Chemical insights into the electronic structure of Fe(II) porphyrin using FCIQMC, DMRG, and generalized active spaces

Oskar Weser1,∗, Leon Freitag2,3, Kai Guther1, Ali Alavi1,4, and Giovanni Li Manni1,∗

1 Max-Planck-Institut für Festkörperforschung, Heisenbergstraße 1, 70569 Stuttgart, Germany
2 ETH Zürich, Laboratorium für Physikalische Chemie, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland
3 Present address: Institute for Theoretical Chemistry, University of Vienna, Währinger Str. 17, 1090 Vienna, Austria
4 University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, United Kingdom
*Corresponding authors: Giovanni Li Manni, giovannilimanni@gmail.com; Oskar Weser, oskar.weser@gmail.com

Abstract

Stochastic-CASSCF and DMRG procedures have been utilized to quantify the role of the electron correlation mechanisms that in a Fe-porphyrin model system are responsible for the differential stabilization of the $^3E_g$ over the $^5A_{1g}$ state. Orbital entanglement diagrams and CI coefficients of the wave function in a localized orbital basis allow for an effective interpretation of the role of charge-transfer configurations.

A preliminary version of the Stochastic Generalized Active Space Self-Consistent Field method has been developed and is here introduced to further assess the $\pi$ backdonation stabilizing effect. By the new method excitations between metal and ligand orbitals can selectively be removed from the complete CI expansion. It is demonstrated that these excitations are key to the differential stabilization of the triplet, effectively leading to a quantitative measure of the correlation enhanced $\pi$ backdonation.

Keywords — Fe porphyrin, FCIQMC, Stochastic-CASSCF, Stochastic-GASSCF, DMRG, $\sigma$ donation/$\pi$ backdonation.
1 Introduction

Fe porphyrins are central building blocks for various enzymes in biochemistry, playing there an important role in redox reactions, and charge and molecular transport\[1, 2, 3, 4\]. Their enzymatic activity is bound to the presence of nearly degenerate electronic states. For example, Fe(II) porphyrin has a $d^6$ electronic configuration with high-spin (quintet), intermediate-spin (triplet), and low-spin (singlet) states close in energy. The relative stability of these states depends on ligand field and many-body correlation effects experienced by the metal center, that, in turn, depend on chemical functionalization and geometry of the conjugated macrocycle.

In a recent work\[5\], the Stochastic-CASSCF method\[6\] has been applied to a simplified Fe(II) porphyrin model complex. Valence 3$d$, nitrogen $\sigma$, and the entire conjugated $\pi$ system electrons have been explicitly correlated in the Stochastic-CASSCF(32, 34) procedure, while variationally optimizing the molecular orbitals. A correlation enhanced $\sigma$ donation/$\pi$ backdonation mechanism has been proposed to explain the relative stability of the intermediate-spin state (triplet) over the high-spin (quintet). According to this mechanism, the electron repulsion at the metal center is differentially reduced for the triplet spin state by electron delocalization induced by charge-transfer (CT) excitations between the 3$d$ orbitals at the metal center and the $\pi$ system of the macrocycle. This mechanism is not captured in smaller active space calculations and it is not accounted for by second order perturbation theory procedures (such as CASPT2) on top of small active space reference wave functions. The correlation enhanced charge transfer mechanism suggests that large active spaces, such as the CAS(32, 34) used in reference 5, are crucial for an accurate description of electron delocalization in metal porphyrins.

In two subsequent works\[7, 8\] the role of semi-core correlation (3$s$3$p$ shell) has been investigated. Phung et al. \[7\] tackled semi-core correlation effects via a composite CASPT2/CC method, while Li Manni and Alavi \[8\] further enlarged the valence (32, 34) active space, including (3$s$, 3$p$) orbitals and their electrons, CAS(40,38). In both cases, inclusion of semi-core correlation had the effect to further stabilize the intermediate spin state over the high spin state.

In this work we analyze in greater detail the electronic structure of the quintet $^5A_{1g}$ and triplet $^3E_g$ states to quantitatively assess the electron correlation mechanisms that lead to the stabilization of the triplet over the quintet spin state.

In this work, the optimized CASSCF(32, 34) wave functions from Ref. 5 have been represented in the one-electron basis of localized orbitals, allowing to classify the leading electronic configurations of the CAS wave functions into on-site (local) excitations and CT excitations, and connect quantum chemical findings to the chemical concepts of $\sigma$ donation and $\pi$ backdonation. Due to the size of the active space and of the corresponding CI expansion the Full Configuration Quantum Monte Carlo (FCIQMC) and the Density Matrix Renormalization Group (DMRG) algorithms have been used as CI eigensolvers, here referred to as Stochastic-CASCI\[6\] and DMRG-CASCI.
Orbital entanglement measures based on DMRG for quantum chemistry provide a clear description of correlation effects that differentially stabilize the $3E_g$ over the $5A_{1g}$ spin state. Furthermore, in order to directly relate the role of the CT configurations to the triplet-quintet spin-gap, a strategy has been designed to selectively allow or forbid ligand-to-metal and metal-to-ligand CT excitations within the CAS(32, 34) configuration interaction expansion, while letting the rest of the wave function relax within the FCIQMC method and orbitals relax self-consistently. This protocol has been inspired by the generalized active space (GAS) procedure with disconnected-GAS subspaces, and is to be considered a preliminary version of the Stochastic-GASSCF method.

These restrictions in the active space wave function directly assess specific electron-correlation mechanisms, for example $\pi$ backdonation, while retaining effects related to mean-field interactions.

The DMRG and FCIQMC methods produce very similar and accurate energy estimates, while providing complementary information towards the understanding of the electronic structure of the investigated model system. On one hand, orbital entanglement measures emerge at no extra computational cost from the DMRG procedure, but would require the additional and expensive sampling of higher order reduced density matrices within the FCIQMC framework. On the other hand, FCIQMC gives direct access to the dominant electronic configurations that characterize the CI wave function of the spin states studied in this work. This would require an extra computationally demanding step in DMRG. In addition, FCIQMC in the stochastic-GASSCF method, allows to enforce restrictions on the type of excitations with ease, which is not straightforward in DMRG. As mentioned above, inclusion of excitation restrictions allows us to quantify the effect of the restricted excitations on the wave function energy and thus their role in stabilizing the triplet spin state.

The remainder of the article is organized as follows. In Section 2 we discuss the important elements of the FCIQMC and DMRG methods, the applied basis sets, and the novel algorithm for Generalized Active Space wave functions in the context of FCIQMC. Results are presented in Section 3 and discussed in Section 4. A summary of the paper and conclusions can be found in Section 5.

2 Computational Details

**Geometry.** All calculations were performed on the same Fe porphyrin model complex as in our earlier work. The pyrrole $\beta$-carbon atoms were removed and remaining bonds were capped with H-atoms. The Fe–N bond length was kept at 1.989 Å which is closer to the triplet’s equilibrium geometry. The molecule was placed into the $xy$ plane with ligand N-atoms on the $x$ and $y$ axes. The point group of the planar molecule is $D_{4h}$, however, calculations in the natural orbital basis were carried out using the $D_{2h}$ Abelian point group, the largest our programs can handle. In the localised basis all calculations have been performed using the $C_1$ point group symmetry.
**Basis Set.** Basis set and integral specifications have been chosen as in our earlier work [5]. Generally contracted atomic natural orbitals (ANO-RCC) of split-valence triple-\(\zeta\) plus polarization quality, were utilized (ANO-RCC-VTZP)[32, 33], resulting in a total of 707 basis functions. Scalar relativistic effects were included with the Douglas-Kroll-Hess integral correction. Evaluation of electron repulsion integrals has been accelerated by using the resolution-of-identity Cholesky decomposition with a threshold of \(1 \times 10^{-4} E_h\).[34, 35, 36, 37, 38]

**Active space.** The active space consists of nine doubly-occupied \(\pi\)-orbitals, seven empty \(\pi^*\)-orbitals, five metal centered 3\(d\)-orbitals, four doubly-occupied \(\sigma_N\)-orbitals, four empty orbitals of the (4s4p) shell, and five empty \(d'\)-orbitals (double-shell orbitals). With the \(d^6\)-configuration of the Fe(II) metal center, this implies a (32, 34) active space. The CASSCF(32, 34) optimized active natural orbitals for the quintet and triplet spin states are depicted in Figure 1 and Figure 2, respectively.

**Localisation procedures.** The natural active orbitals from Reference 5 have been localised using the method introduced by Pipek and Mezey [39]. Furthermore, Procrustes orthogonal transformations[40, 41] of the metal centered active orbitals were performed to make them as similar to the atomic orbital basis as possible. The resulting orbitals for the quintet state are displayed in Figure 3. The triplet localized orbitals are very similar and are reported in the supplementary material. The localised orbital basis neither diagonalize the one body density matrix, so it is not possible to assign occupation numbers, nor do they diagonalize the active space Fock matrix, so it is also not possible to assign orbital energies. Since the CASSCF energy expression is invariant under unitary transformations among the active orbitals, and the orbitals utilized were already the optimized CASSCF(32, 34) orbitals from our earlier work, further orbital reoptimization was not required and only Stochastic-CASCI and DMRG-CASCI calculations were performed in the localised basis. Localized orbitals have been chosen as they represent a convenient basis for analysis of electron correlation effects that involve CT excitations between the metal center and organic macrocycle.

Within the Stochastic-GASSCF method, localized active orbitals were separated into two GAS sub-spaces, namely a (18, 16) subspace, consisting exclusively of (\(\pi, \pi^*\)) orbitals, and a (14, 18) subspace consisting of (3\(d\), \(\sigma_N\), 4s, 4p, \(d'\))-orbitals. No excitations between the two active spaces were allowed (disconnected spaces), and orbitals have been relaxed under the mean-field generated by the GAS wave function[28]. For the Stochastic-GASSCF procedure a convergence threshold of \(1 \times 10^{-4} E_h\) over the energy value was utilized.

**Stochastic-GASSCF method.** GAS wave functions are built by defining a number of active subspaces; within each subspace a full CI expansion is generated, while restricting the number of interspace excitations [28, 42]. GAS spaces are defined “disconnected” if no excitations among them are permitted, while they are defined “connected” if interspace excitations are permitted. In the same GAS wave function both connected and disconnected spaces can exist. In GASSCF, orbitals are variationally optimized under the field generated by
Figure 1: Natural active orbitals for the $^5A_{1g}$ state of the Fe(II) porphyrin model system sorted by irreducible representation and occupation number. The four Gouterman $\pi$-frontier orbitals are labelled with the additional subscript $G$. 
the GAS-CI wave function. As in the CASSCF method, all intraspace orbital rotations in GASSCF (such as GAS1-GAS1, GAS2-GAS2) are “redundant”, as already described by the intra-space excitations of the GAS wave function, and can be excluded from the orbital optimization. Interspace orbital rotations (such as GAS1-GAS2), however, are only partially redundant, as inter-space excitations in the many-body expansion are absent or limited by the GAS rules. Thus, these rotations have to be considered in the orbital optimization step, in addition to inactive-active, inactive-virtual, and active-virtual rotations. Since the GAS restrictions reduce to CAS in the limit of the highest number of interspace excitations, the interspace rotations become more and more redundant for connected spaces with an increasing number of interspace excitations. This has the effect of slowing down the convergence of the GASSCF optimization.

Although highly constrained, GAS wave functions with disconnected spaces are of great theoretical and practical interest. From a practical standpoint they do not suffer of the “orbital redundancies” problems discussed above. Also, one-body density matrices of disconnected GAS spaces are block diagonal and the full diagonalization of them (natural orbitals) is equivalent to block diagonalization of separate GAS subspaces (“pseudo-natural orbitals”). This implies that the diagonalization of the one-body density matrix is an invariant transformation and keeps the GAS spaces separated. For connected GAS subspaces, the off-diagonal elements between orbitals belonging to different GAS spaces do not vanish and diagonalization of the one-body density matrix becomes a non-invariant rotation, because it mixes orbitals from different GAS subspaces. This implies that orbital occupation numbers are not well defined in this case.

From a chemical standpoint disconnected spaces can be utilized to selectively allow or forbid interspace excitations that can be related to specific electron cor-
Figure 3: Localised active orbitals for the $^5A_{1g}$ state of the Fe(II) porphyrin model system. The $\pi$-frontier orbitals $\pi_8, \pi_9, \pi^*_1, \pi^*_2$ get the additional subscript $G$ for Gouterman.
relation mechanisms, such as charge-transfer excitations. Including or removing entire classes of excitations has an impact on relative energies providing a direct and quantitative measure of the role of the class of excitations excluded. This feature will be used in this work to analyze the $\pi$ backdonation mechanism in the Fe(II) porphyrin.

From an algorithmic point of view the GAS ansatz is effectively a restriction of the Hamiltonian within the corresponding complete active space, removing matrix elements that violate the GAS conditions. Thus, GAS constraints can easily be implemented within the FCIQMC framework. Only the sampling of matrix elements under GAS restriction with disconnected spaces will be described here. For a detailed description of the FCIQMC algorithm, we refer to the literature [26, 27, 43, 44].

In FCIQMC, the sampling of the Hamiltonian reduces to randomly picking pairs of determinants $|i\rangle$, $|j\rangle$ with $H_{ij} \neq 0$ for given $|i\rangle$. This so-called excitation generation in practice requires a random selection of one/two electron/s and hole/s from $|i\rangle$ for single/double excitations [26]. This random sampling has to be restricted according to the GAS constraints.

For disconnected spaces where the particle number per space is constant, the excitation generator can easily and efficiently be implemented by random selection of electron(s) and subsequently of orbital(s) within the same GAS subspace.

The first step for the sampling of matrix elements is the random choice between a single or a double excitation, which is unaffected by the GAS approach. For single excitations, the GAS restrictions require the hole to come from the same GAS subspace as the electron; therefore, first an electron is chosen uniformly from all available electrons and then a hole is chosen from the same active subspace with probability, $p(a|i)$, proportional to the matrix element $H_{ij}$ corresponding to the single excitation, namely

$$p(a|i) = \frac{\langle D_a^i | H | D \rangle}{\sum_b \langle D_b^i | H | D \rangle},$$

where $|D_a^i\rangle = \hat{E}_{ai} |D\rangle$, represents a single excitation from the determinant, $|D\rangle$, and the denominator is a normalization factor with respect to all the possible single-excitations that can be generated from $|D\rangle$, such that all probability sum up to 1.

For double excitations, a pair of electrons is chosen uniformly, which also implies the selection of two (or one) active subspaces. From the first active subspace, a hole is then chosen uniformly. In the case of mixed-spin excitations (one $\alpha$-electron and one $\beta$-electron), the spin of the hole does not have to match that of the electron in the same active space and is chosen randomly. This criterion allows for spin-flip excitations. In the Stochastic-GAS code spin-flip excitations can also be avoided forcing the hole in one GAS subspace to have the same spin projection as the electron in the same subspace. The second hole is then picked from the active space of the second electron, again with probability
proportional to the corresponding double excitation matrix element $H_{ij}$

$$p(b|aij) = \frac{\langle D_{ij}^{\text{init}} | H | D \rangle}{\sum_c \langle D_{ij}^{\text{init}} | H | D \rangle}. \quad (2)$$

The spin is chosen to ensure conservation of the total spin projection (the one defined by the reference determinant). In the case where both electrons are coming from the same GAS subspace, this scheme reduces to a heat-bath sampling [45] with the first three steps replaced by a uniform selection.

**FCIQMC setup.** The FCIQMC setup was similar to earlier work[5] with the difference that the recently implemented adaptive shift approach[46] and in case of Stochastic CASCI, the precomputed heat bath excitation generator (PCHB), a variant of the heat bath sampling[45], were employed to speed up calculations. The adaptive shift varies the shift per non-initiator walker to cure the initiator approximation error and speeds up the convergence with respect to walker population. As the expression of the eigenvalue problem within the adaptive shift procedure is modified, also the reduced one- and two-body density matrices are to be sampled in a different way as compared to the conventional FCIQMC algorithm. For this work we used the following non-variational expression for the two-body density matrix elements

$$\Gamma_{pq,rs} = \sum_{D_{\text{init}} \in I} \frac{\langle D_{\text{init}} | \hat{e}_{pqrs} | \Psi_{\text{total}} \rangle}{\langle D_{\text{init}} | \Psi_{\text{total}} \rangle}. \quad (3)$$

where $I$ denotes the set of all initiator determinants and $\Psi_{\text{total}}$ the instantaneous stochastically sampled wave function. The PCHB excitation generator samples from non-uniform probability distributions in a different manner and with different scaling characteristics than the Cauchy Schwartz excitation generator used before.[5] Unlike the adaptive shift approach, PCHB excitation generation does not introduce further approximations to the dynamics[47] and up to stochastic noise the final result for a given number of walkers is not affected, but the time per iteration is reduced.

In the CASCI(32, 34) calculations the walker population was increased up to $2 \times 10^9$ walkers, to reach convergence, while in the Stochastic-GASSCF([18, 16], (14, 18]) calculations a much smaller population of up to $2 \times 10^8$ walkers was sufficient, due to the considerably smaller GASCI expansion that follows from the constraints on the CT excitations. This speedup can be estimated by comparing the different sizes of the configuration spaces. The number of determinants for a given number of electrons $N$, orbitals $n$, and spin projection $S_z$ in an active space is given by

$$\Omega(N, n, S_z) = \binom{n}{N_\alpha} \cdot \binom{n}{N_\beta} = \binom{N}{N_\alpha + S_z} \cdot \binom{N}{N_\beta - S_z}. \quad (4)$$

Considering that only the total spin projection is specified, and that spin projection fluctuations have been permitted within each GAS subspace, the ratio
between CAS and GAS configurations is for the triplet promptly obtained as:

$$\frac{\Omega(\text{CAS})}{\Omega(\text{GAS})} = \frac{\Omega(32, 34, 1)}{\sum_{S_z = -6}^{7} \Omega(18, 16, S_z) \cdot \Omega(14, 18, 1 - S_z)} \approx 14.1 \quad (5)$$

**DMRG setup.** To obtain the most accurate estimate of the exact full CI energy within the DMRG approach, a maximum bond dimension \(m\) value of 10000 was chosen for the natural orbital basis in \(D_{2h}\) point group symmetry. For the localised basis (\(C_1\) point group symmetry) a lower \(m\) value of 4000 had to be chosen due to a larger computational cost. Each DMRG-CASCI calculation has been run with up to 20 sweeps, however the simulation was stopped after the energy difference from the previous sweep has reached \(1 \times 10^{-6} \text{E}_h\).

To avoid convergence to local minima, orbital ordering according to the Fiedler vector of the mutual information matrix [48] and the perturbative correction according to Ref. 49 were employed, the latter one only for the first six sweeps in each calculation. Orbital entanglement measures, i.e. the single-orbital entropy and mutual information (see below) based on DMRG-CI calculations have been computed both for the natural and the localised orbital basis.

**Orbital entanglement.** Consider partitioning the system into two subsystems, \(|a\rangle\) and \(|b\rangle\), where \(|a\rangle\) consists only of one orbital and \(|b\rangle\) of the remainder of the active space. We may obtain the one-orbital reduced density matrix (1o-RDM) by tracing out the states belonging to the environment:

$$\rho_{aa'} = \sum_{b \in |b\rangle} c_{ab} c_{a'b}$$

where \(c_{ab}\) are the CI coefficients for configurations with a given occupation of the subsystems \(a\) and \(b\). Since each orbital may have four possible occupation states: empty, spin-up, spin-down or doubly-occupied, \(\rho\) has a dimension of \(4 \times 4\). We may define the single-orbital entropy \(s(1)_i\) [9, 10, 13, 14] for each orbital \(i\) as a von Neumann entropy term:

$$s(1)_i = -\text{Tr}(\rho \ln \rho) = \sum_{\alpha=1}^{4} \omega_{\alpha,i} \ln \omega_{\alpha,i} \quad (7)$$

where \(\omega_{\alpha,i}\) are the eigenvalues of the 1o-RDM. The single-orbital entropy quantifies the entanglement of the orbital \(i\) with all other orbitals.

Analogously, we may define for an orbital pair \(i, j\) a two-orbital RDM (2o-RDM) and the two-orbital entropy \(s(2)_{ij}\) by placing two orbitals into the subsystem \(a\) in Eq. 6. \(s(2)_{ij}\) describes how two orbitals are entangled with all other orbitals. From the expressions for the single-orbital entropy and two-orbital entropy, we may define the mutual information \(I_{ij}\) as

$$I_{ij} = \frac{1}{2} (s(1)_i + s(1)_j - s(2)_{ij}) (1 - \delta_{ij}) , \quad (8)$$
which describes the entanglement of two orbitals $i$ and $j$ with each other.

**Software.** The Stochastic-CASSCF and Stochastic-GASSCF calculations have been performed within the OpenMolcas chemistry software package[50] interfaced to the NECI program[47] responsible for the FCIQMC CI optimizations.

The DMRG-CASCI calculations were performed with the QCMAquis[51, 52] program, using the same molecular orbital electron integral file (FCIDUMP) produced by OpenMolcas. The entanglement diagrams were produced by the AutoCAS[53, 54] program.

## 3 Results

### Energies and Occupation numbers.**

Stochastic-CASSCF and DMRG-CASCI total energies for the $3E_g$ and $5A_{1g}$ spin states, in the localised and delocalised molecular orbital basis, and Stochastic-GASSCF in the localized orbital basis are summarised in Table 1.

<table>
<thead>
<tr>
<th>State</th>
<th>Stochastic-CASSCF</th>
<th>DMRG-CASCI</th>
<th>Stochastic-GASSCF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Localised (2B walkers)</td>
<td>Delocalised (1B walkers[5])</td>
<td>Localised ($m = 4000$)</td>
</tr>
<tr>
<td>$3E_g$</td>
<td>$-1951.3586(4)$</td>
<td>$-1951.3580(2)$</td>
<td>$-1951.358262$</td>
</tr>
<tr>
<td>$5A_{1g}$</td>
<td>$-1951.3530(3)$</td>
<td>$-1951.3530(2)$</td>
<td>$-1951.352797$</td>
</tr>
<tr>
<td>$\Delta E$</td>
<td>$-3.5(3)$</td>
<td>$-3.1(2)$</td>
<td>$-3.5$</td>
</tr>
</tbody>
</table>

The occupation numbers obtained from our DMRG-CASCI calculations in the delocalised basis, are in very good agreement with our Stochastic CASSCF calculation from previous work[5] and can be found in the Supporting information.

### CI coefficients.

The leading $10^4$ Slater determinants obtained from the FCIQMC dynamic with $2 \times 10^9$ walkers in the localized orbital basis, were grouped into chemically relevant subsets according to the type of excitations they represent with respect to the Hartree-Fock reference determinant (see Ta...
For a set of orbitals $O$ and $P$ the expression $(O \nearrow P)$ denotes the set of all Slater determinants that are generated from the reference by at least one annihilation in $O$ and one creation in $P$. These subsets can be combined among each other to define new sets by forming unions, denoted with $\cup$, intersections, denoted with $\cap$. A subset relationship will be denoted with $\subset$. The variable $\Omega$ denotes the set of all active orbitals in our system. It has to be emphasized, that $(O \nearrow P)$ may contain simultaneous excitations from and into orbitals which are neither contained in $O$ nor $P$. If we require that only electrons in $O$ are annihilated and only in $P$ created, we write $\text{pure}(O \nearrow P)$ for the set of pure excitations.

In Table 2 the weights of these subsets have been reported as normalized values with respect to the entire wave function $(p)$, and in an intermediate-normalization $(\hat{p})$, with respect to the most populated reference determinant. The latter has been utilized to allow a more direct comparison of equivalent excitations in the wave functions of the two spin states considered. In fact, if two multiconfigurational wave functions with different weights on the reference determinants are to be compared, a direct comparison of the normalized weights might be misleading. Even if the weight of an excited determinant is the same in both wave functions, it represents a larger correction for that wave function with a smaller weight in the reference determinant. Intermediate normalization obviates this problem.

**Orbital entanglement.** Figure 4 and Figure 5 show the DMRG-CASCI orbital entanglement diagrams for the $3E_g$ and $5A_{1g}$ spin states in the delocalized (natural orbitals, Figure 1 and Figure 2) and localised orbital basis (Fig. 3), respectively.

**GASSCF density difference.** Total electron density differences between Stochastic-CASSCF and Stochastic-GASSCF wave functions for the $3E_g$ and $5A_{1g}$ spin states are displayed in Figure 6. For GAS wave functions with disconnected spaces it is possible to perform invariant orbital transformation of the GASSCF orbitals into natural orbitals (equivalent to pseudo-natural orbitals) and the electron density is then given by

$$\rho(r) = \sum_{i} \omega_i |\phi_i(r)|^2,$$

where the sum extends over the occupied orbitals (inactive and active), $\phi_i$ refer to the natural orbitals, and $\omega_i$ to their occupation numbers. For inactive orbitals $\omega_i = 2$, and for active orbitals $\omega_i \in [0, 2]$. In the case of a GAS wave function with connected spaces, Equation 9 does not hold and should be replaced with the more general form

$$\rho(r) = \sum_{p,q} D_{pq} \phi^*_p(r) \phi_q(r).$$

**Computational Cost.** Electron integral evaluation and their transformation to MO basis and Super-CI optimization of orbitals within the Stochastic-GASSCF procedure have been performed on a typical desktop workstation. The
Table 2: Percent weight, \( p \), of different subsets of the \( 10^4 \) most populated Slater determinants in the CI expansion for the localised basis. The intermediately normalized weight \( \hat{p} = \frac{p}{p_{\text{ref}}} \) is normalized onto the weight of the reference determinant. The first four sets of excitations do not involve \( \pi \)-orbitals of the macrocycle, and are already contained in the smaller active spaces discussed in Reference [5]. The subsequent sets involve \( \pi \)-orbitals of the macrocycle. The variable \( \pi_G \) indicates the four Gouterman \( \pi \) orbitals, \( \pi_8, \pi_9, \pi^*_1, \pi^*_2 \) (Figure 3.8, 3.9, 3.28, and 3.29). If a set of excitations is a subset of another set in the table, it is indented.

<table>
<thead>
<tr>
<th>Slater-Determinants</th>
<th>( p )</th>
<th>( \hat{p} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( ^5A_{1g} )</td>
<td>( ^3E_g )</td>
</tr>
<tr>
<td>All</td>
<td>89.44</td>
<td>83.26</td>
</tr>
<tr>
<td>Reference</td>
<td>32.21</td>
<td>18.74</td>
</tr>
<tr>
<td>((\sigma_N \nearrow 3d))</td>
<td>12.46</td>
<td>33.10</td>
</tr>
<tr>
<td>((\sigma_{4N} \nearrow 3d_{x^2-y^2}) \subset (\sigma_N \nearrow 3d))</td>
<td>12.20</td>
<td>32.31</td>
</tr>
<tr>
<td>((3d \nearrow 3d))</td>
<td>1.42</td>
<td>16.05</td>
</tr>
<tr>
<td>((3d \nearrow d'))</td>
<td>1.40</td>
<td>2.72</td>
</tr>
<tr>
<td>((\pi \nearrow 4p_z))</td>
<td>22.87</td>
<td>19.40</td>
</tr>
<tr>
<td>((3d \nearrow \pi^*))</td>
<td>1.30</td>
<td>7.56</td>
</tr>
<tr>
<td>((3d \nearrow \pi_G^*))</td>
<td>1.11</td>
<td>6.78</td>
</tr>
<tr>
<td>pure((3d \nearrow \pi_G^*))</td>
<td>0.27</td>
<td>0.63</td>
</tr>
<tr>
<td>((3d \nearrow \pi_G^<em>) \cap \text{pure}(\Omega \nearrow \pi_G^</em> \cup 4s \cup 4p \cup d'))</td>
<td>0.64</td>
<td>1.28</td>
</tr>
<tr>
<td>((\pi \nearrow d'))</td>
<td>8.81</td>
<td>5.82</td>
</tr>
</tbody>
</table>
denote connecting the orbitals show their mutual information value $I_{ij}$. The area
of the grey circles is proportional to the single-orbital entropy $s(1)_i$, the lines
connecting the orbitals show their mutual information value $I_{ij}$. Solid lines
denote $I_{ij}$ values of at least 0.1, dashed lines $I_{ij}$ values of at least 0.01. The
orbital numbering is the one used in Fig. 1 and 2.

Figure 4: DMRG-CASCI entanglement diagrams for the (a) $^3E_g$ and (b) $^5A_{1g}$
spin states, in the natural orbital basis of the Fe(II) porphyrin model. The area
of the grey circles is proportional to the single-orbital entropy $s(1)_i$, the lines
connecting the orbitals show their mutual information value $I_{ij}$. Solid lines
denote $I_{ij}$ values of at least 0.1, dashed lines $I_{ij}$ values of at least 0.01. The
orbital numbering is the one used in Fig. 1 and 2.

Figure 5: DMRG-CASCI orbital entanglement diagrams for the (a) $^3E_g$ and (b) $^5A_{1g}$
spin states of the Fe(II) porphyrin model in the localised orbital basis.
Plot legend as in Fig. 4. The orbital numbering is the one given in Fig. 3.
The two GAS subspaces used in the Stochastic-GASSCF[(18, 16), (14, 18)] are
highlighted with the blue and green background, respectively.
Figure 6: Density difference between Stochastic-CASSCF(32, 34) and Stochastic-GASSCF[(18, 16), (14, 18)] optimized orbitals. We define \( \Delta \rho = \rho(\text{CASSCF}) - \rho(\text{GASSCF}) \). Blue denotes positive and yellow negative values. The iso-surface value is 0.008 Å\(^{-3}\).

FCIQMC and DMRG CI optimizations represent the most demanding part of our calculations and were performed on dedicated computers. The FCIQMC dynamics in the localised basis with \( 2 \times 10^9 \) walkers were performed on the Max-Planck Computing and Data Facility (MPCDF) HPC system Cobra, using 128 nodes, each having two Intel Xeon Gold 6148 (20 cores, 2.4 GHz) processors and 192 GB DDR4 RAM, with the nodes connected through Intel OmniPath interconnect. The calculations used 38 processes and 95 GB of memory per node and ran for 10.8 h resulting in a total of \( 5.25 \times 10^4 \) core hours. Since FCIQMC scales near-linearly with the walker number and is near-linearly parallelized\(^4\), Stochastic-GASSCF calculations with \( 2 \times 10^8 \) walkers required less resources by a factor of 10 and were performed on in-house clusters.

The DMRG-CASCI calculations in the localised basis took 280 h on 32 cores of a node with two Intel Xeon Gold 6138 (2.0 GHz) processors, requiring 560 GB of RAM, resulting in 8960 core hours.

4 Discussion

CASCI energies are invariant under unitary orbital rotations in the active space. In the converged limit, this is also true for FCIQMC and DMRG energy estimates. Upon localisation of the Stochastic-CASSCF natural orbitals the number of determinants in the corresponding CI expansion increases as compared to the CI expansion in the natural orbital basis. This is best illustrated by considering the \( \sigma \) bonding and \( \sigma^* \) antibonding molecular orbitals in the delocalised basis (Figure 1.11 and 1.12). Upon localisation into the corresponding \( \sigma_{4N} \) and \( 3d_{x^2-y^2} \) orbitals (Figure 3.18 and 3.12), the atomic orbital mixing has
to be described at the level of the CI expansion. Additionally, the $D_{2h}$ point group symmetry cannot be utilized in the localized basis; instead the $C_1$ point group symmetry has been employed, that leads to a further increase in the size of the CI expansion. The increased multiconfigurational character of the wave function leads to slower convergence of both FCIQMC and DMRG calculations with respect to the number of walkers and bond dimension value, respectively.

The Stochastic-CASCI calculations in the localised basis, using the adaptive shift procedure and $2 \times 10^9$ walkers, resulted in similar total energies and spin gaps as compared to published values in the natural orbital basis\[8\], where no adaptive shift technique was employed and $1 \times 10^9$ walkers used. A difference of $6 \times 10^{-4} \, E_h$ between localised and delocalised basis is obtained (see Table 1), indicating the good convergence of the FCIQMC dynamics and that a higher walker population is generally necessary when covalent bondings are described by the CI expansion instead of orbital mixing.

For the DMRG calculations, the difference between energies in the localised and delocalised orbital basis is in the order of $7 \times 10^{-4} \, E_h$, and can be explained by considering that a smaller $m$ value had to be employed for the localized basis, due to the increased computational requirements linked to the $C_1$ point group symmetry utilized.

The loss in accuracy under orbital localisation is similar for the FCIQMC and DMRG procedures, and they yield the same spin gaps in the same orbital basis (see Table 1). Not only the energy, but other properties of the wave function, such as the natural orbital occupation numbers, are very similar for the two methods.

Although FCIQMC and DMRG algorithms are conceptually different, they are qualitatively and quantitatively in agreement in the description of the Fe(II) porphyrin model system within the $(32, 34)$ active space. This agreement allows to compare FCIQMC and DMRG properties directly, and analyze orbital entanglement and CI coefficients of various classes of excitations on equal footing.

### Natural Orbital Basis.

The most prominent feature in both $^3E_g$ and $^5A_{1g}$ states are the Gouterman’s orbitals 16, 24, 29, and 33 (see Fig. 1), which show the strongest entanglement for both spin states (Figure 4). As already anticipated by Gouterman\[55\], these orbitals play an important role within the $\pi$ system, and as previously discussed by us\[5\], they are crucial in differentially stabilizing the triplet over the quintet spin states. The entanglement diagrams effectively confirm the crucial role of the Gouterman orbitals. However, the entanglement diagrams also show that other $\pi$ orbitals are strongly correlated (e.g. orbitals 21, 27 and 34) as they are correlated to one or more Gouterman’s orbitals and show large values of single-orbital entropy. Generally speaking the entire system of $\pi$ orbitals is important to capture the important correlation effects related to the aromatic macrocycle.

From the entanglement diagrams of Figure 4 another aspect of the triplet spin state stabilization, already discussed in Ref. 5, is highlighted, namely the
breathing effect of the 3d electrons. This is exemplified by the entanglement between orbitals 22, 23, 24 and 26, 29, 30, respectively. At the first glance, the entanglement within these orbital triples is similar in both states and not very strong. However, orbital 26, which is relevant for the orbital breathing phenomenon, along with orbitals 27, 28, and 31 in the natural orbital basis, are localised in the $^5A_{1g}$ state but delocalised in the $^3E_g$ spin state (see Figure 2). Thus, it is not possible to directly compare the entanglement diagrams of the $^3E_g$ and $^5A_{1g}$ spin states and draw conclusions on the role of orbital breathing and charge transfer processes. The role of orbital breathing and CT mechanisms are better illustrated by the entanglement diagrams in the localised orbital basis (Figure 5), discussed in the next section.

Localised Orbital Basis. As shown in Table 2, the normalized weight ($p$) of the reference determinant of the triplet is substantially smaller than the normalized weight of the reference determinant of the quintet spin state, indicating that correlation effects are in general more pronounced for the triplet. We will discuss metal centered correlation effects first and CT effects between metal and $\pi$ system later.

The $\sigma_{4N}$ and $3d_{z^2-y^2}$ orbitals (12 and 18 in Figure 3), have the largest entanglement (see Figure 5), and describe the $\sigma$ bonding interaction between the metal center and the macrocycle. In the quintet state the $\sigma$-bond entanglement is less pronounced, providing a simple route for an explanation of the geometrical differences between triplet and quintet spin states, with the quintet states in general characterized by larger Fe-N bond distances. This result identically emerges from the inspection of the CI coefficients listed in Table 2. In fact, considering the normalized weights ($p$), the set of excitations from $\sigma_N$ to 3$d$ orbitals makes up 33% of the total wave function in the triplet, and 12% in the quintet state and are dominated by the ($\sigma_{4N} \overset{\sigma}{\rightarrow} 3d_{z^2-y^2}$) excitations. As expected, the correlation within the metal 3$d$ orbitals (orbitals 14–18 in Figure 3) is stronger in the triplet spin state, and it is clearly shown both by the orbital entanglement diagrams and the CI weight of (3$d$ $\not\rightarrow$ 3$d$) excitations. Both the 3$d$ correlation and $\sigma$ donation are already contained in the (8, 6) minimal active space. The double shell correlation of 3$d$ orbitals with the correlating $d'$ functions (19 to 23 in Figure 3) accounts for the 2.72% and 1.40% weight for the triplet and quintet state, respectively, and represents an important factor for a balanced description of the wave functions of the Fe(II) porphyrin, that is captured in the (8, 11) active space utilized by Pierloot and Co-workers.\[56\]

Charge-transfer excitations between metal orbitals and $\pi$ orbitals of the macrocycle are only contained in the large (32, 34) active space. These excitations differentially favour the triplet over the quintet. The entanglement between the $\pi_g$ and $4p_z$ orbitals (9 and 27 in Figure 5) is the second largest entanglement value with a corresponding high CI weight for ($\pi \overset{\pi}{\rightarrow} 4p_z$) in both spin states. Although the normalized weight is slightly higher for the quintet spin state, with 19.40% and 22.87%, for the weight of the triplet and quintet, respectively, in the intermediate normalization this excitation has a larger im-
pact on the wave function of the triplet spin state. The $4p_z$ orbital is empty
in the reference determinant and has the correct symmetry to be used as cor-
relating orbital for the electrons of the $\pi$ system. It is to be expected that this
excitation will reduce in weight if correlating orbitals were included into the
active space. This aspect is currently under investigation.

As the entanglement diagrams (Figure 5) and the CI weight of $(3d \uparrow \pi^*)$
show (Table 2), there is significant $\pi$ backdonation from metal $3d$ orbitals into
the $\pi$ system. This correlation is much more pronounced in the triplet, hence
having a major impact on the spin gap. The $\pi$ backdonation mainly involves
the Gouterman’s orbitals which is illustrated by the CI weight of the restricted
subset $(3d \uparrow \pi^*_G) \subset (3d \uparrow \pi^*)$. This special role of the Gouterman’s orbitals
is also manifested in their single orbital entropy, which is larger than of the
rest of the $\pi$ orbitals. On the other hand it has to be emphasized, that the
CI weight of the pure excitations pure$(3d \uparrow \pi^*_G)$ is significantly lower than the
case where they are coupled to other excitations, which illustrates that the $\pi$
backdonation is not a pure excitation from $3d$ into Gouterman’s orbitals, but a
concerted mechanism. In our previous work [5] there was already evidence for
this observation because natural orbital occupation numbers of Gouterman’s
orbitals were close to 0.0 and 2.0 when they were added to an active space
consisting of solely metal orbitals. Only when the full $\pi$ system was correlated,
the occupation numbers of the Gouterman orbitals changed to approximatively
1.7 and 0.3.[5]

The “breathing” of the correlated $3d$ electrons enables an other complex
correlation mechanism that involves metal $3d$-orbitals, correlating $d'$ orbitals
and $\pi^*$ orbitals. This mechanism is also captured by the entanglement diagrams
(Figure 5) and corresponds to the entanglement of the $3dz^2$, its corresponding
correlating $d'_{xz}$ orbital and the $\pi^*_2$ orbital (Figures 3.15, 3.20, and 3.29)
and, similarly, the $3dyz - d'_{yz} - \pi^*_1$ interaction (Figures 3.16, 3.21, and 3.28).
The entanglement diagram (Figure 5) shows that the orbitals are more entangled
in the triplet state. Also in the CI weights (Table 2) the “breathing” manifests
itself in the weight of the excitations in $(3d \uparrow \pi^*_G) \cap \text{pure}(\Omega \uparrow \pi^*_G \cup d\cup p\cup d')$.
While the weight of this subset is similar for both spin states, the intermediate-
normalized weight of 7% shows, that the “breathing” is only for the triplet a
substantial correction to the reference wave function.

The $\pi \rightarrow d'$ ligand-to-metal charge-transfer excitations are important contrib-
tions to the overall wave function as illustrated by the CI weight of $(\pi \uparrow d')$
(Table 2). Figure 5 shows entangled orbital pairs $\pi_6 - d'_{xz}$ (6 - 20) and $\pi_7 - d'_{yz}$
(7 - 21) as most relevant for these excitations, which provide a mechanism for
the charge delocalisation similar to orbital breathing but working in the opposite
direction, a mechanism that reduces correlation at the level of the macrocycle
via delocalization of charge.

**Stochastic-GASSCF** In the previous paragraph we have discussed the contribu-
tion of correlation effects that involve metal and $\pi$ orbitals by inspection of
orbital entanglement and CI weights of various classes of excitations. In this sec-
tion the Stochastic-GASSCF approach will be utilized to quantitively determine the impact of such mechanisms on spin-state energetics.

As described in the computational details (Section 2) the localised (32, 34) active space was separated into a (18, 16) GAS subspace containing only \((\pi, \pi^*)\) orbitals, and a (14, 18) GAS subspace containing the \((3d, \sigma_N, 4s, 4p, d')\) orbitals. The separation of the two GAS subspaces is visualised in the entanglement diagrams of Figure 5.

As it is expected from the Hylleraas-Undheim-MacDonald theorem, the variational energy is increased upon reduction of the configuration space that follows the GAS restrictions (see Table 1).\cite{57, 58} More interesting is the fact, that the spin gap is reduced by 2.5 kcal mol\(^{-1}\) showing that the triplet is stabilized by correlation effects among electrons of the metal center and of the \(\pi\) system. These effects are similar in magnitude to the semicore correlation effects reported by Phung et al.\cite{7} and Li Manni et al.\cite{8}.

However, the fact that the Stochastic-GASSCF[(18, 16), (14, 18)] still favors the triplet spin state, suggests that the stabilization of the triplet is not exclusively a consequence of many-body correlation effects; it also relates to mean-field effects of the correlated macrocycle. In fact, in the GASSCF procedure, although interspace excitations have been forbidden, the intra-space excitation within the \(\pi\)-system are still considered, and they are capable to change the mean-(ligand-)field effect onto the metal center. This is to be compared to the case where the \(\pi\) system is left outside the active space and treated at the single configurational level of theory. This justifies why, despite CT configurations are removed, the triplet is still lower in energy.

The impact of GASSCF constraints can also be observed in the electron density difference between the CASSCF and GASSCF wave functions \(\Delta \rho_1 := \rho_{\text{CASSCF}} - \rho_{\text{GASSCF}}\) (see Figure 6). In going from GASSCF, where CT configurations are excluded from the wave function, to CASSCF, where all possible configurations are allowed, the density increases in the \(\pi\) system, and decreases in the metal. This in turn polarizes the nitrogen and enhances sigma donation. These effects are present in both spin states, but more pronounced in the triplet. A similar analysis was performed in our previous work \cite{8} where the density difference between the Stochastic-CASSCF(32, 34) and the restricted open-shell Hartree-Fock (ROHF) wave function \(\Delta \rho_2 := \rho_{\text{CASSCF}} - \rho_{\text{ROHF}}\) was considered. As qualitatively expected, the overall difference between the uncorrelated ROHF and CASSCF density is higher than between GASSCF and CASSCF \((\int |\Delta \rho_2(r)| \, dr > \int |\Delta \rho_1(r)| \, dr\). The polarization of the nitrogen relative to the density change in the \(\pi\) system is larger in the case of \(\Delta \rho_1\), while the nitrogen polarization and density changes in the \(\pi\) system are of similar magnitude in the case of \(\Delta \rho_2\). This can be explained by the fact that \((\pi \uparrow \pi^*)\) excitations are already contained in the GASSCF, but not in the uncorrelated ROHF wave function. This intraspace correlation affects the mean field of the macrocycle onto the metal center and differentially stabilizes the triplet as discussed above.
5 Conclusion

In this work, Stochastic-CASSCF, Stochastic-GASSCF and DMRG-CASCI procedures have been applied to an Fe(II) porphyrin model complex in a localized orbital basis, to understand the correlation effects that govern the spin gap between the $^5A_{1g}$ and $^3E_g$ electronic states. The choice of the localized orbitals has been motivated by the possibility to conveniently separate local excitations from CT excitations between metal center and macrocycle.

The FCIQMC and DMRG approaches have provided insights that are in qualitatively and quantitatively good agreement, with a difference of $5 \times 10^{-4} E_h$ in total energies, a negligible difference of $< 1 \times 10^{-1} \text{ kcal mol}^{-1}$ in the spin gap estimates, and very similar natural orbital occupation numbers.

From the two methods, two complementary pieces of information on the role of electron correlation effects in metal-porphyrins were obtained. With FCIQMC it has been possible to directly analyse the leading electronic configurations that characterize the multi-configurational wave function of the $^3E_g$ and $^5A_{1g}$ spin-states, while the DMRG procedure, in the same orbital basis, has produced entanglement diagrams that facilitate the understanding of complex electron correlation effects in metal-porphyrins.

It has been shown that multi-orbital correlation mechanisms between the metal and $\pi$ system, such as the $3d \nearrow d' \nearrow \pi^*$ “breathing” processes are required for a qualitatively correct description of the wave function, and they differentially stabilize the intermediate spin state over the high spin states.

A remarkable finding is about the role of the Gouterman’s orbitals. It has been demonstrated that the $\pi$ backdonation $3d \nearrow \pi^*$ happens mainly into the Gouterman’s orbitals (large values of orbital entanglement), but requires simultaneous excitations in the whole $\pi$-system. This feature is at the core of the necessity to include the whole $\pi$ system of the macrocycle into the active space, and at the core of the failure of second order perturbation procedures described in earlier work.

The Stochastic-GASSCF method with disconnected spaces has been developed within the FCIQMC framework, to quantify the impact of charge-transfer excitations. Within this method, localized metal centered orbitals and $\pi$ orbitals of the macrocycle were separated into two active subspaces. With no interspace excitations, this setup effectively removed the stabilizing CT configurations from the many-body expansion of the two spin states considered, responsible for the $\pi$ backdonation mechanism, and reduced the spin gap from $-3.5 \text{ kcal mol}^{-1}$ to $-1.0 \text{ kcal mol}^{-1}$.

The density difference between the Stochastic-CASSCF and Stochastic-GASSCF wave functions shows that CT correlation effects between the metal and macrocycle amplify $3d \nearrow \pi^*$ backdonation and $\sigma$ donation of nitrogen ligands, and in doing so, confirming the suggested correlation enhanced $\sigma$ donation / $\pi$ backdonation mechanism.

One important question is whether the results obtained in this work can be transferred to the real Fe(II) porphyrin system, with its pyrrole $\beta$-carbon atoms, and spin gaps predictions in adiabatic conditions. The model system without $\beta$-
carbons utilized in this work contains a planar and fully conjugated macrocycle, that mimics the chemical environment around the metal center of real metal-porphyrins. The β-carbons in Fe(II) porphyrin, increase ring-correlation at the level of the π-system of the macrocycle. Thus, further stabilization of the triplet over the quintet is to be expected when considering systems that contain the pyrrole β-carbons.

Adiabatic conditions will lower the energy of the quintet state with respect to the same state at the triplet geometry. The correlation mechanisms described here will still play an important role in differentially stabilizing the intermediate spin state. Nevertheless it might be necessary to include even more correlation effects for a correct description in adiabatic conditions, considering that convergence of the spin-gap with respect to the active space size has not been reached within our calculations, as shown in Figure 2 of Reference [5]. The identification and quantification of further correlation effects are currently being investigated.

Future work on Stochastic-GASSCF will develop in two directions. One goal is to enable interspace excitations which can be used in a similar but more flexible way as the disconnected spaces in this work. Another goal is to apply the approximate, precomputed heat-bath excitation generation and other more efficient sampling strategies in the GAS framework for enhanced performance [45, 59]. Considering the reduced computational costs associated with Stochastic-GASSCF as compared to the corresponding Stochastic-CASSCF wave function optimization, larger active spaces are within reach, providing practical routes for investigating correlation mechanisms.

Acknowledgements

The authors thank Armin Burkhardt and Armin Schuhmacher (MPI FKF) and Max-Planck Computing and Data Facility for keeping the respective computer clusters up and running in the difficult situation of the COVID-19 pandemic and Markus Reiher (ETH Zürich) for helpful discussions.

Funding Information

O.W., A.A. and G.L.M. acknowledge the Max Planck Society for financial support. L. F. acknowledges the Austrian Science Fund (FWF) for the Schrödinger fellowship (Project J 3935-N34).

Conflict of interest

The authors declare no competing financial interest.
The role of charge transfer electron correlation mechanisms that are responsible for the differential stabilization of the $^3E_g$ over the $^5A_{1g}$ state in a Fe porphyrin model system, have been studied by Stochastic-CASSCF, DMRG and a preliminary version of the Stochastic-GASSCF method. Orbital entanglement diagrams and CI coefficients of the wave function in a localised orbital basis allow for an effective interpretation of charge transfer configurations. Within Stochastic-GASSCF, excitations between metal and ligand orbitals have been selectively removed from the complete CI expansion. These excitations are key to the differential stabilization of the triplet and once removed the spin-gap noticeably reduced.

References


[34] Francesco Aquilante, Roland Lindh, and Thomas Bondo Pedersen. Unbi-

[35] Francesco Aquilante, Thomas Bondo Pedersen, and Roland Lindh. Low-


[41] Xenophon. Memorabilia. 371 BC.


