JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2019, 15, 4591−4601

pubs.acs.org/JCTC

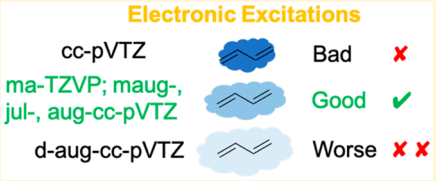
Nature of the 1^1B_u and 2^1A_g Excited States of Butadiene and the Goldilocks Principle of Basis Set Diffuseness

Sijia S. Dong,[§] Laura Gagliardi,*[] and Donald G. Truhlar*[]

Department of Chemistry, Chemical Theory Center, and Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455, United States

Supporting Information

ABSTRACT: Butadiene, being the simplest conjugated organic molecule, has been studied extensively by experiments and various high-level theoretical methods. Previous studies conclude that the complete active space (CAS) self-consistent field (SCF) method was unable to obtain the correct 1^1B_u and 2^1A_g state ordering and that it introduces artificial valence−Rydberg mixing into the 1^1B_u state because of the lack of external correlation. Basis sets and initial guesses specifically constructed for this problem were able to improve the vertical excitation energy of the 1^1B_u state but did not resolve the controversy of the nature of the 1^1B_u state. In the present work, we demonstrate that, using standard intermediately diffuse basis sets such as jul-cc-pVTZ and ma-TZVP, CASSCF is able to obtain the 1^1B_u and 2^1A_g states of *trans*-butadiene with much improved vertical excitation energy and reasonable wave function characteristics, and it provides a good reference wave function (capable of yielding quantitative excitation energies for excited states) for three methods that treat electron correlation in different ways, namely, multiconfiguration pair-density functional theory (MC-PDFT), CAS second-order perturbation theory (PT2), and multistate (MS) CASPT2. We demonstrate that a combined analysis of the orbital second moments, state second moments, and MC-PDFT energy components is a very useful approach in determining excited-state characteristics and assigning states, and we show that basis sets without diffuse functions or with very diffuse basis functions are not stable or accurate in predicting the excited states of butadiene. We show that the 2^1A_g state is valence-like and has an atypical double/single excitation character and the 1^1B_u state has a certain degree of Rydberg character that is not artificial, settling the decades of controversy of the characters of these states.



1. INTRODUCTION

Conjugated organic molecules comprise a large class of molecules that play important roles in nature and in chemical applications such as harvesting light and converting it into chemical, mechanical, or electrical energy as in photosynthesis, vision, light-driven molecular motors, and organic electronics. However, after decades of development, simulating the photochemistry of such systems remains a challenge.

Butadiene is the shortest polyene and the simplest case of a conjugated molecule. The lowest-energy structure is *s-trans*-butadiene, which has been studied extensively both experimentally and theoretically, but the inferences about the nature of its lowest excited states, especially the 1^1B_u and the 2^1A_g states, are still inconclusive. It is widely believed that the 1^1B_u state is valence-like[1] and that the valence−Rydberg mixing of this state observed in some theoretical studies is artificial due to the intrinsic problems of the theoretical methods,[2] such as the unbalanced treatment of correlation energy in multireference singles and doubles configuration interaction (MR-CISD)[3,4] and complete active space self-consistent field (CASSCF)[5] methods. However, an analysis of the experimental transition energies to the 3p Rydberg states and their displacements from their unperturbed values indicates that the 1^1B_u state does have valence−Rydberg mixing, and the Rydberg−valence coupling constant is twice that for cyclopentadiene.[6] The diffuseness of the 2^1A_g state is less controversial. It is believed to be a valence state.[1,7−9] However, it is a dark state and has not been resolved experimentally.[10]

Literature reports indicate that the vertical excitation energy of the 1^1B_u state is overestimated by the CASSCF method. In particular, Serrano-Andrés et al.[7] found that CASSCF significantly overestimates the vertical excitation energy of the 1^1B_u state to such an extent that the vertical excitation energies of the 1^1B_u and 2^1B_u states are in reverse order of their complete active space second-order perturbation theory (CASPT2)[11] energies. Their CASSCF(4,8) calculation [where the (...,...) notation denotes the number of active electrons followed by the number of active orbitals] predicts a vertical excitation energy of 8.54 eV for the 1^1B_u state, while the experimental value is only 5.92 eV.[12] For the 2^1B_u and 3^1B_u states, the vertical excitation energies are predicted to be 6.88 and 7.85 eV, respectively, while the experimental values are 7.07 and 8.00 eV, respectively. They argued that the significant overestimation of the 1^1B_u state vertical excitation energy occurs because CASSCF has inadequate dynamic correlation, and dynamic correlation affects valence states more than Rydberg states. The lower underestimation of the vertical excitation energies of the 2^1B_u and 3^1B_u states would then indicate that the 1^1B_u state is more valence-like than the other

Received: June 4, 2019
Published: July 15, 2019

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601

two ^1B_u states. Because the 1^1B_u state is artificially predicted to be closer in energy to the two more-Rydberg-like ^1B_u states, its valence−Rydberg mixing may be artificial.

In the CASSCF calculations by Serrano-Andrés et al.,[7] the ANO-L[13] basis set with two extra diffuse p functions on carbon was used. Drastically different CASSCF vertical excitation energy values of the 1^1B_u state have been reported when different basis sets were used. Nakayama et al.[14] used a basis set without diffuse functions, namely, QZ3p, which is constructed from the cc-pVQZ basis set (and that is [5s4p3d] for carbon and [3s2p] for hydrogen) and obtained an improved value of 7.73 eV for the 1^1B_u state with a (4,8) active space. Boggio-Pasqua et al.[15] constructed a new basis set, 6-31G*+3p, which is 6-31G* with specially designed 3p functions added to carbon. They obtained an improved 1^1B_u vertical excitation energy of 6.56 eV for a geometry optimized with CASSCF(4,4)/6-31G*. This value is not directly comparable to the previous two because the previous two calculations used the experimental geometry of *s-trans*-butadiene. Cave[16] reported an even more improved CASSCF 1^1B_u vertical excitation energy of 6.48 eV, but no computational details were given.

By examining these results, we hypothesized that the diffuse functions in the basis set may play a key role in predicting the CASSCF vertical excitation energy of the 1^1B_u state, and the present study is designed to sort this out. In particular, our goal is to finally settle the question—for this prototype of conjugated molecules that has been studied since the early days of quantum chemistry—of how does the diffuseness of the basis set affect the character of the states predicted, such as how much Rydberg character the states have. Therefore, in this study, we tested a number of different basis sets to establish a relationship between the diffuseness of the basis sets and the CASSCF vertical excitation energy. A key aspect of our study is that we consider not only the excitation energies but also the second moments and oscillator strengths of the 1^1A_g (ground state), 1^1B_u, and 2^1A_g states of *s-trans*-butadiene. We also tested three post-SCF methods, namely, CASPT2, multistate CASPT2 (MS-CASPT2),[17] and multiconfiguration pair-density functional theory (MC-PDFT),[18,19] using the same CASSCF calculation as a reference, to see how the diffuseness of the basis sets affects their vertical excitation energies.

The MC-PDFT method obtains a reference wave function by a multiconfiguration wave function calculation, calculates the electronic kinetic energy and classical electrostatic energy from the reference wave function, and calculates the other energy contributions from a density functional that is a functional of the total density and on-top pair density arising from the reference wave function. It was developed by aiming to achieve the same level of energetic accuracy as CASPT2 while requiring a much less expensive post-SCF calculation. Using MC-PDFT, we successfully described the vertical excitation energies and oscillator strengths of a carotenoid, the 11-*Z*-retinal protonated Schiff base, whose main component is a polyene.[20] We found that MC-PDFT was more sensitive to the choices of basis sets than CASPT2, but we were able to select a reasonable basis set based on the dipole moment of the molecule. Butadiene, as the shortest polyene, has a ground-state geometry of C_{2h} symmetry and does not have a dipole moment, and the diffuseness of its low-lying excited states is more controversial. Therefore, it is meaningful to study how the diffuseness of the basis sets affects butadiene's MC-PDFT description of its low-lying excited states and how to choose basis sets for a simple polyene with zero dipole moment.

Other theoretical methods could also be considered for in-principle higher accuracy on a small molecule like butadiene, but they are more expensive than the methods discussed here and would be less suitable for studying the wide range of large polyenes; therefore, the focus of this article is to explore these more cost-effective methods and especially to study the question of how to generate a reasonable reference wave function for methods that use the popular CASSCF scheme as the reference to predict the excited states of conjugated organic molecules.

2. COMPUTATIONAL METHODS

In this study, we used the experimental ground-state geometry of *trans*-1,3-butadiene.[21] We also considered a theoretical ground-state geometry optimized by the CCSD(T)/aug-cc-pVQZ method by Watson and Chan.[2] The coordinates can be found in the Supporting Information (SI). These geometries place the molecule in the *xy* plane with C_{2h} symmetry. Only singlets are considered.

We carried out the CASSCF, CASPT2, MS-CASPT2, and MC-PDFT calculations with *OpenMolcas 18.0*.[22−24] We modified *OpenMolcas 17.0* to obtain the second moments of each orbital, including both occupied and virtual orbitals, and distributed this functionality in the public version of *OpenMolcas*. We carried out both state-specific (SS) CASSCF calculations and state-averaged (SA) CASSCF calculations. If *n* states are averaged in the SA-CASSCF calculation, we designate the calculation by SA(*n*)-CASSCF. The MC-PDFT calculations in this work were carried out with the tPBE[18,25] on-top density functional. We carried out the CASPT2 calculations with an imaginary shift[26] of 0.1 E_h (1 E_h = 1 hartree = 27.2114 eV) and an ionization potential−electron affinity (IPEA) shift[27] of 0.25 E_h, which are standard values. Note that the CASPT2 vertical excitation energies for butadiene will tend to be greater if the IPEA shift is greater.[28] Because the major goal of calculating CASPT2 energies in this article is to compare across different basis sets, we used the same shift values for different basis sets and did not attempt to scan through shift values that are different from standard values.

We carried out calculations with two choices of symmetry constraints, in particular, C_{2h} and C_s. For states of A_g symmetry in the C_{2h} point group, we carried out SS-CASSCF and SA(2)-CASSCF calculations to provide the reference wave functions because we are interested in the ground state and the 2^1A_g state. For states of the B_u symmetry in the C_{2h} point group, we carried out SS-CASSCF and SA(3)-CASSCF calculations. With the C_s point group, we carried out SA(*n*)-CASSCF (*n* = 2, 3, or 5) calculations to provide reference wave functions for A′ states, which can be assigned to either A_g or B_u by inspecting the orbitals involved in the transition. For each symmetry constraint, we carried out MC-PDFT and CASPT2 calculations using each CASSCF wave function as the reference wave function, and we carried out MS-CASPT2 calculations using each SA-CASSCF calculation as the reference wave function. A wave function analysis (WFA)[29,30] was carried out using the *WFA* module in *OpenMolcas* for the C_s CASSCF wave functions.

In this study, we chose the (4,8) active space, which includes all of the valence π electrons and π orbitals and two π^* orbitals of both a_u and b_g symmetry. This is the same active space

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601

Table 1. Four Smallest Exponential Parameters of Carbon p Functions for the Basis Sets Compared in This Work

group	basis set[a]	exponents			
group 1	cc-pVTZ	4.1330000	1.2000000	0.3827000	0.1209000
	QZ3p	2.3680000	0.8132000	0.2890000	0.1007000
group 2	6-31+G(d)	1.8812885	0.5442493	0.1687144	0.0438000
	6-311+G(2df,2p)	1.4593300	0.4834560	0.1455850	0.0438000
group 3	6-31G*+3p	0.2864610	0.1687144	0.1110960	0.0389185
group 4	maug-cc-pVTZ	1.2000000	0.3827000	0.1209000	0.0356900
	jul-cc-pVTZ	1.2000000	0.3827000	0.1209000	0.0356900
	aug-cc-pVTZ	1.2000000	0.3827000	0.1209000	0.0356900
	ma-TZVP	0.8143321	0.2888755	0.1005682	0.0335227
group 5	ANO-L	0.3619440	0.1547400	0.0654290	0.0229000
group 6	d-aug′-cc-pVDZ	0.5456000	0.1517000	0.0404100	0.0105000
	d-aug-cc-pVTZ	0.3827000	0.1209000	0.0356900	0.0105000
	DZP-R/DZ	0.0399000	0.0157500	0.0093100	0.0036750
	maug2-cc-pVTZ	0.1209000	0.0356900	0.0080100	0.0028100
	ANO-L2	0.0654290	0.0229000	0.0080100	0.0028100

[a]The basis sets are ranked in descending order of the smallest exponent. If the smallest exponential parameters of the basis sets are the same, we rank the basis sets by the second smallest exponential parameter in descending order.

choice as that in Serrano-Andrés et al.[7] Previous studies have shown that the (4,4) active space that includes only four π^* orbitals is not adequate to recover the π component of the electronic correlation and that more diffuse 3π-like orbitals need to be included in the active space.[7,15]

For basis sets, we considered several different kinds of basis set. Among the standard basis sets, we considered Dunning-type correlation-consistent basis sets with different levels of augmentation, in particular, cc-pVTZ, maug-cc-pVTZ, jul-cc-pVTZ, aug-cc-pVTZ, and d-aug-cc-pVTZ,[31−34] an Ahlrichs-type triple-ζ basis set with minimal augmentation, in particular, ma-TZVP,[35−37] Pople-type 6-31+G(d)[38,39] and 6-311+G-(2df,2p) basis sets,[39−41] and a Roos-type atomic natural orbital (ANO) basis set, namely, ANO-L.[13] We also considered some nonstandard basis sets. The first is the basis set used by Serrano-Andrés et al.,[7] in which two diffuse p functions, with exponential parameters 0.00801 and 0.00281, were added to the carbon atom of the ANO-L basis set; we denote this basis set as ANO-L2. To further study the effect of these extremely diffuse basis functions, we constructed a new basis set by adding these two carbon diffuse functions from Serrano-Andrés et al.[7] to maug-cc-pVTZ; we name this basis set maug2-cc-pVTZ. We also constructed the 6-31G*+3p basis set described by Boggio-Pasqua et al.[15] The exponential parameters and coefficients of the p-type functions of the 6-31G*+3p basis set of carbon are in the SI. Note that they might be (slightly) different from the 6-31G*+3p basis set actually used in Boggio-Pasqua et al.[15] because the authors did not report the full numerical details of their basis set but only described its construction in the text.

The basis sets QZ3p, 6-31G*+3p, d-aug′-cc-pVDZ, DZP-R/DZ, and ANO-L2 are nonstandard basis sets used in the previous studies with which we compare our results, but we did not do additional calculations with these basis sets except for 6-31G*+3p. The basis sets 6-31G*+3p and ANO-L2 are described in the previous paragraph. The QZ3p basis set was used in Nakayama et al.[14] and is constructed from cc-pVQZ (5s4p3d for carbon and 3s2p for hydrogen). The d-aug′-cc-pVDZ basis set was used in the MR-CISD calculations in Dallos and Lischka.[1] It is aug-cc-pVDZ plus the d-augmentation (in addition to standard s and p augmentation) for carbon and cc-pVDZ plus the standard augmentation and d-augmentation on s functions for hydrogen. The DZP-R/DZ basis set is what was used in the symmetry-adapted cluster configuration interaction (SAC-CI) calculations in Saha et al.[9] It is the Huzinaga−Dunning[42] double-ζ basis set with polarization on carbon atoms [4s2p1d]/[2s] plus double-ζ Rydberg functions for 3s, 3p, 3d, 4s, and 4p orbitals of carbon atoms.[43] For convenience, we named it DZP-R/DZ (DZP-R for carbon atoms and DZ for hydrogen atoms).

The second moments $\langle vw \rangle$ (v = x,y,z; w = x,y,z) and $\langle r^2 \rangle$ = $\langle x^2 \rangle$ + $\langle y^2 \rangle$ + $\langle z^2 \rangle$ in Cartesian coordinates were computed to determine the charge distribution of orbitals and states. The second moments of charge distributions presented in this paper include only the electronic contributions (not the nuclear contributions). We use the convention that electron moments are given in atomic units (au) as positive numbers. Roughly speaking, the greater the second moment of an orbital or state, the more diffuse the orbital or state in that coordinate; thus, the second moments provide a measure of Rydberg character or partial Rydberg character or how spread out the charge distribution is.

3. RESULTS AND DISCUSSION

We found that the theoretical geometry (optimized by Watson and Chan[2] using CCSD(T)/aug-cc-pVQZ) gives vertical excitation energies (including CASSCF, CASPT2, MS-CASPT2, and MC-PDFT) below those of the experimental geometry but within 0.03 eV for the 1^1B_u and 2^1A_g states, and it gives the same conclusion as the experimental geometry for the effect of basis sets in this study. Therefore, we focus on results for the widely used experimental geometry of *trans*-1,3-butadiene.

Table 1 summarizes the four smallest exponential parameters of carbon p functions for the basis sets considered in this work. These values provide an indication of how diffuse each basis set is. The smaller the exponential parameters, the more diffuse the basis functions. The table shows that the two basis sets with the diffuse functions from Serrano-Andrés et al.[7] are the most diffuse, and they are followed by DZP-R/DZ, d-aug-cc-pVTZ, and d-aug′-cc-pVDZ. There are two basis sets that do not have diffuse functions: cc-pVTZ and QZ3p. For convenience, the basis sets are grouped based on the

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601

Table 2. Vertical Excitation Energy in eV, CASSCF Oscillator Strength (osc.), Electronic Part of the CASSCF Second Moment z Component $\langle z^2 \rangle$ in au, and the Second Moment $\langle r^2 \rangle$ and $\langle z^2 \rangle$ Values Relative to Those of the SS-CASSCF 1^1A_g ($\Delta\langle r^2 \rangle$ and $\Delta\langle z^2 \rangle$) in au Using Various Basis Sets for 1^1B_u and 2^1A_g States of Butadiene from C_{2h} Calculations

	n[a]	state	CASSCF	CASPT2	MS-CASPT2	MC-PDFT	osc. str.	$\langle z^2 \rangle$	$\Delta\langle z^2 \rangle$	$\Delta\langle r^2 \rangle$
cc-pVTZ	2	2^1A_g	6.79	6.59	6.59	6.69	–	21.6	0.0	1.8
	1	1^1B_u	7.84	6.49	–	4.91	1.01	22.7	1.1	3.4
	3	1^1B_u	8.01	6.41	6.38	4.75	1.02	22.7	1.1	3.2
6-31+G(d)	2	2^1A_g	6.66	6.54	6.54	6.53	–	23.4	1.4	4.4
	1	1^1B_u	7.08	6.70	–	5.76	0.51	32.7	10.7	20.7
	3	1^1B_u	7.20	6.71	6.51	5.83	0.49	33.1	11.1	21.4
6-311+G(2df,2p)	2	2^1A_g	6.69	6.46	6.46	6.48	–	23.1	1.3	4.3
	1	1^1B_u	7.01	6.68	–	5.73	0.52	32.3	10.5	20.2
	3	1^1B_u	7.15	6.71	6.39	5.82	0.49	32.8	11.0	20.8
6-31G*+3p	2	2^1A_g	6.67	6.53	6.53	6.53	–	22.0	1.5	4.8
	1	1^1B_u	6.87	6.75	–	6.03	0.31	39.3	17.3	28.8
	3	1^1B_u	7.09	6.74	6.22	6.15	0.27	40.6	18.6	31.3
maug-cc-pVTZ	2	2^1A_g	6.68	6.45	6.45	6.47	–	23.4	1.7	5.0
	1	1^1B_u	6.83	6.76	–	5.95	0.36	37.3	15.6	27.0
	3	1^1B_u	7.09	6.79	6.02	6.10	0.30	39.0	17.3	29.8
jul-cc-pVTZ	2	2^1A_g	6.68	6.42	6.42	6.46	–	23.3	1.5	4.7
	1	1^1B_u	6.83	6.73	–	5.94	0.37	37.2	15.4	26.9
	3	1^1B_u	7.09	6.75	5.97	6.07	0.32	38.9	17.1	29.8
aug-cc-pVTZ	2	2^1A_g	6.68	6.42	6.42	6.46	–	23.3	1.6	4.7
	1	1^1B_u	6.81	6.75	–	5.97	0.35	38.4	16.7	28.2
	3	1^1B_u	7.09	6.77	5.94	6.10	0.29	40.8	19.1	31.6
ma-TZVP	2	2^1A_g	6.68	6.46	6.46	6.47	–	23.4	1.6	4.9
	1	1^1B_u	6.78	6.78	–	6.03	0.31	39.7	17.9	29.8
	3	1^1B_u	7.05	6.78	6.01	6.14	0.26	41.4	19.6	32.5
ANO-L	2	2^1A_g	6.71	6.52	6.52	6.54	–	22.3	0.4	2.9
	1	1^1B_u	7.12	6.65	–	5.64	0.58	31.7	9.8	17.3
	3	1^1B_u	7.23	6.66	6.28	5.70	0.56	32.2	10.3	18.1
d-aug-cc-pVTZ	2	2^1A_g	6.67	6.41	6.41	6.44	–	23.8	2.1	5.8
	1	1^1B_u	6.48	6.83	–	6.28	0.11	65.4	43.7	67.9
	3	1^1B_u	8.50	6.29	5.96	4.97	0.97	29.8	8.1	9.9
maug2-cc-pVTZ	2	2^1A_g	6.68	6.45	6.45	6.45	–	23.8	2.1	5.9
	1	1^1B_u	6.51	6.84	–	6.31	0.12	72.2	50.5	76.4
	3	1^1B_u	8.55	6.41	5.97	5.05	0.93	42.3	20.6	30.5
ANO-L2[b]		2^1A_g	6.64	6.27			–	23.2	1.3	
	3	1^1B_u	8.54	6.23			0.686	40.9	19.0	

[a] n represents the number of states averaged in the CASSCF calculation. When n = 1, it is a SS-CASSCF calculation. [b] The values are from Serrano-Andrés et al.[7] It was not described in the article whether the CASSCF calculation for the ^1A_g states was state-specific or state-averaged.

similarities in their diffuseness and their behaviors, which will be described in the following paragraphs.

Using one basis set that does not have diffuse functions, cc-pVTZ, eight standard basis sets with diffuse functions, namely, 6-31+G(d), 6-311+G(2df,2p), maug-cc-pVTZ, jul-cc-pVTZ, aug-cc-pVTZ, ma-TZVP, ANO-L, and d-aug-cc-pVTZ, and two constructed basis sets with diffuse functions, 6-31G*+3p and maug2-cc-pVTZ, we calculated the vertical excitation energies (CASSCF, CASPT2, MS-CASPT2, and MC-PDFT), oscillator strengths (CASSCF), and second moments (CASSCF) of butadiene for the states of A_g and B_u symmetry. Selected C_{2h} results are summarized in Table 2, and the complete set of results are in Table S1 (C_{2h}) and Table S4 (C_s) [tables and figures with the prefix S are in the SI]. The absolute energies are in Table S5 (C_{2h}) and Table S6 (C_s). The ANO-L2 results by Serrano-Andrés et al.[7] are listed together with our calculated C_{2h} results in this article because Serrano-Andrés et al. used the C_{2h} symmetry in their calculations. The ground state was 1^1A_g. For the excited states, we focused on the 1^1B_u and 2^1A_g states.

3.1. Basis Set Effect on the Vertical Excitation Energies of the 1^1B_u and 2^1A_g States.

3.1.1. 1^1B_u State. For the 1^1B_u state, two experimental excitation energies for gas-phase *trans*-1,3-butadiene have been reported, 5.92[6,12,44−48] and 6.25 eV,[6] with the former most commonly cited, and reinterpretation of the spectra gives 5.96−6.05 eV.[2] The best theoretical estimates range from 6.1 to 6.32 eV.[1,2,49,50] We take the mean of the minimum and maximum of these theoretical estimates as our best estimate (BE); this gives 6.2 eV with an estimated uncertainty of 0.1 eV, and we estimate errors in all calculated vertical excitation energies by comparison to this BE value; these error estimates are given in Figures 1 and S1.

In the figures, results from our calculations are listed such that the less diffuse basis sets are on the left and the more diffuse basis sets are on the right, where the diffuseness is defined as explained above with regard to Table 1. From Figure 1, it is clear that the calculations with basis functions that are not diffuse enough or are too diffuse have less accurate vertical excitation energies than calculations using basis functions of medium diffuseness. The errors gradually decrease as we

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601

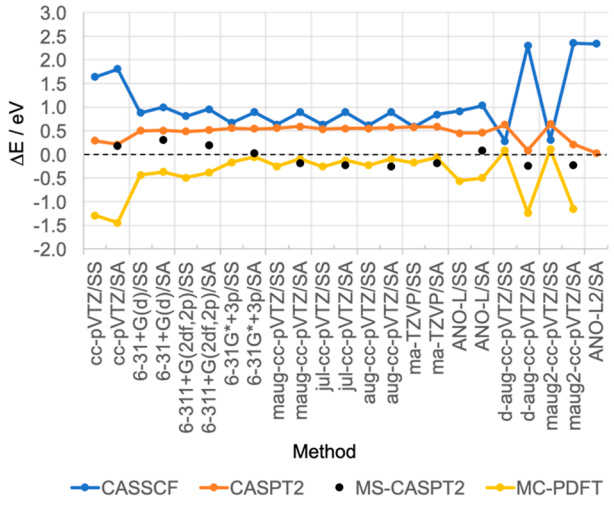


Figure 1. Difference between various theoretical vertical excitation energies (from C_{2h} calculations) and the BE (6.2 ± 0.1 eV) for the 1^1B_u state of butadiene. Methods are labeled "SS" if the reference wave function is from the corresponding state-specific CASSCF calculation, and methods are labeled "SA" if the reference wave function is from the corresponding SA(3)-CASSCF calculation. Those labeled "SS" do not have MS-CASPT2 values.

proceed from group 1 basis sets to group 4 basis sets, and then, they increase in group 5 and even more in group 6 basis sets. In particular, the standard basis sets in group 4, namely, maug-cc-pVTZ, jul-cc-pVTZ, aug-cc-pVTZ, and ma-TZVP, as well as the specifically constructed 6-31G*+3p basis set (group 3), are the most stable and have the smallest errors in vertical excitation energy across CASSCF, CASPT2, MS-CASPT2, and MC-PDFT. For each basis set belonging to groups 2−5, the predicted states from using the SS-CASSCF reference wave function are not significantly different from the predicted states from using the SA-CASSCF reference wave function. However, for the very diffuse d-aug-cc-pVTZ and maug2-cc-pVTZ basis sets, this difference is quite drastic.

We notice that for d-aug-cc-pVTZ and maug2-cc-pVTZ the SS-CASSCF calculation and the MC-PDFT calculation using SS-CASSCF as the reference give very small errors. This is just fortuitous. For the SS-CASSCF calculation of the B_u symmetry, only one state is obtained; therefore, we label it 1^1B_u. However, in the case of d-aug-cc-pVTZ and maug2-cc-pVTZ, this state corresponds to the 2^1B_u state in their respective SA(3)-CASSCF calculations. This assignment is confirmed by comparing the oscillator strengths, the state second moments (Table S1 of the SI), and the orbital second moments (Tables S2 and S3 of the SI) of the states involved. For example, as shown in Table S1, the state second moment $\langle z^2 \rangle$ for the B_u state from SS-CASSCF/d-aug-cc-pVTZ is closer to that of the 2^1B_u state from SA(3)-CASSCF than to that of the 1^1B_u state. To explain this observation, we decomposed the state second moments into their orbital components. As shown in Table S2, for the three B_u states from SA(3)-CASSCF/d-aug-cc-pVTZ, the a_u orbital that is singly occupied (occupation numbers being 0.992, 0.997, and 0.999, respectively) has an orbital second moment $\langle z^2 \rangle$ of 9.6, 54.0, and 76.8 au, respectively. For the B_u state from SS-CASSCF/d-aug-cc-pVTZ, the a_u orbital that is singly occupied (occupation number being 0.993) has a second moment $\langle z^2 \rangle$ of 46.6 au, which has the same order of magnitude as that for the 2^1B_u state from SA(3)-CASSCF. In addition, the 1^1B_u state from SA(3)-CASSCF does have an a_u orbital with a similar amount of diffuseness as the singly occupied a_u orbital of the ^1B_u state from SS-CASSCF, with a $\langle z^2 \rangle$ of 48.3 au, but this orbital is unoccupied (occupation number 0.006). The analysis for maug2-cc-pVTZ is similar. These results demonstrate that the B_u state from SS-CASSCF using the very diffuse basis sets d-aug-cc-pVTZ and maug2-cc-pVTZ corresponds to the 2^1B_u state from SA(3)-CASSCF and not 1^1B_u and that the small errors for d-aug-cc-pVTZ and maug2-cc-pVTZ in Figure 1 are from comparing the excitation energy of the calculated state to the reference energy from a state not corresponding to the calculated state.

The results from using C_s symmetry are similar, as demonstrated by Figure S1. The major difference between the C_s results and the C_{2h} results is the small error from using the SA(5)-CASSCF/maug2-cc-pVTZ reference wave functions in the C_s calculations. This is also caused by comparing the excitation energy of the calculated state to the reference energy from a state not corresponding to the calculated state.

Therefore, we conclude that extremely diffuse basis sets and basis sets without diffuse functions, in particular, those with the smallest carbon p exponential parameter less than ∼0.01 or greater than 0.10, are particularly unsuitable for CASSCF calculations of the 1^1B_u state for butadiene. The best basis sets are group 4 basis sets, which keep the absolute errors of CASSCF, CASPT2, MS-CASPT2, and MC-PDFT using either C_{2h} or C_s symmetry within a maximum of 0.9, 0.6, 0.5, and 0.4 eV, respectively. Remarkably, the absolute MC-PDFT errors for group 4 basis sets are close to those of MS-CASPT2, but across all basis sets tested, they correlate more with CASSCF errors than with CASPT2 or MS-CASPT2. Therefore, we should also avoid using extremely diffuse basis sets for MC-PDFT calculations of the 1^1B_u state for butadiene. Nevertheless, as will be discussed in the next section, MC-PDFT orders the 1^1B_u state below the 2^1A_g state for calculations in both C_{2h} and C_s symmetry, which agrees with the theoretical BEs in Table 3, while CASSCF and CASPT2 fail to order the two states correctly for calculations in C_s symmetry. MS-CASPT2 does not correct the ordering for C_s results unless SA(5)-CASSCF is used to generate the reference wave function (Table S4). Note that CASPT2 was found to be able to predict the ordering of the 1^1B_2 and 2^1A_1 states of *s-cis*-butadiene (C_s symmetry)[51] correctly, as well as the ordering of the 1^1B_u and 2^1A_g states of all-*trans*-1,3,5-hexatriene (C_{2h} symmetry).[52] Differences in the molecules and the particular methodologies used (e.g., active space, number of states averaged, basis set, shift values) are possibilities in causing such differences in the CASPT2 performance. Because C_s calculations have fewer symmetry constraints than C_{2h} calculations, the performance of C_s calculations is considered more realistic and a gauge of potential success in practical calculations on complex systems. Therefore, for *s-trans*-butadiene, although MS-CASPT2 C_s calculations using the SA(5)-CASSCF reference wave functions have less error than MC-PDFT, it is encouraging that MC-PDFT energies are more stable across different reference wave functions than MS-CASPT2 and that MC-PDFT does not need to average over more than three states to obtain the correct ordering of the 1^1B_u and 2^1A_g states.

3.1.2. 2^1A_g State. For the 2^1A_g state, a similar trend is found for the errors in the vertical excitation energies versus the diffuseness of basis set. Figure 2 shows the difference between various theoretical vertical excitation energies from C_{2h} calculations and the best theoretical estimate (6.50 eV by

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601

Table 3. Summary of Selected Literature Results of Vertical Excitation Energies ΔE (eV), Oscillator Strengths (osc. str.), and Second Moments $\langle r^2 \rangle$ or $\langle z^2 \rangle$ (in au) of Butadiene

method[a]		1^1A_g	2^1A_g	1^1B_u	2^1B_u	3^1B_u
theory (MR-AQCC): Dallos and Lischka[1]	ΔE	0.00	–	6.18	–	–
theory (RCA3-F/MR-CISD+Q): Dallos and Lischka[1]	ΔE	0.00	6.55	–	–	–
theory (FCIQMC): Daday et al.[49]	ΔE	0.00	–	6.32 ± 0.03	–	–
theory (EOM-CCSDTQ): Watson and Chan[2]	ΔE	0.00	6.39 ± 0.07	6.21 ± 0.02	–	–
theory (exFCI/AVDZ+CC3/AVQZ): Loos et al.[53]	ΔE	0.00	6.50	–	–	–
theory (EOM-CCSDT/aPVTZ+experiment): Rabidoux et al.[50]	ΔE	0.00	6.24	6.1	–	–
experiment	ΔE	0.00	–	5.92,[b] 6.25[c]	7.07[b]	8.00[b]
	osc. str.	–	–	0.4[d]	–	–
SAC-CI//DZP-R/DZ[e]	ΔE	0.00	6.56	6.33	7.08	7.90
	osc str.	–	–	0.6467	0.2453	0.0016
	$\langle r^2 \rangle$	341.9	351.9	366.1	413.8	587.9
MR-CISD/SA(3)-CAS(4,5)-R/maug-cc-pVTZ[f]	$\Delta E_{\text{MR-CISD}}$	0.00	6.70	6.77	–	–
	$\Delta E_{\text{MR-CISD+Q}}$	0.00	6.67	6.47	–	–
	$\langle z^2 \rangle_{\text{CI}}$	21.6	21.6	26.4	–	–
MR-AQCC/SA(3)-CAS(4,5)-R/maug-cc-pVTZ	ΔE	0.00	6.62	6.27	–	–
	$\langle z^2 \rangle_{\text{AQCC}}$	21.8	23.0	28.1	–	–
MR-CISD/SS-CAS(4,4)/maug-cc-pVDZ	$\Delta E_{\text{MR-CISD}}$	0.00	–	7.13	–	–
	$\Delta E_{\text{MR-CISD+Q}}$	0.00	–	6.82	–	–
	$\langle z^2 \rangle_{\text{CI}}$	21.8	–	31.9	–	–
MR-CISD/SS-CAS(4,4)/d-aug′-cc-pVDZ	$\Delta E_{\text{MR-CISD}}$	0.00	–	7.08	–	–
	$\Delta E_{\text{MR-CISD+Q}}$	0.00	–	6.94	–	–
	$\langle z^2 \rangle_{\text{CI}}$	21.9	–	72.8	–	–
MR-CISD/SA(3)-CAS(4,4)/d-aug′-cc-pVDZ	ΔE_{CASSCF}	0.00	6.67	8.15		
	$\langle z^2 \rangle_{\text{CASSCF}}$	22.07	22.51	22.53		
	$\Delta E_{\text{MR-CISD}}$	0.00	6.74	7.04	–	–
	$\Delta E_{\text{MR-CISD+Q}}$	0.00	6.70	6.62	–	–
	$\langle z^2 \rangle_{\text{CI}}$	22.0	22.3	22.9	–	–
MR-CISD/SA(4)-CAS(4,4)/d-aug′-cc-pVDZ	ΔE_{CASSCF}	0.00	6.63	6.65	8.30	–
	$\langle z^2 \rangle_{\text{CASSCF}}$	21.69	22.19	77.77	27.61	–
	$\Delta E_{\text{MR-CISD}}$	0.00	6.68	6.73	7.40	–
	$\Delta E_{\text{MR-CISD+Q}}$	0.00	6.53	6.69	7.16	–
	$\langle z^2 \rangle_{\text{CI}}$	21.8	22.2	53.9	50.8	–

[a]For each listed theoretical paper, we give the BE. The highest levels of theory on which the listed values are based are summarized in the parentheses. Note that semistochastic heat-bath CI has been used to approximate the full-CI limit using the AVDZ basis set and gives 6.45 eV and 6.58 eV for the vertical excitation energy for the 1^1B_u and 2^1A_g states of butadiene.[61] Because a basis set extrapolation was not carried out, we did not include these values in the list of best estimates. [b]Experimental values.[6,12,44−48] [c]Experimental value from McDiarmid.[6] [d]Experimental value from Brundle and Robin.[54] [e]SAC-CI calculation with the DZP-R/DZ basis set.[9] [f]All MR-CISD and their reference CASSCF results are from Dallos and Lischka.[1] SA(3) calculations averages over states $1^1A_g + 2^1A_g + 1^1B_u$. SA(4) calculations averages over states $1^1A_g + 2^1A_g + 1^1B_u + 2^1B_u$. The $\langle x^2 \rangle$ reported by Dallos and Lischka is equivalent to $\langle z^2 \rangle$ in the current article because the molecule in Dallos and Lischka is in the *yz* plane and our molecule is in the *xy* plane. The "R" in CAS(4,5)-R means only excitations from the reference configurations that have the same symmetry as the states calculated are considered; this follows the definition in Dallos and Lischka.

Loos et al.)[53] of the 2^1A_g state. We use the theoretical BE in lieu of an experimental value because the 2^1A_g state is a dark state and no experimental vertical excitation energy has been reported. CASPT2 and MS-CASPT2 vertical excitation energies differ by only less than 0.001 eV; therefore, we use CASPT2 to represent both results in the discussion. As the basis sets increase in diffuseness, the CASPT2 and MC-PDFT excitation energies decrease and then increase up to ANO-L, with jul-cc-pVTZ and aug-cc-pVTZ at the minimum. As compared to the 1^1B_u state, the 2^1A_g results depend on the diffuse characters of the basis sets to a lesser extent, with a change in vertical excitation energies within a range much smaller than that for the 1^1B_u state. For example, the difference of the 2^1A_g vertical excitation energy between the 6-31G*+3p (group 3) and 6-31+G(d) (group 2) calculations is much smaller than this difference for the 1^1B_u state. Note that the maug2-cc-pVTZ basis set, while having extremely diffuse functions, has errors with magnitudes more similar to those of maug-cc-pVTZ than to those of more diffuse basis sets such as aug-cc-pVTZ. A broad generalization of the observed trends for the 2^1A_g state is that the more diffuse the basis set, the lower the calculated vertical excitation energy for the 2^1A_g state, and again, the basis sets of medium diffuseness give the most accurate results.

For C_s calculations, the basis set dependence of the 2^1A_g state is more prominent than that in C_{2h} calculations but is still not as strong as that for the 1^1B_u state. As shown in Figure S2, two of the most diffuse basis sets, d-aug-cc-pVTZ and maug2-cc-pVTZ, give absolute errors similar to or greater than those of cc-pVTZ, unlike their behavior in C_{2h} calculations. In addition, the MC-PDFT and CASPT2/MS-CASPT2 results for the 2^1A_g state separate, with MC-PDFT errors from C_s calculations being slightly greater than those from the C_{2h} calculations for basis sets in groups 2−4.

Overall, 2^1A_g vertical excitation energies from CASPT2, MS-CASPT2, and MC-PDFT seem to be very close to the

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601

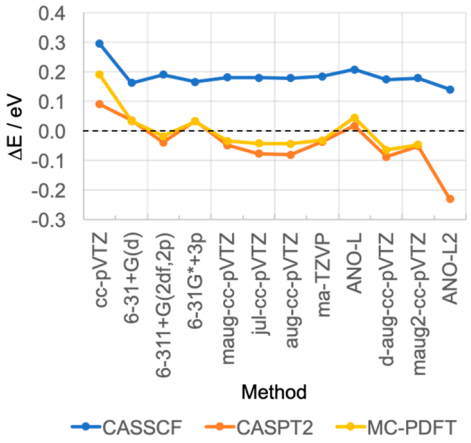


Figure 2. Difference between various theoretical vertical excitation energies (from C_{2h} calculations) and the theoretical BE (6.50 eV) of the 2^1A_g state of butadiene. MS-CASPT2 results overlap with those of CASPT2 and therefore are not shown in this figure. The reference wave function is from the SA(2)-CASSCF calculation.

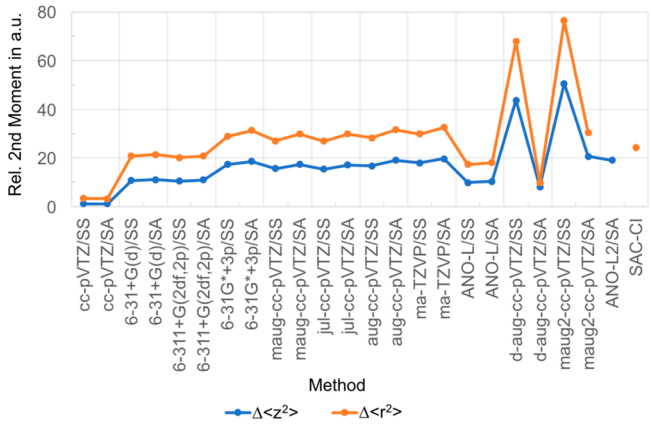


Figure 3. Relative CASSCF second moments ($\Delta\langle r^2\rangle$ and $\Delta\langle z^2\rangle$) of the 1^1B_u state of butadiene (from C_{2h} calculations) using the ground state (1^1A_g from SS-CASSCF) as reference values, as well as the SAC-CI//DZP-R/DZ $\Delta\langle r^2\rangle$ value from Saha et al.[9] Methods labeled "SS" are state-specific CASSCF calculations, and methods labeled "SA" are SA(3)-CASSCF calculations.

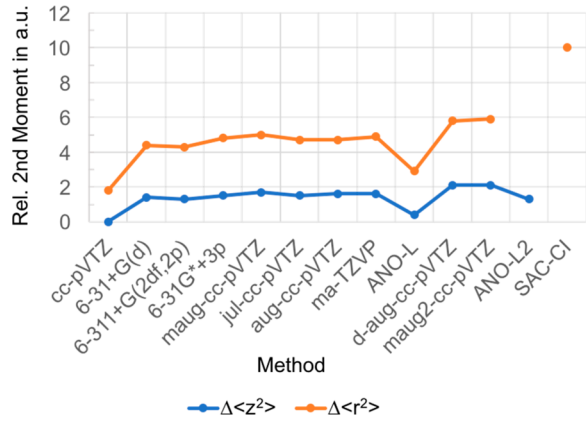


Figure 4. Relative SA(2)-CASSCF second moments ($\Delta\langle r^2\rangle$ and $\Delta\langle z^2\rangle$) of the 2^1A_g state of butadiene (from C_{2h} calculations) using the ground state (1^1A_g from SS-CASSCF) as reference values, as well as the SAC-CI//DZP-R/DZ $\Delta\langle r^2\rangle$ value from Saha et al.[9]

theoretical BE of 6.50 eV, with an absolute error within less than or equal to 0.14 eV for all basis sets tested other than group 1 and group 6 basis sets. Considering CASSCF, the absolute error is still within 0.19 eV for groups 2−4 basis sets. Group 4 basis sets, which are the best for the 1^1B_u state, also perform the best for the 2^1A_g state.

A final comment in this section concerns the differences between our results and results from Loos et al.[53] Loos's 2^1A_g state estimate of 6.50 eV is obtained on the CC3 geometry with a bond length alternation (BLA) of 0.113 Å, whereas the present work and most other work use the experimental geometry with BLA = 0.124 Å. From our calculation, Watson's CCSD(T) geometry (BLA = 0.117 Å) gives a CASPT2 vertical excitation energy 0.01 eV below that from the experimental geometry (BLA = 0.124 Å) for the basis sets that we tested except for cc-pVTZ, which is 0.02 eV below. Loos's 6.50 eV BE is from using exFCI/AVDZ corrected by CC3/AVQZ, which is a different level of theory than CASPT2. Also, Loos's CASPT2 results are from using the (4,4) active space, but we used the (4,8) active space. These differences in the methodology explain the differences of our predicted vertical excitation energies and those from Loos et al.

3.2. Rydberg Characters in the 1^1B_u and 2^1A_g States.

To examine the Rydberg characters in the 1^1B_u and 2^1A_g states, we now turn our attention to the second moments $\langle r^2\rangle$ and $\langle z^2\rangle$ (where z is the direction perpendicular to the plane of the molecule). Because the 1^1A_g state is consistently predicted to be valence-like, a convenient measure of Rydberg character is to compare the second moments of the 1^1B_u and 2^1A_g states to those of the 1^1A_g state values. These comparisons are called relative second moments and are denoted $\Delta\langle r^2\rangle$ and $\Delta\langle z^2\rangle$; they are given in Figures 3 and 4 and Table S1 for C_{2h} and Figures S3 and S4 and Table S4 for C_s.

We find that the basis set dependence of CASSCF second moments shows the same trend as the vertical excitation energy. Figures 3 and 4 and Table 3 also compare our results for the relative second moments to literature values obtained by SAC-CI, MR-CISD, and MR averaged quadratic coupled cluster (AQCC). The MR-CISD and MR-AQCC methods predict the 2^1A_g state to be a little less diffuse than our results, but in general, they agree well. Our CASSCF $\Delta\langle r^2\rangle$ and $\Delta\langle z^2\rangle$ values for the 2^1A_g state are plotted in Figures 4 and S4. Our $\Delta\langle z^2\rangle$ values are in general a little under 2 au (C_{2h}) or 5 au (C_s), while $\Delta\langle z^2\rangle_{CI}$ and $\Delta\langle z^2\rangle_{AQCC}$ are 0−1.2 au. The SAC-CI method predicts the 2^1A_g state to be more diffuse, about 5 au larger in $\Delta\langle r^2\rangle$ than our CASSCF using group-4 basis sets and C_{2h} symmetry and about the same as our SA(5)-CASSCF using group-4 basis sets and C_s symmetry. We conclude that our CASSCF second moments from group-4 basis sets agree well with SAC-CI, MR-CISD, and MR-AQCC.

The MR-CISD $\langle z^2\rangle$ for the 1^1B_u state varies greatly with changes in basis set and the nature of states included in the state-averaging.[1] The SA-CASSCF calculation that is used as the reference wave function for the MR-CISD calculation includes both ^1A_g and ^1B_u states, and MR-CISD tends to predict the 1^1B_u state to be valence-like when the states averaged include only the two valence-like states 1^1A_g and 2^1A_g and 1^1B_u. When a 2^1B_u is included or when the reference wave function is from a SS-CASSCF calculation, the 1^1B_u state is predicted by MR-CISD to have more Rydberg character. Our CASSCF calculations using the C_{2h} symmetry belong to the latter case, where the SA-CASSCF calculation for the ^1B_u states includes three ^1B_u states and the SS-CASSCF considers only one ^1B_u state. Also, we do find the 1^1B_u state to have some

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601

Rydberg character (Figure 3). When we use the C_s symmetry, the SA(3)-CASSCF calculations belong to the former case, where the states averaged include two ^1A_g states and one ^1B_u state. However, unlike MR-CISD, we predict the 1^1B_u state from SA(3)-CASSCF to have Rydberg character to a similar degree as SA(5)-CASSCF (Figure S3), which has more than one ^1B_u state except for cc-pVTZ. Because there is no true Rydberg state to artificially mix with 1^1B_u in SA(3)-CASSCF (C_s) calculations, the Rydberg character predicted in 1^1B_u is interpreted as being real.

In terms of the basis set dependency of the character of 1^1B_u, the MR-CISD/SS-CAS(4,4) calculations using the more diffuse d-aug′-cc-pVDZ (group 6) predicts $\langle z^2 \rangle$ to be ∼40 au larger than that predicted by the less diffuse maug-cc-pVDZ. This is consistent with the basis-set-dependent behavior of CASSCF in predicting the diffuseness of the 1^1B_u state, as shown in Figures 3 and S3, where group 6 basis sets generate at least 20 au larger $\langle z^2 \rangle$ than group 4 basis sets. SAC-CI also predicts some Rydberg character in the 1^1B_u state. CASSCF second moments by basis sets from groups 2, 3, and 4 agree with SAC-CI the best. Although the SAC-CI calculation used a more diffuse basis set, the method is less likely to have artificial mixing of states. Therefore, the degree of Rydberg character from our CASSCF calculations by groups 2−4 is expected to be reliable.

3.3. Basis Set Effect on the 1^1B_u and 2^1A_g States Analyzed by Decomposition of MC-PDFT Energy. To further analyze the characters of the 1^1B_u and 2^1A_g states of butadiene and to rationalize the performance of MC-PDFT as a function of basis set diffuseness, we decomposed the MC-PDFT energy into its components and plotted the energy components for the series of basis sets in Figures S5−S14. We considered the MC-PDFT components used in previous work,[18,55] as shown in

$$E_{\text{MC-PDFT}} = V_{\text{nn}} + \langle \Psi | \hat{T} + \hat{V}_{\text{ne}} | \Psi \rangle + V_{\text{C}}[\rho] + E_{\text{ot}}[\rho, \Pi] \tag{1}$$

in which we define the CASSCF contribution

$$E_{\text{CAS}} = V_{\text{nn}} + \langle \Psi | \hat{T} + \hat{V}_{\text{ne}} | \Psi \rangle + V_{\text{C}}[\rho] \tag{2}$$

the kinetic energy

$$T = \langle \Psi | \hat{T} | \Psi \rangle \tag{3}$$

the nucleus−electron interaction energy

$$E_{\text{ne}} = \langle \Psi | \hat{V}_{\text{ne}} | \Psi \rangle \tag{4}$$

and the classical electronic Coulomb energy

$$E_{\text{C}} = V_{\text{C}}[\rho] \tag{5}$$

and we further decompose the on-top energy E_{ot} into α exchange energy $E_{x\alpha}$, β exchange energy $E_{x\beta}$, and correlation energy E_{corr}, as shown in

$$E_{\text{ot}} = E_{\text{x}} + E_{\text{corr}} = E_{x\alpha} + E_{x\beta} + E_{\text{corr}} \tag{6}$$

These divisions of the on-top energy refer to the contributions of the translated terms in the on-top functional.

As illustrated in Figures S5−S14, as the basis sets become more diffuse, the changes in E_{ne} and E_{C} are most significant among all of these energy components. The absolute values of ΔE_{ne} and ΔE_{C} (where Δ denotes the E_{ne} and E_{C} values of the excited states relative to the ground state) become larger as the basis sets become more diffuse, with the exception of ANO-L. Note that the farther away the electrons are from the nuclei and from each other, the smaller the absolute values of E_{ne} and E_{C}. Because E_{ne} is negative and ΔE_{ne} is positive, a greater $|\Delta E_{\text{ne}}|$ means the excited state is more diffuse. For the classical Coulomb energy, E_{C} is positive and ΔE_{C} is negative, and the same conclusion can be drawn. Therefore, this result confirms the conclusion from analyzing the second moment values, which is that more diffuse basis sets usually predict the excited states to be more diffuse, with 1^1B_u being affected more than 2^1A_g.

The facts that $|\Delta E_{\text{ne}}|$ and $|\Delta E_{\text{C}}|$ are greater for the 1^1B_u state than those for the 2^1A_g state and that ΔT is negative for the 1^1B_u state confirm the belief that the 2^1A_g state is a covalent state and the 1^1B_u state is a valence ionic state.[15] This also explains why the 1^1B_u state is more sensitive to the diffuseness of basis sets and why its predicted vertical excitation energies are overall less accurate than those of the 2^1A_g state, as shown in the previous sections. As discussed in Boggio-Pasqua et al.,[15] when considering both the covalent state and the ionic state, a balanced treatment is needed. Therefore, Boggio-Pasqua et al.[15] constructed the 6-31G*+3p basis set to partially solve this problem. Our calculated vertical excitation energies (Figures 1, 2, S1, and S2) and MC-PDFT energy components (Figures S5−S14) show that the standard basis sets in group 4, i.e., those with an intermediate degree of diffuseness, have similar performances to those of the 6-31G*+3p basis of group 3.

The exception found in ANO-L can be explained by the fact that ANO basis sets are constructed differently from Pople-type, Dunning-type, or Ahlrichs-type basis sets. While ANO-L (group 5) seems to have diffuseness between groups 4 and 6 according to the four smallest exponential parameters of carbon p functions, it predicts excited states with characters more similar to those predicted by group 2 (or an interpolation between groups 1 and 2) than to those predicted by groups 4 or 6. Nevertheless, ANO-L2, which is judged more diffuse than ANO-L by our examination of the small exponential parameters, still predicts more diffuse excited states than ANO-L (Figures 3 and 4). Therefore, we suggest that when determining the diffuseness of basis sets by looking at exponential parameters ANO basis sets should be compared with other basis sets that belong to the ANO family and not directly compared with Pople-type, Dunning-type, or Ahlrichs-type basis sets. Because the errors in ANO-L vertical excitation energies are still roughly between those of groups 4 and 6 and no better than those of group 2 and the errors for ANO-L2 are similar to those of group 6, we still do not recommend ANO-L or ANO-L2 for the investigation of polyenes. There might exist an ANO basis set with diffuseness between ANO-L and ANO-L2 that has performance comparable to that of group 4 basis sets, but searching for such a basis set is outside of the scope of this article.

In addition, comparing the group 6 results in Figures S1, S3, S6, and S7, it is clear that having a more diffuse 1^1B_u state does not necessarily lower the theoretical vertical excitation energy of the 1^1B_u state, which in the case of CASSCF, CASPT2, and MS-CASPT2 would lower the error. This demonstrates that, although the 1^1B_u state has some Rydberg characters, increasing the diffuseness of the basis sets does not increase the accuracy of the prediction once the diffuseness reaches a certain extent for any of the CASSCF, CASPT2, MS-CASPT2, and MC-PDFT methods.

For the exchange and correlation energies from MC-PDFT, although their magnitudes are relatively small compared to E_{ne}

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601

and E_C, Figures S5, S8, S11, and S14 demonstrate that their values correlate well with CASSCF energy contributions, including ΔT, ΔE_{ne}, and ΔE_C. This is rationalized by the fact that E_{ot} depends on the density and on-top pair density calculated from the CASSCF wave function. Nevertheless, as the basis sets become more diffuse, MC-PDFT vertical excitation energies do not seem to be more susceptible to spurious results than CASPT2 or MS-CASPT2. Instead, for group 4 basis sets, MC-PDFT seems to be equally as good as or even better than MS-CASPT2.

3.4. Double Excitation Character of the 1^1B_u and 2^1A_g States. While the 1^1B_u state is believed to have predominantly single excitation character, whether the 2^1A_g state of butadiene has double excitation character is a subject under debate.[56] To address this issue, we carried out a quantitative WFA[29,30] to quantify the amount of double excitation in each state.

In the quantitative WFA, we calculated Ω,[57] which is defined as the spatial integration of the hole density ρ_H or—equivalently—the particle density ρ_E. Plasser et al.[29] derived the following formula for Ω:

$$\Omega = \mathrm{Tr}((\boldsymbol{D}^{0I})^T \boldsymbol{S}\boldsymbol{D}^{0I}\boldsymbol{S})$$

for a general basis set, where $\boldsymbol{S}$ is the overlap matrix and $\boldsymbol{D}^{0I}$ is the one-particle transition density matrix (1TDM). For an orthogonal basis, $\boldsymbol{S}$ is an identity matrix, and Ω becomes the squared Frobenius norm of the 1TDM.[29,58,59] For a general wave function, Ω is considered a measure of single excitation character.

For group 4, the SA(5)-CASSCF C_s wave function yields Ω = 0.55. This value is smaller for less diffuse basis sets (e.g., 0.44 for cc-pVTZ) and larger for more diffuse basis sets (e.g., 0.66 for maug2-cc-pVTZ), with the exception for ANO-L, which is more diffuse but has Ω = 0.47. The SA(3)-CASSCF C_s wave function gives similar results, with Ω = 0.52−0.53 for group 4, 0.42 for cc-pVTZ, 0.58 for maug2-cc-pVTZ, and 0.46 for ANO-L. The promotion number (i.e., the number of promoted electrons) p, is defined as the spatial integral over the attachment or detachment density,[29,60] and it is related to Ω by $p \approx 2 - \Omega$.[30] Therefore, having Ω = 0.55 suggests that the 2^1A_g state has neither typical double excitation character nor typical single excitation character, which is consistent with a recent analysis.[56] The 1^1B_u state has Ω = 0.9 for all basis sets tested, which confirms that it is dominated by a single excitation. The calculated p and Ω values for each basis set are summarized in Table S6.

To rationalize the single/double excitation characters of each state, we select SA(3)-CASSCF/ma-TZVP as an example to discuss the configurations of each state. We choose ma-TZVP because it overall has the smallest CASSCF, CASPT2, and MC-PDFT vertical excitation energy errors in group 4. For convenience, the orbitals in the active space are labeled from orb. 1 to orb. 8. As shown in Tables S7 and S8, the 1^1B_u state is dominated by the HOMO (orb. 2) to LUMO (orb. 3) single excitation, with a small contribution from orb. 2 to orb. 6. This explains the small deviation of Ω from 1 for the 1^1B_u state. For the 1^1B_u state, orb. 6 has the same number of nodes as orb. 3 but has less Rydberg character. This also explains the amount of Rydberg character in this state. For the 2^1A_g state, several configurations contribute non-negligibly, including a direct double excitation from orb. 2 to orb. 3 (Table S8), summing to a mixture of single and double excitations.

4. CONCLUSIONS

It is well-known that one should use well-balanced basis sets to obtain reliable results, but the balance between valence basis functions and diffuse basis functions is not easily determined. In this work, we find that one should not use basis sets that do not have diffuse functions or that have very diffuse basis functions for CASSCF and MC-PDFT calculations of the 1^1B_u state of butadiene. By using standard basis sets with intermediate diffuseness, such as maug-cc-pVTZ, jul-cc-pVTZ, aug-cc-pVTZ, and ma-TZVP, we are able to improve the CASSCF vertical excitation energies of the 1^1B_u and 2^1A_g states and their characters for butadiene. Using this improved CASSCF wave function as the reference, the vertical excitation energies by CASPT2, MS-CASPT2, and MC-PDFT are also improved. Using these intermediately diffuse basis sets, MC-PDFT vertical excitation energies of the 1^1B_u and 2^1A_g states have the same quality as MS-CASPT2, which is a great improvement upon CASSCF with much less cost than CASPT2 or MS-CASPT2. Furthermore, whereas MS-CASPT2 cannot predict the correct ordering of the 1^1B_u and 2^1A_g states, MC-PDFT can predict this ordering correctly. In short, we demonstrate that standard, relatively inexpensive computational methods are able to predict the 1^1B_u and 2^1A_g states of butadiene accurately and that one does not necessarily need to resort to constructing problem-specific basis sets or using higher-level computational methods to correctly describe the excited states of butadiene. We call the dependence of butadiene excited states on basis sets "the Goldilocks principle of basis set diffuseness", which denotes that using just the right amount of diffuseness in the basis set can greatly improve the theoretical prediction of excited states in conjugated organic molecules.

In this study, we demonstrate that a combined analysis of the orbital and state second moments is useful in determining the characters of states and assigning states. We also demonstrate that decomposition of the MC-PDFT energy can further help to confirm the characters of excited states. Our findings agree with the common belief that the 2^1A_g state is valence-like, although 2^1A_g is more diffuse than the 1^1A_g state. Although some previous studies concluded that the 1^1B_u state's valence−Rydberg mixing is artificial, our recommended basis set choices support the conclusion that the 1^1B_u state should have some degree of Rydberg character, consistent with SAC-CI results. We further show that the Rydberg character that we predict in the 1^1B_u state is not artificial. In addition, WFA suggests that the controversial 2^1A_g state has character between single and double excitations. These analyses have settled the decades of controversy of how much Rydberg character the 1^1B_u state has and how much double excitation character the 2^1A_g state has.

Overall, we recommend standard basis sets with diffuse functions on the carbon atom, such as maug-cc-pVTZ, jul-cc-pVTZ, aug-cc-pVTZ, and ma-TZVP, in the CASSCF, CASPT2, MS-CASPT2, and MC-PDFT descriptions of the 1^1B_u and 2^1A_g states of butadiene. These basis sets, whose smallest exponential parameters on the carbon p functions are approximately in the range of 0.03−0.04, give the most balanced treatment of the 1^1B_u and 2^1A_g states. More generally, to avoid spurious results in excited-state calculations, we recommend a combined analysis of the orbital second moments, state second moments, and, if necessary, MC-PDFT energy components to determine the characters of

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601

calculated ground and excited states for any molecule that has low-lying Rydberg states or where the effect of the diffuseness of basis sets has not been well-studied.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00549.

Cartesian coordinates; additional basis set information; tables of vertical excitation energies, state and orbital second moments, and oscillator strengths; figures of vertical excitation energies, second moments, and MC-PDFT energy components; absolute state energies; quantities from wave function analysis; and dominant configurations (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: gagliard@umn.edu (L.G.).
*E-mail: truhlar@umn.edu (D.G.T.).

ORCID

Sijia S. Dong: 0000-0001-8182-6522
Laura Gagliardi: 0000-0001-5227-1396
Donald G. Truhlar: 0000-0002-7742-7294

Present Address

§Materials Science Division and Center for Molecular Engineering, Argonne National Laboratory, Lemont, IL 60439, USA.

Funding

This work was supported by the National Science Foundation under grant CHE−1746186.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Dallos, M.; Lischka, H. A systematic theoretical investigation of the lowest valence-and Rydberg-excited singlet states of trans-butadiene. The character of the $1^1B_u(V)$ state revisited. *Theor. Chem. Acc.* **2004**, *112*, 16−26.
(2) Watson, M. A.; Chan, G. K.-L. Excited states of butadiene to chemical accuracy: Reconciling theory and experiment. *J. Chem. Theory Comput.* **2012**, *8*, 4013−4018.
(3) Werner, H.-J.; Knowles, P. J. An efficient internally contracted multiconfiguration−reference configuration interaction method. *J. Chem. Phys.* **1988**, *89*, 5803−5814.
(4) Knowles, P. J.; Werner, H.-J. An efficient method for the evaluation of coupling coefficients in configuration interaction calculations. *Chem. Phys. Lett.* **1988**, *145*, 514−522.
(5) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **1980**, *48*, 157−173.
(6) McDiarmid, R. An experimental estimate of Rydberg-valence mixing in conjugated dienes. *Chem. Phys. Lett.* **1992**, *188*, 423−426.
(7) Serrano-Andrés, L.; Merchán, M.; Nebot-Gil, I.; Lindh, R.; Roos, B. O. Towards an accurate molecular orbital theory for excited states: Ethene, butadiene, and hexatriene. *J. Chem. Phys.* **1993**, *98*, 3151−3162.
(8) Watts, J. D.; Gwaltney, S. R.; Bartlett, R. J. Coupled-cluster calculations of the excitation energies of ethylene, butadiene, and cyclopentadiene. *J. Chem. Phys.* **1996**, *105*, 6979−6988.
(9) Saha, B.; Ehara, M.; Nakatsuji, H. Singly and doubly excited states of butadiene, acrolein, and glyoxal: Geometries and electronic spectra. *J. Chem. Phys.* **2006**, *125*, 014316.
(10) Schalk, O.; Boguslavskiy, A. E.; Stolow, A. Two-photon excited state dynamics of dark valence, Rydberg, and superexcited states in 1, 3-butadiene. *J. Phys. Chem. Lett.* **2014**, *5*, 560−565.
(11) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* **1990**, *94*, 5483−5488.
(12) Doering, J. P.; McDiarmid, R. Electron impact study of the energy levels of trans-1, 3-butadiene: II. Detailed analysis of valence and Rydberg transitions. *J. Chem. Phys.* **1980**, *73*, 3617−3624.
(13) Widmark, P.-O.; Malmqvist, P.-Å.; Roos, B. O. Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions. *Theor. Chim. Acta.* **1990**, *77*, 291−306.
(14) Nakayama, K.; Nakano, H.; Hirao, K. Theoretical study of the $\pi \rightarrow \pi^*$ excited states of linear polyenes: The energy gap between $1^1B_u^+$ and $2^1A_g^-$ states and their character. *Int. J. Quantum Chem.* **1998**, *66*, 157−175.
(15) Boggio-Pasqua, M.; Bearpark, M. J.; Klene, M.; Robb, M. A. A computational strategy for geometry optimization of ionic and covalent excited states, applied to butadiene and hexatriene. *J. Chem. Phys.* **2004**, *120*, 7849−7860.
(16) Cave, R. J., Ab initio methods for the description of electronically excited states: survey of methods and selected results. In *Modern Electronic Structure Theory and Applications in Organic Chemistry*; Davidson, E. R., Ed.; World Scientific: Singapore, 1997; p 239.
(17) Finley, J.; Malmqvist, P.-Å.; Roos, B. O.; Serrano-Andrés, L. The multi-state CASPT2 method. *Chem. Phys. Lett.* **1998**, *288*, 299−306.
(18) Li Manni, G.; Carlson, R. K.; Luo, S.; Ma, D.; Olsen, J.; Truhlar, D. G.; Gagliardi, L. Multiconfiguration pair-density functional theory. *J. Chem. Theory Comput.* **2014**, *10*, 3669−3680.
(19) Gagliardi, L.; Truhlar, D. G.; Li Manni, G.; Carlson, R. K.; Hoyer, C. E.; Bao, J. L. Multiconfiguration pair-density functional theory: A new way to treat strongly correlated systems. *Acc. Chem. Res.* **2017**, *50*, 66−73.
(20) Dong, S. S.; Gagliardi, L.; Truhlar, D. G. Excitation spectra of retinal by multiconfiguration pair-density functional theory. *Phys. Chem. Chem. Phys.* **2018**, *20*, 7265−7276.
(21) Haugen, W.; Traetteberg, M.; et al. Molecular Structures of 1, 3-Butadiene and 1,3,5-trans-Hexatriene. *Acta Chem. Scand.* **1966**, *20*, 1726.
(22) Veryazov, V.; Widmark, P. O.; Serrano-Andrés, L.; Lindh, R.; Roos, B. O. 2MOLCAS as a development platform for quantum chemistry software. *Int. J. Quantum Chem.* **2004**, *100*, 626−635.
(23) Aquilante, F.; Autschbach, J.; Carlson, R. K.; Chibotaru, L. F.; Delcey, M. G.; De Vico, L.; Ferré, N.; Frutos, L. M.; Gagliardi, L.; Garavelli, M.; et al. Molcas 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table. *J. Comput. Chem.* **2016**, *37*, 506−541.
(24) Galván, I. F.; Vacher, M.; Alavi, A.; Angeli, C.; Autschbach, J.; Bao, J. J.; Bokarev, S. I.; Bogdanov, N. A.; Carlson, R. K.; Chibotaru, L. F.; et al. OpenMolcas: From Source Code to Insight. *Chem. Rxiv: The Preprint Server for Chemistry*; **2019**; https://chemrxiv.org/articles/OpenMolcas_From_Source_Code_to_Insight/8234021 (accessed July 2, 2019).
(25) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(26) Forsberg, N.; Malmqvist, P.-Å. Multiconfiguration perturbation theory with imaginary level shift. *Chem. Phys. Lett.* **1997**, *274*, 196−204.
(27) Ghigo, G.; Roos, B. O.; Malmqvist, P.-Å. A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2). *Chem. Phys. Lett.* **2004**, *396*, 142−149.
(28) Zobel, J. P.; Nogueira, J. J.; González, L. The IPEA dilemma in CASPT2. *Chemical science* **2017**, *8*, 1482−1499.
(29) Plasser, F.; Wormit, M.; Dreuw, A. New tools for the systematic analysis and visualization of electronic excitations. I. Formalism. *J. Chem. Phys.* **2014**, *141*, 024106.

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601

(30) Plasser, F.; Bäppler, S. A.; Wormit, M.; Dreuw, A. New tools for the systematic analysis and visualization of electronic excitations. II. Applications. *J. Chem. Phys.* **2014**, *141*, 024107.
(31) Dunning, T. H., Jr Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(32) Kendall, R. A.; Dunning, T. H., Jr; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(33) Papajak, E.; Leverentz, H. R.; Zheng, J.; Truhlar, D. G. Efficient diffuse basis sets: cc-pV*x*Z+ and maug-cc-pV*x*Z. *J. Chem. Theory Comput.* **2009**, *5*, 1197−1202.
(34) Papajak, E.; Truhlar, D. G. Convergent partially augmented basis sets for post-Hartree- Fock calculations of molecular properties and reaction barrier heights. *J. Chem. Theory Comput.* **2011**, *7*, 10−18.
(35) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(36) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057−1065.
(37) Zheng, J.; Xu, X.; Truhlar, D. G. Minimally augmented Karlsruhe basis sets. *Theor. Chem. Acc.* **2011**, *128*, 295−305.
(38) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta.* **1973**, *28*, 213−222.
(39) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+ G basis set for first-row elements, Li-F. *J. Comput. Chem.* **1983**, *4*, 294−301.
(40) Krishnan, R. B. J. S.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650−654.
(41) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265−3269.
(42) Dunning, T. H., Jr Gaussian basis functions for use in molecular calculations. I. Contraction of (9s5p) atomic basis sets for the first row atoms. *J. Chem. Phys.* **1970**, *53*, 2823−2833.
(43) Dunning, T. H., Jr.; Hay, P. J. Gaussian basis sets for molecular calculations. In *Methods of Electronic Structure Theory*, Schaefer, H. F., III, Ed.; Plenum: New York, 1977; Vol. 3, p 1.
(44) Doering, J.; McDiarmid, R. 100 eV electron impact study of 1, 3-butadiene. *J. Chem. Phys.* **1981**, *75*, 2477−2478.
(45) Mosher, O. A.; Flicker, W. M.; Kuppermann, A. Electronic spectroscopy of s-trans 1, 3-butadiene by electron impact. *J. Chem. Phys.* **1973**, *59*, 6502−6511.
(46) Mosher, O. A.; Flicker, W. M.; Kuppermann, A. Triplet states in 1, 3-butadiene. *Chem. Phys. Lett.* **1973**, *19*, 332−333.
(47) McDiarmid, R. On the ultraviolet spectrum of trans-1, 3-butadiene. *J. Chem. Phys.* **1976**, *64*, 514−521.
(48) Flicker, W. M.; Mosher, O. A.; Kuppermann, A. Electron-impact investigation of excited singlet states in 1, 3-butadiene. *Chem. Phys.* **1978**, *30*, 307−314.
(49) Daday, C.; Smart, S.; Booth, G. H.; Alavi, A.; Filippi, C. Full configuration interaction excitations of ethene and butadiene: Resolution of an ancient question. *J. Chem. Theory Comput.* **2012**, *8*, 4441−4451.
(50) Rabidoux, S. M.; Cave, R. J.; Stanton, J. F. A Non-Adiabatic Investigation of the Electronic Spectroscopy of trans-1, 3-Butadiene. *J. Phys. Chem. A* **2019**, *123*, 3255−3271.
(51) Celani, P.; Bernardi, F.; Olivucci, M.; Robb, M. A. Excited state reaction pathways for s cis buta 1, 3 diene. *J. Chem. Phys.* **1995**, *102*, 5733−5742.
(52) Garavelli, M.; Celani, P.; Bernardi, F.; Robb, M. A.; Olivucci, M. Force fields for "ultrafast" photochemistry: The S2 (1Bu)→ S1 (2Ag)→ S0 (1Ag) reaction path for all-trans-Hexa-1, 3, 5-triene. *J. Am. Chem. Soc.* **1997**, *119*, 11487−11494.
(53) Loos, P.-F.; Boggio-Pasqua, M.; Scemama, A.; Caffarel, M.; Jacquemin, D. Reference energies for double excitations. *J. Chem. Theory Comput.* **2019**, *15*, 1939−1956.
(54) Brundle, C. R.; Robin, M. B. Nonplanarity in hexafluorobutadiene as revealed by photoelectron and optical spectroscopy. *J. Am. Chem. Soc.* **1970**, *92*, 5550−5555.
(55) Sharma, P.; Bernales, V.; Truhlar, D. G.; Gagliardi, L. Valence $\pi\pi^*$ excitations in benzene studied by multiconfiguration pair-density functional theory. *J. Phys. Chem. Lett.* **2019**, *10*, 75−81.
(56) Shu, Y.; Truhlar, D. G. Doubly excited character or static correlation of the reference state in the controversial 2^1A_g State of *trans*-butadiene? *J. Am. Chem. Soc.* **2017**, *139*, 13770−13778.
(57) Plasser, F.; Lischka, H. Analysis of excitonic and charge transfer interactions from quantum chemical calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2777−2789.
(58) Feng, X.; Luzanov, A. V.; Krylov, A. I. Fission of entangled spins: An electronic structure perspective. *J. Phys. Chem. Lett.* **2013**, *4*, 3845−3852.
(59) Skolnik, J. T.; Mazziotti, D. A. Cumulant reduced density matrices as measures of statistical dependence and entanglement between electronic quantum domains with application to photosynthetic light harvesting. *Phys. Rev. A: At., Mol., Opt. Phys.* **2013**, *88*, 032517.
(60) Head-Gordon, M.; Grana, A. M.; Maurice, D.; White, C. A. Analysis of electronic transitions as the difference of electron attachment and detachment densities. *J. Phys. Chem.* **1995**, *99*, 14261−14270.
(61) Chien, A. D.; Holmes, A. A.; Otten, M.; Umrigar, C. J.; Sharma, S.; Zimmerman, P. M. Excited states of methylene, polyenes, and ozone from heat-bath configuration interaction. *J. Phys. Chem. A* **2018**, *122*, 2714−2722.

DOI: 10.1021/acs.jctc.9b00549
J. Chem. Theory Comput. 2019, 15, 4591−4601