Sparse Representations in Electronic Structure Theory

Edward Hohenstein

SLAC National Accelerator Laboratory

Project Overview

SciDAC: Designing Photocatalysts Through Scalable Quantum Mechanics and Dynamics

Todd Martinez[1] (PI), Henry Van Den Bedem[2] (co-PI), T.J. Lane[2] (co-PI), Alex Aiken[1] (co-PI), Lexing Ying[1] (co-PI), Kunle Olukotun[1] (co-PI), Possu Huang[1] (co-PI), Ron Dror[1] (co-PI), Edward Hohenstein[2] (co-PI)

1. Stanford University, 2. SLAC National Accelerator Laboratory

- Improve computational modeling tools for modern architectures to enable design of improved photocatalysts

- **Major Thrusts**
  - Computational frameworks and algorithms for parallel execution on exascale architectures (Alex Aiken, Kunle Olukotun, Lexing Ying)
    - Legion/Regent, DeLite
  - New methods for *ab initio* and QM/MM molecular dynamics (Ed Hohenstein, Todd Martínez)
  - Development of new algorithms for protein design with applications to photoactivated enzymes (Ron Dror, Possu Huang, TJ Lane, Henry van den Bedem)
    - Protein design with novel cofactors
    - QM/MM investigations of photoenzymes

Fatty Acid Photodecarboxylase

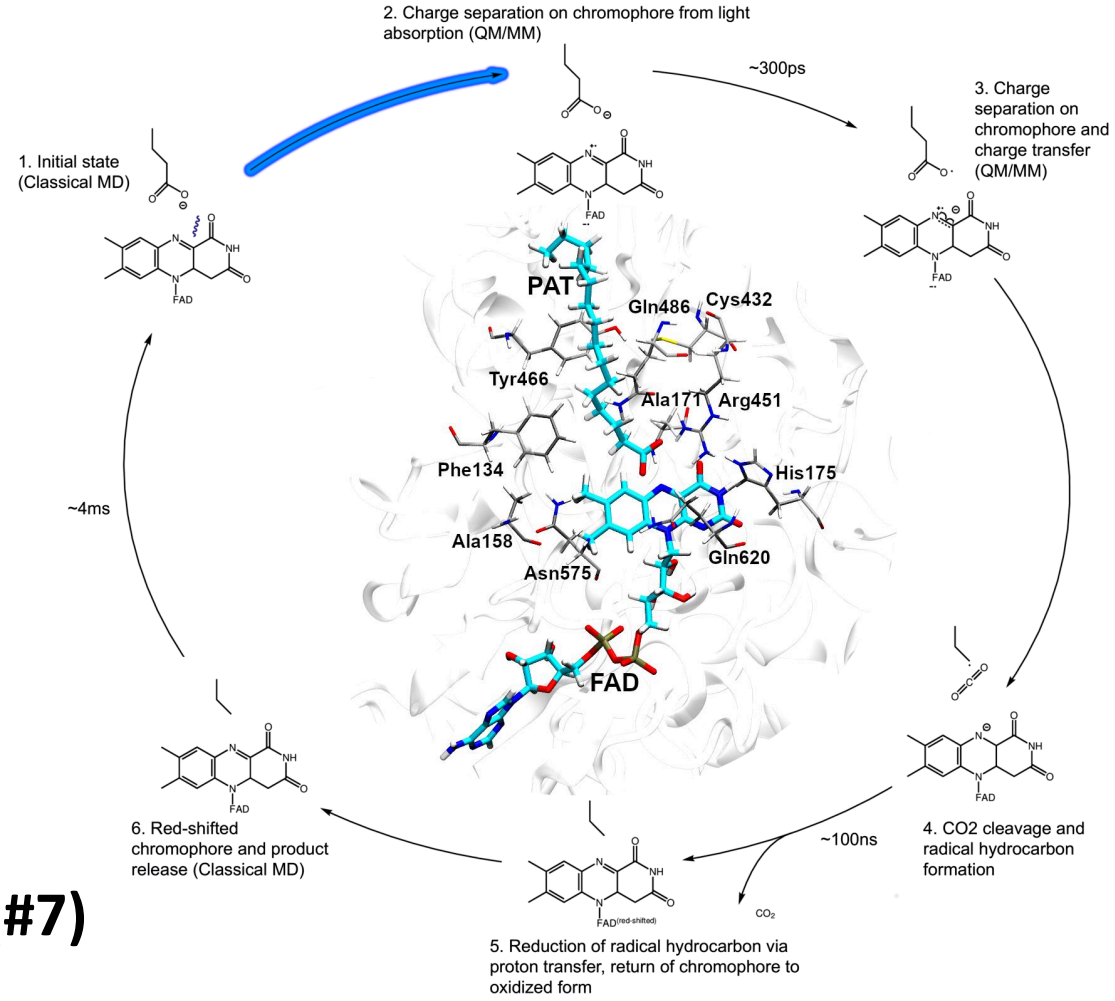
- Photo-activated production of alkanes
- Requirement for light: not yet clear
- Mechanism: not known
- Project goal: QM calculations to predict mechanism, verify by comparing to spectroscopy, crystallography

- Like all photoenzymes, FAP contains an “antenna” molecule (cofactor) that absorbs photons and plays a key role in the chemistry

- QM/MM calculations of excited state mechanism are in progress

- Large QM regions (300 atoms): Need fast and accurate QM!

See Alice Walker’s poster (#7)



Primitives for Electronic Structure Theory

- Electron repulsion integrals

$$(\mu\nu|\lambda\sigma) = \int\int \phi_\mu(r_1)\phi_\nu(r_1) r_{12}^{-1} \phi_\lambda(r_2)\phi_\sigma(r_2) \mathrm{d}r_1 \mathrm{d}r_2$$

- Coulomb matrix

$$J^{(\mathrm{D})}_{\mu\nu} = \sum_{\lambda\sigma} (\mu\nu|\lambda\sigma) D_{\lambda\sigma}$$

- Exchange matrix

$$K^{(\mathrm{D})}_{\mu\nu} = \sum_{\lambda\sigma} (\mu\lambda|\nu\sigma) D_{\lambda\sigma}$$

- Coulomb derivative

$$J(\mathrm{A},\mathrm{B})^\xi = \sum_{\mu\nu\lambda\sigma} \frac{\partial}{\partial\xi}(\mu\nu|\lambda\sigma) A_{\mu\nu} B_{\lambda\sigma}$$

- Exchange derivative

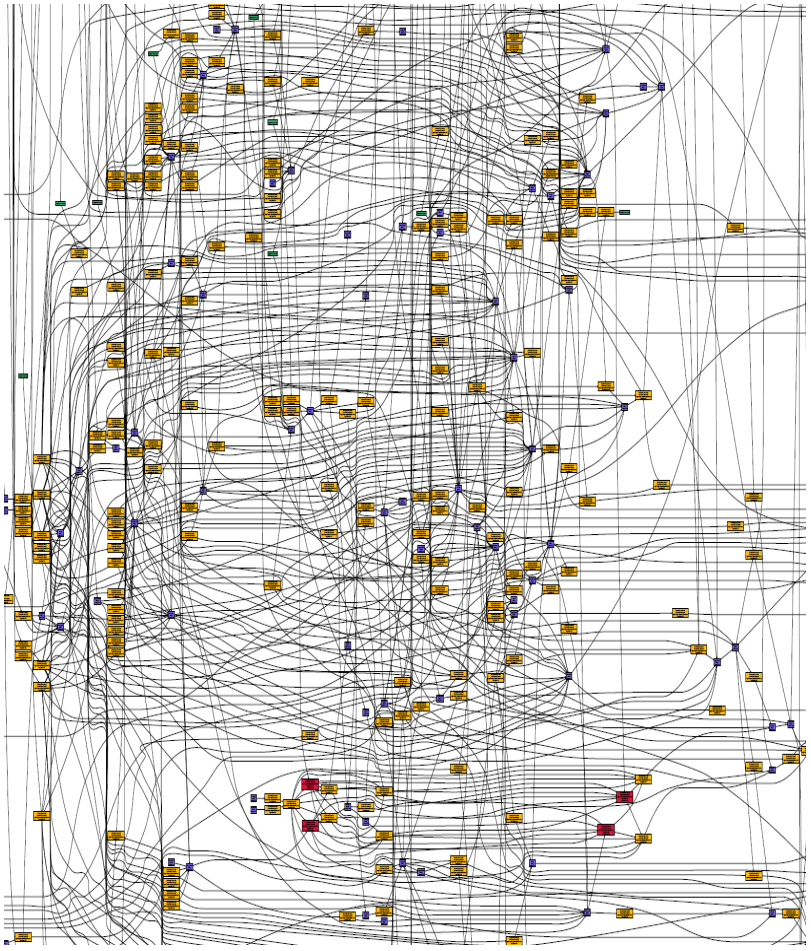
$$K(\mathrm{A},\mathrm{B})^\xi = \sum_{\mu\nu\lambda\sigma} \frac{\partial}{\partial\xi}(\mu\lambda|\nu\sigma) A_{\mu\nu} B_{\lambda\sigma}$$

- *All* electronic structure methods can be formulated in terms of these primitives
  - Some can be formulated efficiently
    - HF/DFT
    - CIS/TDDFT
    - CASCI, CASSCF, etc.

- We have fast, hand tuned, single node, multi-GPU versions (TeraChem)
- We want automatically generated multi-node, multi-GPU versions

Legion/Regent

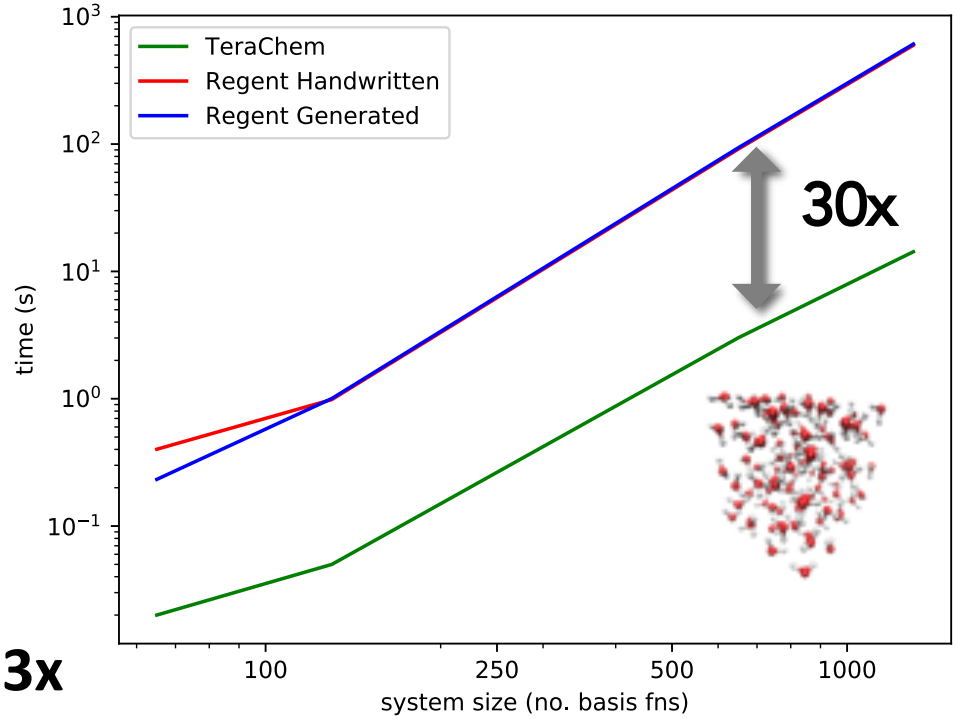
- Developed by Alex Aiken's group (Stanford)
- Task-based programming model
  - Heterogeneous machines
  - Distributed memory
- Key features
  - Programs are written without specifying where computations will run and where data will be placed
  - Separate mapping phase allows program to be tuned to a particular machine
  - CUDA kernels for GPU execution can be automatically generated
- Development of J/K matrices is underway
  - Target: multi-node, multi-GPU

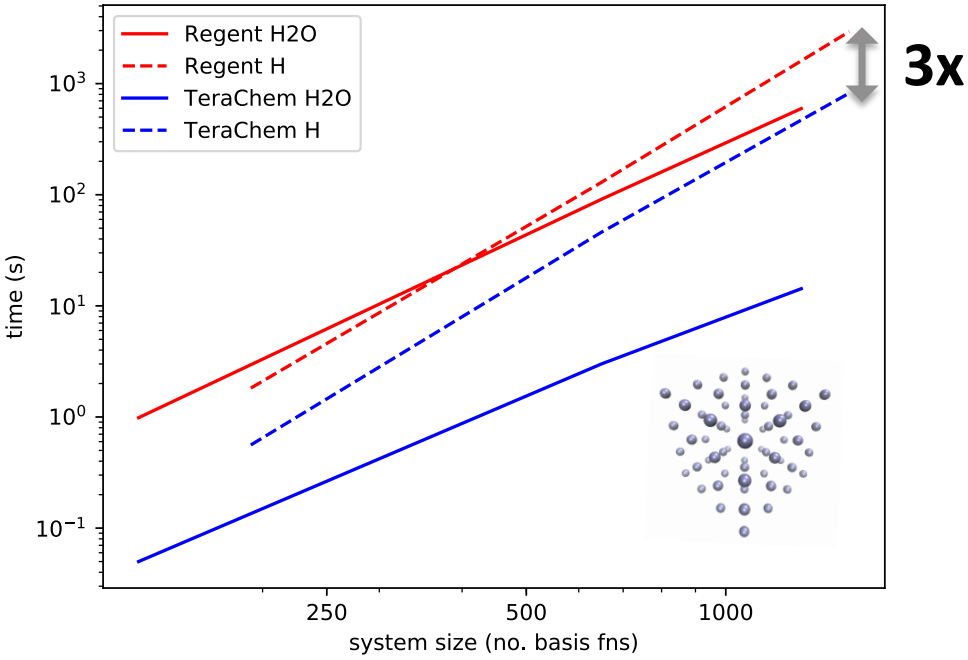


Task graph for one step of a simple application

Coulomb Matrices in Legion/Regent

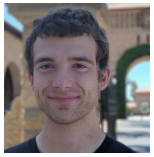
- Handwritten and generated GPU kernels are equally efficient
- 30x slower than TeraChem for boxes of water molecules
- Only 3x slower for grids of hydrogen atoms





- Culprit seems to be a liberal use of atomic memory operations
  - Now only 6x slower for boxes of water molecules

Grace Johnson



Ellis Hoag

Exploiting Rank-Sparsity in Correlated Methods

- Calculations of J/K matrices exploit spatial sparsity and streaming architectures
- What about correlated methods? Configuration interaction, coupled-cluster, etc.
  - Not always amenable to formulations in terms of J and K
- Find rank-sparsity in the wavefunction coefficients

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}$$

Diagonal
Rank = number nonzero diagonal elements

Assume rank is small and find best low rank approximation
that reproduces known entries in the matrix…

Rank-Reduced Full CI

- Full Configuration Interaction:
  - Exact solution to Schrodinger's equation in a finite basis
  - Scales factorially with system size

$$|\Psi_{\mathrm{CI}}\rangle = \sum_I c_I |\Phi_I\rangle$$

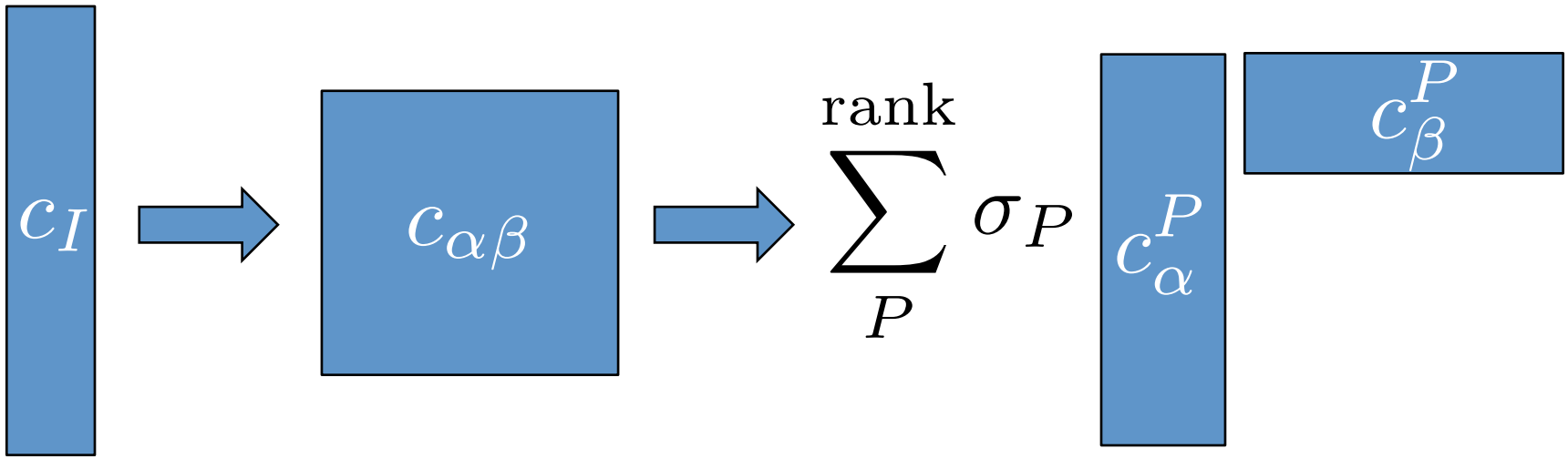
All possible arrangements of electrons in orbitals

- Write in terms of alpha and beta components

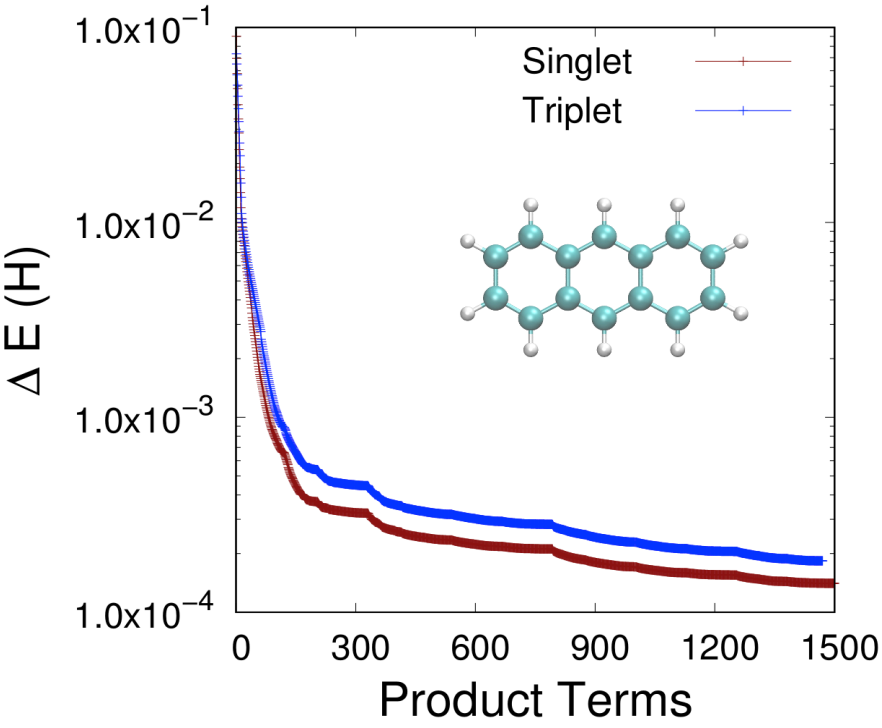
$$|\Psi_{\mathrm{CI}}\rangle = \sum_{\alpha\beta} c_{\alpha\beta} |\Phi_{\alpha\beta}\rangle$$

Direct product of all possible arrangements of α electrons in α orbitals, β electrons in β orbitals

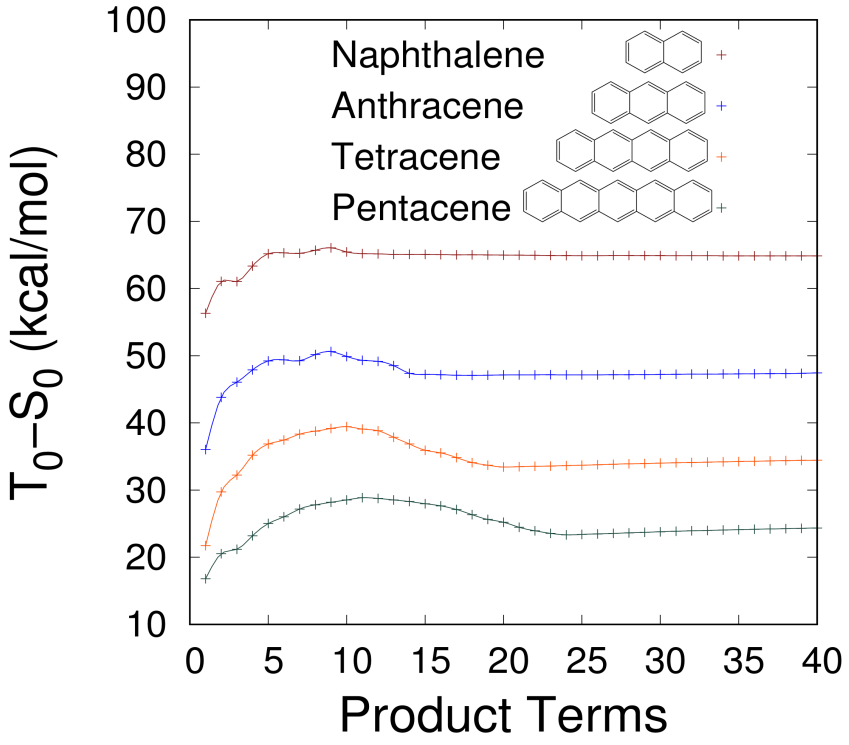
- Rearrange a vector to a matrix (and approximate as a rank-sparse matrix)



Rank-Reduced Full CI



50 terms are sufficient for
1 kcal/mol accuracy

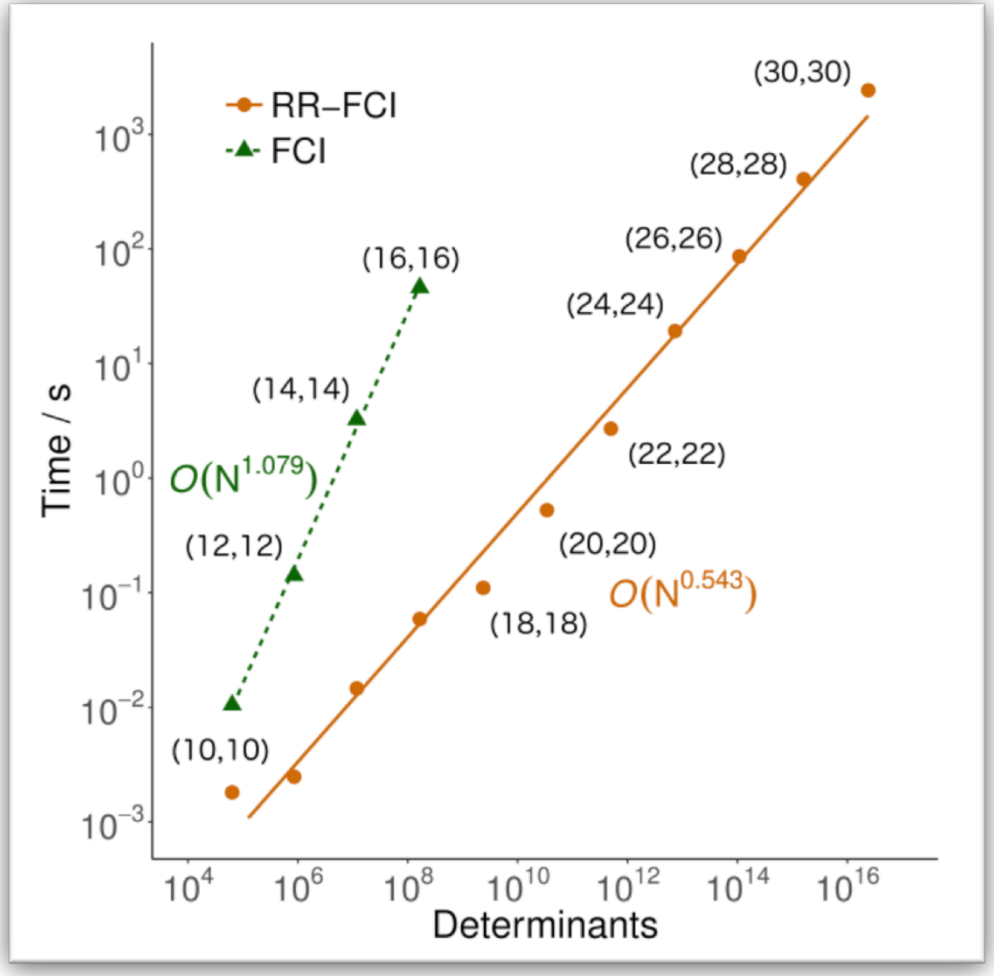


Energy differences converge
more rapidly

Scott Fales

B.S. Fales, S. Seritan, N. Settje, B.G. Levine, H. Koch, T.J. Martínez, *J. Chem. Theory Comput.* **14**, 4139 (2018).

Rank-Reduced Full CI



$\mathcal{O}(N!)$ reduces to $\mathcal{O}(\sqrt{N!})$

Scott Fales

B.S. Fales, S. Seritan, N. Settje, B.G. Levine, H. Koch, T.J. Martínez, *J. Chem. Theory Comput.* **14**, 4139 (2018).

Rank-Reduced Coupled Cluster

- Coupled-cluster (CCSD) wavefunction ansatz

$$|\Psi_{\mathrm{CC}}\rangle = e^{\hat{T}}|\Phi_0\rangle$$

$$\hat{T} = \hat{T}_1 + \hat{T}_2 = \sum_{ia} t_i^a a_a^\dagger a_i + \frac{1}{4}\sum_{ijab} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_j a_i$$

- The wavefunctions contains O(N^4) parameters (storage bottleneck)
- CCSD amplitude equations, scales as O(N^6) (compute bottleneck)

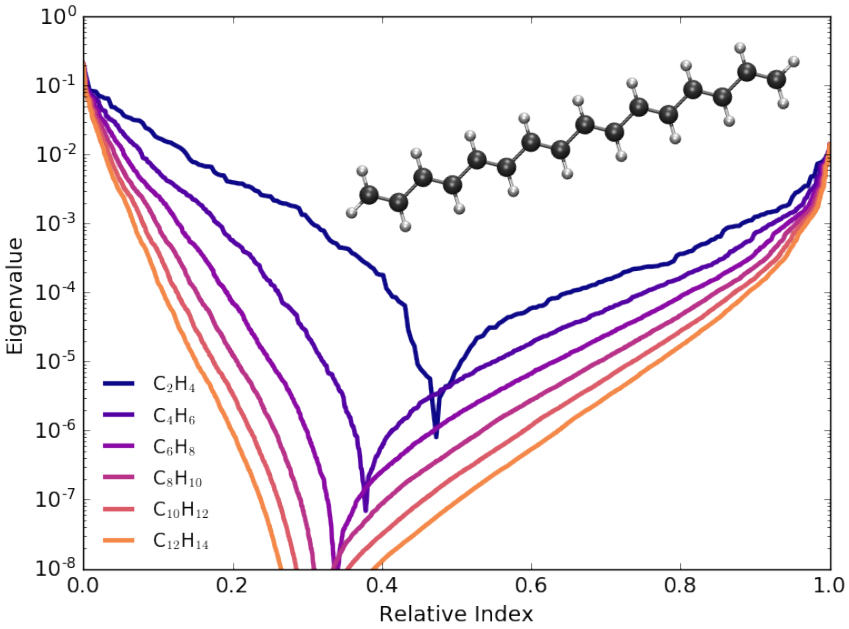
$$\langle \Phi_i^a | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Phi_0 \rangle = 0 \qquad \langle \Phi_{ij}^{ab} | e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Phi_0 \rangle = 0$$

- Can we compress the CCSD amplitudes?

$$t_{ij}^{ab} = \sum_{PQ} U_{ia}^P T^{PQ} U_{jb}^Q$$

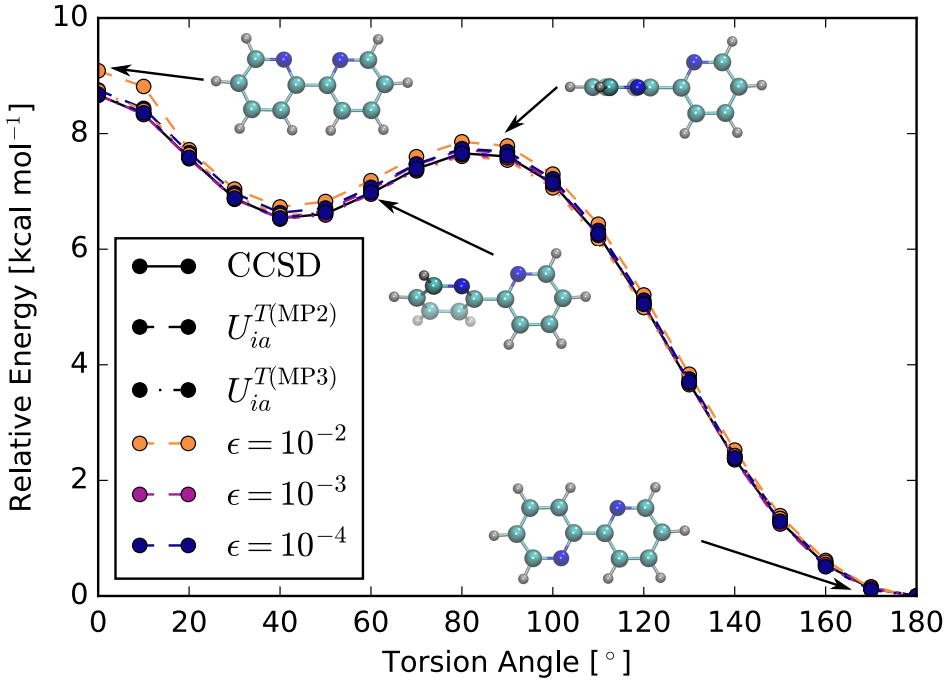
- Solve for the amplitudes in a compressed representation

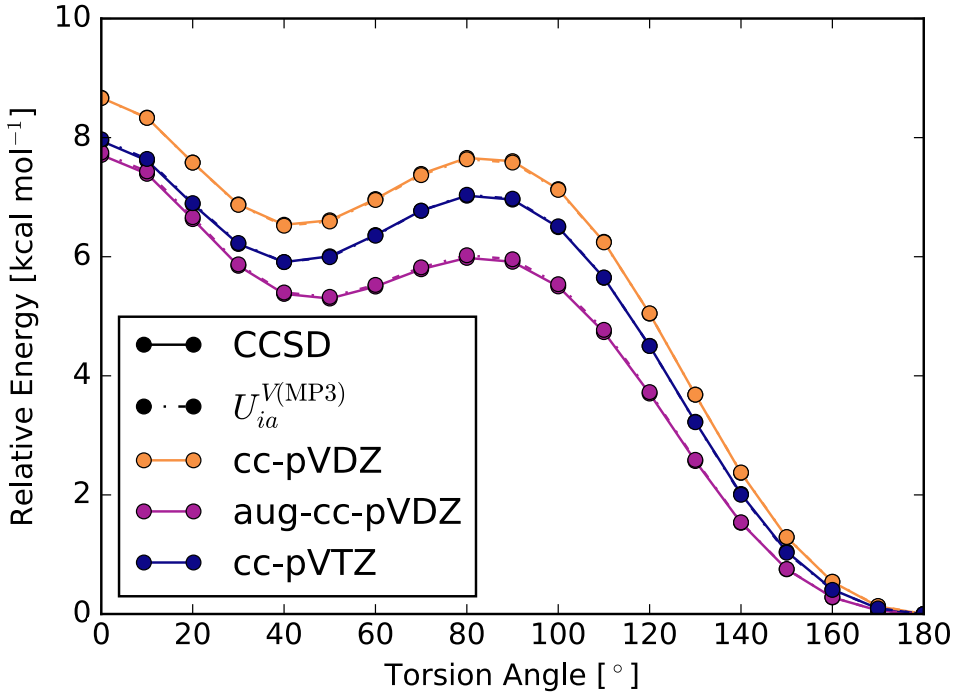
See my poster (#7)



Does it work?

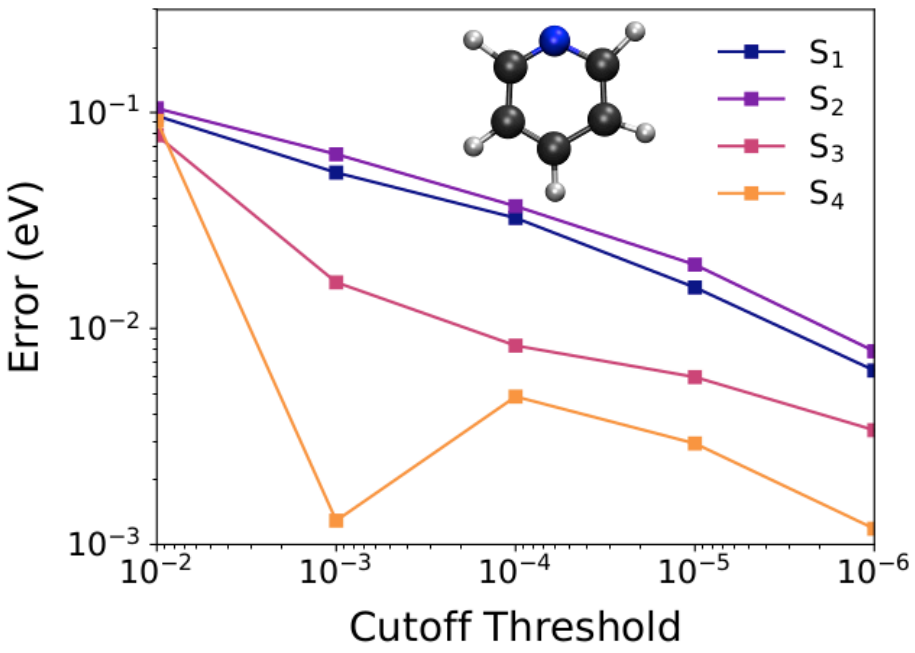
- Can obtain sub-mH accuracy in absolute energies
- Relative energies are more forgiving than absolute energies
- Compression increases in larger basis sets
  - With recommended cutoffs:
    - *ov* space is compressed to 10% of the original size
    - Only 1% of the doubles parameters are needed

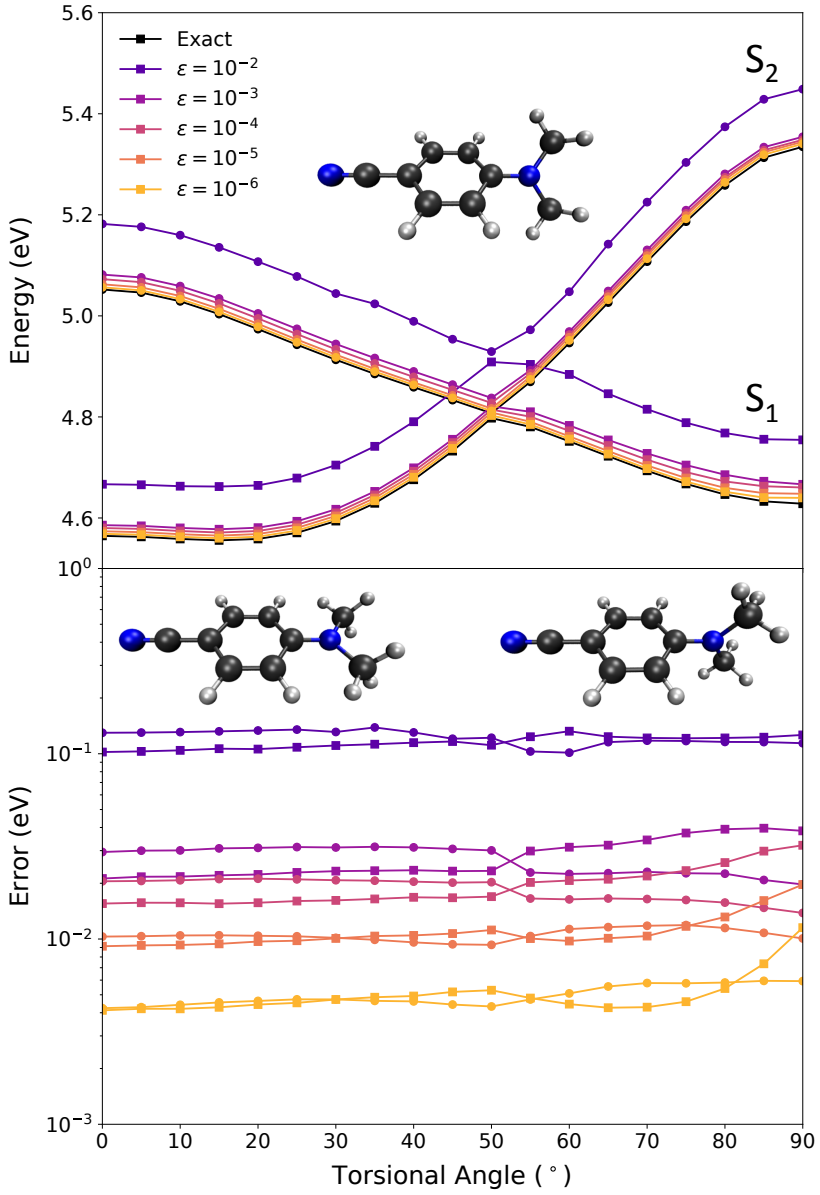




R.M. Parrish, Y. Zhao, E.G. Hohenstein, and T.J. Martinez, *J. Chem. Phys.* **150**, 164118 (2019).

Electronic excited states





- Equation-of-motion CCSD
- Similar compression as on the ground state
  - 10% of the *ov* space
  - 1% of the doubles parameters
    - For 20-30 atom systems
  - Roughly 0.01 eV accuracy

E.G. Hohenstein, Y. Zhao, R.M. Parrish, and T.J. Martinez, *J. Chem. Phys.*, submitted.

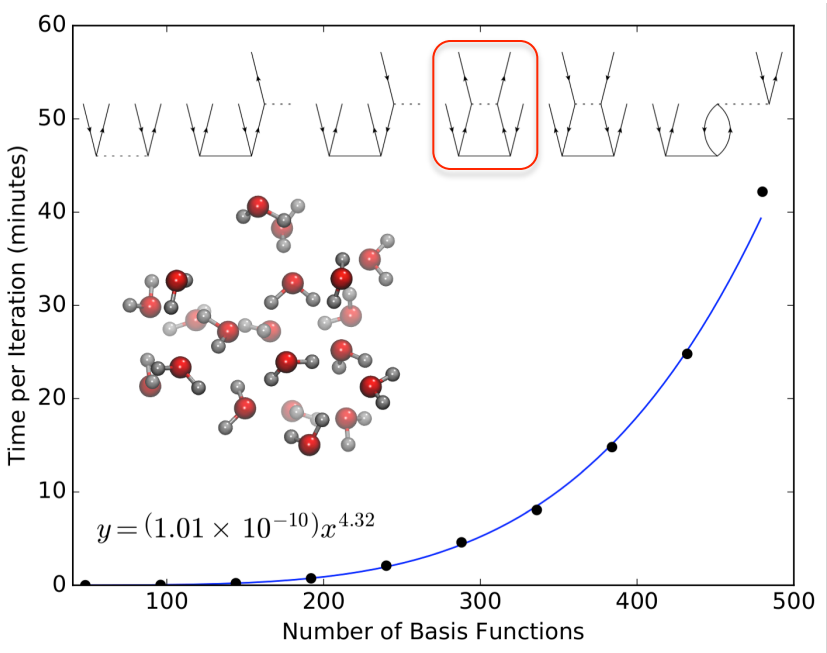
The Path Forward

- Develop an implementation in Legion
- RR-CCSD factorization can reduce the scaling of linear terms to O(N^5)
  - Linearized CCD, for example:

$$\langle \Phi_{ij}^{ab} | \hat{H} + [\hat{H}, \hat{T}_2] | \Phi_0 \rangle = 0$$

  - (Right) Timings of RR-LCCD/cc-pVDZ
  - Promising for EOM-CCSD
    - R is a linear operator!

Rate-limiting contribution to the amplitude equations



- Introduce Tensor HyperContraction factorization of the amplitudes

$$U_{ia}^{A} \approx \sum_{P} X_i^P X_a^P Y_A^P$$

- *PQ* auxiliary indices allow separation of orbital indices
- Amplitude equations solved in the low-rank *AB* space

$$t_{ij}^{ab} \approx \sum_{PQ} \sum_{AB} X_i^P X_a^P \underbrace{Y_A^P T^{AB} Y_B^Q}_{T^{PQ}} X_j^Q X_b^Q$$

- Will allow reduction of computational complexity to *O*(N^4) and storage to *O*(N^2)

Triples Corrections to CCSD with Tensor HyperContraction

- High accuracy requires triples corrections: O(N^7) scaling
  - CCSD(T): ground state energies
  - CC3: excitation energies

- Apply THC factorizations of integrals and amplitudes to reduce the scaling:

$$(ij|kl) \approx \sum_{PQ} X_i^P X_j^P Z^{PQ} X_k^Q X_l^Q$$

$$t_{ij}^{ab} \approx \sum_{PQ} \tau_i^P \tau_a^P T^{PQ} \tau_j^Q \tau_b^Q$$

- Representative contribution to the (T) correction:

$$E \leftarrow \sum_{ijkabc} \left(\sum_d \left[t_{ij}^{ad} (ck|bd) \right] \sum_e \left[t_{ji}^{ce} (ak|be) \right] \Delta_{ijk}^{abc} \right)$$

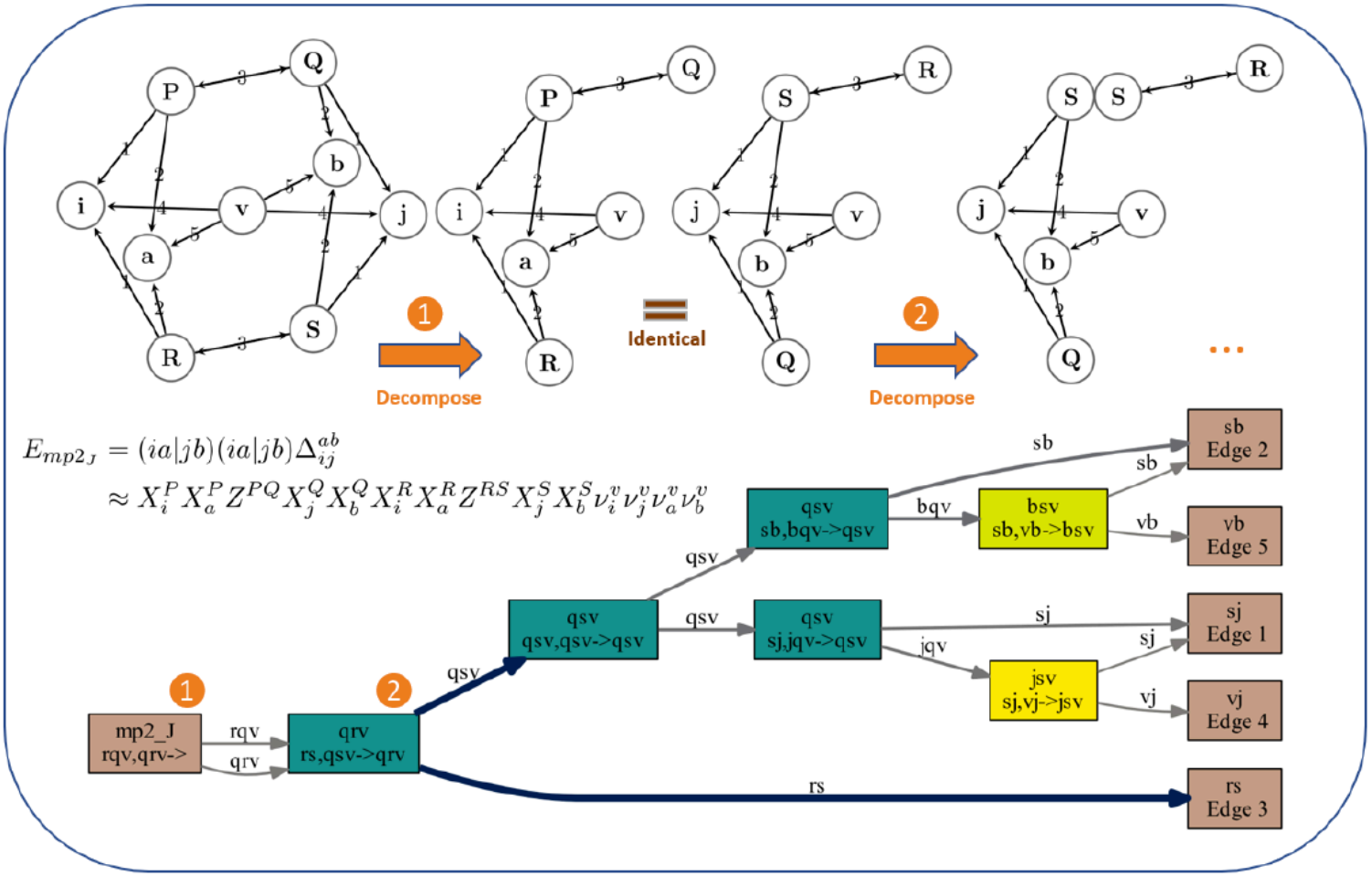
- With Tensor HyperContraction:

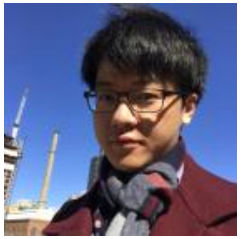
$$E \leftarrow \sum_{ijkabcdeABCDPQRSw} \tau_i^A \tau_a^A T^{AB} \tau_j^B \tau_d^B X_c^P X_k^P Z^{PQ} X_b^Q X_d^Q \tau_j^C \tau_c^C T^{CD} \tau_i^D \tau_e^D X_a^R X_k^R Z^{RS} X_b^S X_e^S \delta_i^w \delta_j^w \delta_k^w \delta_a^w \delta_b^w \delta_c^u$$

- We want to apply THC to CCSD(T), potential for O(N^5) scaling…but:
  - 124 contributions to the (T) correction
  - 4 × 10^{26} different ways to implement each term

Automatic Factorization and Code Generation

- Represent factorized expressions as a graph
- Determine the optimal factorization (for each particular problem size)
  - Minimize FLOPs, minimize memory footprint
- Automatically generate code implementing the factorization (in Regent)



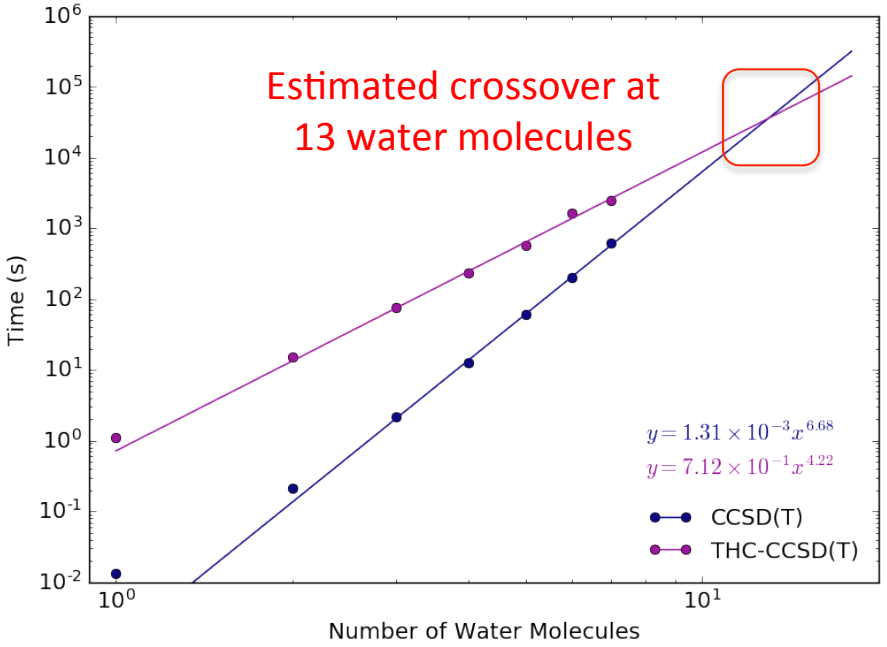


Yao Zhao

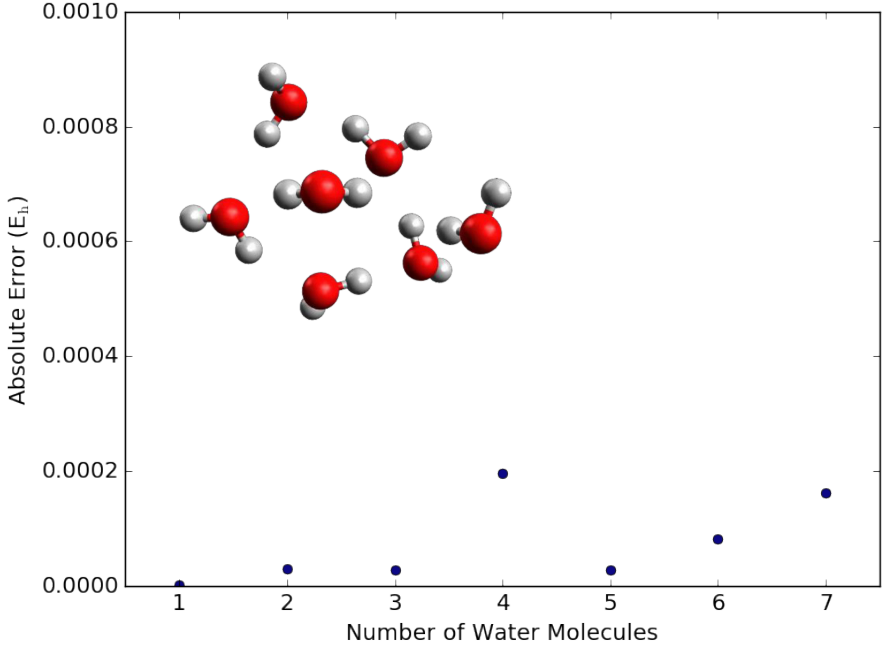
THC-CCSD(T): Accuracy and Timings

- Verified that all 124 terms can be evaluated with $O(N^5)$ effort or less
- Conventional implementation: $O(o^3v^4)$
- THC factorization: $O(v^4N_{grid})$
- Crossover with between THC and conventional implementation
  - Estimated around 13 water molecules
- THC errors in (T) correction below 2.0×10^{-4} E_h
- Current implementation requires further optimization to reduce memory footprint
- Automatic generation of Regent code

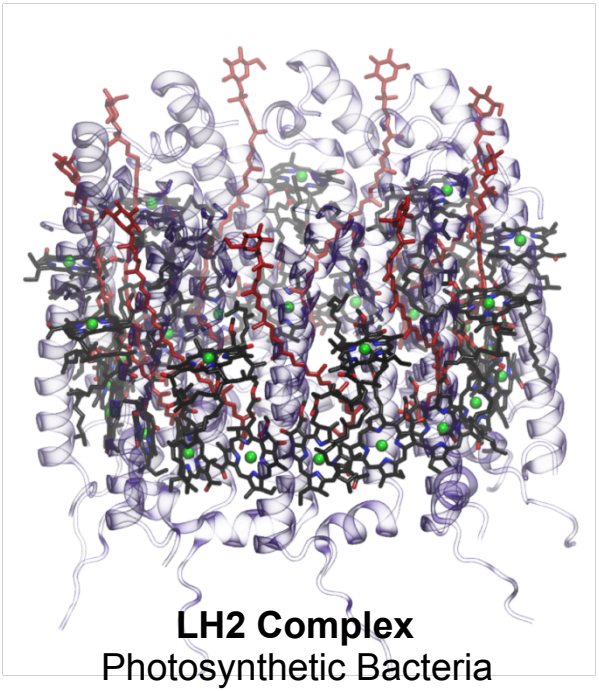
Timings of (THC-)CCSD(T)/cc-pVDZ



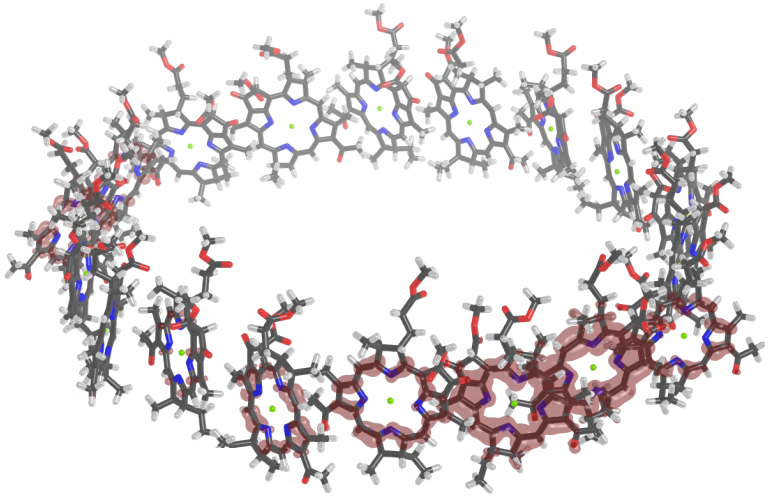
Errors of THC-CCSD(T)/cc-pVDZ



Sparse fragment-based models: *ab initio* exciton model

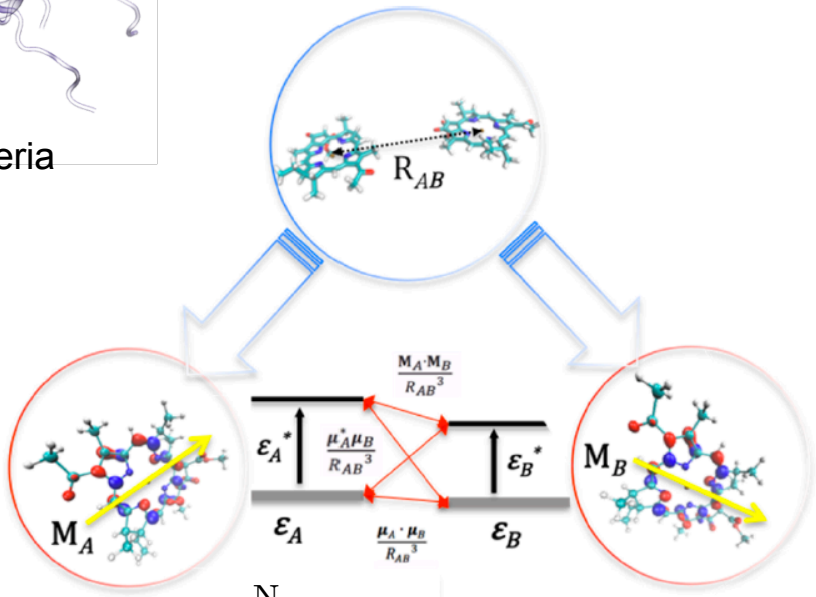


LH2 Complex
Photosynthetic Bacteria



B850 Ring

Exciton Model



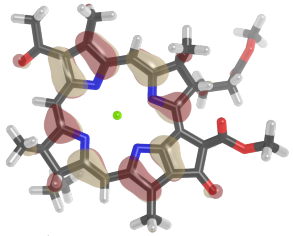
$$\hat{H} = \sum_{i}^{N} E_i |i\rangle\langle i| + \sum_{i \neq j} V_{ij} |i\rangle\langle j|$$

Exact solution scales exponentially:

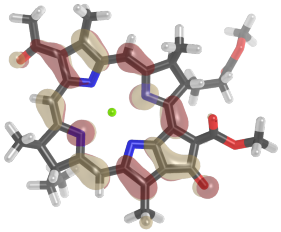
N chromophores, 2 states each, $O(2^N)$

Quantum Sparsity?

Monomer States (TDDFT):



$S_0^A \rightarrow |0_A\rangle$



$S_1^A \rightarrow |1_A\rangle$

1x Qubit per Monomer!!

Map the excitonic Hamiltonian to a Generalized spin-lattice Hamiltonian:

$$\hat{H} \equiv \mathcal{E}\hat{I} + \sum_A \mathcal{Z}_A \hat{Z}_A + \mathcal{X}_A \hat{X}_A$$

$$+ \sum_{A>B} \mathcal{XX}_{AB} \hat{X}_A \otimes \hat{X}_B + \mathcal{XZ}_{AB} \hat{X}_A \otimes \hat{Z}_B$$

$$+ \mathcal{ZX}_{AB} \hat{Z}_A \otimes \hat{X}_B + \mathcal{ZZ}_{AB} \hat{Z}_A \otimes \hat{Z}_B$$

Rob Parrish

Peter McMahon

MC-VQE Ansatz:

$$|\Psi_\Theta\rangle \equiv \hat{U} \sum_{\Theta'} |\Phi_{\Theta'}\rangle V_{\Theta'\Theta}$$

(2) State-Averaged MC-VQE Entangler (Quantum)

(3) Subspace Eigenstates (Classical)

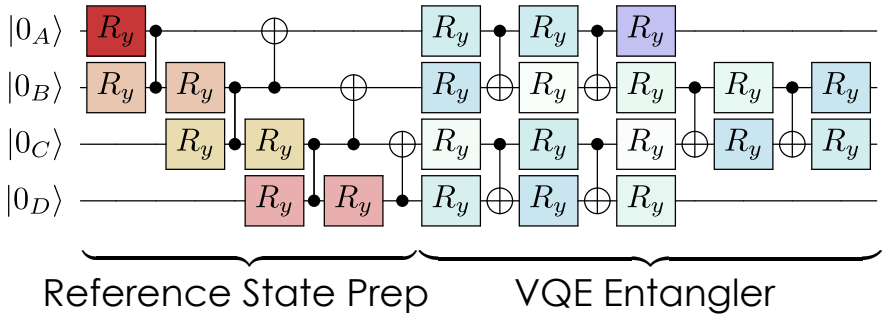
(1) Contracted Reference States (Classical)

State-Averaged Energy:

$$\bar{E} \equiv \frac{1}{N_\Theta} \sum_\Theta^{N_\Theta} \langle \Phi_\Theta | \hat{U}^\dagger \hat{H} \hat{U} | \Phi_\Theta \rangle = \frac{1}{N_\Theta} \sum_\Theta^{N_\Theta} E_\Theta$$

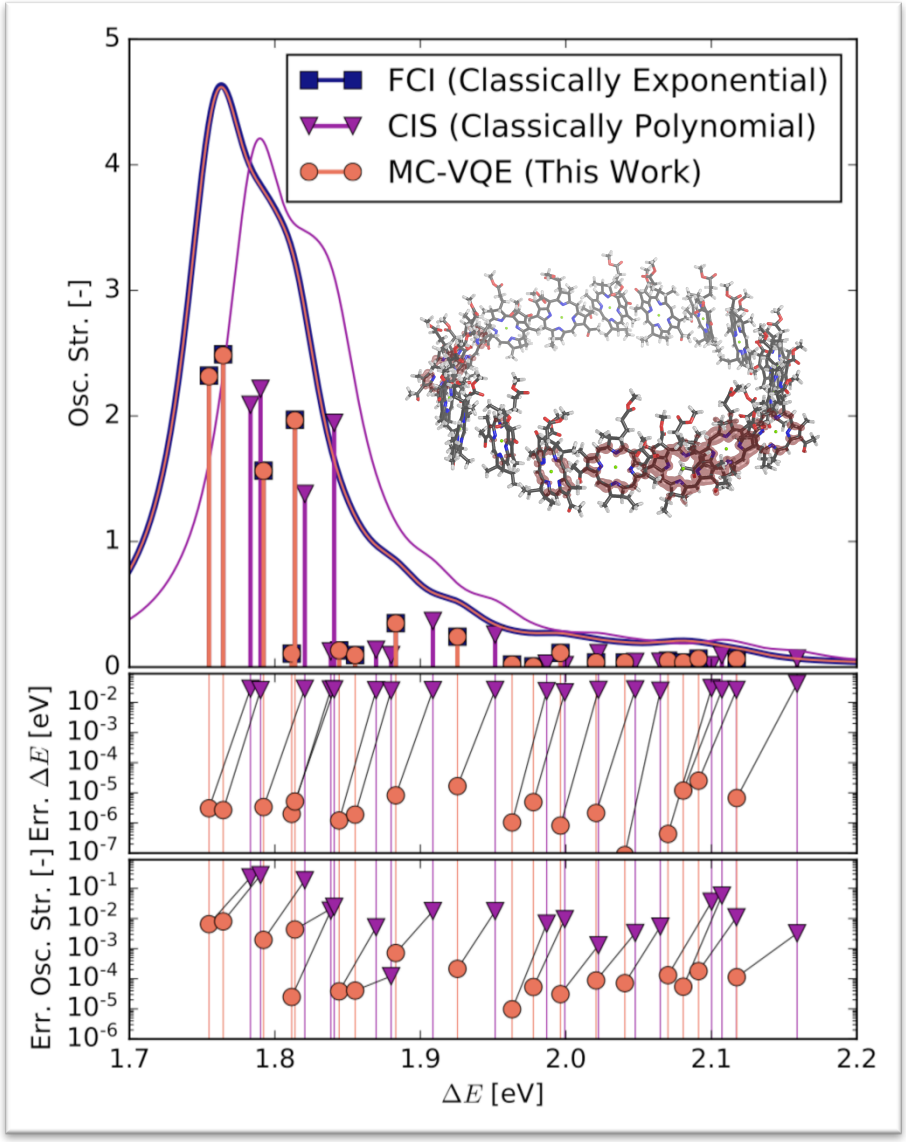
Contracted, Entangled Hamiltonian:

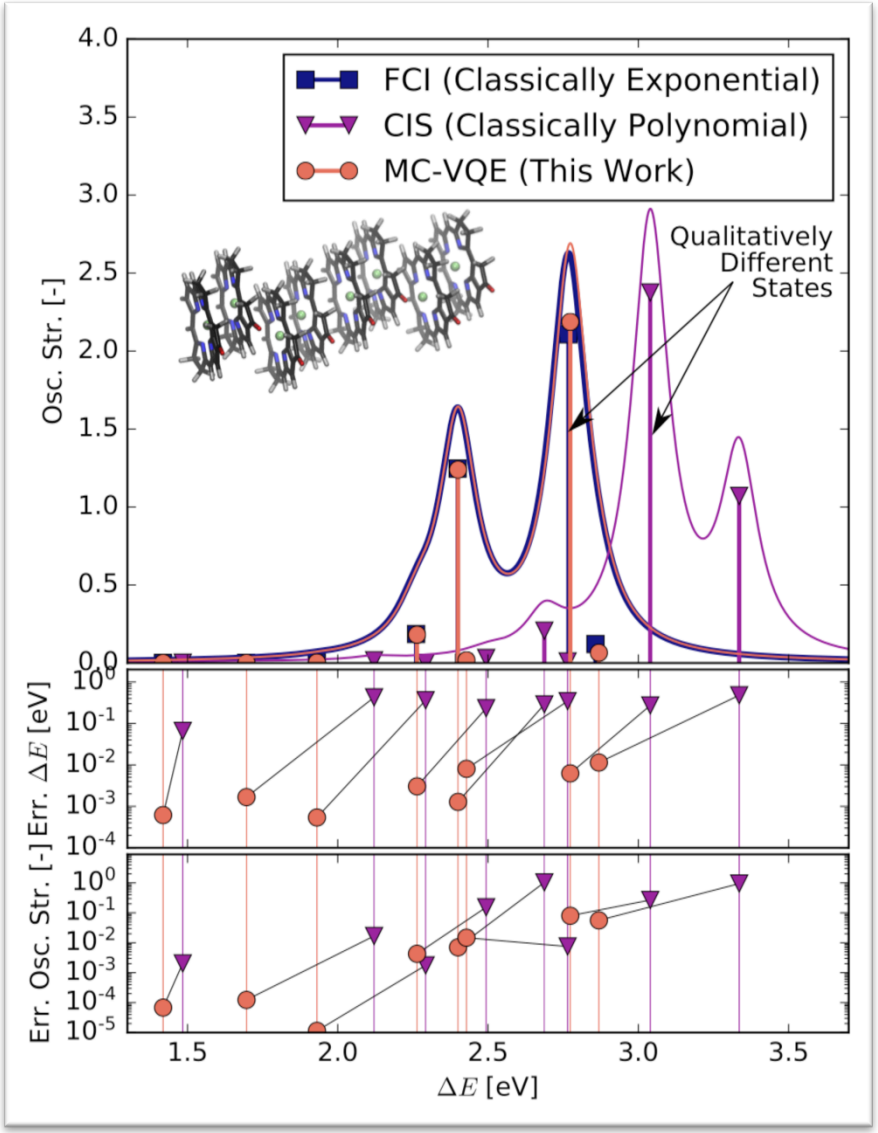
$$H_{\Theta\Theta'} \equiv \langle \Phi_\Theta | \hat{U}^\dagger \hat{H} \hat{U} | \Phi_{\Theta'} \rangle$$



R.M. Parrish, E.G. Hohenstein, P.L. McMahon, and T.J. Martinez, *Phys. Rev. Lett.* **122**, 230401 (2019).

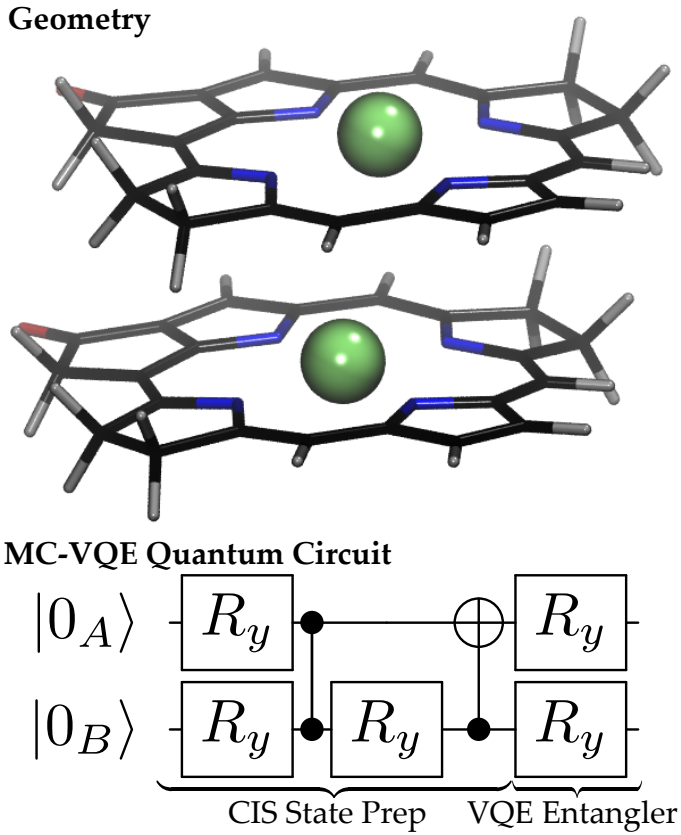
Quantum algorithms for excitonic systems

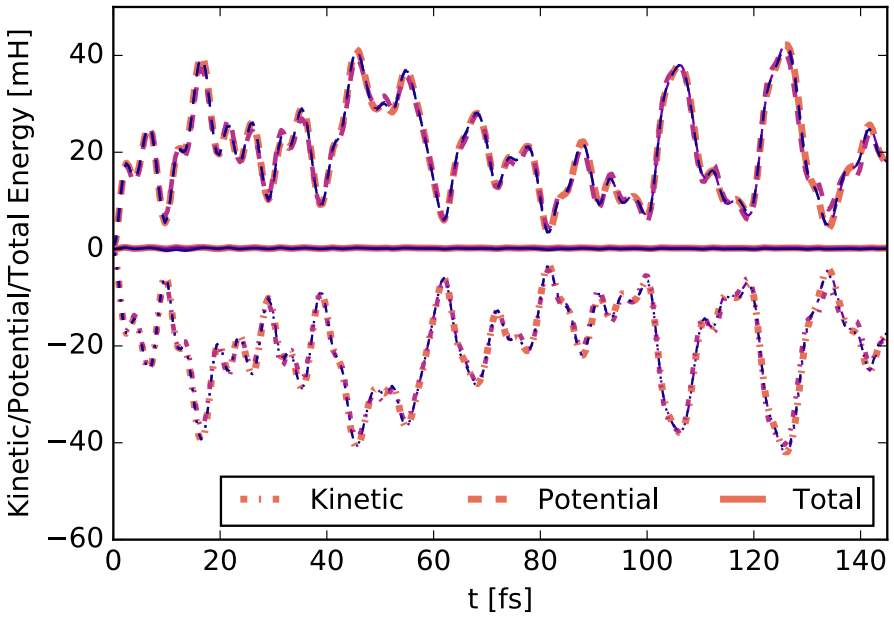




R.M. Parrish, E.G. Hohenstein, P.L. McMahon, and T.J. Martinez, *Phys. Rev. Lett.* **122**, 230401 (2019).

Quantum Dynamics on Quantum Computers





- Calculation of matrix elements use classical GPU-accelerated codes
  - …and derivatives of those matrix elements
- Energies of excitonic states from MC-VQE
  - …include response of the MC-VQE parameters for analytic gradients

R.M. Parrish, E.G. Hohenstein, P.L. McMahon, and T.J. Martinez, arXiv:1906.08728

Acknowledgements

- Todd Martinez (SLAC)
- Henry Van Den Bedem (SLAC)
- T.J. Lane (SLAC)
- Alex Aiken (SLAC)
- Lexing Ying (Stanford)
- Kunle Olukotun (Stanford)
- Possu Huang (Stanford)
- Ron Dror (Stanford)
- Rob Parrish (QCWare)
- Peter McMahon (Cornell)
- Alice Walker
- Grace Johnson
- Ellis Hoag
- Scott Fales
- Stefan Seritan
- Nick Settje
- Ben Levine (MSU)
- Henrik Koch (Pisa)
- Yao Zhao

SciDAC: Designing Photocatalysts Through Scalable Quantum Mechanics and Dynamics

SciDAC
Scientific Discovery through Advanced Computing

U.S. DEPARTMENT OF ENERGY | Office of Science