# Communication: A mean field platform for excited state quantum chemistry

Jacqueline A. R. Shea, and Eric Neuscamman

Citation: The Journal of Chemical Physics **149**, 081101 (2018); doi: 10.1063/1.5045056
View online: https://doi.org/10.1063/1.5045056
View Table of Contents: http://aip.scitation.org/toc/jcp/149/8
Published by the American Institute of Physics

# Communication: A mean field platform for excited state quantum chemistry

Jacqueline A. R. Shea[1] and Eric Neuscamman[1,2,a)]
[1]*Department of Chemistry, University of California, Berkeley, California 94720, USA*
[2]*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

(Received 15 June 2018; accepted 10 August 2018; published online 22 August 2018)

We present a mean field theory for excited states that is broadly analogous to ground state Hartree-Fock theory. Like Hartree-Fock, our approach is deterministic, state-specific, applies a variational principle to a minimally correlated ansatz, produces energy stationary points, relaxes the orbital basis, has a Fock-build cost-scaling, and can serve as the foundation for correlation methods such as perturbation theory and coupled cluster theory. To emphasize this last point, we pair our mean field approach with an excited state analog of second order Møller-Plesset theory and demonstrate that in water, formaldehyde, neon, and stretched lithium fluoride, the resulting accuracy far exceeds that of configuration interaction singles and rivals that of equation of motion coupled cluster. *Published by AIP Publishing.* https://doi.org/10.1063/1.5045056

In a nutshell, Hartree-Fock (HF) theory[1] applies the ground state variational principle to a wave function ansatz that includes only those correlations that are absolutely necessary to produce a qualitatively correct description of the electrons in a simple molecular ground state. Indeed, HF theory's Slater determinant hews as closely as possible to a classical mean field state, in which the particles would be completely uncorrelated, while accommodating the Pauli correlations that must be included when describing electrons. This simplicity keeps HF theory (relatively) affordable as it proceeds to minimize its variational principle and make the energy stationary with respect to changes in the molecular orbital basis. That HF theory has for decades been the central platform on which high-accuracy weak correlation treatments are built[2,3] is a reminder of how valuable a qualitatively correct minimally correlated wave function with relaxed orbitals can be. For strongly correlated ground states in which an unambiguous minimal active space exists, the story is much the same, but with complete active space self-consistent field (CASSCF)[4–7] and its descendants[8–15] replacing HF theory as the basic platform upon which weak correlation methods are built.[16–30] In this paper, we develop an affordable mean field platform for simple excited states while also providing an initial weak correlation treatment and a discussion of how the approach can be generalized to strongly correlated excitations.

Unlike the situation for ground states, even zeroth order descriptions of most excited states require more than Pauli correlations. For example, consider the open shell singlet of a simple HOMO → LUMO excitation. In this state, two opposite-spin electrons are strongly correlated with each other so as not to occupy the same orbital at the same time, implying that a minimally correlated excited state ansatz must incorporate correlations not present in a Slater determinant. An obvious candidate for this job is a configuration interaction singles (CIS) ansatz[31] which, to stay in line with ground state mean field methods like HF and CASSCF, has had its energy made stationary with respect to relaxations of the molecular orbital basis. While this direction has of course been explored before, previous approaches have tended to approximate the orbital relaxation in order to suppress cost.[32–35] For example, Subotnik's OO-CIS approach[32] achieves its efficiency by employing an incomplete and HF-approximated Newton-Raphson optimization. Inspired by the need for fully relaxed excited state orbitals in situations like charge transfer and core spectroscopy, we seek here to achieve them at the same cost-scaling as CIS by combining recent progress in excited state variational principles[36–39] with a compound application of automatic differentiation (AD). To demonstrate the efficacy of this approach as a mean field platform on which to build excited state correlation methods, we use it as the basis for an excited state analog of second order Møller-Plesset (MP2) theory.[3]

As we are looking to describe an excited state, minimizing the energy

$$E = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{1}$$

in pursuit of the ground state variational principle will not guarantee convergence to the desired state. Instead, we will seek to minimize the Lagrangian

$$L = W + \vec{\mu}\cdot\frac{\partial E}{\partial \vec{\nu}} \tag{2}$$

in which the excited state variational principle[40]

$$W = \frac{\langle\Psi|(\omega - \hat{H})^2|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{3}$$

guarantees that a sufficiently flexible wave function will converge to the exact (excited) energy eigenstate with energy

a)eneuscamman@berkeley.edu

closest to $\omega$ while the Lagrange multipliers $\vec{\mu}$ guarantee that an energy stationary point with respect to the wave function variables $\vec{\nu}$ will be achieved even when working with an approximate ansatz. While we could have done without $W$ if we merely wished to ensure that an energy stationary point was reached, its presence guarantees that in the limit of a sufficiently flexible wave function, the global minimum of $L$ will be the stationary point corresponding to the desired excited state. Although we do not pursue it here, one can imagine augmenting $W$ with functions of the dipole moment or other observables in order to differentiate between states that are energetically degenerate.

To make the minimization of this Lagrangian affordable, we must deal both with the difficulty of the $\hat{H}^2$ term and with the fact that derivatives of $L$ with respect to the variational parameters $\vec{\nu}$ (e.g., the CI coefficients and the orbital rotation variables) lead to second derivatives of the energy. To avoid $\hat{H}^2$, we can resolve an identity in the basis of Slater determinants

$$W = \frac{1}{\langle\Psi|\Psi\rangle}\sum_I \langle\Psi|(\omega - \hat{H})|I\rangle\langle I|(\omega - \hat{H})|\Psi\rangle \qquad (4)$$

and, following the approach of coupled cluster (CC) theory,[2] make an approximation in which we restrict our attention (and the range of the sum) to the most chemically relevant corner of Hilbert space, in this case the span of the closed shell "ground state" determinant and the singles excitations. Note that $W$ need not be evaluated exactly for the theory to operate correctly, as its role is merely to guide the optimization into the correct stationary point. So long as $W$ provides a sufficiently strong nudge to get us close, the Lagrange multiplier term in Eq. (2) will ensure convergence to the stationary point. This type of "nudged" convergence to a stationary point has been seen before in the application of full configuration interaction quantum Monte Carlo to excited states,[41] where a heavily approximated projection operator that could only serve to nudge the state propagation away from the lower states was sufficient to converge the imaginary time evolution onto excited states. When an approximate nudge is insufficient, our formulation allows for a systematic retreat to safety via increasingly accurate evaluations of $W$, but we stress that in our initial testing this has yet to prove necessary. Indeed, in all cases tested so far, the even more aggressive approximation

$$W \approx (\omega - E)^2 \qquad (5)$$

leads the minimization of $L$ to converge on to the same stationary point as when truncating the sum in Eq. (4). If this equivalence is maintained after more extensive testing in the future, there would be a strong simplicity argument in favor of employing Eq. (5).

Unlike the $\hat{H}^2$ terms, the challenge of second energy derivatives can be overcome by a compound use of automatic differentiation (AD),[42] which ensures that all of the derivatives of a many-input-single-output function can be evaluated for a small constant multiple of the cost of evaluating the function itself. As AD's use in quantum chemistry is still in its infancy,[43–45] let us briefly explain the principles so that its usefulness for our purposes is clear. First, consider that

TABLE I. Expressions for the derivatives of the four basic arithmetic functions in terms of the (presumably stored) values of their inputs $a$ and $b$ and output $f$.

| $f(a,b)$ | $a+b$ | $a-b$ | $ab$ | $a/b$ |
|---|---|---|---|---|
| $\partial f/\partial a$ | 1 | 1 | $b$ | $1/b$ |
| $\partial f/\partial b$ | 1 | −1 | $a$ | $-f/b$ |

many complicated functions, such as the CIS energy, can be written as a graph in which each node is one of the four basic binary operations shown in Table I in which two input quantities $a$ and $b$ go in and the output quantity $f$ that comes out may then become one of the inputs for one or more of the other nodes in the graph. One can compute the overall function value $g(\vec{x})$ by traversing the graph, starting from the dangling edges that are the inputs $\vec{x}$ and moving forward through all the nodes until the final output $g(\vec{x})$ is reached. Now, if one can afford to store the outputs of each node in the graph, it becomes possible to evaluate all of $g$'s first derivatives with respect to the elements of $\vec{x}$ for a cost that is a small constant multiple of the evaluation cost of $g$ via a sort of reverse traversal of the graph.[42] Crucially, if one considers a given node and assumes that they already know the partial derivative of $g$ with respect to the output $f$ of that node, then the chain rule and the simple derivative formulas in Table I ensure that the partial derivatives of $g$ with respect to the node's inputs $a$ and $b$ can be evaluated via three or fewer binary arithmetic operations. Starting at the final node and working backwards, one finds that the number of operations required to get all the derivatives of $g$ with respect to all the intermediates and all the elements of $\vec{x}$ is not worse than four times the number of nodes in the graph, and so the cost to get all the derivatives $\partial g/\partial\vec{x}$ is a small constant multiple of the cost to evaluate $g$ itself.

This approach, known as reverse accumulation, provides a straightforward if tedious recipe for constructing a low cost implementation of analytic derivatives, and we may, for example, apply it to the CIS energy $E$ to obtain an efficient function for the energy first derivatives that appear in Eq. (2). Folding this logic over on itself, we recognize that thanks to the dot product of these efficient derivatives with $\vec{\mu}$, $L$ itself can be implemented as a many-input-single-output function of the variables $\vec{\nu}$ and $\vec{\mu}$ whose cost is a small constant multiple of that of $E$. By a second, compound application of reverse accumulation, we may thus arrive at an implementation that delivers analytic derivatives of $L$ with respect to both $\vec{\nu}$ and $\vec{\mu}$ for a cost that is also a small constant multiple of that of a CIS energy evaluation and with an additional memory requirement for intermediate storage that scales only as the square of the system size. Note that, in practice, setting up reverse accumulation is tedious work that can and has been automated in many software packages. In the present study, we have leveraged the machine learning community's rapid progress in reverse accumulation software by employing the TensorFlow[46] framework to evaluate our Lagrangian's derivative vector, the norm of which we minimize by a quasi-Newton approach.[47] As the cost of a CIS energy evaluation scales as the cost of a Fock matrix build, so does evaluation

of $L$ and the necessary derivatives, leading our excited state mean field (ESMF) approach to have the same cost scaling as CIS.

While we have motivated this Lagrangian approach with the prospects of relaxing the orbitals in CIS, the logic supporting its construction is much more general, and indeed $L$ can in principle be employed efficiently with any ansatz for which $E$ and a reasonable approximation to $W$ can be efficiently evaluated. $L$ could, for example, be used as a more rigorous alternative to maximum overlap methods[48] when optimizing a CASSCF wave function for an individual excited state. Although we are quite curious about this possibility, we do not pursue strongly correlated excited state treatments in this study. Instead, we focus on delivering a fully orbital-relaxed CIS wave function and testing its ability to act as a platform for excited state correlation treatments in the same way HF theory does for ground states.

To this end, we employ the CIS-like ansatz

$$|\Phi\rangle = e^{\hat{X}}\left(c_0|0\rangle + \sum_{ia} c_{ia}|_i^a\rangle + \sum_{\bar{i}\bar{a}} c_{\bar{i}\bar{a}}|_{\bar{i}}^{\bar{a}}\rangle\right), \tag{6}$$

in which excitations are labeled by alpha ($i$, $a$) or beta ($\bar{i}$, $\bar{a}$) indices and the closed shell determinant $|0\rangle$ is included to help the orbital-relaxed excited state better maintain orthogonality to the RHF ground state. The vector of variables $\vec{v}$ that we optimize via $L$ includes the coefficients $c_0$, $c_{ia}$, and $c_{\bar{i}\bar{a}}$, as well as the elements of the matrix $X$ that defines the orbital rotation operator,

$$\hat{X} = \sum_{p<q} X_{pq}\left(\hat{a}_p^+\hat{a}_q - \hat{a}_q^+\hat{a}_p\right), \tag{7}$$

which for the present study we constrain so as to keep the orbitals spin-restricted. Although we do not explore the possibility in the present study, this approach could be generalized to work with a CASSCF wave function by replacing the CIS expansion in Eq. (6) with the CASSCF CI expansion. Given the much greater size of such CI expansions, it may in that case be more effective to use $L$ only for the orbital rotation optimization and to instead rely on modern CI solvers to keep the energy stationary with respect to the CI coefficients.

While the excited state mean field (ESMF) ansatz $|\Phi\rangle$ is more flexible than CIS and might therefore be expected to be more accurate, one should remember that, due to the significant effect of weak correlation on energetics, HF theory itself is quite poor quantitatively even when it is qualitatively a good zeroth order wave function. By the same reasoning, total energies and energy differences from ESMF wave functions (which revert to RHF when $\omega$ targets the ground state) are not likely to be competitive in accuracy with methods that incorporate correlation effects, such as equation of motion coupled cluster singles and doubles (EOM-CCSD).[49]

As in the ground state, we are interested in the ESMF wave function not as a destination in and of itself, but rather as a reliable platform upon which to construct correlation treatments that can reasonably hope to achieve more quantitative accuracy. Considering first the prospects for a coupled cluster theory, we note that due to the natural termination of the Baker-Campbell-Hausdorff (BCH) expansion in traditional, similarity-transformed CC singles and doubles (CCSD),[2] the usual approach of projecting the CC eigenvalue equation

$$\left(e^{-\hat{T}}\hat{H}e^{\hat{T}} - E\right)|\Phi\rangle = 0 \tag{8}$$

into the space of low order (in this case internally contracted) excitations from the reference will, as in the ground state theory, lead to a polynomially complex system of equations for the cluster amplitudes. However, due to the fact that bare triples excitations would be present within the internally contracted doubles, the cost scaling of such an approach would, although still polynomial, be substantially higher than in ground state CCSD. More enticing is the prospect of leveraging the fact that the potentially long-range nature of the excitation and the orbital relaxations it induces should already be accounted for in the ESMF reference state, allowing lower-cost local CC approaches[50–53] to focus on what they do best: treating short-ranged weak correlation.

As intriguing as CC methods may be, the principle of Occam's razor suggests that an analog of the much simpler MP2 theory would be a wiser starting point for investigations into post-ESMF correlation methods. We can arrive at just such a method, which we will denote as ESMP2, by applying standard Rayleigh-Schrödinger perturbation theory to the zeroth order Hamiltonian

$$\hat{H}_0 = \hat{R}(\hat{F} - \hat{H})\hat{R} + \hat{P}\hat{H}\hat{P} + \hat{Q}\hat{F}\hat{Q}, \tag{9}$$

where $\hat{F}$ is the Fock operator with respect to the ESMF one-body density matrix, $\hat{R} = |\Phi\rangle\langle\Phi|$, $\hat{P}$ is the projector to the span of $|0\rangle$ and the singles excitations in the ESMF orbital basis, and $\hat{Q} = 1 - \hat{P}$. Note that $\hat{P}|\Phi\rangle = |\Phi\rangle$ and $\hat{Q}|\Phi\rangle = 0$. This choice of $\hat{H}_0$ leads to the MP2-like zeroth order relationship,

$$\hat{H}_0|\Phi\rangle = E_0|\Phi\rangle = \langle\Phi|\hat{F}|\Phi\rangle|\Phi\rangle, \tag{10}$$

and indeed we see that when we set $\omega$ so as to target the ground state, $|\Phi\rangle$ becomes the RHF state and ESMP2 simplifies to MP2. In the excited state case, as in traditional MP2, the first order wave function contains no determinants with fewer than two excitations, which is a consequence of including the $\hat{P}\hat{H}\hat{P}$ term in $\hat{H}_0$. We also maintain the relationship $E_{\mathrm{ESMF}} = E_0 + E_1$ in direct analogy to the MP2 relationship $E_{\mathrm{HF}} = E_0 + E_1$. As the Fock operator is not diagonal for excited states, the first order amplitudes on doubles and triples are found by inverting $\hat{F} - E_0$ via the minimal residual Krylov subspace method, for which the MP2-style denominators are an excellent preconditioner. Note that this approach produces a fully excited-state-specific first order wave function, as opposed to the CIS(D) method where the triples are a product of the CIS coefficients and the ground state MP2 amplitudes.[56,57]

While we have, for the sake of simplifying development and testing, written our pilot ESMP2 implementation in a fully uncontracted form whose triples part has an $N^7$ cost scaling, multiple avenues exist for recovering the $N^5$ scaling of traditional MP2. On the one hand, we could exploit sparsity in the ESMF coefficient matrix, which we observe to have only a handful of elements that are not small. Indeed, we have tested this idea by setting all but three of the elements to zero when solving the ESMP2 linear equation and evaluating its second

TABLE II. Comparisons for singlet excitations. For EOM-CC(2,3), we report excitation energies in eV, with other methods' results reported as excitation energy errors in eV relative to EOM-CC(2,3) and summarized in terms of mean unsigned error (UE) and maximum UE. For stretched LiF, transitions are relative to the closed shell ionic state and are labeled by the F → Li orbitals involved (the bond is aligned along the $z$ axis). For other cases, transitions are relative to the ground state. Traditional methods were evaluated with Molpro[54] and QChem.[55]

| State | EOM CC(2,3) | CIS | OO-CIS | CIS(D) | ESMF | ESMP2 | EOM CCSD |
|---|---|---|---|---|---|---|---|
| $H_2O$, cc-pVDZ, $r$ = 0.9614 Å, $a$ = 104.4° | | | | | | | |
| $n \to \sigma^*$ | 8.22 | 0.96 | −0.35 | −0.18 | −0.74 | −0.01 | −0.08 |
| $n \to \pi^*$ | 10.25 | 0.70 | −0.32 | −0.11 | −0.77 | 0.00 | −0.06 |
| $\sigma \to \sigma^*$ | 10.86 | 0.94 | −0.26 | −0.17 | −0.73 | −0.05 | −0.06 |
| $\sigma \to \pi^*$ | 12.93 | 0.65 | −0.26 | −0.10 | −0.83 | −0.03 | −0.04 |
| $\pi \to \sigma^*$ | 14.82 | 0.18 | −0.47 | −0.06 | −0.82 | −0.04 | −0.01 |
| $CH_2O$, cc-pVDZ, geometry: B3LYP/cc-pVTZ | | | | | | | |
| $n \to \pi^*$ | 4.22 | 0.40 | −0.14 | 0.01 | −0.98 | −0.09 | −0.05 |
| $\pi \to \pi^*$ | 10.02 | 0.26 | −0.30 | 0.46 | −1.12 | −0.06 | 0.12 |
| $n \to \sigma^*$ | 8.70 | 1.68 | 0.50 | −0.46 | −0.71 | 0.09 | −0.08 |
| LiF, cc-pVDZ, $r$ = 8.0 Å | | | | | | | |
| $2p_x \to 2s$ | −2.68 | 1.49 | −3.03 | −1.44 | −1.02 | 0.10 | −0.22 |
| $2p_z \to 2s$ | −2.68 | 1.44 | −1.24 | −1.31 | −1.02 | 0.10 | −0.22 |
| $2p_z \to 2p_z$ | −0.84 | 1.48 | −3.20 | −1.48 | −1.02 | 0.10 | −0.22 |
| Ne, cc-pVTZ | | | | | | | |
| $2s \to 3p$ | 64.32 | 2.66 | 1.74 | 0.41 | 1.36 | 0.34 | 0.06 |
| Mean UE | | 1.07 | 0.98 | 0.51 | 0.93 | 0.08 | 0.10 |
| Max UE | | 2.66 | 3.20 | 1.48 | 1.36 | 0.34 | 0.22 |

order energy for the molecules discussed below and found that excitation energy predictions are not strongly affected. On the other hand, we could follow the internally contracted approach of Evangelista and co-workers.[58]

Table II reveals that, like its HF cousin in the ground state, the ESMF approximation does not confer quantitative accuracy. This behavior can be understood as a direct consequence of our design goal of hewing as closely as possible to a classical mean field theory. By including only those correlations that are absolutely necessary to realize a fermionic excited state (namely, Pauli correlations and the open shell correlation), ESMF is missing all weak correlations and so, like HF, does not produce quantitative energies. The fact that ESMF tends to underestimate excitation energies can also be understood in terms of what correlations are included. Indeed, in creating their open-shell arrangements, ESMF gives each excited state roughly one pair's worth of electron correlation, and so these states' energies are biased low compared to that of the closed shell ground state. Thus, in direct analogy to HF theory, the simplicity and mean field nature of ESMF prevents it from delivering accurate energetics on its own, but this of course was not the intention. What is more important is the question of whether ESMF can match HF theory's ability to act as a foundation for correlation methods, a question that our early results appear to answer strongly in the affirmative.

The data reveal that in water, formaldehyde, stretched lithium fluoride, and neon, ESMP2 rivals EOM-CCSD in accuracy. Its errors are typically at least a factor of five smaller than CIS, and it substantially outperforms CIS(D) in the charge transfer states of stretched LiF. Our primary explanation for this success is that the reference wave function's mean-field quality of having fully relaxed orbitals places it at a similar "distance" from the correct wave function as for HF in the ground state, with the subtle effects of weak electron correlation being all that is missing. By contrast, CIS(D), which is also inspired by and closely entwined with MP2 theory, is asking its perturbation to capture both correlation effects and orbital relaxation. While the former are typically small in systems that are not strongly correlated, the size of the latter is much more system dependent. Table II shows that in the low lying transitions of water, in which the overall spatial distribution of electrons is not greatly changed, both CIS(D) and ESMP2 are highly accurate. In stretched LiF, however, the transitions convert between ions and neutral atoms, and as one would expect these large charge density changes lead to strong orbital relaxation effects, as revealed by comparing the CIS and ESMF energies. These relaxations are much more difficult to treat perturbatively and, to make matters worse, cause the closed-shell-state MP2 amplitudes that all CIS(D) states rely on to be less appropriate for the open-shell states. Looked at from this perspective, it is not surprising that ESMP2, thanks to its orbital-relaxed reference and fully state-specific perturbation, delivers more uniform accuracy across charge-transfer, valence, and Rydberg states alike. Indeed, when compared to the high-level benchmark of EOM-CC(2,3),[59] the maximum unsigned error of ESMP2 is less than the mean unsigned error for CIS(D).

We can also compare our results with more recent attempts to provide orbital relaxations for CIS. The OO-CIS method, for example, provides relaxations via a single Newton-Raphson orbital optimization step in which the HF Hessian is used as an approximation to the CIS orbital Hessian.[32] As OO-CIS lacks a treatment for weak correlation, it is not surprising that in $H_2O$, $CH_2O$, and Ne, its accuracy is for most states better than CIS (which has neither orbital relaxation nor a weak correlation treatment) but worse than CIS(D), EOM-CCSD, and ESMP2, as shown in Table II. The large errors that OO-CIS makes in LiF can be understood as a consequence of its Hessian approximation, as LiF's smallest HF Hessian eigenvalues are about an order of magnitude smaller than those in the other molecules, and so when this Hessian is inverted to get the OO-CIS orbital relaxation, the resulting Newton-Raphson step is much too large. In this system at least, it appears that the HF Hessian is not an effective surrogate for the CIS Hessian. Unlike OO-CIS and our approach, the variationally orbital-adapted CIS (VOA-CIS) method seeks to provide orbital relaxations through specially chosen expansions of the configuration interaction space.[34] However, there is at present no single prescription for defining the expansion (the developers explore at least eight possibilities in their initial paper[34]) and so instead of making an extensive and hard-to-interpret direct comparison, we will point out the developers' conclusion that "most of the time, VOA-CIS closely follows CIS(D)" and their data that shows that it is not unusual for VOA-CIS to be in error by between 0.5 eV and 1 eV for single excitation energies.[34] In summary, our preliminary testing

shows ESMP2, with its inclusion of both orbital relaxation and state-specific correlation, to be closer in its behavior to EOM-CCSD than to previous attempts at augmenting the CIS wave function.

We have presented an excited state mean field theory and investigated its potential as a platform on which to build excited state correlation treatments. Like HF theory, the method relies on making a minimally correlated wave function's energy stationary with respect to orbital relaxations. While HF does this for a Slater determinant, we do so for a CIS-like wave function in order to accommodate the basic structure of simple excitations, and so the cost of our mean field optimization has the same scaling with system size as CIS. Unlike HF theory, our approach incorporates an excited state variational principle into its Lagrangian so that ground and excited states are treated equally. In exploring the method's potential as a platform for correlation treatments, we have constructed an excited state analog of MP2 theory and found that, in initial tests, its accuracy rivals that of EOM-CCSD.

Looking to the future, there are a number of important questions to consider about this excited state mean field approach. First, how will it fare when applied to larger and more complicated charge transfer systems, core excitations, and Rydberg states? Second, will its advantages be maintained when paired with more complex and expensive methods such as coupled cluster and the complete active space self-consistent field? Finally, can this minimally correlated, excited-state-specific wave function be usefully employed as a replacement for the Slater determinant in an excited-state-specific generalization of the Kohn-Sham approximation? We look forward to exploring these and other questions that will undoubtedly arise in the context of excited state mean field theory.

We thank Joe Subotnik for many helpful discussions, and we acknowledge funding through the Early Career Research Program of the Office of Science, Office of Basic Energy Sciences, the U.S. Department of Energy, Grant No. DE-SC0017869. Calculations were performed both on our own desktop computers and at the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

[1] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications, Mineola, N.Y., 1996).
[2] R. J. Bartlett and M. Musial, Rev. Mod. Phys. **79**, 291 (2007).
[3] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic Structure Theory* (John Wiley & Sons, Ltd., West Sussex, England, 2000).
[4] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, and S. T. Elbert, Chem. Phys. **71**, 41 (1982).
[5] H.-J. Werner and P. J. Knowles, J. Chem. Phys. **82**, 5053 (1985).
[6] P. J. Knowles and H.-J. Werner, Chem. Phys. Lett. **115**, 259 (1985).
[7] B. O. Roos, Adv. Chem. Phys. **69**, 399 (1987).
[8] J. Olsen, B. O. Roos, P. Jørgensen, and H. J. A. Jensen, J. Chem. Phys. **89**, 2185 (1988).
[9] P.-Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer, and L. Gagliardi, J. Chem. Phys. **128**, 204109 (2008).
[10] D. Ghosh, J. Hachmann, T. Yanai, and G. K.-L. Chan, J. Chem. Phys. **128**, 144117 (2008).
[11] D. Zgid and M. Nooijen, J. Chem. Phys. **128**, 144116 (2008).
[12] G. Li Manni, D. Ma, F. Aquilante, J. Olsen, and L. Gagliardi, J. Chem. Theory Comput. **9**, 3375 (2013).
[13] R. E. Thomas, Q. Sun, A. Alavi, and G. H. Booth, J. Chem. Theory Comput. **11**, 5316 (2015).
[14] G. Li Manni, S. D. Smart, and A. Alavi, J. Chem. Theory Comput. **12**, 1245 (2016).
[15] J. E. T. Smith, B. Mussard, A. A. Holmes, and S. Sharma, J. Chem. Theory Comput. **13**, 5468 (2017).
[16] S. R. Langhoff and E. R. Davidson, Int. J. Quantum Chem. **8**, 61 (1974).
[17] H.-J. Werner and E. A. Reinsch, J. Chem. Phys. **76**, 3144 (1982).
[18] K. Andersson, P.-Å. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).
[19] K. Andersson, P.-Å. Malmqvist, and B. O. Roos, J. Chem. Phys. **96**, 1218 (1992).
[20] R. J. Gdanitz and R. Ahlrichs, Chem. Phys. Lett. **143**, 413 (1988).
[21] H.-J. Werner and P. J. Knowles, Theor. Chim. Acta **78**, 175 (1990).
[22] P. G. Szalay and R. J. Bartlett, Chem. Phys. Lett. **214**, 481 (1993).
[23] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu, J. Chem. Phys. **114**, 10252 (2001).
[24] C. Angeli, R. Cimiraglia, and J.-P. Malrieu, J. Chem. Phys. **117**, 9138 (2002).
[25] C. Angeli, M. Pastore, and R. Cimiraglia, Theor. Chem. Acc. **117**, 743 (2007).
[26] E. Neuscamman, T. Yanai, and G. K.-L. Chan, Int. Rev. Phys. Chem. **29**, 231 (2010).
[27] G. Li Manni, R. K. Carlson, S. Luo, D. Ma, J. Olsen, D. G. Truhlar, and L. Gagliardi, J. Chem. Theory Comput. **10**, 3669 (2014).
[28] S. Sharma and A. Alavi, J. Chem. Phys. **143**, 102815 (2015).
[29] M. K. MacLeod and T. Shiozaki, J. Chem. Phys. **142**, 051103 (2015).
[30] G. Gidofalvi and D. A. Mazziotti, J. Chem. Phys. **129**, 134108 (2008).
[31] A. Dreuw and M. Head-Gordon, Chem. Rev. **105**, 4009 (2005).
[32] X. Liu, S. Fatehi, Y. Shao, B. S. Veldkamp, and J. E. Subotnik, J. Chem. Phys. **136**, 161101 (2012).
[33] X. Liu, Q. Ou, E. Alguire, and J. E. Subotnik, J. Chem. Phys. **138**, 221105 (2013).
[34] X. Liu and J. E. Subotnik, J. Chem. Theory Comput. **10**, 1004 (2014).
[35] B. S. Veldkamp, X. Liu, M. R. Wasielewski, J. E. Subotnik, and M. A. Ratner, J. Phys. Chem. A **119**, 253 (2015).
[36] L. Zhao and E. Neuscamman, J. Chem. Theory Comput. **12**, 3436 (2016).
[37] P. J. Robinson, S. D. Pineda Flores, and E. Neuscamman, J. Chem. Phys. **147**, 164114 (2017).
[38] J. A. R. Shea and E. Neuscamman, J. Chem. Theory Comput. **13**, 6078 (2017).
[39] H.-Z. Ye, M. Welborn, N. D. Ricke, and T. Van Voorhis, J. Chem. Phys. **147**, 214104 (2017).
[40] J. H. Choi, C. F. Lebeda, and R. P. Messmer, Chem. Phys. Lett. **5**, 503 (1970).
[41] N. S. Blunt, S. D. Smart, G. H. Booth, and A. Alavi, J. Chem. Theory Comput. **143**, 134117 (2015).
[42] A. Griewank and A. Walther, *Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation*, 2nd ed. (Society for Industrial and Applied Mathematics, Philadelphia, 2008).
[43] S. Sorella and L. Capriotti, J. Chem. Phys. **133**, 234111 (2010).
[44] E. Neuscamman, J. Chem. Phys. **139**, 194105 (2013).
[45] T. Tamayo-Mendoza, C. Kreisbeck, R. Lindh, and A. Aspuru-Guzik, ACS Cent. Sci. **4**, 559 (2018).
[46] See https://www.tensorflow.org for information about the TensorFlow software package; accessed 2 June 2018.
[47] J. Nocedal, Math. Comput. **35**, 773 (1980).
[48] A. T. B. Gilbert, N. A. Besley, and P. M. W. Gill, J. Phys. Chem. A **112**, 13164 (2008).
[49] A. I. Krylov, Annu. Rev. Phys. Chem. **59**, 433 (2008).
[50] M. Schütz and H.-J. Werner, J. Chem. Phys. **114**, 661 (2001).
[51] J. E. Subotnik, A. Sodt, and M. Head-Gordon, J. Chem. Phys. **125**, 074116 (2006).
[52] J. Yang, G. K.-L. Chan, F. R. Manby, M. Schütz, and H.-J. Werner, J. Chem. Phys. **136**, 144105 (2012).
[53] C. Riplinger and F. Neese, J. Chem. Phys. **138**, 034106 (2013).
[54] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 242 (2012).
[55] Y. Shao *et al.*, Mol. Phys. **113**, 184 (2015).
[56] M. Head-Gordon, R. J. Rico, M. Oumi, and T. J. Lee, Chem. Phys. Lett. **219**, 21 (1994).
[57] Y. M. Rhee and M. Head-Gordon, J. Phys. Chem. A **111**, 5314 (2007).
[58] C. Li, P. Verma, K. P. Hannon, and F. A. Evangelista, J. Chem. Phys. **147**, 074107 (2017).
[59] S. Hirata, M. Nooijen, and R. J. Bartlett, Chem. Phys. Lett. **326**, 255 (2000).