

Reliable approximation in electronic structure methods: from real space partitioning to quantum Monte Carlo approaches

Amanda Dumi
Seminar at Oak Ridge National Laboratory
8/17/2022

When is DFT not enough?

- DFT citation/publications demonstrate it's applicability and often reliable performance
- In strongly correlated regimes, or when functional and system physics do not agree

Reduce cost through various system reductions

Sampling Configurations

Quantum Monte Carlo:
applications and
developments

Hilbert Space

selected CI as approximate
trial wave functions

Real Space

Determining molecular
fragments with
unsupervised machine
learning (UML)

Reduce cost through various system reductions

Sampling Configurations

Quantum Monte Carlo:
applications and
developments

Hilbert Space

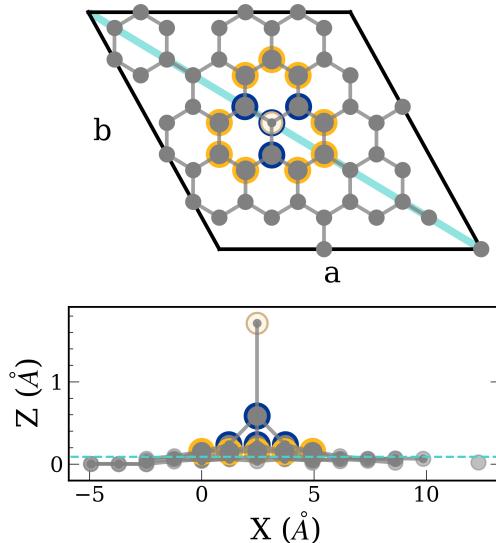
selected CI as approximate
trial wave functions

Real Space

Determining molecular
fragments with
unsupervised machine
learning (UML)

QMC application: periodic absorption¹

H on graphene



Method	Binding energy (meV)
DMC	-691 ± 19
PW91	-810 to -830, -870
PBE	-790, -840, -980

- functional dependent binding energy
- a need for benchmark values

1. A. Dumi et al, “The Binding of Atomic Hydrogen on Graphene from Density Functional Theory and Diffusion Monte Carlo Calculations,” *The Journal of Chemical Physics*, 156.14 (2022), 144702 DOI:10.1063/5.0085982 .

Diffusion Monte Carlo

- Recast the Schrodinger equation in imaginary time

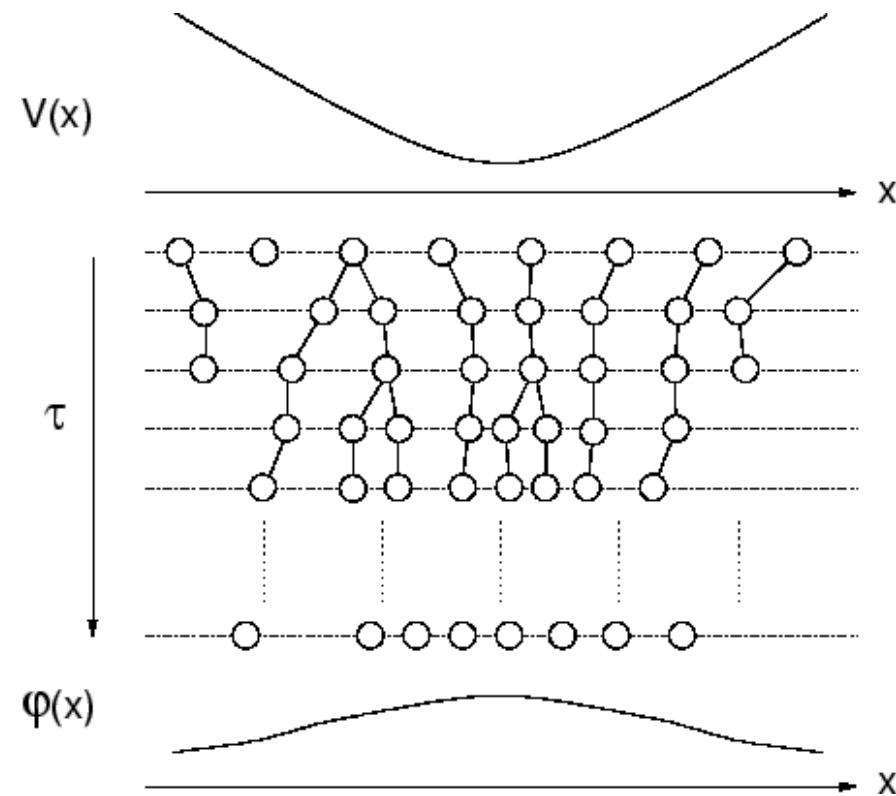
$$\frac{\partial |\Psi\rangle}{\partial \tau} = -\hat{H} |\Psi\rangle$$

- A formal solution to this:

$$|\psi(\tau_1 + \delta\tau)\rangle = e^{-\hat{H}\delta\tau} |\psi(\tau_1)\rangle$$

Anything non-orthogonal to the ground state decays exponentially

$$\lim_{\tau \rightarrow \infty} |\psi(\tau)\rangle = c_0 e^{-\epsilon_0 \tau} |\phi_0\rangle$$



$$-\frac{\partial \psi(\mathbf{R}, \tau)}{\partial \tau} =$$

$$\left[\sum_{i=1}^N -\frac{1}{2} \nabla_i^2 \psi(\mathbf{R}, \tau) \right] \text{ diffusion term}$$

+

$$(V(\mathbf{R}) - E_T) \psi(\mathbf{R}, \tau) \text{ branching term}$$

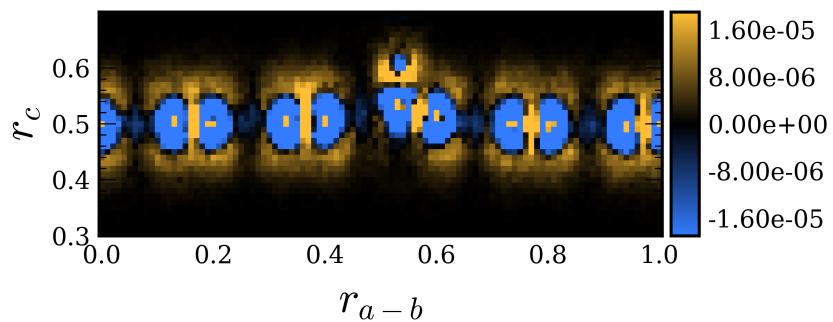
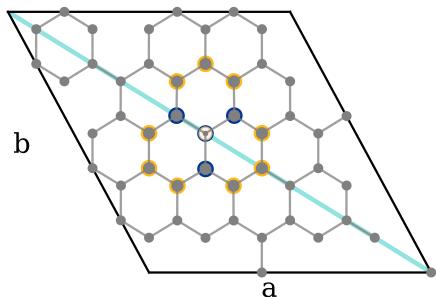
Importance sampling

- use of trial wavefunction for efficient sampling.

Fixed-node approximation

- Antisymmetry of fermions causes sampling problems
- Solution: fix nodes of trial wavefunction
- necessitates accurate nodal surface of trial wavefunction

QMC application: periodic absorption H on graphene



Method	Binding energy (meV)
PBE	-820
PBE	-871
PBE0	-851 (-800)
HSE	-794 (-743)
DMC	-691 ± 19
PW91	-810 to -830, -870
PBE	-790, -840, -980

- QMCPACK binding energies provide benchmark.
- Hybrid functionals are very close.
- Even with close binding energies, density seems to disagree in bonding region.

Reduce cost through various system reductions

Sampling Configurations

Quantum Monte Carlo:
applications and
developments

Hilbert Space

selected CI as approximate
trial wave functions

Real Space

Determining molecular
fragments with
unsupervised machine
learning (UML)

Reduce cost through various system reductions

Sampling Configurations

Quantum Monte Carlo:
applications and
developments

Hilbert Space

selected CI as approximate
trial wave functions

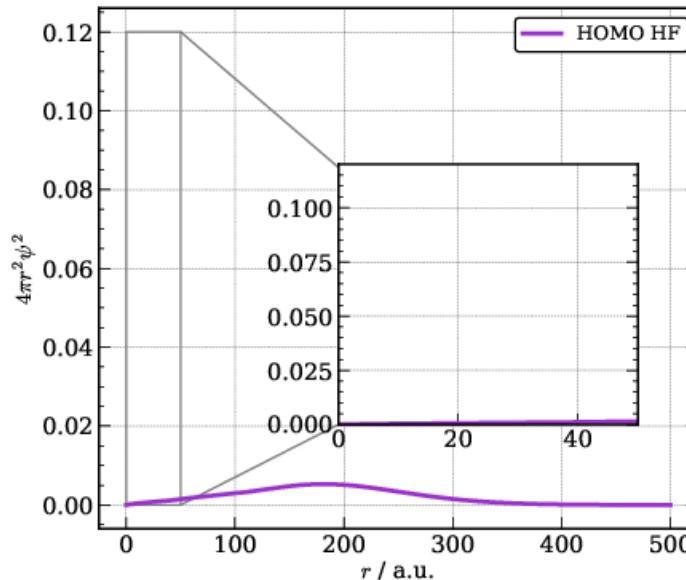
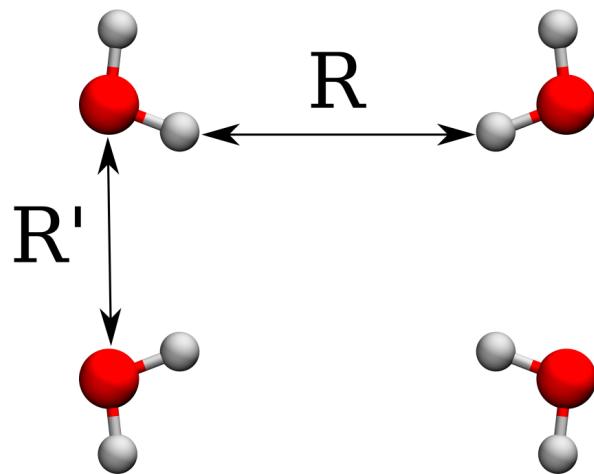
Real Space

Determining molecular
fragments with
unsupervised machine
learning (UML)

QMC application: non-valence correlation bound anion¹

NVCB rely on accurate description of correlation to bind

$$E_{corr} = E_{true} - E_{HF}$$



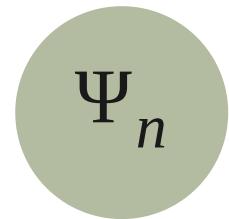
1. S. Upadhyay et al, “The Role of High-Order Electron Correlation Effects in a Model System for Non-Valence Correlation-Bound Anions,” *The Journal of Chemical Physics*, 153.22 (2020), 224118 DOI:10.1063/5.0030942 .

Selected-Cl

$$\Psi_n = \sum c_I^{(n)} |D_I\rangle$$

Selected-Cl

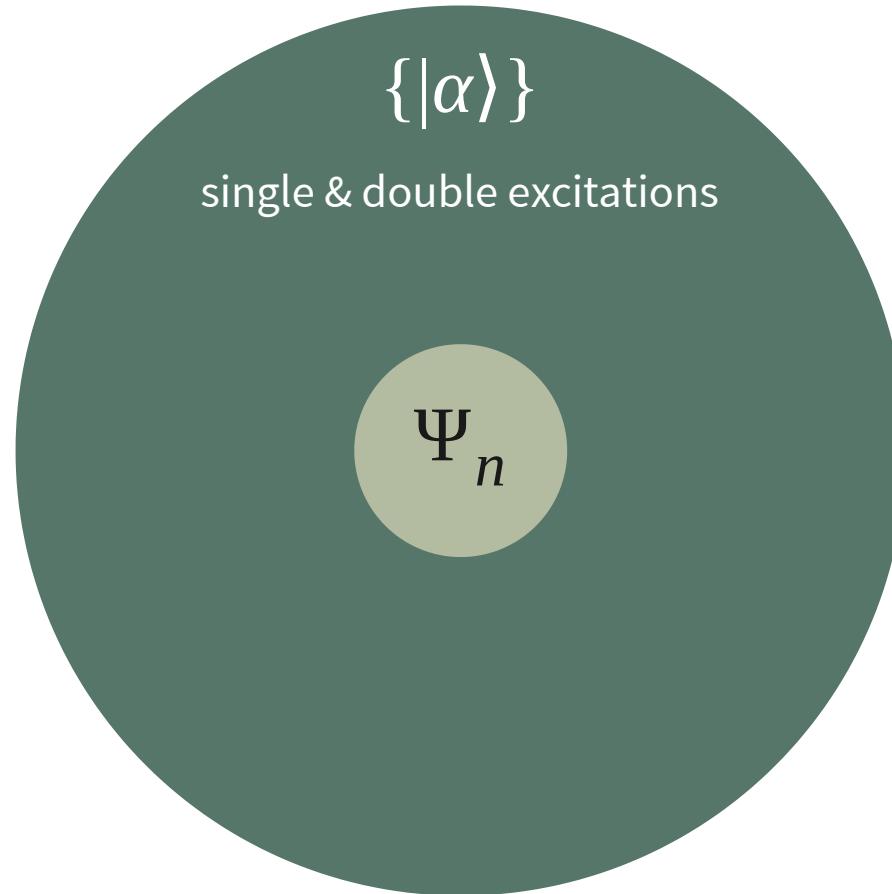
$$\Psi_n = \sum c_I^{(n)} |D_I\rangle$$


$$\Psi_n$$

Selected-Cl

$$\Psi_n = \sum c_I^{(n)} |D_I\rangle$$

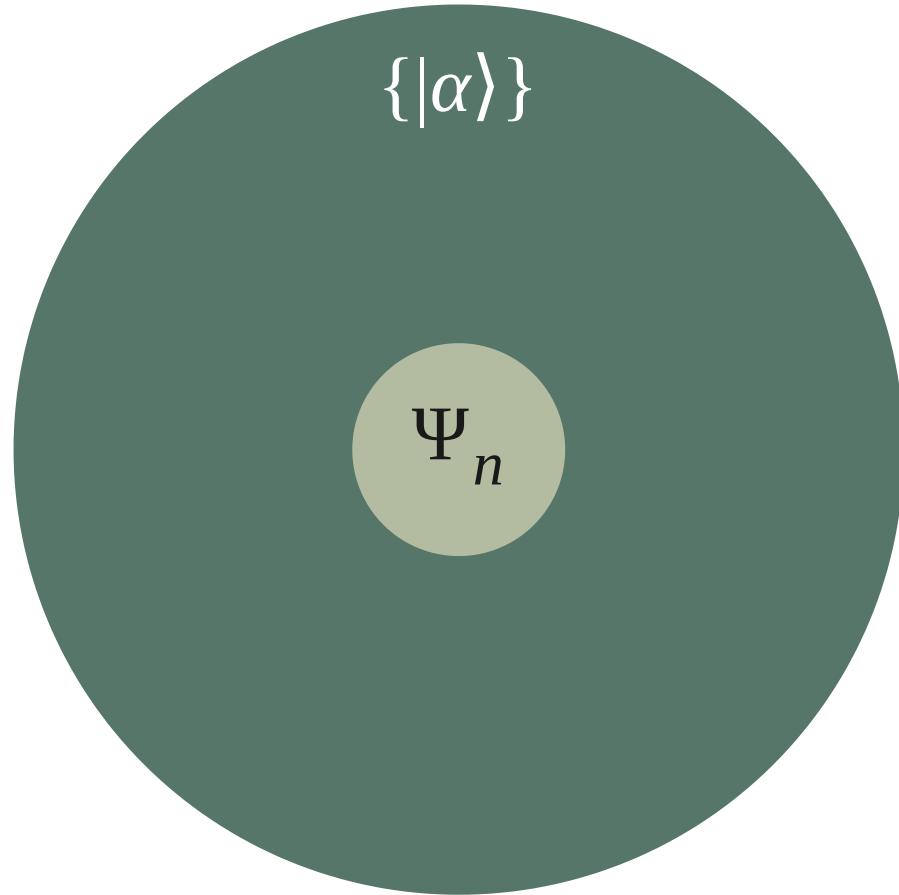
$$e_\alpha = \frac{\langle \Psi^{(n)} | \hat{H} | \alpha \rangle^2}{E^{(n)} - \langle \alpha | \hat{H} | \alpha \rangle}.$$



Selected-Cl

$$\Psi_n = \sum c_I^{(n)} |D_I\rangle$$

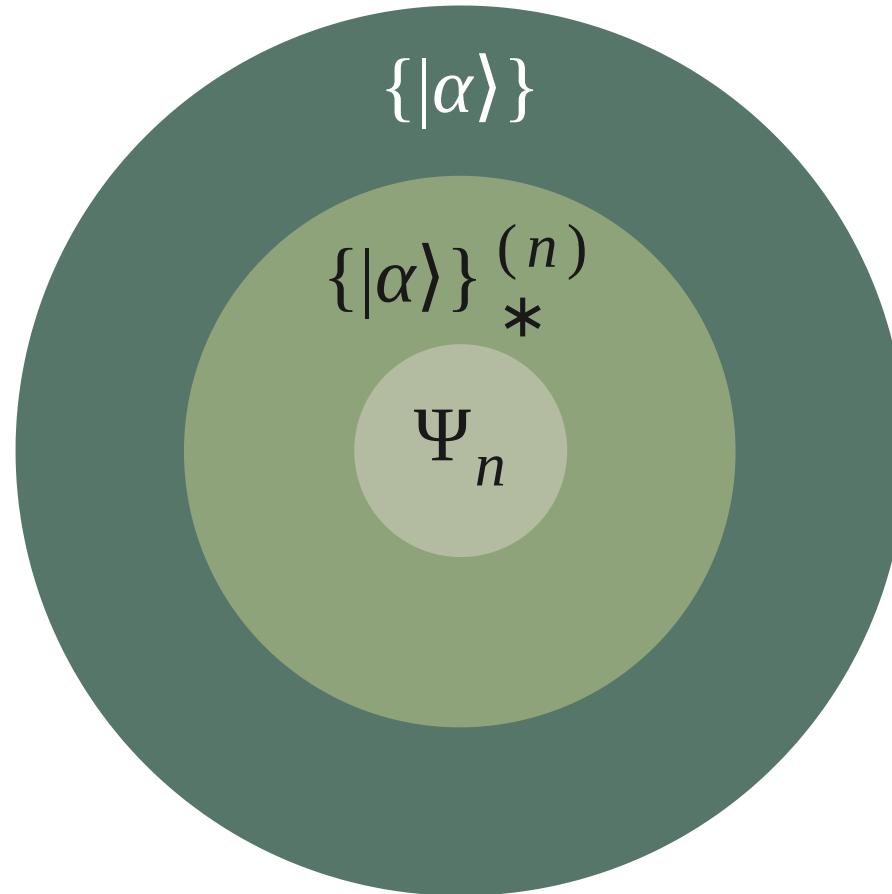
$$e_\alpha = \frac{\langle \Psi^{(n)} | \hat{H} | \alpha \rangle^2}{E^{(n)} - \langle \alpha | \hat{H} | \alpha \rangle}.$$



Selected-Cl

$$\Psi_n = \sum c_I^{(n)} |D_I\rangle$$

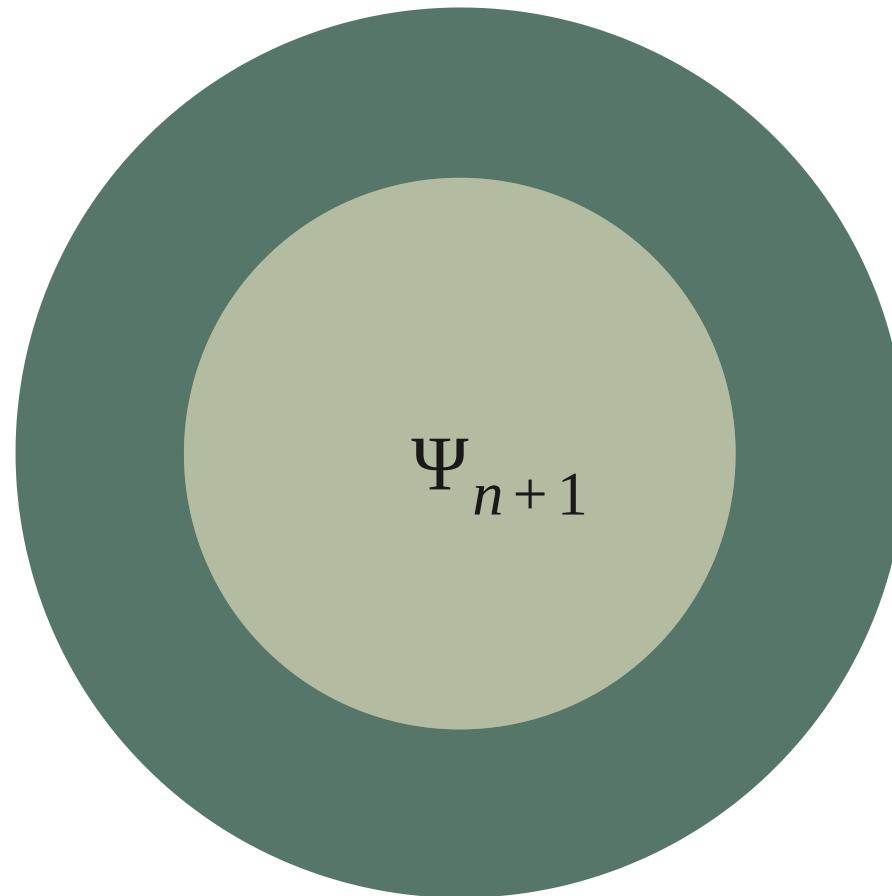
$$e_\alpha = \frac{\langle \Psi^{(n)} | \hat{H} | \alpha \rangle^2}{E^{(n)} - \langle \alpha | \hat{H} | \alpha \rangle}.$$



Selected-Cl

$$\Psi_n = \sum c_I^{(n)} |D_I\rangle$$

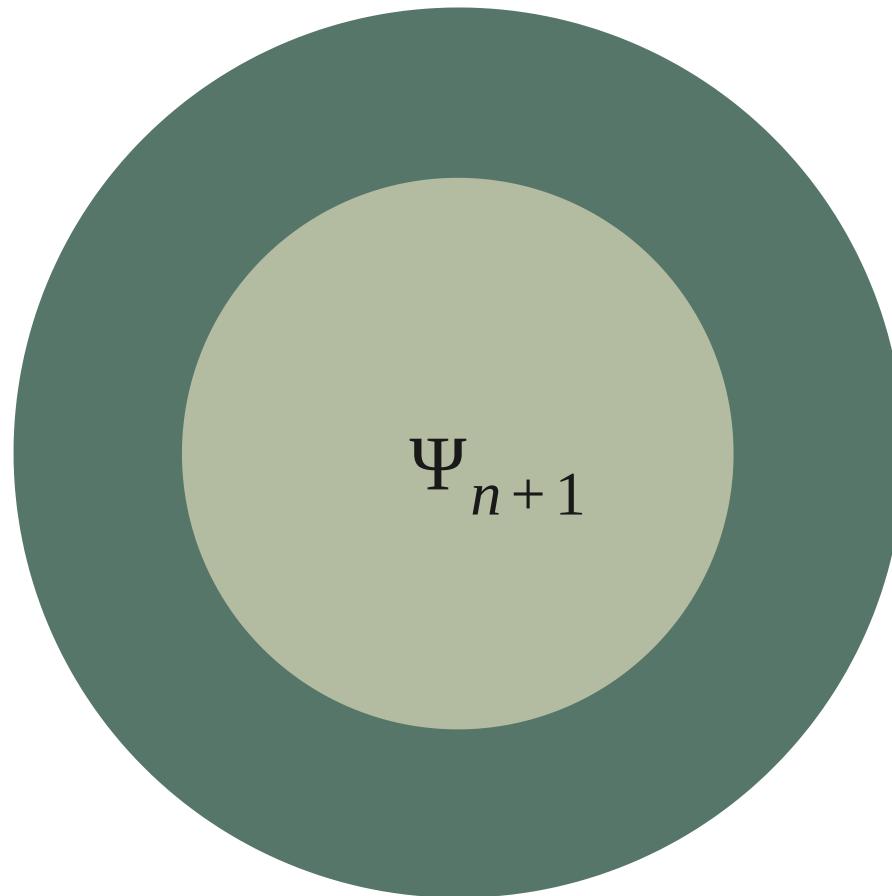
$$e_\alpha = \frac{\langle \Psi^{(n)} | \hat{H} | \alpha \rangle^2}{E^{(n)} - \langle \alpha | \hat{H} | \alpha \rangle}.$$



Selected-Cl

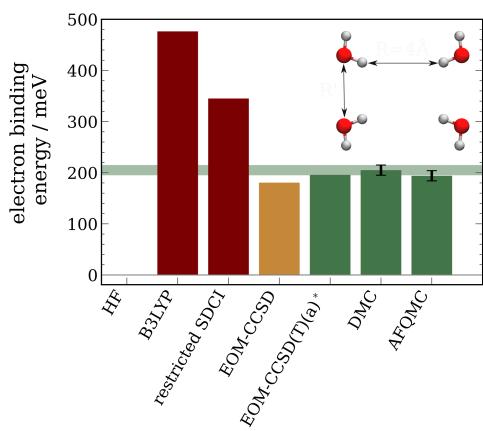
$$\Psi_n = \sum c_I^{(n)} |D_I\rangle$$

$$e_\alpha = \frac{\langle \Psi^{(n)} | \hat{H} | \alpha \rangle^2}{E^{(n)} - \langle \alpha | \hat{H} | \alpha \rangle}.$$



$(\text{H}_2\text{O})_4$ Results

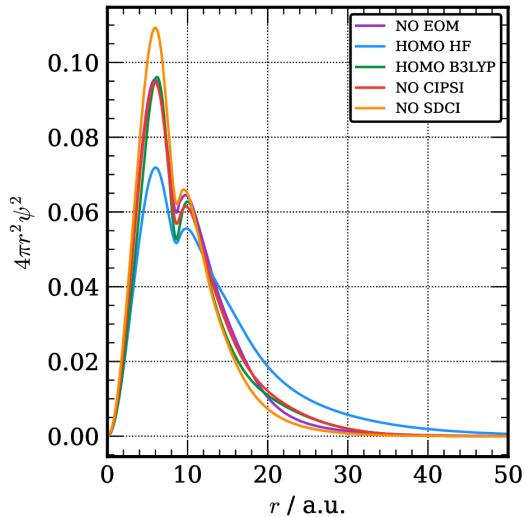
QMC for weak e^- binding



wave function	basis set	EBE (meV)
D/HF	aug-cc-pVDZ+7s7p	183 ± 10
SD/HF	aug-cc-pVDZ	176 ± 12
SD/HF	cc-pVDZ	-528 ± 25
SD/B3LYP	aug-cc-pVDZ+7s7p	212 ± 11
SD/HF(N)//SD/NO SDCI(A)	aug-cc-pVDZ+7s7p	205 ± 10
SD/HF(N)//MD/NO SDCI(A)	aug-cc-pVDZ+7s7p	202 ± 12
MD/CIPSI NO	aug-cc-pVDZ+3s1p	190 ± 9

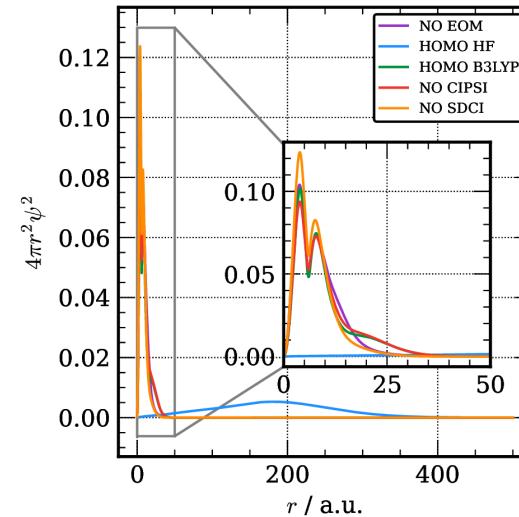
- Summary of selected-CI success/challenges and overall performance of other methods.

$R = 7$



- HF binds electron correlation not essential
- still orbital shape has noticeable changes
- DFT can recover those changes.

$R = 4$



- NCVB regime
- HF does not bind
- CIPSI shape changes slight, impacts energy.
- SDCI shape seems wrong

Reduce cost through various system reductions

Sampling Configurations

Quantum Monte Carlo:
applications and
developments

Hilbert Space

selected CI as approximate
trial wave functions

Real Space

Determining molecular
fragments with
unsupervised machine
learning (UML)

Reduce cost through various system reductions

Sampling Configurations

Quantum Monte Carlo:
applications and
developments

Hilbert Space

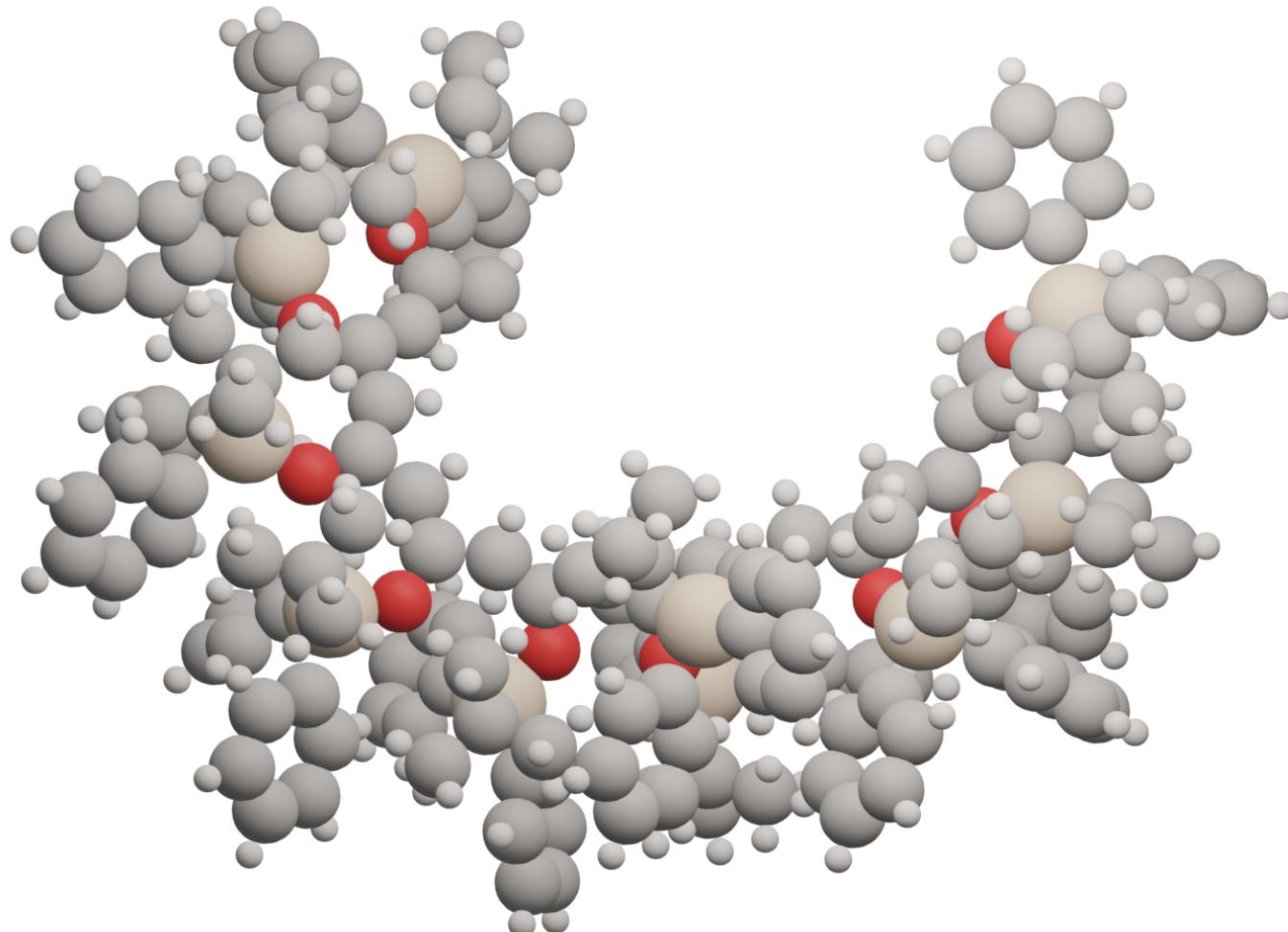
selected CI as approximate
trial wave functions

Real Space

Determining molecular
fragments with
unsupervised machine
learning (UML)

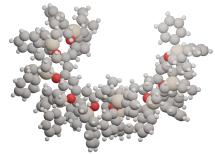
Fragmenting with unsupervised machine learning

Problem

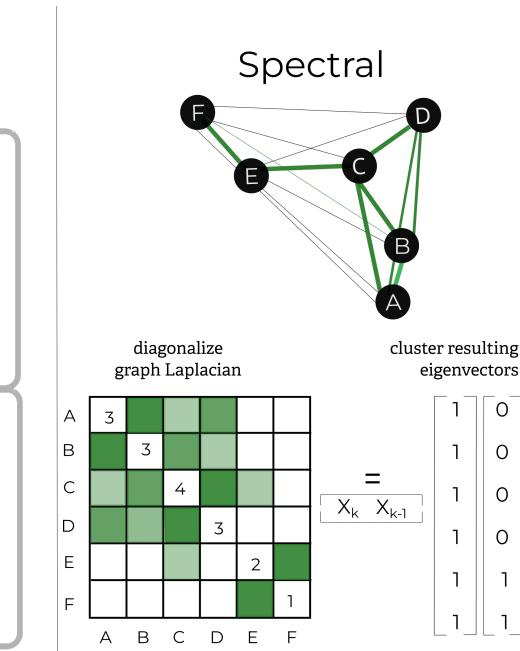
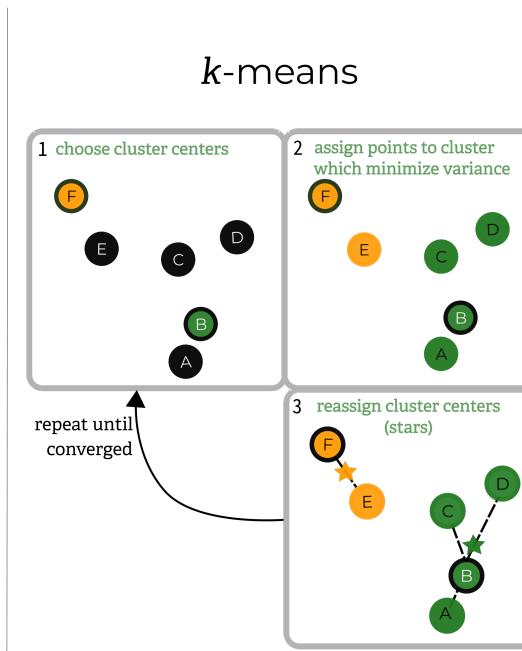
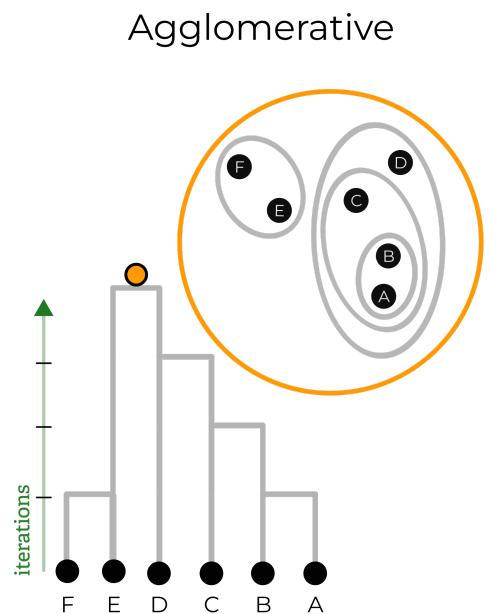


Fragmenting with unsupervised machine learning

Problem

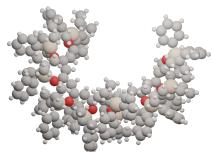


Approach



Fragmenting with unsupervised machine learning

Problem



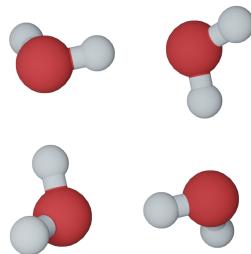
Approach

	structure only	electronic structure derived
coarse graining information	Cartesian $D_{AB} = \frac{1}{R_{AB}}$	Mayer bond matrix (Mbm) $B_{AB}^{\text{Mayer}} = \sum_s^{\text{on A}} \sum_t^{\text{on B}} (\mathbf{PS})_{st} (\mathbf{PS})_{ts}$
	covalent radii (cr) $D_{AB} = \begin{cases} 1 & \text{if } R_{AB} \leq 1.1(A_{cr} + B_{cr}) \\ 0 & \text{otherwise} \end{cases}$	rounded Mbam (rMbm) $D_{AB} = \begin{cases} \lceil B_{AB} \rceil & \text{if } \{B_{AB}\} \geq 0.5 \\ \lfloor B_{AB} \rfloor & \text{otherwise} \end{cases}$

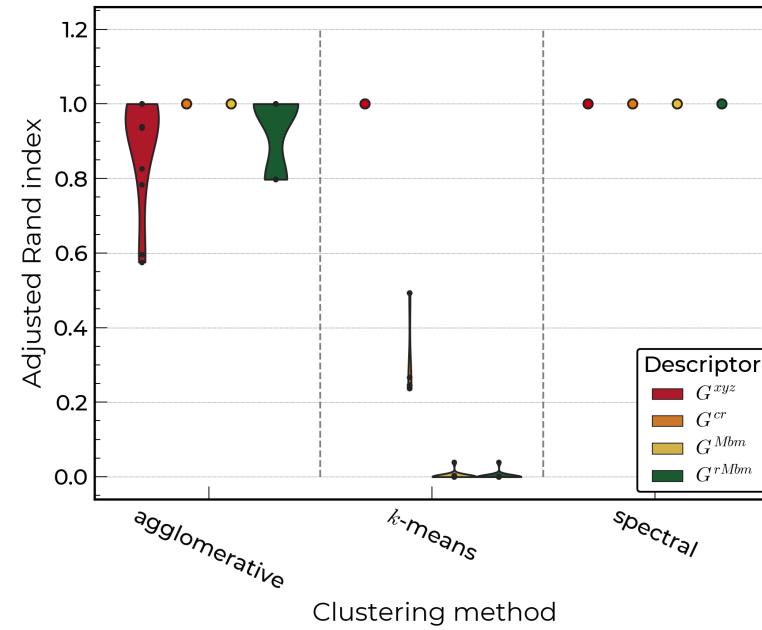
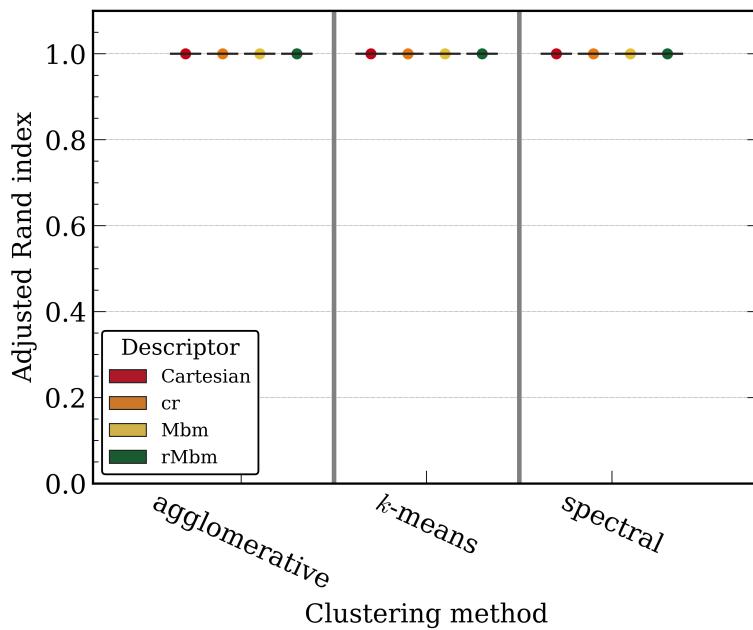
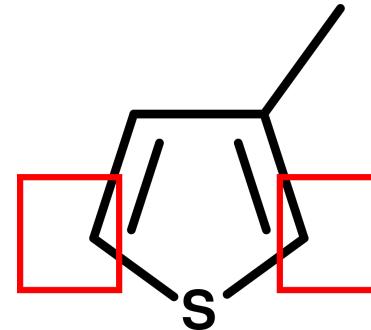
Partitioning straightforward cases

adjusted Rand index =1 → expected fragmentation

- Wales water cluster database

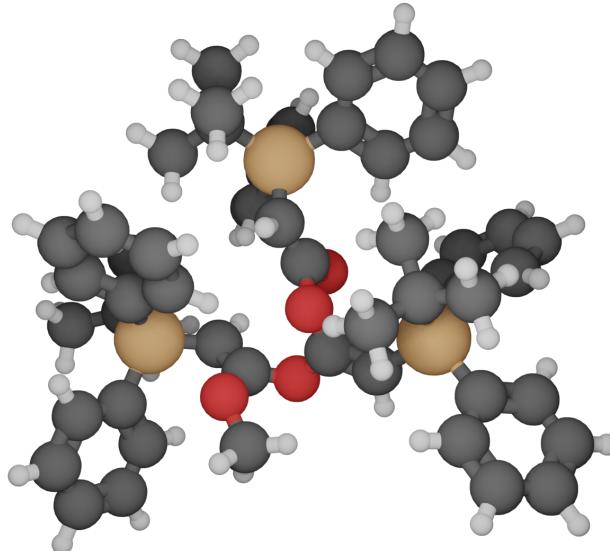
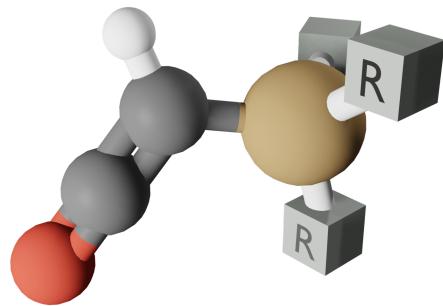


- methylthiophene tetramers



Summary:

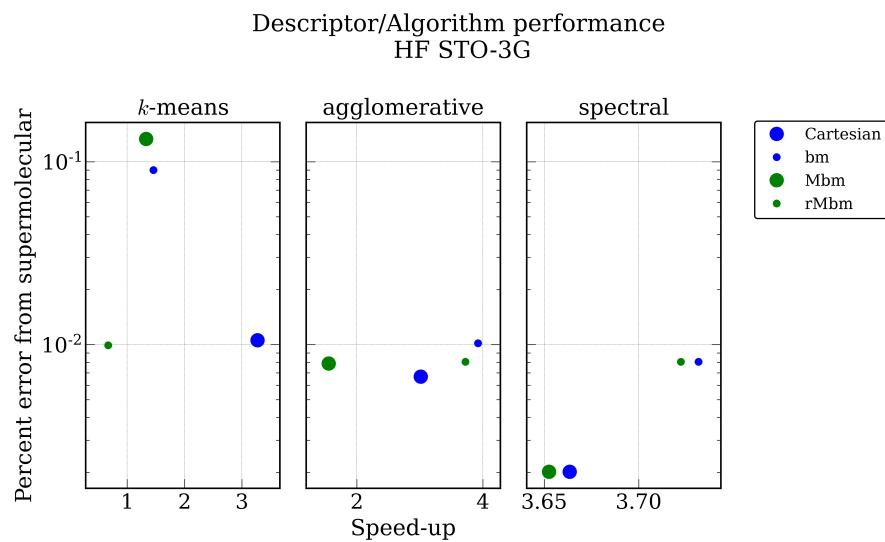
A more ambiguous case: silyl ketene



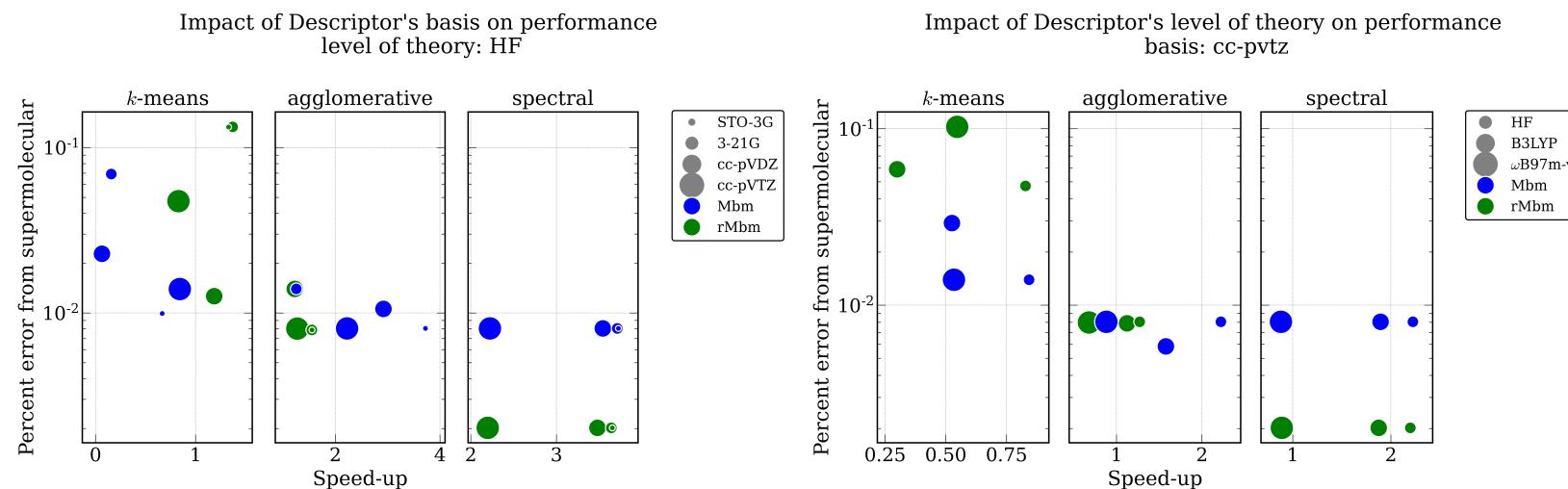
Computational details:

- vacuum embedding: $E_{\text{tot}} = \sum_{i=1}^{N_{\text{frag}}} E_i$
 - E_{frag} = Fragments with hydrogen caps - H only values

A more ambiguous case: silyl ketene



- Spectral clustering is recommended
 - gives lowest error/best speed-up
 - for clustering a which is insensitive to descriptor



Acknowledgements

- Dr. Kenneth D. Jordan
- Dr. Daniel S. Lambrecht
- Dr. Luke Shulenburger, Dr. Raymond Clay
- Co-authors

Computational Resources

- Center for Research computing, University of Pittsburgh
- Argonne Leadership Computing Facility
- Sandia National Laboratories

Funding:

- ACS: DNI-63213
- NSF: 1807638
- Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Conclusions

In regimes of strong electron correlation, mean field methods may be problematic and require need alternatives

Presented on three approaches:

- QMC: application to H_2O_4 and H on graphene
 → systems where DFT varies.
- SCI methods in H_2O_4 as trial wave function
- UML MF: offer an approach to determining molecular fragmentation automatically, with very little information

References

- Albert P. Bartók, Risi Kondor, & Gábor Csányi, “On Representing Chemical Environments,” *Phys. Rev. B*, 87 (2013), 184115DOI:[10.1103/PhysRevB.87.184115](https://doi.org/10.1103/PhysRevB.87.184115)
- Amanda Dumi, Shiv Upadhyay, Leonardo Bernasconi, Hyeondeok Shin, Anouar Benali, & Kenneth D. Jordan, “The Binding of Atomic Hydrogen on Graphene from Density Functional Theory and Diffusion Monte Carlo Calculations,” *The Journal of Chemical Physics*, 156.14 (2022), 144702DOI:[10.1063/5.0085982](https://doi.org/10.1063/5.0085982)
- A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, & G. J. Tucker, “Spectral Neighbor Analysis Method for Automated Generation of Quantum-Accurate Interatomic Potentials,” *Journal of Computational Physics*, 285 (2015), 316–30DOI:<https://doi.org/10.1016/j.jcp.2014.12.018>
- Shiv Upadhyay, Amanda Dumi, James Shee, & Kenneth D. Jordan, “The Role of High-Order Electron Correlation Effects in a Model System for Non-Valence Correlation-Bound Anions,” *The Journal of Chemical Physics*, 153.22 (2020), 224118DOI:[10.1063/5.0030942](https://doi.org/10.1063/5.0030942)