

Multi-state extrapolation of UV/Vis absorption spectra with QM/QM hybrid methods

Sijin Ren, and Marco Caricato

Citation: *The Journal of Chemical Physics* **144**, 184102 (2016);

View online: <https://doi.org/10.1063/1.4948471>

View Table of Contents: <http://aip.scitation.org/toc/jcp/144/18>

Published by the [American Institute of Physics](#)

Articles you may be interested in

[Toward reliable density functional methods without adjustable parameters: The PBE0 model](#)

The Journal of Chemical Physics **110**, 6158 (1999); 10.1063/1.478522

[Point charge embedding for ONIOM excited states calculations](#)

The Journal of Chemical Physics **145**, 224109 (2016); 10.1063/1.4972000

[Long-range interactions from the many-pair expansion: A different avenue to dispersion in DFT](#)

The Journal of Chemical Physics **146**, 024111 (2017); 10.1063/1.4973728

[Efficient algorithm for multiconfiguration pair-density functional theory with application to the heterolytic dissociation energy of ferrocene](#)

The Journal of Chemical Physics **146**, 034101 (2017); 10.1063/1.4973709

[Assessment of low-scaling approximations to the equation of motion coupled-cluster singles and doubles equations](#)

The Journal of Chemical Physics **141**, 164116 (2014); 10.1063/1.4898709

[Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange](#)

The Journal of Chemical Physics **119**, 2943 (2003); 10.1063/1.1590951



The banner features a dark blue background with a grid pattern. On the left is a circular logo containing a molecular structure with green, blue, and orange atoms. On the right are three circular icons: a top one with a red and blue contour plot, a middle one with a red and blue contour plot, and a bottom one with a colorful grid pattern. The text 'JCP Communications' is centered in white, with a 'Read Now!' button below it.

JCP Communications

[Read Now!](#)

Multi-state extrapolation of UV/Vis absorption spectra with QM/QM hybrid methods

Sijin Ren and Marco Caricato^{a)}

Department of Chemistry, University of Kansas, 1251 Wescoe Hall Dr., Lawrence, Kansas 66045, USA

(Received 19 February 2016; accepted 20 April 2016; published online 9 May 2016)

In this work, we present a simple approach to simulate absorption spectra from hybrid QM/QM calculations. The goal is to obtain reliable spectra for compounds that are too large to be treated efficiently at a high level of theory. The present approach is based on the extrapolation of the entire absorption spectrum obtained by individual subcalculations. Our program locates the main spectral features in each subcalculation, e.g., band peaks and shoulders, and fits them to Gaussian functions. Each Gaussian is then extrapolated with a formula similar to that of ONIOM (Our own N-layered Integrated molecular Orbital molecular Mechanics). However, information about individual excitations is not necessary so that difficult state-matching across subcalculations is avoided. This multi-state extrapolation thus requires relatively low implementation effort while affording maximum flexibility in the choice of methods to be combined in the hybrid approach. The test calculations show the efficacy and robustness of this methodology in reproducing the spectrum computed for the entire molecule at a high level of theory. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4948471>]

I. INTRODUCTION

Theoretical simulations of UV/Vis absorption spectra of large chromophores have become an important tool for the design of new compounds in materials and renewable energy science. For instance, it is important to develop compounds that present a large absorption in the visible region where the sun emits most photons so that large amounts of solar energy can be harvested.^{1,2} The photochemical characteristics of chromophores and dyes can be tuned by proper choices of substituent groups. Hence, accurate theoretical simulations of such effects largely benefit the design process by screening promising candidates.³ The most successful tool to perform such calculations is time-dependent density functional theory (TD-DFT), since it provides the best compromise between accuracy and computational effort. However, an important limitation of DFT is that the approximate functionals available perform differently for different compounds, and a sensible choice for each case is often difficult.^{3,4} Thus, much effort has been devoted to the benchmarking and calibration of approximate functionals.⁵⁻⁹ On the other hand, more reliable methods may simply not be affordable in practical situations.

A possible strategy to overcome this difficulty is to use multi-scale or hybrid techniques. In general, the goal of multi-scale methods is to reduce the computational cost by partitioning the system into regions, where the core is treated at a high level of theory while the rest is treated at a lower and less computationally demanding level. This reduction in cost should be obtained without a significant loss in accuracy compared to the calculation on the entire system at the high level of theory. An example of such partitioning is shown

in Figure 1, where “model” refers to the core region, and “real” refers to the entire system. Such approaches have been used extensively in biochemical applications where, for instance, the reactive pocket of a protein is described quantum mechanically (QM) while the rest is treated classically. These methods can also combine two QM methods, QM/QM, and they have been used in many applications.¹⁰⁻¹⁶ The use of QM/QM methods has also been extended in recent times to excited state calculations.¹⁷⁻²⁴

One of the most successful hybrid methods is ONIOM (Our own N-layered Integrated molecular Orbital molecular Mechanics),²⁵⁻³⁰ which is briefly described in Sec. II. The main feature of ONIOM is that the combined energy is expressed as an extrapolation of the energy computed for the real and model systems, see Figure 1. Thus, if no electronic embedding schemes are used,³¹⁻³⁴ these subcalculations are independent and the method requires no specific implementation. This also permits maximum flexibility in the choice of the high and low levels of theory. We showed that ONIOM can accurately reproduce excitation energies and transition properties computed at a high level of theory at a fraction of the computational cost.³⁵⁻³⁷ However, a drawback of this approach is that its success depends on the correct matching of states between the subcalculations. This is relatively easy for low-lying, isolated, and bright transitions, but it becomes a significant issue if one is interested in various regions of the spectrum.

In this work, we propose a proof-of-concept approach to extrapolate multiple states at once or, in other words, to extrapolate the entire spectrum within an energy range. The extrapolation formula is borrowed from ONIOM, but no state-matching is required. Therefore, this approach maintains attractive ONIOM features (e.g., simplicity, favorable cost/accuracy ratio) while avoiding any connection

^{a)}Electronic mail: mcaricato@ku.edu

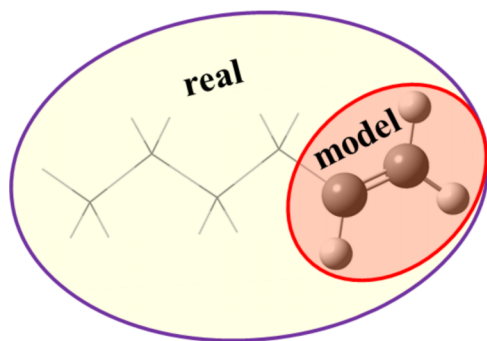


FIG. 1. A sample system partitioning scheme for two layered ONIOM in which the entire system is labeled as real, and the core system is labeled as model.

to a particular description of individual excited states. Perfect state-matching would indeed provide a direct interpretation of the extrapolated spectrum in terms of characteristic transition types (e.g., $\pi \rightarrow \pi^*$). Unfortunately, a typical band is composed of a relatively large number of excitations with varying oscillator strength, especially in higher energy regions. This often makes a definite assignment difficult even with standard methods, and it renders the use of QM/QM hybrid methods like ONIOM virtually impossible. The reason is that states that are close in energy often present various degrees of orbital mixing that are different across levels of theory and model vs. real systems. It is this mixing that prevents a clear state-matching, thus precluding the extrapolation in the higher energy regions. Our method bypasses this limitation by extrapolating spectral features (i.e., bands and shoulders) rather than individual states. However, once the multi-state extrapolation is performed, a qualitative analysis of the types of transitions that form a particular band can always be obtained by examination of the corresponding bands of the subcalculations that were used for the extrapolation. An important application of this multistate extrapolation is the design of new dyes for solar energy harvesting. As mentioned above, a critical goal in this field is the tuning of the substituents on the main chromophoric moiety for maximum absorption of sunlight. Our method may allow to explore a large number of options by treating the core moiety at a high level of theory and the substituents at a lower level without sacrificing accuracy. These calculations would provide valuable screening information of promising candidates for sunlight harvesting dyes, which could then be synthesized and tested in actual devices. From a computational perspective, our method may permit, for instance, the use of global functionals as low level methods without worrying about the unphysical low-lying excitations with small oscillator strength (typical of these functionals), since these will not contribute to the overall spectrum. The computer program that we developed to perform the extrapolation can be downloaded from our group web site,³⁸ and it can read the necessary information for the extrapolation (i.e., excitation energies and related oscillator strength) from a simple text file. As shown by our test calculations, this proof-of-concept approach is quite robust and provides spectra that are in very good agreement with the corresponding high level calculations on the entire system.

The paper is organized as follows. Section II reviews the ONIOM method and discusses the multi-state extrapolation. Details of the calculations are presented in Section III, while test calculations are described in Section IV. Section V reports an overall discussion of the results and final remarks.

II. THEORY

In this work, we propose to simulate UV-Vis spectra of large chromophores through a multi-state extrapolation of spectra obtained by combining two QM levels of theory, following a strategy similar to the ONIOM method.^{25–30,35–37} Hence, let us briefly review ONIOM for two layers and introduce the notation that we will use throughout the paper. Following the layer separation in Figure 1, open valencies resulting from the severing of covalent bonds are capped with link atoms (usually hydrogens), e.g., $X-Y \rightarrow X-H$. The link atom bond length is based on the ratio between the current $X-Y$ length and its typical value. The energy extrapolation is performed combining the energies from separate subcalculations,

$$E_{ext} = E_{real}^{low} + E_{model}^{high} - E_{model}^{low}, \quad (1)$$

where *high* and *low* refer to the levels of theory, and *model* and *real* to the system partition as in Figure 1. If no electronic embedding is introduced, the three subcalculations are truly independent. The same extrapolation can also be used to compute excitation energies,

$$\omega_{ext} = \omega_{real}^{low} + \omega_{model}^{high} - \omega_{model}^{low}. \quad (2)$$

Finally, transition properties are evaluated by extrapolation of transition moments with formulas equivalent to Eqs. (1) and (2).³⁷ In previous studies,^{35–37} we showed that excitation energies and properties can be accurately evaluated with ONIOM with great savings in computational cost, and that the use of link atoms does not affect the results when the model system is sensibly chosen. Despite these promising results and the simplicity of the extrapolation in Eq. (2), a straightforward application of ONIOM to compute excitation energies is difficult. The issue resides in the proper matching of states across subcalculations. Indeed, a clear classification of an excitation is often difficult, and clear-cut cases such as very bright $\pi \rightarrow \pi^*$ excitations with large oscillator strength (f) are not the norm. On the contrary, matching based on oscillator strength, energy ordering, or even orbital contribution can be ambiguous due to considerable differences between levels of theory or between model and real systems.

Thus, here we propose to abandon individual states and attempt to directly extrapolate entire absorption bands. In particular, we concentrate on three key parameters for the definition of an extrapolated band i : position ω_i^{ext} , height ϵ_i^{ext} , and half-width σ_i^{ext} . These parameters, collectively called $\{P^{ext}\}$, are obtained through an ONIOM-type extrapolation from individual subcalculations as

$$P^{ext} = P_{real}^{low} + P_{model}^{high} - P_{model}^{low}. \quad (3)$$

Once the P^{ext} s are computed with Eq. (3), a Gaussian envelope is assigned to the extrapolated band as

$$\epsilon^{ext}(\omega) = \epsilon_i^{ext} \exp \left[- \left(\frac{\omega - \omega_i^{ext}}{\sigma_i^{ext}} \right)^2 \right]. \quad (4)$$

The final extrapolated spectrum is then built from the sum of the individual extrapolated bands. In this approach, the spectral bands from each subcalculation may include any number of excited states with different oscillator strengths. Thus, the issue of individual state-matching is transformed into the issue of finding and matching a relatively small number of bands from each subcalculation (more precisely, finding and matching the band parameters $\{P^{sub}\}$). Although defining and matching bands may still be ambiguous in certain cases, as we shall see in Section IV, it is still a much simpler and better defined problem than matching a large number of states individually. We note that information about the nature of the states that mostly contribute to a band in the extrapolated spectrum can always be recovered from an analysis of the subcalculations spectra. This qualitative analysis can help in the characterization of the band contributions while still avoiding the cumbersome state-matching in the extrapolation.

The first step is to simulate the spectrum for each subcalculation. This is achieved by assigning a Gaussian function to each excitation, which is proportional to the extinction coefficient through³⁹

$$\epsilon(\omega) = \frac{D_i \omega_i}{4 \times 2.926 \times 10^{-39} \sqrt{\pi} \sigma} \exp \left[- \left(\frac{\omega - \omega_i}{\sigma} \right)^2 \right], \quad (5)$$

where ω_i is the excitation energy in eV, σ is the half-width at ϵ_{max}/e (we set $\sigma = 0.4$ eV), and D_i is the dipole strength (in cgs units), which is related to the oscillator strength (f_i) by

$$D_i = \frac{f_i}{3.7922 \times 10^{33} \omega_i}. \quad (6)$$

The final spectrum is obtained by summing all of the Gaussian functions. Note that this choice of fitting is not unique, and others can be selected without altering the procedure for obtaining $\{P^{sub}\}$ described below.

Once the full spectrum for a subcalculation is obtained, our program performs a scan to find each peak, and their position ω_i^{sub} and height ϵ_i^{sub} are stored. Then, σ_i^{sub} is taken for each peak as the half-width at the height of ϵ_i^{sub}/e . If two bands are too close to each other, an alternative value of σ is taken at $2\epsilon_i^{sub}/e$. A particularly delicate point is the treatment of band shoulders. These cannot be well represented by a single Gaussian function for the main band peak, so a different Gaussian is assigned to each shoulder. The detection of shoulders in the subcalculation bands is performed with a second scan after the primary Gaussian parameters for the peaks have been assigned. The shoulder detection scheme is shown in Figure 2. The shoulder detection is performed on both the right- and the left-hand sides of each band peak. As shown in Figure 2 for a right-hand search, starting from position 1 the difference between the assigned Gaussian and the actual spectrum (ΔX) is monitored. If $\Delta X > \alpha$, where α is a parameter, the presence of a shoulder is detected. At this point, the height of the shoulder is obtained by scanning

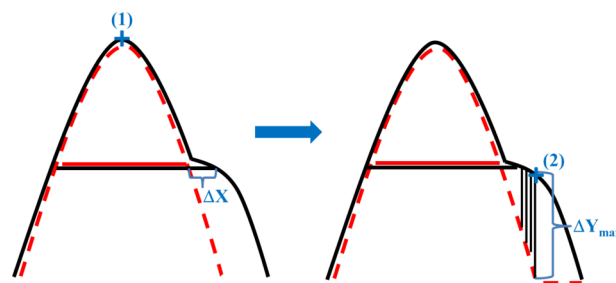


FIG. 2. A schematic representation of shoulder detection. The dotted red curve represents the fitted Gaussian assigned to the mean peak (1). ΔX is the width difference between the fitted Gaussian and the actual spectrum, and it is used to assign the presence of a shoulder (2). ΔY is the height of the shoulder at (2).

the difference in height between the actual spectrum and the Gaussian function assigned to the main peak (ΔY in Figure 2). The position for the shoulder Gaussian is chosen at the largest value of ΔY , i.e., point 2 in Figure 2. We use $\alpha = 0.1$ in this work. The band width parameter (σ) of the new Gaussian is taken as the width difference between the peak Gaussian and the actual band at shoulder height/ e . However, since a shoulder is described by a single Gaussian that is very close to the main peak, the direct sum of the two Gaussian functions (one for the peak and one for the shoulder) would overestimate the height for both peak and shoulder. Therefore, we implemented an iterative optimization of both height and width of the two Gaussians so that their sum matches the original height (ϵ_1, ϵ_2) and half-width (σ_1, σ_2) parameters.

Once the main features (peak/shoulder) of each subcalculation spectrum are found and a Gaussian function is assigned to each of them (i.e., $\{P^{sub}\}$), the extrapolated Gaussian functions are obtained using Eqs. (3) and (4). In our program, we can extrapolate a subset of the bands, and a sanity check is implemented so that the number of requested bands is equal or smaller than those found in the subcalculations. The complete implementation steps are outlined in Figure 3.

Fitting each band as one Gaussian

1. Form separate spectra for sub-calculations (rl, mh, ml)
2. Locate peaks (position and height) on each spectra
3. Locate band width parameter for each peak
4. Fit each band into a Gaussian
5. Check for existence of shoulder under each peak
6. Add Gaussian to describe shoulder
7. Locate Gaussian (position, width, and height)
8. Optimize the heights of shoulder and its main peak iteratively

Spectra Extrapolation

1. Extrapolate each of the parameters (P = position, height, width)
2. Sum all extrapolated Gaussians to form an extrapolated spectra

FIG. 3. Scheme of the steps involved in the spectra extrapolation.

III. COMPUTATIONAL DETAILS

All calculations were performed with a development version of the GAUSSIAN suite of programs.⁴⁰ The geometries were optimized at the CAM-B3LYP⁴¹/6-311++G** level of theory and are available in the supplementary material.⁴² The same level is used for the *target* and *model/high* excited state calculations using the usual linear response approach.⁴³⁻⁴⁵ The *target* spectra, i.e., the spectra obtained considering the entire molecule and the target level of theory, are used as reference for the comparison of the extrapolated and subcalculation spectra. We chose CAM-B3LYP because it has shown the ability to provide a balanced description of excitations of different nature (e.g., valence and Rydberg), thanks to its range separation.⁵⁻⁷ In the layer definition, we follow the standard ONIOM approach to use hydrogen link atoms to cap severed covalent bonds. We test two methods as *low* level of theory: CIS⁴⁶ and B3LYP⁴⁷⁻⁴⁹ with the 6-31+G* basis set. However, we report only the CIS results in the main text since the B3LYP results are qualitatively the same and can be found in the supplementary material.⁴²

IV. RESULTS

We apply our method on seven test molecules shown in Figure 4: 1-hexene (**1**), 1,3-decadiene (**2**), 2-nonenal (**3**), 1-cyanoazulene (**4**), 1-isocyanoazulene (**5**), betaine-30 (**6**), and a methylamino-carbonyl-substituted betaine (**7**). The size of the main chromophoric moiety is different across the test set so that we can evaluate the performance of our method in different conditions. In the following, the model system is shown in a ball-and-stick representation, while the other layer is shown in a tube-frame representation. The figures also show all spectra (i.e., target, subcalculations, and extrapolated) overlaid on each other for a direct comparison of the band positions, heights, and widths. Insets also show the target and subcalculation spectra individually, where each excited state is represented as a stick with length proportional to the oscillator strength. In this way, we are able to visualize the number and the relative importance of the individual excitations that really contribute to each band.

The results for (**1**) are shown in Figure 5. In this case, the chromophore is the alkene group. We consider ten states for the target and each subcalculation. As shown by the stick spectra

in the insets of the figure, only a few states provide significant contribution to the final spectrum. Additionally, a shoulder in the target calculation is present to the right of the main peak (i.e., at higher energy). The same band is found as individual peaks in the subcalculations. The extrapolation is able to correctly handle these differences so that the extrapolated spectrum is in excellent agreement with the target. Indeed, the position, height, and width of both peak and shoulder of the target spectrum are better reproduced by the extrapolated spectrum than by any of the subcalculations.

For (**2**), shown in Figure 6, two conjugated double bonds form the main chromophore. 100 states are used to build the spectrum. There is only one main peak within this subset of states, which is mainly due to one $\pi \rightarrow \pi^*$ excitation. However, a large number of states with small oscillator strength are also present that form a small tail in the 6.5-7 eV region. The extrapolated spectrum is able to describe both the large peak and the tail with excellent accuracy. Figure 6 clearly shows that all subcalculations individually do not reproduce the main peak in position or height. For this case in particular, the advantage of extrapolating bands rather than individual states is particularly evident for the tail part of the spectrum where many states with small oscillator strength contribute.

For (**3**), shown in Figure 7, we consider the conjugated carbonyl and alkene groups as the chromophore. Considering 50 states produces three spectral features: one intense peak at lower energy, and two moderate peaks. The extrapolated spectrum provides the best match for the first, intense band both in terms of position and intensity. The two small bands, on the other hand, are merged together into a tail of reasonable intensity. This is due to the fact that the two shallow peaks strongly overlap, and the two Gaussians assigned to them merge together into one. It is not surprising that close, shallow peaks are problematic for our approach. Nonetheless, we find the simulation of this region acceptable at this stage, especially compared to that of the individual subcalculations, see Figure 7. One way to improve the description of these cases may be to extend the iterative optimization of the shoulders described in Section II to the entire spectrum in a recursive manner.

Test cases (**4**) and (**5**) are two isomers: 1-cyanoazulene and 1-isocyanoazulene.⁵⁰ The results for these molecules are shown in Figure 8. They represent an unusual choice of model system since the core layer (azulene) is larger than the second

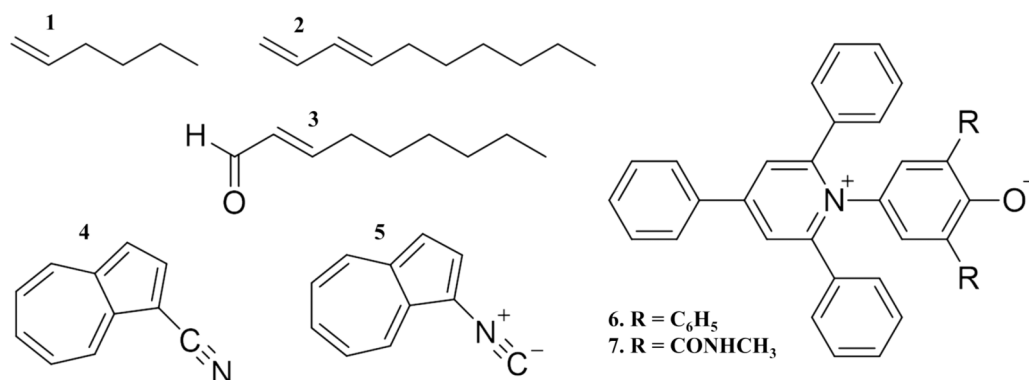


FIG. 4. Test molecules used in evaluating the method.

CAM-B3LYP/6-311++G** : CIS/6-31+G*

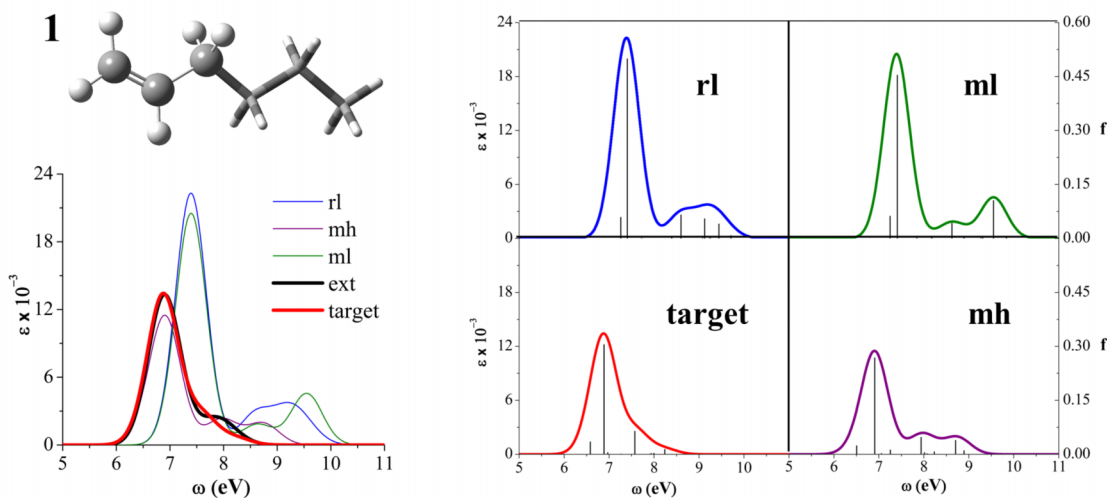


FIG. 5. Absorption spectra for 1-hexene (**1**). The ball-and-stick representation describes the model system, while the tube-frame representation indicates the rest of the molecule. The insets report the subcalculations and target spectra with the stick representation of individual excitations. rl: real/low, ml: model/low, mh: model/high, ext: extrapolated.

layer (substituent). However, they are interesting because they allow us to study the effect of different substituents on the absorption of the main chromophore. Indeed, azulene is a particularly versatile compound whose photochemical characteristics can be finely tuned by substitution in any of the eight possible positions.⁵⁰ Here we want to probe how our approach handles relatively small differences due to similar substituents. We considered 100 states for the target and each subcalculation to simulate the spectrum. As shown in the insets in Figure 8, there are a handful of states with large oscillator strength, but the band structure in the 7-8 eV and 9-10 eV regions is due to a large number of states with moderate value of oscillator strength. The main effect of the substituents is to add a small band at low energy (<4 eV), and to modify the higher energy bands in the 6-10 eV region, whereas the main peak at 5 eV is clearly due to azulene since it is qualitatively similar in both molecules and across subcalculations. From

a comparison of the *target* insets of Figure 8, the effect of the different substituents in (**4**) and (**5**) is evident in the 7-9 eV region. We also point out that the small differences between the model subcalculations in the two molecules are due to the slightly different link atom bond lengths (1.025 and 1.080 Å, respectively) induced by the different substituents. The extrapolated spectrum for the cyanoazulene is in very good agreement with the target for the main peak as well as for the higher energy region. The extrapolated data improves the agreement both in terms of position and intensity of the bands compared to the subcalculations. The agreement for the isocyano isomer is excellent in the energy region below 5 eV, but it is not quite as good in the higher energy region. In the 6-8 eV region, the first peak is slightly underestimated while the second is slightly overestimated. However, this description is overall closer to the target spectrum than any of the subcalculations. Moreover, the extrapolation is able to

CAM-B3LYP/6-311++G** : CIS/6-31+G*

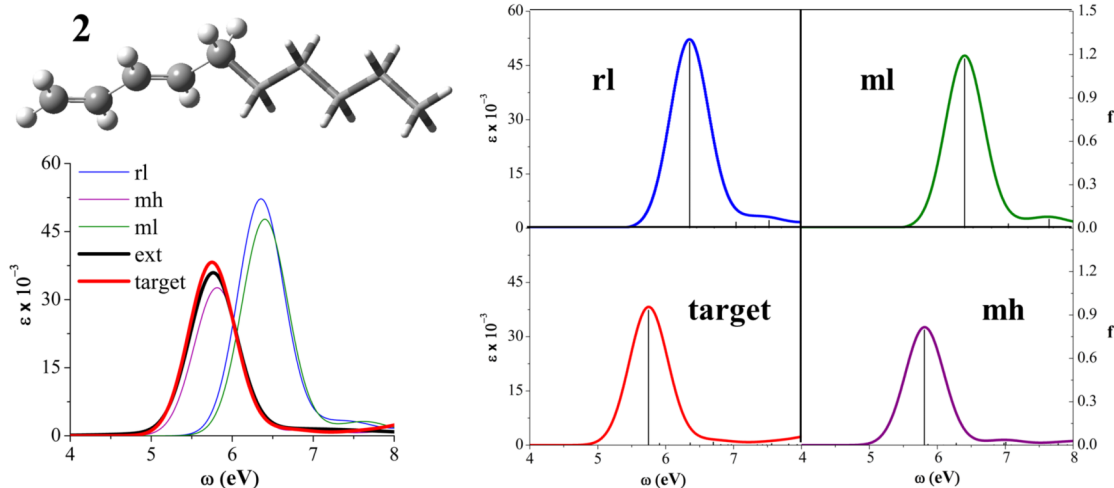


FIG. 6. Absorption spectra for 1,3-decadiene (**2**). The ball-and-stick representation describes the model system, while the tube-frame representation indicates the rest of the molecule. The insets report the subcalculations and target spectra with the stick representation of individual excitations. rl: real/low, ml: model/low, mh: model/high, ext: extrapolated.

CAM-B3LYP/6-311++G** : CIS/6-31+G*

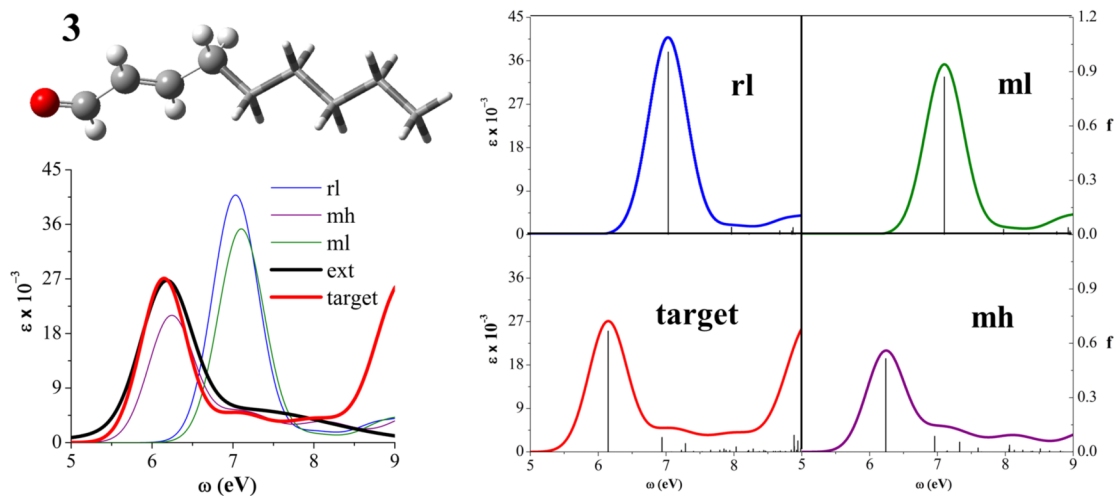
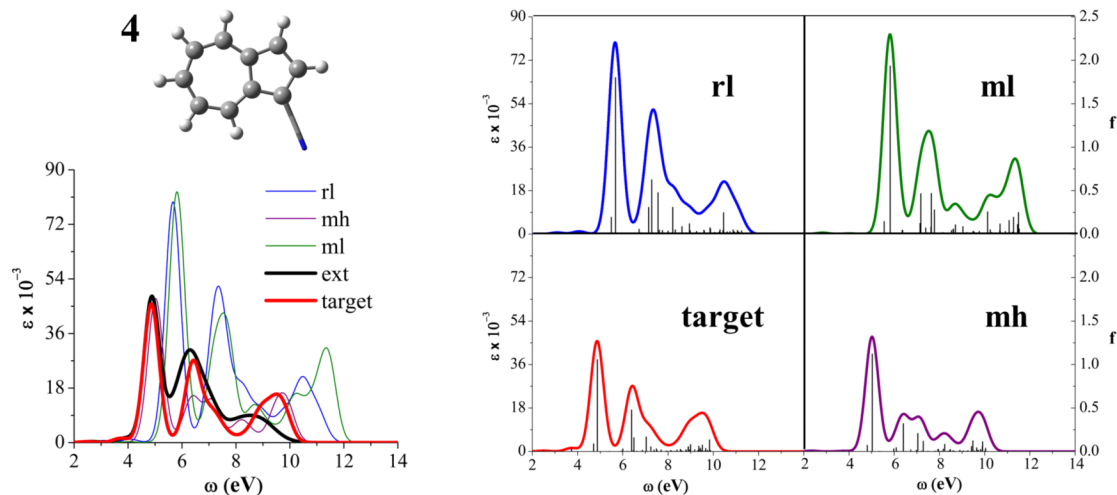


FIG. 7. Absorption spectra for 2-nonenal (**3**). The ball-and-stick representation describes the model system, while the tube-frame representation indicates the rest of the molecule. The insets report the subcalculations and target spectra with the stick representation of individual excitations. rl: real/low, ml: model/low, mh: model/high, ext: extrapolated.

CAM-B3LYP/6-311++G** : CIS/6-31+G*



CAM-B3LYP/6-311++G** : CIS/6-31+G*

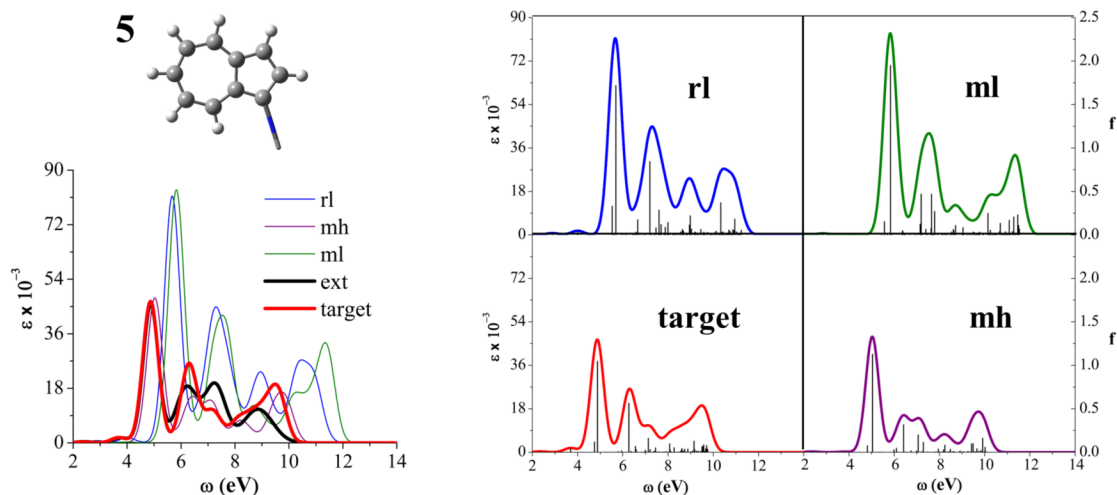


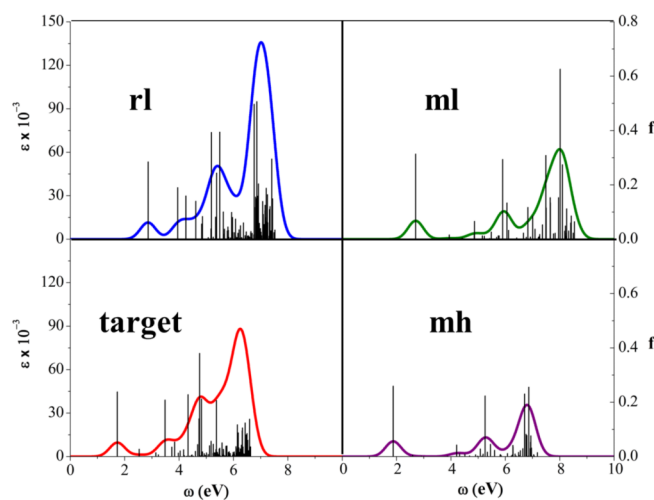
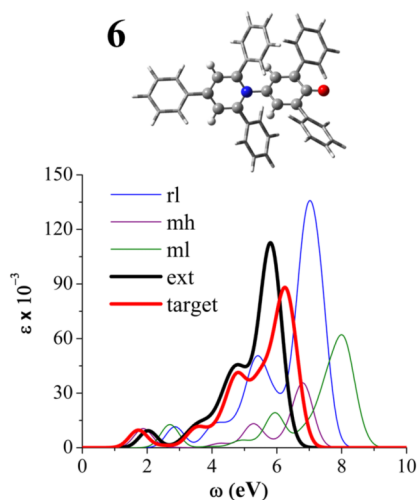
FIG. 8. Absorption spectra for 1-cyanoazulene (**4**, top) and 1-isocyanoazulene (**5**, bottom). The ball-and-stick representation describes the model system, while the tube-frame representation indicates the rest of the molecule. The insets report the subcalculations and target spectra with the stick representation of individual excitations. rl: real/low, ml: model/low, mh: model/high, ext: extrapolated.

capture the differences between the spectra of compounds (**4**) and (**5**) qualitatively and quantitatively.

Betaine-30 (**6**) is a large conjugated molecule, which has been used to define a polarity scale for solvents due to the sensitivity of its absorption spectrum to the environment.⁵¹ The layer separation and simulated spectra are shown in Figure 9 (top). This is a particularly challenging case because there is no clear way to define a layer separation. The model system shown in the figure is directly conjugated, but weaker conjugation with the phenyl substituents may be important for the absorption spectrum. Additionally, the phenyl substituents are chromophores themselves, and they may contribute to the spectrum independently. We considered 100 states for the simulation of the spectrum, which results in a peak below 2 eV, a peak/shoulder around 3.5 eV, a peak around 5 eV, and a peak around 6 eV. The latter two are the result of many states with moderate oscillator strength. The model system calculations are able to reproduce the low energy peak, but the remaining parts of the spectrum are shifted at higher energies compared to the target calculation, which indicates that the effect of

the substituents is significant. The first band is reproduced well by the extrapolation despite the overestimation of the low level subcalculations. The second and third features (at 3.5 and 5 eV, respectively) are in excellent agreement with the target and show considerable improvement with respect to the subcalculations. The last band is slightly overestimated in intensity and underestimated in position, which is due to the increasing contributions of the substituents. Nonetheless, the agreement with the target is remarkable considering the performance of the subcalculations. This case shows the robustness of the extrapolation approach even when the core layer definition is pushed towards the limit of acceptability. Compound (**7**) is a variant of betaine 30 where two of the phenyl rings are substituted with a methylamino-carbonyl group, see Figure 9 (bottom). The substitution considerably changes the target spectrum both in terms of position and intensity of the bands. For instance, the shoulder at 3.5 eV disappears, and the last peak is considerably less intense. We used the same model system as for (**6**) so that the model subcalculations are basically the same as before. The

CAM-B3LYP/6-311++G** : CIS/6-31+G*



CAM-B3LYP/6-311++G** : CIS/6-31+G*

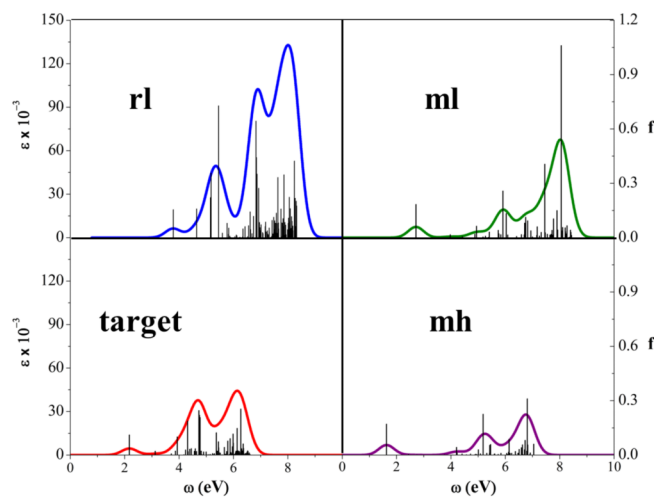
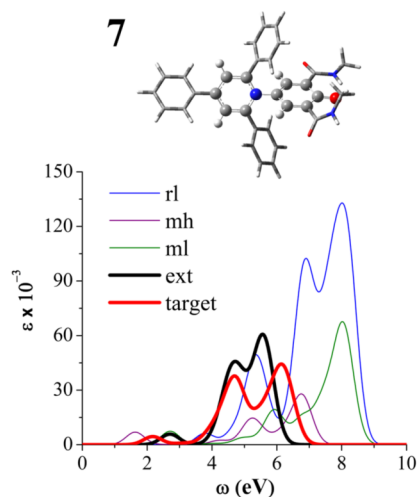


FIG. 9. Absorption spectra for betaine-30 (**6**) and a substituted betaine (**7**). The ball-and-stick representation describes the model system, while the tube-frame representation indicates the rest of the molecule. The insets report the subcalculations and target spectra with the stick representation of individual excitations. rl: real/low, ml: model/low, mh: model/high, ext: extrapolated.

extrapolated spectrum is able to reproduce these modifications, providing a performance that is comparable to that for betaine 30: the first peak is slightly shifted towards higher energy, the second peak is well reproduced, and the last peak is shifted towards lower energy and overestimated in intensity. As before, the extrapolated spectrum is in better qualitative and quantitative agreement with the target one than any of the subcalculations.

V. DISCUSSION AND CONCLUSIONS

In this work, we present an extrapolation approach for the simulation of absorption spectra with hybrid QM/QM methods. The extrapolation is based on the ONIOM formula, but instead of considering individual states, we extrapolate entire bands. Thus, we shift the problem of matching individual states, which is easy only for few low-lying bright states, to that of matching spectral features that are fewer in number. One of the key points is to recognize these features. This is easy for peaks but less so for shoulders, and in Section II we discuss how we propose to handle this issue. The results in Section IV, where the method is applied to a number of test cases, show the ability of the band-extrapolation idea to reproduce spectra computed at higher levels of theory on the entire molecule.

The main issue for the success of the method is the proper layer separation. This is not specific to our case, but it is common to all multi-layer methods. In previous work, we suggested a number of guidelines for the proper choice of the model system for excited state calculations,^{35–37} and the same guidelines apply here. An issue that is specific to our method is the overlap between bands. Since we assign a Gaussian function to each spectral feature with extrapolated values of position, height, and width, overlap between adjacent Gaussian functions may lead to intensity overestimation in the extrapolated band. This is particularly problematic for shoulders and their relative main peak. To solve, or at least alleviate this problem, we suggest an iterative procedure for the definition of the optimal choice of these parameters within each subcalculation. The final parameters are then extrapolated in the usual way, see Eq. (3). This provides excellent results for test molecules (1), (2), and (4). However, band overlap can also happen between peaks, and we have not addressed this issue yet. This results in overestimations of certain bands, e.g., in molecule (3), although the overall shape of the spectrum is still well described. A promising approach to overcome this problem is to extend the iterative process used for shoulders to the entire spectrum recursively, but we leave this development for future work. Finally, test molecules (6) and (7) indicate that the extrapolation method is robust even when the layer separation is not optimal.

The range of applicability of this method is connected to the choice of model system. Since the latter is meant to be the main chromophoric moiety in the molecule, its excitations are likely to be concentrated in the low energy region of the spectrum. Thus, this is the region where the extrapolation method will work best. At higher energies, excitations from the second layer will contribute significantly to the overall spectrum, and the extrapolation may fail. However, this is not

a strong limitation since practical applications in materials and energy research are focused on the low energy region. Therefore, we expect that this extrapolation approach will be useful, for instance, in the design of efficient dyes where the absorption spectrum is tuned by different ligands.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Award No. EPS-0903806 and matching support from the State of Kansas through the Kansas Board of Regents.

- ¹A. Hagfeldt, G. Boschloo, L. Sun, L. Kloo, and H. Pettersson, *Chem. Rev.* **110**, 6595 (2010).
- ²B. E. Hardin, H. J. Snaith, and M. D. McGehee, *Nat. Photonics* **6**, 162 (2012).
- ³M. Pastore, S. Fantacci, and F. De Angelis, *J. Phys. Chem. C* **117**, 3685 (2013).
- ⁴P. Dev, S. Agrawal, and N. J. English, *J. Chem. Phys.* **136**, 224301 (2012).
- ⁵M. Caricato, G. W. Trucks, M. J. Frisch, and K. B. Wiberg, *J. Chem. Theory Comput.* **6**, 370 (2010).
- ⁶M. Caricato, G. W. Trucks, M. J. Frisch, and K. B. Wiberg, *J. Chem. Theory Comput.* **7**, 456 (2011).
- ⁷D. Jacquemin, A. Planchat, C. Adamo, and B. Mennucci, *J. Chem. Theory Comput.* **8**, 2359 (2012).
- ⁸D. Jacquemin, I. Duchemin, and X. Blase, *J. Chem. Theory Comput.* **11**, 5340 (2015).
- ⁹D. Jacquemin and C. Adamo, "Computational Molecular Electronic Spectroscopy with TD-DFT," in *Density-Functional Methods for Excited States* (Springer International Publishing, Cham, 2016), pp. 347–375.
- ¹⁰C. Huang, M. Pavone, and E. A. Carter, *J. Chem. Phys.* **134**, 154110 (2011).
- ¹¹C. Huang and E. A. Carter, *J. Chem. Phys.* **135**, 194104 (2011).
- ¹²F. Libisch, C. Huang, and E. A. Carter, *Acc. Chem. Res.* **47**, 2768 (2014).
- ¹³J. D. Goodpaster, T. A. Barnes, and T. F. Miller, *J. Chem. Phys.* **134**, 164108 (2011).
- ¹⁴J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, *J. Chem. Phys.* **140**, 18A507 (2014).
- ¹⁵T. A. Barnes, J. D. Goodpaster, F. R. Manby, and T. F. Miller III, *J. Chem. Phys.* **139**, 024103 (2013).
- ¹⁶M. E. Fornace, J. Lee, K. Miyamoto, F. R. Manby, and T. F. Miller III, *J. Chem. Theory Comput.* **11**, 568 (2015).
- ¹⁷C. Daday, C. König, O. Valsson, J. Neugebauer, and C. Filippi, *J. Chem. Theory Comput.* **9**, 2355 (2013).
- ¹⁸J. Neugebauer and E. J. Baerends, *J. Phys. Chem. A* **110**, 8786 (2006).
- ¹⁹A. S. P. Gomes and C. R. Jacob, *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **108**, 222 (2012).
- ²⁰S. Höfener, A. S. P. Gomes, and L. Visscher, *J. Chem. Phys.* **139**, 104106 (2013).
- ²¹S. Höfener, A. Severo Pereira Gomes, and L. Visscher, *J. Chem. Phys.* **136**, 044104 (2012).
- ²²S. Höfener and L. Visscher, *J. Chem. Phys.* **137**, 204120 (2012).
- ²³Y. G. Khait and M. R. Hoffmann, *J. Chem. Phys.* **133**, 044107 (2010).
- ²⁴T. A. Wesolowski, *J. Chem. Phys.* **140**, 18A530 (2014).
- ²⁵M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *J. Phys. Chem.* **100**, 19357 (1996).
- ²⁶T. Vreven and K. Morokuma, *J. Chem. Phys.* **111** (1999).
- ²⁷T. Vreven, K. S. Byun, I. Komaromi, S. Dapprich, J. A. Montgomery, Jr., K. Morokuma, and M. J. Frisch, *J. Chem. Theory Comput.* **2**, 815 (2006).
- ²⁸K. Morokuma, Q. Wang, and T. Vreven, *J. Chem. Theory Comput.* **2**, 1317 (2006).
- ²⁹L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding, and K. Morokuma, *Chem. Rev.* **115**, 5678 (2015).
- ³⁰H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch, and T. Vreven, *J. Chem. Phys.* **128**, 034107 (2008).
- ³¹K. V. Jovan Jose and K. Raghavachari, *J. Chem. Theory Comput.* **10**, 4351 (2014).
- ³²H. P. Hratchian, A. V. Krukau, P. V. Parandekar, M. J. Frisch, and K. Raghavachari, *J. Chem. Phys.* **135**, 014105 (2011).
- ³³N. J. Mayhall, K. Raghavachari, and H. P. Hratchian, *J. Chem. Phys.* **132**, 114107 (2010).

- ³⁴N. J. Mayhall and K. Raghavachari, *J. Chem. Theory Comput.* **6**, 3131 (2010).
- ³⁵M. Caricato, T. Vreven, G. W. Trucks, M. J. Frisch, and K. B. Wiberg, *J. Chem. Phys.* **131**, 134105 (2009).
- ³⁶M. Caricato, T. Vreven, G. W. Trucks, and M. J. Frisch, *J. Chem. Phys.* **133** (2010).
- ³⁷M. Caricato, T. Vreven, G. W. Trucks, and M. J. Frisch, *J. Chem. Theory Comput.* **7**, 180 (2011).
- ³⁸See <https://caricatogroup.ku.edu/software> for our multi-state extrapolation code; accessed 19 February 2016.
- ³⁹N. Harada, S.-M. L. Chen, and K. Nakanishi, *J. Am. Chem. Soc.* **97**, 5345 (1975).
- ⁴⁰M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, P. V. Parandekar, N. J. Mayhall, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian Development Version, Revision H.09+, Gaussian, Inc., Wallingford, CT, 2010.
- ⁴¹T. Yanai, D. P. Tew, and N. C. Handy, *Chem. Phys. Lett.* **393**, 51 (2004).
- ⁴²See supplementary material at <http://dx.doi.org/10.1063/1.4948471> for optimized geometries of test molecules, and for results obtained with B3LYP as low level method.
- ⁴³R. Bauernschmitt and R. Ahlrichs, *Chem. Phys. Lett.* **256**, 454 (1996).
- ⁴⁴M. E. Casida, C. Jamorski, K. C. Casida, and D. R. Salahub, *J. Chem. Phys.* **108**, 4439 (1998).
- ⁴⁵E. R. Stratmann, G. E. Scuseria, and M. J. Frisch, *J. Chem. Phys.* **109**, 8218 (1998).
- ⁴⁶J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, *J. Phys. Chem.* **96**, 135 (1992).
- ⁴⁷A. D. Becke, *J. Chem. Phys.* **98**, 1372 (1993).
- ⁴⁸A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- ⁴⁹C. T. Lee, W. T. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ⁵⁰R. E. Robinson, T. C. Holovics, S. F. Deplazes, D. R. Powell, G. H. Lushington, W. H. Thompson, and M. V. Barybin, *Organometallics* **24**, 2386 (2005).
- ⁵¹C. Reichardt and T. Welton, *Solvents and Solvent Effects in Organic Chemistry*, 4th ed. (Wiley-VCH, 2011).