AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the accepted manuscript version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Gunnar Schmitz and Ove Christiansen (2017). Accuracy of Frequencies Obtained with the Aid of Explicitly Correlated Wave Function Based Methods. In *Journal of Chemical Theory and Computation* 13(8), 3602-3613.

DOI: <https://doi.org/10.1021/acs.jctc.7b00476>

Publication metadata

Title:	Accuracy of Frequencies Obtained with the Aid of Explicitly Correlated Wave Function Based Methods
Author(s):	Gunnar Schmitz and Ove Christiansen
Journal:	<i>Molecular Cell</i>
DOI/Link:	https://doi.org/10.1021/acs.jctc.7b00476
Document version:	Accepted manuscript (post-print)

© The authors 2017. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Journal of Chemical Theory and Computation* 13(8), 3602-3613. <https://doi.org/10.1021/acs.jctc.7b00476>

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

Accuracy of frequencies obtained with the aid of explicitly correlated wave function based methods

Gunnar Schmitz* and Ove Christiansen

Department of Chemistry, Aarhus University

E-mail: Gunnar.Schmitz@chem.au.dk

Abstract

We assess the basis set convergence of harmonic frequencies using different explicitly correlated wave function based methods. All commonly available CCSD(T) variants as well as MP2-F12 and MP4(F12*) are considered and a hierarchy of the different approaches is established. As for reaction and atomization energies CCSD(F12*)(T*) is a close approximation to CCSD(F12)(T*) and clearly superior to the other tested approximations. The used scaling for the triples correction enhances the accuracy relative to CCSD(F12*)(T) especially for small basis sets and is very attractive since no additional computational costs are added. However this scaling slightly breaks size consistency and therefore we additionally study the accuracy of CCSD(F12*)(T*) and CCSD(F12*)(T) in the context of calculating anharmonic frequencies to check if this causes problems in the generation of the potential energy surface (PES). We find a fast basis set convergence for harmonic and anharmonic frequencies. Already in the cc-pVDZ-F12 basis the RMSD to the CBS-limit is only around 4–5 cm⁻¹.

Introduction

Vibrational spectroscopy plays an important role in understanding phenomena in different fields of chemistry ranging from industrial to biological applications. Theoretical chemistry is therefore expected to provide tools with accurate predictive power to assist the scientists in these areas. Over the last decade methods like Vibrational Perturbation Theory¹⁻³ (VPT), Vibrational Configuration Interaction⁴⁻⁶ (VCI) or Vibrational Coupled Cluster⁶⁻⁸ (VCC) gained a lot of interest, since they provide a highly accurate description of anharmonic frequencies. However these methods need accurate potential surfaces (PESs) and a sufficiently accurate description of harmonic frequencies, which are the major contribution to anharmonic frequencies. Therefore very accurate methods like Coupled Cluster (CC) methods e.g. Coupled Cluster singles and doubles with inclusion of the perturbative triples correction (CCSD(T)) for calculating the electronic structure are required. But the problems of these methods like the slow basis set convergence directly translates to the computation of harmonic and anharmonic frequencies. The basis set convergence problem can be ameliorated if terms with an explicit dependence on the inter electronic distance are included in the electronic wave functions. This allows a better description of the electron electron cusp. Nowadays the most widely used explicitly correlated methods are the R12 and F12 methods, which go back to work by Kutzelnigg and Klopper,⁹⁻¹¹ where the Hartree Fock (HF) reference is augmented by pair functionals χ_{ij} called geminals. The geminals have an explicit dependence on r_{12} via a correlation factor $f(r_{12})$. In R12 theory it is given as $f(r_{12}) = r_{12}$ and in F12 theory as $f_{12}(r_{12}) = -\frac{1}{\gamma} \cdot \exp(-\gamma r_{12})$, where γ is a parameter which is optimized for a given basis set and usually close to one. One usually observes that the usage of F12 theory enhances the basis set convergence of the correlation energy in such a way that with a basis set of cardinal number X results are obtained with an accuracy as in conventional methods with a basis sets of cardinal number $X + 2$.

Therefore explicitly correlated methods are of great interest to calculate vibrational spectra and pave the way to use smaller basis sets and still obtain highly accurate PES. Recently

Martin et al.¹² studied the convergence of some explicitly CCSD(T) methods for calculating harmonic frequencies and earlier Rauhut et al.¹³ investigated the same issue for CCSD-F12a(T) and CCSD-F12b(T) and later used explicitly correlated CCSD(T) for application studies.^{14,15} Furthermore the work of Tew et al. in Ref. 16 is to mention, where they did a similar assessment. We extend these studies by taking more variants of explicitly correlated CCSD(T) methods and additional explicitly correlated methods like MP2-F12 and MP4(F12*) into account. The later one is defined as a fourth order approximation to CCSD(F12*(T)).

Furthermore we use less approximate methods as internal reference. Our work complements already available studies on the basis set convergence for molecular properties on the CC level like in Refs. 17–19 by focusing especially on explicit correlated methods. However, we limit ourself to the CCSD(T) level and for the influence of higher order excitations we refer to the work of Heckert et al.¹⁸ or Ruden et al.¹⁹

As Martin et al.¹² we explore the usage of the scaled triples correction (T*), which approximately introduces explicitly correlated effects in this contribution. Usually only the explicitly correlated CCSD variant is combined with the conventional triples correction. But because in the presence of explicitly correlated terms the doubles amplitudes are typically slightly smaller in magnitude, the triples correction is less accurate and is known to often limit the accuracy. Knizia et al. therefore proposed to scale the conventional triples correction with the ratio of the MP2-F12 and MP2 energy to lift this problem.²⁰ As all explicitly correlated contributions the scaling vanishes in the complete basis set (CBS) limit, but this approach slightly breaks size consistency and this could be a problem in the generation of PESs, which we also want to explore in the current study. The issue of missing size consistency has been noted before. For example in Ref. 21 of Ahnen et al. a supermolecule approach is favored, when calculating interaction energies and also in Ref. 22 the scaling is kept fixed when calculating interaction energies to resolve this issue. We note that Köhn reported in Ref. 23 and 24 an explicitly correlated triples correction. However it is not straightforward to implement

without code generators and therefore not available in many program packages. The scaling correction is therefore at the moment the most used pragmatic choice since it is easy to implement and comes at no extra cost.

To construct the PES we use the so called n -mode representation in normal coordinates, where cuts through the hypersurface are used including terms, where just $1, 2, \dots, n$ coordinates are varied at a time. We note that several authors also report results using a full dimensional description.^{25–30} This paper is organized as follows: First the selected test set is presented. Afterwards the computational setup is discussed in detail. The actual analysis is then separated into 4 parts. In the first part we asses the accuracy of harmonic frequencies obtained with explicitly correlated CCSD(T) variants with only a minimal number of approximations. The best performing method from this initial study is then chosen as internal reference to study the accuracy of more approximate methods in part 2. In part 3 the accuracy relative to anharmonic frequencies is studied and simple corrections like scaling factors are considered. In the last part anharmonic corrections with means of VCC are calculated and we explore the basis set convergence in this context and if the triples scaling can cause negative effects.

Theoretical methods

Table 1: Summary of some of the characteristics of the used methods.

Method	Fixed Amplitude	Projector	F12 contribution	Add. approx to CCSD(F12)
CCSD(F12)	No	Full	In CC equations	-
CCSD(F12)-SP	Yes	Full	In CC equations	-
CCSD(F12*)	Yes	Full + approx.	In CC equations	Terms up to 3th order and approx. higher order
CCSD[F12]	Yes	Full	In CC equations	Terms up to 3th order
CCSD-F12b	Yes	approx.	In CC equations	Terms up to 3th order
CCSD(2) _{F12}	Yes	Full	Perturbative	
CCSD(2*) _{F12}	Yes	Full	Perturbative	extended Brillouin condition

Before discussing the chosen test set and the results, we give a brief summary of the used theoretical methods in our study. The main focus of this work is to compare different explicitly correlated methods and especially their CCSD(T) variants. As already outlined in the introduction the conventional triples correction is usually combined with an explicitly correlated CCSD variant. In order to approximately include explicitly correlated effects in the triples correction, one can scale this correction, for example by the ratio of the MP2-F12 and MP2 correlation energy giving rise to the (T*) approximation. Using either the (T) or (T*) correction the main difference is the underlying CCSD variant, which should be described in the following. CCSD(R12) introduced by Fliegel et al. in the R12 Framework³¹ using the correlation factor $f(r_{12}) = r_{12}$ and later extended to F12 Theory by Tew et al.³² using the correlation factor $f_{12}(r_{12}) = -\frac{1}{\gamma} \cdot \exp(-\gamma r_{12})$ ³³ is known to be very close to the full CCSD-R12/CCSD-F12. In the following we want to discuss the key differences of the approximations to CCSD(F12) used in this work. However this can only be a short overview and we refer for the interested reader for example to the review of Hättig et al.³⁴ and to the very recent review on perspectives and challenges in F12 theory by Grüneis et al.³⁵ In modern implementations the geminal coefficients are usually fixed by invoking the cusp condition.^{36,37} The postfix SP or Fix is often used to emphasize its usage. However in some methods like CCSD(F12*) this postfix is omitted, since this methods were presented using this approximation and are not rigorously tested without it.

For the available CCSD variants we can loosely define two categories. In the first one terms in the CCSD-F12 equations are neglected based on arguments originating from the connection to perturbation theory or a careful cost analysis.^{38,39} In the second category the explicit electron correlation is incorporated solely by means of perturbation theory.⁴⁰ The methods CCSD[F12] and CCSD(F12*)¹ of Hättig et al.³⁸ and the methods CCSD-F12a and CCSD-F12b of Adler et al.³⁹ fall in the first category. However in CCSD-F12a and CCSD-F12b for all terms not already included in MP2-F12 a more approximate strong orthogonality

¹We note that CCSD(F12*) is by some authors referred to as CCSD-F12c.

projector is used. This projector is used to keep the geminals orthogonal to the orbital basis and due to the invoked approximations it removes additionally explicitly correlated terms, which give a minor contribution. Additionally some terms are discarded in CCSD-F12a and CCSD-F12b due to efficiency reasons. For CCSD[F12] and CCSD(F12*) explicitly correlated terms are mainly neglected based on the order in perturbation theory where they contribute to the energy and the same projector as for CCSD(F12) is used for most contributions. A more approximate projector is however also used for CCSD(F12*) for contributions in fourth or higher order to capture the most important higher order contributions while staying computational efficient. The method $\text{CCSD}(2^*)_{\overline{\text{F12}}}$ introduced by Torheyden and Valeev⁴⁰ and later extended to $\text{CCSD}(2)_{\overline{\text{F12}}}$ by Hättig et al.³⁸ falls in the second category.² The later method is more accurate since it includes coupling elements between the orbital basis and the Complementary Auxiliary basis set (CABS), which vanish if one assumes that the extended Brillouin condition is fulfilled. Table 1 summarizes these methods and indicates the used approximations.

Test set

In the following section the accuracy of harmonic vibrational frequencies obtained with $\text{CCSD}(\text{F12}^*)(\text{T}^*)$ will be discussed. For that purpose we use the F1 test set from Ref. 41, from which we exclude all open shell molecules and some special cases. Therefore we refer to this subset as F1-sub1 and list all molecules in Table 2. In the supporting material additional information like the optimized structures and frequencies can be found. This test set contains 110 molecules and 865 frequencies. Furthermore an additional F1-sub2 subset is formed containing 13 molecules and 74 harmonic frequencies, for which highly accurate references are available, which are deduced from experiment or from high level ab initio calculation which aim at experimental accuracy. The F1-sub1 test set therefore gives good

²We adopt the notation from Ref. 38, where $\text{CCSD}(2)_{\overline{\text{F12}}}$ was introduced as a modification to the method $\text{CCSD}(2)_{\overline{\text{R12}}}$ introduced in Ref. 40. The previous reported (more approximate) method is referred to as $\text{CCSD}(2^*)_{\overline{\text{F12}}}$.

insight into the agreement with experiments including anharmonic effects while the F1-sub2 test set assess the accuracy in relation to "true" harmonic frequencies. The molecules forming the reduced set are given in Table 3. Since the F1-sub1 set is much larger an in-depth analysis using various basis sets is only done for the F1-sub2 set and the most cost effective methods are chosen for the larger one.

Table 2: List of the molecules in the F1-sub1 test set.

Molecules
AlCl ₃ B ₂ H ₆ BF ₃ BH ₃ CO C ₂ H ₂ C ₂ H ₄ C ₂ N ₂ cC ₂ H ₂ N cC ₂ H ₄ O cC ₃ H ₆ CH ₂ CCCH ₂ CH ₂ CCHCl CH ₂ CCl ₂ CH ₂ CHCHO CH ₂ Cl ₂ CH ₃ CCCH ₃ CH ₃ CCH CH ₃ CH ₂ Cl CH ₃ CH ₂ CN CH ₃ CH ₂ F CH ₃ CH ₃ CH ₃ CHO CH ₃ Cl CH ₃ CN CH ₃ COCH ₃ CH ₃ COF CH ₃ COOH CH ₃ F CH ₃ NC CH ₃ NH ₂ CH ₃ NNCH ₃ CH ₃ OCH ₃ CH ₃ OH CH ₃ SiH ₃ CH ₄ CHCl ₃ CHF ₃ cis-Butadiene cis-CHClCHCl cis-CHFCHF cis-OCCO Cl ₂ O ClCCCl ClCCH ClCN ClF ₃ ClNO ClNO ₂ ClNS CO ₂ COCl ₂ COClF COF ₂ COS CS ₂ CScI ₂ CSF ₂ F ₂ F ₂ NH F ₂ O F ₂ SO FCN gauche-CH ₂ ClCH ₂ Cl H ₂ CO H ₂ O H ₂ O ₂ H ₂ S H ₂ S ₂ HCCCCH HCCCH ₂ Cl HCCCH ₂ F HCCF HCN HCOOH HCOOHCH ₃ HNCO HNO ₃ HOCl HOF LiF N ₂ N ₂ F ₂ N ₃ H NCl ₂ F NClF ₂ NF ₃ NH ₃ NNO NOF NSF PCl ₃ PF ₃ PH ₃ S ₂ F ₂ SCl ₂ SiH ₂ Cl ₂ SiH ₃ CCH SiH ₃ Cl SiH ₃ F SiH ₄ SiHCl ₃ SiHF ₃ SO ₂ SO ₃ SOCl ₂ trans-CHCClCHCl trans-Butadiene trans-CH ₂ ClCH ₂ Cl trans-CHFCHF trans-OCCO

Table 3: List of the molecules in the F1-sub2 test set (in arbitrary order) with the references for the harmonic frequencies.

Molecules
H ₂ O ⁴² HCN ⁴³ CO ₂ ⁴⁴ C ₂ H ₂ ⁴⁵ H ₂ S ⁴⁶ H ₂ CO ⁴⁷ PH ₃ ⁴⁸ CH ₃ OH ⁴⁹ NH ₃ ⁵⁰ N ₂ ⁵¹ F ₂ ⁵¹ CH ₄ ⁵² C ₂ H ₄ ⁵³

Computational details

All electronic structure calculations were carried out using the TURBOMOLE program package⁵⁴ and for the anharmonic frequency calculations the MidasCpp program package⁵⁵ was used. The CCSD(F12*)(T*) calculations were done with the ccscdf12 module and the cc-pVXZ-F12 basis set family of Peterson and co-workers was used.⁵⁶ For DFT reference calculation the def2-TZVPP basis⁵⁷ was applied and for some semi empirical calculations the

minix basis set family⁵⁸ was used. The geometry optimization and the calculation of the Hessian were done using numerical differentiation. The step length for the gradient was 0.005 au while for the hessian a step length of 0.01 au was chosen. The displacements were carried out in symmetry adapted coordinates. Throughout this work the frozen core approximation was used for calculating the correlation energy. All orbitals with an energy below -3 Hartree were excluded from the correlation treatment. In the F12 calculations we applied the complementary auxiliary basis set (CABS) approach. As CABS thresholds we used a value of 10^{-8} . We note that potentially small discontinuities can appear due to CABS orbitals moving above or below the threshold for linear dependence. We did however not observe any practical problems in this regard.

Additionally some DFT reference calculations were carried out, where representatives from each main category was chosen: PBE⁵⁹ for functionals using the Generalized Gradient Approximation (GGA), B3-LYP^{60,61} for hybrid functionals and TPSS⁶² for meta-GGAs. Furthermore we also included the HF-3c⁵⁸ method in our study to represent semi-empirical methods.

Anharmonic frequencies were calculated by means of VCC[3] response theory⁶³ using potentials with at most 3 mode couplings. The primitive modal basis consists of B-Splines of order 10, where the density of the localized basis is set to 0.8 as described in Ref. 64. To generate the PES an adaptive density guided approach (ADGA)⁶⁵ to select the grid points based on physical information was used. The threshold are chosen in such a way that this adaptive approach is converged when the relative change in the potential from one iteration to the other is less than 10^{-2} and the absolute change less than 10^{-6} .

Accuracy for vibrational frequencies using numerical differentiation

The numerical differentiation introduces an additional error in the frequency calculation. Intensive numerical tests suggest to use a step length of 0.01 au for the hessian calculation and 0.005 au for the gradient calculation to keep the error sufficiently small. Furthermore

the SCF calculation, the density and in CC calculation the residual should be converged tightly. In the current cases the SCF energy is converged up to the 10th digit after the comma (\$scfconv) as well as the root mean square of the density matrix (\$denconv). The CC residual (\$oconv) is at least converged to the 9th digit after the comma.

To give an estimate for the expected influence of the numerical differentiation, the harmonic frequencies are compared with semi-numerical (analytical gradient and numerical 2nd derivatives) ones on the MP2/cc-pVDZ-F12 level of theory. For that purpose the F1-sub2 as well as the larger F1-sub1 set is used. In Table 4 the root-means-square-deviation (RMSD), the mean (MEAN), the mean absolute deviation (MAD) and the standard deviation (STD) is listed for both sets, where as reference the semi-numerical frequencies are used. The important error marks are below 1 cm^{-1} and taking a confidence interval of 2 STD we can estimate an error range of 1.2 cm^{-1} . As complementary information Table 5 lists these error measures for both test sets with respect to the experimental values. This indirect way of assessing the accuracy of the fully numerical frequencies yields a bit smaller errors indicating some error cancellation. The overall accuracy should only be slightly effected by using numerical Hessians and the observed errors should be transferable to the subsidiary studies. They might be a bit larger since in contrast to MP2 the doubles amplitudes in CC calculation are only solved up to a certain precision while the MP2 amplitudes are exact up to machine precision. This additional numerical noise also effects the final accuracy, but should be sufficiently small.

Table 4: Error statistic (in cm^{-1}) for the harmonic frequencies on the MP2/cc-pVDZ-F12 level of theory using fully numerical frequencies instead of semi numerical frequencies.

test set	RMSD	MEAN	MAD	STD
F1-sub1	0.60	-0.09	0.15	0.60
F1-sub2	0.48	-0.26	0.28	0.46

Table 5: Error statistic (in cm^{-1}) for the harmonic frequencies on the MP2/cc-pVDZ-F12 level of theory using fully numerical and semi numerical frequencies.

test set	type	RMSD	MEAN	MAD	STD
F1-sub1	fully numerical	85.50	32.33	42.45	63.35
F1-sub1	semi numerical	85.63	32.42	42.46	63.47
	Δ	-0.13	-0.09	-0.01	-0.12
F1-sub2	fully numerical	37.54	-2.40	34.34	33.63
F1-sub2	semi numerical	37.52	-2.15	34.16	33.48
	Δ	0.02	-0.25	0.18	0.15

Results and discussion

Accuracy relative to experiment

Table 6: Error statistic (in cm^{-1}) for the harmonic frequencies relative to experiment using different methods and basis sets for the F1-sub2 test set. The root mean square deviation (RMSD), the mean error (MEAN), the mean absolute error (MAD), the standard deviation (STD), the maximal (MAX) and minimal (MIN) error as well as the RANGE = (MAX - MIN) is given. Additionally the molecule with the largest error for a frequency is listed in the column Problematic case.

Method	basis set	RMSD	MEAN	MAD	STD	MAX	MIN	RANGE	Problematic case
CCSD(F12)(T*)	cc-pVDZ-F12	5.61	1.49	3.88	6.00	15.20	-15.20	30.4	F ₂
CCSD(F12)(T*)	cc-pVTZ-F12	4.61	0.86	2.94	4.82	11.88	-13.08	24.96	PH ₃
CCSD(F12)(T*)	cc-pVQZ-F12	4.61	1.21	3.02	4.91	13.40	-12.52	25.92	F ₂
CCSD(F12)(T)	cc-pVDZ-F12	7.20	6.02	7.00	8.86	31.46	-12.86	44.32	F ₂
CCSD(F12)(T)	cc-pVTZ-F12	5.05	2.70	4.16	5.87	19.17	-12.23	31.4	F ₂
CCSD(F12)(T)	cc-pVQZ-F12	4.84	2.15	3.70	5.43	17.33	-12.01	29.34	F ₂
CCSD(F12)-SP(T*)	cc-pVDZ-F12	5.53	1.50	3.71	5.90	17.24	-13.58	30.82	F ₂
CCSD(F12)-SP(T*)	cc-pVTZ-F12	4.69	1.44	3.29	5.08	13.96	-12.97	26.93	F ₂
CCSD(F12)-SP(T*)	cc-pVQZ-F12	4.71	1.67	3.34	5.16	15.06	-12.47	27.53	F ₂
CCSD(F12*)(T*)	cc-pVDZ-F12	5.45	1.29	3.52	5.75	15.32	-13.75	29.07	F ₂
CCSD(F12*)(T*)	cc-pVTZ-F12	4.69	1.41	3.27	5.07	13.85	-12.97	26.82	F ₂
CCSD(F12*)(T*)	cc-pVQZ-F12	4.71	1.66	3.34	5.16	15.06	-12.46	27.52	F ₂

In this section we focus on the accuracy relative to experiment. In this analysis only methods with a minimum number of approximations are considered and the influence of further approximations will be carefully examined. The error statistics are listed in Table 6 for different basis sets. The influence of the triples scaling (T*) is evident. Especially in small basis sets the RMSD is significantly reduced. In larger basis sets the influence decreases,

but is still observable. This lead us to the conclusion to only use the (T*) correction in the subsequent studies on harmonic frequencies since it causes no extra cost in comparison to the (T) correction, but improves the results significantly in small basis sets and also slightly in large basis sets. Additionally the influence of the fixed amplitude approach is small as expected and will be used for the rest of this work, where the CCSD(F12)(T*)/cc-pVQZ results serve as internal standard, which defines a good estimate for the CCSD(T) CBS limit. It is important to note that the convergence in the basis set sequence cc-pVDZ-F12, cc-pVTZ-F12, cc-pVQZ-F12 is extremely fast. Already the results in the cc-pVDZ-F12 basis are very close to the chosen reference. The RMSD is slightly above 5 cm⁻¹. The remaining disagreement to experiment is related to missing higher order excitations and relativistic effects. We note that one can extract from the work of Heckert and Gauss^{17,18,66} that the largest errors for CCSD(T)-F12 fundamentals compared to experiment are for vibrations involving double or triple bonds, where post (T) correlation is significant.

Although the overall agreement in our current study is good, there are in a few cases still some larger deviations. For F₂ errors around 15 cm⁻¹ are observed and those for N₂ and PH₃ lie around 5 cm⁻¹. These three molecules are consistently problematic cases for all tested methods and basis sets, where the errors for F₂ are in nearly all cases the largest ones.

Accuracy relative to the CCSD(T) basis set limit

The next step is to benchmark more approximate CCSD(F12)(T) variants against the CCSD(F12)(T*)/cc-pVQZ-F12 reference. In all these cases the fixed amplitude Ansatz is used as well as the scaled triples correction. The explicitly correlated CCSD variants, MP4(F12*), MP2-F12 and some common DFT functionals as well as the empiric HF-3c method is also assessed for comparison. The results are summarized in Table 7. For CCSD(F12*)(T*) a fast convergence to the reference is observed as shown in Figure 1 where the error distributions for the different basis sets are depicted. For cc-pVTZ-F12 and cc-pVQZ-F12 the distributions are quite narrow. But also the small cc-pVDZ-F12 basis enables

Table 7: Error statistic (in cm^{-1}) for the harmonic frequencies using different methods and basis sets for the F1-sub2 test set and the CCSD(F12)(T*)/cc-pVQZ-F12 reference. Additionally the molecule with the largest error for a frequency is listed in the column Problematic case.

Method	Basis set	RMSD	MEAN	MAD	RANGE	Problematic case
CCSD(F12)(T*)	cc-pVDZ-F12	3.02	0.28	2.10	18.15	C ₂ H ₄
CCSD(F12)(T*)	cc-pVTZ-F12	0.86	-0.35	0.51	4.76	C ₂ H ₂
CCSD(F12*)(T*)	cc-pVDZ-F12	2.70	0.08	1.78	16.21	C ₂ H ₄
CCSD(F12*)(T*)	cc-pVTZ-F12	0.61	0.20	0.33	3.15	C ₂ H ₄
CCSD(F12*)(T*)	cc-pVQZ-F12	0.40	0.45	0.45	1.77	F ₂
CCSD[F12](T*)	cc-pVDZ-F12	5.58	-5.66	5.86	27.19	F ₂
CCSD[F12](T*)	cc-pVTZ-F12	1.92	-2.11	2.11	9.98	F ₂
CCSD[F12](T*)	cc-pVQZ-F12	0.59	-0.69	0.69	3.57	F ₂
CCSD-F12b(T*)	cc-pVDZ-F12	5.89	-5.18	5.70	24.51	F ₂
CCSD-F12b(T*)	cc-pVTZ-F12	2.71	-2.95	2.95	11.65	F ₂
CCSD-F12b(T*)	cc-pVQZ-F12	0.77	-0.90	0.90	4.12	F ₂
CCSD(2) _{F12} (T*)	cc-pVDZ-F12	5.70	-5.58	5.81	25.90	F ₂
CCSD(2) _{F12} (T*)	cc-pVTZ-F12	2.71	-2.71	2.71	12.23	F ₂
CCSD(2) _{F12} (T*)	cc-pVQZ-F12	0.86	-0.88	0.88	3.96	F ₂
CCSD(2*) _{F12} (T*)	cc-pVDZ-F12	10.09	-5.90	8.33	56.71	F ₂
CCSD(2*) _{F12} (T*)	cc-pVTZ-F12	2.35	-1.97	2.05	13.31	F ₂
CCSD(2*) _{F12} (T*)	cc-pVQZ-F12	1.01	-0.50	0.91	8.0	F ₂
CCSD(F12*)	cc-pVDZ-F12	28.58	34.05	34.05	92.26	F ₂
CCSD(F12*)	cc-pVTZ-F12	29.40	34.86	34.86	92.90	F ₂
CCSD(F12*)	cc-pVQZ-F12	29.77	35.22	35.22	94.98	F ₂
MP4(F12*)	cc-pVDZ-F12	24.73	-16.21	21.89	166.6	N ₂
MP4(F12*)	cc-pVDZ-F12	25.29	-19.29	22.24	166.63	N ₂
MP4(F12*)	cc-pVDZ-F12	25.89	-20.08	22.75	169.21	N ₂
MP2-F12	cc-pVDZ-F12	37.34	6.08	34.25	237.45	N ₂
MP2-F12	cc-pVTZ-F12	36.92	7.03	33.35	232.65	N ₂
MP2-F12	cc-pVQZ-F12	37.03	7.32	33.51	233.36	N ₂
MP2	cc-pVDZ-F12	39.52	-0.29	36.56	255.05	N ₂
MP2	cc-pVTZ-F12	33.54	0.70	30.58	226.37	N ₂
MP2	cc-pVQZ-F12	35.51	3.84	32.55	234.64	N ₂
CCSD(T)	cc-pVDZ-F12	19.16	-12.96	14.99	90.56	C ₂ H ₂
CCSD(T)	cc-pVTZ-F12	10.3	-9.73	9.73	26.14	C ₂ H ₂
CCSD(T)	cc-pVQZ-F12	4.92	-4.64	4.64	12.61	H ₂ O
PBE	def2-TZVPP	67.95	-44.97	55.77	238.45	H ₂ CO
B3-LYP	def2-TZVPP	31.18	12.01	32.13	189.37	F ₂
TPSS	def2-TZVPP	47.63	-20.16	34.98	210.12	H ₂ O
TPSSH	def2-TZVPP	30.20	8.98	26.96	189.2	F ₂
HF-3c	minix	323.97	286.36	286.36	645.81	F ₂

accurate results. We note that in the cc-pVQZ-F12 basis in principle all CCSD(T) methods enable the same accuracy since the deviations fall below remaining error due to numerical differentiation, which can be expected to be around 1 cm^{-1} .

Interestingly CCSD(F12*)(T*) shows slightly improved error statistics in comparison to the parent CCSD(F12)(T*) method. But this effect is small due to some error cancellation. Concerning the accuracy, CCSD(F12*)(T*) is directly followed by CCSD[F12](T*). The RMSD for CCSD[F12](T*) is approximately larger by a factor of two than that for CCSD(F12*)(T*) for the same basis set. This decline in accuracy does not match the only slightly increased cost when using CCSD(F12*)(T*) compared to CCSD[F12](T*). For the present test set all CCSD(T) methods differ only slightly in computational costs. Taking calculations on CH₃OH in the cc-pVQZ-F12 basis as an example a single point calculation using 2 threads on a Intel Xeon E5-2680 CPU with 2.50 GHz takes around 36 minutes for the cheapest calculation (CCSD(2*)_{F12}(T*)) and 38 minutes for the most expensive one (CCSD(F12*)(T*)). In general for larger systems the (T) correction will dominate and the differences for the different CCSD variants are negligible. The results for CCSD-F12b(T*)³⁹ and CCSD(2)_{F12}(T*) are quite similar and the gathered data is not sufficient to really state which method provides more accurate results. However CCSD(2)_{F12}(T*) is computationally slightly more efficient and might therefore be the preferred choice. In contrast CCSD(2*)_{F12}(T*) should only be used in combination with large basis sets. It is important to note that CCSD(F12)(T*) and CCSD(F12*)(T*), already in the cc-pVDZ-F12 basis provide an accuracy comparable with the results in the cc-pVTZ-F12 basis in most of the other CCSD(T) variants.

Compared to the CCSD(T) methods the RMSD for MP2-F12 and MP4(F12*) is significantly larger, but still the basis set convergence is fast. The results do not significantly change by increasing the basis set size. The intrinsic error due to the selected method is larger than the basis set incompleteness error (BSIE) and therefore the convergence appears to be fast, since the intrinsic error dominates. It is noteworthy that using MP4(F12*) and MP2-F12 we observe some very large errors which affect the overall statistics significantly. For example

at the MP4(F12*) level the frequency in N₂ can not be described better than by an error of 150 cm⁻¹.

For the explicitly correlated methods the following hierarchy concerning the accuracy can therefore be concluded:

$$\text{MP2-F12} < \text{MP4(F12*)} \approx \text{CCSD(F12*)} < \text{CCSD(2*)}_{\overline{\text{F12}}}(\text{T*}) < \text{CCSD(2)}_{\overline{\text{F12}}}(\text{T*}) \approx < \text{CCSD-F12b(T*)} < \text{CCSD[F12]}(\text{T*}) < \text{CCSD(F12*)}(\text{T*})$$

To reach a fairly good accuracy at least CCSD(2) _{$\overline{\text{F12}}$} (T*) should be used. The basis set convergence of conventional CCSD(T) is significantly slower.

Taking a look at the conventional CCSD(T) we observe as usual that with F12 theory in a basis of cardinal number X results are obtained with an accuracy as in conventional methods with a basis sets of cardinal number $X + 2$. That this rule of thumb also holds for calculating harmonic frequencies is not new and was already reported by several authors like Rauhut and co-workers¹³⁻¹⁵ or Martin et al.,¹² but for the sake of completeness we want to demonstrate this important findings again. Furthermore it impressively shows the benefits of using F12 Theory with even modest basis sets. In the small cc-pVDZ-F12 basis we obtain as accurate results as in a conventional calculation in the cc-pVQZ-F12 basis with CCSD(F12*)(T*).

All methods without triples correction and MP4(F12*) are significantly inferior to the so far discussed methods. For the MPn family of methods, the picture is mixed. The frequencies can be quite close to the reference values, but there are exceptions with large deviations, which affect the overall statistics. If CCSD(F12*)(T*) calculations are not feasible due to the size of a molecule one might fall back to MP2-F12 as a compromise, but our data suggests that the accuracy is not significantly increased in comparison to conventional MP2, since the intrinsic error is much larger than the basis set incompleteness error.

Most DFT functionals performs worse than the wave function based methods as seen in Table 7. However TPSSH and B3-LYP enable a similar accuracy to the results obtained with

CCSD(F12*). Nevertheless, this small test set shows the large variety of possible outcomes when using different functionals. Obtaining systematically high accuracy is still a problem when using DFT.

Taking a look at the range of errors, one has to note that it can become quite large for some methods. Therefore it is interesting to check if this large range is only due to outliers or if the distribution of errors is generally broader. For that purpose a histogram of the errors in the frequencies is depicted in Fig.2 for CCSD(F12*)(T*), CCSD(F12*), MP4(F12*) and MP2-F12. Although for the given range the plot for CCSD(F12*)(T*) is simply a line, we keep it for comparison to the other methods. For CCSD(F12*) and MP2-F12 and MP4(F12*) outliers can be observed. For CCSD(F12*) only on the positive side of the x axis, for MP4(F12*) and MP2-F12 primarily on the negative side, but for MP2-F12 the distribution is broader towards the positive side. For MP2-F12 and MP4(F12*) as already pointed out N₂ is a problematic case and still errors around 150 cm⁻¹ remain. The other outermost outlier is F₂ and one frequency of HCN. F₂ is a difficult case for nearly all tested methods.

Since a substantial fraction of CC based vibrational calculations in the literature opposed to R12/F12 relies on extrapolation techniques, it is interesting to also compare this approach with the already tested explicitly correlated methods. This is also interesting from the perspective that analytical derivatives are harder to implement for F12 CC methods compared to conventional CC methods. For this purpose we did extrapolations from the cc-pVDZ-F12/cc-pVTZ-F12 (CBS(23)) as well as cc-pVTZ-F12/cc-pVQZ-F12 (CBS(34)) basis sets. We choose to continue with the F12 basis sets as compared to the smaller cc-pVXZ basis sets conventionally used, as we eventually want to compare with the F12 results. We considered to extrapolate only the correlation energy using the equation of Halkier et al.⁶⁷

$$E_{\text{corr}}^{\infty}(X, Y) = \frac{X^3 E_{\text{corr}}(X) - Y^3 E_{\text{corr}}(Y)}{X^3 - Y^3} \quad (1)$$

while keeping the HF energy of the larger basis and to extrapolate also the HF energy with

the two point extrapolation of Petersson and co-workers:⁶⁸

$$E_{\text{HF}}^{\infty}(X, Y) = \frac{e^{-a\sqrt{Y}}E_{\text{HF}}(X) - e^{-a\sqrt{X}}E_{\text{HF}}(Y)}{e^{-a\sqrt{Y}} - e^{-a\sqrt{X}}} \quad , \quad (2)$$

where $X < Y$ and a is set to 6.30 as recommended in Ref. 68. The error statistics with respect to the CCSD(F12)(T*)/cc-pVQZ-F12 reference can be found in Table 8. It is evident that the influence of the extrapolation of the HF part has virtually no effect for the CBS(23) extrapolation and a very minor effect for the CBS(34) one. Compared to the explicitly correlated results even the CBS(34) extrapolation does with a RMSD around 2.5 not yield significantly better results than with CCSD(F12*)(T*)/cc-pVDZ-F12, where we observe a RMSD of 3.02. Also the comparison to the conventional calculations shows that in the reported case extrapolation does not significantly improve the accuracy of the larger basis. In case of the CBS(23) extrapolation we even observe slightly larger errors than for CCSD(T)/cc-pVDZ-F12. But in case of CBS(34) the error statistic is however improved compared to the CCSD(T)/cc-pVQZ-F12 results and the RMSD roughly reduced by a factor of two.

We are aware that we restricted ourself to only the simplest extrapolation schemes, but nevertheless this demonstrates that the inclusion of explicitly correlated terms is the preferred choice for systematically increasing the accuracy in the calculation of harmonic frequencies. Most likely we would with a CBS(45) extrapolation enter the target regime, but the computational costs can then not compete with the CCSD(F12*)(T*)/cc-pVQZ-F12 calculations.

Table 8: Error statistic (in cm^{-1}) for the harmonic frequencies using different extrapolation schemes. Additionally the molecule with the largest error for a frequency is listed in the column Problematic case.

Method	HF extrapol?	RMSD	MEAN	MAD	RANGE	Problematic case
CCSD(T)/CBS(23)	No	11.12	-11.47	11.47	43.05	N ₂
CCSD(T)/CBS(34)	No	2.73	-1.20	1.78	13.32	NH ₃
CCSD(T)/CBS(23)	Yes	11.13	-11.55	11.55	44.98	N ₂
CCSD(T)/CBS(34)	Yes	2.56	-1.12	1.75	12.60	NH ₃

Figure 1: Error distribution of $\text{CCSD}(\text{F12}^*)(\text{T}^*)$ relative to the $\text{CCSD}(\text{F12})(\text{T})/\text{cc-pVQZ-F12}$ reference.

Accuracy in relation to anharmonic frequencies

Direct comparison and empiric corrections

The next step is to use the experimental frequencies (including anharmonicities) of the F1-sub1 set for a further assessment of $\text{CCSD}(\text{F12}^*)(\text{T}^*)$, which showed good performance while still being computational feasible. Initially we compare experimental anharmonic frequencies, with calculated harmonic frequencies. We do this not because it is theoretically attractive, but because this is a common approach as well as starting point for the pragmatic approach of using scaling factors. In this analysis (see Table 9) we again also include some popular DFT functionals, MP2 and the semi empiric HF-3c method as comparison. In contrast to the results from Table 7 the RMSD is approximately increased by a factor of 13 for $\text{CCSD}(\text{F12}^*)(\text{T}^*)$. Such large deviations may be expected due to the neglect of anharmonic effects. For the other methods the error statistics change less dramatic. For the tested DFT functionals as well the RMSD increases. In contrast to the previous part the DFT functionals perform better than the wave function based methods. This is due to the larger errors in the harmonic frequencies, which compensate errors due to anharmonic effects. The results

(a) CCSD(F12*)(T*)/QZ-F12

(b) CCSD(F12*)/QZ-F12

(c) MP4(F12*)/QZ-F12

(d) MP2-F12/QZ-F12

Figure 2: Distribution of the mean error for some of the tested electronic structure methods.

for the PBE functional are a good example. The raw RMSD for PBE is smaller than for any other tested method. However when we compared to harmonic frequencies, PBE was the functional with the largest errors.

For HF-3c the error statistics for harmonic and anharmonic frequencies are nearly the same pointing out the large error bar for this method. Since the agreement of the coupled cluster results with the reference values is very good in case of harmonic frequencies, this results give an impression on the influence of anharmonic effects. This also gives the possibility to calculate scaling factors, which are empirical factors to correct for basis set incompleteness,

anharmonic effects and other effects not present in the used model. As in Ref.⁴¹ the scaling factors are obtained by minimizing the residual

$$\Delta = \sum_i (\lambda \nu_i^{theor} - \nu_i^{expt})^2 \quad (3)$$

in a least square fit sense, which leads to the simple expression

$$\lambda = \frac{\sum_i \nu_i^{theor} \nu_i^{expt}}{\sum_i (\nu_i^{theor})^2} \quad (4)$$

for the scaling factors. Table 9 lists the factors for the different tested methods. Although the scaling factor offers a simple correction it is only an ad hoc repair and we do not recommend the general usage. The over all accuracy can be improved, but still the RMSD is around 35–40 cm⁻¹ and will likely be quite different when a different test set is used. The results do show that when such pragmatic approaches are used the differences between methods are much reduced. The lowest RMSD_{opt} are for example very similar for CCSD(F12*)(T*)/cc-pVDZ-F12 and B3-LYP/def2-TZVPP and TPSSH/def2-TZVPP. The calculated scaling factors are in agreement with that recently reported factors by Kesharwani et al.⁶⁹ However they report a smaller RMSD of 24.14 cm⁻¹ after optimization for CCSD(F12*)(T*)/cc-pVQZ-F12. But since in both studies different test sets are used a direct comparison is difficult and just indicates the range of possible outcomes if different test sets are used.

Correct inclusion of anharmonic effects

For a valid treatment to calculate anharmonic frequencies vibrational a serious attempt to solve the vibrational Schrödinger equation is mandatory. For example through self-consistent field methods (VSCF) with subsequent treatment of vibrational correlation effect using variational methods (VCI) or coupled cluster methods (VCC). To address this issue some of the

³For this basis set the test set had to be slightly decreased and contains 100 instead of 110 molecules. (See supportive information for details).

Table 9: Root mean square deviation for the experimental frequencies (in cm^{-1}) using different methods. The scaling factors are listed as well as the root mean square deviation using the scaling factor.

Method	Basis set	RMSD	λ	RMSD_{opt}
CCSD(F12*)(T*)	cc-pVDZ-F12	70.34	0.96402	33.95
CCSD(F12*)(T*)	cc-pVTZ-F12	70.81	0.96376	34.07
CCSD(F12*)(T*)	cc-pVQZ-F12 ³	68.79	0.96439	36.30
MP2	cc-pVDZ-F12	85.63	0.95667	46.46
PBE	def2-TZVPP	41.76	0.99454	40.77
B3-LYP	def2-TZVPP	63.71	0.96868	34.79
TPSS	def2-TZVPP	48.55	0.98218	38.16
TPSSH	def2-TZVPP	61.87	0.96977	34.35
HF-3c	minix	315.73	0.84458	84.90

molecules (H_2O , CO_2 , H_2S , H_2CO , N_2 , F_2) in the F1-sub2 set are picked out for a further analysis with means of VCC[3] calculations using potentials with at most 3 mode couplings. For the molecules in the test set with 3 or less atoms this leads to a full dimensional description of the PES and the vibrational wave function. With the exception of N_2 and F_2 we included a mass-dependent kinetic energy contribution to the vibrational Hamiltonian, which is usually referred to as Watson correction. This correction has for light molecules usually a substantial impact and we also observed a correction up to 15 cm^{-1} . As described in Ref. 70 we used a simplified form of the Watson kinetic energy term by approximating the elements of the μ tensor with their equilibrium values. For some of the test molecules we confirmed that the errors due to this approximation are in the sub wave-number regime. The electronic structure single point calculations were done at the CCSD(F12*)(T*) and CCSD(F12*)(T) level of theory using the cc-pVDZ-F12, cc-pVTZ-F12 and cc-pVQZ-F12 basis.

To study the basis set convergence we consider the results obtained with the PES using the cc-pVQZ-F12 basis as internal reference. Table 10 lists the RMSD for the different molecules to this reference and experiment. Using the cc-pVDZ-F12 basis at maximum a RMSD of 5 cm^{-1} can be observed, but there are in fact also cases with a RMSD around 1 cm^{-1} or below. However for F_2 the origin of a low RMSD below 1 cm^{-1} is due to error cancellations in the

small basis and it increases when the basis set is increased. If one uses the cc-pVTZ-F12 basis in the PES construction the RMSD is typically below 1 cm^{-1} or slightly above. This trend is consistent with the observation made during the calculation of harmonic frequencies. Also Kahn et al.⁷¹ demonstrated in their calculation of cubic and quartic force constants for water using explicitly correlated methods that the enhanced basis set convergence takes over to the calculation of anharmonic frequencies. Nevertheless, we see that the RMSDs for the small basis set are likely to be a bit larger for the anharmonic treatment than in the context of calculating harmonic frequencies. The comparison to experiment shows that in most cases it is possible to reach a fairly good accuracy with a RMSD around 3 cm^{-1} . The remaining disagreement is likely related to missing higher order excitations, missing core correlation, and relativistic effects in the potential.

Complementary to these findings we repeated this calculations with a PES constructed using CCSD(F12*)(T) calculations. The results can be found in Table 11. It is evident that the basis set convergence is slower and even in the cc-pVTZ-F12 basis the RMSD to the cc-pVQZ-F12 results is typically above 1 cm^{-1} . If one takes a look at the RMSD to experimental values the basis set convergence of the frequencies is also slower.

There are also examples, which show in the cc-pVDZ-F12 basis already a quite good agreement, but this agreement is than due to error cancellation. As an example we observe for CO_2 better results using the PES constructed from CCSD(F12*)(T*) calculations in the same basis. However this better agreement is attributed to error cancellation. If the basis set is increased the RMSD converges towards the results of CCSD(F12*)(T*), which indicates that also in the context of calculating anharmonic frequencies CCSD(F12*)(T*) is the preferred choice over CCSD(F12*)(T).

Overtones

So far we considered fundamental frequencies and found that the triples scaling enhances the accuracy in the potential. Besides fundamentals also overtones can be of interest. Compared

Table 10: Root mean square deviation of the anharmonic fundamental frequencies (in cm^{-1}) obtained with different PESs varying the basis set in the electronic structure calculation. The RMSDs to the cc-pVQZ-F12 results as well as to experiment (RMSD_{exp}) are given. The single point calculations were carried out on the CCSD(F12*)(T*) level of theory.

Molecule	PES	RMSD	RMSD_{exp}
H ₂ O	cc-pVDZ-F12	1.04	3.71
H ₂ O	cc-pVTZ-F12	0.64	3.12
H ₂ O	cc-pVQZ-F12	0.00	3.03
CO ₂	cc-pVDZ-F12	4.43	3.52
CO ₂	cc-pVTZ-F12	0.22	1.03
CO ₂	cc-pVQZ-F12	0.00	0.97
H ₂ S	cc-pVDZ-F12	5.05	10.35
H ₂ S	cc-pVTZ-F12	0.86	6.31
H ₂ S	cc-pVQZ-F12	0.00	5.65
H ₂ CO	cc-pVDZ-F12	3.32	2.69
H ₂ CO	cc-pVTZ-F12	0.66	3.07
H ₂ CO	cc-pVQZ-F12	0.00	3.45
N ₂	cc-pVDZ-F12	3.03	3.29
N ₂	cc-pVTZ-F12	0.01	6.33
N ₂	cc-pVQZ-F12	0.00	6.32
F ₂	cc-pVDZ-F12	0.43	14.99
F ₂	cc-pVTZ-F12	1.05	13.50
F ₂	cc-pVQZ-F12	0.00	14.55

Table 11: Root mean square deviation of the anharmonic fundamental frequencies (in cm^{-1}) obtained with different PESs varying the basis set in the electronic structure calculation. The RMSDs to the cc-pVQZ-F12 results as well as to experiment (RMSD_{exp}) are given. The single point calculations were carried out on the CCSD(F12*)(T) level of theory.

Molecule	PES	RMSD	RMSD_{exp}
H ₂ O	cc-pVDZ-F12	2.70	6.19
H ₂ O	cc-pVTZ-F12	1.24	5.17
H ₂ O	cc-pVQZ-F12	0.00	4.00
CO ₂	cc-pVDZ-F12	1.31	1.91
CO ₂	cc-pVTZ-F12	1.09	3.03
CO ₂	cc-pVQZ-F12	0.00	1.97
H ₂ S	cc-pVDZ-F12	7.98	14.11
H ₂ S	cc-pVTZ-F12	1.51	7.80
H ₂ S	cc-pVQZ-F12	0.00	6.48
H ₂ CO	cc-pVDZ-F12	3.21	5.61
H ₂ CO	cc-pVTZ-F12	0.81	4.62
H ₂ CO	cc-pVQZ-F12	0.00	4.20
N ₂	cc-pVDZ-F12	2.09	9.65
N ₂	cc-pVTZ-F12	1.37	8.93
N ₂	cc-pVQZ-F12	0.00	7.56
F ₂	cc-pVDZ-F12	12.65	31.35
F ₂	cc-pVTZ-F12	2.36	21.06
F ₂	cc-pVQZ-F12	0.00	18.70

to fundamentals the overtones will probe wider parts of the potential. As an additional test we therefore calculated for N_2 and FH 10 excitation energies and studied again the basis set convergence in the PES using $\text{CCSD}(\text{F12}^*)(\text{T}^*)$ and $\text{CCSD}(\text{F12}^*)(\text{T})$. Figure 3 shows a bar chart with the different errors with respect to the selected cc-pVQZ-F12 reference for both electronic structure methods. Going to higher and higher excitation energies the errors increase. This has two reasons: First the (excitation) energy is an extensive property and therefore its error has to be extensive as well and increases together with the energy. Second it becomes more and more difficult to describe the overtones.

Like for the already studied cases the triples scaling improves the convergence and reduces the errors to the reference. However one has also to note that for the case of N_2 the errors for the cc-pVDZ-F12 basis have a different sign than for the other cases and that the errors first increase and afterwards decrease again. Although the results with the PES obtained from $\text{CCSD}(\text{F12}^*)(\text{T})$ energies show larger errors, they are more systematic in this regard. Since one is often only interested in the first few overtones, we conclude from the current observations that the triples scaling does not introduce large artifacts and do give a cost effective improvement. Although there is a potential issue due to the scaling, it is probably not of practical significance in most applications. From the study on harmonic frequencies it is evident that N_2 is a very problematic case for MP2 and MP4 and still we do not observe a major break down due to the scaling of the ratio of the MP2-F12 and MP2 correlation energy here, where we probe a wider range of the PES.

(a) N_2

(b) FH

Figure 3: Error in the excitation energies to the chosen cc-pVQZ-F12 reference for PES obtained with CCSD(F12*)(T*) and CCSD(F12*)(T) calculations.

Outlook and Conclusion

We studied the usage of explicitly correlated wave function based methods in the context of calculating harmonic and anharmonic frequencies. For calculation of harmonic frequencies a hierarchy of the different tested wave function based methods could be established. The observed trend is in agreement with similar studies dealing with reaction and atomization energies. CCSD(F12*)(T) is a close approximation to the parent CCSD(F12)(T) method and its accuracy can be enhanced in the CCSD(F12*)(T*) method if the triples correction is scaled by the ratio of the MP2-F12 and MP2 correlation energy to approximately include explicitly correlated effects also in the triples excitations. However the scaling leads for super molecules to unphysical results and more systematic models to include explicitly correlated terms in the triples correction are in these cases required. As often observed the basis set convergence is very fast and results in the cc-pVTZ-F12 basis are only slightly inferior than calculations in the larger cc-pVQZ-F12 basis. Also the usage of the cc-pVDZ-F12 basis allows results with a RMSD below 4 cm^{-1} to the CBS limit. If CCSD(F12*)(T*) calculations are not feasible a natural choice is to use MP2-F12 calculations as a replacement. However we find that due to the large intrinsic MP2 error the accuracy is not significantly increased in comparison to conventional MP2.

For the calculation of anharmonic frequencies the observed basis set dependence of the harmonic frequencies is to a large extent transferable. For the cc-pVDZ-F12 basis a RMSD around 5 cm^{-1} was found, while in the harmonic case it is below 4 cm^{-1} . Also for the calculation of anharmonic frequencies we find that CCSD(F12*)(T*) is superior to CCSD(F12*)(T) if fundamentals and overtones are considered. For CCSD(F12*)(T) the basis set convergence is much slower and the results converge to the CCSD(F12*)(T*) results. This indicates that the triples scaling causes no problem and only enhances the convergence. Due to the findings we plan to report further studies on anharmonic frequencies using the CCSD(F12*)(T*)/cc-pVDZ-F12 calculations for the generation of the PES since it allows accurate frequencies while still being economical for systems larger than in the current study. Furthermore the

obtained harmonic frequencies and geometries on the CCSD(F12*)(T*)/cc-pVQZ-F12 level can be used for further assessments of more approximate methods.

Acknowledgment

G.S. thanks Nora Graf for implementing the numerical gradient and hessian in TURBO-MOLE and granting the permission to use her code.

Furthermore we thank the Villum foundation for financial support. In addition O.C. acknowledges support from the Lundbeck Foundation, the Danish e-infrastructure Cooperation (DeiC), and the Danish Council for Independent Research through a Sapere Aude III grant (DFR – 4002-00015).

Additional information

All obtained structures and frequencies are assembled in the supporting material.

References

- (1) Clabo, D.; Allen, W. D.; Remington, R. B.; Yamaguchi, Y.; Schaefer, H. F. A Systematic Study of Molecular Vibrational Anharmonicity and Vibration-Rotation Interaction by Self-Consistent-Field Higher-Derivative Methods. Asymmetric Top Molecules. *Chem. Phys.* **1988**, *123*, 187–239.
- (2) Barone, V. Anharmonic Vibrational Properties by a Fully Automated Second-Order Perturbative Approach. *J. Chem. Phys.* **2005**, *122*, 014108–1–014108–9.
- (3) Vazquez, J.; Stanton, J. F. Simple(r) Algebraic Equation for Transition Moments of Fundamental Transitions in Vibrational Second-Order Perturbation Theory. *Mol. Phys.* **2006**, *104*, 377–388.

- (4) Bowman, J. M.; Christoffel, K.; Tobin, F. Application of SCF-SI Theory to Vibrational Motion in Polyatomic Molecules. *The Journal of Physical Chemistry* **1979**, *83*, 905–912.
- (5) Carter, S.; Bowman, J. M.; Handy, N. C. Extensions and Tests of “Multimode”: A Code to obtain Accurate Vibration/Rotation Energies of Many-Mode Molecules. *Theor. Chem. Acc.* **1998**, *100*, 191–198.
- (6) Christiansen, O. Vibrational Coupled Cluster Theory. *J. Chem. Phys.* **2004**, *120*, 2149–2159.
- (7) Prasad, M. D. Time-Dependent Coupled Cluster Method: A New Approach to the Calculation of Molecular Absorption Spectra. *J. Chem. Phys.* **1988**, *88*, 7005–7010.
- (8) Fauchaux, J. A.; Hirata, S. Higher-Order Diagrammatic Vibrational Coupled-Cluster Theory. *J. Chem. Phys.* **2015**, *143*, 134105.
- (9) Kutzelnigg, W. r12-Dependent Terms in the Wave Function as Closed Sums of Partial Wave Amplitudes for Large l. *Theor. Chim. Acta* **1985**, *68*, 445–469.
- (10) Klopper, W.; Kutzelnigg, W. MP2-R12 Calculations on the Relative Stability of Carbocations. *The Journal of Physical Chemistry* **1990**, *94*, 5625–5630.
- (11) Klopper, W.; Samson, C. C. M. Explicitly Correlated Second-Order Møller-Plesset Methods with Auxiliary Basis Sets. *J. Chem. Phys.* **2002**, *116*, 6397–6410.
- (12) Martin, J. M. L.; Kesharwani, M. K. Assessment of CCSD(T)-F12 Approximations and Basis Sets for Harmonic Vibrational Frequencies. *J. Chem. Theory Comput.* **2014**, *10*, 2085–2090, PMID: 26580535.
- (13) Rauhut, G.; Knizia, G.; Werner, H.-J. Accurate Calculation of Vibrational Frequencies using Explicitly Correlated Coupled-Cluster Theory. *J. Chem. Phys.* **2009**, *130*, 054105.

- (14) Rodriguez-Betancourt, V.-M.; Quezada-Navarro, V.-M.; Neff, M.; Rauhut, G. Anharmonic Frequencies of [F, C, N, X] Isomers (X = O, S) obtained from Explicitly Correlated Coupled-Cluster Calculations. *Chem. Phys.* **2011**, *387*, 1 – 4.
- (15) Botschwina, P.; Oswald, R.; Rauhut, G. Explicitly Correlated Coupled Cluster Calculations for the Propargyl Cation (H₂C₃H⁺) and Related Species. *Phys. Chem. Chem. Phys.* **2011**, *13*, 7921–7929.
- (16) Tew, D. P.; Klopper, W.; Bachorz, R. A.; Hättig, C. *Handbook of High-resolution Spectroscopy*; John Wiley & Sons, Ltd, 2011.
- (17) Tew, D. P.; Klopper, W.; Heckert, M.; Gauss, J. Basis Set Limit CCSD(T) Harmonic Vibrational Frequencies. *J. Phys. Chem. A* **2007**, *111*, 11242–11248, PMID: 17511434.
- (18) Heckert, M.; Kállay, M.; Gauss, J. Molecular equilibrium geometries based on coupled-cluster calculations including quadruple excitations. *Mol. Phys.* **2005**, *103*, 2109–2115.
- (19) Cortez, M. H.; Brinkmann, N. R.; Polik, W. F.; Taylor, P. R.; Bomble, Y. J.; Stanton, J. F. Factors Contributing to the Accuracy of Harmonic Force Field Calculations for Water. *J. Chem. Theory Comput.* **2007**, *3*, 1267–1274, PMID: 26633200.
- (20) Knizia, G.; Adler, T. B.; Werner, H.-J. Simplified CCSD(T)-F12 Methods: Theory and Benchmarks. *J. Chem. Phys.* **2009**, *130*, 054104.
- (21) Ahnen, S.; Hehn, A.-S.; Vogiatzis, K. D.; Trachsel, M. A.; Leutwyler, S.; Klopper, W. Accurate computations of the structures and binding energies of the imidazole...benzene and pyrrole...benzene complexes. *Chem. Phys.* **2014**, *441*, 17 – 22.
- (22) Marchetti, O.; Werner, H.-J. Accurate Calculations of Intermolecular Interaction Energies Using Explicitly Correlated Coupled Cluster Wave Functions and a Dispersion-Weighted MP2 Method. *J. Phys. Chem. A* **2009**, *113*, 11580–11585, PMID: 19807061.

- (23) Köhn, A. Explicitly Correlated Connected Triple Excitations in Coupled-Cluster Theory. *J. Chem. Phys.* **2009**, *130*, 131101.
- (24) Köhn, A. Explicitly Correlated Coupled-Cluster Theory using Cusp Conditions. II. Treatment of Connected Triple Excitations. *J. Chem. Phys.* **2010**, *133*, 174118.
- (25) Wang, Y.; Braams, B. J.; Bowman, J. M.; Carter, S.; Tew, D. P. Full-dimensional quantum calculations of ground-state tunneling splitting of malonaldehyde using an accurate ab initio potential energy surface. *J. Chem. Phys.* **2008**, *128*, 224314.
- (26) Qu, C.; Bowman, J. M. An ab initio potential energy surface for the formic acid dimer: zero-point energy, selected anharmonic fundamental energies, and ground-state tunneling splitting calculated in relaxed 1-4-mode subspaces. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24835–24840.
- (27) Mizukami, W.; Habershon, S.; Tew, D. P. A compact and accurate semi-global potential energy surface for malonaldehyde from constrained least squares regression. *J. Chem. Phys.* **2014**, *141*, 144310.
- (28) Braams, B. J.; Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577–606.
- (29) Manzhos, S.; Wang, X.; Dawes, R.; Carrington, T. A Nested Molecule-Independent Neural Network Approach for High-Quality Potential Fits. *J. Phys. Chem. A* **2006**, *110*, 5295–5304.
- (30) Yurchenko, S. N.; Breidung, J.; Thiel, W. Vibrational spectrum of BiH₃: Six-dimensional variational calculations on high-level ab initio potential energy surfaces. *Theor. Chem. Acc.* **2005**, *114*, 333–340.
- (31) Fliegl, H.; Klopper, W.; Hättig, C. Coupled-Cluster Theory with Simplified Linear-r12 Corrections: The CCSD(R12) Model. *J. Chem. Phys.* **2005**, *122*, 084107.

- (32) Tew, D. P.; Klopper, W. New Correlation Factors for Explicitly Correlated Electronic Wave Functions. *J. Chem. Phys.* **2005**, *123*, 074101–074101.
- (33) Ten-no, S. Initiation of explicitly correlated Slater-type geminal theory. *Chem. Phys. Lett.* **2004**, *398*, 56 – 61.
- (34) Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. Explicitly Correlated Electrons in Molecules. *Chem. Rev.* **2012**, *112*, 4–74.
- (35) Grüneis, A.; Hirata, S.; ya Ohnishi, Y.; Ten-no, S. Perspective: Explicitly correlated electronic structure theory for complex systems. *J. Chem. Phys.* **2017**, *146*, 080901.
- (36) Ten-no, S. Explicitly Correlated Second Order Perturbation Theory: Introduction of a Rational Generator and Numerical Quadratures. *J. Chem. Phys.* **2004**, *121*, 117–129.
- (37) Ten-no, S. New Implementation of Second-Order Møller-Plesset Perturbation Theory with an Analytic Slater-Type Geminal. *J. Chem. Phys.* **2007**, *126*, 014108.
- (38) Hättig, C.; Tew, D. P.; Köhn, A. Communications: Accurate and Efficient Approximations to Explicitly Correlated Coupled-Cluster Singles and Doubles, CCSD-F12. *J. Chem. Phys.* **2010**, *132*, 231102.
- (39) Adler, T. B.; Knizia, G.; Werner, H.-J. A Simple and Efficient CCSD(T)-F12 Approximation. *J. Chem. Phys.* **2007**, *127*, 221106.
- (40) Torheyden, M.; Valeev, E. F. Variational Formulation of Perturbative Explicitly-Correlated Coupled-Cluster Methods. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3410–3420.
- (41) Merrick, J. P.; Moran, D.; Radom, L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A* **2007**, *111*, 11683–11700.
- (42) Strey, G. Extremal Properties of Force Constants. *J. Mol. Spectrosc.* **1967**, *24*, 87–99.

- (43) Strey, G.; Mills, I. The Anharmonic Force Field and Equilibrium Structure of HCN and HCP. *Mol. Phys.* **1973**, *26*, 129–138.
- (44) Gershikov, A.; Spiridonov, V. Anharmonic Force Field of CO₂ as Determined by a Gas-Phase Electron Diffraction Study. *J. Mol. Struct.* **1983**, *96*, 141–149.
- (45) Strey, G.; Mills, I. Anharmonic force field of acetylene. *J. Mol. Spectrosc.* **1976**, *59*, 103 – 115.
- (46) Jr., H. C. A.; Plyler, E. K. Infrared Spectrum of Hydrogen Sulfide. *J. Chem. Phys.* **1956**, *25*, 1132–1136.
- (47) Yachmenev, A.; Yurchenko, S. N.; Jensen, P.; Thiel, W. A new spectroscopical potential energy surface for formaldehyde in its ground electronic state. *J. Chem. Phys.* **2011**, *134*, 244307.
- (48) Wang, D.; Shi, Q.; Zhu, Q.-S. An Ab Initio Quartic Force Field of PH₃. *J. Chem. Phys.* **2000**, *112*, 9624–9631.
- (49) Hänninen, V.; Halonew, L. Calculation of Spectroscopic Parameters and Vibrational Overtones of Methanol. *Mol. Phys.* **2003**, *101*, 2907–2916.
- (50) Huang, X.; Schwenke, D. W.; Lee, T. J. Rovibrational spectra of ammonia. I. Unprecedented accuracy of a potential energy surface used with nonadiabatic corrections. *J. Chem. Phys.* **2011**, *134*, 044320.
- (51) Huber, K.; Herzberg, G. Constants of Diatomic Molecules (data prepared by J.W. Gallagher and R.D. Johnson, III). 1979.
- (52) Majumder, M.; Hegger, S. E.; Dawes, R.; Manzhos, S.; Wang, X.-G.; Jr., C. T.; Li, J.; Guo, H. Explicitly correlated MRCI-F12 potential energy surfaces for methane fit with several permutation invariant schemes and full-dimensional vibrational calculations. *Mol. Phys.* **2015**, *113*, 1823–1833.

- (53) Martin, J. M.; Taylor, P. R. The geometry, vibrational frequencies, and total atomization energy of ethylene. A calibration study. *Chem. Phys. Lett.* **1996**, *248*, 336 – 344.
- (54) TURBOMOLE V7.1 2016, a Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2017, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>.
- (55) Christiansen, O. et al. MidasCpp (Molecular Interactions, Dynamics and Simulation Chemistry Program Package in C++). 2017; www.chem.au.dk/midas.
- (56) Peterson, K. A.; Adler, T. B.; Werner, H.-J. Systematically Convergent Basis Sets for Explicitly Correlated Wavefunctions: The Atoms H, He, B-Ne, and Al-Ar. *J. Chem. Phys.* **2008**, *128*, 084102.
- (57) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (58) Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J. Comput. Chem.* **2013**, *34*, 1672–1685.
- (59) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (60) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (61) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

- (62) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- (63) Seidler, P.; Christiansen, O. Vibrational Excitation Energies from Vibrational Coupled Cluster Response Theory. *J. Chem. Phys.* **2007**, *126*, 204101.
- (64) Toffoli, D.; Sparta, M.; Christiansen, O. Accurate Multimode Vibrational Calculations using a B-Spline Basis: Theory, Tests and Application to Dioxirane and Diazirine. *Mol. Phys.* **2011**, *109*, 673–685.
- (65) Sparta, M.; Toffoli, D.; Christiansen, O. An Adaptive Density-Guided Approach for the Generation of Potential Energy Surfaces of Polyatomic Molecules. *Theor. Chem. Acc.* **2009**, *123*, 413–429.
- (66) Heckert, M.; Kállay, M.; Tew, D. P.; Klopper, W.; Gauss, J. Basis-set extrapolation techniques for the accurate calculation of molecular equilibrium geometries using coupled-cluster theory. *J. Chem. Phys.* **2006**, *125*, 044108.
- (67) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. Basis-Set Convergence of the Energy in Molecular Hartree-Fock Calculations. *Chem. Phys. Lett.* **1999**, *302*, 437–446.
- (68) Zhong, S.; Barnes, E. C.; Petersson, G. A. Uniformly Convergent n-Tuple- ζ Augmented Polarized (nZaP) Basis Sets for Complete Basis Set Extrapolation. I. Self-Consistent Field Energies. *J. Chem. Phys.* **2008**, *129*, 184116.
- (69) Kesharwani, M. K.; Brauer, B.; Martin, J. M. L. Frequency and Zero-Point Vibrational Energy Scale Factors for Double-Hybrid Density Functionals (and Other Selected Methods): Can Anharmonic Force Fields Be Avoided? *J. Phys. Chem. A* **2015**, *119*, 1701–1714.

- (70) Toffoli, D.; Kongsted, J.; Christiansen, O. Automatic Generation of Potential Energy and Property Surfaces of Polyatomic Molecules in Normal Coordinates. *J. Chem. Phys.* **2007**, *127*, 204106.
- (71) Kahn, K.; Kirtman, B.; Noga, J.; Ten-no, S. Anharmonic vibrational analysis of water with traditional and explicitly correlated coupled cluster methods. *J. Chem. Phys.* **2010**, *133*, 074106.