

# **A state-specific multireference coupled-cluster method based on the bivariational principle**

Tilman Bodenstien<sup>1</sup> and Simen Kvaal<sup>1, a)</sup>

*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry,  
University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo,  
Norway*

(Dated: 2 December 2020)

A state-specific multireference coupled-cluster method based on Arponen’s bivariational principle is presented, the bivar-MRCC method. The method is based on single-reference theory, and therefore has a relatively straightforward formulation and modest computational complexity. The main difference from established methods is the bivariational formulation, in which independent parameterizations of the wavefunction (ket) and its complex conjugate (bra) are made. Importantly, this allows manifest multiplicative separability of the state (exact in the extended bivar-MRECC version of the method, and approximate otherwise), and additive separability of the energy, while preserving polynomial scaling of the working equations. A feature of the bivariational principle is that the formal bra and ket references can be included as bivariational parameters, which eliminates much of the bias towards the formal reference. A pilot implementation is described, and extensive benchmark calculations on several standard problems are performed. The results from the bivar-MRCC method are comparable to established state-specific multireference methods. Considering the relative affordability of the bivar-MRCC method, it may become a practical tool for non-experts.

---

<sup>a)</sup>Electronic mail: [simen.kvaal@kjemi.uio.no](mailto:simen.kvaal@kjemi.uio.no).

# I. INTRODUCTION

In this article, we demonstrate how Arponen’s bivariational principle<sup>1,2</sup> (BIVP) can be employed to derive a state-specific multireference coupled-cluster (MRCC) method for electronic-structure theory, avoiding many of the problems associated with the currently established state-specific methods, such as sufficiency conditions, non-commuting cluster operators, and so on. The present *proof-of-concept* method is based on single-reference theory, and uses a complete-active space (CAS) approach, but avoids, at least in principle, a bias towards an arbitrary formal reference via an optional bivariational optimization. Thus, the method is not a “genuine” multireference method, but should be nearly free of the problems associated with reference bias. We name the method *the bivariational (state-specific) multireference coupled-cluster method* (bivar-MRCC). When reference optimization is included, we name it the *orbital-adaptive* bivariational multireference coupled-cluster (bivar-OAMRCC) method. In the same manner as standard single-reference coupled-cluster theory can be viewed as an approximation to Arponen’s extended coupled-cluster (ECC) method, we also obtain an extended version, bivar-(OA)MRECC. The bivariational approach allows a manifestly multiplicatively separable state parameterization, automatically providing size-consistency of the energy and computed properties, including size-intensivity of excitation energies, whilst being of relative simplicity. Moreover, the bivariational MRCC ansatz should be amenable to relatively straightforward mathematical analysis, e.g., *a priori* error analysis, such as done previously for the ECC method.<sup>3</sup> Hence, this approach has the potential of being a powerful quantum chemical tool usable for the non-expert.

Arponen’s bivariational approach is top-down, starting with potentially *different* but *exact* parameterizations for both a bra and a ket vector  $\langle \tilde{\Psi} |$  and  $|\Psi\rangle$ . The exact Schrödinger equation is then obtained by requiring the bivariate Rayleigh quotient (or expectation value functional)  $\langle \tilde{\Psi} | H | \Psi \rangle / \langle \tilde{\Psi} | \Psi \rangle$  to be stationary. Approximations are in turn obtained by truncating the state parameters, i.e., a *nonlinear Galerkin approach* in the language of numerical analysis.<sup>4</sup> Mathematical statements of the convergence of the computed results can be made from this top-down approach using basic results from non-linear functional analysis.<sup>3,4</sup> We stress that, while there are “two wavefunctions” in bivariational approaches, they form a *unique state* approximation  $\rho = |\Psi\rangle \langle \tilde{\Psi} | / \langle \tilde{\Psi} | \Psi \rangle$ . Since this state is obtained variationally, expectation values are obtained in a straightforward manner using the Hellmann–Feynman

theorem.<sup>5</sup> Equations for excited states and response theory are also readily formulated.

In his original analysis of conventional single-reference CC theory (and the introduction of the ECC method),<sup>1,2</sup> Arponen used the bivariational approach to write down an expectation value functional which is today known as the *coupled-cluster Lagrangian*<sup>6–8</sup> in the quantum chemistry community. The latter term originates from the work of Helgaker and Jørgensen who independently discovered this expression.<sup>6,7</sup> Compared to Arponen’s formalism, the conventional derivation of the CC Lagrangian can be claimed to be bottom-up: Starting from the projected similarity-transformed Schrödinger equation, one realizes that its approximate fulfillment via projection is a constrained optimization of the CC energy, and that the corresponding Lagrangian can be conveniently written as an expectation value using an auxiliary bra vector involving the Lagrange multipliers. Thus, in a sense, the bivariational point of view is now *standard* in quantum chemistry, but its power is not fully utilized: the conventional view is very “ket centric”, while the bivariational top-down approach places equal importance to the bra and ket, the left and right Schrödinger equations, and hence all state parameters. Indeed, for general bivariational methods, the standard notion of a “projection manifold” is not meaningful, since the stationary conditions do not decouple bra and ket Schrödinger equations. Finally, let us remark, that all current MRCC theories are similarly focused on the ket side, being based on projections of a similarity transformed Schrödinger equation (or the Bloch equation for state-universal theories). A complete overview of existing state-specific MRCC approaches is beyond the scope of this work. We direct the reader towards the excellent reviews by Köhn *et al.*,<sup>9</sup> Lyakh *et al.*,<sup>10</sup> as well as the perspective article by Evangelista.<sup>11</sup>

The bivar-MRCC method resembles the complete-active space coupled-cluster (CASCC) method pioneered by Oliphant, Adamowicz, and Piecuch.<sup>12–14</sup> Indeed, the ket ansatz is identical. However, whereas CASCC is based on the projection of the corresponding ket Schrödinger equation, we instead provide an exact bra parameterization and appeal to the BIVP. Moreover, the BIVP allows optimization of the formal reference by means of non-orthogonal orbital rotations akin to the non-orthogonal orbital-optimized CC (NOCC) method developed by Pedersen and coworkers.<sup>15</sup> For systems with multireference character, this may lead to significantly more compact wave function representations, in particular of single-reference type.<sup>16,17</sup> Care is taken so that both the bra and the ket vectors are separable, and a balanced treatment of the model space (i.e., the CAS) is obtained for the bra

and the ket.

We present first numerical benchmark calculations for the bivar-MRCC and bivar-MRECC methods, performed with a full-configuration interaction (FCI) based pilot implementation. As a multireference method should be reasonably accurate for both single- and multireference problems, we opted for an example which incorporates both, namely the insertion of Be into  $\text{H}_2$ , a standard example for testing novel multireference coupled-cluster methods, since it is computationally feasible even for complicated methods.<sup>10,18,19</sup> We also performed numerical experiments on the dissociation of the HF,  $\text{H}_8$ , and  $\text{N}_2$  molecules. Whenever possible, we compare our results with MRCC calculations presented in the literature.

The article is organized as follows: In Section II we discuss the BIVP and its Galerkin approximation. We outline how a bivariational method can be analyzed mathematically in order to obtain *a priori* error estimates for the Galerkin approximations. In Section III we introduce the bivar-MRCC method, including the bivariational optimization of the formal reference. We discuss size-consistency in the sense of separability of the energy and the state in Section IV, and prove that bivar-MR(E)CC is separable. In Section V we discuss our implementation of the bivar-MRCC method, before we present some numerical results in Section VI. Finally, in Section VII we present our conclusions and future perspectives.

## II. THE BIVARIATIONAL PRINCIPLE

A complete mathematical exposition of the present material, including the exact infinite-dimensional case, is out of scope for the present article, and will be presented elsewhere. The current treatment is compatible with a finite-dimensional Hilbert space  $\mathcal{H}$ , or alternatively a bounded Hamiltonian  $H$ . Virtually all Hamiltonians of interest in molecular quantum mechanics are unbounded as they contain a kinetic energy term. On the other hand, whenever one thinks of a finite-dimensional FCI model as “exact”, the present setting is sufficient.

### A. Bivariate Rayleigh quotient

The BIVP, introduced by Arponen in his seminal coupled-cluster treatise,<sup>1,2</sup> and also studied by Löwdin around the same time,<sup>20</sup> is a generalization of the Rayleigh–Ritz vari-

ational principle to Hamiltonians  $H$  that are not necessarily self-adjoint, even if the most important application is to these Hamiltonians. The approach introduces, in addition to the usual ket vector  $|\Psi\rangle$ , the dual vector  $\langle\tilde{\Psi}|$  as a truly independent variable, since relaxing the requirement  $H = H^\dagger$  makes the left and right eigenvectors independent. The starting point is then the bivariate Rayleigh quotient

$$\mathcal{E}(\tilde{\Psi}, \Psi) = \frac{\langle\tilde{\Psi}|H|\Psi\rangle}{\langle\tilde{\Psi}|\Psi\rangle}, \quad (1)$$

which is stationary if and only if

$$H|\Psi\rangle = E|\Psi\rangle, \quad \langle\tilde{\Psi}|H = E\langle\tilde{\Psi}|, \quad \langle\tilde{\Psi}|\Psi\rangle \neq 0, \quad (2)$$

where  $E = \mathcal{E}(\tilde{\Psi}, \Psi)$ . This is the bivariational principle. The basic idea is to introduce potentially *different* approximations to  $\langle\tilde{\Psi}|$  and  $|\Psi\rangle$ , a flexibility which turns out to be very useful. However, as the bivariate Rayleigh quotient is not below bounded, unlike the usual variational Rayleigh quotient for a self-adjoint  $H$ , one cannot insert any trial bra-ket pair into  $\mathcal{E}$  and hope for a meaningful result. In Sec. II B we outline how bivariational approximations can be made.

A potentially confusing aspect of the BIVP is the fact that we now have “two wavefunctions”. However, the *state* formed is unique, i.e., it is a non-Hermitian rank-one density operator  $\rho = |\Psi\rangle\langle\tilde{\Psi}| / \langle\tilde{\Psi}|\Psi\rangle$ . Since  $\rho$  is determined variationally, the Hellmann–Feynman approach<sup>5</sup> can be used to define expectation values of arbitrary observables  $A$ , i.e.,

$$\langle A \rangle \equiv \text{Tr } \rho A = \frac{\langle\tilde{\Psi}|A|\Psi\rangle}{\langle\tilde{\Psi}|\Psi\rangle}. \quad (3)$$

By introducing the *time-dependent* BIVP,<sup>2,21</sup> we can take the leap to the time domain. The bra and ket time-dependent Schrödinger equations are obtained as stationary points of the action-like integral

$$\mathcal{S} = \int_0^T \langle\tilde{\Psi}(t)|(i\partial_t - H)|\Psi(t)\rangle dt, \quad (4)$$

where we have assumed, for simplicity, the normalization  $\langle\tilde{\Psi}(t)|\Psi(t)\rangle = 1$ . This opens up the route to not only response theory<sup>22</sup> and the approximation of excited states,<sup>2,23</sup> but also real-time propagation of quantum systems far from the ground-state.<sup>24–27</sup>

## B. Parameterization maps and discretization

As mentioned, the bivariate Rayleigh quotient is not below bounded. In fact, even the ground state is a saddle point. In lieu of the below boundedness, a parameterization of the pair  $(\tilde{\Psi}, \Psi)$  must be made that can get arbitrarily close to the exact ground state.

Suppose that we are given a parameterization map  $\chi : \tilde{V} \oplus V \rightarrow \tilde{\mathcal{H}} \oplus \mathcal{H}$ , where  $V$  is some Hilbert space, and where  $\tilde{\mathcal{H}}$  (resp.  $\tilde{V}$ ) is the dual space of  $\mathcal{H}$  (resp.  $V$ ), i.e., the space of bra vectors. The map  $\chi$  is assumed to be smooth and with a smooth inverse near a ground-state pair  $(\tilde{\Psi}_*, \Psi_*)$ , i.e., the parameterization is *exact* near the ground state. The map induces an energy expectation value functional  $\mathcal{E}_\chi : \tilde{V} \oplus V \rightarrow \mathbb{C}$ , with  $\mathcal{E}_\chi(\tilde{v}, v) = \mathcal{E}(\tilde{\Psi}(\tilde{v}, v), \Psi(\tilde{v}, v))$ , which is smooth, and whose critical points are in one-to-one correspondence with the solutions of Eq. (2) that can be reached with  $\chi$ . In particular, the ground state is parameterized by a point  $(\tilde{v}_*, v_*) \in \tilde{V} \oplus V$ . It follows that the Schrödinger equation and its dual can be written: Find  $(\tilde{v}_*, v_*) \in \tilde{V} \oplus V$  such that

$$\frac{\partial}{\partial \tilde{v}} \mathcal{E}_\chi(\tilde{v}_*, v_*) = 0, \quad \frac{\partial}{\partial v} \mathcal{E}_\chi(\tilde{v}_*, v_*) = 0. \quad (5)$$

A bivariational approximation is now obtained by a Galerkin approach defined by projection in the space  $V$ , i.e., we restrict Eq. (5) to the space  $\tilde{V}_d \oplus V_d$ , where the subscript  $d$  is a discretization parameter: Find  $(\tilde{v}_{d*}, v_{d*}) \in \tilde{V}_d \oplus V_d$  such that

$$P_{V_d} \frac{\partial}{\partial \tilde{v}_d} \mathcal{E}_\chi(\tilde{v}_{d*}, v_{d*}) = 0, \quad P_{\tilde{V}_d} \frac{\partial}{\partial v_d} \mathcal{E}_\chi(\tilde{v}_{d*}, v_{d*}) = 0. \quad (6)$$

Here,  $P_{V_d}$  is the projector onto  $V_d$ . In any Galerkin approach, it is assumed that any  $v \in V$  can be approximated sufficiently well by the projections  $v_d = P_{V_d} v \in V_d$ , i.e.,  $\|v_d - v\| \rightarrow 0$  as  $d \rightarrow \infty$ , and similarly for the dual element. The example to keep in mind is that of  $V$  being the space of cluster amplitudes (or operators), and  $V_d$  being a a single (S), double (D), etc., truncation. The limit  $d \rightarrow \infty$  corresponds to the untruncated limit (and also the basis set limit in the infinite dimensional case). Typically,  $V$  consists of excitation amplitude vectors, and  $\tilde{V}$  consists of de-excitation amplitude vectors.

## C. Local strong monotonicity analysis

Several questions arise: First, does the discrete bivariational Schrödinger equation (6) have a solution? Is this solution unique? As  $d \rightarrow \infty$ , do the solution  $(\tilde{v}_{d*}, v_{d*})$  and the

corresponding energy  $E_d = \mathcal{E}_d(\tilde{v}_{d*}, v_{d*})$  converge to the exact solution  $(\tilde{v}_*, v_*)$  and energy  $E_* = \mathcal{E}_\chi(\tilde{v}_*, v_*)$ , respectively? For an approximation that is variational in the usual sense, these questions are easily answered due to the below boundedness of the Rayleigh quotient.<sup>28</sup> However, for a bivariational approximation an analysis must be made to obtain sufficient conditions for convergence. In this section, which can be skipped on first reading, we outline how this can be done, even if the actual error analysis for bivar-MR(E)CC is relegated to future work.

The outline is extracted from an error analysis for Arponen’s ECC method obtained by Laestadius and Kvaal.<sup>3,29</sup> Under reasonable conditions the ECC method was shown to be convergent, implying the same results for the standard coupled-cluster method. For an analysis of standard CC without the BIVP, see the works of Rohwedder and Schneider.<sup>30,31</sup>

A key analysis tool for studying CC theory<sup>3,29–33</sup> was that of local strong monotonicity<sup>4</sup> of the *flipped gradient*  $F : \tilde{V} \oplus V \rightarrow \tilde{V} \oplus V$  given by

$$F(\tilde{v}, v) = \left( \frac{\partial}{\partial v} \mathcal{E}_\chi(\tilde{v}, v), \frac{\partial}{\partial \tilde{v}} \mathcal{E}_\chi(\tilde{v}, v) \right). \quad (7)$$

Local strong monotonicity can be rephrased as the Jacobian of  $F$  being positive definite at the ground state. Flipping the gradient is motivated by the following: The simple gradient of the original bivariate Rayleigh quotient  $\mathcal{E}$  is *not* monotone, since every eigenvalue is a saddle point. However, it can be demonstrated that flipping the gradient turns the ground-state saddle point into something like a local minimum under reasonable conditions on the Hamiltonian. Thus, it makes sense to consider Eq. (7) and find conditions on  $\chi$  such that local strong monotonicity is inherited.

For such an analysis, it is easiest to work with a parameterization in which the normalizations of the bra and ket are fixed. In the setting of ECC, the bra-ket pair is normalized according to  $\langle \tilde{\Psi} | \Psi \rangle = \langle \Phi_0 | \Psi \rangle = 1$ , where  $|\Phi_0\rangle$  is the reference determinant in single-reference theory.

When  $F$  is locally strongly monotone near the ground-state, Zarantonello’s Theorem<sup>3,4,29</sup> on local form implies  $v_{d*} \rightarrow v_*$  and  $\tilde{v}_{d*} \rightarrow \tilde{v}_*$  as  $d \rightarrow \infty$ . The critical point formulation of the Schrödinger equation immediately implies a quadratic error estimate,

$$|E_* - E_d| \leq C(\|v_* - v_{d*}\|^2 + \|\tilde{v}_* - \tilde{v}_{d*}\|^2), \quad (8)$$

for some constant  $C$ , which holds for sufficiently large  $d$ , i.e., for sufficiently large Galerkin subspaces  $V_d$ . We remark, that this is a *local* result. There may be solutions of the discrete

equations that are not related to exact solutions, and the local convergence is only achieved sufficiently far into the Galerkin sequence of spaces, i.e., for large enough  $d$ .

### III. STATE-SPECIFIC MULTIREFERENCE FORMULATION

#### A. Bra and ket model spaces

As for all multireference methods, we need a specification of a *model space*  $\mathcal{H}_0$  and an *external space*  $\mathcal{H}_{\text{ext}}$ , together forming the *computational*  $N$ -electron space  $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_{\text{ext}}$ . We take a standard complete-active space (CAS) approach to generating the model space<sup>28</sup>: Given a collection of linearly independent spin-orbitals  $\{\varphi_x\}$ , a basis for the computational (FCI) space  $\mathcal{H}$  is generated by all possible  $N$ -electron Slater determinants formed by these. The model space is spanned by a subset of these determinants defined as follows: we partition  $\{\varphi_x\}$  into *inactive* spin-orbitals  $\{\varphi_i\}$ , *active* spin-orbitals  $\{\varphi_p\}$ , and *secondary* spin-orbitals  $\{\varphi_a\}$ . In the model-space determinants, elements of  $\{\varphi_i\}$  are always occupied, those of  $\{\varphi_p\}$  can be occupied and unoccupied, while elements of  $\{\varphi_a\}$  are always unoccupied, see Fig. 1. Since the external space  $\mathcal{H}_{\text{ext}}$  is the orthogonal complement of the model space, a determinantal basis is given by the model-space determinants with at least one spin-orbital replaced by an element in  $\{\varphi_i\} \cup \{\varphi_a\}$ . These spin-orbitals are therefore called *external* in this work.

As we are going to employ single-reference formalism, we partition the active spin-orbitals as  $\{\varphi_p\} = \{\varphi_i\} \cup \{\varphi_a\}$ , such that the formal reference  $|\Phi_0\rangle$  is given by populating the elements of  $\{\varphi_i\} \cup \{\varphi_a\}$ , of which there are  $N$  in total. A single-reference cluster operator then excites electrons from this set into the set  $\{\varphi_p\} \cup \{\varphi_a\}$ . An *internal* cluster operator excites only from  $\{\varphi_i\}$  to  $\{\varphi_a\}$ , while an *external* cluster operator excites at least one electron from  $\{\varphi_i\}$  or into  $\{\varphi_a\}$ , i.e., at least one label is external.

The computational space is now generated by the set of spin-orbitals  $\{\varphi_x\} = \{\varphi_i\} \cup \{\varphi_p\} \cup \{\varphi_a\} \cup \{\varphi_\alpha\}$ . We do *not* assume that this set is orthonormal. Instead, we will assume biorthogonality with the dual spin-orbitals that define the dual computational Hilbert space  $\tilde{\mathcal{H}} = \tilde{\mathcal{H}}_0 \oplus \tilde{\mathcal{H}}_{\text{ext}}$ . That is, we introduce dual spin-orbitals  $\{\tilde{\varphi}_x\} = \{\tilde{\varphi}_i\} \cup \{\tilde{\varphi}_p\} \cup \{\tilde{\varphi}_a\} \cup \{\tilde{\varphi}_\alpha\}$  that generate  $\tilde{\mathcal{H}}_0$  and  $\tilde{\mathcal{H}}_{\text{ext}}$ , in the same manner as above. The number of functions in each subset must match the number of elements in the corresponding primal subsets.



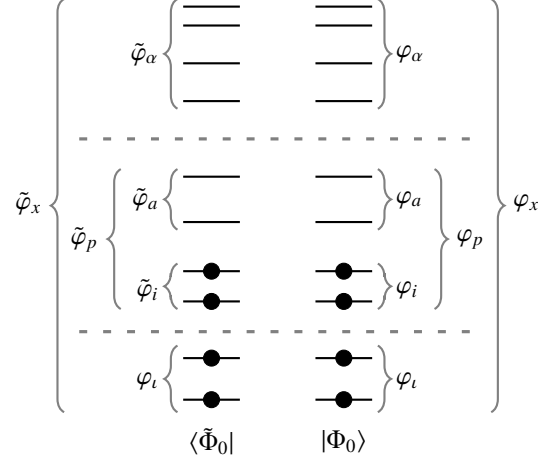


FIG. 1. Illustration of the composition of the bra and ket single-particle spaces in terms of biorthogonal spin-orbitals, including the index conventions used in this work. The dashed lines indicate the extent of the CAS. The bra and ket formal references are shown.

Biorthogonality means  $\langle \tilde{\varphi}_x | \varphi_y \rangle = \delta_{xy}$ . The dual formal reference  $\langle \tilde{\Phi}_0 |$  is defined by occupying all the elements of  $\{\tilde{\varphi}_\iota\} \cup \{\tilde{\varphi}_i\}$ , from which we obtain  $\langle \tilde{\Phi}_0 | \Phi_0 \rangle = 1$ . Indeed, we have in general  $\langle \tilde{\Phi}_\mu | \Phi_\nu \rangle = \delta_{\mu\nu}$ , where  $\mu$  is a generic numbering scheme of the Slater determinant basis.

A remark on the relation between the dual and primal spin-orbital sets is natural at this point. The space  $\tilde{\mathcal{H}}$  is the complex-conjugate space of  $\mathcal{H}$ , and therefore  $\{\tilde{\varphi}_x\}$  must span the same single-particle space as  $\{\varphi_x\}$ . One set therefore uniquely determines the other, and an orthonormal or partially orthonormal basis set is certainly one possibility. However, as the (active) spin-orbitals are to be bivariationally optimized, which may result in non-orthonormal spin-orbitals, it is convenient to keep the fully biorthogonal notation.

A general model space ket can be written

$$|\Psi_0\rangle = \sum_{\mu \in \text{CAS}} |\Phi_\mu\rangle c_\mu, \quad (9)$$

where  $\mu \in \text{CAS}$  indicates that the sum is over the CAS determinant basis. Similarly, a general model space bra can be written

$$\langle \tilde{\Psi}_0 | = \sum_{\nu \in \text{CAS}} d_\nu \langle \tilde{\Phi}_\nu |. \quad (10)$$

We note that  $\langle \tilde{\Psi}_0 | \Psi_0 \rangle = \sum_{\mu \in \text{CAS}} d_\mu c_\mu = d^T c$ , where we introduce matrix notation for the amplitudes.

## B. Bra and ket parameterizations in Hilbert space

In the internally contracted MRCC (ic-MRCC) scheme,<sup>9-11</sup> a general wavefunction  $|\Psi\rangle \in \mathcal{H}$  is written

$$|\Psi\rangle = e^{T_{\text{ic-MRCC}}} |\Psi_0\rangle, \quad (11)$$

where  $T_{\text{ic-MRCC}} = \sum_{\mu \in \text{CAS}} T^{(\mu)}$  is a general cluster operator whose components  $T^{(\mu)}$  are single-reference cluster operators relative to the model-space determinant  $|\Phi_\mu\rangle$ . We note that  $T^{(\mu)}$  is in general not unique, and that  $[T^{(\mu)}, T^{(\nu)}] \neq 0$ . These are among the basic problems of MRCC which we would like to avoid.

Suppose now that  $|\Psi\rangle \in \mathcal{H}$  has a nonzero component along the formal reference  $|\Phi_0\rangle$ . We can then use single-reference CC theory to *uniquely* write  $|\Psi\rangle = \langle \tilde{\Phi}_0 | \Psi \rangle \exp(T_{\text{SR}}) |\Phi_0\rangle$ , where  $T_{\text{SR}} = T_0 + T$  is a single-reference cluster operator. The term  $T_0$  contains precisely those excitations that stay within the model space, leaving  $T$  as the external excitations, i.e., those which have at least one external label. Exploiting this, we obtain

$$|\Psi\rangle = e^T |\Psi_0\rangle, \quad (12)$$

where  $\langle \tilde{\Phi}_0 | \Psi \rangle \exp(T_0)$  has been converted to a CAS expansion.

Similar to the above considerations, a general bra  $\langle \Omega |$  with nonzero component along  $\langle \tilde{\Phi}_0 |$  can be written  $\langle \Omega | = \langle \tilde{\Psi}_0 | e^S$ , where  $S$  is an external de-excitation operator (excitation operator for bras). In the spirit of Arponen's ECC method, we can postmultiply with the invertible operator  $e^{-T}$  to obtain the bra parameterization

$$\langle \tilde{\Psi} | = \langle \tilde{\Psi}_0 | e^S e^{-T}. \quad (13)$$

This is valid so long as  $\langle \tilde{\Psi} | e^T |\Phi_0\rangle \neq 0$ , which puts a very mild restriction on the exact ground-state bra. We note that  $\langle \tilde{\Psi} | \Psi \rangle = \langle \tilde{\Psi}_0 | \Psi_0 \rangle = d^T c$ .

This completes the specification of an exact parameterization map  $(\tilde{\Psi}, \Psi) = \chi(s, t, d, c)$ , where  $s$  and  $t$  are the amplitudes of  $S$  and  $T$ , respectively. Here, the dependence on  $\{\varphi_x\}$  and  $\{\tilde{\varphi}_x\}$  is implicit in the determinantal basis. Plugging into the bivariate Rayleigh quotient, we obtain the energy functional of bivar-MRECC,

$$\begin{aligned} \mathcal{E}_{\text{bivar-MRECC}}(s, t, d, c) \\ = \langle \tilde{\Phi}_0 | DC | \Phi_0 \rangle^{-1} \langle \tilde{\Phi}_0 | D e^S e^{-T} H e^T C | \Phi_0 \rangle, \end{aligned} \quad (14)$$

where we introduced the CAS cluster operators  $C$  and (de-excitation) operator  $D$ . We remark, that in Arponen's ECC, a further coordinate change  $(t, s) = (t(s, s'), s)$  is made, where  $s'$  are the amplitudes of a cluster operator  $S'$  defined by  $\langle \Phi_\mu | S' | \Phi_0 \rangle = \langle \Phi_\mu | e^{S'} T | \Phi_0 \rangle$ . This is done in order to ensure a certain linkedness structure of the diagram series, and also has the implication that the *time-dependent* Schrödinger equation takes the form of a canonical Hamiltonian system.<sup>2,29,34</sup> We will not further explore the ECC flavor of the multireference ansatz here, but instead reparameterize  $\Lambda = e^S - 1$  to obtain the energy functional of the bivar-MRCC method,

$$\begin{aligned} \mathcal{E}_{\text{bivar-MRCC}}(\lambda, t, d, c) &= \langle \tilde{\Phi}_0 | DC | \Phi_0 \rangle^{-1} \langle \tilde{\Phi}_0 | D(1 + \Lambda) e^{-T} H e^T C | \Phi_0 \rangle \\ &= (d^T c)^{-1} d^T K(t, \lambda) c, \end{aligned} \quad (15)$$

where  $K(t, \lambda) = [K(t, \lambda)_{\mu\nu}] = [\langle \tilde{\Phi}_\mu | (1 + \Lambda) e^{-T} H e^T | \Phi_\nu \rangle]$  can be considered an effective CAS Hamiltonian.

### C. Truncation schemes

We briefly consider Galerkin schemes for the bivar-MR(E)CC method, i.e., cluster operator truncations. The model space expansion coefficients  $c$  and  $d$  are in this work never truncated. The untruncated cluster operators read

$$T = \sum_{\mu \in \text{ext}} t_\mu X_\mu, \quad \Lambda = \sum_{\mu \in \text{ext}} \lambda_\mu Y_\mu, \quad (16)$$

where  $\mu \in \text{ext}$  denotes a general external excitation index. We first introduce the usual truncation scheme in terms of external singles (S), doubles (D), etc., relative to the formal reference. The result is the SD... $K(n, m)$  truncation scheme, built from a CAS with  $n$  electrons in  $m$  (spatial) orbitals, and external single and double excitations, etc., up to  $K$ -fold excitations. This simple scheme is not sufficient for most multireference cases, and in Sec. VI we show some illustrative numerical results.

In order to obtain a more balanced description for all model space states, in particular when degeneracies are present, it is necessary that the cluster operators include excitations out of all model space determinants.<sup>35</sup> The simplest choice is the first-order interaction space (FOIS)<sup>10,35,36</sup> defined by all single and double excitations relative to the *model space* into

the external space. Inclusion of the FOIS in the cluster operators ensures that computed energies will be correct through second order in perturbation theory. This is so, because the excitations in the FOIS are precisely those that are coupled to the model space  $\mathcal{H}_0$  by two-body Hamiltonians. Thus, the FOIS consists of selected external doubles, triples, quadruples, and so on. The resulting truncation will be denoted  $\text{SD}\dots K(n,m)\text{FOIS}$ . We remark that the same approach is employed in the CASCC method and also simulates the excitation manifold used in non-contracted MR methods.

#### D. Working equations

We proceed to discuss the stationary conditions of the bivar-MRCC energy. The equations for the extended version are similarly obtained, and omitted here. Differentiation of Eq. (15) with respect to the CAS amplitudes  $d_\mu$  and  $c_\mu$  yields, respectively, right and left eigenvalue equations for the effective Hamiltonian matrix  $K = K(t, \lambda)$ ,

$$Kc = Ec, \quad \text{and} \quad K^T d = Ed, \quad (17a)$$

as well as the condition  $d^T c \neq 0$ . Here,  $E = E(t, \lambda) = \mathcal{E}_{\text{bivar-MRCC}}(\lambda, t, d, c)$  can, in the regime of weak dynamical correlation, be taken to be the smallest eigenvalue. However, in practice the ground-state solution may correspond to a higher eigenvalue; see Sections V and VI. Without loss, we can assume that  $d^T c = 1$  at the solution.

Differentiation with respect to  $\lambda_\mu$  gives a ( $\lambda$ -independent) equation for  $t$ ,

$$\Omega_\mu(t, d, c) := \langle \tilde{\Phi}_\mu | D e^{-T} H e^T C | \Phi_0 \rangle = 0. \quad (17b)$$

Finally, differentiation with respect to  $t_\mu$  gives a linear equation for  $\lambda$ ,

$$\begin{aligned} \tilde{\Omega}_\mu(\lambda, t, d, c) := & \langle \tilde{\Phi}_0 | D [e^{-T} H e^T, X_\mu] C | \Phi_0 \rangle \\ & + \sum_{\nu \in \text{ext}} \langle \tilde{\Phi}_\nu | D [e^{-T} H e^T, X_\mu] C | \Phi_0 \rangle \lambda_\nu = 0. \end{aligned} \quad (17c)$$

The  $t$ -equations (17b) and the  $\lambda$ -equations (17c) are similar in structure as the corresponding equations in standard singlereference CC theory<sup>8</sup>.

#### E. Bivariational optimization of reference

We now describe the orbital-adaptive element of the bivar-MRCC method, i.e., the bivar-OAMRCC formulation. In order to alleviate the arbitrariness of the formal bra and ket

reference determinants, we introduce optional orbital rotations in the model space as bi-variational parameters, i.e., we let the active occupied orbitals  $\{\varphi_i\}$  and  $\{\tilde{\varphi}_i\}$  be variational parameters in  $\mathcal{E}_{\text{bivar-MRCC}}$ . The spin-orbitals are free to vary within the active space. Since the singlereference CC ansatz is invariant under separate rotations of occupied and virtual orbitals, it is sufficient to consider orbital variations of the form

$$\varphi_p \longrightarrow \sum_q \varphi_q (e^\kappa)_{qp}, \quad \text{and} \quad \tilde{\varphi}_q \longrightarrow \sum_p (e^{-\kappa})_{qp} \tilde{\varphi}_p, \quad (18)$$

where  $\kappa = [\kappa_{pq}]$  is a non-singular matrix with  $\kappa_{ij} = \kappa_{ab} = 0$ . The transformation preserves biorthogonality of the single-particle functions.

The determinants transform as

$$|\Phi_\mu\rangle \longrightarrow e^{\hat{\kappa}} |\Phi_\mu\rangle, \quad \text{and} \quad \langle \tilde{\Phi}_\mu| \longrightarrow \langle \tilde{\Phi}_\mu| e^{-\hat{\kappa}}, \quad (19)$$

with

$$\hat{\kappa} = \sum_{ia} \kappa_{ai} c_a^\dagger \tilde{c}_i - \kappa_{ia} c_i^\dagger \tilde{c}_a \equiv \hat{\kappa}_+ - \hat{\kappa}_-. \quad (20)$$

Here,  $\tilde{c}_q$  is the destruction operator associated with the dual spin-orbital  $\tilde{\varphi}_q$ , and  $c_p^\dagger$  is the creation operator associated with  $\varphi_p$ . The non-zero matrix elements of  $\kappa$  are all independent, and we can express the energy functional in terms of  $\kappa$ , given an arbitrary fixed “guess” of orbitals, viz.,

$$\begin{aligned} & \mathcal{E}_{\text{bivar-OAMRCC}}(\lambda, t, d, c, \kappa_-, \kappa_+) \\ &= \langle \tilde{\Phi}_0 | DC | \Phi_0 \rangle^{-1} \langle \tilde{\Phi}_0 | D(1 + \Lambda) e^{-T} e^{-\hat{\kappa}} H e^{\hat{\kappa}} e^T C | \Phi_0 \rangle. \end{aligned} \quad (21)$$

Assuming that the current basis is actually the solution, i.e.,  $\kappa = 0$  is the critical point, we obtain stationary conditions from the first-order term of the Baker–Campbell–Hausdorff (BCH) series of  $e^{-\hat{\kappa}} H e^{\hat{\kappa}}$ ,

$$0 = \langle \tilde{\Phi}_0 | D[(1 + \Lambda) e^{-T} H e^T, c_a^\dagger \tilde{c}_i] C | \Phi_0 \rangle, \quad (22a)$$

and

$$0 = \langle \tilde{\Phi}_0 | D[(1 + \Lambda) e^{-T} H e^T, c_i^\dagger \tilde{c}_a] C | \Phi_0 \rangle, \quad (22b)$$

for every pair of  $(a, i)$ .

It should be noted that, in the full, untruncated bivar-MRCC model, *all* the matrix elements of  $\kappa$  are redundant, since we are at the FCI limit by construction of the parameterization. (Note that single excitations are not removed in the spin-orbital optimization.) Thus, reference optimization only makes sense in a truncated bivar-MRCC model. Moreover, in the case where the truncation leads to an accurate dynamical correlation representation, it can be expected that the orbital dependence is weak.

#### IV. SIZE-CONSISTENCY

We now discuss the size-consistency properties of bivar-MR(E)CC in terms of the concept of additive and multiplicative separability. While not exactly equivalent to size-consistency, the notion of separability is natural in the context of both variational and bivariational approximations.

##### A. Additive and multiplicative separability

We begin with some general considerations of separability for the bivariational principle, as briefly indicated by Arponen and coworkers in the context of ECC and conventional single-reference CC.<sup>37</sup> The notion of separability goes back to Primas, see Ref. 38 and references therein. Consider the subdivision of the  $N$ -electron system, with Hilbert space  $\mathcal{H}$ , into two noninteracting subsystems of  $N_A$  and  $N_B$  electrons. We assume that  $\mathcal{H}$  and its dual  $\tilde{\mathcal{H}}$  are generated by biorthogonal spin-orbital sets  $\{\varphi_x\}$  and  $\{\tilde{\varphi}_x\}$ , respectively, which we for simplicity take to be finite. We assume that the subsystem Hilbert spaces  $\mathcal{H}_A$  and  $\mathcal{H}_B$  and their duals are generated by partitioning  $\{\varphi_x\} = \{\varphi_x^A\} \cup \{\varphi_x^B\}$  and  $\{\tilde{\varphi}_x\} = \{\tilde{\varphi}_x^A\} \cup \{\tilde{\varphi}_x^B\}$ , respectively. We write  $H_A$  and  $H_B$  for the subsystem Hamiltonians, and let  $H = H_A + H_B$  be the full supersystem Hamiltonian.

The exact problem is *separable* in the following sense: Consider bra and ket solutions of the Schrödinger equation for  $H_A$  and  $H_B$ , i.e.,

$$H_A |\Psi_A\rangle = E_A |\Psi_A\rangle, \quad H_B |\Psi_B\rangle = E_B |\Psi_B\rangle, \quad (23)$$

and

$$\langle \tilde{\Psi}_A | H_A = E_A \langle \tilde{\Psi}_A |, \quad \langle \tilde{\Psi}_B | H_B = E_B \langle \tilde{\Psi}_B |. \quad (24)$$

These bras and kets are critical points of the energy functionals  $\mathcal{E}_A(\tilde{\Psi}, \Psi) = \langle \tilde{\Psi} | H_A | \Psi \rangle / \langle \tilde{\Psi} | \Psi \rangle$  and  $\mathcal{E}_B(\tilde{\Psi}, \Psi) = \langle \tilde{\Psi} | H_B | \Psi \rangle / \langle \tilde{\Psi} | \Psi \rangle$ , and it is immediate that for the supersystem,  $|\Psi\rangle = |\Psi_A\rangle \wedge |\Psi_B\rangle$  and  $\langle \tilde{\Psi} | = \langle \tilde{\Psi}_A | \wedge \langle \tilde{\Psi}_B |$  (where  $\wedge$  denotes the antisymmetrized tensor product) are exact solutions of

$$H |\Psi\rangle = E |\Psi\rangle, \quad \langle \tilde{\Psi} | H = E \langle \tilde{\Psi} |, \quad (25)$$

with  $E = E_A + E_B$ . In other words, we have a stationary point of the supersystem energy functional  $\mathcal{E}(\tilde{\Psi}, \Psi) = \langle \tilde{\Psi} | (H_A + H_B) | \Psi \rangle / \langle \tilde{\Psi} | \Psi \rangle$ . We therefore obtain *multiplicatively separable* bra and ket eigenvectors with *additively separable* energies, and we say that the exact theory is *separable* with respect to the given subsystem partitioning. It is to be noted, that symmetries such as total spin can be broken by the product state. Additionally, we have not ruled out supersystem stationary states that are not separable in the above sense.

We define a given bivariational approximation to be separable with respect to a subsystem partitioning if the critical points  $(\tilde{\Psi}_A, \Psi_A)$  and  $(\tilde{\Psi}_B, \Psi_B)$  obtained with the approximation implies that  $(\tilde{\Psi}, \Psi) = (\tilde{\Psi}_A \wedge \tilde{\Psi}_B, \Psi_A \wedge \Psi_B)$  is a critical point for the same approximation applied to the supersystem, with critical value  $E = E_A + E_B$ . In particular, the products must lie in the supersystem approximation manifold. For a given method, there are details to be filled in, such as the choice of model spaces for both the subsystems and the supersystem.<sup>39</sup>

The presently discussed concept of separability contains no notion of a process of fragmentation, i.e., the subsystems are chosen arbitrarily, and there is no residual interaction between  $A$  and  $B$  to be nullified by distance. Separability is therefore *different* from size-consistency, where the consideration along a reaction path determines the subsystems. If only energies were considered, size-consistency is the more stringent criterion. On the other hand, separability also requires the states to be multiplicatively separable, and is therefore more stringent than mere additive separability of the energy. Finally, the present separability concept does not take into account that in a practical setting, the spin-orbital subsets are obtained from some mean-field procedure such as Hartree–Fock or CASSCF, and that the resulting model spaces may not separate accordingly.

## B. Size-intensivity of excited state-energies

For a separable method, expectation values of observables on the form  $O = O_A + O_B$  are automatically additively separable due to the Hellmann–Feynman theorem. Moreover,

excitation energies in the sense of linear-response theory are automatically size-intensive for a large class of bra and ket parameterizations. A detailed study will be presented in a future publication, and we here only sketch the outline of the argument valid for the exact case, which highlights the role of separability. The starting point is the action-like functional (4): Expansion of the state history  $(\tilde{\Psi}(t), \Psi(t))$  around the ground state state to second order gives an effective action-like functional for the first-order perturbations of the bra and ket. The resulting equations of motion gives a linear-response type eigenvalue problem for excitation energies and states.<sup>23,40</sup> Applying the multiplicative separability of the stationary state will yield an eigenvalue problem block diagonal with respect to excitations within  $A$  and  $B$ , the blocks being the corresponding subsystem eigenvalue problems. Thus, excitation energies on  $A$  and  $B$  separately will be exactly reproduced. Additionally, there is a block corresponding to joint excitations on  $A$  and  $B$ , and its solution will be products of individual excitations on  $A$  and  $B$ . Thus, excitation energies are exactly size-intensive. An example method where this argument holds true is Arponen's ECC.<sup>23</sup> This should be contrasted with equation-of-motion CC (EOM-CC), which is known to be equivalent to linear-response excitation energies for conventional single-reference CC theory. Here, the joint excitation energies for  $A$  and  $B$  are *not* additively separable,<sup>28</sup> even if the separate excitation energies on  $A$  and  $B$  are exactly reproduced in the supersystem.<sup>40</sup>

### C. Separability of bivar-MRECC

We here demonstrate separability of the bivar-MRECC method. For simplicity, we treat only the non-orbital adaptive case, but the results readily generalize. Consider non-interacting subsystems  $A$  and  $B$  defined as previously. Additionally, the subsystem spin-orbitals are subdivided according to  $\{\varphi_x^S\} = \{\varphi_\iota^S\} \cup \{\varphi_i^S\} \cup \{\varphi_a^S\} \cup \{\varphi_\alpha^S\}$ , where  $S \in \{A, B\}$ , and we require  $\{\varphi_\iota\} = \{\varphi_\iota^A\} \cup \{\varphi_\iota^B\}$ ,  $\{\varphi_i\} = \{\varphi_i^A\} \cup \{\varphi_i^B\}$ ,  $\{\varphi_a\} = \{\varphi_a^A\} \cup \{\varphi_a^B\}$ , and  $\{\varphi_\alpha\} = \{\varphi_\alpha^A\} \cup \{\varphi_\alpha^B\}$ . The bra spin-orbitals are similarly partitioned.

A straightforward calculation shows that the product of subsystem bivar-MRECC bras and kets become

$$|\Psi\rangle = |\Psi_A\rangle \wedge |\Psi_B\rangle = e^{T_A+T_B} C_A C_B |\Phi_0\rangle, \quad (26a)$$

and

$$\langle\tilde{\Psi}| = \langle\tilde{\Psi}_A| \wedge \langle\tilde{\Psi}_B| = \langle\tilde{\Phi}_0| D_A D_B e^{S_A+S_B} e^{-T_A-T_B}. \quad (26b)$$



It holds that  $C_A C_B$  and  $D_A D_B$  are CAS-only cluster operators with non-vanishing reference components, and that  $T_A + T_B$  and  $S_A + S_B$  are external cluster operators for the supersystem. Moreover, the tensor product of the FOIS for subsystems  $A$  and  $B$  form a subspace of the supersystem FOIS, and a  $K$ -fold external excitation on each subsystem is a  $K$ -fold external excitation in the supersystem. Thus, the product bra and ket lie in the supersystem biv-MRECC approximation manifold.

Since any excitation/de-excitation on  $B$  commutes with  $H_A$  and vice versa, we obtain

$$\frac{\langle \tilde{\Psi}_A \wedge \tilde{\Psi}_B | H | \Psi_A \wedge \Psi_B \rangle}{\langle \tilde{\Psi}_A \wedge \tilde{\Psi}_B | \Psi_A \wedge \Psi_B \rangle} = \frac{\langle \tilde{\Psi}_A | H_A | \Psi_A \rangle}{\langle \tilde{\Psi}_A | \Psi_A \rangle} + \frac{\langle \tilde{\Psi}_B | H_B | \Psi_B \rangle}{\langle \tilde{\Psi}_B | \Psi_B \rangle}. \quad (27)$$

It then follows that  $(\tilde{\Psi}_A \wedge \tilde{\Psi}_B, \Psi_A \wedge \Psi_B)$  is a critical point for the supersystem biv-MRECC energy functional, and we conclude that biv-MRECC is separable for noninteracting subsystems  $A$  and  $B$ .

#### D. Partial separability of biv-MRCC

Regarding the non-extended biv-MRCC method, the product of the subsystem bra functions are not in the approximation manifold for the supersystem due to the linear occurrence of  $\Lambda_A$  and  $\Lambda_B$ . Indeed, the product bra function is

$$\langle \tilde{\Psi}_A | \wedge \langle \tilde{\Psi}_B | = \langle \tilde{\Phi}_0 | D_A D_B (1 + \Lambda_A + \Lambda_B + \Lambda_A \Lambda_B) e^{-T_A - T_B}, \quad (28)$$

and the operator  $\Lambda_A \Lambda_B$  contains higher excitations than do  $\Lambda_A$  and  $\Lambda_B$ .

However, the linearity in  $\Lambda_A$  and  $\Lambda_B$  in the subsystem bras implies that the kets are in fact multiplicatively separable (just as conventional single-reference CC theory), and that the energy is additively separable. Thus,  $|\Psi\rangle = |\Psi_A\rangle \wedge |\Psi_B\rangle$  solves the amplitude equations (17b) for  $H = H_A + H_B$  with energy  $E_A + E_B$ . Regarding the supersystem bra solution, it can be shown that the solution to Eq. (17c) for the supersystem is on the form  $\Lambda = \Lambda_A + \Lambda_B + \Lambda_{AB}$ , where the mixed operator differ from  $\Lambda_A \Lambda_B$  in general. However, we obtain that for observables that separate as  $O = O_A + O_B$ , the operator  $\Lambda_{AB}$  does not contribute to the expectation value, which therefore becomes separable.

Turning to the supersystem CAS cluster operators  $C$  and  $D$ , we note that the operator  $\Lambda_{AB}$  in the previous paragraph does not contribute to the effective Hamiltonian, and we obtain the separation

$$K(t_A + t_B, \lambda_A + \lambda_B + \lambda_{AB}) = K_A(t_A, \lambda_A) \otimes I_B + I_A \otimes K_B(t_B, \lambda_B), \quad (29)$$

where  $I_A$  and  $I_B$  are the identity operators on  $\mathcal{H}_A$  and  $\mathcal{H}_B$ , respectively. It follows that the supersystem CAS cluster operators that solve Eq. (17a) become  $C = C_A C_B$  and  $D = D_A D_B$ .

We conclude that bivar-MRCC is almost separable: The energy and all separable observables are additively separable, and the ket is multiplicatively separable. The bra is almost separable, in the sense that  $\Lambda = \Lambda_A + \Lambda_B + \Lambda_{AB}$ , with  $\Lambda_{AB} \neq \Lambda_A \Lambda_B$ .

## E. Formal comparison with CASCC

It is instructive to compare the bivar-MR(E)CC method to the CASCC method of Adamowicz and coworkers, which can be defined in terms of the Lagrangian-like functional

$$\begin{aligned} \mathcal{E}_{\text{CASCC}}(\lambda, t, d, c) \\ = \langle \Phi_0 | DC | \Phi_0 \rangle^{-1} \langle \Phi_0 | (\Lambda + D) e^{-T} H e^T C | \Phi_0 \rangle, \end{aligned} \quad (30)$$

and we have  $|\Psi\rangle = e^T C |\Phi_0\rangle$  as in bivar-MR(E)CC, and  $\langle \tilde{\Psi} | = \langle \Phi_0 | (\Lambda + D) e^{-T}$ . This functional, where orthonormal spin-orbitals are assumed, is derived from the projection of the similarity transformed Schrödinger equation  $e^{-T} H e^T C |\Phi_0\rangle = E C |\Phi_0\rangle$ . While the ket parameterization is identical to bivar-MRCC, the bra is quite different, and the approximate separability as outlined for the bivar-MRCC method is lost. Furthermore, in CASCC the  $d$ -amplitudes cannot be interpreted as the left-eigenvector of the effective Hamiltonian, even if  $c$  is a right-eigenvector. Thus, while the CASCC method is cheaper and more straightforward to implement, we conjecture that the multiplicative structure of the bivar-MR(E)CC methods for both the bra and ket will have a strong impact on computed properties, response theory, and in particular excitation energies.

## V. IMPLEMENTATION

An implementation of the working equations (17) for the bivar-MRCC and bivar-MRECC methods requires solving a non-symmetric CI, a CC-type and, in the case where orbital-adaptivity is included via Eq. (22), a mean-field problem. All these are coupled. The simplest approach is an iterative approach, where these subproblems are solved in turn.

In our pilot implementation, determinant based FCI technology is used, i.e., wave functions are expanded in a FCI basis, and matrix elements over general excitation operators are decomposed into contributions of type  $\langle \tilde{\Phi}_\mu | c_x^\dagger \tilde{c}_y | \Phi_\nu \rangle$  by inserting the FCI identity.<sup>41</sup> All

operations on FCI vectors are shared-memory parallelized, and expectation values are computed by evaluating inner products. For simplicity, spin symmetry is exploited only in the computation of matrix-vector products involving the Hamiltonian (also known as  $\sigma$ -vectors). Up to four vectors of full length are kept in memory, and no efficient CC or CI partitioning techniques are implemented,<sup>28,41,42</sup> thus limiting the scope of applicability to performing benchmark studies on small model systems. A more efficient implementation is currently under development, and uses the fact that single-reference technology can be used to compute the contributions that feature the same model space determinant (generated by  $D$  and  $C$  operators) on the left and right hand side of Eq. (17) and Eq. (22), respectively. This approach has been discussed for Mukherjee’s state-specific Mk-MRCC method with singles and doubles (Mk-MRCCSD)<sup>43</sup> and leads to an effective scaling of  $\mathcal{O}(n_{\text{MS}}n_{\text{occ}}^2n_{\text{virt}}^4)$  for these terms, with  $n_{\text{MS}}$  being the number of model space determinants, and  $\mathcal{O}(n_{\text{occ}}^2n_{\text{virt}}^4)$  the scaling of the conventional CCSD method. The remaining contributions are very sparse and automated code generation together with tensor contraction technology can be used to compute these efficiently.<sup>44–46</sup> Regarding the orbital adaptive variants, efficient techniques as recently discussed by Olsen and co-workers<sup>16,47,48</sup> will be considered.

The external cluster operators used in the present implementation are defined by the  $\text{SD}\dots K(n, m)$  and  $\text{SD}\dots K(n, m)\text{FOIS}$  schemes, see Section III C. A collective  $K$ -fold excitation index  $\mu \in \text{ext}$  is represented by the  $2K$ -tuple

$$\begin{aligned} \mu \rightarrow (\iota_1^{(\mu)}, \dots, \iota_M^{(\mu)}, i_{M+1}^{(\mu)}, \dots, i_K^{(\mu)}, \\ a_1^{(\mu)}, \dots, a_{K-M'}^{(\mu)}, \alpha_{K-M'+1}^{(\mu)}, \dots, \alpha_K^{(\mu)}), \end{aligned} \quad (31)$$

with  $\iota_1^{(\mu)} > \dots > \iota_M^{(\mu)}$ ,  $i_{M+1}^{(\mu)} > \dots > i_K^{(\mu)}$ ,  $a_1^{(\mu)} > \dots > a_{K-M'}^{(\mu)}$ , and  $\alpha_{K-M'+1}^{(\mu)} > \dots > \alpha_K^{(\mu)}$ . In Eq. (31), internal amplitudes (defined by  $M = M' = 0$ ) are excluded. For computations with nonzero inactive orbitals (i.e.,  $M > 0$ ), linear dependencies (e.g., induced by double excitations containing active-active “spectator excitations”<sup>49,50</sup>) can occur in the amplitude equation iterations, and are removed efficiently by orthogonalization (*vide infra*). In each step of the mean-field optimization Eq. (22), the integrals with active indices are transformed according to Eq. (18), ensuring that  $X_\mu$  is (Hermitian) adjoint to  $Y_\mu$  at any step in the computation.

In our current implementation, the maximum excitation rank is not limited, in principle allowing for arbitrary order cluster operators. Excitations from the model space to the

FOIS are implemented by redefinition in terms of excitations with respect to the reference determinant. Thus, a bivar-MR(E)CCSD( $n,m$ )FOIS cluster operator contains maximally  $(n+2)$ -fold excitations with respect to the reference determinant.<sup>10</sup> When the CAS is large, the FOIS will add a considerable number of amplitudes. To reduce computational cost, these excitations may be restricted according to the weight of the model space determinant in Eq. (10) and Eq. (9), i.e., excitations out of a determinant with negligible weight in the model space bra and ket may be neglected (cf. Section IV, SI for a numerical example).

The computation of the bivar-MR(E)CC wave functions and energies is performed iteratively (see Fig. 2): First, based on a (fixed) model space definition, the Hamiltonian is diagonalized in this subspace to give the model-space wave functions (9) and (10). Based on these CASCI vectors, the (initial) reference state and reference determinant are defined. Using this definition, the doubles part of the  $t$ -amplitude vector is populated with second-order Møller–Plesset (MP2) values. Then, the  $t$ - and  $\lambda$ -equations are solved iteratively. In the case of bivar-MRECC,  $t$  and  $\lambda$  are optimized simultaneously. If orbital adaptivity is considered, the orbitals are optimized either before or after solving the CC problem by solving Eq. (22), and the integrals are transformed into a basis where all  $\kappa_{pq} = 0$ .<sup>51</sup> Finally, the matrix  $K(t, \lambda)$  (Eq. (15)) is constructed within the model space and diagonalized to give the updated model space vectors  $c$  and  $d$ . The amplitude vectors  $t$  and  $\lambda$  and the CI coefficients are re-optimized until convergence, which is typically achieved in 3 to 10 (outer) iterations.

During iterations, the character of the reference wave functions Eq. (9) and Eq. (10) is preserved by choosing those eigenvectors of Eq. (17a) that have the largest overlap with the  $c$  and  $d$  vectors from the previous iteration. This avoids problems created by root flipping. Thus, the initial choice immediately after the CASCI step defines the nature of the state that is optimized. Similarly to the CASCC(sw) method<sup>52</sup>, the reference determinant is allowed to change dynamically during iterations. Whenever this happens, the definition of the cluster operators is reset and the amplitudes are reinitialized to MP2 values.

Convergence of the  $t$ - and  $\lambda$ -equations is accelerated by using a quasi-Newton-Raphson update together with the direct inversion in the iterative subspace technique.<sup>53</sup> If the set of inactive orbitals is nonempty, the amplitude vectors are orthogonalized using Cholesky decomposition of the metric (which is observed to be close to symmetric in most cases)  $S_{\mu\nu} = \langle \tilde{\Psi}_0 | Y_\mu X_\nu | \Psi_0 \rangle \approx \sum_\rho L_{\mu\rho} L_{\nu\rho}$ .<sup>54–57</sup> The amplitude update is then given by  $\Delta t_\mu = -\sum_\nu L_{\mu\nu}^{-1} \Omega_\nu(t, d, c) / \Delta_\nu$ , where  $\Delta_\nu$  is the MP $n$  energy denominator. The  $\lambda$ -amplitudes are

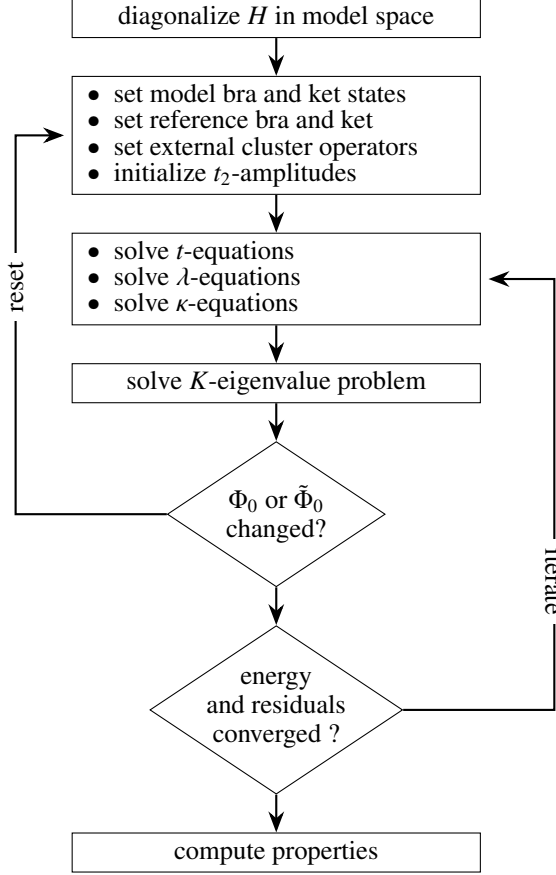


FIG. 2. Iteration scheme for a bivar-MR(E)CC computation. After the initial diagonalization of  $H$  in the CAS model space and the consequent initialization of bivar-MR(E)CC variables, the main loop solves the various working equations in turn. If the reference changes after the  $K$ -matrix diagonalization, the bivar-MR(E)CC variables need to be reset. The iteration truncates after the energy changes less than a given tolerance and the residual norms are sufficiently small. See text for further details.

updated similarly. The CI-expansion vectors are updated with aid of a minimum polynomial extrapolation technique.<sup>58</sup> In the bivar-MRECC case, a flexible microiteration scheme is employed.

## VI. NUMERICAL RESULTS

Accuracy is one of the most important requirements for a novel MRCC method. Other (weaker) conditions, like size-consistency, favorable computational scaling, orbital invariance, symmetry preservation, natural reduction to SRCC for single-reference cases, and

tractable access to molecular properties have also been formulated.<sup>10</sup> Moreover, being partially motivated from *mathematical* arguments and based on an unconventional formulation of quantum mechanics, the bivar-MRCC method should be tested with respect to *physical predictions*, i.e., expectation values as defined in Eq. (3). For this reason, we did not only investigate the accuracy of absolute energies, but also the quality of the actual density operators  $\rho = |\Psi\rangle\langle\tilde{\Psi}| / \langle\tilde{\Psi}|\Psi\rangle$  compared to FCI results. Furthermore we emphasize, that our scope is not to present the performance of the method under ideal conditions, but rather study the behaviour using set-ups typically found in everyday and sub-optimal applications, e.g., by using different reference orbitals. To this end, absolute energies, spin-expectation values, dipole moments and density operators have been computed using different orbital sets and compared to FCI results. Whenever possible, computed quantities are compared to the results of other MRCC methods found in the literature.

### A. Error measures relative to full CI

Absolute energies are compared to the respective FCI values by evaluating the difference  $\Delta E_{\text{FCI}} = \text{Tr}(\rho H) - E_{\text{FCI}}$ . In order to quantify the accuracy of the density operator, the Frobenius norm  $\|\delta\rho_{\text{FCI}}\|_{\text{F}}^2 = \text{Tr}((\rho - \rho_{\text{FCI}})^\dagger(\rho - \rho_{\text{FCI}}))$  has been calculated, a standard coherence and entanglement measurement in quantum information theory.<sup>59,60</sup> Small values of  $\|\delta\rho_{\text{FCI}}\|_{\text{F}}$  indicate that  $\rho$  is a good approximation to the FCI state. Since spin-symmetry can be directly related to the quality of the approximate bra and ket,<sup>61</sup> the total-spin contamination  $\Delta S^2_{\text{FCI}} = (\text{Tr}(\rho S^2) - S(S+1))/\hbar^2$  is used as an additional accuracy measure. Finally, the accuracy of dipole moments is expressed by  $\|\Delta m_{\text{FCI}}\|_2^2 = \sum_{i=1}^3 (\text{Tr}(\rho m_i) - m_{i,\text{FCI}})^2 / (ea_0)^2$ , where  $m_i$  denotes the  $i$ -th component of the electronic dipole operator.

Statistical errors are computed using the following definitions: deviation  $\Delta x_i = x_i - x_i^{\text{FCI}}$ , mean deviation,  $\overline{\Delta x_i} = \sum_{i=1}^n \Delta x_i / n$ , mean absolute deviation  $\text{MAD}(\Delta x_i) = \sum_{i=1}^n |\Delta x_i| / n$ , maximum absolute deviation  $\text{MAX}(\Delta x_i) = \max_{i=1,\dots,n} |\Delta x_i|$ , standard deviation  $\text{STD}^2(\Delta x_i) = \sum_{i=1}^n (\Delta x_i - \overline{\Delta x_i})^2 / n$  and non-parallelity error  $\text{NPE}(\Delta x_i) = \max_{i=1,\dots,n} \Delta x_i - \min_{i=1,\dots,n} \Delta x_i$ .

## B. Model systems

A multireference method should be reasonably accurate for both single-reference and multireference problems. We therefore opted for studying the novel bivar-MR(E)CC methods using a model system providing both. The computational investigation of the potential curve of the symmetrical insertion of Be into  $\text{H}_2$  has been described comprehensively<sup>18</sup> and serves as a standard example for testing novel MRCC methods, since it is computationally feasible even for complicated methods.<sup>10,19</sup> Moreover, it features a lot of problems for quantum chemical methods like severe multireference character, level crossings and change of leading determinant along the potential curve.<sup>9</sup> Therefore the performance of the bivar-MR(E)CC methods has mainly been tested using this system. Additionally, the chemical bond-breaking of the hydrogen fluoride molecule in the ground state and the widely-used  $\text{H}_8$  model system<sup>35,62</sup> have been studied using the bivar-MRCC(2,2)FOIS method yielding highly accurate results. For example, H–F bond breaking with 12 points in  $1.0 \leq R_{\text{H-F}} \leq 5.0 a_0$  yielded  $\text{MAD}(\Delta E_{\text{FCI}}) = 1.12 mE_{\text{H}}$ ,  $\text{MAX}(\Delta E_{\text{FCI}}) = 1.27 mE_{\text{H}}$ ,  $\text{STD}(\Delta E_{\text{FCI}}) = 0.06 mE_{\text{H}}$ , and  $\text{NPE}(\Delta E_{\text{FCI}}) = 0.22 mE_{\text{H}}$  (cf. Section II, SI). However, since the electronic structures of HF and  $\text{H}_8$  are less complicated than the one of the  $\text{BeH}_2$  system, the results are not discussed in detail here, but can be found in the Supplementary Information. Furthermore, we computed the potential curve for the  $\text{N}_2$  dissociation using the bivar-MRCCSD(6,6)FOIS model ( $\text{MAD}(\Delta E_{\text{FCI}}) = 0.13 mE_{\text{H}}$ ,  $\text{MAX}(\Delta E_{\text{FCI}}) = 0.19 mE_{\text{H}}$ ,  $\text{STD}(\Delta E_{\text{FCI}}) = 0.03 mE_{\text{H}}$ , and  $\text{NPE}(\Delta E_{\text{FCI}}) = 0.11 mE_{\text{H}}$ , cf. Section IV, SI), yielding similarly accurate results as described in Ref. 63 in the context of the MRexpT and SRMRCC methods, even for restricted FOIS excitations. However, due to the small basis, combined with the large (400 determinants) model space employed, these computations are *a priori* expected to be highly accurate and therefore not discussed in detail. Finally, we conducted a numerical size-consistency test consisting of two  $\text{H}_2$  molecules at quasi-infinite distance<sup>49</sup> yielding size-consistent energies (cf. Section IV D and SI Section V) within the thresholds employed in the calculations.

## C. Technical details

In order to be able to compare to other MRCC methods and owing to computational restrictions, the same parameters regarding geometry and basis set described in Refs. 19,

52, 62, and 63 have been used for the BeH<sub>2</sub> (10s3p/3s2p and 4s/2s basis), HF (DZV basis), H<sub>8</sub> (minimal basis) and N<sub>2</sub> (6-31G basis) models, respectively. The ( $C_1$  as well as  $C_{2v}$ ) CASSCF orbitals have been computed using the Bochum-suite of *ab initio* wave function programs.<sup>57,64,65</sup> The FCI computations are performed with a local program based on a quasi-relativistic CI code<sup>66</sup> and verified against the results presented in Ref. 19. If not otherwise mentioned, energies and residuals were converged to thresholds  $10^{-6}$  a.u. and  $10^{-4}$  a.u., respectively. Amplitudes with absolute value smaller than  $10^{-10}$  a.u. were neglected. If not mentioned otherwise, FOIS excitations are not restricted. In all computations discussed here, all electrons were correlated.

#### D. BeH<sub>2</sub> full CI results

The symmetric insertion geometry of the system is parameterized using the distance  $x$  from the Be atom to the H<sub>2</sub> moiety, with  $x = 0$  referring to the linear arrangement.<sup>19</sup> (Note that the molecule has been rotated in space, thereby interchanging  $b_1$  and  $b_2$  irreducible representations.) For  $0 < x \leq 4 a_0$ , the system comprises the symmetry of the  $C_{2v}$  point-group. The FCI energies for the first 10 states with  $M_S = 0$  are shown in Fig. 3. (The mapping of the  $C_1$  FCI states  $\Gamma_1, \Gamma_2$ , etc., onto the corresponding states in  $C_{2v}$  is given in the SI.) Apparently, the nature of the  $C_1$  ground state changes along the insertion pathway. The respective  $C_1$  ground state  $\Gamma_1$  ( $^1A_1$  or  $^3B_1$  in  $C_{2v}$ ) is dominated by the appropriate linear combinations of the following four determinants:

$$\begin{aligned} |\Phi_1\rangle &= |(1a_1)^2(2a_1)^2(1b_1)^2\rangle, \\ |\Phi_2\rangle &= |(1a_1)^2(2a_1)^2(3a_1)^2\rangle, \\ |\Phi_{3,4}\rangle &= |(1a_1)^2(2a_1)^2(3a_1)^1(1b_1)^1\rangle, \end{aligned} \tag{32}$$

where the exponent denotes the (spatial) orbital occupancy. Concerning the cusp by  $x = 2.75 a_0$ , the FCI wave function of the  $^1A_1$  state (CAS(2,2)SCF orbitals) constitutes  $\sim 52\%$  of  $|\Phi_1\rangle$  and  $\sim 39\%$  of  $|\Phi_2\rangle$ , making it more suitable for a single-reference based MRCC description than a 50:50 mixture. Furthermore, there are small contributions from other determinants, namely the  $^1A_1$ -symmetric combinations of  $|(1a_1)^2(2a_1)^1(3a_1)^1(1b_1)^1(2b_1)^1\rangle$  ( $\sim 2\%$ ) and  $|(1a_1)^2(2a_1)^2(1b_2)^2\rangle$  ( $\sim 2\%$ ) which should be included in an accurate correlation treatment (cf. Section III, SI).



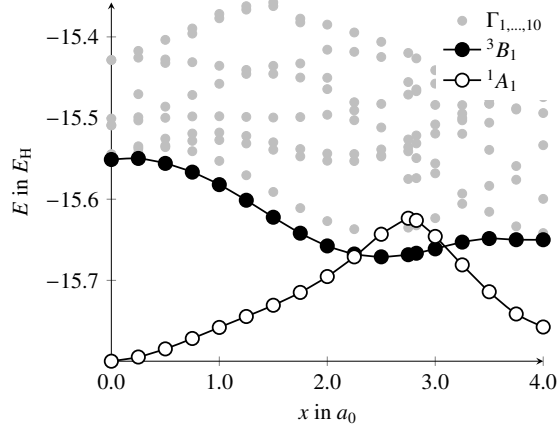


FIG. 3. FCI potential curves for the  $C_{2v}$ -symmetric insertion of Be into  $H_2$ . Note the region where the  $^1A_1$  state becomes an excited state.

### E. Absolute energies

The potential energy curve of the  $^1A_1$  state has been investigated using the bivar-MRCCSD method with 2-in-2 (CAS(2,2)) and 4-in-6 active spaces (CAS(4,6)), based on the respective CASSCF orbitals. The CAS(2,2) is spanned by the four determinants given in Eq. (32), i.e., the  $1b_1$  and  $3a_1$  orbitals are chosen active. For CAS(4,6), the doubly occupied  $2a_1$  orbital and three virtual orbitals are added based on orbital energy<sup>36</sup>. The reference wave functions for the bivar-MR(E)CCSD computations using the CAS(2,2) model space have been constructed in the following way (The same procedure has also been applied for the computations based on the CAS(4,6) model space): Diagonalizing both the Hamiltonian and the  $K$ -matrix in this space yields four states, that in  $C_{2v}$  transform as two  $^1A_1$ , one  $^1B_1$ , and one  $^3B_1$  state. The initial values for the expansion vectors  $c$  and  $d$  were chosen such that they correspond to the one of the two the  $^1A_1$  (CASCI) states being lowest in energy. The reference wave functions were then updated during the iterative procedure, preserving the  $^1A_1$  nature of the expansion vectors throughout by an overlap criterion (cf. Section V). Note, that for  $2.25 \leq x \leq 3.0 a_0$ , this means that the optimization has been performed for an excited state.

The results can be found in Fig. 4. The bivar-MRCCSD(2,2) results are close to the values obtained with single-reference CCSD using  $|\Phi_1\rangle$ ,<sup>19</sup> indicating that these calculations lack important dynamical correlation contributions from other determinants, in particular doubly excited determinants relative to  $|\Phi_2\rangle$ .<sup>35</sup> Including single- and double excitations for all four

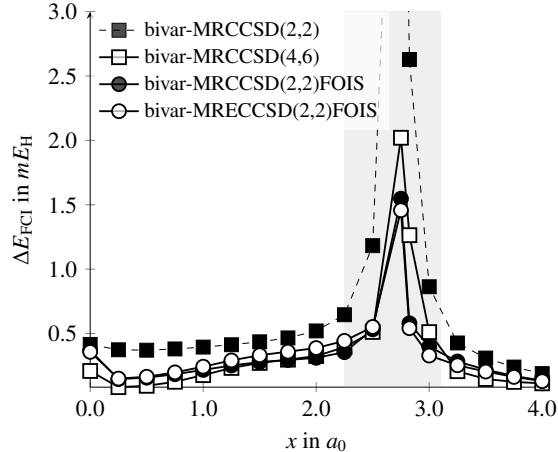


FIG. 4. bivar-MR(E)CCSD energy differences with respect to FCI values for the  $^1A_1$  state of  $\text{BeH}_2$  based on the respective CASSCF orbitals. The light gray background indicates the region where  $^1A_1$  is an excited state.

model space determinants (bivar-MRCCSD(2,2)FOIS) improves the treatment significantly and decreases the maximum error to just under the “chemical accuracy limit” of 1 kcal/mol ( $\sim 1.594 mE_H$ ). The results of the extended variant bivar-MRECCSD(2,2)FOIS are very similar (cf. Table I), slightly superior in the multireference, but slightly inferior in the single-reference region. The same findings have been described for the single-reference coupled cluster and extended coupled cluster methods using the same model system.<sup>19</sup> Increasing the model space to include 225 determinants (CAS(4,6)), but neglecting the FOIS (bivar-MRCCSD(4,6)) yields results similar to the four determinant model space variant with additional FOIS (bivar-MRCCSD(2,2)FOIS).

All curves show a discontinuity at  $x = 2.75 a_0$  which can be traced back to the complicated nature of the FCI wave function at this geometry, constituting an almost 50:50 mixture of the determinants  $|\Phi_1\rangle$  and  $|\Phi_2\rangle$  as described in Section VI D. The effect of bivariational reference optimization by orbital optimization at this point will be discussed in Section VI G.

A comparison to other MRCC methods using identical basis/geometry setup is shown in Fig. 5. The ic-MRCCSD method uses a sophisticated internally contracted ansatz where the cluster operator includes terms from all model space states, see Eq. (11).<sup>49,55</sup> It can therefore be assumed to be similar or more accurate than the single-reference ansatz used in bivar-MRCCSD. The latest CASCCSD(sw) method is closest to our current bivar-MRCCSD(2,2)FOIS method but much more accurate in the region where  $^1A_1$  is an excited

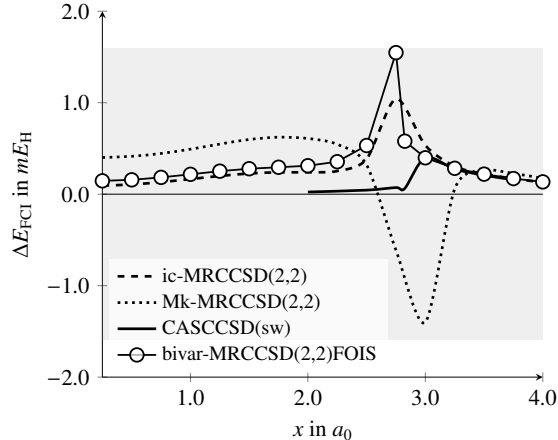


FIG. 5. Method comparison showing energy differences of the  $^1A_1$  state of  $\text{BeH}_2$  with respect to FCI results of different MRCC methods. The gray area indicates the region of "chemical accuracy", i.e.,  $|\Delta E_{\text{FCI}}| \leq 1$  kcal/mol.

state in  $C_1$ .<sup>52</sup> This discrepancy might primarily be traced back to the different model space reference wave function used. In test computations on a  $\text{H}_8$  model system, the CASCCSD(sw) and MRCCSD(2,2)FOIS results were very similar, with a slight superiority of MRCCSD(2,2)FOIS in the multireference region (cf. Section I, SI). Additionally, the results using the established Mk-MRCCSD method<sup>67</sup> are shown. However, being an Jeziorski–Monkhorst type method<sup>50</sup>, a direct comparison is complicated, and we merely note that the overall accuracy is good despite the instability in the region where  $^1A_1$  is not the ground state. Finally, we also like to mention that the MRCCSD method from Kállay, Szalay, and Surján<sup>68</sup> ( $\Delta E_{\text{FCI}} = 1.890 mE_{\text{H}}$  at  $x = 2.75 a_0$ ) and the MRexpT method from Hanrath *et al.*<sup>69</sup> ( $\text{MAD}(\Delta E_{\text{FCI}}) = 0.591 mE_{\text{H}}$ ,  $\text{MAX}(\Delta E_{\text{FCI}}) = 1.693 mE_{\text{H}}$ ,  $\text{NPE}(\Delta E_{\text{FCI}}) = 1.663 mE_{\text{H}}$ ) are very accurate. However, the reported values are based on SCF orbitals and/or use a different basis set and are therefore not shown here. Altogether, most MRCC methods discussed, including the novel bivar-MRCCSD model, demonstrate chemical accuracy, i.e.,  $|\Delta E_{\text{FCI}}| \leq 1$  kcal/mol for this model system.

## F. Density operators

Absolute energies are not good indicators of accuracy in general, particularly when the desired state is not the ground state. In Fig. 6, the energy differences of several bivar-

MRCCSD(2,2)FOIS computations with respect to FCI values are shown. These computations differ only in the orbitals used for constructing the computational Hilbert space, including the external space. Apparently, by using the “right” set of orbitals one can get very close to and even below the FCI energy, in particular inside the region where the desired state is not the ground state. Thus, in order to assess whether the right value has been obtained for the right reason, a more reliable characteristic has to be used.

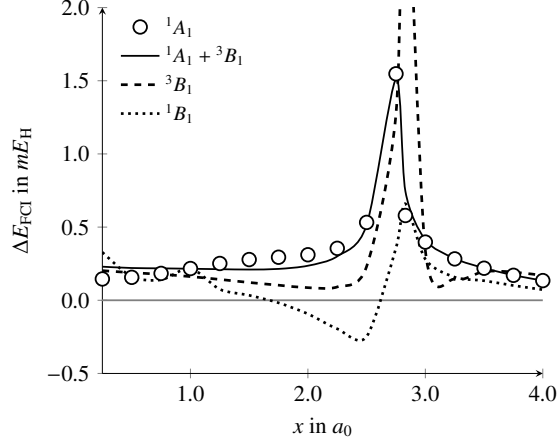


FIG. 6. bivar-MRCCSD(2,2)FOIS energy differences with respect to FCI values for the  $^1A_1$  state of  $\text{BeH}_2$  based on CAS(2,2)SCF orbitals optimized for different electronic states.

To this end, the Frobenius norm  $\|\delta\rho_{\text{FCI}}\|_F$  of the difference density operator (cf. Section VIA) of the  $^1A_1$  state has been computed for the bivar-MRCCSD(2,2)FOIS method along the  $\text{BeH}_2$  potential curve using CAS(2,2)SCF orbitals optimized for different electronic states. This data is summarized in Tab. I. Additionally, values for the bivar-MRECCSD(2,2)FOIS and bivar-MRCCSD(4,6) variants are shown for comparison.

The mean absolute deviations of both the bivar-MRCCSD(2,2)FOIS energy and density errors computed over the entire potential curve using different orbitals are similar, but the maximal absolute deviations differ significantly. Consider for example the errors at  $x = 2.75 a_0$ . While the energy error of the computation based on orbitals for  $^1B_1$  is very small ( $0.358 mE_H$ ), the error in the density operator is rather large (0.206) when compared to the errors obtained with  $^1A_1$  orbitals ( $1.547 mE_H$  and 0.052). This can be resolved by analysing the CAS(2,2)SCF wave functions: The  $^1B_1$  state is composed of the open-shell determinants  $|\Phi_3\rangle$  and  $|\Phi_4\rangle$  (cf. Eq. (32)), while in  $^1A_1$ , the weights of the closed-shell determinants  $|\Phi_1\rangle$  and  $|\Phi_2\rangle$  are large. Thus, optimizing orbitals for the  $^1B_1$  (or  $^3B_1$ ) state has

a significant impact on the  $^1A_1$  wave function without contaminating the overall symmetry of the wave function.

Based on the density error analysis, it can thus be concluded that the energy errors of the computations with orbitals optimized for  $^1B_1$  and  $^3B_1$  are unreliable. In contrast, the values obtained with state-specific ( $^1A_1$ ) and state-average (50:50 mixture of  $^1A_1$  and  $^3B_1$  states) orbitals are considerably smaller, i.e., they represent the FCI state better. Considering actual applications, we note that typically one of the latter two orbital sets will be used<sup>36</sup> – both of which have been demonstrated to be accurate for the right reason.

TABLE I. Comparison of energy and density operator error characteristics of the BeH<sub>2</sub> potential curve computed with bivar-MRCCSD(2,2)FOIS based on CAS(2,2)SCF orbitals optimized for different states. For comparison, bivar-MRECCSD(2,2)FOIS and bivar-MRCCSD(4,6) values are presented in the last two rows.

Orbitals	$\Delta E_{\text{FCI}} (mE_{\text{H}})$			$  \delta\rho_{\text{FCI}}  _{\text{F}}$		
	MAD	MAX	2.75 $a_0$	MAD	MAX	2.75 $a_0$
$^1A_1$	0.356	1.547	1.547	0.020	0.052	0.052
$^1A_1 + ^3B_1$ <sup>a</sup>	0.356	1.511	1.511	0.021	0.040	0.033
$^3B_1$	0.411	3.172	1.272	0.033	0.199	0.108
$^1B_1$	0.192	0.659	0.358	0.037	0.206	0.206
$^1A_1$ <sup>b</sup>	0.364	1.458	1.458	0.021	0.050	0.050
$^1A_1$ <sup>c</sup>	0.406	2.020	2.020	0.023	0.078	0.078

<sup>a</sup> 50:50 mixture

<sup>b</sup> bivar-MRECCSD(2,2)FOIS

<sup>c</sup> bivar-MRCCSD(4,6)

As a second measure for density operator accuracy, the spin-contamination  $\Delta S^2_{\text{FCI}}$  has been computed. In the present implementation, only the symmetry  $[H, S_z] = 0$  is exploited, total-spin conservation is not enforced. However, using a qualitatively correct model-space bra and ket is generally thought to be able to reduce the spin-contamination in the correlated wave function substantially.<sup>10</sup> For all singlet states studied, the computed spin-contamination was negligible with  $\Delta S^2_{\text{FCI}} \ll 10^{-3}$ . Concerning the triplet  $^3B_1$  state, the errors are relatively small for bivar-MRCCSD(2,2)FOIS computations ( $\text{MAD}(\Delta S^2_{\text{FCI}}) < 0.02$ ,

$\text{MAX}(\Delta S^2_{\text{FCI}}) < 0.07$ , cf. Section III, SI) compared to the values discussed in the context of single-reference CC methods.<sup>61,70,71</sup> These errors decrease further with increasing active space size ( $\text{MAD}(\Delta S^2_{\text{FCI}}) < 0.01$ ,  $\text{MAX}(\Delta S^2_{\text{FCI}}) < 0.04$  for bivar-MRCCSD(4,6)). Thus, the errors in the expectation value of the total-spin operator induced by the coupled cluster expansion used in the bivar-MRCC methods are, at least for this case, insignificant.

## G. Orbital optimization

In the orbital-adaptive variant bivar-OAMRCC, active-active non-orthogonal orbital rotations are introduced via Eq. (18). To investigate the effect of this mean-field orbital optimization, test computations have been performed on the  $\text{BeH}_2$  system in the multireference region ( $x = 2.75 a_0$ ) using the bivar-OAMRCCSD(2,2)FOIS and bivar-OAMRCCSD(4,6) models. In all calculations the results have been found to be very close to the results obtained without orbital optimization, i.e., below the convergence threshold. For the computations using the small CAS(2,2) model space, this is due to almost vanishing gradient norms at every iteration. Concerning the CAS(4,6) based computations, we have tested two different iteration schemes, employing a different ordering of solving the mean-field and the CC problem, respectively. The gradient norms are not small in this case and the results are discussed in Section III in the supplementary information.

However, the insensitivity of the energy towards bivariational optimization of the reference may be due to the limited size of this system. The SD-FOIS truncation of the external amplitudes is already very large for a six-electron system. Recall, that at the FCI limit all orbital rotations are redundant. Moreover, the choice of the reference should play a major role in general, since the conventional argument against MRCC methods based on single-reference theory is its bias towards the formal reference. Thus, we conjecture that for larger systems, in particular extended systems, orbital optimization will play a much larger role.

## H. Molecular properties

To gain some insight into the accuracy of first-order properties computed with the bivar-MRCC method, the dipole moments of the  $^1A_1$  state of  $\text{BeH}_2$  along the PES have been computed using Eq. (3). The electronic dipole moment integrals were taken from a local

version of the TURBOMOLE program package.<sup>72</sup> The computed values are compared to the corresponding FCI results by evaluating the error measure  $\|\Delta m_{\text{FCI}}\|_2$  (cf. Section VIA). For comparison, (orbitally unrelaxed) single-reference CC dipole moments have been computed using the CFOUR program package.<sup>73</sup> The mean absolute and maximum absolute deviations are depicted in Fig. 7, the individual values can be found in the SI. The coupled-cluster with singles and doubles (CCSD) and coupled-cluster with singles, doubles, and triples (CCSDT) dipole moments based on restricted Hartree-Fock orbitals are very accurate for single-reference systems, but less accurate in the multireference region.<sup>74</sup> The bivar-MRCCSD computations improve upon the CAS(2,2)SCF reference values significantly, and even outperform the RHF-CCSDT method in this example.

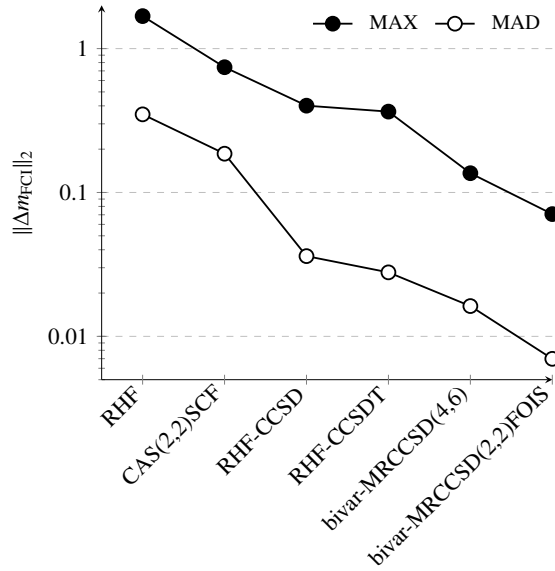


FIG. 7. Error characteristics of the electronic dipole moment vector compared to FCI values for the potential curve of the  $^1A_1$  state of  $\text{BeH}_2$ .

## VII. CONCLUSION

In this article, we have introduced a state-specific multireference coupled-cluster method based on Arponen’s bivariational principle, the bivar-MRCC method. An extended version, bivar-MRECC, was also discussed, similar to Arponen’s extended CC method. The bivar-MRCC method is wholly based on singlereference theory, has modest complexity, and avoids some of the problems associated with established multireference methods. For example, all

cluster operators commute, and there is no need for sufficiency conditions as is needed in, e.g., Mukherjee’s state-specific method. The method requires a formal reference much like the CASCC method of Oliphant, Adamowicz, and Piecuch, but the bias can to an extent be eliminated using bivariational optimization of the reference.

The bivar-MR(E)CC methods are demonstrated to be size-consistent in the sense of additive separability of the energy under a partitioning of the  $N$ -electron system into non-interacting subsystems. Moreover, bivar-MRECC is separable also in the sense that the state is multiplicatively separable. For bivar-MRCC the separability is only approximate for the bra. We expect this approximate separability for bivar-MRCC to give high accuracy of excited states and response properties.

A pilot implementation has been described, and extensive benchmark calculations on the insertion of a Be atom into  $H_2$  has been performed. Therein, the method has been demonstrated to be very accurate, i.e., within the desired range of chemical accuracy, and an analysis of the density- and total-spin operator shows that the state description is indeed very accurate despite that the ansatz requires “two wave functions”. All in all, the bivar-MRCC method seems to perform equally well as established state-specific MRCC methods do. While the pilot implementation is based on full-configuration interaction methodology to facilitate rapid development of a flexible program, a more efficient and optimally scaling implementation has been outlined. Such an implementation will open up the possibility for applications closer to the state-of-the-art, including transition metal chemistry and luminescence phenomena.

The systems studied exhibited only weak dependence on the orbital rotations in the working equations. We conjecture that for such small systems as were studied, the first-order interaction space is sufficient to describe the majority of dynamical correlation, which means that the bivar-MRCC state is near the FCI state, and in this limit the orbital rotations are redundant. We furthermore conjecture that orbital rotations will play a larger role for larger systems.

From the point of view of theory, the natural continuation of this work is the derivation of response theory and theory for excited states, which the bivariational approach allows in a relatively straightforward manner. Moreover, the time-dependent bivariational principle combined with biorthogonal orbital-optimization allows an *ab initio* dynamics method suitable to, say, study molecules under the influence of intense laser pulses, charge migration,



charge transfer, and other situations where the system evolves far away from the ground-state. Multireference character of the resulting state can be expected to be significant, and is a major challenge of state-of-the-art methods today.

The bivariational formulation of bivar-MRCC has an important advantage in that a mathematical *a priori* error analysis is possible. The major challenge is finding the right assumptions on the system Hamiltonian and model space to facilitate a monotonicity analysis. If these assumptions are also reasonable in a wide range of situations, the bivar-MRCC method gains a distinct advantage over other MRCC theories, for which few mathematical results exist.

The modest complexity of the bivar-MRCC method allows extending the field of application far beyond the simple benchmark calculations presented here, once an efficient implementation is in place. We conclude that the bivar-MRCC method has potential to become a useful and practical tool in many areas of quantum molecular sciences, also for non-experts.

## ACKNOWLEDGMENTS

This work has received funding from the Research Council of Norway (RCN) under CoE Grant Nos. 287906 and 262695 (Hylleraas Centre for Quantum Molecular Sciences), and from ERC-STG-2014 under grant agreement No 639508. We further thank the Norwegian Metacenter for Computational Science for support via NOTUR, project no. NN4654K. The authors thank A. Laestadius, M. E. Harding and T. B. Pedersen for constructive remarks.

## DATA AVAILABILITY

The data that supports the findings of this study are available within the article and its supplementary material.

## REFERENCES

- <sup>1</sup>J. Arponen, J. Phys. G: Nucl. Phys. **8**, L129 (1982).
- <sup>2</sup>J. Arponen, Ann. Phys. **151**, 311 (1983).
- <sup>3</sup>A. Laestadius and S. Kvaal, SIAM J. Numer. Anal. **56**, 660 (2018).

- <sup>4</sup>E. Zeidler, *Nonlinear Functional Analysis and its Application II/B: Nonlinear Monotone Operators* (Springer, New York, Heidelberg, Berlin, 1990).
- <sup>5</sup>R. P. Feynman, Phys. Rev. **56**, 340 (1939).
- <sup>6</sup>T. Helgaker and P. Jørgensen, Theor. Chim. Acta **75**, 111 (1989).
- <sup>7</sup>T. Helgaker and P. Jørgensen, Adv. Quant. Chem. **19**, 183 (1988).
- <sup>8</sup>H. Koch, H. J. Aa. Jensen, P. Jørgensen, T. Helgaker, G. E. Scuseria, and H. F. Schaefer, J. Chem. Phys. **92**, 4924 (1990).
- <sup>9</sup>A. Köhn, M. Hanauer, L. Mück, T.-C. Jagau, and J. Gauss, WIREs Comput. Mol. Sci. **3**, 176 (2012).
- <sup>10</sup>D. I. Lyakh, M. Musiał, V. F. Lotrich, and R. J. Bartlett, Chem. Rev. **112**, 182 (2012).
- <sup>11</sup>F. A. Evangelista, J. Chem. Phys. **149**, 030901 (2018).
- <sup>12</sup>N. Oliphant and L. Adamowicz, J. Chem. Phys. **94**, 1229 (1991).
- <sup>13</sup>N. Oliphant and L. Adamowicz, J. Chem. Phys. **96**, 3759 (1992).
- <sup>14</sup>P. Piecuch, N. Oliphant, and L. Adamowicz, J. Chem. Phys. **99**, 1875 (1993).
- <sup>15</sup>T. B. Pedersen, B. Fernández, and H. Koch, J. Chem. Phys. **114**, 6983 (2001).
- <sup>16</sup>J. Olsen, J. Chem. Phys. **143**, 114102 (2015).
- <sup>17</sup>P. C. Hiberty and S. Shaik, J. Comput. Chem. **28**, 137 (2007).
- <sup>18</sup>G. D. Purvis, R. Shephard, F. B. Brown, and R. J. Bartlett, Int. J. Quantum Chem. **23**, 835 (1983).
- <sup>19</sup>F. A. Evangelista, J. Chem. Phys. **134**, 224102 (2011).
- <sup>20</sup>P.-O. Löwdin, J. Math. Phys. **24**, 70 (1983).
- <sup>21</sup>P. Chernoff and J. Marsden, *Properties of Infinite Dimensional Hamiltonian Systems* (Springer, 1974).
- <sup>22</sup>H. Koch and P. Jørgensen, J. Chem. Phys. **93**, 3333 (1990).
- <sup>23</sup>J. Arponen, R. Bishop, and E. Pajanne, Phys. Rev. A **36**, 2539 (1987).
- <sup>24</sup>S. Kvaal, J. Chem. Phys. **136**, 194109 (2012).
- <sup>25</sup>T. Sato, H. Pathak, Y. Orimo, and K. Ishikawa, J. Chem. Phys. **148**, 051101 (2018).
- <sup>26</sup>T. Pedersen and S. Kvaal, J. Chem. Phys. **150**, 144106 (2019).
- <sup>27</sup>H. E. Kristiansen, Ø. S. Schøyen, S. Kvaal, and T. B. Pedersen, J. Chem. Phys. **152**, 071102 (2020).
- <sup>28</sup>T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, 2002).

- <sup>29</sup>S. Kvaal, A. Laestadius, and T. Bodenstein, “Guaranteed convergence for a class of coupled-cluster methods based on arponen’s extended theory,” (2020), to appear in Mol. Phys., arXiv:2003.06796 [physics.chem-ph].
- <sup>30</sup>T. Rohwedder, ESAIM: Math. Mod. Num. Anal. **47**, 421 (2013).
- <sup>31</sup>T. Rohwedder and R. Schneider, ESAIM: Math. Mod. Num. Anal. **47**, 1553 (2013).
- <sup>32</sup>A. Laestadius and F. M. Faulstich, Mol. Phys. **117**, 2362 (2019).
- <sup>33</sup>F. M. Faulstich, A. Laestadius, Ö. Legeza, R. Schneider, and S. Kvaal, SIAM J. Numer. Anal. **57**, 2579 (2019).
- <sup>34</sup>J. Arponen, R. Bishop, and E. Pajanne, Phys. Rev. A **36**, 2519 (1987).
- <sup>35</sup>L. Adamowicz, J.-P. Malrieu, and V. V. Ivanov, J. Chem. Phys. **112**, 10075 (2000).
- <sup>36</sup>B. Roos, R. Lindh, P.-Å. Malmqvist, V. Veryazov, and P.-O. Widmark, *Multiconfigurational Quantum Chemistry* (John Wiley & Sons, Inc., 2016).
- <sup>37</sup>J. Arponen, Theor. Chim. Acta **80**, 149 (1991).
- <sup>38</sup>R. Chowdhuri, D. Mukherjee, and M. Prasad, in *Aspects of Many-Body Effects in Molecules and Extended Systems*, Lecture Notes in Chemistry, Vol. 50, edited by D. Mukherjee (Springer, 1988) pp. 3–33.
- <sup>39</sup>M. Nooijen, K. Shamasundar, and D. Mukherjee, Mol. Phys. **103**, 2277 (2005).
- <sup>40</sup>H. Koch, H. J. Aa. Jensen, P. Jørgensen, and T. Helgaker, J. Chem. Phys. **93**, 3345 (1990).
- <sup>41</sup>P. J. Knowles and N. C. Handy, Comp. Phys. Comm. **54**, 75 (1989).
- <sup>42</sup>J. Gauss and J. F. Stanton, J. Chem. Phys. **103**, 3561 (1995).
- <sup>43</sup>E. Prochnow, F. A. Evangelista, H. F. Schaefer, W. D. Allen, and J. Gauss, J. Chem. Phys. **131**, 064109 (2009).
- <sup>44</sup>S. Hirata, J. Phys. Chem. A **107**, 9887 (2003).
- <sup>45</sup>E. Solomonik, D. Matthews, J. R. Hammond, J. F. Stanton, and J. Demmel, J. Parallel Distr. Com. **74**, 3176 (2014).
- <sup>46</sup>R. Schutski, J. Zhao, T. M. Henderson, and G. E. Scuseria, J. Chem. Phys. **147**, 184113 (2017).
- <sup>47</sup>P.-Å. Malmqvist, Int. J. Quant. Chem. **30**, 479 (1986).
- <sup>48</sup>S. Kähler and J. Olsen, J. Chem. Phys. **147**, 174106 (2017).
- <sup>49</sup>M. Hanauer and A. Köhn, J. Chem. Phys. **134**, 204111 (2011).
- <sup>50</sup>B. Jeziorski and H. J. Monkhorst, Phys. Rev. A **24**, 1668 (1981).

- <sup>51</sup>R. H. Myhre, J. Chem. Phys. **148**, 094110 (2018).
- <sup>52</sup>I. A. Zaporozhets, V. V. Ivanow, D. I. Lyakh, and L. Adamowicz, J. Chem. Phys. **143**, 024109 (2015).
- <sup>53</sup>P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
- <sup>54</sup>H.-J. Werner and P. J. Knowles, J. Chem. Phys. **89**, 5803 (1988).
- <sup>55</sup>F. A. Evangelista and J. Gauss, J. Chem. Phys. **134**, 114102 (2011).
- <sup>56</sup>P. K. Samanta and A. Köhn, J. Chem. Phys. **149**, 064101 (2018).
- <sup>57</sup>R. F. Fink and V. Staemmler, Theor. Chim. Acta **87**, 129 (1993).
- <sup>58</sup>S. Cabay and L. W. Jackson, SIAM J. Numer. Anal. **13**, 734 (1976).
- <sup>59</sup>A. Torokhti and P. Howlett, eds., “Computational methods for modelling of nonlinear systems,” (Elsevier, 2007) Chap. 7, pp. 291–378.
- <sup>60</sup>Y. Yao, G. H. Dong, X. Xiao, and C. P. Sun, Sci. Rep. **6**, 32010 (2016).
- <sup>61</sup>A. I. Krylov, J. Chem. Phys. **113**, 6052 (2000).
- <sup>62</sup>K. Jankowski, L. Meissner, and J. Wasilewski, Int. J. Quant. Chem. **28**, 931 (1985).
- <sup>63</sup>A. Engels-Putzka and M. Hanrath, Mol. Phys. **107**, 143 (2009).
- <sup>64</sup>U. Meier and V. Staemmler, Theor. Chim. Acta **76**, 95 (1989).
- <sup>65</sup>V. Staemmler, Theor. Chim. Acta **45**, 89 (1977).
- <sup>66</sup>T. Bodenstein, *Entwicklung und Anwendung von Multireferenzverfahren zur Beschreibung magnetischer Eigenschaften von Metallkomplexen*, Ph.D. thesis, Karlsruhe Institute of Technology (2015).
- <sup>67</sup>U. S. Mahapatra, B. Datta, and D. Mukherjee, J. Chem. Phys. **110**, 6171 (1999).
- <sup>68</sup>M. Kállay, P. G. Szalay, and P. R. Surján, J. Chem. Phys. **117**, 980 (2002).
- <sup>69</sup>M. Hanrath, J. Chem. Phys. **123**, 084102 (2005).
- <sup>70</sup>W. Chen and H. B. Schlegel, J. Chem. Phys. **101**, 5957 (1994).
- <sup>71</sup>J. F. Stanton, J. Chem. Phys. **101**, 371 (1994).
- <sup>72</sup>C. van Wüllen, J. Comp. Chem. **32**, 1195 (2011).
- <sup>73</sup>J. F. Stanton, J. Gauss, L. Cheng, M. E. Harding, D. A. Matthews, and P. G. Szalay, “CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package,” With contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, M. Hilgenberg, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, T. Kirsch, K. Klein, W.J. Lauderdale, F. Lipparini, T. Metzroth, L.A. Mück, D.P. O’Neill, D.R.

Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowitz, A. Tajti, 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>.

<sup>74</sup>T. Lee and G. Scuseria, in *Quantum Mechanical Electronic Structure Calculations With Chemical Accuracy*, edited by S. Langhoff (Kluwer Academic Publishers, Dordrecht, 1995) pp. 47–108.