Theor Chem Acc (2011) 129:727–746
DOI 10.1007/s00214-011-0954-4

FEATURE ARTICLE

Thermodynamic limit and size-consistent design

So Hirata

Received: 4 February 2011 / Accepted: 26 April 2011 / Published online: 1 June 2011

Abstract This article aims to dispel confusions about the definition of size consistency as well as some incompatibility that exists between different criteria for judging whether an electronic structure theory is size consistent and thus yields energies and other quantities having correct asymptotic size dependence. It introduces extensive and intensive diagram theorems, which provide unambiguous sufficient conditions for size consistency for extensive and intensive quantities, respectively, stipulated in terms of diagrammatic topology and vertex makeup. The underlying algebraic size-consistency criterion is described, which relies on the polynomial dependence of terms in the formalism on the number of wave vector sampling points in Brillouin-zone integrations. The physical meanings of two types of normalization of excitation amplitudes in electron-correlation theories, namely, the intermediate and standard normalization, are revealed. The amplitudes of the operator that describes an extensive quantity (the extensive operator) are subject to the intermediate normalization, while those of the operator that corresponds to an intensive quantity (the intensive operator) must be normalized. The article also introduces the extensive-intensive consistency theorem which specifies the relationship between the spaces of determinants reached by the extensive and intensive operators in a size-consistent method for intensive quantities. Furthermore, a more fundamental question is addressed as to what makes energies extensive and thus an application of thermodynamics to chemistry valid. It is shown that the energy of an electrically neutral, periodic, non-metallic solid is extensive. On this basis, a strictly size-consistent redefinition of the Hartree–Fock theory is proposed.

Keywords Size consistency · Electronic structure theory · Diagrams · Thermodynamic limit · Extensive diagram theorem · Intensive diagram theorem

1 Significance and definitions

An observable thermochemical quantity is either extensive or intensive [1]. Extensive quantities include energy, entropy, and mass and are asymptotically proportional to the volume of a chemical system. Intensive quantities such as temperature, pressure, and chemical potential are asymptotically independent of the volume. Computational methods that are meant to predict these quantities accurately must, in the very least, ensure that the predicted values satisfy these asymptotic conditions. The methods that do not meet this basic requirement (size consistency) suffer from ever increasing errors in their prediction with volume and reduce to being nonsensical in the thermodynamic (infinite-volume) limit. Size consistency is thus one of the most important design principles of computational methods for chemistry and a distinguishing feature of genuine many-body techniques. The definition of size consistency we adopt here is given below.

Definition 1 (*size consistency*) [2]: A size-consistent method yields a total energy (or some other extensive quantity) of an electrically neutral, periodic solid which is asymptotically proportional to its volume ("size") under the periodic boundary condition and an excitation energy (or some other intensive quantity) in the same system

---

S. Hirata (✉)
Department of Chemistry,
University of Illinois at Urbana-Champaign,
600 South Mathews Avenue, Urbana, IL 61801, USA
e-mail: sohirata@illinois.edu

which is asymptotically constant with size. When varying the size, the number density should be kept constant.

The term “size extensivity” is used in this article to mean size consistency for extensive quantities and the term “size intensivity” to mean size consistency for intensive quantities. For example, a size-consistent excited-state method must be size extensive for total energies and size intensive for excitation energies.

In chemistry, there have been two size-consistency criteria [2–4], which are, however, not only inequivalent but also mutually inconsistent with each other. One is the K-dependence [4–9] and related diagrammatic criteria [2, 5, 7, 10–12], which are the only accepted ones in condensed matter and nuclear physics. The other is the supermolecule criterion [3, 13, 14], which has been said to have the advantage of being applicable numerically [3]. These are summarized as follows.

Definition 2 (*the K-dependence criterion*): The number (K) of wave vector (k vector) sampling points in the Brillouin zone (BZ) is a direct measure of system size under the periodic boundary condition (see below). Size consistency of a method can be proved or disproved by analyzing the polynomial dependence of terms in its formulas on K. Each of the equations defining a size-consistent method must consist only of terms with the same K dependence. Furthermore, the equations defining extensive and intensive observables must have the asymptotic K^1 and K^0 dependence, respectively.

Definition 3 (*the supermolecule criterion*): Let E_{A} and E_{B} be the total energies (or other extensive quantities) of isolated molecules A and B, respectively, and E_{AB} be the total energy of supermolecule AB that consists of molecules A and B between which there is no interaction. A size-consistent method must yield total energies of these three molecules so that they satisfy the additive separability condition: $E_{\mathrm{AB}} = E_{\mathrm{A}} + E_{\mathrm{B}}$. Each of the excitation energies (or other intensive quantities) of the isolated constituent molecules obtained by a size-consistent method must be found among the excitation energies of the supermolecule.

It should be cautioned that, in the literature, the term size extensivity is often associated with the K-dependence criterion and the term size consistency with the supermolecule criterion, which is not the nomenclature we adopt. Our standpoint is that there is only one valid criterion of size consistency and it is the K-dependence and related diagrammatic criteria. Since it is entirely possible for a method to be size consistent for intensive quantities but not for extensive quantities or vice versa, we do not use the terms size extensivity and size consistency interchangeably as has been done elsewhere.

2 Supermolecule criterion

What is wrong with the supermolecule criterion? One of its conceptual as well as practical problems is that it is not always possible to prevent molecules A and B, which are presumably infinitely separated, from interacting. Suppose A and B are different molecules with n_{A} and n_{B} electrons, respectively. An electron can transfer from B to A as soon as supermolecule AB is formed if the electron affinity of A exceeds the ionization energy of B. In this circumstance, a size-consistent method will fail the supermolecule test because

$$E_{\mathrm{AB}(n_{\mathrm{A}}+n_{\mathrm{B}})} = E_{\mathrm{A}(n_{\mathrm{A}}+1)} + E_{\mathrm{B}(n_{\mathrm{B}}-1)} \neq E_{\mathrm{A}(n_{\mathrm{A}})} + E_{\mathrm{B}(n_{\mathrm{B}})}, \qquad (1)$$

where the numbers of electrons are given in the parentheses.

One may argue that we should consider a supermolecule made of identical molecules so that there cannot be electron transfer. However, because the orbitals of one of the constituent molecules and those of the other are now degenerate, the issue of orbital invariance comes into play [4, 10]. If the total energies obtained by the method are not invariant with respect to rotation among occupied orbitals and/or to rotation among virtual orbitals, the total energies of the supermolecule cannot be determined uniquely because degenerate orbitals are freely rotated among themselves. In other words, a size-consistent method can fail the supermolecule test if the method lacks the property of orbital invariance. Orbital invariance is an important property, but it is an altogether different issue from size consistency.

The previous examples are the cases when size-consistent methods fail to satisfy the supermolecule criterion. We can also come up with the converse examples. Configuration-interaction singles and doubles (CISD) is not size consistent but is nevertheless exact for up to two-electron systems. CISD, therefore, satisfies the supermolecule criterion for any supermolecule with two electrons or less.

Another example is a supermolecule of two nuclei with no electrons. Any electronic structure method passes this supermolecule test and is judged size consistent because $E_{\mathrm{AB}} = E_{\mathrm{A}} = E_{\mathrm{B}} = 0$, which is of course a nonsensical conclusion. One may argue that this is a far-fetched example with no physical significance. To the contrary, it touches upon the fundamental question about the origin of extensivity or the existence of a thermodynamic limit of energy density [15–20] discussed in Sect. 5. The total energy of a lattice of like-charged particles grows as $V^{5/3}$ with volume V and is, therefore, non-extensive. The supermolecule criterion for this highly charged system is doubly wrong as it suggests that any method is size extensive for total energies when the latter are not even supposed to be extensive!

The existence of these counterexamples indicates that the supermolecule criterion is neither a sufficient nor necessary condition for size consistency. It cannot be applied uniformly to the formalism (as opposed to an implementation) of a method and it gives unreliable results depending on the details such as the choice of supermolecule, the distribution of electrons among constituent molecules, the particular excited state considered, whether or not the method lacks orbital invariance, and so forth, and is thus misleading. Even when a method is shown to pass many such tests, it only provides a piece of circumstantial evidence for the exact proportionality of an extensive quantity (or exact constancy of an intensive quantity) with respect to the number of non-interacting particles, which falls short of the definition of size consistency given above, which pertains to the *asymptotic* behavior of extensive and intensive quantities for a system of *interacting* particles.

Finally, it is sometimes stated incorrectly that for a method to be size consistent, a wave function of supermolecule AB must be multiplicatively separable into wave functions of molecules A and B, namely,

$$\Psi_{\mathrm{AB}} = \Psi_{\mathrm{A}}\Psi_{\mathrm{B}}. \tag{2}$$

As we shall show later, this is not necessary for the energy to be additively separable for the supermolecule. For instance, the first-order Møller–Plesset perturbation (MP1) wave function does not have an exponential structure and is not multiplicatively separable. Nevertheless, its corresponding second-order Møller–Plesset perturbation (MP2) energy remains size consistent (see Sect. 4.1).

3 *K*-dependence criterion

3.1 Localized and delocalized bases

In a spatially localized basis, the total energy can be defined as a sum of local energies of molecular fragments, each of which is computed using the local basis around the fragment. An excitation energy is obtained as one occurring in a molecular fragment using its local basis. In this way, the total energy is by construction extensive and excitation energies intensive regardless of how the local energies and local excitation energies are obtained. In other words, it is more instructive to use a completely delocalized basis when determining the intrinsic size consistency or its lack of a method.

Symmetry-adapted orbitals for an infinitely extended, periodic system such as canonical HF orbitals form a delocalized basis. The canonical HF spin-orbital in the pth band with the wave vector k_p in a one-dimensional periodic system is given by

$$\varphi_{pk_p}(\boldsymbol{r}) = K^{-1/2}\sum_{\mu}\sum_{m} C_{pk_p}^{\mu k_p}\exp(imk_p a)\chi_\mu(\boldsymbol{r}-m\boldsymbol{a}), \tag{3}$$

where $C_{pk_p}^{\mu k_p}$ is an expansion coefficient, $\boldsymbol{a}$ is the lattice vector (with length a), and $\chi_\mu(\boldsymbol{r}-m\boldsymbol{a})$ is the μth atomic orbital (AO) centered in the mth unit cell. In the underlying periodic boundary condition, an infinite chain is regarded as a mathematical ring of K unit cells. Adopting the unit of length in which $a = \pi$, k_p can take one of the following K distinct values:

$$k_p = \frac{2n}{K}, \quad n = 0, \ldots, K-1. \tag{4}$$

Hence, K is a direct measure of the system's volume (size) and the polynomial dependence on K (the scaling property) of a quantity gives its size dependence [9]. The factor of $K^{-1/2}$ in Eq. (3) arises from the normalization of φ_{pk_p}.

The Fock and antisymmetrized two-electron integrals in this basis are given by

$$f_{qk_q}^{pk_p} = K^0\sum_{\mu,\nu}\sum_{m} C_{pk_p}^{\mu k_p *}C_{qk_q}^{\nu k_q}\exp(imk_q a) f_{\nu(m)}^{\mu(0)}, \tag{5}$$

$$\begin{aligned} v_{rk_r sk_s}^{pk_p qk_q} = K^{-1}\sum_{\kappa,\lambda,\mu,\nu}\sum_{m_1,m_2,m_3} & C_{pk_p}^{\kappa k_p *}C_{qk_q}^{\lambda k_q *}C_{rk_r}^{\mu k_r}C_{sk_s}^{\nu k_s} \\ & \times \exp\{i(-m_1k_q + m_2k_r + m_3k_s)a\} v_{\mu(m_2)\nu(m_3)}^{\kappa(0)\lambda(m_1)}, \end{aligned} \tag{6}$$

where $f_{\nu(m)}^{\mu(0)}$ and $v_{\mu(m_2)\nu(m_3)}^{\kappa(0)\lambda(m_1)}$ are Fock and antisymmetrized two-electron integrals in the AO basis. The quantities in the delocalized basis vanish identically unless appropriate momentum conservation conditions are satisfied, which are

$$k_p = k_q + 2n, \tag{7}$$

$$k_p + k_q = k_r + k_s + 2n, \tag{8}$$

respectively, for $f_{qk_q}^{pk_p}$ and $v_{rk_r sk_s}^{pk_p qk_q}$, where n is an integer.

Equations (5) and (6) indicate that $f_{qk_q}^{pk_p}$ and $v_{rk_r sk_s}^{pk_p qk_q}$ exhibit the asymptotic size dependence of K^0 and K^{-1}, respectively [9], insofar as the lattice sums in these equations converge at finite, K-independent values. As shown in Sect. 5, with proper definitions of these integrals, this is indeed the case for a non-metallic solid with an electrically neutral unit cell. The K^0 scaling of $f_{qk_q}^{pk_p}$ and K^{-1} scaling of $v_{rk_r sk_s}^{pk_p qk_q}$ constitute the basis of the following analyses [5].

3.2 Hartree–Fock

Let us use this K-dependence argument to prove the size consistency (i.e., extensivity of the total energies in the ground states and intensivity of the orbital energies) of the HF method. The total energy (which includes the nuclear repulsion energy; see Sect. 5) is given by

$$E_{\mathrm{HF}} = \sum_{i}\sum_{k_i} f_{ik_i}^{ik_i} - \frac{1}{2}\sum_{i,j}\sum_{k_i,k_j} v_{ik_ijk_j}^{ik_ijk_j}. \tag{9}$$

We use in the above equation and hereafter the convention in which i, j, k, and l label occupied energy bands, a, b, c, and d virtual energy bands, and p, q, r, s, t, and u either. The first term has the summation over k_i, which gives rise to a factor of K^1, and $f_{ik_i}^{ik_i}$, which is a K^0 quantity. This term is thus proportional to K^1 or proportional to size and is, therefore, extensive. Similarly, the second term can be shown to be a K^1 quantity and extensive because the two-fold k summation gives rise to a factor of K^2 and $v_{ik_ijk_j}^{ik_ijk_j}$ scales as K^{-1}.

The HF orbital energies, $f_{ik_i}^{ik_i}$ and $f_{ak_a}^{ak_a}$, are related to ionization energies and electron affinities. They form energy bands. The orbital energy differences, $f_{ak_a}^{ak_a} - f_{ik_i}^{ik_i}$, may also be considered approximations to excitation energies. Since $f_{qk_q}^{pk_p}$ scales as K^0, the HF approximations to these are intensive, as they should be. Together, the HF method is size consistent. It may be emphasized that the above proves the asymptotic correct size dependence of observables for a system of interacting particles. An extension to two- and three-dimensional solids is trivial.

Equation (9) can be diagrammatically expressed as follows [5, 12]:

E_{HF} =  − . (10)

In these diagrams of the Hugenholtz type, each dot symbol is called a vertex and represents either an integral or an excitation amplitude (collectively called a tensor). Each arc or line (called an edge) denotes a spin-orbital index. The number of edges attached to a vertex is called the order of the vertex. An edge connecting two vertexes signifies a contraction of two tensors by the common spin-orbital index represented by the edge. In this particular case, the vertexes are drawn as filled circles, each of which denoting either a Fock integral if two edges are associated with it or a two-electron integral if four edges are attached. The first term in the right-hand side of Eq. (9) has only one tensor (f) with two indexes, which happen to be the same (ik_i). The corresponding diagram (above, left) has one vertex (filled circle) and its two edges meet to make one loop. The second term of Eq. (9) is converted to one vertex (v) with two loops (ik_i and jk_j) (above, right). The diagrams are closed and connected.

Definition 4 (*closed and open diagrams*): An edge is called external if one of its termini is not attached to any vertex. Otherwise, an edge is internal. A diagram is closed if it contains only internal edges and open if it has one or more external edges.

Definition 5 (*connected and disconnected diagrams*): A diagram is connected if there is a path between any two vertexes [21]. Otherwise, it is disconnected.

In the following, we will show that the criteria of size consistency using the K-dependence argument can be cast into some rules on the topology of the diagrams and the makeup of their vertexes. We do not document here the precise rules of diagrammatic evaluations, which are not essential for our purpose and can be found in March et al. [5] as well as in Shavitt and Bartlett [12].

3.3 Coupled-cluster doubles

Let us now turn to the question of size extensivity of coupled-cluster doubles (CCD) for correlation energies in the ground states [10, 22]. In this approximation [11], a wave function is obtained by acting an exponential excitation operator on a reference wave function, which we take to be a HF wave function, $|\Phi_0\rangle$. Thus,

$$\begin{aligned}|\Psi\rangle &= \exp\left(\hat{T}_2\right)|\Phi_0\rangle \\ &= \left(1 + \hat{T}_2 + \hat{T}_2^2/2! + \hat{T}_2^3/3! + \ldots\right)|\Phi_0\rangle,\end{aligned} \tag{11}$$

with the two-electron excitation operator defined by its action,

$$\hat{T}_2|\Phi_0\rangle = \frac{1}{(2!)^2}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a} t_{ik_ijk_j}^{ak_abk_b}|\Phi_{ik_ijk_j}^{ak_abk_b}\rangle, \tag{12}$$

where $|\Phi_{ik_ijk_j}^{ak_abk_b}\rangle$ is a doubly excited determinant. Note that only three of the k vectors of a non-vanishing amplitude, $t_{ik_ijk_j}^{ak_abk_b}$, are linearly independent and appear as the summation indexes. Equation (11) is an example of the use of the intermediate normalization [14]: $|\Psi\rangle$ is not normalized, but $|\Phi_0\rangle$ is. In Sect. 4.1, we will reveal the significance of this normalization scheme in relation to size extensivity.

To arrive at the equations that determine the correlation energy and amplitudes, we insist that $|\Psi\rangle$ satisfy the Schrödinger equation in the space spanned by $|\Phi_0\rangle$ and its doubly excited determinants, namely,

$$\langle\Phi_0|\hat{H}\exp\left(\hat{T}_2\right)|\Phi_0\rangle = \langle\Phi_0|E_{\mathrm{CCD}}\exp\left(\hat{T}_2\right)|\Phi_0\rangle, \tag{13}$$

$$\langle\Phi_{ik_ijk_j}^{ak_abk_b}|\hat{H}\exp\left(\hat{T}_2\right)|\Phi_0\rangle = \langle\Phi_{ik_ijk_j}^{ak_abk_b}|E_{\mathrm{CCD}}\exp\left(\hat{T}_2\right)|\Phi_0\rangle. \tag{14}$$

This projection in conjunction with the intermediate normalization is not the only way to define $t_{ik_ijk_j}^{ak_abk_b}$, but it is the most successful one for a good reason, which will be discussed in Sect. 4.1. The Hamiltonian has up to two-body interactions and, therefore, does not allow determinants that differ by more than two spin-orbitals to interact. Using this fact, we can simplify the above two equations to

$$\langle\Phi_0|\hat{H}\big(1+\hat{T}_2\big)|\Phi_0\rangle = E_{\mathrm{CCD}}, \tag{15}$$

$$\langle\Phi^{ak_abk_b}_{ik_ijk_j}|\hat{H}\big(1+\hat{T}_2+\hat{T}_2^2/2!\big)|\Phi_0\rangle = E_{\mathrm{CCD}}t^{ak_abk_b}_{ik_ijk_j}. \tag{16}$$

When the integrals are evaluated, they reduce to the sum-of-product expressions,

$$E_{\mathrm{HF}} + \frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a} v^{ik_ijk_j}_{ak_abk_b}t^{ak_abk_b}_{ik_ijk_j} = E_{\mathrm{CCD}} \tag{17}$$

and

$$\begin{aligned}
&v^{ak_abk_b}_{ik_ijk_j} - P_{i/j}\sum_{k} f^{kk_k}_{ik_i}t^{ak_abk_b}_{kk_kjk_j} + P_{a/b}\sum_{c} f^{ak_a}_{ck_c}t^{ck_cbk_b}_{ik_ijk_j}\\
&\quad + \frac{1}{2}\sum_{k,l}\sum_{k_k} v^{kk_klk_l}_{ik_ijk_j}t^{ak_abk_b}_{kk_klk_l} + \frac{1}{2}\sum_{c,d}\sum_{k_c} v^{ak_abk_b}_{ck_cdk_d}t^{ck_cdk_d}_{ik_ijk_j}\\
&\quad + P_{a/b}P_{i/j}\sum_{k,c}\sum_{k_k} v^{ak_akk_k}_{ik_ick_c}t^{ck_cbk_b}_{kk_kjk_j}\\
&\quad + \frac{1}{4}\sum_{k,l,c,d}\sum_{k_k,k_c} v^{kk_klk_l}_{ck_cdk_d}t^{ck_cdk_d}_{ik_ijk_j}t^{ak_abk_b}_{kk_klk_l}\\
&\quad - \frac{1}{2}P_{a/b}P_{i/j}\sum_{k,l,c,d}\sum_{k_k,k_c} v^{kk_klk_l}_{ck_cdk_d}t^{ak_ack_c}_{kk_kjk_j}t^{dk_dbk_b}_{ik_ilk_l}\\
&\quad - \frac{1}{2}P_{a/b}\sum_{k,l,c,d}\sum_{k_k,k_c} v^{kk_klk_l}_{ck_cdk_d}t^{ck_cbk_b}_{kk_klk_l}t^{ak_adk_d}_{ik_ijk_j}\\
&\quad - \frac{1}{2}P_{i/j}\sum_{k,l,c,d}\sum_{k_k,k_c} v^{kk_klk_l}_{ck_cdk_d}t^{ck_cdk_d}_{kk_kjk_j}t^{ak_abk_b}_{ik_ilk_l}\\
&\quad + E_{\mathrm{HF}}t^{ak_abk_b}_{ik_ijk_j} + \frac{1}{4}\sum_{k,l,c,d}\sum_{k_k,k_l,k_c} v^{kk_klk_l}_{ck_cdk_d}t^{ck_cdk_d}_{kk_klk_l}t^{ak_abk_b}_{ik_ijk_j}\\
&\quad = E_{\mathrm{CCD}}t^{ak_abk_b}_{ik_ijk_j},
\end{aligned} \tag{18}$$

where P's are index permutation operators, whose precise definitions [12] are unimportant here.

Equation (18) is solved for the amplitudes of the $\hat{T}_2$ operator, which are then substituted in Eq. (17) to yield the CCD energy. For the method to be size extensive, all terms in Eq. (18) must exhibit the mutually consistent K dependence and, furthermore, all terms in Eq. (17) must scale as K^1. Surprisingly, the first of these conditions is not met, at least manifestly, by Eq. (18).

Let us first analyze Eq. (17). As we have established in the previous subsection, E_{HF} is extensive. We, therefore, only need the second term to scale as K^1 and, for this to be the case, we must require $t^{ak_abk_b}_{ik_ijk_j}$ to be an asymptotic K^{-1} quantity because $v^{ik_ijk_j}_{ak_abk_b}$ scales as K^{-1} and the three-fold k summation gives rise to a factor of K^3. However, if we assume this, we find that the last two terms in the left-hand side of Eq. (18) as well as the right-hand side scale differently (as K^0) from the rest (as K^{-1}).

This apparent violation of size extensivity can be removed by noticing that these K^0 terms cancel with each other exactly between left- and right-hand sides. Using Eq. (17), we can rewrite the last two terms of the left-hand side of Eq. (18) as

$$\left\{E_{\mathrm{HF}} + \frac{1}{4}\sum_{k,l,c,d}\sum_{k_k,k_l,k_c} v^{kk_klk_l}_{ck_cdk_d}t^{ck_cdk_d}_{kk_klk_l}\right\}t^{ak_abk_b}_{ik_ijk_j} = E_{\mathrm{CCD}}t^{ak_abk_b}_{ik_ijk_j}. \tag{19}$$

Thus, their K dependence turns out to be inconsequential to our analysis. Note that these terms with inconsistent K dependence are disconnected (see below), which algebraically means that they are simple products rather than contractions with at least one common contraction index. Eliminating these mutually canceling, disconnected terms, we can rewrite Eq. (18) into a manifestly size-extensive form:

$$\begin{aligned}
&v^{ak_abk_b}_{ik_ijk_j} - P_{i/j}\sum_{k} f^{kk_k}_{ik_i}t^{ak_abk_b}_{kk_kjk_j} + P_{a/b}\sum_{c} f^{ak_a}_{ck_c}t^{ck_cbk_b}_{ik_ijk_j}\\
&\quad + \frac{1}{2}\sum_{k,l}\sum_{k_k} v^{kk_klk_l}_{ik_ijk_j}t^{ak_abk_b}_{kk_klk_l} + \frac{1}{2}\sum_{c,d}\sum_{k_c} v^{ak_abk_b}_{ck_cdk_d}t^{ck_cdk_d}_{ik_ijk_j}\\
&\quad + P_{a/b}P_{i/j}\sum_{k,c}\sum_{k_k} v^{ak_akk_k}_{ik_ick_c}t^{ck_cbk_b}_{kk_kjk_j}\\
&\quad + \frac{1}{4}\sum_{k,l,c,d}\sum_{k_k,k_c} v^{kk_klk_l}_{ck_cdk_d}t^{ck_cdk_d}_{ik_ijk_j}t^{ak_abk_b}_{kk_klk_l}\\
&\quad - \frac{1}{2}P_{a/b}P_{i/j}\sum_{k,l,c,d}\sum_{k_k,k_c} v^{kk_klk_l}_{ck_cdk_d}t^{ak_ack_c}_{kk_kjk_j}t^{dk_dbk_b}_{ik_ilk_l}\\
&\quad - \frac{1}{2}P_{a/b}\sum_{k,l,c,d}\sum_{k_k,k_c} v^{kk_klk_l}_{ck_cdk_d}t^{ck_cbk_b}_{kk_klk_l}t^{ak_adk_d}_{ik_ijk_j}\\
&\quad - \frac{1}{2}P_{i/j}\sum_{k,l,c,d}\sum_{k_k,k_c} v^{kk_klk_l}_{ck_cdk_d}t^{ck_cdk_d}_{kk_kjk_j}t^{ak_abk_b}_{ik_ilk_l} = 0.
\end{aligned} \tag{20}$$

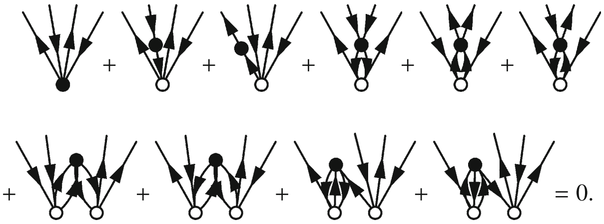
Every term in the left-hand side scales as K^{-1} insofar as the K^{-1} dependence of $t^{ak_abk_b}_{ik_ijk_j}$ is assumed. This last assumption, which is validated by Eq. (20), in turn proves the extensivity of the CCD energy given by Eq. (17).

Equation (17) is diagrammatically expressed by closed, connected diagrams:

$E_{\mathrm{HF}} +$  $= E_{\mathrm{CCD}},$ (21)

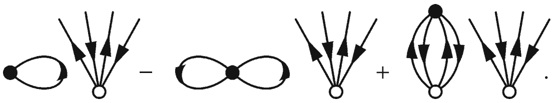
where the open circle with four edges is $t^{ak_abk_b}_{ik_ijk_j}$ and the filled circle is $v^{ik_ijk_j}_{ak_abk_b}$. An upwards (downwards) edge denotes a virtual (occupied) spin-orbital and an edge with ambiguous directionality as in Eq. (10) corresponds to an occupied spin-orbital. Hence, the above diagram graphically represents the contraction of $t^{ak_a bk_b}_{ik_i jk_j}$ and $v^{ik_i jk_j}_{ak_a bk_b}$ with four common indexes, ik_i, jk_j, ak_a, and bk_b, and corresponds to the second term in Eq. (17). For each vertex, we must associate a momentum conservation condition so that the corresponding tensor is non-vanishing. The satisfaction of the momentum conservation condition for $t^{ak_a bk_b}_{ik_i jk_j}$ guarantees the same for $v^{ik_i jk_j}_{ak_a bk_b}$ and vice versa. Generally, n vertexes introduce n momentum conservation conditions. If and only if the diagram is connected, one of these n conditions is redundant and automatically satisfied if the other $n-1$ conditions are satisfied.

The graphical representation of Eq. (20) consists of open diagrams with exactly four external edges. It is given by



(22)

Each diagram is connected. In contrast, the terms that have the incorrect K dependence and have been eliminated after mutual cancelation are disconnected:



(23)

As we have established, each diagram in Eq. (22) scales as K^{-1}, whereas each in Eq. (23) as K^0. We can generalize these observations into the following rule:

Rule 1 (*the K dependence of extensive diagrams*) [5, 7, 8]: A connected diagram with no intensive vertex (see Rule 3 for the definition of an intensive vertex) and n external edges scales as $K^{1-n/2}$. This applies to a single vertex and to a connected diagram that consists of two or more vertexes as well as to a connected subdiagram in an overall disconnected diagram.

Before discussing how this rule comes about, let us apply it to the diagrams in the three equations immediately above. The total energy diagrams in Eqs. (10) and (21) are closed and connected. Thus, according to this rule, they must scale as $K^{1-0/2} = K^1$ and are extensive. All diagrams in Eq. (22) are open, connected diagrams with four external edges. The rule predicts that they scale consistently as $K^{1-4/2} = K^{-1}$. On the other hand, the disconnected diagrams in Eq. (23) are composed of two connected subdiagrams, one scaling as K^1 and the other K^{-1}, according to the rule. They, therefore, inevitably scale as K^0 and differently from open, connected diagrams with the same number of external edges (K^{-1}).

Clearly, size extensivity of a method is judged by inspecting the topology of diagrams that appear in its formalism.

Rule 2 (*the extensive diagram theorem*) [5, 7, 8]: A method is size extensive if the following two conditions are met. Firstly, its expression for an extensive quantity can be cast into the form that consists of only connected diagrams with no intensive vertex. Secondly, its equations that determine excitation amplitudes can be brought to the form that consists of only connected diagrams with no intensive vertex or, less commonly, of only disconnected diagrams with no intensive vertex having an equal number of connected subdiagrams. This criterion is equivalent to that based on the K-dependence criterion.

This is an immediate consequence of Rule 1 and is a sufficient condition for size extensivity. Recognize that an equation can only contain diagrams with an equal number of external edges. The K dependence of a disconnected diagram that consists of m connected subdiagrams with $n_1, \ldots, n_m$ external edges is

$$\prod_{i}^{m} K^{1-n_i/2} = K^{m-n/2}, \tag{24}$$

where n is the total number of external edges of the diagram. Hence, for all diagrams in an equation to have the same K dependence, they may be disconnected but must consist of the same number (m) of connected subdiagrams. When $m = 1$, all diagrams are connected. The terms in an energy expression must be closed ($n = 0$) and scale as K^1 and thus we require $m = 1$ (connected). To avoid misunderstanding, this is not the celebrated linked-cluster theorem of Goldstone [23], which is certainly based on the special case ($m = 1$) of the above rule, but aims at proving the size extensivity of the Møller–Plesset perturbation (MP) series to all orders.

Let us return to the question of how Rule 1 can be derived [5, 7, 8]. The rule is proved by mathematical induction for connected diagrams with two or more vertexes. The basis is that it holds for single extensive vertexes (see Rule 3 for the definition of an extensive vertex), which is indeed the case with $f^{pk_p}_{qk_q}$ and $v^{pk_p qk_q}_{rk_r sk_s}$. Consider a connected diagram that consists of m vertexes with orders $e_1, \ldots, e_m$ and n external edges. The number of internal edges l is given by [21]

$$l = \sum_{i=1}^{m} e_i/2 - n/2. \tag{25}$$

There are m momentum conservation conditions associated with m vertexes, but one of them is redundant because of the connectedness. Therefore, among l internal edges and the same number of k vectors, $m-1$ of them are constrained by the $m-1$ non-redundant momentum conservation conditions. Consequently, the number of k vectors that appear as the summation indexes (h) is given by

$$h = l - (m-1). \tag{26}$$

Remembering that a single vertex of order e_i scales as $K^{1-e_i/2}$, we find the exponent of the K dependence of the diagram to be

$$\begin{aligned}\sum_{i=1}^{m}(1-e_i/2)+h &= \sum_{i=1}^{m}(1-e_i/2)\\ &\quad+\sum_{i=1}^{m}e_i/2-n/2-(m-1)=1-n/2.\end{aligned} \tag{27}$$

This proves Rule 1.

3.4 Configuration-interaction doubles

Configuration-interaction doubles (CID) is not size extensive for correlation energies in the ground states [12, 14]. This can be proved readily by slightly modifying the equations in the previous subsection. A CID wave function is written as

$$|\Psi\rangle = \left(1+\hat{T}_2\right)|\Phi_0\rangle, \tag{28}$$

where $|\Phi_0\rangle$ is a HF wave function. Just as in CCD, we can use the intermediate normalization and projection to arrive at the following equations to solve for $\{t^{ak_abk_b}_{ik_ijk_j}\}$ and E_{CID}:

$$\langle\Phi_0|\hat{H}\left(1+\hat{T}_2\right)|\Phi_0\rangle = E_{\mathrm{CID}}, \tag{29}$$

$$\langle\Phi^{ak_abk_b}_{ik_ijk_j}|\hat{H}\left(1+\hat{T}_2\right)|\Phi_0\rangle = E_{\mathrm{CID}}t^{ak_abk_b}_{ik_ijk_j}. \tag{30}$$

Alternatively, one can use the variational principle and minimize $\langle\Psi|\hat{H}|\Psi\rangle/\langle\Psi|\Psi\rangle$ with respect to variation in $\{t^{ak_abk_b}_{ik_ijk_j}\}$. This leads to the identical set of equations as shown above.

Expanded, they become the following sum-of-product expressions:

$$E_{\mathrm{HF}}+\frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a}v^{ik_ijk_j}_{ak_abk_b}t^{ak_abk_b}_{ik_ijk_j}=E_{\mathrm{CID}} \tag{31}$$

and

$$\begin{aligned}&v^{ak_abk_b}_{ik_ijk_j}-P_{i/j}\sum_{k}f^{kk_k}_{ik_i}t^{ak_abk_b}_{kk_kjk_j}+P_{a/b}\sum_{c}f^{ak_a}_{ck_c}t^{ck_cbk_b}_{ik_ijk_j}\\ &\quad+\frac{1}{2}\sum_{k,l}\sum_{k_k}v^{kk_klk_l}_{ik_ijk_j}t^{ak_abk_b}_{kk_klk_l}+\frac{1}{2}\sum_{c,d}\sum_{k_c}v^{ak_abk_b}_{ck_cdk_d}t^{ck_cdk_d}_{ik_ijk_j}\\ &\quad+P_{a/b}P_{i/j}\sum_{k,c}\sum_{k_k}v^{ak_akk_k}_{ik_ick_c}t^{ck_cbk_b}_{kk_kjk_j}\\ &\quad+E_{\mathrm{HF}}t^{ak_abk_b}_{ik_ijk_j}=E_{\mathrm{CID}}t^{ak_abk_b}_{ik_ijk_j}.\end{aligned} \tag{32}$$

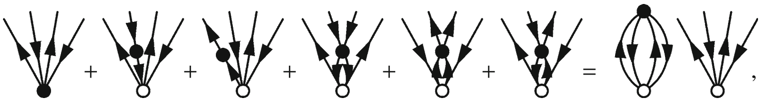
Note that Eq. (31) that defines E_{CID} and Eq. (17) for E_{CCD} have the identical form; hence, the energy expression that arises from the intermediate normalization and projection does not by itself cause the lack of size extensivity in CID. The material difference must exist in the amplitude equations, Eqs. (32) and (18).

Let the asymptotic dependence of $t^{ak_abk_b}_{ik_ijk_j}$ be K^n. The first term in Eq. (32) scales as K^{-1}. The second through the penultimate terms are asymptotically proportional to K^n, whereas the last term and the right-hand side scale as K^{n+1}. In the limit $K\to\infty$, the K^n terms are infinitesimally small as compared to the K^{n+1} terms and hence,

$$\begin{aligned}t^{ak_abk_b}_{ik_ijk_j}&=v^{ak_abk_b}_{ik_ijk_j}/(E_{\mathrm{CID}}-E_{\mathrm{HF}})\\ &=4v^{ak_abk_b}_{ik_ijk_j}\Big/\sum_{k,l,c,d}\sum_{k_k,k_l,k_c}v^{kk_klk_l}_{ck_cdk_d}t^{ck_cdk_d}_{kk_klk_l},\end{aligned} \tag{33}$$

where we have used Eq. (31) in the second equality. The leftmost of this equation scales as K^n by assumption, whereas the rightmost as K^{-n-3}. Therefore, $t^{ak_abk_b}_{ik_ijk_j}$ is a $K^{-3/2}$ quantity, an ominous sign as it already violates Rule 1. Substituting this K dependence in Eq. (31), we find the CID correlation energy, $E_{\mathrm{CID}}-E_{\mathrm{HF}}$, to have the non-physical size dependence of $K^{1/2}$ [12, 14]. This proves the lack of size extensivity in CID for correlation energies. In the limit $K\to\infty, E_{\mathrm{CID}}=E_{\mathrm{HF}}$.

The diagrammatic representation of Eq. (31) is identical to Eq. (21) ("CCD" should be read "CID") and is connected. That of Eq. (32) is



(34)

from which mutually canceling contributions have already been deleted. With Rule 2, the lack of size extensivity in this equation can be seen immediately. The disconnected right-hand side scales differently from the connected left-hand side. Does neglecting the right-hand side restore size extensivity in CID? The answer is yes and the method is now called linearized CCD (LCCD) or D-MBPT(∞) [14, 24]. LCCD does not have a wave function, but is clearly size extensive from Rule 2. This underscores the inadequacy of defining size consistency on the basis of multiplicative separability of wave functions because some methods do not even have a wave function.

3.5 Perturbation theories

We partition the Hamiltonian $\hat{H}$ into a zeroth-order part $\hat{H}_0$ and the remainder $\hat{V}$,

$$\hat{H} = \hat{H}_0 + \hat{V}, \tag{35}$$

where $\hat{H}_0$ is chosen such that we have all of its eigenfunctions and eigenvalues that are size consistent. One such $\hat{H}_0$ is the sum of Fock operators, which satisfies

$$\hat{H}_0|\Phi_0\rangle = E_0|\Phi_0\rangle, \tag{36}$$

$$\hat{H}_0|\Phi^{ak_a}_{ik_i}\rangle = \left(E_0 + f^{ak_a}_{ak_a} - f^{ik_i}_{ik_i}\right)|\Phi^{ak_a}_{ik_i}\rangle, \tag{37}$$

$$\hat{H}_0|\Phi^{ak_abk_b}_{ik_ijk_j}\rangle = \left(E_0 + f^{ak_a}_{ak_a} + f^{bk_b}_{bk_b} - f^{ik_i}_{ik_i} - f^{jk_j}_{jk_j}\right)|\Phi^{ak_abk_b}_{ik_ijk_j}\rangle, \tag{38}$$

etc., with E_0 being the sum of all occupied orbital energies,

$$E_0 = \sum_i \sum_{k_i} f^{ik_i}_{ik_i}. \tag{39}$$

Clearly, E_0 is extensive (K^1) and an energy difference (excitation energy) between an excited state and the ground state is intensive (K^0).

We seek approximate solutions of the Schrödinger equation,

$$\hat{H}|\Psi\rangle = E|\Psi\rangle, \tag{40}$$

employing, once again, the intermediate normalization,

$$\langle\Phi_0|\Psi\rangle = 1, \tag{41}$$

which implies

$$\langle\Phi_0|\hat{H}|\Psi\rangle = E, \tag{42}$$

$$\langle\Phi_0|\hat{V}|\Psi\rangle = E - E_0. \tag{43}$$

For future convenience, we introduce the projection operator, $\hat{Q}$, defined by

$$\hat{Q} = 1 - |\Phi_0\rangle\langle\Phi_0|. \tag{44}$$

Using Eq. (35), we can rewrite the Schrödinger equation into the form [5],

$$\left(E_{\mathrm{LS}} - \hat{H}_0\right)|\Psi\rangle = \left(E_{\mathrm{LS}} - E + \hat{V}\right)|\Psi\rangle, \tag{45}$$

which is true for any arbitrary value of the level shift parameter, E_{LS}. This in conjunction with the intermediate normalization leads to the perturbation expansion of the wave function and energy,

$$|\Psi\rangle = |\Phi_0\rangle + |\Psi^{(1)}\rangle + |\Psi^{(2)}\rangle + \ldots, \tag{46}$$

$$E = E_0 + E^{(1)} + E^{(2)} + \ldots, \tag{47}$$

with

$$|\Psi^{(n)}\rangle = \left\{\hat{Q}\left(E_{\mathrm{LS}} - \hat{H}_0\right)^{-1}\hat{Q}\left(E_{\mathrm{LS}} - E + \hat{V}\right)\right\}^n|\Phi_0\rangle, \tag{48}$$

$$E^{(n)} = \langle\Phi_0|\hat{V}|\Psi^{(n-1)}\rangle. \tag{49}$$

One of the important objectives here is to find a suitable value of E_{LS} that ensures rapid convergence of these series, while maintaining the correct size dependence for each truncated sum.

One obvious choice is $E_{\mathrm{LS}} = E$, which defines the Brillouin–Wigner (BW) perturbation theory [5, 12], a method well-known to be not size extensive and that has, therefore, never been widely used. The first- and second-order energies (BW1 and BW2) are respectively given by

$$E^{(1)}_{\mathrm{BW}} = \langle\Phi_0|\hat{V}|\Phi_0\rangle = -\frac{1}{2}\sum_{i,j}\sum_{k_i,k_j} v^{ik_ijk_j}_{ik_ijk_j}, \tag{50}$$

$$\begin{aligned} E^{(2)}_{\mathrm{BW}} &= \frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a} \frac{\langle\Phi_0|\hat{V}|\Phi^{ak_abk_b}_{ik_ijk_j}\rangle\langle\Phi^{ak_abk_b}_{ik_ijk_j}|\hat{V}|\Phi_0\rangle}{E - E_0 + f^{ik_i}_{ik_i} + f^{jk_j}_{jk_j} - f^{ak_a}_{ak_a} - f^{bk_b}_{bk_b}} \\ &= \frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a} \frac{v^{ik_ijk_j}_{ak_abk_b} v^{ak_abk_b}_{ik_ijk_j}}{E - E_0 + f^{ik_i}_{ik_i} + f^{jk_j}_{jk_j} - f^{ak_a}_{ak_a} - f^{bk_b}_{bk_b}}. \end{aligned} \tag{51}$$

The first-order energy, $E^{(1)}_{\mathrm{BW}}$, scales as K^1 (extensive) and, together with E_0, it restores the HF energy. The second-order energy is not well-defined as it contains the unknown E in the denominator. If we make the approximation $E - E_0 \approx E^{(1)}_{\mathrm{BW}} + E^{(2)}_{\mathrm{BW}}$, the denominator scales as K^1, the numerator as K^{-2}, and the three-fold k summation gives rise to a factor of K^3, making the overall scaling of $E^{(2)}_{\mathrm{BW}}$ to be non-physical K^0 [12]. If we instead use $E - E_0 \approx E^{(2)}_{\mathrm{BW}}$, an argument similar to the one used in CID leads to the conclusion that $E^{(2)}_{\mathrm{BW}}$ scales as $K^{1/2}$, which is still non-physical [12]. In either case, the second-order energy per unit cell, $E^{(2)}_{\mathrm{BW}}/K$, tends to zero as $K \to \infty$. Note that the presence of an extensive quantity, $E - E_0$, in the denominator is responsible for the lack of size extensivity.

A more useful choice, which leads to the MP series [25], is $E_{\mathrm{LS}} = E_0$ [5, 12]. The first- and second-order wave functions are given by

$$|\Psi^{(1)}_{\mathrm{MP}}\rangle = \frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a}\frac{|\Phi^{ak_abk_b}_{ik_ijk_j}\rangle\langle\Phi^{ak_abk_b}_{ik_ijk_j}|\hat{V}|\Phi_0\rangle}{f^{ik_i}_{ik_i}+f^{jk_j}_{jk_j}-f^{ak_a}_{ak_a}-f^{bk_b}_{bk_b}}, \tag{52}$$

$$\begin{aligned}|\Psi^{(2)}_{\mathrm{MP}}\rangle &= \frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a}\frac{|\Phi^{ak_abk_b}_{ik_ijk_j}\rangle\langle\Phi^{ak_abk_b}_{ik_ijk_j}|\hat{V}-E^{(1)}_{\mathrm{MP}}|\Psi^{(1)}\rangle}{f^{ik_i}_{ik_i}+f^{jk_j}_{jk_j}-f^{ak_a}_{ak_a}-f^{bk_b}_{bk_b}}\\ &= \frac{1}{16}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a}\sum_{k,l,c,d}\sum_{k_k,k_l,k_c}\frac{|\Phi^{ak_abk_b}_{ik_ijk_j}\rangle\langle\Phi^{ak_abk_b}_{ik_ijk_j}|\hat{V}-E^{(1)}_{\mathrm{MP}}|\Phi^{ck_cdk_d}_{kk_klk_l}\rangle\langle\Phi^{ck_cdk_d}_{kk_klk_l}|\hat{V}|\Phi_0\rangle}{\left(f^{ik_i}_{ik_i}+f^{jk_j}_{jk_j}-f^{ak_a}_{ak_a}-f^{bk_b}_{bk_b}\right)\left(f^{kk_k}_{kk_k}+f^{lk_l}_{lk_l}-f^{ck_c}_{ck_c}-f^{dk_d}_{dk_d}\right)}.\end{aligned} \tag{53}$$

To maintain consistency in the perturbation order, we have used $E = E_0$ in Eq. (52) and $E = E_0 + E^{(1)}_{\mathrm{MP}}$ in Eq. (53). In contrast to BW2, the denominators do not contain an extensive quantity and they scale as K^0.

The corresponding energies [13, 25–29] are

$$E^{(1)}_{\mathrm{MP}} = \langle\Phi_0|\hat{V}|\Phi_0\rangle = -\frac{1}{2}\sum_{i,j}\sum_{k_i,k_j} v^{ik_ijk_j}_{ik_ijk_j}, \tag{54}$$

$$E^{(2)}_{\mathrm{MP}} = \frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a}\frac{\langle\Phi_0|\hat{V}|\Phi^{ak_abk_b}_{ik_ijk_j}\rangle\langle\Phi^{ak_abk_b}_{ik_ijk_j}|\hat{V}|\Phi_0\rangle}{f^{ik_i}_{ik_i}+f^{jk_j}_{jk_j}-f^{ak_a}_{ak_a}-f^{bk_b}_{bk_b}} = \frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a}\frac{v^{ik_ijk_j}_{ak_abk_b}v^{ak_abk_b}_{ik_ijk_j}}{f^{ik_i}_{ik_i}+f^{jk_j}_{jk_j}-f^{ak_a}_{ak_a}-f^{bk_b}_{bk_b}}, \tag{55}$$

and

$$\begin{aligned}E^{(3)}_{\mathrm{MP}} &= \frac{1}{16}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a}\sum_{k,l,c,d}\sum_{k_k,k_l,k_c}\frac{\langle\Phi_0|\hat{V}|\Phi^{ak_abk_b}_{ik_ijk_j}\rangle\langle\Phi^{ak_abk_b}_{ik_ijk_j}|\hat{V}-E^{(1)}_{\mathrm{MP}}|\Phi^{ck_cdk_d}_{kk_klk_l}\rangle\langle\Phi^{ck_cdk_d}_{kk_klk_l}|\hat{V}|\Phi_0\rangle}{\left(f^{ik_i}_{ik_i}+f^{jk_j}_{jk_j}-f^{ak_a}_{ak_a}-f^{bk_b}_{bk_b}\right)\left(f^{kk_k}_{kk_k}+f^{lk_l}_{lk_l}-f^{ck_c}_{ck_c}-f^{dk_d}_{dk_d}\right)}\\ &= \frac{1}{8}\sum_{i,j,k,l,a,b}\sum_{k_i,k_j,k_k,k_a}\frac{v^{ik_ijk_j}_{ak_abk_b}v^{kk_klk_l}_{ik_ijk_j}v^{ak_abk_b}_{kk_klk_l}}{\left(f^{ik_i}_{ik_i}+f^{jk_j}_{jk_j}-f^{ak_a}_{ak_a}-f^{bk_b}_{bk_b}\right)\left(f^{kk_k}_{kk_k}+f^{lk_l}_{lk_l}-f^{ak_a}_{ak_a}-f^{bk_b}_{bk_b}\right)}\\ &\quad+\frac{1}{8}\sum_{i,j,a,b,c,d}\sum_{k_i,k_j,k_a,k_c}\frac{v^{ik_ijk_j}_{ak_abk_b}v^{ak_abk_b}_{ck_cdk_d}v^{ck_cdk_d}_{ik_ijk_j}}{\left(f^{ik_i}_{ik_i}+f^{jk_j}_{jk_j}-f^{ak_a}_{ak_a}-f^{bk_b}_{bk_b}\right)\left(f^{ik_i}_{ik_i}+f^{jk_j}_{jk_j}-f^{ck_c}_{ck_c}-f^{dk_d}_{dk_d}\right)}\\ &\quad+\sum_{i,j,k,a,b,c}\sum_{k_i,k_j,k_k,k_a}\frac{v^{ik_ijk_j}_{ak_abk_b}v^{kk_kbk_b}_{ck_cjk_j}v^{ak_ack_c}_{ik_ikk_k}}{\left(f^{ik_i}_{ik_i}+f^{jk_j}_{jk_j}-f^{ak_a}_{ak_a}-f^{bk_b}_{bk_b}\right)\left(f^{ik_i}_{ik_i}+f^{kk_k}_{kk_k}-f^{ak_a}_{ak_a}-f^{ck_c}_{ck_c}\right)}.\end{aligned} \tag{56}$$

In the second equality of Eq. (56), we have used the fact that the following term cancels exactly with a contribution with the same magnitude and opposite sign:

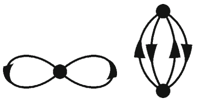
$$-\frac{E_{\mathrm{MP}}^{(1)}}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a}\frac{\langle\Phi_0|\hat{V}|\Phi_{ik_ijk_j}^{ak_abk_b}\rangle\langle\Phi_{ik_ijk_j}^{ak_abk_b}|\hat{V}|\Phi_0\rangle}{\left(f_{ik_i}^{ik_i}+f_{jk_j}^{jk_j}-f_{ak_a}^{ak_a}-f_{bk_b}^{bk_b}\right)^2}. \qquad (57)$$

It is straightforward to confirm the K^1 dependence of $E_{\mathrm{MP}}^{(1)}$, $E_{\mathrm{MP}}^{(2)}$, and $E_{\mathrm{MP}}^{(3)}$ and the non-physical K^2 dependence of Eq. (57).

Diagrammatically, the above equations are expressed as

$E_{\mathrm{MP}}^{(1)} = -$ , $E_{\mathrm{MP}}^{(2)} =$ , $E_{\mathrm{MP}}^{(3)} =$  +  + , (58)

which are connected. The canceling contribution, on the other hand, is disconnected as

. (59)

Hence, the correlation energies from MP1, MP2, and MP3 are extensive. Systematic cancelation of disconnected terms in the MP series at all orders was speculated by Brueckner [30] and proved by Goldstone [23]. The proof is known as the linked-cluster theorem [23], of which the time-independent and time-dependent proofs are found in Shavitt and Bartlett [12] and in March et al. [5] respectively.

The low-order MP correlation energies can be recovered by making a systematic approximation to CCD, Eqs. (17) and (20). The correlation energies are given by the identical formula, but with approximate amplitudes,

$$\frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a} v_{ak_abk_b}^{ik_ijk_j}\,{}^{(n-1)}t_{ik_ijk_j}^{ak_abk_b} = E_{\mathrm{MP}}^{(n)}, \quad n = 2,3, \qquad (60)$$

where ${}^{(n-1)}t_{ik_ijk_j}^{ak_abk_b}$ can be obtained recursively by solving the following equation, which linearizes the CCD equation, namely, drops the terms with more than one amplitudes:

$$\begin{aligned} & v_{ik_ijk_j}^{ak_abk_b} - P_{i/j}\sum_k f_{ik_i}^{kk_k}\,{}^{(n)}t_{kk_kjk_j}^{ak_abk_b} + P_{a/b}\sum_c f_{ck_c}^{ak_a}\,{}^{(n)}t_{ik_ijk_j}^{ck_cbk_b} \\ & \quad + \frac{1}{2}\sum_{k,l}\sum_{k_k} v_{ik_ijk_j}^{kk_klk_l}\,{}^{(n-1)}t_{kk_klk_l}^{ak_abk_b} \\ & \quad + \frac{1}{2}\sum_{c,d}\sum_{k_c} v_{ck_cdk_d}^{ak_abk_b}\,{}^{(n-1)}t_{ik_ijk_j}^{ck_cdk_d} \\ & \quad + P_{a/b}P_{i/j}\sum_{k,c}\sum_{k_k} v_{ik_ick_c}^{ak_akk_k}\,{}^{(n-1)}t_{kk_kjk_j}^{ck_cbk_b} = 0 \end{aligned} \qquad (61)$$

and ${}^{(0)}t_{kk_kjk_j}^{ak_abk_b} = 0$. In this viewpoint, unlike the conventional one based on Rayleigh–Schrödinger perturbation theory, the low-order MP wave functions are understood to have the same exponential structure as CCD and the diagrammatic energy expression is manifestly connected.

3.6 Configuration-interaction singles for excited states

An analysis analogous to Sect. 3.4 for configuration-interaction singles (CIS) [31, 32] reveals that its total energy for the ground state using a HF wave function as the reference is the total energy of HF itself. Thus, CIS captures no electron correlation. This must, however, be distinguished from CID, whose correlation energy decreases as $K^{-1/2}$. The CIS correlation energy, unlike the CID counterpart, remains zero regardless of size and its K dependence is not incorrect, though it may be considered irrelevant. In this sense, CIS is size extensive for total energies of the ground states.

A CIS excited-state wave function is defined by

$$|\Psi\rangle = \hat{C}_1|\Phi_0\rangle, \qquad (62)$$

where $\hat{C}_1$ is a one-electron excitation operator,

$$\hat{C}_1|\Phi_0\rangle = \sum_{i,a}\sum_{k_i} c_{ik_i}^{ak_a}|\Phi_{ik_i}^{ak_a}\rangle. \qquad (63)$$

The excited-state wave function, $|\Psi\rangle$, and all singly excited determinants used as the expansion basis must have the same total wave vector,

$$\kappa \equiv k_a - k_i. \qquad (64)$$

When $\kappa = 0$, the transition is called direct and, otherwise, indirect. Equation (62) is incompatible with the intermediate normalization and so we must use the standard normalization,

$$1 = \langle\Phi_0|\hat{C}_1^\dagger\hat{C}_1|\Phi_0\rangle = \sum_{i,a}\sum_{k_i}|c_{ik_i}^{ak_a}|^2. \qquad (65)$$

This indicates that $c_{ik_i}^{ak_a}$ is a $K^{-1/2}$ quantity. The above equation is diagrammed as

$1 =$ , (66) where we have used a different symbol (open square) for the vertex of $c_{ik_i}^{ak_a}$ to underscore the difference in K dependence of the amplitudes between intensive and extensive operators. We define these two types of operators and the corresponding vertexes and amplitudes:

Rule 3 (*intensive and extensive operators*): An intensive operator is the one whose amplitudes are normalized, while the amplitudes of an extensive operator are intermediately normalized. Their corresponding vertexes and amplitudes are called intensive and extensive vertexes and amplitudes. The Hamiltonian is an extensive operator.

Throughout this article, we associate letter T with extensive operators and C with intensive operators. We can generalize Rule 1 as follows:

Rule 4 (*the K dependence of intensive diagrams*): A connected diagram with n external edges, m intensive vertexes, and any number of extensive vertexes scales as $K^{1-n/2-m/2}$.

The open diagram of $c_{ik_i}^{ak_a}$ has two external edges and scales as $K^{-1/2}$ and hence it satisfies Rule 4. Note also that the open diagram of $t_{ik_ijk_j}^{ak_abk_b}$ in CID scales as $K^{-3/2}$ and, therefore, satisfies Rule 4 but not Rule 1. This indicates that its $\hat{T}_2$ is in fact an intensive operator and unsuitable for describing extensive correlation effects in the ground state, as we have found.

The projection or the variational principle leads to the equation to be solved for $c_{ik_i}^{ak_a}$:

$$\langle \Phi_{ik_i}^{ak_a} | \hat{H}\hat{C}_1 | \Phi_0 \rangle = E_{\text{CIS}}^{(n)} c_{ik_i}^{ak_a}, \quad (67)$$

where n labels an electronic state with $n = 0$ corresponding to the ground state. Expanding this, we arrive at the sum-of-product expression,

$$-\sum_j f_{ik_i}^{jk_j} c_{jk_j}^{ak_a} + \sum_b f_{bk_b}^{ak_a} c_{ik_i}^{bk_b} + \sum_{j,b} \sum_{k_j} v_{ik_ibk_b}^{ak_ajk_j} c_{jk_j}^{bk_b} + E_{\text{HF}} c_{ik_i}^{ak_a} = E_{\text{CIS}}^{(n)} c_{ik_i}^{ak_a}, \quad (68)$$

where

$$k_b - k_j = k_a - k_i \equiv \kappa \quad (69)$$

needs to be satisfied in the third term to prevent $v_{ik_ibk_b}^{ak_ajk_j}$ from vanishing and, similarly, $k_a - k_j = k_b - k_i = \kappa$ in the first two terms. The terms in the CIS equation as shown above have inconsistent K dependence: The first through third terms scale as $K^{-1/2}$, whereas the last term and the right-hand side as $K^{1/2}$. This inconsistency can be readily resolved by noticing

$$E_{\text{CIS}}^{(n)} = E_{\text{CIS}}^{(0)} + \omega_{\text{CIS}}^{(n)}, \quad (70)$$

$$E_{\text{CIS}}^{(0)} = E_{\text{HF}}, \quad (71)$$

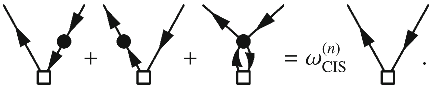
where $\omega_{\text{CIS}}^{(n)}$ is an excitation energy of the nth state, which should be a K^0 quantity. Using this, we can rewrite Eq. (68) to

$$-\sum_j f_{ik_i}^{jk_j} c_{jk_j}^{ak_a} + \sum_b f_{bk_b}^{ak_a} c_{ik_i}^{bk_b} + \sum_{j,b} \sum_{k_j} v_{ik_ibk_b}^{ak_ajk_j} c_{jk_j}^{bk_b} = \omega_{\text{CIS}}^{(n)} c_{ik_i}^{ak_a}, \quad (72)$$

which has the consistent $K^{-1/2}$ dependence. Thus, CIS is size intensive (K^0) for $\omega_{\text{CIS}}^{(n)}$ and size extensive (K^1) for $E_{\text{CIS}}^{(n)}$. Multiplying the above equation by $c_{ik_i}^{ak_a *}$ and summing over i, a, and k_i, we obtain

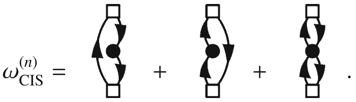
$$\omega_{\text{CIS}}^{(n)} = -\sum_{i,j,a} \sum_{k_i} c_{ik_i}^{ak_a *} f_{ik_i}^{jk_j} c_{jk_j}^{ak_a} + \sum_{i,a,b} \sum_{k_i} c_{ik_i}^{ak_a *} f_{bk_b}^{ak_a} c_{ik_i}^{bk_b} + \sum_{i,j,a,b} \sum_{k_i,k_j} c_{ik_i}^{ak_a *} v_{ik_ibk_b}^{ak_ajk_j} c_{jk_j}^{bk_b}. \quad (73)$$

Diagrammatically, Eq. (72) is



(74)

All the diagrams scale consistently as $K^{-1/2}$, even though the right-hand side is disconnected. The diagrams for Eq. (73) are connected and scale correctly as K^0:



(75)

We state the following rule for size intensivity.

Rule 5 (*the intensive diagram theorem*): A method is size intensive if the following two conditions are met. Firstly, its expression for an intensive quantity can be cast into the form that consists of connected diagrams with exactly two intensive vertexes. Secondly, its equations that determine intensive (excitation, ionization, electron-attachment, etc.) amplitudes can be brought to the form that consists of connected diagrams with exactly one intensive vertex or of disconnected diagrams with one open, connected subdiagram having one intensive vertex plus any number of closed, connected subdiagrams having two intensive vertexes.

This is a sufficient, but more restrictive condition for size intensivity than the corresponding K-dependence criterion. The derivation of the above rule proceeds as follows. From Rule 4, we find that a closed, connected diagram with two intensive vertexes scales as $K^{1-0/2-2/2} = K^0$ and is intensive, which corresponds to the second sentence of the above rule. An open, connected diagram with one intensive vertex scales as $K^{1-n/2-1/2} = K^{1/2-n/2}$, where n is the number of external edges. An amplitude equation that contains only such diagrams should, therefore, have the consistent K dependence of $K^{1/2-n/2}$. Furthermore, one can attach any number of closed, connected subdiagrams with two intensive vertexes (K^0 quantities) to each of these open diagrams without altering their $K^{1/2-n/2}$ scaling. Size-intensive methods can thus have disconnected diagrams and they usually do.

For an excited state accessible by a direct transition ($\kappa = 0$), one may choose to include $|\Phi_0\rangle$ among the expansion basis,

$$|\Psi\rangle = c_0|\Phi_0\rangle + \sum_{i,a}\sum_{k_i} c^{ak_a}_{ik_i}|\Phi^{ak_a}_{ik_i}\rangle, \tag{76}$$

where c_0 scales as K^0 regardless of whether the intermediate normalization ($c_0 = 1$) or the standard normalization,

$$1 = |c_0|^2 + \sum_{i,a}\sum_{k_i} |c^{ak_a}_{ik_i}|^2, \tag{77}$$

is adopted. The ground state is now included in the determinant space for the diagonalization of the Hamiltonian matrix. The projection and the variation both lead to the same set of equations,

$$\sum_{i,a}\sum_{k_i} f^{ik_i}_{ak_a} c^{ak_a}_{ik_i} = \omega^{(n)}_{\mathrm{CIS}} c_0, \tag{78}$$

and

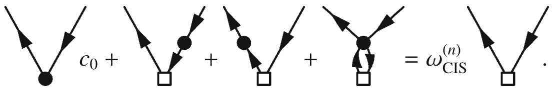
$$f^{ak_a}_{ik_i} c_0 - \sum_{j} f^{jk_j}_{ik_i} c^{ak_a}_{jk_j} + \sum_{b} f^{ak_a}_{bk_b} c^{bk_b}_{ik_i} + \sum_{j,b}\sum_{k_j} v^{ak_ajk_j}_{ik_ibk_b} c^{bk_b}_{jk_j} = \omega^{(n)}_{\mathrm{CIS}} c^{ak_a}_{ik_i}, \tag{79}$$

where we have used $E^{(n)}_{\mathrm{CIS}} = E_{\mathrm{HF}} + \omega^{(n)}_{\mathrm{CIS}}$. The right-hand side of Eq. (78) scales as K^0. For the left-hand side to scale in the same way, $c^{ak_a}_{ik_i}$ needs to be a K^{-1} quantity. This causes the first term in Eq. (79) to scale differently (as K^0) from the rest of the equation (as K^{-1}). To restore consistent K dependence, we must, therefore, require that $f^{ak_a}_{ik_i} = f^{ik_i}_{ak_a} = 0$ (Brillouin's theorem), namely, that the reference wave function is HF [14].

The diagrammatic representations of these equations are

$= \omega^{(n)}_{\mathrm{CIS}} c_0$ (80)

and



$c_0 + \quad + \quad + \quad = \omega^{(n)}_{\mathrm{CIS}}$. (81)

In the size-consistency analysis, the presence of c_0 can be ignored because of its K^0 dependence. The left-hand side of the first equation is connected, but it has only one intensive vertex and thus violates the intensive diagram theorem. The first diagram in the second equation, while connected, has no intensive vertex and does not conform to the theorem, either. Therefore, they apparently do not define a size-intensive method unless these violating diagrams are made to vanish by, in this case, the HF reference and Brillouin's theorem.

3.7 Coupled-cluster doubles for excited states

Let us turn to the question of size intensivity of excited-state CCD, namely, equation-of-motion CCD (EOM-CCD) [11, 12, 33–38]. Although an excited-state method without a single excitation operator is not useful in practice, we use this simple method to clarify its size dependence. From the discussion in Sect. 3.3, we can formulate CCD in a manifestly connected form:

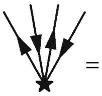
$$\langle\Phi_0|\bar{H}|\Phi_0\rangle = E^{(0)}_{\mathrm{CCD}}, \tag{82}$$

$$\langle\Phi^{ak_abk_b}_{ik_ijk_j}|\bar{H}|\Phi_0\rangle = 0, \tag{83}$$

with

$$\bar{H} = \left[\hat{H}\exp\left(\hat{T}_2\right)\right]_{\mathrm{C}}, \tag{84}$$

where the subscript "C" means that the operators in the bracket are diagrammatically connected and the superscript "(0)" indicates that it pertains to the ground state. Using pentagram symbols to denote vertexes of $\bar{H}$, we can express Eq. (83) diagrammatically as



$= 0.$ (85)

EOM-CCD is a CID using this CCD effective Hamiltonian, $\bar{H}$. Thus, we solve

$$\langle\Phi^{ak_abk_b}_{ik_ijk_j}|\bar{H}\hat{C}_2|\Phi_0\rangle = E^{(n)}_{\mathrm{CCD}} c^{ak_abk_b}_{ik_ijk_j}, \tag{86}$$

for $E^{(n)}_{\mathrm{CCD}}$ and $\{c^{ak_abk_b}_{ik_ijk_j}\}$, where n labels an electronic state with $n = 0$ corresponding to the ground state. The intensive amplitudes are subject to the standard normalization:

$$\sum_{i,j,a,b}\sum_{k_i,k_j,k_a} |c^{ak_abk_b}_{ik_ijk_j}|^2 = 1. \tag{87}$$

Hence, $c^{ak_abk_b}_{ik_ijk_j}$ is a $K^{-3/2}$ quantity. Equation (86) is expanded as

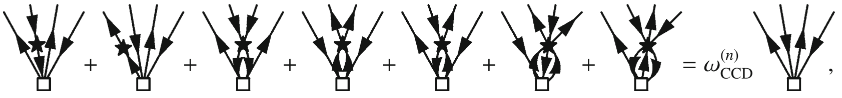
$$\begin{aligned}
&\bar{v}^{ak_abk_b}_{ik_ijk_j} - P_{i/j}\sum_{k}\bar{f}^{kk_k}_{ik_i}c^{ak_abk_b}_{kk_kjk_j} + P_{a/b}\sum_{c}\bar{f}^{ak_a}_{ck_c}c^{ck_cbk_b}_{ik_ijk_j}\\
&\quad+\frac{1}{2}\sum_{k,l}\sum_{k_k}\bar{v}^{kk_klk_l}_{ik_ijk_j}c^{ak_abk_b}_{kk_klk_l} + \frac{1}{2}\sum_{c,d}\sum_{k_c}\bar{v}^{ak_abk_b}_{ck_cdk_d}c^{ck_cdk_d}_{ik_ijk_j}\\
&\quad+P_{a/b}P_{i/j}\sum_{k,c}\sum_{k_k}\bar{v}^{ak_akk_k}_{ik_ick_c}c^{ck_cbk_b}_{kk_kjk_j}\\
&\quad+\frac{1}{2}P_{i/j}\sum_{k,c,d}\sum_{k_k,k_c}\bar{w}^{ak_abk_bkk_k}_{ik_ick_cdk_d}c^{ck_cdk_d}_{jk_jkk_k}\\
&\quad-\frac{1}{2}P_{a/b}\sum_{k,l,c}\sum_{k_k,k_l}\bar{w}^{ak_akk_klk_l}_{ik_ijk_jck_c}c^{bk_bck_c}_{kk_klk_l}\\
&\quad+E^{(0)}_{\text{CCD}}c^{ak_abk_b}_{ik_ijk_j} = E^{(n)}_{\text{CCD}}c^{ak_abk_b}_{ik_ijk_j},
\end{aligned}\tag{88}$$

where the integrals with overbars come from $\bar{H}$, but, since they are connected, they scale exactly the same way as the counterparts without overbars, according to Rule 1. Hence, $\bar{f}^{pk_p}_{qk_q}$, $\bar{v}^{pk_pqk_q}_{rk_rsk_s}$, and $\bar{w}^{pk_pqk_qrk_r}_{sk_stk_tuk_u}$ scale as K^0, K^{-1}, and K^{-2}, respectively. Note that CCD's $\bar{H}$ (unlike $\hat{H}$) has up to a six-electron excitation operator and up to a two-electron deexcitation operator. The three-electron excitation operator in $\bar{H}$ gives rise to the three-electron integrals denoted by $\bar{w}$.

This equation does not display consistent scaling. The first term in the left-hand side scales as K^{-1} and the second through the penultimate terms as $K^{-3/2}$, whereas the last term and the right-hand side are $K^{-1/2}$ quantities. The inconsistencies are resolved by using Eq. (85), namely, $\bar{v}^{ak_abk_b}_{ik_ijk_j} = 0$, as well as $E^{(n)}_{\text{CCD}} - E^{(0)}_{\text{CCD}} = \omega^{(n)}_{\text{CCD}}$. Thus,

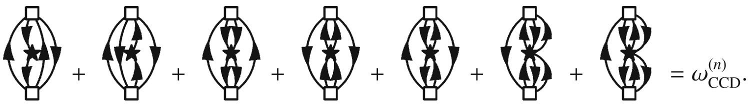
$$\begin{aligned}
&- P_{i/j}\sum_{k}\bar{f}^{kk_k}_{ik_i}c^{ak_abk_b}_{kk_kjk_j} + P_{a/b}\sum_{c}\bar{f}^{ak_a}_{ck_c}c^{ck_cbk_b}_{ik_ijk_j}\\
&\quad+\frac{1}{2}\sum_{k,l}\sum_{k_k}\bar{v}^{kk_klk_l}_{ik_ijk_j}c^{ak_abk_b}_{kk_klk_l} + \frac{1}{2}\sum_{c,d}\sum_{k_c}\bar{v}^{ak_abk_b}_{ck_cdk_d}c^{ck_cdk_d}_{ik_ijk_j}\\
&\quad+P_{a/b}P_{i/j}\sum_{k,c}\sum_{k_k}\bar{v}^{ak_akk_k}_{ik_ick_c}c^{ck_cbk_b}_{kk_kjk_j}\\
&\quad+\frac{1}{2}P_{i/j}\sum_{k,c,d}\sum_{k_k,k_c}\bar{w}^{ak_abk_bkk_k}_{ik_ick_cdk_d}c^{ck_cdk_d}_{jk_jkk_k}\\
&\quad-\frac{1}{2}P_{a/b}\sum_{k,l,c}\sum_{k_k,k_l}\bar{w}^{ak_akk_klk_l}_{ik_ijk_jck_c}c^{bk_bck_c}_{kk_klk_l} = \omega^{(n)}_{\text{CCD}}c^{ak_abk_b}_{ik_ijk_j},
\end{aligned}\tag{89}$$

which scales consistently as $K^{-3/2}$. Diagrammatically, this becomes



(90)

where

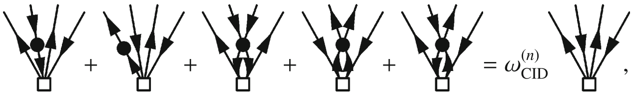


(91)

These diagrammatic equations satisfy the intensive diagram theorem. Hence, EOM-CCD is size extensive for the total energy, $E^{(n)}_{\text{CCD}}$, and size intensive for the excitation energy, $\omega^{(n)}_{\text{CCD}}$.

3.8 Configuration-interaction doubles for excited states

Is CID size intensive for excitation energies? Excluding $|\Phi_0\rangle$ from the basis, we arrive at the CID equation for an excitation energy, $\omega^{(n)}_{\text{CID}} = E^{(n)}_{\text{CID}} - E_{\text{HF}}$, which can be expressed diagrammatically as

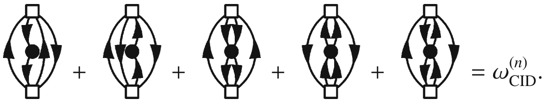


(92)

where an open square represents $c^{ak_abk_b}_{ik_ijk_j}$, which should now be normalized as

$$\sum_{i,j,a,b}\sum_{k_i,k_j,k_a}|c^{ak_abk_b}_{ik_ijk_j}|^2 = 1 \tag{93}$$

and be asymptotically proportional to $K^{-3/2}$. Closing Eq. (92) by the vertex of $c^{ak_abk_b*}_{ik_ijk_j}$, we obtain the diagrammatic expression for $\omega^{(n)}_{\text{CID}}$,



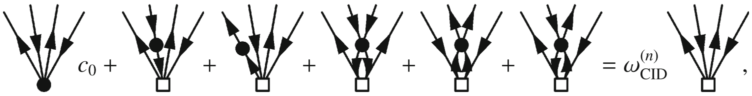
(94)

Understandably, these diagrammatic equations are subsets of the corresponding ones of EOM-CCD with each pentagram vertex (representing parts of $\bar{H}$) replaced by a filled circle vertex ($\hat{H}$) and, therefore, satisfy the intensive diagram theorem. CID is in fact size intensive for excitation energies insofar as $|\Phi_0\rangle$ is not among the basis functions.

If $|\Phi_0\rangle$ is included, we must solve two sets of equations instead of one. They are

$$= \omega^{(n)}_{\mathrm{CID}} c_0, \tag{95}$$

and



(96)

both of which violate the intensive diagram theorem. CID is, therefore, not size intensive if $|\Phi_0\rangle$ is among the basis functions. The violation arises from the fact that the equality analogous to Eq. (85) in EOM-CCD or to $f^{ak_a}_{ik_i} = f^{ik_i}_{ak_a} = 0$ in CIS with the HF reference does not hold in CID. In other words, $\hat{C}_2$ is responsible for simultaneously capturing both the extensive correlation energy and intensive excitation energies, which is not achievable in a size-consistent fashion.

4 Size-consistent design

The extensive and intensive diagram theorems and underlying K-dependence criterion provide an unambiguous, foolproof, and convenient basis on which to decide whether a method is size consistent, once its formalism—not an implementation—is given. However, it does not necessarily offer any physical insight as to why some choices of approximations ensure cancelation of terms that violate the criterion. In this section, we consider this question of size-consistent design and make several concrete propositions.

4.1 Size extensivity

Let us recognize that the correlation energies of CCD, MP2, and CID are all expressed in the identical form,

$$E = E_{\mathrm{HF}} + \frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a} v^{ik_ijk_j}_{ak_abk_b} t^{ak_abk_b}_{ik_ijk_j}, \tag{97}$$

which is diagrammatically connected. This energy expression is a consequence of the intermediate normalization,

$$|\Psi\rangle = \left(1 + \hat{T}_2\right)|\Phi_0\rangle, \tag{98}$$

and the associated projection. The MP1 and CID wave functions share the same structure given above. While the CCD wave function has the different, exponential form,

$$|\Psi\rangle = \exp\left(\hat{T}_2\right)|\Phi_0\rangle, \tag{99}$$

as far as the derivation of the above energy expression is concerned, one might as well define it by Eq. (98). Thus, the lack of size extensivity in CID cannot be inferred from the above energy expression alone. It primarily comes from the amplitude equations, which, in conjunction with the energy expression, cause the amplitudes of $\hat{T}_2$ to have the incorrect K dependence as an extensive operator. It is also wrong to conclude that MP2 is not size extensive for correlation energies just because the corresponding MP1 wave function has a CID structure. In fact, CCD and MP2 are size extensive for correlation energies given by Eq. (97) because their $t^{ak_abk_b}_{ik_ijk_j}$ scales correctly as K^{-1} (Rule 1). CID is not size extensive for correlation energies because the corresponding amplitudes scale incorrectly as $K^{-3/2}$.

Let us consider an alternative energy expression suggested by the variational principle,

$$E = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle}. \tag{100}$$

Substituting the MP1 wave function into the numerator, we obtain [5]

$$\langle\Psi|\hat{H}|\Psi\rangle = \langle\Phi_0|\left(1 + \hat{T}_2^{\dagger}\right)\hat{H}\left(1 + \hat{T}_2\right)|\Phi_0\rangle \tag{101}$$

$$= \langle\Phi_0|\hat{H}|\Phi_0\rangle + 2\langle\Phi_0|\hat{H}\hat{T}_2|\Phi_0\rangle + \langle\Phi_0|\hat{T}_2^{\dagger}\hat{H}\hat{T}_2|\Phi_0\rangle \tag{102}$$

$$= E_{\mathrm{HF}} + 2E^{(2)}_{\mathrm{MP}} + \langle\Phi_0|\hat{H}|\Phi_0\rangle\langle\Phi_0|\hat{T}_2^{\dagger}\hat{T}_2|\Phi_0\rangle + \langle\Phi_0|\hat{T}_2^{\dagger}\hat{H}\hat{T}_2|\Phi_0\rangle_{\mathrm{C}} \tag{103}$$

$$= E_{\mathrm{HF}}\left(1 + \langle\Phi_0|\hat{T}_2^{\dagger}\hat{T}_2|\Phi_0\rangle\right) + 2E^{(2)}_{\mathrm{MP}} + \langle\Phi_0|\hat{T}_2^{\dagger}\hat{V}\hat{T}_2|\Phi_0\rangle_{\mathrm{C}} + \langle\Phi_0|\hat{T}_2^{\dagger}\hat{H}_0\hat{T}_2|\Phi_0\rangle_{\mathrm{C}} \tag{104}$$

$$= E_{\mathrm{HF}}\left(1 + \langle\Phi_0|\hat{T}_2^{\dagger}\hat{T}_2|\Phi_0\rangle\right) + E^{(2)}_{\mathrm{MP}} + E^{(3)}_{\mathrm{MP}}, \tag{105}$$

where, in the last equality, the following relationships [5] are used:

$$\langle\Phi_0|\hat{T}_2^{\dagger}\hat{V}\hat{T}_2|\Phi_0\rangle_{\mathrm{C}} = \langle\Phi_0|\hat{V}\frac{\hat{Q}}{E_0 - \hat{H}_0}\hat{V}\frac{\hat{Q}}{E_0 - \hat{H}_0}\hat{V}|\Phi_0\rangle_{\mathrm{C}} = E^{(3)}_{\mathrm{MP}}, \tag{106}$$

$$\langle\Phi_0|\hat{T}_2^{\dagger}\hat{H}_0\hat{T}_2|\Phi_0\rangle_{\mathrm{C}} = \langle\Phi_0|\hat{T}_2^{\dagger}\left(\hat{H}_0 - E_0\right)\hat{T}_2|\Phi_0\rangle_{\mathrm{C}} = -\langle\Phi_0|\hat{V}\frac{\hat{Q}}{E_0 - \hat{H}_0}\hat{V}|\Phi_0\rangle_{\mathrm{C}} = -E^{(2)}_{\mathrm{MP}}. \tag{107}$$

With the same substitution, the denominator is

$$\langle\Psi|\Psi\rangle = 1 + \langle\Phi_0|\hat{T}_2^\dagger\hat{T}_2|\Phi_0\rangle = 1 + \frac{1}{4}\sum_{i,j,a,b}\sum_{k_i,k_j,k_a}|t_{ik_ijk_j}^{ak_abk_b}|^2, \qquad (108)$$

which scales as K^1 because $t_{ik_ijk_j}^{ak_abk_b}$ is a K^{-1} quantity. Hence, the correlation contribution in

$$E = E_{\mathrm{HF}} + \frac{E_{\mathrm{MP}}^{(2)} + E_{\mathrm{MP}}^{(3)}}{1 + \langle\Phi_0|\hat{T}_2^\dagger\hat{T}_2|\Phi_0\rangle} \qquad (109)$$

displays the non-physical K^0 dependence because both the numerator and denominator scale as K^1. With the standard normalization and expectation value, the MP1 wave function no longer gives extensive correlation energies even though $t_{ik_ijk_j}^{ak_abk_b}$ scales correctly!

When we substitute the CCD wave function into the expectation value, Eq. (100), we find [39]

$$E = \frac{\langle\Phi_0|\exp\left(\hat{T}_2^\dagger\right)\hat{H}\exp\left(\hat{T}_2\right)|\Phi_0\rangle}{\langle\Phi_0|\exp\left(\hat{T}_2^\dagger\right)\exp\left(\hat{T}_2\right)|\Phi_0\rangle} \qquad (110)$$

$$= E_{\mathrm{HF}} + \langle\Phi_0|\exp\left(\hat{T}_2^\dagger\right)\hat{H}\exp\left(\hat{T}_2\right)|\Phi_0\rangle_{\mathrm{C}}, \qquad (111)$$

which is manifestly diagrammatically connected and yields an extensive value when $t_{ik_ijk_j}^{ak_abk_b}$ scales correctly as K^{-1}, though it differs from the value obtained by projection, Eq. (97). CCD is, therefore, size extensive for correlation energies regardless of whether they are defined by projection or as expectation values.

In CID, the projection and expectation values coincide. Hence,

$$E = \frac{\langle\Phi_0|\left(1 + \hat{T}_2^\dagger\right)\hat{H}\left(1 + \hat{T}_2\right)|\Phi_0\rangle}{\langle\Phi_0|\left(1 + \hat{T}_2^\dagger\right)\left(1 + \hat{T}_2\right)|\Phi_0\rangle} \qquad (112)$$

$$= E_{\mathrm{HF}} + \langle\Phi_0|\hat{H}\hat{T}_2|\Phi_0\rangle, \qquad (113)$$

where we have used Eqs. (29) and (30) in the second equality. Since $t_{ik_ijk_j}^{ak_abk_b}$ determined by CID does not exhibit the correct scaling, the correlation energies are not extensive regardless. These observations are summarized in Table 1 and lead us to the following proposition:

Table 1 Size extensivity of various methods for correlation energies in the ground states

	CCD	MP2	CID
Intermediate normalization and projection	yes	yes	no
Normalization and expectation value	yes	no	no

Proposition 1 (*the intermediate normalization and size extensivity*): *The intermediate normalization and associated projection ensure a formally extensive energy expression.*

The amplitudes that enter this expression, however, must also scale correctly with K (according to Rule 1) for the energy to be indeed extensive, but this is not guaranteed by the intermediate normalization.

What follows is an argument that supports the use of the intermediate normalization for extensive quantities. Let us consider a periodic extended system in which there are no inter-cell interactions [40]. Let us also assume that the reference wave function is HF and thus invariant to rotation among occupied orbitals and among virtual orbitals. The wave function in the intermediate normalization, Eq. (98), becomes

$$|\Psi\rangle = |\Phi_0\rangle + \sum_{n=1}^{K}|\Phi_0^{[1]}\cdots\hat{T}_2^{[n]}\Phi_0^{[n]}\cdots\Phi_0^{[K]}\rangle, \qquad (114)$$

where $|\Phi_0^{[n]}\rangle$ is a HF wave function of the nth unit cell composed of the Wannier orbitals localized around that cell,

$$\eta_p(\boldsymbol{r} - n\boldsymbol{a}) = K^{-1/2}\sum_{k_p}\exp(-ink_pa)\varphi_{pk_p}(\boldsymbol{r}), \qquad (115)$$

and $\hat{T}_2^{[n]}$ is a double excitation operator acting only on the function in the nth cell. In the basis of the Wannier orbitals, the amplitudes of $\hat{T}_2^{[n]}$ and two-electron integrals are given by

$$t_{ij}^{ab} = K^{-2}\sum_{k_i,k_j,k_a}t_{ik_ijk_j}^{ak_abk_b}, \qquad (116)$$

$$v_{ab}^{ij} = K^{-2}\sum_{k_i,k_j,k_a}v_{ak_abk_b}^{ik_ijk_j}. \qquad (117)$$

The factors of K^{-2} in these equations originate from the normalization coefficient $K^{-1/2}$ in Eq. (115). The quantities in the left-hand sides do not carry any cell index because of the periodicity of the extended system and the absence of any inter-cell interactions. The correct scaling of these local quantities is K^0 and this is satisfied insofar as $t_{ik_ijk_j}^{ak_abk_b}$ scales correctly as K^{-1}.

Since the Hamiltonian is additively separable in this case,

$$\hat{H} = \sum_{n=1}^{K}\hat{H}^{[n]}, \qquad (118)$$

the total energy obtained by projection is given by

$$E = \langle\Phi_0|\hat{H}|\Psi\rangle = E_{\mathrm{HF}} + \sum_{n=1}^{K}\langle\Phi_0^{[n]}|\hat{H}^{[n]}\hat{T}_2^{[n]}|\Phi_0^{[n]}\rangle$$

$$= E_{\mathrm{HF}} + \frac{K}{4}\sum_{i,j,a,b}v_{ab}^{ij}t_{ij}^{ab}, \qquad (119)$$

where we have used $\hat{H}^{[m]}|\Psi^{[n]}\rangle = 0$ if $m \neq n$. The correlation energy (the very last term) is manifestly extensive. This logic can be readily generalized to the situation in which there are inter-cell interactions. The intermediate normalization and projection, therefore, implicitly exploit the locality of chemical interactions (see the next section) and thereby ensure extensive energy expressions, even though its wave function, Eq. (114), is not multiplicatively separable.

The expectation value of the Hamiltonian in the above wave function yields

$$E = E_{\mathrm{HF}} + \frac{X}{Y}, \tag{120}$$

with

$$X = \frac{K}{4}\sum_{i,j,a,b} v^{ij}_{ab} t^{ab}_{ij} + \frac{K}{8}\sum_{i,j,k,l,a,b} t^{ab*}_{ij} v^{kl}_{ij} t^{ab}_{kl} + \frac{K}{8}\sum_{i,j,a,b,c,d} t^{ab*}_{ij} v^{ab}_{cd} t^{cd}_{ij} + K \sum_{i,j,k,a,b,c} t^{ab*}_{ij} v^{kb}_{cj} t^{ac}_{ik}, \tag{121}$$

$$Y = 1 + \frac{K}{4}\sum_{i,j,a,b} |t^{ab}_{ij}|^2.$$

The correlation energy, therefore, scales as non-physical K^0, merely confirming what we have already established in Eq. (109). This observation also suggests the following:

Proposition 2 (*the standard normalization and size intensivity*): *The use of the standard normalization and the expectation value of the Hamiltonian is suitable for describing an intensive quantity.*

The first two propositions are simply the converse of Rule 3.

An intermediately normalized, exponential wave function of the same system is given by

$$|\Psi\rangle = \prod_{n=1}^{K} |\Psi^{[n]}\rangle, \tag{122}$$

$$|\Psi^{[n]}\rangle = \exp\left(\hat{T}_2^{[n]}\right)|\Phi_0^{[n]}\rangle. \tag{123}$$

The overall wave function, $|\Psi\rangle$, is multiplicatively separable into individual unit-cell wave functions, $\{|\Phi_0^{[n]}\rangle\}$. The projection of this onto the reference wave function leads to the identical energy expression given above as Eq. (119), which is extensive. The expectation value of the Hamiltonian in this wave function is also extensive, which can be verified easily:

$$E = \frac{\sum_{n=1}^{K} \langle\Psi^{[1]}|\Psi^{[1]}\rangle \ldots \langle\Psi^{[n]}|\hat{H}^{[n]}|\Psi^{[n]}\rangle \ldots \langle\Psi^{[K]}|\Psi^{[K]}\rangle}{\prod_{n=1}^{K}\langle\Psi^{[n]}|\Psi^{[n]}\rangle} = \sum_{n=1}^{K} \frac{\langle\Psi^{[n]}|\hat{H}^{[n]}|\Psi^{[n]}\rangle}{\langle\Psi^{[n]}|\Psi^{[n]}\rangle}. \tag{124}$$

What would be a guideline for obtaining the correctly scaling amplitudes of an extensive operator? An observation from the analyses in the previous section is the following:

Proposition 3 (*non-canceling extensive quantities in amplitudes*): *The presence of a non-canceling extensive quantity in the equation that determines extensive amplitudes tends to cause the violation of size extensivity.*

For example, the BW1 wave function has a correlation energy of the whole system in the denominator of its definition, which causes the corresponding BW2 energy, Eq. (51), to be non-extensive. Likewise, the MP2 wave function has an extensive correlation energy in the numerator of its definition, Eq. (53), giving rise to the term, though canceling out eventually, that scales incorrectly. Diagrammatically, any non-canceling extensive quantity becomes a closed subdiagram. A multiplication or a division by such quantities leads to disconnected diagrams with closed subdiagrams (which are called "unlinked"). This inevitably violates the extensive diagram theorem (Rule 2).

4.2 Size intensivity

Let us define two complementary spaces of determinants, which prove useful in the following discussion.

Definition 6 (*extensive and intensive determinant spaces*): The extensive determinant space of an operator, $\hat{C}$, is the space spanned by the determinants reached by the action of $\hat{C}^{\dagger}\hat{C}$ on $|\Phi_0\rangle$. The intensive determinant space of $\hat{C}$ is the space of the determinants accessible by $\hat{C}$ from $|\Phi_0\rangle$.

For instance, the extensive determinant space of $\hat{C}_0 + \hat{C}_2$ consists of $|\Phi_0\rangle$ and all doubly excited determinants, where $\hat{C}_n$ denotes an n-electron excitation operator. The extensive determinant space of $\hat{C}_2$ is just $|\Phi_0\rangle$.

CIS, EOM-CCD, and CID can all be cast into a CI problem with a diagrammatically connected, generalized Hamiltonian, $\tilde{H}$,

$$\hat{P}\tilde{H}\hat{C}|\Phi_0\rangle = \hat{P}E\hat{C}|\Phi_0\rangle, \tag{125}$$

where $\hat{P}$ is the projector onto the intensive determinant space of $\hat{C}$. In CIS, $\tilde{H} = \hat{H}$ and $\hat{C} = \hat{C}_0 + \hat{C}_1$ or $\hat{C} = \hat{C}_1$. Either of these definitions of $\hat{C}$ leads to size-intensive CIS with the HF reference. In EOM-CCD, $\tilde{H} = \bar{H}$ and $\hat{C} = \hat{C}_0 + \hat{C}_2$ or $\hat{C} = \hat{C}_2$. By virtue of Eq. (85), EOM-CCD is also size intensive with either choice of $\hat{C}$. On the other hand, CID ($\tilde{H} = \hat{H}$) has been found to be size intensive for excitation energies when $\hat{C} = \hat{C}_2$, but not so when

$\hat{C} = \hat{C}_0 + \hat{C}_2$. What is then the general guideline to obtain a size-intensive excited-state method?

Notice that the only way in which the left-hand side of Eq. (125) is diagrammatically disconnected is

$$\left(\hat{P}_x\tilde{H}|\Phi_0\rangle\right)\left(\hat{P}_y\hat{C}|\Phi_0\rangle\right), \qquad (126)$$

where $\hat{P} = \hat{P}_x + \hat{P}_y$ because $\tilde{H}$ is assumed to be by itself diagrammatically connected. Since $\hat{P}_y$ can be a projector onto any subspace of the intensive determinant space, $\hat{P}_x$ is necessarily the projector onto the extensive determinant space of $\hat{C}$ defined above.

The analyses of size intensivity in the previous section suggest that, in size-intensive formalisms, these disconnected contributions must vanish, namely, $\hat{P}_x\tilde{H}|\Phi_0\rangle = 0$ except in the space of $|\Phi_0\rangle$. In the latter space, $\tilde{H}|\Phi_0\rangle$ can remain nonzero, yielding the following disconnected contribution:

$$\left(\langle\Phi_0|\tilde{H}|\Phi_0\rangle\right)\left(\hat{P}\hat{C}|\Phi_0\rangle\right). \qquad (127)$$

In other words, the CI problem must have the trivial solution, $\hat{C} = 1$, corresponding to the ground state, within the extensive determinant space of $\hat{C}$,

$$\hat{P}_x\tilde{H}|\Phi_0\rangle = \hat{P}_x E_0|\Phi_0\rangle, \qquad (128)$$

where $E_0 = \langle\Phi_0|\tilde{H}|\Phi_0\rangle$.

Assuming that Eq. (128) is satisfied and subtracting Eq. (127) from Eq. (125), we obtain

$$\hat{P}\tilde{H}\hat{C}|\Phi_0\rangle_{\mathrm{C}} = \omega\hat{P}\hat{C}|\Phi_0\rangle, \qquad (129)$$

where $\omega = E - E_0$ and subscript "C" means that the left-hand side is diagrammatically connected. This furthermore leads to

$$\langle\Phi_0|\hat{C}^{\dagger}\tilde{H}\hat{C}|\Phi_0\rangle_{\mathrm{C}} = \omega, \qquad (130)$$

where the amplitudes of the intensive operator, $\hat{C}$, are assumed to be normalized in accordance with Proposition 2. These equations succinctly outline the structures of size-intensive CIS and EOM-CCD equations and satisfy, in the most general way, the intensive diagram theorem. We thus summarize this observation as follows:

Rule 6 (*the extensive-intensive consistency theorem*): When a method for excited, ionized, or electron-attached states is defined as a CI problem in the intensive determinant space of $\hat{C}$ with some general, diagrammatically connected Hamiltonian, $\tilde{H}$, for the method to be size intensive, the CI problem must have the trivial solution, $\hat{C} = 1$ or $\tilde{H}|\Phi_0\rangle = E_0|\Phi_0\rangle$, corresponding to the ground state, within the extensive determinant space of $\hat{C}$.

That this condition is met by CIS and EOM-CCD can be readily verified. Let us consider EOM-CCD. If $|\Phi_0\rangle$ is not included in the intensive determinant space, namely, $\hat{C} = \hat{C}_2$, the extensive determinant space is $|\Phi_0\rangle$ and $\bar{H}|\Phi_0\rangle = E_{\mathrm{CCD}}|\Phi_0\rangle$ is satisfied in this space. If $|\Phi_0\rangle$ is instead included in the diagonalization space, the extensive determinant space becomes $|\Phi_0\rangle$ plus all doubly excited determinants. In this space, $\bar{H}|\Phi_0\rangle = E_{\mathrm{CCD}}|\Phi_0\rangle$ amounts exactly to the CCD energy and amplitude equations, Eqs. (82) and (83), which are again satisfied. EOM-CCD is, therefore, size intensive for excitation energies regardless of whether $\hat{C} = \hat{C}_2$ or $\hat{C} = \hat{C}_0 + \hat{C}_2$. CIS with the HF reference can likewise be shown to be size intensive with or without $|\Phi_0\rangle$ in the intensive determinant space.

Let us consider CID next. If $|\Phi_0\rangle$ is not included in the intensive determinant space, the extensive determinant space is $|\Phi_0\rangle$ and $\hat{H}|\Phi_0\rangle = E_{\mathrm{HF}}|\Phi_0\rangle$ is a trivial solution within that space. CID is, therefore, size intensive with this choice of the intensive determinant space. If $|\Phi_0\rangle$ is included, the extensive determinant space is $|\Phi_0\rangle$ plus all doubly excited determinants. In this space, $\hat{C} = 1$ is not a solution of the CI equation. Hence, CID is not size intensive in the latter choice of the intensive determinant space.

Let us define EOM-MP2 by EOM-CCD with the approximation, $t_{ik_ijk_j}^{ak_abk_b} \approx {}^{(1)}t_{ik_ijk_j}^{ak_abk_b}$; see Eqs. (60) and (61). This method is, again, introduced for pedagogical purposes as it does not have any single excitation operator. $\{{}^{(1)}t_{ik_ijk_j}^{ak_abk_b}\}$ are not the solutions of the CCD equations. For EOM-MP2 to be size intensive, therefore, $|\Phi_0\rangle$ needs to be excluded from the intensive determinant space so that the extensive determinant space becomes just $|\Phi_0\rangle$.

Some approximations of EOM-CC have been proposed in which the spaces of the determinants accessible by $\hat{T}$ and $\hat{C}$ do not match [41]. For instance, a CI problem using the CCSD effective Hamiltonian, $\bar{H}$, is solved in the space of $|\Phi_0\rangle$ and its singly, doubly, and triply excited determinants: $\hat{C} = \hat{C}_0 + \hat{C}_1 + \hat{C}_2 + \hat{C}_3$. The extensive determinant space of $\hat{C}$ consists of $|\Phi_0\rangle$ and its singly, doubly, and triply excited states. Since the CCSD equation, $\bar{H}|\Phi_0\rangle = E_{\mathrm{CCSD}}|\Phi_0\rangle$, is satisfied only up to the space of doubly excited determinants, but not in the space of triply excited determinants, this method is not size intensive. When, on the other hand, $\hat{C}_0$ is excluded from $\hat{C}$, the extensive determinant space becomes the one accessible by $\hat{C}_0 + \hat{C}_1 + \hat{C}_2$ from $|\Phi_0\rangle$. With this choice of $\hat{C}$, the method is size intensive.

EOM-CC has been applied to ionized [42, 43], electron-attached [42, 44], and even spin-flipped states [45]. For instance, ionization-potential EOM-CCSD (IP-EOM-CCSD) [43] solves a CI problem with the CCSD effective Hamiltonian within the space of determinants obtained by one-hole (1*h*) and two-hole-one-particle (2*h*1*p*) creation in $|\Phi_0\rangle$.

The extensive determinant space of the corresponding hole-particle operator is $|\Phi_0\rangle$ plus all singly excited determinants, in which the CCSD equation, $\bar{H}|\Phi_0\rangle = E_{\mathrm{CCSD}}|\Phi_0\rangle$, is satisfied. Therefore, IP-EOM-CCSD(1*h*,2*h*1*p*) is size intensive for ionization energies. IP-EOM-CCSD (1*h*, 2*h*1*p*, 3*h*2*p*) is also size intensive as its extensive determinant space is $|\Phi_0\rangle$ plus the set of all singly and doubly excited determinants, in which the CCSD equation is still satisfied.

In physical terms, what Rule 6 suggests is as follows:

Proposition 4 (*the separation of extensive and intensive quantities*): *The correlation energy in the ground state should be exhausted by an extensive operator in the determinant space in which an intensive operator subsequently describes excitations or other intensive processes. If the space acted on by the intensive operator is greater than that by the extensive operator, the intensive operator can be used to capture simultaneously the extensive correlation energy in the ground state and intensive transition energies. If this happens, the method tends to lack size consistency.*

Therefore, the order in which extensive and intensive operators are applied matters; the application of an extensive operator must precede that of an intensive operator, as may also be suggested by linear and nonlinear response theories [34, 35, 37, 46]. A number of methods seem to violate this provision and, consequently, not size consistent [47–54].

5 The origin of extensivity

Extensivity of energies of usual chemical systems is an experimental observation and has been treated as a postulate in this article and in thermochemistry [1]. It is by no means a trivial property. Outside and within chemistry, one can find systems whose energies are not necessarily extensive. For example, the energy of a uniform mass density bound by gravitational forces increases asymptotically as $V^{5/3}$ with volume V in three dimensions, as $V^{3/2}$ in two dimensions, and as $V \ln V$ in one dimension, all of which are non-extensive. The energy of a uniform electron gas without a neutralizing positive charge background also grows as $V^{5/3}$ in three dimensions. These counterexamples suggest that the extensivity of energy or the existence of a thermodynamic (infinite-volume) limit of energy density [15–20] relies on some tacit assumptions. Most importantly, the system must be electrically neutral for the thermodynamic limit to exist [17].

With this assumption, we can divide the effective interactions between two unit cells in an infinitely extended, electrically neutral, periodic, non-metallic solid into the electrostatic (E_{E}), exchange (E_{X}), and correlation (E_{C}) interactions. We can show [55–60] that these three interactions display asymptotic distance (r) decay that can be expressed as

$$E_{\mathrm{E}} = (\boldsymbol{\mu}\cdot\boldsymbol{\mu})r^{-3} - 3(\boldsymbol{\mu}\cdot\boldsymbol{r})^2 r^{-5}, \tag{131}$$

$$E_{\mathrm{X}} = C_1 \exp(-C_2 r), \tag{132}$$

$$E_{\mathrm{C}} = C_3 r^{-6}, \tag{133}$$

where $\boldsymbol{\mu}$ is the permanent dipole moment of the unit cell, C_1 is a constant, C_2 is a positive constant, and C_3 is a negative constant. Since the unit cells are neutral, the leading electrostatic interaction is of dipole–dipole type as shown in the first of these equations [9, 55].

The exchange [56–58] and correlation [59, 60] interactions decay sufficiently rapidly that their lattice sums in one, two, and three-dimensional solids should converge at finite, intensive values, which are the respective energies per unit cell. The electrostatic interaction, as given in Eq. (131), asymptotically decays as r^{-3} and appears to cause a logarithmic divergence in its lattice sum of a three-dimensional solid. However, this can be shown to be avoided by rearranging the lattice sum spherically and approximating its long-range part by the integral,

$$\begin{aligned}&\frac{1}{V_0}\int_{r_0}^{\infty}\int_0^{\pi}\int_0^{2\pi}\left\{(\boldsymbol{\mu}\cdot\boldsymbol{\mu})r^{-3} - 3(\boldsymbol{\mu}\cdot\boldsymbol{r})^2 r^{-5}\right\}r^2 dr \sin\theta d\theta d\phi \\ &= \frac{4\pi}{V_0}\int_{r_0}^{\infty}\mu^2 r^{-1} dr \\ &\quad - \frac{2\pi}{V_0}\int_{r_0}^{\infty}\mu^2 r^{-1} dr \int_0^{\pi} 3\cos^2\theta \sin\theta d\theta = 0,\end{aligned} \tag{134}$$

where V_0 is the volume of the unit cell and r_0 is the onset of the asymptotic regime. This proves the intensivity of the energy per unit cell and the extensivity of the energy of a non-metallic solid with electrically neutral unit cells. We, thus, state the following rule:

Rule 7 (*the thermodynamic interaction asymptote theorem*): The energy of a periodic system of interacting particles is extensive and obeys the laws of thermodynamics if the effective inter-cell interaction is pairwise and decays asymptotically as $(1 - 3\cos^2\theta)r^{-3}$ or more rapidly with distance r and inclination θ.

This is a weaker (thus more preferable) condition than the one introduced earlier stating that the asymptote of $r^{-3-\epsilon}$ with $\epsilon > 0$ is necessary to ensure the existence of a thermodynamic limit [15, 19]. It is also only a sufficient condition and, if not satisfied, can mean that the energy is non-extensive.

Rule 8 (*the existence of a thermodynamic limit of energy density*): The effective inter-cell interactions in an electrically neutral, periodic, non-metallic solid decay at least as rapidly as the thermodynamic interaction asymptote and, therefore, its energy is extensive.

The exchange interactions in metals decay much more slowly than Eq. (132) and may become dominant in the limit $r \to \infty$. Therefore, the above argument leading to Rule 8 does not apply to metals. The proof of the extensivity of energies for metals will be given in a forthcoming article [61].

The foregoing is a simpler, but more restricted proof of the existence of a thermodynamic limit of energy density. This question has been posed in rigorous mathematical terms and proven by Lebowitz and Lieb for real matter consisting of mobile electrons and nuclei [17], by Fefferman for perfect crystals [62], and by Hainzl et al. for crystals with defects [63, 64].

6 Size consistency of the HF theory redux

If a total energy is not necessarily extensive, how can various electronic structure methods be shown to be size consistent? Let us reexamine the K dependence of $f_{qk_q}^{pk_p}$ and $v_{rk_rsk_s}^{pk_pqk_q}$, upon which the foregoing analysis on size consistency is based. They are defined by

$$f_{qk_q}^{pk_p} = -\frac{1}{2}\int \varphi_{pk_p}^{*}(\boldsymbol{r})\nabla^2\varphi_{qk_q}(\boldsymbol{r})d\boldsymbol{r} + \sum_{i}\sum_{k_i} v_{qk_qik_i}^{pk_pik_i}, \tag{135}$$

$$v_{rk_rsk_s}^{pk_pqk_q} = \iint \frac{\tilde{\zeta}_{pk_prk_r}(\boldsymbol{r}_1)\tilde{\zeta}_{qk_qsk_s}(\boldsymbol{r}_2)}{r_{12}}d\boldsymbol{r}_1 d\boldsymbol{r}_2 - \iint \frac{\zeta_{pk_psk_s}(\boldsymbol{r}_1)\zeta_{qk_qrk_r}(\boldsymbol{r}_2)}{r_{12}}d\boldsymbol{r}_1 d\boldsymbol{r}_2, \tag{136}$$

where the orbital-pair densities, $\tilde{\zeta}_{pk_pqk_q}(\boldsymbol{r})$ and $\zeta_{pk_pqk_q}(\boldsymbol{r})$, are given by

$$\zeta_{pk_pqk_q}(\boldsymbol{r}) = \varphi_{pk_p}^{*}(\boldsymbol{r})\varphi_{qk_q}(\boldsymbol{r}), \tag{137}$$

$$\tilde{\zeta}_{pk_pqk_q}(\boldsymbol{r}) = \zeta_{pk_pqk_q}(\boldsymbol{r}) - \delta_{pq}\delta_{k_pk_q}\frac{\rho_{\mathrm{N}}(\boldsymbol{r})}{NK}, \tag{138}$$

with the nuclear charge density,

$$\rho_{\mathrm{N}}(\boldsymbol{r}) = \sum_{I}\sum_{m} Z_I \delta(\boldsymbol{r} - \boldsymbol{r}_I + m\boldsymbol{a}). \tag{139}$$

Here, N is the number of electrons per unit cell, I labels the nucleus with charge Z_I centered at $\boldsymbol{r}_I$, and m is the unit-cell index. Note that the second term in the right-hand side of Eq. (138) ensures the charge neutrality of $\tilde{\zeta}_{pk_pqk_q}(\boldsymbol{r})$,

$$\int \tilde{\zeta}_{pk_pqk_q}(\boldsymbol{r})d\boldsymbol{r} = 0. \tag{140}$$

Therefore, $\tilde{\zeta}_{pk_pqk_q}(\boldsymbol{r})$ can have unit-cell dipole and higher moments, but not unit-cell charges. The magnitudes of these dipole and higher multipole moments are proportional to K^{-1} because the orbitals are normalized in the whole volume. Both terms of Eq. (136) for the types of $v_{rk_rsk_s}^{pk_pqk_q}$ that appear in HF and electron-correlation methods physically represent sums of electrostatic interactions in the lattice of the dipole and higher multipole moments. According to Rule 7, these lattice sums are extensive and thus scale as $(K^{-1})^2K^1 = K^{-1}$. This proves the K^{-1} dependence of $v_{rk_rsk_s}^{pk_pqk_q}$.

The kinetic-energy integrals in the definition of $f_{qk_q}^{pk_p}$ scale as K^0, which can be inferred from the fact that the orbitals are normalized in the whole volume. The second term is also a K^0 quantity because of the K^{-1} dependence of $v_{qk_qik_i}^{pk_pik_i}$. This proves the K^0 dependence of $f_{qk_q}^{pk_p}$. Substituting Eqs. (135) and (136) into the HF energy expression, Eq. (9), we recover the correct result,

$$\begin{aligned} E_{\mathrm{HF}} = &-\frac{1}{2}\sum_{i}\sum_{k_i}\int \varphi_{ik_i}^{*}(\boldsymbol{r})\nabla^2\varphi_{ik_i}(\boldsymbol{r})d\boldsymbol{r} \\ &+\frac{1}{2}\sum_{i,j}\sum_{k_i,k_j}\iint \frac{|\varphi_{ik_i}(\boldsymbol{r}_1)|^2|\varphi_{jk_j}(\boldsymbol{r}_2)|^2}{r_{12}}d\boldsymbol{r}_1 d\boldsymbol{r}_2 \\ &-\sum_{i}\sum_{k_i}\sum_{I}\sum_{m}\int \frac{Z_I|\varphi_{ik_i}(\boldsymbol{r})|^2}{|\boldsymbol{r} - \boldsymbol{r}_I + m\boldsymbol{a}|}d\boldsymbol{r} \\ &+\frac{1}{2}\sum_{I,J}\sum_{m_1,m_2}\frac{Z_IZ_J}{|\boldsymbol{r}_I - m_1\boldsymbol{a} - \boldsymbol{r}_J + m_2\boldsymbol{a}|} \\ &-\frac{1}{2}\sum_{i,j}\sum_{k_i,k_j}\iint \frac{\varphi_{ik_i}^{*}(\boldsymbol{r}_1)\varphi_{jk_j}(\boldsymbol{r}_1)\varphi_{jk_j}^{*}(\boldsymbol{r}_2)\varphi_{ik_i}(\boldsymbol{r}_2)}{r_{12}}d\boldsymbol{r}_1 d\boldsymbol{r}_2. \end{aligned} \tag{141}$$

It should be understood that the contributions with vanishing denominators are excluded from the nuclear-repulsion energy (the fourth term). Three observations are in order. Firstly, terms in this energy expression are not necessarily individually extensive; The second, third, and fourth terms are non-extensive and only their sum is possibly extensive. Secondly, the particular definitions of $f_{qk_q}^{pk_p}$ and $v_{rk_rsk_s}^{pk_pqk_q}$ given in Eqs. (135) and (136) are crucial for establishing the size extensivity of HF and electron-correlation methods and they depend critically on the local electrical neutrality expected in a system with an extensive total energy. Thirdly, these definitions of $f_{qk_q}^{pk_p}$ and $v_{rk_rsk_s}^{pk_pqk_q}$ differ from the conventional definitions widely adopted by chemists, if not from those of physicists. This last observation leads to the following proposition:

Proposition 5 (*the size-consistent Hamiltonian*): *The correct definitions of the one- and two-electron integrals in the Hamiltonian are given by Eqs.* (135)–(138). *For size consistency, the HF equations must be revised with these definitions.*

These new, correct definitions of $f_{qk_q}^{pk_p}$ and $v_{rk_rsk_s}^{pk_pqk_q}$ do not alter the HF energy, HF molecular orbitals, or correlation energies, but they make the lattice sums in $f_{qk_q}^{pk_p}$ converge much more rapidly [65].

Acknowledgments Dr. Rodney J. Bartlett (Gainesville), Dr. Marcel Nooijen (Waterloo), and Dr. Yu-ya Ohnishi (Urbana) are thanked for their critical reading of this manuscript and many useful comments. Dr. David Chandler (Berkeley) and Dr. Elliott H. Lieb (Princeton) are thanked for their assistance with the literature on the extensivity of energy. This work has been supported by the National Science Foundation (CHE-1118616 and OCI-1102418) and the Department of Energy (DE-FG02-11ER16211). The author is a Camille Dreyfus Teacher-Scholar and an Alumni Research Scholar of University of Illinois.

References

1. Chandler D (1987) Introduction to modern statistical mechanics. Oxford University Press, New York
2. Bartlett RJ (1981) Ann Rev Phys Chem 32:359
3. Nooijen M, Shamasundar KR, Mukherjee D (2005) Mol Phys 103:2277
4. Hirata S, Keçeli M, Yagi K (2010) J Chem Phys 133:034109
5. March NH, Young WH, Sampanthar S (1967) The many-body problem in quantum mechanics. Cambridge University Press, Cambridge
6. Deleuze M, Delhalle J, Pickup BT, Calais J-L (1992) Phys Rev B 46:15668
7. Deleuze M, Delhalle J, Pickup BT, Calais J-L (1995) Adv Quantum Chem 26:35
8. Nooijen M, Bartlett RJ (1997) Int J Quantum Chem 63:601
9. Hirata S (2009) Phys Chem Chem Phys 11:8397
10. Bartlett RJ, Purvis GD (1978) Int J Quantum Chem 14:561
11. Bartlett RJ, Musiał M (2007) Rev Mod Phys 79:291
12. Shavitt I, Bartlett RJ (2009) Many-body methods in chemistry and physics. Cambridge University Press, Cambridge
13. Pople JA, Binkley JS, Seeger R (1976) Int J Quantum Chem S10:1
14. Szabo A, Ostlund NS (1989) Modern quantum chemistry. McGraw-Hill, New York
15. Fisher ME, Ruelle D (1966) J Math Phys 7:260
16. Dyson FJ, Lenard A (1967) J Math Phys 8:423
17. Lebowitz JL, Lieb EH (1969) Phys Rev Lett 22:631
18. Lieb EH, Lebowitz JL (1972) Adv Math 9:316
19. Lieb EH (1976) Rev Mod Phys 48:553
20. Thirring W (1990) Foundations Phys 20:1103
21. Chartrand G (1985) Introductory graph theory. Dover, Mineola
22. Pople JA, Krishnan R, Schlegel HB, Binkley JS (1978) Int J Quantum Chem 14:545
23. Goldstone J (1957) Proc Roy Soc London A 239:267
24. Bartlett RJ, Shavitt I, Purvis GD (1979) J Chem Phys 71:281
25. Møller C, Plesset MS (1934) Phys Rev 46:618
26. Bartlett RJ, Silver DM (1975) J Chem Phys 62:3258
27. Kunz AB (1972) Phys Rev B 6:606
28. Suhai S (1983) Chem Phys Lett 96:619
29. Suhai S (1983) Phys Rev B 27:3506
30. Brueckner KA (1955) Phys Rev 100:36
31. Foresman JB, Head-Gordon M, Pople JA, Frisch MJ (1992) J Phys Chem 96:135
32. Hirata S, Head-Gordon M, Bartlett RJ (1999) J Chem Phys 111:10774
33. Nakatsuji H, Hirao K (1978) J Chem Phys 68:2053
34. Koch H, Jørgensen P (1990) J Chem Phys 93:3333
35. Koch H, Jensen HJA, Jørgensen P, Helgaker T (1990) J Chem Phys 93:3345
36. Stanton JF, Bartlett RJ (1993) J Chem Phys 98:7029
37. Rico RJ, Head-Gordon M (1993) Chem Phys Lett 213:224
38. Katagiri H (2005) J Chem Phys 122:224901
39. Čížek J (1969) Adv Chem Phys 14:35
40. Taylor PR (1994) In: Roos BO (ed) Lecture notes in quantum chemistry II. Springer, Berlin, pp 125–202
41. Nakatsuji H (1991) Chem Phys Lett 177:331
42. Nakatsuji H (1991) J Chem Phys 94:6716
43. Stanton JF, Gauss J (1994) J Chem Phys 101:8938
44. Nooijen M, Bartlett RJ (1995) J Chem Phys 102:3629
45. Krylov AI (2006) Acc Chem Res 39:83
46. Hirata S, Fan P-D, Shiozaki T, Shigeta Y (2008) In: Leszczynski J, Shukla M (eds) Radiation induced molecular phenomena in nucleic acid: a comprehensive theoretical and experimental analysis. Springer, Berlin, pp 15–64
47. Hubač I (1996) In: Tsipis A, Popov VS, Herschbach DR, Avery JS (eds) New methods in quantum chemistry. Kluwer, Dordrecht, pp 183–202
48. Mášik J, Hubač I (1997) In: McWeeny R, Maruani J, Smeyers YG, Wilson S (eds) Quantum systems in chemistry and physics: trends in methods and applications. Kluwer, Dordrecht, pp 283–308
49. Hubač I, Mášik J, Mach P, Urban J, Babinec P (1999) In: Leszczynski J (eds) Computational chemistry: reviews of current trends. World Scientific, Singapore, pp 1–48
50. Mášik J, Hubač I (1999) Adv Quantum Chem 31:75
51. Kowalski K, Piecuch P (2000) J Chem Phys 113:5644
52. Kowalski K, Piecuch P (2001) Chem Phys Lett 344:165
53. Hirata S, Nooijen M, Grabowski I, Bartlett RJ (2001) J Chem Phys 114:3919
54. Hirata S, Nooijen M, Grabowski I, Bartlett RJ (2001) J Chem Phys 115:3967
55. Delhalle J, Piela L, Brédas J-L, André J-M (1980) Phys Rev B 22:6254
56. des Cloizeaux J (1964) Phys Rev 135:A685
57. Monkhorst HJ, Kertesz M (1981) Phys Rev B 24:3015
58. Piela L, André J-M, Fripiat JG, Delhalle J (1981) Chem Phys Lett 77:143
59. Ayala P, Scuseria GE (1999) J Chem Phys 110:3660
60. Hirata S, Podeszwa R, Tobita M, Bartlett RJ (2004) J Chem Phys 120:2581
61. Hirata S, Ohnishi Y-Y (2011) unpublished
62. Fefferman C (1985) Comm Math Phys 98:289
63. Hainzl C, Lewin M, Solovej JP (2009) Adv Math 221:454
64. Hainzl C, Lewin M, Solovej JP (2009) Adv Math 221:488
65. Ohnishi Y-Y, Hirata S (2011) unpublished