Pavlo O. Dral

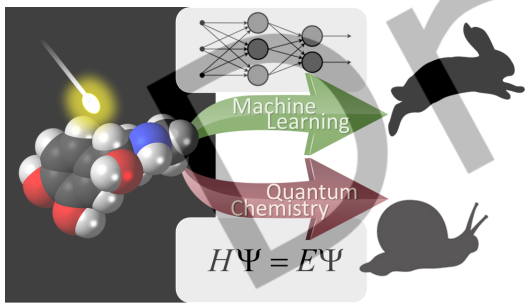
AI computational chemistry

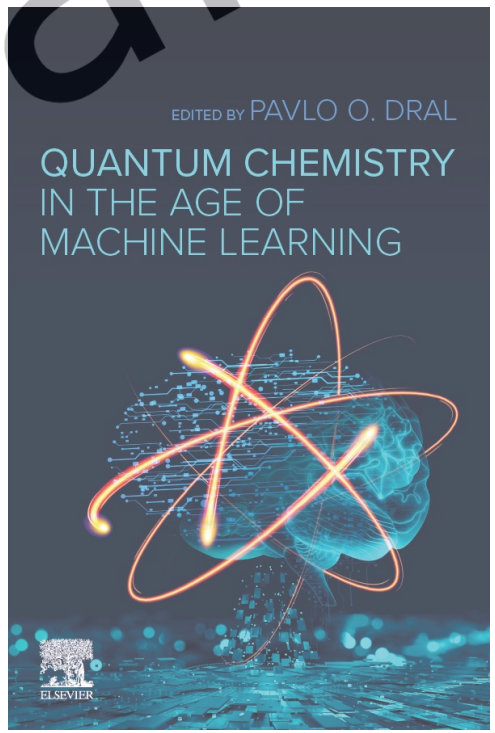


Professor | Outstanding Youth (Overseas)

Email: dral@xmu.edu.cn

Research Areas: artificial intelligence, quantum chemistry, dynamics, excited states, semi-empirical methods





XACS
Xiamen Atomistic Computing Suite

nature communications
CHEMISTRY AND MATERIALS SCIENCES
TOP 25
READ ARTICLES OF 2021
OFFICIAL AUTHOR

2021-Present: Xiamen University, Full Professor
2019-2021: Xiamen University, Associate Professor
2013-2019: Max-Planck-Institut für Kohlenforschung, Postdoc
2010-2013: University of Erlangen-Nürnberg, Ph.D.

MLatom.com

Research Interests:

Our research transforms chemical simulations by developing novel AI methods and providing software and cloud computing platforms.

✓Selected papers:

AI platform: ***J. Chem. Theory Comput.*** *2024*, *20*, *1193*
AI-quantum dynamics: ***Nat. Commun.*** *2022*, *13*, *1930*
AI-quantum mechanics: ***Nat. Commun.*** *2021*, *12*, *7022*
AI-excited states: ***Nat. Rev. Chem.*** *2021*, *5*, *388*
AI force fields: ***Chem. Sci.*** *2021*, *12*: *14396*

Group website: dr-dral.com

Pavlo O. Dral

Lecture 1

Quantum chemical levels and machine learning

Pavlo O. Dral

22 April 2024

Task 1.1. Calculate the bond length in H_2 molecule as fast as possible.

14:48

Questions:
- method you used
- bond length you obtained
- how much time it took to complete the calculations

Send answers on WeChat!

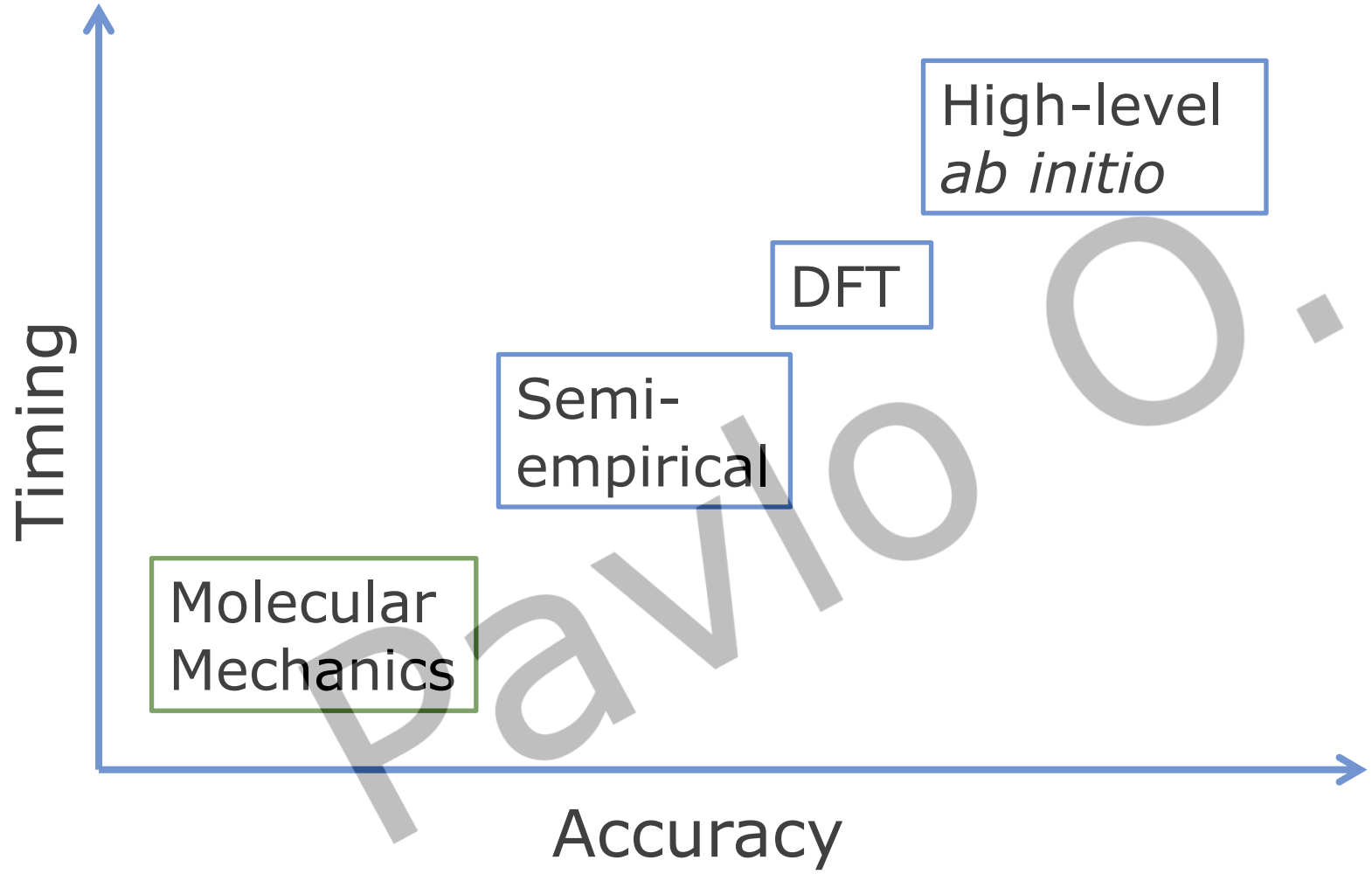
10 min

Task 1.2. Calculate the bond length in H_2 molecule as accurate as possible.

Questions:
- method you used
- bond length you obtained
- how much time it took to complete the calculations

Pavlo O. Dral

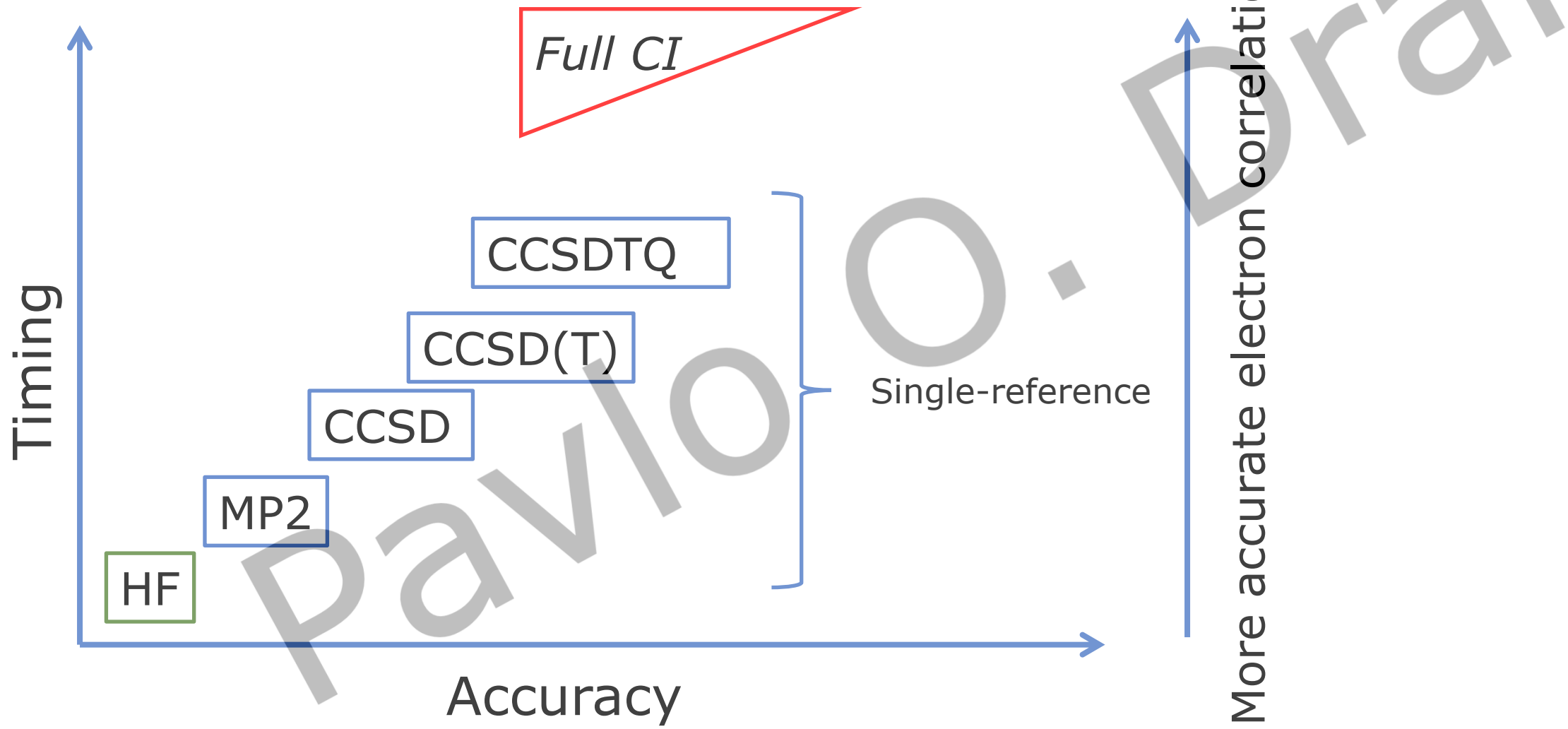
Quantum chemistry approximations



Pavlo O. Dral

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

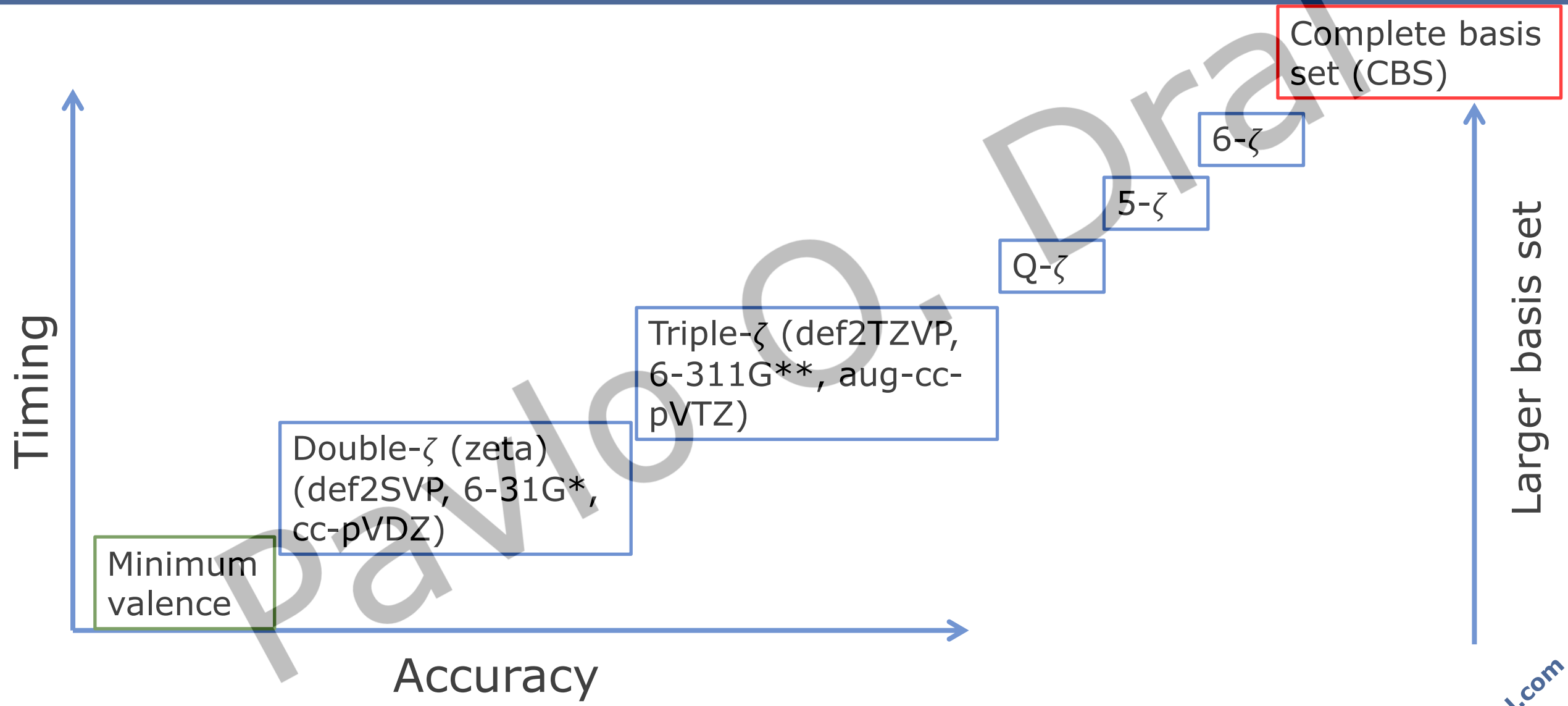
Quantum chemistry approximations



Pavlo O. Dral

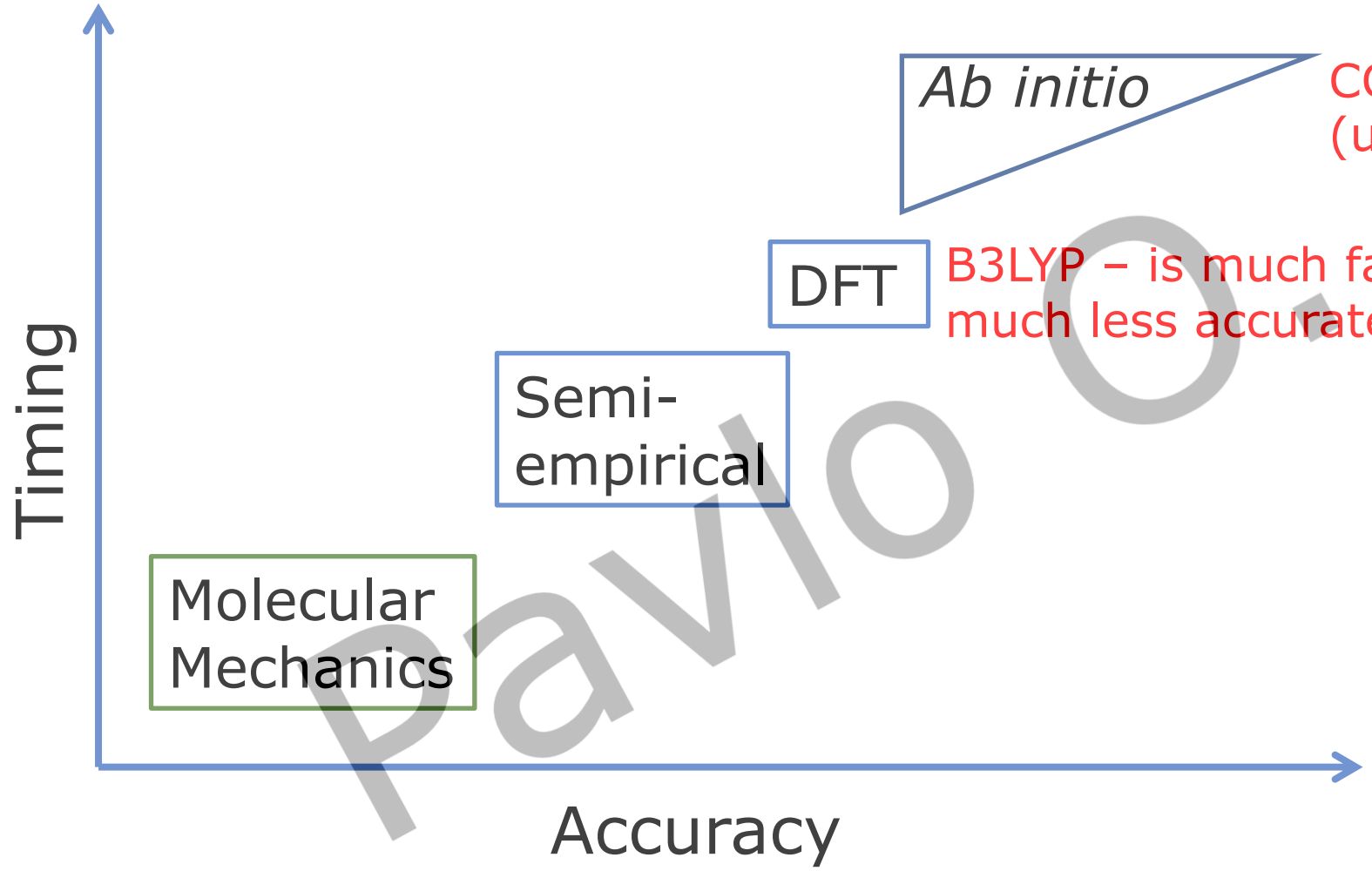
P. O. Dral, *J. Phys. Chem. Lett.* **2020**, *11*, 2336

Quantum chemistry approximations



Pavlo O. Dral

Quantum chemistry approximations



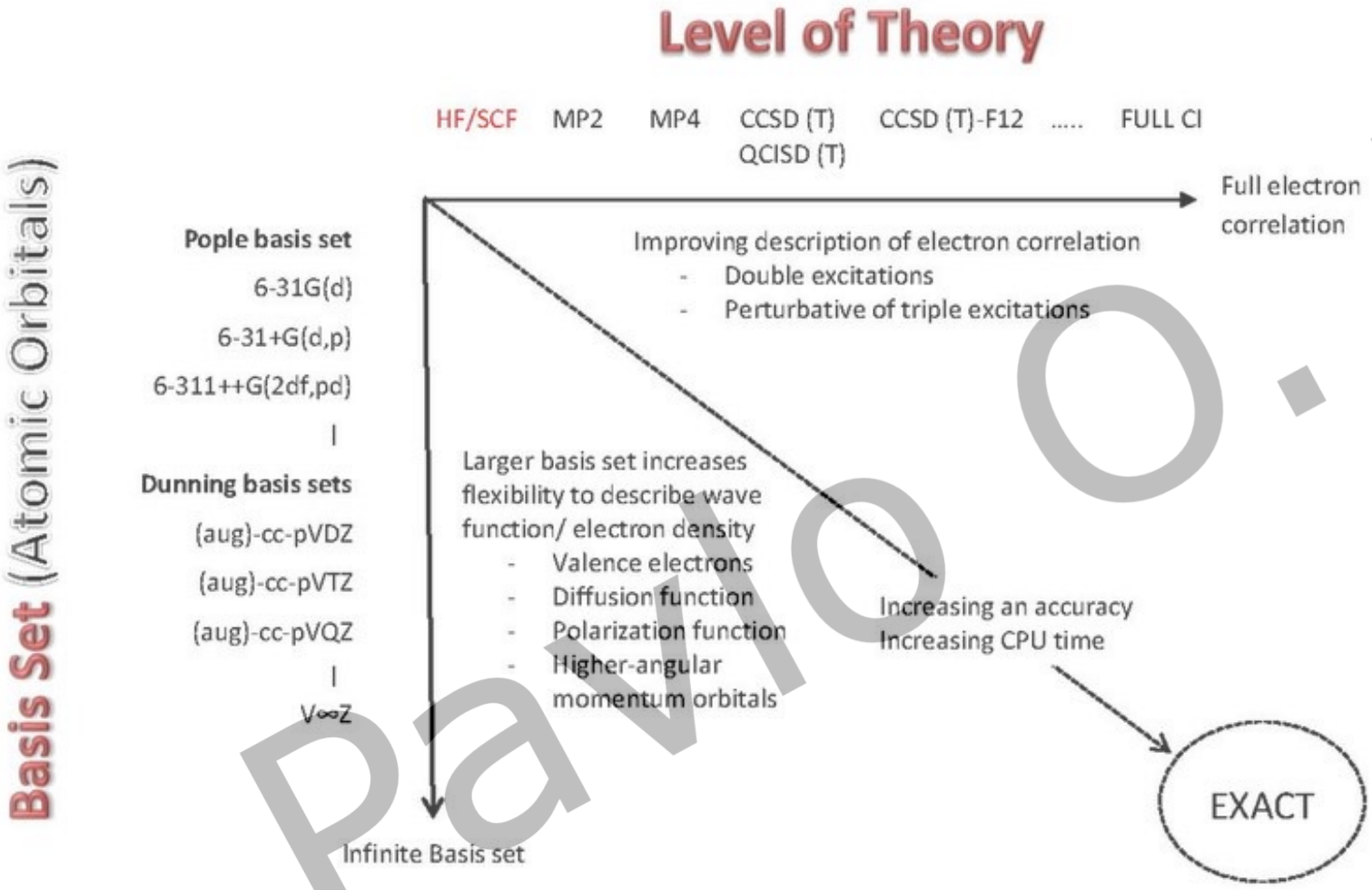
CCSD(T) – gold standard, very slow (use it if you can!)

B3LYP – is much faster than the CCSD(T) but is much less accurate

Pavlo O. Dral

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

Pople's diagram





John Pople, Nobel prize, 1998

Pavlo O. Dral

Panisanun, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

The experimental bond length in H_2 is 0.7414 Å

Task 1.2. Calculate the bond length in H_2 molecule as accurate as possible.

Questions:
- method you used
- bond length you obtained
- how much time it took to complete the calculations

Pavlo O. Dral

The experimental bond length in H_2 is 0.7414 Å

Even for such a simple molecule, rigorous quantum chemical (QC) calculations take substantial time to get nearly perfect agreement with experiment.

E.g. FCI/aug-cc-pV6Z: 0.7415 Å, ~5 (CPU) days

Pavlo O. Dral

Task 1.3. Calculate the bond length in H_2 molecule with machine learning (ML) (see instructions below).

Questions:

- bond length you obtained
- how much time it took to complete the calculations

Send answers on WeChat!

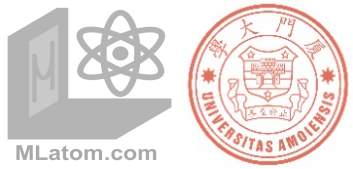
8 min

MLatom@XACS input file (not Gaussian!):

```
geomopt                 # Request geometry optimization
MLmodelType=KREG        # of the KREG type
MLmodelIn=energies.unf # in energies.unf file
XYZfile='
2

H              0.0000000000000          0.0000000000000          0.0000000000000
H              0.0000000000000          0.0000000000000          0.8000000000000
'
optXYZ=eq_KREG.xyz      # optimized geometry output
```

Pavlo O. Dral



Job Information

* Job Name: h2_opt_ml

Job Location: from_job_submitter

Job Type: XACS (auto detect) | XMVB | XEDA | MLatom | Gaussian | Mlatom_dev | Turbomole

Input File

or edit XACS input file:

```
1 geomopt                # Request geometry optimization
2 MLmodelType=KREG       # of the KREG type
3 MLmodelIn=energies.unf # in energies.unf file
4 XYZfile='
5 2
6
7 H            0.0000000000000          0.0000000000000          0.
8 H            0.0000000000000          0.0000000000000          0.
9 '
10 optXYZ=eq_KREG.xyz     # optimized geometry output
```

Additional auxiliary file (optional)

auxiliary file(s)

Name	Operation
energies.unf	DELETE

Submit

Pavlo O. Dral

Input File

or edit XACS input file:

```
1 geomopt                 # Request geometry optimization
2 MLmodelType=KREG        # of the KREG type
3 MLmodelIn=energies.unf # in energies.unf file
4 XYZfile=h2_init.xyz
5 optXYZ=eq_KREG.xyz      # optimized geometry output
```

Additional auxiliary file (optional)

auxiliary file(s)

Name	Operation
h2_init.xyz	DELETE
energies.unf	DELETE

Pavlo O. Dral

Task 1.3. Calculate the bond length in H_2 molecule with machine learning (ML) (see instructions below).

Questions:

- bond length you obtained
- how much time it took to complete the calculations

Note that MLatom can also do quantum chemical calculations, e.g.:

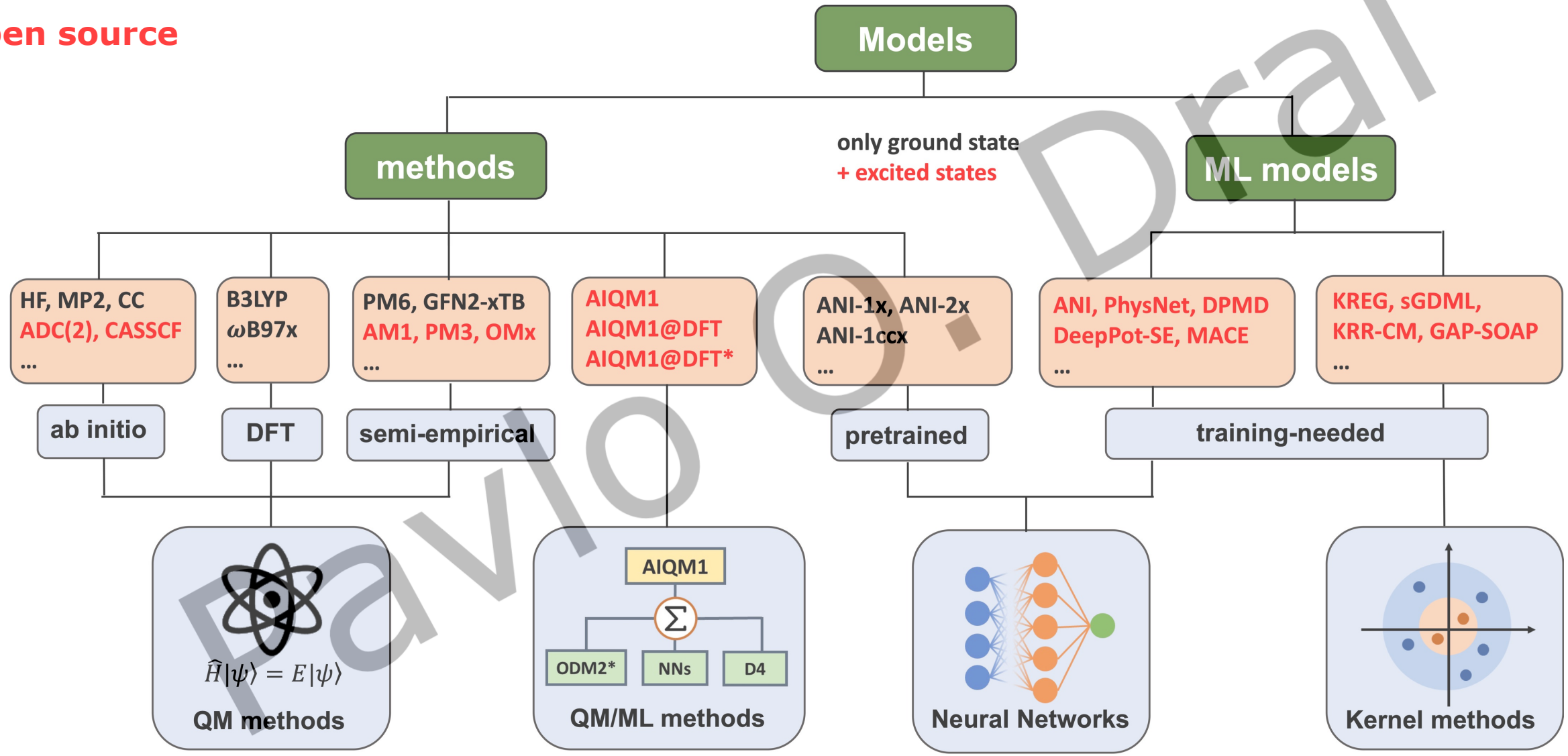
```
geomopt                  # Request geometry optimization
method=CCSD(T)/cc-pVDZ
qmprog=gaussian
XYZfile='
2

H              0.0000000000000          0.0000000000000          0.0000000000000
H              0.0000000000000          0.0000000000000          0.8000000000000
'
optXYZ=eq_KREG.xyz       # optimized geometry output
```

Pavlo O. Dral

MLatom: Ecosystem for AI-enhanced computational chemistry

Open source



Pavlo O. Dral

P. O. Dral, F. Ge, Y.-F. Hou, P. Zheng, Y. Chen, M. Barbatti, O. Isayev, C. Wang, B.-X. Xue, M. Pinheiro Jr, Y. Su, Y. Dai, Y. Chen, S. Zhang, L. Zhang, A. Ullah, Q. Zhang, Y. Ou. *J. Chem. Theory Comput.* **2024,** *20*, 1193

Extras!

MLQD

A Package for Quantum Dissipative Dynamics with Machine Learning
by Arif Ullah, Anhui University

[MLQD: A. Ullah, P. O. Dral. *Comput. Phys. Commun*. **2024**, *294*, 108940]

XACScloud.com

Quantum chemical programs:

Gaussian, GAMESS, ORCA, PySCF, Turbomole, COLUMBUS

Interfaces

Semi-empirical quantum chemical programs: xTB, SPARROW, MNDO

Machine learning programs: PhysNet, GAP, TORCHANI, SCINE, hyperopt

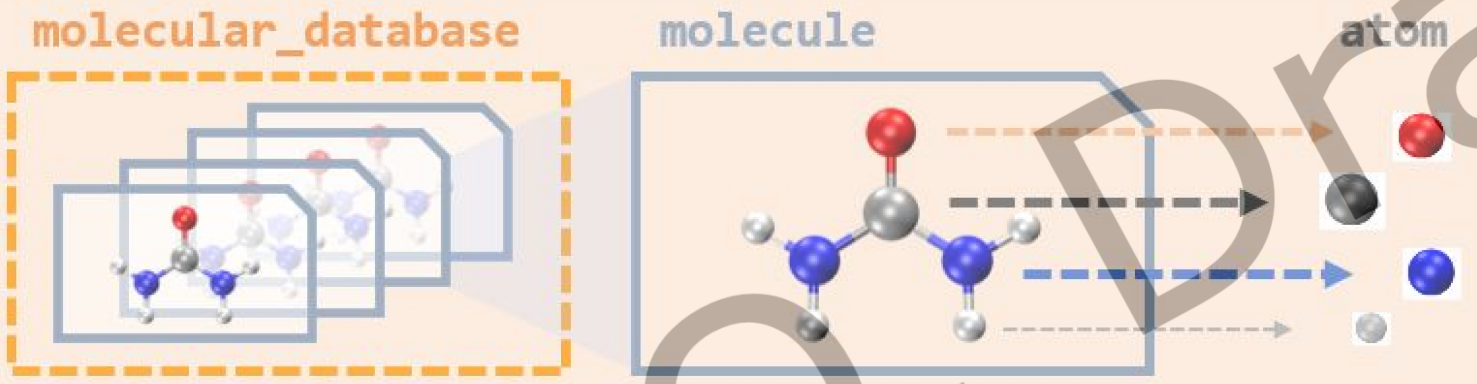
Dynamics and other atomistic simulation: ASE, NX newtonx.org

Pavlo O. Dral

Open source

Implementation and use

- Data in different formats and types **V3**



- Python API **V3**

```
import mlatom as ml
```

- Input file

```
ANI-1ccx
geomopt
xyzfile=init.xyz
optxyz=opt.xyz
```

- Command line

```
> $mlatom ANI-1ccx geomopt xyzfile=init.xyz optxyz=opt.xyz
```

Pavlo O. Dral

* 'V3' marks implementations released in MLatom 3

MLatom: Program for AI-enhanced computational chemistry

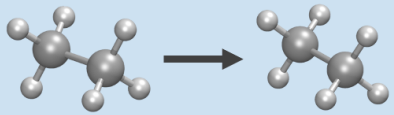
XACS
Xiamen Atomistic Computing Suite
XACScloud.com

MLatom.com

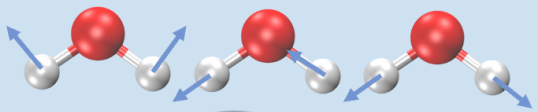
Open source

Single point calculations — Energies, forces, Hessian matrix...

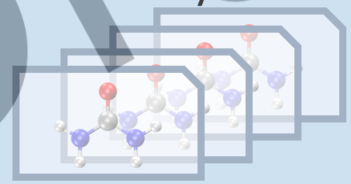
Geometry optimizations



Frequency calculations



Molecular dynamics

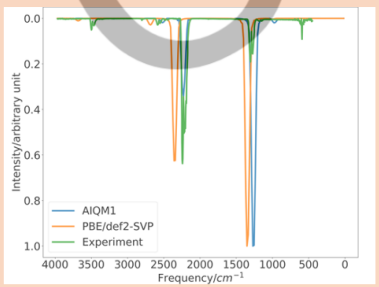


Thermochemistry calculations

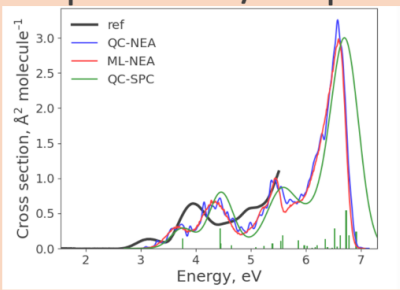
Heat of formation:

$$\Delta H_{\mathrm{at},T} = \left[\sum_A H_T(A)\right] - H_T$$

(Ro)vibrational spectra



One-photon UV/vis spectra



Properties & spectra

Simulations

Quantum dissipative dynamics

Two-photon absorption spectra

Pavlo O. Dral

P. O. Dral, F. Ge, Y.-F. Hou, P. Zheng, Y. Chen, M. Barbatti, O. Isayev, C. Wang, B.-X. Xue, M. Pinheiro Jr, Y. Su, Y. Dai, Y. Chen, S. Zhang, L. Zhang, A. Ullah, Q. Zhang, Y. Ou. *J. Chem. Theory Comput.* **2024,** *20*, 1193

XACS
Xiamen Atomistic Computing Suite
XACScloud.com

https://xacs.xmu.edu.cn/software/mlatom

MLatom.com

xacs.xmu.edu.cn/docs/mlatom-cn/

XACS Docs

XACS Docs | XACS Home | Products | Support | About | Cloud Computing

- XMVB
- XEDA
- MLatom
- MLatom-CN
- Cloud Computing
- Workshop-2023
- Toruń workshop 2023

MLatom

搜索文档

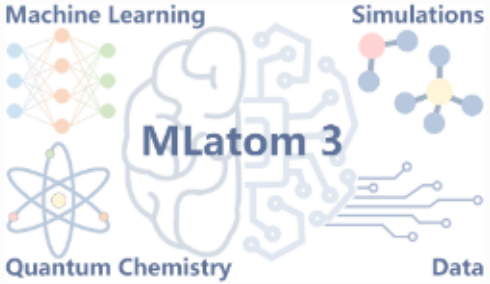
- 概览
  - 概览
  - 许可及引用
  - 新版本
  - 相关信息
  - 联系与支持
- 设置
  - 安装
  - 源代码
  - 使用
  - 在XACS云平台上运行
- 教程
  - 开始

欢迎来到MLatc

查看页面源码

欢迎来到M 档!

MLatom 3是一个程序包 大功能来增强典型的计算化学 模拟，并创建复杂的工作流程。该 开源 软件包为用 户可以使用命令行选项、输入文件或脚本(使用MLatom作为Python包)在其计算机上 行模拟 XACScloud.com 的计算服务。用户可以进行 能量 和 热化学性质、 几何优化、 分子动力学 和 量子动力学 模拟，并通过 ML、量子力学和组合模型 模拟 （旋转）振动光谱 、 单光子紫外/可见吸收光谱 和 双光子吸收光谱 模拟。用户可以从包含预训练ML模型和量子力学近似(如接近耦合簇精度的AIQM1)的广泛方法库中进行选择，还可以使用各种ML算法构建自己的模型。广泛使用多种先进的软件包和库的接口使MLatom具有巨大灵活性。



下面的视频总结了MLatom的各项功能

Pavlo O. Dral

dr-dral.com

Bond length in H_2

Experiment: 0.7414 Å

Quantum Chemistry: (FCI/aug-cc-pV6Z) 0.7415 Å, ~5 CPU-days

Machine learning (ML): ??? Å, time?

Pavlo O. Dral

Bond length in H_2

Experiment:	0.7414 Å
Quantum Chemistry: (FCI/aug-cc-pV6Z)	0.7415 Å, ~5 CPU-days
Machine learning (ML):	0.7415 Å, ~0.3 seconds

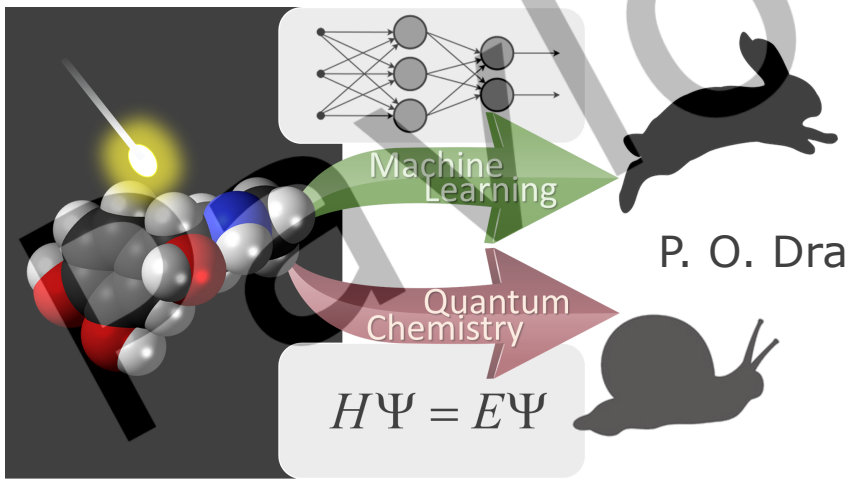
Pavlo O. Dral

Bond length in H_2

Experiment: 0.7414 Å

Quantum Chemistry: (FCI/aug-cc-pV6Z) 0.7415 Å, ~5 CPU-days

Machine learning (ML): 0.7415 Å, ~0.3 seconds

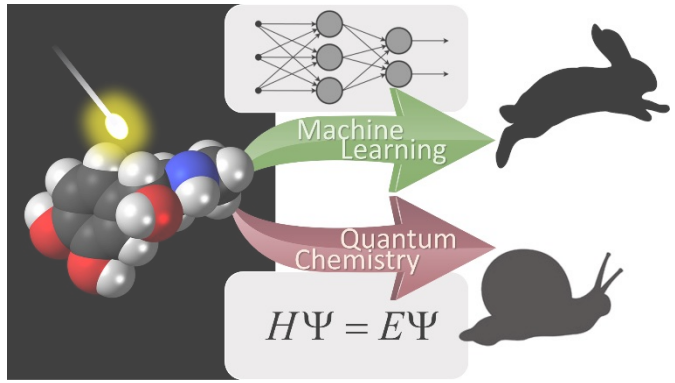


P. O. Dral, M. Barbatti, *Nat. Rev. Chem.* **2021,** *5*, 388

Pavlo O. Dral

Machine learning in quantum chemistry

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



Timing

Accuracy

Ab initio

ML-ab initio

DFT

ML-DFT

Semi-empirical

ML-SQC

Machine Learning

Molecular Mechanics

General-purpose (no training required):

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

ANI-1ccx ANI-1ccx: J. S. Smith, *et al*. *Nat. Commun.* **2019,** *10*, 2903

Pavlo O. Dral

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

Quantum chemistry approximations



Timing

Accuracy

Ab initio

CCSD(T) – gold standard, very slow (use it if you can!)

DFT

B3LYP – is much faster than the CCSD(T) but is much less accurate

Semi-empirical

Molecular Mechanics

Do not use B3LYP if you can use AIQM1!

AIQM1*

- **Accuracy of CCSD(T)**
- **Orders of magnitude faster than B3LYP**

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

*CHNO elements only – extensions on the way

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

Pavlo O. Dral

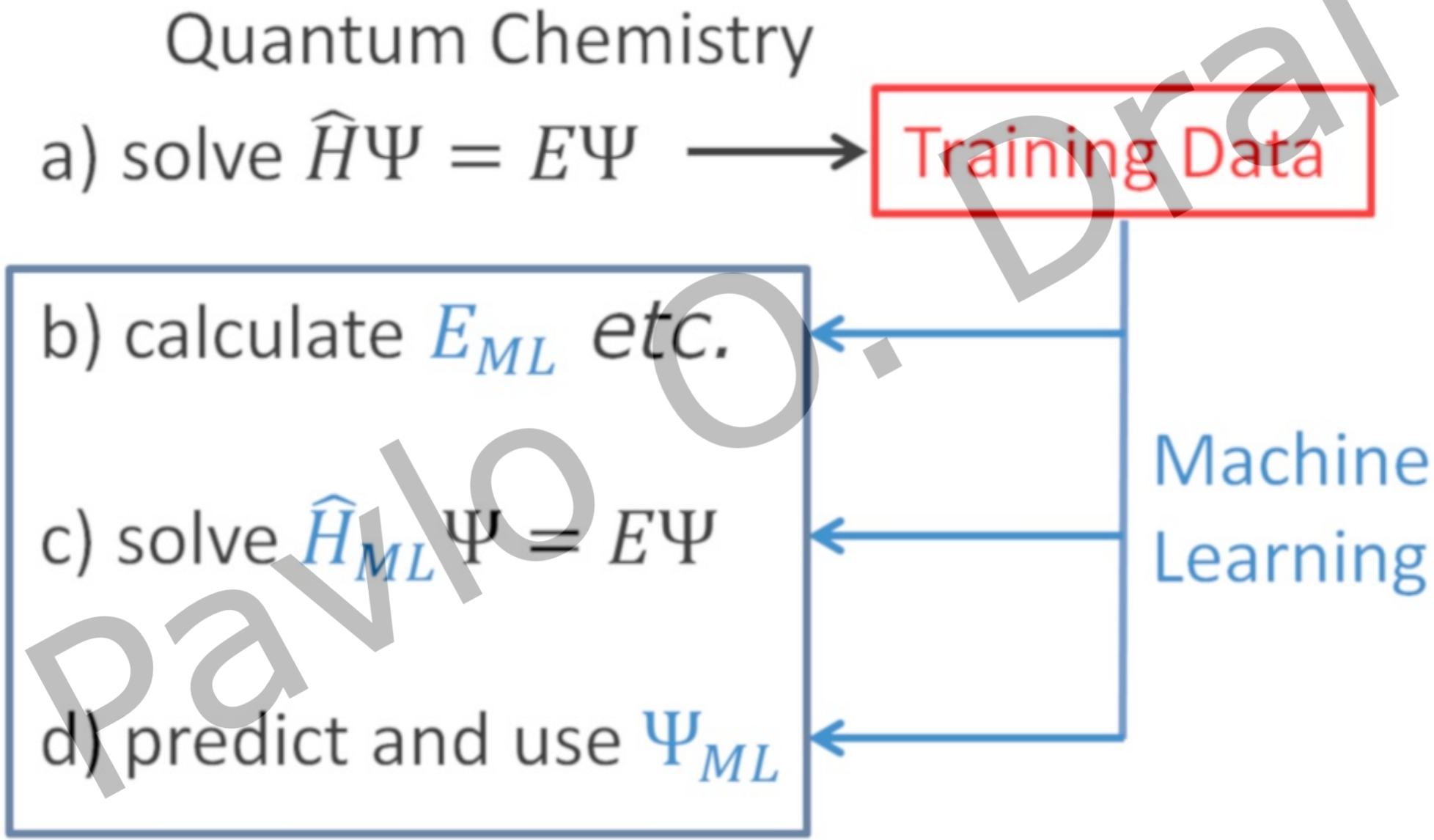
Machine learning

The name ‘Machine learning’ (ML) implies that machines try to learn from [Big] data by themselves **without being programmed explicitly** by humans.

ML is an application of a broader artificial intelligence (AI) field to practical problems.

Pavlo O. Dral

Machine learning in quantum chemistry



Pavlo O. Dral

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

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



Training data

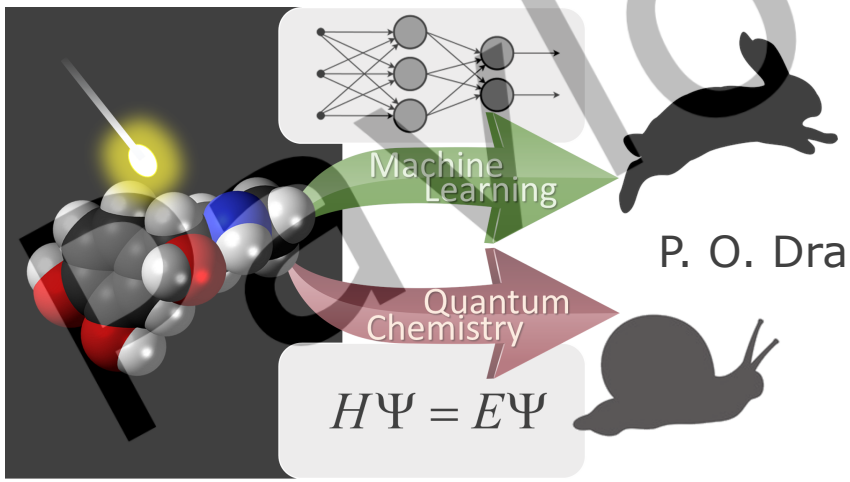
Pavlo O. Dral

Bond length in H_2

Experiment: 0.7414 Å

Quantum Chemistry: (FCI/aug-cc-pV6Z) 0.7415 Å, ~5 CPU-days

Machine learning (ML): 0.7415 Å, ~0.3 seconds



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

© Pavlo O. Dral

How to get such an ML model?

- data
  - energies at FCI: E_FCI_451.dat)
  - XYZ geometries (h2.xyz)
- ML model
  - KREG (train.inp)

Pavlo O. Dral

Task 1.4. Train the ML model for the H_2 molecule (see instructions below) and obtain the bond length with this model.

Questions:

- how much time did it take to train the model?
- what are the training and validation errors?
- what is the bond length of H_2 obtained with your model?

```
createMLmodel              # Specify the task for MLatom
MLmodelType=KREG           # Specify the model type
MLmodelOut=energies.unf # Save model in energies.unf
XYZfile=h2.xyz             # File with XYZ geometries
Yfile=E_FCI_451.dat        # The file with FCI energies
sigma=opt                  # Optimize hyperparameter sigma
lgSigmaL=-4                # Lower bound of log2(sigma)
lambda=opt                 # Optimize hyperparameter lambda
```

Send answers on WeChat!

10 min

Pavlo O. Dral

Homework 1.1. Calculate bond lengths and angle in H_2O as accurately as you can.

Questions:
1. method you used
2. bond lengths and angle you obtained
3. how much time it took to complete the calculations

Homework 1.2.
1. How to visualize xyz geometries (i.e., molecules in files .xyz as in the lectures)?
2. What are other molecular representation in addition to xyz coordinates?

Pavlo O. Dral