pubs.acs.org/JACS

Why Is Quantum Chemistry So Complicated?

Jack Simons*

Cite This: J. Am. Che	m. Soc. 2023, 145, 4343–4354	🔇 Read C	Dnline	
ACCESS	LIII Metrics & More		E Article Recommendations	

ABSTRACT: The myriad tools of quantum chemistry are now widely used by a diverse community of chemists, biologists, physicists, and material scientists. The large number of methods (e.g., Hartree-Fock, density functional theory, configuration interaction, perturbation theory, coupled-clusters, equations of motion, Green's functions, and more) and the multitude of atomic orbital basis sets often give rise to consternation and confusion. In this Perspective, I explain why quantum chemistry has so many different methods and why researchers should understand their relative strengths and weaknesses. I explain how chemistry's use of orbitals and the need for wave functions to be antisymmetric causes computational-effort scaling proportional to the cube or higher power of the number of orbitals. I also illustrate how the fact that the Schrödinger equation's energies are extensive makes it difficult to extract intensive properties such as bond and excitation energies, ionization potentials, and electron affinities.

1. INTRODUCTION

In this Perspective, I am not trying to explain things to quantum chemistry (QC) experts. Instead, I am attempting to address experimental colleagues who make use of QC to help interpret their data but who often ask "Why are there so many acronyms, different competing methods, all these different basis sets, and such confusing jargon in your field?", "Why do my calculations take so long?", and "What method should I use?" To answer, I focus on three issues: (i) the extreme difficulties the Schrödinger equation (SE) presents, (ii) numerical precision challenges in connecting the SE's extensive energies to intensive properties of experimental interest, and (iii) how orbitals and antisymmetric wave functions (i.e., Slater determinants) give rise to extreme (third power and higher) scaling in computational effort as the number of orbitals varies. I explain how the computational efforts in various steps of a QC calculation scale, since this knowledge is essential for designing appropriate and practical QC calculations. Finally, I show how one can estimate how the computer time involved in a calculation changes upon moving from one basis set to another within various QC methods.

What is involved when a researcher wants to determine the electronic energy at some specified geometry and in some specified electronic state using QC? In principle, the researcher seeks a solution to the electronic Schrödinger equation:

$$\left[-\frac{1}{2} \sum_{i=1}^{N} \nabla_{r_{i}}^{2} + \sum_{i=1}^{N} \left\{ \sum_{a=1}^{N_{\text{Nucc}}} \frac{-Z_{a}}{|r_{i} - R_{a}|} + \frac{1}{2} \sum_{j=1; \neq i}^{N} \frac{1}{|r_{i} - r_{j}|} \right\} \right] \Psi$$

$$= E \Psi$$
(1)

Here, I use atomic units where the mass of the electron m_{e} , the constant \hbar , and the unit of charge *e* are all of unit magnitude. In such units, the energies occur in Hartrees, each being equivalent to 27.21 eV, r_i is the position of the *i*th electron, Z_a is the charge of the *a*th nucleus located at R_a , N is the number of electrons in the system, and N_{Nuc} is the number of nuclei.

Consider what this entails for an ethylene molecule with 6 nuclei and 16 electrons. This is a second-order differential equation involving $3 \times N = 48$ spatial coordinates. The kinetic energy (first term) and electron-nuclear Coulomb interaction (second term) involve sums, one for each electron. If they were the only factors present, this would allow one to introduce an orbital-product form for the wave function Ψ . However, the electron–electron interaction potential (third term) depends on the positions of pairs of electrons, so such a product form is not entirely appropriate. Thus, one is faced with a non-separable second-order partial differential equation in 48 dimensions with solutions Ψ that are odd under the interchange of any two electrons.

There is not just one solution to this SE; there are infinitely many solutions. One relates to the ground electronic state, others describe excited states (e.g., $\pi\pi^*$, $\sigma\pi^*$, $\sigma\sigma^*$ states, etc.), while others relate to states of the ethylene cation plus a free electron. There are also solutions for doubly, triple, etc. ionized ethylene all the way up to states with the nuclei holding no bound electrons but with 16 free electrons.

This illustrates how overwhelming and essentially impossible the task would be if one were really trying to solve the SE for all of its energies and wave functions, but that is not what is done in practice. Instead, QC introduces tools that allow one to arrive at reasonable approximations to a limited number of electronic states. For example, one might want to focus on the ground electronic state, a few excited electronic states, and the lowestenergy state of the corresponding cation. It is by limiting the scope and by living with limited but reasonable precision that QC provides

Published: February 14, 2023



useful results, but much of QC's complexity is a result of how it goes about doing so.

2. ORBITALS AND ANTISYMMETRY ARE INVOLVED IN THE NUMERICAL COMPLEXITY

If the SE's Hamiltonian were additive, as in

$$\left[-\frac{1}{2}\sum_{i=1}^{N}\nabla_{r_{i}}^{2}+\sum_{i=1}^{N}\left\{\sum_{a=1}^{N_{\text{Nuc}}}\frac{-Z_{a}}{|r_{i}-R_{a}|}+V(r_{i})\right\}\right]\Psi_{K}^{0}=E_{K}^{0}\Psi_{K}^{0}$$
(2)

then any solution $\{\Psi_K^0\}$ and $\{E_K^0\}$ (labeled with the index *K* to identify the state and the superscript 0 to denote "approximate") could be written, respectively, as a Slater determinant of *N* spinorbitals (i.e., spatial orbitals multiplied by α or β spin functions),

$$\Psi_{K}^{0} = |\phi_{K1}(1)\phi_{K2}(2)\phi_{K3}(3)\phi_{K4}(4)\dots\phi_{KN}(N)|$$
(3)

and as a sum of orbital energies $\varepsilon_i(K)$,

ът

$$E_K^0 = \sum_{j=1}^N \varepsilon_j(K) \tag{4}$$

Two of QC's most commonly employed methods, Hartree– Fock (HF) and the Kohn–Sham (KS) version of density functional theory (DFT), are of this form but differ in terms of the particular potential $V(r_i)$ they employ. HF uses a Coulombminus-exchange function for V, whereas DFT uses a Coulomb potential plus a so-called exchange-correlation functional, as I discuss later.

However, the SE also contains the two-electron Coulomb repulsion terms. The most common paths QC uses to connect to the true *N*-electron wave functions and energies are perturbation theory (PT), variational methods (VM), and the coupled-cluster (CC) method. In PT and most VMs, the wave function is expanded as a sum of Slater determinants $\{\Psi_k^0\}$:

$$\Psi = \sum_{K=1}^{N_{det}} C_K \Psi_K^0 \tag{5}$$

where N_{det} is the number of determinants.

In variational approaches¹ (e.g., configuration interaction (CI) and multi-configuration self-consistent field (MCSCF)), an expression for the energy E_r

$$E = \frac{\langle \sum_{K} C_{K} \Psi_{K}^{0} | H | \sum_{K} C_{K} \Psi_{K}^{0} \rangle}{\langle \sum_{K} C_{K} \Psi_{K}^{0} | \sum_{K} C_{K} \Psi_{K}^{0} \rangle}$$
(6)

is minimized with respect to the parameters in the wave function. In CI, the parameters are the $\{C_K\}$ coefficients, while in MCSCF they include the linear-combination-of-atomic-orbitals-to-form-molecular-orbital (LCAO-MO) coefficients $\{C_{k,m}\}$ that describe the MOs in terms of atomic basis functions (AOs) $\{\chi_u\}$:

$$\phi_k(r) = \sum_{\mu=1}^M C_{k,\mu} \chi_\mu(r) [\alpha \text{ or } \beta]$$
(7)

In the MCSCF case, this minimization,

$$\frac{\partial}{\partial C_{I}} \langle \Psi | H | \Psi \rangle = 0 \tag{8}$$

$$\frac{\partial}{\partial C_{j,\mu}} \langle \Psi | H | \Psi \rangle = 0 \tag{9}$$

leads to an eigenvalue equation of dimension $N_{det} \times N_{det}$ for the $\{C_K\}$:

pubs.acs.org/JACS

$$\sum_{K=1}^{N_{det}} \langle \Psi_j^0 | H | \Psi_K^0 \rangle C_K = E C_J$$
(10)

as well as equations similar to (but more complicated than) the HF equations for the LCAO-MO coefficients. These two types of equations are coupled, meaning $\{C_K\}$ appears in the latter and $\{C_{j,\nu}\}$ appears in the former, so they are solved iteratively. In the CI case, only eq 10 is solved because the LCAO-MO coefficients are assumed to have earlier been determined (e.g., in a HF calculation).

In perturbation approaches,² the difference *W* between the electron–electron Coulomb repulsion and the one-electron potential used to approximate it is scaled by an amount $0 \le \lambda \le 1$, and the wave function and total energy are expanded in powers of this parameter λ to write the SE as

$$[H^{0} + \lambda W] \sum_{K} \{ C_{K}^{0} + \lambda C_{K}^{1} + \lambda^{2} C_{K}^{2} + ... \} \Psi_{K}^{0} =$$

$$\{ E^{0} + \lambda E^{1} + \lambda^{2} E^{2} + ... \} \sum_{K} \{ C_{K}^{0} + \lambda C_{K}^{1} + \lambda^{2} C_{K}^{2} + ... \} \Psi_{K}^{0}$$
(11)

Collecting terms of equal power in λ on the left- and right-hand sides of this equation results in a series of equations, one for each order. These equations are solved order-by-order to produce the various-order contributions to the energy and to the wave function's { C_K } coefficients. For example,

$$\langle \Psi_J^0 | H^0 | \Psi_J^0 \rangle = E^0 \tag{12}$$

$$\langle \Psi_J^0 | W | \Psi_J^0 \rangle = E^1 \tag{13}$$

and

$$\langle \Psi^0_J | W | \Psi^1_J \rangle = E^2 \tag{14}$$

are the equations for the zeroth- through second-order energies, and

$$C_J^1 = \frac{\langle \Psi_J^0 | W | \Psi_K^0 \rangle}{E_J^0 - E_K^0}$$
(15)

gives the first-order coefficients, as a result of which

$$E_J^2 = \sum_{K \neq J} \frac{|\langle \Psi_J^0 | W | \Psi_K^0 \rangle|^2}{E_J^0 - E_K^0}$$
(16)

A primary issue is whether the perturbation expansion converges, and there are cases in which it does not³ (e.g., when a bond length is elongated toward homolytic cleavage). The particular form of PT where the perturbation is the difference between the true electron–electron interaction and the HF Coulomb-minus-exchange terms is referred to as Møller–Plesset PT (MPPT) or many-body PT (MBPT). In this variant of PT, the starting-point (zeroth-order) wave function is a single Slater determinant. Notice that the PT coefficient and energy expressions involve Hamiltonian matrix elements between pairs of Slater determinants, a feature that is common to CI, PT, MCSCF, and CC, and one that gives rise to much computational complexity.

In CC theory,⁴ the wave function is written not as a linear combination of Slater determinants but in terms of the exponential of an operator T that is itself a sum of terms,

$$T = 1 + T1 + T2 + T3 + \dots$$
(17)

that act on a reference wave function Ψ_J^0 (taken to be a single Slater determinant in most CC theory) to generate single (T^1) , double (T^2) , triple (T^3) , etc. promotions (often called excitations) in which one, two, three, etc. spin-orbitals occupied in Ψ_J^0 are removed and replaced by spin-orbitals that were not occupied in Ψ_J^0 .

The CC wave function is written as

$$\Psi = \exp(T)\Psi_I^0 \tag{18}$$

and CC theory expresses the SE as

$$\exp(-T)H\,\exp(T)\Psi_J^0 = E\Psi_J^0 \tag{19}$$

and expands the exponentials to obtain a rewritten SE,

where the termination at the fourth-power terms is exact. These CC equations involve products of T operators up through the fourth power, and each T^n operator is written as a product of amplitudes that are the primary variables of this theory and (second-quantized⁵) operators that generate the *n*-fold orbital excitations.

CC theory multiplies the above SE on the left by $\langle \Psi_{J}^{0}|$ to generate an equation for determining *E* and on the left by singly, doubly, etc. excited determinants ($\langle S\Psi_{J}^{0}|$, $\langle D\Psi_{J}^{0}|$, etc.) to generate the equations for determining the amplitudes appearing in the T^{1} , T^{2} , etc. factors. This gives rise to Hamiltonian (and more complicated since they involve both *H* and *T*) matrix elements between pairs of Slater determinants (i.e., $|\Psi_{J}^{0}\rangle$ on the right and $\langle \Psi_{J}^{0}|$, $\langle S\Psi_{J}^{0}|$, $\langle D\Psi_{J}^{0}|$, etc. on the left). Thus the CC equations are fourth-order algebraic equations whose solutions give the amplitudes as well as the energy *E* of the particular state of interest having been specified by the reference wave function Ψ_{J}^{0} .

The matrix elements between Slater determinants, which also appear in PT and VT methods, are expressed in terms of oneand two-electron integrals involving the MOs occupied in the two determinants, such as $\langle \phi_j(r)| - \frac{1}{2} \nabla^2 - \sum_a \frac{Z_a}{|r-R_a|} |\phi_k(r)\rangle$ and $\langle \phi_j(r)\phi_{j'}(r')| \frac{1}{|r-r'|} |\phi_k(r)\phi_{k'}(r')\rangle$. Three of the more constraining bottlenecks of QC involve (i) evaluating the integrals in the AO basis, (ii) transforming these integrals into the MO basis, and (iii) expressing the matrix elements between pairs of Slater determinants in terms of these MO-based integrals. Later, I give concrete examples to illustrate these difficulties.

3. ATOMIC ORBITAL BASIS SETS

The QC computer codes compute⁶ one- and two- electron integrals $\langle \chi_{\mu}(r)| - \frac{1}{2}\nabla^2 - \sum_a \frac{Z_a}{|r-R_a|} |\chi_{\nu}(r)\rangle$ and $\langle \chi_{\mu}(r)\chi_{\lambda}(r')| \frac{1}{|r-r'|} |\chi_{\nu}(r)\phi_{\gamma}(r')\rangle$, where $\{\chi_{\mu}(r)\}$ are the AO basis functions. The most commonly used AOs are of Gaussian form $Dx^a y^b z^c e^{-\alpha r^2}$,

where *D* is a normalization constant and the integers *a*, *b*, and *c* define the orbital's spatial orientation (e.g., *a*,*b*,*c* = 0,0,0 for spherical; 1,0,0, 0,1,0, and 0,0,1 for *x*-, *y*-, and *z*-directed; 1,1,0, 0,1,1, and 1,0,1 for *xy*-, *yz*-, and *xz*-directed, etc.). The exponent α regulates the radial extent/size of the orbital, and *r* is the distance to the electron occupying this orbital from the center on which the orbital resides.

3.1. The Cusp Problem. These primitive Gaussian-type orbitals (PGTOs) are used primarily because the one- and two-electron integrals can efficiently be evaluated¹ for them. However, they don't have the proper "cusp" that quantum mechanics demands. The true wave function must have a slope that is non-zero and given by

$$\frac{\partial\varphi}{\partial r} = -Z\varphi \tag{21}$$

as *r* approaches a nucleus of charge *Z*. PGTOs don't obey this cusp condition because $e^{-\alpha r^2}$ causes $\frac{\partial \varphi}{\partial r}$ to vanish at $r \rightarrow 0$.

QC attempts to address the improper behavior of PGTOs near nuclei by combining several PGTOs of varying radial extent to form a contracted GTO function that is "peaked" near the nucleus:

$$\chi_{\mu}(r) = \sum_{\mu'=1}^{M_{\rm prim}} d_{\mu,\mu'} \chi_{\rm prim,\mu'}$$
(22)

Although this contraction step is especially important for describing the core orbitals that are strongly peaked at nuclear centers, it is also used to describe some valence-level GTOs.

3.2. Core and Valence Gaussian-Type Orbitals (GTOs). I will use a carbon atom to illustrate how the contraction coefficients $\{d_{\mu,\mu'}\}$ are determined. In one approach, a HF calculation on the ground state of the C-atom is carried out using a basis consisting of PGTOs, and the LCAO-MO coefficients of the *1s* HF MO are used as the core *1s* GTO's $\{d_{\mu,\mu'}\}$. Often, just one contracted GTO is employed for treating the core *1s* GTO.

In this same approach, a valence 2s and a set of three (x,y,z) 2p contracted GTOs would be formed by using the HF LCAO-MO coefficients of the HF 2s and 2p orbitals as their $\{d_{\mu,\mu'}\}$. However, unlike the core 1s GTO, more than one 2s and one set of 2p GTOs are created. For the first GTOs, one (double-zeta), two (triple-zeta), or more sets of 2s and 2p(x,y,z) GTOs are created, with these additional functions usually being more radially diffuse than in the first GTO. These additional GTOs might consist of single PGTOs, or they might be contractions of a few PGTOs (the basis set library⁷ will contain such information). The net result of such a GTO basis formation would be one contracted 1s GTO and two (double-zeta), three (triple-zeta), or more sets of additional s and p(x,y,z) valence orbitals.

For example, in a 6-31G basis, the core 1s GTO involves a contraction of six PGTOs (to mimic the cusp behavior of the 1s orbital) as well as two sets of valence GTOs. In the valence space, one s and one set of p(x,y,z) functions are each a contraction of three PGTOs, while the other valence GTOs have one s and one set of p(x,y,z) functions each consisting of one PGTO. I refer the reader to ref 7 and the YouTube video⁸ I created that more fully explain such matters as well as for an explanation of the notations used to label various basis sets.

The bottom line in this example is that the original PGTO basis contains 6 + 3 + 1 = 10 *s*-symmetry PGTOs and three sets of 3 + 1 = 4 *p*-symmetry PGTO functions, for a total of 22

PGTOs. The number of contracted GTOs is 1 + 2 = 3 for the *s* functions and two sets of p(x,y,z) functions, or 9 GTOs in total. The difference between 22 PGTOs and 9 GTOs might not seem like a big deal but it is because much of QC's computational cost scales as a high power of the number of final GTO basis functions.

There is another commonly used path for constructing contracted GTO basis sets; it is labeled cc (standing for correlation-consistent) and uses a correlated rather than a HF level calculation for determining the $\{d_{\mu,\mu'}\}$. Again, I refer the reader to ref 7 and my YouTube video⁸ for more details on the nomenclature used in such cases and for information about where one can find various basis sets listed in a form that can be used as input to a QC program.

These basis functions' primary purpose is to describe the spatial distributions of the electrons that occupy the molecule's core and valence orbitals. The purpose of using multiple valence GTOs is to allow for the MOs to adopt different radial extents in various chemical environments. For example, the bonding π MO of ethylene will certainly involve a linear combination of *p*-symmetry π GTOs on the two carbon atoms with LCAO-MO coefficients of the same sign on the two atoms. But, in 1,1-dichloro-ethylene, the π MO will likely be more radially compact near the carbon atom. This can be accomplished by forming the bonding π MO,

$$\varphi = \sum_{j} C_{j}^{\text{left}} \chi_{j}^{\text{left}} + \sum_{j} C_{j}^{\text{right}} \chi_{j}^{\text{right}}$$
(23)

using different LCAO-MO coefficients C_j^{left} on the left carbon's GTOs $\{\chi_j^{\text{left}}\}$ than on the right GTOs $\{\chi_j^{\text{right}}\}$. By more heavily weighting the GTOs with larger α -values, the LCAO-MO process can make the MO more compact on the carbon holding the two chlorine atoms.

In addition to the core and valence bases, some basis sets for heavier elements include effective core potentials (ECPs). In these cases, the core electrons (e.g., 1s and 2s and 2p for a Clatom) are not even treated; their presence is described by the ECP, and the remainder of the basis is used to treat the other electrons (e.g., 3s and 3p for Cl). Many ECPs also take into consideration the effect of relativistic contraction of inner-shell orbitals on the valence-shell orbitals.

3.3. Polarization and Diffuse GTOs. In addition to core and valence GTOs, the basis often also includes two other kinds of functions-polarization functions and diffuse functions. I will first explain what a polarization function is and what it is used for, again using the π bond in ethylene as an example. If the AO basis set contained only *p*-symmetry AOs on the two C-atoms, this bonding MO would be limited to be of the character shown in Figure 1 (a). However, if the basis also included on each Catom a d-symmetry GTO that had a similar radial size as the pbasis function, the p and d functions could be combined in the LCAO-MO process as shown in Figure 1 (b) to produce a π bond as shown in Figure 1 (c). In this case, the polarization functions allow the conventional p functions to bend (i.e., be polarized) to the left or to the right. One can also add psymmetry polarization functions (e.g., to ethylene's four H-atom 1s valence AOs) to allow s-symmetry AOs to polarize as shown in Figure 1 (d).

It is important to emphasize that these polarization functions are of similar radial size to the valence AOs they are polarizing. So, the C-atom d polarization function shown in Figure 1 (b) is not appropriate for describing a 3d excited orbital of the C-atom; it is, instead, similar in size to the C-atom's 2p orbital. Likewise, the p



Figure 1. Illustration of how polarization functions can allow π (in (a), (b), and (c)) and σ (in d) orbitals to distort.

polarization function shown in Figure 1 (d) would be of similar radial size to the H-atom's 1s orbital if it were used to polarize the H-atom's valence 1s orbital; it would not be appropriate for describing a 2p excited state of the H-atom.

GTO basis sets often also include so-called diffuse GTOs. These functions have smaller orbital exponents α than in core or valence GTOs and are especially crucial to include when studying anions having very small electron binding energies. For studying dipole-bound anions or Rydberg states, even more diffuse basis functions are needed, and they usually need to be manually added to a standard basis set.

I should note that the polarization and diffuse functions serve additional roles. The former are certainly important for describing the polarization of a molecule's electron cloud induced either by the presence of another molecule or by an external electric field. The latter are crucial when treating polarization and dispersion (i.e., van der Waals) interactions.

3.4. Why So Many Different GTO Basis Sets? The fact that the number (M) of contracted GTOs usually far exceeds the number of electrons causes the high-power scaling of the computer time needed (e.g., M^3 to M^5 or worse) to be a major difficulty. This strongly encourages workers designing optimal GTOs to keep M as small as possible. This means that the contractions used to form the core and valence GTOs from PGTOs have to be carefully designed and the number of valence, polarization, and diffuse functions kept to a minimum for the task at hand. Also, different tasks require different basis sets. For example, if one is not studying a molecule in which a very diffuse MO is expected to occur, there likely is no need to include diffuse basis functions unless one is also studying the molecule's dispersion interaction with another or its physical polarization by an outside influence. If one is studying hydrocarbons having a variety of substituents, a valence basis with enough flexibility to characterize differences in local electronegativity is necessary. Moreover, if one is seeking high accuracy in, for example, a CC calculation, one would not want to use a basis containing few valence-type GTOs and no polarization functions; more GTOs offer more flexibility in allowing the wave function to become optimal. To gain appreciation for how calculated molecular geometries and energies (i.e., bond energies, ionization energies, etc.) vary when using different basis sets and different methods, I refer you to

Figures 15.4, 15.7, 15.12, and 15.16 of ref 1, where the distributions of errors in such quantities for various basis sets and methods are shown.

Knowing the number (M) of AO basis functions is important because this determines how much computer time the calculation will take moving from one basis to another. In Table 1, I illustrate how the basis set size can greatly exceed the number of valence electrons by listing M for several commonly used bases for a first-row atom. In most QC programs, the total number of basis functions for the molecule being studied is given in the output information, so the easiest way⁹ to determine M is by initiating a QC calculation.

 Table 1. Basis Set Size M for a Variety of Common Bases for

 Each First-Row Atom

Basis	6-31++G	6-31++G*	6-311++G	6-311++G*
M	13	19	17	23
Basis	aug-cc-p	/DZ aug-	cc-pVTZ	aug-cc-pVQZ
М	23	0	46	80

THE INTEGRAL TRANSFORMATION SCALES AS HIGH AS M⁵

The QC computer codes evaluate the AO-based two-electron integrals $\langle \chi_{\mu}(1)\chi_{\lambda}(2)|\frac{1}{|r-r|}|\chi_{\nu}(1)\chi_{\gamma}(2)\rangle$ and the corresponding one-electron producing M^4 such. The code then performs a series of four steps to transform the integrals to the MO basis. First, an intermediate four-indexed array *X* is formed from the AO integrals:

$$\sum_{\nu=1}^{M} \langle \chi_{\mu}(1)\chi_{\gamma}(2)|\frac{1}{r_{1,2}}|\chi_{\rho}(1)\chi_{\nu}(2)\rangle C_{l,\nu} = X_{\mu,\gamma;\rho,l}$$
(24)

This step involves M multiplications of an AO-based twoelectron integral by an LCAO-MO coefficient for each of the N_l MOs labeled l and for each of the three indices μ , γ , ρ , thus requiring a total of $N_l \times M^4$ computations.

The second step involves $N_l \times N_k \times M^3$ calculations and forms an array *Y*:

$$\sum_{\rho=1}^{M} X_{\mu,\gamma;\rho,l} C_{k,\rho} = Y_{\mu,\gamma;k,l}$$
(25)

The third step involves $N_l \times N_k \times N_n \times M^2$ calculations and produces an array Z:

$$\sum_{\gamma=1}^{M} Y_{\mu,\gamma;k,l} C_{n,\gamma} = Z_{\mu,n;k,l}$$
(26)

The final step requires $N_l \times N_k \times N_n \times N_m \times M$ calculations and produces the two-electron integrals in the MO basis:

$$\sum_{\rho=1}^{M} Z_{\mu,n;k,l} C_{m,\mu} = \langle \phi_m(r) \phi_n(r') | \frac{1}{|r-r'|} | \phi_k(r) \phi_l(r') \rangle$$
$$\equiv \langle m, n \mid k, l \rangle$$
(27)

Notice that if one or more of the N_l , N_k , N_n , and N_m (call it N'') is not proportional to M, the overall scaling of this process will be $N'' \times M^4$; on the other hand, if all four of these numbers

are proportional to M, the scaling is M^5 . In any event, this step presents one of QC's biggest computational challenges.

The computer time (CPU) needed in a QC calculation is the sum of the time needed to carry out all the steps, but often there is one step that is most taxing, in which case we quote the scaling of that step. The MOs occupied in the reference wave function are *O* in number. For HF and KS-DFT, O = N, as is the case for MPPT and CC, because all four approaches use a single Slater determinant of *N* spin-orbitals as their starting point. In CI and MCSCF, O > N because there are two or more Slater determinants in these wave functions. An AO basis having *M* functions produces these *O* MOs as well as M - O = V virtual MOs.

For a given molecule and a given method, if one changes from a basis having M AOs to another with M' AOs, O does not change but V does. In such cases, the CPU time scales as some power of (M' - O)/(M - O). On the other hand, using a given class of basis set on two different molecules, both O and M will be different, so the CPU time will vary as a power of (O'/O)times a power of (M' - O')/(M - O). The former scaling is usually referred to as basis-set scaling and the latter as moleculesize scaling. For larger basis sets, where $M \gg O$, $(M' - O')/(M - O) \approx M'/M$ will hold, and this is how this scaling is usually denoted.

When it comes to molecule-size scaling, it is important to note that techniques such as localized orbitals^{10–12} and methods for approximating two-electron integrals^{13,14} have proven useful in reducing the scaling problems in larger molecules and clusters.

5. EXAMPLES OF HOW VARIOUS QC STEPS SCALE WITH BASIS SIZE

Assume that a researcher wants to find the energies of the lowest two electronic states of ethylene and the energy of the ethylene cation at the equilibrium geometry $\{R_a\}$ of the neutral. Assume the researcher has chosen to use a 6-31++G(d,p) basis set for the C- and H-atoms that together contain 62 GTOs, and has chosen to first carry out a HF or DFT calculation to optimize the geometry and, in the HF case, to subsequently treat the correlation energy using PT with second-order Møller–Plesset perturbation theory (MP2), and to search for the excited state's energy using configuration interaction (CI). The worker will tell the program how to occupy the orbitals with the available electrons; let's assume that the researcher has specified a state with 8 α and 8 β electrons coupled to give a singlet state and with the 16 electrons occupying the 8 lowest-energy HF or DFT orbitals.

The program first performs a series of HF or DFT calculations in which a geometry optimization algorithm is employed to generate a series of steps (i.e., geometry changes) that lead to lower and lower total electronic energies. All QC codes have a variety of such geometry optimization modules¹⁵ and the worker may have to specify one. Some use the gradient (i.e., first derivative with respect to displacements of the nuclei) of the HF or DFT energy to find a path leading downhill, hopefully to the desired minimum; others also use second-derivative information on this search.

If all goes well, the result would be a minimum-energy geometry corresponding to $H_2C=CH_2$ and a corresponding HF or DFT energy. If the user also specified that the program compute the MP2 energy at this same geometry, the program would do so. To estimate the energy of the ground state of the ethylene cation, the user could either use Koopmans' theorem (KT) to make this estimate as minus the HF orbital energy of the

 π orbital or ask the program to compute the HF, DFT, or MP2 energy for the system with only 15 electrons (e.g., with the π orbital β electron removed) coupled to generate a doublet spin state. I will now walk through the steps involved in this project, pointing out the computational-scaling issues that arise.

5.1. The Hartree–Fock Step. The wave function is approximated by a single Slater determinant involving *N* spinorbitals:

$$\Psi_{\rm HF} = |\phi_1(1)\phi_2(2)\phi_3(3)\phi_4(4)...\phi_N(N)| \tag{28}$$

If the spin-orbitals occur in spatially identical pairs of orbitals, as they would in the calculation being considered here, this is a spin-restricted HF (RHF) calculation. If each spin-orbital were allowed to have its own LCAO-MO coefficients (e.g., the $\pi\alpha$ spin-orbital having different LCAO-MO coefficients than the $\pi\beta$ spin-orbital), QC calls this a spin-unrestricted HF (UHF) calculation. I'll say more about UHF later.

With each of the spin-orbitals expressed as a linear combination of the AO basis functions multiplied by LCAO-MO coefficients, this leads to the HF eigenvalue problem,

$$\sum_{\nu} \langle \chi_{\mu} | F | \chi_{\nu} \rangle C_{j,\nu} = \varepsilon_j \sum_{\nu} \langle \chi_{\mu} | \chi_{\nu} \rangle C_{j,\nu}$$
⁽²⁹⁾

where $C_{j,\nu}$ is the expansion coefficient of the *j*th MO involving the AO χ_{ν} , ε_{j} is the HF orbital energy of that MO, $\langle \chi_{\mu} | \chi_{\nu} \rangle$ is the overlap integral between the two AOs, and $\langle \chi_{\mu} | F | \chi_{\nu} \rangle$ is the corresponding matrix element of the Fock operator, defined as

$$\langle \chi_{\mu} | F | \chi_{\nu} \rangle = \langle \chi_{\mu} | - \frac{1}{2} \nabla^{2} - \sum_{a} \frac{Z_{a}}{|r - R_{a}|} | \chi_{\nu} \rangle$$

$$+ \frac{1}{2} \sum_{k=1}^{N} \sum_{\gamma=1}^{M} \sum_{\rho=1}^{M} \left[\langle \chi_{\mu}(1) \chi_{\gamma}(2) | \frac{1}{r_{1,2}} | \chi_{\nu}(1) \chi_{\rho}(2) \rangle - \langle \chi_{\mu}(1) \chi_{\gamma}(2) | \frac{1}{r_{1,2}} | \chi_{\rho}(1) \chi_{\nu}(2) \rangle \right] C_{k,\gamma} C_{k,\rho}$$

$$(30)$$

Here, k runs over the N occupied spin-orbitals. We see the kinetic energy, the interaction of the electron with each of the nuclei, and the Coulomb and exchange energies of the orbital (labeled j) interacting with the N - 1 other electrons occupying spin-orbitals.

The Fock and overlap matrices have dimensions $M \times M$, and the HF eigenvalue equations are solved iteratively since the matrix elements of the Fock operator involve knowledge of $C_{j,\nu}$ for the *N* occupied spin-orbitals. Each iteration involves solving an $M \times M$ matrix eigenvalue problem for *M* MO energies and *M* sets of LCAO-MO coefficients; of these, *N* belong to the occupied spin-orbitals while M - N = V belong to virtual spinorbitals.

Two significant computational costs arise in this process. The one- and two-electron integrals must be computed in the AO basis, requiring CPU time proportional to M^4 . In some approaches, these same integrals must be stored for future use (the HF process must be iterated, and these same integrals are needed in subsequent QC steps); this requires storage proportional to M^4 . If the integrals are not stored, they must be re-calculated in each iteration of the HF process.

Second, at least the N eigenvalues and eigenvectors belonging to the occupied spin-orbitals must be found, needing CPU time scaling as $N \times M^2$. In subsequent studies beyond the HF level, one usually needs to know all or most of the orbital energies and LCAO-MO coefficients, not just those belonging to the N occupied HF spin-orbitals, so M^5 scaling of this step is possible.

Notice that HF does not require that the integrals be transformed from the AO to the MO basis. If the researcher had used DFT rather than HF in this step of the calculation, things would have progressed much as described above, with overall M^4 scaling from the AO integral evaluation. However, in the MP2 and CI steps and in MCSCF and CC calculations, this integral transformation step is (usually) necessary.

5.2. The MP2 Step. To carry out the MP2 calculation, one needs access to a subset of the two-electron integrals in the MO orbital basis, and the MP2-level energy is the HF energy plus a correction,

$$\Delta E_{\rm MP2} = \sum_{i < j=1}^{N} \sum_{m < n=N+1}^{M} \frac{|\langle m, n \mid i, j \rangle - \langle m, n \mid j, i \rangle|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_m - \varepsilon_n}$$
(31)

where the indices *i* and *j* label the *N* occupied spin-orbitals and *m* and *n* label the (M - N) unoccupied spin-orbitals. Because the number of occupied MOs (i and j) is not proportional to *M*, only $N \times M^4$ calculations are required in the integral transformation for MP2. Also, if one does not wish to include correlation of the core orbitals, the indices *i* and *j* can be further limited, as a result of which even fewer of the two-electron integrals need be transformed. Nevertheless, integral transformation remains the limiting factor.

As an alternative to using MP2 to estimate the correlation energy, CI or MCSCF could be used. For these methods one usually needs most if not all of the MO-based two-electron integrals, but in these cases at least one of the indices i, j, k, and l does not grow with M, so the integral transformation CPU time can scale as $O \times M^4$.

In CC methods, there are steps having even higher-power scaling. For example, when including only single- and double-excitation operators (T^1 and T^2 ; termed CCSD), the calculation time, aside from the integral transformation step, scales as $N^2 \times M^4$, where N = O is the number of orbitals occupied in the reference Slater determinant. If triple excitation operators (T^3) are treated perturbatively as in CCSD(T), the calculation scales as $N^3 \times M^4$. For a fixed molecule but varying the basis set, the scaling of either CC method is (M'/M)⁴, but for a fixed basis type and different molecules, the scaling is (N'/N)² or ${}^3 \times (M'/M)^4$. So, the effort in such CC calculations can exceed the potential M^5 scaling of the integral transformation.

5.3. DFT Offers an Alternative That Includes Correlation, with M^4 and Potentially M^3 Scaling. The geometry optimization and HF- and MP2-level energy calculations detailed above could instead be achieved using KS-DFT. Analytical gradient and Hessian expressions are available to use in the geometry optimization step, and DFT's energy includes correlation. In the most common variants of DFT,^{16,17} one solves the KS equations for the KS spin-orbitals ϕ_i and orbital energies ε_v

$$\left[-\frac{1}{2}\nabla^2 - \sum_{a=1}^{N_{\text{Nuc}}} \frac{Z_a}{|r - R_a|} + U(\rho(r))\right] \phi_i(r) = \varepsilon_i \phi_i(r)$$
(32)

using an exchange-correlation functional $U(\rho(r))$ that depends only on the electron density $\rho(r)$ at the location r. $U(\rho(r))$ is not a simple function but a functional that usually depends on spatial gradients of $\rho(r)$, integrals involving $\rho(r)$, and fractional powers of $\rho(r)$. The sum of the squares of the N occupied spin-orbitals gives $\rho(r)$ at each iteration of the KS equation.

pubs.acs.org/JACS

The ϕ_i are expanded in an AO basis, which leads to an eigenvalue problem,

$$\sum_{\nu} \langle \chi_{\mu} | F_{\text{KS}} | \chi_{\nu} \rangle C_{j,\nu} = \varepsilon_j \sum_{\nu} \langle \chi_{\mu} | \chi_{\nu} \rangle C_{j,\nu}$$
(33)

where

$$\begin{aligned} \langle \chi_{\mu} | F_{\text{KS}} | \chi_{\nu} \rangle &= \langle \chi_{\mu} | - \frac{1}{2} \nabla^2 - \sum_{a} \frac{Z_a}{|r - R_a|} | \chi_{\nu} \rangle \\ &+ \int \rho(r') \chi_{\mu}(r) \chi_{\nu}(r) \frac{1}{|r - r'|} \, \mathrm{d}r \, \mathrm{d}r' \\ &+ \int \chi_{\mu}(r) U(\rho(r)) \chi_{\nu}(r) \, \mathrm{d}r \end{aligned}$$
(34)

The first two terms are the kinetic energy and electron-nuclei Coulomb interactions. The third term gives the matrix element involving the Coulomb repulsion of an electron at r interacting with the full electron density $\rho(r')$. This includes the Coulomb interaction of the electron at r with its own density; this is the so-called self-interaction, which should not be present.

The final term in the KS equation involves the exchangecorrelation functional $U(\rho(r))$ that fills two roles: it attempts to correct for the self-interaction just discussed, and it expresses the correlation energy. The main challenge in developing DFT has been in creating functional forms for $U(\rho(r))$ that are computationally tractable, physically meaningful, and accurate. The myriad of functionals¹⁸ that have been put forth, and are still under active development, is one of DFT's contributions to the complications of QC.

Solving the $M \times M$ -dimension KS eigenvalue equations requires CPU time proportional to M^3 , but it does not necessarily require calculation of the same M^4 AO-based twoelectron integrals that HF theory does, nor does it require transformation of the integrals to the MO basis. DFT expresses the KS orbitals in the same kind of LCAO-MO expression as in HF, but in some variants it expresses the density $\rho(r')$ as a separate linear combination of AOs (a so-called auxiliary set of M' AOs is used for this expansion),

$$\rho(r) = \sum_{\gamma=1}^{M'} c_{\gamma} \chi_{\gamma}'(r)$$
(35)

where the expansion coefficients $\{c_{\mu}\}$ of the density can be obtained from the LCAO-MO coefficients of the *N* occupied KS spin-orbitals. This allows the third term in the KS equation to be rewritten as

$$\int \rho(r')\chi_{\mu}(r)\chi_{\nu}(r)\frac{1}{|r-r'|} dr dr' = \sum_{\gamma=1}^{M'} c_{\gamma} \int \chi_{\gamma}'(r')\chi_{\mu}(r)\chi_{\nu}(r)\frac{1}{|r-r'|} dr dr'$$
(36)

These integrals have only three AO indices, so there are only $M' \times M^2$ of them, where M' is the number of auxiliary basis functions, thus giving *cubic rather than quartic scaling* in the integral-evaluation step.

A similar path can be followed for the fourth term in the KS equations. For example, if $U(\rho(r))$ involves a term of the form $-A\rho^{1/3}(r)$, the function $\rho^{1/3}(r)$ could also be expanded in terms of auxiliary AOs,

$$\rho^{1/3}(r) = \sum_{\mu=1}^{M'} f_{\mu} \chi_{\mu}'(r)$$
(37)

and again the f_{μ} coefficients can be related to the occupiedorbitals' LCAO-MO coefficients. The fourth term in the KS equations can then be expressed as

$$\int \chi_{\mu}(r) U(\rho(r)) \chi_{\nu}(r) \, \mathrm{d}r = -\sum_{\gamma=1}^{M'} A f_{\gamma} \int \chi_{\mu}(r) \chi_{\gamma}'(r) \chi_{\nu}(r) \, \mathrm{d}r$$
(38)

thus retaining the $M' \times M^2$ cubic scaling. I note that most DFT calculations include in the exchange-correlation functional exchange integrals as in HF (discussed later), thus returning one to M^4 overall scaling.

5.4. Evaluating the Ionization Potential. The QC researcher also wants to determine the energy of the ethylene cation. This could be achieved by subtracting the neutral molecule's HF, MP2, or DFT total energy from the cation's HF, MP2, or DFT energy. Of course, this path would require a separate HF, MP2, or DFT calculation on the cation thus essentially doubling the computer time involved.

Alternatively, one could use Koopmans' theorem¹⁹ and estimate the vertical ionization potential (IP) as minus the energy ε_{π} of the π orbital. This theorem was derived by analytically subtracting the Hamiltonian expectation value of the HF Slater determinant having 16 electrons from the expectation value of the Slater determinant in which one of the electrons in the π orbital has been removed:

$$\langle |\phi_1(1)\phi_2(2)\phi_3(3)\phi_4(4)...\phi_{\pi}(15)\phi_{\pi}(16)|H |\phi_1(1)\phi_2(2)\phi_3(3)\phi_4(4)...\phi_{\pi}(15)\phi_{\pi}(16)|\rangle - \langle |\phi_1(1)\phi_2(2)\phi_3(3)\phi_4(4)...\phi_{\pi}(15)|H |\phi_1(1)\phi_2(2)\phi_3(3)\phi_4(4)...\phi_{\pi}(15)\phi|\rangle = -IP$$
(39)

The difference in these two expectation values turns out to be ε_{π} :

$$\varepsilon_{\pi} = \sum_{\mu,\nu=1}^{M} C_{i,\mu} C_{i,\nu} \langle \chi_{\mu} | F | \chi_{\nu} \rangle = \langle \phi_{i}^{i} | -\frac{1}{2} \nabla^{2} - \sum_{a} \frac{Z_{a}}{|r-R_{a}|} | \phi_{i} \rangle + \sum_{k=1}^{N} \left[\langle \phi_{i}(1)\phi_{k}(2) | \frac{1}{r_{1,2}} | \phi_{i}(1)\phi_{k}(2) \rangle - \langle \phi_{i}(1)\phi_{k}(2) | \frac{1}{r_{1,2}} | \phi_{k}(1)\phi_{i}(2) \rangle \right]$$
(40)

where $i = \pi$.

The other occupied orbitals' energies ε_i give estimates of IPs for removing an electron from each of those MOs. In principle, the orbital energies ε_m belonging to unoccupied HF orbitals can give estimates for the electron affinities (EAs) associated with adding an electron to those MOs, but one has to be very careful in identifying which virtual orbital is chosen,²⁰ especially when considering orbitals having positive energies. When using a basis set with diffuse functions, it is common for some/many of the lower positive-energy virtual orbitals to correspond to pseudocontinuum orbitals that approximate a free electron plus the neutral molecule. Reference 20 details how to avoid such problems.

In contrast with HF, DFT orbital energies²¹ do not offer the same connection to IPs and EAs. The energy of the highest occupied KS MO gives the correct IP if one has the exact functional $U(\rho(r))$, but the virtual KS MOs' energies are not good estimates to EAs (but the virtual-to-occupied KS orbital energy differences often do give decent excitation energies).

5.5. The Energy of the Excited State. If the excited state of ethylene of interest is of a different spatial or spin symmetry than the ground state, then a HF calculation followed by an MP2 evaluation of the correlation energy, or a DFT calculation, can suffice. In the HF or DFT case, the program occupies the spinorbitals in a manner that produces the desired spatial and spin symmetry. For example, if the triplet $\pi\pi^*$ state were under consideration, the worker could tell the program to place the 15th and 16th electrons into $\pi\alpha$ and $\pi^*\alpha$ spin-orbitals, respectively. The HF or DFT code could then produce an energy for this triplet $\pi\pi^*$ state, and the energy of the ground state could be subtracted from it to estimate the vertical (i.e., at the specific geometry) excitation energy. The MP2 energy correction for this excited state could be evaluated as discussed above, and this value could be used to form an MP2-level estimate of the excitation energy. I should note that, to study a singlet $\pi\pi^*$ state, this approach would not work because the most basic approximate wave function for such cases would require two Slater determinants.

If the QC researcher wants to determine the energy of an excited state of the same spatial and spin symmetry as the ground state, then a different approach is needed, and CI offers such a route. When using CI, the program forms several Slater determinants, all having the same spatial and spin symmetry as the ground state, that differ from the dominant determinant of the ground state through the promotion of one, two, three, or more electrons. Once the CI code is instructed as to which single, double, etc. promotions to include, it forms the relevant Slater determinants and then forms the Hamiltonian matrix elements of eq 10 between all pairs of Slater determinants. *These matrix elements are expressed in terms of the MO-based integrals, so the CI calculation will require this potentially M⁵ calculation step.*

The dimension of the CI Hamiltonian matrix thus formed is $N_{det} \times N_{det}$ and to find N_{States} eigenvalues requires $N_{States} \times N_{det}^2$ steps, which can get out of control if N_{det} becomes very large. For this reason, most CI calculations involve careful selection (e.g., by limiting the orbital energy ranges for electron promotions and selecting only determinants that couple strongly with the most dominant determinants) of which determinants to include. Moreover, special techniques²² are used to solve for the N_{States} eigenvalues of such large matrices. The results of a CI calculation in which the lowest two states are extracted as solutions of the $N_{det} \times N_{det}$ eigenvalue problem are the energies of these two states (the ground state and the lowest excited state of the same symmetry) and their corresponding CI expansion coefficients { C_I }.

6. WHAT ARE DYNAMIC AND STATIC ELECTRON CORRELATIONS?

The PT, CI, MCSCF, and CC methods include correlation effects by including more than one Slater determinant in the wave function. DFT treats correlation through its exchangecorrelation functional. Both routes present challenges, but I begin by discussing the former to explain what dynamic and static correlations are. pubs.acs.org/JACS

One might think the most important determinants to add to the dominant Slater determinant would be those involving single promotions (e.g., determinants like $|\phi_1(1)\phi_2(2)\phi_3(3)\phi_K(4)...$ $\phi_N(N)|$ in which ϕ_4 has been replaced by ϕ_K). However, this is not the case, because a combination of the dominant determinant and one involving a single promotion is equivalent to a single determinant in which the fourth spin-orbital is a combination of ϕ_4 and ϕ_K :

$$C_{A}|\phi_{1}(1)\phi_{2}(2)\phi_{3}(3)\phi_{4}(4)...\phi_{N}(N)|$$

$$+ C_{B}|\phi_{1}(1)\phi_{2}(2)\phi_{3}(3)\phi_{K}(4)...\phi_{N}(N)|$$

$$= |\phi_{1}(1)\phi_{2}(2)\phi_{3}(3)[C_{A}\phi_{4}(4) + C_{B}\phi_{K}(4)]...\phi_{N}(N)|$$
(41)

If the energy has already been minimized with respect to the LCAO-MO coefficients of the fourth spin-orbital, the optimal value of C_B will be zero. This does not mean that singly excited determinants will have zero amplitudes if one also includes doubly and/or higher promotions; it just illustrates why the single promotions are usually not very important—because the occupied orbitals have already been optimized.

Doubly promoted determinants are actually the most important to include, which I illustrate by considering a wave function consisting of two determinants, with the second formed by taking the doubly occupied orbital ϕ in the first and replacing this orbital by a different orbital ϕ' (e.g., $\phi = \pi$ and $\phi' = \pi^*$ in the ethylene situation). It can be shown that such a combination of two determinants, with coefficients C_1 and $-C_2$, can be rewritten as follows:

$$\Psi = C_1 \dots \phi \alpha \phi \beta \dots I - C_2 \dots \phi' \alpha \phi' \beta \dots I$$
$$= \frac{C_1}{2} [\dots (\phi - x\phi') \alpha (\phi + x\phi') \beta \dots I$$
$$- \dots (\phi - x\phi') \beta (\phi + x\phi') \alpha \dots I]$$
(42)

where

$$x = \sqrt{\frac{C_2}{C_1}} \tag{43}$$

Instead of viewing the function as one determinant having both electrons in ϕ and another having both electrons in ϕ' , with $|C_1|^2$ giving the probability of the first and $|C_2|^2$ giving the probability of the second, one can take another viewpoint. This same function can be viewed as one electron being in one polarized orbital,

$$\phi_{\!\!+} = \frac{\phi + x\phi'}{\sqrt{1 + x^2}} \tag{44}$$

while the other electron resides in another polarized orbital,

$$\phi_{-} = \frac{\phi - x\phi'}{\sqrt{1 + x^2}}$$
(45)

The factor of $\sqrt{1 + x^2}$ makes them normalized.

In Figure 2, I show how the π and π^* MOs of ethylene combine to form a polarized orbital pair. The $\phi_+ = \phi + x\phi'$ orbital is polarized to give more electron density on the left atom, while $\phi_- = \phi - x\phi'$ is polarized to give more electron density on the right atom. Note that the polarized orbital pairs need not involve the polarization basis functions I discussed earlier

(they were of d character); polarized orbitals pairs and polarization basis functions need not be the same thing and are not in this case.



Figure 2. Illustration of how mixing the bonding π and anti-bonding π^* orbitals gives rise to left- and right-polarized orbitals.

When the ethylene molecule is at or near its equilibrium geometry with the π bond intact, the coefficient x turns out to be small but non-zero. In this case, the polarized orbitals describe what QC calls <u>dynamical correlation</u>, in which the two electrons are avoiding one another. Such correlation corrections to the total energy are often tenths of an eV per electron pair for valence orbitals of first-row atoms. <u>Static/essential correlation</u> arises when the π bond is broken by rotating the CH₂ groups toward 90°, where x becomes unity, meaning the polarized orbitals have evolved into one localized on the left atom with the other on the right. This is how two determinants generate singlet-coupled radical centers upon homolytic cleavage of the π bond.

Although MCSCF, PT, CI, and CC calculations describe correlations through the introduction of doubly excited determinants (although CC theory does so via its T^2 operator), they differ in an important manner. PT and CC theory focuses on finding the total electronic energy of one electronic state through the choice of the single-determinant reference function that is assumed to dominate the final wave function. In contrast, CI and MCSCF can provide more than one solution of their matrix eigenvalue problem, offering a route to the energies of more than a single electronic state; also, they don't assume that one Slater determinant is dominant in any solution. All of these methods suffer the potential M^{δ} plague of integral transformation plus the $N_{States} \times N_{det}^2$ eigenvalue-solution step of CI and MCSCF or the steps scaling as high as M^{δ} or M^7 in CC.

In contrast, DFT treats electron correlation through its exchange-correlation functional $U(\rho(r))$ that fills two roles: correcting for the self-interaction error and expressing the correlation energy (in some functionals^{23,24} also including long-range dispersion terms). Largely for scaling that does not exceed M^4 and for accuracy that has resulted from developments and testing of exchange-correlation functionals, DFT has become one of QC's most widely used methods.

However, it is important to be aware of some weaknesses that remain in DFT (although they are being addressed by a multitude of researchers). In addition to the issue of KS orbital energies not offering the same connection to IPs and EAs as Koopmans' theorem, there are two more I wish to highlight: how the exchange-correlation functionals handle the self-interaction problem and to what extent they are capable of treating the essential/static correlations that arise when bonds are homolytically cleaved.

In hybrid DFT methods, the self-interaction is addressed by adding into the KS Fock matrix terms that explicitly are of the exchange form, weighted by a fractional parameter 0 < X < 1:

$$-X\sum_{\gamma,\kappa}\sum_{j}C_{j,\gamma}C_{j,\kappa}\int\chi_{\gamma}(r')\chi_{\kappa}(r)\chi_{\mu}(r)\chi_{\nu}(r')\frac{1}{|r-r'|} dr dr'$$
(46)

and then scaling the exchange portion of $U(\rho(r))$ by the counterbalancing factor (1.0 - X). Following this path brings one back to having two-electron integrals with four indices and thus to M^4 scaling but still avoids the M^5 integral transformation.

As an example²⁵ illustrating the importance of properly correcting for self-interaction errors, consider dissociation of the H_2^+ cation into $H + H^+$ (analogous problems arise in dissociations of other bonds involving odd numbers of electrons). In this case, most common KS-DFT methods produce a total electron density

$$\rho(r) = |\sigma_g(r)|^2 \tag{47}$$

as the square of the singly occupied σ_g orbital. As such, the net charge on each of the two nuclear centers is $+^{1}/_{2}$; that is, the electron density is half on the left center and half on the right. This produces in the DFT energy profile a Coulomb repulsion between the two $+^{1}/_{2}$ charges at large *R*, which, of course, is wrong. The hybrid approach discussed above tries to remedy this but turns out not to do so fully.

DFT has also had difficulty with static correlation arising in homolytic bond cleavage,²⁶ as in dissociation of H_2 into two Hatoms (similar issues arise, for example, in rotating the two H_2C units in ethylene to break the π bond). The simplest qualitatively correct model involves two electronic configurations,

$$\Psi = C_1 |\sigma \alpha(1) \sigma \beta(2)| - C_2 |\sigma^* \alpha(1) \sigma^* \beta(2)|$$
(48)

where the bonding and anti-bonding MOs are

$$\sigma = c\chi_{\text{left}} + c\chi_{\text{right}} \quad \text{and} \quad \sigma^* = c'\chi_{\text{left}} - c'\chi_{\text{right}} \tag{49}$$

and χ_{left} and χ_{right} are comprised of AO basis functions on the two nuclei. This wave function allows the polarized orbital pair to evolve into χ_{left} and χ_{right} to produce a singlet-coupled pair of radical centers when the bond breaks.

The problem that DFT has in dealing with this is that it expresses the total density, $\rho(r') = \sum_{j=1}^{N} |\phi_j(r')|^2$ as a sum over densities for the *N* occupied KS spin-orbitals in its single Slater determinant. However, most available KS procedures are not able to find a single spatial orbital to use in a singlet-coupling manner within a single Slater determinant that can produce the electron density of the above two-determinant function:

$$\rho(r) = 2C_1^2[|\sigma(r)|^2 + x^2|\sigma^*(r)|^2]$$
(50)

where

111

$$x = \sqrt{\frac{C_2}{C_1}} \tag{51}$$

A single determinant

$$\Psi = \left| \phi_{+}(r_{1}) \alpha \phi_{-}(r_{2}) \beta \right| \tag{52}$$

involving the polarized orbital pair

Table 2. Scaling of Computationally Dominant Components with Number of AOs (M) and Occupied MOs (N) for AO-Integral Evaluation (*eval.*), AO-to-MO Integral Transformation (*trans.*), and Matrix Eigenvalue-Solving (*eigenval.*) Steps or Steps Intrinsic to the Method

HF, DFT MP(2) CCSD CCSD(T) CI, MCSCF CIS, RPA, TDDFT	CCSD-EOM-EE, IP, EA
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$N^{2}(M - N)^{4}$ for CCSD step $(NM)^{2}$ for EE or M^{2} for IP, EA

$$\phi_{\pm} = \frac{\sigma \pm x \sigma^*}{\sqrt{1 + x^2}} \tag{53}$$

does have the correct density, but this determinant is not a singlet; it is a mixture of singlet and triplet spin functions. Generalizations^{26,27} of DFT that make use of more than N spin-orbitals have been introduced that show significant progress in dealing with this issue but most DFT methods still suffer in this manner.

One can still determine homolytic bond energies using DFT by carrying out a DFT calculation on the reactant molecule (e.g., H_2), carrying out separate DFT calculations on each of the fragments obtained when the bond is cleaved, and then subtracting these energies. However, it is difficult to form a reliable energy vs bond-length curve for dissociating H₂ by using DFT energies at H-H distances ranging from equilibrium through large-R. If the two occupied spin-orbitals are not required to have identical spatial character, a full potential curve can be obtained, but there is a problem. For H₂ this uses a determinant of the form $|\sigma \alpha(1)\sigma' \beta(2)|$ in which the two spatial orbitals can be different combinations of χ_{left} and χ_{right} . Such socalled unrestricted (UKS) wave functions can be used to generate an energy plot that evolves from near the equilibrium bond length (where the energy minimization gives $\sigma' = \sigma$) out to large-*R* (where σ' becomes χ_{right} and σ becomes χ_{left}). The problem with this approach is that, at some critical distance, σ and σ' suddenly become different. The energy plot does not undergo a discontinuity, but its slope does, and the spin eigenvalue is no longer singlet. Such "kinks" in the UKS (or UHF) energy surfaces are not fully satisfactory descriptions.

7. THE SCHRÖDINGER ENERGIES ARE EXTENSIVE, BUT MOST PROPERTIES WE ARE INTERESTED IN ARE INTENSIVE—THIS IS A PROBLEM

I will use two molecules to illustrate this issue, both described using a $6-31++G^{**}$ basis—ethylene and the H₂N-(Ala)₁₀-COOH polypeptide. For ethylene, the HF and MP2 energies are -2123.54 eV and -2131.46 eV, respectfully, and there are 62 GTOs in the basis. For the polypeptide, the HF and MP2 energies are $-68\,969.52$ eV and $-69\,182.78$ eV, and the basis has 1281 GTOs.

Each of these energies approximates the total energy of the molecule with its electrons attached relative to the energy of that molecule with all its electrons stripped away. I don't know of any experiment that actually measures such energies! Usually, experiments measure intensive quantities such as bond energies, activation energies, excitation energies, IPs, and EAs.

To compute intensive energies as differences in two Schrödinger total energies presents a numerical-precision challenge. For example, the polypeptide's energy is ca. 32 times that of ethylene, which is not too surprising since it has far more electrons than ethylene. However, the bond energies, IPs, and other intensive energies of these two molecules don't differ by much. To cleave ethylene's double bond costs 6.4 eV, while to break a C-C single bond near the middle of the polypeptide requires 3.9 eV.

As the molecular size grows, the total (extensive) energy increases correspondingly but the magnitudes of the intensive energies do not change correspondingly. Therefore, to reliably compute the intensive properties by subtracting two total energies requires that the extensive energies be computed to higher and higher precision as the molecular size grows. For example, to determine the 3.9 eV bond energy of the polypeptide, the total energy ($-69\,182.78\,eV$) has to be precise to 6 digits. Clearly, this situation becomes futile for very large molecules and is why QC's basic ingredients (e.g., integrals and matrix eigenvalues) need to be computed using doubleprecision (or higher) arithmetic even for small or medium-size molecules. This suggests that trying to calculate eigenvalues of the Schrödinger equation head-on might not offer an optimal connection to experiments.

The good news is that methods exist within QC for directly calculating intensive properties, although most of them still require that an approximate wave function (HF, KS-DFT, CI, MP, MCSCF, or CC) first be calculated, even though its energy is not used to determine the desired intensive property. As a result, the computational scaling of the wave function determination remains. The properties that can be evaluated in this manner include excitation energies (EEs), IPs, EAs, polarizabilities, and more. However, there are properties such as reaction activation energies, reaction ΔE and ΔH values, and non-covalent intermolecular interactions that are not subject to these direct-calculation approaches; they still need to be determined through the Schrödinger energy-difference route.

Most of these so-called response methods express the intensive properties as solutions to matrix eigenvalue problems. For IPs and EAs, the matrices have dimension $M \times M$ and are embodied in equations-of-motion (EOM) or Green's function methods whose reference wave functions have various levels of rigor, including HF (i.e., Koopmans' theorem), MP2,²⁸ MP3,^{29–31} up to the CC level,³² and others^{33,34} relate to KS-DFT orbitals in their underlying foundation. For EEs, the matrices have (larger) dimension $(N \times M) \times (N \times M)$ that are usually treated using iterative techniques.^{22,35} These methods³⁶ include CI including single excitations (CIS), time-dependent DFT (TDDFT), the Random-Phase Approximation (RPA), and CC-based^{37,38} EOM theories. All of these response methods produce intensive energies because their working equations are derived by analytically subtracting all terms common to the extensive energies of the ground and excited states (for EEs) or the neutral and ionic states (for IPs and EAs). Because they produce the response in energy that accompanies displacement of a nuclear center, I would add to these theories the well-known analytical gradient (and higher derivatives) methods^{39,15} that have proven extremely useful in finding equilibrium geometries and transition states.

Journal of the American Chemical Society

8. SUMMARY

I hope this Perspective allows the reader to better understand why QC is so complicated (because electrons move in a way that is described by quantum, not classical, mechanics), what QC's various methods are and how they describe the electrons' correlated motions, where the computational and conceptual complexities in QC arise, and how to estimate how the computing time of your QC calculation will change if you change basis sets or move from one molecule to another with the same basis type. [Refer back to section 4 for information on how these two distinct cases are handled.] In Table 2, I remind you about the dominant scaling in various methods to be of assistance.

Finally, let me touch briefly on "what method and basis should I use?" It depends on what you want to do. To determine equilibrium geometries and harmonic vibrational frequencies, it would make sense to begin by using HF or KS-DFT within a modest basis. If you know that the state of interest is multiconfigurational, it would be better to replace HF/KS-DFT by MCSCF. This would be essential when studying a transition state whose barrier involves an avoided crossing. In carrying out the geometry optimization, you should make use of analytical gradients, and to obtain the harmonic vibrational frequencies, analytical second-derivatives could be used. The resultant geometry and frequency data could then be used to estimate entropies and enthalpies. After that, it would be wise to expand the basis and move beyond HF to include correlation using MP, CI, and CC if feasible, ultimately reaching the most precise method and best basis you can afford to employ; this is where you need to use knowledge of scaling to help decide what is feasible. This strategy can also be followed to study reaction paths and bond dissociation energies, but when faced with a barrier resulting from an avoided crossing, it would be wise to test results obtained with any single-determinant-based method (e.g., MP or CC) by also using a multi-determinant-based method (e.g., MCSCF or CI).

If you are interested in EEs, IPs, or EAs, it would be wise to first try CIS or TDDFT for EEs or Koopmans' theorem or Green's function for IPs or EAs, again starting with a modest basis. For EEs, CIS is often a wise first step as it is not very computationally taxing, although it should be viewed as only that, since it is not treating electron correlation. If an excited state is expected to have a large charge-transfer character, TDDFT should be avoided as it has difficulties in these situations⁴⁰ (the self-interaction problem). After that, the basis should be enhanced and one should consider using EOM-CClevel EE, IP, or EA methods if possible, again using knowledge of scaling to guide you.

QC continues to undergo rapid growth in terms of the number of people using it, while growing computing power (e.g., GPUs, quantum computers, etc.) and new methods (e.g., machine learning, artificial intelligence, density matrix renormalization group, Monte-Carlo, etc.) promise to offer even more capability. In my opinion, these evolutions are likely to produce even greater "complications" than QC currently displays. However, I think scholars from the experimental and other theoretical chemistry communities will continue to be able to surmount such difficulties, even if they also continue to complain about them, even if only because of QC's utility.

In line with the tone of this Perspective article, I am not giving an exhaustive list of literature references. Rather, I provide references primarily to articles or chapters of an overview or review nature that I found especially pedagogical and that I think offer an efficient avenue for the readers I am attempting to reach to gain more in-depth insights. Within these references, the reader can find ample sources to the earlier works by many of the pioneering workers and to their efforts over the ensuing years.

AUTHOR INFORMATION

Corresponding Author

Jack Simons – Henry Eyring Center for Theoretical Chemistry, Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States; orcid.org/0000-0001-8722-184X; Email: jack.simons@utah.edu

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.2c13042

Notes

The author declares no competing financial interest.

REFERENCES

(1) See Chapters 10–12 in Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic Structure Theory*; Wiley & Sons: New York, 2000. This advanced and very thorough textbook covers much of the technical foundations of wave function-based QC.

(2) See Chapter 14 of the book cited in ref 1.

(3) See pp 193-195 of the book cited in ref 1.

(4) Bartlett, R. J. Coupled-Cluster Theory: An Overview of Recent Developments. In *Modern Electronic Structure Theory*, Part II; Yarkony, D. R., Ed.; World Scientific Publishing, 1995; pp 1047–1131. Also, see Chapter 23 of the book cited in ref 1.

(5) Jørgensen, P.; Simons, J. Second Quantization-Based Methods in Quantum Chemistry; Academic Press: New York, 1981.

(6) See Chapter 9 of the book cited in ref 1.

(7) See, for example, https://www.basissetexchange.org/.

(8) The video on basis sets can be found at https://www.youtube. com/watch?v=q_tjk3YvvjE These two links provide even more indepth information on basis set nomenclature: https://gaussian.com/ basissets/ and https://www.basissetexchange.org/. More videos on various aspects of electronic structure theory can be accessed at https:// simons.hec.utah.edu/YouTubeLinks.html.

(9) Alternatively, https://cccbdb.nist.gov/nbf.asp is a web site where you can enter your molecule's formula and it will return the *M* values for various AO basis sets (if it exists in their database). If your molecule does not appear there, you can use the site to determine the *M* values for each of the atoms or functional groups in your molecule and then sum them up.

(10) Subotnik, J. E.; Shao, Y.; Liang, W.; Head-Gordon. An efficient method for calculating maxima of homogeneous functions of orthogonal matrices: Applications to localized occupied orbitals. *J. Chem. Phys.* **2004**, *121*, 9220–9229.

(11) Liakos, D. G.; Sparta, M.; Kesharwani, K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *J. Chem. Theor. Comput.* **2015**, *11*, 1525–1539.

(12) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *J. Chem. Phys.* **2009**, *130*, 114108. (13) See sections 9.11–9.13 in ref 1.

(14) Several of these approaches (resolution of identity (RI), Cholesky, and more) are discussed in Almlöf, J. Direct Methods in Electronic Structure Theory. In *Modern Electronic Structure Theory*; Yarkony, D. R., Ed.; World Scientific Publishing, 1995; pp 110–151.

(15) Schlegel, H. B. Geometry Optimization on Potential Energy Surfaces. In *Modern Electronic Structure Theory*; Yarkony, D. R., Ed.; World Scientific Publishing, 1995; pp 459–500.

(16) Becke, A. Exchange-Correlation Approximations in Density-Functional Theory, In *Modern Electronic Structure Theory*; Yarkony, D. R., Ed.; World Scientific Publishing, 1995; pp 1022–1046. (17) Baerends, E. J.; Gritsenko, O. V. A Quantum Chemical View of Density Functional Theory. J. Phys. Chem. A 1997, 101, 5383-5403.
(18) Verma, P.; Truhlar, D. G. Status and Challenges of Density Functional Theory. Trends Chem. 2020, 2, 302-318.

(19) See pp 454–458 of the book cited in ref 1.

(20) Anusiewicz, I.; Skurski, P.; Simons, J. Finding Valence Antibonding Levels while Avoiding Rydberg, Pseudo-continuum, and Dipole-Bound Orbitals. J. Am. Chem. Soc. **2022**, 144 (25), 11348– 11363.

(21) Stowasser, R.; Hoffmann, R. What do the Kohn-Sham orbitals and eigenvalues mean? J. Am. Chem. Soc. **1999**, 121, 3414–3420.

(22) Davidson, E. R.; Thompson, W. J. Monster Matrices: Their Eigenvalues and Eigenvectors. *Comput. Phys.* **1993**, *7*, 519–522.

(23) Grimme, S. Semiempirical GGA-Type Density Function Constructed with a Long-Range Dispersion Correction. J. Comput. Chem. 2006, 27, 1787–1799.

(24) Ehrlich, S.; Moellmann, J.; Grimme, S. Dispersion-Corrected Density Functional Theory for Aromatic Interactions in Complex Systems. *Acc. Chem. Res.* **2013**, *46*, 916–926.

(25) Chermette, H.; Ciofini, I.; Mariotti, F.; Daul, C. Correct dissociation behavior of radical ions such as H_2^+ in density functional calculations. *J. Chem. Phys.* **2001**, *114*, 1447–1453.

(26) Fuchs, M.; Niquet, Y.-M.; Gonze, X.; Burke, K. Describing static correlation in bond dissociation by Kohn-Sham density functional theory. *J. Chem. Phys.* **2005**, *122*, 094116.

(27) Gagliardi, L.; Truhlar, D. G.; Li Manni, G.; Carlson, R. K.; Hoyer, C. E.; Bao, J. L. Multiconfigurational Pair-Density Functional Theory: A New Way to Treat Strongly Correlated Systems. *Acc. Chem. Res.* **2017**, *50*, 66–73.

(28) Doll, J. P.; Reinhardt, W. P. Many-bond Green's Functions for Finite, Nonuniform Systems: Applications to Closed Shell Atoms. *J. Chem. Phys.* **1972**, *57*, 1169–1184.

(29) Ortiz, J. V. Dyson orbital concepts for description of electrons in molecules. J. Chem. Phys. **2020**, 153, 070902.

(30) Simons, J.; Smith, W. D. Theory of the Electron Affinities of Small Molecules. J. Chem. Phys. **1972**, *57*, 3787–3792.

(31) Jørgensen, P.; Simons, J. A Complete Treatment of the Electron Propagator through Third Order. J. Chem. Phys. 1975, 63, 5302-5304.
(32) Nooijen, M.; Bartlett, R. J. Equation of motion coupled cluster

method for electron attachment. J. Chem. Phys. **1995**, 102, 3629–3647. (33) Govoni, M.; Galli, G. Large Scale GW Calculations. J. Chem.

Theory Compt. 2015, 11, 2680–2698.

(34) Onida, G.; Reining, L.; Rubio, A. Electronic excitations: density functional versus many-body Green's function approaches. *Rev. Mod. Phys.* **2002**, *74*, 601–659.

(35) Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. Toward a systematic molecular orbital theory for excited states. *J. Chem. Phys.* **1992**, *96*, 135–149.

(36) See, for example, Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Excited-State Electronic Structure with Configuration Interaction Singles and Tamm-Dancoff Time-Dependent Density Functional Theory on Graphical Processing Units. *J. Chem. Theory Comput.* **2011**, *7*, 1814–1823.

(37) Stanton, J. F.; Bartlett, R. J. The equation of motion coupledcluster method: A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *J. Chem. Phys.* **1993**, *98*, 7029–7039.

(38) Krylov, A. I. Eqations-of-Motion Coupled-Cluster Methods for Open-Shell and Electronically Excited Species: The Hitchhiker's Guide to Fock Space. *Annu. Rev. Phys. Chem.* **2008**, *59*, 433–462.

(39) Csaszar, A. G., Fogarasi, G., Schaefer, H. F., III, Szalay, P. G., Eds. Molecular Quantum Mechanics: Analytical Gradients and Beyond, Selected Papers of Peter Pulay; Eötvös L. University (ELTE) Institute of Chemistry: Budapest, Hungary, 2007.

(40) Dreuw, A.; Head-Gordon, M. Failure of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excited States: The Zincbacteriochlorin-Bacteriochlorin and Bacteriochlorophyll-Spheroidene Complexes. J. Am. Chem. Soc. **2004**, 126, 4007– 4016.

Recommended by ACS

Directional Ionic Bonds

pubs.acs.org/JACS

Illia Hutskalov, Ilija Čorić, et al.	
APRIL 07, 2023	
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY	READ 🗹

Hydrogen Bonding with Hydridic Hydrogen–Experimental Low-Temperature IR and Computational Study: Is a Revised Definition of Hydrogen Bonding Appropriate?

Svatopluk Civiš, Pavel Hobza, et al.

APRIL 10, 2023	
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY	READ 🗹

Quid Pro Flow

Andrea Laybourn, Anna G. Slater, et al.	
EBRUARY 14, 2023	
OURNAL OF THE AMERICAN CHEMICAL SOCIETY	READ 🗹

Quantum Alchemy Based Bonding Trends and Their Link to Hammett's Equation and Pauling's Electronegativity Model

Michael J. Sahre, O. Anatole von Lilienfeld, et al.

MARCH 02, 2023	
OURNAL OF THE AMERICAN CHEMICAL SOCIETY	READ 🗹

Get More Suggestions >