

## Can G4-like composite *Ab Initio* methods accurately predict vibrational harmonic frequencies?

Emmanouil Semidalas & Jan M. L. Martin

To cite this article: Emmanouil Semidalas & Jan M. L. Martin (29 Sep 2023): Can G4-like composite *Ab Initio* methods accurately predict vibrational harmonic frequencies?, Molecular Physics, DOI: [10.1080/00268976.2023.2263593](https://doi.org/10.1080/00268976.2023.2263593)

To link to this article: <https://doi.org/10.1080/00268976.2023.2263593>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



View supplementary material [↗](#)



Published online: 29 Sep 2023.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

# Can G4-like composite *Ab Initio* methods accurately predict vibrational harmonic frequencies?

Emmanouil Semidalas and Jan M. L. Martin

Dept. of Molecular Chemistry and Materials Science, Weizmann Institute of Science, Rehovot, Israel

## ABSTRACT

Minimally empirical G4-like composite wavefunction theories [E. Semidalas and J. M. L. Martin, *J. Chem. Theory Comput.* **16**, 4238–4255 and 7507–7524 (2020)] trained against the large and chemically diverse GMTKN55 benchmark suite have demonstrated both accuracy and cost-effectiveness in predicting thermochemistry, barrier heights, and noncovalent interaction energies. Here, we assess the spectroscopic accuracy of top-performing methods: G4-*n*, cc-G4-*n*, and G4-*n*-F12, and validate them against explicitly correlated coupled-cluster CCSD(T\*)(F12\*) harmonic vibrational frequencies and experimental data from the HFREQ2014 dataset, of small first- and second-row polyatomics. G4-T is three times more accurate than plain CCSD(T)/def2-TZVP, while G4-T<sub>ano</sub> is two times superior to CCSD(T)/ano-pVTZ. Combining CCSD(T)/ano-pVTZ with MP2-F12 in a parameter-free composite scheme results to a root-mean-square deviation of 5 cm<sup>-1</sup> relative to experiment, comparable to CCSD(T) at the complete basis set limit. Application to the harmonic frequencies of benzene reveals a significant advantage of composites with ANO basis sets – MP2/ano-pV<sub>m</sub>Z and [CCSD(T)-MP2]/ano-pVTZ (*m* = Q or 5) – over similar protocols based on CCSD(T)/def2-TZVP. Overall, G4-type composite energy schemes, particularly when combined with ANO basis sets in CCSD(T), are accurate and comparatively inexpensive tools for computational vibrational spectroscopy.

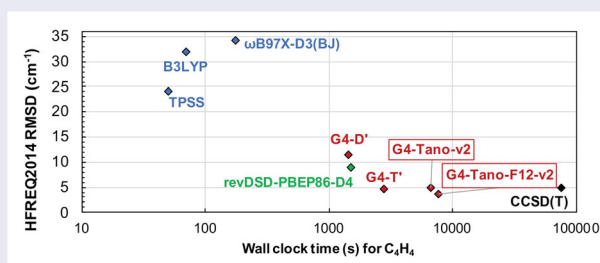
## ARTICLE HISTORY

Received 14 August 2023

Accepted 21 September 2023

## KEYWORDS

Coupled-cluster theory; vibrational spectroscopy; composite wavefunction methods






## 1. Introduction

In the realm of computational spectroscopy [1], accurate prediction of molecular vibrations has long been a valuable tool in chemistry, biochemistry, and materials science [2–7]. To achieve spectroscopic accuracy, defined as an error of less than 1 cm<sup>-1</sup> from gas-phase vibrations [8, 9], it may be necessary to venture beyond the ‘gold-standard’ CCSD(T) method [10], i.e. coupled-cluster with single, double, and perturbative triple excitations, and employ higher-order correlation approaches that include quadruple excitations, such as in the CCSDT(Q) method [11]. Additionally, the slower basis set convergence [12–14] of the correlation energy can quickly lead

to computational constraints, especially for the electronic structure of larger molecules.

In response to the challenges that lie with computationally expensive *ab initio* methods, the composite wave function theories (cWFTs) have emerged as promising alternatives (for a very recent review, see Karton [15]). These methods combine high-level treatments of electron correlation through additivity approximations, economical basis set extrapolations, and often empirical corrections. The end result is a robust approach aimed at the accuracy of CCSD(T) at the complete basis set limit (CBS). Examples of such cWFTs include the Pople’s group Gaussian-*n* theories (G<sub>n</sub>) [16–21], the CBS models

**CONTACT** Jan M. L. Martin  gershom@weizmann.ac.il  Dept. of Molecular Chemistry and Materials Science, Weizmann Institute of Science, Rehovot 7610001, Israel

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/00268976.2023.2263593>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

by Petersson and co-workers [22–24], the Weizmann- $n$  theory ( $n = 1$  and 2) by the Martin group [25–27], and the ccCA (correlation consistent composite approach) of Wilson and coworkers [28–30].

Among the more economical composite energy schemes, G4 [19] and G4(MP2) [20, 31, 32] approaches fit in. The former combines MP $n$  ( $n = 2$  and 4) and CCSD(T) methods, while the latter reduces cost by omitting the MP4 step. In 2011, Radom and co-workers introduced the G4(MP2)-6X protocol [33], an improved G4(MP2) variant, featuring six empirical parameters for correlation energies and another six for the high-level correction. Building on this, Chan et al. [34] shifted from Pople- to Karlsruhe-type basis sets in their G4(MP2)-XK approach. Inspired by their work, we presented an hierarchy of G4-type cWFTs [35, 36], validated against the energetics of the chemically large GMTKN55 benchmark suite of Goerigk, Grimme, and coworkers [37] (general main-group thermochemistry, kinetics, and noncovalent interactions, 55 subsets)

Another avenue for improving cost-effectiveness and accuracy in cWFTs is using explicitly correlated theory [38–41], where  $r_{12}$  terms that depend on the interelectronic distance, e.g. of the form  $[1 - \exp(\gamma/r_{12})]/\gamma$ , are added in the wave function. This inclusion accelerates basis set convergence [42, 43], with R12/F12 methods typically requiring 2–3 additional basis set cardinal numbers or ‘zetas’ compared to conventional calculations [38, 39, 41]. Presently, numerous *explicitly correlated* composite thermochemical protocols have been reported, including the W4-F12 [44], ccCA-F12 [45], G4- $m$ -F12 [35], and SVECV-f12 theories [46].

For harmonic vibrational frequencies it is generally accepted that valence-only CCSD(T) suffices, particularly in systems with low static correlation that are dominated by a single reference determinant [9, 47]. This arises from a fortuitous error compensation between the approximate treatment of the triples term, the missing core-valence correlation, and the neglect of higher order excitations.

To approach the full CI limit with composite energy schemes, post-CCSD(T) terms must be included. This has been successfully demonstrated in the well-known HEAT- $n$  approaches [48–51], the Feller-Peterson-Dixon (FPD) model, [52–54], the Wilson’s group ccCA, [28–30], and the Weizmann- $n$  ( $n = 3$  and 4) theories [55–57] by the Martin group. We recently examined [58] the importance of post-CCSD(T) corrections in cWFTs, particularly CCSDT(Q)<sub>A</sub> [11], for spectroscopic constants in heavy-atom diatomics at different static correlation regimes, and reported accurate predictions, including ozone vibrational frequencies.

There has been a fair amount of work on post-CCSD(T) cWFT methods in the context of vibrational spectroscopy. We note, inter alia, the 2005 work of Heckert et al. [59, 60] and Puzzarini et al. [61] on accurate geometries *viz.* rotational constants. Ruden et al. [62] considered quadruples and quintuples terms in CCSD(T)-based composite schemes for harmonic frequencies of HF, N<sub>2</sub>, F<sub>2</sub>, and CO, while Karton and Martin [8] applied pointwise W4 theory (and truncations thereof, as well as the enhanced W4.3 theory) to spectroscopic constants and electric properties of 28 first- and second-row diatomics, as well as several polyatomics [8]. The spectroscopic constants of formaldehyde were obtained by Schaefer and co-workers [63] through CCSDT(Q)-based focal-point analysis [64–67], while Zhu and Xu [68] reported static polarisabilities at CCSD(T)/CBS. Huang and Lee [69], and later Lee and coworkers [70, 71] explored the CcCR methodologies (‘C’ stands for CBS, complete basis set; ‘cC’ for core correlation; and ‘R’ for relativistic effects) for determining fundamental vibrational frequencies [70] and anharmonic rotational constants [71].

In contrast, as well as in comparison to combining high-quality harmonic frequencies with DFT-level anharmonic force fields (see, e.g. [3] and references therein; Ref. [72]), effort toward an economical cWFT approximation to CCSD(T) harmonic frequencies has been fairly limited. An unfairly overlooked paper by Betens and coworkers [73] considered the combination of MP2 in larger basis sets with CCSD(T)–MP2 in smaller ones. Barone and coworkers [74] introduced what they termed their ‘cheap’ approximation, which augments an MP2/cc-pVT,QZ CBS extrapolation (the notation means ‘from cc-pVTZ and cc-pVQZ’) with diffuse function, CCSD(T)–MP2, and core-valence corrections.

The purpose of this study is to assess whether G4-type composite energy schemes can be a viable alternative to large basis set CCSD(T) harmonic vibrational frequency calculations. We shall validate these G4-type methods against basis set limit extrapolated CCSD(T) frequencies, as well as CCSD(T\*)(F12\*)/VQZ-F12 calculated harmonic vibrational frequencies and experimental ones of 31 molecules from the HFREQ2014 dataset [75]. As a proof of principle, cWFTs are then applied to the difficult [76] harmonic force field of benzene.

## 2. Computational details

All calculations were performed on the Chemfarm HPC cluster of the Faculty of Chemistry at the Weizmann Institute, mostly using the MOLPRO 2022.3 [77] electronic structure programme system. Built on top of

**Table 1.** Abbreviations for methods, basis sets, and other terms in the present work.

Acronym	Meaning	Reference
MP2	second-order Møller-Plesset perturbation theory (PT)	[79]
MP2-F12	explicitly correlated second-order Møller-Plesset PT	[85]
VPT2	second-order vibrational PT	[107]
CCSD(T)	coupled-cluster with single, double, and quasiperturbative triple excitations	[10, 82]
CCSD(T*)(F12*)	ditto but explicitly correlated coupled-cluster theory with the F12 geminal in the Hättig-Kohn-Tew (F12*) approximation to CCSD-F12 and the Marchetti-Werner scaling for (T)	[91, 108]
cWFT	composite wavefunction theory	[16–21, 109]
G <sub>n</sub>	Gaussian- <i>n</i> theory	
G4- <i>n</i>	G4-like theory with CCSD(T) contributions from def2-SVPD or def2-TZVP basis sets ( <i>n</i> = D,T)	
cc-G4- <i>n</i>	G4-like theory with core-valence correlation at the MP2 Level ( <i>n</i> = D,T)	[36]
G4- <i>n</i> -F12	explicitly correlated RI-MP2-F12-based G4-like theory	[35]
def2- <i>n</i> ZVPP	Weigend-Ahlrichs def2 basis set family ( <i>n</i> = T,Q)	[92]
cc-pVnZ	correlation-consistent basis set family for valence correlation ( <i>n</i> = D,T,Q)	[96–98]
cc-pVnZ-F12	correlation-consistent basis set family for use in explicitly correlated calculations ( <i>n</i> = D,T,Q)	
ano-pVnZ	atomic natural orbital basis sets ( <i>n</i> = D,T,Q,5)	[95]
GMTKN55	general main-group thermochemistry, kinetics, and noncovalent interactions, 55 subsets	[37]
WTMAD2	weighted mean absolute deviation (type 2)	[37]

the ALASKA integral derivative package [78], canonical MP2 [79] analytical derivatives (Ref. [80] and references therein; see Ref. [81] for the specific MOLPRO implementation) and canonical CCSD(T) [10, 82] analytical first derivatives (Ref. [49] and references therein) were evaluated with nondegenerate symmetry enabled, while force constant matrices (Hessians) were evaluated semi-numerically using central differences of gradients. For verification purposes, MP2 Hessians in the same basis set were also calculated analytically [83] using Gaussian 16; [84] we found harmonic frequencies from the analytical and semi-numerical Hessians to differ by only on the order of  $0.03 \text{ cm}^{-1}$ , which is negligible in the context of this work.

The explicitly correlated density-fitted DF-MP2-F12 method [85] was employed with analytic gradients [86, 87] and the 3\*C(FIX, HY1) Ansatz, in which the extended Brillouin condition is assumed and the "HY1" hybrid approximation is used for matrix elements [88] over the F12 geminal [89], together with fixed geminal amplitudes [89, 90]. The CCSD(T)(F12\*) [91] geometry optimizations and frequency calculations were carried out fully numerically for want of an analytical gradient.

In our study, we employed various basis set families, including the Weigend-Ahlrichs def2 family [92]: def2-SVP, def2-nZVP, and def2-nZVPP, (*n* = T and Q), along with their augmented alternatives with diffuse functions def2-SVPD and def2-nZVPPD (*n* = T and Q) [93]. The combination of def2-nZVPP on hydrogen atoms and def2-nZVPPD on main group elements is denoted as def2-nZVPPD'. Among the atomic natural orbital (ANOs) basis sets pioneered by Almlöf and

Taylor [94], we chose the ano-pVnZ (*n* = D, T, Q, 5) of Neese and Valeev [95], as well as their aug-ano-pVnZ (diffuse function augmented) and saug-ano-pVnZ (minimally augmented) variants from the same reference. Table 1 provides a list of abbreviations for methods and basis sets used in this study.

Among the correlation consistent basis set family, we consider the cc-pVnZ and aug-cc-pVnZ basis sets (*n* = D, T, Q, 5) [96–98] on hydrogen and the first row, and the (aug-)-cc-pV(n+d)Z basis sets [99] on second-row elements, which include an additional tight *d* function as was previously found [100, 101] to be important when these elements are in high oxidation states. (In this work, the largest impact is seen for SO<sub>2</sub>.) Additionally, for calculations including inner-shell correlation, we employed the core-valence weighted aug-cc-pwCVnZ (*n* = T and Q) basis sets [102]. The shorthand haVnZ+d refers in this paper to the combination of cc-pVnZ on hydrogen with aug-cc-pVnZ on first-row atoms and aug-cc-pV(n+d)Z on second-row atoms.

Aside from the orbital basis set (OBS) employed in a standard explicitly correlated calculation with density-fitting, there are three additional auxiliary basis sets (ABS): the 'JKFit' basis set for the Coulomb and exchange integrals, the 'MP2Fit' basis set for density fitting in MP2, and the 'CABS' also known as complementary auxiliary basis set [103, 104]. We utilised the cc-pVnZ-F12 [105] (*n* = T and Q) basis sets as OBS, along with the default cc-pVnZ-F12/JKFit and cc-pVnZ-F12/MP2Fit in MOLPRO as JKFit and MP2Fit ABSs, respectively. For CABS, we used Yousaf and Peterson's cc-pVnZ-F12/OptRI [106]. Slater-type geminal terms of the F12 form [1-exp ( $\gamma$

**Table 2.** Molecules considered in the HFREQ2014 dataset.

BH<sub>3</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>4</sub>, CCl<sub>2</sub>, CF<sub>2</sub>, CH<sub>3</sub>OH, CH<sub>4</sub>, Cl<sub>2</sub>, ClCN,  
 ClF, CO, CO<sub>2</sub>, CS, CS<sub>2</sub>, F<sub>2</sub>, H<sub>2</sub>CO, H<sub>2</sub>CS, H<sub>2</sub>O,  
 H<sub>2</sub>S, HCl, HCN, HF, HNO, HOCl, N<sub>2</sub>, NH<sub>3</sub>, NNO,  
 OCS, PH<sub>3</sub>, SiO, SO<sub>2</sub>

$r_{12})/\gamma$  were used with a  $\beta$  geminal exponent of 1.0 for both triple- and quadruple- $\zeta$  OBS, as recommended in Table V of Ref. [90]. In the text below, ‘VnZ-F12’ signifies the cc-pVnZ-F12 basis sets.

To validate the accuracy of composite schemes, we used the HFREQ2014 dataset [75] of harmonic frequencies for small molecules (Table 2). Error statistics were estimated relative to CCSD(T\*)(F12\*)/VQZ-F12 calculations (reference) and experimental values from ref. [75] and references therein. On a related note, Mehta et al. [110] considered the same CCSD(T\*)(F12\*)/VQZ-F12 reference for HFREQ2014 in their study on the performance of double-hybrid density functional theory for molecular vibrations; they also carried out CCSD(T) calculations there for comparison, but excluded some of the HFREQ2014 species such as F<sub>2</sub>, HNO, and CF<sub>2</sub>. Also, Head-Gordon and coworkers [111] recently introduced analytical second derivatives of VV10 dispersion corrected [112] containing density functionals and evaluated their predictive accuracy for harmonic frequencies across various molecular systems including those in the HFREQ2014 dataset. They concluded that while the VV10-enhanced DFT functionals offered no advantage for small-molecule vibrational spectra, but a significant improvement was seen in vibrational spectra of noncovalent complexes.

Geometries were optimised using the total electronic energy as the target function for each cWFT method, employing the numerical gradient. That was accomplished through the optg procedure [113] in MOLPRO. The optimizations were completed once the maximum gradient component was less than  $10^{-5}$  hartree/bohr, the optimisation step was less than  $10^{-5}$  bohr, and the change in total energy from the previous iteration was less than  $10^{-11}$  hartree. In numerical gradients and Hessians, the default stepsize of 0.01 a.u. was used, unless otherwise noted (see Table 3 below for cases where a 0.005 a.u. value was used). The cutoffs of two-electron integrals were set to  $10^{-20}$  for screening and  $10^{-18}$  for the prefactor test. The total energy in subsequent calculations of force constant calculations was converged within  $10^{-12}$  E<sub>h</sub>.

We provide an implementation of `hfreq` for automated geometry optimisation and calculation of harmonic vibrational frequencies with composite energy schemes at the following Github link ‘<https://github.com/msemidalas/hfreq>’. `hfreq` is written in Python and the Git repository contains several sample files to reproduce the results of this work. In this version, geometry optimisation employs analytic gradients, while Hessians are computed numerically in MOLPRO. The mass-weighted Hessian is diagonalised in Psi4 [114]. Figure 1 graphically shows the automated procedure for the evaluation of harmonic frequencies.

A reviewer highlighted a very recent paper by Jensen [115], in which three methods for extrapolating vibrational frequencies are discussed. The first, ‘Opt-xpol’, involves geometry optimisation by minimising the basis

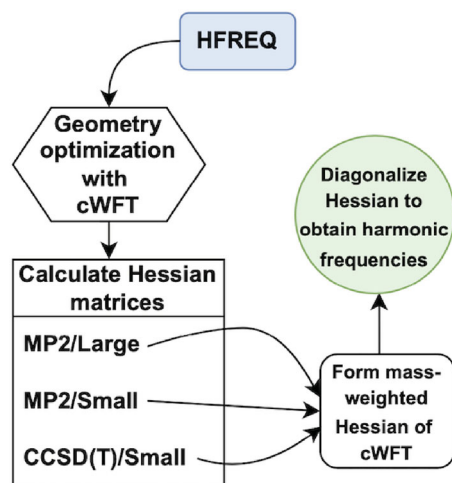
**Table 3.** Root-mean-square deviations and mean-absolute deviations (cm<sup>-1</sup>) of calculated harmonic frequencies at the CCSD(T) level from CCSD(T\*)(F12\*)/VQZ-F12 calculations (referred to as CCSD(T)/CBS) and experiment for the HFREQ2014 dataset.

	Errors relative to					Errors relative to					Errors relative to			
	CCSD(T)/CBS		Expt			CCSD(T)/CBS		Expt			CCSD(T)/CBS		Expt	
	RMSD	MAD	RMSD	MAD		CCSD(T)	RMSD	MAD	RMSD		MAD	CCSD(T)	RMSD	MAD
def2-SVP	31.06	23.82	30.36	23.22	VDZ+d	32.65	20.68	31.31	19.82	ano-pVDZ	28.71 <sup>a</sup>	18.89	26.99 <sup>a</sup>	17.41
def2-SVPD	28.13	18.51	26.53	17.96	haVDZ+d	37.93 <sup>b</sup>	27.63	36.72 <sup>b</sup>	27.13	saug-ano-pVDZ	25.35	17.65	23.99	16.64
										aug-ano-pVDZ	25.18	18.98	23.86	18.30
def2-TZVP	16.17 <sup>a</sup>	10.76	15.41 <sup>a</sup>	9.68	VTZ+d	11.24	7.60	11.75	7.63	ano-pVTZ	10.66 <sup>a</sup>	7.17	10.09 <sup>a</sup>	6.68
def2-TZVPP	8.89	6.35	8.90	6.43	haVTZ+d	11.66	9.51	11.03	9.15	saug-ano-pVTZ	9.03	6.11	8.37	5.78
										aug-ano-pVTZ	8.83	6.58	8.62	6.40
def2-QZVP	4.13	2.64	5.75	3.94	VQZ+d	5.62	2.95	7.06	4.15	ano-pVQZ	5.10 <sup>a</sup>	3.34	5.89 <sup>a</sup>	4.13
def2-QZVPP		ditto			haVQZ+d	4.09	3.23	4.87	3.89	saug-ano-pVQZ	4.19	2.73	4.95	3.20
										aug-ano-pVQZ	3.62	2.48	4.51	2.87
					V5Z+d					ano-pV5Z	2.81	1.99	4.71	3.53
					haV5Z+d	1.99	1.48	4.40	3.28	saug-ano-pV5Z	2.46	1.49	4.32	2.86
										aug-ano-pV5Z	2.52	1.26	4.51	2.08
CCSD(T*)(F12*)/ cc-pVDZ-F12	2.75	2.00	5.42	4.01	CCSD(T)/ haVQ5Z+d	0.99	0.65	4.50	3.14	CCSD(T*)(F12*)/ cc-pVTZ-F12	0.70	0.50	4.69	3.63

CBS refers to the CCSD(T) complete basis set limit

<sup>a</sup> Using a step size of 0.005 a.u. during numerical differentiation leads to minor increases in RMSDs by 0.01–0.03 cm<sup>-1</sup>. <sup>b</sup> 33.84 and 32.15 cm<sup>-1</sup>, respectively, if the acetylene bending frequencies are excluded.





**Figure 1.** hfreq implementation workflow.

set-extrapolated energy, followed by frequency calculation from the extrapolated Hessian at the optimised geometry: this parallels our approach in the present work and in the hfreq code. The second, ‘v-xpol’, directly extrapolates vibrational frequencies from optimised geometries using two different basis sets, an approach explored earlier by Varandas [116] and Broda and colleagues [117–119]. The ‘H-xpol’ approach directly extrapolates optimised Hessians, regardless of reference geometries. Jensen’s findings show that all three approaches yield similar results for small molecules using double-triple  $\zeta$  extrapolation in cc-pVnZ basis sets at wB97X-D [120] and MP2 levels. However, for H-bonded complexes, ‘H-xpol’ yields unsatisfactory results with extrapolation from pcseg-0 and pcseg-1 basis sets [121], as well as from pcseg-1 and pcseg-2, likely due to poor reference geometries.

### 3. Results and discussion

#### 3.1. Performance of CCSD(T) for harmonic frequencies

Assessing CCSD(T) for harmonic frequencies is the first step to estimate the accuracy of composite energy schemes. Kesharwani and Martin reported that valence-only CCSD(T) at the complete basis set limit is about  $5\text{ cm}^{-1}$  as accurate as the experimental harmonic frequencies for the HFREQ2014 dataset [75]. (For the avoidance of doubt, we should stress that these experimental data are truly harmonic, obtained from fitting series expansion in the vibrational quantum numbers to many vibrational band origins.) Explicitly-correlated CCSD(T\*)(F12\*) with VQZ-F12 achieves a root-mean-square deviation (RMSD) of  $4.7\text{ cm}^{-1}$  compared to experiment; T\* denotes pointwise Marchetti-Werner

scaling [108] of the triples. In a recent DFT study of harmonic frequencies [110], very similar results were obtained with fairly large basis sets, such as VQZ-F12, between either the point-wise scaling T\* or scaling the triples term by a constant factor ( $T_s$ ) [122].

Table 3 presents the error statistics of CCSD(T) with different classes of basis sets compared to reference calculated harmonic frequencies and experimental values for HFREQ2014; the error distribution is also depicted as a ‘box and whiskers plot’ in Figure 2.

First of all, concerning the reference, we could have made two basically equivalent choices, as CCSD(T\*)(F12\*)/VQZ-F12 and pointwise CCSD(T)/haVQ,5Z+d extrapolation differ from each other by just  $1.0\text{ cm}^{-1}$  RMS. We have selected the former throughout. (For the avoidance of doubt, doubly and triply degenerate frequencies are assigned weights of 2 and 3, respectively, in the statistics.)

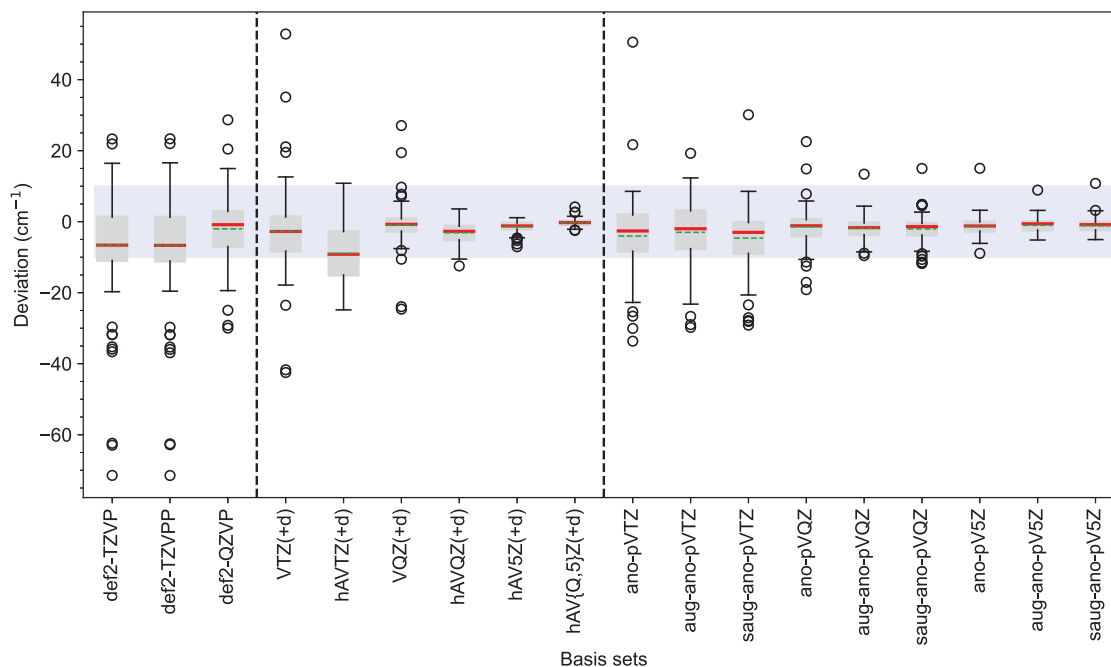
Second, the addition of tight  $d$  functions to the second-row aug-cc-pVnZ basis sets has a very significant effect in  $\text{SO}_2$  for smaller  $n$ . At first sight, no similar phenomenon is seen for the ANO basis sets; however, the primitive  $d$  functions that make up the  $d$  symmetry ANOs have exponents 5.0755, 2.1833, 0.9392, 0.404, and 0.1738; hence, the high-exponent space is already adequately covered in the primitives.

Third, in both ANO and correlation consistent families, augmented basis sets have better statistics than unaugmented ones (with the exception of cc-pVDZ+d).

Fourth, among the ANO family, the fully augmented aug-ano-pVnZ basis sets have slightly better error statistics than the more economical ‘minimally augmented’ saug-ano-pVnZ basis sets.

Quadruple- $\zeta$  quality basis sets from all families outperform  $n = D$  and T members. The correlation-consistent haVQZ+d and the ANO set saug-ano-pVQZ yield similar RMSDs of  $4.9$  and  $5.0\text{ cm}^{-1}$  compared to experiment. For triple- $\zeta$  and especially double- $\zeta$ , the ANO basis sets are markedly superior. Somewhat surprisingly, def2-TZVPP and def2-QZVPP appear to be superior to their cc-pVnZ counterparts, and similar error reductions occur as more zetas are added, with  $n = Q$  having a small  $\text{RMSD}_{\text{exp}}$  ( $5.8\text{ cm}^{-1}$ ). (We note that def2-QZVPP and def2-QZVPP are equivalent for the molecules considered here, and that the step-size choice for numerical differentiation, 0.01 or 0.005 a.u., has no significant effect on the calculated frequencies.)

CCSD(T) with def2-TZVPP is in nearly thrice worse agreement with experiment than the complete basis set limit. Given that def2-TZVPP is used in the high-level correction [CCSD(T)-MP2] because of its lower cost in the G4-type cWFTs, any improvements of the error



**Figure 2.** Box-and-whisker plot for the deviations of harmonic vibrational frequencies at CCSD(T) with various basis sets from CCSD(T\*)(F12\*)/VQZ-F12 for the HFREQ2014 dataset. The interquartile range (IQR) is the difference between the third and first quartiles ( $IQR = Q_3 - Q_1$ ). The upper whisker extends up to  $Q_3 + 1.5 \cdot IQR$ , while the lower whisker extends down to  $Q_1 - 1.5 \cdot IQR$ , and outliers are shown outside the whiskers. The median is indicated by a red line, while a green dotted line represents the mean. The blue band indicates  $\pm 10 \text{ cm}^{-1}$ .

statistics over ‘pure’ CCSD(T)/def2-TZVP would make these cWFTs useful for spectroscopy.

ANO basis sets outperform similarly-sized correlation consistent basis sets with notable improvements in  $RMSD_{CBS}$ , which are 7.2, 1.7, and  $1 \text{ cm}^{-1}$  over VnZ with  $n = D, T$ , and  $Q$ , respectively. The lowest errors are obtained for ano-pV5Z in CCSD(T) with an RMSD of  $2.8 \text{ cm}^{-1}$  relative to reference. Another possibility is that the ANO basis sets – which are better equipped than small-medium sized correlation consistent basis sets for predicting accurately harmonic frequencies [76, 123, 124] – may offer advantages in composite wave function schemes as discussed below.

The hypersensitivity to the basis set of the acetylene bending frequencies was first noted by Lee and coworkers [125] and analysed in detail in Refs. [126, 127] as an intramolecular BSSE (basis set superposition error) problem. Another frequency that exhibits basis set hypersensitivity is the umbrella mode of ammonia – which is a conspicuous outlier even for ano-pV5Z, less so for saug-ano-pV5Z.

### 3.2. Composite wave function theory approaches

Composite wave function theory has paved the way for cost-effective computations without significantly compromising accuracy. Various classes of additivity schemes

have been studied previously, successfully predicting geometrical parameters [59, 60], rotational constants [61], and vibrational frequencies [58, 62, 63, 70, 71, 128].

The cost-effectiveness in these methods is achieved by combining various levels of electron correlation treatments using additive approximations, basis set extrapolations, and, when applicable, empirical corrections. By way of illustration, consider the following expression:

$$E = MP2/LARGE + [CCSD(T)/SMALL - MP2/SMALL]$$

Now, if we express MP2 as  $HF + E2$  and CCSD(T) as  $HF + E2 + HLC$ , we can simplify it further, as

$$E = HF/LARGE + E2/LARGE + HF/SMALL + E2/SMALL + HLC/SMALL - HF/SMALL - MP2/SMALL$$

where SMALL and LARGE correspond to two different basis set sizes. Simplifying this equation, we get

$$E = HF/LARGE + E2/LARGE + HLC/SMALL.$$

Therefore, the basis set for HF is effectively the same as LARGE for the MP2 correlation contribution, ensuring there is no mismatch as the other HF contributions cancel out.

Table 4 details error statistics for harmonic frequencies in the HFREQ2014 species, comparing them to calculated and experimental data. Detailed equations of our G4-type composite schemes have been previously provided in Refs. [35, 36] and the top performing G4-n, cc-G4-n, and G4-n-F12 methods in prediction of reaction

**Table 4.** Root-mean-square deviations and mean absolute deviations ( $\text{cm}^{-1}$ ) of calculated harmonic frequencies with various composite wave function schemes from computed CCSD(T\*)(F12\*)/VQZ-F12 calculations (referred to as CCSD(T)/CBS) and experiment for the HFREQ2014 dataset.

cWFT method	components				Errors relative to						analytic derivatives <sup>a</sup>	
	MP2	$N_{bas}$	CCSD(T)	$N_{bas}$	CCSD(T)/CBS		Expt.		$c_1^b$	$c_2 = c_3^c$		
					RMSD	MAD	RMSD	MAD				
G4-D-v0	def2-SVSP,TZVPPD	36, 142	def2-SVSP	36	26.07	21.99	27.29	22.62	−0.07	1.05	G	G
G4-D	def2-T,QZVPPD	142, 258	def2-SVP	48	11.53	8.28	10.49	7.69	0.00	0.96	G	G
G4-D-v2	ditto	ditto	def2-SVPD	72	8.20	6.13	7.83	5.83	0.00	0.96	G	G
G4-T	ditto	ditto	def2-TZVP	86	4.65	3.32	5.41	4.35	0.57	1.05	G	G
G4-T	ditto	ditto	def2-TZVP	86	4.62	3.34	5.37	4.23	0.57	1.05	H	E
G4-D'	def2-T,QZVPPD'	130, 246	def2-SVP	48	11.54	8.32	10.52	7.71	0.00	0.96	G	G
G4-D'-v2	ditto	ditto	def2-SVPD	72	8.23	6.16	7.87	5.86	0.00	0.96	G	G
G4-T'	ditto	ditto	def2-TZVP	86	4.74	3.45	5.49	4.42	0.57	1.05	G	G
G4-T'	ditto	ditto	def2-TZVP	86	4.73	3.46	5.53	4.40	0.57	1.05	H	E
cc-G4(FC)-D	awCVT,QZ	210,402	def2-SVP	48	13.54	9.93	12.95	9.81	1.09	1.12	G	G
cc-G4-D	awCVT,QZ + CV	ditto	def2-SVP	48	11.49	7.80	11.29	7.84	1.09	1.12	G	G
cc-G4(FC)-D-v2	awCVT,QZ	ditto	def2-SVPD	72	9.10	7.06	9.28	7.07	1.09	1.12	G	G
cc-G4-D-v2	awCVT,QZ + CV	ditto	def2-SVPD	72	6.91	4.70	7.77	5.56	1.09	1.12	G	G
cc-G4(FC)-T	awCVT,QZ	ditto	def2-TZVP	86	5.30	4.24	5.45	4.36	0.63	1.03	G	G
cc-G4-T	awCVT,QZ + CV	ditto	def2-TZVP	86	4.25	3.05	5.07	3.84	0.63	1.03	G	G
G4-T <sub>ano</sub> -v1	ano-pVQZ	230	ano-pVTZ	116	6.65	4.49	6.55	4.65	−	−	G	G
G4-T <sub>ano</sub> -v2	ano-pV5Z	302	ano-pVTZ	116	4.98	3.75	5.14	3.93	−	−	G	G
cWFT method	MP2-F12	$N_{bas}$	CCSD(T)	$N_{bas}$	RMSD	MAD	RMSD	MAD	$c_1$	$c_2 = c_3$	MP2-F12	CCSD(T)
G4-T <sub>ano</sub> -F12-v0	VDZ-F12	96	ano-pVTZ	116	10.92	7.73	9.94	7.84	−	−	G	G
G4-D <sub>ano</sub> -F12-v0	VDZ-F12	96	ano-pVDZ	48	13.54	9.94	12.27	9.58	−	−	G	G
G4-T <sub>ano</sub> -F12-v1	VTZ-F12	178	ano-pVTZ	116	4.97	3.56	5.21	3.93	−	−	G	G
G4-D <sub>ano</sub> -F12-v1	VTZ-F12	178	ano-pVDZ	48	7.99	5.72	7.36	5.62	−	−	G	G
G4-T <sub>ano</sub> -F12-v2	VQZ-F12	310	ano-pVTZ	116	3.74	2.93	4.58	3.45	−	−	G	G
G4-D <sub>ano</sub> -F12-v2	VQZ-F12	310	ano-pVDZ	48	7.04	5.29	6.72	5.16	−	−	G	G
W1 (Val.)					1.96	1.40	4.76	3.54			G	G
W2 (Val.)					1.15	0.70	4.67	3.26			G	G

<sup>a</sup>G: Gradient, H: Hessian, and E: Energy is doubly differentiated numerically.<sup>b</sup> $c_1$  is the extrapolation coefficient of MP2 correlation energies.<sup>c</sup> $c_2$  and  $c_3$  are scaling parameters for  $E_{CCSD-MP2}$  and  $E_{(T)}$  terms, respectively.

energetics are now validated for harmonic frequencies. Some of these approaches are parameter-free while others achieve optimal results with a maximum of two fitted parameters. In addition, we consider the performance of W1<sub>val</sub> and W2<sub>val</sub> theories, i.e. Weizmann-n theories [25–27] including only valence correlation.

It can be rightly argued that, beyond small molecules where spectral inversion is comparatively easy, fundamental frequencies are more relevant for practical applications than harmonic frequencies. However, as shown in Table 4 of Ref. [129], even for CCSD(T)/CBS simple scaling of harmonic frequencies carries an intrinsic error of about 25  $\text{cm}^{-1}$ , comparable to the uncertainty in hybrid DFT harmonic frequencies. The use of dual or multiple scaling factors for different frequency ranges at semi-arbitrary cutoff points (e.g. [130–133]) is a half-measure at best; second-order rotation-vibration perturbation theory (VPT2) [107] will require a semidiagonal quartic force field. Schneider and Thiel [134] as far back as 1989 (in the context of semiempirical MO theory) pointed out that all the required force constants can be obtained by finite differences (in normal coordinates) of analytical second derivatives: this may be a viable approach for the present cWFT methods, or cWFT

harmonic frequencies may be combined with DFT anharmonic force fields as demonstrated by Boese and Martin for the azabenzenes.

### 3.2.1. G4-type cWFTs based on CCSD(T)/def2-TZVP

The most accurate result, using a def2 basis set for the CCSD(T) part, materialises for G4-T with  $\text{RMSD}_{\text{CCSD(T)/CBS}}$  of 4.7  $\text{cm}^{-1}$  and  $\text{RMSD}_{\text{exp}}$  of 5.4  $\text{cm}^{-1}$ . In other words, not materially different from CCSD(T) at the valence CBS limit. Fundamentally, the same result is obtained if the force constants are obtained fully numerically at CCSD(T) and analytically at MP2. The lower-cost approach, G4-D, based on def2-SVP in CCSD(T), is three times worse in accuracy, while in G4-D-v2 with def2-SVPD, the error statistics are cut in half. The largest errors occur in the  $\pi_g$  and  $\pi_u$  degenerate bending modes of acetylene, which G4-D and G4-D-v2 underestimate by 26.9 and 39.9  $\text{cm}^{-1}$ , while the accurate G4-T falls short by only 0.7  $\text{cm}^{-1}$ .

An important technical note for wave function calculations using basis sets augmented with diffuse functions, particularly for linear molecules, is that employing such basis sets, e.g. def2-nZVPPD ( $n = \text{T, Q}$ ) for acetylene,



may result in an overlap matrix  $S$  with very small eigenvalues. Gaussian implements a form of SVD (singular value decomposition) in which the eigenvectors of  $S$  with eigenvalues below a cutoff (default value:  $10^{-6}$ ) are discarded. This leads to truncation of the virtual orbital space, which may not be consistent across the surface – or even along a single normal mode displacement – and hence may cause erratic harmonic frequencies. In the present work, this occurred for the bending frequencies of acetylenes. To address this, we disabled the ‘SVD screening’ in Gaussian using `IOp(3/32)=2`; MOLPRO has no such screening in the first place, but for acetylene, `def2-nZVPPD` led to  $S$  eigenvalues below the default `THROVL` =  $10^{-8}$ , and lowering `THROVL` brought on numerical issues that required severely tightening the integral evaluation cutoffs. No such problems were seen if the ‘D’ functions were retained on the heavy atoms but not on H, which we denote `def2-nZVPPD'` and applied for acetylene.

Substituting  $E_{2/\text{def2-T,QZVPPD}}$  in G4 type approaches, where diffuse functions are omitted on hydrogen atoms, has no significant effect on error statistics when compared to reference or experiment. G4-T' stands out as the most accurate with an  $\text{RMSD}_{\text{CBS}}$  of  $4.74\text{ cm}^{-1}$ , a mere  $0.11\text{ cm}^{-1}$  above regular G4-T. Moreover, opting for `def2-T,QZVPPD'` in G4-type approaches effectively addresses concerns related to small eigenvalues in the overlap matrix, particularly for linear molecules, without compromising accuracy.

In our initial studies on G4-like cWFTs [35, 36] we did not explore a D,T extrapolation in MP2, like  $E_{2/\text{def2-SVSP,TZVPPD}} + [\text{CCSD(T)-MP2}]/\text{def2-SVSP}$ , where `def2-SVSP` means no p functions on hydrogen atoms, assuming it would be less accurate than other cWFTs. We trained such a G4-D-v0 composite on the GMTKN55 dataset and found a high WTMAD2 (weighted mean absolute deviation) value of  $4.77\text{ kcal/mol}$  compared to  $\text{CCSD(T)}/\text{CBS}$  data, extracted from the ACCDB database [135], or higher level from our earlier work. The RMSD values for harmonic frequencies are unacceptably high, reaching  $26.07$  and  $27.29\text{ cm}^{-1}$  compared to  $\text{CCSD(T)}/\text{CBS}$  and experimental values: this is on par with (much cheaper) hybrid DFT functionals such as B3LYP and TPSS0 [129]. However, standard  $\text{CCSD(T)}/\text{def2-SVSP}$  is far less accurate than G4-D-v0, with corresponding RMSDs of  $61.37$  and  $60.46\text{ cm}^{-1}$ . Clearly, there is no advantage in a D,T extrapolation and if  $10\text{ cm}^{-1}$  errors are acceptable then empirical spin-scaled double-hybrid functionals, such as DSD-PBEP86-D3(BJ) [136] and `revDSD-PBEP86-D4` [137], represent a much more cost-effective alternate.

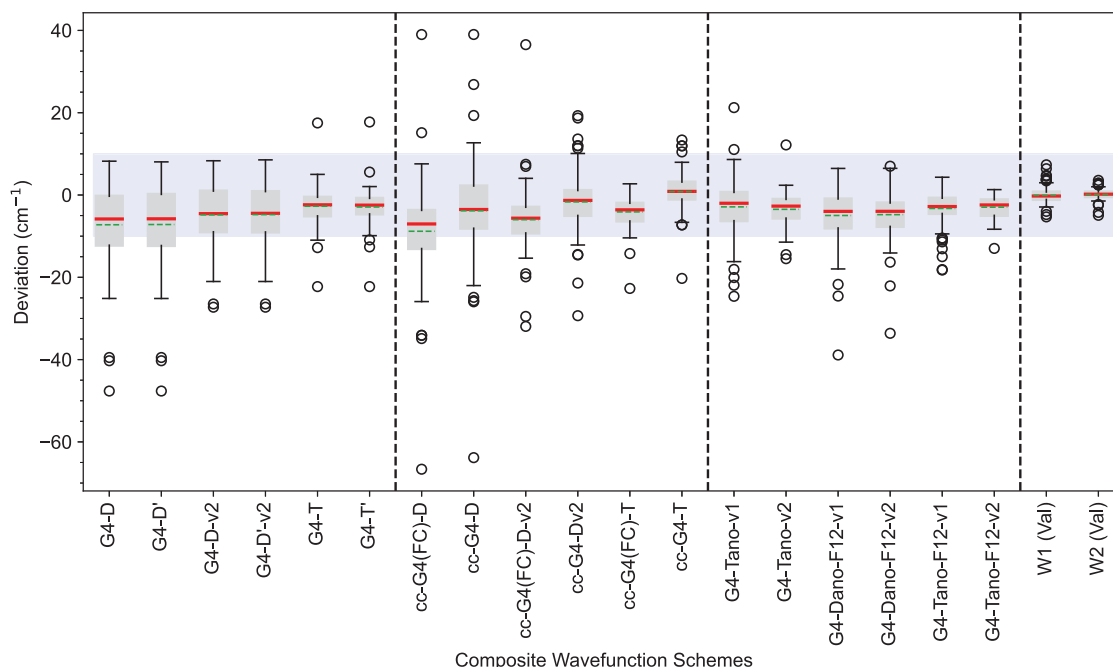
Additionally, we examined the impact of using VDZ-F12 in MP2-F12 instead of VTZ-F12. We saw RMSD

values relative to  $\text{CCSD(T)}/\text{CBS}$  rise high as  $13.54$  and  $10.92\text{ cm}^{-1}$  with MP2-F12/VDZ-F12 and either  $[\text{CCSD(T)-MP2}]/\text{ano-pVDZ}$  or `ano-pVTZ`, respectively, which is on par with the performance of double hybrid functionals.

What about the effects of core-core and core-valence correlation? To answer that, we consider the correlation consistent augmented `aug-cc-pwCVnZ` basis sets that provide the necessary radial and angular flexibility in the core-valence region. In 2018, Sylvetsky and Martin [138] showed that a `awCVT,QZ` basis set extrapolation at  $\text{CCSD(T)}$  proved sufficient and captured the most significant part of electron correlation. Here, we follow a two-step approach to assess those effects. Firstly, we check the effects of increased basis set radial flexibility for valence correlation by replacing `def2-T,QZVPPD` with `awCVT,QZ` at the MP2 level, while only correlating the valence electrons. In doing so, no material gains in accuracy are seen, but rather the reverse. The  $\text{RMSD}_{\text{CCSD(T)}/\text{CBS}}$  increases by  $2\text{ cm}^{-1}$  for `cc-G4(FC)-D`,  $0.9\text{ cm}^{-1}$  for `cc-G4(FC)-D-v2`, and  $0.6\text{ cm}^{-1}$  for `cc-G4-T`, each compared to the corresponding G4-n. Secondly, we correlate all electrons in MP2 and the results showed the CV correlation improves the RMSD by  $2.0\text{ cm}^{-1}$  in `cc-G4-D` relative to `cc-G4(FC)-D`, with the former achieving identical accuracy to G4-D. Further improvement of  $2.2\text{ cm}^{-1}$  occurs with `def2-SVPD` in  $\text{CCSD(T)}$ . The lowest RMSD of  $4.25\text{ cm}^{-1}$  is found for `cc-G4-T`, this result represents a  $1.0\text{ cm}^{-1}$  amelioration over valence-only `cc-G4(FC)-T`, and even surpassing G4-T by  $0.3\text{ cm}^{-1}$ .

For higher accuracy regimes, we refer the reader to our recent study [58] on ground-state spectroscopic constants of diatomic molecules from post- $\text{CCSD(T)}$  up to  $\text{CCSDTQ56}$ . We showed there that  $2\text{ cm}^{-1}$  accuracy is achievable on a semi-routine basis (see Table 5 in ref. [58]), but that this requires *both* post- $\text{CCSD(T)}$  valence correlation correction at least at the  $\text{CCSDT(Q)}_{\Lambda}$  level *and* core-valence correlation corrections at the  $\text{CCSD(T)}$  level. (Including each on its own will actually make agreement *worse*, as valence  $\text{CCSD(T)}$  benefits from a felicitous error compensation.) We have repeated the Dunham analyses [139] from Ref. [58] without the scalar relativistic correction. The largest individual difference in  $\omega_e$  is seen for HCl ( $-4.3\text{ cm}^{-1}$ ) followed by  $-2.8\text{ cm}^{-1}$  for HF, but most effects are on the order of  $1\text{ cm}^{-1}$  or less. Thus, the RMSD on  $\omega_e$  increased only mildly, from  $2.10$  to  $2.55\text{ cm}^{-1}$ , while the effect on other spectroscopic constants was negligible: obviously, compared to  $5$ – $10\text{ cm}^{-1}$  RMSDs for more approximate methods, such an increase is entirely negligible. Needless to say, this will no longer be the case for heavy p-block compounds.

We attempted to add diagonal Born-Oppenheimer corrections at the  $\text{CCSD}/\text{aug-cc-pVTZ}$  level using the



**Figure 3.** Box-and-whisker plot for harmonic frequency deviations of composite wave function schemes from CCSD(T\*)(F12\*)/VQZ-F12 for the HFREQ2014 dataset. Plot description details are the same as for Figure 2.

implementation [140] in CFOUR [141]. While corrections may exceed  $1 \text{ cm}^{-1}$  for  $\text{H}_2$ ,  $\text{BH}$ , and the like, for heavier diatomics they are negligible compared to other remaining error sources.

### 3.2.2. ANO basis sets for G4-type cWFTs

ano-pVnZ basis sets are now considered in the CCSD(T) part of cWFTs for the harmonic frequencies in HFREQ 2014. The energy expressions of these cWFTs are simply derived from the sum of the total MP2 energy with a larger basis set and the higher level [CCSD(T)-MP2]/ano-pVTZ correction, without introducing empirical parameters. Using  $E_{2/\text{ano-pVQZ}}$  in G4-T<sub>ano</sub>-v1, we find an RMSD of  $6.6 \text{ cm}^{-1}$  w.r.t both reference and experiment (see Table 4 and Figure 3). Closer agreement to CCSD(T) at the complete basis set limit is achieved with ano-pV5Z in MP2, leading to the G4-T<sub>ano</sub>-v2 variant, which shows better  $\text{RMSD}_{\text{exp}}$  than G4-T' and G4-T by approximately  $0.3 \text{ cm}^{-1}$ .

In our earlier work [35], we found that MP2-F12-based methods, such as the parameter-free cc-G4-F12-T, yielded the lowest WTMAD2 values ( $\sim 1.0 \text{ kcal/mol}$ ) for the energetics of the GMTKN55 benchmark suite [37]. That prompts the question whether the predicted harmonic frequencies can become more accurate by substituting explicitly correlated MP2-F12 for conventional MP2 in composite energy schemes. Among the tested MP2-F12-based variants, G4-T<sub>ano</sub>-F12-v2 is the most accurate with an RMSD of  $3.7 \text{ cm}^{-1}$  relative to reference

calculated frequencies, compared to  $7 \text{ cm}^{-1}$  for G4-D<sub>ano</sub>-F12-v2. Reducing the basis set size to  $n = \text{T}$  from  $n = \text{Q}$  in MP2-F12/VnZ-F12 worsens  $\text{RMSD}_{\text{CCSD(T)}/\text{CBS}}$  by  $1.2 \text{ cm}^{-1}$  for G4-T<sub>ano</sub>-F12 and by  $0.95 \text{ cm}^{-1}$  for G4-D<sub>ano</sub>-F12. Clearly, combining MP2-F12 correlation with CCSD(T)/ano-pVnZ presents an attractive option for accurate vibrational frequencies in parameter-free cWFTs. The calculated frequencies can be inspected in the Supporting Information.

Finally, we note that the Weizmann-n methods, W1 and W2, lead to the lowest RMSDs relative to the calculated reference harmonic frequencies, at  $1.96$  and  $1.15 \text{ cm}^{-1}$ , respectively. There is no systematic improvement in  $\text{RMSD}_{\text{exp}}$  values over the most accurate ANO-based method, indicating that the convergence towards CCSD(T)/CBS has been achieved.

### 3.3. Harmonic frequencies of benzene: successes and limitations of cWFTs

In 1997, Martin, Taylor, and Lee [76] computed the CCSD(T) geometry and harmonic force field of  $\text{C}_6\text{H}_6$ , and noted that the two out-of-plane ring modes  $\omega_4$  and  $\omega_5$  exhibit a more pronounced form of the same hypersensitivity as seen for the acetylene bending frequencies [125] and traced to intramolecular BSSE in Ref. [126]. (Moran et al. [127] later extended this discussion to benzene.) Limitations of available computers at the time precluded going to larger basis sets such as haVQZ, but since ANOs minimise the BSSE for a given contracted size,

**Table 5.** Calculated and experimentally derived harmonic frequencies (in  $\text{cm}^{-1}$ ) for the benzene molecule.

		Expt. Ref. [142]	CCSD(T)/			CCSD(T*)(F12*)/			MP2-F12/		
			ano-pV5Z	ano-pVQZ	ano-pVTZ	VT,QZ-F12	VQZ-F12	VTZ-F12	VDZ-F12 +[CCSD(T)-MP2]/def2-TZVP	VTZ-F12	VQZ-F12
$\omega_1$	$a_{1g}$	1010.0	1006.6	1005.5	1003.3	1006.3	1006.5	1007.1	1006.3	1007.8	1005.6
$\omega_2$	$a_{1g}$	3218.0	3204.3	3208.0	3209.9	3205.0	3205.4	3206.5	3209.0	3208.9	3204.9
$\omega_3$	$a_{2g}$	1392.0	1380.5	1378.3	1374.9	1382.8	1382.0	1379.7	1374.0	1378.1	1380.1
$\omega_4$	$b_{2g}$	717.0	712.4	711.6	708.4	709.1	705.4	695.2	679.7	687.4	695.7
$\omega_5$	$b_{2g}$	1012.0	1012.4	1010.1	1006.8	1013.4	1009.5	998.7	977.5	992.4	998.9
$\omega_6$	$e_{2g}$	617.0	611.5	611.0	610.3	611.8	611.8	611.7	610.4	610.6	610.6
$\omega_7$	$e_{2g}$	3210.0	3179.6	3181.5	3184.6	3180.6	3180.9	3181.8	3184.1	3183.9	3179.8
$\omega_8$	$e_{2g}$	1645.0	1639.3	1638.5	1637.0	1639.5	1639.6	1639.9	1640.2	1640.6	1638.4
$\omega_9$	$e_{2g}$	1197.0	1191.8	1191.4	1192.3	1192.1	1192.1	1192.2	1192.5	1190.5	1191.1
$\omega_{10}$	$e_{1g}$	861.0	863.4	863.1	863.6	863.5	862.7	860.5	856.4	857.6	860.4
$\omega_{11}$	$a_{2u}$	683.0	683.6	683.9	685.9	683.6	683.2	682.1	676.3	678.6	681.1
$\omega_{12}$	$b_{1u}$	1030.0	1024.6	1022.9	1019.4	1025.7	1025.4	1024.5	1022.3	1022.1	1022.4
$\omega_{13}$	$b_{1u}$	—	3169.9	3171.4	3174.4	3171.1	3171.4	3172.3	3173.2	3174.4	3170.3
$\omega_{14}$	$b_{2u}$	1338.0	1328.6	1329.1	1328.3	1329.1	1329.7	1331.4	1319.9	1323.9	1321.1
$\omega_{15}$	$b_{2u}$	1163.0	1158.4	1158.5	1160.5	1158.4	1158.5	1158.9	1158.0	1156.3	1156.9
$\omega_{16}$	$e_{2u}$	406.0	406.3	406.4	405.8	405.6	405.3	404.6	399.6	403.2	403.5
$\omega_{17}$	$e_{2u}$	987.0	987.0	985.3	983.0	987.2	985.9	982.3	977.5	984.0	986.0
$\omega_{18}$	$e_{1u}$	1057.0	1057.2	1056.2	1054.8	1057.4	1057.4	1057.5	1055.0	1057.0	1056.2
$\omega_{19}$	$e_{1u}$	1522.0	1511.3	1509.5	1506.6	1512.6	1512.5	1512.2	1513.2	1511.6	1511.2
$\omega_{20}$	$e_{1u}$	3212.0	3195.1	3197.9	3200.7	3195.9	3196.3	3197.3	3199.2	3199.8	3195.5
RMSD (cm <sup>−1</sup> )		REF		1.53	3.62	0.99	1.64	4.45	9.93	6.65	4.38
without $\omega_4, \omega_5$		REF		1.52	3.51	0.79	0.90	1.99	4.90	3.26	2.02
			G4-T	MP2-F12/ VDZ-F12	MP2-F12/ VTZ-F12 +[CCSD(T)-MP2]/ano-pVTZ <sup>a</sup>			MP2-F12/ VQZ-F12	MP2/ ano-pVQZ +[CCSD(T)-MP2]/ano-pVTZ	MP2/ ano-pV5Z +[CCSD(T)-MP2]/ano-pVTZ	
$\omega_1$	$a_{1g}$	1018.0		1006.3 (0.1)		1007.8 (0.1)		1009 (−0.6)	1007.8		1010.1
$\omega_2$	$a_{1g}$	3201.4		3209.8 (−0.6)		3209.7 (−0.6)		3205.7 (−0.6)	3207.7		3202.7
$\omega_3$	$a_{2g}$	1380.0		1373.3 (0)		1377.4 (0)		1379.4 (0)	1376.3		1378.4
$\omega_4$	$b_{2g}$	738.8		685.4 (−0.6)		695.5 (−0.6)		703.4 (−0.7)	709.5		710.1
$\omega_5$	$b_{2g}$	1004.5		986.5 (0)		1002.3 (−0.6)		1005.6 (2.9)	1004.4		1005.0
$\omega_6$	$e_{2g}$	608.7		610.7 (0)		610.9 (0)		610.9 (0)	610.1		610.5
$\omega_7$	$e_{2g}$	3172.9		3185 (−0.7)		3184.8 (−0.7)		3180.6 (−0.8)	3180.7		3177.5
$\omega_8$	$e_{2g}$	1636.7		1639.8 (0.4)		1640.2 (0.4)		1638 (0.4)	1637.1		1637.1
$\omega_9$	$e_{2g}$	1190.3		1192.2 (−0.2)		1190.2 (−0.2)		1190.8 (−0.2)	1190.3		1190.4
$\omega_{10}$	$e_{1g}$	861.3		857.8 (−0.6)		859 (−0.6)		861.8 (−0.6)	861.7		861.9
$\omega_{11}$	$a_{2u}$	680.9		681 (−0.6)		680 (−0.6)		682.4 (−0.6)	682.9		682.5
$\omega_{12}$	$b_{1u}$	1028.8		1023 (−0.3)		1022.9 (−0.3)		1023.1 (−0.3)	1021.0		1022.5
$\omega_{13}$	$b_{1u}$	3167.0		3174.1 (−0.7)		3175.2 (−0.7)		3171.2 (−0.7)	3170.5		3167.8
$\omega_{14}$	$b_{2u}$	1322.1		1323.7 (−4.8)		1327.7 (−4.9)		1324.9 (−4.8)	1321.7		1319.6
$\omega_{15}$	$b_{2u}$	1155.4		1157.7 (−0.9)		1156 (−0.9)		1156.6 (−0.9)	1156.6		1156.0
$\omega_{16}$	$e_{2u}$	401.7		400 (−0.5)		403.5 (−0.5)		403.8 (−0.5)	405.3		404.9
$\omega_{17}$	$e_{2u}$	989.5		976.6 (−0.5)		983.2 (−0.4)		985.2 (−0.4)	983.4		985.0
$\omega_{18}$	$e_{1u}$	1055.5		1055 (−0.1)		1057 (−0.1)		1056.2 (−0.1)	1055.1		1055.7
$\omega_{19}$	$e_{1u}$	1508.9		1512.3 (0.2)		1510.7 (0.2)		1510.3 (0.2)	1507.6		1509.1
$\omega_{20}$	$e_{1u}$	3191.2		3200 (−0.7)		3200.6 (−0.7)		3196.3 (−0.7)	3197.3		3193.2
RMSD (cm <sup>−1</sup> )		6.37		8.18 (0.25)		4.77 (0.22)		2.57 (0.31)	2.98		2.82
without $\omega_4, \omega_5$		4.04		4.67 (0.32)		3.25 (0.16)		1.58 (0.65)	2.63		2.52

(a) Values in parentheses indicate deviations between cWFTs having their [CCSD(T)-MP2] term scaled by 1.04382 and parameter-free cWFTs ( $c_{[\text{CCSD(T)}-\text{MP2}]} = 1$ ).

ANO4321 was attempted and found to be resilient than cc-pVTZ.

As extracting a full set of experimental harmonic frequencies and anharmonicity constants for such a large molecule would require a staggering number of vibrational band origins, the available ‘experimental’ harmonic frequencies of Miani et al. [142] are in truth semi-experimental (a term introduced in the spectroscopic realm by Jean Demaison [143]), namely, from combining experimental fundamentals with a DFT calculated quartic force field. (We note in passing that results

of a ‘blind challenge’ on the ground-state correlation energy of benzene were recently reported [144].)

Hence benzene would appear to be a good ‘proof of concept’ for the application of composite WFTs to harmonic frequencies of not-so-small molecules. Owing to the high symmetry, we can actually carry out CCSD(T) and CCSD(T\*)(F12\*) calculations close to the basis set limit, giving us a realistic reference. Any cWFT that would reproduce the harmonic frequencies well might be a good candidate for an anharmonic force field.

Table 5 showcases the calculated harmonic frequencies using ‘pure’ coupled-cluster methods and their corresponding cWFTs. For the reference, we consider CCSD(T)/ano-pV5Z harmonic frequencies. Geometry optimizations were carried out in  $D_{2h}$  point group symmetry and the Hessian was obtained through the method of finite differences.

Overall, the RMSDs consistently improve with larger basis sets, ranging from  $3.6\text{ cm}^{-1}$  for CCSD(T)/ano-pVTZ to  $1.5\text{ cm}^{-1}$  for ano-pVQZ. CCSD(T\*)(F12\*)/cc-pVTZ-F12 has an RMSD of  $4.5\text{ cm}^{-1}$ , which drops to  $1.6\text{ cm}^{-1}$  for cc-pVQZ-F12. Much of that is due to the two problematic  $\omega_4$  and  $\omega_5$  modes, however: conspicuous discrepancies of  $-17$  and  $-14\text{ cm}^{-1}$ , respectively, are observed at the CCSD(T\*)(F12\*)/VTZ-F12 level; these however are drastically reduced to  $\{-7, -3\}\text{ cm}^{-1}$ , for VQZ-F12, and still further to  $\{-3, +1\}\text{ cm}^{-1}$  for VT,QZ-F12 extrapolation. Statistics without them are RMSD =  $1.9$  for VTZ-F12 and  $0.9\text{ cm}^{-1}$  for VQZ-F12, which is more in line with HFREQ2014. A T,Q extrapolation (with an exponent of 4.5960, Table X in Hill et al. [90]) yields RMSD =  $1.0\text{ cm}^{-1}$  including all modes, and  $0.8\text{ cm}^{-1}$  excluding  $\omega_4$  and  $\omega_5$ .

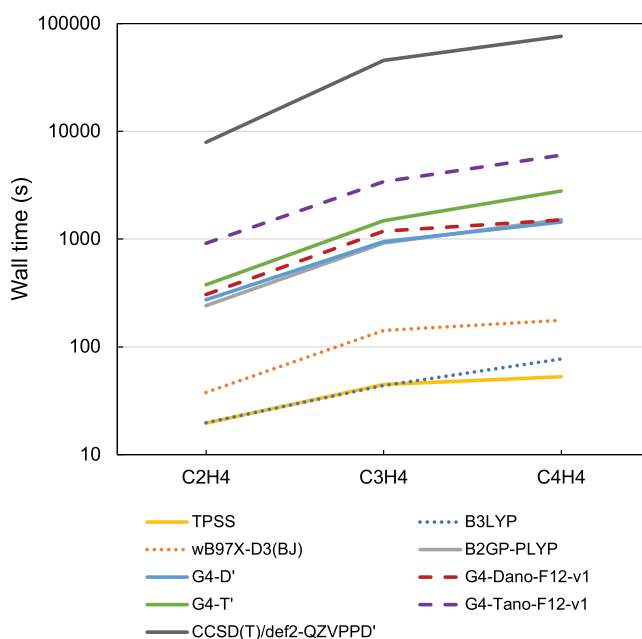
Next, we estimate the accuracy of composite schemes for benzene. The most accurate results are obtained in parameter-free methods based on a high level correction [CCSD(T)-MP2/ano-pVTZ] that is combined with either  $E_{2/\text{ano-pVQZ}}$  (RMSD =  $3.0\text{ cm}^{-1}$ ) or  $E_{2/\text{ano-pV5Z}}$  (RMSD =  $2.8\text{ cm}^{-1}$ ). Similar gains are observed for the MP2-F12-based G4-T<sub>ano</sub>-F12-v2, which exhibits an RMSD of  $2.6\text{ cm}^{-1}$ , and this result is  $1.0\text{ cm}^{-1}$  lower than plain CCSD(T)/ano-pVTZ. Scaling the triples term  $E_{(T)}$  does more harm than good, resulting in an increase of  $\sim 0.2\text{ cm}^{-1}$  for combinations of MP2-F12/VnZ-F12 with CCSD(T)/ano-pVTZ.

For the low-cost G4-T, with an RMSD of  $6.4\text{ cm}^{-1}$ , the  $\omega_4$  and  $\omega_5$  modes are underestimated by  $-27$  and  $-8\text{ cm}^{-1}$ , respectively. Introducing MP2-F12 correlation, instead of conventional MP2, leads to a slight deterioration of  $0.2\text{ cm}^{-1}$  with VTZ-F12, but a significant improvement of  $2.0\text{ cm}^{-1}$  occurs with VQZ-F12.

Consequently, for accurate harmonic frequencies it is recommended to combine MP2-F12/VnZ-F12 ( $n = \text{T or Q}$ ) with CCSD(T)/ano-pVTZ, as such parameter-free cWFTs offer substantial gains.

### 3.4. Timing comparison of cWFT and DFT methods

The reviewers requested a timing comparison of the present and alternative approaches. Figure 4 depicts ‘wall clock’ times of selected cWFT and DFT methods for semi-numerical evaluation of harmonic frequencies obtained using MOLPRO from non-stationary



**Figure 4.** Visual representation of wall clock times for selected cWFT and DFT methods. Note that the y axis is logarithmic.

geometries. All calculations ran on identical architecture nodes using 16 physical cores of an Intel Xeon Gold 5320 CPU with a maximum memory of 46.4 GB per core.

A notable speedup occurs with increasing molecular size when using cWFTs compared to ‘brute force’ CCSD(T) in the largest basis set of the composite scheme, the latter achieves speedups for cyclobutadiene by a factor of 53 for G4-D’ and 27 for G4-T’. For ethylene, those ratios decrease to 29 and 21, respectively.

The MP2-F12-based method G4-D<sub>ano</sub>-F12-v1 is on par with G4-D’. However, the former exploits the accelerated basis set convergence at the MP2-F12 level, resulting to a  $3.55\text{ cm}^{-1}$  improvement in RMSD for HFREQ2014. Moreover, using the triple- $\zeta$  basis set VTZ-F12 in MP2-F12 makes G4-T<sub>ano</sub>-F12-v1 nearly three times more expensive than with VDZ-F12. Despite the increased cost, G4-T<sub>ano</sub>-F12-v1 is almost an order of magnitude less expensive than CCSD(T)/def2-QZVPPD’.

To conclude, we assess the computational costs of DFT for harmonic frequencies alongside cWFTs; in order to keep the timing comparison fair, we carried out these calculations using MOLPRO. Specifically, as representative examples for timing, we picked the meta-GGA functional TPSS [145] and the hybrid functionals B3LYP [146, 147] and  $\omega\text{B97X-D3(BJ)}$  [148], as well as the double-hybrid B2GP-PLYP [149], which will have a similar cost as DSD-PBEP86-D3(BJ) [136] or revDSD-PBEP86-D4 [137] (which presently do not have analytical gradients in MOLPRO). Notably, G4-D’ and B2GP-PLYP exhibit similar costs. Hybrid GGA functionals, such as



$\omega$ B97X-D3(BJ) and  $\omega$ B97X-V [150] present the lowest-cost option, if ca.  $30\text{ cm}^{-1}$  accuracy is acceptable.

#### 4. Conclusions

In this study, we have thoroughly examined the performance of various coupled-cluster composite wave function approaches (cWFT) for harmonic frequencies. Our investigation involved the development of extrapolation formulas for force constants, while also enabling geometry optimisation and harmonic frequency calculations through an implementation we provide.

We have validated the top-performing composite energy schemes, based on previous evaluations for reaction energetics in Refs. [35, 36] using the large GMTKN55 test suite, against the harmonic frequencies in the HFREQ2014 dataset from CCSD(T\*)(F12\*)/VQZ-F12 calculations and experimental data. G4-T is three times more accurate than plain CCSD(T)/def2-TZVP, while G4-T<sub>ano</sub> is twice as accurate as CCSD(T)/ano-pVTZ. Notably, ANO basis sets combined with explicitly correlated MP2-F12, such as G4-T<sub>ano</sub>-F12, show promising performance, achieving accuracy of  $5\text{ cm}^{-1}$  compared to the experiment, and they are on par with the accuracy of CCSD(T) at the complete basis set limit. Following closely was our standard G4-T approach, built upon def2-TZVP for the high level correction. Additionally, the Weizmann-n theories, W1 and W2, delivered the most accurate results when compared to the calculated reference harmonic vibrational frequencies.

The addition of diffuse functions on hydrogen does not materially help performance for neutral molecules, and in fact causes significant near-linear-dependence issues. In codes that eliminate ‘near-singular’ eigenvectors of the overlap matrix (i.e. those for which the eigenvalue drops below a threshold), adding superfluous basis functions in general – and diffuse functions where they are unneeded in particular – can cause discontinuities on a correlated potential energy surface as orbitals drop in and out of the virtual space. When carrying out (semi)numerical frequency calculations, this can cause erratic results, as we observed here for acetylene.

In summary, we recommend the following:

- If an accuracy of  $20\text{--}30\text{ cm}^{-1}$  is sufficient, or if anharmonicity’s deviation from a simple scaling factor exceeds that level (and an anharmonic force field is not a practical option), then consider a DFT option such as  $\omega$ B97M-V [151] or the even more economical B97M-V [152].
- If  $10\text{ cm}^{-1}$  is satisfactory, an empirical double hybrid like DSD-PBEP86 or revDSD-PBEP86 may be the right choice.
- For  $4\text{--}5\text{ cm}^{-1}$  accuracy, consider present G4-type approaches including G4-T<sub>ano</sub>-v2 and G4-T<sub>ano</sub>-F12-v1. Both methods share the same high-level correction [CCSD(T)-MP2]/ano-pVTZ, but G4-T<sub>ano</sub>-v2 is combined with MP2/ano-pV5Z and G4-T<sub>ano</sub>-F12-v1 uses MP2-F12/VTZ-F12. These two composite schemes offer similar accuracy and computational cost, and are suitable for larger molecules, particularly if analytic second derivatives at MP2 and CCSD(T) are available. It is worth noting that no empirical scaling parameters were employed in these top-performing approaches.
- For higher accuracy within the range of  $1\text{--}2\text{ cm}^{-1}$ , it is important to extend beyond CCSD(T) as well as consider relativistic effects and the impact of diagonal Born-Oppenheimer corrections, especially in the case of hydrides.

#### Acknowledgments

The authors would like to thank Dr. Mark Vilensky for technical assistance with ChemFarm, and Dr. Margarita Shepelenko for critical reading of the manuscript prior to submission. ES thanks the Feinberg Graduate School (Weizmann Institute of Science) for a doctoral fellowship and the Onassis Foundation (Scholarship ID: FZP 052-2/2021-2022).

#### Disclosure statement

No potential conflict of interest was reported by the author(s).

#### Funding

Research at Weizmann was supported by the Israel Science Foundation (grant 1969/20), by the Minerva Foundation (grant 2020/05), and by a research grant from the Artificial Intelligence for Smart Materials Research Fund, in Memory of Dr. Uriel Arnon.

#### Supporting information and data availability statement

A Microsoft Excel Workbook with all calculated harmonic frequencies for the HFREQ2014 dataset and the benzene molecule can be found at the DOI <http://doi.org/10.34933/2b615f91-8187-47c1-a29b-86705932634a>, or the shortened URL <http://tinyurl.com/G4compfreq>. The hfreq implementation for geometry optimisation and calculation of harmonic vibrational frequencies with composite energy schemes is available on GitHub at <http://github.com/msemidalas/hfreq>. Additional data can be obtained from the authors upon reasonable request.

#### References

- [1] S. Yurchenko, *Computational Spectroscopy of Polyatomic Molecules* (CRC Press, Boca Raton, 2023).



- [2] C.R. Baiz, B. Błasiak, J. Bredenbeck, M. Cho, J. -H. Choi, S.A. Corcelli, A.G. Dijkstra, C. -J. Feng, S. Garrett-Roe, N. -H. Ge, M.W.D. Hanson-Heine, J.D. Hirst, T.L.C. Jansen, K. Kwac, K.J. Kubarych, C.H. Londergan, H. Maekawa, M. Reppert, S. Saito, S. Roy, J.L. Skinner, G. Stock, J.E. Straub, M.C. Thielges, K. Tomlinaga, A. Tokmakoff, H. Torii, L. Wang, L.J. Webb and M.T. Zanni, *Chem. Rev.* **120**, 7152–7218 (2020). doi:10.1021/acs.chemrev.9b00813
- [3] V. Barone, S. Alessandrini, M. Biczysko, J.R. Cheeseman, D.C. Clary, A.B. McCoy, R.J. DiRisio, F. Neese, M. Melosso and C. Puzzarini, *Nat. Rev. Methods Primers* **1**, 38 (2021). doi:10.1038/s43586-021-00034-1
- [4] A. Paul and T. Birol, *Annu. Rev. Mater. Res.* **49**, 31–52 (2019). doi:10.1146/annurev-matsci-070218-121825
- [5] C. Puzzarini, J. Bloino, N. Tasinato and V. Barone, *Chem. Rev.* **119**, 8131–8191 (2019). doi:10.1021/acs.chemrev.9b00007
- [6] C. Puzzarini, J.F. Stanton and J. Gauss, *Int. Rev. Phys. Chem.* **29**, 273–367 (2010). doi:10.1080/01442351003643401
- [7] M. Steinhauser and S. Hiermaier, *Int. J. Mol. Sci.* **10**, 5135–5216 (2009). doi:10.3390/ijms10125135
- [8] A. Karton and J.M. Martin, *J. Chem. Phys.* **133**, 1–17 (2010). doi:10.1063/1.3489113
- [9] J.F. Stanton and J. Gauss, *Adv. Chem. Phys.* **125**, 101–146 (2003). doi:10.1002/0471428027.ch2
- [10] K. Raghavachari, G.W. Trucks, J.A. Pople and M. Head-Gordon, *Chem. Phys. Lett.* **157**, 479–483 (1989). doi:10.1016/S0009-2614(89)87395-6
- [11] M. Kállay and J. Gauss, *J. Chem. Phys.* **123**, 214105 (2005). doi:10.1063/1.2121589
- [12] R.N. Hill, *J. Chem. Phys.* **83**, 1173–1196 (1985). doi:10.1063/1.449481
- [13] W. Kutzelnigg and J.D. Morgan, *J. Chem. Phys.* **96**, 4484–4508 (1992). doi:10.1063/1.462811
- [14] C. Schwartz, *Phys. Rev.* **126**, 1015–1019 (1962). doi:10.1103/PhysRev.126.1015
- [15] A. Karton, *Annual Reports in Computational Chemistry* **18**, 123–166 (2022). doi:10.1016/bs.arcc.2022.09.003
- [16] A.G. Baboul, L.A. Curtiss, P.C. Redfern and K. Raghavachari, *J. Chem. Phys.* **110**, 7650–7657 (1999). doi:10.1063/1.478676
- [17] L.A. Curtiss, M.P. McGrath, J. -P. Blaudeau, N.E. Davis, R.C. Binning and L. Radom, *J. Chem. Phys.* **103**, 6104–6113 (1995). doi:10.1063/1.470438
- [18] L.A. Curtiss, K. Raghavachari, G.W. Trucks and J.A. Pople, *J. Chem. Phys.* **94**, 7221–7230 (1991). doi:10.1063/1.460205
- [19] L.A. Curtiss, P.C. Redfern and K. Raghavachari, *J. Chem. Phys.* **126**, 084108 (2007a). doi:10.1063/1.2436888
- [20] L.A. Curtiss, P.C. Redfern and K. Raghavachari, *J. Chem. Phys.* **127**, 124105 (2007b). doi:10.1063/1.2770701
- [21] J.A. Pople, M. Head-Gordon, D.J. Fox, K. Raghavachari and L.A. Curtiss, *J. Chem. Phys.* **90**, 5622–5629 (1989). doi:10.1063/1.456415
- [22] J.A. Montgomery, M.J. Frisch, J.W. Ochterski and G.A. Petersson, *J. Chem. Phys.* **110**, 2822–2827 (1999). doi:10.1063/1.477924
- [23] J.A. Montgomery, M.J. Frisch, J.W. Ochterski and G.A. Petersson, *J. Chem. Phys.* **112**, 6532–6542 (2000). doi:10.1063/1.481224
- [24] G.A. Petersson, in *Quantum-Mechanical Prediction of Thermochemical Data*, edited by J. Cioslowski (Kluwer Academic Publishers, Dordrecht, 2001), pp. 99–130.
- [25] J.M.L. Martin and G. de Oliveira, *J. Chem. Phys.* **111**, 1843–1856 (1999). doi:10.1063/1.479454
- [26] J.M.L. Martin and S. Parthiban, in *Quantum-Mechanical Predict. Thermochem. Data*, edited by J. Cioslowski, Vol. 22 of *Understanding Chemical Reactivity* (Kluwer Academic Publishers, Dordrecht, 2002), pp. 31–65.
- [27] S. Parthiban and J.M.L. Martin, *J. Chem. Phys.* **114**, 6014–6029 (2001). doi:10.1063/1.1356014
- [28] N.J. DeYonker, T.R. Cundari and A.K. Wilson, *J. Chem. Phys.* **124**, 114104 (2006). doi:10.1063/1.2173988
- [29] N.J. DeYonker, B.R. Wilson, A.W. Pierpont, T.R. Cundari and A.K. Wilson, *Mol. Phys.* **107**, 1107–1121 (2009). doi:10.1080/00268970902744359
- [30] C. Peterson, D. Penchoff and A. Wilson, *Annual Reports in Computational Chemistry* **12**, 3–45 (2016). doi:10.1016/bs.arcc.2016.04.001
- [31] L.A. Curtiss, K. Raghavachari and J.A. Pople, *J. Chem. Phys.* **98**, 1293–1298 (1993). doi:10.1063/1.464297
- [32] L.A. Curtiss, P.C. Redfern, K. Raghavachari, V. Rassolov and J.A. Pople, *J. Chem. Phys.* **110**, 4703–4709 (1999). doi:10.1063/1.478385
- [33] B. Chan, J. Deng and L. Radom, *J. Chem. Theor. Comput.* **7**, 112–120 (2011). doi:10.1021/ct100542x
- [34] B. Chan, A. Karton and K. Raghavachari, *J. Chem. Theor. Comput.* **15**, 4478–4484 (2019). doi:10.1021/acs.jctc.9b00449
- [35] E. Semidalas and J.M.L. Martin, *J. Chem. Theor. Comput.* **16**, 7507–7524 (2020a). doi:10.1021/acs.jctc.0c01106
- [36] E. Semidalas and J.M.L. Martin, *J. Chem. Theor. Comput.* **16**, 4238–4255 (2020b). doi:10.1021/acs.jctc.0c00189
- [37] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Phys. Chem. Chem. Phys.* **19**, 32184–32215 (2017). doi:10.1039/C7CP04913G
- [38] C. Hättig, W. Klopper, A. Köhn and D.P. Tew, *Chem. Rev.* **112**, 4–74 (2012). doi:10.1021/cr200168z
- [39] L. Kong, F.A. Bischoff and E.F. Valeev, *Chem. Rev.* **112**, 75–107 (2012). doi:10.1021/cr200204r
- [40] S. Ten-no, *Theor. Chem. Acc.* **131**, 1070 (2012). doi:10.1007/s00214-011-1070-1
- [41] S. Ten-no and J. Noga, *WIREs Comput. Mol. Sci.* **2**, 114–125 (2012). doi:10.1002/wcms.68
- [42] W. Klopper and W. Kutzelnigg, *Chem. Phys. Lett.* **134**, 17–22 (1987). doi:10.1016/0009-2614(87)80005-2
- [43] W. Kutzelnigg and W. Klopper, *J. Chem. Phys.* **94**, 1985–2001 (1991). doi:10.1063/1.459921
- [44] A. Karton and J.M.L. Martin, *J. Chem. Phys.* **136**, 124114 (2012). doi:10.1063/1.3697678
- [45] A. Mahler and A.K. Wilson, *J. Chem. Theor. Comput.* **9**, 1402–1407 (2013). doi:10.1021/ct300956e
- [46] O.N. Ventura, M. Kieninger, A. Katz, M. Vega-Tejido, M. Segovia and K. Irving, *Int. J. Quantum Chem.* **121**, 1–19 (2021). doi:10.1002/qua.26745
- [47] J.F. Stanton, *Chem. Phys. Lett.* **281**, 130–134 (1997). doi:10.1016/S0009-2614(97)01144-5
- [48] Y.J. Bomble, J. Vázquez, M. Kállay, C. Michauk, P.G. Szalay, A.G. Császár, J. Gauss and J.F. Stanton, *J. Chem. Phys.* **125**, 064108 (2006). doi:10.1063/1.2206789

- [49] M.E. Harding, J. Vázquez, B. Ruscic, A.K. Wilson, J. Gauss and J.F. Stanton, *J. Chem. Phys.* **128**, 114111 (2008). doi:10.1063/1.2835612
- [50] A. Tajti, P.G. Szalay, A.G. Császár, M. Kállay, J. Gauss, E.F. Valeev, B.A. Flowers, J. Vázquez and J.F. Stanton, *J. Chem. Phys.* **121**, 11599–11613 (2004). doi:10.1063/1.1811608
- [51] J.H. Thorpe, C.A. Lopez, T.L. Nguyen, J.H. Baraban, D.H. Bross, B. Ruscic and J.F. Stanton, *J. Chem. Phys.* **150**, 224102 (2019). doi:10.1063/1.5095937
- [52] D.A. Dixon, D. Feller and K.A. Peterson, *Annual Reports in Computational Chemistry* **8**, 1–28 (2012). doi:10.1016/B978-0-444-59440-2.00001-6
- [53] D. Feller, K. Peterson and D. Dixon, *Annual Reports in Computational Chemistry* **12**, 47–78 (2016). doi:10.1016/bs.arcc.2016.02.001
- [54] D. Feller, K.A. Peterson and B. Ruscic, *Theor. Chem. Acc.* **133**, 1407 (2014). doi:10.1007/s00214-013-1407-z
- [55] A.D. Boese, M. Oren, O. Atasoylu, J.M.L. Martin, M. Kállay and J. Gauss, *J. Chem. Phys.* **120**, 4129–4141 (2004). doi:10.1063/1.1638736
- [56] A. Karton, E. Rabinovich, J.M.L. Martin and B. Ruscic, *J. Chem. Phys.* **125**, 144108 (2006). doi:10.1063/1.2348881
- [57] N. Sylvetsky, K.A. Peterson, A. Karton and J.M.L. Martin, *J. Chem. Phys.* **144**, 214101 (2016). doi:10.1063/1.4952410
- [58] M. Spiegel, E. Semidalas, J.M.L. Martin, M.R. Bentley and J.F. Stanton, *Mol. Phys.*, e2252114 (2023). doi:10.1080/00268976.2023.2252114
- [59] M. Heckert, M. Kállay and J. Gauss, *Mol. Phys.* **103**, 2109–2115 (2005). doi:10.1080/00268970500083416
- [60] M. Heckert, M. Kállay, D.P. Tew, W. Klopper and J. Gauss, *J. Chem. Phys.* **125**, 044108 (2006). doi:10.1063/1.2217732
- [61] C. Puzzarini, M. Heckert and J. Gauss, *J. Chem. Phys.* **128** (2008). doi:10.1063/1.2912941
- [62] T.A. Ruden, T. Helgaker, P. Jørgensen and J. Olsen, *J. Chem. Phys.* **121**, 5874–5884 (2004). doi:10.1063/1.1780155
- [63] W.J. Morgan, D.A. Matthews, M. Ringholm, J. Agarwal, J.Z. Gong, K. Ruud, W.D. Allen, J.F. Stanton and H.F. Schaefer, *J. Chem. Theor. Comput.* **14**, 1333–1350 (2018). doi:10.1021/acs.jctc.7b01138
- [64] W.D. Allen, A.L.L. East and A.G. Császár, in *Structures and Conformations of Non-Rigid Molecules*, edited by J. Laane, M. Dakkouri, B. van der Veken, and H. Oberhammer (Springer Netherlands, Dordrecht, 1993), pp. 343–373.
- [65] A.G. Császár, W.D. Allen and H.F. Schaefer, *J. Chem. Phys.* **108**, 9751–9764 (1998). doi:10.1063/1.476449
- [66] A.L.L. East and W.D. Allen, *J. Chem. Phys.* **99**, 4638–4650 (1993). doi:10.1063/1.466062
- [67] A.L.L. East, C.S. Johnson and W.D. Allen, *J. Chem. Phys.* **98**, 1299–1328 (1993). doi:10.1063/1.464298
- [68] Z. Zhu and X. Xu, *J. Chem. Theor. Comput.* **19**, 3112–3122 (2023). doi:10.1021/acs.jctc.3c00025
- [69] X. Huang and T.J. Lee, *J. Chem. Phys.* **129**, 044312 (2008). doi:10.1063/1.2957488
- [70] R.C. Fortenberry and T.J. Lee, *Annual Reports in Computational Chemistry* **15**, 173–202 (2019). doi:10.1016/bs.arcc.2019.08.006
- [71] M.B. Gardner, B.R. Westbrook, R.C. Fortenberry and T.J. Lee, *Spectrochim. Acta A* **248**, 119184 (2021). doi:10.1016/j.saa.2020.119184
- [72] A.D. Boese and J.M.L. Martin, *J. Phys. Chem. A* **108**, 3085–3096 (2004). doi:10.1021/jp0369589
- [73] Y. Fan, J. Ho and R.P. Bettens, *J. Phys. Chem. A* **110**, 2796–2800 (2006). doi:10.1021/jp0562330
- [74] V. Barone, M. Biczysko, J. Bloino and C. Puzzarini, *J. Chem. Phys.* **141** (2014). doi:10.1063/1.4887357
- [75] J.M.L. Martin and M.K. Kesharwani, *J. Chem. Theor. Comput.* **10**, 2085–2090 (2014). doi:10.1021/ct500174q
- [76] J.M. Martin, P.R. Taylor and T.J. Lee, *Chem. Phys. Lett.* **275**, 414–422 (1997). doi:10.1016/S0009-2614(97)00735-5
- [77] H.-J. Werner, P.J. Knowles, F.R. Manby, J.A. Black, K. Doll, A. Heßelmann, D. Kats, A. Köhn, T. Korona, D.A. Kreplin, Q. Ma, T.F. Miller, A. Mitrushchenkov, K.A. Peterson, I. Polyak, G. Rauhut and M. Sibaev, *J. Chem. Phys.* **152**, 144107 (2020). doi:10.1063/5.0005081
- [78] R. Lindh, *Theor. Chim. Acta* **85**, 423–440 (1993). doi:10.1007/BF01112982
- [79] C. Möller and M.S. Plesset, *Phys. Rev.* **46**, 618–622 (1934). doi:10.1103/PhysRev.46.618
- [80] M. Head-Gordon, *Mol. Phys.* **96**, 673–679 (1999). doi:10.1080/00268979909483003
- [81] A. El-Azhary, G. Rauhut, P. Pulay and H.-J. Werner, *J. Chem. Phys.* **108**, 5185–5193 (1998). doi:10.1063/1.475955
- [82] J.D. Watts, J. Gauss and R.J. Bartlett, *J. Chem. Phys.* **98**, 8718–8733 (1993). doi:10.1063/1.464480
- [83] M. Head-Gordon and T. Head-Gordon, *Chem. Phys. Lett.* **220**, 122–128 (1994). doi:10.1016/0009-2614(94)00116-2
- [84] Gaussian 16, Revision C.01, M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman and D.J. Fox, Gaussian, Inc., Wallingford, CT (2016). <http://www.gaussian.com>.
- [85] J.C. Womack and F.R. Manby, *J. Chem. Phys.* **140**, 044118 (2014). doi:10.1063/1.4863136
- [86] W. Györfy, G. Knizia and H.-J. Werner, *J. Chem. Phys.* **147**, 214101 (2017). doi:10.1063/1.5003065
- [87] W. Györfy and H.-J. Werner, *J. Chem. Phys.* **148**, 114104 (2018). doi:10.1063/1.5020436
- [88] H.-J. Werner, T.B. Adler and F.R. Manby, *J. Chem. Phys.* **126**, 164102 (2007). doi:10.1063/1.2712434
- [89] S. Ten-no, *Chem. Phys. Lett.* **398**, 56–61 (2004). doi:10.1016/j.cplett.2004.09.041

- [90] J.G. Hill, K.A. Peterson, G. Knizia and H. -J. Werner, *J. Chem. Phys.* **131**, 194105 (2009). doi:10.1063/1.3265857
- [91] C. Hättig, D.P. Tew and A. Köhn, *J. Chem. Phys.* **132**, 231102 (2010). doi:10.1063/1.3442368
- [92] F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005). doi:10.1039/b508541a
- [93] D. Rappoport and F. Furche, *J. Chem. Phys.* **133**, 134105 (2010). doi:10.1063/1.3484283
- [94] J. Almlöf and P.R. Taylor, *J. Chem. Phys.* **86**, 4070–4077 (1987). doi:10.1063/1.451917
- [95] F. Neese and E.F. Valeev, *J. Chem. Theor. Comput.* **7**, 33–43 (2011). doi:10.1021/ct100396y
- [96] T.H. Dunning, *J. Chem. Phys.* **90**, 1007–1023 (1989). doi:10.1063/1.456153
- [97] R.A. Kendall, T.H. Dunning and R.J. Harrison, *J. Chem. Phys.* **96**, 6796–6806 (1992). doi:10.1063/1.462569
- [98] D.E. Woon and T.H. Dunning, *J. Chem. Phys.* **98**, 1358–1371 (1993). doi:10.1063/1.464303
- [99] T.H. Dunning, K.A. Peterson and A.K. Wilson, *J. Chem. Phys.* **114**, 9244–9253 (2001). doi:10.1063/1.1367373
- [100] J.M. Martin, *J. Mol. Struct. THEOCHEM* **771**, 19–26 (2006). doi:10.1016/j.theochem.2006.03.035
- [101] J.M.L. Martin, *J. Chem. Phys.* **108**, 2791–2800 (1998). doi:10.1063/1.475670
- [102] K.A. Peterson and T.H. Dunning, *J. Chem. Phys.* **117**, 10548–10560 (2002). doi:10.1063/1.1520138
- [103] W. Klopper and C.C.M. Samson, *J. Chem. Phys.* **116**, 6397–6410 (2002). doi:10.1063/1.1461814
- [104] E.F. Valeev, *Chem. Phys. Lett.* **395**, 190–195 (2004). doi:10.1016/j.cplett.2004.07.061
- [105] K.A. Peterson, T.B. Adler and H. -J. Werner, *J. Chem. Phys.* **128**, 084102 (2008). doi:10.1063/1.2831537
- [106] K.E. Yousaf and K.A. Peterson, *Chem. Phys. Lett.* **476**, 303–307 (2009). doi:10.1016/j.cplett.2009.06.003
- [107] H.H. Nielsen, *Rev. Mod. Phys.* **23**, 90–136 (1951). doi:10.1103/RevModPhys.23.90
- [108] O. Marchetti and H.-J. Werner, *J. Phys. Chem. A* **113**, 11580–11585 (2009). doi:10.1021/jp9059467
- [109] L.A. Curtiss, K. Raghavachari, P.C. Redfern, V. Rassolov and J.A. Pople, *J. Chem. Phys.* **109**, 7764–7776 (1998). doi:10.1063/1.477422
- [110] N. Mehta, G. Santra and J.M.L. Martin, *Can. J. Chem.* **101**, 656–663 (2023). doi:10.1139/cjc-2022-0277
- [111] J. Liang, X. Feng, X. Liu and M. Head-Gordon, *J. Chem. Phys.* **158**, 204109 (2023). doi:10.1063/5.0152838
- [112] O.A. Vydrov and T. Van Voorhis, *J. Chem. Phys.* **133**, 244103 (2010). doi:10.1063/1.3521275
- [113] F. Eckert, P. Pulay and H.J. Werner, *J. Comput. Chem.* **18**, 1473–1483 (1997). doi:10.1002/(SICI)1096-987X(199709)18:12 < 1473::AID-JCC5 > 3.0.CO;2-G
- [114] D.G.A. Smith, L.A. Burns, A.C. Simmonett, R.M. Parrish, M.C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A.M. James, S. Lehtola, J.P. Misiewicz, M. Scheurer, R.A. Shaw, J.B. Schriber, Y. Xie, Z.L. Glick, D.A. Sirianni, J.S. O'Brien, J.M. Waldrop, A. Kumar, E.G. Hohenstein, B.P. Pritchard, B.R. Brooks, H.F. Schaefer, A.Y. Sokolov, K. Patkowski, A.E. DePrince, U. Bozkaya, R.A. King, F.A. Evangelista, J.M. Turney, T.D. Crawford and C.D. Sherrill, *J. Chem. Phys.* **152**, 184108 (2020). doi:10.1063/5.0006002
- [115] F. Jensen, *J. Phys. Chem. A* **127**, 2859–2863 (2023). doi:10.1021/acs.jpca.3c00064
- [116] A.J.C. Varandas, *J. Chem. Phys.* **157**, 174110 (2022). doi:10.1063/5.0122268
- [117] M.A. Broda, A. Buczek, T. Kupka and J. Kaminsky, *Vib. Spectrosc.* **63**, 432–439 (2012). doi:10.1016/j.vibspec.2012.09.005
- [118] A. Buczek, T. Kupka and M.A. Broda, *J. Mol. Model.* **17**, 2265–2274 (2011a). doi:10.1007/s00894-010-0944-9
- [119] A. Buczek, T. Kupka and M.A. Broda, *J. Mol. Model.* **17**, 2029–2040 (2011b). doi:10.1007/s00894-010-0913-3
- [120] J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.* **10**, 6615–6620 (2008). doi:10.1039/b810189b
- [121] F. Jensen, *J. Chem. Phys.* **115**, 9113–9125 (2001). doi:10.1063/1.1413524
- [122] K.A. Peterson, M.K. Kesharwani and J.M. Martin, *Mol. Phys.* **113**, 1551–1558 (2015). doi:10.1080/00268976.2014.985755
- [123] L. McCaslin and J. Stanton, *Mol. Phys.* **111**, 1492–1496 (2013). doi:10.1080/00268976.2013.811303
- [124] H. Schneider, K.M. Vogelhuber, F. Schinle, J.F. Stanton and J.M. Weber, *J. Phys. Chem. A* **112**, 7498–7506 (2008). doi:10.1021/jp800124s
- [125] E.D. Simandiras, J.E. Rice, T.J. Lee, R.D. Amos and N.C. Handy, *J. Chem. Phys.* **88**, 3187–3195 (1988). doi:10.1063/1.453963
- [126] J.M.L. Martin, T.J. Lee and P.R. Taylor, *J. Chem. Phys.* **108**, 676–691 (1998). doi:10.1063/1.475429
- [127] D. Moran, A.C. Simmonett, F.E. Leach, W.D. Allen, P.V.R. Schleyer and H.F. Schaefer, *J. Am. Chem. Soc.* **128**, 9342–9343 (2006). doi:10.1021/ja0630285
- [128] P.M. Nelson, Z.L. Glick and C.D. Sherrill, *J. Chem. Phys.* **159**, 094104 (2023). doi:10.1063/5.0168608
- [129] M.K. Kesharwani, B. Brauer and J.M.L. Martin, *J. Phys. Chem. A* **119**, 1701–1714 (2015). doi:10.1021/jp508422u
- [130] M.L. Laury, S.E. Boesch, I. Haken, P. Sinha, R.A. Wheeler and A.K. Wilson, *J. Comput. Chem.* **32**, 2339–2347 (2011). doi:10.1002/jcc.21811
- [131] M.L. Laury, M.J. Carlson and A.K. Wilson, *J. Comput. Chem.* **33**, 2380–2387 (2012). doi:10.1002/jcc.23073
- [132] J.C. Zapata Trujillo and L.K. McKemmish, *J. Phys. Chem. A* **126**, 4100–4122 (2022). doi:10.1021/acs.jpca.2c01438
- [133] J.C. Zapata Trujillo and L.K. McKemmish, *J. Phys. Chem. A* **127**, 1715–1735 (2023). doi:10.1021/acs.jpca.2c06908
- [134] W. Schneider and W. Thiel, *Chem. Phys. Lett.* **157**, 367–373 (1989). doi:10.1016/0009-2614(89)87263-X
- [135] P. Morgante and R. Peverati, *J. Comput. Chem.* **40**, 839–848 (2019). doi:10.1002/jcc.25761
- [136] S. Kozuch and J.M.L. Martin, *J. Comput. Chem.* **34**, 2327–2344 (2013). doi:10.1002/jcc.23391
- [137] G. Santra, N. Sylvetsky and J.M.L. Martin, *J. Phys. Chem. A* **123**, 5129–5143 (2019). doi:10.1021/acs.jpca.9b03157
- [138] N. Sylvetsky and J.M.L. Martin, *Mol. Phys.* **117**, 1078–1087 (2019). doi:10.1080/00268976.2018.1478140
- [139] J.L. Dunham, *Phys. Rev.* **41**, 713–720 (1932). doi:10.1103/PhysRev.41.713
- [140] J. Gauss, A. Tajti, M. Kállay, J.F. Stanton and P.G. Szalay, *J. Chem. Phys.* **125**, 144111 (2006). doi:10.1063/1.2356465
- [141] D.A. Matthews, L. Cheng, M.E. Harding, F. Lipparini, S. Stopkowicz, T. -C. Jagau, P.G. Szalay, J. Gauss

- and J.F. Stanton, *J. Chem. Phys.* **152**, 214108 (2020). doi:[10.1063/5.0004837](https://doi.org/10.1063/5.0004837)
- [142] A. Miani, E. Cané, P. Palmieri, A. Trombetti and N.C. Handy, *J. Chem. Phys.* **112**, 248–259 (2000). doi:[10.1063/1.480577](https://doi.org/10.1063/1.480577)
- [143] J. Demaison, *Mol. Phys.* **105**, 3109–3138 (2007). doi:[10.1080/00268970701765811](https://doi.org/10.1080/00268970701765811)
- [144] J.J. Eriksen, T.A. Anderson, J.E. Deustua, K. Ghanem, D. Hait, M.R. Hoffmann, S. Lee, D.S. Levine, I. Magoulas, J. Shen, N.M. Tubman, K.B. Whaley, E. Xu, Y. Yao, N. Zhang, A. Alavi, G.K. -L. Chan, M. Head-Gordon, W. Liu, P. Piecuch, S. Sharma, S.L. Ten-no, C.J. Umrigar and J. Gauss, *J. Phys. Chem. Lett.* **11**, 8922–8929 (2020). doi:[10.1021/acs.jpclett.0c02621](https://doi.org/10.1021/acs.jpclett.0c02621)
- [145] J. Tao, J.P. Perdew, V.N. Staroverov and G.E. Scuseria, *Phys. Rev. Lett.* **91**, 146401 (2003). doi:[10.1103/PhysRevLett.91.146401](https://doi.org/10.1103/PhysRevLett.91.146401)
- [146] A.D. Becke, *J. Chem. Phys.* **98**, 1372–1377 (1993). doi:[10.1063/1.464304](https://doi.org/10.1063/1.464304)
- [147] C. Lee, W. Yang and R.G. Parr, *Phys. Rev. B.* **37**, 785–789 (1988). doi:[10.1103/PhysRevB.37.785](https://doi.org/10.1103/PhysRevB.37.785)
- [148] Y.-S. Lin, G.-D. Li, S.-P. Mao and J.-D. Chai, *J. Chem. Theory. Comput.* **9**, 263–272 (2013). doi:[10.1021/ct300715s](https://doi.org/10.1021/ct300715s)
- [149] A. Karton, A. Tarnopolsky, J.-F. Lamère, G.C. Schatz and J.M.L. Martin, *J. Phys. Chem. A* **112**, 12868–12886 (2008). doi:[10.1021/jp801805p](https://doi.org/10.1021/jp801805p)
- [150] N. Mardirossian and M. Head-Gordon, *Phys. Chem. Chem. Phys.* **16**, 9904 (2014). doi:[10.1039/c3cp54374a](https://doi.org/10.1039/c3cp54374a)
- [151] N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.* **144**, 214110 (2016). doi:[10.1063/1.4952647](https://doi.org/10.1063/1.4952647)
- [152] N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.* **142**, 074111 (2015). doi:[10.1063/1.4907719](https://doi.org/10.1063/1.4907719)