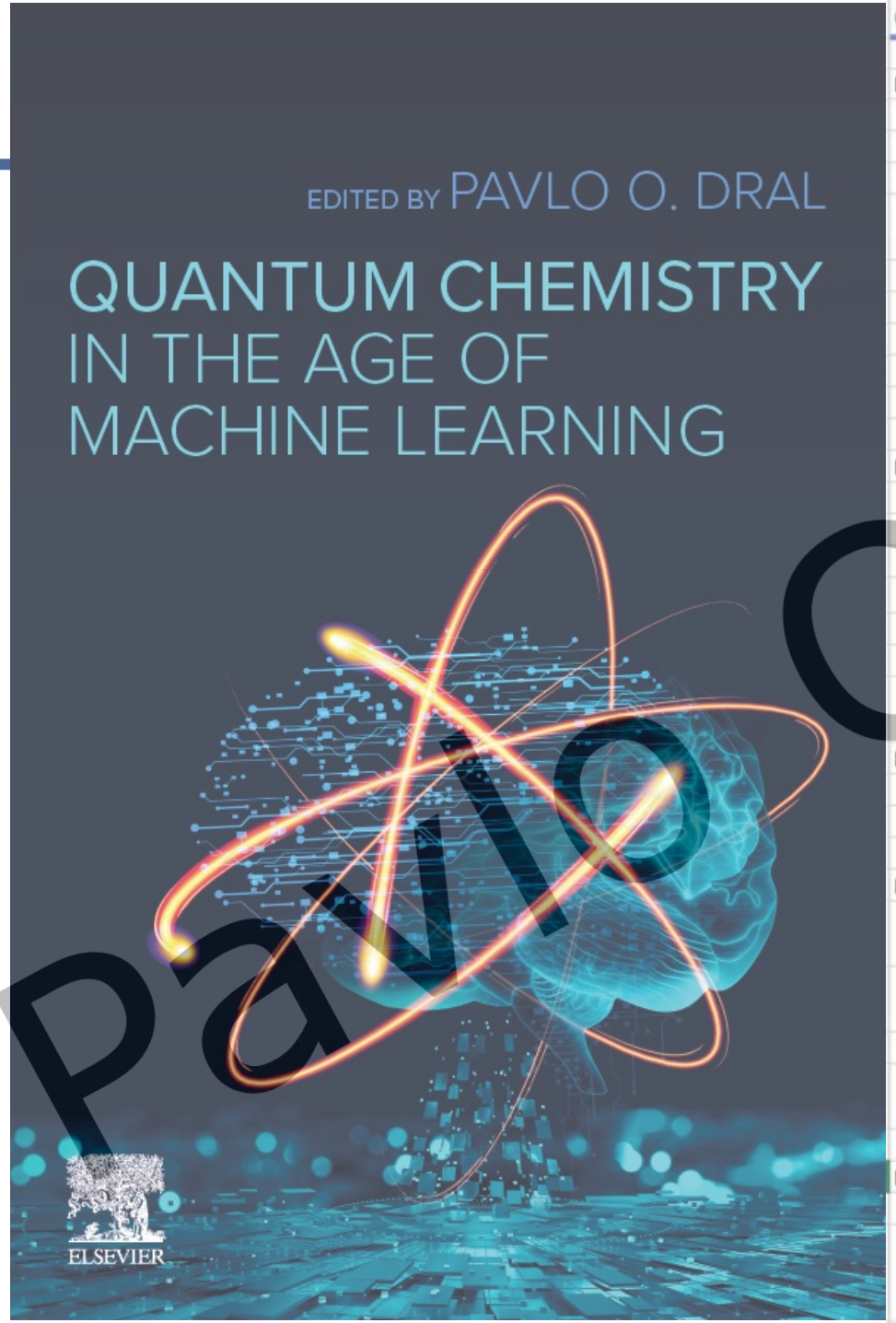
TORCHANI SPARROW

ASE hyperopt

XACS
Xiamen Atomistic Computing Suite
XACScloud.com

MLatom.com

Published in September 2022



Chapter	Title
	Preface
Part 1	**Introduction**
1	Very brief introduction to quantum chemistry
2	Density functional theory
3	Semiempirical quantum mechanical methods
4	From small molecules to solid-state materials: A brief discourse on an example of carbon compounds
5	Basics of dynamics
6	Machine learning: An overview
7	Unsupervised learning
8	Neural networks
9	Kernel methods
10	Bayesian inference
Part 2	**Machine learning potentials**
11	Potentials based on linear models
12	Neural network potentials
13	Kernel method potentials
14	Constructing machine learning potentials with active learning
15	Excited-state dynamics with machine learning
16	Machine learning for vibrational spectroscopy
17	Molecular structure optimizations with Gaussian process regression
Part 3	**Machine learning of quantum chemical properties**
18	Learning electron densities
19	Learning dipole moments and polarizabilities
20	Learning excited-state properties
Part 4	**Machine learning-improved quantum chemical methods**
21	Learning from multiple quantum chemical methods: Δ-learning, transfer learning, co-kriging, and beyond
22	Data-driven acceleration of coupled-cluster and perturbation theory methods
23	Redesigning density functional theory with machine learning
24	Improving semiempirical quantum mechanical methods with machine learning
25	Machine learning wavefunction
Part 5	**Analysis of Big Data**
26	Analysis of nonadiabatic molecular dynamics trajectories
27	Design of organic materials with tailored optical properties: Predicting quantum-chemical polarizabilities and derived quantities

27 chapters
65 authors!

Example on MLatom@XACS cloud

Case study

This case study uses the KREG model to fit the PES of the hydrogen molecule and then uses this KMP to get the optimized bond length of H_2. All the data, input files, and instructions are also described at https://github.com/dralgroup/MLinQCbook22-KMP. The calculations will be done with the MLatom package [11,22,23] for atomistic machine learning simulations using the data set with 451 Cartesian (XYZ) geometries of H_2 and full configuration interaction (FCI) energies calculated with the aug-cc-pV6Z basis set. This data set is taken from Ref. [32].

The case study consists of two simple steps.

Step 1. Training the KREG potential

To train the KREG potential, we need both geometries and reference energies, which come with the data set. In addition, we need to choose the initial guess for the equilibrium geometry to obtain $r^{\mathrm{eq}}_{a,b\neq a}$ terms in the RE descriptor as defined by Eq. (6). Here we choose a geometry with H—H bond length of 0.8 Å as an initial guess of the equilibrium structure saved in the `eq.xyz` file:

```
2
H 0.000000 0.000000 0.000000
H 0.000000 0.000000 0.800000
```

After that we can train the KREG model on the entire data set with MLatom using the following input file (the comments are given after "#" symbol):

Book chapter 13:

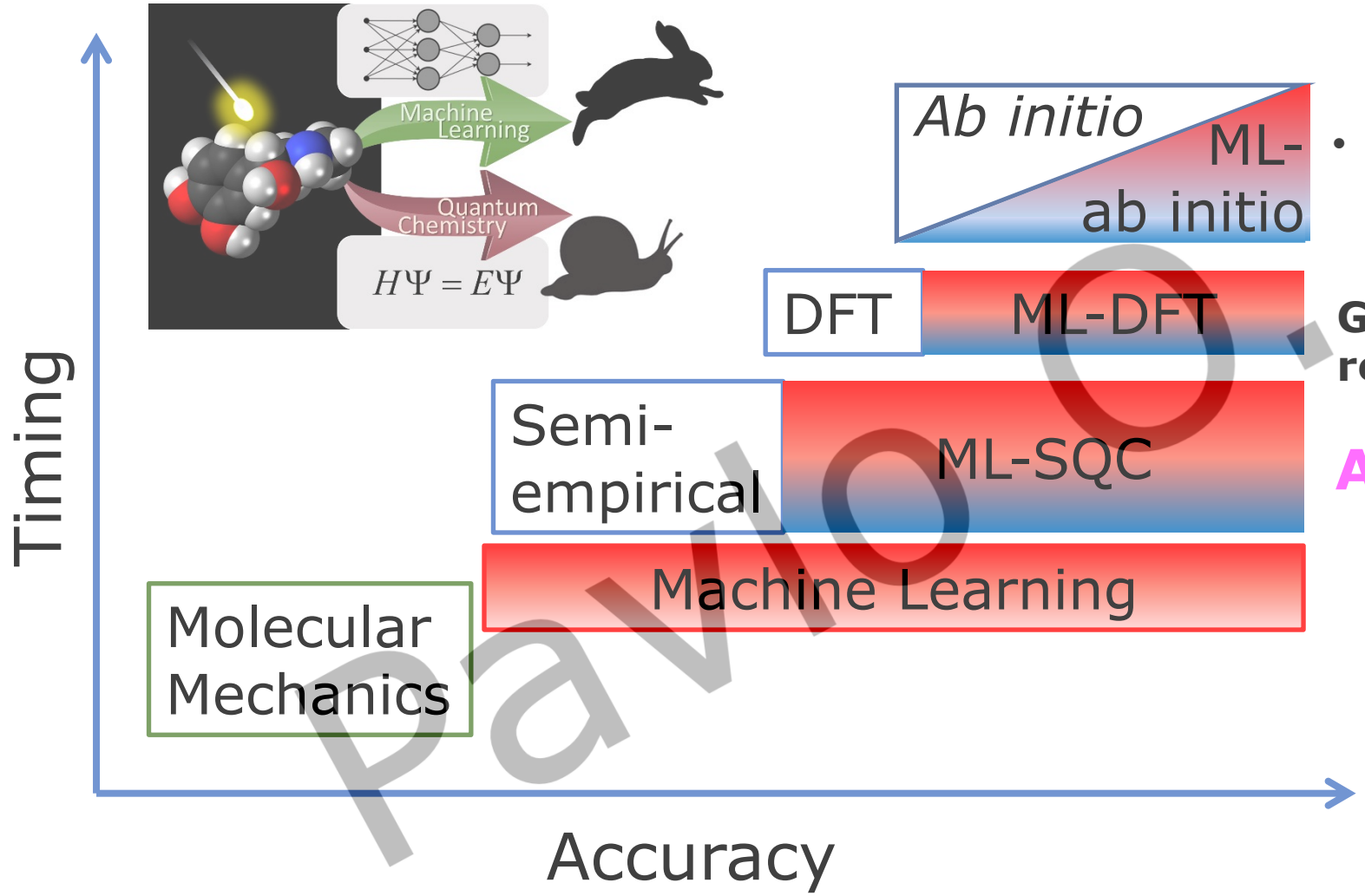
Yi-Fan Hou and Pavlo O. Dral. Kernel method potentials.

In Quantum Chemistry in the Age of Machine Learning,
Pavlo O. Dral, Ed. Elsevier: 2023.

DOI: 10.1016/B978-0-323-90049-2.00020-2.

Machine learning in quantum chemistry

P. O. Dral, M. Barbatti, *Nat. Rev. Chem.* **2021,** *5*, 388



Concepts:

- Delta-learning
  R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld. *JCTC* **2015**
- Learning Hamiltonian
  P. O. Dral, O. A. von Lilienfeld, W. Thiel. *JCTC* **2015,** *11*, 2120

General-purpose (no training required):

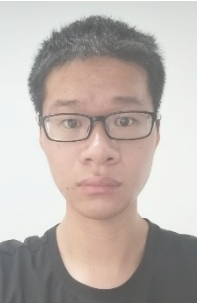
AIQM1 AIQM1: P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

P. O. Dral, *J. Phys. Chem. Lett.* **2020**, *11*, 2336

Accuracy vs transferability vs cost

Wei Wu

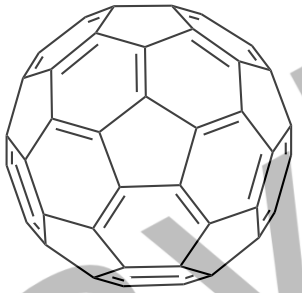
Olexandr Isayev



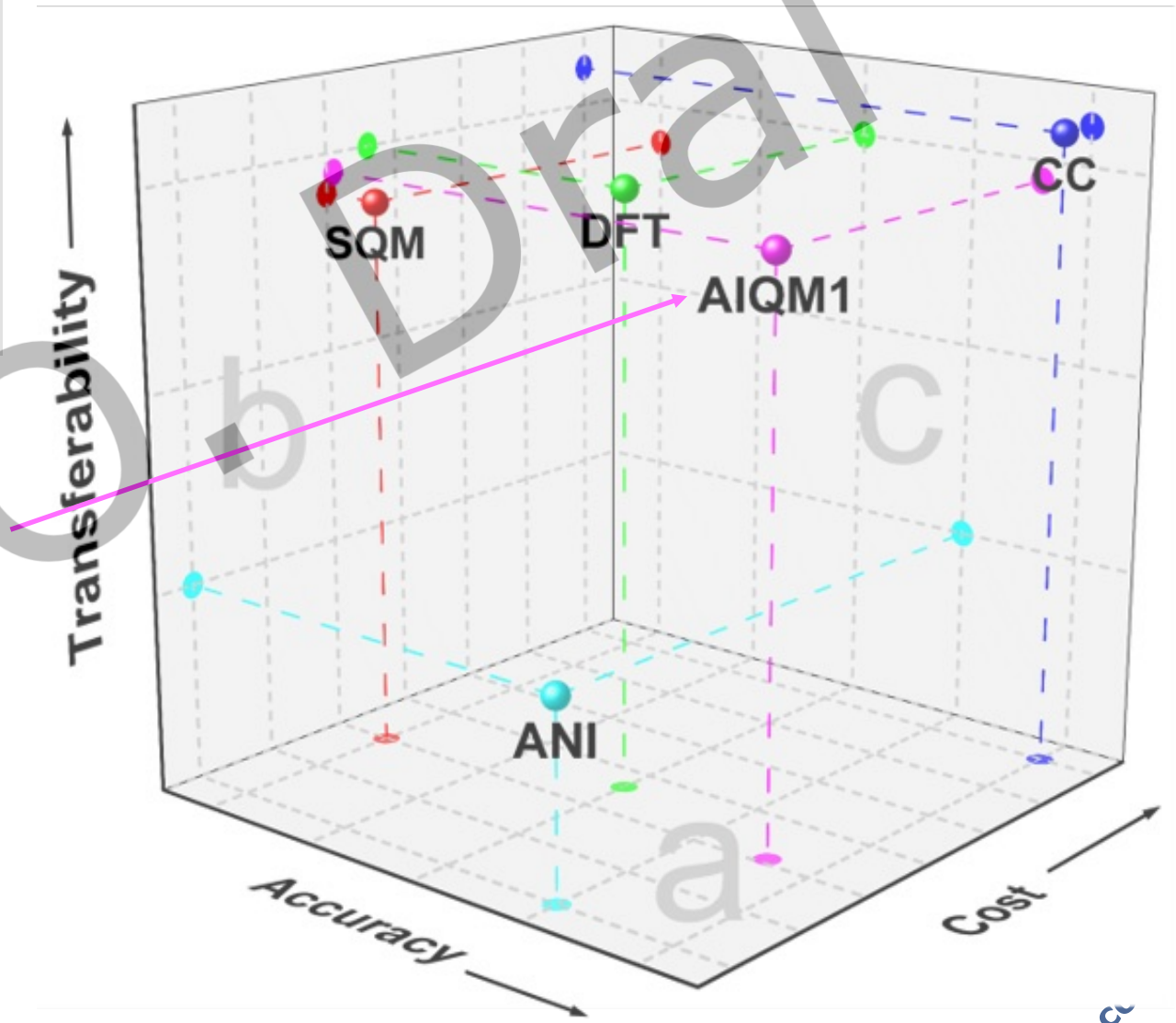
Peikun Zheng

Roman Zubatyuk

$$\text{AIQM1} = \text{SQM} + \Delta^{\text{CCSD(T)/CBS}}_{\text{SQM}} + \text{D4}$$



Intelligence-Enhanced Quantum Chemical Method 1
AIQM1



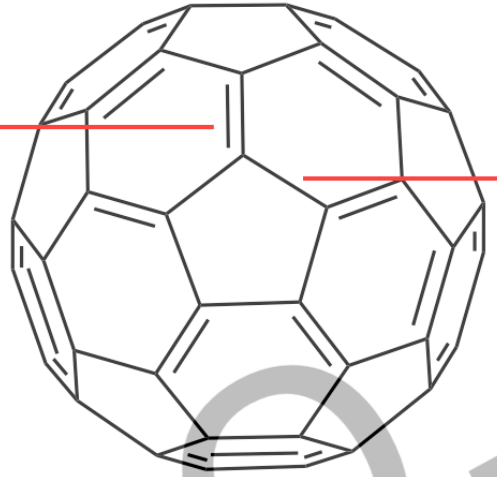
CCSD(T)*/CBS SP: 69 h /15 CPUs
DFT opt: 30 min /32 CPUs
AIQM1 opt: 14 sec / 1 CPU

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral. Artificial Intelligence-Enhanced Quantum Chemical Method with Broad Applicability. *Nat. Commun.* **2021**, *12*, 7022

Ground-state geometries

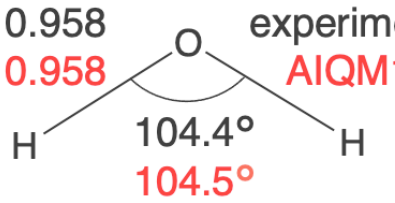
a

1.355–1.401 (experiment)
1.393 (AIQM1)
1.388 (AIQM1@DFT(*))
1.379 (ωB97X(-D4)/ def2-TZVPP)
1.451 (ANI-1ccx)



1.432–1.467 (experiment)
1.467 (AIQM1)
1.464 (AIQM1@DFT(*))
1.448 (ωB97X(-D4)/ def2-TZVPP)
1.451 (ANI-1ccx)

b H_2O in CHNO benchmark



experiment
AIQM1

c MGHBL9 and MGNHBL11 benchmark

$H_3C–H$	$H–H$	$HO–H$				
1.086	0.741	0.957				
1.086	0.726	0.958				
$HC≡C–H$	$N≡C–H$	$^{\bullet}O–H$	$^{-}N=N^{+}=O$		$OC=O$	
1.203 1.063	1.153 1.065	0.970	1.128 1.184		1.160	
1.205 1.064	1.152 1.065	0.959	1.120 1.186		1.160	
$O=C(H)–H$	$H_2N–H$		$N≡N$		$^{-}C≡O^{+}$	
1.203 1.101	1.012	theoretical reference	1.098		1.128	
1.205 1.101	1.012	AIQM1	1.097		1.125	

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

Heats of formation

data set	ODM2	B3LYP/ 6-31G*	ωB97X/ 6-31G*	ωB97X-D/ 6-31G*	ωB97X/ def2-TZVPP	ωB97X-D4/ def2-TZVPP	ANI-1ccx	AIQM1 @DFT*	AIQM1 @DFT	AIQM1	CCSD(T)* /CBS
	energies, kcal/mol										
CHNO	2.64	6.71	4.10	3.84	3.21	2.76	—	2.49	2.12	0.87	—
G3/99	3.04	8.53	3.46	3.22	4.18	3.20	—	2.83	2.06	0.88	—
ISOMERS44 (ΔH_f)	1.16	8.08	3.57	3.53	4.52	3.78	—	3.00	2.27	0.42	—
ISOMERS44 (ΔH_r)	0.70	2.29	1.45	1.31	1.19	1.10	1.68	0.95	0.89	0.50	—
IsoL6/11	1.48	5.26	3.83	3.36	1.75	1.64	1.46	1.65	1.55	0.62	0.47
HC7/11	5.37	6.44	16.90	13.98	6.83	7.10	2.53	8.89	9.16	1.43	1.57

Ground-state properties of neutral, closed-shell compounds
(heats of formation, reaction enthalpies, and ZPVE-exclusive reaction energies)

Torsion	0.74	0.55	0.30	0.29	0.20	0.19	0.23	0.23	0.23	0.19	0.05
	bond lengths, Å										
CHNO	0.015	0.006	0.008	0.007	0.010	0.010	0.011	0.010	0.010	0.007	—
MGHBL9	0.023	0.007	0.006	0.005	0.002	0.002	0.047	0.011	0.011	0.004	—
MGNHBL11	0.026	0.006	0.003	0.002	0.008	0.008	0.004	0.008	0.008	0.002	—
	bond angles, °										
CHNO	2.04	0.70	0.68	0.64	0.68	0.68	1.00	0.77	0.77	0.70	—
	dihedral angles, °										
CHNO	4.07	5.20	4.68	6.10	7.12	7.11	5.86	2.14	2.14	2.31	—

P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral, *Nat. Commun.* **2021**, *12*, 7022

Revising experimental heats of formation

CHNO data set

vinylacetylene

original: 73
revised: 70.4
AIQM1: 69.1
G4: 69.1

acetylacetone

original: –91
revised: –89.9
AIQM1: –87.3
G4: –87.6

1,3-pyrimidine

exper.: 46.8
AIQM1: 44.5
G4: 44.3

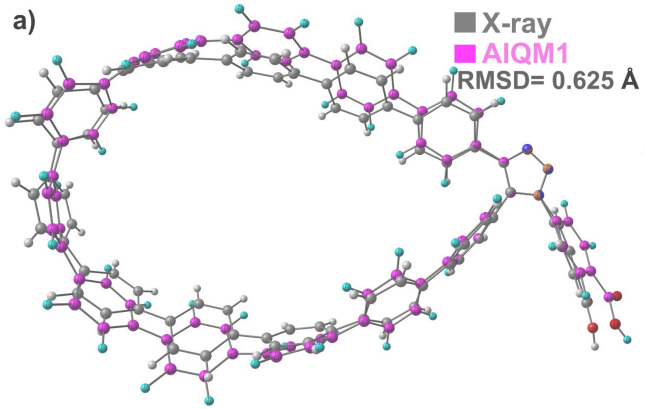
acrolein

exper.: –18
AIQM1: –16.0
G4: –15.9

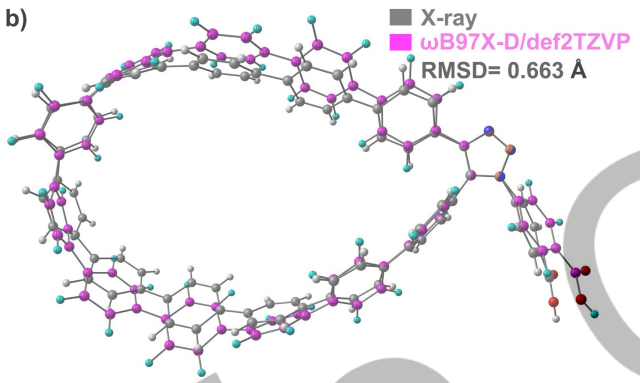
P. Zheng, W. Yang, W. Wu, O. Isayev, P. O. Dral, *J. Phys. Chem. Lett.* **2022**, *13*, 3479

AIQM1 investigation of cycloparaphenylenes

❖ **Quality of theoretically optimized geometries**

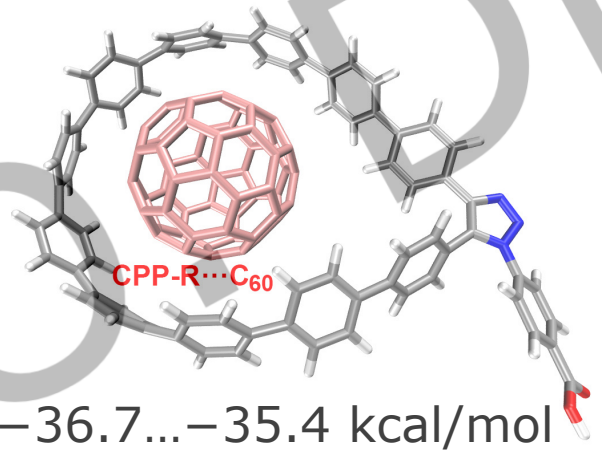


2 CPU-minutes

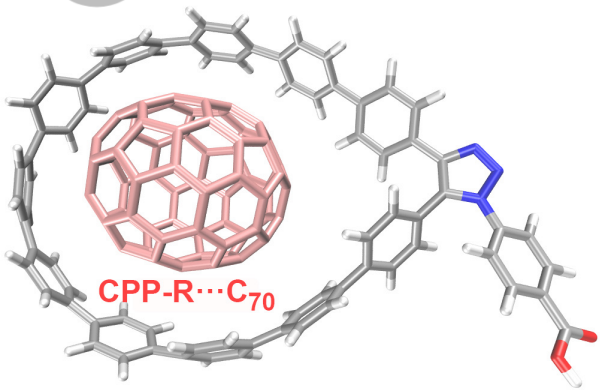


20 CPU-hours

❖ **Interaction with fullerenes at AIQM1**



−36.7…−35.4 kcal/mol
15 CPU-minutes



−41.9…−38.3 kcal/mol
30 CPU-minutes

AIQM1/CIS also correctly predicted that the fluorescence of cycloparaphenylenes is quenched after complexation with fullerenes

T. A. Schaub, A. Zieleniewska, R. Kaur, M. Minameyer, W. Yang, C. M. Schüßlbauer, L. Zhang, M. Freiberger, L. N. Zakharov, T. Drewello, P. O. Dral, D. Guldi, R. Jasti. Tunable Macrocyclic Polyparaphenylene Nanolassos via Copper-Free Click Chemistry. *Chem. Eur. J.* **2023**.

Unpublished: Infrared spectra with AIQM1

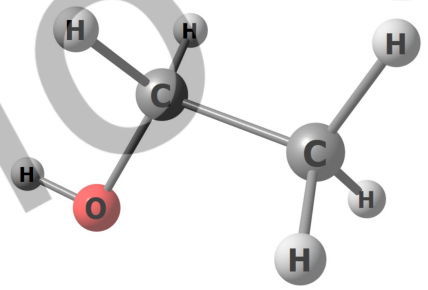
Long MD trajectory is needed!

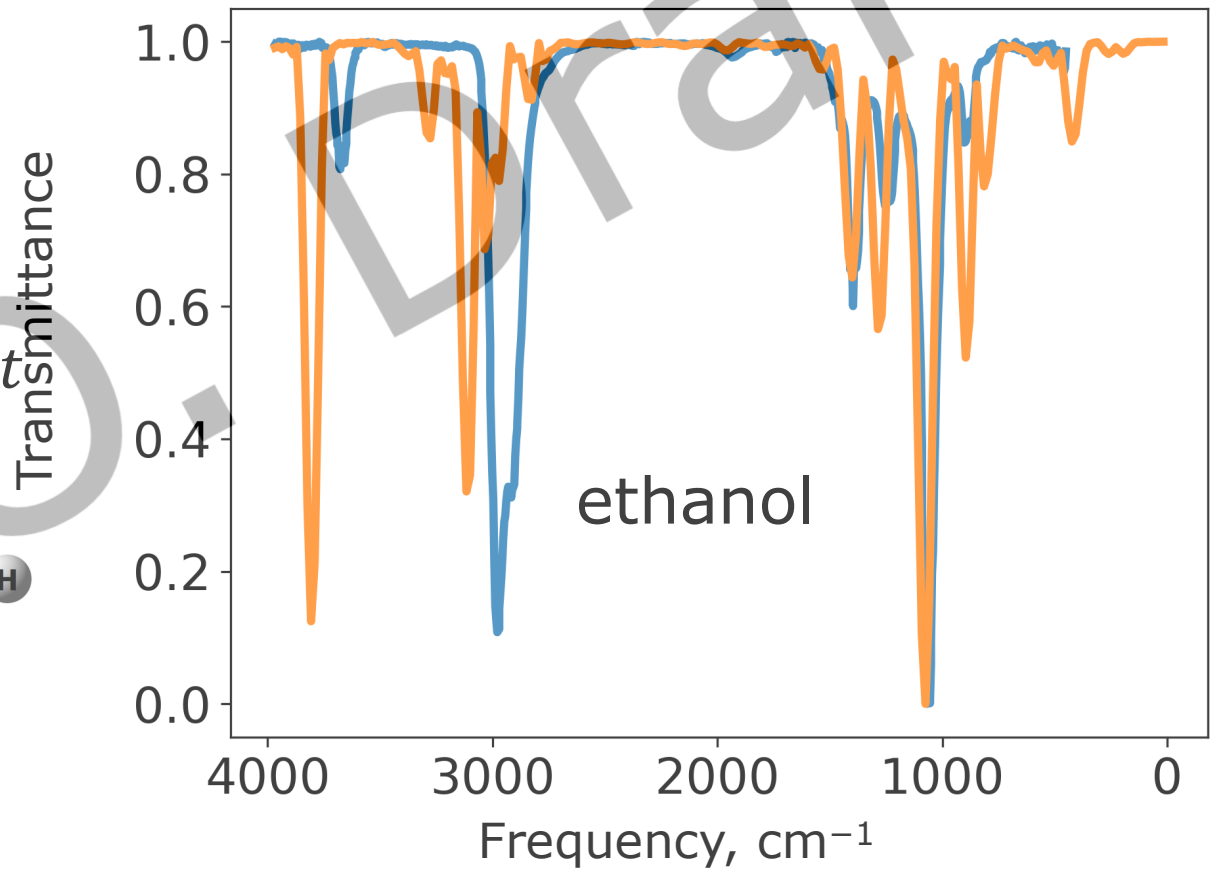
IR intensities:[1,2]

$$P(\omega) \propto \int \langle \dot{\boldsymbol{\mu}}(\tau)\dot{\boldsymbol{\mu}}(t+\tau)\rangle_\tau e^{-i\omega t}\mathrm{d}t$$

$$P(\omega) \propto \omega \tanh\left(\frac{\hbar\omega}{2k_B T}\right) \int \langle \boldsymbol{\mu}(\tau)\boldsymbol{\mu}(t+\tau)\rangle_\tau e^{-i\omega t}\mathrm{d}t$$

Implementation by
Yi-Fan Hou





Experiment (NIST)
AIQM1 MD (by Yifan Hou, unpublished)
200 ps trajectory at 300 K

[1] Phys. Chem. Chem. Phys. **2013,** 15, 6608

[2] Phys. Chem. Chem. Phys. **2016**, 28325

What do we need for molecular dynamics?

To propagate molecular dynamics:

- We just need forces →
  can be obtained efficiently from machine learning potentials

$$F_{A,d} = M_A a_{A,d}$$

$$F = -\frac{\partial E}{\partial r}$$

$$a_{A,d} = F_{A,d}\frac{1}{M_A}$$

$$\frac{\partial^2 \mathbf{r}_{A,d}}{\partial t^2} = -\frac{\partial E}{\partial \mathbf{r}_{A,d}}\frac{1}{M_A}$$

$$\frac{\partial^2 x_{A,d}}{\partial^2 t} = F_{A,d}\frac{1}{M_A}$$

$$F = -\frac{\partial E}{\partial x_{A,d}} = -\frac{\partial E^{\text{machine learning potential}}(\mathbf{x})}{\partial x_{A,d}}$$

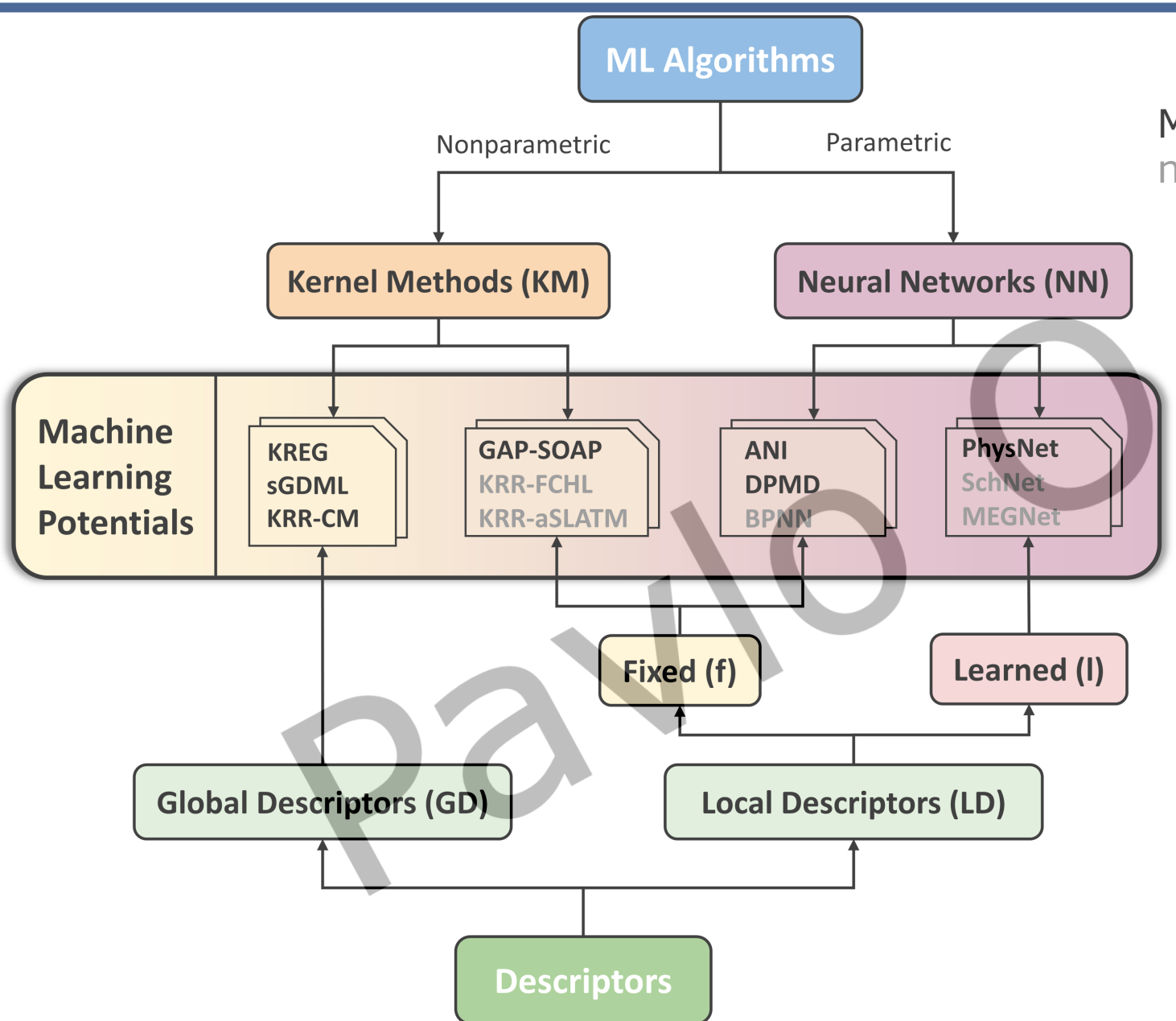
Velocity Verlet algorithm:

$$\mathbf{x}(t+\Delta t) = \mathbf{x}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\boldsymbol{a}(t)\Delta t^2$$

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \frac{1}{2}[\boldsymbol{a}(t) + \boldsymbol{a}(t+\Delta t)]\Delta t,$$

where $\mathbf{x}$ is the coordinate, $\mathbf{v}$ – the velocity, $\boldsymbol{a}$ – the acceleration, t – the time and

Categories of machine learning potentials



Models implemented/interfaced in MLatom
not implemented in MLatom

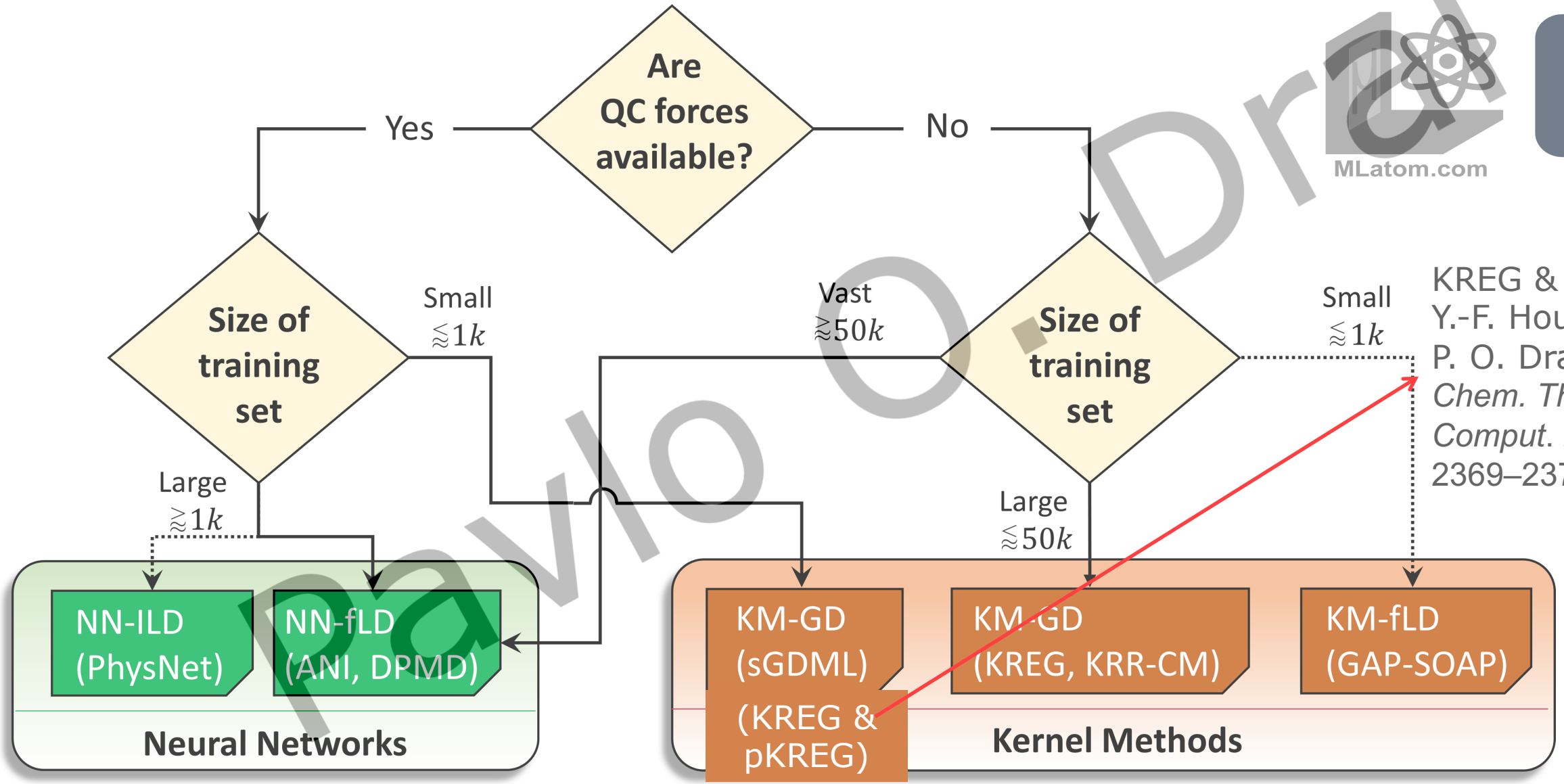
MLatom.com

XACS

M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413

P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou, M. Pinheiro Jr, J. Huang, M. Barbatti. *Top. Curr. Chem.* **2021,** *379*, 27

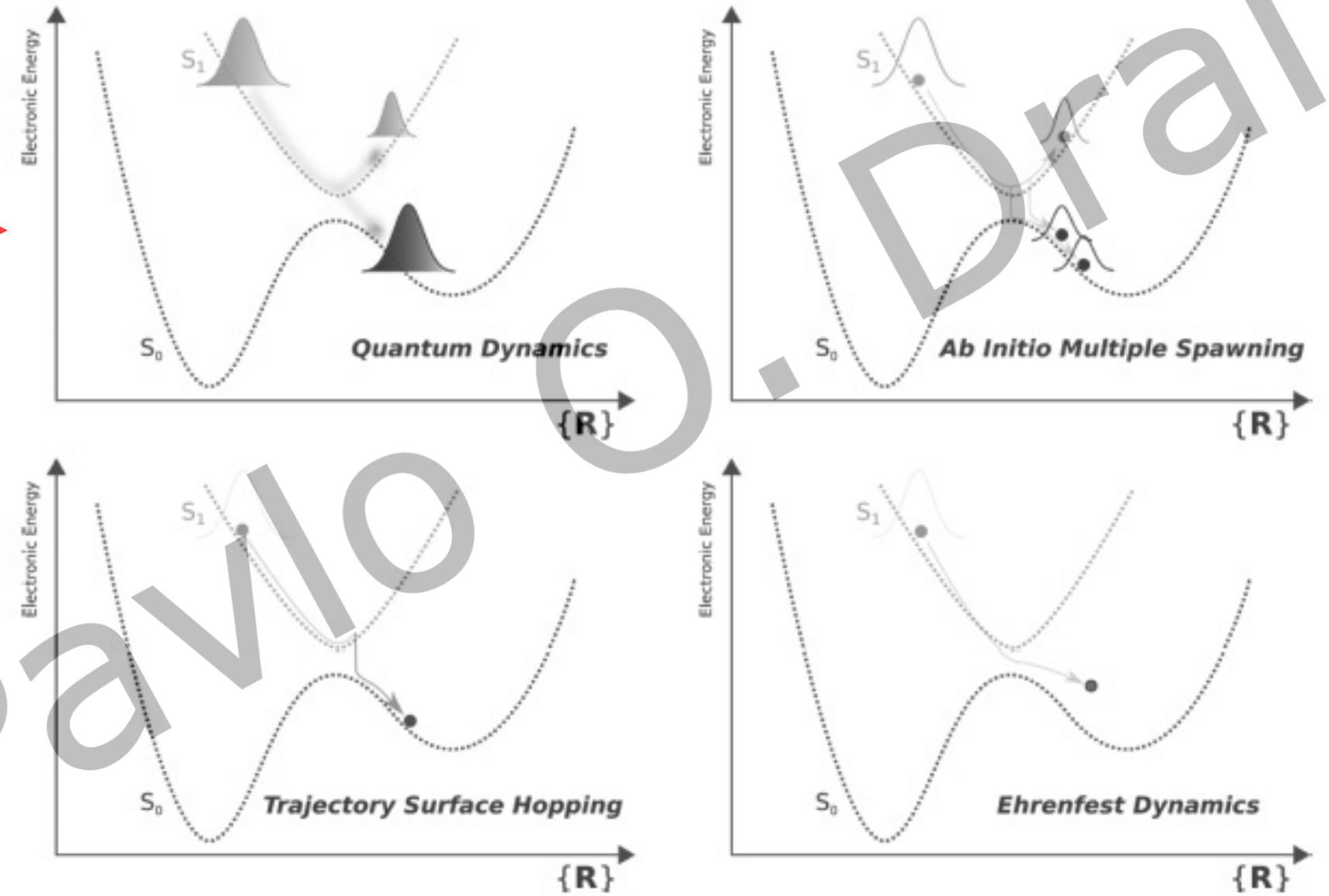
Choosing the 'Right' ML Potential



KREG & pKREG: Y.-F. Hou, F. Ge, P. O. Dral. *J. Chem. Theory Comput.* **2023**, *19*, 2369–2379

M. Pinheiro Jr, F. Ge, N. Ferré, P. O. Dral, M. Barbatti. *Chem. Sci.* **2021**, *12*, 14396–14413

Many flavors for quantum dynamics (QD)



B. F. E. Curchod, T. J. Martínez, *Chem. Rev.* **2018,** *118*, 3305

Machine learning to speed up dynamics

$$\cancel{\boldsymbol{\rho}(t) = f[\boldsymbol{\rho}(t - \Delta t)]}$$

dynamics propagation is

- ~~computationally expensive~~

dynamics

ML dynamics

$$\boldsymbol{\rho}(t) = f^{\mathrm{ML}}[\rho(t - \Delta t)]$$

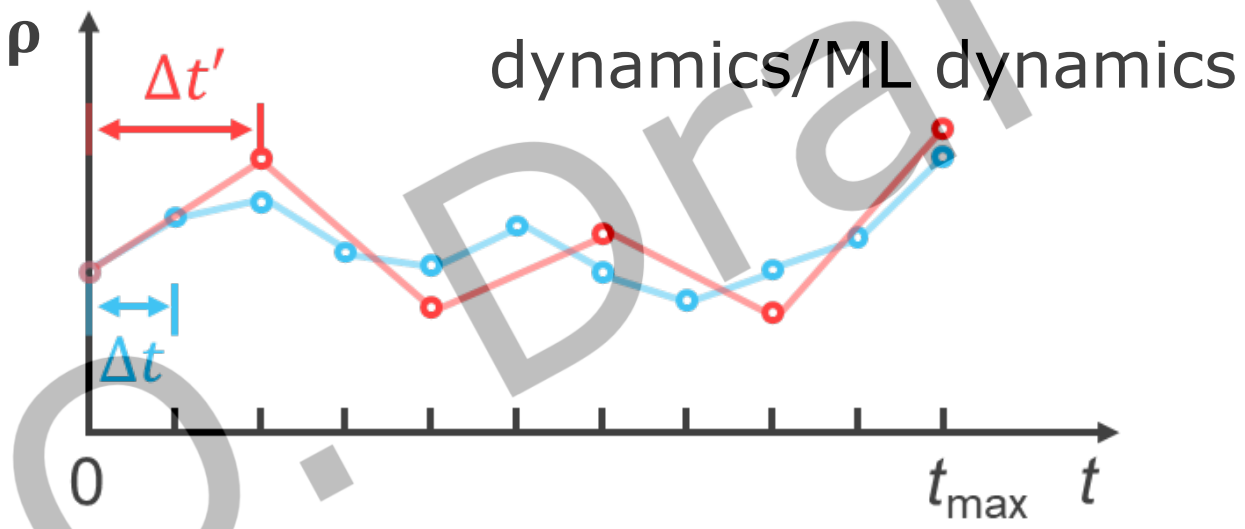
L. E. H. Rodrıguez, A. Ullah, K. J. R. Espinosa, P. O. Dral, A. A. Kananenka. *Mach. Learn. Sci. Technol.* **2022,** *3*, 045016

Can we do better?

$$\cancel{\boldsymbol{\rho}(t) = f[\boldsymbol{\rho}(t - \Delta t)]}$$

dynamics propagation is

- ~~computationally expensive~~
- recursive (iterative)



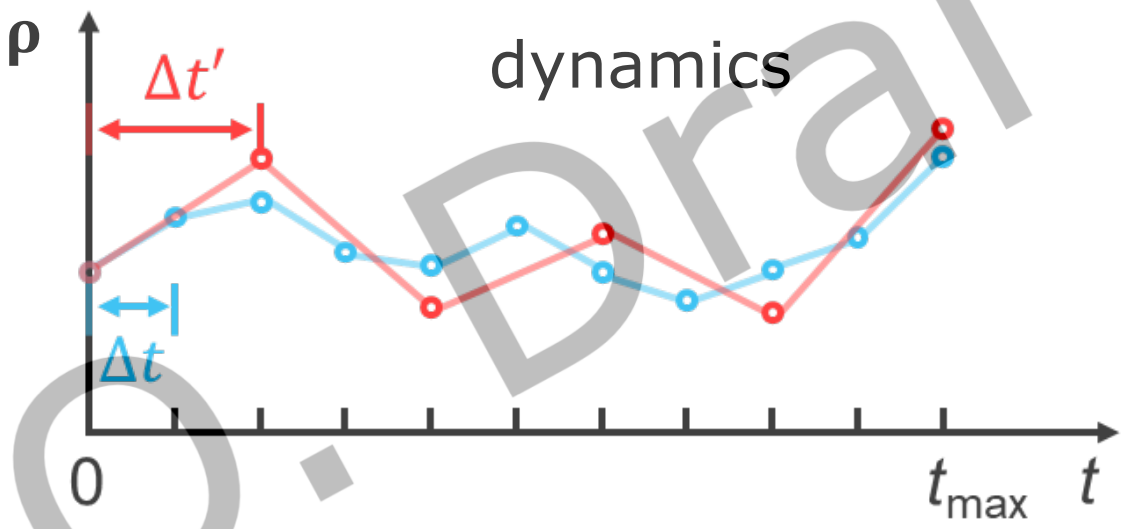
Can we do better?

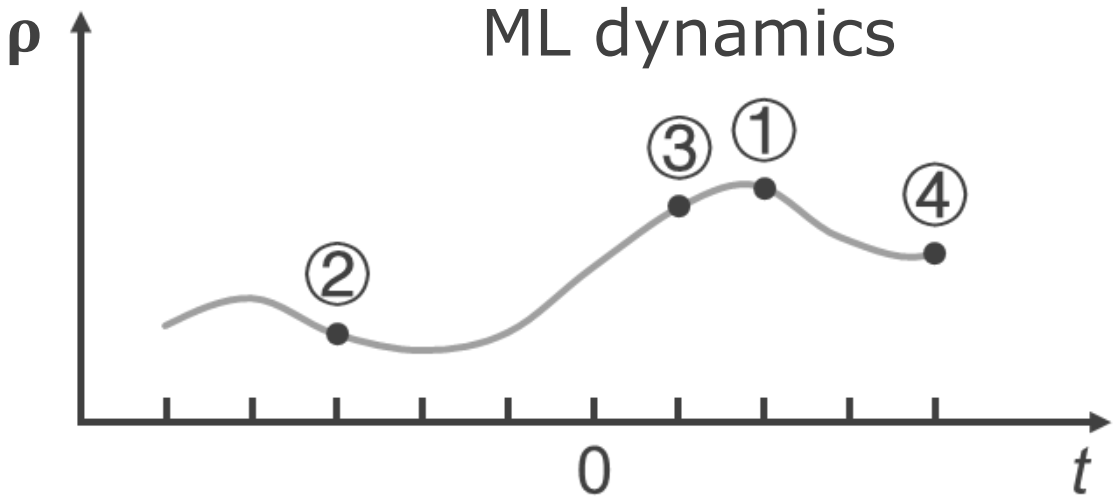
$$\cancel{\boldsymbol{\rho}(t) = f[\boldsymbol{\rho}(t - \Delta t)]}$$

dynamics propagation is

- ~~computationally expensive~~
- ~~recursive (iterative)~~

$$\boldsymbol{\rho}(t) = f[t; \text{other parameters}]$$



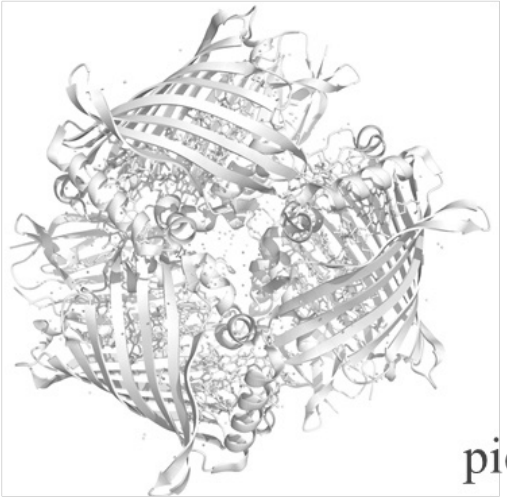


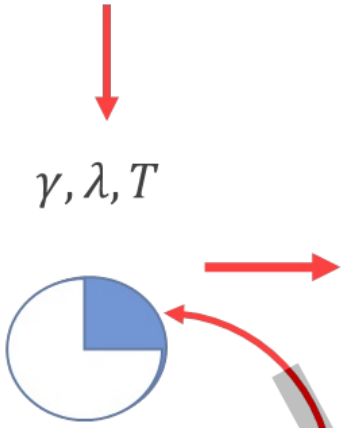
AI-QD (artificial intelligence-based quantum dynamics)

γ = characteristic frequency
λ = reorganization energy
T = temperature

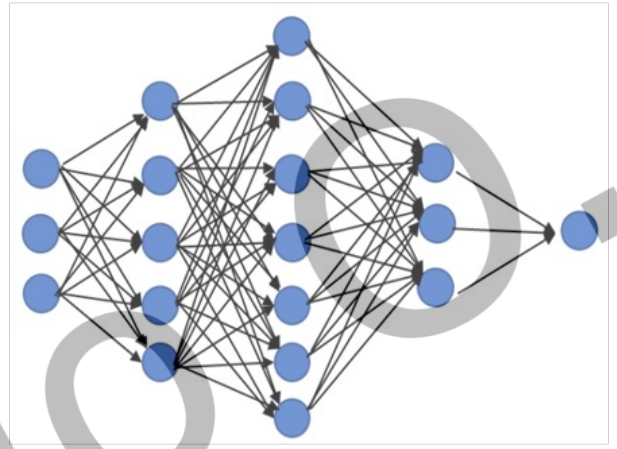
$$\boldsymbol{\rho}(\text{time}) = f[\text{time}; \text{simulation parameters}]$$

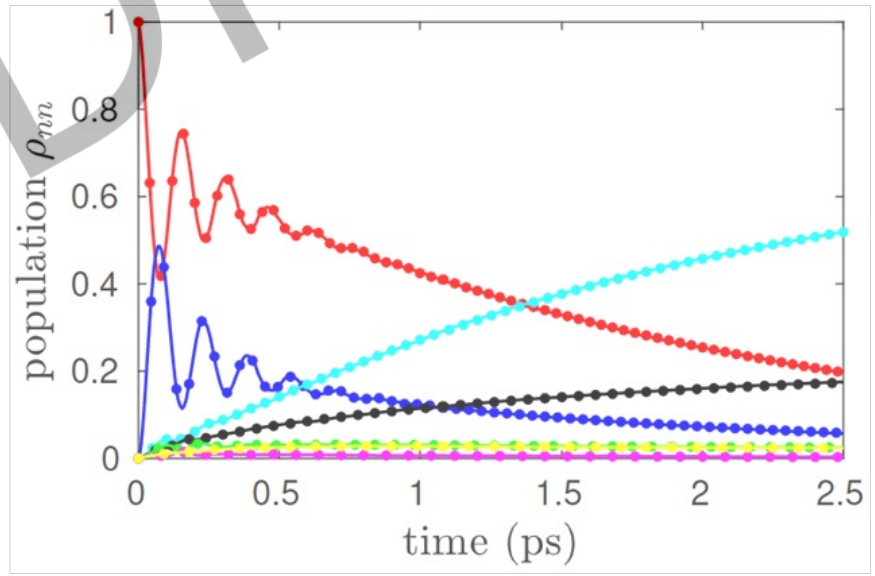
PDB code: 3ENI





2.5ps



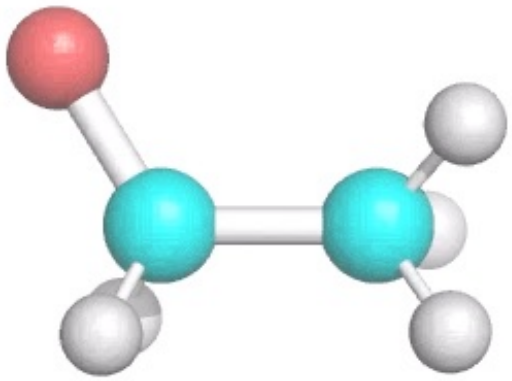


Dots: Reference
Line: AI-QD

7-sites Fenna–Matthews–Olson (FMO) complex

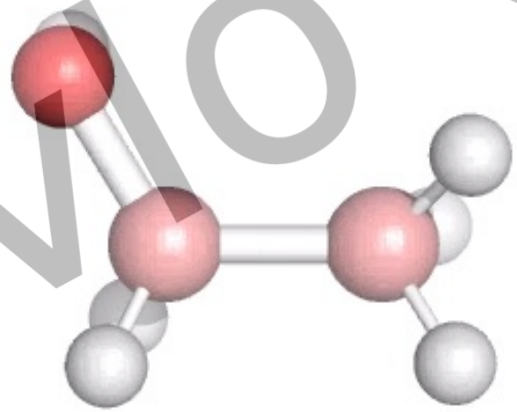
A. Ullah, P. O. Dral. Predicting the future of excitation energy transfer in light-harvesting complex with artificial intelligence-based quantum dynamics. *Nat. Commun.* **2022**, *13*, 1930

3D MD



1 CPU day per 2.5 ps

4D–A^2I



0.05 fs

1 ns per 1 hour wall-clock time on a single RTX 3080Ti GPU card

F. Ge, L. Zhang, A. Ullah, P. O. Dral. Four-dimensional-spacetime atomistic artificial intelligence models. **March 17, 2022,** preprint: http://doi.org/10.26434/chemrxiv-2022-qf75v

Can we do even better?

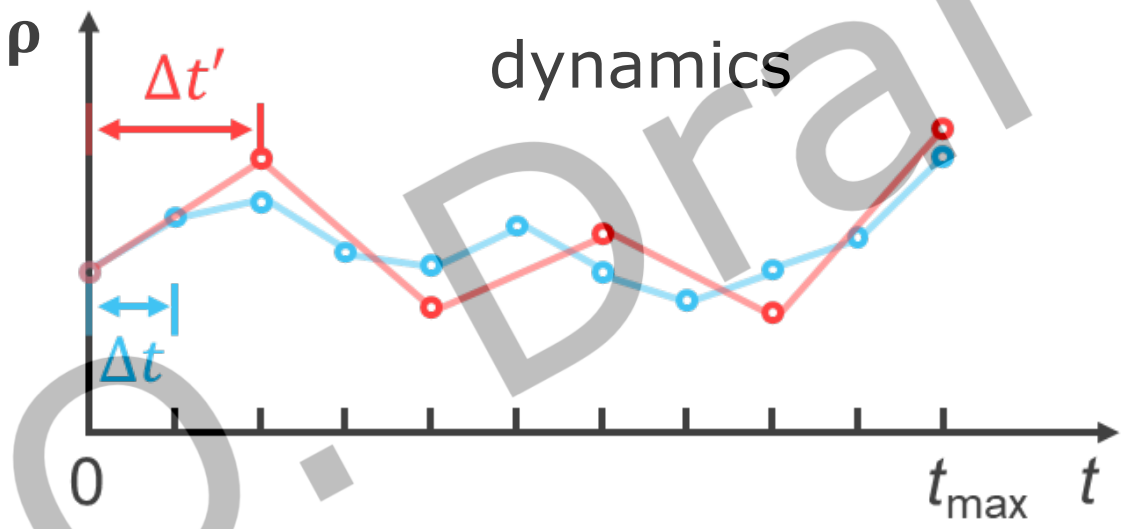
$$\cancel{\boldsymbol{\rho}(t) = f[\boldsymbol{\rho}(t - \Delta t)]}$$

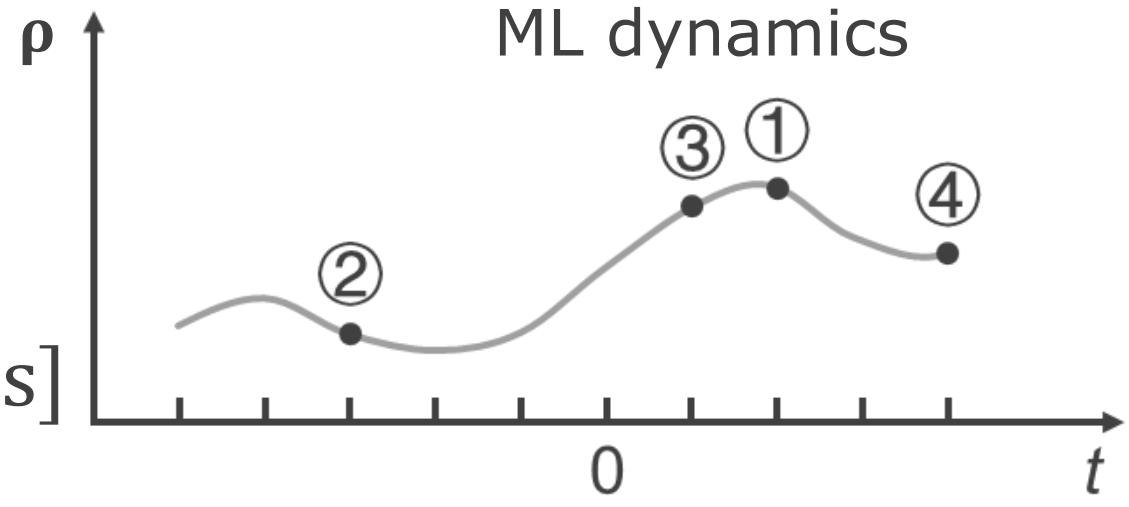
dynamics propagation is

- ~~computationally expensive~~
- ~~recursive (iterative)~~

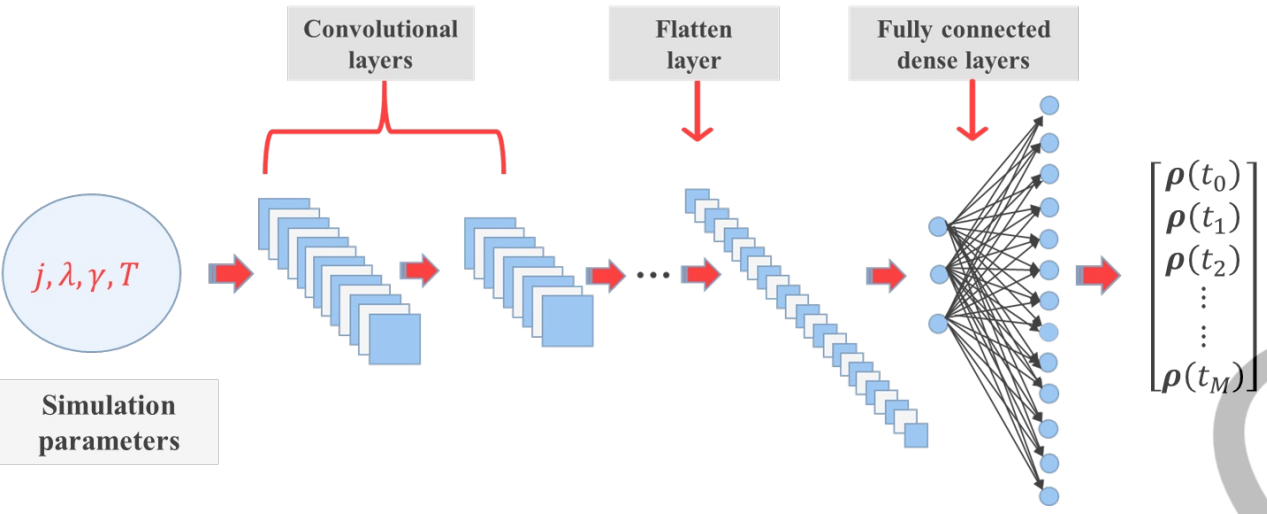
$$\cancel{\boldsymbol{\rho}(t) = f[t; \text{other parameters}]}$$

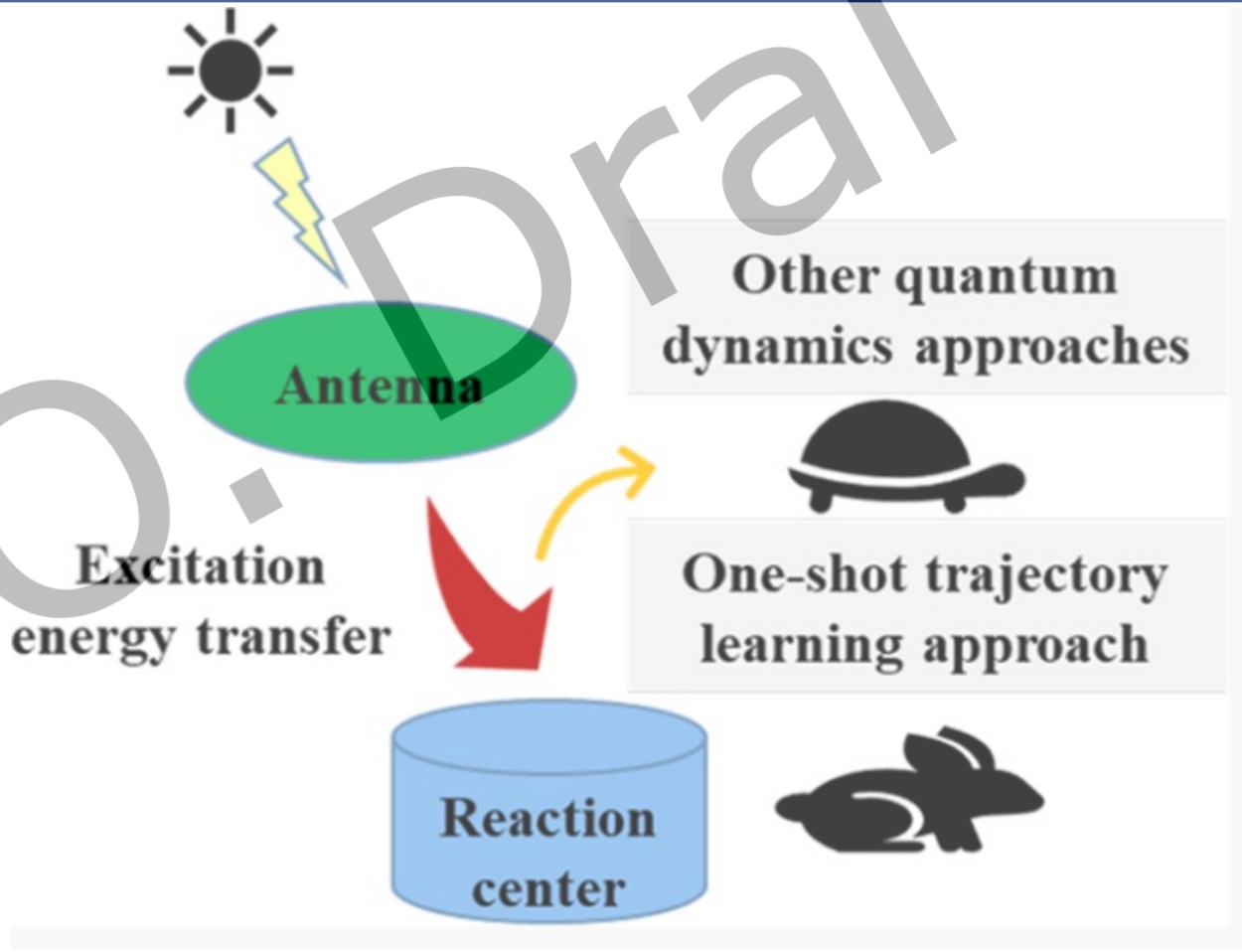
$$\{\boldsymbol{\rho}(t_k)\}_{k=0}^{M} = f[\text{other parameters}]$$





One-Shot Trajectory Learning (OSTL)





$$\{\boldsymbol{\rho}(t_k)\}_{k=0}^{M} = f[\text{other parameters}]$$

- **10 ps long dynamics in just 70 ms**
- good for massive simulation in parameter space

A. Ullah, P. O. Dral. *J. Phys. Chem. Lett.* **2022**, 6037

Thank you for your attention!

Artificial intelligence in computational chemistry

January 2023

Artificial intelligence firmly entere
latest advances in application of ar
topics include, but not limited to,
machine learning, extracting chem
quantum chemical methods impro
artificial intelligence methods for

Guest editors:

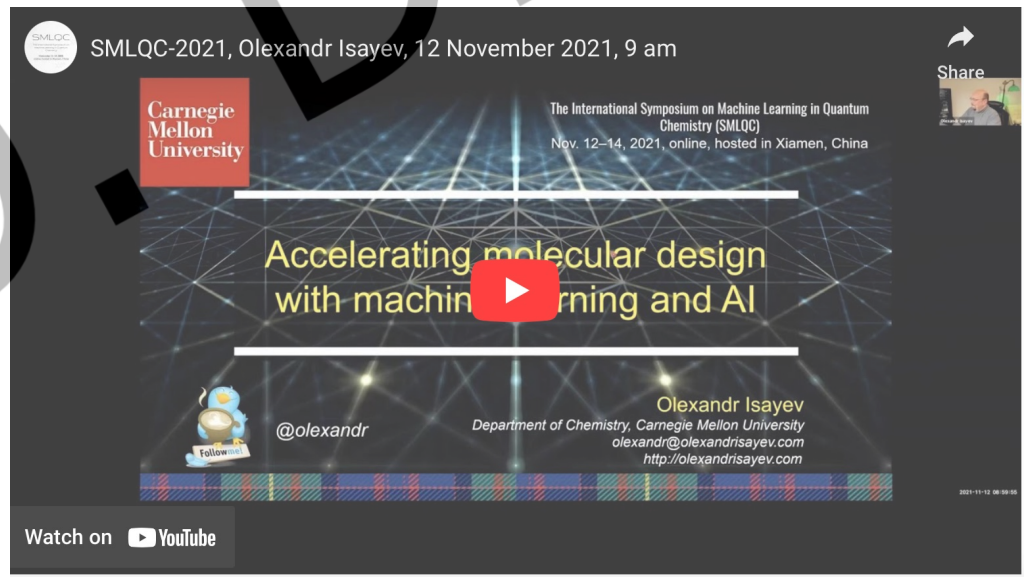
Pavlo O. Dral, PhD
Xiamen University, Xiamen, China
(Machine learning in chemistry, Quantum chemistry, Excited states, Electronic structure simulations, Dynamics)

Joel M. Bowman, PhD
Emory University, Atlanta, Georgia, United States of America
(Machine learning potential energy surfaces, vibrational and reaction dynamics)

Feliu Maseras, PhD
Institute of Chemical Research of Catalonia, Tarragona, Spain
(Computational homogeneous catalysis; Quantum mechanics/Molecular mechanics methods; Statistical analysis of computational results)

Fang Liu, PhD
Emory University, Atlanta, Georgia, United States of America
(quantum chemistry, solvent effects, machine learning, excited states)

Symposia and seminars on machine learning in quantum chemistry



SMLQC-2023 to be held in Uppsala, Sweden (Organizer: Roland Lindh)

More info with videos & news:
http://smlqc.mlatom.com