

Generalization of the Tensor Product Selected CI Method for Molecular Excited States

Published as part of *The Journal of Physical Chemistry A* virtual special issue “Krishnan Raghavachari Festschrift”.

Nicole M. Braunscheidel,[§] Vibin Abraham,[§] and Nicholas J. Mayhall*



Cite This: *J. Phys. Chem. A* 2023, 127, 8179–8193



Read Online

ACCESS |



Metrics & More



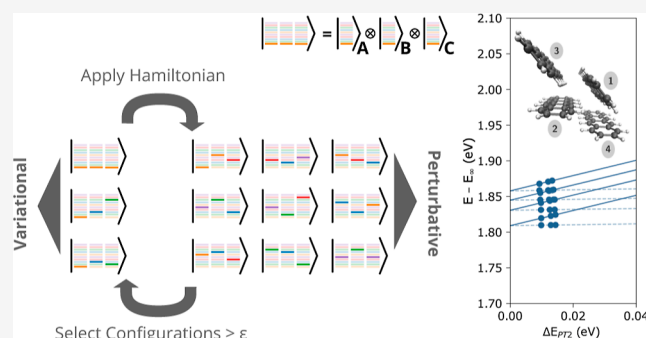
Article Recommendations



Supporting Information

ABSTRACT: In a recent paper [*JCTC*, 2020, 16, 6098], we introduced a new approach for accurately approximating full CI ground states in large electronic active-spaces called Tensor Product Selected CI (TPSCI). In TPSCI, a large orbital active space is first partitioned into disjoint sets (clusters) for which the exact, local many-body eigenstates are obtained. Tensor products of these locally correlated many-body states are taken as the basis for the full, global Hilbert space. By folding correlation into the basis states themselves, the low-energy eigenstates become increasingly sparse, creating a more compact selected CI expansion. While we demonstrated that this approach can improve accuracy for a variety of systems, there is even greater potential for applications to excited states, particularly those which have some excited-state character.

In this paper, we report on the accuracy of TPSCI for excited states, including a far more efficient implementation in the Julia programming language. In traditional SCI methods that use a Slater determinant basis, accurate excitation energies are obtained only after a linear extrapolation and at a large computational cost. We find that TPSCI with perturbative corrections provides accurate excitation energies for several excited states of various polycyclic aromatic hydrocarbons with respect to the extrapolated result (i.e., near exact result). Further, we use TPSCI to report highly accurate estimates of the lowest 31 eigenstates for a tetracene tetramer system with an active space of 40 electrons in 40 orbitals, giving direct access to the initial bright states and the resulting 18 doubly excited (biexcitonic) states.



1. INTRODUCTION

Electronic excited states play an important role in a vast number of technologically relevant processes ranging from solar cells to sensing, artificial photosynthesis, and beyond. Theoretical simulations are key for the interpretation and prediction of spectra, lending detailed support to experiments. However, not all excited states can be easily simulated computationally. Traditional theoretical methods that depend on single excitations (common to all linear-response methods) like time-dependent density functional theory (TDDFT)^{1–3} often fail to properly describe charge-transfer (CT) states^{4–6} and require an additional doubly excited component to capture the presence of doubly excited states.^{7–9} Even more sophisticated methods like equation of motion coupled cluster with singles and doubles (EOM-CCSD)^{10,11} can fail for doubly excited states with errors of around 1 eV,^{12,13} requiring higher excitations to produce accurate results. In order to provide qualitatively correct descriptions of two-electron excitations, multireference methods, such as complete active space self-consistent field (CASSCF),¹⁴ complete active space second-order perturbation

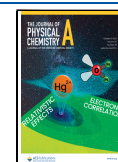
theory (CASPT2),^{15,16} or multireference configuration interaction (MRCI),^{17,18} are required. However, these methods cannot be used for active spaces larger than about 20 orbitals with 20 electrons. It is also very difficult to select active orbitals for state averaging when the ground and excited states differ significantly in dipole moment, usually seen in cases with charge-transfer excitations.

Selected configuration interaction (SCI)¹⁹-based approaches have been recently used to calculate accurate estimates for vertical excitation energies,^{20–23} double excitations,¹³ doublet–doublet transitions in radicals,²⁴ and excited-state dipole moments and oscillator strengths.²⁵ Motivated by the fact that low-energy eigenstates often have most of their weight on a

Received: May 12, 2023

Revised: August 21, 2023

Published: September 21, 2023



relatively small subspace of determinants, SCI techniques attempt a bottom-up discovery of this space of “important” Slater determinants. For weakly correlated systems, SCI provides an incredibly efficient approach for obtaining near-FCI estimates of the ground and excited state energies. However, the computational cost of SCI approaches is heavily dependent on the amount of correlation present, as this necessarily increases the dimension of the important subspace of determinants. Although these cited applications have been on small to medium molecules, the SCI variational spaces for these systems are already in the tens of millions. For larger systems, the problem will quickly become intractable for SCI-based approaches.

Fortunately, the dimension of the important variational space is not an intrinsic characteristic of a given Hamiltonian but is rather a basis-dependent quantity. For a trivial example, consider the case where one first rotates the basis into the exact eigenbasis. In this basis, the relevant variational space has a dimension equal to one. As such, it is possible to decrease the size of the variational space by “simply” choosing a more appropriate basis in which to represent the problem. With orbital rotations being the simplest change of basis possible where the many-electron transformation is parametrized by simple one-electron functions, SCI calculations are often performed using the natural orbitals computed from either a cheaper SCI calculation or other single reference methods like CCSD or MP2. Even though this does generally lead to a smaller variational dimension compared to using canonical Hartree–Fock orbitals, the improvements are often rather limited to a factor of 2 or so.^{26,27} Recently, orbital optimization has also been proposed to improve the SCI energies with respect to the number of determinants in a given orbital space.^{27–30}

Along this direction, we recently introduced a new method called Tensor Product Selected Configuration Interaction (TPSCI)³¹ which defines a SCI algorithm not in a Slater determinant basis but rather, in a basis of tensor product states of locally entangled many-body wave functions. This amounts to a change of basis, where many-body rotations are applied locally to the basis of Slater determinants, folding in local electron correlation into the basis functions themselves.

Traditional Selected CI methods are memory limited due to the size of the variational space needed to reach a target accuracy. In TPSCI, our goal is to trade off some run time (TPSCI calculations are significantly slower than Slater determinant methods) for reduced memory requirements (TPSCI variational spaces are generally much smaller than those of Slater determinant methods).

Other methods such as active space decomposition (ASD)^{32,33} and rank-1 matrix product states^{34,35} also have a similar framework, operating in a similar tensor product space. In ASD, the rapid growth of the Hilbert space was controlled with a low-rank matrix product state (MPS) approximation instead of the sparsity-based approximation used in our current work. While an MPS approximation can be effective for compressing a state, it does impose an often artificial one-dimensional entanglement structure. In the rank-1 matrix product state method, the global states are written as a linear combination of entangled states, similar to TPSCI but mainly focusing on disjointed molecular units. A broad list of methods exists which focuses on forming the wave function of the full system in this clustered framework, including Block Correlated Coupled Cluster (BCCC)³⁶ and the related Tensor Product State Coupled Electron Pair-Type Approximations (TPS-

CEPA),³⁷ the cMF-based coupled cluster,³⁸ the ab initio Frenkel–Davydov model,^{39,40} the renormalized exciton model (REM),^{41,42} the Block Interaction Product State (BIPS),⁴³ the comb-Tensor network states based approach by Li,⁴⁴ and the generalized and localized active space methods.^{45–47}

In this work, we extend our recently proposed TPSCI methodology³¹ to provide near-FCI approximations to relatively large manifolds of excited states in a limited basis of active orbitals.

2. METHODS

The core strategy in TPSCI is to build a localized representation that increases the sparsity of the target global eigenstates. Let us start by assuming that our orbital active space permits partitioning into smaller, disjoint active spaces (referred to as “clusters” throughout). While clusters can be defined through different considerations (locality, orbital entanglement,⁴⁸ symmetry etc.), the general guideline is that intracluster interactions should be stronger than intercluster interactions.

Within each cluster, we want to define a many-body transformation⁴⁹ that accounts for all relevant local correlations. In principle, one can obtain such a transformation by simply diagonalizing the local Hamiltonian (the terms that remain after the removal of operators that act outside of the cluster). However, this explicitly neglects the influence of neighboring clusters on the composition of our many-body transformations. We instead include the influence of intercluster interactions in a mean-field fashion by adopting the cluster Mean Field (cMF) method that was introduced by Scuseria and co-workers^{38,50} and explored by Gagliardi and co-workers.^{51,52}

This mean-field treatment arises (analogously to the Hartree–Fock theory) by variationally minimizing the energy of a single tensor product state (TPS) with respect to both orbital and local many-body rotations (defined by a set of local configuration interaction coefficients). As such, cMF can be understood as a CASSCF problem with multiple active spaces, similar to generalized active space or occupation-restricted active space methods.^{45,46,53} We will express the cMF ground-state wave function as

$$|\psi_0\rangle = |0_I\rangle|0_J\rangle \dots |0_N\rangle = |0_I 0_J \dots 0_N\rangle \quad (1)$$

where $I, J, \dots$ label clusters, and $|0_I\rangle$ is the lowest energy eigenstate of the cMF effective Hamiltonian on cluster I

$$\begin{aligned} \hat{H}_I^{\text{eff}} = & \sum_{pq} h_{pq} \hat{p}^\dagger \hat{q} + \frac{1}{2} \sum_{pqrs} \langle pq|rs \rangle \hat{p}^\dagger \hat{q}^\dagger \hat{q} \hat{r} \\ & + \sum_{pr} \sum_{J \neq I} \sum_{qs} \langle pq||rs \rangle \gamma_{qs}^J \hat{p}^\dagger \hat{r} \end{aligned} \quad (2)$$

where $\hat{p}$ ($\hat{p}^\dagger$) are the Fermionic annihilation (creation) operators on orbital p , γ_{qs}^J is an element of the one-particle reduced density matrix (1RDM) on cluster J , and h_{pq} , $\langle pq|rs \rangle$, and $\langle pq||rs \rangle$ are the one-electron, simple two-electron, and antisymmetrized two-electron integrals, respectively. The local cMF effective Hamiltonian (arising naturally from tracing out the remaining clusters) commutes with $\hat{N}$, $\hat{S}_z$, and $\hat{S}^2$, and as such, the cluster states, $|\alpha_I\rangle$, automatically preserve particle number and spin symmetries. Because the $\hat{H}_I^{\text{eff}}$ depends on all other clusters via the 1RDM, this must be solved self-consistently. The similarities between $\hat{H}_I^{\text{eff}}$ and the traditional

Fock operator also extend to our ability to define a perturbation theory, as introduced in ref 50, and discussed later. For small clusters, this many-body transformation can simply be defined through the exact diagonalization (FCI) of $\hat{H}_i^{\text{eff}}$, although approximate eigenstates would be needed for larger clusters.

In order to span the full Hilbert space of the global system, we must separately diagonalize eq 2 in all possible sectors of the cluster's local Fock space. The global states can then be represented in the tensor product basis of cMF eigenstates

$$|\Psi\rangle = \sum_{\alpha} \sum_{\beta} \dots \sum_{\omega} c_{\alpha,\beta,\dots,\omega} |\alpha_1\rangle |\beta_2\rangle \dots |\omega_N\rangle \quad (3)$$

where $c_{\alpha,\beta,\dots,\omega}$ is the coefficient tensor, and $|\alpha_i\rangle$ is an eigenvector of eq 2.

The focus of this paper is to develop and test an excited-state generalization of our selected CI procedure (TPSCI)³¹ which algorithmically builds a sparse approximation to eq 3. The remaining theory section is organized as follows: In Section 2.1, we discuss clustering and how to generate initial cluster states by diagonalizing local Hamiltonians, Section 2.2 provides details about the matrix element evaluation, and finally, in Section 2.3, we explain in detail the steps of the TPSCI algorithm.

2.1. Generating Initial Cluster States. **2.1.1. cMF Orbital Optimization.** Because the cMF orbitals are optimized in addition to the cluster state coefficients, the final definition of the clusters is ultimately determined uniquely by the variational principle. However, a good initial guess is often necessary for the reliable orbital convergence of cMF, similar to CASSCF. There are several ways to generate an initial guess for orbital clusters, and the best choice is often system dependent. The most straightforward (yet tedious) approach would be to localize the active space orbitals and manually put them into clusters. Alternatively, one could use more automated strategies (based on graph theoretic algorithms or embedding-inspired approaches). Exploring and comparing these options will be the focus of future studies.

In this work, we use a simple DIIS procedure to optimize the orbitals,^{54,55} where the orbital rotation gradient is taken as the error vector. For each set of orbitals, the local CI coefficients are optimized self-consistently, and the optimized 1RDMs and 2RDMs are used to construct the new orbital gradient. As such, our optimization is a two-step procedure consisting of an inner "CI" optimization and an outer "orbital" optimization. Inclusion of the orbital Hessian and directly coupling orbital and CI coefficient degrees of freedom would significantly improve the convergence. However, for this paper, the cMF is not the computational bottleneck, so we defer this to future work.

2.1.2. Initial Computation of the Local Cluster State Basis. The result of the cMF calculation is not only the variationally best tensor product state but also a set of cluster-local effective Hamiltonians dressed in the mean-field interactions of the other clusters. We chose the eigenvectors of these effective Hamiltonians as our initial cluster state basis. While the default setting in our Julia implementation includes all possible electron numbers for a given cluster, functionality has been added to allow the user to define a net change (δ_e) in particle number for each cluster. This removes the cost associated with Fock sectors that will ultimately be insignificant in the final TPSCI wave function. Once the allowed particle number subspaces (Fock sectors) are defined, the eigenvectors of the local cMF effective Hamiltonians are obtained in each cluster. A total of M (a user-defined parameter) eigenvectors are computed for each Fock

sector and saved in memory as basis vectors. Because these eigenvectors diagonalize a local Hamiltonian, all of the local correlations are folded into the basis vectors.

2.1.2.1. Spin Completeness of the Basis. In our initial TPSCI paper,³¹ we computed the lowest M states for each requested sector of Fock space, treating each m_s block independently. Because the different m_s blocks have different dimensions, truncating with a fixed M necessarily introduces spin-contamination into the global basis. While this was not a significant problem in our first paper focusing on ground states, for excited states, spin-contamination can become more significant. To reduce this spin-contamination in the final TPSCI state (and to save both computational time and memory), in this newer implementation, we simply generate the high- and low- m_s components by directly applying the spin raising and lowering operators, S^+ and S^- , to the $m_s = 0$ (even number of electrons) and $m_s = \frac{1}{2}$ (odd number of electrons) eigenstates. This ensures that all m_s components are included for each cluster state computed, such that M state truncation does not break the $\hat{S}_z$ symmetry. Figure 1 shows a depiction of this process.

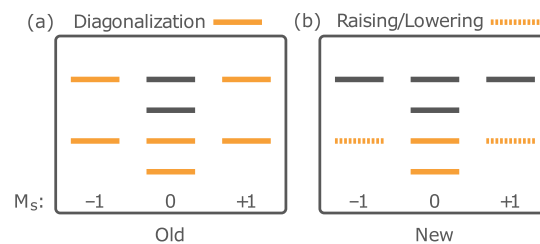


Figure 1. Comparison between the (a) previous approach and (b) new approach for obtaining cluster states across different m_s sectors. Gray (discarded states). Solid orange (states obtained via diagonalization). Dashed orange (states obtained by application of ladder operators). Here, we have an example of where a user selected to keep 2 states ($M = 2$), and a cluster which has a singlet, triplet, singlet, and triplet state ordering. In the old approach (a), our truncation would have incomplete treatment of the second triplet state, whereas the new approach (b) is spin-complete.

While this approach ensures that truncating M does not create spin-contamination, there is still the possibility of creating spin-contamination in the SCI selection. This is a direct analogy to the situation with spin-contamination in determinant-based selected CI codes. We have added the ability to add the important TPS needed for achieving spin-complete wave functions by simply adding the dominant contributions from the $\hat{S}^2$ residual vector, $|r_{\hat{S}^2}\rangle = (\hat{S}^2 - \langle\hat{S}^2\rangle)|\Psi^s\rangle$. However, for future work, we will likely attempt to generalize the recent work from Scemama and co-workers to also reduce the variational dimension.⁵⁶

2.2. Matrix Element Evaluation. Diagonalization directly on the basis of tensor product states requires us to evaluate Hamiltonian matrix elements between arbitrary tensor product states. To save on time-complexity during the Hamiltonian evaluation, we precompute the representation of all the relevant local operators in the cluster basis. Following the relevant notation used in the ASD work,³² we refer to them as Γ tensors. To aid in the explanation of these tensors, we first introduce the standard electronic Hamiltonian in second quantization

$$\hat{H} = \sum_{pq} h_{pq} \hat{p}^\dagger \hat{q} + \frac{1}{2} \sum_{pqrs} \langle pq|rs \rangle \hat{p}^\dagger \hat{q}^\dagger \hat{s} \hat{r} \quad (4)$$

where $\hat{p}^\dagger$ and $\hat{q}$ are the Fermionic creation and annihilation operators, and h_{pq} and $\langle pq|rs \rangle$ are the one and two electron integrals, respectively. We partition the electronic Hamiltonian into one, two, three, and four cluster contributions, defined by the number of distinct clusters being acted upon

$$\begin{aligned} \hat{H} = & \sum_I \hat{H}_I + \sum_{I<J} \hat{H}_{IJ} \\ & + \sum_{I<J<K} \hat{H}_{IJK} + \sum_{I<J<K<L} \hat{H}_{IJKL} \end{aligned} \quad (5)$$

where H_I has all creation and annihilation operators in cluster I , H_{IJ} has operators in both clusters I and J and so on. The full set of terms for each of these n -cluster interactions is included in the [Supporting Information](#). At most, we can only have four cluster interactions since we have at most four Fermionic operators.

Each of these terms involves a contraction of the two-electron integrals with the appropriate Γ tensors. Therefore, we can precompute these terms and store them in a dictionary in memory for later access. For example, if we have a local state β in cluster I where operators $\hat{p}^\dagger$, $\hat{q}^\dagger$, and $\hat{r}$ act on cluster I , its associated Γ tensor is the following

$${}^I\Gamma_{pqr}^{\beta'\beta} = \langle \beta'_I | \hat{p}^\dagger \hat{q}^\dagger \hat{r} | \beta_I \rangle \quad (6)$$

This is also an example of the largest rank Γ tensor that our implementation will store in memory. For large clusters and large M values (number of local cluster states), these can become a memory bottleneck. It is, in principle, possible to avoid the storage of these five-index tensors since they can only contribute to two-cluster terms; however, we have not found the need yet.

These gamma tensors are contracted with the integrals during the computation of each Hamiltonian matrix element. For example, the following $\hat{H}_{IJ}$ term would provide the following contribution to the $\langle \psi | \hat{H}_{IJ} | \psi \rangle$ matrix element

$$\begin{aligned} \langle \psi | \hat{H}_{IJ} | \psi \rangle \leftarrow & - (-1)^\chi \prod_{K \neq I, J} \delta_{\omega_K, \omega'_K} \\ & \times \sum_{pqr \in I} \sum_{s \in J} \langle pq|rs \rangle {}^I\Gamma_{pqr}^{\beta'\beta} \Gamma_s^{\gamma'\gamma} \end{aligned} \quad (7)$$

where $\chi = \sum_{K=I}^{J-1} N_K$ and accounts for the sign by summing over the number of electrons in each cluster between the two active clusters, and $\delta_{\omega_K, \omega'_K}$ arises from the orthonormality between states ω and ω' on cluster K . There is an additional negative sign that arises from the anticommutator relationship when you switch the two annihilation operators $\hat{s}$ and $\hat{r}$, since the operators must be adjacent to the cluster they are acting upon.

The orthonormality of the cluster states creates sparsity in the Hamiltonian, such that we only need to compute contributions between tensor product states that have identical inactive clusters states. Analogous to the Slater–Condon rules, only tensor product states that differ by less than 5 clusters can be coupled by the Hamiltonian.

2.3. Algorithm. We start by first listing the overall steps for the TPSCI algorithm (which can also be seen in [Figure 2](#)) and then follow with a more detailed discussion of each step. We also

include a table of the required user-defined parameters for a TPSCI calculation in [Table 1](#).

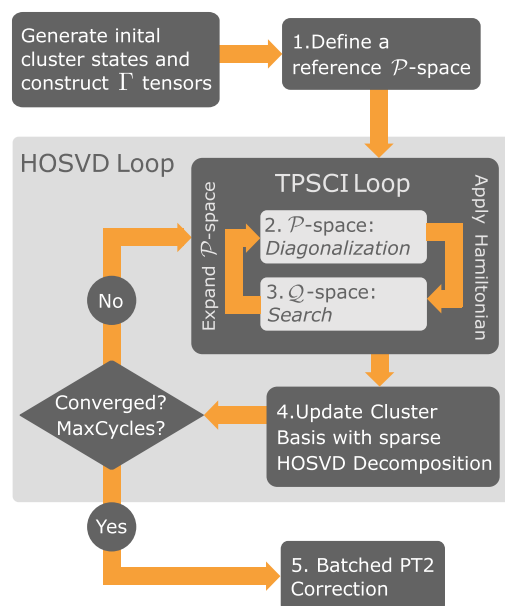


Figure 2. Flowchart of the TPSCI algorithm including the HOSVD loop.

Table 1. Table of Definitions of Parameters Used to Define a TPSCI Calculation

N	number of clusters
R	number of global eigenvectors requested
M	maximum number of cluster states in any given sector of Fock space for any cluster
δ_c	range of Fock sectors for each cluster to include; for example, if cluster I has 10 electrons in the cMF reference, then compute cluster states for $10 - \delta_c \rightarrow 10 + \delta_c$
ϵ_{CIPSI}	threshold for discarding first-order TPS coefficients; coefficients larger than this value will be included in the variational space
ϵ_{FOIS}	threshold for screening when computing the first-order interaction space; values larger than this will be included when computing the first-order wave function

Steps of a TPSCI calculation for computing R states:

- 1 Define a reference $\mathcal{P}$ -space with dimension of at least R . ([Section 2.3.1](#))
- 2 Diagonalize the Hamiltonian in the $\mathcal{P}$ -space and collect lowest R eigenstates. ([Section 2.3.2](#))
- 3 Search Q -space perturbatively and expand $\mathcal{P}$ -space. If converged, continue, else return to step 2. ([Section 2.3.3](#))
- 4 Update cluster basis with sparse higher-order singular value decomposition (HOSVD) decomposition. If converged, continue, else return to step 2. ([Section 2.3.4](#))
- 5 Compute a state-specific PT2 energy correction. ([Section 2.3.5](#))

2.3.1. Define a Reference $\mathcal{P}$ Space. For ground-state TPSCI calculations, the cMF wave function often serves as a sufficient initial $\mathcal{P}$ space. However, for excited states, it is often helpful to specify an initial $\mathcal{P}$ space that qualitatively describes the target states.

If the system were to be fully decoupled such that there were no interactions between clusters, then the full Hamiltonian would be diagonal in the TPS basis. Additionally, the low-energy

spectrum would be dominated by “excitonic” states, those states where every cluster is in its ground state except for a single (or pair) that is excited. However, as the clusters become more strongly interacting, the low energy spectrum can develop greater weight on higher exciton-rank tensor products. For weakly to moderately interacting clusters, the excitonic basis provides a qualitatively correct description of the target excited states and thus is an excellent initial $\mathcal{P}$ space for starting the TPSCI procedure. The single excitonic basis for a given cluster can be written as

$$|\psi_{\lambda_L}\rangle = |0_I, 0_J, \dots, \lambda_L, \dots, 0_N\rangle \quad (8)$$

where cluster L is in its singly excited-state λ . For very weakly interacting systems, one would expect the low-energy states to be primarily represented as linear combinations of these single excitonic states. For comparative purposes, we will refer to such a method as the TPS-single exciton (TPS-SE). This is equivalent to the so-called Block correlated CI method described by Li and co-workers.³⁶ Although the TPS-SE results will not generally be accurate since it lacks all interactions with higher excited configurations (e.g., charge-transfer excitations), the TPS-SE method provides a very effective way to initialize the TPSCI calculation with a qualitatively correct initial $\mathcal{P}$ space. Further, for situations where we expect biexcitons to contribute to the final wave functions (see Section 3.2), the user can also directly add these configurations to the starting wave function.

We provide a comparison of the TPS-SE with TPSCI for one of the systems we studied (P1) in the Supporting Information, which demonstrates that one does generally need to go beyond TPS-SE for accurate excited states.

2.3.2. $\mathcal{P}$ -Space: Diagonalization. Once the variational space is defined, we build the Hamiltonian from eq 5 in the $\mathcal{P}$ space and diagonalize. As described above, the required matrix element evaluation is much more expensive than traditional Slater determinant methods due to two main reasons: (i) the loss of sparsity of the Hamiltonian matrix and (ii) the need to contract the integrals with the precomputed Γ tensors mentioned in Section 2.2. Because the Hamiltonian matrix storage usually constitutes a memory bottleneck, we have implemented the option for either a full matrix build or a matrix-vector product build for use in a Krylov solver. However, while the matrix-vector algorithm significantly reduces the memory requirements, it is much slower because it recomputes the matrix elements for each Lanczos iteration. As such, if allowed by memory, our current implementation defaults to the full Hamiltonian matrix build. After we build and diagonalize the Hamiltonian, we have a set of variational states that are a sum of tensor product states, $|\mathcal{P}^s\rangle = \sum_i c_i^s |\mathcal{P}_i\rangle$ and a variational energy, E_0 .

2.3.3. $\mathcal{Q}$ -Space: Search. To obtain the first-order interacting space (FOIS), we calculate the action of the Hamiltonian on the set of tensor product states, $\{\mathcal{P}_i\}$, in the current variational space.

$$\begin{aligned} |\mathcal{Q}_j^s\rangle &= \sum_i |\mathcal{Q}_j\rangle \langle \mathcal{Q}_j | \hat{H} | \mathcal{P}_i \rangle c_i^s \\ &= |\mathcal{Q}_j\rangle b_j^s \end{aligned} \quad (9)$$

The states, $|\mathcal{Q}_j\rangle$, run over all tensor product states that can be reached by the Hamiltonian from the current variational space, excluding the variational space. Since the Hamiltonian is not sparse on the TPS basis, the action of the Hamiltonian on the TPS states can become very costly. Therefore, we have

implemented a series of screening and prescreening techniques based on a user-defined threshold, ϵ_{FOIS} , where we delete components $|\mathcal{Q}_j\rangle$, if $\max_s |b_j^s| \leq \epsilon_{\text{FOIS}}$. We then collect the resulting non-negligible configurations that lie in the $\mathcal{Q}$ space.

Consistent with the original Slater determinant Configuration Interaction Perturbatively Selected Iteratively (CIPSI)¹⁹ method, we compute the first-order correction to our current variational state(s) to determine which new degrees of freedom should be added to our variational space. In our work, we use a generalization of the Barycentric Møller–Plesset¹⁹ (MP) perturbation theory using the cMF effective Hamiltonian (eq 2), which is explicitly described in Appendix A.

Once the first-order coefficients are computed for each state, $|\mathcal{P}^s\rangle$

$$c_j^{s(1)} = \sum_i \frac{\langle \mathcal{Q}_j | \hat{H} | \mathcal{P}_i \rangle c_i^{s(0)}}{\Delta E_j^{(0)}} \quad (10)$$

any $\mathcal{Q}$ space configuration with a perturbative coefficient greater than $\epsilon_{\text{CIPSI}} (\max_s |c_j^{s(1)}| > \epsilon_{\text{CIPSI}})$ is added to the $\mathcal{P}$ space.⁵⁷ If no additional TPS states are added to the variational space, then the TPSCI protocol is considered converged.

2.3.4. Update Cluster Basis with HOSVD Decomposition. Once the TPSCI wave function has converged (i.e., no additional TPS states are required in the variational space), we can optionally update the cluster basis using a quantum number-preserving Tucker decomposition called a higher-order SVD decomposition (HOSVD)

$$\mathcal{T}_{\alpha,\beta,\dots,\gamma} = C_{i,j,\dots,d} U_{\alpha,i} U_{\beta,j} \dots U_{\gamma,d} \quad (11)$$

where $\alpha, \beta, \dots, \gamma$ are each specific to a cluster, and $C_{i,j,\dots,d}$ is the core tensor which is formed by a change of basis from α to i, β to j , etc. Because we are using the HOSVD to only rotate the cluster basis and not truncate the space,^{31,58} each U is a unitary matrix in the vector space of its specified cluster. These unitary matrices are local many-body rotations which can be directly obtained from individual singular value decompositions (SVD) along the associated axis, e.g.,

$$\mathcal{T}_{\alpha,\beta,\dots,\gamma} = U_{\alpha,i} \Sigma_i V_{i,\beta,\dots,\gamma} \quad (12)$$

or equivalently, by diagonalizing the cluster reduced density matrix (cluster-RDM) which is obtained by tracing out the remaining clusters from the converged TPSCI wave function.

$$\rho_{\alpha\alpha'} = c(\alpha, \beta, \dots, \gamma) c(\alpha', \beta, \dots, \gamma) \quad (13)$$

where $c(\alpha, \beta, \dots, \gamma)$ is the TPS coefficient vector. We note that, in practice, we want to preserve certain local quantum numbers (particle number and spin projection). As such, we only block-diagonalize the cluster-RDM within each quantum number subspace. This ensures that the global wave function retains the proper eigenstates of both $\hat{N}$ and $\hat{S}_z$.

When moving to a multistate problem, there are various ways to complete this HOSVD to obtain the tucker factors (U). One option is to decompose each state into its own basis. However, this state-specific approach would be extremely complex, making it difficult to reliably compute energy differences and transition properties between states. Instead, we compute a single global basis in a state-averaged way. To create this global basis, we simply average the cluster-RDMs from each TPSCI eigenvector

$$\rho_{\alpha\alpha'} = \frac{1}{R} \sum_s \sum_{\beta, \dots, \gamma} c(\alpha, \beta, \dots, \gamma)^s c(\alpha', \beta, \dots, \gamma)^s \quad (14)$$

where s is denoting the state. We can then diagonalize $\rho_{\alpha\alpha'}$ to obtain the tucker factors for cluster α

$$\rho_{\alpha\alpha'} = U_{\alpha,i} g_i U_{\alpha',i} \quad (15)$$

We view the use of the HOSVD as optional, analogous to the use of natural orbitals in conventional Slater determinant selected CI calculations. As such, it is obtained iteratively, where cheaper calculations provide states that are decomposed to produce more compact representations for subsequent calculations with tighter thresholds, ϵ_{CIPSI} . We refer to this computational protocol of systematically tightening the thresholds after one (or more) HOSVD steps as “HOSVD bootstrapping” in the [Results and Discussion](#) section.

2.3.5. Batched PT2 Energy Correction. Even though the Selected-CI algorithm captures most of the static correlation with the CI expansion, it does not capture the dynamic correlation efficiently enough to produce near FCI accuracy results. The inclusion of the missing dynamic correlation is usually carried out using a state-specific PT2 correction. It is important to note that we are only referring to dynamical correlation inside of the active space; out-of-active space correlations would still need to be accounted for (through either downfolding, PT2, or adiabatic connection type schemes) to enable direct comparison to the experiment. As mentioned in [Section 2.3.3](#), in the cMF basis, we choose a Barycentric Møller–Plesset¹⁹ (MP) type partitioning in this work. Whereas computing the first-order wave function can quickly become a memory bottleneck due to the vast size of the Q space, the energy computation has no inherent memory demand.

For computing the PT2 energy correction, we have implemented a parallelized batched algorithm, where we compute a small segment, or batch, of the first-order wave function, then contract it to evaluate the energy, discarding the state before moving to the next segment. Our current implementation batches over what we refer to as `FockConfig`'s or unique distributions of particles across clusters. This approach is analogous to the determinant-based approach described in [ref 59](#). While this does offer system-dependent speedups, the scaling is far from optimal. The reason is that by parallelizing over Fock space configurations, we have rather poor load balancing due to the fact that some Fock space configurations have many more configurations than others. Improvements to our batching will be the focus of future work.

3. RESULTS AND DISCUSSION

We investigated the efficiency of the TPSCI approach for excited states by mainly focusing on polycyclic aromatic hydrocarbons (PAH). These systems have been chosen for three reasons: (i) they provide a straightforward approach to orbital clustering, allowing us to defer the more complicated clustering patterns to focused future work, (ii) we have already begun to understand the ground-state behavior in our previous paper,³¹ and (iii) because they are chemically interesting in terms of novel material in synthesizing chiral nanographenes,⁶⁰ twisted carbon nanobelts,⁶¹ and carbon-based electronic devices⁶² etc. Benchmarking a wider variety of chemical systems will be the focus of follow-up papers.

The first few systems ([Section 3.1](#)) constitute a set of π conjugated systems that can be grouped into clusters of six orbitals which simply differ in their connectivity. The last

example is a tetracene tetramer, which is noncovalently bound and supports interesting multiexcitonic states. For all systems, we compute accurate estimates of both the ground state and a large number of excited states.

For the PAH systems, we use geometries optimized at the B3LYP/cc-pVDZ⁶³ level of theory. The active space for each PAH system P1–P5 is generated by extracting the molecular orbitals that have a significant $2p_z$ atomic orbital character, then localized using the Boys⁶⁴ localization method. We use the 6-31G*⁶⁵ basis for the singlet fission tetracene tetramer calculations. The active space for the tetracene tetramer system is generated by first obtaining a set of natural orbitals obtained by diagonalizing a state-averaged CIS density (CIS-NO),⁶⁶ then these are localized using the Pipek-Mezey⁶⁷ localization method. All semistochastic heat-bath CI (SHCI) calculations were performed with Arrow.^{68–71} The integrals for all calculations were generated using the PySCF package,⁷² and the cMF and TPSCI calculations were performed with our open-source Julia⁷³ packages `ClusterMeanField.jl`⁷⁴ and `FermiCG`.⁷⁵ The geometries for all of the systems are included in the [Supporting Information](#).

It is important to note that the calculations reported in this section are not capable of being directly compared to experiments. While we believe they are highly accurate inside the active spaces, more work is needed to provide direct comparisons with experiments. In particular, it will be necessary to include the missing sigma-bond correlation,⁷⁶ dynamic correlation outside of the active space, larger basis set effects, and vibronic effects to make sure that our calculations are directly comparable to the experiment.

We note that the thresholding used in the original ground-state TPSCI work³¹ pruned by using the probability and hence was square of the ϵ_{CIPSI} in this work. The current work prunes on the absolute value of the first-order coefficients to be more consistent with other selected CI codes.

3.1. PAH Systems. We present four medium sized PAH systems (P1–P4) and one larger system (P5) ([Figure 3](#)). Taking the π space as the active space, P1–P4 have an active space of size 24 electrons in 24 orbitals (4 clusters), and P5 has an active space of 36 electrons in 36 orbitals (6 clusters). Considering that the low-energy excited states of benzene consist of one singlet state and three triplet states, in our calculations on P1–P4, we compute 16 total excited states, while for P5 we compute 24 excited states (i.e., four states per cluster).

3.1.1. Smaller PAH Systems (P1–P4). In [Figure 4](#), we present the extrapolation of the ground and 16 excited states for systems P1–P4 using TPSCI where the ground state is shown in navy blue, triplets in blue, and singlets are colored orange.

As is commonly done in selected CI calculations, we assume a linear relationship between the PT2 energy correction and variational energy (i.e., the larger the cutoff, the cheaper the selected CI calculation, therefore more energy correction will be required). The extrapolated results in [Figure 4](#) were computed by first converging to the tightest ϵ_{CIPSI} possible (4×10^{-4} for P1, P3, and P4 and 6×10^{-4} for P2) through the HOSVD bootstrapping approach. The additional cheaper points for extrapolation were obtained by deleting TPs with a coefficient smaller than a specified epsilon value and then recomputing the eigenvectors and PT2 corrections in these successively smaller variational spaces. We note that the same cluster basis is used for each point in the extrapolations, i.e., we do not perform any additional HOSVD for the extrapolation points. This allows us to track states and monitor if any root flips are observed across

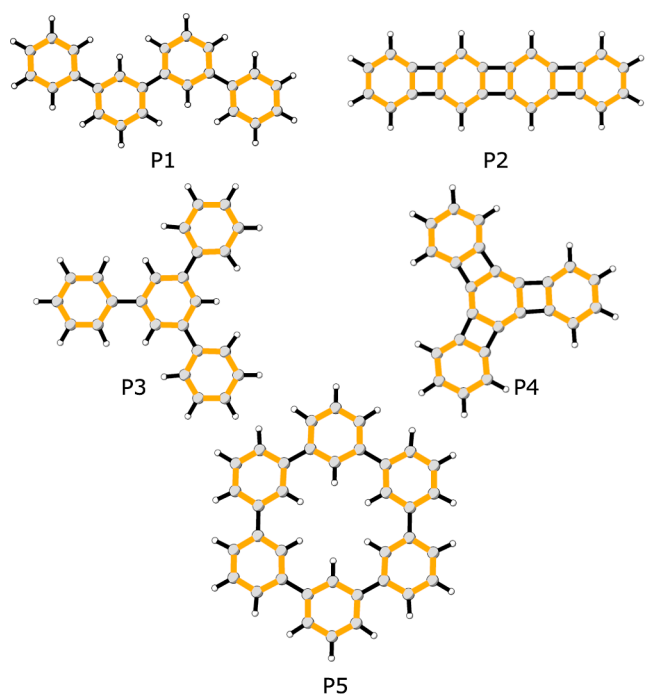


Figure 3. PAH systems used for the excitation energies. Each gold highlighted region corresponds to a separate cluster.

extrapolation points. The point where the extrapolated lines cross the y -axis (i.e., where the variational energy is predicted to have a zero PT2 correction) is our best estimate of the FCI energies. Therefore, the closer the variational energies are to the extrapolated result (y -axis), the more reliable the calculation.

For each of the P1–P4 systems, the ground-state variational estimate converges much faster than the excited states, as seen from Figure 4. In future work, we plan to investigate the use of iso-PT2 selection schemes to help improve the convergence of excited states.^{77,78}

The TPSCI results for the singly connected systems, P1 and P3, converge much faster than those for the P2 and P4 systems. This is to be expected given the fact that each cluster is connected by two bonds instead of one, leading states in the P2 and P4 systems to develop significantly more intercluster entanglement.

Overall, we see that qualitatively, the low-energy electronic structure of the clusters is retained when the system is more weakly coupled than otherwise. For instance, for P1, P3, and P4, there are three triplets for every singly excited benzene unit (4 singlets and 12 triplets). This is the same ratio as that found in the isolated benzene structure. In contrast, for the P2 system, we observe 7 singlets and 9 triplets within the lowest 16 states. We interpret this increase in singlet contribution to arise from the increased interactions between the clusters, which provides more ability for the electronic structure to delocalize between clusters.

Although P4 also has clusters which are connected by two bonds, the nonlinear geometry prevents the qualitative reorganization of the electronic structure such that there are still 4 singlets and 12 triplets. Further, unlike P2, the singlet–triplet gap is not significantly lowered compared to that of P1 or P3.

We note that in the P2 extrapolated graph, we observe a very steep slope for one of the states around 4.0 eV. This could indicate that this state was not converged tightly enough for extrapolation. Alternatively, this might have arisen from the

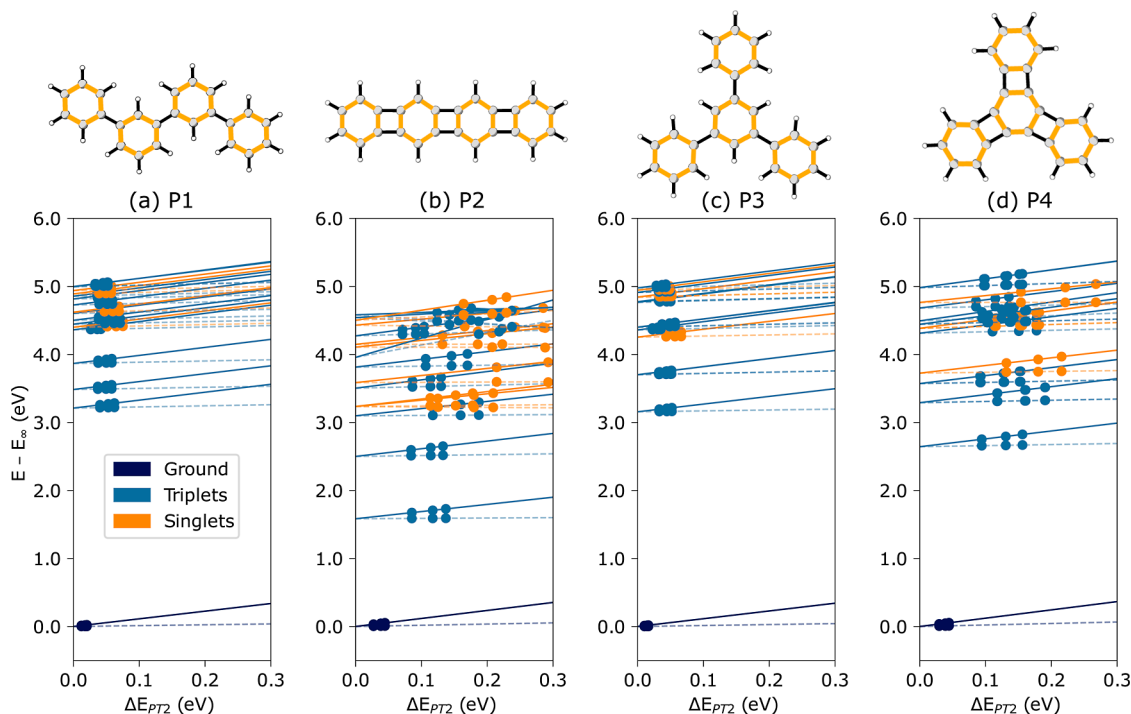


Figure 4. Extrapolation of the ground state and 16 excited states for the medium-sized PAH systems: (a) P1, (b) P2, (c) P3, and (d) P4, studied using TPSCI (with HOSVD bootstrapping). After the bootstrapping, the TPSCI wave function coefficients are clipped at the larger thresholds to obtain the additional points in the extrapolation to plot against the PT2 energy correction with root tracking. All energies are shifted by the extrapolated ground-state TPSCI energy, so the ground state converges to 0 eV. Variational energy fit (solid lines). PT2 energy fit (dashed lines). ($\epsilon_{\text{CIPSI}} = n \times 10^{-4}$ with $n = 4, 6, 8$ for P1, P3, and P4. $\epsilon_{\text{CIPSI}} = n \times 10^{-4}$ with $n = 6, 8, 10$ for P2).

manner in which we apply perturbation theory. As mentioned above, we are currently using a nondegenerate PT2 formalism, which can create problems in cases of near degeneracy. In the follow-up work, we plan on implementing a quasi-degenerate formalism,^{79–84} following a strategy similar to our recent work,^{85,86} to better understand the current results and to safeguard against such issues in the future.

3.1.2. Larger PAH (P5). As a larger example of a π -conjugated system, we also consider P5, which has an active space of 36 electrons in 36 orbitals that is partitioned into 6 clusters. Similar to the case before, we expect three triplets and one singlet for every singly excited cluster, giving a total of 25 states. We present the extrapolations of both TPSCI and semistochastic heat bath CI (SHCI) in Figure 5. The linear extrapolation has been shown

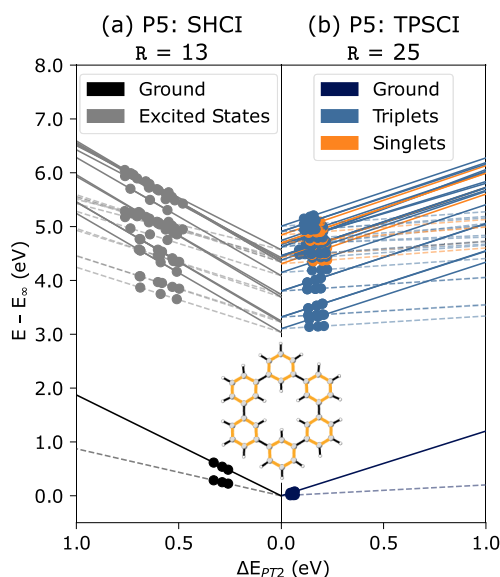


Figure 5. Extrapolation for the P5 molecule using the (a) SHCI and (b) TPSCI methods, respectively. R denotes the number of roots: 13 for SHCI and 25 for TPSCI. (TPSCI $\epsilon_{\text{CIPSI}} = n \times 10^{-4}$ with $n = 4, 6, 8$ and SHCI $\epsilon_{\text{CIPSI}} = n \times 10^{-5}$ with $n = 5, 7, 10$).

previously in the literature to generate overestimated energies.²⁰ A quadratic fit is recommended in these cases, but for comparison, we use a linear fit for both methods.

In order to label the eigenstates, we compute the expectation value of $\hat{S}^2$ for each of the TPSCI states. Although the TPSCI results are rather tightly converged, nearly degenerate states can mix arbitrarily, leading to a few instances of nontrivial spin-contamination. However, the extent of this is generally small enough that it does not prevent us from labeling the states.⁸⁷ Using SHCI, we were not able to compute all 25 states because the variational space grew to be too large to fit in memory. The largest calculation we were able to obtain was for 13 roots. Further, we did not have access to $\langle \hat{S}^2 \rangle$ values for SHCI, so we were not able to label the resulting states, and thus they are simply left gray in Figure 5.

In addition to the plots in Figure 5, we also present these results in Table 2. Here, we report the variational excitation energies (ω_{var}), magnitude of the PT2 correction to the excitation energies ($\Delta\omega_{PT2} = \omega_{PT2} - \omega_{\text{var}}$), and the extrapolated excitation energies (ω_∞). To better highlight the accuracy of the perturbatively corrected results, we also present the extrap-

olation corrections ($\Delta\omega_\infty = \omega_\infty - \omega_{PT2}$) for the excitation energies.

For all excited states computed, the TPSCI variational energy is closer to its extrapolated result than the corresponding variational HCI result. This is a consequence of folding local correlations directly into the TPS basis. Not only do the TPSCI results have smaller PT2 corrections ($\Delta\omega_{PT2}$) compared to SHCI, but more importantly, the extrapolation correction is significantly smaller than the PT2 correction for each state, $\Delta\omega_{PT2} > \Delta\omega_\infty$. In contrast, this is not the case for the SHCI results, where the extrapolation corrections are consistently larger than the PT2 corrections. For all excitation energies, the magnitude of $\Delta\omega_{PT2}$ for SHCI is around a factor of 3 times that of TPSCI.

The fact that the TPSCI variational (and perturbative) results are closer to the extrapolated values lends greater confidence to the extrapolated values. This is extra important in situations where different methods yield extrapolations that differ nontrivially, as seen in Figure 5. While the overall features are similar between SHCI and TPSCI, the extrapolated values differ by a non-negligible amount (up to around 100 meV). Because of the fact that our extrapolation is smaller, we expect that the TPSCI extrapolations are closer to the exact FCI results than are the SHCI extrapolations.⁸⁸

3.2. Singlet Fission: Tetracene Tetramer. Singlet fission is a multichromophoric process in which a bright singlet excited state is converted into two lower energy triplets. The mechanism involves an entangled multiexciton singlet state, $^1(TT)$.⁸⁹ While the $^1(TT)$ state is likely the first multiexciton state to be accessed, due to spin conservation, it has been recently shown that the triplet and quintet multiexcitons, $^3(TT)$ and $^5(TT)$, also play an important role in the separation process.⁹⁰ Because of the intrinsic two-electron nature of the multiexcitonic state, it is difficult to compute all three spin states of the multiexciton, the initial singlet excitation, and the final triplet states on equal footing. However, because of the underlying product structure of the target states,⁹¹ tensor product state methods offer unique advantages. Since the chromophores are naturally partitioned into different clusters, a diabatic basis can be naturally formed using the cluster states.³³ Here we test our tensor product-based method on a tetracene tetramer taken from a tetracene crystal that exhibits this singlet fission process.

To construct an orbital active space that accurately represents the targeted states, we performed a CIS calculation for the first four singlets and triplets, then built a state-averaged one particle reduced density matrix (1RDM) and diagonalized to obtain the natural orbitals.⁹² Using the eigenvalues of the state-averaged 1RDM, we take the 40 most correlated orbitals (those with the most fractional occupations) as our active space. While a larger active space would have been possible in principle, this is the largest active space that was tractable when treating each chromophore (10 electrons in 10 orbitals) with an exact FCI cluster solver. In future work, we will report on a restricted active space configuration interaction (RASCI) cluster solver to increase the size of the clusters (and thus active spaces) treatable. After defining the (40o, 40e) orbital active space, we constructed an initial guess through localization, then variationally optimized the cluster orbitals with cMF, defined by 4 clusters each with 10 electrons ($5\alpha + 5\beta$) in 10 orbitals.

3.2.1. Extrapolation. We used the same technique that was used for the PAH systems to obtain the extrapolated plots seen in Figure 6. In subplot (a) of Figure 6, we show 31 states that were calculated using TPSCI. We label the states based on the

Table 2. Excitation Energies (eV) and Wavefunction Dimension for the Most Accurate Calculation Reported for the P5 System Using TPSCI and SHCI ($\epsilon_{\text{CIPSI}} = 4 \times 10^{-4}$ for TPSCI and $\epsilon_{\text{CIPSI}} = 5 \times 10^{-5}$ for SHCI)^a

state	TPSCI				SHCI			
	dimension: 112,788				dimension: 1,741,084			
	ω_{Var}	$\Delta\omega_{\text{PT2}}$	$\Delta\omega_{\infty}$	ω_{∞}	ω_{Var}	$\Delta\omega_{\text{PT2}}$	$\Delta\omega_{\infty}$	ω_{∞}
1	3.22	−0.09	−0.02	3.10	3.68	−0.25	−0.39	3.04
2	3.43	−0.09	−0.02	3.32	3.86	−0.24	−0.37	3.25
3	3.43	−0.09	−0.02	3.32	3.93	−0.27	−0.43	3.22
4	3.91	−0.08	−0.02	3.80	4.37	−0.25	−0.38	3.74
5	3.91	−0.08	−0.02	3.80	4.40	−0.27	−0.44	3.69
6	4.28	−0.11	−0.03	4.15	4.82	−0.30	−0.44	4.09
7	4.46	−0.11	−0.04	4.31	4.94	−0.22	−0.34	4.38
8	4.52	−0.04	−0.01	4.46	5.00	−0.26	−0.40	4.34
9	4.52	−0.11	−0.03	4.38	5.03	−0.31	−0.29	4.43
10	4.58	−0.11	−0.03	4.44	5.07	−0.27	−0.40	4.39
11	4.58	−0.11	−0.04	4.44	5.11	−0.30	−0.47	4.34
12	4.59	−0.10	−0.03	4.45	5.15	−0.29	−0.29	4.57
13	4.59	−0.10	−0.03	4.46				
14	4.74	−0.08	−0.04	4.62				
15	4.74	−0.08	−0.04	4.62				
16	4.77	−0.09	−0.01	4.67				
17	4.78	−0.09	−0.01	4.68				
18	4.83	−0.10	−0.03	4.70				
19	4.83	−0.10	−0.03	4.70				
20	4.87	−0.09	−0.03	4.74				
21	4.97	−0.09	−0.03	4.85				
22	5.00	−0.07	−0.02	4.91				
23	5.00	−0.07	−0.02	4.91				
24	5.09	−0.07	−0.02	5.01				

^aLineally extrapolated results were obtained from PT2 energy corrections. ω_{Var} is the variational excitation energy; $\Delta\omega_{\text{PT2}}$ is the PT2 energy correction to the excitation energy; $\Delta\omega_{\infty}$ is the extrapolation correction; and ω_{∞} is the extrapolated excitation energy.

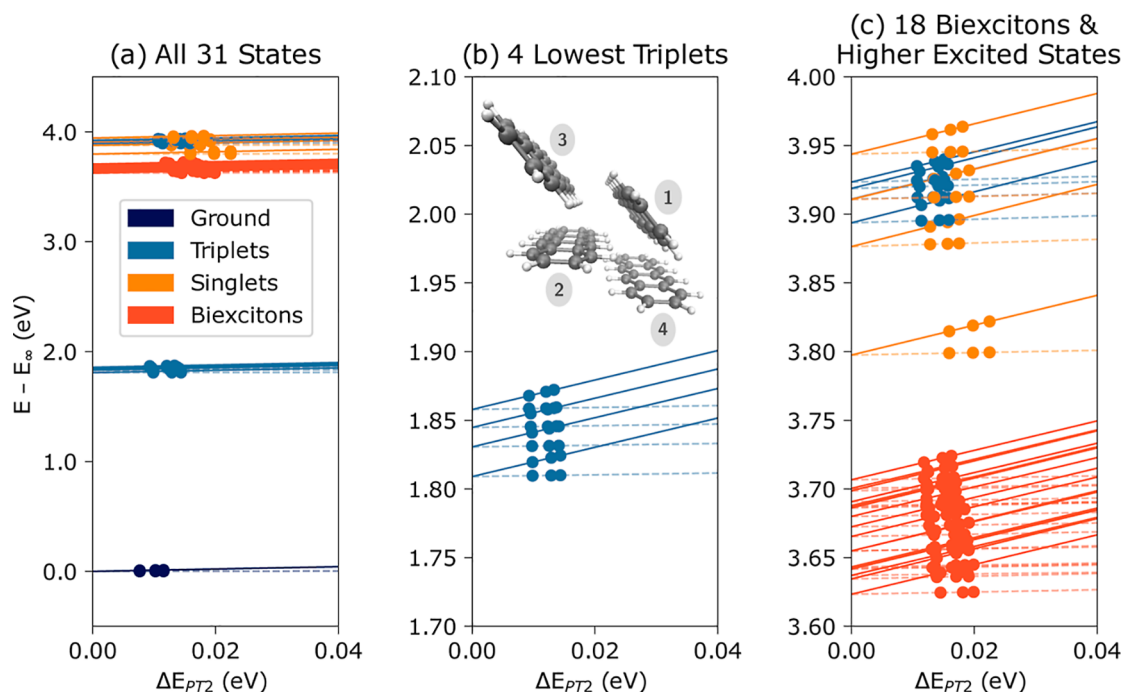


Figure 6. Extrapolated results using HOSVD bootstrapping then clip at larger thresholds to obtain extrapolation for tetracene tetramer singlet-fission example with root tracking. (a) Full spectra with 31 roots shown. (b) Middle section of the energy spectra with 4 triplet excited states and cluster labels of the tetracene tetramer. (c) Top portion of the spectra with the remaining 26 roots shown.

expectation value of the $\hat{S}^2$ operator and the dominant Fock space configurations in each eigenstate. The singlet ground state is denoted in navy blue; triplets are in blue, singlets (bright states) are in orange, and the biexcitons are in red. The orientation of the tetracene tetramer is shown in a herringbone lattice where the chromophores are stacked and then shifted slightly from one another. We observe faster convergence for the lower excited states compared to the biexcitons and higher single excitons, which follows the intuition that the higher excited state manifold is generally more entangled, thus requiring more TPS configurations to converge. In subplot (b) of Figure 6, we reduce the energy scale to highlight the four lowest energy triplet states. Because these states are largely “singly excitonic” in nature, their rate of convergence is similar to that of the ground state.

In subplot (c), the energy scale is changed to highlight the higher energy states, and all 18 biexcitons, four singlets, and four higher excited triplet states (primarily superpositions of T2 single excitons) are shown. As expected, the biexciton spectrum (shown in red) is relatively dense compared with the higher energy states.

Due to the fact that we have used a fixed orbital active space of 40 orbitals and neglected external correlation, our excitation energies are expected to be significantly overestimated compared to the experiment. For example, the experimental value of the bright state lies at 2.35 eV.⁹³ The significant difference between the experimental value and our computed results is primarily due to both the inadequate composition of our active space and missing vibronic effects, not from errors within the active space.

In the future, we would plan to use TPSCI as a CASSCF solver, allowing us to use state averaging so that our active space orbitals treat the ground and excited states on an equal footing. We note that the orbital optimization during the cMF calculation only mixes the active 40 orbitals among themselves, but this could be extended to mix all of the orbitals. In addition to orbital optimization, we will also consider the inclusion of dynamical correlation via an operator downfolding (e.g., DUCC^{94,95}) or by doing a PT2 type correction. Finally, we can also increase the size of our active space beyond this 40 orbital example. This will require a more efficient cluster state solver to allow us to exceed the 10 orbitals per cluster used in this calculation. A RASCI cluster solver enabling larger clusters will be reported in a subsequent article.

3.2.2. Wave Function Analysis. To analyze the TPSCI wave function, we have access to the expectation value of $\hat{S}^2$, number of important configurations in each Fock space, and the associated weight of that Fock space in the overall TPS wave function. In Table 3, we present a summarized version of the wave function analysis. For each state, we list the state label, $\langle \hat{S}^2 \rangle$, the variational excitation energies, the PT2 corrections, and the overall percent charge-transfer character.

The state label is defined by the $\langle \hat{S}^2 \rangle$ and the dominant Fock space configurations. While the states were generally easy to label, the presence of near-degeneracies between states of different spin multiplicities creates difficulties in resolving spin states accurately, as our variational space would need to be converged to within the energy gap between the states. While we could always “un-mix” them manually by just diagonalizing the $2 \times 2\hat{S}^2$ matrix, we have not investigated that in this study.

By analyzing the total weight of Fock sectors having clusters with different numbers of electrons, we can quantify the amount

Table 3. Results for all 31 Eigenstates of Tetracene Tetramer with Associated Labels Based on Expectation Values of the $\hat{S}^2$ Operator $\langle \hat{S}^2 \rangle$, Variational Excitation Energies (ω_{Var}), PT2 Energy Corrections (ω_{PT2}) for Excitation Energies in eV, and Percentage of Charge Transfer (% CT) for all 31 Eigenstates in the TPSCI Wavefunction

state	label	$\langle \hat{S}^2 \rangle$	ω_{Var}	ω_{PT2}	% CT
1	S ₀	0.000	0	0	0.09
2	T ₁	2.000	1.811	0.002	1.36
3	T ₁	2.000	1.833	0.002	0.17
4	T ₁	2.000	1.847	0.002	0.40
5	T ₁	2.000	1.860	0.002	0.36
6	¹ (TT)	0.001	3.631	0.007	3.81
7	¹ (TT)	0.091	3.642	0.006	1.95
8	³ (TT)	1.919	3.643	0.006	1.53
9	⁵ (TT)	5.945	3.648	0.006	1.36
10	¹ (TT)	0.234	3.649	0.006	4.23
11	³ (TT)	1.811	3.650	0.006	3.02
12	³ (TT)	1.838	3.661	0.005	3.01
13	¹ (TT)	0.163	3.661	0.006	0.98
14	³ (TT)	2.000	3.672	0.006	0.65
15	⁵ (TT)	5.999	3.678	0.005	1.21
16	⁵ (TT)	5.998	3.685	0.005	0.79
17	⁵ (TT)	5.995	3.691	0.005	0.52
18	¹ (TT)	0.302	3.692	0.005	0.52
19	³ (TT)	1.718	3.693	0.005	0.50
20	⁵ (TT)	5.986	3.696	0.005	0.39
21	¹ (TT)	0.050	3.704	0.005	0.96
22	³ (TT)	1.961	3.705	0.005	0.89
23	⁵ (TT)	5.989	3.711	0.004	0.42
24	S ₁	0.000	3.807	0.008	10.80
25	S ₁	0.000	3.883	0.005	1.92
26	T ₂	2.000	3.898	0.004	2.03
27	T ₂	2.000	3.915	0.003	0.36
28	S ₁	0.001	3.917	0.006	2.55
29	T ₂	2.000	3.923	0.003	0.98
30	T ₂	2.000	3.927	0.003	2.32
31	S ₁	0.000	3.950	0.005	5.03

of charge transfer present in a given state. Following this approach, we observe a significant amount of charge transfer present in the first bright state (state 24) with 10.8% charge transfer. We analyze this charge-transfer character in the first singlet excited state more carefully in Table 4, where we list charge-transfer Fock space contributions that are overall greater than 0.001 in the final TPSCI wave function. As shown in this table, the localized representation of the TPSCI method makes the analysis more direct. Not only can we quantify the amount of CT character, but we can also further decompose it into individual CT contributions. For example, in Table 4 we can see that while charge transfers between clusters 1 and 2 are of a “charge resonance” type, where the transfers are equal in both directions, the CT interactions between clusters 1 and 4 are more asymmetrical, with more electron density moving from 1 to 4 than in the opposite direction.

4. CONCLUSIONS

In this work, we generalize our Tensor Product Selected Configuration Interaction algorithm to enable the computation of excited states. TPSCI has the ability to provide extremely

Table 4. Charge-Transfer Wavefunction Analysis for First Singlet Excited State^a

Fock space (α, β)	#configs	weight	CT character
(5,5)(5,5)(5,5)(5,5)	11,157	0.87	no CT
(4,5)(5,5)(5,5)(6,5)	1150	0.018	1 $\rightarrow$ 4 (α)
(5,4)(5,5)(5,5)(5,6)	1135	0.018	1 $\rightarrow$ 4 (β)
(5,4)(5,6)(5,5)(5,5)	856	0.017	1 $\rightarrow$ 2 (β)
(4,5)(6,5)(5,5)(5,5)	897	0.016	1 $\rightarrow$ 2 (α)
(6,5)(4,5)(5,5)(5,5)	843	0.015	2 $\rightarrow$ 1 (α)
(5,6)(5,4)(5,5)(5,5)	876	0.015	2 $\rightarrow$ 1 (β)
(6,5)(5,5)(5,5)(4,5)	922	0.004	4 $\rightarrow$ 1 (α)
(5,6)(5,5)(5,5)(5,4)	865	0.004	4 $\rightarrow$ 1 (β)
(5,5)(6,5)(4,5)(5,5)	1114	0.004	3 $\rightarrow$ 2 (α)
(5,5)(5,6)(5,4)(5,5)	1099	0.004	3 $\rightarrow$ 2 (β)
(5,5)(4,5)(6,5)(5,5)	951	0.003	2 $\rightarrow$ 3 (α)
(5,5)(5,4)(5,6)(5,5)	938	0.003	2 $\rightarrow$ 3 (β)

^aThese are the charge-transfer Fock space configurations that contribute with a weight of >0.001 and are in descending order by their contributions.

accurate (near-exact FCI) results for strongly correlated systems that would otherwise be intractable for Slater determinant-based methods. We use our own extrapolated results within the limited basis of active orbitals as a benchmarking tool for our TPSCI method, as no experimental or FCI results can be obtained for the numerous excited states computed.

We demonstrated the accuracy of TPSCI for excited states on a series of small PAH molecules with active spaces of 24 electrons in 24 orbitals. The excitation energies are within 1 kcal/mol or 0.043 eV once the state-specific PT2 correction is added for the P1, P3, and P4 systems when compared to our nearly exact extrapolated results. The P2 system was a very challenging system due to the additional connectivity that increased intercluster interactions. Nonetheless, all but four higher energy excited states were within 0.043 eV of the extrapolated results. We see from the extrapolated results for P2, one of the higher excited states has an extreme slope, suggesting the further development of a quasidegenerate PT2 formulation. We then extend TPSCI to one larger PAH system, P5, with an active space of 36 electrons in 36 orbitals, and compare it to a semistochastic heat bath CI (SHCI). All TPSCI excitation energies for P5 are extremely close to the nearly exact extrapolated result. When we compare TPSCI to SHCI for the P5 system, we are able to calculate an additional 12 states with TPSCI. Furthermore, all variational excitation energies with TPSCI are closer to their respective extrapolated results compared to how far away the HCI excitation energies are from their extrapolated values. TPSCI also has smaller PT2 corrections when compared with SHCI for all states.

After testing TPSCI on smaller PAH systems and comparing it to SHCI, we investigated TPSCI's ability to compute "beyond-dimer model" singlet fission excited states of a tetracene tetramer cluster. We chose an activated space of 40 electrons in 40 orbitals with a total of four clusters (one for each tetracene). Our model has all three spin states of the dark multiexciton state as well as the singlet excited states. We calculated the ground state and 30 excited states (eight triplets, four singlets, and 18 biexcitons). All variational excitation energies are extremely accurate and even closer to exact results with the PT2 energy correction (only a 0.001 eV difference). We are able to label our states from both the $\langle \hat{S}^2 \rangle$ values and TPS wave function analysis.

In addition to accurately solving large active spaces for several roots, the TPSCI wave function further allows analysis of charge-transfer characteristics and multiexciton states. This analysis will be extended to produce quantitated diabatic bases and subsequent effective Hamiltonians. The TPS representation also makes analysis in terms of quantum information quantities, such as von Neumann entropy, very natural. These directions, in addition to the construction of properties and RDMs, will be the focus of future work.

In order to extend TPSCI to larger active spaces, it will be necessary to use an approximate solver within the cluster, like the restricted active space approach, which will be the focus of a future article. In addition to improved cluster solvers, automation of the orbital clustering is also needed to minimize the amount of user input needed to setup a calculation. Even though we report excitation energies near the exact limit within the active space, we have not yet included any influence from the higher lying virtual orbitals, which is necessary for recovering dynamic correlation. Including this external dynamic correlation will be the focus of future work through either downfolding or PT2 treatments. Orbital optimization with the TPSCI method is also a possible future direction to provide CASSCF values for large active spaces. Even without these suggested future directions, TPSCI has the ability to study ground states, excited states, charge-transfer states, and multiexciton states for large, strongly correlated systems and hopes to serve as an accurate method to benchmark against systems that are intractable with FCI.

■ APPENDIX A

Definition of Perturbation Theory

The perturbation theory used in this work is defined by using Löwdin's partitioning theory. We seek a correction to the zeroth-order wavefunction for state s , which is constructed as a linear combination of TPSs that lie within the $\mathcal{P}$ space. We refer to this reference state as $|\mathcal{P}^s\rangle$. To partition the Hamiltonian for perturbative treatment

$$\hat{H} = \hat{H}^{(0)} + \lambda \hat{H}^{(1)} \quad (\text{A1})$$

we wish to choose a partitioning where the zeroth-order contribution contains the full Hamiltonian in the $\mathcal{P}$ space but an approximate, diagonal Hamiltonian in the $\mathcal{Q}$ space. This is achieved by the following partitioning

$$\hat{H}^{(0)} = \left(\begin{array}{c|c} \hat{H} & 0 \\ \hline 0 & \hat{F}_D^{\text{cMF}} + \langle \mathcal{P}^s | \hat{V}^{\text{cMF}} | \mathcal{P}^s \rangle \end{array} \right) \quad (\text{A2})$$

$$\hat{H}^{(1)} = \left(\begin{array}{c|c} 0 & \hat{H} \\ \hline \hat{H} & \hat{H} - \hat{F}_D^{\text{cMF}} - \langle \mathcal{P}^s | \hat{V}^{\text{cMF}} | \mathcal{P}^s \rangle \end{array} \right), \quad (\text{A3})$$

where $\hat{H} = \hat{F}_D^{\text{cMF}} + \hat{V}^{\text{cMF}}$, and

$$\hat{F}^{\text{cMF}} = \sum_I \hat{H}_I^{\text{eff}} \quad (\text{A4})$$

and where the subscript D ($\hat{F}_D^{\text{cMF}}$) indicates the diagonal of the operator (this is only consequential if one is working in the HOSVD basis because the cMF effective Hamiltonian is already diagonal in the cMF basis). This is referred to as "barycentric" partitioning because the zeroth-order Hamiltonian contains the

reference state expectation value of the “Fock-like” cMF Hamiltonian.

With this partitioning, the expression for the first-order coefficients becomes

$$c_j^{s(1)} = \frac{\langle Q_j | \hat{H} | \mathcal{P}^s \rangle}{\langle \mathcal{P}^s | \hat{F}^{\text{cMF}} | \mathcal{P}^s \rangle - \langle Q_j | \hat{F}_D^{\text{cMF}} | Q_j \rangle} \quad (\text{A5})$$

While other partitionings are likely to work better (Epstein–Nesbet^{96,97} or even a quasidegenerate formulation of the above approach^{85,86}), we defer consideration of different partitionings to future work.

APPENDIX B

Computing the First-Order Wavefunction

The first-order correction shows up in two distinct places in the TPSCI algorithm: Step 3 and Step 5 in Section 2.3. In Step 5, the state-specific PT2 energy correction is computed, but since we only compute the final second-order energy, we do not actually ever need the full first-order wavefunction. Consequently, we obtain only small batches of the first-order wavefunction (external configurations that share the same FockConfig), then compute a batch's contribution to the energy, then immediately discard the corresponding first-order amplitudes. In this way, computing the PT2 correction in Step 5 does not really create a memory bottleneck. However, in Step 3 when we perturbatively search the Q -space to expand our $\mathcal{P}$ -space, we do need to compute the first-order wavefunction. To do this, we use prescreening to reduce the number of terms we have to consider.

To build the first-order wavefunction, we apply the Hamiltonian to our variational space, but because the Hamiltonian connects a single TPS with up to a quartic number of new TPS's ($O(M^4)$), the first-order interaction space quickly becomes intractable to store in memory. For example, consider a single 4-body term in the Hamiltonian (i.e., each Fermionic creation/annihilation operator is acting on a different cluster) applied to state s , $|\Psi^s\rangle$, is given below as

$$|\sigma^s\rangle = \hat{H}|\Psi^s\rangle \quad (\text{B1})$$

$$|\sigma^s\rangle \leftarrow \hat{H}_{1,2,3,4}|\Psi^s\rangle \quad (\text{B2})$$

$$\begin{aligned} \sigma_{\alpha'\beta'\gamma'\delta'\epsilon}^s & \leftarrow \langle pqlrs | \langle \Gamma_{\hat{p}}^{\alpha'\alpha} | \alpha' \rangle \langle \alpha' | \alpha \otimes \Gamma_{\hat{q}}^{\beta'\beta} | \beta' \rangle \langle \beta' | \beta \otimes \Gamma_{\hat{s}}^{\gamma'\gamma} | \gamma' \rangle \\ & \langle \gamma' | \gamma \otimes \Gamma_{\hat{r}}^{\delta'\delta} | \delta' \rangle \langle \delta' | \delta \rangle | \alpha \beta \gamma \delta \epsilon \rangle c_{\alpha\beta\gamma\delta\epsilon}^s \end{aligned} \quad (\text{B3})$$

$$\sigma_{\alpha'\beta'\gamma'\delta'\epsilon}^s \leftarrow \langle pqlrs | \Gamma_{\hat{p}}^{\alpha'\alpha} \Gamma_{\hat{q}}^{\beta'\beta} \Gamma_{\hat{s}}^{\gamma'\gamma} \Gamma_{\hat{r}}^{\delta'\delta} c_{\alpha\beta\gamma\delta\epsilon}^s = H_{\alpha\beta\gamma\delta}^{\alpha'\beta'\gamma'\delta'} c_{\alpha\beta\gamma\delta\epsilon}^s \quad (\text{B4})$$

Because the variational states, $|\Psi^s\rangle$, are always represented in a sparse basis, this operation is performed element wise over the configurations in the variational space. To denote this, we will rewrite the above equation, highlighting the fact that the right-hand side indices are specified

$$\sigma_{****\epsilon}^s \leftarrow \sum_{pqrs} \langle pqlrs | \Gamma_{\hat{p}}^* \Gamma_{\hat{q}}^* \Gamma_{\hat{s}}^* \Gamma_{\hat{r}}^* c^s \quad (\text{B5})$$

where the * symbol is used to denote the full range of values of the associated index. The above equation reveals the potential bottleneck of computing the first-order wavefunction. Suppose $M = 400$, (i.e., the number needed to keep all states for a six

orbital six electron cluster), a single TPS in the variational space could couple to such a large number of configurations that it would require 200 Gb of memory to simply store a single contribution, $\sigma_{****\epsilon}^s$. To avoid this, we can perform a series of screenings based on the following inequalities.

Assuming the following relationship

$$A_{ijkl} = B_{abcd} C_{ai} D_{bj} E_{ck} F_{dl} \quad (\text{B6})$$

FOIS screening inequality 1

$$|A_{****}|_{\infty} \leq \sum_{abcd} |B_{abcd}| \cdot |C_{a*}|_{\infty} \cdot |D_{b*}|_{\infty} \cdot |E_{c*}|_{\infty} \cdot |F_{d*}|_{\infty} \quad (\text{B7})$$

FOIS screening inequality 2

$$|A_{****}|_{\infty} \leq \sum_{abcd} |B_{abcd}| \cdot |C_{a*}|_{\infty} \cdot |D_{b*}|_{\infty} \cdot |E_{c*}|_{\infty} \cdot |F_{dl}| \quad (\text{B8})$$

Using inequality 1 allows us to determine if an entire block of contributions will have values all smaller than a user-specified threshold ϵ_{FOIS} , allowing us to avoid the computation altogether. Inequality 2 allows us to predetermine which cluster state indices are capable of contributing. This means that after determining if the entire block is not negligible (by using inequality 1), we can prune the number of cluster states, creating a smaller, effective M value, so that the terms we actually compute in eq B5 have a lower percentage of discarded values. After prescreening the indices α' , β' , γ' , and δ' , we compute the screened block of σ contributions. Finally, we filter out all values with magnitudes less than ϵ_{FOIS} , before storing the contribution to σ in memory. While the screening does incur some overhead for tighter values of ϵ_{FOIS} , for the looser values that are often used in Step 3 (e.g., $\epsilon_{\text{FOIS}} = 1e - 5$ or $1e - 6$), the screening can significantly speed up a calculation without having a significant impact on the results. In future work, we will study the interplay of thresholds and performance to better understand how the current screening procedure works and to potentially improve on this rather straightforward approach.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpca.3c03161>.

Tensor product state single exciton data analysis for the P1 polyaromatic hydrocarbon system, TPSCI versus TPSCI with HOSVD decomposition extrapolations for the tetracene tetramer system, extrapolation plots to show convergence of the user defined parameter: maximum number of roots (M), variational extrapolations of medium-sized PAH systems (P1–P4) on a smaller x -axis, enumeration of all distinct hamiltonian terms (PDF) xyz geometry coordinates for all systems (TXT)

AUTHOR INFORMATION

Corresponding Author

Nicholas J. Mayhall – Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24060, United States; orcid.org/0000-0002-1312-9781; Email: nmayhall@vt.edu

Authors

Nicole M. Braunscheidel – Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24060, United States; orcid.org/0000-0001-7555-9173

Vibin Abraham — Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jpca.3c03161>

Author Contributions

[§]N.M.B. and V.A. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation (Award no. 1752612).

REFERENCES

- (1) Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.
- (2) Marques, M.; Gross, E. TIME-DEPENDENT DENSITY FUNCTIONAL THEORY. *Annu. Rev. Phys. Chem.* **2004**, *55*, 427–455.
- (3) Furche, F.; Ahlrichs, R. Adiabatic time-dependent density functional methods for excited state properties. *J. Chem. Phys.* **2002**, *117*, 7433–7447.
- (4) Dreuw, A.; Head-Gordon, M. Failure of time-dependent density functional theory for long-range charge-transfer excited states: the zincbacteriochlorin- bacteriochlorin and bacteriochlorophyll- spheroidene complexes. *J. Am. Chem. Soc.* **2004**, *126*, 4007–4016.
- (5) Levine, B. G.; Ko, C.; Quenneville, J.; Martínez, T. J. Conical intersections and double excitations in time-dependent density functional theory. *Mol. Phys.* **2006**, *104*, 1039–1051.
- (6) Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation energies in density functional theory: An evaluation and a diagnostic test. *J. Chem. Phys.* **2008**, *128*, 044118.
- (7) Maitra, N. T. Double and Charge-Transfer Excitations in Time-Dependent Density Functional Theory. *Annu. Rev. Phys. Chem.* **2021**, *73*, 117.
- (8) Maitra, N. T.; Zhang, F.; Cave, R. J.; Burke, K. Double excitations within time-dependent density functional theory linear response. *J. Chem. Phys.* **2004**, *120*, 5932–5937.
- (9) Elliott, P.; Goldson, S.; Canahui, C.; Maitra, N. T. Perspectives on double-excitations in TDDFT. *Chem. Phys.* **2011**, *391*, 110–119.
- (10) Geertsens, J.; Rittby, M.; Bartlett, R. J. The equation-of-motion coupled-cluster method: Excitation energies of Be and CO. *Chem. Phys. Lett.* **1989**, *164*, 57–62.
- (11) Comeau, D. C.; Bartlett, R. J. The equation-of-motion coupled-cluster method. Applications to open-and closed-shell reference states. *Chem. Phys. Lett.* **1993**, *207*, 414–423.
- (12) Watts, J. D.; Gwaltney, S. R.; Bartlett, R. J. Coupled-cluster calculations of the excitation energies of ethylene, butadiene, and cyclopentadiene. *J. Chem. Phys.* **1996**, *105*, 6979–6988.
- (13) Loos, P.-F.; Boggio-Pasqua, M.; Scemama, A.; Caffarel, M.; Jacquemin, D. Reference Energies for Double Excitations. *J. Chem. Theory Comput.* **2019**, *15*, 1939–1956.
- (14) Roos, B. O.; Taylor, P. R.; Sigbahn, P. E. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **1980**, *48*, 157–173.
- (15) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.
- (16) Finley, J.; Malmqvist, P. Å.; Roos, B. O.; Serrano-Andrés, L. The multi-state CASPT2 method. *Chem. Phys. Lett.* **1998**, *288*, 299–306.
- (17) Siegbahn, P. E. M. The externally contracted CI method applied to N₂. *Int. J. Quantum Chem.* **1983**, *23*, 1869–1889.
- (18) Werner, H.; Reinsch, E. The self-consistent electron pairs method for multiconfiguration reference state functions. *J. Chem. Phys.* **1982**, *76*, 3144–3156.
- (19) Huron, B.; Malrieu, J. P.; Rancurel, P. Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions. *J. Chem. Phys.* **1973**, *58*, 5745–5759.
- (20) Holmes, A. A.; Umrigar, C. J.; Sharma, S. Excited states using semistochastic heat-bath configuration interaction. *J. Chem. Phys.* **2017**, *147*, 164111.
- (21) Loos, P.-F.; Scemama, A.; Blondel, A.; Garniron, Y.; Caffarel, M.; Jacquemin, D. A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks. *J. Chem. Theory Comput.* **2018**, *14*, 4360–4379.
- (22) Chien, A. D.; Holmes, A. A.; Otten, M.; Umrigar, C. J.; Sharma, S.; Zimmerman, P. M. Excited States of Methylene, Polyenes, and Ozone from Heat-Bath Configuration Interaction. *J. Phys. Chem. A* **2018**, *122*, 2714–2722.
- (23) Loos, P.-F.; Lipparini, F.; Boggio-Pasqua, M.; Scemama, A.; Jacquemin, D. A Mountaineering Strategy to Excited States: Highly Accurate Energies and Benchmarks for Medium Sized Molecules. *J. Chem. Theory Comput.* **2020**, *16*, 1711–1741.
- (24) Loos, P.-F.; Scemama, A.; Boggio-Pasqua, M.; Jacquemin, D. Mountaineering strategy to excited states: Highly accurate energies and benchmarks for exotic molecules and radicals. *J. Chem. Theory Comput.* **2020**, *16*, 3720–3736.
- (25) Damour, Y.; Quintero-Monsebaiz, R.; Caffarel, M.; Jacquemin, D.; Kossoski, F.; Scemama, A.; Loos, P. F. Ground- and Excited-State Dipole Moments and Oscillator Strengths of Full Configuration Interaction Quality. *J. Chem. Theory Comput.* **2023**, *19*, 221–234.
- (26) Zhang, N.; Liu, W.; Hoffmann, M. R. Iterative Configuration Interaction with Selection. *J. Chem. Theory Comput.* **2020**, *16*, 2296–2316.
- (27) Levine, D. S.; Hait, D.; Tubman, N. M.; Lehtola, S.; Whaley, K. B.; Head-Gordon, M. CASSCF with Extremely Large Active Spaces Using the Adaptive Sampling Configuration Interaction Method. *J. Chem. Theory Comput.* **2020**, *16*, 2340–2354.
- (28) Smith, J. E. T.; Mussard, B.; Holmes, A. A.; Sharma, S. Cheap and Near Exact CASSCF with Large Active Spaces. *J. Chem. Theory Comput.* **2017**, *13*, 5468–5478.
- (29) Yao, Y.; Umrigar, C. J. Orbital Optimization in Selected Configuration Interaction Methods. *J. Chem. Theory Comput.* **2021**, *17*, 4183–4194.
- (30) Damour, Y.; Vêril, M.; Kossoski, F.; Caffarel, M.; Jacquemin, D.; Scemama, A.; Loos, P.-F. Accurate full configuration interaction correlation energy estimates for five- and six-membered rings. *J. Chem. Phys.* **2021**, *155*, 134104.
- (31) Abraham, V.; Mayhall, N. J. Selected Configuration Interaction in a Basis of Cluster State Tensor Products. *J. Chem. Theory Comput.* **2020**, *16*, 6098–6113.
- (32) Parker, S. M.; Shiozaki, T. Communication: Active Space Decomposition with Multiple Sites: Density Matrix Renormalization Group Algorithm. *J. Chem. Phys.* **2014**, *141*, 211102.
- (33) Parker, S. M.; Seideman, T.; Ratner, M. A.; Shiozaki, T. Communication: Active-Space Decomposition for Molecular Dimers. *J. Chem. Phys.* **2013**, *139*, 021108.
- (34) Nishio, S.; Kurashige, Y. Rank-one basis made from matrix-product states for a low-rank approximation of molecular aggregates. *J. Chem. Phys.* **2019**, *151*, 084110.
- (35) Nishio, S.; Kurashige, Y. Importance of dynamical electron correlation in diabatic couplings of electron-exchange processes. *J. Chem. Phys.* **2022**, *156*, 114107.
- (36) Li, S. Block-correlated coupled cluster theory: The general formulation and its application to the antiferromagnetic Heisenberg model. *J. Chem. Phys.* **2004**, *120*, 5017–5026.
- (37) Abraham, V.; Mayhall, N. J. Coupled Electron Pair-Type Approximations for Tensor Product State Wave Functions. *J. Chem. Theory Comput.* **2022**, *18*, 4856–4864.
- (38) Papastathopoulos-Katsaros, A.; Jiménez-Hoyos, C. A.; Henderson, T. M.; Scuseria, G. E. Coupled Cluster and Perturbation Theories Based on a Cluster Mean-Field Reference Applied to Strongly Correlated Spin Systems. *J. Chem. Theory Comput.* **2022**, *18*, 4293–4303.

- (39) Morrison, A. F.; You, Z.-Q.; Herbert, J. M. Ab Initio Implementation of the Frenkel-Davydov Exciton Model: A Naturally Parallelizable Approach to Computing Collective Excitations in Crystals and Aggregates. *J. Chem. Theory Comput.* **2014**, *10*, 5366–5376.
- (40) Morrison, A. F.; Herbert, J. M. Low-Scaling Quantum Chemistry Approach to Excited-State Properties via an ab Initio Exciton Model: Application to Excitation Energy Transfer in a Self-Assembled Nanotube. *J. Phys. Chem. Lett.* **2015**, *6*, 4390–4396.
- (41) Al Hajj, M.; Malrieu, J.-P.; Guihéry, N. Renormalized excitonic method in terms of block excitations: Application to spin lattices. *Phys. Rev. B* **2005**, *72*, 224412.
- (42) Ma, Y.; Liu, Y.; Ma, H. A new fragment-based approach for calculating electronic excitation energies of large systems. *J. Chem. Phys.* **2012**, *136*, 024113.
- (43) Wang, K.; Xie, Z.; Luo, Z.; Ma, H. Low-Scaling Excited State Calculation Using the Block Interaction Product State. *J. Phys. Chem. Lett.* **2022**, *13*, 462–470.
- (44) Li, Z. Expressibility of comb tensor network states (CTNS) for the P-cluster and the FeMo-cofactor of nitrogenase. *Electron. Struct.* **2021**, *3*, 014001.
- (45) Ma, D.; Li Manni, G.; Gagliardi, L. The generalized active space concept in multiconfigurational self-consistent field methods. *J. Chem. Phys.* **2011**, *135*, 044128.
- (46) Weser, O.; Guthrie, K.; Ghanem, K.; Li Manni, G. Stochastic Generalized Active Space Self-Consistent Field: Theory and Application. *J. Chem. Theory Comput.* **2022**, *18*, 251–272.
- (47) Hermes, M. R.; Pandharkar, R.; Gagliardi, L. Variational localized active space self-consistent field method. *J. Chem. Theory Comput.* **2020**, *16*, 4923–4937.
- (48) Barcza, G.; Legeza, O.; Marti, K. H.; Reiher, M. Quantum-information analysis of electronic states of different molecular structures. *Phys. Rev. A* **2011**, *83*, 012508.
- (49) This transformation can be unitary if we retain the size of the local Hilbert space by not discarding any degrees of freedom. However, more generally, and for larger clusters, this transformation will discard high energy states, making it an isometry mapping a large space to a smaller space, retaining lengths.
- (50) Jiménez-Hoyos, C. A.; Scuseria, G. E. Cluster-based mean-field and perturbative description of strongly correlated fermion systems: Application to the one- and two-dimensional Hubbard model. *Phys. Rev. B* **2015**, *92*, 085101.
- (51) Hermes, M. R.; Pandharkar, R.; Gagliardi, L. Variational Localized Active Space Self-Consistent Field Method. *J. Chem. Theory Comput.* **2020**, *16*, 4923–4937.
- (52) Pandharkar, R.; Hermes, M. R.; Cramer, C. J.; Gagliardi, L. Localized Active Space-State Interaction: a Multireference Method for Chemical Insight. *J. Chem. Theory Comput.* **2022**, *18*, 6557–6566.
- (53) Ivanic, J. Direct configuration interaction and multiconfigurational self-consistent-field method for multiple active spaces with variable occupations. I. Method. *J. Chem. Phys.* **2003**, *119*, 9364–9376.
- (54) Abraham, V.; Mayhall, N. J. Cluster many-body expansion: A many-body expansion of the electron correlation energy about a cluster mean field reference. *J. Chem. Phys.* **2021**, *155*, 054101.
- (55) Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett.* **1980**, *73*, 393–398.
- (56) Chilkuri, V. G.; Applencourt, T.; Gasperich, K.; Loos, P.-F.; Scemama, A. New Electron Correlation Methods and their Applications, and Use of Atomic Orbitals with Exponential Asymptotes. In *Adv. Quantum Chem.*; Musial, M., Hoggan, P. E., Eds.; Academic Press, 2021; Vol. 83, pp 65–81.
- (57) Note that the threshold used in the earlier TPSCI work³¹ was pruning the probabilities instead of the coefficients like we do in this work.
- (58) Mayhall, N. J. Using Higher-Order Singular Value Decomposition To Define Weakly Coupled and Strongly Correlated Clusters: The n-Body Tucker Approximation. *J. Chem. Theory Comput.* **2017**, *13*, 4818–4828.
- (59) Tubman, N. M.; Levine, D. S.; Hait, D.; Head-Gordon, M.; Whaley, K. B. An efficient deterministic perturbation theory for selected configuration interaction methods. arXiv:1808.02049, 2018. <https://doi.org/10.48550/arXiv.1808.02049>.
- (60) Reger, D.; Haines, P.; Amsharov, K. Y.; Schmidt, J. A.; Ullrich, T.; Bönisch, S.; Hampel, F.; Görling, A.; Nelson, J.; Jelfs, K. E.; et al. A Family of Superhelicenes: Easily Tunable, Chiral Nanographenes by Merging Helicity with Planar π Systems. *Angew. Chem., Int. Ed.* **2021**, *60*, 18073–18081.
- (61) Fan, W.; Matsuno, T.; Han, Y.; Wang, X.; Zhou, Q.; Isobe, H.; Wu, J. Synthesis and Chiral Resolution of Twisted Carbon Nanobelts. *J. Am. Chem. Soc.*, *143*, 15924..
- (62) Fan, Q.; Yan, L.; Tripp, M. W.; Krejčí, O.; Dimosthenous, S.; Kachel, S. R.; Chen, M.; Foster, A. S.; Koert, U.; Liljeroth, P.; et al. Biphenylene network: A nonbenzenoid carbon allotrope. *Science* **2021**, *372*, 852–856.
- (63) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (64) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.
- (65) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.
- (66) Shu, Y.; Hohenstein, E. G.; Levine, B. G. Configuration interaction singles natural orbitals: an orbital basis for an efficient and size intensive multireference description of electronic excited states. *J. Chem. Phys.* **2015**, *142*, 024102.
- (67) Lehtola, S.; Jónsson, H. Pipek–Mezey Orbital Localization Using Various Partial Charge Estimates. *J. Chem. Theory Comput.* **2014**, *10*, 642–649.
- (68) Li, J.; Otten, M.; Holmes, A. A.; Sharma, S.; Umrigar, C. J. Fast Semistochastic Heat-Bath Configuration Interaction. *J. Chem. Phys.* **2018**, *149*, 214110.
- (69) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595–1604.
- (70) Holmes, A. A.; Tubman, N. M.; Umrigar, C. J. Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 3674–3680.
- (71) Yao, Y.; Umrigar, C. J. Orbital Optimization in Selected Configuration Interaction Methods. *J. Chem. Theory Comput.* **2021**, *17*, 4183–4194.
- (72) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z.-H.; et al. Recent developments in the PySCF program package. *J. Chem. Phys.* **2020**, *153*, 024109.
- (73) Bezanson, J.; Edelman, A.; Karpinski, S.; Shah, V. B. Julia: A fresh approach to numerical computing. *SIAM Rev.* **2017**, *59*, 65–98.
- (74) Mayhall, N. J.; Abraham, V.; Braunschweig, N. M. ClusterMeanField.jl, 2023. <https://github.com/nmayhall-vt/ClusterMeanField.jl>, (accessed July 21, 2023).
- (75) Mayhall, N. J.; Abraham, V.; Braunschweig, N. M. FermiCG.jl, 2023. <https://github.com/mayhallgroup/FermiCG>, (accessed July 21, 2023).
- (76) Angeli, C. On the nature of the $\pi \rightarrow \pi^*$ ionic excited states: The V state of ethene as a prototype. *J. Comput. Chem.* **2009**, *30*, 1319–1333.
- (77) Dash, M.; Feldt, J.; Moroni, S.; Scemama, A.; Filippi, C. Excited States with Selected Configuration Interaction-Quantum Monte Carlo: Chemically Accurate Excitation Energies and Geometries. *J. Chem. Theory Comput.* **2019**, *15*, 4896–4906.
- (78) Dash, M.; Moroni, S.; Filippi, C.; Scemama, A. Tailoring CIPSI Expansions for QMC Calculations of Electronic Excitations: The Case Study of Thiophene. *J. Chem. Theory Comput.* **2021**, *17*, 3426–3434.
- (79) Löwdin, P.-O. A Note on the Quantum-Mechanical Perturbation Theory. *J. Chem. Phys.* **1951**, *19*, 1396–1401.

- (80) Löwdin, P. O. Studies in Perturbation Theory. IV. Solution of Eigenvalue Problem by Projection Operator Formalism. *J. Math. Phys.* **1962**, *3*, 969–982.
- (81) Gershgorin, Z.; Shavitt, I. An application of perturbation theory ideas in configuration interaction calculations. *Int. J. Quantum Chem.* **1968**, *2*, 751–759.
- (82) Angeli, C.; Borini, S.; Cestari, M.; Cimiraglia, R. A quasidegenerate formulation of the second order n -electron valence state perturbation theory approach. *J. Chem. Phys.* **2004**, *121*, 4043–4049.
- (83) Shavitt, I.; Redmon, L. T. Quasidegenerate perturbation theories. A canonical van Vleck formalism and its relationship to other approaches. *J. Chem. Phys.* **1980**, *73*, 5711–5717.
- (84) Sharma, S.; Jeanmairet, G.; Alavi, A. Quasi-degenerate perturbation theory using matrix product states. *J. Chem. Phys.* **2016**, *144*, 034103.
- (85) Mayhall, N. J.; Head-Gordon, M. Increasing Spin-Flips and Decreasing Cost: Perturbative Corrections for External Singles to the Complete Active Space Spin Flip Model for Low-Lying Excited States and Strong Correlation. *J. Chem. Phys.* **2014**, *141*, 044112.
- (86) Mayhall, N. J.; Goldey, M.; Head-Gordon, M. A Quasidegenerate Second-Order Perturbation Theory Approximation to RAS- n SF for Excited States and Strong Correlations. *J. Chem. Theory Comput.* **2014**, *10*, 589–599.
- (87) We could also approximate the eigenstates by diagonalizing $\hat{H} + \alpha S^2$ for some small value of α to break the near degeneracies between states of different spin, however we have not done this here.
- (88) We have tested increasing the number of cluster states and found that our TPSCI results did not change significantly, ruling out that any truncation effects were at play.
- (89) Smith, M. B.; Michl, J. Recent Advances in Singlet Fission. *Annu. Rev. Phys. Chem.* **2013**, *64*, 361–386.
- (90) Johnson, J. C. Open questions on the photophysics of ultrafast singlet fission. *Commun. Chem.* **2021**, *4*, 85.
- (91) Morrison, A. F.; Herbert, J. M. Evidence for Singlet Fission Driven by Vibronic Coherence in Crystalline Tetracene. *J. Phys. Chem. Lett.* **2017**, *8*, 1442–1448.
- (92) Shu, Y.; Hohenstein, E. G.; Levine, B. G. Configuration interaction singles natural orbitals: An orbital basis for an efficient and size intensive multireference description of electronic excited states. *J. Chem. Phys.* **2015**, *142*, 024102.
- (93) Wilson, M. W.; Rao, A.; Johnson, K.; Gélinas, S.; di Pietro, R.; Clark, J.; Friend, R. H. Temperature-independent singlet exciton fission in tetracene. *J. Am. Chem. Soc.* **2013**, *135*, 16680–16688.
- (94) Bauman, N. P.; Bylaska, E. J.; Krishnamoorthy, S.; Low, G. H.; Wiebe, N.; Granade, C. E.; Roetteler, M.; Troyer, M.; Kowalski, K. Downfolding of many-body Hamiltonians using active-space models: Extension of the sub-system embedding sub-algebras approach to unitary coupled cluster formalisms. *J. Chem. Phys.* **2019**, *151*, 014107.
- (95) Kowalski, K.; Bauman, N. P. Sub-system quantum dynamics using coupled cluster downfolding techniques. *J. Chem. Phys.* **2020**, *152*, 244127.
- (96) Epstein, P. S. The Stark effect from the point of view of Schroedinger's quantum theory. *Phys. Rev.* **1926**, *28*, 695–710.
- (97) Nesbet, R. K. Configuration interaction in orbital theories. *Proc. R. Soc. A: Math. Phys. Sci.* **1955**, *230*, 312–321.