# The MRCC program system: Accurate quantum chemistry from water to proteins

Mihály Kállay[a], Péter R. Nagy, Dávid Mester, Zoltán Rolik, Gyula Samu, József Csontos, József Csóka, P. Bernát Szabó, László Gyevi-Nagy, Bence Hégely, István Ladjánszki, Lóránt Szegedy[b], Bence Ladóczki[c], Klára Petrov, Máté Farkas[d], Pál D. Mezei[e], Ádám Ganyecz[1]

*Department of Physical Chemistry and Materials Science,*

*Budapest University of Technology and Economics, H-1521 Budapest, P.O.Box 91,*

*Hungary*

(Dated: 16 January 2020)

---

[a] Electronic mail: kallay@mail.bme.hu

[b] Present address: Universität Hamburg, Bundesstraße 55, D-20146 Hamburg

[c] Present address: Graduate School of System Informatics, Kobe University, Nada-ku, Kobe 657-8501, Japan

[d] Present address: Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics and Informatics, University of Gdansk, 80-952 Gdansk, Poland; and International Centre for Theory of Quantum Information, University of Gdansk, 80-308 Gdansk, Poland

[e] Present address: Department of Chemistry, University of Basel, Basel 4056, Switzerland

Mrcc is a package of *ab initio* and density functional quantum chemistry programs for accurate electronic structure calculations. The suite has efficient implementations of both low- and high-level correlation methods, such as second-order Møller–Plesset (MP2), random-phase approximation (RPA), second-order algebraic-diagrammatic construction [ADC(2)], coupled-cluster (CC), configuration interaction (CI), and related techniques. It has a state-of-the-art CC singles and doubles with perturbative triples [CCSD(T)] code, and its specialties, the arbitrary-order iterative and perturbative CC methods developed by automated programming tools enable achieving convergence with regard to the level of correlation. The package also offers a collection of multi-reference CC and CI approaches. Efficient implementations of density functional theory (DFT) and more advanced combined DFT-wave function approaches are also available. Its other special features, the highly-competitive linear-scaling local correlation schemes, allow for MP2, RPA, ADC(2), CCSD(T), and higher-order CC calculations for extended systems. Local correlation calculations can be considerably accelerated by multi-level approximations and DFT-embedding techniques, and an interface to molecular dynamics software is provided for quantum mechanics/molecular mechanics (QM/MM) calculations. All components of Mrcc support shared-memory parallelism, and multi-node parallelization is also available for various methods. For academic purposes, the package is available free of charge.

## I. INTRODUCTION

MRCC is a versatile suite of quantum chemistry programs focusing on *ab initio* electron correlation and density functional methods. The package also serves as a platform of method development aiming at extending the reach of highly reliable methods for the accurate and efficient simulation of molecular systems. The development of the MRCC program system was started at the Eötvös University, Budapest, in 2000. Between 2002 and 2004 the program development was continued at the University of Mainz, Germany. From late 2004, the program has been developed at the Budapest University of Technology and Economics. The program has been officially distributed since 2005 and currently has more than 500 registered users.

The original focus of the development was on the automated implementation of high-order coupled-cluster (CC) and configuration interaction (CI) methods. The special techniques based on the strings of spin-orbital indices facilitated the programming of arbitrary single-reference CC methods, such as CCSD, CCSDT, CCSDTQ, CCSDTQP and so on[1] (see Tables I and IV for the meaning of acronyms). Utilizing the efficient high-order CC implementations, it became feasible to get reliable estimates for the correlation contributions beyond CCSD(T) for smaller molecules. Several high-accuracy computational protocols were designed based on this code, which were capable of approaching experimental accuracy for thermochemical quantities or even providing more accurate results than the corresponding experiments (for representative examples see Refs. 2–6). Subsequently, hierarchies of perturbative high-order CC models, for instance, CCSDT(Q), were developed[7,8], using which the post-CCSD(T) correlation contributions could be computed more economically. This allowed not only for the development of more advanced thermochemistry protocols (see, e.g., Refs. 9–12) but also for high-accuracy calculations for relatively large systems, such as the benzene molecule[13]. While the original program used the standard nonrelativistic Hamiltonian, later it was also combined with various relativistic approaches thereby extending the scope of high-order CC methods to smaller molecules composed of heavy elements[14].

The automated programming tools also enabled the rapid implementation and testing of sophisticated active-space and multi-reference CC (MRCC) approaches. First, the infrastructure for the separate treatment of active and inactive orbital spaces was established, and a general active-space CC algorithm was developed[15]. Later, a real state-selective MRCC

ansatz was coded[16], and its improved variants were also proposed and tested[17,18]. We note that the name of the program suite also stems from the MRCC acronym. The first users of the code dubbed it Mrcc, and this name stuck to the program, even if its feature list has been considerably extended since then.

Besides energy calculations, significant effort was devoted to the implementation of properties for general CC methods. Analytic gradient[19], Hessian[20], and third derivative algorithms[21] were developed for arbitrary CC approaches. Based on these analytic derivatives, numerous properties were made available, such as dipole and higher moments, geometry[19], vibrational frequencies, NMR chemical shifts[20], magnetizabilities[22], static and frequency-dependent polarizabilities[23], electronic $g$-tensors[24], diagonal Born–Oppenheimer corrections (DBOCs)[25], hyperpolarizabilities[21], and Raman intensities[26]. For the accurate calculations of excitation energies and excited-state first-order and transition properties a linear response code was also written[27].

Originally, Mrcc was not a standalone program. It was interfaced to other quantum chemistry codes for transformed molecular orbital (MO) integrals, only the correlation calculations were performed by it, and MO density matrices were returned in the case of property calculations. However, our later research projects entailed the development of Mrcc to a versatile standalone suite. A new generation of programs was added including four- and three-center atomic orbital (AO) electron repulsion integral (ERI) codes[28–30]. Hartree–Fock (HF), Kohn–Sham (KS), and multiconfigurational (MC) self-consistent field (SCF) programs were written[28]. Numerous density functional theory (DFT) functionals were added[31,32]. Efficient algorithms for standard quantum chemical approaches were implemented including, for example, the MP2[33], dRPA, SOSEX[31], CIS[34], CIS(D)[35], TDDFT[34], CC2[36], ADC(2)[35], and CCSD(T)[37,38] theories. The infrastructure for geometry optimizations, vibrational frequencies, and other properties was built up, and several wave function analysis tools were added.

In the past decade, the direction of the program development was mainly determined by our interest in the acceleration of accurate quantum chemical calculations. Efficient cost-reduction techniques were elaborated for general CC[39], CC2[36], and ADC(2)[35] methods relying on active-space and orbital transformation techniques. A new approach proposed for the cost reduction of density fitting (DF) methods was also implemented[33]. For large molecules, a new class of local correlation methods was developed including the local

MP2[40], the local dRPA[31], and the local natural orbital (LNO) CCSD(T) and higher-order CC approaches[28,37,41–43]. These local correlation methods turned out to be highly competitive with similar models advocated by other groups[43,44] and enable correlated calculations for thousands of atoms[31,40,42,43]. These advancements also necessitated the improvement of our HF code, which had become the bottleneck in the local correlation calculations. Considerable effort was invested in speeding up the SCF calculations and handling the linear dependencies of the basis sets, and low-order-scaling exchange builder algorithms were coded[40,43]. Concerning excited states, cubic-scaling CIS and TDDFT methods were developed utilizing local fitting domains[34], while, for correlated excited-states methods, such as ADC(2), a sublinear scaling local correlation approach was added[45].

Parallel to the above efforts we also concentrated on the improvement of DFT methods. Besides the design of new "pure" functionals[32], the main focus was on combined DFT-wave function models[46–52]. Several new types of double hybrid (DH) functionals based on dRPA and SOSEX were developed[46,47,49,50] and implemented in Mrcc together with the MP2-based DHs put forward by other groups. For excited states, our new TDDFT-ADC(2) composite approach[48] is available in addition to the CIS(D)-based excited-state DHs. Several multiscale techniques are also provided, such as our DFT embedding approach proposed recently[51,52], or the well-established ONIOM (our own n-layered integrated molecular orbital and molecular mechanics) model. For very large systems, the computations can be further accelerated by the standard quantum mechanics/molecular mechanics (QM/MM) technique via an interface to molecular dynamics (MD) software[53].

In this publication, we review the various features of the Mrcc package, make recommendations concerning the usage of the methods, and provide brief benchmark data. Finally, the future directions of the code development are outlined.

## II. FEATURES, RECOMMENDATIONS, AND BENCHMARKS

Here, we discuss the capabilities of Mrcc grouped according to the various types of methods. For convenience, the list of the currently available methods is also presented in Tables I, IV, and V, while representative benchmark results are compiled in Tables II and III. The geometries for the test systems employed in the benchmark calculations can be found in our previous publications[2,31,40,42]. The core electrons were kept frozen in all correlation

and excited-state calculations. If not noted otherwise, the reported timings are wall-clock times measured on a machine with 128 GB of main memory and a 6-core 3.5 GHz Intel Xeon E5-1650 processor.

## A.  Hartree–Fock and multiconfigurational SCF

MRCC provides RHF, UHF, and ROHF programs. All of these can use both the conventional four-center ERIs-based algorithms and the DF technique. In any case, the SCF calculation can be run in disk-based mode, when the precomputed ERIs are stored on disk and retrieved in each iteration step, or in direct mode, when the ERIs are computed on the fly. Unless the disk input/output (I/O) is very efficient, the direct algorithms are usually faster as the integrals are computed in parallel on several CPU cores. The use of four-center integrals is recommended only when highly accurate calculations, such as high-order CC calculations are carried out. Otherwise, especially for bigger molecules, the much faster DF-HF algorithms should be used. For large systems, all types of DF-HF calculations can be further sped up by local DF approaches (see Sect. II E). The SCF code is equipped with the standard convergence acceleration techniques, such as the direct inversion in the iterative subspace (DIIS), damping of density matrices, and level shifting. The iterations can be started from various initial guesses or restarted from smaller basis set calculations. Analytic gradients are available for RHF and UHF for both the conventional and the DF approaches.

To illustrate the efficiency of our SCF code, we performed benchmark calculations for a DNA fragment of two adenine-thymine base pairs ($DNA_2$ for short) of 128 atoms (see Table II). The DF-HF calculations required 11.6, 48.7, 47.7, and 179.6 minutes per iteration step with the cc-pVDZ, aug-cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets (see Sect. II K for the definition of basis sets). 12 (14) SCF cycles were required to reach convergence with the cc-pV$X$Z (aug-cc-pVDZ) basis sets starting the iteration from the superposition of atomic densities initial guess. The calculation of analytic gradients takes 13.9, 55.2, 57.4, and 339.2 minutes with the above basis sets, respectively.

For more complicated cases quadratic SCF algorithms are implemented. These are based on the Newton method or its trust region extensions utilizing augmented Hessian approaches[54], which can also be combined with line search. The quadratic SCF algorithms

TABLE I: Methods implemented in Mrcc for ground states.

| Method | | References | | Grad.[c] |
|---|---|---|---|---|
| Abbreviation | Full name/description | Meth.[a] | Imp.[b] | |
| RHF | restricted Hartree–Fock | 55 | 28 | + |
| UHF | unrestricted Hartree–Fock | 55 | 28 | + |
| ROHF | restricted open-shell Hartree–Fock | 56 | 28 | − |
| MCSCF | multiconfigurational self-consistent field | 54 | 57 | − |
| RKS | restricted Kohn–Sham | 58 | 31 | + |
| UKS | unrestricted Kohn–Sham | 58 | 31 | + |
| ROKS | restricted open-shell Kohn–Sham | 58,59 | 31 | − |
| DH-DFT | double hybrid DFT | 60 | 46,47 | + |
| MP2 | second-order Møller–Plesset perturbation theory | 61 | 33 | + |
| SCS-MP2 | spin-component scaled MP2 | 62 | 33 | − |
| SOS-MP2 | scaled opposite-spin MP2 | 63 | 33 | − |
| dRPA | direct random-phase approximation | 64 | 31,33 | − |
| SOSEX | second-order screened exchange | 65 | 31 | − |
| rPT2 | renormalized second-order perturbation theory | 66 | 49 | − |
| RPA | random-phase approximation | 67 | 31 | − |
| RPAX2 | approximate RPA with exchange | 68 | 46 | − |
| CC2 | second-order CC singles and doubles | 69 | 36 | − |
| SCS-CC2 | spin-component scaled CC2 | 70 | 36,48 | − |
| SOS-CC2 | scaled opposite-spin CC2 | 71 | 36,48 | − |
| CCSD | CC singles and doubles | 72,73 | 28,38 | + |
| CCSDT | CC singles, doubles, and triples | 74 | 1 | + |
| CCSDTQ | CC singles, doubles, triples, and quadruples | 75 | 1 | + |
| CCSDTQP | CC singles, doubles, triples, quadruples, and pentuples | 1 | 1 | + |
| CC($n$) | CC method including up to $n$-tuple excitations | 1 | 1 | + |
| CCSD(T) | CC singles and doubles with perturbative triples | 76 | 28,38 | − |
| CCSDT(Q) | CCSDT with perturbative quadruples | 8,77 | 7,8 | − |
| CCSDTQ(P) | CCSDTQ with perturbative pentuples | 7,8 | 7,8 | − |
| CC($n-1$)($n$) | CC($n-1$) method with perturbative $n$-tuple excitations | 7,8 | 7 | − |

| | | | | |
|---|---|---|---|---|
| CCSD[T] | CCSD with perturbative triples, "bracket T" | 78 | 7,8 | − |
| CCSDT[Q] | CCSDT with perturbative quadruples, "bracket Q" | 79 | 7,8 | − |
| CC$(n-1)[n]$ | generalization of CCSD[T] to $n$-tuple excitations | 7,8 | 7,8 | − |
| CCSD(T)$_\Lambda$ | CCSD with perturbative asymmetric (T) correction | 80,81 | 7 | − |
| CCSDT(Q)$_\Lambda$ | CCSDT with perturbative asymmetric (Q) correction | 7 | 7 | − |
| CC$(n-1)(n)_\Lambda$ | generalization of CCSD(T)$_\Lambda$ to $n$-tuple excitations | 7 | 7 | − |
| CCSDT-1a | CCSD with iterative approximate triples, ansatz 1a | 78,82 | 7 | − |
| CCSDTQ-1a | CCSDT with iterative approximate quadruples, ansatz 1a | 79 | 7 | − |
| CC$(n)$-1a | generalization of CCSDT-1a to $n$-tuple excitations | 7 | 7 | − |
| CCSDT-1b | CCSD with iterative approximate triples, ansatz 1b | 78 | 7 | − |
| CCSDTQ-1b | CCSDT with iterative approximate quadruples, ansatz 1b | 7 | 7 | − |
| CC$(n)$-1b | generalization of CCSDT-1b to $n$-tuple excitations | 7 | 7 | − |
| CC3 | CCSD with iterative approximate triples, CC3 ansatz | 83 | 7 | − |
| CC4 | CCSDT with iterative approximate quadruples, CC3 ansatz | 7 | 7 | − |
| CC$n$ | generalization of CC3 to $n$-tuple excitations | 7 | 7 | − |
| CCSDT-3 | CCSD with iterative approximate triples, ansatz 3 | 78 | 7 | − |
| CCSDTQ-3 | CCSDT with iterative approximate quadruples, ansatz 3 | 7 | 7 | − |
| CC$(n)$-3 | generalization of CCSDT-3 to $n$-tuple excitations | 7 | 7 | − |
| CISD | CI singles and doubles | 55 | 1 | + |
| CISDT | CI singles, doubles, and triples | 55 | 1 | + |
| CISDTQ | CI singles, doubles, triples, and quadruples | 55 | 1 | + |
| CI$(n)$ | CI method including up to $n$-tuple excitations | 55 | 1 | + |
| FCI | full CI | 55 | 1 | + |
| MRCCSD | multireference CC singles and doubles | 84,85 | 15 | + |
| MRCC$(n)$ | multireference CC including up to $n$-tuple excitations | 15,85 | 15 | + |
| MRCISD | multireference CI singles and doubles | 86 | 15 | + |
| MRCI$(n)$ | multireference CI including up to $n$-tuple excitations | 15,86 | 15 | + |

[a]References describing the methodological developments [b]References describing the implementation in Mrcc [c]Availability of analytic gradients

are robust but rather expensive, roughly an order of magnitude more costly than standard

SCF with DIIS as the Newton equations are solved in each iteration step with the full Hessian. A less expensive quasi-Newton scheme based on the Broyden–Fletcher–Goldfarb–Shanno (BFGS) update is also provided, but it is less robust. Relying on the quadratic SCF infrastructure and our string-based FCI code[1], an MCSCF program was also developed. Currently it supports only complete active spaces.

TABLE II. Representative timings (wall-clock times in minutes) using SCF and low-order correlation methods for the $DNA_2$ molecule with the cc-pVDZ, aug-cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets including 1332, 2224, 3016, and 5748 basis functions, respectively. For (NAF-)dRPA and MP2, the computation times required for the SCF iteration, and for the calculation and transformation of the integrals are also included.

| Method | Basis set | | | |
| --- | --- | --- | --- | --- |
| | cc-pVDZ | aug-cc-pVDZ | cc-pVTZ | cc-pVQZ |
| DF-HF | 139 | 682 | 572 | 2155 |
| PBE0 | 142 | 690 | 612 | 2174 |
| dRPA | 359 | 1207 | 3220 | 13594 |
| DF-MP2 | 173 | 867 | 913 | 4622 |
| Dual-basis DF-HF | - | 188 | 174 | 674 |
| Local DF-HF, algorithm 1 | 20 | 310 | 91 | 426 |
| Local DF-HF, algorithm 2 | 13 | 277 | 61 | 293 |
| NAF-dRPA | 54 | 416 | 1416 | 6837 |

## B.   Density functional theory

Mrcc provides a wide range of density functional approximations including both KS and post-KS approaches. KS calculations can be carried out in the RKS, UKS, and ROKS approximations in both disk-based and integral-direct mode, utilizing either conventional or DF algorithms. As DF SCF is anyway more efficient than the conventional one, and the DF-KS approach, if a pure, i.e., non-hybrid functional is used, is cubically scaling, DF is default for KS calculations. Concerning density functionals, local density approximation (LDA), generalized gradient approximation (GGA), meta-GGA (depending on either kinetic-

energy density or the Laplacian of the density), hybrid, range-separated hybrid (RSH), and van der Waals (vdW) functionals are available. MRCC has about two dozens of built-in functionals[87,88] including the most popular ones. In addition, an interface was developed to the LIBXC library of density functionals[89,90], which enables the use of several hundreds of functionals. Empirical dispersion corrections can be calculated using the DFT-D3 approach of Grimme and co-workers[91,92]. KS calculations can also make use of the quadratic SCF algorithms mentioned in Sect. II A and, for hybrid functionals, can be accelerated by local DF approaches (see Sect. II E). Analytic gradients are implemented for RKS and UKS with LDA, GGA, meta-GGA (depending only on kinetic-energy density), and vdW functionals, and with empirical dispersion corrections.

The program is also furnished with efficient implementations of DH functionals and post-KS approaches. Several frequently-used MP2-based DHs are provided, together with analytic gradients, currently only for closed shell systems. Post-KS methods include dRPA and its improved variants, such as SOSEX, RPA with exchange, and rPT2 (see Table I). A special feature of MRCC is the availability of DH approaches utilizing correlation methods other than MP2. Our dRPA-based dRPA75 approach[46] as well as its spin-component scaled (SCS)[47] and range-separated[50] variants scale as the fourth power of the system size but outperform the genuine MP2-based DHs in most cases and can successfully cope with long-term correlation effects without empirical corrections[93]. Further DFT-wave function blends based on SOSEX[49] and ADC(2)[48] are also available (see also Sect. II D for the latter), and a flexible framework is also provided to facilitate the design of new composite DFT approaches. The implementations of all DH and post-KS methods rely on the DF approximation, and the computation times of the corresponding calculations can be significantly reduced by the natural auxiliary function (NAF) technique[33] (see Sect. II E), and, for ground states, even linear scaling with the system size can be achieved with local correlation approaches (see Sect. II G).

For the DNA$_2$ system, an iteration step of a KS calculation with the Perdew–Burke–Ernzerhof (PBE) hybrid functional (PBE0)[94,95] requires 11.9, 49.3, 51.0, and 181.2 minutes with the cc-pVDZ, aug-cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets, respectively (see Table II). This means that, as expected, a KS calculation is only slightly more expensive than the corresponding HF SCF run (cf. HF results in Sect. II A). The situation for DFT analytic gradients is also similar. KS calculations with pure functionals are also relatively efficient,

but the computation of the Coulomb term is not yet fully optimized in Mrcc. The efficiency of MP2-based DH calculations is also determined by that of the corresponding MP2 computations, see Sect. II C for the relevant benchmark results. To demonstrate the performance of our implementations of dRPA and dRPA-based composite schemes, we carried out dRPA calculations for the above test cases with PBE orbitals. The relevant timings including the KS SCF run and the determination of the exact exchange energy are presented in Table II. A dRPA-based DH is about as costly as the corresponding dRPA method, while a SOSEX calculation is roughly as expensive as the underlying dRPA and an MP2 calculation (see Sect. II C for the relevant MP2 timings).

## C.   Ab initio correlation methods

A wide variety of *ab initio* correlation methods are offered by Mrcc. For a detailed list of correlation models available for ground states see Table I. The second-order approaches include MP2 and CC2 and their SCS and scaled opposite-spin (SOS) variants[33,36]. The implementation of both relies on the DF technique, and the SOS versions also utilize Laplace transform for the energy denominators, which reduces their scaling to $N^4$, where $N$ is a measure of the system size. While MP2 is a ground-state theory, CC2 can be used for both ground and excited states. However, CC2 does not offer any advantage for ground states over MP2 or other second-order methods, thus, it is only recommended for excited states (see Sect. II D for further details). The MP2 code is efficient and can use RHF and UHF orbitals, and the calculation of analytic gradients is also possible with RHF orbitals.

The highly-optimized, hand-coded CCSD(T) program of Mrcc is one of the most efficient CCSD(T) codes at the moment, if not the most efficient one[28,37,38]. It can utilize both conventional four-center ERIs and the DF technique. In the latter case, not only the four-index integral transformation is avoided, but the four-center integrals are assembled on the fly, and the calculations are almost completely free of disk I/O. Both multi-threaded and multi-node parallelism are supported, and the parallelization efficiency is excellent (see Sect. II L for the parallelization). Unless there are overriding reasons, e.g., high-accuracy is required, the usage of the DF-CCSD(T) scheme is advocated.

One of the specialties of Mrcc is the open-ended CC code which enables the treatment of arbitrary excitations in the CC calculations. In contrast to the correlation methods

listed so far, which were coded "by hand", the open-ended CC algorithms were developed by automated techniques built on strings[1]. Not only iterative CC methods including up to $n$-tuple excitations, that is, CCSD, CCSDT, CCSDTQ, CCSDTQP, ..., are available[1], but hierarchies of efficient perturbative approximations thereof, for instance, CCSD(T), CCSDT(Q), CCSDTQ(P), ..., are also provided[7,8] (see Table I for a complete list). These hierarchies facilitate the approximation of the FCI correlation energy with arbitrary accuracy and are extensively employed in high-accuracy calculations. A useful rule of thumb is that the highest excitations should be treated with a non-iterative perturbative model. For example, the best practice to evaluate the contribution of quadruple excitations is to run CCSDT(Q) instead of full CCSDTQ, whereas, for pentuples, the CCSDTQ(P)$_\Lambda$ method is the best choice[7]. Please also notice that the ROHF reference function requires special treatment with perturbative CC approaches[8]. Concerning the basis set requirement of these approaches, experience shows that the basis set convergence of these higher-order methods is about as slow as that for CCSD(T)[10]. Nevertheless, a reasonable estimate for the higher-order contributions can be gained with double- and triple-$\zeta$ basis sets. For further details on how to use these methods in high-accuracy calculations, the interested reader is referred to the literature describing the corresponding model chemistries (see, e.g., Refs. 2–6,96). We note in passing that, as a spin-off of higher-order CC methods, arbitrary CI approaches including full CI are implemented as well. We also mention that the high-order CC and CI approaches are, in principle, also available with DF, but it does practically not influence the speed of the corresponding calculations. Hence, the use of conventional four-center ERIs is suggested for such methods.

Exploiting the automated tools developed for higher excitations, a general state-selective MRCC method was also implemented[15] utilizing the ansatz originally proposed in Refs. 84 and 85. The approach, which is an active-space single-reference model rather than a true MRCC method, uses a limited set of higher excitations selected by restricting particular indices mimicking the excitation manifold covered in a MRCI calculation. As the energy is not invariant to the choice of the reference determinant, the method is only recommended when a dominant determinant can be identified in the wave function. Another limitation of the ansatz is its relatively high computation time due to the large number of terms in the working equations. Consequently, it is only applicable to benchmark calculations for small molecules. Similar to the single-reference methods, MRCI calculations are also possible as a

byproduct. Note that this is an uncontracted, i.e., not internally contracted, MRCI. Since our algorithm is a determinant-based one with limited spin adaptation[1,15], it is relatively inefficient in comparison to fully spin-adapted MRCI implementations.

For all single- and multireference CC and CI methods, any kind of orbitals, that is, RHF, UHF, ROHF, RKS, UKS, ROKS, and MCSCF, can be used. Nonetheless, the usage of HF and MCSCF orbitals is recommended for the single- and multireference approaches, respectively. Analytic gradients are implemented for arbitrary non-perturbative single- and multireference CC and CI methods with RHF, UHF, and ROHF orbitals. Analytic gradient calculations with MCSCF orbitals are possible via the interface to the CoLUMBUS package. Analytic Hessians (linear response functions) can be computed for the aforementioned categories of methods using the CFOUR interface. Analytic third derivatives (quadratic response functions) are available for arbitrary non-perturbative CC models with the latter interface (see Sect. II M for the details about the interfaces). Higher-order CC and CI, MRCC, and MRCI energy and derivative calculations are also parallelized utilizing both multi-threaded and multi-node techniques.

The performance of our MP2 code is again illustrated for $DNA_2$, and the corresponding timings can be found in Table II. The efficiency of the DF-CCSD(T) implementation is demonstrated for the glycine molecule. With the cc-pVDZ, cc-pVTZ, cc-pVQZ, and cc-pV5Z basis set of 95, 220, 425, and 730 basis functions the CCSD(T) runs took 0.2, 4.7, 83.3, and 954.6 minutes, respectively, including the integral transformations. Further extensive benchmarks can be found in Ref. 38. To gain some insight into the computational requirements for higher-order CC methods, we carried out calculations for water. As it is evident from Table III, the perturbative models are efficient approximations to the parent iterative methods. The evaluation of analytic gradients for the iterative approaches is roughly twice as costly as the calculation of energies. For the parallelization efficiency of CC methods see Sect. II L.

## D. Electronically excited states

Except for the perturbative CC methods, the excited-state analogue of practically all approaches available for ground states can be found in MRCC. For a detailed list of excited-state methods see Table IV.

TABLE III. Representative timings (wall-clock times in minutes) using high-order CC methods for the water molecule with the cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets including 24, 58, and 115 basis functions, respectively. The timings for the LR methods include the solution of the ground-state CC equations and the LR-CC equations for one singlet excited state.

| Method | Basis set | | |
|---|---|---|---|
| | cc-pVDZ | cc-pVTZ | cc-pVQZ |
| CCSDT | 0.1 | 1.2 | 12.1 |
| CCSDT(Q) | 0.1 | 2.7 | 62.7 |
| CCSDTQ | 0.5 | 31.0 | 843.6 |
| CCSDTQ(P)$_\Lambda$ | 2.5 | 204.0 | - |
| CCSDTQP | 3.0 | 751.0 | - |
| CCSDTQ, MP2 NOs | 0.1 | 0.8 | 1.5 |
| LR-CCSDT | 0.1 | 2.3 | 29.8 |
| LR-CCSDTQ | 1.6 | 92.6 | 2632.7 |

The low-order wave function methods, such as CIS and MCSCF are normally not used alone for excited-state calculations but are utilized in more advanced correlated methods for generating the initial guess or the MOs. Instead, the reasonably accurate TDDFT approaches are recommended. The latter are implemented in the well established adiabatic approximation[97] and currently available for LDA and GGA functionals and their hybrids and RSHs. The TDDFT equations can also be solved with the Tamm–Dancoff approximation, which roughly halves the computation time at the expense of moderate loss of accuracy[98]. The implementation of CIS, TDDFT, and related methods exploits DF, and the computation times can be reduced employing local DF tricks[34] (Sect. II F).

The second-order correlation methods include CIS(D), CIS(D$_\infty$), and ADC(2), which are the excited-state analogues of MP2 in some sense, and LR-CC2, which is the excited-state extension of CC2 via linear response (LR) theory[35,36]. The corresponding SCS and SOS variants are also available[45]. The implementation of the second-order approaches makes use of the DF technique, and both disk-based and integral-direct algorithms can be executed. The SOS variants also utilize Laplace transform for the energy denominators of double excitation amplitudes and scale as $N^4$, while the scaling of the other models is $N^5$. On top

TABLE IV: Methods implemented in MRCC for excited electronic states.

| Method | | References | | Grad.[c] |
| Abbreviation | Full name/description | Meth.[a] | Imp.[b] | |
|---|---|---|---|---|
| CIS | configuration interaction singles | 100 | 34 | + |
| MCSCF | multiconfigurational self-consistent field | 54 | 57 | − |
| TDHF | time-dependent Hartree–Fock | 101 | 34 | − |
| TDDFT | time-dependent DFT | 97 | 34 | − |
| TDA-TDDFT | Tamm–Dancoff approximation TDDFT | 98 | 34 | − |
| DH-TDDFT | double hybrid TDDFT | 99 | 48 | − |
| CIS(D) | CIS with perturbative doubles | 102 | 36 | − |
| SCS-CIS(D) | spin-component scaled CIS(D) | 103,104 | 36,48 | − |
| SOS-CIS(D) | scaled opposite-spin CIS(D) | 104 | 36,48 | − |
| CIS($D_\infty$) | CIS with iterative approximate doubles | 105 | 36 | − |
| SCS-CIS($D_\infty$) | spin-component scaled CIS($D_\infty$) | 70 | 36,48 | − |
| SOS-CIS($D_\infty$) | scaled opposite-spin CIS($D_\infty$) | 71 | 36,48 | − |
| ADC(2) | second-order algebraic-diagrammatic construction | 106 | 35 | − |
| SCS-ADC(2) | spin-component scaled ADC(2) | 70 | 35,48 | − |
| SOS-ADC(2) | scaled opposite-spin ADC(2) | 71 | 35,48 | − |
| LR-CC2 | linear response CC2 | 69 | 36 | − |
| LR-SCS-CC2 | linear response SCS-CC2 | 70 | 36,48 | − |
| LR-SOS-CC2 | linear response SOS-CC2 | 71 | 36,48 | − |
| LR-CCSD | linear response CCSD | 73 | 1,27 | + |
| LR-CCSDT | linear response CCSDT | 74 | 1,27 | + |
| LR-CCSDTQ | linear response CCSDTQ | 75 | 1,27 | + |
| LR-CCSDTQP | linear response CCSDTQP | 1 | 1,27 | + |
| LR-CC($n$) | linear response CC($n$) | 1 | 1,27 | + |
| CISD | CI singles and doubles | 55 | 1 | + |
| CISDT | CI singles, doubles, and triples | 55 | 1 | + |
| CISDTQ | CI singles, doubles, triples, and quadruples | 55 | 1 | + |
| CI($n$) | CI method including up to $n$-tuple excitations | 55 | 1 | + |

| FCI | full CI | | 55 | 1 | + |
|---|---|---|---|---|---|
| LR-MRCCSD | linear response MRCCSD | | 84,85 | 15,27 | + |
| LR-MRCC($n$) | linear response MRCC($n$) | | 15,85 | 15,27 | + |
| MRCISD | multireference CI singles and doubles | | 86 | 15 | + |
| MRCI($n$) | multireference CI including up to $n$-tuple excitations | | 15,86 | 15 | + |

[a]References describing the methodological developments [b]References describing the implementation in Mrcc [c]Availability of analytic gradients

of the second-order wave function methods, several time-dependent DH approaches were also implemented. Besides the previous CIS(D)-based DHs[99], a special feature of Mrcc is the availability of ADC(2)-based TDDFT-wave function composites.

Regarding the performance of the second-order methods, one should keep in mind that the iterative methods, for valence states, yield excitation energies that are generally more accurate than those obtained with the significantly more expensive LR-CCSD, while, for Rydberg-type excitations, the error is considerably larger[107,108]. However, this problem can be remedied to a large extent by spin-scaling[109]. Taking into account that, in contrast to the other methods, ADC(2) is Hermitian, and consequently the calculation of ADC(2) spectral intensities is considerably cheaper, ADC(2) and especially its spin-scaled variants are recommended for spectrum calculations. The performance of the perturbative CIS(D) model is somewhat weaker for excitation energies compared to the iterative approaches, and the CIS(D) transition moments are only evaluated at the CIS level. The above features are also inherited by the excited-state DH methods derived from CIS(D) and ADC(2), and we recommend our recent SOS-ADC(2)-based composite model, which outperforms both the second-order wave function and the TDDFT approaches[48]. We also note that the second-order methods and the corresponding DHs can be accelerated by NAF, natural orbital (NO) (Sect. II F), and local correlation techniques (Sect. II G).

Our automated programming tools were also utilized to extend the arbitrary-order single- and multireference CC approaches to excited states invoking LR-CC theory[110], which is equivalent to the equation-of-motion (EOM) CC approach[111] for excitation energies[27]. The higher-order LR-CC methods are well suited for the high-accuracy calculation of spectral properties of small molecules and also for benchmarking lower-level approaches. We note that, at the moment, LR-CCSD calculations can only be carried out by the string-based

algorithms. Thus, the LR-CCSD implementation is not as efficient as our ground-state CCSD code or other hand-coded LR-CCSD/EOM-CCSD programs[111]. For the LR-MRCC and for the CI methods the same remarks apply as for the parent ground-state models (see Sect. II C).

Ground to excited-state transition moments can be evaluated with CIS, TD-HF, TDA, TDDFT, CIS(D), ADC(2), the spin-scaled versions of the latter two methods, and arbitrary single- and multireference LR-CC and CI methods. For the latter, excited to excited-state transition moments can also be computed. The presently available spectral intensities include oscillator and rotator strengths both in the length and the velocity gauge. Analytic gradients are implemented for arbitrary single- and multireference LR-CC and CI methods.

The performance of CIS, TDDFT, DH-TDDFT, and second-order wave function methods are well documented in our recent publications[34–36,48]. For example, a TDDFT calculation for the three lowest singlet states of the 38-atom lauric acid molecule using the aug-cc-pVTZ basis (1196 AOs) and integral-direct algorithm takes 44 minutes. For 14 excited states of the azobenzene molecule of 24 atoms the ADC(2) calculation with the aug-cc-pVTZ basis set (874 basis functions) requires 363 minutes[35]. The spectrum calculations with the DH-DFT approaches are only slightly more expensive than those with the underlying second-order correlation methods[48].

To illustrate the computational requirements for LR-CC calculations, we performed LR-CCSD, LR-CCSDT, and LR-CCSDTQ calculations for a singlet excited state of water. The LR-CCSD excitation energy calculations required 0.02, 0.4, 4.7, and 55.8 minutes including the time for the solution of the CCSD equations with the cc-pVDZ, cc-pVTZ, cc-pVQZ, and cc-pV5Z bases (24, 58, 115, and 201 functions), respectively. The runtimes for LR-CCSDT and LR-CCSDTQ are collected in Table III. The evaluation of analytic gradients or transition moments necessitates the solution of an additional, somewhat more expensive system of equations for the ground state and two other ones for the excited state. As a consequence, the corresponding computation times for gradients and transition probabilities would be roughly 3- to 4-times longer.

## E.  Reduced-cost techniques for ground states

Over the years, significant effort has been devoted to speed up the various electronic structure calculations relying on dual basis set approximations, local DF, orbital transformation techniques, compression of the auxiliary basis, and local correlation approaches. The methods for which reduced-cost algorithms are provided are collected in Table V.

A simple cost-reducing technique for the ground-state calculations is the dual basis set approach, which approximates the SCF energy by expanding the energy in Taylor series and retaining only the zeroth and first order terms[52]. The algorithm implemented in MRCC, which is closely related to the method of Head-Gordon *et al.*[112,113], starts with the solution of the SCF equations in a small basis set, then the resulting density matrix is projected into a large basis set, purified, and finally, a single Fock matrix is constructed and diagonalized in the large basis to evaluate the approximate SCF energy. If the two basis sets are chosen in a way that the small basis is a subset of the large one constructed by dropping the functions of highest angular momentum of the large basis, the procedure results in very accurate energies at the cost of about two SCF iterations with the large basis. The errors are still tolerable if the truncation scheme is more aggressive. Besides the reduced computational time, the approach also diminishes SCF convergence issues related to the linear dependencies of large basis sets. As a demonstrative example, let us consider the DF-HF calculations of Sect. II A for $DNA_2$ with the dual basis set approach (see Table II). If the large (small) basis set is chosen as aug-cc-pVDZ (cc-pVDZ), cc-pVTZ (cc-pVDZ), and cc-pVQZ (cc-pVTZ), the speedup factors with respect to the calculations carried out solely with the large basis set are 3.6, 3.0, and 3.2, respectively.

The local DF approximations utilize that spatially localized electron density distributions can be expanded in a relatively small fitting basis instead of the entire auxiliary basis of the system[40,114–116]. If applied to the scaling reduction of the exchange term in HF- or KS-SCF calculations, the MOs are localized in each iteration step, and a domain of auxiliary functions is constructed for each localized MO. The corresponding integrals of the MOs are evaluated with this limited fitting basis thereby reducing the scaling of the SCF to cubic (algorithm 1). Alternatively, MO-specific AO domains can also be used at the evaluation of the exchange term, which further decreases the scaling to asymptotically linear (algorithm 2). Note that the computational costs of the Coulomb term are not affected in either

TABLE V. Reduced-cost and reduced-scaling algorithms implemented in Mrcc.

| Cost-reduction technique | Methods[a] | References |
|---|---|---|
| Dual basis set | RHF, UHF, ROHF, RKS, UKS, ROKS | 52 |
| Local density fitting | RHF, UHF, ROHF, RKS, UKS, ROKS | 28,40 |
| | CIS, TDHF, TDDFT, TDA-TDDFT | 34 |
| Natural auxiliary functions | MP2, dRPA, SOSEX, CCSD(T) | 33 |
| | LR-CC2 | 36 |
| | CIS, CIS(D), ADC(2), CIS(D$_\infty$) | 35 |
| | CIS(D)- and ADC(2)-based DH-TDDFT | 48 |
| Natural orbitals | CCSD, CCSD(T), CCSDT, CCSDT(Q), ... | 39 |
| | LR-CC2 | 36 |
| | ADC(2), CIS(D$_\infty$) | 35 |
| Combined natural orbitals– | LR-CC2 | 36 |
| natural auxiliary functions | ADC(2), CIS(D$_\infty$) | 35 |
| Local correlation | MP2, DH-DFT | 40 |
| | dRPA, SOSEX | 31 |
| | CCSD, CCSD(T), CCSDT, CCSDT(Q), ... | 28,37,41–43 |
| | ADC(2), CIS(D$_\infty$), LR-CC2 | 45 |

[a]The cost-reduction techniques are also available for the spin-scaled versions of MP2, CIS(D), ADC(2), CIS(D$_\infty$), and CC2.

cases, thus, the overall scaling of the SCF is cubic, or at best asymptotically quadratic. Nevertheless, the local fitting approach drastically reduces the costs of SCF calculations. The effect of algorithm 1 is sensible even for relatively small systems of a couple of dozens of atoms, while algorithm 2 is recommended for considerably larger molecules: beyond at least 100 atoms if basis sets excluding diffuse functions are used, and for very large molecules of

several hundreds of atoms if diffuse AOs are also added to the basis set.

To illustrate the efficacy of the local DF, we also performed the DF-HF calculations reported in Sect. II A for $DNA_2$ with the local fitting algorithms (see Table II). Using algorithm 1, speedup factors of about 5-7 (2.2) can be achieved with the cc-pV$XZ$ (aug-cc-pVDZ) basis sets with respect to conventional DF-HF. The corresponding speedups with algorithm 2 are 7-11 (2.5). The local DF algorithms also allow for SCF calculations for very large systems. Our cutting-edge HF calculations for the lipid transfer protein (1023 atoms) with the def2-QZVP basis set (44712 AOs)[43] and for the HIV-1 integrase enzyme (2380 atoms) with the def2-TZVP basis (44035 AOs)[42] are, to the best of our knowledge, the largest HF calculations ever carried out on a single CPU.

The computational expenses of high-order correlation methods, such as CCSD(T), CCSDT, CCSDT(Q), CCSDTQ, ..., can be systematically reduced by orbital transformation techniques, where the MOs are subjected to an appropriate transformation, and the resulting basis is truncated. In MRCC, the frozen NO[117] and the optimized virtual orbital (OVO)[118,119] techniques are implemented[39]. In the former case, the one-particle MP2 density matrix is constructed and diagonalized, and the NOs, i.e., the eigenvectors of the matrix, are selected considering the corresponding occupation numbers, i.e., the eigenvalues. In the OVO approach, an overlap functional for lower-level wave functions expanded in the full and the reduced virtual spaces is defined and maximized. There is no remarkable difference between the performance of the MP2 NO and OVO approaches, but because of its simplicity, the former can be recommended. Another argument in favor of MP2 NOs is that the convergence of the energy with the truncation threshold is more or less systematic, therefore the energies can be well extrapolated. Using the MP2 NO approach, speedups of up to several orders of magnitude can be achieved depending on the basis set. For instance, the CCSDTQ calculations for water with the cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets presented in Sect. II C can be performed 5-, 36-, and 556-times faster, respectively (see Table III). Thanks to its efficiency, the MP2 NO technique helped us to estimate the correlation contribution of quadruple excitations for as large molecules as naphthalene and anthracene[39]. Additional benchmarks for the accuracy and efficiency of the scheme are available in Ref. 39.

The natural auxiliary function approach is a useful tool for the cost reduction of DF methods[33]. Here, virtually a singular value decomposition of the three-center integral list

required for the approximation of the four-center ERIs is performed. The important singular vectors are kept and interpreted as the elements of a transformed auxiliary basis. As the NAF technique reduces the size of the auxiliary basis, those methods benefit strongly that scale nonlinearly with the number of fitting functions. Among the methods implemented in MRCC, dRPA and the related models fall in this category, which are quadratic in the auxiliary basis size. For example, the dRPA calculations for $DNA_2$ using the NAF approximation (see Table II) are 6.6-, 2.9-, 2.3-, and 2.0-times faster with the cc-pVDZ, aug-cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets, respectively, with respect to the original calculations. For further benchmarks regarding the accuracy and efficiency of the NAF approximation, see Ref. 33.

## F.   Reduced-cost techniques for excited states

The local DF approximation was also extended to the CIS, TDHF, as well as hybrid TDDFT and TDA methods. A robust algorithm for the construction of local fitting domains for excited-state calculations was presented[34]. The approximation can dramatically decrease the computation time required for the HF exchange contribution, while it is practically error-free, and the resulting schemes scale as the third power of the system size. Our numerical results show that average speedup factors of 2 to 4 can be attained even for molecules of 50 to 100 atoms. For example, a single calculation for the three lowest excited states of the 98-atom D21L6 dye molecule using the aug-cc-pVDZ (aug-cc-pVTZ) basis set and the PBE0 functional takes 134 (652) minutes (8-core 1.7 GHz Intel Xeon E5-2609 v4 processor, 128 GB of main memory)[34]. For bigger systems of localized electronic structure, significantly larger speedups can be gained. We also demonstrated that the approximation enables excited-state calculations even for molecules of 1000 atoms on a single CPU with double-$\zeta$ basis sets including diffuse functions[34].

The generalization of the NAF approach to excited-state calculations is a fairly straight-forward task as it does not use any information about the excitation. Of the excited-state models, the NAF approximation is available for CIS, CIS(D), CIS($D_\infty$), ADC(2), CC2, (TDA-)TDDFT, and DH-TDDFT. On the basis of our benchmark results, with the default threshold determined for the aug-cc-pVTZ basis set, the size of the auxiliary basis can be halved regardless of the system, while the excitation energies can be recovered within 2 meV[36,48]. It means that the time required for the calculations can be cut roughly by half

since the NAF construction has only a minor overhead. For example, an ADC(2) [SOS-ADC(2)] calculation for the lowest singlet excited states of a 36-atom flavone derivative (1196 AOs) takes 92 (48) minutes, of which the NAF construction is only 2 minutes[48]. The corresponding canonical calculation lasts 188 (86) minutes.

More significant gains could be obtained by the NO approximation, however, the choice of the wave function to calculate the one-particle density matrix is not trivial. For this purpose, we proposed an excited-state specific "state-averaged" density matrix which is the average of the MP2 and CIS(D) densities[36]. With this approximation, the size of the virtual space can be halved at the expense of a mean absolute error of 0.02 eV in the calculated excitation energies for the CC2 and ADC(2) methods compared to the canonical results, while that for the oscillator strengths is 0.001 in the case of ADC(2)[35,36]. The NO approach can be effectively combined with the NAF approximation, and an average speedup of more than an order of magnitude can be achieved using the aug-cc-pVTZ basis set. The favorable speedup and memory reduction enable us to perform calculations for extended molecular systems of up to 100 atoms. For example, an ADC(2) calculation for the four lowest excited states of the D21L6 dye molecule (98 atoms and 3412 AOs) takes 116 hours[35]. It is important to emphasize that each of the elaborated approximations is robust and can be applied in a black-box manner.

## G. Local correlation methods

The methods for which local correlation approximations are available in MRCC include MP2, dRPA, the corresponding double-hybrid density functionals, and any CC methods available via either the efficient, hand-coded CCSD(T) implementation or the open-ended general-order CC programs (see Table V and Refs. 28,31,37,40–43,51). Currently only single-reference methods built on closed-shell HF or KS determinants are released, but the restricted open-shell extension of the above methods is also implemented and will be made available in the near future. The common characteristic of our local approaches is that correlation energies are obtained as the sum of orbital contributions evaluated in orbital-specific MO bases (domain approximation), which approach originates from the so-called cluster-in-molecule scheme.[120] The occupied space of the domain consists of strong pair localized MOs (LMOs) having correlation energy contributions higher than a threshold (pair

approximation), while the correlation energies of distant pairs are obtained at a multipole approximated MP2 level[40]. For perturbative approaches, such as MP2 and the (T) contribution of CCSD(T), each wave function amplitude is evaluated only once without redundancy employing effective Laplace-transform based expressions[31,37,40]. For iterative methods, such as dRPA or CCSD, the amplitudes required for the domain energy evaluation interact with all amplitudes which have occupied indices appearing in the strong pair list of the domain. This ensures the proper relaxation of the amplitudes with neighboring wave function components.[28,31,42] The more demanding dRPA, CCSD, CCSD(T), CCSDT, CCSDT(Q), ... domain calculations are accelerated using local natural orbitals (LNOs)[28,42] computed from the available domain MP2 amplitudes. This key and so far unique combination of techniques also appears in the nomenclature of the methods: LNO-dRPA, LNO-CCSD(T), etc. Furthermore, DF and the related NAF[33] technique are employed throughout to speed up the required integral transformation steps.

This strategy leads to asymptotically linear-scaling operation counts and asymptotically constant memory and disk space requirements for systems of non-metallic electronic structure. The linear-scaling region has been achieved also for realistic, three-dimensional systems, for instance, for proteins containing hundreds or even a few thousands of atoms[42,43]. Additionally, the independence of the domain computations can be exploited for embarrassingly parallel execution and also for checkpointing, thus extensive calculations are effectively restartable.[42] By identifying the symmetry equivalent domains, speedups up to the rank of the point group can be attained even for non-Abelian groups[41,42]. We have also implemented various multi-level local correlation schemes in which high-level and environmental domains are automatically identified via a high-level layer atom list specified by the user.[51,52] With the multi-level correlation technique more accurate methods, such as LNO-CCSD(T) or even LNO-CCSDT(Q) can be employed for the chemically active atoms, and cost-effective schemes, e.g., local MP2 (LMP2), can be selected for the correlation energy contribution of the environment. For more details and applications, see Sect. II H and Refs. 51 and 52. It is important to note that neither the multi-level local correlation nor any part of the above local correlation algorithms rely on empirical approaches, such as bond-cutting and capping, manual or distance-based fragment/domain determination, scaling parameters, etc. All algorithms are completely *ab initio* and black box, the orbital lists are constructed automatically, the approximations are governed via pre-determined thresholds and adopt to the

complexity of the systems' wave function.

The accuracy of the above local approximations has been extensively benchmarked for LMP2[40], LNO-dRPA[31], LNO-CCSD[43], LNO-CCSD(T)[42,43], up to LNO-CCSDT(Q)[41]. Composite threshold combinations were also developed (Loose, Normal, Tight, etc.), which enable black-box convergence tests and are useful to estimate the remaining local error with respect to the conventional reference, because of the systematic convergence of the above local approximations. For instance, when the conventional CCSD(T) reference is available, the mean absolute (maximum) errors of the default (Normal) LNO-CCSD(T) correlation energy errors are 0.02–0.07% (0.08-0.15%), while Tight settings provided maximum errors below 0.09% in every case tested so far. Our test sets included the most challenging compilation of systems used for local CCSD(T) methods so far collecting 26 molecules of 30–63 atoms and employing (augmented) triple-$\zeta$ or even quadruple-$\zeta$ AO bases.[42] Energy differences, such as reaction, conformation, and non-covalent interaction energies, benchmarked against the conventional CCSD(T) reference are also excellent. The mean absolute (maximum) errors of Normal LNO-CCSD(T) were found to be 0.1-0.3 (0.6-1.0) kcal/mol, while Tight settings provided 0.1 (0.5) kcal/mol accuracy even if LNO-CCSD(T) energies were extrapolated towards the complete basis set (CBS) limit.[43] In these tests, the accuracy of the LNO-CCSD(T) scheme was found highly competitive compared to analogous alternatives.[43] Independently, Paulechka and Kazakov benchmarked three reduced-cost CCSD(T) implementations against canonical CCSD(T) references and found a previous, much less optimized version of LNO-CCSD(T) to be the most accurate and efficient when aiming at accurate formation enthalpies.[44]

For more extended systems, when the conventional reference is not available, we recommend to perform convergence tests in order to estimate both the local and the basis set incompleteness (BSI) errors since, depending on the property of interest, absolute errors might increase extensively with the system size. A sample convergence study is presented in Fig. 1, where an intermediate step in the biosynthesis of cholesterol is considered. The reaction was identified as highly challenging for local correlation methods[121] since lanosterol (educt) and (S)-2,3-oxidosqualene (product) are markedly different, separated by many elementary steps of the net reaction; hence, one cannot rely on error compensation between the educt and product. Even for such complicated examples the Tight results and especially the CCSD(T) energy estimates based on the Normal and Tight values are converged to a
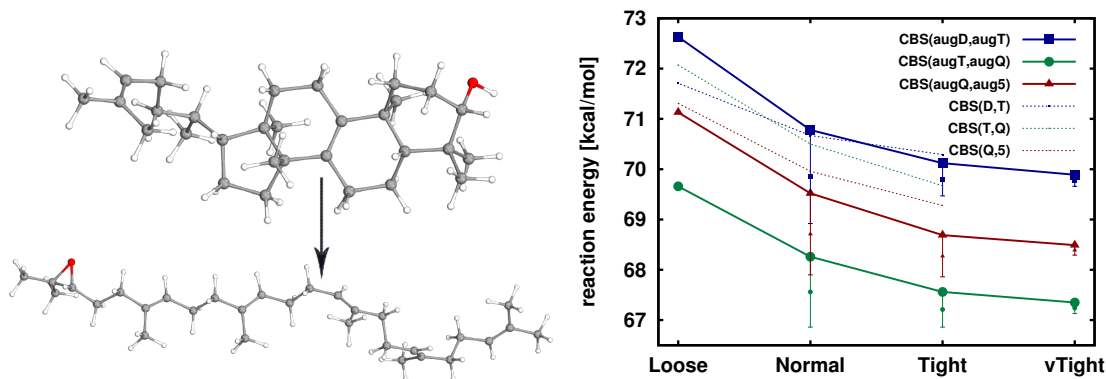
FIG. 1. Left panel: isomerization of lanosterol (educt) to (S)-2,3-oxidosqualene (product). Right panel: convergence of LNO-CCSD(T) reaction energies (kcal/mol) with the local approximation hierarchy (Loose to very Tight), and AO basis set completeness [CBS$(X,X+1)$ extrapolation[122] using (aug-)cc-pV$XZ$ basis sets with $X$=D, T, Q, and 5]. Smaller-sized symbols and the corresponding error bars represent an estimation for the converged LNO-CCSD(T) results and their estimated local errors.[43]

few tenths of a kcal/mol (Fig. 1, right panel). Extensive details and more examples on the CCSD(T) energy and local error estimates are given in Ref. 43. It is also important to highlight the relatively slow convergence of the CCSD(T) correlation energies with the AO basis set size. In general, we recommend the use of at least triple-$\zeta$ and quadruple-$\zeta$ basis sets in conjunction with basis set extrapolation approaches[122] for accurate CC energies. In order to facilitate such large-scale computations, significant efforts have been invested to eliminate all numerical issues stemming from near-linear dependencies of such basis sets when they are employed for extended molecules. This feature is currently also a unique property of the LNO-CC implementation.[43] The convergence patterns of the local correlation energies are found to be relatively independent of the chosen AO basis set as it is apparent also in Fig. 1, which can be exploited for further cost reduction using composite energy expressions. BSI corrections can be evaluated using lower level local correlation treatments or lower level correlation methods. For instance, the triple-$\zeta$ Tight LNO-CCSD(T) energy combined with a triple- and quadruple-$\zeta$ extrapolated Normal LMP2 BSI correction turned out to be a particularly efficient combination[123], especially if employed with the dual basis approach of Sect. II E.

Besides the above benchmark studies, LNO-CCSD(T) has already been employed in vari-

ous real-life applications. For example, the interaction of ATP with a hydroxyflavone derivative fluorophore[124], the many-body interactions in water clusters[125], the thermodynamic properties of menthol isomers[126], the reaction mechanisms of a Michael-addition reaction[127], and the curing of epoxy resins[128] were studied, and accurate enthalpies of formation were computed for organic compounds[44,129].

The local correlation programs were designed to run effectively on both single CPU workstations using multi-threaded parallelization and small to large computer clusters via an additional multi-node parallel layer. Integral-direct, in-core, and memory-economic algorithms were also implemented, which almost completely eliminate disk I/O operations up to the LNO-CCSD(T) level and facilitate exceptionally low per-node memory-requirements. Higher-order, e.g., LNO-CCSDT(Q) calculations still require extensive storage space and about 2-3 orders of magnitude more operations than LNO-CCSD(T), while LMP2, especially for large systems of above 100 atoms, runs comparably with or even faster than DF-HF. We performed benchmarks for a system containing 90 atoms[43], which plays a key role in an organocatalytic addition reaction[127]. Using a single, six-core CPU, LNO-CCSD(T) calculations took 1.6 (4.6) days employing the aug-cc-pVTZ (aug-cc-pVQZ) basis set and required only 7 (11) GBs of memory[43]. Since these runtimes are only about 5-6 times longer than that of the corresponding DF-HF run, LNO-CCSD(T) has become a feasible alternative of density functional approximations including exact exchange, and LNO-CCSD(T) calculations are usually not rate determining if employed in thermochemistry protocols including geometry optimizations and harmonic frequency evaluations performed with density functionals above rung three. To illustrate the capabilities of the current implementation, we note that LNO-CCSD(T)/aug-cc-pV5Z calculations are feasible for molecules of above 100 atoms, e.g., for the challenging systems appearing in the L7 test set[130]. Moreover, LNO-CCSD(T)/def2-QZVP calculations were performed studying the ligand binding energy of a lipid transfer protein[43] containing 1023 atoms and almost 45000 AOs (see Fig. 2). Consequently, especially in combination with the tools of Sect. II H, "gold standard" quality LNO-CCSD(T) energetics can be obtained also for biochemical systems.

The scope of local correlation methods was also expanded to excited states[45]. A general domain construction scheme was developed for selecting the important MOs for both the excitation and the electron correlation, and it was combined with the NO- and NAF-based cost reduction techniques outlined in Sect. II F. Our test calculations showed that the
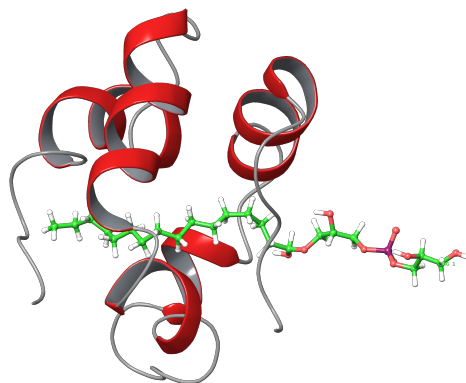
FIG. 2. Lipid transfer protein (PDB ID: 1N89)[131], including 1023 atoms, a ligand of 79 atoms, and 44712 AOs with the def2-QZVP basis set. The LNO-CCSD(T)/def2-QZVP run took 18 days for the complex.[43]

additional errors originating from the local approximations are highly acceptable (2-4 meV). Relying on this framework, local ADC(2), CIS(D$_\infty$), and CC2 methods are provided. While the scaling of the final correlated calculations is sublinear, the rate-determining step is the solution of the formally cubic-scaling CIS problem[34]. The approach is already competitive for relatively small systems constituted of fewer than 100 atoms but also enables us to carry out calculations for systems of up to 400 atoms using triple-$\zeta$ basis sets with diffuse functions. For instance, a local ADC(2) calculation for the four lowest excited states of the D21L6 dye molecule using the aug-cc-pVTZ basis set (cf. Sect. II F) required only 43 hours[45]. One of the largest systems considered so far is the 311-atom WW-6 solar-cell dye (aug-cc-pVTZ basis, 10604 AOs). For the lowest excited state of this molecule, the calculation took 210 hours, and the CIS part required 60% of the total wall-clock time[45].

## H.  Multilevel methods

If one is interested in only a small part of the whole system, but the environment effects are not negligible, further cost reduction can be achieved by utilizing multilevel approximations. In Mrcc, the simple ONIOM scheme of Morokuma and co-workers[132], the projector-based embedding (PbE) scheme of Manby and Miller[133], and its Huzinaga-equation based extension[51], which is numerically very close to the original PbE, are available. Furthermore, the mechanical and electronic embedding schemes of QM/MM can be invoked via an interface to the Amber MD program[53]. Analytic gradients are provided for the ONIOM and

the QM/MM approaches, while QM/MM MD simulations can be carried out with Amber.

The ONIOM method in Mrcc includes only the IMOMO (integrated MO+MO) variant, however, all the above discussed DFT and wave function-based approaches (and their gradients, if available) can be utilized. A special feature of the implementation is that the program copes with an arbitrary number of layers in a black-box manner as the bonds across subsystem borders and the necessary capping atoms (in the case of covalently separated subsystems) are automatically identified based on the localized MOs from the low-level calculation for the real system.

The PbE approaches, which are free from capping atoms and utilize a frozen embedding potential for the determination of the density in the region of interest, can only be utilized for single point energy calculations with the implemented ground-state methods. Typically a GGA functional is recommended as a low-level method and a density functional of a higher rung or a wave function-based approach should be chosen as a high-level technique to maximize the accuracy/cost ratio. One should keep in mind that, in these calculations, the evaluation of the exact exchange or the correlation energy within the embedded subsystem are the most time consuming steps. The latter can be especially demanding if one embeds a method that scales with a higher power of the number of virtual orbitals, for example, CCSD(T), as the system is partitioned only in the occupied MO space, but the size of the virtual MO manifold is not reduced. The costs of the SCF calculations in the embedded subsystem can be substantially decreased via the local density fitting algorithms (see Sect. II E). In addition, a special feature of our PbE implementation, the use of local correlation approaches (see Sect. II G) in the high level region, also bypasses the second problem by truncating the virtual space. The computational costs of the PbE calculations can be further alleviated by the dual basis set embedding approach[52], where the PbE protocol is performed with a small basis set and then a large basis set is applied to calculate a first-order correction for each of the energy terms. The use of mixed basis sets as large basis, where a bigger basis is only employed on the atoms of the embedded subsystem, further accelerates the dual basis set calculations and also efficiently eliminates the spurious densities, which commonly occur in simple calculations carried out with mixed basis sets.

In the spirit of the focused models, all the above mentioned multilevel approximations can be combined with MM to extend the IMOMO scheme into the full ONIOM approach or to form DFT-in-DFT-in-MM and wave function-in-DFT-in-MM embeddings. Further-

more, the previously mentioned multilevel local correlation technique can also be combined with MM, ONIOM, or PbE. To illustrate the power of the multilevel methods, we mention our QM/MM calculations presented in Ref. 52 performed with a sufficiently large QM subsystem (571 atoms) on the biologically important methylation reaction catalyzed by the catechol-*O*-methyltransferase enzyme (3420 atoms). The LNO-CCSD(T)-in-PBE-D3-in-MM and LNO-CCSD(T)-in-LMP2-in-MM calculations required 2.83 and 3.81 days with the cc-pVTZ basis set (12594 basis functions) and the local fitting procedure (algorithm 1) for the SCF, which mean 2.23 and 1.65 speedups, respectively, with respect to the full LNO-CCSD(T)-in-MM/cc-pVTZ calculation, which took 6.3 days. If an intermediate LMP2 layer is inserted between the CC and DFT layers, i.e., the multilevel approaches are combined as LNO-CCSD(T)-in-LMP2-in-PBE-D3-in-MM, the computational time (speedup) is 2.41 days (2.61). The simulation time can be further reduced with the dual basis approach to 1.21 days. The computational complexity of these calculations is rather high, however, the required efforts for the input preparation are not higher than those for the typical setup of a QM/MM calculation: basically an atom list has to be specified for each separate layer.

## I.  Geometry optimizations and property calculations

Optimization of molecular geometries and evaluation of first-order properties are possible with any method implemented in Mrcc for which analytic gradients are available (see Tables I and IV). Presently, only unrestricted geometry optimizations can be performed in redundant internal or Cartesian coordinates using the BFGS algorithm. The implemented first-order properties include dipole, quadrupole, and octapole moments as well as the electric field at the atomic centers. Harmonic vibrational frequencies and infrared intensities can be computed with the aforementioned methods via numerically differentiated analytic gradients. At the frequency calculations ideal gas thermodynamic properties are also evaluated invoking the rigid rotor-harmonic oscillator approximation. These features can be recommended for property calculations for bigger molecules using lower-level methods for which efficient gradients are implemented, such as DF-HF, DFT, MP2, and DH-DFT. They are also useful for calculating accurate properties for relatively small systems with high-level CC approaches.

MRCC is also equipped with various wave function analysis tools. In addition to the standard Mulliken and Löwdin atomic charges[134], partial charges can also be computed with the promising intrinsic atomic orbital (IAO) approach[135]. Bond orders can be estimated according to the definition of Mayer[136], and the atomic decomposition of the SCF and correlation energies can also be determined. The localization of MOs is possible with the Boys[137], Pipek–Mezey[138], Cholesky decomposition[139], and intrinsic bond orbital (IBO)[135] algorithms.

A wide variety of further properties can be computed with arbitrary CC and CI methods via the interface to the CFOUR quantum chemistry suite (see also Sect. II M). Geometry optimizations can be performed, and practically all the first-order properties that are implemented in CFOUR, such as multipole moments, electric field gradients, etc., can be computed[19]. Similar holds for the second-order properties accessible in CFOUR as well. Besides harmonic vibrational frequencies[20], NMR chemical shifts[20], static and frequency-dependent electric dipole polarizabilities[23], magnetizabilities and rotational $g$-tensors[22], electronic $g$-tensors[24], spin-spin coupling constants, and spin rotation constants have been tested so far. The analytic Hessian code was also modified to enable the evaluation of DBOCs[25]. Among higher-order properties, static and frequency-dependent electric-dipole first (general, second-harmonic-generation, optical-rectification) hyperpolarizabilities[21] and Raman intensities[26] are available. Anharmonic force fields and the related spectroscopic properties can also be computed via numerical differentiation of analytic gradients or Hessians. These facilities are practical for obtaining benchmark-quality properties for smaller molecules.

We note that property calculations are also possible with several types of relativistic Hamiltonians, see Sect. II J for more details.


## J.  Relativistic calculations

Currently, MRCC, in standalone mode, supports only the standard nonrelativistic Hamiltonian, and the consideration of relativity is only possible via the use of effective core potentials (ECPs). However, the MOLPRO, CFOUR, and DIRAC interfaces enable various levels of relativistic treatments for general (LR-)CC and CI models (see also Sect. II M)[14,140]. While, with MOLPRO, only relativistic single-point energy calculations can be performed with the Douglas–Kroll–Hess (DKH) Hamiltonians, several relativistic Hamiltonians are at

our disposal, and also property calculations are possible with the latter two interfaces.

With CFOUR, exact two-component (X2C) and spin-free Dirac–Coulomb (DC) calculations can be carried out. The evaluation of mass-velocity and Darwin corrections is also possible using the density matrices computed with the nonrelativistic Hamiltonian for the iterative (LR-)CC and CI methods listed in Tables I and IV. With the latter methods, analytic gradients can also be computed, and geometry optimizations and first-order property calculations can be performed with the X2C treatment.

The general (LR-)CC and CI codes of MRCC were also adapted for four-component calculations using Kramers-restricted Dirac–Fock orbitals and Abelian double-group symmetry[14]. The necessary MO integrals can be taken over from the DIRAC suite via an interface which enables calculations with the full DC and several approximate relativistic Hamiltonians. Single-point energy, unrelaxed first-order property, and transition moment calculations are available with all (LR-)CC and CI methods. Numerous properties implemented in DIRAC can be evaluated, including rather exotic ones.

While the correlation calculations with the DKH, X2C, or spin-free DC Hamiltonians are about as expensive as the corresponding nonrelativistic calculations, the full four-component treatment is significantly more demanding. For example, the CCSDTQ calculations of Sect. II C for water using the DC Hamiltonian require 41.0 and 344.0 minutes with the cc-pVDZ and cc-pVTZ basis sets, respectively. That is, the four-component relativistic calculations are 77- and 11-times slower, respectively. Unless the considered phenomenon definitely necessitates the full four-component treatment, such calculations can only be recommended for systems containing heavy elements or if very high accuracy is desired. Otherwise the more economical relativistic approaches are suggested.


## K.    Basis sets and basis set optimization

MRCC uses Gaussians as atomic orbital basis functions. A special feature is the handy infrastructure for bond functions, and an experimental elliptical Gaussian code is also available[141]. By default, most of the popular basis set families are provided; for instance, Dunning's (augmented) correlation consistent (core-)valence sets [(aug-)cc-pV$XZ$ and (aug-)cc-pCV$XZ$, $X$ = D, T, Q, 5, 6][142,143], the Gaussian basis sets of Pople and co-workers (6-31G\*, 6-311G\*, 6-31G\*\*, 6-31++G\*\*, ...)[144,145], the def2 Gaussian basis sets of Weigend

and Ahlrichs (def2-SVP, def2-TZVP, def2-TZVPP, ...)[146], and the F12 (cc-pV$X$Z-F12, $X =$ D, T, Q)[147] and the correlation consistent pseudo-potential (cc-pV$X$Z-PP and aug-cc-pV$X$Z-PP, $X =$ D, T, Q, 5)[148,149] basis sets of Peterson and co-workers. For calculations employing the DF approximation, the auxiliary bases for the Dunning- and Karlsruhe-type AO bases are available for both Coulomb/exchange[150] and correlation[151,152] fitting. For heavier elements various widespread ECPs are implemented including the LANL2DZ ECPs of Hay and Wadt[153], the Stuttgart–Köln ECPs for the def2 basis sets[154], and the Stuttgart–Köln multiconfiguration Dirac–Hartree–Fock-adjusted ECPs[155]. Further AO basis sets, auxiliary bases, and ECPs can be obtained from the Basis Set Exchange[156–158].

A flexible framework is also provided for the optimization of basis sets including not only conventional Gaussian bases but also bond functions[141]. The optimization can be performed with any method for which single-point energy calculations are available (see Table I). Both unconstrained and constrained optimizations can be carried out. In the latter case, particular exponents and coefficients or all parameters for a given angular momentum shell are kept fixed during the optimization. A convenient feature is that several basis sets can also be optimized at a time, e.g., the basis sets of two or more atoms in a molecule. In this way the complete variational optimization of all the wave function parameters of a molecular system can also be performed.

As MRCC is designed for accurate electronic structure calculations, we recommend the usage of the correlation consistent and Karlsruhe-type basis sets. The latter sets are more economical and most adequate for DFT or low-order correlation calculations, but the former bases should be used in high-order CC calculations. The optimization of a new basis set is an option when, e.g., the basis set of the selected family does not exist for an element. At the optimization, the protocols developed for the given basis set type should be followed.


## L.   Parallelization

All components of MRCC support shared-memory parallelism. For larger arrays, the linear algebra is performed by multi-threaded BLAS (Basic Linear Algebra Subprograms) and LAPACK (Linear Algebra Package) routines, and the expensive ones of the remaining loops are also parallelized using the OpenMP (Open Multi-Processing) technology. The parallelization efficiencies for SCF and low-level correlation methods are reasonable. For

example, the speedup factors for the $DNA_2$ DF-HF/cc-pVTZ calculation of Sect. II A are 1.7, 3.1, and 3.8 on 2, 4, and 6 cores, respectively, and similar tendencies can be observed for DFT and low-level approaches. The SCF algorithms, the optimized CCSD(T) program, and the open-ended CC code also benefit from multi-node parallelization via replicated-memory techniques. This approach enables the use of modern computer clusters consisting of multiple compute nodes for SCF and CC calculations while also ensuring the best node-level performance.
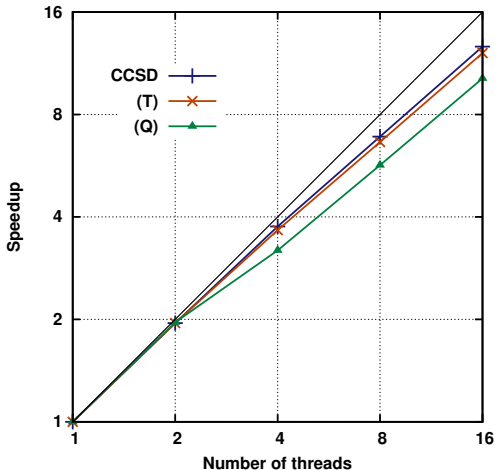


FIG. 3. Multi-threaded scaling of MRCC. The CCSD calculations were carried out on a uracil trimer with the aug-cc-pVDZ basis set, the (T) correction was calculated for an $(H_2O)_{10}$ cluster using the cc-pVDZ basis set, the (Q) results were obtained for a water molecule with the cc-pVTZ basis set.

The parallel scaling of the hand-coded CCSD(T) program[38] is illustrated on a uracil trimer system using the aug-cc-pVDZ basis set for the CCSD calculation and an $(H_2O)_{10}$ cluster using the cc-pVDZ basis set for the (T) correction. The parallel performance of the general CC program is demonstrated on a water molecule for the CCSDT(Q) method using the cc-pVTZ basis set. The speedups of the CCSD calculation and the (T) and (Q) corrections were measured on nodes containing two 8-core Intel Xeon E5-2650 v2 CPUs and 100 GB of memory. The multi-threaded and multi-node speedups, compared to the one-thread and the one-node measurements, are depicted in Figs. 3 and 4, respectively. The CCSDT and other iterative high-order CC calculations do not scale well with either the number of threads or the number of nodes because of the large number of memory operations, the sequential disk
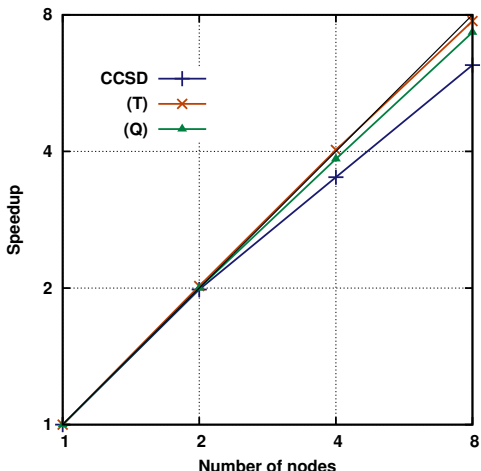
33

FIG. 4. Multi-node scaling of MRCC. See the caption of Fig. 3 for details of the calculations.

I/O, and the more extensive network communication. However, the CCSD code and the most expensive perturbative corrections [(T) and (Q)] exhibit excellent strong scaling both in multi-threaded calculations and in multi-node runs up to at least 8 compute nodes. The perturbative corrections show an almost linear scaling with the number of nodes. The multi-threaded speedups of the CCSD, the (T), and the (Q) calculations are between 10 and 13 corresponding to a $65-80\%$ parallel efficiency, which is probably close to the limit achievable with this particular hardware. The speedup and the corresponding parallel efficiency of the the CCSD code on 8 nodes are about 6.2 and 75%. More performance benchmarks are provided for the hand-coded DF-CCSD(T) implementation in Ref. 38.

The excellent scaling of MRCC made it possible to determine the CCSD(T) correlation energies of some medium sized molecules (31-43 atoms)[38] with triple- and quadruple-$\zeta$ quality basis sets. These include one of the largest CCSD(T) calculations (1569 AOs) ever published, which took about 68 hours on 224 cores of 8 CPUs. These examples illustrate that CCSD(T)/CBS calculations can now be performed relatively routinely up to about 30 atoms using state-of-the-art implementations and a handful of modern, many-core CPUs.

## M.   Interfaces

While MRCC is basically a standalone code, it was interfaced to other quantum chemistry, MD, and molecular visualization software. These interfaces considerably broaden the capabilities of our package and also facilitate the use of its special features by the users of other

programs. The interface[53] with the AMBER[159,160] MD program, which enables QM/MM optimizations and MD simulations, was presented in Sect. II H and is not detailed here. Solvation effects can be modeled by the polarizable continuum model (PCM)[161] via an interface to the PCMSOLVER library[162–164]. For the sake of completeness we also mention the interfaces to the LIBXC library[89,90] and the DFTD3 program of Grimme and co-workers[91,92], which are introduced in Sect. II B and not discussed here.

The interfaces to quantum chemistry suites CFOUR[165], COLUMBUS[166], DIRAC[167], MOLPRO[168], ORCA[169], and PSI4[170] enable single-point energy calculations with the general (LR-)CC and CI codes of MRCC using the transformed MO integrals computed by these programs. In addition, property calculations are also possible with CFOUR, COLUMBUS, and DIRAC. The CFOUR, MOLPRO, ORCA, and PSI4 interfaces are rather automated, and MRCC can simply be executed by setting the corresponding keywords in the input files of these packages. The use of the other interfaces needs more extensive user intervention. The most versatile interface is the one with CFOUR, which allows for the computation of numerous properties (Sect. III) and the usage of relativistic Hamiltonians (Sect. II J). With COLUMBUS, energy and first-order property calculations, geometry optimizations, and harmonic frequency computations can be performed with HF and MCSCF reference functions. The detailed description of the interface with the relativistic DIRAC code can be found in Sect. II J.

For the pre- and post processing of molecular structures and wave functions, an interface was also developed with the MOLDEN program[171]. It can be used for drawing molecular geometries, tracking geometry optimizations, and visualizing MOs, densities, and molecular vibrations. The interface file printed in Molden format can also be processed by several other molecular visualization packages.

## N. Technicalities

MRCC is distributed via the website www.mrcc.hu and is released at least once a year. The website provides detailed manuals, and a forum is also established where the users can discuss questions related to the package. For academic researchers, MRCC is available free of charge.

The suite is supplied as precompiled binaries, which are ready for use. Optionally, one can also install the program from the source code, which is also accessible at our website. For

a basic installation, a Fortran 90 and a C compiler as well as working BLAS and LAPACK libraries are necessary. The best performance can be expected using Intel's compilers and Math Kernel Library (MKL). The additional programs or libraries, such as Libxc, should be installed separately.

The usage of Mrcc is very simple: just a single input file and the execution of a driver program are required. We also endeavored to keep the structure of the input as simple and flexible as possible. Besides the molecular structure, only the type of the calculation and the AO basis set are needed to be specified in a minimal input. All other parameters of the calculation, such as convergence thresholds, auxiliary basis sets, integration grids, or truncation thresholds for low-cost methods, are automatically selected employing sensible defaults. The molecular geometry can be specified in various formats, which are compatible to those employed by other quantum chemistry codes. The output provides detailed information for the user to be able to track the calculation, but the final results are always highlighted at the end.

## III. OUTLOOK

In the near future, the program development will still be influenced by our research interests, especially by our intention to develop accurate but low-cost quantum chemical methods. Along this line, we envisage the further development of low-order scaling SCF algorithms, distributed-memory parallel schemes for local correlation methods, analytic gradients and properties for local correlation approaches, improved embedding models, analytic gradients and properties for embedding methods, explicitly correlated local correlation methods, new composite DFT-wave function models, and analytic gradients for reduced-cost ground- and excited-state methods. Furthermore, we also plan to improve the infrastructure of our program and implement further standard quantum chemical models. In particular, we intend to implement improved geometry optimization and transition-state search algorithms; analytic gradients for TDDFT, ADC(2), and related methods; further molecular properties; and further solvation models; and we also plan to supplement our manual with a detailed tutorial. These features will be available in later versions of Mrcc.

## ACKNOWLEDGMENTS

## REFERENCES

[1] M. Kállay and P. R. Surján, J. Chem. Phys. **115**, 2945 (2001).

[2] 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 (2004).

[3] A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay and J. Gauss, J. Chem. Phys. **120**, 4129 (2004).

[4] A. Karton, E. Rabinovich, J. M. L. Martin and B. Ruscic, J. Chem. Phys. **125**, 144108 (2006).

[5] D. Feller, K. A. Peterson and D. A. Dixon, J. Chem. Phys. **129**, 204105 (2008).

[6] D. Bakowies, J. Chem. Phys. **130**(14), 144113 (2009).

[7] M. Kállay and J. Gauss, J. Chem. Phys. **123**, 214105 (2005).

[8] M. Kállay and J. Gauss, J. Chem. Phys. **129**, 144101 (2008).

[9] Y. J. Bomble, J. Vazquéz, M. Kállay, C. Michauk, P. G. Szalay, A. G. Császár, J. Gauss and J. F. Stanton, J. Chem. Phys. **125**, 064108 (2006).

[10] A. Karton, P. R. Taylor and J. M. L. Martin, J. Chem. Phys. **127**, 064104 (2007).

[11] A. Karton, I. Kaminker and J. M. L. Martin, J. Phys. Chem. A. **113**(26), 7610 (2009).

[12] A. Karton, Chem. Phys. Lett. **737**, 136810 (2019).

[13] M. E. Harding, J. Vázquez, J. Gauss, J. F. Stanton and M. Kállay, J. Chem. Phys. **135**, 044513 (2011).

[14] H. S. Nataraj, M. Kállay and L. Visscher, J. Chem. Phys. **133**, 234109 (2010).

[15] M. Kállay, P. G. Szalay and P. R. Surján, J. Chem. Phys. **117**, 980 (2002).

[16] S. Das, D. Mukherjee and M. Kállay, J. Chem. Phys. **132**, 074103 (2010).

[17] S. Das, M. Kállay and D. Mukherjee, J. Chem. Phys. **133**, 234110 (2010).

[18] S. Das, M. Kállay and D. Mukherjee, Chem. Phys. **392**, 83 (2012).

[19] M. Kállay, J. Gauss and P. G. Szalay, J. Chem. Phys. **119**, 2991 (2003).

[20] M. Kállay and J. Gauss, J. Chem. Phys. **120**, 6841 (2004).

[21] D. P. O'Neill, M. Kállay and J. Gauss, J. Chem. Phys. **127**, 134109 (2007).

[22] J. Gauss, K. Ruud and M. Kállay, J. Chem. Phys. **127**, 074101 (2007).

[23] M. Kállay and J. Gauss, J. Mol. Struct. (THEOCHEM). **768**, 71 (2006).

[24] J. Gauss, M. Kállay and F. Neese, J. Phys. Chem. A. **113**, 11541 (2009).

[25] J. Gauss, A. Tajti, M. Kállay, J. F. Stanton and P. G. Szalay, J. Chem. Phys. **125**, 144111 (2006).

[26] D. P. O'Neill, M. Kállay and J. Gauss, Mol. Phys. **105**, 2447 (2007).

[27] M. Kállay and J. Gauss, J. Chem. Phys. **121**, 9257 (2004).

[28] Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki and M. Kállay, J. Chem. Phys. **139**, 094105 (2013).

[29] G. Samu and M. Kállay, J. Chem. Phys. **146**, 204101 (2017).

[30] G. Samu and M. Kállay, J. Chem. Phys. **149**, 124101 (2018).

[31] M. Kállay, J. Chem. Phys. **142**, 204105 (2015).

[32] P. D. Mezei, G. I. Csonka and M. Kállay, J. Chem. Theory Comput. **14**, 2469 (2018).

[33] M. Kállay, J. Chem. Phys. **141**, 244113 (2014).

[34] D. Mester and M. Kállay, J. Chem. Theory Comput. **15**, 1690 (2019).

[35] D. Mester, P. R. Nagy and M. Kállay, J. Chem. Phys. **148**, 094111 (2018).

[36] D. Mester, P. R. Nagy and M. Kállay, J. Chem. Phys. **146**, 194102 (2017).

[37] P. R. Nagy and M. Kállay, J. Chem. Phys. **146**, 214106 (2017).

[38] L. Gyevi-Nagy, M. Kállay and P. R. Nagy, J. Chem. Theory Comput. **16**, 366 (2020).

[39] Z. Rolik and M. Kállay, J. Chem. Phys. **134**, 124111 (2011).

[40] P. R. Nagy, G. Samu and M. Kállay, J. Chem. Theory Comput. **12**, 4897 (2016).

[41] Z. Rolik and M. Kállay, J. Chem. Phys. **135**, 104111 (2011).

[42] P. R. Nagy, G. Samu and M. Kállay, J. Chem. Theory Comput. **14**, 4193 (2018).

[43] P. R. Nagy and M. Kállay, J. Chem. Theory Comput. **15**, 5275 (2019).

[44] E. Paulechka and A. Kazakov, J. Chem. Theory Comput. **14**, 5920 (2018).

[45] D. Mester, P. Nagy and M. Kállay, J. Chem. Theory Comput. **15**, 6111 (2019).

[46] P. D. Mezei, G. I. Csonka, A. Ruzsinszky and M. Kállay, J. Chem. Theory Comput. **11**, 4615 (2015).

[47] P. D. Mezei, G. I. Csonka, A. Ruzsinszky and M. Kállay, J. Chem. Theory Comput. **13**, 796 (2017).

[48] D. Mester and M. Kállay, J. Chem. Theory Comput. **15**, 4440 (2019).

[49] P. D. Mezei, A. Ruzsinszky and M. Kállay, J. Chem. Theory Comput. **15**, 6607 (2019).

[50] P. D. Mezei and M. Kállay, J. Chem. Theory Comput. **15**, 6678 (2019).

[51] B. Hégely, P. R. Nagy, G. G. Ferenczy and M. Kállay, J. Chem. Phys. **145**, 064107 (2016).

[52] B. Hégely, P. R. Nagy and M. Kállay, J. Chem. Theory Comput. **14**, 4600 (2018).

[53] B. Hégely, F. Bogár, G. G. Ferenczy and M. Kállay, Theor. Chem. Acc. **134**, 132 (2015).

[54] T. Helgaker, P. Jørgensen and J. Olsen. *Molecular Electronic Structure Theory.* Wiley, Chichester, 2000.

[55] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry.* McGraw-Hill, New York, 1989.

[56] C. C. J. Roothan, Rev. Mod. Phys. **32**, 179 (1960).

[57] J. Csóka and M. Kállay. to be published, 2020.

[58] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[59] M. Filatov and S. Shaik, Chem. Phys. Lett. **288**, 689 (1998).

[60] S. Grimme, J. Chem. Phys. **124**, 034108 (2006).

[61] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[62] S. Grimme, J. Chem. Phys. **118**, 9095 (2003).

[63] Y. Jung, R. C. Lochan, A. D. Dutoi and M. Head-Gordon, J. Chem. Phys. **121**, 9793 (2004).

[64] F. Furche, Phys. Rev. B. **64**, 195120 (2001).

[65] A. Grüneis, M. Marsman, J. Harl, L. Schimka and G. Kresse, J. Chem. Phys. **131**, 154115 (2009).

[66] X. Ren, P. Rinke, G. E. Scuseria and M. Scheffler, Phys. Rev. B. **88**, 035120 (2013).

[67] A. Szabo and N. S. Ostlund, J. Chem. Phys. **67**, 4351 (1977).

[68] A. Heßelmann, Phys. Rev. A. **85**, 012517 (2012).

[69] O. Christiansen, H. Koch and P. Jørgensen, Chem. Phys. Lett. **243**, 409 (1995).

[70] A. Hellweg, S. A. Grün and C. Hättig, Phys. Chem. Chem. Phys. **10**, 4119 (2008).

[71] N. O. C. Winter and C. Hättig, J. Chem. Phys. **134**, 184101 (2011).

[72] J. Čížek, J. Chem. Phys. **45**, 4256 (1966).

[73] G. D. Purvis III and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982).

[74] P. Piecuch, S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **110**, 6103 (1999).

[75] S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **97**, 4282 (1992).

[76] K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[77] Y. J. Bomble, J. F. Stanton, M. Kállay and J. Gauss, J. Chem. Phys. **123**, 054101 (2005).

[78] M. Urban, J. Noga, S. J. Cole and R. J. Bartlett, J. Chem. Phys. **83**, 4041 (1985).

[79] S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **108**, 9221 (1998).

[80] S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **108**, 5243 (1998).

[81] T. D. Crawford and J. F. Stanton, Int. J. Quantum Chem. **70**, 601 (1998).

[82] Y. S. Lee, S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **81**, 5906 (1984).

[83] H. Koch, O. Christiansen, P. Jørgensen, A. M. Sánchez de Merás and T. Helgaker, J. Chem. Phys. **106**, 1808 (1997).

[84] N. Oliphant and L. Adamowicz, J. Chem. Phys. **94**, 1229 (1991).

[85] P. Piecuch, N. Oliphant and L. Adamowicz, J. Chem. Phys. **99**, 1875 (1993).

[86] R. J. Buenker and S. D. Peyerimhoff, Theor. Chem. Acc. **35**, 33 (1974).

[87] Functionals were obtained from the Density Functional Repository as developed and distributed by the Quantum Chemistry Group, CCLRC Daresbury Laboratory, Daresbury, Cheshire, WA4 4AD United Kingdom. Contact Huub van Dam (h.j.j.vandam@dl.ac.uk) or Paul Sherwood for further information.

[88] R. Strange, F. R. Manby and P. J. Knowles, Comput. Phys. Commun. **136**, 310 (2001).

[89] M. A. L. Marques, M. J. T. Oliveira and T. Burnus, Comput. Phys. Commun. **183**, 2272 (2012).

[90] https://www.tddft.org/programs/libxc/.

[91] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[92]S. Grimme, S. Ehrlich and L. Goerigk, J. Comput. Chem. **32**, 1456 (2011).

[93]B. Brauer, M. K. Kesharwani, S. Kozuch and J. M. L. Martin, Phys. Chem. Chem. Phys. **18**, 20905 (2016).

[94]J. P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[95]J. P. Perdew, M. Ernzerhof and K. Burke, J. Chem. Phys. **105**, 9982 (1996).

[96]A. Ganyecz, M. Kállay and J. Csontos, J. Chem. Theory Comput. **13**, 4193 (2017).

[97]M. E. Casida, Recent advances in density functional methods, in *Computational Chemistry: Reviews of Current Trends*, edited by D. P. Chong, volume 1. World Scientific, Singapore, 1999.

[98]S. Hirata and M. Head-Gordon, Chem. Phys. Lett. **314**, 291 (1999).

[99]S. Grimme and F. Neese, J. Chem. Phys. **127**, 154116 (2007).

[100]J. B. Foresman, M. Head-Gordon, J. A. Pople and M. J. Frisch, J. Phys. Chem. **96**, 135 (1992).

[101]A. D. McLachlan and A. M. Ball, Rev. Mod. Phys. **36**, 844 (1964).

[102]M. Head-Gordon, R. J. Rico, M. Oumi and T. J. Lee, Chem. Phys. Lett. **219**, 21 (1994).

[103]S. Grimme and E. I. Izgorodina, Chem. Phys. **305**, 223 (2004).

[104]Y. M. Rhee and M. Head-Gordon, J. Phys. Chem. A. **111**, 5314 (2007).

[105]M. Head-Gordon, M. Oumi and D. Maurice, Mol. Phys. **96**, 593 (1999).

[106]J. Schirmer, Phys. Rev. A. **26**, 2395 (1982).

[107]D. Kánnár, A. Tajti and P. G. Szalay, J. Chem. Theory Comput. **13**, 202 (2017).

[108]D. Kánnár and P. G. Szalay, J. Chem. Theory Comput. **10**, 3757 (2014).

[109]A. Tajti and P. G. Szalay, J. Chem. Theory Comput. **15**, 5523 (2019).

[110]H. Koch and P. Jørgensen, J. Chem. Phys. **93**, 3333 (1990).

[111]J. F. Stanton and R. J. Bartlett, J. Chem. Phys. **98**, 7029 (1993).

[112]W. Liang and M. Head-Gordon, J. Phys. Chem. A. **108**, 3206 (2004).

[113]R. P. Steele, R. A. DiStasio Jr. and M. Head-Gordon, J. Chem. Theory Comput. **5**, 1560 (2009).

[114]R. Polly, H.-J. Werner, F. R. Manby and P. J. Knowles, Mol. Phys. **102**, 2311 (2004).

[115]C. Köppl and H.-J. Werner, J. Chem. Theory Comput. **12**, 3122 (2016).

[116]D. Mejía-Rodríguez and A. M. Köster, J. Chem. Phys. **141**, 124114 (2014).

[117]T. R. Barr and E. R. Davidson, Phys. Rev. A. **1**, 644 (1970).

[118]L. Adamowicz and R. J. Bartlett, J. Chem. Phys. **86**, 6314 (1987).

[119] P. Neogrády, M. Pitoňák and M. Urban, Mol. Phys. **103**, 2141 (2005).

[120] W. Li, Z. Ni and S. Li, Mol. Phys. **114**, 1447 (2016).

[121] M. Schwilk, D. Usvyat and H.-J. Werner, J. Chem. Phys. **142**, 121102 (2015).

[122] T. Helgaker, W. Klopper, H. Koch and J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[123] P. R. Nagy and M. Kállay, J. Chem. Theory Comput. **16**, in preparation (2019).

[124] M. Bojtár, P. Z. Janzsó-Berend, D. Mester, D. Hessz, M. Kállay, M. Kubinyi and I. Bitter, Beilstein J. Org. Chem. **14**, 747 (2018).

[125] I. Bakó, I. Mayer, A. Hamza and L. Pusztai, J. Mol. Liq. **285**, 171 (2019).

[126] V. Štejfa, A. Bazyleva, M. Fulem, J. Rohlíček, E. Skořepová, K. Růžička and A. V. Blokhin, J. Chem. Thermodyn. **131**, 524 (2019).

[127] T. Földes, Á. Madarász, Á. Révész, Z. Dobi, S. Varga, A. Hamza, P. R. Nagy, P. M. Pihko and I. Pápai, J. Am. Chem. Soc. **139**, 17052 (2017).

[128] L. Kárpáti, Á. Ganyecz, T. Nagy, G. Hamar, E. Banka, M. Kállay and V. Vargha, Polym. Bull. pp. doi:10.1007/s00289–019–02987–6 (2019).

[129] E. Paulechka and A. Kazakov, J. Chem. Eng. Data. **64**, 4863 (2019).

[130] R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay and P. Hobza, J. Chem. Theory Comput. **9**, 3364 (2013).

[131] J.-L. Pons, F. de Lamotte, M.-F. Gautier and M.-A. Delsuc, J. Biol. Chem. **278**, 14249 (2003).

[132] F. Maseras and K. Morokuma, J. Comput. Chem. **16**, 1170 (1995).

[133] F. R. Manby, M. Stella, J. D. Goodpaster and T. F. Miller III, J. Chem. Theory Comput. **8**, 2564 (2012).

[134] R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).

[135] G. Knizia, J. Chem. Theory Comput. **9**, 4834 (2013).

[136] I. Mayer, Chem. Phys. Lett. **97**, 270 (1983).

[137] J. M. Foster and S. F. Boys, Rev. Mod. Phys. **32**, 300 (1960).

[138] J. Pipek and P. Mezey, J. Chem. Phys. **90**, 4916 (1989).

[139] F. Aquilante, T. B. Pedersen, A. M. Sánchez de Merás and H. Koch, J. Chem. Phys. **125**, 174101 (2006).

[140] M. Kállay, H. S. Nataraj, B. K. Sahoo, B. P. Das and L. Visscher, Phys. Rev. A. **83**, 030503(R) (2011).

[141] D. Mester, J. Csontos and M. Kállay, Theor. Chem. Acc. **134**, 74 (2015).

[142]T. H. Dunning Jr., J. Chem. Phys. **90**, 1007 (1989).

[143]D. E. Woon and T. H. Dunning Jr., J. Chem. Phys. **103**, 4572 (1995).

[144]P. C. Hariharan and J. A. Pople, Theor. Chem. Acc. **28**, 213 (1973).

[145]R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).

[146]F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[147]K. A. Peterson, T. B. Adler and H.-J. Werner, J. Chem. Phys. **128**, 084102 (2008).

[148]K. A. Peterson, J. Chem. Phys. **119**, 11099 (2003).

[149]K. A. Peterson, D. Figgen, E. Goll, H. Stoll and M. Dolg, J. Chem. Phys. **119**, 11113 (2003).

[150]F. Weigend, J. Comput. Chem. **29**, 167 (2008).

[151]F. Weigend, M. Häser, H. Patzelt and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).

[152]F. Weigend, A. Köhn and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).

[153]P. J. Hay and W. R. Wadt, J. Chem. Phys. **82**, 270 (1985).

[154]D. Andrae, U. Häußermann, M. Dolg, H. Stoll and H. Preuß, Theor. Chem. Acc. **77**, 123 (1990).

[155]B. Metz, M. Schweizer, H. Stoll, M. Dolg and W. Liu, Theor. Chem. Acc. **104**, 22 (2000).

[156]Basis Set Exchange, https://www.basissetexchange.org/.

[157]D. J. Feller, J. Comput. Chem. **17**, 1571 (1996).

[158]B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson and T. L. Windus, J. Chem. Inf. Model. **59**, 4814 (2019).

[159]D. A. Case, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Götz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, and P. A. Kollman. AMBER 2017. University of California, San Francisco, 2017.

[160]R. Salomon-Ferrer, D. A. Case and R. C. Walker, Wiley Interdiscip. Rev. Comput. Mol. Sci. **3**, 198 (2013).

[161]J. Tomasi, B. Mennucci and R. Cammi, Chem. Rev. **105**, 2999 (2005).

[162]R. Di Remigio, K. Mozgawa, H. Cao, V. Weijo and L. Frediani, J. Chem. Phys. **144**, 124103 (2016).

[163] R. Di Remigio, A. H. Steindal, K. Mozgawa, V. Weijo, H. Cao and L. Frediani, Int. J. Quantum Chem. **119**, e25685 (2019).

[164] PCMSolver, an open-source library for the polarizable continuum model electrostatic problem, written by R. Di Remigio, L. Frediani and contributors (see http://pcmsolver.readthedocs.io/).

[165] Cfour, a quantum chemical program package written by J. F. Stanton, J. Gauss, M. E. Harding, P. G. Szalay with contributions from A. A. Auer, R. J. Bartlett, U. Benedikt, C. Berger, D. E. Bernholdt, Y. J. Bomble, O. Christiansen, M. Heckert, O. Heun, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, K. Klein, W. J. Lauderdale, D. A. Matthews, T. Metzroth, D. P. O'Neill, D. R. Price, E. Prochnow, K. Ruud, F. Schiffmann, S. Stopkowicz, J. Vázquez, F. Wang, J. D. Watts and the integral packages MOLECULE (J. Almlöf and P. R. Taylor), PROPS (P. R. Taylor), ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de.

[166] H. Lischka, R. Shepard, I. Shavitt, F. B. Brown, R. M. Pitzer, R. Ahlrichs, H.-J. Böhm, A. H. H. Chang, D. C. Comeau, R. Gdanitz, H. Dachsel, M. Dallos, C. Erhard, M. Ernzerhof, G. Gawboy, P. Höchtl, S. Irle, G. Kedziora, T. Kovar, T. Müller, V. Parasuk, M. Pepper, P. Scharf, H. Schiffer, M. Schindler, M. Schüler, E. Stahlberg, P. G. Szalay, and J.-G. Zhao. Columbus, an ab initio electronic structure program, release 5.9, 2001.

[167] Dirac08, a relativistic ab initio electronic structure program, development version (2010), written by L. Visscher, H. J. Aa. Jensen, and T. Saue, with new contributions from R. Bast, S. Dubillard, K. G. Dyall, U. Ekström, E. Eliav, T. Fleig, A. S. P. Gomes, T. U. Helgaker, J. Henriksson, M. Iliaš, Ch. R. Jacob, S. Knecht, P. Norman, J. Olsen, M. Pernpointner, K. Ruud, P. Sałek, and J. Sikkema (see http://dirac.chem.sdu.dk).

[168] H.-J. Wernera, P. J. Knowles, G. Knizia, F. R. Manby and M. Schütz, Wiley Interdiscip. Rev. Comput. Mol. Sci. **2**, 242 (2012).

[169] F. Neese, Wiley Interdiscip. Rev. Comput. Mol. Sci. **8**, e1327 (2018).

[170] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince III, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer III, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, J. Chem. Theory Comput. **13**, 3185

(2017).

G. Schaftenaar and J. H. Noordik, J. Comput.-Aided Mol. Design. **14**, 123 (2000).