From simple molecules to nanotubes. Reliable predictions of ionization potentials from the $\Delta$MP2-SCS methods

Szymon Śmiga$^1$, Sylwia Siecińska and Ireneusz Grabowski$^2$

Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Grudziadzka 5, 87-100 Toruń, Poland

$^1$ Author to whom any correspondence should be addressed.

E-mail: szsmiga@fizyka.umk.pl

Keywords: ionization potential, $\Delta$MP2 method, DNA/RNA bases, nanotubes, spin-component-scaled, diagonal, frequency-independent approximation

Abstract

The vertical ionization potentials for systems of various sizes, ranging from simple molecules, DNA/RNA bases, donor and acceptor organic molecular systems as well as nanotubes are calculated using the $\Delta$MP2-SCS family of methods [Śmiga et al 2018 J. Chem. Theory Comput. 14 4780–90]. We have shown that for all investigated cases, the $\Delta$MP2-SCS methods, being almost cost free single-step post-Hartree–Fock calculation, provide very accurate vertical ionization potentials comparable in quality with state-of-art outer valence Green’s function methods. Moreover, we show that a combination of the $\Delta$MP2-SCS methods with the resolution of identity technique is effective and reliable an alternative to the semi-empirical density functional theory methods and Green’s function-based calculations of ionization potentials for large molecular systems such as silicon-based or organic-based solar cells, for which the IP-EOM-CCSD calculations are too expensive for routine calculations.

1. Introduction

A photo-voltaic effect is a physical process allowing to convert light into electricity in solar cell devices. The effect mostly occurs in the semiconducting materials characterized by small energy difference (the energy gap) between the valence and the conduction bands or between the highest occupied electron energy and lowest unoccupied electron energy levels, respectively.

One of the key parameters of solar cell devices is the power conversion efficiency (PCE) ratio [1] which gives the information about the ability to convert light into electricity. The PCE ratio is proportional to open-circuit voltage ($V_{oc}$) which, as was recently shown [2], is in turn governed by energy difference between the value of first electron attachment (EA) (or negative lowest unoccupied molecular orbital energy (LUMO) $E_{A} = -\varepsilon^{LUMO}_{A}$) and first ionization potential (IP) (or negative highest occupied molecular orbital (HOMO) energy $E_{D} = -\varepsilon^{HOMO}_{D}$) of the acceptor and donor material, respectively. However, according to reference [3], in order to identify potential candidates for photovoltaic materials, one basically needs only to know the value of their first IP.

The IP of a molecule is also a fundamental quantity in other contexts such as charge transfer and transport processes, photo-electron spectroscopy and especially molecular electronics, where together with EA attracts great interest in applied and theoretical research [1, 4–6]. Thus, the ability of accurate prediction (with the accuracy in range of 0.1–0.3 eV) of these two intrinsic parameters has become a key importance in the context of a new developments in these field [7–13].

To date, many useful approaches have been established allowing to calculate EAs and IPs with various accuracy and cost [11, 14]. Unfortunately, the size of the molecules used e.g. in the construction of silicon-based or organic-based solar cells [15–18] limits the application of first-principle $ab$ initio methods in modeling of such structures. The utilization of state-of-art IP/EA equation-of-motion coupled-cluster (IP/EA-EOM-CC) [19–21] or electron propagator (EP) theory [22–47] methods (providing high accuracy
results) become especially difficult, particularly when one needs to perform calculations in larger basis sets. On the other hand, application of simple and relatively cheap density functional theory (DFT) methods often do not lead to accurate predictions. For example, reference [49] reports that standard DFT error. In contrast, the application of simple and relatively cheap density functional theory (DFT) leads to accurate predictions of IPs for large molecular systems due to delocalization effects. As shown in figure 1 of reference [11], the results become especially difficult, particularly when one needs to perform calculations in larger basis sets. The optimized self-energy terms have been optimized with respect to VIPs benchmark data set of fifty variously sized molecules (for more details see reference [11]). The optimized self-energy terms justify the construction of one parameter variant which is computationally more efficient.

In the above, indices $i, j$ indicate occupied orbitals, $a, b$ indicate unoccupied ones ($p, q$ indices will be used for any orbitals), while $\sigma$ and $\tau$ are spin indices.

The $c_{SS}$ and $c_{OS}$ parameters scaling the same spin (SS) and opposite spin (OS) parts of the second-order self-energy, respectively, have been optimized with respect to VIPs benchmark data set of fifty variously sized molecules (for more details see reference [11]). The optimized $c_{SS}$ and $c_{OS}$ parameter values, having strictly property (i.e. IP) based nature, define herein the ΔMP2-SCS(IP) ($c_{SS} = 0.67$ and $c_{OS} = 0.38$) and ΔMP2-SOS(IP) ($c_{OS} = 1.0$ and $c_{SS} = 0.0$) methods. It was shown [11] that both can predict the values of VIP with accuracy comparable with the state-of-art EP methods such as outer valence Green’s function (OVGF) [63].

We note that, similarly to the construction of other EP methods [64, 65] (e.g. OVGF methods), the introduction of scaling parameters in equation (2) is justified based on the following empirical observations: (1) it is generally known that the first, non-vanishing, second-order self-energy term ($\Sigma^{(2)}_{pp}$) (ΔMP2/EP2 method) leads to a large overestimation of Koopmans orbital energies. Addition of the third order $\Sigma^{(3)}_{pp}$ term contribution (ΔMP3/EP3 method), in turn, leads the opposite trend [64, 66]; (2) the experimental IPs values quite often lie between second-order and third-order ΔMP/EP values. Instead of rescaling third-order term (like in OVGF methods), one can disregard computationally demanding $\Sigma^{(3)}_{pp}$ term and scale only the second-order one [11, 56, 57, 67]. As shown in figure 1 of reference [11], the rescaling of $\Sigma^{(2)}_{pp}$ self-energy term leads to a significant reduction of overestimation in ΔMP2 method. Furthermore, the SCS decomposition allows to exploit mutual, almost linear relation between OS and SS self-energy terms which justify the construction of one parameter variant which is computationally more favorable.

The methods and results presented in reference [11] are encouraging, however, in order to demonstrate the real predictive power and robustness of ΔMP2-SCS(IP) and ΔMP2-SOS(IP) approaches, larger assessment is required. Therefore, in the present study, we have employed both ΔMP2-SCS methods for the calculation of VIP for variously sized systems ranging from simple molecules to large organic photo-voltaic molecules and single-wall carbon nanotubes (SWCNT) and have made the most stringent performance analysis of our new methods providing a thorough benchmark of them in comparison with other approaches.
2. Computational details

All EP method calculations (OVGF, P3, P3+, EP3, EP2) [66] have been performed with the Gaussian 16 package. [68] The IPs with low pole strength (PS) (lower than 0.85) have been ignored in subsequent analysis. The IP-EOM calculations have been performed with the NWChem software [69]. If not otherwise stated the ΔMP2 and the spin-resolved ΔMP2 methods [ΔMP2-SCS(IP) and ΔMP2-SOS(IP)] results have been obtained with our local version of the ACESII package [70]. In the case of large molecular systems reported in sections 3.3 and 3.4 all ΔMP2 calculations have been performed with locally modified version of PSI4 quantum chemistry packages [71] where the resolution of identity (RI) [72, 73] was employed in the calculation of electron repulsion integrals. As was shown in references [74, 75] the utilization of RI approximation allows to substantially reduce the computational bottleneck of EP methods, i.e. atomic- to molecular-orbital transformation scaling as $O(N^3)$, without significant loss of accuracy. If not otherwise stated, all calculations have been performed using Dunning [76] cc-pVXZ family of basis sets (X = D, T). In the case of RI approximation, the auxiliary basis have been chosen as a counterpart of original basis set [77]. Likewise in other studies [11, 54, 58, 78], the CBS results reported in tables 2–4 have been estimated using the two point inverse cubic formula [79]. For more details we refer the reader to section 3.

3. Results

3.1. Vertical valence ionization potentials

In our previous study (see table 6 in reference [11]) we have shown that ΔMP2-SCS(IP) and ΔMP2-SOS(IP) methods are able to provide accurate results not only for the first IP but also for all valence IPs. However, the conclusion was based on a few simple molecules, and more broader assessment in this context should be made. In the recent study [9] Ranasinghe et al have presented a reliable benchmark data set of 1468 valence VIPs computed for 155 molecules using IP-EOM method with singles and doubles [19, 80] (IP-EOM-CCSD) in cc-pVTZ basis set. The data set, consisting of two subsets denoted as CGB and NIST was compiled from molecules taken from references [81] and [82], respectively. We will refer to these names in the following.

In order to assess the methods on the same footing (lack of basis set error compensation), we have performed the ΔMP2-SCS(IP) and ΔMP2-SOS(IP) calculations using the same computational setup as in reference [9], namely cc-pVTZ basis sets and geometries. Additionally, for comparison, we also computed the same quantities for several EP approaches.

In table 1 we report MSE and MAE calculated with respect to IP-EOM-CCSD/cc-pVTZ data for all methods mentioned above. We remark, that in the evaluation of errors we have considered only the states where all methods yield well converged results (PS ≥ 0.85).

Inspection of the data reveals that in the case of valence IPs, the ΔMP2-SCS(IP) and ΔMP2-SOS(IP) methods yield MAEs comparable with much more sophisticated and numerically demanding OVGF approaches. In the case of CGB data set, the smallest error is observed for P3+ method (MAE = 0.16 eV). In the same time all OVGF methods give MAEs in the range of 0.19 to 0.27 eV. The bare EP2 and ΔMP2 methods (without any parametrization) overestimate significantly the IPs whereas the application of EP3 method leads to the opposite trend. Similar trends are observed for NIST subset. Here the best performance is provided by OVGF-N method yielding MAE = 0.17 eV. This is closely followed by P3 (MAE = 0.18 eV) and P3+ (MAE = 0.2 eV) results and other OVGF methods which give the MAEs in the range of 0.19 to 0.23 eV. We remark that ΔMP2-SCS(IP) and ΔMP2-SOS(IP) give, in this case, only slightly worse performance with the MAEs which are only 0.05 ~ 0.07 eV larger than those for OVGF-N method. To
conclude the valence IPs assessment, we report (in parenthesis) the MAEs computed for all states reported in reference [9] for ΔMP2-SCS(IP) and ΔMP2-SCS(IP) methods. This step allows to directly compare the performance of both IP-optimized ΔMP2-SCS methods with the results obtained for CGB subset gathered in table 2 in reference [9]. One can note that both methods yield MAEs which are almost twice as good as those obtained using QTP [83] family of DFT functionals, which were specially designed and parametrized to provide best performance for VIPs. Moreover we note that both methods outperform the CCSD results obtained from extended Koopmans theorem [84–86], which surprisingly gives, in this case, MAE = 0.8 eV, probably mainly due to the lack of inclusion of higher excitations [87].

For the first IPs (HOMO column in table 1) the performance of the methods is quite similar for both subsets. The ΔMP2-SOS(IP) yield for both sets best performance MAE of 0.16 eV and 0.15 eV for CBG and NIST, respectively. These are closely followed by ΔMP2-SCS(IP) and P3+- methods results giving only slightly larger MAE of about 0.01 ~ 0.04 eV. Again the CGB HOMO IPs can be directly compared to those gathered in table 2 in reference [9]. One can note that both ΔMP2-SCS(IP) and ΔMP2-SOS(IP) outperform significantly IPs obtained using standard DFT methods and they are almost twice as good as IPs calculated by QTP [83] functionals.

Summarizing this part, we can conclude that the broad assessment conducted has proven that at a small fraction of computational cost (see table 1 in reference [58]), the ΔMP2-SCS(IP) and ΔMP2-SOS(IP) methods provide results which are inline with more advanced EP approaches such as OVGF or P3/P3+ for all valence VIPs and are much better than standard and more sophisticated DFT functionals [9].

### 3.2. Ionization potentials of nucleic acid bases

Another interesting application of the ΔMP2-SCS methods was performed for five DNA/RNA nucleobases (depicted in figure 1), namely, guanine, adenine, cytosine, thymine, and uracil. As discussed in references [90–92], the accurate determination of the ionization energies of these molecules plays an important role toward a better understanding of electrostatic interaction of nucleobases with other molecules, or e.g. phenomena related photo damage or repair of the genetic materials [92]. The initial DNA/RNA geometries have been taken from reference [88] and re-optimized at the B3LYP/6-31G(2df, p) [93–95] level of theory. To enable a broad comparison, also in this case, we have performed calculation for all methods listed in table 1 at the CBS limit (see section 2 for more details). Unfortunately, the direct comparison with experimental results is rather difficult due to large dispersion of data [91, 92]. Thus as a reference we have
Table 2. First two VIPs (in eV) of nucleic acid bases obtained for various methods at CBS limit. First column reports experimental IP values from reference [89] whereas the second one, the same values but including ZPE and geometry relaxation corrections. In the last two lines we report the MAE and MARE calculated with respect to IP-EOM-CCSD/CBS data. The best results are marked bold.

|        | Expt. \(^a\) | Expt.-C \(^b\) | IP-EOM \(^c\) | EP2 | EP3 | P3 | P3+ | OVGF | ΔMP2 |
|--------|--------------|----------------|---------------|-----|-----|----|-----|------|------|
|        |              |                |               | N   | A   | B  | C   | -SOS(IP) | -SCS(IP) |
| Adenine| 8.26         | 8.45           | 8.39          | 8.16 | 8.72 | 8.62 | 8.39 | 8.20 | 8.19 | 8.39 | 8.14 | 8.36 | 8.27 |
|        | 9.41         | 9.13           | 9.79          | 9.82 | 9.69 | 9.41 | 9.38 | 9.41 | 9.58 | 8.96 | 9.43 | 9.46 |
| Cytosine| 8.68         | 8.83           | 8.86          | 8.52 | 9.00 | 8.68 | 8.69 | 8.68 | 8.85 | 8.41 | 8.90 | 8.83 |
|        | 9.58         | 9.08           | 9.77          | 9.75 | 9.64 | 9.42 | 9.40 | 9.42 | 9.58 | 8.86 | 9.52 | 9.56 |
| Guanine| 7.77         | 7.99           | 8.08          | 7.75 | 8.14 | 8.36 | 8.25 | 8.08 | 7.85 | 7.82 | 8.02 | 7.68 | 8.02 | 7.95 |
|        | 9.84         | 9.44           | 10.08         | 10.13| 10.00| 9.69 | 9.70 | 9.69 | 9.88 | 9.25 | 9.81 | 9.82 |
| Thyamine| 8.87         | 9.15           | 9.25          | 9.50 | 9.41 | 8.95 | 9.04 | 8.95 | 9.16 | 8.86 | 9.34 | 9.24 |
|        | 10.28        | 10.02          | 10.97         | 10.83| 10.69| 10.57| 10.52| 10.57| 10.70| 9.71 | 10.36| 10.64|
| Uracil | 9.32         | 9.58           | 9.66          | 9.39 | 9.82 | 9.38 | 9.47 | 9.38 | 9.58 | 9.30 | 9.73 | 9.65 |
|        | 10.37        | 10.12          | 11.06         | 10.92| 10.79| 10.67| 10.61| 10.67| 10.80| 9.82 | 10.64| 10.73|
|        | MAE (eV)     | 0.32           | 0.25          | 0.33 | 0.22 | 0.17 | 0.18 | 0.21 | 0.13 | 0.47 | 0.10 | 0.11 |
|        | MARE (%)     | 3.38           | 2.54          | 3.52 | 2.30 | 1.71 | 1.93 | 2.26 | 1.33 | 5.00 | 0.98 | 1.15 |

\(^a\) Adiabatic IP data measured by photo ionization mass-spectrometry in gas phase from reference [89].
\(^b\) Experimental data from reference [89] corrected for ZPE and geometry relaxation corrections calculated at B3LYP/6-31G(2df, p) C = Δrelax + ZPEc − ZPEn; note, that same correction was also applied for second IP (see reference [54]).
\(^c\) IP-EOM-CCSD results, all electron are correlated.
Table 3. First IPs (in eV) obtained for several donor and acceptor molecules [54, 97, 98] and methods. If not otherwise noted, the calculations are reported at CBS [79] limit obtained from cc-pVDZ and cc-pVTZ data. The MAE and MARE are calculated with respect to experimental molecular data. Core electrons have been kept frozen in all calculations. The best results are marked bold.

|        | OVGF | ∆MP2 |      |      |      |      |      |
|--------|------|------|------|------|------|------|------|
|        | Expt. | EP3  | P3   | P3+  | N    | A    | B    | C    | --- | -SOS(IP) | -SCS(IP) |
| **H2Pc** | 6.41a | 6.17 | 6.89 | 6.79 | 6.27 | 5.92 | 5.96 | 6.27 | 6.66 | 6.04     | 6.20     |
| **Pentacene** | 6.61a | 6.52 | 6.78 | 6.66 | 6.42 | 6.15 | 6.20 | 6.43 | 6.27 | 6.24     | 6.42     |
| **Benzo[a]pyrene** | 7.12a | 7.15 | 7.39 | 7.27 | 7.03 | 6.77 | 6.81 | 7.03 | 6.83 | 6.99     | 6.85     |
| **MgOEP** | 6.19a | 6.01 | 6.36 | 6.24 | 5.90 | 5.65 | 5.76 | 5.90 | 6.28 | 6.02     | 5.96     |
| **DCV3T** | 6.09a | 7.79 | 8.24 | 8.07 | 7.77 | 7.53 | 7.34 | 7.67 | 7.41 | 7.68     | 7.84     |
| **DIP** | 5.35a | 6.99 | 7.25 | 7.10 | 6.83 | 6.56 | 6.59 | 6.83 | 6.36 | 6.64     | 6.79     |
| **6T** | 4.7a  | 6.82 | 7.12 | 6.97 | 6.68 | 6.46 | 6.41 | 6.68 | 6.39 | 6.57     | 6.77     |
| **Rubrene** | 6.52a | 6.41c | 6.52c | 6.32c | 5.91c | 5.86c | 5.91c | 6.12c | 5.52 (5.96)f | 5.89 (6.16)f | 6.02 (6.35)f |
| **C60** | 7.57a | 8.42c | 8.26d | 7.91c | 7.60c | 7.41c | 7.60c | 7.79c | 6.73 (7.03)f | 7.25 (7.43)f | 7.30 (7.53)f |
| **C70** | 7.47a | 7.84d | 7.76d | 7.51d | 7.26d | 7.17d | 7.26d | 7.40d | 6.50 (6.92)f | 7.06 (7.27)f | 7.08 (7.37)f |
| **PCBM** | 7.17a | 7.59d | 7.46d | 7.22d | 6.99d | 6.90d | 6.99d | 7.14d | 6.18 (6.66)f | 6.82 (7.05)f | 6.87 (7.17)f |

|        | MAEe |      |      |      |      |      |      |      | 0.27 | 0.33 | 0.21 | 0.16 | 0.37 | 0.31 | **0.14** | 0.60 (0.39)f | 0.34 (0.23)f | 0.29 (0.15)f |
|        | MARE (%) |      |      |      |      |      |      |      | 3.72 | 5.08 | 3.11 | 2.41 | 5.52 | 4.62 | **2.12** | 8.48 (5.60)f | 5.03 (3.46)f | 4.28 (2.28)f |

---

* experimental values from reference [97–100] and reference therein.
  
*bulk experimental values from reference [97] and reference therein.
  
*results obtained with cc-pVDZ [76] basis set.
  
*results obtained with 6-31G [95, 101] basis set.
  
*errors are calculated only for (underline) systems where molecular experimental values were available.
  
*CBS results.
Table 4. VIPs for various methods obtained at CBS limit for
benzo[a]pyrene, C60 and MgOEP molecular systems. The last lines
report MAE and MARE calculated with respect to experimental data
(expt.) The best results are marked bold.

|          | Expt. a | CN/23b | CP3/23b | CP3+/23b | — | SOS(IP) | SCS(IP) |
|----------|---------|--------|---------|----------|---|---------|---------|
| Benzo[a]pyrene |        |        |         |          |   |         |         |
| Π1       | 7.12    | 7.12   | 7.34    | 7.24     | 6.83 | 6.99    | 6.85    |
| Π2       | 8.05    | 8.05   | 8.28    | 8.17     | 7.68 | 7.85    | 7.80    |
| Π3       | 8.79    | 8.78   | 9.00    | 8.89     | 8.34 | 8.66    | 8.65    |
| Π4       | 8.97    | 9.00   | 9.25    | 9.09     | 8.36 | 8.87    | 8.84    |
| Π5       | 9.51    | 9.62   | 9.78    | 9.66     | 8.99 | 9.43    | 9.50    |
| C60      |         |        |         |          |   |         |         |
| Hυ       | 7.57    | 7.82   | 8.48    | 8.13     | 7.03 | 7.53    | 7.43    |
| HG       | 8.95    | 9.32   | 9.86    | 9.47     | 8.13 | 8.81    | 8.81    |
| MgOEP    |         |        |         |          |   |         |         |
| A1       | 6.19    | 6.26   | 6.72    | 6.60     | 6.28 | 6.02    | 5.96    |
| A2       | 6.57    | 6.45   | 6.89    | 6.73     | 6.05 | 6.28    | 6.24    |
| B2       | 7.42    | 7.87   | 8.00    | 7.73     | 6.39 | 7.49    | 7.45    |
| E        | 7.63    | 8.00   | 8.10    | 7.83     | 6.46 | 7.56    | 7.55    |
| A2       | 8.30    | 8.22   | 8.32    | 8.05     | 6.68 | 7.79    | 7.78    |
| MAE      | 0.16    | 0.41   | 0.25    | 0.67     | 0.16 | 0.19    | 0.19    |
| MARE (%) | 1.99    | 5.34   | 3.28    | 8.36     | 2.09 | 2.48    | 2.48    |

a experimental data from reference [98–100].
b see reference [54] for the description of methods and results. The
OVGF-N, P3, and P3+ cc-pVDZ results have been corrected using EP2
cc-pVDZ and cc-pVTZ results extrapolated to CBS.

utilized IP-EOM-CCSD [19, 21, 80] results extrapolated to CBS limit as in references [78, 79] which yield
quite good agreement with other IP-EOM calculations [90, 96].

In table 2 we present first two VIPs for aforementioned DNA and RNA nucleobases together with MAE
and mean average relative error (MARE) calculated with respect to reference IP-EOM-CCSD data. First one
can note that IP-EOM-CCSD reference results (also from other studies [90, 96]) are in good agreement with
experimental data from reference [89] corrected for zero-point energy (ZPE) and geometry relaxation
corrections calculated at B3LYP/6-31G(2df, p) level. This finding additionally confirms the quality of
IP-EOM-CCSD data. Further inspection of table 2 reveals the excellent performance of ΔMP2-SOS(IP) and
ΔMP2-SCS(IP) methods which yield in this case relatively small errors of MAE = 0.10 eV
(MARE = 0.98%) and MAE = 0.11 eV (MARE = 1.15%), respectively. Second best result is provided by
OVGF-C method with MAE of 0.13 eV and MARE = 1.33%. Similar good performance is observed for
other OVGF and P3+ method which give MAE in the range of 0.17 to 0.22 eV. Surprisingly, the
performance of the P3 method, in contrast to the results reported in section 3.1 a n d o t h e r w o r k related to
VIP of nucleobases [90, 92], is rather poor and comparable with EP2 method, which gives quite similar
MAE = 0.32 eV and MARE of 3.38%. The behavior of P3 method can be explained by investigation of MAE
with respect to the basis set size, as was done in reference [58]. The larger basis set the bigger discrepancy
from reference results are observed for P3 method [58]. For small basis set, on the other hand, the P3
performs relatively well and thus benefits from basis set deficiencies. This was actually the case in both
previous studies [90, 92] regarding DNA and RNA nucleobases. Therefore, to conclude, in order to obtain
reliable results for a particular method one need also to test its convergence with respect to the basis set size.

3.3. Large photo voltaic molecules
In reference [11] we have employed the ΔMP2-SCS methods for calculation of IPs of 24 medium sized
organic acceptor molecules [102] that are particularly interesting from the standpoint of their future
application in organic photo voltaic devices [15–18]. Due to the small computational cost (especially the
variant combined with RI technique) of the ΔMP2-SCS methods their potential applicability to identify
this kind of systems belongs to one of the most important advantages of these approaches. Thus, in order to
discuss the capabilities and limitations of the present methods, we make additional assessment for several
large acceptors or donors organic molecules (listed in table 3 and depicted in figure 2), often used in
organic electronics [54, 97, 98]. The geometries of all structures have been re-optimized at the B3LYP/6-31G(2df, p) [93–95] level of theory.

In table 3 we report first IPs values extrapolated to CBS limit, calculated using various EP and \( \Delta \text{MP2} \) methods. The core orbitals were kept frozen in all EP calculations. In the case of rubrene, \text{C}_{60}, \text{C}_{70}, \text{and} \text{PCBM systems for all} \( \Delta \text{MP2} \) methods we report two values. First is obtained with the same basis set as for EP methods (see table 3 for more details) and the other (in the parenthesis) is the value obtained at CBS limit.

Inspection of MAE and MARE (calculated with respect to the molecular experimental data) reveals that the best performance among EP methods is given by OVGF-C which yields MAE of 0.14 eV (MARE = 2.12%). It is closely followed by OVGF-N results with MAE of 0.16 (MARE = 2.41%) and later by P3+ method which yields MAE of 0.21 eV (MARE = 3.11%). The worst performance is given by P3, EP3, OVGF-A, and B methods whose MAEs range between 0.27 and 0.37 eV (MARE between 3.72% and 5.52%). Turning now attention to \( \Delta \text{MP2} \)-methods results gives similar picture to those reported in sections 3.1 and 3.2 i.e. the \( \Delta \text{MP2-SCS} \) methods improve upon the unscaled \( \Delta \text{MP2} \) variant which yield MAE of 0.60 eV (MARE = 8.48%). The \( \Delta \text{MP2-SOS(IP)} \) and \( \Delta \text{MP2-SCS(IP)} \) give here, in turn, MAE = 0.34 eV (MARE = 5.03%) and MAE = 0.29 eV (MARE = 4.28%), respectively, being quite similar to ones reported by OVGF-B/C and P3 methods.

Due to very small computational cost of \( \Delta \text{MP2} \) methods, we have also been able to calculate CBS results (reported in parenthesis) for rubrene, \text{C}_{60}, \text{C}_{70}, \text{and} \text{PCBM systems what was not possible for other EP approaches. In this case we see, quite significant reduction of errors for all} \( \Delta \text{MP2} \) methods (\( \Delta \text{MP2: MAE} = 0.39 \text{ eV, MARE} = 5.6\%; \Delta \text{MP2-SOS(IP): MAE} = 0.23 \text{ eV, MARE} = 3.46\%; \Delta \text{MP2-SCS(IP): MAE} = 0.15 \text{ eV, MARE} = 2.28\%)). This is well known fact that for small basis sets (like employed here 6-31G and cc-pVDZ) the EP2 or \( \Delta \text{MP2} \) gives quite large errors with respect to the experimental data [104]. This feature is unfortunately inherited by \( \Delta \text{MP2-SCS} \) methods what can be clearly seen in table 3.

For three systems where molecular gas phase experimental values were unavailable, namely DCV3T, DIP, and 6T, we report the bulk phase experimental data. One can immediately note the large difference between the bulk phase and computed IPs values. The latter have much better agreement with the molecular experimental results as indicated in reference [97]. The difference between gas and bulk phase experimental data are mostly dictated by the polarization effect of bulk material due to the molecular ion which remains in the solid after an electron is removed [105]. We note, that also in this case, the quality of results given by \( \Delta \text{MP2-SCS} \) methods is not worse than other EP approaches such as OVGF-N and C.

One can note that table 3 does not report the EP2 IP values. This is due to the fact that the PS for most systems was found to be lower than 0.85 for many calculated first IPs. This finding is important in the context of recently published composite electron propagator (CEP) [54] method where authors propose focal point analysis [54, 106] to the calculation of IPs. In the CEP method, the high-level EP calculations (e.g. OVGF, P3, etc) are performed using quite small basis sets and then improved using incremental

---

**Figure 2.** Graphical representation of donor and acceptor molecules [54, 97, 98] from table 3: (a) sexithienyl (6T); (b) benzo[a]pyrene; (c) \text{C}_{60}; (d) DCV3T; (e) diindenoperylene (DIP); (f) Mg-octaethylporphyrin (MgOEP); (g) phthalocyanine (H2Pc); (h) rubrene; (i) pentacene; (j) PCBM; (k) \text{C}_{70}. Color coding: H-light gray, C-dark gray, N-blue, O-red, Mg-light green, S-yellow.
Table 5. Comparison of IPs (in eV) obtained for several SWCNTs (C_{20n}H_{20}, with \( n = 2–4 \)) and computational methods in 6-31G basis set. The MAE and MARE are calculated with respect to IP-EOM-CCSD/6-31G data from reference [103]. Core electrons have been kept frozen in all calculations. The best results are marked bold.

| \( n \) | State | IP-EOM* | EP2 | EP3 | P3+ | N | A | B | C | --- | --- | --- |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | \( 1^2 \)A\(_{g} \) | 5.97 | 5.63 | 6.12 | 6.17 | 6.07 | 5.84 | 5.83 | 5.84 | 5.96 | 5.51 | 5.92 | 5.90 |
| 2 | \( 1^2 \)A\(_{u} \) | 6.91 | 6.46 | 7.05 | 6.95 | 6.86 | 6.78 | 6.72 | 6.87 | 6.32 | 6.91 | 6.81 |
| 3 | \( 1^2 \)A\(_{u} \) | 5.08 | 4.74 | 5.26 | 5.39 | 5.26 | 4.94 | 4.91 | 4.94 | 5.08 | 4.63 | 5.04 | 4.96 |
| 4 | \( 1^2 \)A\(_{u} \) | 6.72 | 6.09 | 7.21 | 6.99 | 6.78 | 6.67 | 6.62 | 6.67 | 6.81 | 5.76 | 6.55 | 6.57 |
| 3 | \( 1^2 \)A\(_{u} \) | 5.13 | 4.79 | 5.39 | 5.54 | 5.38 | 5.17 | 5.46 | 5.02 | 5.17 | 4.69 | 5.13 | 5.00 |
| 4 | \( 1^2 \)A\(_{u} \) | 5.55 | 5.20 | 5.89 | 5.99 | 5.82 | 5.48 | 5.46 | 5.48 | 5.63 | 5.02 | 5.48 | 5.45 |

MAE 0.41 0.26 0.28 0.15 0.09 0.17 0.11 0.04 0.57 0.06 0.11
MARE (%) 6.88 4.42 4.96 2.74 1.59 2.99 1.85 0.71 9.59 0.90 1.96

*IP-EOM-CCSD results.

Figure 3. Comparison of first VIP of several SWCNTs (C_{20n}H_{20}, with \( n = 2–10 \)) calculated with several methods.

Corrections based on EP2 method results obtained for both smaller and larger basis sets. The approach was found to be extremely accurate in the IP predictions for organic photo-voltaic molecules. In table 4 we report the comparison of the \( \Delta MP2-SCS(IP) \) and \( \Delta MP2-SOS(IP) \) methods and the best CEP results reported in reference [54] for benzo[a]pyrene, C_{60} and MgOEP. One can note that the quality of \( \Delta MP2-SCS(IP) \) results is similar to the best CN/23 CEP method, being at the same time computationally much less demanding due to avoidance of performing high-level calculations. Thus, we conclude that also in this case the \( \Delta MP2-SCS \) methods are robust and provide high-quality results.

3.4. Carbon nanotubes

Since their discovery, the SWCNTs became recognized as potential candidates for materials for organic solar cells devices [107], due to their various interesting electronic properties [108]. In several recent studies [103, 109, 110], the impact of the nanotube length on the IPs was studied using various theoretical approaches e.g. IP-EOM-CCSD [103] or several DFT methods [110]. Thus, in this section, as a final test, we perform a similar assessment of \( \Delta MP2-SCS \) methods. To this end, we have considered SWCNTs of various lengths using geometries from reference [111] (C_{20n}H_{20}, with \( n = 2–10 \)).

First, we investigate the accuracy of EP and \( \Delta MP2 \) type methods in the prediction of the first two highest IPs. As a reference we utilized the IP-EOM-CCSD data taken from reference [103]. All calculations have been performed using the 6-31G basis set [95, 101] keeping the core electrons frozen. The results are reported in table 5 only for \( n = 2–4 \) due to the large computational cost of EP methods for larger SWCNTs.

Inspection of table 5 shows again that the IP-optimized \( \Delta MP2-SCS(IP) \) and \( \Delta MP2-SOS(IP) \) methods give similar accuracy to the OVGF and P3+ approaches yielding MAEs of 0.06 eV (MARE = 0.9%) and 0.11 eV (MARE = 1.96%) for \( \Delta MP2-SOS(IP) \) and \( \Delta MP2-SCS(IP) \), respectively. Not surprisingly the worst performance, is given by standard \( \Delta MP2 \) (MAE = 0.57 eV, MARE = 9.59%) and EP2 (MAE = 0.41 eV, MARE = 6.88%) methods, followed by EP3 (MAE = 0.26 eV, MARE = 4.42%) and P3 (MAE = 0.28 eV, MARE = 4.96%) results.
As was shown in references [103, 110] the IP-EOM-CCSD SWCNTs IPs data are quite sensitive to the quality of basis set used in the calculations. For example the difference between 6-31G and aug-cc-pVDZ [76] C60H20 results is roughly of 0.5 eV. Thus, in order to investigate basis set dependence of ΔMP2-SCS methods result in figure 3 we also report the value of first IP as a function of SWCNT length. As a reference, we report the IP-EOM-CCSD data taken from reference [110]. For comparison, we report also the ωB97XD results obtained without and with IP-tuned value of ω (denoted with asterisk-ωB97XD*). Likewise in reference [110] all data have been obtained in may-cc-pVDZ [112] basis set. As we can see for n = 2–6 the latter data are in quite good agreement with the reference IP-EOM-CCSD values. Thus, for n = 7–10 (where IP-EOM-CCSD results are unavailable) the ωB97XD* IPs are used as a reference.

From figure 3 we clearly see that the ΔMP2-SCS(IP) and ΔMP2-SOS(IP) results present similar qualitative behavior to IP-EOM-CCSD data even for C80H20 where ωB97XD* shows almost same IP values as for C60H20 and C100H20 SWCNTs. Furthermore, one can note that both ΔMP2 methods differ in results by 0.25 ~ 0.40 eV from reference IP-EOM-CCSD values for n = 2–6. Similar behavior is observed for ωB97XD (without IP-tuning) method results what indicate that optimization of ω, which is not always straightforward task especially for complex molecular systems [113], is crucial for obtaining reliable IP values.

For larger carbon nanotubes (n = 7–10) the agreement of ΔMP2 results with ωB97XD* is more noticeable, especially for ΔMP2-SOS(IP) method. This assessment indicates that both ΔMP2-SCS approaches also, in this case, can provide good qualitative predictions despite their simplicity.

4. Conclusions

In this study we have performed the comprehensive assessment of the ΔMP2-SCS methods for large set of small and medium size molecules, DNA/RNA bases, donor and acceptor organic molecular systems, and finally SWCNTs.

The analysis of the results clearly shows that the ΔMP2-SCS family of methods are able to provide IPs comparable in accuracy with state-of-art EP approaches such as OVGF, P3+ or CEP methods with substantially lower computational cost (O(N^3)), also outperforming the standard DFT methods.

Additionally the combination of our ΔMP2-SCS(IP) and ΔMP2-SOS(IP) methods with the RI technique, provides excellent results for a variety of large molecular systems what is really attractive from the standpoint of the search and development of new organovoltaic materials for solar cell devices. Therefore, the ΔMP2-SCS(IP) and ΔMP2-SOS(IP) methods can be used as a simple but reliable and practical method for routine calculation of accurate VIPs in broad applications.

Finally, we remark, that despite the fact, that the ΔMP2-SCS methods provide results which differ from the reference data i.e. the IP-EOM-CCSD and experimental ones of only 0.1–0.2 eV, there is a still room for improvement e.g. by utilization in equation (2) the system dependent cOS and cSS parameters [51, 52, 62, 114–116] which will be the object of our further study.

Acknowledgments

We thank Manuel Díaz-Tinoco and Kamel Alimi for optimized geometries used in section 3.3. SS thanks the Polish National Science Center for the partial financial support under Grant No. 2016/21/D/ST4/00903. Calculations have been partially carried out in Wroclaw Centre for Networking and Supercomputing (http://www.wcss.pl), Grant No. 499.

ORCID iDs

Szymon Śmiga  https://orcid.org/0000-0002-5941-5409
Ireneusz Grabowski  https://orcid.org/0000-0002-5661-5659

References

[1] Lee S-H, Kim J-H, Shim T-H and Park J-G 2009 Electron. Mater. Lett. 5 47
[2] Mihailetchi V D, Blom P W M, Hummelen J C and Rispens M T 2003 J. Appl. Phys. 94 6849
[3] Louis E, San-Fabian E, Díaz-García M A, Chiappe G and Vergés J A 2017 J. Phys. Chem. Lett. 8 2445
[4] Yang F, Sun K and Forrest S 2007 Adv. Mater. 19 4166
[5] Nitzan A and Ratner M A 2003 Science 300 1384
[6] Brabec C J, Sariciftci N S and Hummelen J C 2001 Adv. Funct. Mater. 11 15
[71] Parrish R M et al 2017 J. Chem. Theory Comput. 13 3185
[72] Feyereisen M, Fitzgerald G and Komornicki A 1993 Chem. Phys. Lett. 208 359
[73] Vahtras O, Almlof J and Feyereisen M 1993 Chem. Phys. Lett. 213 514
[74] Flores-Moreno R and Ortiz J V 2009 J. Chem. Phys. 131 124110
[75] Lew-Yee J F H, Flores-Moreno R, Morales J L and del Campo J M 2020 J. Chem. Theory Comput. 16 1597
[76] Dunning T H 1989 J. Chem. Phys. 90 1090
[77] Weigend F, Köhn A and Hättig C 2002 J. Chem. Phys. 116 3175
[78] Bokhan D and Ten-no S 2010 J. Chem. Phys. 132 021101
[79] Helgaker T, Klopper W, Koch H and Noga J 1997 J. Chem. Phys. 106 9639
[80] Stanton J F and Gauss J 1999 J. Chem. Phys. 111 8785
[81] Wängler F, Köhn A and Hattig C 2002 J. Chem. Phys. 116 3175
[82] Bozkaya U and Ünal A 2018 J. Phys. Chem. A 122 4375
[83] Kumar D, Dutta A K and Manohar P U 2017 J. Chem. Sci. 129 1611
[84] Orlow V, Smirnov A and Varshavsky Y 1976 Tetrahedron Lett. 17 4377
[85] Bravaya K B, Kostko O, Dolgikh S, Landau A, Ahmed M and Krylov A I 2010 J. Phys. Chem. A 114 12305
[86] Faber C, Attaccalite C, Olevano V, Runge E and Blase X 2011 Phys. Rev. B 83 115123
[87] Roca-Sanjúan D, Rubio M, Merchán M and Serrano-Andrés I 2006 J. Chem. Phys. 125 084302
[88] Becke A D 1993 J. Chem. Phys. 98 5648
[89] Krishnan R, Binkley J S, Seeger R and Pople J A 1980 J. Chem. Phys. 72 650–4
[90] Hehe W J, Ditchfield R and Pople J A 1972 J. Chem. Phys. 56 2257–61
[91] Golubeva A A and Krylov A I 2009 Phys. Chem. Chem. Phys. 11 1503
[92] Dargouth A A M H M, Casida M E, Taoiali W, Alimi K, Ljungberg M P, Koval P, Sánchez-Portal D and Fonster D 2015 Computation 3 616
[93] Zakrzewski V G, Dolgounitcheva O and Ortiz J V 2008 J. Chem. Phys. 129 104106
[94] Akiyama I, Li K C, LeBreton P R, Fu P P and Harvey R G 1979 J. Phys. Chem. 83 2997
[95] Kitagawa S, Morishima I, Yonezawa T and Sato N 1979 Inorg. Chem. 18 1345
[96] Ditchfield R, Hehe W J and Pople J A 1971 J. Chem. Phys. 54 724–8
[97] Richard R, Marshall M, Dolgounitcheva O, Ortiz J V, Brédas J, Marom N and David S 2016 J. Chem. Theory Comput. 12 595
[98] Peng B, Govind N, Apra E, Klemm M, Hammond J R and Kowalski K 2017 J. Chem. Phys. A 121 1328
[99] Corzo H H, Galano A, Dolgounitcheva O, Zakrzewski V G and Ortiz J V 2015 J. Phys. Chem. A 119 8813
[100] Takahashi T, Harada Y, Sato N, Seki K, Inokuchi H and Fujisawa S 1979 Bull. Chem. Soc. Japan 52 380
[101] Cataldo S, Salice P, Menza E and Pignataro B 2012 Energy Environ. Sci. 5 5919
[102] Baughman R H, Zakhidov A A and de Heer W A 2002 Science 297 787
[103] Zhou Z, Steigerwald M, Hybertsen M, Brus L and Friesner R A 2004 J. Am. Chem. Soc. 126 3597
[104] Zhou B, Hu Z, Jiang Y, He X, Sun Z and Sun H 2018 J. Phys. Condens. Matter 30 215301
[105] Petrushenko I K and Ivanov N A 2013 Physica E 54 262
[106] Davidson E R 1996 Chem. Phys. Lett. 260 514
[107] Knight J, Wang X, Gallandi L, Dolgounitcheva O, Ren X, Ortiz J V, Rinke P, Körsdörfer T and Marom N 2016 J. Chem. Theory Comput. 12 615
[108] Szabados A 2006 J. Chem. Phys. 125 214105
[109] Fink R F 2010 J. Chem. Phys. 133 174113
[110] Śmiga S and Fabiano E 2017 Phys. Chem. Chem. Phys. 19 30249