

# MSINDO-sCIS: A New Method for the Calculation of Excited States of Large Molecules

Immanuel Gadaczek,\* Katharina Krause, Kim Julia Hintze, and Thomas Bredow

Institut für Physikalische und Theoretische Chemie, Universität Bonn, Wegelerstr. 12, 53115 Bonn, Germany

**S** Supporting Information

**ABSTRACT:** Theoretical background, parametrization, and performance of the semiempirical configuration interaction singles (CIS) method MSINDO-sCIS designed for the calculation of optical spectra of large organic molecules are presented. The CIS Hamiltonian is modified by scaling of the Coulomb and exchange integrals and a semiempirical correction. For a recently proposed benchmark set of 28 medium-sized organic molecules, vertical excitation energies for singlet and triplet states are calculated and statistically evaluated. A full reparameterization of the MSINDO method for both ground and excited state properties was necessary. The results of the reparameterized MSINDO-sCIS method are compared to the currently best semiempirical method for excited states, OM3-CISDTQ, and to other standard methods, such as MNDO and INDO/S. The mean absolute deviation with respect to the theoretical best estimates (TBEs) for MSINDO-sCIS is 0.44 eV, comparable to the OM3 method but significantly smaller than for INDO/S. The computational effort is strongly reduced compared to OM3-CISDTQ and OM3-MRCISD, since only single excitations are taken into account. Higher excitations are implicitly included by parametrization and an empirical correction term. By application of the Davidson–Liu block diagonalization method, high computational efficiency is achieved. Furthermore, it is demonstrated that the MSINDO-sCIS method correctly describes charge-transfer (CT) states that represent a problem for time-dependent density functional theory (TD-DFT) methods.

## 1. INTRODUCTION

During recent decades, there has been substantial progress in the experimental and theoretical characterization of electronically excited states. New experimental techniques provide better insight into photophysical processes in molecules.<sup>2,3</sup> On the other hand, more and more efficient and accurate theoretical methods for the description of excited states have been developed.<sup>4,5</sup> Multireference configuration interaction (MR-CI),<sup>6</sup> multistate complete active-space second order perturbation theory (MS-CASPT2),<sup>7</sup> and coupled cluster methods (CC2,<sup>8</sup> CCSDT,<sup>9</sup> CC3<sup>10</sup>) are well-established and highly accurate but, at the same time, extremely costly. Recently published benchmark calculations demonstrate the advance of CCx<sup>11</sup> methods. Even with relatively small basis sets, they provide reasonable agreement with more elaborate ab initio approaches. Since these highly accurate methods are only applicable for small molecules, time-dependent density functional theory (TD-DFT)<sup>12</sup> has become one of the most popular methods for spectra prediction and the description of excited states due to its reliability at low computational cost in most cases. Yet the application of TD-DFT is at present limited to systems with a few hundred atoms. Although TD-DFT is an attractive choice for the description of excited states, there are well-documented problems,<sup>13,14</sup> in particular for charge-transfer (CT) states. Even hybrid methods and the perturbative corrected double-hybrid methods<sup>15</sup> do not give a quantitatively correct description. Moreover, the accuracy of TD-DFT is limited in general. Recently, Thiel et al.<sup>16</sup> have shown that the vertical excitation energies obtained from TD-DFT have mean absolute deviations of 0.3–0.5 eV from the theoretical best estimates (TBEs). This raises the question of how, e.g., novel organic solar cells, which normally consist of

large donor and acceptor species with many hundred atoms, can be treated theoretically, when no applicable method exists. For such problems, a reliable method which is also efficient is desirable. In this context, semiempirical methods have come back into focus for the description of excited states in larger molecules.<sup>17</sup> Due to the integral approximations made in all semiempirical methods, the computational effort is decreased by orders of magnitude<sup>18,19</sup> compared to first-principles methods. Therefore systems with up to thousands of atoms can be treated. The error introduced by the vast integral approximations is compensated by calibration of the method against experimental reference data. Most semiempirical methods are parametrized to reproduce ground-state properties, which does not guarantee that the excited-state properties are also reliable. The INDO/S (intermediate neglect of differential overlap for spectroscopy) method is well-known to provide accurate results for vertical excitation energies.<sup>20</sup> Therefore, this method is still frequently used for the calculation of optical absorption spectra for large organic molecules and transition-metal compounds.<sup>21</sup> Due to the well-balanced parameters, as developed for the INDO/S and the reparameterized INDO/S2<sup>22</sup> methods, the calculated excitation energies are in good agreement with experimental results. Although these methods seem to be superior over standard quantum-chemical methods, they have certain limitations. INDO/S makes use of the configuration interaction singles (CIS) method<sup>23</sup> for the treatment of excited states. The lack of higher excitations restricts the INDO/S-CIS method to states which are dominated by single excitations. This is problematic if a

Received: June 22, 2011

Published: September 21, 2011

mixing of double excitations is dominating the electronic transition. Due to the use of a minimal basis set, Rydberg states cannot be treated. The main reason why INDO/S is not used in photochemistry, however, is the design of the method. INDO/S—CI targets spectroscopy but is not parametrized for ground-state properties, and potential energy surfaces (PES) may not be reliable. Therefore, the application to the calculation of fluorescence spectra and photoreactions is limited. This has led to some more elaborate concepts, where higher excitations are included. It has been recently shown by Silva-Junior and Thiel that semiempirical methods, especially the OMx<sup>24–26</sup> CISDTQ and MRCISD<sup>27</sup> approaches, give promising results for the calculation of excited states.<sup>17</sup> Accurate results for the exploration of PES for both ground and excited states have been reported. Therefore, this is from our point of view the most reliable semiempirical method for the calculation of excited states at the moment. However, due to the calculation of NDDO-type two-electron—two-center integrals<sup>28,29</sup> and the inclusion of double, triple, and quadruple excitations, the computational effort is dramatically increased with respect to INDO/S, even though the implementation of the GUGA leads to a speedup in the calculation of the CI matrix elements.<sup>27</sup>

This leads to our starting point. In our present implementation, we do not go beyond the explicit calculation of singles excitations. This gives a method similar to the INDO/S case, but at the full CIS level, where all possible single excitations are taken into account. In our implementation, we take advantage of the orthogonalization correction in the MSINDO method,<sup>30</sup> which is similar to the OMx methods. In the parametrization, we include both ground and excited states properties. In order to account for cases where higher excitations play an important role, we introduced a semiempirical correction term.

In this article, we present the basic equations of our MSINDO-sCIS method and then show the results of our method for the benchmark set of Thiel. We will show that our method shows similar accuracy compared to the OMx-CISDTQ methods for most compounds.

## 2. METHOD

In this section, we describe the implementation of the CIS equations into the MSINDO method. The first part reviews the CIS theory and the implementation of our empirical correction. In the second part, we discuss the implementation of the method with an efficient matrix diagonalization, where all semiempirical considerations are introduced. In the last subsections, we will describe the basic ideas of our parametrization for ground and excited states, including charge-transfer states.

**2.1. Theory.** Within the CIS approach, the excited-state wave function is formed in terms of the Hartree—Fock orbitals. With  $N$  occupied and  $M$  virtual orbitals,  $N \times M$  determinants are created by interchanging all pairs of occupied and virtual orbitals. Linear combinations of these determinants form the CIS wave function of the excited states. For closed-shell systems, the CIS wave function can be formed in terms of a singlet or triplet configuration state function (CSF):

$$|^{1,3}\Psi_I\rangle = \sum_i^{\text{occ}} \sum_a^{\text{vir}} t_i^a |^{1,3}\Phi_a^i\rangle \quad (1)$$

The unique set of the expansion factors  $\{t_i^a\}$ —the CIS amplitudes—defines the wave function of the given excited state  $I$ . Variational solution of the Schrödinger equation with this wave

function leads to an eigenvalue problem, where the CIS matrix  $H$  has to be diagonalized:

$$^M H_{ia,jb} = \langle ^M \Phi_a^i | \hat{H} | ^M \Phi_b^j \rangle \quad M = 1, 3 \quad (2)$$

Since the original MSINDO parameters have been optimized for the ground-state properties,<sup>30</sup> we introduced two additional parameters for an improved parametrization of the excited states. These parameters are scaling factors for the two-electron integrals in the CIS matrices ( $c_1$  and  $c_2$  for singlets,  $c_T$  for triplets):

$$^1 H_{ia,jb} = \delta_{ij} \delta_{ab} [\varepsilon_a - \varepsilon_i - d_{ia}^{\text{corr}}] + 2c_1 (ia|jb) - c_2 (ij|ab) \quad (3)$$

Considering that higher excitations have a large influence on certain totally symmetric excited states, we improved the CIS energy using a semiempirical correction. The correction term  $d_{ia}^{\text{corr}}$  (eq 4) is an empirical correction for totally symmetric single excited states that may have a strong mixing with double excitations  $ii \rightarrow aa$  from the ground state.

$$d_{ia}^{\text{corr}} = |(aa|ia) + (ii|ia)| \quad (4)$$

It is well-known that perturbative doubles corrections to the excited CIS states—the so-called CIS(D) method<sup>31,32</sup>—give a significant improvement in the excitation energies. In principle, the ground state and all excited states should be corrected by electron correlation due to mixing with higher excitations. In CIS(D), this is approximated by second-order perturbation theory that takes into account double excitations with regard to ground and excited state determinants. The full implementation of doubles correction is an  $O(N^5)$  process, which would counteract the philosophy of semiempirical methods. While the ground state is implicitly correlated in semiempirical methods, it is not clear how much this is the case for excited states.<sup>33</sup> Only in totally symmetric excited states can double excitations of the type  $ii \rightarrow aa$  and also higher excitations based hereupon mix with the singly excited determinants, which leads to a lowering of the corresponding excitation energy. Since the corresponding coupling integrals are in general larger in absolute value than those of other excitations  $ij \rightarrow aa$ ,  $ii \rightarrow ab$ , and  $ij \rightarrow ab$ , which may belong to a different irreducible representation, a correction is more important for totally symmetric excited states. This is in line with the deviations of calculated excitation energies at the ab initio CIS(D) level compared to CC2, CCSD, and CC3. For this purpose, we performed ab initio CIS(D) calculations for *E*-butadiene as a test molecule with the optimized MSINDO structures and aug-cc-TZVP basis sets and compared them to previous results obtained at CCn level with the same basis set.<sup>34</sup> The largest deviation (0.9 eV for CIS(D)) was observed for the singlet  $A_g$  state, whereas singlet  $B_u$  and triplet  $A_g$  and  $B_u$  states coincide within 0.2–0.3 eV. In preliminary test calculations, it was found that this approach led to the best overall agreement with the TBEs. Therefore, we decided to empirically correct only these kinds of excited states. Without the correction term (eq 4) we observed in some cases, that also the ordering of the excited states was incorrect.  $d_{ia}^{\text{corr}}$  is exactly zero if orbitals  $i$  and  $a$  do not belong to the same irreducible representation. It is not necessary to make a similar correction for triplet states due to spin selection rules.

$$^3 H_{ia,jb} = \delta_{ij} \delta_{ab} [\varepsilon_a - \varepsilon_i] - c_T (ij|ab) \quad (5)$$

Solving eq 2 gives then the excitation energy of a given state (eqs 6 and 7):

$${}^1E_{\text{CIS}} - E_{\text{RHF}} = \sum_i^{\text{occ}} \sum_a^{\text{vir}} t_i^a \left\{ t_i^a (\varepsilon_a - \varepsilon_i - d_{ia}^{\text{corr}}) + \sum_j^{\text{occ}} \sum_b^{\text{vir}} t_j^b [2c_1(ia|jb) - c_2(ij|ab)] \right\} \quad (6)$$

$${}^3E_{\text{CIS}} - E_{\text{RHF}} = \sum_i^{\text{occ}} \sum_a^{\text{vir}} t_i^a \left\{ t_i^a (\varepsilon_a - \varepsilon_i) - \sum_j^{\text{occ}} \sum_b^{\text{vir}} t_j^b c_T(ij|ab) \right\} \quad (7)$$

This approach follows the original ideas of Grimme<sup>35</sup> in his DFT/SCI method, where a scaling of the integrals gave a significant improvement of the vertical excitation energies. Different from the expressions in ref 35 where only a single scale factor  $c = c_2 = c_T$  is applied and  $c_1 = 1$  is used as a constant, we treat all parameters  $c_1$ ,  $c_2$ , and  $c_T$  as adjustable parameters in order to improve the calculated excitation energies.

The empirical shift introduced for Rydberg states and core excited states<sup>35</sup> is not applied in our implementation, since our semiempirical method is not intended to be applied to these kinds of problems. The scaling in eqs 3 and 5 does not affect the Fock matrix, and therefore these parameters are decoupled from the ground state.

**2.2. Matrix Diagonalization.** For the matrix diagonalization necessary to solve the CIS equations, we use the Davidson–Liu block diagonalization method,<sup>36</sup> which has been shown to be very efficient, especially for large sparse matrices.<sup>37</sup> This iterative procedure starts with the guess vectors  $|\mathbf{b}_i\rangle$ , which are expanded to form the best approximation of the CIS wave function. All eigenvectors of the CIS matrix are expanded in an  $L$ -dimensional orthonormal subspace of the eigenvectors:

$$|\mathbf{x}^k\rangle = \sum_i^L \alpha_i^k |\mathbf{b}_i\rangle \quad (8)$$

where  $|\mathbf{x}^k\rangle$  is the exact eigenvector and  $\alpha_i^k$  represents the expansion coefficients. With these new basis vectors, a projected CIS matrix is formed:

$$\langle \mathbf{b}_i | \hat{H} | \mathbf{b}_j \rangle \quad 1 \leq i, j \leq L \quad (9)$$

The expansion coefficients  $\alpha^k$  are given by the eigenvectors of this matrix with the eigenvalues  $\rho_k$  and the eigenvectors  $|\mathbf{c}^k\rangle$ . The approximated eigenvectors  $|\mathbf{c}^k\rangle$  are corrected by  $|\delta^k\rangle$ :

$$|\mathbf{c}^k\rangle - |\delta^k\rangle = |\mathbf{x}^k\rangle \quad (10)$$

The set of  $|\delta^k\rangle$  is directly related to the residual vectors  $|\mathbf{r}^k\rangle$ :

$$(\mathbf{H} - \lambda^k) |\delta^k\rangle = -(\mathbf{H} - \lambda^k) |\mathbf{x}^k\rangle = -|\mathbf{r}^k\rangle \quad (11)$$

The residuals are calculated by connecting eqs 8 and 11. Using the definition of the  $\sigma$  (eq 9) vectors, one obtains

$$|\mathbf{r}^k\rangle = \sum_i^L \alpha_i^k (|\sigma_i\rangle - \rho_k |\mathbf{b}_i\rangle) \quad (12)$$

Since the CIS matrix is sparse and dominated by the diagonal elements, the correction vector is approximated by

$$|\delta^k\rangle \approx -(\mathbf{D} - \rho^k \mathbf{E})^{-1} |\mathbf{r}^k\rangle \quad (13)$$

where  $\mathbf{D}$  is an arbitrary diagonal matrix, which is connected to  $\mathbf{H}$ . In our implementation, we do not use the diagonal elements of  $\mathbf{H}$ . Since these are dominated by the orbital energy differences, we simply set

$$D_{ia,ia} = \varepsilon_a - \varepsilon_i \quad (14)$$

This leads to a significantly faster calculation of the diagonal matrix  $\mathbf{D}$ , which improves the overall performance of the Davidson algorithm despite an increase in the number of iterations. The obtained correction vectors  $|\delta^k\rangle$  are then normalized with respect to the existing set of expansion vectors  $|\mathbf{b}_i\rangle$  by a modified Gram–Schmidt orthogonalization.<sup>38</sup> This leads to an increased size of the expansion space in every step of the algorithm up to a preselected threshold. As a standard criterion for convergence, we have chosen  $10^{-8}$  au for the eigenvalues and  $10^{-6}$  au for the norm of the residuals. In each step, the  $\sigma$  vectors are calculated according to eq 9. With a restricted Hartree–Fock reference, the CSFs are used, and two independent sets of  $\sigma$  vectors are defined:

$${}^1\sigma_a^i = (E_{\text{RHF}} - \varepsilon_i + \varepsilon_a - d_{ia}^{\text{corr}}) t_i^a + \sum_j^{\text{occ}} \sum_b^{\text{vir}} t_j^b [2c_1(ia|jb) - c_2(ij|ab)] \quad (15)$$

$${}^3\sigma_a^i = (E_{\text{RHF}} - \varepsilon_i + \varepsilon_a) t_i^a - \sum_j^{\text{occ}} \sum_b^{\text{vir}} t_j^b c_T(ij|ab) \quad (16)$$

By defining pseudo-Fock matrices of the form

$${}^1\tilde{F}_{ia} = \sum_j^{\text{occ}} \sum_b^{\text{vir}} t_j^b [2c_1(ia|jb) - c_2(ij|ab)] \quad (17)$$

$${}^3\tilde{F}_{ia} = -\sum_j^{\text{occ}} \sum_b^{\text{vir}} t_j^b c_T(ij|ab) \quad (18)$$

the  $\sigma$  vectors are much more efficiently calculated in the atomic orbital (AO) basis. This leads to the AO transformed pseudo-Fock matrix, which is given in the general case as

$${}^1\tilde{F}_{\mu\nu} = \sum_{\lambda\sigma} T_{\lambda\sigma} [2c_1(\mu\nu|\lambda\sigma) - c_2(\mu\lambda|\nu\sigma)] \quad (19)$$

$${}^3\tilde{F}_{\mu\nu} = \sum_{\lambda\sigma} T_{\lambda\sigma} c_T(\mu\lambda|\nu\sigma) \quad (20)$$

where  $T_{\lambda\sigma}$  are AO-transformed CIS amplitudes. Within the INDO approximation,<sup>39</sup> three kinds of elements of the pseudo-Fock matrix have to be distinguished, the diagonal elements, the intra-atomic off-diagonal blocks, and the two-center terms:

$${}^1F_{\mu\nu} = \begin{cases} \sum_{\lambda\sigma \in A} T_{\lambda\sigma} [2c_1(\mu\mu|\lambda\sigma) - c_2(\mu\lambda|\mu\sigma)] & \text{for } \mu = \nu \in A \\ + \sum_{\lambda \in B \neq A} T_{\lambda\lambda} c_1(\mu\mu|\lambda\lambda) & \\ \sum_{\lambda\sigma \in A} T_{\lambda\sigma} [2c_1(\mu\nu|\lambda\sigma) - c_2(\mu\lambda|\nu\sigma)] & \text{for } \mu, \nu \in A \\ -T_{\mu\nu} c_2(\mu\mu|\nu\nu) & \text{for } \mu \in A, \nu \in B \end{cases} \quad (21)$$



$${}^3F_{\mu\nu} = \begin{cases} -\sum_{\lambda\sigma\in A} T_{\lambda\sigma}c_T(\mu\lambda|\mu\sigma) & \text{for } \mu = \nu \\ -\sum_{\lambda\sigma\in A} T_{\lambda\sigma}c_T(\mu\lambda|\nu\sigma) & \text{for } \mu, \nu \in A \\ -T_{\mu\nu}c_T(\mu\mu|\nu\nu) & \text{for } \mu \in A, \nu \in B \end{cases} \quad (22)$$

Necessary matrix operations of the Davidson–Liu approach are efficiently performed by BLAS<sup>40,41</sup> routines in our implementation.

**2.3. Parameterization.** In order to obtain accurate values for the excitation energies, the whole MSINDO parameter set, which has repeatedly demonstrated its reliability,<sup>42–45</sup> has been reoptimized. This affects not only the excited states but also the ground-state properties. After augmenting the original MSINDO reference set,<sup>46</sup> the ground-state properties and some additional vertical excitation energies were optimized with respect to accurate reference data by varying the parameters in a nonlinear minimization algorithm, where the sums of least-squares errors are minimized.<sup>47</sup> The one-center two-electron integrals are evaluated in terms of the Slater–Condon parameters  $F^n$  and  $G^n$  ( $n = 0–3$ ). Following the ideas of Zerner and Ridley,<sup>20</sup> these are treated as adjustable parameters in the present MSINDO-sCIS implementation at variance with the agreed MSINDO method<sup>46</sup> (see Table 6, Supporting Information). This was necessary, since Slater–Condon factors play an important role in spectroscopy.<sup>48</sup> In the original MSINDO version,<sup>46</sup> the Slater–Condon factors were calculated analytically with a special set of orbital exponents (for a better comparison, their values are given in Table 6 in the Supporting Information). In the present approach, it was necessary to treat the  $F^n$  and  $G^n$  as independent empirical parameters in order to obtain reasonable agreement with experimental results for both ground and excited state properties. We had, to some extent, to give up the concept of a physical interpretation of the  $F$  and  $G$  factors. The  $F_{ss}^0$  values are expected to increase within the second row, but the optimized value for C is larger than that for N, O, and F. The obtained values of the Slater–Condon factors  $F_{ss}^0$ ,  $F_{sp}^0$ ,  $F_{pp}^0$ ,  $F_{pp}^2$ , and  $G_{sp}^1$  for the second-row elements and additionally  $F_{sd}^0$ ,  $F_{pd}^0$ ,  $G_{pd}^1$ ,  $G_{sd}^2$ , and  $G_{pd}^3$  for the third-row element sulfur are presented in Table 6 in the Supporting Information. Furthermore, we allowed the correction factors for the orthogonalization  $f^{B,orth}$  (see ref 46) to change. In the original version, these were formally preparameterized and then treated as constants, depending on the angular momentum of the orbital (1, 0.75, and 0.5). Additionally, the ionization potentials  $I$  and the shielding parameters, which are introduced by a distance-dependent exponential function with the exponent  $\kappa_{EP}$ , were also included in the parametrization. The last parameters, which were reoptimized, are the resonance integral parameters  $K_i$ , which depend on the local symmetry ( $s\sigma$ ,  $p\sigma$ ,  $p\pi$ ,  $d\sigma$ ,  $d\pi$ , or  $d\delta$ ) in a diatomic coordinate system. The complete new set of these parameters compared to the old parameters for common elements in organic molecules can be found in Table 7 in the Supporting Information. The excited-state parameters, which are introduced in the calculation of the excited states (eqs 3 and 5), have the following values:

$$c_i = \begin{cases} 0.9225 & \text{for } i = 1 \\ 0.9993 & \text{for } i = 2 \\ 0.8944 & \text{for } i = T \end{cases} \quad (23)$$

In the original reference,<sup>35</sup> the parameters  $c_1$ ,  $c_2$ , and  $c_T$  had the values 1.0, 0.317, and 0.317. However, these parameters have been obtained for density functional theory, which is completely

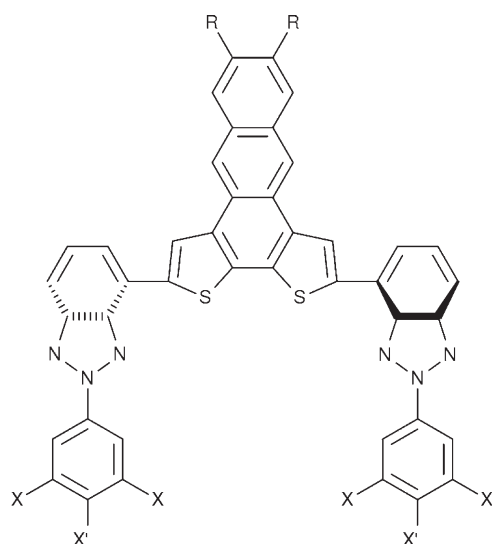
different from our present approach. Since both  $c_1$  and  $c_2$  are close to 1, the artificial problem of self-interaction (incomplete cancellation of Coulomb and exchange integrals) is nearly avoided for singlet states. These excited state parameters are in the spirit of DFT, where only a few global parameters are used to obtain accurate results, rather than in the spirit of semiempirical methods. But our statistical evaluation shows (see section 5.2) that it is not necessary to introduce bond-specific parameters for excited states. In summary, we have taken four steps in our parametrization. The first one is the inclusion of Slater–Condon factors in the parametrization. The second is the introduction of three global parameters for the excited states, which scale the Coulomb and exchange integrals. Third, an element-specific scaling of the orthogonalization correction is introduced. And the last step is the empirical correction  $d^{corr}$  (eq 4), which improves the excitation energies of totally symmetric excitations. Those ideas eventually define a new method, which is intended to be efficient and accurate. In a first step, only selected elements of the first two periods H, C, N, O, F, and, additionally, S have been reparameterized. Further work, the reparameterization of the remaining second and third row elements including the third-row transition metals, is in progress.

### 3. GENERAL CONSIDERATIONS

Ground-state geometry optimizations and vertical excitation energy calculations are carried out with the new version MSINDO-sCIS. Different from Thiel's statistical evaluation,<sup>17</sup> we did not take the geometries from MP2/6-31G\* calculations, because we want to demonstrate that the MSINDO-sCIS method is capable of providing reliable results for both kinds of properties. In preliminary tests, vertical excitation energies with MP2 geometries were calculated for selected molecules of the test set. The small change in the geometries did not affect the vertical excitation energies for both singlet and triplet states significantly (changes were below 0.1 eV). Since semiempirical methods employ a minimal valence basis set, Rydberg states or states with a valence–Rydberg mixing cannot be described properly; accordingly, Thiel's benchmark set<sup>1</sup> contains only valence excited states. Within the MSINDO-sCIS approach, all possible single excitations of the valence orbitals are taken into account in the active space. For the excited-state assignment, the electronically excited states are first classified according to their point group symmetry. Afterward, the most significant orbitals of the excitation are used for a further inspection of the kind of transition. With this information, it is possible to assign the states and distinguish between  $\pi-\pi^*$ ,  $n-\pi^*$ , and  $\sigma-\pi^*$  transitions.

### 4. COMPUTATIONAL PERFORMANCE

As a benchmark for the computational performance of our approach, we have calculated the first eight excited states of selected molecules of the class of bis(phenyl-benzotriazole) dithiophenes (see Figure 1) and compared the MSINDO-sCIS CPU timings to TD-DFT calculations (see Table 1). These molecules are expected to have charge-transfer states in the range of visible light, so that they should work in a solar cell device.<sup>49</sup> The TD-DFT calculations (B3LYP/TZV(P)) were carried out using the ORCA 2.7.0 program package.<sup>50</sup> The computational performance for both methods strongly depends on the number of Davidson cycles. Comparing the timings of MSINDO-sCIS and TD-B3LYP (shown in Table 1) gives a ratio of up to 1:7500 on a single Quad-Core AMD 2378 Opteron 2.4 GHz processor.



**Figure 1.** Class of molecules used for performance test in comparison with TDDFT (B3LYP/TZVP).

**Table 1.** MSINDO-sCIS and TD-DFT Calculation Timings for the Different Types of Bis(phenyl-benzotriazole) dithiophenes (see Figure 1)<sup>a</sup>

molecule	MSINDO-sCIS	B3LYP/TZVP
X' = H, X = NO <sub>2</sub> , R = OC(C <sub>2</sub> H <sub>5</sub> )C <sub>5</sub> H <sub>11</sub>	315s (N = 20)	11952s (N = 05)
X' = H, X = NO <sub>2</sub> , R = isoprop	195s (N = 20)	74986s (N = 05)
X' = H, X = NO <sub>2</sub> , R = CH <sub>3</sub>	150s (N = 19)	61303s (N = 06)
X' = H, X = F, R = OC(C <sub>2</sub> H <sub>5</sub> )C <sub>5</sub> H <sub>11</sub>	238s (N = 20)	259435s (N = 06)
X' = F, X = H, R = OC(C <sub>2</sub> H <sub>5</sub> )C <sub>5</sub> H <sub>11</sub>	190s (N = 20)	244928s (N = 12)
X' = X = H, R = NO <sub>2</sub>	89s (N = 17)	665596s (N = 52)
X' = X = H, R = phenyl	155s (N = 20)	230268s (N = 10)
X' = X = H, R = COOH	104s (N = 19)	166497s (N = 14)

<sup>a</sup>The values in parentheses are the number N of Davidson iterations.

This superior speed of MSINDO-sCIS allows applications of the excited-state description of really large molecules containing more than 1000 atoms, which are not feasible with TD-DFT or even more accurate methods.

## 5. RESULTS

In this section, we will compare ground- and excited-state properties calculated with the new method MSINDO-sCIS with experimental and other theoretical approaches. We will start with the ground state properties, where we will show that the new parameter set is as reliable as the original set developed by Ahlswede and Jug.<sup>46</sup> This is followed by a detailed description of the vertical excitation energies, where all excitations from the benchmark set of Silva-Junior et al.<sup>1,11</sup> were taken into account. The MSINDO results are then compared to the theoretical best estimates,<sup>1</sup> TD-B3LYP,<sup>16</sup> INDO and OMx results.<sup>17</sup> It should be mentioned that the comparison with the OM3-CISDTQ results

**Table 2.** MSINDO Mean Absolute Errors for Ground- and Excited-State Properties<sup>a</sup>

property	unit	no. references	parameter set	
			old <sup>b</sup>	new <sup>c</sup>
$\Delta_f H$	kcal/mol	89	5.25	4.97
R	Å	210	0.013	0.016
$\varphi$	deg	94	1.69	1.76
IP	eV	83	0.43	0.53
$\mu$	D	49	0.34	0.35
$\Delta E_{0-n}$	eV	36		0.34

<sup>a</sup>Heats of formation  $\Delta_f H$ , bond lengths R, bond angles  $\varphi$ , ionization potentials IP, dipole moments  $\mu$ , and vertical excitation energies  $\Delta E_{0-n}$ . In the old parameterization, no excited states were included. <sup>b</sup>Reference 46. <sup>c</sup>Present work.

is just for benchmarking purposes, since the common used method for excited states is OM2/MRCISD, which has the same accuracy.<sup>17</sup>

**5.1. Ground State Properties.** To avoid the well-known problems of INDO/S for ground-state geometries, while being parametrized for excited states, we included ground-state properties in our parametrization. A comparison of the results obtained with new MSINDO-sCIS parameters with those of the previous version is given in Table 2. Distances R and angles  $\varphi$  are only slightly less accurate than in the 1999 version of MSINDO.<sup>46,42</sup> On the other hand, it can be seen from Table 2 that the new set of parameters even slightly improves the heats of formation, while the ionization potentials are slightly less accurate than before. But in the overall view of these results, it is obvious that we now have reliable parameters for both ground and excited states.

**5.2. Accuracy of Vertical Excitation Energies.** **5.2.1. Vertical Singlet Excitations.** The benchmark of the calculated singlet excitation energies with respect to the TBES<sup>1</sup> is presented in Table 3. We will discuss the results of all groups of compounds in the following paragraphs.

**Unsaturated Aliphatic Hydrocarbons. Ethene.** The energy of the singlet  $\pi-\pi^*$  state of ethene (TBE-2 is 7.80 eV) is underestimated by 0.31 eV with MSINDO-sCIS. This error is larger than that of the OMx-CISDTQ suite of methods (0.02–0.05 eV) but much less compared to standard MNDO methods, where the energy is underestimated by more than 1 eV, and also compared to the INDO/S method, which overestimates the energy by more than 0.5 eV.

**E-Butadiene, E-Hexatriene, and E-Octatetraene.** The most interesting point in this series of C<sub>2n</sub>H<sub>2n+2</sub> polyenes is the gap between the bright 1B<sub>u</sub> and the 2A<sub>g</sub> state. In the benchmark set, this gap is reduced with increasing n until the ordering is reversed. Due to the large contributions from double excitations (33%)<sup>51</sup> and higher excited configurations (20% within CISDTQ)<sup>17</sup> for the A<sub>g</sub> state, this problem cannot be solved by parametrization. The MSINDO-sCIS errors for the 2A<sub>g</sub> states of E-butadiene, E-hexatriene, and E-octatetraene are 0.29 eV, 1.42 eV, and 1.31 eV, respectively. These relatively large errors indicate that our approach—including the empirical correction term (eq 4)—is not accurate for excited states with significant double excitation contributions. Even compared to the INDO/S method, where the errors are 0.32 eV, 0.81 eV, and 0.76 eV for the 2A<sub>g</sub> states, MSINDO is even slightly inferior.

**Cyclopropene.** The excited states of strained ring systems such as cyclopropene are problematic for standard semiempirical

**Table 3. Vertical Excitation Energies  $\Delta E$  [eV] for Singlet States: MSINDO Results Compared with the Theoretical Best Estimates TBE-2 from Ref 1**

molecule	state	type	TBE-2	MSINDO
ethene	$1^1B_{1u}$	$\pi-\pi^*$	7.80	7.49
E-butadiene	$1^1B_u$	$\pi-\pi^*$	6.18	6.09
	$2^1A_g$	$\pi-\pi^*$	6.55	6.84
E-hexatriene	$1^1B_u$	$\pi-\pi^*$	5.10	5.33
	$2^1A_g$	$\pi-\pi^*$	5.09	6.51
E-octatetraene	$1^1B_u$	$\pi-\pi^*$	4.66	4.88
	$2^1A_g$	$\pi-\pi^*$	4.47	5.78
cyclopropene	$1^1B_1$	$\pi-\pi^*$	6.67	6.07
	$1^1B_2$	$\pi-\pi^*$	6.68	5.97
cyclopentadiene	$1^1B_2$	$\pi-\pi^*$	5.55	5.33
	$2^1A_1$	$\pi-\pi^*$	6.28	6.57
norbornadiene	$1^1A_2$	$\pi-\pi^*$	5.37	5.42
	$1^1B_2$	$\pi-\pi^*$	6.21	5.69
benzene	$1^1B_{2u}$	$\pi-\pi^*$	5.08	5.34
	$1^1B_{1u}$	$\pi-\pi^*$	6.54	5.93
	$1^1E_{1u}$	$\pi-\pi^*$	7.13	7.02
	$1^1E_{2g}$	$\pi-\pi^*$	8.15	8.11
naphthalene	$1^1B_{3u}$	$\pi-\pi^*$	4.25	4.59
	$1^1B_{2u}$	$\pi-\pi^*$	4.82	5.40
	$2^1A_g$	$\pi-\pi^*$	5.90	6.03
	$1^1B_{1g}$	$\pi-\pi^*$	5.75	6.26
	$2^1B_{3u}$	$\pi-\pi^*$	6.11	6.42
	$2^1B_{2u}$	$\pi-\pi^*$	6.36	6.26
	$2^1B_{1g}$	$\pi-\pi^*$	6.46	7.73
	$3^1A_g$	$\pi-\pi^*$	6.49	6.77
furan	$1^1B_2$	$\pi-\pi^*$	6.32	5.59
	$2^1A_1$	$\pi-\pi^*$	6.57	6.27
	$3^1A_1$	$\pi-\pi^*$	8.13	6.63
pyrrole	$2^1A_1$	$\pi-\pi^*$	6.37	5.92
	$1^1B_2$	$\pi-\pi^*$	6.57	5.69
	$3^1A_1$	$\pi-\pi^*$	7.91	6.45
imidazole	$2^1A'$	$\pi-\pi^*$	6.25	5.45
	$1^1A''$	$n-\pi^*$	6.65	6.29
	$3^1A'$	$\pi-\pi^*$	6.73	6.13
pyridine	$1^1B_2$	$\pi-\pi^*$	4.85	5.89
	$1^1B_1$	$n-\pi^*$	4.59	4.66
	$1^1A_2$	$n-\pi^*$	5.11	5.29
	$2^1A_1$	$\pi-\pi^*$	6.26	5.54
	$2^1B_2$	$\pi-\pi^*$	7.27	6.99
	$3^1A_1$	$\pi-\pi^*$	7.18	7.10
pyrazine	$1^1B_{3u}$	$n-\pi^*$	4.13	3.44
	$1^1A_u$	$n-\pi^*$	4.98	4.66
	$1^1B_{2u}$	$\pi-\pi^*$	4.97	5.45
	$1^1B_{2g}$	$n-\pi^*$	5.65	5.46
	$1^1B_{1g}$	$n-\pi^*$	6.69	7.05
	$1^1B_{1u}$	$\pi-\pi^*$	6.83	5.58
	$2^1B_{2u}$	$\pi-\pi^*$	7.81	8.39
	$2^1B_{1u}$	$\pi-\pi^*$	7.86	5.98
pyrimidine	$1^1B_1$	$n-\pi^*$	4.43	4.41
	$1^1A_2$	$n-\pi^*$	4.85	4.74
	$1^1B_2$	$\pi-\pi^*$	5.34	5.95
	$2^1A_1$	$\pi-\pi^*$	6.82	6.03

**Table 3. Continued**

molecule	state	type	TBE-2	MSINDO
pyridazine	$1^1B_1$	$n-\pi^*$	3.85	3.85
	$1^1A_2$	$n-\pi^*$	4.44	4.38
	$2^1A_1$	$\pi-\pi^*$	5.20	5.07
	$2^1A_2$	$n-\pi^*$	5.66	5.11
s-triazine	$1^1A_1''$	$n-\pi^*$	4.70	4.79
	$1^1A_2''$	$n-\pi^*$	4.71	5.56
	$1^1E''$	$n-\pi^*$	4.75	4.93
	$1^1A_2'$	$\pi-\pi^*$	5.71	7.07
s-tetrazine	$1^1B_{3u}$	$n-\pi^*$	2.46	2.65
	$1^1A_u$	$n-\pi^*$	3.78	3.50
	$1^1B_{1g}$	$n-\pi^*$	4.87	5.40
	$1^1B_{2u}$	$\pi-\pi^*$	5.08	5.21
	$1^1B_{2g}$	$n-\pi^*$	5.28	5.26
	$2^1A_u$	$n-\pi^*$	5.39	5.19
formaldehyde	$1^1A_2$	$n-\pi^*$	3.88	3.74
	$1^1B_1$	$\sigma-\pi^*$	9.04	9.38
	$2^1A_1$	$\pi-\pi^*$	9.29	8.96
acetone	$1^1A_2$	$n-\pi^*$	4.38	4.37
	$1^1B_1$	$\sigma-\pi^*$	9.04	8.42
	$2^1A_1$	$\pi-\pi^*$	8.90	6.71
p-benzoquinone	$1^1B_{1g}$	$n-\pi^*$	2.74	2.62
	$1^1A_u$	$n-\pi^*$	2.86	3.17
	$1^1B_{3g}$	$\pi-\pi^*$	4.44	5.02
	$1^1B_{1u}$	$\pi-\pi^*$	5.47	5.92
	$1^1B_{3u}$	$n-\pi^*$	5.55	5.47
	$2^1B_{3g}$	$\pi-\pi^*$	7.16	6.20
formamide	$1^1A''$	$n-\pi^*$	5.55	5.40
	$2^1A'$	$\pi-\pi^*$	7.35	7.23
acetamide	$1^1A''$	$n-\pi^*$	5.62	5.59
	$2^1A'$	$\pi-\pi^*$	7.14	7.47
propanamide	$1^1A''$	$n-\pi^*$	5.65	5.36
	$2^1A'$	$\pi-\pi^*$	7.09	7.30
cytosine	$2^1A'$	$\pi-\pi^*$	4.66	4.50
	$1^1A''$	$n-\pi^*$	4.87	4.98
	$2^1A''$	$n-\pi^*$	5.26	5.18
	$3^1A'$	$\pi-\pi^*$	5.62	5.73
thymine	$1^1A''$	$n-\pi^*$	4.82	4.40
	$2^1A'$	$\pi-\pi^*$	5.20	4.69
	$3^1A'$	$\pi-\pi^*$	6.27	5.44
	$2^1A''$	$n-\pi^*$	6.16	5.53
	$4^1A'$	$\pi-\pi^*$	6.53	6.70
uracil	$1^1A''$	$n-\pi^*$	5.00	4.88
	$2^1A'$	$\pi-\pi^*$	5.25	5.01
	$3^1A'$	$\pi-\pi^*$	6.26	5.80
	$2^1A''$	$n-\pi^*$	6.10	5.84
	$4^1A'$	$\pi-\pi^*$	6.70	6.08
	$3^1A''$	$n-\pi^*$	6.56	6.21
adenine	$1^1A''$	$n-\pi^*$	5.12	5.12
	$2^1A'$	$\pi-\pi^*$	5.25	5.32
	$3^1A'$	$\pi-\pi^*$	5.25	5.60
	$2^1A''$	$n-\pi^*$	5.75	5.82

methods. Here, MSINDO-sCIS provides the best results within the considered semiempirical methods. While the OMx methods



still underestimate the  $^1B_1$  state by 0.43–0.83 eV, the MSINDO-sCIS error is not larger than 0.6 eV. For the problematic  $^1B_2$  state (TBE: 6.68 eV), the MSINDO-sCIS error of –0.61 eV is still high. But still, MSINDO-sCIS performs better than standard MNDO methods.<sup>17</sup> It has been pointed out by Thiel and Silva-Junior<sup>17</sup> that this state has significant contributions from diffuse orbitals<sup>52</sup> and therefore may be problematic for methods using minimal basis sets.

**Cyclopentadiene.** For the first two excited singlet states of cyclopentadiene, MSINDO-sCIS slightly underestimates the  $^1B_2$  state (TBE 5.55 eV) by 0.22 eV and overestimates the  $^1A_1$  (TBE 6.28 eV) state by 0.29 eV. This is a slightly better result than that obtained by OMx, where the error is between 0.41 and 0.79 eV.<sup>17</sup>

**Norbornadiene.** Comparing the results for norbornadiene shows that MSINDO reproduces the first excited  $A_2$  state well with an error of 0.05 eV. For the second excitation with  $B_2$  symmetry, MSINDO-sCIS underestimates the TBE by 0.52 eV. The absolute error is comparable to the OMx methods that overestimate the excitation energy for both states by more than 0.5 eV. It can be seen from the INDO/S results<sup>17</sup> that the underestimation of the excitation energies in this molecules seems to be a typical INDO problem. But the underestimation in the case of MSINDO is less than for INDO/S ( $A_2$  error, –0.87 eV;  $B_2$  error, –0.67 eV<sup>17</sup>).

**Aromatic Hydrocarbons and Heterocycles. Benzene.** For benzene, four excited singlet states in the range of 5.0–8.2 eV are found in the benchmark set.<sup>1</sup> For three of the states, MSINDO-sCIS calculations show good agreement with TBE-2. The first excited singlet state  $B_{2u}$  as well as the two higher excited states ( $E_{1u}$  and  $E_{2g}$ ) are well reproduced by MSINDO-sCIS with errors of 0.26 eV, 0.11 eV, and 0.04 eV. Only the  $B_{1u}$  state differs more from the TBE-2 by –0.49 eV. The severe problem of the  $\sigma-\sigma^*$  contamination in the INDO/S method<sup>17</sup> is apparently diminished with our ansatz. The OMx methods show similar performance with errors between 0.1 and 0.67 eV. Therefore, it can be concluded that the  $B_{1u}$  state is in general problematic to describe within semiempirical methods.

**Naphthalene.** The lower excited states of naphthalene are reasonably reproduced, while higher states are less accurate. The doubles mixing for the  $A_g$  states, which has been discussed before,<sup>17</sup> is well described with our empirical correction method. Comparing the errors of the  $A_g$  states of MSINDO-sCIS (0.13–0.28 eV) with the OMx errors (0.55–0.74 eV) shows that our simple correction in connection with a good parametrization can even outperform the explicit calculation of higher excited determinants (CISDTQ). But the main difference from the OMx method is that MSINDO-sCIS overestimates the excitation energy, while OMx excitation energies are too low. The MSINDO-sCIS errors are unfortunately much larger for higher excited states of a given irreducible representation. For example, the difference for the  $2B_{1g}$  state is 1.27 eV, which is considerably larger than for OMx (0.15–0.22 eV) and also INDO/S (0.07 eV). This seems to be a general trend.

**Furan.** The singlet excitation energies of furan obtained with MSINDO-sCIS show the same errors as the other semiempirical methods. The error range of 0.3–1.5 eV is similar to the OMx methods. It is an improvement over the INDO/S and INDO/S2 methods, where errors up to 2.0 eV are obtained.

**Pyrrrole.** For pyrrrole, which is isoelectronic with furan, the results for the excitation energies are similar. The error range is somewhat larger (0.65–1.46 eV), and again all energies are underestimated. This is similar to other semiempirical methods.<sup>17</sup>

**Imidazole.** The imidazole spectrum consists of two  $\pi-\pi^*$  transitions and one  $n-\pi^*$  transition. The  $n-\pi^*$  transition is well reproduced by MSINDO-sCIS with an error of 0.26 eV, but the  $\pi-\pi^*$  transitions are underestimated by approximately 0.8 eV. This leads to a wrong ordering of the excitations, but the MSINDO-sCIS errors are still smaller than those of other semiempirical methods.

**Pyridine.** Due to the break of symmetry by substituting one carbon atom by nitrogen in benzene, the four  $\pi-\pi^*$  transitions split into six transitions with  $A_1$  and  $B_2$  symmetry in pyridine. The lowest four of them are included in the benchmark set. Comparing the TBEs with the MSINDO-sCIS results, a large scattering of the excitation energies is observed. The errors are between 0.05 and 1.04 eV. The scattering is higher than with OMx or INDO/S, and even the  $1B_2 - 2A_1$  ordering is wrong. But here, the higher excitations are in better agreement with the TBE-2 benchmark results. In particular, the two  $n-\pi^*$  transitions are described rather well. Both values are overestimated (0.07 and 0.18 eV) but still close to the benchmark results.

**Pyrazine, Pyrimidine, and Pyridazine.** The performance of MSINDO-sCIS with the azabenzenes with two nitrogen atoms is similar to that for pyridine. The exception is pyridazine, where the excitation energies are underestimated by less than 0.2 eV. The pyrazine results, where eight reference energies are in the benchmark set, are very inhomogeneous. Except for the  $1^1B_{2g}$  state, all energies are underestimated by more than 0.5 eV. The state ordering is in general parallel to the ab initio data, except for the change of the  $1^1B_{1g}$  and  $1^1B_{1u}$  states. The same effect can be observed in the OMx benchmark calculations of Thiel and Silva-Junior.<sup>17</sup>

**s-Triazine.** The three lowest excitations for s-triazine are nearly degenerate  $n-\pi^*$  transitions (4.70–4.75 eV). Here, MSINDO-sCIS overestimates the second excitation by 0.85 eV. Since the other two states are overestimated by only 0.09–0.18 eV, this also results in a loss of the near-degeneracy, similar to the results of AM1 and PM3.<sup>17</sup> The  $\pi-\pi^*$  excitation energy is overestimated by more than 1.0 eV. This is the opposite of all other semiempirical methods, where all energies are strongly underestimated, except for OM3, where the error is only –0.02 eV.

**s-Tetrazine.** The excited states of s-tetrazine are well described by MSINDO-sCIS, except for the third  $n-\pi^*$  excitation (error: 0.53 eV). The errors for the other states are all below 0.3 eV, which is superior to all other semiempirical methods, where the errors are typically over 0.5 eV. Even the OMx methods have problems with the  $n-\pi^*$  excitation, where the maximum error is over 1.0 eV.

**Aldehydes, Ketones, and Amides. Formaldehyde and Acetone.** Formaldehyde and acetone have nearly the same electronic spectrum. The results obtained by MSINDO-sCIS are in good agreement with those of TBEs-2 (errors of 0.01–0.34 eV), except for the  $A_1$  states (errors of 0.37–2.19 eV), which is problematic, since the energy is strongly underestimated in the case of acetone. Even though this is the maximum error in the complete benchmark set, this is an improvement compared with the INDO/S and also the INDO/S2 results, where the energy of these states is overestimated by 3–4 eV. Here, the OMx methods perform much better with errors between 0.39 and 0.82 eV.

**p-Benzoquinone.** The two lowest singlet excited states are dark states of the  $n-\pi^*$  type. MSINDO-sCIS reproduces the values of these states well. The errors are below 0.25 eV. The higher lying  $n-\pi^*$  state with  $B_{3u}$  symmetry is also correctly described with an error of 0.08 eV. The  $\pi-\pi^*$  states show typical

**Table 4. Vertical Excitation Energies  $\Delta E$  [eV] for Triplet States: MSINDO Results Compared with the Theoretical Best Estimates TBE-2 from Ref 1**

molecule	state	type	TBE-2	MSINDO
ethene	$1^3B_{1u}$	$\pi-\pi^*$	4.50	3.90
E-butadiene	$1^3B_u$	$\pi-\pi^*$	3.20	3.04
	$1^3A_g$	$\pi-\pi^*$	5.08	4.31
E-hexatriene	$1^3B_u$	$\pi-\pi^*$	2.40	2.62
	$1^3A_g$	$\pi-\pi^*$	4.15	3.69
E-octatetraene	$1^3B_u$	$\pi-\pi^*$	2.20	2.40
	$1^3A_g$	$\pi-\pi^*$	3.55	3.23
cyclopropene	$1^3B_2$	$\pi-\pi^*$	4.28	3.48
	$1^3B_1$	$\sigma-\pi^*$	6.40	5.93
cyclopentadiene	$1^3B_2$	$\pi-\pi^*$	3.26	2.82
	$1^3A_1$	$\pi-\pi^*$	5.09	4.09
norbornadiene	$1^3A_2$	$\pi-\pi^*$	3.68	3.17
	$1^3B_2$	$\pi-\pi^*$	4.16	3.14
benzene	$1^3B_{1u}$	$\pi-\pi^*$	4.15	3.30
	$1^3E_{1u}$	$\pi-\pi^*$	4.86	5.36
	$1^3B_{2u}$	$\pi-\pi^*$	5.88	6.32
	$1^3E_{2g}$	$\pi-\pi^*$	7.51	7.26
naphthalene	$1^3B_{2u}$	$\pi-\pi^*$	3.09	2.82
	$1^3B_{3u}$	$\pi-\pi^*$	4.09	4.56
	$1^3B_{1g}$	$\pi-\pi^*$	4.42	3.93
	$2^3B_{2u}$	$\pi-\pi^*$	4.56	4.90
	$2^3B_{3u}$	$\pi-\pi^*$	4.92	5.75
	$1^3A_g$	$\pi-\pi^*$	5.42	5.10
	$2^3B_{1g}$	$\pi-\pi^*$	6.12	6.89
	$2^3A_g$	$\pi-\pi^*$	6.17	7.15
	$3^3A_g$	$\pi-\pi^*$	6.65	7.40
	$3^3B_{1g}$	$\pi-\pi^*$	6.67	7.16
furan	$1^3B_2$	$\pi-\pi^*$	4.11	3.18
	$1^3A_1$	$\pi-\pi^*$	5.43	4.26
pyrrole	$1^3B_2$	$\pi-\pi^*$	4.44	3.30
	$1^3A_1$	$\pi-\pi^*$	5.42	6.13
imidazole	$1^3A'$	$\pi-\pi^*$	4.65	3.59
	$2^3A'$	$\pi-\pi^*$	5.64	4.94
	$1^3A''$	$n-\pi^*$	6.25	5.81
	$3^3A'$	$\pi-\pi^*$	6.38	6.15
pyridine	$1^3A_1$	$\pi-\pi^*$	4.06	3.71
	$1^3B_1$	$n-\pi^*$	4.25	4.33
	$1^3B_2$	$\pi-\pi^*$	4.64	5.16
	$2^3A_1$	$\pi-\pi^*$	4.91	5.62
	$1^3A_2$	$n-\pi^*$	5.28	5.40
	$2^3B_2$	$\pi-\pi^*$	6.08	6.35
s-tetrazine	$1^3B_{3u}$	$n-\pi^*$	1.87	2.05
	$1^3A_u$	$n-\pi^*$	3.49	3.31
	$1^3B_{1g}$	$n-\pi^*$	4.18	4.48
	$1^3B_{1u}$	$\pi-\pi^*$	4.36	3.86
	$1^3B_{2u}$	$\pi-\pi^*$	4.39	4.45
	$1^3B_{2g}$	$n-\pi^*$	4.89	4.77
	$2^3A_u$	$n-\pi^*$	4.96	4.92
	$2^3B_{1u}$	$\pi-\pi^*$	5.32	6.49
formaldehyde	$1^3A_2$	$\pi-\pi^*$	3.50	3.81
	$1^3A_1$	$\pi-\pi^*$	5.87	6.57
acetone	$1^3A_2$	$n-\pi^*$	4.05	4.51

**Table 4. Continued**

molecule	state	type	TBE-2	MSINDO
p-benzoquinone	$1^3A_1$	$\pi-\pi^*$	6.07	6.64
	$1^3B_{1g}$	$n-\pi^*$	2.50	2.70
	$1^3A_u$	$n-\pi^*$	2.61	3.22
	$1^3B_{1u}$	$\pi-\pi^*$	3.02	3.30
formamide	$1^3B_{3g}$	$\pi-\pi^*$	3.37	3.44
	$1^3A''$	$n-\pi^*$	5.28	5.60
acetamide	$1^3A'$	$\pi-\pi^*$	5.69	6.40
	$1^3A''$	$n-\pi^*$	5.35	5.80
propanamide	$1^3A'$	$\pi-\pi^*$	5.71	6.63
	$1^3A''$	$n-\pi^*$	5.38	5.54
	$1^3A'$	$\pi-\pi^*$	6.08	6.43

errors of round about 0.5 eV, which is similar to all other semiempirical methods.

**Formamide, Acetamide, and Propanamide.** The series of the three smallest amides—formamide, acetamide, and propanamide—is included in the benchmark set. All of these molecules have a similar spectrum, with a low lying  $n-\pi^*$  transition as the first excited singlet state. The errors of MSINDO-sCIS are between 0.03 and 0.29 eV, different from all other semiempirical methods, which strongly underestimate these states (errors of 0.4–1.1 eV). The errors for the  $\pi-\pi^*$  transitions (0.12–0.34 eV) are slightly higher than for the  $n-\pi^*$  state but still acceptable.

**Nucleobases.** *Cytosine, Thymine, Adenine, and Uracil.* The four nucleobases cytosine, thymine, adenine, and uracil play an important role in biochemistry. They are suitable representatives for the whole class of building blocks in larger biomolecules. All four benchmark states of cytosine are well reproduced by MSINDO-sCIS; their errors are in the range of 0.08–0.16 eV. For this molecule, the OMx methods have errors of 0.27–0.68 eV. Again, it can be observed that the error for  $n-\pi^*$  transitions is lower than for the  $\pi-\pi^*$  transitions. Thymine is a more problematic case; here, MSINDO-sCIS errors are 0.17–0.83 eV. This is comparable to the OMx results. The excited-state energies of uracil are underestimated by 0.12–0.65 eV. Again, we observe that the lower states are better described than the higher states. The vertical excitation energies of adenine are overestimated by 0.08–0.35 eV. Here, we find a correct state ordering compared to the TBE-2.

**5.2.2. Vertical Triplet Excitations.** The benchmark set of the calculated triplet excitations can be found in Table 4. We will discuss all of the results in the following paragraphs.

**Unsaturated Aliphatic Hydrocarbons.** The triplet state energies of the unsaturated hydrocarbons are underestimated for the smaller chains and the rings. For the larger chains (hexatriene and octatetraene), the energies of the first excited triplet states are overestimated, while the second excited triplet state energies are underestimated. The errors are in the range of 0.16–1.1 eV. This is similar to that for the OMx methods. Compared to common NDDO methods (MNDO, AM1, and PM3) and to INDO/S, where the excitation energies are underestimated by up to 3 eV, MSINDO-sCIS is an improvement.

**Aromatic Hydrocarbons and Heterocycles.** The vertical triplet excitations of benzene are described within an acceptable error range (0.25–0.85 eV) by MSINDO-sCIS. The  $^3E_{1u}$  and the  $^3B_{1u}$  state energies are underestimated, while the other states are slightly overestimated. The state ordering is correct and the errors decrease when the excitation energy is raised. For naphthalene,



**Table 5.** Deviations of Vertical Excitation Energies in eV for Singlet and Triplet Excited States from TBE for TD-B3LYP, INDO/S, and OM3-CISDTQ<sup>a</sup>

singlet states (count = 103) <sup>b</sup>	TD-B3LYP/TZVP <sup>c</sup>	INDO/S <sup>d</sup>	OM3-CISDTQ <sup>d</sup>	MSINDO-sCIS
mean error [eV]	−0.07	−0.23	−0.22	−0.10
mean abs. error [eV]	0.27	0.51	0.45	0.44
std. dev. [eV]	0.33	0.70	0.54	0.59
max.(+) dev. [eV]	1.02	2.79	1.76	2.19
max.(−) dev. [eV]	0.75	1.45	1.19	1.42
triplet states (count = 63)	TD-B3LYP/TZVP <sup>c</sup>	INDO/S <sup>d</sup>	OM3-CISDTQ <sup>d</sup>	MSINDO-sCIS
mean error [eV]	−0.45	−0.31	−0.26	−0.01
mean abs. error [eV]	0.45	0.65	0.45	0.50
std. dev. [eV]	0.49	0.86	0.54	0.59
max.(+) dev. [eV]		2.49	1.08	1.17
max.(−) dev. [eV]	0.93	2.01	1.17	1.17

<sup>a</sup> MSINDO-sCIS errors are given with respect to TBE-2. <sup>b</sup> Count for TBE-1 is 104 for OM3 and TD-B3LYP/TZVP and 103 for INDO/S. <sup>c</sup> TBE-1 values taken from ref 16. <sup>d</sup> TBE-1 values taken from ref 17.

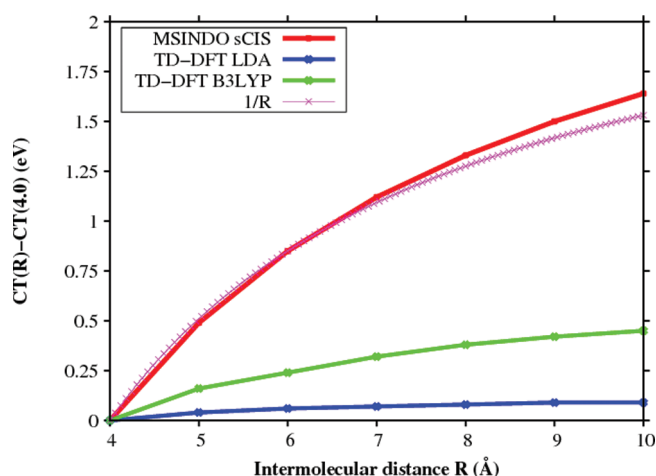
the errors scatter in the range of 0.27–0.88 eV. This is within the OMx error range and slightly better than the INDO/S results. The excited triplet states of furan are strongly underestimated by 0.93–1.17 eV but still less than in the case of the INDO/S method. Even INDO/S2 underestimates the second excited triplet state by more than 1.5 eV.

The lowest triplet states of pyrrole and imidazole are strongly underestimated by up to 1.0 eV. At the same time, the  $1^3A_1$  state of pyrrole is underestimated by 0.67 eV, which leads to a  $1^3B_2-1^3A_1$  splitting of 1.45 eV compared to 0.98 eV in TBE-2. The OMx methods result in a more correct  $1^3B_2-1^3A_1$  splitting between 0.75 and 0.89 eV. The errors for pyridine (0.08–0.52 eV) are comparably low and can compete with the OMx results, which are obtained on the CISDTQ level. Here, we see again that the totally symmetric excitation has the largest difference from the TBE value. The same holds for s-tetrazine (errors are 0.04–1.17 eV), where the MSINDO-sCIS results are closer to the TBE-2 values than any other semiempirical method, except for the highest vertical triplet excitation of s-tetrazine in the benchmark set. Here, we have a large error of 1.17 eV.

**Aldehydes, Ketones, and Amides.** The vertical triplet excitation energies of all aldehydes, ketones, and amides are overestimated by 0.07–0.88 eV. The qualitative state ordering is correct for all cases. Formaldehyde and acetone triplet excited states are described with acceptable errors comparable to OMx. Within the MSINDO-sCIS approach, we observe the general trend of an overestimation and no scattering as for all other semiempirical methods. Compared to OMx methods, the same behavior is observed for p-benzoquinone within the MSINDO-sCIS method. All states of this molecule are within a maximum overestimation of 0.2 eV except for the  $A_u$  state, where an error of 0.61 eV is obtained with MSINDO-sCIS. This is similar to that for the OMx methods, where the error for this state is in the range of 0.41–0.61 eV. In the case of amides, we see again the effect that the  $n-\pi^*$  transitions are described with a much smaller error than the  $\pi-\pi^*$  transitions. The errors are nearly twice as large for the  $\pi-\pi^*$  transitions. This is an effect which is not observed in the OMx or INDO/S methods. In the OMx methods, the  $n-\pi^*$  transitions are underestimated by nearly the same amount by which the  $\pi-\pi^*$  transitions are overestimated. This makes a simple shift of the values impossible for the OMx methods. Within the MSINDO-sCIS theory for amides, aldehydes, and

ketones, the values could in principle be shifted by a small amount to match it with experimental data.

**5.3. Statistical Evaluation.** To classify the MSINDO-sCIS method within the mainframe of computationally feasible methods, we compared the statistical evaluation to three methods. We have chosen INDO/S because of the conceptual equivalence, OM3-CISDTQ, because it has been demonstrated that this semiempirical method is currently the most accurate for the description of excited states, and TD-B3LYP/TZVP, because of its popularity in excited-state calculations. A statistical overview is given in Table 5. Since the benchmark results for TBE-1<sup>11</sup> and the more recent TBE-2<sup>1</sup> do not differ that much, the OM3,<sup>17</sup> INDO/S,<sup>17</sup> and TD-B3LYP<sup>16</sup> statistics available in the literature may be compared to the MSINDO-sCIS results benchmarked to TBE-2. It can be seen that for singlet states the MSINDO-sCIS method provides the same accuracy as the OM3-CISDTQ method. The mean errors are half those of the OM3 method, while the mean absolute errors are nearly the same. The standard deviation for singlets is comparable to those of the OM3 and lower compared to those of the INDO/S method. The maximum  $+/-$  deviations of MSINDO-sCIS for singlets are higher than in the OM3 method but still smaller than for the INDO/S method. The outlier in the MSINDO-sCIS method is the  $2^1A_1$  state of acetone, which is a  $\pi-\pi^*$  transition. The totally symmetric excitations are in general the problematic states in our CIS method, indicating that the simple correction term (eq 4) is not very accurate. A comparison with the INDO/S method shows that MSINDO-sCIS is slightly better in all statistical aspects, while TD-B3LYP/TZVP is in turn better by a factor of 1.5 to 2 in all values. But in the overall view on the important statistical points, it can be seen that the MSINDO-sCIS accuracy for singlets is sufficient. For triplet excitations, MSINDO-sCIS is well behaved. Although no higher excitations than singlets are explicitly included, MSINDO-sCIS is on a similar level as OM3-CISDTQ. Even the maximum  $+/-$  deviations are comparable to those of OM3-CISDTQ. Comparing the MSINDO-sCIS statistics with those of INDO/S shows that, although the methods are conceptionally similar, MSINDO-sCIS is superior. The comparison with TD-B3LYP shows that MSINDO-sCIS can compete with TD-DFT for triplet states. Here, MSINDO-sCIS has—similar to OM3-CISDTQ—only a slightly larger standard



**Figure 2.** Lowest excitation energy of a charge transfer (CT) state in the  $C_2H_4-C_2F_4$  complex.

deviation, which results in a comparable reliability to that of TD-B3LYP, although the maximum  $+/-$  deviations are higher.

## 6. CHARGE-TRANSFER STATES

The standard approach for excited state calculations, TD-DFT, has well-known problems with charge transfer (CT) states.<sup>13–15</sup> In order to compare our present approach with TD-DFT, we studied the common benchmark system, the  $C_2H_4-C_2F_4$  complex.<sup>53</sup> There is a high-lying CT state at around 13 eV, where one electron of ethene is transferred to the tetra-fluoro-ethene. To visualize the results, we plotted the excitation energy against the distance between both molecules (see Figure 2). For the TD-DFT calculations, we used the ORCA program package.<sup>50</sup> Starting with optimized structures of ethene and tetra-fluoro-ethene, the excitation energies were calculated starting with a distance of 4 Å. This distance was increased up to 10 Å with a step value of 1 Å. The energy of the CT state with respect to its value at  $R = 4$  Å was then plotted against the distance. It can be seen from Figure 2 that MSINDO-sCIS gives the correct  $1/R$  behavior, while the TD-DFT methods fail to give the correct description. This is a typical problem in TD-DFT and cannot even be solved by using double hybrid methods.<sup>15</sup> Within the ab initio CIS theory, on the other hand, this problem is totally absent. Therefore, it should not appear in semiempirical methods that are based on Hartree–Fock theory. Since we have introduced scaling parameters in the description of the excited state (eq 6), it was necessary to ensure that the CT error does not occur in our method. But according to the present results, the  $1/R$  behavior is still correctly reproduced with MSINDO-sCIS.

## 7. SUMMARY AND CONCLUSIONS

We have introduced a novel method for the calculation of excited states at a semiempirical level. Since the parametrization included ground state properties, the MSINDO-sCIS method yields reliable results for both ground and excited states. This is an improvement over the common INDO/S methods, which focus on excited states. We have demonstrated that the vertical excitation energies obtained with MSINDO-sCIS are in reasonable agreement with the TBEs, comparable to the OM3 methods. Most errors are in the range of 0.1–0.6 eV, with a trend toward larger errors for higher excitations. Compared to TD-B3LYP, the

MSINDO-sCIS method has a larger mean error, but it has two major advantages. First, the calculation times are orders of magnitude smaller, and second, the charge-transfer error is not present. A comparison of the accuracy with TD-B3LYP shows that MSINDO-sCIS can even compete for triplets. This is quite surprising, but it shows that the parameters are well balanced for the exchange part. For the triplet case, MSINDO-sCIS is superior to all other semiempirical methods except the OMx methods. Furthermore, due to the use of the Davidson–Liu algorithm, MSINDO-sCIS is computationally more efficient. Conceptually MSINDO-sCIS is below the OMx level, because OMx methods include NDDO integrals and also higher excitations. But we showed that a careful parametrization yields comparable results, although some of the parameters lost their physical significance. Therefore, it cannot be excluded that the new parametrization gives unbalanced results for systems that are quite different from those not included in the reference set. A couple of outliers have been observed in the present study, e.g., for the  $2^1A_1$  state of acetone. This may be an indicator for an unbalanced treatment of the excited state energies. There are other outliers where MSINDO-sCIS gives large errors for excited states with large doubles contributions (for example, in the polyenes). However, the overall performance is quite satisfactory.

Although analytical gradients for all sorts of semiempirical wave functions have been available for a number of years,<sup>54</sup> it is another advantage of the present approach that analytical gradients of a CIS wave function are much easier to implement<sup>55</sup> and faster to calculate than for higher excited determinants. This opens an efficient way for the calculation of electronic spectra, including vibrational coupling and excited-state geometries. Therefore, future applications to technically important systems, e.g., organic solar cells, are planned, where CT states play an important role and the molecules consist of several hundred atoms. Here, the larger errors for high-lying states do not play an important role since usually only the few lowest states are of interest.

Due to the implementation of the cyclic cluster model in MSINDO,<sup>56</sup> the calculation of excited state properties of solids and surfaces is another subject of future research with MSINDO-sCIS.<sup>57</sup>

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Tables 6 and 7. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [gadaczek@thch.uni-bonn.de](mailto:gadaczek@thch.uni-bonn.de).

## ■ ACKNOWLEDGMENT

The financial support by the Deutsche Forschungsgemeinschaft (SFB 813, project A2) is gratefully acknowledged.

## ■ REFERENCES

- (1) Silva-Junior, M.; Schreiber, M.; Sauer, S.; Thiel, W. *J. Chem. Phys.* **2010**, *133*, 174318.
- (2) Erman, P. In *Molecular Spectroscopy*; Barrow, R. F., Long, D. A., Sheridan, J., Eds.; The Royal Society of Chemistry: London, 1979; Vol. 6; pp 174–231.

- (3) Alerstam, E.; Andersson-Engels, S.; Svensson, T. *Optics Express* **2008**, *16*, 10440–10454.
- (4) Chattopadhyay, S.; Pahari, D.; Mahapatra, U.; Mukherjee, D. *Comput. Chem.: Rev. Curr. Trends* **2005**, *9*, 121.
- (5) Grimme, S. *Reviews in Computational Chemistry*; John Wiley & Sons, Inc.: New York, 2004; pp 153–218.
- (6) Buenker, R. J.; Peyerimhoff, S. D.; Butscher, W. *Mol. Phys.* **1978**, *35*, 771–791.
- (7) Finley, J.; Malmqvist, P.-A.; Roos, B. O.; Serrano-Andrés, L. *Chem. Phys. Lett.* **1998**, *288*, 299–306.
- (8) Christiansen, O.; Koch, H.; Jorgensen, P. *Chem. Phys. Lett.* **1995**, *243*, 409–418.
- (9) Koch, H.; Christiansen, O.; Jorgensen, P.; de Meras, A. M. S.; Helgaker, T. J. *Chem. Phys.* **1997**, *106*, 1808–1818.
- (10) Christiansen, O.; Koch, H.; Jorgensen, P. J. *Chem. Phys.* **1995**, *103*, 7429–7441.
- (11) Schreiber, M.; Silva-Junior, M.; Sauer, S.; Thiel, W. J. *Chem. Phys.* **2008**, *128*, 134110.
- (12) Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *52*, 997.
- (13) Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009–4037.
- (14) Casida, M. E. *THEOCHEM* **2009**, *914*, 3–18.
- (15) Grimme, S.; Neese, F. J. *Chem. Phys.* **2007**, *127*, 154116.
- (16) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. J. *Chem. Phys.* **2008**, *129*, 104103.
- (17) Silva-Junior, M.; Thiel, W. J. *Chem. Theory Comput.* **2010**, *6*, 1546–1564.
- (18) Jug, K. *Theor. Chim. Acta* **1969**, *14*, 91–135.
- (19) Thiel, W. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 559–580.
- (20) Ridley, J.; Zerner, M. C. *Theor. Chim. Acta* **1973**, *32*, 111–134.
- (21) Zerner, M. C.; Loew, G. H.; Kirchner, R. F.; Mueller-Westerhoff, U. T. J. *Am. Chem. Soc.* **1980**, *102*, 589–599.
- (22) Li, J.; Williams, B.; Cramer, C. J.; Truhlar, D. G. J. *Chem. Phys.* **1999**, *110*, 724–733.
- (23) Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *J. Phys. Chem.* **1992**, *96*, 135–149.
- (24) Kolb, M.; Thiel, W. J. *Comput. Chem.* **1993**, *14*, 775–789.
- (25) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.
- (26) Otte, N.; Scholten, M.; Thiel, W. J. *Phys. Chem. A* **2007**, *111*, 5751–5755.
- (27) Koslowski, A.; Beck, M. E.; Thiel, W. J. *Comput. Chem.* **2003**, *24*, 714–726.
- (28) Dewar, M. J. S.; Thiel, W. *Theor. Chim. Acta* **1977**, *46*, 89.
- (29) Dewar, M. J. S.; Thiel, W. J. *Am. Chem. Soc.* **1977**, *99*, 4899–4907.
- (30) Bredow, T.; Jug, K. In *MSINDO, Electronic Encyclopedia of Computational Chemistry*; v. Ragué Schleyer, P., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; Wiley: New York, 2004.
- (31) Head-Gordon, M.; Rico, R.; Oumi, M.; Lee, T. *Chem. Phys. Lett.* **1994**, *219*, 21–29.
- (32) Head-Gordon, M.; Oumi, M.; Maurice, D. *Mol. Phys.* **1999**, *96*, 593–602.
- (33) Baker, J. D.; Zerner, M. C. *Chem. Phys. Lett.* **1990**, *175*, 192–196.
- (34) Silva-Junior, M. R.; Sauer, S. P.; Schreiber, M.; Thiel, W. *Mol. Phys.* **2010**, *108*, 453–465.
- (35) Grimme, S. *Chem. Phys. Lett.* **1996**, *259*, 128–137.
- (36) Davidson, E. R. J. *Comput. Phys.* **1975**, *17*, 87–94.
- (37) Leininger, M. L.; Sherrill, C. D.; Allen, W. D.; Schaefer, H. F. S., III. *J. Comput. Chem.* **2001**, *22*, 1574–1589.
- (38) Hoffmann, W. *Computing* **1989**, *41*, 335–348.
- (39) Bacon, A. D.; Zerner, M. C. *Theor. Chim. Acta* **1979**, *53*, 21–54.
- (40) Blackford, L. S.; et al. *ACM Trans. Mater. Software* **2002**, *28*, 135–151.
- (41) Dongarra, J. J. *Int. J. High Perform. Appl. Supercomput.* **2002**, *16*, 1–199.
- (42) Ahlswede, B.; Jug, K. J. *Comput. Chem.* **1999**, *20*, 572–578.
- (43) Jug, K.; Kunert, L.; Köster, A. *Theor. Chem. Acc.* **2000**, *104*, 417–425.
- (44) Adachi, M.; Bredow, T.; Jug, K. *Dyes Pigments* **2004**, *63*, 225–230.
- (45) Ferro, N.; Bredow, T. J. *Comput. Chem.* **2010**, *31*, 1063–1079.
- (46) Ahlswede, B.; Jug, K. J. *Comput. Chem.* **1999**, *20*, 563–571.
- (47) Bartels, R. H. University of Texas and Center for Numerical Analysis and Report CNA-44. University of Texas: Austin, TX, 1972.
- (48) Slater, J. *Quantum Theory of Atomic Structure*; McGraw-Hill: New York, 1960; Vol. 1; pp 339–442.
- (49) Höger, S. Personal communications, 2010.
- (50) Petrenko, T.; Krylova, O.; Neese, F.; Sokolowski, M. *New J. Phys.* **2009**, *11*, 015001.
- (51) Head, J. *Int. J. Quantum Chem.* **2003**, *95*, 580–592.
- (52) González-Luque, R.; Merchán, M.; Roos, B. Z. *Phys. D* **1996**, *36*, 311–316.
- (53) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. J. *Chem. Phys.* **2003**, *119*, 2943–2946.
- (54) Patchkovskii, S.; Koslowski, A.; Thiel, W. *Theor. Chem. Acc.* **2005**, *114*, 84–89.
- (55) Gadaczek, I.; Krause, K.; Hintze, K.; Bredow, T. Manuscript in preparation, 2011.
- (56) Bredow, T.; Geudtner, G.; Jug, K. J. *Comput. Chem.* **2001**, *22*, 89–101.
- (57) Gadaczek, I.; Hintze, K.; Bredow, T. Manuscript in preparation, 2011.