New Computational Chemistry Methods, Database and Challenges: an Overview

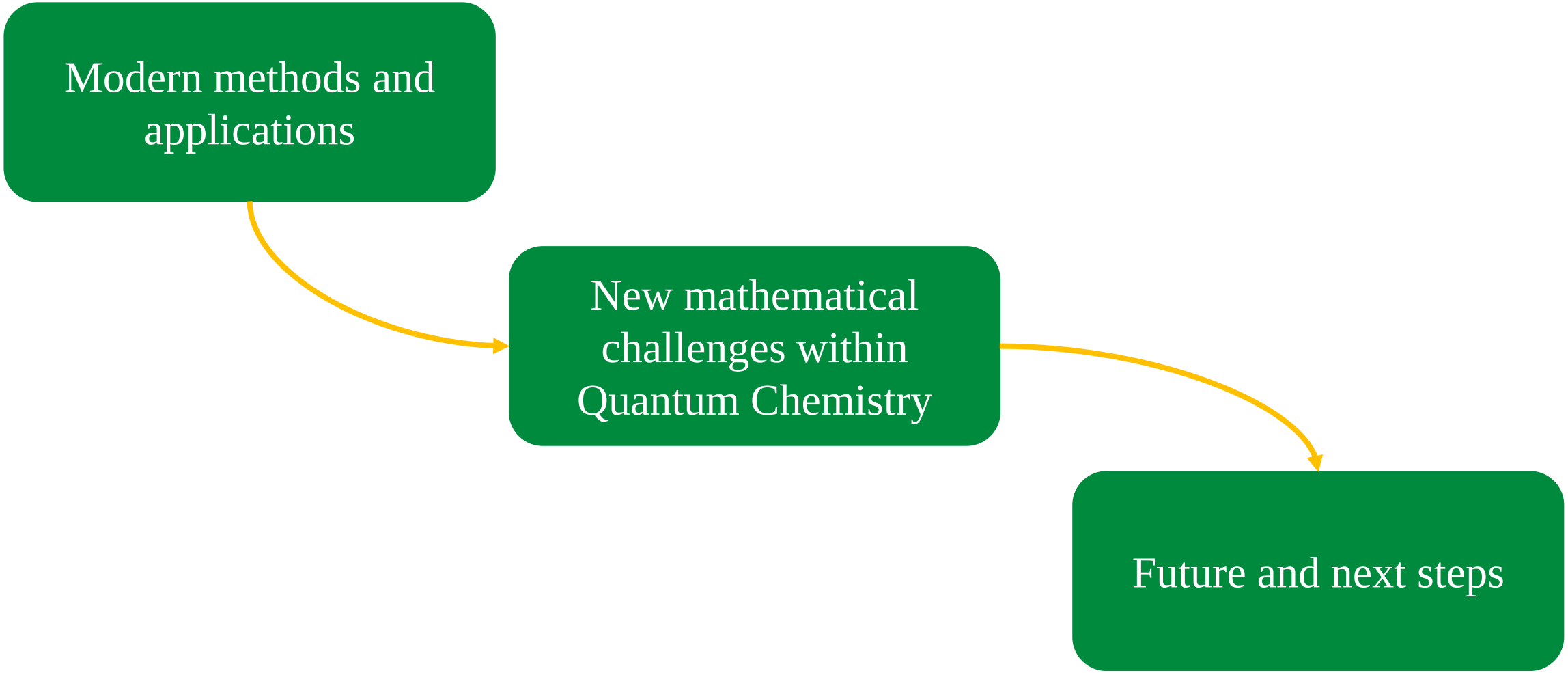
M.Sc. Student Victor Hugo Malamace da Silva
Supervisors: Dr. Phillip Choi and Dr. Stanislav R. Stoyanov

Early Stages and History

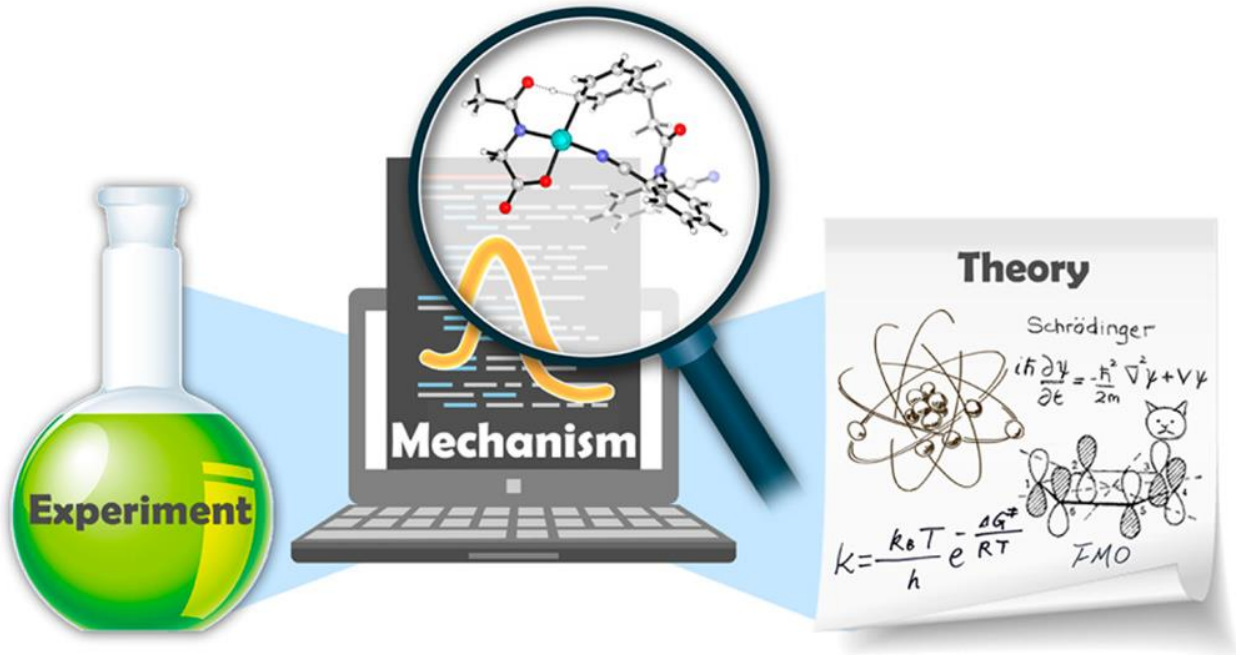
$$H\psi = E\psi$$

- Molecular DFT calculations;
- Molecular Dynamics;
- Solid State;
- Monte Carlo and conformation analysis;
- Transition State and reaction investigation;
- Docking and Protein Simulation;

New Perspective by Three Questions



Modern methods and applications



Application:

- Economical and environmental favorable approach;
- Several fast analysis and data processing;
- Increasing precision and accuracy, leading to a more reliable data;

Perspective:

- New approaches for quantum methods, with new approximations and machine learning;
- New motivations for the use of computational chemistry, based on industry demand or to enhance the knowledge of a chemical theory;

Modern methods and applications

- Investigation of mutagenicity and toxicity;
- Development of new methods to reduce the use of live animal;
- Computational chemistry as a inexpensive, fast and safer approach for this kind of analysis;
- Use of Hard and Soft, Acid and Base theory to study the toxicity;

$$\eta = [E_{LUMO} - E_{HOMO}]/2 \quad \text{(Hardness)}$$

$$\sigma = 1/\eta \quad \text{(Softness)}$$

$$\mu = [E_{LUMO} + E_{HOMO}]/2 \quad \text{(Chemical Potential)}$$

$$\omega = \mu^2/2\eta \quad \text{(Electrophilic index)}$$

$$\omega^- = \eta_A(\mu_A - \mu_B)^2/2(\eta_A + \eta_B)^2] \quad \text{(Nucleophilicity index)}$$

Chemical Research in Toxicology

pubs.acs.org/crt Persp

Density Functional Theory in the Prediction of Mutagenicity: A Perspective

Piers A. Townsend and Matthew N. Grayson*

Modern methods and applications

- Electron Ionization Mass Spectra;
- Development of new methods to increase the accuracy of the computational methods;
- Application on forensic drug investigation, pharmacokinetics and doping analysis;
- May be beneficial for a machine learning approach;

Wang *et al. J Cheminform* (2020) 12:63
https://doi.org/10.1186/s13321-020-00470-3

Journal of Cheminformatics

RESEARCH ARTICLE Open Access

Predicting in silico electron ionization mass spectra using quantum chemistry

Shunyang Wang[1,2], Tobias Kind[1], Dean J. Tantillo[2] and Oliver Fiehn[1*]

ACS OMEGA

Article

Cite This: *ACS Omega* 2019, 4, 15120−15133

http://pubs.acs.org/journal/acsodf

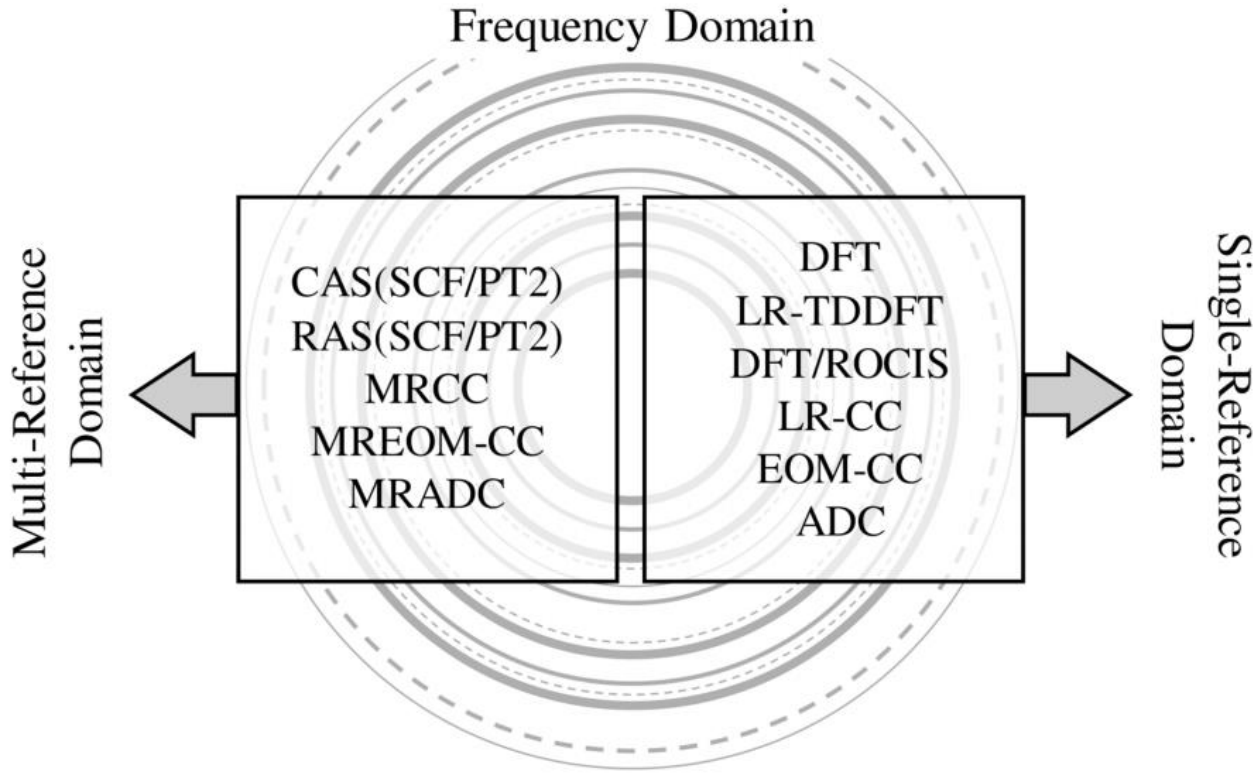
Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB Methods

Jeroen Koopman and Stefan Grimme*

Mulliken Center for Theoretical Chemistry, Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany

Modern methods and applications

- X-ray Spectroscopy;
- Challenges on core-excited states and electric field;
- Use of DFT calculation as cost-effectively approach to determine core-excitation energies;

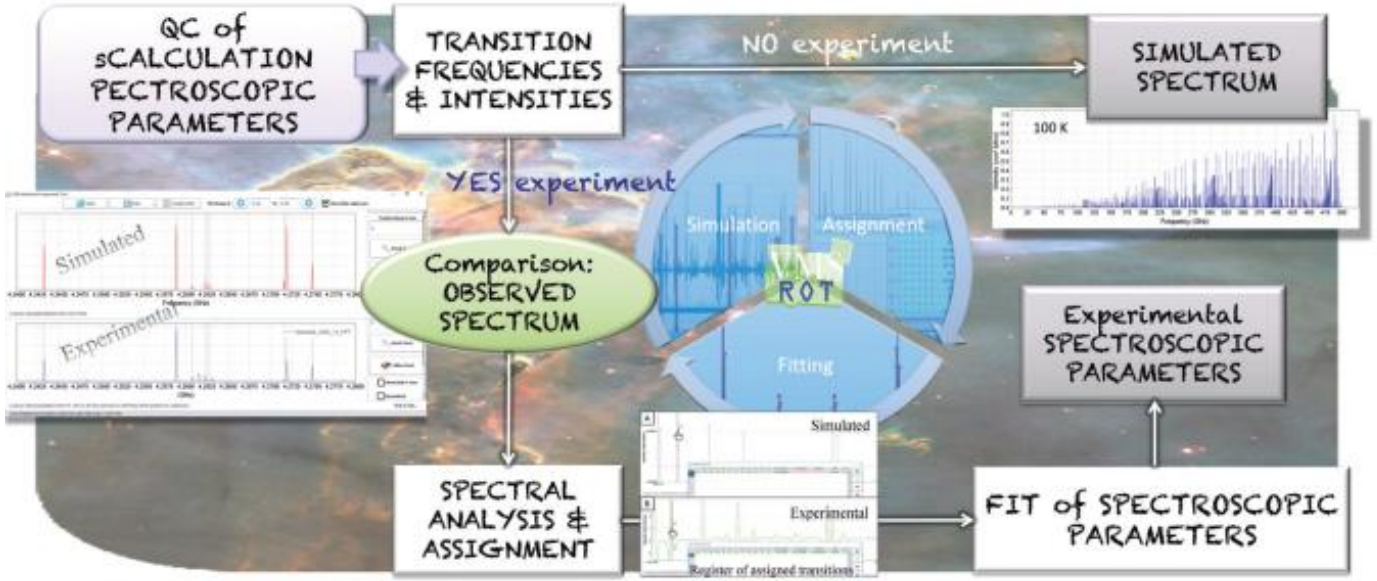


Progress in the Theory of X-ray Spectroscopy: From Quantum Chemistry to Machine Learning and Ultrafast Dynamics

C. D. Rankine* and T. J. Penfold*

Modern methods and applications

- Astrochemistry;
- Use of quantum chemistry for spectroscopy analysis of interstellar complex organic molecules;
- Increase the knowledge of well-known reaction routes and chemical properties;



The challenging playground of astrochemistry: an integrated rotational spectroscopy – quantum chemistry strategy

Cristina Puzzarini *[a] and Vincenzo Barone [b]

New mathematical challenges within Quantum Chemistry

- New approach:
  - Non Born-Oppenheimer approximation;
  - Nuclei treated the same way as electrons (Nuclei-electron Orbital);
  - Improvement on the multicomponent wave function and non equilibrium calculations;

- Geometry and energy;
- Excited State properties;
- Nonadiabatic dynamics;

Nuclear–electronic orbital methods: Foundations and prospects

Cite as: J. Chem. Phys. 155, 030901 (2021); doi: 10.1063/5.0053576
Submitted: 8 April 2021 • Accepted: 17 May 2021 •
Published Online: 15 July 2021

View Online Export Citation CrossMark

Sharon Hammes-Schiffer[a)]

AFFILIATIONS

Department of Chemistry, Yale University, New Haven, Connecticut 06520, USA

New mathematical challenges within Quantum Chemistry

- Nuclear - electronic orbital (NEO) and Hartree-Fock method;

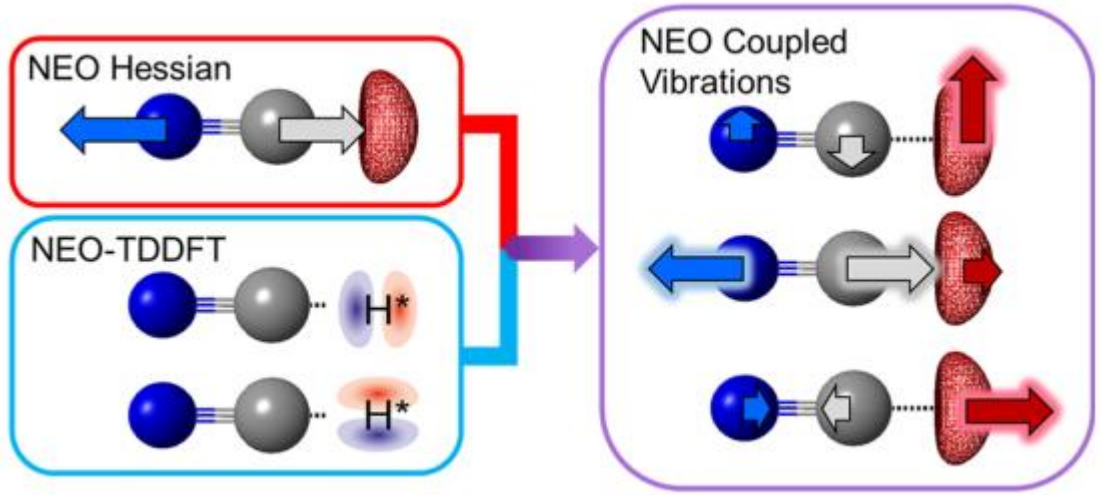
$$\mathbf{F}^{e}\mathbf{C}^{e} = \mathbf{S}^{e}\mathbf{C}^{e}\boldsymbol{\varepsilon}^{e},$$

$$\mathbf{F}^{p}\mathbf{C}^{p} = \mathbf{S}^{p}\mathbf{C}^{p}\boldsymbol{\varepsilon}^{p}$$

  - F = Fock matrix;
  - C = Orbital Coefficient matrix;
  - S = Overlap matrix;
  - ε = Orbital Energy Matrix;
- Strongly coupled;
- NEO energy path provides may be useful to track the movement of protons on a reaction;
- Development of NEO-DFT;

$$E[\rho^{e},\rho^{p}] = E_{ext}[\rho^{e},\rho^{p}] + E_{ref}[\rho^{e},\rho^{p}] + E_{exc}[\rho^{e}] + E_{pxc}[\rho^{p}] + E_{epc}[\rho^{e},\rho^{p}],$$

- Limitations:
  - Time dependent DFT and nonadiabatic calculation are still a challenge for this kind of approach;
  - Computational cost due to less approximation;
- One alternative to a better implementation to this method is the use of Machine Learning;



New mathematical challenges within Quantum Chemistry

- New mathematical approach using more than one basis set;
- Alternative to wavefunction-based methods and improvement to DFT calculations;
- Better performance when combining triple-zeta basis sets;

$$E_{2B} = E(\text{DFT/B1}) + c_1\,[E(\text{DFT/B2}) - E(\text{DFT/B1})\,] + E_{SO}$$

c_1 – coefficient determinend by minimizing the mean unsigned error;

E_{SO} – spin - orbital Energy

Multicoefficient Density Functional Theory (MC−DFT)

Jien-Lian Chen, Yi-Lun Sun, Kuo-Jui Wu, and Wei-Ping Hu*

Department of Chemistry and Biochemistry, National Chung Cheng University, Chia-Yi 621, Taiwan

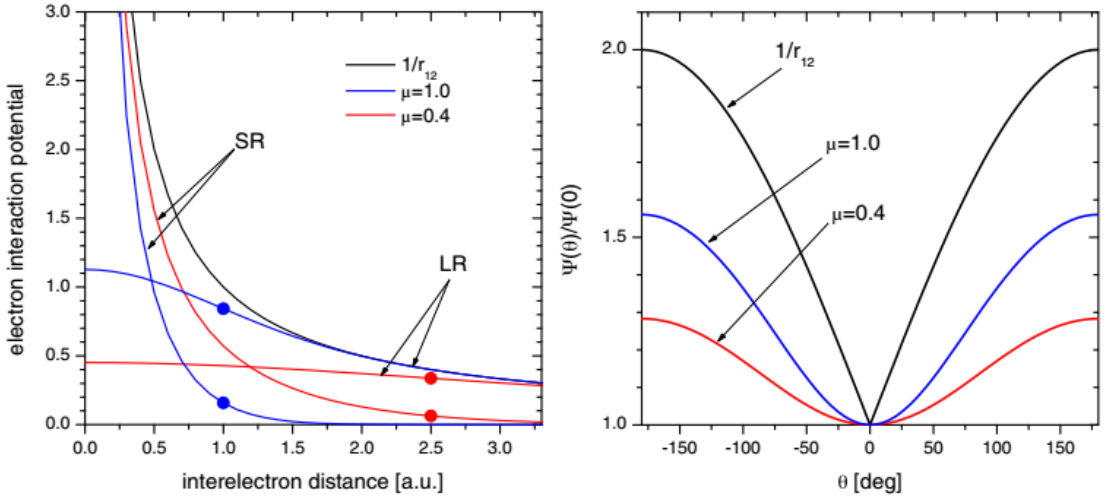
Received: July 26, 2007; In Final Form: October 1, 2007

New mathematical challenges within Quantum Chemistry

- Range-separated MC-DFT presents several benefits;
- Combination of density function methods and wave function methods for a better description;
- Use of the methods is range dependent;
- Improvement to long-range interaction and spin-symmetry problems;

$$\hat{H}^{LR} = \hat{T} + \hat{V}_{ne} + \hat{V}_{ee}^{LR} + \hat{V}^{SR}[\rho_{\Psi^{LR}}]$$

- LR – Long range; SR – Short Range;



$$\mu = \frac{R_e}{2}$$

$\mu \rightarrow range\ cutoff\ parameter$

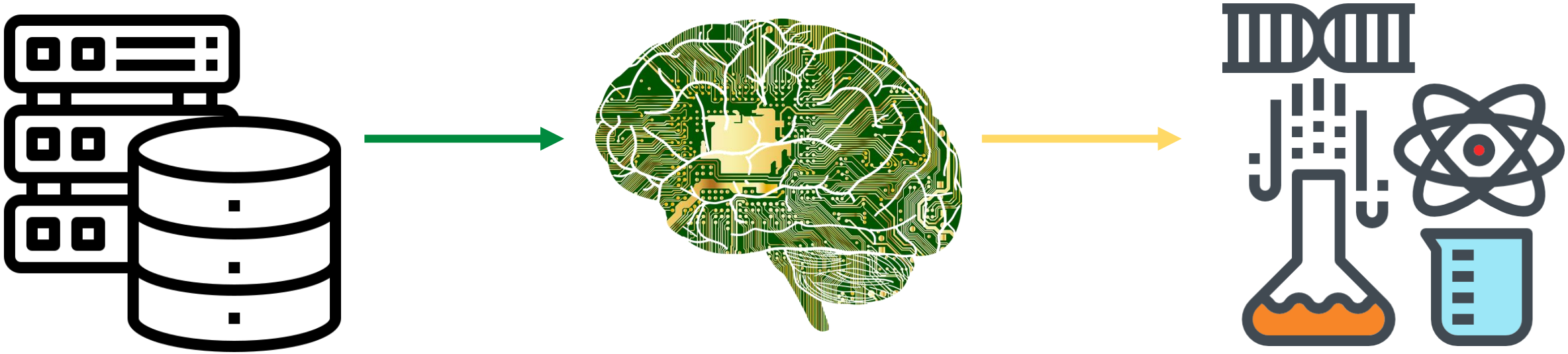
$R_e \rightarrow interatomic\ distance$

Range-separated multiconfigurational density functional theory methods

Katarzyna Pernal[1] | Michał Hapka[1,2]

Machine Learning and Database

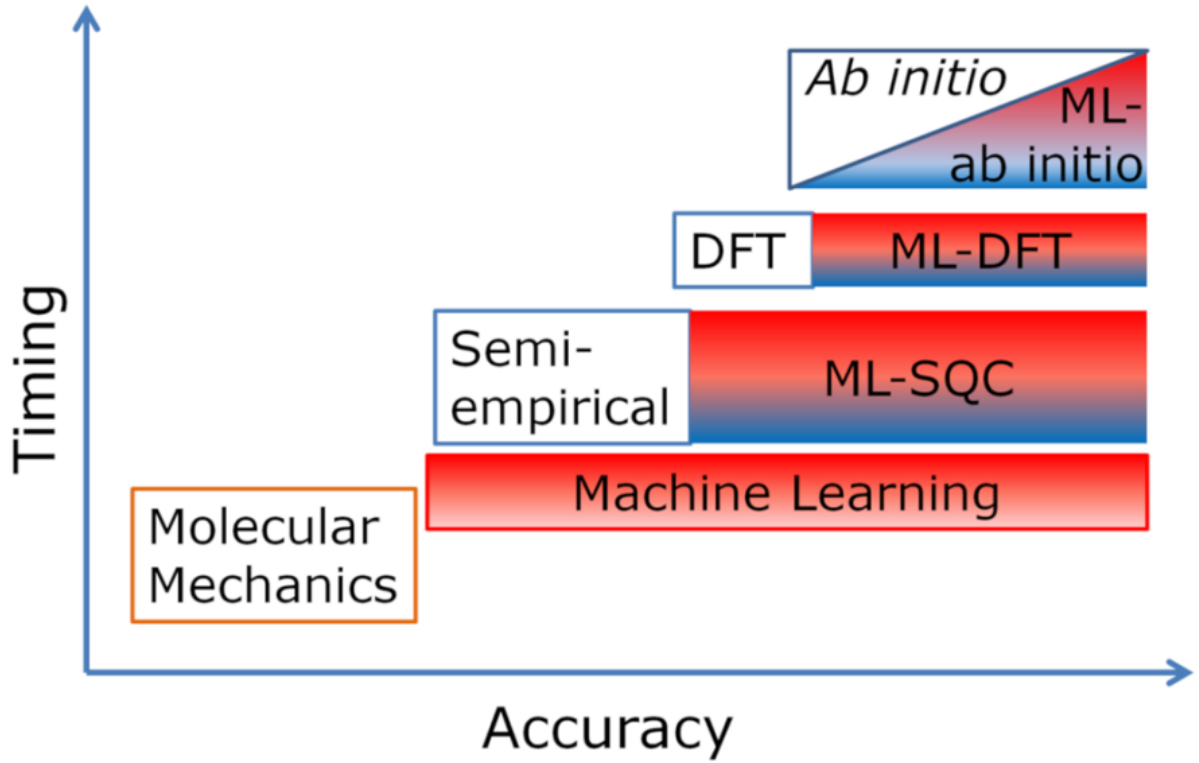
- Required for the Machine Learning approach;
- Improve the knowledge on properties behaviour;
- How to approach the data provided;



Future and next steps

Machine Learning (ML):

- Main modern advance for quantum chemistry;
- Supervised ML - high demand of data;
- Wide range of application and increasing accuracy;



THE JOURNAL OF PHYSICAL CHEMISTRY Letters

pubs.acs.org/JPCL Perspective

Quantum Chemistry in the Age of Machine Learning

Pavlo O. Dral*

Future and next steps

Machine Learning (ML):

- Machine Learning as a strategy to enhance the structure simulation;
- GEOM database for conformers;
- Special challenge for polymers and their size;
- Less expensive approach to calculate excited-state;
- Limitations:
  - Must incorporate data, physical principles and constrains;
  - Amount of data – requiring thousands or millions of points;

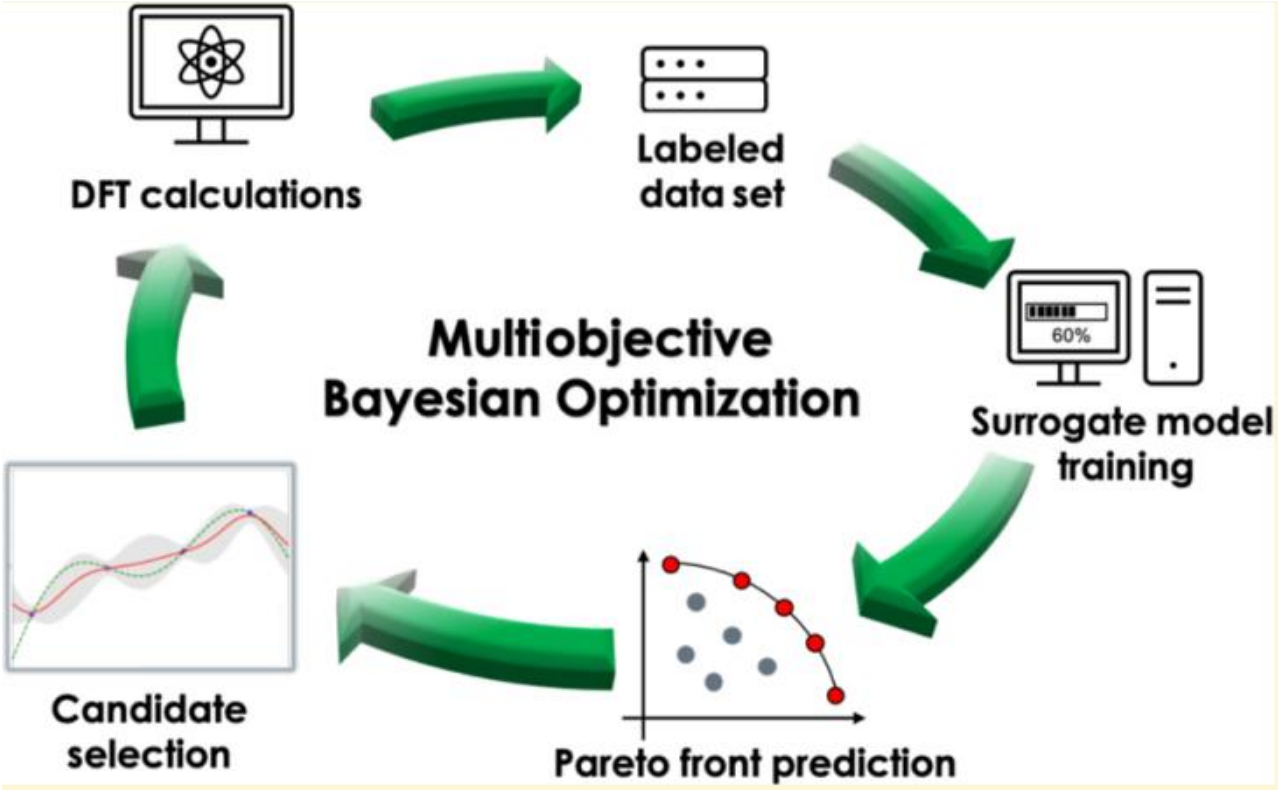
Learning Matter: Materials Design with Machine Learning and Atomistic Simulations

Simon Axelrod, Daniel Schwalbe-Koda, Somesh Mohapatra, James Damewood, Kevin P. Greenman, and Rafael Gómez-Bombarelli*

Future and next steps

Multiobjective Bayesian Optimization (MBO):

- Aim to locate better Redox flow batteries materials;
- MBO as an approach of Machine Learning to analyze the data;
- Provide a substantial amount of data for graphs analysis;



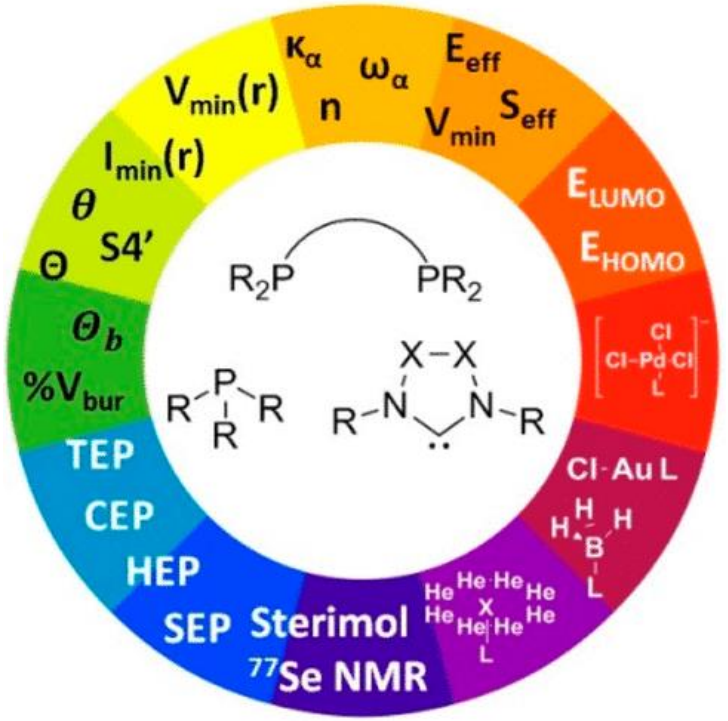
Discovery of Energy Storage Molecular Materials Using Quantum Chemistry-Guided Multiobjective Bayesian Optimization

Garvit Agarwal, Hieu A. Doan, Lily A. Robertson, Lu Zhang, and Rajeev S. Assary*

Future and next steps

Ligand Knowledge Base:

- Building a database of descriptors;
- Enhance the activity of organometallics catalysts;
- Statistical prediction and analysis;
- Data-led approach, with experimental and computational data combined, for high-quality response;



Building a Toolbox for the Analysis and Prediction of Ligand and Catalyst Effects in Organometallic Catalysis

Published as part of the Accounts of Chemical Research special issue "Data Science Meets Chemistry".

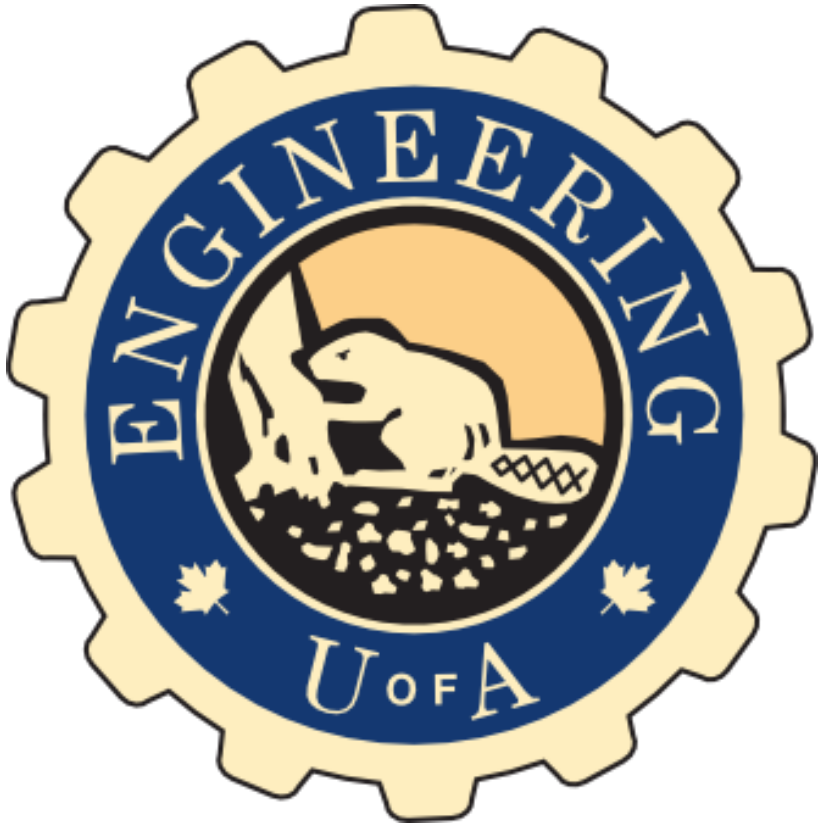
Derek J. Durand and Natalie Fey*

Final Remarks

- Computational chemistry seen as a reliable, cheap and greener approach to chemical problems;
- New methods and application are alligned with the improvement on the hardware and methods;
- Machine Learning as the next step for a more accurate calculation and faster;
- Databases containing different properties and parameters are now required for this type of approach;

The next step?

- Quantum Computers and the new age of calculation;



Thank you!