



Molecular Physics

An International Journal at the Interface Between Chemistry and Physics

ISSN: 0026-8976 (Print) 1362-3028 (Online) Journal homepage: <http://www.tandfonline.com/loi/tmph20>

Reflections on size-extensivity, size-consistency and generalized extensivity in many-body theory

Marcel Nooijen , K. R. Shamasundar & Debashis Mukherjee

To cite this article: Marcel Nooijen , K. R. Shamasundar & Debashis Mukherjee (2005) Reflections on size-extensivity, size-consistency and generalized extensivity in many-body theory, Molecular Physics, 103:15-16, 2277-2298, DOI: [10.1080/00268970500083952](https://doi.org/10.1080/00268970500083952)

To link to this article: <https://doi.org/10.1080/00268970500083952>



Published online: 21 Feb 2007.



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



Article views: 482



Citing articles: 76 View citing articles [↗](#)

Reflections on size-extensivity, size-consistency and generalized extensivity in many-body theory

MARCEL NOOIJEN*[†], K. R. SHAMASUNDAR[†] and DEBASHIS MUKHERJEE[‡]

[†]Department of Chemistry, University of Waterloo, Waterloo, N2L 3G1 Ontario, Canada

[‡]Indian Association for the Cultivation of Science, Calcutta, India

(Received 1 August 2004; in final form 5 January 2005)

An overview is presented of the notions of size-extensivity and size-consistency, which play an important role in the discussion of many-body methods in physics and chemistry. We also introduce the concept of generalized extensivity, which has been used before implicitly in the literature, and provide an operational definition, which can be tested numerically in a similar way as size-consistency. A numerical example illustrates the concept of generalized extensivity as applied to the EOM-CCSD, STEOM-CCSD and extended-STEOM-CCSD electronic structure methods for excited states. In another line of thought it is argued that algebraic proofs are often more easily constructed than diagrammatic proofs to demonstrate proper separation into non-interacting fragments. The algebraic line of reasoning is illustrated for single reference methods and applied to discuss separability properties for a variety of Hilbert space multi-reference coupled cluster methods.

Keywords: Size-extensivity; Size-consistency; Generalized extensivity; Many-body theory

1. Introduction

Size-extensivity, size-intensivity and size-consistency, or in brief ‘scaling properties’, are important concepts in many-body physics and quantum chemistry [1–6]. In the solid state, where one deals with infinite systems, the proper scaling of computational methods is mandatory. Likewise, for the ever larger molecular systems that are being computed nowadays, proper scaling of the energy with the size of the system, referred to as size-extensivity, is vital.

The notion of size-extensivity [7, 8] is one of the recurring themes in the work of Rodney J. Bartlett as the distinguishing feature of many-body perturbation theory and coupled cluster methods. Diagrammatic techniques [9–12], another prominent feature of his work, e.g. [13, 14], play an important role in the theoretical developments, as it suffices to demonstrate the linkedness or connectedness of the quantity of interest to prove that it scales correctly. Size-extensivity is an important component of the ‘Systematic Treatment of Electron Correlation’ and in this paper we will reflect on some of the questions it continues to raise.

The twin notion of size-consistency [15], which indicates the additive separability of the energy for

a molecular system that consists of two physically separated, non-interacting, units, is another powerful scaling concept. It has the advantage that it is easily checked in practice, by performing a calculation on the compound system and the subunits separately, and comparing the energy of the compound system with the sum of energies of the isolated fragments. Size-consistency can therefore be given a simple operational definition. As a result, one can easily speak of size-consistency also for methods like density functional theory in which the notion of ‘connected’ does not have a meaning.

Many-body methods have the power to unify these different scaling definitions through the notion of connected and linked diagrams [6, 16–18]. However, it is fair to say that the proofs of connectedness are often rather non-trivial, in particular for multi-reference approaches, and even for the diagram aficionado. We will try to show here that scaling properties for non-interacting systems (or size-consistency) are often easier to demonstrate directly, without reference to the connectivity of the theory. Under certain conditions, to be discussed below, size-consistency implies size-extensivity, and then methods that satisfy the operational definition of size-consistency can also be expected to yield consistent accuracy for systems of increasing size. Diagrammatic analysis leads to the notion of generalized extensivity, which refers to the scaling of

*Corresponding author. Email: nooijen@uwaterloo.ca

the correlation energy if the orbitals are partitioned into two arbitrary subsets, and the fluctuation potential corresponding to mixed integrals is switched to zero. Like size-consistency, generalized extensivity can be given an operational definition and this provides a rigorous numerical test of the connectivity of the energy for general classes of theories based on second quantization. While at first sight the notions of size-consistency and generalized extensivity appear closely related, for multi-reference situations the partitioning of the orbital space into two separate subspaces is not related to the selection of the model space when discussing generalized extensivity, and therefore it provides a more stringent requirement than size-consistency.

In this paper we will first briefly discuss a variety of different situations where scaling properties are of interest and we will try to provide a historical perspective. The notions of size-extensivity and size-consistency are used very broadly in the literature and we prefer to speak more specifically of a physical quantity that is to scale properly in the context of a particular type of physical system or state (e.g. open or closed shell). In the remaining sections we will primarily focus on size-consistency, considering the case of a compound system consisting of two non-interacting units, possibly both open shell. We will show that it is often quite easy to show that the solutions for the isolated subsystem can be used to construct the solution for the compound system. One then needs to demonstrate that this reconstructed solution conforms to the parameterization of the wave function according to the particulars of the method, and that the working equations for the compound system are satisfied, given the solution for the isolated fragments. Using elementary operator techniques it is typically quite easy to show if a method behaves properly in this fashion. The proofs are transparent in our opinion and capture the basic physics of the situation: the presence of another fragment at infinity should not affect the basic properties for the system of interest.

We will see that, mathematically, the equations for the compound system either factorize into two products, each associated with an isolated subsystem, or the equations reduce to the equations for the isolated subsystem (the other factor effectively being unity). This latter condition conforms to the notion of connectedness and reflects the additive separability of the Hamiltonian. The factorization of equations reflects the multiplicative nature of the wave operator and it requires explicitly a multiplicative ansatz for the wave function. For this reason this picture is probably particularly suitable for Coupled Cluster type wave functions, while it does not hold, for example, for many-body perturbation theory. Both pictures have their

appeal though. A number of the results presented in this paper are well known and are repeated here for pedagogical reasons. Some new results (or new derivations of known properties) are included here also, in particular regarding multi-reference coupled cluster methods and the discussion of generalized extensivity. We will employ the water molecule with slightly stretched OH bond lengths to provide a numerical illustration of the concept of generalized extensivity and its operational definition as applied to selected coupled cluster methods to treat excited states.

2. General historical perspective on scaling and separability properties

The origins of many-body methods go back to the study of nuclear matter [19, 20], the electron gas [21], and infinite periodic systems [9]. The notions of size-extensivity and size-intensivity and the relation to connected diagrams were established in these fields [22, 23] and have been transferred to quantum chemistry [7, 8]. The notion of size-consistency [15] and the discussion of non-interacting systems is a chemical concept. Below, we will provide a brief summary of these concepts, the tools that are used to address the theoretical questions involved, and we will try to show the connection between the two concepts. We will distinguish between various forms of scaling properties (even if they are often referred to by the same name in the literature) and mention some of the limitations and open questions from a formal mathematical point of view. In section 2.3 we will provide an operational definition of generalized extensivity which provides the most stringent scaling requirements.

2.1. Scaling of properties for infinite periodic systems and very large molecules

- (i) The model system used to discuss *size-extensivity* is an infinite insulating periodic system with a closed-shell unit cell. Using a transparent counting argument it can be shown using k -space integration (where the number of k -points is related to the number of unit cells in the Born-von-Karman (BvK) repeating superunit) that connected diagrams scale proportionally to the number of units in the Born-von-Karman cell. The infinite BvK limit can be taken leading to a definite property per unit cell. This is nicely explained in the book by March *et al.* on many-body techniques [9]. Additional discussions and applications of these types of proof, in particular for inhomogeneous

periodic systems, can be found in relatively recent studies [17, 18].

- (ii) The basic model problem used to discuss *size-intensivity* concerns, for example, ionization potentials or excitation energies for similar periodic systems as under (i). Historically, Green's function or propagator techniques are used in the discussion [9, 10, 24]. Again, connected diagrams play a pivotal role, and the counting argument using k -space integration over complete bands is used [17, 18]. Linked diagrams for the propagator (no disconnected closed parts; every piece in the diagram carries uncontracted lines) imply size-intensive energy differences, meaning that the excitation energies become independent of the number of units in the BvK cell. The proof of the scaling is built upon the ground state energy discussed in (i) for a closed-shell unit (completely filled bands).
- (iii) We are not aware of a rigorous proof of scaling properties for metallic systems, in which one cannot integrate over completely filled bands. Open-shell systems are typically discussed in comparison to a ground state in which the unit cell is closed shell, as is the case under (ii).
- (iv) Large systems. It is quite hard to discuss or even define size-extensivity rigorously for a large but not periodic system. In order to take a limit, one has to grow a repeating unit in the form of a rigid chain and show that the change in energy becomes independent of the chain length in the limit. For example, we might consider the molecule $R_1-(CH_2)_n-R_2$, and define the energy E_n as the energy of the molecule with n geometrically identical CH_2 units. Size-extensivity then implies $\lim_{n \rightarrow \infty} (E_{n+1} - E_n) = C$, a constant. As far as we know, no formal proofs along these lines have been tried (and they would still not address the general molecular system which does not have a repeat unit). The same holds for size-intensive properties like ionization potentials which would become independent of the chain length in the limit. Let us emphasize here that one could have a site (or chromophor) in the system that itself is not repeated. The excitation localized on the chromophor should, however, become independent of the number of some repeat unit in the system in the limit.

Despite these somewhat hairsplitting formal issues there is little doubt in the community that methods that scale correctly for infinite periodic systems also scale correctly (whatever this means precisely) for mathematically less well-defined systems. In many-body theory the notion of size-extensivity is closely tied to connected

diagrams and it is a logical step therefore to use connectedness of the energy *as the definition* for size-extensivity in many-body theory. Such a definition applies to particular theoretical methods and does not refer to particular types of physical systems. Rodney J. Bartlett has been a strong advocate of this definition throughout his career and the papers where this definition originated are widely quoted [7, 8]. Likewise, the linked property of the many-body Green's function can be used to define size-intensivity [18]. There is no doubt (we think) that scaling properties translate properly to large but non-periodic systems, where we can take no meaningful mathematical infinite limit.

Let us note here that it is quite difficult to design a *practical* test for size-extensivity. Calculations of oligomers of increasing size have been reported (e.g. [17, 25–27]), although the interest is not so much a check of scaling properties but to identify when the periodic regime is reached. For example, at what size do ionization potentials become independent of the size of the system, or when does the increment in total energy, or the polarizability, become 'constant'? One would necessarily use methods that scale properly with size, and not use the calculations to test methods (but see [17]). Therefore, a formal theoretical definition of size-extensivity is required, rather than an operational one. Of course, this is also a limitation. Density functional theory is widely used in solid state physics and is clearly 'size-extensive', but the definition in terms of connected diagrams is not applicable to DFT. Likewise, few theoreticians in quantum chemistry will like our definition of a size-intensive quantity in terms of a linked Green's function: how does one decide then on the size-intensivity of a theory that is not cast in terms of Green's functions? It serves to prove the point that it is useful to have an operational definition that applies to all computational methods, and this leads us to the next topic.

2.2. Scaling of properties of non-interacting subsystems

The model system used to discuss size-consistency consists of a subsystem A in the presence of a subsystem B that is physically separated from system A such that no interaction exists between the two subsystems. The Hamiltonian of the compound system can be expressed as a direct sum operator $\hat{H}_{AB} = (\hat{H}_A \otimes \mathbf{1}_B) + (\mathbf{1}_A \otimes \hat{H}_B) \equiv \hat{H}_A + \hat{H}_B$, and the computed energy of the combined system should be the sum of the energies of the subsystems in isolation: $E_{AB} = E_A + E_B$. This is called additive separability [3]. The wave function of the compound system can be taken to be the antisymmetrized product of the wave functions in isolation, although this might violate certain symmetries,

e.g. spin-symmetry. If the computed wave function satisfies this antisymmetrized product property it is called multiplicatively separable [3]. There are many methods in quantum chemistry in which the energy scales properly (and hence also properties calculated as energy derivatives), but the underlying wave function (if there even is one, cf. DFT) does not. A good example is perturbation theory, where the energy is additively separable but the wave function is not multiplicatively separable.

The proper scaling of the energy is of most concern. Let us emphasize here that we have an *operational definition* of additive separability of the energy. Every computational method in quantum chemistry evaluates the (relative) energy of the system and a particular method can always be put to the test: calculate the energies for the compound system and the isolated subsystems and check the result.

In formal discussions of the fragment energies and wave functions it is convenient to use orbitals for the compound system that are localized on the individual fragments. It may happen that a method is *not* invariant under rotation of orbitals in specific subspaces (e.g. occupied and virtual orbitals), and this causes difficulties in the discussion. This is true, for example, for CEPA-type methods [3, 28], which are unambiguously defined only if the generation of the orbitals is specified. Such methods lacking certain orbital invariances can only satisfy the separability properties if they are based on localized orbitals.

It may also happen that one cannot localize the orbitals under the freedom of the given orbital invariances. An example is the separation of a closed-shell molecule in open-shell fragments using restricted Hartree–Fock orbitals. Such methods will typically fail to satisfy the separability property independent of other formal aspects. In these cases the reference energy itself is not size-consistent and this is typically not rectified by the inclusion of electron correlation (full CI is an exception of course). In the discussion below we will assume that one can, in principle, use localized orbitals, without affecting the result. This in turn implies that the reference state, or the model space in multi-reference situations, that is used in the many-body correlation approaches has the appropriate product character (to be discussed more precisely below). Three cases may be distinguished.

(i) *The states on fragments A and B are both closed shell (or the compound system is single reference open shell).* If the energy can be expressed in terms of connected diagrams it will be additively separable. The key issue is that the fragments and the compound system are both described at the single reference level. Hence the scaling behaviour holds for closed-shell systems but also, for example, for high spin single reference states on

the compound system. Of course, the energy may be additively separable even if there are no underlying connected diagrams. For example, the energy in DFT calculations is additively separable for two closed-shell subsystems. In the context of many-body theory the relation between additive separability and connected diagrams can be made even stronger. If the energy has a diagrammatic expansion and it is additively separable, the energy will necessarily consist of connected diagrams only [5, 29, 30]. Often this additive separability property is taken as *the definition* of size-extensivity [1, 5, 8], since it follows by induction that the energy is additively separable for an arbitrary number of non-interacting closed-shell fragments, if it is separable for two. However, let us emphasize here that this definition concerns extensivity regarding non-interacting systems, and this is different from the definitions that rely on a repeating unit in an *interacting* system, discussed previously. Because of the 1–1 relation to connected diagrams, if and only if a method satisfies the *practical* test of additive separability for two closed-shell systems, it will also be size-extensive. The situation is more complicated if (multi-reference) open-shell fragments are involved. Let us also refer to section 2.3 below for further discussion.

If the fragments are closed shell (or the compound system is single reference open shell) it means that one can always use orbitals that are localized on the fragments to describe the compound system, provided the methodology satisfies the necessary invariances. This is where the power of connected diagrams comes into play. Let us assume that matrix elements of the elementary operators that are the building blocks of the diagrams are all vanishing if they contain orbital labels from both fragments A and B. This notion is equivalent to the assumption that all elementary operators entering the diagram are connected, meaning that operator components with mixed orbital labels vanish identically. It then follows that all orbital labels in a connected diagram *have* to be located on the same fragment, in order to give a non-vanishing contribution. Connected diagrams carrying mixed labels will always vanish identically as this implies that at least one of the elementary building blocks in the diagram has mixed labels (for further discussion, see, for example, [31]). Therefore, the energy is just a sum over the contributions for system A alone, and that for system B alone, and the energy is additively separable. The same observation holds for all additively separable quantities: they have to consist of connected diagrams only. Not all observable quantities are connected in this sense, however. For example, the physical operator \hat{S}^2 is non-local, it does have non-vanishing mixed matrix elements, and the expectation value of \hat{S}^2 is not additively separable.

(ii) *One of the fragments is multi-reference open shell, the other fragment is closed shell.* Again the orbitals in the calculation can be localized on the fragments. If the energy is additively separable it means that, for example, excitation energies or ionization potentials on the MR open-shell fragment are size-intensive: they are independent of the presence of the fragment at infinity [32]. Moreover, it follows that the energies will be size-intensive also if an arbitrary number of non-interacting closed-shell systems are added to the system. From a diagrammatic perspective it is necessary and sufficient that the energy does not contain disconnected fully contracted parts. In the terminology of Mukherjee, such methods would be (at least) core-extensive [6, 16]. Again, one typically extrapolates the finding for the separable system and infers that properties are size-intensive also with respect to infinite periodic interacting systems. This can be proved for many-body methods, using diagrammatic techniques [18], but the notion would apply more generally, for example, to time-dependent DFT methods. Again there is a very simple practical test to see if a method is separable in this sense.

(iii) *Both fragments are multi-reference open shell.* This is the most difficult case. In order to perform a meaningful formal analysis, one will need to use localized orbitals, and the analysis will be invalid if the orbitals cannot be localized, without affecting the result. In practice, orbitals need not be localized if suitable orbital invariance relations hold, but it is important that, in principle, they can be (and without affecting the result). This separability condition into two open-shell fragments is relevant for the breaking of chemical bonds, for simultaneous excitations on a molecule, or in charge transfer excitations. The requirement that the energy scales properly for a compound system of two open-shell fragments puts a stringent test on many-body methods, and perhaps it is sufficient if methods that are used to describe such chemical phenomena satisfy the test to a 'large extent'. We do not consider this most strict test an absolute requirement, as it is not a necessary condition for methods to be accurate for large systems, for example.

The treatment of multi-reference open-shell systems typically involves a diagonalization step of a Hamiltonian-like operator over a suitably chosen 'model' space. This might be the bare Hamiltonian defined over a set of determinants that differ in their occupation of a set of active orbitals, such as, for example, in CASPT2, where perturbative corrections are applied subsequently to the zeroth-order multi-configurational wave function obtained from the diagonalization step. Alternatively, the operator to be diagonalized might be a similarity transformed

Hamiltonian, as is the case in Equation-of-Motion Coupled-Cluster theory. In this case, part of the dynamical correlation effects are included through the similarity transformation, while a mix of dynamical and non-dynamical correlation effects are obtained through diagonalization over a sizeable configuration space. A third alternative would involve the diagonalization of a small effective Hamiltonian that contains implicitly all of the effects of dynamical correlation. Diagonalization of the effective Hamiltonian includes primarily non-dynamical correlation effects. These various possibilities lead to a complicated array of formal scaling behaviours if the compound state of interest consists essentially of the antisymmetrized product of two multi-reference open-shell states on non-interacting subsystems.

If the diagonalization space is a direct product of the model spaces for the individual subsystems *and* the Hamiltonian-like operator used in the diagonalization is the direct sum of the subsystem operators, results in this step will be size-consistent. This is a sufficient condition but it is not a necessary requirement. It is also possible that the effective operator used in the diagonalization is designed to have a special structure that leads to size-consistent results for at least a selection of eigenvalues of the operator, even if the diagonalization space is not a direct product space. An intricate example of this situation is provided by the modification introduced for the Equation-of-Motion Coupled-Cluster Theory by Meissner and Bartlett [33], or, relatedly, extended Similarity Transformed Equation-of-Motion Coupled-Cluster theory (extended-STEOM) [34, 35].

It is tempting to assume that if a multi-reference theory satisfies the requirements that it separates correctly into two open-shell fragments *and* it separates correctly into an open-shell and a closed-shell fragment that the method will satisfy all desirable scaling properties. However, the proper separation of a system into two non-interacting fragments (either closed or open shell) is actually not sufficient as might be appreciated from the following counterexample: consider taking the bare Hamiltonian as the effective (additively separable) Hamiltonian over an incomplete model space. Diagonalizing the Hamiltonian over the direct product of the model spaces for a compound system would yield exactly size-consistent results (consider, for example, our demonstration for the CID method to be discussed in section 3). This is true even if one would consider the separation into a multi-dimensional and one-dimensional model space, which would satisfy the separability condition into closed- and open-shell systems. However, such a method does not necessarily satisfy generalized extensivity (to be discussed below) or even size-extensivity. For example, we might take

as the model space for the open-shell fragment all singles and double excitations out of a single reference determinant and the reference determinant itself. The method (i.e. CISD) is not size-extensive, but, by a *suitable choice* of model space for the compound system, this method will satisfy all imposed separability conditions. This observation indicates that we have to be quite careful to draw conclusions for multi-reference methods with general model spaces that might be construed to satisfy separability conditions, *but do not necessarily satisfy extensivity*.

2.3. Generalized extensivity for arbitrary partitioning of the orbitals into two non-interacting subsets

As indicated above, size-consistency is a very useful concept because it has a simple operational definition. The same is not true for size-extensivity, and moreover we found it impossible to give a rigorous, physically motivated, definition of size-extensivity for a general non-periodic system. It is possible to generalize the non-interacting concept for second quantized Hamiltonians such that it allows a more general and *operational* definition of extensivity. The definition below has been used implicitly in work by Mukherjee *et al.* [6, 16], but the precise definition, or its operational nature, has not been emphasized in the literature.

Let us consider a Hamiltonian expressed in normal order with respect to some reference determinant $|\Phi_0\rangle$

$$H = h_0 + \sum_{p,q} f_q^p \{\hat{p}^\dagger \hat{q}\} + \frac{1}{4} \sum_{p,q,r,s} V_{rs}^{pq} \{\hat{p}^\dagger \hat{q}^\dagger \hat{r} \hat{s}\},$$

where the curly brackets indicate that the second quantized operators act in normal order [12]. In the above, $h_0 = \langle \Phi_0 | \hat{H} | \Phi_0 \rangle$ is the energy of the reference determinant, while f_q^p is the corresponding Fock matrix. Let us further assume for simplicity that the Fock matrix is diagonal $f_q^p = \varepsilon_p \delta_q^p$, and let us define a zeroth-order Hamiltonian $H_0 = h_0 + \sum_{p,q} \varepsilon_p \{\hat{p}^\dagger \hat{p}\}$ and a perturbation $\hat{V} = (1/4) \sum_{p,q,r,s} V_{rs}^{pq} \{\hat{p}^\dagger \hat{q}^\dagger \hat{r} \hat{s}\}$. To discuss the concept of generalized extensivity we partition the orbitals into two *arbitrary* subsets R and S. Using this partitioning we can define the perturbation to consist of three parts

$$\begin{aligned} \hat{V} &= \hat{V}_R \oplus \hat{V}_S \oplus \hat{V}_{RS}, \\ \hat{V}_R &= \frac{1}{4} \sum_{p,q,r,s \in R} V_{rs}^{pq} \{\hat{p}^\dagger \hat{q}^\dagger \hat{r} \hat{s}\}, \\ \hat{V}_S &= \frac{1}{4} \sum_{p,q,r,s \in S} V_{rs}^{pq} \{\hat{p}^\dagger \hat{q}^\dagger \hat{r} \hat{s}\}, \\ \hat{V}_{RS} &= \hat{V} - \hat{V}_R - \hat{V}_S. \end{aligned}$$

In order to check for generalized extensivity we set all mixed matrix elements that contain orbitals from both sets R and S, which define \hat{V}_{RS} , to zero and we perform four calculations using precisely the same method, reference space and so forth. First use the Hamiltonian \hat{H}_0 which defines the energy $E_{RS}^0 = E_R^0 + E_S^0$ for a particular symmetry and spin adapted eigenstate $|\Phi_{RS}^0\rangle$ of \hat{H}_0 . Let us emphasize that $|\Phi_{RS}^0\rangle$ can be multi-determinantal, but it would correspond to a single configuration state function in CI language. The zeroth-order energies are just given by the sum of the Hartree–Fock orbital energies of the occupied (spatial) orbitals in $|\Phi_{RS}^0\rangle$. In the second calculation we use the Hamiltonian $\hat{H}_0 + \hat{V}_R$, and converge to the state that is closest to $|\Phi_{RS}^0\rangle$. In particular, the occupation of the orbitals in subsystem S would remain unchanged. This would yield energy E_{R/S_0} . Next use $\hat{H}_0 + \hat{V}_S$ defining the energy $E_{R^0/S}$, and finally use $\hat{H}_0 + \hat{V}_R + \hat{V}_S$ which yields E_{RS} . Operationally these calculations can simply be done by setting the required two-electron integrals to zero, and homing in on a correlated state that is initially described by $|\Phi_{RS}^0\rangle$. A method would satisfy the requirement of general extensivity if and only if the total energies satisfy $E_{RS} = E_{R/S_0} + E_{R^0/S} - E_{RS}^0$, for *arbitrary* partitions of the orbitals into two subsets R and S. This requirement is a very strong demand on many-body theories as the *model spaces* in multi-reference theories, for example, are *not* related to the arbitrary orbital partitioning. In particular, the model space may not be a direct product space for arbitrary orbital partitionings. In practice, some care has to be exercised to define an ‘arbitrary’ orbital partitioning that is properly discriminating. Obviously, it is not very productive to consider *all* possible partitionings, but one should consider the worst possible scenario. If a method satisfies the test under (any) such conditions it satisfies generalized extensivity. If there are orbital partitions where at least some of the eigenvalues fail the test, the method does not satisfy generalized extensivity, at least for some of the roots. It is possible that only certain roots of a particular character will always satisfy the test, independent of the orbital partitioning. Such is the case, for example, for the principle IPs in IP-EOMCCSD calculations: they satisfy generalized extensivity. Shake-up states having 2h–1p character that would involve orbitals from both partition R and partition S would violate the test [16]. In general, therefore, one can only establish generalized extensivity for a certain class of roots of the equation.

In the test for generalized extensivity the model system is mathematically rather than physically defined. A difference with size-consistency is that the nuclear–electron part is not affected in the generalized extensivity test, while in the size-consistency test one discards the

physical system B completely when calculating system A in isolation. In the generalized extensivity test the partitioning is made by a division of the orbitals, and interactions are only switched off after the mean field effects are fully incorporated. This strategy is motivated primarily by pragmatic considerations. In this way it is well possible to carry out the test, even if the zeroth-order energies are very different from the correlated energies. A numerical example is presented in section 5.

A physically motivated example that illustrates the concept concerns the separation of core and valence contributions to correlation energies. The orbitals are partitioned into two subsets: one set of orbitals (occupied and virtual) that describes core-correlation effects and another set of orbitals that describes valence correlation effects. If a method satisfies generalized extensivity the correlation energy rigorously separates into two parts upon neglecting core–valence interactions. In fact, for this property to hold the weaker condition of core-extensivity is already sufficient [16]. This example shows that the notion of extensivity is important already for ‘small’ systems: we can meaningfully speak of core-extensivity of the energy for atoms, for example. Moreover, this type of extensivity is very important for the accuracy of a method. The accuracy of CISD methods significantly deteriorates with the size of the atom under consideration, while CCSD methods are generally unaffected.

If a method is extensive in the general sense it means that if the energy is expressed as a perturbation series with respect to a single determinantal vacuum, the energy will necessarily consist of connected diagrams only. Single reference CC theory and many-body perturbation theory satisfy generalized extensivity, but equation-of-motion coupled-cluster theory does not, even though (some of) the excitation energies are size-consistent and the method is size-intensive for extended systems. Likewise, the notion of general extensivity puts a strong requirement on multi-reference CC theories, and theories that are called extensive, or additively separable by Mukherjee *et al.* [6, 16] or by Meissner *et al.* [33, 36], really satisfy the generalized extensivity concept. To the best of our knowledge, generalized extensivity has never been demonstrated numerically for multi-reference coupled-cluster methods, and we plan to provide examples in the near future that illustrate the concept and which will establish if this notion, that motivated much of the formal work in the field, is numerically important. In section 5 we will illustrate generalized extensivity for coupled-cluster methods for excited states. The generalized extensivity concept is addressed in a general fashion, but the conclusions that can be drawn from these examples are similar to the scaling properties for charge-transfer excitations,

which have been discussed before in the literature (e.g. [37]). The notion of a charge-transfer excitation over a very long distance is a somewhat artificial construct, and the importance of generalized extensivity as a criterion for the accuracy of a method has not been widely appreciated, we believe. In section 5 we show that generalized extensivity tests indicate that the property is quite important and the violation of this property can be used to rationalize the somewhat poorer results that are obtained from EOM-CCSD calculations for valence excited states.

Let us briefly reiterate the different types of scaling properties and the definitions used in this paper. Size-extensivity and size-intensivity refer to properties that scale linearly or become independent of the size of the system in the thermodynamic limit: we need to consider an infinite system with a well-defined repeat unit to have a well-defined mathematical definition. Size-consistency refers loosely to the additive separability of the energy for states describing physically separated, non-interacting systems. We can typically speak only of size-consistency for particular classes of solutions (states) of a particular method/calculation. Equation-of-Motion Coupled-Cluster theory is a clear example of the subtleties that can occur. The choice of model space and so forth is typically defined precisely for the non-interacting system of interest, if size-consistency is tested in practice. The above descriptions of size-extensivity and size-consistency apply to quite general models used in quantum chemistry (e.g. *ab initio* wave function methods, explicitly correlated methods, density functional theory, or semi-empirical approaches). Generalized extensivity, the third scaling criterion, in contrast applies only to methods that use a methodology based on a one-particle product basis to describe many-electron states, at least using the present definitions. If it is possible to make a formal diagrammatic perturbation expansion of the energy based on a single determinantal vacuum state, the energy would consist of connected diagrams only, if (and only if) a theory is to satisfy generalized extensivity. Generalized extensivity automatically implies that size-extensivity is satisfied, and that size-consistency is satisfied if the vacuum determinant referred to above can be expressed as a single determinant in terms of localized orbitals (the usual condition in connection with physically separated non-interacting systems). The operational definition of generalized extensivity can be applied to any method based on a one-particle product basis and requires that a particular energy is additively separable under an arbitrary partitioning of the orbitals, while keeping all other computational parameters fixed (model space, excitation manifold, operators included, and so forth). The roots to be compared in the various

calculations that are needed to validate generalized extensivity, $E_{RS} = E_{R/S_0} + E_{R_0/S} - E_{RS}^0$, have to be commensurate. We think this seldom presents a problem in practice, but a rigorous mathematical definition of what it means to be ‘commensurate’ is somewhat hard to give.

In the following two sections we focus mainly on size-consistency or additive separability of the energy for physically separated, non-interacting systems. This property is most straightforward to analyse by purely algebraic means, and it is perhaps the most intuitively obvious property to be discussed for molecular systems. Our analysis is based on the use of the techniques of second quantization, as in this fashion the antisymmetry requirement of the wave function can be incorporated in a trivial manner. This does present a limitation as the concept of size-consistency is certainly more generally applicable.

3. Outline of a general algebraic strategy for proving separability of the energy and amplitude equations, and consideration of some simple examples

The orbital/operator space is partitioned into two non-interacting subspaces A and B. Non-interacting means that all one- and two-electron integrals that involve indices from both subsystems are necessarily zero. This definition implies that we use orbitals that are localized on the subsystems. Next, we assume we satisfy a set of working equations for the isolated subsystems A and B, and we use these solutions to try to show that in this case also the equation for the composite system is satisfied with the same values for the wave function parameters and energies. If the method is separable in this sense, the proof typically follows in a straightforward fashion. No explicit use is made of the notions of connected and linked, and these concepts typically do not enter the proofs, which are phrased in a completely algebraic way.

The proofs are most conveniently expressed through operator expressions evaluated over the true vacuum. Then no issues arise with antisymmetry, due to the use of second quantization (see e.g. [31]). This requires perhaps some further discussion. It is often said that the wave function for non-interacting systems is multiplicatively separable. However, if the wave function is required to be antisymmetric this is certainly not true in the strict mathematical sense of the word. However, using the language of second quantization we can say that the wave function might be expressed as $\hat{A}\hat{B}|vac\rangle$, where \hat{A} is a general second quantized operator containing annihilation and creation operators corresponding to orbitals on subsystem A only, and likewise for operator \hat{B} . Hence at the operator level, the wave

function is described as product: antisymmetry is built in automatically.

In our discussion below we will use as compact a notation as possible, since we will need to indicate the subsystem, and this would quickly render a full notation rather cumbersome, obscuring the basic simplicity of the manipulations. Let us consider a well-known example to illustrate the basic mechanism and notation.

3.1. Size-consistency for Single Reference Coupled-Cluster Theory (SRCC)

Let us consider the analysis of the single reference coupled-cluster method in some detail as a pedagogical example. The generic SRCC equations are expressed as

$$\begin{aligned} \langle \hat{\mu}^\dagger \hat{x} e^{-\hat{T}} \hat{H} e^{\hat{T}} \hat{\mu} \rangle &= 0, \quad \forall \hat{x}, \\ E &= \langle \hat{\mu}^\dagger e^{-\hat{T}} \hat{H} e^{\hat{T}} \hat{\mu} \rangle. \end{aligned} \quad (1)$$

In our notation $\hat{\mu}|vac\rangle = |\Phi_\mu\rangle$ defines the reference determinant, and $\hat{\mu}$ denotes a string of creation operators. The symbol \hat{x} denotes a de-excitation operator with respect to the reference determinant, and the cluster operator is implicitly defined as $\hat{T} = \sum_x t_x \hat{x}^\dagger$. We use a somewhat compressed notation to facilitate the separability analysis. The symbol $\forall x$ is to be loosely interpreted as ‘for all proper operators associated with a particular approximation’. We hope the loss of accuracy due to this notation is compensated by an increased transparency and focus on essentials.

From the generic equations we extract equations for the subsystems A and B as follows:

$$\begin{aligned} \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_A} \hat{H}_A e^{\hat{T}_A} \hat{\mu}_A \rangle &= 0, \quad \forall x_A \\ E_A &= \langle \hat{\mu}_A^\dagger e^{-\hat{T}_A} \hat{H}_A e^{\hat{T}_A} \hat{\mu}_A \rangle \\ \text{and } \langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_B} \hat{H}_B e^{\hat{T}_B} \hat{\mu}_B \rangle &= 0, \quad \forall x_B \\ E_B &= \langle \hat{\mu}_B^\dagger e^{-\hat{T}_B} \hat{H}_B e^{\hat{T}_B} \hat{\mu}_B \rangle. \end{aligned} \quad (2)$$

The subsystem operators are used to express the operators for the compound equations:

$$\hat{\mu}_{AB} = \hat{\mu}_A \hat{\mu}_B, \quad \hat{H}_{AB} = \hat{H}_A + \hat{H}_B, \quad \hat{T}_{AB} = \hat{T}_A + \hat{T}_B. \quad (3)$$

This means that the wave function for the compound system

$$\begin{aligned} e^{\hat{T}_{AB}} \hat{\mu}_{AB} |vac\rangle &= e^{\hat{T}_A + \hat{T}_B} \hat{\mu}_A \hat{\mu}_B |vac\rangle \\ &= (e^{\hat{T}_A} \hat{\mu}_A) (e^{\hat{T}_B} \hat{\mu}_B) |vac\rangle \end{aligned}$$

is multiplicatively separable in the operator sense as discussed above. Here we use $e^{\hat{T}_A + \hat{T}_B} = e^{\hat{T}_A} e^{\hat{T}_B}$. The cluster operators on the different subsystems commute as they involve different orbitals, and have an even number of annihilation/creation operators. The cluster operator \hat{T}_B also commutes with $\hat{\mu}_A$, independent of the number of creation operators in the string $\hat{\mu}_A$, and conversely for the operator \hat{T}_A and operator string $\hat{\mu}_B$. Substituting the relevant parts of equation (3), the equations for the compound system AB are given by

$$\begin{aligned} \langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A e^{-\hat{T}_A} e^{-\hat{T}_B} (\hat{H}_A + \hat{H}_B) e^{\hat{T}_A} e^{\hat{T}_B} \hat{\mu}_B \hat{\mu}_A \rangle &= 0, \quad \forall x_A, \\ \langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_A} e^{-\hat{T}_B} (\hat{H}_A + \hat{H}_B) e^{\hat{T}_A} e^{\hat{T}_B} \hat{\mu}_B \hat{\mu}_A \rangle &= 0, \quad \forall x_B, \\ \langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A \hat{x}_B e^{-\hat{T}_A} e^{-\hat{T}_B} (\hat{H}_A + \hat{H}_B) e^{\hat{T}_A} e^{\hat{T}_B} \hat{\mu}_B \hat{\mu}_A \rangle &= 0, \quad \forall x_A, x_B, \\ E_{AB} &= \langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger e^{-\hat{T}_A} e^{-\hat{T}_B} (\hat{H}_A + \hat{H}_B) e^{\hat{T}_A} e^{\hat{T}_B} \hat{\mu}_B \hat{\mu}_A \rangle. \end{aligned} \quad (4)$$

The manifold of excitation operators is split into three parts: \hat{x}_A , excitations on system A only; \hat{x}_B , excitations on system B only; and mixed excitations $\hat{x}_A \hat{x}_B$. The latter are actually more numerous than needed in say CCSD, but it is easily shown that all of these matrix elements vanish, if the subsystem equations are satisfied. In the proof we will only use that the operators that belong to different subsystems commute. All operators have an even number of second quantized operators, except, possibly, the strings of operators that define the reference state. We might assume here for convenience that these also have an even number of operators, and commute therefore. In the actual manipulations we will always pull $\hat{\mu}_A$ through both $\hat{\mu}_B^\dagger \hat{\mu}_B$ such that there are no sign worries and the proofs are generally valid. In the vacuum expectation values only fully contracted expressions remain and they can always be partitioned into two groups: the operators on system A contract among each other and the operators on system B do likewise. It then follows that

$$\begin{aligned} &\langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A e^{-\hat{T}_A} e^{-\hat{T}_B} (\hat{H}_A + \hat{H}_B) e^{\hat{T}_A} e^{\hat{T}_B} \hat{\mu}_B \hat{\mu}_A \rangle \\ &= \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_A} \hat{H}_A e^{\hat{T}_A} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger e^{-\hat{T}_B} e^{\hat{T}_B} \hat{\mu}_B \rangle \\ &\quad + \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_A} e^{\hat{T}_A} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger e^{-\hat{T}_B} \hat{H}_B e^{\hat{T}_B} \hat{\mu}_B \rangle \\ &= \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_A} \hat{H}_A e^{\hat{T}_A} \hat{\mu}_A \rangle = 0, \quad \forall x_A. \end{aligned} \quad (5)$$

The reduction for the equations on system B is of course analogous, while for the mixed equations we obtain

$$\begin{aligned} &\langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A \hat{x}_B e^{-\hat{T}_A} e^{-\hat{T}_B} (\hat{H}_A + \hat{H}_B) e^{\hat{T}_A} e^{\hat{T}_B} \hat{\mu}_B \hat{\mu}_A \rangle \\ &= \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_A} \hat{H}_A e^{\hat{T}_A} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_B} e^{\hat{T}_B} \hat{\mu}_B \rangle \\ &\quad + \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_A} e^{\hat{T}_A} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_B} \hat{H}_B e^{\hat{T}_B} \hat{\mu}_B \rangle \\ &= \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_A} \hat{H}_A e^{\hat{T}_A} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B \hat{\mu}_B \rangle \\ &\quad + \langle \hat{\mu}_A^\dagger \hat{x}_A \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_B} \hat{H}_B e^{\hat{T}_B} \hat{\mu}_B \rangle = 0, \quad \forall x_A, x_B. \end{aligned} \quad (6)$$

Each of the terms in these equations is zero for two reasons: each of the factors vanishes. For completeness we should also consider mixed excitation operators that do not preserve the number of electrons on each fragment. It is easily seen that in both factors there are no fully contracted terms and there will always be remaining uncontracted annihilation or creation operators, and the vacuum expectation value necessarily vanishes. Since this is generally the case we will not again consider mixed operators of this type.

Finally, the energy reduces to

$$\begin{aligned} E_{AB} &= \langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger e^{-\hat{T}_A} e^{-\hat{T}_B} (\hat{H}_A + \hat{H}_B) e^{\hat{T}_A} e^{\hat{T}_B} \hat{\mu}_A \hat{\mu}_B \rangle \\ &= \langle \hat{\mu}_A^\dagger e^{-\hat{T}_A} \hat{H}_A e^{\hat{T}_A} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger e^{-\hat{T}_B} e^{\hat{T}_B} \hat{\mu}_B \rangle \\ &\quad + \langle \hat{\mu}_A^\dagger e^{-\hat{T}_A} e^{\hat{T}_A} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger e^{-\hat{T}_B} \hat{H}_B e^{\hat{T}_B} \hat{\mu}_B \rangle \\ &= E_A + E_B. \end{aligned} \quad (7)$$

Let us reiterate the basic steps in this proof as we will use the technique a number of times below. We start from a generic equation, which immediately defines the subsystem equations. These subsystem equations are assumed solved, and we use the solutions in the equations for the compound system and construct (i.e. use an educated guess for) the compound solution, assuming that one does not need mixed operators, or, occasionally, that these can be expressed in terms of the solution for the subsystem. This is the basic underlying principle in the analysis of non-interacting subsystems: the isolated fragments should provide all the information. The only difficult step might be the construction of the solution for the compound system. All remaining parts of the proof are simple rearrangements that solely rely on the validity of the commutation rules.

The above formulation of coupled-cluster theory would lead to connected diagrams only. We can formulate the CC equations in an alternative way, and the diagrammatic proof of size-consistency would then require a judicious cancellation of disconnected

terms. Using the present means of analysis the proof is equally straightforward. The generic algebraic equation reads

$$\begin{aligned} \langle \hat{\mu}^\dagger \hat{x}(\hat{H} - E)e^{\hat{T}} \hat{\mu} \rangle &= 0, \quad \forall x, \\ \langle \hat{\mu}^\dagger (\hat{H} - E)e^{\hat{T}} \hat{\mu} \rangle &= 0. \end{aligned} \quad (8)$$

Let us consider the compound equation with mixed excitation operators

$$\begin{aligned} &\langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A \hat{x}_B (\hat{H}_A + \hat{H}_B - E_A - E_B) e^{\hat{T}_A} e^{\hat{T}_B} \hat{\mu}_B \hat{\mu}_A \rangle \\ &= \langle \hat{\mu}_A^\dagger \hat{x}_A (\hat{H}_A - E_A) e^{\hat{T}_A} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B e^{\hat{T}_B} \hat{\mu}_B \rangle \\ &\quad + \langle \hat{\mu}_A^\dagger \hat{x}_A e^{\hat{T}_A} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B (\hat{H}_B - E_B) e^{\hat{T}_B} \hat{\mu}_B \rangle = 0. \end{aligned} \quad (9)$$

In this case only the factors that include the Hamiltonian vanish. The other factor has some non-zero value. This is true if \hat{x}_A is a de-excitation operator, but also if it is the unit operator (defining the energy equation). This single mixed equation therefore contains all relevant cases if we interpret the de-excitation operator in a slightly more general way (de-excitation on A only ($\hat{x}_B = 1$), de-excitation on subsystem B only ($\hat{x}_A = 1$), de-excitation on both, and finally the energy equation ($\hat{x}_A = \hat{x}_B = 1$)). The crucial aspect is that the result is a sum of terms that can be written as factors involving the operators on system A and system B, and that one of the factors always contains the subsystem equations. It is the direct product form of the compound equations that does the trick. No cancellation of disconnected terms is required. This analysis is not meant to suggest that the analysis in terms of connected diagrams is not useful. There is simply more than one way to arrive at the desired result.

3.2. Configuration interaction with double excitations only (CID)

Let us use these techniques to prove that CID is *not* separable. The generic CID equations can be phrased as follows, using intermediate normalization and with a suitable interpretation of the operators involved:

$$\begin{aligned} \langle \hat{\mu}^\dagger \hat{x}(\hat{H} - E)(1 + \hat{C}) \hat{\mu} \rangle &= 0, \quad \forall x, \\ E &= \langle \hat{\mu}^\dagger \hat{H}(1 + \hat{C}) \hat{\mu} \rangle. \end{aligned} \quad (10)$$

If we use that for the compound system $\hat{C}_{AB} = \hat{C}_A + \hat{C}_B + \hat{C}_A \hat{C}_B$ then the equations *would* be separable. However, this means that mixed quadruple

excitations have to be included. Adding and subtracting this quadratic quadruple term to the CID equations isolates the error in the original equations. Let us consider the case of de-excitation on fragment A:

$$\begin{aligned} &\langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A (\hat{H}_A + \hat{H}_B - E_A - E_B) \\ &\quad \times (1 + \hat{C}_A + \hat{C}_B + \hat{C}_A \hat{C}_B - \hat{C}_A \hat{C}_B) \hat{\mu}_B \hat{\mu}_A \rangle \\ &= \langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A (\hat{H}_A + \hat{H}_B - E_A - E_B)(1 + \hat{C}_A)(1 + \hat{C}_B) \hat{\mu}_B \hat{\mu}_A \rangle \\ &\quad - \langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A (\hat{H}_A + \hat{H}_B - E_A - E_B)(\hat{C}_A \hat{C}_B) \hat{\mu}_B \hat{\mu}_A \rangle \\ &= \langle \hat{\mu}_A^\dagger \hat{x}_A (\hat{H}_A - E_A)(1 + \hat{C}_A) \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B (1 + \hat{C}_B) \hat{\mu}_B \rangle \\ &\quad + \langle \hat{\mu}_A^\dagger \hat{x}_A (1 + \hat{C}_A) \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B (\hat{H}_B - E_B)(1 + \hat{C}_B) \hat{\mu}_B \rangle \\ &\quad - \langle \hat{\mu}_A^\dagger \hat{x}_A (\hat{H}_A - E_A) \hat{C}_A \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{C}_B \hat{\mu}_B \rangle \\ &\quad - \langle \hat{\mu}_A^\dagger \hat{x}_A \hat{C}_A \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger (\hat{H}_B - E_B) \hat{C}_B \hat{\mu}_B \rangle \\ &= \langle \hat{\mu}_A^\dagger \hat{x}_A \hat{C}_A \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger (\hat{H}_B - E_B) \hat{\mu}_B \rangle. \end{aligned} \quad (11)$$

As can be seen, proving that methods are *not* size-consistent can be more involved than showing that they are, using the present technique. It is not claimed that this is a particularly transparent proof, and perhaps a fair amount of hindsight was needed to construct it. We also deliberately considered the case of CID (rather than CISD), as then the ideal solution for the wave function restricted to doubles only is easily reconstructed.

Let us consider one more example case: equation-of-motion coupled-cluster theory [38, 39] or, equivalently, Coupled-Cluster Linear Response Theory [32, 40, 41]. This theory has the interesting property that it scales correctly if one considers as subsystems one ground state and one excited state, but it does not scale correctly if excited states are considered on both subsystems [32]. These states include so-called singly excited charge transfer states in which the excitation energy should reduce to the sum of the EOMCC ionization potential on one fragment and the EOMCC electron affinity on the other, as discussed for example in [33, 37]. Meissner and Bartlett have shown how EOMCC can be modified to restore additive separability of singly excited charge-transfer excitation energies. The different scaling properties for different types of excitation indicate that one definition of ‘size-consistency’ really does not suffice. One has to consider the various cases individually, as they may behave differently.

Another confusing issue concerns size-intensity/size-extensivity. For extended periodic systems the

results from CCLRT (which uses a size-intensive definition for transition moments in contrast to EOMCC [42]) scale properly, for example because the equations can also be derived from a Green's function or propagator point of view [43, 44]. The linked diagram approximation for the Coupled-Cluster Green's function then ensures proper scaling of energies and transition moments [18]. This is in apparent conflict with the statement that excitation energies in EOMCC or CCLRT are not 'extensive' as charge-transfer excitations are not additively separable. The problem lies in the use of the word extensive for different properties. The excitation energy in EOMCC contains disconnected diagrams when analysed from a single reference perspective, and EOMCC violates generalized extensivity, even though excitation energies are size-intensive for infinite systems. As mentioned before we prefer to define various scaling properties in the context of specific physical/mathematical model systems, as these definitions are most clear-cut. The theoretical notion of connected diagrams can then be used to demonstrate and understand the various aspects of scaling issues. It is interesting to observe that the EOMCC method can be used to illustrate the subtleties of such a wide variety of scaling requirements.

Let us restrict ourselves here to the basic demonstration that EOMCC excitation energies do not depend on the presence of another non-interacting fragment in its ground state. The generic EOMCC equations can be stated as the CC equations considered before, and in addition the EOM equations for excited states:

$$\langle \hat{\mu}^\dagger \hat{x} e^{-\hat{T}} (\hat{H} - E) e^{\hat{T}} \hat{C} \hat{\mu} \rangle = 0, \forall x. \quad (12)$$

Considering the compound system where fragment A is excited, while fragment B is in the ground state, we find for the energy $E_{AB} = E_A + E_B^0$, where E_B^0 denotes the coupled-cluster energy on fragment B. Let us examine the mixed equation for the compound system, as this contains all the information we need

$$\begin{aligned} & \langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A \hat{x}_B e^{-\hat{T}_A} e^{-\hat{T}_B} (\hat{H}_A + \hat{H}_B - E_A - E_B^0) \\ & \quad \times e^{\hat{T}_A} e^{\hat{T}_B} \hat{C}_A \hat{\mu}_B \hat{\mu}_A \rangle \\ &= \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_A} (\hat{H}_A - E_A) e^{\hat{T}_A} \hat{C}_A \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_B} e^{\hat{T}_B} \hat{\mu}_B \rangle \\ &+ \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_A} e^{\hat{T}_A} \hat{C}_A \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_B} (\hat{H}_B - E_B^0) e^{\hat{T}_B} \hat{\mu}_B \rangle. \end{aligned} \quad (13)$$

We can consider three cases. If the operator $\hat{x}_B = 1$ then we find the EOMCC equations from the first term, while

the second term vanishes as E_B^0 equals the CC ground state energy for system B. If, on the other hand, $\hat{x}_A = 1$, while \hat{x}_B is a true excitation, then the first term vanishes due to the orthogonality condition on system B, while the second equation contains the CC equations for system B. It can be verified that the energy E_B^0 does not contribute to the last factor in the equation, which shows this factor is equivalent to the SRCC equation for system B. Likewise, for the truly mixed term it is easily verified that at least one of the factors in each term is zero. As was the case for the alternative CC approach, the factorization of each term of the equation into subsystem parts is the crucial simplification. It reflects the direct product structure of the non-interacting problem.

We can argue in a similar fashion as we did for the CI doubles method that EOMCC is not separable if *both* subsystems are excited. In this case, quadruples are needed to define the product wave function. For the case of singly excited charge-transfer states, one would require triple excitations: 2 hole–1 particle operators on the IP site and 2 particle–1 hole operators on the EA fragment. As in the case of the CI doubles equations, to actually demonstrate that the method is not size-consistent is more troublesome. We refer to the literature [16, 32].

4. Consideration of separability conditions for multi-reference Hilbert space Coupled-Cluster methods in an incomplete model space

In this section we will consider separability issues for Hilbert space multi-reference CC methods [45]. In the literature, explicit proofs are typically constructed [6, 45, 46], but often these proofs are rather non-trivial, *and* they focus on generalized extensivity in the sense of connected diagram expansion of the energy in a single reference framework. Here we focus instead on size-consistency of the energies and using the present operator technique the proofs are fairly transparent in our opinion. In addition, we will emphasize that size-consistency and generalized extensivity are really different requirements for the case of two open-shell fragments. Generalized extensivity implies size-consistency provided that the compound model space is the direct product of the subsystem model spaces, analogous to the condition of separability of the reference determinant in the single reference case.

Let us consider the Jeziorski–Monkhorst equations [45] in our current notation:

$$\langle \hat{\mu}^\dagger \hat{x} e^{-\hat{T}_\mu} \hat{H} e^{\hat{T}_\mu} \hat{\mu} \rangle - \sum_\nu \langle \hat{\mu}^\dagger \hat{x} e^{-T_\nu} e^{T_\nu} \hat{\nu} \rangle H_{\nu\mu}^{\text{eff}} = 0, \quad \forall \mu, x. \quad (14)$$

The symbols $\hat{\mu}, \hat{\nu}$ indicate strings of creation operators that generate the corresponding reference determinants when acting on the vacuum. The determinants $|\Phi_\mu\rangle = \hat{\mu}|vac\rangle$ span the multi-dimensional model space, while \hat{x} indicates a de-excitation operator that is used in the projection manifold. The cluster operator $\hat{T}_\mu = \sum_k t_k^\mu \hat{x}_k^\dagger$ typically excites out of the model space onto single and double excitations from the reference determinant $\hat{\mu}|vac\rangle$. For incomplete model spaces the cluster operator may in addition contain components that excite within the model space, depending on the precise formulation. The energies in the Jeziorski–Monkhorst approach are obtained from a projected equation involving a Hamiltonian and a metric matrix (this will be a useful formulation to discuss the incomplete model space):

$$\sum_\lambda \langle \hat{\mu}^\dagger \hat{H} e^{T_\lambda} \hat{\lambda} \rangle_{c_\lambda} = \sum_\lambda \langle \hat{\mu}^\dagger e^{T_\lambda} \hat{\lambda} \rangle_{c_\lambda} E, \quad (15)$$

or, equivalently,

$$\sum_\lambda \langle \hat{\mu}^\dagger (\hat{H} - E) e^{T_\lambda} \hat{\lambda} \rangle_{c_\lambda} = 0. \quad (16)$$

This equation can be used to define the effective Hamiltonian that enters the amplitude equations. If we define

$$\begin{aligned} \tilde{H}_{\mu\lambda} &= \langle \Phi_\mu | \hat{H} e^{\hat{T}_\lambda} | \Phi_\lambda \rangle, \\ S_{\mu\lambda} &= \langle \Phi_\mu | e^{\hat{T}_\lambda} | \Phi_\lambda \rangle = \langle \Phi_\mu | e^{-\hat{T}_\mu} e^{\hat{T}_\lambda} | \Phi_\lambda \rangle, \end{aligned} \quad (17)$$

the effective Hamiltonian is defined as $H_{\lambda\mu}^{\text{eff}} = (S^{-1} \tilde{H})_{\lambda\mu}$. Diagonalizing this effective Hamiltonian provides the same energies and model space eigenfunctions as obtained from the generalized eigenvalue problem. For complete model spaces, or when intermediate normalization is imposed, the overlap matrix \mathbf{S} is the unit matrix, while the effective Hamiltonian reduces to the traditional expression $H_{\lambda\mu}^{\text{eff}} = \tilde{H}_{\lambda\mu} = \langle \Phi_\lambda | \hat{H} e^{\hat{T}_\mu} | \Phi_\mu \rangle$.

For the compound system consisting of two non-interacting fragments the model space is defined to be the direct product of the fragment model spaces, and is spanned by the determinants $\hat{\mu}_A \hat{\mu}_B |vac\rangle$. The crucial property that has to be satisfied by the compound effective Hamiltonian is that it is the direct sum of the fragment effective Hamiltonians $H_{AB}^{\text{eff}} = H_A^{\text{eff}} + H_B^{\text{eff}}$, or, more precisely, the matrix elements of the effective Hamiltonian need to satisfy

$$H_{\mu_A, \mu_B, \lambda_A, \lambda_B}^{\text{eff}} = H_{\mu_A, \lambda_A}^{\text{eff}} \delta_{\mu_B, \lambda_B} + \delta_{\mu_A, \lambda_A} H_{\mu_B, \lambda_B}^{\text{eff}}. \quad (18)$$

If this is the case it easily follows that the energies of the compound system are size-consistent, while the model space wave functions are multiplicatively separable:

$$\begin{aligned} & \sum_{\lambda_A, \lambda_B} H_{\mu_A, \mu_B, \lambda_A, \lambda_B}^{\text{eff}} c_{\lambda_A} c_{\lambda_B} \\ &= \sum_{\lambda_A, \lambda_B} \left(H_{\mu_A, \lambda_A}^{\text{eff}} \delta_{\mu_B, \lambda_B} + \delta_{\mu_A, \lambda_A} H_{\mu_B, \lambda_B}^{\text{eff}} \right) c_{\lambda_A} c_{\lambda_B} \\ &= \left(\sum_{\lambda_A} H_{\mu_A, \lambda_A}^{\text{eff}} c_{\lambda_A} \right) c_{\mu_B} + c_{\mu_A} \left(\sum_{\lambda_B} H_{\mu_B, \lambda_B}^{\text{eff}} c_{\lambda_B} \right) \\ &= (E_A + E_B) c_{\lambda_A} c_{\mu_A}. \end{aligned} \quad (19)$$

To demonstrate that this is indeed the important aspect for size-consistency, we will show that the Hilbert space MRCC equations for the compound system are satisfied if the effective Hamiltonian is additively separable, independent of the precise definition of the cluster operators, or normalization condition. Later we will show that the effective Hamiltonian is indeed additively separable for three different variants of Hilbert space MRCC.

Let us consider the equations for the composite system obtained by projection against the mixed excitation $\hat{x}_A \hat{x}_B$, as this can be used to deduce all proper subcases of the equations

$$\begin{aligned} & \langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A \hat{x}_B e^{-\hat{T}_{A\mu}} e^{-\hat{T}_{B\mu}} (\hat{H}_A + \hat{H}_B) e^{\hat{T}_{A\mu}} e^{\hat{T}_{B\mu}} \hat{\mu}_A \hat{\mu}_B \rangle \\ & - \sum_{\nu_A, \nu_B} \langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A \hat{x}_B e^{-T_{\mu A}} e^{-T_{\mu B}} e^{T_{\nu A}} e^{T_{\nu B}} \hat{\nu}_A \hat{\nu}_B \rangle \\ & \times H_{\nu_A, \nu_B, \mu_A, \mu_B}^{\text{eff}} = 0. \end{aligned} \quad (20)$$

This yields upon inserting the additively separable form of the effective Hamiltonian followed by the usual rearrangements:

$$\begin{aligned} & \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_{A\mu}} \hat{H}_A e^{\hat{T}_{A\mu}} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_{B\mu}} \hat{H}_B e^{\hat{T}_{B\mu}} \hat{\mu}_B \rangle \\ & + \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_{A\mu}} e^{\hat{T}_{A\mu}} \hat{\mu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_{B\mu}} \hat{H}_B e^{\hat{T}_{B\mu}} \hat{\mu}_B \rangle \\ & - \sum_{\nu_A, \nu_B} \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-T_{\mu A}} e^{T_{\nu A}} \hat{\nu}_A \rangle \langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_{B\mu}} e^{\hat{T}_{B\nu}} \hat{\nu}_B \rangle \\ & \times \left(H_{\nu_A, \mu_A}^{\text{eff}} \delta_{\nu_B, \mu_B} + \delta_{\nu_A, \mu_A} H_{\nu_B, \mu_B}^{\text{eff}} \right) \\ & = \left(\langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_{A\mu}} \hat{H}_A e^{\hat{T}_{A\mu}} \hat{\mu}_A \rangle \right. \\ & \quad \left. - \sum_{\nu_A} \langle \hat{\mu}_A^\dagger \hat{x}_A e^{-T_{\mu A}} e^{T_{\nu A}} \hat{\nu}_A \rangle H_{\nu_A, \mu_A}^{\text{eff}} \right) \langle \hat{\mu}_B^\dagger \hat{x}_B \hat{\mu}_B \rangle \\ & + \langle \hat{\mu}_A^\dagger \hat{x}_A \hat{\mu}_A \rangle \left(\langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_{B\mu}} \hat{H}_B e^{\hat{T}_{B\mu}} \hat{\mu}_B \rangle \right. \\ & \quad \left. - \sum_{\nu_B} \langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_{B\mu}} e^{\hat{T}_{B\nu}} \hat{\nu}_B \rangle H_{\nu_B, \mu_B}^{\text{eff}} \right). \end{aligned} \quad (21)$$

It is seen that the compound equations are satisfied if the equations for the subsystem in isolation are satisfied. This observation holds also in the case $\hat{x}_A = 1$, or $x_B = 1$, and even in the case that $\hat{x}_A = \hat{x}_B = 1$. It is easily verified that the equations also obey size-consistency if one would start from the linked version of the equations

$$\langle \hat{\mu}^\dagger \hat{x} \hat{H} e^{\hat{T}_\mu} \hat{\mu} \rangle - \sum_v \langle \hat{\mu}^\dagger \hat{x} e^{T_v} \hat{v} \rangle H_{v\mu}^{\text{eff}} = 0, \quad \forall \mu, x, \quad (22)$$

and we obtain the analogous situation to the single reference CC equations, which were analysed in two different ways. The above shows that the Hilbert space CC equations are satisfied for the compound system in terms of the solutions for the isolated fragments, provided the effective Hamiltonian is additively separable and the compound model space is the direct product of the fragment model spaces.

The remaining issue to consider is this additive separability of the effective Hamiltonian, which depends on some further details in the defining equations. We will consider three variations on the theme of general model space Hilbert space MRCC (GMS-HS-MRCC). In the first variant, which has not been explicitly discussed in the literature to the best of our knowledge, the cluster operators do not contain any excitations within the model space: $\langle \Phi_\mu | \hat{T}_\lambda | \Phi_\lambda \rangle = 0, \forall \mu, \lambda$. This generally implies that for the incomplete model space intermediate normalization does not hold $S_{\mu\lambda} = \langle \Phi_\mu | e^{T_\lambda} | \Phi_\lambda \rangle \neq \delta_{\mu\lambda}$. The effective Hamiltonian is defined as discussed above: $H_{\lambda\mu}^{\text{eff}} = (S^{-1} \hat{H})_{\lambda\mu}$. The second approach is discussed in the original paper by Jeziorski and Monkhorst [45], and has recently been extensively investigated by Li and Paldus [47–49]. The definition of the cluster operator is extended such that the wave operator satisfies intermediate normalization: $S_{\mu\lambda} = \langle \Phi_\mu | e^{\hat{T}_\lambda} | \Phi_\lambda \rangle = \delta_{\mu\lambda}$. This means that there may be elements in the cluster operator that excite between model space determinants and their coefficients exactly negate the product of disconnected excitation operators, each of which typically excite out of the model space. A simple example would be provided by the two-dimensional model space $|a\bar{a}\rangle$ and $|b\bar{b}\rangle$. The single excitations with coefficients $t_a^b, t_a^{\bar{b}}$ excite out of the model space when acting on $|a\bar{a}\rangle$, but their product leads precisely to the second model space determinant. To impose intermediate normalization, Li and Paldus include the additional cluster operator that excites between the model space determinants with coefficient $t_{a\bar{a}}^{b\bar{b}} = -t_a^b t_a^{\bar{b}}$. In this way they maintain intermediate normalization, but the cluster operator contains manifestly disconnected parts. Li and Paldus refer to this definition of the cluster amplitudes that can excite

within the model space as ‘satisfying the C condition’. For large model spaces this prescription may require extending the cluster operator with potentially high rank excitation operators in order to satisfy this C condition, or, equivalently, intermediate normalization, which can make the implementation of the general scheme rather cumbersome. Since the overlap matrix S in intermediate normalization is unity by definition, the effective Hamiltonian reduces to $H_{\mu\lambda}^{\text{eff}} = \langle \Phi_\mu | H e^{T_\lambda} | \Phi_\lambda \rangle$, which is used in the Jeziorski–Monkhorst and Li and Paldus papers.

A third version of a GMS Hilbert space MRCC theory has been proposed by Meissner, Kucharski and Bartlett [36, 46] and independently by Mukherjee *et al.* [50, 51], both following work by Mukherjee *et al.* for the Fock space approach in an incomplete model space [52, 53]. The difference with the above variants concerns partly the definition of the effective Hamiltonian, which is defined as follows:

$$\sum_\lambda \langle \hat{v}^\dagger e^{-T_\mu} e^{T_\lambda} \hat{\lambda} \rangle H_{\lambda\mu}^{\text{eff}} = \langle \hat{v}^\dagger e^{-T_\mu} \hat{H} e^{T_\mu} \hat{\mu} \rangle, \quad \forall v, \mu. \quad (23)$$

This equation is very similar to the equation used in the first variant

$$\sum_\lambda \langle \hat{v}^\dagger e^{T_\lambda} \hat{\lambda} \rangle H_{\lambda\mu}^{\text{eff}} = \langle \hat{v}^\dagger \hat{H} e^{T_\mu} \hat{\mu} \rangle, \quad \forall v, \mu,$$

yielding slight differences only if $\langle \Phi_v | e^{-T_\mu}$ contains determinants that were not present in the original projection manifold (for example, singles and double excitations with respect to $|\Phi_\mu\rangle$ plus the model space itself). We will show below that any of the above definitions yields an additively separable effective Hamiltonian and size-consistent energies.

Let us first discuss a second difference in the Meissner/Mukherjee approach that pertains to the notion of generalized extensivity and which provided the motivation underlying the approach. In general, connectivity and additive separability of the effective Hamiltonian are not sufficient for the more stringent property of generalized extensivity, as we discussed in section 2.3 using the example of the direct product of the CISD model space for two non-interacting subsystems. This issue was recognized a long time ago and much work has been done by Mukherjee [5, 6, 52, 53], Meissner [36, 46, 54] and others to illustrate this aspect, and to overcome this difficulty. In their work on the general model space Hilbert space Coupled-Cluster method, Meissner/Mukherjee impose additional equations on the effective Hamiltonian: all ‘extensivity-violating’ components in the effective Hamiltonian corresponding to quasi-open operators are required to

be zero, and additional ‘extensivity-correcting’ cluster operators of the same second-quantized form are needed in order to satisfy these additional conditions. A second quantized operator in $H_{\lambda\mu}^{\text{eff}}$ is extensivity-violating if the result when acting on $|\Phi_\mu\rangle$ lies within the model space, but the result would lie outside the model space by acting with the second quantized operator on at least one of the other model determinants. These operators in the effective Hamiltonian are responsible for violating the generalized extensivity of the energies upon diagonalization of the effective Hamiltonian. If the extensivity-violating operators in $H_{\lambda\mu}^{\text{eff}}$ are set to zero, it means the effective Hamiltonian is a so-called closed operator: acting with any of the second quantized operators in $H_{\lambda\mu}^{\text{eff}}$ (with non-zero amplitude) on any model determinant cannot excite out of the model space. The extensivity-violating operators are a subset of the so-called ‘quasi-open’ operators, which are operators that contain exclusively active orbital annihilation and creation operators, but which excite outside of the model space by acting on at least one of the model space determinants. Those quasi-open operators which excite *within* the model space when acting on some model space determinants are responsible for the violation of generalized extensivity. The quasi-open concept is based on an active orbital space and an underlying complete model space. This concept is not strictly needed to discuss extensivity issues, and using the terminology of general model spaces one does not introduce active orbitals, only the model space determinants [47].

The cluster operator in the Meissner/Mukherjee approach is extended with corresponding extensivity-correcting operators that are identical in form as the components equated to zero in the effective Hamiltonian. The additional cluster amplitudes provide the freedom to equate the extensivity-violating components of the effective Hamiltonian to zero. Moreover, inclusion of these additional cluster operators results in solutions for the cluster amplitudes that are connected, or satisfy generalized extensivity. The full theory is a delicate balance, such that generalized extensivity is satisfied, independent of the partitioning of the orbitals in non-interacting subsets, and independent of the direct product structure of the model space. The cluster operator, including the extensivity-correcting quasi-open components, contains components that excite within the model space as in the Li–Paldus theory, but it does not satisfy intermediate normalization in general. This method is different from the previous two schemes. Meissner *et al.* [46] and Mukherjee *et al.* [50] have shown explicitly using diagrammatic techniques that this formulation of Hilbert space MRCC leads to a connected Hamiltonian that satisfies generalized extensivity. The effective Hamiltonian would be additively separable,

independently of the chosen partitioning of the orbitals. Moreover, they demonstrated that the energies satisfy generalized extensivity: upon diagonalization of the effective Hamiltonian the total energy can be expressed in terms of connected diagrams only. This connectedness of the total energy is based on a single reference description of the state, in the traditional way. This is a much stronger condition than size-consistency of the energy. Below we will show that the extensivity-correcting operators in the wave operator are not essential to ensure the additive separability of the effective Hamiltonian or size-consistency (which assumes that the model space is a direct product space). They are important for generalized extensivity in the connected diagram sense as discussed by Meissner *et al.* [36, 46] and Mukherjee *et al.* [50, 51]. In practice, the condition that the extensivity-violating components of the effective Hamiltonian are equated to zero may mean that the corresponding quasi-open cluster amplitudes take on rather large values, and this may deteriorate the accuracy of the results, even if the approach is rigorously extensive.

Let us return to our algebraic discussion of size-consistency. As noted before, the crucial aspect in connection to size-consistency is the additive separability of the effective Hamiltonian. For the first two approaches the effective Hamiltonian for the compound system is defined through

$$\begin{aligned}\tilde{H}_{\mu A, \mu B, \lambda A, \lambda B} &= \left\langle \mu_A^\dagger \mu_B^\dagger (\hat{H}_A + \hat{H}_B) e^{T_{\lambda A}} e^{T_{\lambda B}} \lambda_B \lambda_A \right\rangle \\ &= \left\langle \mu_A^\dagger \hat{H}_A e^{T_{\lambda A}} \lambda_A \right\rangle \left\langle \mu_B^\dagger e^{T_{\lambda B}} \lambda_B \right\rangle \\ &\quad + \left\langle \mu_A^\dagger e^{T_{\lambda A}} \lambda_A \right\rangle \left\langle \mu_B^\dagger \hat{H}_B e^{T_{\lambda B}} \lambda_B \right\rangle \\ &= \tilde{H}_{\mu A, \lambda A} S_{\mu B, \lambda B} + S_{\mu A, \lambda A} \tilde{H}_{\mu B, \lambda B}, \\ S_{\mu A, \mu B, \lambda A, \lambda B} &= \left\langle \mu_A^\dagger \mu_B^\dagger e^{T_{\lambda A}} e^{T_{\lambda B}} \lambda_B \lambda_A \right\rangle \\ &= \left\langle \mu_A^\dagger e^{T_{\lambda A}} \lambda_A \right\rangle \left\langle \mu_B^\dagger e^{T_{\lambda B}} \lambda_B \right\rangle = S_{\mu A, \lambda A} S_{\mu B, \lambda B}.\end{aligned}\tag{24}$$

It then follows that the effective Hamiltonian is given by

$$\begin{aligned}H_{\mu A, \mu B, \lambda A, \lambda B}^{\text{eff}} &= S_{\mu A, \mu B, \nu A, \nu B}^{-1} \tilde{H}_{\nu A, \nu B, \lambda A, \lambda B} \\ &= H_{\mu A, \lambda A}^{\text{eff}} \delta_{\mu B, \lambda B} + \delta_{\mu A, \lambda A} H_{\mu B, \lambda B}^{\text{eff}}.\end{aligned}\tag{25}$$

The above relation states that for a compound non-interacting system the effective Hamiltonian is additively separable: $H_{AB}^{\text{eff}} = H_A^{\text{eff}} + H_B^{\text{eff}}$.

Let us also show algebraically that the Meissner/Mukherjee definition of the effective Hamiltonian satisfies the property that the compound effective Hamiltonian is the direct sum of the fragment effective Hamiltonians. Let us re-emphasize that this property

does not depend on the additional conditions for the extensivity-violating components of the Hamiltonian. The latter conditions are required to prove connectedness of the energy and generalized extensivity. The additivity of the effective Hamiltonian is most easily shown by postulating this structure of the effective Hamiltonian and showing the consistency of the hypothesis:

$$\begin{aligned} & \sum_{\lambda} \left\langle \hat{v}_A^{\dagger} \hat{v}_B^{\dagger} e^{-T_{\mu A}} e^{-T_{\mu B}} e^{T_{\lambda A}} e^{T_{\lambda B}} \hat{\lambda}_A \hat{\lambda}_B \right\rangle \\ & \quad \times \left(H_{\lambda A \mu A}^{\text{eff}} \delta_{\lambda B \mu B} + \delta_{\lambda A \lambda A} H_{\lambda B \mu B}^{\text{eff}} \right) \\ & \quad - \left\langle \hat{v}_A^{\dagger} \hat{v}_B^{\dagger} e^{-T_{\mu A}} e^{-T_{\mu B}} \hat{H} e^{T_{\mu A}} e^{T_{\mu B}} \hat{\mu}_A \hat{\mu}_B \right\rangle = 0, \quad (26) \end{aligned}$$

or

$$\begin{aligned} & \sum_{\lambda A, \lambda B} \left\langle \hat{v}_A^{\dagger} e^{-T_{\mu A}} e^{T_{\lambda A}} \hat{\lambda}_A \right\rangle H_{\lambda A \mu A}^{\text{eff}} \left\langle \hat{v}_B^{\dagger} e^{-T_{\mu B}} e^{T_{\lambda B}} \hat{\lambda}_B \right\rangle (\delta_{\lambda B \mu B}) \\ & \quad - \left\langle \hat{v}_A^{\dagger} e^{-T_{\mu A}} \hat{H}_A e^{T_{\mu A}} \hat{\mu}_A \right\rangle \left\langle \hat{v}_B^{\dagger} e^{-T_{\mu B}} e^{T_{\mu B}} \hat{\mu}_B \right\rangle \\ & \quad + \sum_{\lambda A, \lambda B} \left\langle \hat{v}_A^{\dagger} e^{-T_{\mu A}} e^{T_{\lambda A}} \hat{\lambda}_A \right\rangle \delta_{\lambda A \mu A} \left\langle \hat{v}_B^{\dagger} e^{-T_{\mu B}} e^{T_{\lambda B}} \hat{\lambda}_B \right\rangle (H_{\lambda B \mu B}^{\text{eff}}) \\ & \quad - \left\langle \hat{v}_A^{\dagger} e^{-T_{\mu A}} e^{T_{\mu A}} \hat{\mu}_A \right\rangle \left\langle \hat{v}_B^{\dagger} e^{-T_{\mu B}} \hat{H}_B e^{T_{\mu B}} \hat{\mu}_B \right\rangle = 0, \end{aligned}$$

or

$$\begin{aligned} & \left(\sum_{\lambda A} \left\langle \hat{v}_A^{\dagger} e^{-T_{\mu A}} e^{T_{\lambda A}} \hat{\lambda}_A \right\rangle H_{\lambda A \mu A}^{\text{eff}} - \left\langle \hat{v}_A^{\dagger} e^{-T_{\mu A}} \hat{H}_A e^{T_{\mu A}} \hat{\mu}_A \right\rangle \right) \delta_{\nu B \mu B} \\ & \quad + \delta_{\nu A \mu A} \left(\sum_{\lambda B} \left\langle \hat{v}_B^{\dagger} e^{-T_{\mu B}} e^{T_{\lambda B}} \hat{\lambda}_B \right\rangle H_{\lambda B \mu B}^{\text{eff}} \right. \\ & \quad \left. - \left\langle \hat{v}_B^{\dagger} e^{-T_{\mu B}} \hat{H}_B e^{T_{\mu B}} \hat{\mu}_B \right\rangle \right) = 0, \quad (27) \end{aligned}$$

which shows that indeed the direct sum form of the effective Hamiltonian satisfies the equation for the compound effective Hamiltonian in the Meissner/Mukherjee formulation.

Let us note that the above proof for the general incomplete model space includes the case that one of the fragments is open shell, while the second fragment is closed shell, or, more generally, has a one-dimensional model space. It is sufficient that the Hilbert space MRCC equations reduce to the single reference CC equations for the one-dimensional model space. It then follows that all three size-consistency variations discussed in section 2 hold for all three realizations of the Hilbert space CC method.

Li and Paldus [55] very recently proved the above separability property for their GMS Hilbert space MRCC approach, using a virtually identical technique

as is used in this paper. However, their proof is based on the use of intermediate normalization. Here we have shown that the theory can be formulated more generally, although one does need to adjust the definition of the effective Hamiltonian accordingly, if intermediate normalization does not hold. On the contrary, Mukherjee *et al.* have stressed [5, 6, 53] that intermediate normalization *has* to be abandoned for general incomplete model spaces in order to satisfy generalized extensivity, while keeping the Jeziorski–Monkhorst parameterization of the wave function. This is, however, a far stronger requirement than size-consistency.

The results presented in this paper and in the paper by Li and Paldus [55] clearly show that the effective Hamiltonian and the cluster amplitudes are additively separable for non-interacting systems (in the conventional algebraic sense used in this paper), and this is also unambiguously supported by numerical results [55]. The proof of additive separability for the cluster amplitudes and effective Hamiltonian explicitly assume that the model space is a direct product space, however. Additive separability should not be expected to hold if *arbitrary* partitions of the orbital space are made, which are not related to the choice of model space. In general, the effective Hamiltonian may contain disconnected diagrams and correspondingly might not satisfy generalized extensivity, meaning that the effective Hamiltonian itself is not additively separable for arbitrary partitionings of the orbital space. Even if we assume for the sake of argument that the cluster amplitudes would be connected, the effective Hamiltonian, in general, does contain diagrammatically disconnected terms, as indicated below

$$\begin{aligned} \hat{H}_{\lambda \mu}^{\text{eff}} &= \left\langle \lambda \left| \hat{H} e^{\hat{T}_{\mu}} \right| \mu \right\rangle = \left\langle \lambda \left| e^{\hat{T}_{\mu}} e^{-\hat{T}_{\mu}} \hat{H} e^{\hat{T}_{\mu}} \right| \mu \right\rangle \\ &= \left\langle \lambda \left| e^{\hat{T}_{\mu}} \hat{H}^{(\mu)} \right| \mu \right\rangle. \quad (28) \end{aligned}$$

It is possible that $\hat{H}^{(\mu)}$ excites out of the model space, while $e^{\hat{T}_{\mu}}$ brings the result back to the model space. These are clearly disconnected contributions, and there is no reason for them to vanish in general. This conclusion is in apparent conflict with the diagrammatic analysis presented by Paldus *et al.* [56], who phrase the theory in terms of connected diagrams only. This contradiction, however, can be attributed to a graphical representation of the projection operator as a topologically connected quantity in their work, rather than as a set of open lines, which would lead to disconnected contributions to the effective Hamiltonian.

It is somewhat surprising perhaps that the effective Hamiltonian and cluster amplitudes are additively

separable in regard to the physically non-interacting system, while they do contain manifestly disconnected diagrams. Some insight is provided by the cluster operators themselves. Even though the extra cluster operators included to maintain intermediate normalization are disconnected (see our specific example of intermediate normalization in the Li–Paldus theory), they *are* localized on one subsystem. This follows because the overlap matrix for the compound system factorizes

$$S_{\lambda A, \lambda B, \mu A, \mu B} = \left\langle \hat{\lambda}_A^\dagger e^{\hat{T}_{\mu A}} \hat{\mu}_A \right\rangle \left\langle \hat{\lambda}_B^\dagger e^{\hat{T}_{\mu B}} \hat{\mu}_B \right\rangle = \delta_{\lambda A, \mu A} \delta_{\lambda B, \mu B}.$$

It follows that there are no cluster amplitudes with mixed labels. There is a special qualification to the cluster operator such that the labels do not run indiscriminately over all possible values, and this leads to unfamiliar territory for diagrammatic analysis. From a physical point of view the components of the operator \hat{T} are always localized on one subsystem and they might be considered ‘connected’ from this point of view. Likewise, each component of the effective Hamiltonian is always localized on one subsystem, even if it is diagrammatically a disconnected quantity. Diagrammatic analysis does not help in this case to establish size-consistency, which is better proven by algebraic means.

Irrespective of the additive separability or even connectivity of the effective Hamiltonian and cluster operators in the Li–Paldus theory, the effective Hamiltonian may contain extensivity-violating components that are not explicitly equated to zero, and we anticipate that the energies obtained by diagonalizing the effective Hamiltonian will not satisfy generalized extensivity. The property of generalized extensivity is more amenable to diagrammatic analysis and requires operators to be connected.

For completeness, the proper separability of the state selective MRCC approach described by Mahapatra *et al.* [57] can also be demonstrated in virtually the same manner. The generic equations are given by [57]

$$\left\langle \hat{\mu}^\dagger \hat{x} e^{-\hat{T}_\mu} \hat{H} e^{\hat{T}_\mu} \hat{\mu} \right\rangle c_\mu + \sum_{\nu \neq \mu} \left\langle \hat{\mu}^\dagger \hat{x} e^{-T_\mu} e^{T_\nu} \hat{\mu} \right\rangle H_{\mu\nu}^{\text{eff}} c_\nu = 0, \quad \forall \mu, x, \quad (29)$$

while we use the same equations as before to define the effective Hamiltonian. The effective Hamiltonian should properly be referred to as an intermediate Hamiltonian here [58–62], as only one root of the equations is required to be accurate [57]. Mahapatra *et al.* initially restricted themselves to a complete model space, but similar strategies as in the Jeziorski–Monkhorst Hilbert

space approach can be used to define the general model space variant [63]. The compound equations read

$$\begin{aligned} & \left\langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A \hat{x}_B e^{-\hat{T}_{\Lambda\mu}} e^{-\hat{T}_{B\mu}} (\hat{H}_A + \hat{H}_B) e^{\hat{T}_{\Lambda\mu}} e^{\hat{T}_{B\mu}} \hat{\mu}_A \hat{\mu}_B \right\rangle c_{\mu A} c_{\mu B} \\ & - \sum_{\nu A, \nu B \neq \mu A, \mu B} \left\langle \hat{\mu}_A^\dagger \hat{\mu}_B^\dagger \hat{x}_A \hat{x}_B e^{-T_{\mu A}} e^{-T_{\mu B}} e^{T_{\nu A}} e^{T_{\nu B}} \hat{\mu}_A \hat{\mu}_B \right\rangle \\ & \times H_{\mu A \mu B, \nu A \nu B}^{\text{eff}} c_{\nu A} c_{\nu B} = 0, \end{aligned} \quad (30)$$

or

$$\begin{aligned} & \left(\left\langle \hat{\mu}_A^\dagger \hat{x}_A e^{-\hat{T}_{\Lambda\mu}} \hat{H}_A e^{\hat{T}_{\Lambda\mu}} \hat{\mu}_A \right\rangle \right. \\ & - \sum_{\nu A \neq \mu A} \left\langle \hat{\mu}_A^\dagger \hat{x}_A e^{-T_{\mu A}} e^{T_{\nu A}} \hat{\mu}_A \right\rangle H_{\mu A, \nu A}^{\text{eff}} \left. \right) c_{\mu A} \left\langle \hat{\mu}_B^\dagger \hat{x}_B \hat{\mu}_B \right\rangle c_{\mu B} \\ & + \left\langle \hat{\mu}_A^\dagger \hat{x}_A \hat{\mu}_A \right\rangle c_{\mu A} \left(\left\langle \hat{\mu}_B^\dagger \hat{x}_B e^{-\hat{T}_{B\mu}} \hat{H}_B e^{\hat{T}_{B\mu}} \hat{\mu}_B \right\rangle c_{\mu B} \right. \\ & \left. - \sum_{\nu B \neq \mu B} \left\langle \hat{\mu}_B^\dagger \hat{x}_B e^{-T_{\mu B}} e^{T_{\nu B}} \hat{\mu}_B \right\rangle H_{\mu B, \nu B}^{\text{eff}} \left. \right) c_{\nu B} = 0. \end{aligned} \quad (31)$$

It is easily verified that the approach yields size-consistent results following essentially the same argument as before. Mukherjee *et al.* have demonstrated [57] that the approach satisfies generalized extensivity also, provided that extensivity-violating components of the effective Hamiltonian are equated to zero.

5. Numerical test results of generalized extensivity

In this section we will illustrate the generalized extensivity test discussed in section 2.3 by performing several different types of calculations of the excitation energies of the water molecule. In order to clearly show the effects, the OH bond in water is stretched a little, to ensure that correlation effects and extensivity errors are significant, when they occur. The water molecule in the calculations has C_{2v} symmetry and is defined by an OH bond length of 1.50 Å and a HOH angle of 105°. We used Sadlej’s polarized basis set [64], containing 42 basis functions in total. To test generalized extensivity the orbitals are partitioned into two subsets. Subset R contains all orbitals of a_1 and a_2 symmetry, while subset S contains the orbitals of b_1 and b_2 symmetry. After the Hartree–Fock calculation and the transformation of the integrals, all two-electron integrals that involve orbitals from both subsets are artificially set to zero in all of the calculations discussed below. In principle, the orbital subsets could be chosen arbitrarily, but grouping the orbitals according to their symmetry labels facilitates the discussion. In order to test

generalized extensivity we perform four calculations for a given method:

1. Keep all two-electron integrals, except the mixed R–S integrals which have been set to zero already ($\rightarrow E_{RS}$).
2. Set all two-electron integrals that involve orbitals in subset S to zero ($\rightarrow E_{R/S_0}$).
3. Set all two-electron integrals that involve orbitals in subset R to zero ($\rightarrow E_{R_0/S}$).
4. Set all two-electron integrals to zero, which defines \hat{H}_0 ($\rightarrow E_{RS}^0$).

If we calculate the energies for a particular state the extensivity error is defined by $\Delta^{\text{extensivity}} = E_{RS} - (E_{R/S_0} + E_{R_0/S} - E_{RS}^0) = E_{RS} + E_{RS}^0 - (E_{R/S_0} + E_{R_0/S})$. It will be no surprise that the CCSD ground state energy satisfies $E_{RS} = E_{R/S_0} + E_{R_0/S} - E_{RS}^0$, and the extensivity error is zero. Since the reference Hartree–Fock calculation is the same in all calculations, and the correlation energy in the H_0 calculation is zero, this implies that the correlation energy for the ground state is additively separable.

All calculations of excitation energies discussed below are based on a CCSD calculation and they are core-extensive, which means that we can simply focus on the excitation energies themselves, rather than total energies. In the first test we examine generalized extensivity for the EOM-CCSD approach (see table 1).

The excitation can either be a ‘localized’ excitation, meaning that the excitation is confined to one subsystem (R or S) only, or it may be a mixed excitation, indicating that every non-zero excitation coefficient has mixed labels. Due to the symmetry criterion used in the partitioning of the orbitals, the excited states of A_1 and A_2 symmetry are ‘localized’ excitations, while the excitations that are of B_1 or B_2 symmetry are mixed excitations (or ‘charge-transfer’ excitations) that involve excitation between the two subsets. The localized

excitation energies of A symmetry reduce either to the full result (if the two-electron integrals are present), or they reduce to the zeroth-order result (which is just the difference of the Hartree–Fock orbital energies). For all of these localized excitations the extensivity error is found to be zero. However, for the mixed excitations of B symmetry, there are substantial extensivity errors of the order of 0.05 to 0.15 eV for this particular system. Let us emphasize that the excitation energies in the blocks of A symmetry are not extensive either, but the present test case cannot reveal the error due to the particulars of the orbital partitioning. Choosing a different orbital partitioning by dividing, for example, the occupied and (selected) virtual orbitals in the same irrep into different groups would reveal the extensivity error for these excitations also. In general, excitations in EOM-CCSD explicitly violate the extensivity test only if the zeroth-order occupied and virtual orbital associated with a particular excitation would be classified to fall into different subsets.

The results from EOM-CCSD can be contrasted with the results from a STEOM-CCSD calculation [37]. In STEOM, a sequence of transformations is carried out that involves a ground state CCSD calculation, a calculation of principal ionization potentials using IP-EOM-CCSD and a calculation of principal electron affinities using EA-EOM-CCSD. Using this information a connected transformed Hamiltonian is obtained which is diagonalized over single excitations only. The STEOM scheme is closely related to FSCE, and differs only in some details, which do not affect issues regarding extensivity. Conclusions drawn for STEOM will hold equally well for FSCE, therefore. In the STEOM calculation reported below the active IP orbitals include the $3a_1, 1b_1, 1b_2$ orbitals, while all virtual orbitals below 9 eV are included in the active EA set, which comprises $4a_1, 3b_1, 2b_2$ and one orbital of a_2 symmetry. The STEOM excitation energies can be

Table 1. Generalized extensivity test of EOM-CCSD excitation energies for the stretched water molecule in the Sadlej basis set.

Symmetry	Orbital character	RS	R	S	H_0	$R + S - H_0$	Extensivity error $RS - (R + S - H_0)$
1A_1	$3a_1 \rightarrow 4a_1$	7.110	7.110	13.408	13.408	7.110	0
	$3a_1 \rightarrow 5a_1$	10.597	10.597	16.536	16.536	10.597	0
	$1b_1 \rightarrow 2b_1$	11.531	15.222	11.531	15.222	11.531	0
1B_2	$1b_2 \rightarrow 4a_1$	11.294	13.657	11.247	13.734	11.167	0.124
	$1b_2 \rightarrow 5a_1$	14.336	16.671	14.374	16.861	14.184	0.152
	$1b_2 \rightarrow 6a_1$	15.068	17.505	15.177	17.664	15.018	0.050
1B_1	$3a_1 \rightarrow 2b_1$	13.745	13.792	14.306	14.419	13.678	0.067
	$1b_1 \rightarrow 4a_1$	14.147	14.134	14.156	14.211	14.079	0.068
	$3a_1 \rightarrow 3b_1$	15.830	15.984	16.336	16.612	15.709	0.121
1A_2	$1b_2 \rightarrow 2b_1$	4.616	14.745	4.616	14.745	4.616	0
	$1b_2 \rightarrow 3b_1$	9.2844	16.937	9.284	16.937	9.284	0

Table 2. Generalized extensivity test of STEOM-CCSD excitation energies (in eV) for the stretched water molecule in the Sadlej basis set.

Symmetry	Orbital character	RS	R	S	H ₀	R + S – H ₀	Extensivity error
¹ A ₁	3a ₁ → 4a ₁	7.563	7.563	13.408	13.408	7.563	0
	3a ₁ → 5a ₁	10.108	10.108	16.535	16.535	10.108	0
	1b ₁ → 2b ₁	11.654	15.222	11.654	15.222	11.654	0
¹ B ₂	1b ₂ → 4a ₁	11.170	13.657	11.247	13.734	11.170	0
	1b ₂ → 5a ₁	14.184	16.671	14.374	16.861	14.184	0
	1b ₂ → 6a ₁	15.018	17.505	15.177	17.664	15.018	0
¹ B ₁	3a ₁ → 2b ₁	13.678	13.792	14.306	14.419	13.678	0
	1b ₁ → 4a ₁	14.079	14.134	14.156	14.211	14.079	0
	3a ₁ → 3b ₁	15.709	15.984	16.336	16.612	15.709	0
¹ A ₂	1b ₂ → 2b ₁	5.006	14.745	5.006	14.745	5.006	0
	1b ₂ → 3b ₁	9.263	16.937	9.263	16.937	9.263	0

proven to satisfy generalized extensivity (as does FSCC), and this is confirmed in table 2: STEOM satisfies the generalized extensivity test in all cases, and this will be independent of the choice of the active space, or orbital partitioning.

For the mixed excitations the STEOM results reduce to the sum of IP on one system and EA on the other, since there is no coupling between the ‘correlated’ hole and particle. If a set of two-electron integrals is set to zero, the correlated IP/EA from the IP/EA-EOM-CCSD calculation is replaced by the Hartree–Fock orbital energy, and it all adds up to the correct extensive results. It is interesting to note that the EOM-CCSD result that is obtained in the additive fashion under the heading $R + S - H_0$ in table 1 nicely agrees with the STEOM result: the additive EOMCC results from the R-only and S-only calculations are precisely obtained as correlated IP+uncorrelated EA or *vice versa*, and accumulating the result provides the (correct) STEOM result. In a single EOM-CC calculation of the RS system this would require the inclusion of triples, however, which are included implicitly in STEOM. Let us note here that Meissner’s dressing of EOMCC [33] would also result precisely in the STEOM-CCSD results for these ‘charge-transfer’ excitations. The results from Meissner’s approach would be different from STEOM for the localized excitations of A symmetry, however, in which the hole and particle are interacting.

One can argue that the well-known error of EOM-CCSD for valence excited states is related to the violation of generalized extensivity. This is rigorously true if hole and particle are not interacting (the mixed cases below), but it can be argued in a similar fashion for states that are not of the charge-transfer type. The argument provides further confirmation that the implicit triples in STEOM, which are necessary for extensivity, are also important in practice, as, in general, STEOM

yields more accurate results for valence excitations [65]. Valence excitations have the characteristic feature that both the 2h1p IP amplitudes and the 2p1h EA amplitudes corresponding to the orbitals involved in the excitation are sizeable. If this is the case the product of these \hat{S}_2 amplitudes is relatively important, indicating a significance of triple excitations in EOM-CC, and a significant generalized extensivity error.

Let us also consider the extended STEOM approach [34, 35]. In this calculation the same doubly transformed Hamiltonian $\hat{G} = \{e^{\hat{S}}\}^{-1} e^{-\hat{T}} \hat{H} e^{\hat{T}} \{e^{\hat{S}}\}$ is used as in STEOM, but now this connected transformed Hamiltonian is diagonalized over single and double excitations. In principle, the extended STEOM approach does not satisfy generalized extensivity because operators exist in the transformed Hamiltonian which excite from singles to doubles (staying within the model or diagonalization space), but which excite out of the model space when acting on doubly excited determinants. However, in extended STEOM the matrix elements of such extensivity-violating operators might be very small as the dominant contributions are pushed to three-body components in the transformed Hamiltonian, or, alternatively, they involve excitations of inactive orbitals. In order to clearly demonstrate the violation of extensivity in the extended-STEOM approach we choose a smaller active space for the virtual space than in the previous STEOM calculation, consisting of the (4, 5)a₁, 2b₁, 2b₂ and the 1a₂ orbitals. The results are collected in table 3.

It is seen that the excitations of B symmetry do not satisfy generalized extensivity, and the error is significant (but always less than 0.01 eV) for excitations that involve the 6a₁ or 3b₁ orbitals, which are *not* included in the STEOM active space. For the ‘charge-transfer’ excitations of B symmetry the extended STEOM results are very close to STEOM-CCSD, and they are virtually

Table 3. Generalized extensivity test of extended STEOM-CCSD excitation energies (in eV) for the stretched water molecule in the Sadlej basis set.

Symmetry	Orbital character	RS	R	S	H ₀	R + S – H ₀	Extensivity error
¹ A ₁	3a ₁ → 4a ₁	7.062	7.062	13.408	13.408	7.062	0
	3a ₁ → 5a ₁	10.603	10.603	16.535	16.535	10.603	0
	1b ₁ → 2b ₁	11.492	15.222	11.492	15.222	11.492	0
¹ B ₂	1b ₂ → 4a ₁	11.17	13.657	11.247	13.734	11.17	0
	1b ₂ → 5a ₁	14.189	16.671	14.374	16.861	14.184	–0.005
	1b ₂ → 6a ₁	14.936	17.415	15.177	17.664	14.928	–0.008
¹ B ₁						0	0
	3a ₁ → 2b ₁	13.678	13.792	14.306	14.419	13.679	0.001
	1b ₁ → 4a ₁	14.079	14.134	14.156	14.211	14.079	0
¹ A ₂	3a ₁ → 3b ₁	15.688	15.984	16.314	16.612	15.686	–0.002
	1b ₂ → 2b ₁	4.593	14.745	4.593	14.745	4.593	0
	1b ₂ → 3b ₁	9.274	16.937	9.274	16.937	9.274	0

identical if a larger active space is employed (not shown). On the other hand, it is interesting (and somewhat surprising to us) that the extended STEOM energies are much closer to the EOM-CCSD results than to STEOM for the A₁ and A₂ excitations. For this example, the extensivity error in EOM-CCSD is clearly less important than the neglect of three-body operators in the transformed Hamiltonian in STEOM. Let us note here that the case of stretched water is not particularly suitable for the STEOM approach as the amplitudes that enter the transformation are rather large. For this same reason also extended STEOM results may not be fully accurate. It appears to us that the stretched water excitation spectrum might actually be quite a useful benchmark case for calibrating excitation spectra. Let us reiterate that all results in this section were obtained by artificially switching all integrals of mixed *a/b* character to zero. Let us for the sake of completeness consider the problem of stretched water for the original Hamiltonian, keeping all integrals. As illustrated in table 4, the rather large deviations between the excitation energies obtained by different methods remain for the true stretched water case, although they are smaller than in the extensivity test case.

In most cases we find a fairly smooth convergence of the energies as we run through the sequence EOM-CCSD, STEOM-CCSD, STEOM + D-CCSD, and extended STEOM-CCSD. In this table, STEOM + D indicates a non-iterative perturbative correction to STEOM that is based on a partitioning of the extended STEOM Hamiltonian, keeping only the Fock matrix in the doubles–doubles block of the final transformed Hamiltonian. This method has not before been reported in the literature, but it can provide a useful gauge of the accuracy of STEOM for a relatively modest computational cost. It would be useful to compare the results

Table 4. Comparison of EOM-CCSD, STEOM-CCSD, perturbatively corrected STEOM + D and extended STEOM excitation energies (in eV) for stretched water in the PBS basis set.

Symmetry	Orbital character	EOM	STEOM	STEOM + D	Extended STEOM
¹ B ₂	1b ₂ → 4a ₁	3.20	3.12	2.96	2.79
¹ A ₂	1b ₂ → 2b ₁	4.39	4.23	4.13	3.98
¹ A ₁	3a ₁ → 4a ₁	6.52	6.47	6.37	6.21
¹ B ₁	3a ₁ → 2b ₁	6.85	6.79	6.73	6.53
¹ B ₁	1b ₁ → 4a ₁	8.33	8.22	8.30	8.20
¹ B ₂	1b ₂ → 5a ₁	8.59	8.46	8.52	8.51
¹ A ₂	1b ₂ → 3b ₁	8.92	8.90	8.90	8.86
¹ A ₁	1b ₂ → 2b ₂	9.25	9.13	9.13	9.03
¹ B ₂	1b ₂ → 6a ₁	9.29	9.21	9.20	9.20

in table 4 against EOM-CCSDT as no fully reliable result is available for these excitation energies at present.

Let us briefly surmise the potential of the generalized extensivity test. This test can be used to indicate if methods satisfy generalized extensivity or not, and to assess the magnitude of the effect. However, care has to be exercised in interpreting the result: if results are shown to violate generalized extensivity a conclusion is reached. However, if the test appears to indicate that a particular method satisfies the test, the orbital partitioning may not have probed the method appropriately for the states considered. It is essential that zeroth-order states are considered that correspond to mixed orbital excitations. This statement might be illustrated by the general extensivity test applied to the IP-EOM-CCSD method. This method will satisfy generalized extensivity for principle IPs as it is impossible to create a mixed orbital zeroth-order ionization. However, the IP-EOM-CCSD approach will violate generalized extensivity for

mixed 2h1p states, consisting for example of 1h ionization on one subsystem and a 1h1p excitation on the other. This would be analogous to EE-EOM-CCSD for mixed excitations. It follows that certain classes of states may be described correctly (e.g. principle IPs), independent of the partitioning of the orbital space, while others may violate generalized extensivity for certain orbital partitionings (e.g. mixed shake-up states).

Let us reiterate here that the EOM-CCSD method is core-extensive *and* size-extensive/intensive for extended systems as it does not contain disconnected closed diagrams. Generalized extensivity is a more demanding property. For the excitation methods considered above, in principle the same test could be carried out using the size-consistency test and investigating an appropriate set of non-interacting charge-transfer states to probe extensivity. For the STEOM and EOMCC methods such tests have been presented before [37]. However, in practice it may be hard to find low-lying charge-transfer excitations to illustrate the concepts. This appears to be quite straightforward for the present test case, and this may be even more relevant for multi-reference situations which can easily be plagued by intruder problems as high-lying excitations are probed in the numerical size-consistency test [55]. Moreover, in the case of multi-reference methods the model space ought *not* to be a complete direct product space of mixed excitations, as this would constitute a test of size-consistency rather than generalized extensivity. In the generalized extensivity test the partitioning of the Hamiltonian can be made independent of the model space, and the test would be properly discriminating of generalized extensivity if zeroth-order states have mixed orbital character. The size of the extensivity error would also provide some insight whether generalized extensivity is of practical importance for multi-reference methods, which generally use rather small diagonalization spaces. The violation of extensivity is generally a cumulative effect that tends to be more severe in larger basis sets, for example. However, this aspect does not play a role in the diagonalization of the effective Hamiltonian, as the dimension of the model space would be independent of the atomic basis set.

6. Summary

We presented a general survey of the notions of size-extensivity, size-consistency and generalized extensivity, which are crucial ingredients of formal many-body theory. Using algebraic operator techniques we demonstrate a convenient way to carry out the proof that a variety of coupled-cluster methods satisfy the requirement that the energy is additively separable upon

consideration of a compound system of two non-interacting Hamiltonians. The proofs are demonstrated to be quite straightforward. In the proofs for multi-reference CC methods we explicitly assume that the incomplete model space is separable, meaning that the model space for the compound system is the direct product of the fragment model spaces. This is quite a strong condition itself, and it suffices to show then that the effective Hamiltonian for the compound system is the direct sum of the fragment effective Hamiltonians, to prove size-consistency. We discuss three variants of Hilbert space multi-reference theories, which are each size-consistent: (i) the approach by Li and Paldus which includes active space cluster components such that the overall wave operator satisfies intermediate normalization; (ii) the approach by Meissner/Mukherjee which includes extensivity-correcting quasi-open components in the cluster operator, which are determined by equating the corresponding extensivity-violating quasi-open components of the effective Hamiltonian to zero; and (iii) a third approach in which the active space components of the cluster operator are equated to zero, while the effective Hamiltonian is defined using a metric matrix. Neither approach (ii) nor (iii) satisfies intermediate normalization, while keeping the general Jeziorski–Monkhort parameterization for the target wave functions.

While size-consistency in conjunction with the direct product structure of the model space implies additive separability of the effective Hamiltonian for physically non-interacting subsystems, it does not imply generalized extensivity of the resulting energies, with respect to arbitrary partitionings of the orbitals independent of the model space. Further analysis (as discussed in the works by Meissner and Mukherjee) is required to establish this more elusive property.

Yet another possibility is to use an intermediate Hamiltonian formalism [58–62]. It is sufficient (but not necessary; see, for example, Meissner [66, 67]) for generalized extensivity of the resulting energies that the intermediate Hamiltonian is defined over a complete model space, and that it is connected. Using the techniques developed in this paper the additive separability (or, equivalently, connectivity in the case of the complete model space) of the intermediate Hamiltonian is easily shown, and then we can build upon the general properties of the complete model space to prove generalized extensivity. A detailed discussion of the theory of intermediate Hamiltonians is beyond the scope of this paper, but it appears to us that this might provide a robust way to obtain accurate and general extensive results in multi-reference situations.

We have described an operational definition of generalized extensivity which can probe extensivity in

the sense of a single reference connected diagram theorem, and we have shown a numerical example that illustrates the concepts for the EOM-CCSD, STEOM-CCSD and extended STEOM-CCSD approaches, which show a particularly rich pallet of scaling behaviours. For these methods, generalized extensivity is demonstrated to be important, as the lack thereof leads to errors in excitation energies of up to about 0.2 eV. However, for these cases the number of extensivity-violating operators can be rather large, and their cumulative effect is responsible for the deterioration of the results. The practical importance of generalized extensivity is not entirely obvious for the general multi-reference case. In particular, for small model spaces the requirements of size-consistency and core-extensivity may provide sufficient constraints in practice to not jeopardize the accuracy of the results. The operational test of generalized extensivity in principle will allow us to address this question, which we leave for future investigations.

Acknowledgements

Extensivity and connected diagrams are cornerstones of the work of Rodney J. Bartlett and it is our great pleasure to contribute this paper to this volume celebrating his 60th birthday. This work was supported by a Discovery Grant of the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

- [1] H. Primas. In *Modern Quantum Chemistry*, Vol. 2, O. Sinanoglu (Ed.), Academic Press, New York (1965).
- [2] B.H. Brandow. *Rev. mod. Phys.*, **39**, 771 (1967).
- [3] W. Kutzelnigg. In *Modern Theoretical Chemistry*, H.F. Schaefer (Ed.), Plenum Press, New York (1977).
- [4] I. Lindgren, D. Mukherjee. *Phys. Rep.*, **151**, 93 (1987).
- [5] R. Chowdhuri, D. Mukherjee, M.D. Prasad. In *Aspects of Many-body Effects in Molecules and Extended Systems*, Vol. 50, D. Mukherjee (Ed.), Springer, Berlin (1989).
- [6] D. Mukherjee, S. Pal. *Adv. quant. Chem.*, **20**, 291 (1989).
- [7] R.J. Bartlett, G.D. Purvis. *Int. J. quant. Chem.*, **14**, 561 (1978).
- [8] R.J. Bartlett. *A. Rev. Phys. Chem.*, **32**, 359 (1981).
- [9] N.H. March, W.H. Young, S. Sampanthar. *The Many-body Problem in Quantum Mechanics*, Cambridge University Press, Cambridge (1967).
- [10] A.L. Fetter, J.D. Walecka. *The Many-body Problem in Quantum Mechanics*, McGraw-Hill, New York (1971).
- [11] S. Raimes. *Many-electron Theory*, North-Holland, Amsterdam (1972).
- [12] I. Lindgren, J. Morrison. *Atomic Many-body Theory*, Springer, Berlin (1982).
- [13] S.A. Kucharski, R.J. Bartlett. *Adv. quant. Chem.*, **18**, 281 (1986).
- [14] R.J. Bartlett. In *Modern Electronic Structure Theory*, Vol. 2, D.R. Yarkony (Ed.), World Scientific, Singapore (1995).
- [15] J.A. Pople. *Int. J. quant. Chem. Symp.*, **10**, 1 (1976).
- [16] D. Mukhopadhyay, S. Mukhopadhyay, R. Chauduri, D. Mukherjee. *Theor. Chim. Acta*, **80**, 441 (1991).
- [17] M. Deleuze, J. Delhalle, B.T. Pickup, J.L. Calais. *Phys. Rev. B*, **46**, 15668 (1992).
- [18] M. Nooijen, R.J. Bartlett. *Int. J. quant. Chem.*, **63**, 601 (1997).
- [19] K.A. Brueckner. *Phys. Rev. B*, **97**, 1353 (1955).
- [20] K.A. Brueckner. *Phys. Rev. B*, **100**, 36 (1955).
- [21] M. Gell-Mann, K.A. Brueckner. *Phys. Rev. B*, **106**, 364 (1957).
- [22] J. Hubbard. *Proc. R. Soc. A*, **240**, 539 (1957).
- [23] J. Goldstone. *Proc. R. Soc. A*, **239**, 267 (1957).
- [24] A.A. Abrikosov, L.P. Gorkov, I.E. Dzyaloshinski. *Methods of Quantum Field Theory in Statistical Physics*, Prentice Hall, Englewood Cliffs, NJ (1963).
- [25] M. Vracko, B. Champagne, D.M. Mosley, J.-M. Andre. *J. chem. Phys.*, **102**, 6831 (1995).
- [26] E.A. Perpete, B. Champagne, B. Kirtman. *J. chem. Phys.*, **107**, 2463 (1997).
- [27] B. Champagne, D. Jacquemin, J.-M. Andre, B. Kirtman. *J. phys. Chem.*, **101**, 3158 (1997).
- [28] A.C. Hurley. *Electron Correlation in Small Molecules*, Academic Press, New York (1976).
- [29] W. Kutzelnigg, D. Mukherjee, S. Koch. *J. chem. Phys.*, **87**, 5902 (1987).
- [30] D. Mukherjee, W. Kutzelnigg, S. Koch. *J. chem. Phys.*, **87**, 5911 (1987).
- [31] M. Nooijen. *The Coupled Cluster Green's Function*, Vrije Universiteit, Amsterdam (1992).
- [32] H. Koch, H.J. Jensen, P. Jorgensen, T. Helgaker. *J. chem. Phys.*, **93**, 3345 (1990).
- [33] L. Meissner, R.J. Bartlett. *J. chem. Phys.*, **102**, 7490 (1995).
- [34] M. Nooijen, V. Lotrich. *J. chem. Phys.*, **113**, 494 (2000).
- [35] M. Nooijen. *J. phys. Chem. A*, **104**, 4553 (2000).
- [36] L. Meissner, S. Kucharski, R.J. Bartlett. *J. chem. Phys.*, **91**, 6187 (1989).
- [37] M. Nooijen, R.J. Bartlett. *J. chem. Phys.*, **107**, 6812 (1997).
- [38] J. Geertsen, C.M.L. Rittby, R.J. Bartlett. *Chem. Phys. Lett.*, **164**, 57 (1989).
- [39] J.F. Stanton, R.J. Bartlett. *J. chem. Phys.*, **98**, 7029 (1993).
- [40] H.J. Monkhorst. *Int. J. quant. Chem. Symp.*, **11**, 421 (1977).
- [41] D. Mukherjee, P.K. Mukherjee. *Chem. Phys.*, **39**, 325 (1979).
- [42] H. Koch, R. Kobayashi, A.S. Demeras, P. Jorgensen. *J. chem. Phys.*, **100**, 4393 (1994).
- [43] M. Nooijen, J.G. Snijders. *Int. J. quant. Chem. Symp.*, **26**, 55 (1992).
- [44] M. Nooijen, J.G. Snijders. *Int. J. quant. Chem.*, **48**, 15 (1993).
- [45] B. Jeziorski, H.J. Monkhorst. *Phys. Rev. A*, **24**, 1668 (1981).
- [46] L. Meissner, R.J. Bartlett. *J. chem. Phys.*, **92**, 561 (1990).
- [47] X. Li, J. Paldus. *J. chem. Phys.*, **119**, 5320 (2003).
- [48] X. Li, J. Paldus. *J. chem. Phys.*, **119**, 5334 (2003).
- [49] X. Li, J. Paldus. *J. chem. Phys.*, **119**, 5346 (2003).

- [50] D. Mukhopadhyay, D. Mukherjee. *Chem. Phys. Lett.*, **163**, 171 (1989).
- [51] D. Mukhopadhyay, D. Mukherjee. *Chem. Phys. Lett.*, **177**, 441 (1991).
- [52] D. Mukherjee. *Chem. Phys. Lett.*, **125**, 207 (1986).
- [53] D. Mukherjee. *Int. J. quant. Chem. Symp.*, **20**, 409 (1986).
- [54] L. Meissner, K. Jankowski. *Int. J. quant. Chem.*, **36**, 705 (1989).
- [55] X. Li, J. Paldus. *Int. J. quant. Chem.*, submitted.
- [56] J. Paldus, X. Li, D.K. Petraco. *J. Math. Chem.*, **35**, 215 (2004).
- [57] U.S. Mahapatra, B. Datta, D. Mukherjee. *J. chem. Phys.*, **110**, 6171 (1998).
- [58] J.-P. Malrieu, P. Durand, J.-P. Daudey. *J. Phys. B*, **18**, 809 (1985).
- [59] S. Koch. *Theor. chim. Acta*, **81**, 169 (1991).
- [60] D. Mukhopadhyay, B. Datta, D. Mukherjee. *Chem. Phys. Lett.*, **197**, 236 (1992).
- [61] L. Meissner, M. Nooijen. *J. chem. Phys.*, **102**, 9604 (1995).
- [62] L. Meissner, P. Malinowski. *Phys. Rev. A*, **61**, 062510 (2000).
- [63] D. Pahari, S. Chattopadhyay, S. Das, D. Mukherjee. *Chem. Phys. Lett.*, **381**, 223 (2003).
- [64] A. Sadlej. *Theor. chim. Acta*, **79**, 123 (1991).
- [65] M. Nooijen. *Spectrochim. Acta A*, **55**, 539 (1999).
- [66] L. Meissner. *Chem. Phys. Lett.*, **255**, 244 (1996).
- [67] L. Meissner. *Collect. Czech. Chem. Commun.*, **68**, 105 (2003).