Nuclear Gradients of Near-Exact Complete Active Space Self-Consistent Field Wave Functions

James E. T. Smith\textsuperscript{1,2}, Joonho Lee\textsuperscript{3} and Sandeep Sharma\textsuperscript{2}

\textsuperscript{1}Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010 USA
\textsuperscript{2}Department of Chemistry, University of Colorado Boulder, Boulder, Colorado 80309, USA
\textsuperscript{3}Department of Chemistry, Columbia University, New York, New York 10027, USA

In this paper, we study the nuclear gradients of heat bath configuration interaction self-consistent field (HCISCF) wave functions and use them to optimize molecular geometries for various molecules. We show that the HCISCF nuclear gradients are fairly insensitive to the size of the "selected" variational space, which allows us to reduce the computational cost without introducing significant error. The ability of HCISCF to treat larger active spaces combined with the flexibility for users to control the computational cost makes the method very attractive for studying strongly correlated systems which require a larger active space than possible with complete active space self-consistent field (CASSCF). Finally, we study the realistic catalyst, Fe(PDI), and highlight some of the challenges this system poses for density functional theory (DFT). We demonstrate how HCISCF can clarify the energetic stability of geometries obtained from DFT when the results are strongly dependent on the functional. We also use the HCISCF gradients to optimize geometries for this species and show that the triplet potential energy surface is much more sensitive to the nuclear coordinates than the singlet surface.

I. INTRODUCTION

Systems that contain transition metals, covalent bond breaking, and electronically excited states are often challenging for theoretical methods because their electronic structure can be dominated by more than one electronic configuration. To study such systems we require multi-reference approaches like the complete active space self-consistent field (CASSCF) method. In CASSCF, we perform the full configuration interaction (FCI) procedure on a subset of the molecular orbitals termed the active space. The FCI expansion enumerates all possible configurations for a given set of orbitals and scales combinatorially with the size of the active space. This steep cost limits the size of active spaces that can be treated to roughly 22 electrons in 22 orbitals, which we abbreviate as (22e,22o). A substantial amount of research in quantum chemistry is devoted to reducing the cost of configuration interaction (CI) methods through a variety of approximations: restricted active space, generalized active space, density matrix renormalization, selected configuration interaction (SCI), FCI quantum monte carlo. The heat bath configuration interaction (HCI) is a particularly efficient implementation of the SCI method. These methods can correlate more than 40 electrons in 40 orbitals in routine calculations, making the study of larger and more complex systems accessible.

With the ability to treat larger active spaces, extending these algorithms to calculate molecular properties other than single-point energy is vital for connecting theoretical and experimental research. One such property of interest is the nuclear gradients of the electronic energy which are crucial for chemical applications because they are used to obtain minimum energy geometries, transition states, reaction pathways, and are used in ab initio molecular dynamics. Many of the approximate CI/CASSCF schemes have already been extended to calculate the nuclear gradient and in this work we extend the family of HCI methods.

The recently-developed HCI algorithm has proved to be an efficient approximation of FCI where the size of the CI expansion is controlled with a single user parameter $\epsilon_1$. This parameter allows one to continuously grow the wave function from a single determinant, e.g. Hartree-Fock (HF), to all configurations in the FCI expansion. The energy of this wave function can be corrected with a perturbative correction which is often necessary to reduce the error, relative to FCI, to an acceptable level. In recent work, two of the authors extended the HCI algorithm to be used in CASSCF-like calculations and implemented this method in the open source quantum chemistry package PySCF. This work showed that the multireference orbitals obtained from HCISCF were relatively insensitive to the quality of the HCI wave function and the perturbative correction was unnecessary to obtain accurate orbitals. In this work, we refer to orbitals as multireference when they are optimal, i.e. minimize the energy, for a set of determinants. In practice, this insensitivity is advantageous because it means that accurate multireference orbitals can be obtained from relatively cheap HCI calculations. In general, other properties may be also be insensitive to the accuracy of the HCI wave function, which motivates this work and the study of nuclear gradients.

---

\textsuperscript{a)Electronic mail: jsmith@flatironinstitute.org}
\textsuperscript{b)Electronic mail: limusoonho@gmail.com}
The rest of this paper is organized as follows. First, we briefly review HCI theory and its self-consistent variant along with analytical gradients of CI wave functions. Next, we discuss the gradients of CASSCF-like wave functions where HCI is used as the CI solver and verify that they converge smoothly with the HCI parameter $\epsilon_1$. Then, we highlight several strategies to improve the quality of the gradients when increasing the size of the CI expansion is not feasible, e.g., when memory is limited. Finally, we demonstrate the use of these gradients to find equilibrium geometries and adiabatic singlet-triplet gaps for the model catalytic system (PDI)Fe-N$_2$. We then briefly review HCI theory and its self-consistent variational wave function, then a second-order correction to the energy is computed using multireference Epstein-Nesbet perturbation theory (PT). Like other selected CI (SCI) + PT methods, it derives from FCI and linearly parametrizes the wave function $|\Psi\rangle = \sum_i c_i |D_i\rangle$. Unlike FCI, where the expansion is a set of all possible determinants, SCI methods express the wave function as a set of important determinants. This allows SCI+PT methods to correlate much larger systems than possible with FCI.

II. THEORY

A. Heat-bath Configuration Interaction Algorithm

HCI is a two-step procedure consisting of a variational calculation and a perturbative correction. In the first stage of the HCI algorithm, a set of important determinants is iteratively built and used to expand the variational wave function, then a second-order correction to the energy is computed using multireference Epstein-Nesbet perturbation theory (PT). Like other selected CI (SCI) + PT methods, it derives from FCI and linearly parametrizes the wave function $|\Psi\rangle = \sum_i c_i |D_i\rangle$. Unlike FCI, where the expansion is a set of all possible determinants, SCI methods express the wave function as a set of important determinants. This allows SCI+PT methods to correlate much larger systems than possible with FCI.

1. Variational Stage

In the first step, the HCI algorithm identifies a set of important determinants that approximate the full many-body wave function and we refer to this set as the variational space ($V$). At the beginning of the variational step, the user specifies a small number of determinants to add to $V$ and usually this includes only the Hartree-Fock determinant. Then the algorithm adds all connected determinants $|D_a\rangle$ which satisfy the following importance criteria

$$\max_{|D_i\rangle \in V} |H_{ai}c_i| > \epsilon_1$$

where $|D_i\rangle$ is a determinant already in the variational space, $H_{ai} = \langle D_a | H | D_i \rangle$, $c_i$ is the amplitude of the $|D_i\rangle$, and $\epsilon_1$ is a user-specified parameter that controls the size of the variational space, i.e., the accuracy of the variational wave function. A smaller $\epsilon_1$ means that the $V$ is closer to the FCI determinant space and in the limit that $\epsilon_1$ goes to 0, we recover our FCI wave function and energy. We repeat this selection of important determinants and grow the variational space until the HCI energy converges.

The CIPSI selection criteria, $\sum_{|D_i\rangle \in V} H_{ai}c_i > \epsilon_{\text{CIPSI}}$, is more complicated and uses first order perturbation theory to determine whether a determinant should be added to the variational space. Despite its similarity to the CIPSI selection criteria, the HCI importance function offers several key advantages over the original CIPSI algorithm. Since the matrix elements between determinants connected by double-excitations are just the two-electron integral elements (and a parity factor of +1 or -1), evaluating the HCI selection criteria requires no calculations. While avoiding calculations is advantageous, the real gain in performance comes from the fact that the two-electron integrals are stored and sorted once before the variational stage. Selecting important determinants only requires iterating through the sorted two-electrons integrals and the algorithm stops once those values fall below a certain threshold. As a result, this means that the algorithm never even "touches" the unimportant connected determinants, which often outnumber the important ones by many orders of magnitude. We note that HCI was inspired by challenges in the original formulation of CIPSI and more modern implementations have shown substantial improvements compared to their predecessors.

2. Perturbative Stage

We then correct the variational energy using a second order correction from Epstein-Nesbet perturbation theory, evaluated as

$$E_2(\epsilon_2) = \sum_{|D_a\rangle \in V} \left( \sum_{|D_i\rangle \in V} H_{ai}c_i \right)^2$$

where the inner summation is "screened" in a similar manner to the variational stage. Analogously, the perturbative expansion grows as $\epsilon_2$ becomes smaller and in the limit $\epsilon_2 \rightarrow 0$ the perturbative correction becomes exact.

While screening the inner sum in Eq. (2) mitigates the large memory requirement for calculating a perturbative correction it does not eliminate it entirely. Calculating an accurate correction requires a small $\epsilon_2$ value and for large systems, evaluating Eq. (2) deterministically becomes prohibitively expensive. In these cases, we can reduce the memory requirement further by calculating $E_2(\epsilon_2)$ semi-stochastically. By that we mean that we calculate as much of the correction deterministically as computational resources allow and then calculate the remainder stochastically. This procedure is advantageous compared to a purely stochastic one because the portion of $E_2(\epsilon_2)$ calculated stochastically is smaller and it takes fewer iterations to converge the stochastic error. We refer to calculations that use this procedure as semi-stochastic.
FIG. 1: The model Fe(PDI) complex. Ar = 2,6-dimethylphenyl. The numbering convention is the same as Stieber et al.\textsuperscript{71} for consistency.

HCI (SHCI) to distinguish them from HCl calculations that only use the variational step. There is another even more efficient scheme, which was recently introduced by one of the authors of this paper, but we do not use that scheme in this work.\textsuperscript{72}

B. Gradients of HCISCF Wave Functions

We briefly review analytical energy gradients of variational wave functions. For a detailed discussion on the subject, we refer the reader to the review by Yamaguchi and Schaefer.\textsuperscript{76} We start with the electronic Hamiltonian in second quantization:

\[
\hat{H} = \sum_{ij} h_{ij} a_i^{\dagger} a_j + \sum_{ijkl} g_{ijkl} a_i^{\dagger} a_j^{\dagger} a_k a_l \tag{3}
\]

We define the one- and two-body integrals in the usual way:

\[
h_{ij} = \int \phi_i^*(x) \left( -\frac{1}{2} \nabla^2 - \sum_l \frac{Z_l}{r_{il}} \right) \phi_j(x) dx
\]

\[
g_{ijkl} = \int \phi_i^*(x_1) \phi_j^*(x_2) \phi_k(x_1) \phi_l(x_2) dx_1 dx_2 \tag{4}
\]

where \(\phi\) are orbitals, \(Z_l\) are the nuclear charges, \(r_{il}\) the electron-nuclear separations, and \(r_{12}\) the electron electron separation. We use \(i, j, k,\) and \(l\) as orthogonal orbital indices and \(\sigma, \sigma'\) as spin indices. In most cases, these are molecular orbitals (MOs), which are linear combinations of atomic orbitals (AOs) \(|i\rangle = \sum_{\mu} C_{i\mu}^a |\mu\rangle\). For the remainder of this work, we use the following convention for orbital indices: Greek letters, e.g. \(\mu\) and \(\nu\) refer to AOs, while \(i, j\) refer to MOs.

The active space Hamiltonian is a function of the basis set, the MO coefficients \(C\), and the nuclear coordinates \(\{a_i\}\) while the electronic wave function \(|\Psi\rangle\) is a function of the parameters \(\{c_i\}\) making the electronic energy a function of all four. Expanding the nuclear gradients in terms of these variables we get:

\[
dE_{\text{elec}}/da = \frac{\partial E_{\text{elec}}}{\partial a} + \sum_{i\mu} \frac{\partial E_{\text{elec}}}{\partial C_{i\mu}} \frac{dC_{i\mu}}{da} + \sum_i \frac{\partial E_{\text{elec}}}{\partial c_i} \frac{dc_i}{da} \tag{5}
\]

The first term is the contribution from the basis set, i.e. the atomic orbitals, the second term is from the MO coefficients, and the final term is from the wave function parameters. For variational wave functions, \(\frac{\partial E_{\text{elec}}}{\partial C_{i\mu}} = 0\) and \(\frac{\partial E_{\text{elec}}}{\partial c_i} = 0\) so the last two terms vanish. In other words, the nuclear gradients should have no contribution from changes in MO or CI coefficients. To derive an explicit form for the gradients, we start by expressing the electronic energy in second-quantized form:

\[
E_{\text{elec}} = \sum_{ij} h_{ij} \gamma_{ij} + \sum_{ijkl} g_{ijkl} \Gamma_{ijkl} \tag{6}
\]

where \(\gamma_{ij} = \sum_{\sigma} \langle \Psi | a_i^{\dagger} a_j^{\dagger} | \Psi \rangle\) is the one-body reduced density matrix (RDM) and \(\Gamma_{ijkl} = \)
\[ \sum_{\sigma\sigma'} \langle \Psi | a_{ij}^{|\dagger} a_{j\sigma}^{|\dagger} a_{k\sigma'} a_{l\alpha} | \Psi \rangle \text{ is the two-body RDM.} \]

Taking derivative of Eq. (6) with respect to a general nuclear coordinate \( a \) generates terms including \( \frac{\partial \gamma_{ij}}{\partial a} \) and \( \frac{\partial \Gamma_{ijkl}}{\partial a} \). Since the RDMs are just functions of the CI coefficients, we know we can ignore their derivatives because the final term in Eq. (5) vanishes. As a result, we arrive at a concise expression for the nuclear gradients of the energy:

\[
\frac{dE_{\text{elec}}}{da} = \sum_{ij} \frac{dh_{ij}}{da} \gamma_{ij} + \sum_{ijkl} \frac{dg_{ijkl}}{da} \Gamma_{ijkl} \quad (7)
\]

Expanding this equation further, we can separate the terms relating to the MO response:

\[
\frac{dE_{\text{elec}}}{da} = \sum_{ij} h_{ij}^{a} \gamma_{ij} + \sum_{ijkl} g_{ijkl}^{a} \Gamma_{ijkl} + \sum_{ij} U_{ij}^{a} (X_{ij} - X_{ji}) - \sum_{ij} S_{ij}^{a} X_{ij} \quad (8)
\]

Where \( h_{ij}^{a} \), \( g_{ijkl}^{a} \), and \( S_{ij}^{a} \) are the "skeleton" derivatives of the one-electron, two-electron and overlap integrals defined as

\[
\begin{align*}
    h_{ij}^{a} &= \sum_{\mu\nu} C_{\mu}^{i} C_{\nu}^{j} \frac{dh_{\mu\nu}}{da} \\
    g_{ijkl}^{a} &= \sum_{\mu\nu\rho\sigma} C_{\mu}^{i} C_{\nu}^{j} C_{\rho}^{k} C_{\sigma}^{l} \frac{dg_{\mu\nu\rho\sigma}}{da} \\
    S_{ij}^{a} &= \sum_{\mu\nu} C_{\mu}^{i} C_{\nu}^{j} \frac{dS_{\mu\nu}}{da}
\end{align*}
\]

\( U_{ij}^{a} \) is called the orbital response and is defined by

\[
\frac{dC_{\mu}^{i}}{da} = \sum_{m} C_{\mu}^{i} C_{\mu}^{m} \quad (10)
\]

It can be determined by solving the Coupled Perturbed Hartree-Fock (CPHF) equations. Finally, \( X_{ij} \) is the Lagrangian from the Generalized Brillouin Theorem, defined as

\[
X_{ij} = \sum_{m} h_{im} \gamma_{jm} + 2 \sum_{mkl} g_{imkl} \Gamma_{jmkl} \quad (11)
\]

In general, calculating the gradients of a CI-type wave function like Eq. (7) requires solving the CPHF equations for each nuclear degree of freedom or the z-vector equations. For CASSCF, we can simplify Eq. (8) to Eq. (12) and avoid solving the CPHF or z-vector equations. We direct the reader to the derivation of Equation (133) in the excellent review by Yamaguchi and Schaefer. All three terms in Eq. (12) arise from response of the basis set to changes in nuclear coordinates and no terms from the response of MO or CI coefficients are present.

\[
\frac{dE_{\text{elec}}}{da} = \sum_{ij} h_{ij}^{a} \gamma_{ij} + \sum_{ijkl} g_{ijkl}^{a} \Gamma_{ijkl} - \sum_{ij} S_{ij}^{a} X_{ij} \quad (12)
\]

This derivation hinges on the fact that solving the CASSCF equations is equivalent to symmetrizing \( X_{ij} \), i.e. \( X_{ij} = X_{ji} \). For approximate CASSCF schemes like HCISCF and DMRGSCF, \( X_{ij} \) is not guaranteed to be symmetric and since we ignore the third term in Eq. (12), it is not exact. In practice, one can include rotation among the active space orbitals, which we refer to as active-active (AA) rotations during the orbital optimization to eliminate the error of using Eq. (12) with approximate schemes like DMRGSCF and HCISCF. Despite the formal lack of exactness, we highlight that if these approximate CI methods are sufficiently converged, the expression for the gradients is a justified approximation. If these methods are not converged to the FCI limit, Eq. (12) must include a response term due to the AA rotations to be formally exact.

### III. COMPUTATIONAL DETAILS

For all DFT geometry optimizations, we use Gaussian16 with the def2-TZVP basis set and density fitting to reduce the computational cost. We use the default convergence thresholds in the Gaussian16 package. In addition to finding optimal geometries, we used harmonic frequency analysis to confirm that the reported geometries are all genuine local minima, i.e. contain no imaginary frequencies. We perform all other quantum chemical calculations using PySCF and checked several of the mean-field calculations using Q-Chem. For HCISCF calculations we use our own implementation of the HCI algorithm, called Dice in tandem with PySCF. The nuclear gradients are calculated in PySCF using the RDMs obtained from Dice and geomeTRIC is used to perform the geometry updates. Due to the challenging nature of Fe(PDI) problem we relaxed the geometry convergence thresholds for all optimizations of the Fe(PDI) species, which are each double the default thresholds used by both Gaussian16 and geomeTRIC. Our convergence thresholds are: 2 \( \cdot \) 10\(^{-6} \) Ha for the energy (\( \tau_{e} \)), 6 \( \cdot \) 10\(^{-4} \) Ha/Bohr for the root mean squared (RMS) of the nuclear gradient (\( \tau_{g_{\text{rms}}} \)), 9 \( \cdot \) 10\(^{-4} \) Ha/Bohr for the maximum of the nuclear gradient (\( \tau_{g_{\text{max}}} \)), 2.4 \( \cdot \) 10\(^{-3} \) Å for the RMS of the atomic displacement (\( \tau_{d_{\text{rms}}} \)), and 3.6 \( \cdot \) 10\(^{-3} \) Å for the maximum atomic displacement (\( \tau_{d_{\text{max}}} \)).

For the Fe(PDI) species, we use a numerical procedure to select the active space to reduce the heuristics which typically plague active space calculations. All multireference calculations use the cc-pVDZ basis set. We outline the procedure for selecting the active space orbitals below:
1. Run UKS calculation using the PBE0 functional\textsuperscript{31,34} to generate an initial guess for Step 2.

2. Run UHF calculation to generate spin orbitals without correlation from an exchange-correlation functional.

3. Use the procedure outlined in Ref.\textsuperscript{85} to generate unrestricted natural orbitals (UNOs)\textsuperscript{86}

4. Use the UHF natural orbital occupation numbers (NOONs) along with the spin density information from the UHF calculation to ensure that the targeted state shows strong correlation and qualitatively correct spin density.

5. Use the UNOs in a "loose" HCI calculation where the $\epsilon_1 = 10^{-4}$ Ha and the active space is (100e,100o). We select the orbitals based on their NOON so the 50 orbitals below the HONO and 50 above the LUNO are part of the active space.

6. Use the 1-RDM from the previous step to calculate new NOs and sort based on NOONs.

7. Use these orbitals in all subsequent HCISCF calculations, selecting the orbitals solely based on their NOONs analogous to the way used in Step 5. We select the orbitals in pairs (i.e. one occupied and one virtual) around the HONO-LUNO gap. For all Fe(PDI) calculations, we use an active space of (40e,40o), which includes more than those recommended by the criteria UNO methods, i.e. $0.02 < \text{NOON} < 1.98$\textsuperscript{86}

We emphasize that Steps 1 and 2 are sensitive to the optimization strategy used for the SCF procedure, e.g. DIIS\textsuperscript{57,58} or Newton's method. We surveyed a range of choices for these strategies in both steps and select the best pair of strategies for each species. By "best" here, we mean that the SCF solution is the lowest in energy, stable with respect to its variational parameters, and has non-negligible spin-density on the Fe atom. All results from this survey are shown and described in Appendix C. While we have tried to search thoroughly for the lowest energy UHF solution, we note that it is possible that even lower energy solutions exist in this complicated energy landscape. From our experience these different SCF solutions could yield different multireference orbitals, i.e. distinct multiconfigurational self consistent field (MCSCF) wave functions, and as a result different optimized molecular structures.

For all geometry comparisons we use Kabsch's algorithm\textsuperscript{59} to align geometries and report the difference in geometry as $\frac{||\Delta||}{\sqrt{n_a}}$ where $\Delta$ is the difference in aligned nuclear coordinates and $n_a$ is the number of atoms. For brevity, we refer to this quantity as RMSD for the remainder of the paper. To align the geometries we use the open source package rmsd\textsuperscript{60}

The input scripts necessary to reproduce all of the results from this work are publicly available on GitHub at https://github.com/jamesETsmi/thi/uci_nuc_gradients.

IV. RESULTS

A. HCISCF Gradients

In this subsection, we discuss the errors that arise when using Eq. (12) for approximate CASSCF wave functions. In Fig. 2 we compare different variants of HCISCF gradients to CASSCF gradients, all using the cc-pVDZ basis set\textsuperscript{52} As mentioned earlier, the HCISCF gradients approximate CASSCF gradients because HCI is an approximation of FCI and as a result, the energy is not invariant for AA orbital rotation. In our discussions of gradients, we note that all gradients discussed suffer from basis set incompleteness errors, but since that is constant throughout the calculations we do not discuss it further.

There are two ways we can eliminate the error of this approximation: we can include AA rotations in the MCSCF optimization or we can converge the HCI wave function to the FCI limit. Including the AA rotations in the MCSCF optimization removes the need for the $z$-vector equation completely and ensures that Eq. (12) is exact. However it can significantly increase the number of MCSCF iterations required to reach convergence. AA rotations make the MCSCF procedure more challenging because the corresponding parameters are strongly coupled to the CI coefficients. Due to these challenges, optimizing AA rotations is often the most expensive way to improve the gradients. In the second approach, we can either increase the size of the variational space or include a perturbative correction and increase the size of the perturbative space to better converge an HCI wave function. By better converging to the FCI limit, we eliminate any effect of AA rotations on the energy, and as a result any need for the $z$-vector equation, again ensuring that Eq. (12) is exact. It is more cost-efficient to increase the number of configurations treated perturbatively and if we are memory limited this may be the only option. When performing CASSCF calculations using only the variational step of HCI, we refer to these as vHCISCF wave functions and when we use both the variational and the perturbative, we refer to them as HCISCF. In Fig. 2 we compare the different flavors of HCI-based gradients for the N$_2$ and Sc$_2$ dimers using active spaces of (10e,8o) and (6e,18o) respectively. We start with an intentionally inaccurate vHCISCF wave function, to leave considerable room for improvement, and show that adding AA rotations or PT correction can substantially decrease the gradient error.

Although AA rotations and the PT correction in our MCSCF optimization can reduce the error in our nuclear gradients, they do so at a cost that is often difficult to predict. Both options increase the number of MCSCF iterations required for convergence and with AA rotations the
gradient convergence tolerances used by the Gaussian16 package during geometry optimizations. For all wave functions we observe a steady convergence of the gradient to the exact (CASSCF) value. The rate of convergence differs for the two species, but given the larger Hilbert space and generally more complex nature of Sc2, this is not surprising. For the typical range of \( \epsilon_1 \) values used in research applications (\( \epsilon_1 = 10^{-4} - 10^{-5} \) Ha) the HCISCF gradients agree well with their CASSCF counterparts.

These encouraging results allow us to reduce the computational time required to calculate nuclear gradients of vHCISCF without significantly reducing their accuracy. For the remainder of the paper, all gradients are calculated with vHCISCF and no AA rotations, unless otherwise specified.

calculations may not converge at all. One other strategy is to use a smaller \( \epsilon_1 \) value, which increases the number of determinants in the HCI wave function. Using \( \epsilon_1 \), we can control the accuracy of our gradients, compared to CASSCF, and we find that we can use a relatively large value of \( \epsilon_1 \), i.e. a smaller number of determinants in our wave function, and still produce accurate gradients. For the same two systems as above, we examine the convergence of the analytic gradients with respect to \( \epsilon_1 \), i.e. the size of the variational space. Again we use the cc-pVDZ basis set and Fig. 3 shows the results for both systems. For context we show the loose, default, and tight

FIG. 2: Errors in gradient relative to CASSCF for various flavors of HCI-based wave functions for (a) \( \text{N}_2 \) \( \text{CAS}(10e,8o)/\text{cc-pVDZ} \) at a bond length of 1 Å and (b) \( \text{Sc}_2 \) \( \text{CAS}(6e,18o)/\text{cc-pVDZ} \) at 2.38 Å. Both species use values of \( \epsilon_1 = 5 \cdot 10^{-3} \) and \( \epsilon_2 = 10^{-10} \) Ha. The error in vHCISCF here is high due to the deliberate use of a very approximate \( \epsilon_1 \) parameter. This is done to highlight the ability of AA rotations and the perturbative correction to effectively reduce the overall error in the gradients.

FIG. 3: The error in the gradient as a function of \( \epsilon_1 \) for several HCI-based wave functions. As the wave function converges to the CASSCF result the improvements from AA rotations and PT contribution make a negligible difference in the gradient.
B. Geometry Optimization of Fe(PDI) using HCISCF.

We use two geometries reported by Ortuño and Cramer which they obtained using DFT: the first, \(^{1}\text{BS}(1,1)\), which we call A, optimized for the singlet spin symmetry and the second, \(^{3}\text{BS}(3,1)\), which we call C, optimized for the triplet. They used the A geometry for their multireference calculations. Both structures were calculated using unrestricted M06-L with the def2-TZVP basis and density fitting. During initial studies on this species, when starting a triplet geometry optimization from A, we found another stable triplet geometry using the same functional and basis set, which we refer to as B. In this work, we use the shorthand of (M,X) to describe a state with multiplicity M at geometry X, where \(X \in \{A, B, C\}\).

Given the challenging nature of Fe(PDI), we anticipated that our DFT results would be sensitive to choices of functional and starting geometry and attempted to quantify this sensitivity by studying the system with several DFT functionals. We chose several of the Minnesota functionals (M06-2X, M06-L, and MN15) because of their success with challenging transition metal complexes. Since we found a new stable triplet geometry (B) when starting triplet geometry optimization from A using the M06-L functional, we started triplet geometry optimizations from both A and C and found again that the final optimized structure was sensitive to the starting geometry. To further compare the effect of the functional choice, we illustrate the effect on energy and the spin density on the Fe atom in Fig. 4. We found that both the choice of functional and starting geometry can lead to qualitatively different results. The left panel demonstrates that the identity of the ground state depends on the amount of exact exchange included in the functional and the right panel highlights that spin density on the iron atom ranges from two to four unpaired electrons.

NOONs are often used to assess the amount of static correlation for a system and values close to 1.5 and 0.5 indicate a substantial amount of this type of correlation. As described earlier, it is common practice to include NOs where \(0.02 < \text{NOON} < 1.98\) in the active space. In Table I, we see that the NOONs which would be selected by this criteria differ by functional, adding to the uncertainty when performing active space calculations after DFT. Finally, the optimized geometries with DFT are also different and we compare the geometries in the Appendix, see Table II.

1. Impact of the Initial Geometries

Since we hope to calculate accurate singlet-triplet (S-T) gaps which we can qualitatively compare to experiments, it is challenging to "decide" which geometries and functionals we should use for multireference calculations. To address this, we can use SHCI energies to determine which DFT geometry is the most stable in an unbiased manner and use these geometries in the subsequent study of the S-T gaps. For each geometry (A, B, and C), we used the procedure outlined above to select the active space of size (40e,40o) and then ran single point vHCISCF calculations followed by a single tight SHCI calculation. We deliberately correlate a larger number of orbitals than would be selected by the NOON number of orbitals than would be selected by the NOON numerical thresholds mentioned earlier in order to confidently include all strongly correlated orbitals in the active space. For the vHCISCF calculations we used a value of \(\epsilon_1 = 7.5 \cdot 10^{-5}\) Ha. Given the sensitivity of the MCSCF solutions to the input orbitals, we tested the several values of \(\epsilon_1\) and found that the MCSCF orbitals converge faster than energy with respect to this parameter agreeing with previous work by several of the authors.

We report this analysis in the Appendix D. For the final SHCI calculation we used several values of \(\epsilon_1\) to extrapolate to the FCI limit, the tightest one being \(10^{-5}\) Ha, and \(\epsilon_2 = 5 \cdot 10^{-8}\) Ha. We detail this extrapolation procedure in Appendix A. Extrapolations for singlet and triplet states at all three geometries are shown in Fig. 5.

Fig. 5 indicates that A is the best structure to start geometry optimization from for the singlet spin symmetry and C for the triplet. While this process is more clear than comparing geometries as a function of DFT functionals, it was not as unambiguous as we had expected. The fact that the extrapolations of (3,B) and (3,C) cross demonstrates the nontrivial nature of our conclusion that C is better triplet starting geometry than B. We highlight that geometries A and B are similar (RMSD is 2.8 \(\cdot 10^{-2}\) Å) and the singlet energies for the two geometries agree to within 1 mHa. However, the triplet energies at these two geometries differ by roughly 30 mHa. For completeness, we optimized the Fe(PDI) structure starting from both multiplicities at all three initial geometries for a total of six vHCISCF geometry optimizations. In all vHCISCF gradients used in geometry optimizations,

| Functional | Mult. | Geom. | HONO-1 | HONO | LUNO | LUNO+1 |
|------------|-------|-------|--------|-------|-------|--------|
| M06-2X     | 1     | A     | 1.325  | 1.000 | 1.000 | 0.675  |
| M06-2X     | 3     | A     | 1.992  | 1.000 | 1.000 | 0.008  |
| M06-2X     | 3     | C     | 1.188  | 1.000 | 1.000 | 0.812  |
| M06-L      | 1     | A     | 1.921  | 1.000 | 1.000 | 0.079  |
| M06-L      | 3     | A     | 1.980  | 1.000 | 1.000 | 0.019  |
| M06-L      | 3     | C     | 1.566  | 1.000 | 1.000 | 0.434  |
| MN15       | 1     | A     | 1.617  | 1.000 | 1.000 | 0.383  |
| MN15       | 3     | A     | 1.983  | 1.000 | 1.000 | 0.017  |
| MN15       | 3     | C     | 1.686  | 1.000 | 1.000 | 0.314  |

TABLE I: The natural orbital occupation numbers (NOONs) for the systems studied with DFT as a function of starting geometry and functional. The geometries listed here are the starting geometries from Ortuño and Cramer. Here HONO is the highest occupied NO and LUNO is the lowest unoccupied NO.
we use Eq. (12) without including any AA rotations or response terms mentioned in Section II making the gradients approximate. This approach is consistent with several previous works where gradients of approximate CASSCF methods were calculated. We believe that it is important to quantify the effect of this choice and intend to study it in a future work.

2. Comparing vHCISCF Optimized Geometries

Before comparing the effect of initial geometry, we mention that the RMSD between initial and final geometries for all vHCISCF optimization are between 0.1-0.4 Å and the detailed values are shown in Table II. When comparing optimized geometries starting from different
FIG. 6: The extrapolated SHCI energies at the final (i.e., "relaxed") geometries. Similar to the extrapolated energies at the initial geometry, the singlet energies of A and B are very similar, while their triplet energies differ by more than 30 mHa.

| Geometries | Singlet RMSD | Triplet RMSD |
|------------|--------------|--------------|
| A          | 1.7 · 10⁻¹   | 1.3 · 10⁻¹   |
| B          | 1.9 · 10⁻¹   | 1.5 · 10⁻¹   |
| C          | 3.9 · 10⁻¹   | 3.4 · 10⁻¹   |

TABLE II: Comparing the change in geometry after optimization for the singlet and triplet species. All RMSD values are in Å and are calculated using the rmsd package.

initial coordinates, the structural difference shrink for all singlet states after vHCISCF geometry optimization. In particular, states starting from A and B get closer after geometry optimization and nearly agree within the loose tolerance used by Gaussian16 (6.7 · 10⁻³ Å). For triplet states, the geometries all diverge from each other, albeit only slightly. We show a numerical comparison of the geometries in Table III.

Next we examine the differences between the experimental and theoretical structures of Fe(PDI). For consistency, we use the same numbering scheme as Stieber et al. Fig. 2 shows error in the initial and final geometries of state (1,A). To reiterate, all initial geometry were optimized with symmetry-broken DFT, while the final geometries was obtained with vHCISCF. We compare select bond lengths from both geometries to experimental ones reported in Ref. 71 which were obtained from crystallographic experiments. We only report a subset of the bond lengths because (1,A) is only a model system for the experimental complex and they differ slightly in the composition. Despite using multireference methods, we found that the optimized geometries did not outperform the M06-L, MN15, and M06-2X. However, the DFT geometry optimizations use def-TZVP while the vHCISCF calculations use cc-pVDZ. The def-TZVP basis is considerably larger than cc-pVDZ and should a priori provide a better description of geometry. Due to computational limitations, we were not able to run vHCISCF calculations using def-TZVP. We note that the agreement between the DFT-optimized and experimental geometries may be fortuitous since the Fe(PDI) complex contains more alkyl groups and the experimental evidence suggests that the electronic structure is sensitive to these changes. However, a further investigation is necessary to test this claim. The other geometry optimizations starting from other geometries and/or spin symmetry showed similar behavior and we report them in Appendix.
3. Singlet-Triplet Gaps

Ref. [71] argued that the ground state is a singlet with a thermally accessible triplet state based on the strong temperature dependence of NMR data for species similar to Fe(PDI). However, the set of compounds they report in Table 2 of their paper differ by the alkyl (methyl, ethyl, or isopropyl) groups attached to C(2) and C(8). Given the noticeable influence of the functional groups on this temperature dependence, it is not clear if those same arguments apply to our model species. Ref. [96] also suggested that the S-T gap of Fe(PDI) small and reported vertical excitation values of 2.2 kcal/mol using RASSCF and 3.4 kcal/mol using RASPT2. Their RAS calculations correlated 22 electrons in 22 orbitals and use a mixed basis where a triple-zeta basis is used for Fe, a double-zeta basis is used for C and N, and a minimal basis is used for H.

Our calculations agree with previous experimental and theoretical work that the ground state of the Fe(PDI) complex is a singlet. However, we found that the S-T was larger than previously predicted and was not thermally accessible. As alluded to earlier, there are several factors which could explain this difference such as the missing functional groups in our model system as well as different correlated methods and basis sets. When comparing to the experimentally obtained structure, vHCISCF performs similarly, albeit slightly worse, than the three DFT functionals studied. Fe(PDI) is a challenging system even for accurate multireference methods and we encourage others in the community to use it as a benchmark system going forward.

VI. DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://github.com/jamesETsmith/hci_nuc_gradients.

VII. ACKNOWLEDGEMENTS

This work utilized resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. S.S. acknowledges funding from the NSF award CHE1800584 and the Sloan research fellowship. J.E.T.S. gratefully acknowledges support from a fellowship through The Molecular Sciences Software Institute under NSF Grant ACI-1547580. J.E.T.S. was also supported by the Flatiron Institute.
FIG. 7: The errors in initial and final geometry compared to experimentally obtained bond lengths for state (1, A). See Fig. for a numerically labelled diagram Fe(PDI). We report the analogous figures for other spin states and starting geometries in Appendix E.

Appendix A: Extrapolation Procedure

All calculations used for extrapolation use $\epsilon_2 = 5 \cdot 10^{-8}$ Ha and ten different values of $\epsilon_1$ ranging from $5 \cdot 10^{-5}$ to $10^{-5}$ Ha.

1. Run all SHCI calculations
2. Collect the PT correction ($E_2$) (our independent variable) and total energy ($E_{SHCI}$)
3. Fit linear and quadratic curves to $E_{SHCI}$ as a function of $E_2$
4. We estimate the error in this extrapolated energy as the standard deviation of the curve fitting parameter representing the y-intercept.

Appendix B: DFT Geometry Optimization

In addition to the energy and spin density on the Fe center displayed, we show the numerical values for several properties of the Fe(PDI) complex as a function of starting geometry and functional. In addition to the energy and Fe spin density shown in Fig. we show the spin contamination and the RMSD compared to the singlet and triplet geometries from Ortuño and Cramer in Table B.1.

Appendix C: Stable UHF Solutions

Since the UDFT and UHF calculations used to prepare orbitals for the multireference calculations were so sensitive to their optimization settings, we surveyed a large number of these settings to find the lowest energy state possible. We emphasize that such care is needed for this species because MCSCF optimizations are sensitive to the initial orbitals for this species. We examined six different optimization options for the DFT stage and three for the UHF stage leading to a total of 108 ways to prepare the orbitals. In Fig. we show a subset of these calculations where the final UHF produced a stable wave function with respect to the orbital rotation parameters.

Appendix D: MCSCF Sensitivity to $\epsilon_1$

We investigated the sensitivity of the vHCISCF orbitals to the HCI parameter, $\epsilon_1$, analogous to the analysis
### TABLE B.1: Comparing the DFT results as a function of starting geometry and functional.

| Functional | Multiplicity | Starting Geometry | Energy (Ha) | Fe Spin Density | \( \langle S^2 \rangle \) | RMSD (Å) |
|------------|--------------|-------------------|-------------|----------------|-----------------|----------|
| M06-2X     | 1            | A                 | -2506.345333 | 1.949442       | 1.9345          | 1.1 \cdot 10^{-1} |
| M06-2X     | 3            | A                 | -2506.323491 | 1.132163       | 2.0419          | 7.8 \cdot 10^{-1}  |
| M06-2X     | 3            | C                 | -2506.364616 | 3.721854       | 3.9329          | 3.3 \cdot 10^{-1}  |
| M06-L      | 1            | A                 | -2506.768403 | 1.454747       | 1.2123          | 6.6 \cdot 10^{-4}  |
| M06-L      | 3            | A                 | -2506.768703 | 1.457988       | 2.0696          | 7.8 \cdot 10^{-1}  |
| M06-L      | 3            | C                 | -2506.753225 | 3.120475       | 2.9871          | 4.4 \cdot 10^{-3}  |
| MN15       | 1            | A                 | -2505.529508 | 1.767367       | 1.6642          | 3.3 \cdot 10^{-2}  |
| MN15       | 3            | A                 | -2505.524226 | 1.254893       | 2.0519          | 7.4 \cdot 10^{-1}  |
| MN15       | 3            | C                 | -2505.526531 | 2.040751       | 2.5562          | 7.8 \cdot 10^{-1}  |

The RMSD values are calculated relative to broken symmetry geometries Ortuño and Cramer reported. All singlets are compared to their "BS(1,1)" geometry and triplets were compared to "BS(1,3)."

**FIG. C.1:** A summary of the various stable UHF solutions as a function of geometry, multiplicity, as well as SCF optimization options. The "Fast Newton" strategies are heuristic options implemented in PySCF where a second order optimization is performed in a minimal basis and then projected to the desired basis to accelerate the calculation. No stable solutions were found using DIIS in the UHF optimization.

from Smith et al.\[56\] This process involves two steps: first, converging the vHCISCF orbitals with respect to several different values of \( \epsilon_1 \) and second, running an SHCI calculation using these orbitals with identical settings. Despite the fact that Fe(PDI) is more multireference than the Fe(porphyrin) system studies in Ref. \[56\] we find that the orbitals are relatively insensitive to \( \epsilon_1 \) and show the results in Fig. D.1.

### Appendix E: Geometry Comparisons

We show the error between several theoretical results and the experimentally determined bond lengths for a subset of bond lengths near the central Fe atom. For all states, we compare vHCISCF to M06-L and for (1,A), (3,A), and (3,C) we compare to MN15 and M06-2X as well, since those calculations were already performed during our earlier analysis. In general, vHCISCF performs similarly to the ensemble of DFT results, but does not
FIG. D.1: We report the sensitivity of the vHCISCF orbitals with respect to $\epsilon_1$. In the two step procedure we first optimize the vHCISCF orbitals and then run a final "tight" SHCI calculations using those orbitals to obtain a more accurate energy. The $\epsilon_1$ on the x-axis is for the vHCISCF calculations and indicates the accuracy/quality of the wave function with more accurate wave functions on the left. Our results show that the SHCI energy is relatively insensitive to the $\epsilon_1$ used during the vHCISCF optimization.

FIG. E.1: Comparing errors in initial and final geometry compared to experimentally obtained bond lengths for state $(1, B)$.

FIG. E.2: Comparing errors in initial and final geometry compared to experimentally obtained bond lengths for state $(1, C)$.

FIG. E.3: Comparing errors in initial and final geometry compared to experimentally obtained bond lengths for state $(3, A)$.

FIG. E.4: Comparing errors in initial and final geometry compared to experimentally obtained bond lengths for state $(3, B)$.

typically provide the most accurate bond lengths.

1B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, “A Complete Active Space SCF Method (CASSCF) Using a Density Matrix Formulated Super-CI Approach,” Chem. Phys. 48, 1680 (1980).

2J. Olsen, B. O. Roos, P. Jørgensen, and H. J. A. Jensen, “Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces,” J. Chem. Phys. 89, 2185–2192 (1988).

3P. A. Malmqvist, A. Rendell, and B. O. Roos, “The restricted active space self-consistent-field method, implemented with a split graph unitary group approach,” J. Phys. Chem. 94, 5477–5482 (1990).
FIG. E.5: Comparing errors in initial and final geometry compared to experimentally obtained bond lengths for state (3,C).
J. E. T. Smith, B. Mussard, A. A. Holmes, and S. Sharma, (2017), arXiv:1610.06660v2.

Ration Interaction Algorithm Inspired by Heat-Bath Sampling,”

A. A. Holmes, N. M. Tubman, and C. J. Umrigar, “Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction in Slater determinant space,” J. Chem. Phys. 145, 044112 (2016), arXiv:1605.02686.

Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, “Hybrid stochastic-deterministic calculation of the second-order perturbative contribution of multireference perturbation theory,” J. Chem. Phys. 147, 34101 (2017).

N. M. Tubman, D. S. Levine, D. Halt, M. Head-Gordon, and K. B. Whaley, “Modern Approaches to Exact Diagonalization and Selected Configuration Interaction with the Adaptive Sampling CI Method,” J. Chem. Theory Comput. 16, 2139–2159 (2020), arXiv:1807.00821.

D. L. Cleland, G. H. Booth, and A. Alavi, “Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum Monte Carlo method,” J. Chem. Phys. 132 (2010), 10.1063/1.3392277.

P. R. Petruzielo, A. A. Holmes, H. J. Chuangli, M. P. Nightingale, and C. J. Umrigar, “Semistochastic projector monte carlo method,” Phys. Rev. Lett. 109, 1–5 (2012), arXiv:arXiv:1207.6138v2.

R. E. Thomas, Q. Sun, A. Alavi, and G. H. Booth, “Stochastic Multireference Self-Consistent Field Theory,” J. Chem. Theory Comput. 11, 5316–5325 (2015), arXiv:1510.03635.

A. A. Holmes, N. M. Tubman, and C. J. Umrigar, “Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling,” J. Chem. Theory Comput. 12, 3674–3680 (2016).

S. Sharma, A. A. Holmes, G. Jeannedart, A. Alavi, and C. J. Umrigar, “Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory,” J. Chem. Theory Comput. 13, 1595–1604 (2017), arXiv:1610.06600v2.

T. E. F. Smith, B. Mussard, A. A. Holmes, and S. Sharma, “Cheap and Near Exact CASSCF with Large Active Spaces,” J. Chem. Theory Comput. 13, 5406–5478 (2017), arXiv:1708.07544.

B. F. E. Curchod and T. J. Martinez, “Ab Initio Nonadiabatic Quantum Molecular Dynamics,” Chem. Rev. 118, 3385–3396 (2018).

P. O. Giresh and C. Filippi, “Ground- and Excited-State Geometry Optimization of Small Organic Molecules with Quantum Monte Carlo,” J. Chem. Theory Comput. 9, 5513–5525 (2013).

M. Dash, J. Feldt, S. Moroni, A. Scemama, and C. Filippi, “Excited States with Selected Configuration Interaction-Quantum Monte Carlo: Chemically Accurate Excitation Energies and Geometries,” J. Chem. Theory Comput. 15, 4896–4906 (2019).

P. Liu, Y. Kurashige, T. Yanai, and K. Morokuma, “Multireference Ab Initio Density Matrix Renormalization Group (DMRG) -CASSCF and DMRG-CASPT2 Study on the Photochromic Ring Opening of Spiropyran,” J. Chem. Theory Comput. 9, 4462–4469 (2013).

W. Hu and G. K. L. Chan, “Excited-State Geometry Optimization with the Density Matrix Renormalization Group, as Applied to Polynes,” J. Chem. Theory Comput. 11, 3000–3009 (2015), arXiv:1502.07731.

E. Maradzhie, G. Gidofalvi, J. M. Turney, H. F. Schafer, and A. E. DePrince, “Analytic Energy Gradients for Variational Two-Electron Reduced-Density-Matrix-Driven Complete Active Space Self-Consistent Field Theory,” J. Chem. Theory Comput. 13, 4113–4122 (2017).

A. W. Schlimmen and D. A. Mazzotti, “Analytical gradients of variational reduced-density-matrix and wavefunction-based methods from an overlap-weighted semidefinite program,” J. Chem. Theory Comput. 14 (2018), 10.1063/1.5043104.

M. Dash, S. Moroni, A. Scemama, and C. Filippi, “Perturbative Self-Selected Configuration-Interaction Wave Functions for Efficient Geometry Optimization in Quantum Monte Carlo,” J. Chem. Theory Comput. 14, 4176–4182 (2018), arXiv:1804.09610.

J. W. Park, “Second-Order Orbital Optimization with Large Active Space Using Adaptive Sampling Configuration Interaction (ASCI) and Its Application to Molecular Geometry Optimization Jae Woo Park,” J. Chem. Theory Comput. 17, 1522–1534 (2021).

J. W. Park, “Near-Exact CASSCF-Level Geometry Optimization with a Large Active Space using Adaptive Sampling Configuration Interaction Self-Consistent Field Corrected with Second-Order Perturbation Theory (ASCI-SCF-PT2),” J. Chem. Theory Comput. 17, 4092–4104 (2021).

Q. Sun, “Libcint: An efficient general integral library for Gaussian basis functions,” J. Comput. Chem. 36, 1664–1671 (2015), arXiv:1112.0643.

Q. Sun, J. Yang, and G. K.-L. Chan, “A general second order complete active space self-consistent-field solver for large-scale systems,” Chem. Phys. Lett. 683, 291–297 (2017), arXiv:1610.08394.

Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. L. Chan, “ pyscfc: the Python-based simulations of chemistry framework,” Wiley Interdiscip. Rev. Comput. Mol. Sci. 8 (2018), 10.1002/wcms.1340, arXiv:1701.08223.

Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. E. Kirsch, Y. Gao, S. Guo, J. Hermann, M. R. Hersnes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitleff, M. J. Williamson, S. Wouters, Yang, J. M. Yu, Zhu, T. C. Berkelbach, S. Sharma, A. Sokolov, and G. K.-L. Chan, “Recent developments in the PySFC program package,” J. Chem. Phys. 153, 024109 (2020), arXiv:2002.12531.

S. C. E. Stiebel, C. Milsmann, J. M. Hoyt, Z. R. Turner, K. D. Finkelstein, K. Wieghardt, S. Debeer, and P. J. Chirik, “Bis(imino)pyridine iron dinitrogen compounds revisited: Differences in electronic structure between four- and five-coordinate iron(III) complexes,” Inorg. Chem. 51, 3770–3785 (2012).

P. S. Epstein, “The Stark effect from the point of view of Schroedinger’s quantum theory,” Phys. Rev. 28, 695–710 (1926).

R. K. Nesbet, “Configuration interaction in orbital theories,” Proc. R. Soc. Ser. A 230, 312–321 (1955).
M. Caffarel, P. F. Loos, E. Giner, and A. Scemama, “Quantum Package 2.0: An Open-Source Determinant-Derived Suite of Programs,” J. Chem. Theory Comput. 15, 3591–3609 (2019).

J. Li, M. Otten, A. A. Holmes, S. Sharma, and C. J. Umrigar, “Fast Semistochastic Heat–Bath Configuration Interaction,” J. Chem. Phys. 149, 1–12 (2018) [arXiv:1809.04600].

Y. Yamaguchi and H. F. Schaefer, “Analytic Derivative Methods in Molecular Electronic Structure Theory: A New Dimension to Quantum Chemistry and its Applications to Spectroscopy,” in H. Handb. High-resolution Spectrosc. (John Wiley & Sons, Ltd., 2011) pp. 325–362.

J. Gerratt and I. M. Mills, “Force constants and dipole-moment derivatives of molecules from perturbed hartree-fock calculations,” J. Chem. Phys. 49, 1719–1729 (1968).

N. C. Handy and H. F. Schaefer, “On the evaluation of analytic energy derivatives for correlated wave functions,” J. Chem. Phys. 81, 5031–5033 (1984).

M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, W. N. Del Bene, M. A. Li, X. L. Li, M. Caricato, M. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. J. Montgomery, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Healy, D. E. Kaege, N. L. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, “Gaussian 16 Rev. C. 01 Release Notes,” (2016).

E. Epifanovsky, A. T. Gilbert, X. Feng, J. Lee, Y. Mao, N. Marderiossian, F. Pokhillo, A. F. White, M. P. Coons, A. L. Dempwolf, Z. Gan, D. Hart, P. R. Horn, L. D. Jacobson, I. Kallman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z. Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Algurier, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cordero-Sabicha, S. Daugast, M. De Werginçois, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, F. T. Fang, S. Faraji, Q. Feng, T. Friedhoff, J. Guyvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. Hansen-Heine, P. H. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Honck, X. Huang, K. Hu, B. C. Huynh, M. Ivanov, A. Jássz, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khitsyn, J. Kim, G. Kis, P. Klumzinger, Z. Koszor-Benda, J. K. Koh, D. Kosenkov, L. Koulish, T. Kowalczuk, C. M. Krauter, K. Kue, A. Kraus, T. Kubo, T. Kubi, I. Ladjászai, A. Landau, R. V. Lawler, D. Lefrançois, S. Lehtola, R. R. Li, Y. P. Li, J. Lieb, M. Liebenthal, H. H. Lin, Y. S. Lin, F. Liu, K. Y. Liu, M. Löröbersger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S. P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. Menger, J. M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, E. Epifanovsky, G. Gidofalvi, and A. Eugene Deprince, “Analytic Energy Gradients for Variational Two-Electron Reduced-Density Matrix Methods within the Density Fitting Approximation,” J. Chem. Theory Comput. 15, 276–289 (2019).

H. S. Yu, X. He, S. L. Li, and D. G. Truhlar, “MN15: A Kohn-Sham global-hybrid exchange-correlation density functional with broad accuracy for multi-reference and single-reference systems and noncooperative interactions,” J. Chem. Phys. 125 (2006), 10.1063/1.2270903.

J. W. Mullinax and D. G. Truhlar, “The M06 suite of density functionals for main group thermochecmistry, thermochecmical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function,” Theor. Chem. Acc. 120, 215–241 (2008).

Y. C. Su, A. T. Moore, J. J. S. de Miranda, J. T. C. Vilar, J. P. Perdew, M. C. B. Neto, K. N. E. de Souza, and D. G. Truhlar, “MN15: A Kohn-Sham global-hybrid exchange-correlation density functional with broad accuracy for multi-reference and single-reference systems and noncooperative interactions,” J. Chem. Phys. 125 (2006), 10.1063/1.2270903.

J. W. Mullinax, E. Epifanovsky, G. Gidofalvi, and A. Eugene Deprince, “Analytic Energy Gradients for Variational Two-Electron Reduced-Density Matrix Methods within the Density Fitting Approximation,” J. Chem. Theory Comput. 15, 276–289 (2019).

P. M. Zimmerman and A. E. Rask, “Evaluation of full valence correlation energies and gradients,” J. Chem. Phys. 150 (2019), 10.1063/1.5100534.

M. A. Ortuno and C. J. Cramer, “Multireference Electronic Structures of Fe-Pyridine(diimine) Complexes over Multiple Oxidation States,” J. Chem. Phys. A 121, 5932–5939 (2017).