Title
Excited state mean-field theory without automatic differentiation

Permalink
https://escholarship.org/uc/item/4b35806b

Journal
The Journal of Chemical Physics, 152(20)

ISSN
0021-9606

Authors
Zhao, Luning
Neuscamman, Eric

Publication Date
2020-05-29

DOI
10.1063/5.0003438

Peer reviewed

Excited State Mean-Field Theory without Automatic Differentiation

Luning Zhao[1] and Eric Neuscamman[2, 3, a)]

[1)]*Department of Chemistry, University of Washington, Seattle, Washington 98195, USA*
[2)]*Department of Chemistry, University of California, Berkeley, California 94720, USA*
[3)]*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720,*
USA

(Dated: 2 April 2020)

We present a formulation of excited state mean-field theory in which the derivatives with respect to the wave function parameters needed for wave function optimization (not to be confused with nuclear derivatives) are expressed analytically in terms of a collection of Fock-like matrices. By avoiding the use of automatic differentiation and grouping Fock builds together, we find that the number of times we must access the memory-intensive two-electron integrals can be greatly reduced. Furthermore, the new formulation allows the theory to exploit existing strategies for efficient Fock matrix construction. We demonstrate this advantage explicitly via the shell-pair screening strategy, with which we achieve a cubic overall cost scaling. Using this more efficient implementation, we also examine the theory's ability to predict charge redistribution during charge transfer excitations. Using coupled cluster as a benchmark, we find that by capturing orbital relaxation effects and avoiding self-interaction errors, excited state mean field theory out-performs other low-cost methods when predicting the charge density changes of charge transfer excitations.

I. INTRODUCTION

Whether one is talking about light harvesting,[1,2] photocatalysts,[3,4] core spectroscopy,[5–7] metal-to-ligand charge transfer,[8,9] or non-adiabatic dynamics,[10,11] reliable predictions of excited state properties are extremely valuable. However, the leading theoretical methods for modeling excited states are fundamentally more approximate than their ground state counterparts due to their reliance on additional approximations. Linear response (LR) methods, for example, assume that the excited state is in some way close to the ground state in state space, leading in practice to a situation in which crucial orbital relaxation effects[12–15] are simply absent in configuration interaction singles (CIS)[16] and time-dependent density functional theory (TDDFT)[16–18] and only treated in a limited manner in singles and doubles equation-of-motion Coupled Cluster (EOM-CCSD) theory.[19] These shortcomings contribute to the difficulties that CIS and TDDFT have with charge transfer (CT) states[13,15] and to the eV-sized errors that EOM-CCSD often makes for doubly excited states.[20,21] As these difficulties arise in part from a lack of orbital relaxation following an excitation, the development of excited state methods that have fully-relaxed, excited-state-specific orbitals is strongly desirable.

Towards this end, the recently-introduced excited state mean-field (ESMF) theory[22] attempts to provide a minimally-correlated reference state for single excitations in which the orbitals are fully relaxed. In some aspects, ESMF is similar to the ΔSCF approach[23–26] and the restricted open-shell Kohn Sham (ROKS) approach,[27,28] but in contrast to these methods it delivers a completely spin-pure wave function by construction and can handle states in which multiple singly-excited configurations are present in a superposition. ESMF is also similar to the restricted active space self-consistent field (RASSCF) method[29,30] since both methods optimize orbitals

for a set of selected configurations. However, these two approaches have important differences. While RASSCF typically approaches excited states via state-averaging, ESMF is a wholly excited-state-specific method. In addition, while ESMF avoids the need to define an active space, this simplicity comes at a price: it is not appropriate for strongly correlated settings such as conical intersections. Like Hartree-Fock (HF) theory,[31,32] the energetic accuracy of ESMF itself is limited due to missing correlation effects, which previous work has shown to cause a bias towards underestimating excitation energies.[33] Nonetheless, ESMF acts as a powerful starting point for perturbation theory[33] and can be used to construct a novel form of density functional theory.[15] The perturbation theory results are particularly exciting, with preliminary testing showing an accuracy competitive with EOM-CCSD.[33]

Like σ-SCF[34,35] and some quantum Monte Carlo approaches to excited states,[21,36–38] ESMF takes a variational approach to orbital relaxation in which it minimizes a function whose global minimum is the desired excited state. However, unlike these approaches, the generalized variational principle (GVP) that ESMF minimizes is defined in terms of the norm of the energy gradient with respect to wave function parameters such as CI coefficients and orbital rotation matrices,[33] and so the optimization requires some information about second derivatives of the energy with respect to the wave function variables. On the bright side, automatic differentiation (AD) via TensorFlow[39] or any other general-purpose AD framework can be employed to evaluate the necessary derivatives while maintaining the same Fock-matrix-build cost-scaling of an ESMF energy evaluation. On a less positive note, this approach requires the memory-intensive two-electron integrals (TEIs) to be accessed many times during each evaluation of the GVP's gradient and makes it difficult to take advantage of acceleration techniques like shell-pair screening,[40] density fitting,[41] and tensor hyper contraction.[42–44] Thus, while AD has been extremely helpful in quickly developing an initial implementation for testing ESMF theory (see Appendix for examples of cumbersome expressions that it allows one to avoid), it severely limits the practical efficiency of the ap-

[a)]Electronic mail: eneuscamman@berkeley.edu.

proach. In this study, we develop explicit analytic expressions for the objective function's gradient with respect to wave function variables that simplify into a collection of nine Fock-like matrix builds that can be carried out together (thus minimizing the number of times the TEIs need to be accessed) and which can benefit straightforwardly from acceleration methods. The result is a dramatic speedup compared to our previous TensorFlow implementation, thus allowing ESMF theory to be used in significantly larger systems. Now, to avoid confusion between the analytic gradients we are discussing with other types of analytic gradients, let us be very clear: there are no nuclear gradients in this study. All of the gradients we investigate are with respect to wave function variables such as configuration coefficients or orbital rotation parameters.

We will begin with a brief review of ESMF theory, including the wave function ansatz and the GVP employed in its optimization. We will then discuss why the TensorFlow implementation of the energy derivatives is inefficient, after which we derive the analytic energy derivative expressions and show how they can be formulated using Fock-builds. Following these theoretical developments, we test the efficiency in two charge-transfer systems and demonstrate a speedup of two orders of magnitude relative to our previous implementation. We then turn to an investigation of charge transfer in a minimally-solvated environment that would not have been possible with the previous AD-based approach to evaluating the GVP derivatives. Encouraged by the results, we round out our results section with additional testing of charge density changes in four other charge transfer examples. Finally, we conclude with a summary and some comments on future directions.

II. THEORY

A. ESMF Method

In ESMF, the wave function ansatz is written as

$$|\Psi\rangle = e^{\hat{X}}\left(c_0|\Phi\rangle + \sum_{ia}\sigma_{ia}\hat{a}^\dagger_{a\alpha}\hat{a}_{i\alpha}|\Phi\rangle + \tau_{ia}\hat{a}^\dagger_{a\beta}\hat{a}_{i\beta}|\Phi\rangle\right) \quad (1)$$

in which $|\Phi\rangle$ is the restricted Hartree-Fock determinant, and the coefficients σ_{ia} and τ_{ia} correspond to excitations of an alpha-spin and a beta-spin electron, from the ith occupied orbital to the ath virtual orbital. The operator $\hat{X}$ is defined as,

$$\hat{X} = \sum_{pq}X_{pq}\hat{a}^\dagger_p\hat{a}_q \quad (2)$$

in which X is an anti-symmetric matrix so that the orbital rotation operator $\hat{U} = \exp(-\hat{X})$ is unitary. Note that this orbital rotation will be state specific, which means that different states' wave functions will not be strictly orthogonal. While this trait has not prevented ESMF from acting as a powerful reference for perturbation theory[33] or from accurately predicting charge density changes (see Section III D), it can cause some concern, and so we have for the sake of completeness included the overlaps between ground and excited states for

the systems we study in the Supplemental Materials. Here, we simply note that these overlaps are quite small.

Although there is an argument to be made that a single open-shell configuration state function should be seen as the minimally correlated reference function for excited states, we choose to include all singly-excited configurations in order to handle states in which two or more of these configurations exist in a superposition, as for example occurs in the low-energy spectrum of N_2. Although such states are technically multi-configurational, they certainly do not require a general-purpose strongly correlated ansatz and indeed are already treated at a qualitatively correct level by CIS. That said, the open shell character of ESMF (in which two electrons correlate their positions so as to not reside in the same spatial orbital simultaneously) does involve more correlation than HF theory and so it is a step further away from a true mean-field theory in which no correlation is present at all.

In the initial development of ESMF theory,[22] the ansatz was optimized using the Lagrangian-based objective function,

$$L_{\vec{\lambda}} = W + \vec{\lambda}\cdot\nabla E \quad (3)$$

in which W was seen as an approximated excited state variational principle whose purpose is to guide the optimization to the energy stationary point associated with a particular excited state and $\vec{\lambda}$ is a set of Lagrange multipliers that ensure that the optimized wave function is indeed an energy stationary point. The approximate variational principle was chosen as

$$W = (\omega - E)^2 \approx \frac{\langle\Psi|(\omega-H)^2|\Psi\rangle}{\langle\Psi|\Psi\rangle} \quad (4)$$

which, although successful in initial testing on small molecules, was found to have multiple shortcomings. First, the target function is not bounded from below with respect to Lagrangian multipliers, making simple quasi-Newton methods difficult to use directly. Instead, the expression $|\nabla L_{\vec{\lambda}}|^2$ was minimized, which further increased the computational cost by necessitating an additional layer of automatic differentiation (still the right cost scaling, but now containing components that formally involve triple derivatives of the energy). Further, a more recent study[33] found that this approach can show poor numerical conditioning, sometimes requiring hundreds of quasi-Newton iterations to converge.

In order to address these two problems, a finite-difference Newton-Raphson (NR) method was developed for an objective function based not on Lagrange multipliers but instead on a GVP.[33]

$$L_{\mu\chi} = \chi\left(\mu(\omega-E)^2 + (1-\mu)|\nabla E|^2\right) + (1-\chi)E \quad (5)$$

Starting with χ set to 1, μ is gradually reduced to zero during the optimization so as to ensure convergence to the stationary point with energy closest to ω. Unlike the target function in Equation 3, this approach is bounded from below, allowing both NR and quasi-Newton methods to be employed without the need for an additional layer of AD. In addition, one can switch χ to 0 close to convergence and rely on stationary-point methods like NR, thus potentially benefiting from even

more efficient gradients and such method's super-linear convergence. In practice, initial testing has shown this approach to be more robust and more efficient than Equation 3 in a variety of systems.[33]

B. Analytic Derivatives

Even though this new approach improved the optimization's efficiency, its implementation still relied on AD for the derivatives of both the energy and $L_{\mu\chi}$, which, although convenient, leads to an unnecessarily high prefactor in the method's cost due to frequent access of the TEIs and the handling of the TEIs as a dense 4-index array without the efficiencies that accelerated Fock-build methods enjoy. In order to address this source of inefficiency, we will now derive explicit expressions for the analytic derivatives of the ESMF energy and objective function and show that they can be formulated as a set of Fock matrix builds. We will focus on the special case of singlet excited states, whose wave function can be written as

$$|\Psi\rangle = e^{\hat{X}} \left(c_0 |\Phi\rangle + \sum_{ia} \sigma_{ia} \left(\hat{a}_{a\alpha}^{\dagger} \hat{a}_{i\alpha} + \hat{a}_{a\beta}^{\dagger} \hat{a}_{i\beta} \right) |\Phi\rangle \right)$$

$$= e^{\hat{X}} \left(c_0 |\Phi\rangle + |\Psi_{\mathrm{CIS}}\rangle \right) \tag{6}$$

in which the alpha and beta electron excitations have the same coefficients and we have grouped the linear combination of singly excited determinants to the CIS ($|\Psi_{\mathrm{CIS}}\rangle$) wave function. Although the derivation will be based on singlet excited states, a generalization to the triplet case is straightforward.

1. Notation

Before we derive ESMF energy, target function, and derivatives, we introduce the notations used in our derivations. Orbital index i, j, k denote occupied orbitals. Index a, b, c denote virtual orbitals. We use the indices p, q, r, s for general orbitals. We denote the first N_o rows of the $\mathbf{V}$ matrix as Θ, where N_o is the total number of occupied orbitals. The last N_v rows of $\mathbf{V}$ are denoted as the matrix Γ, where N_v is the number of virtual orbitals. Similarly, we denote the corresponding blocks of the matrix $\mathbf{M}$ (see Table I) as $\mathbf{R}$ and Φ.

We define the "generalized" Coulomb, exchange, and Fock matrices as,

$$J[\mathbf{D}]_{pq} = \sum_{rs} D_{rs}(rs|pq)$$

$$K[\mathbf{D}]_{pq} = \sum_{rs} D_{rs}(pr|qs) \tag{7}$$

$$F[\mathbf{D}]_{pq} = 2J[D]_{pq} - K[D]_{pq}$$

in which $(rs|pq)$ are the two-electron integrals in the atomic orbital basis in 1122 order, and $\mathbf{D}$ is a "generalized" density matrix which is not necessarily symmetric. For a summary of the notation for the scalar and matrix quantities we use, as well as for some matrix operations, see Table I.

TABLE I: Summary of Notation

Description	Notation
One-electron Integrals in AO Basis	$\mathbf{G}$
Two-electron Integrals in AO Basis	$(pq\|rs)$
RHF Orbital Coefficients	$\mathbf{C}$
RHF Determinant Coefficient	c_0
Excited Determinant Coefficients	σ
ESMF Orbital Coefficients	$\mathbf{V} = \mathbf{U}^T\mathbf{C}$
c_0 Lagrangian Multiplier	μ_0
σ Lagrangian Multiplier	μ
X Lagrangian Multiplier	$\mathbf{M}$
Transformed U Lagrangian Multiplier	$\mathbf{W} = \mathbf{M}^T\mathbf{C}$
First N_o Rows of $\mathbf{V}$	Θ
Last N_v Rows of $\mathbf{V}$	Γ
First N_o Rows of $\mathbf{W}$	$\mathbf{R}$
Last N_v Rows of $\mathbf{W}$	Φ
General Fock Matrix	$\mathbf{F}[\mathbf{D}]$
Wave Function Square Norm	$N_2 = c_0^2 + 2\sum_{ia} c_{ia}^2$
A Matrix	$\Gamma^T\sigma^T\sigma\Gamma - \Theta^T\sigma\sigma^T\Theta$
B Matrix	$\Theta^T\sigma^T\sigma\Gamma - \mathbf{R}^T\sigma\sigma^T\Theta$
Matrix Trace	$\mathrm{Tr}[\mathbf{O}] = \sum_p O_{pp}$
Matrix Inner Product	$\mathbf{O} \cdot \mathbf{V} = \sum_{pq} O_{pq} V_{pq}$

2. ESMF Energy

The ESMF energy is computed as

$$E = \left(c_0^2 \left\langle \Phi | e^{-\hat{X}} \hat{H} e^{\hat{X}} | \Phi \right\rangle + 2c_0 \left\langle \Phi | e^{-\hat{X}} \hat{H} e^{\hat{X}} | \Psi_{\mathrm{CIS}} \right\rangle \right.$$
$$\left. + \left\langle \Psi_{\mathrm{CIS}} | e^{-\hat{X}} \hat{H} e^{\hat{X}} | \Psi_{\mathrm{CIS}} \right\rangle \right) / N_2 \tag{8}$$

By splitting the Hamiltonian into one- and two-body operators, the energy can be formulated as

$$E = \frac{E_1 + E_2}{N_2} \tag{9}$$

The ESMF energy can now be evaluated in the same formalism as CIS energy with rotated molecular orbital coefficients as,

$$\mathbf{V} = \mathbf{U}^T\mathbf{C} \tag{10}$$

in which the one- and two-body components are

$$E_1 = 2N_2\mathrm{Tr}[\Theta\mathbf{G}\Theta^T] + 4c_0\mathrm{Tr}[\Theta\mathbf{G}\Gamma^T\sigma^T]$$
$$+ 2\mathrm{Tr}[\sigma\Gamma\mathbf{G}\Gamma^T\sigma^T - \Theta\mathbf{G}\Theta^T\sigma\sigma^T] \tag{11}$$

and

$$E_2 = N_2\mathbf{F}[\Theta^T\Theta] \cdot (\Theta^T\Theta) + 4c_0\mathbf{F}[\Theta^T\Theta] \cdot (\Theta^T\sigma\Gamma)$$
$$+ 2\mathbf{F}[\Theta^T\Theta] \cdot \mathbf{A} + 2\mathbf{F}[\Theta^T\sigma\Gamma] \cdot (\Theta^T\Theta) \tag{12}$$

respectively. The first terms in the one- and two-body energy expression come from the contribution of the un-excited determinant with rotated orbitals. The second terms are due to the cross terms between excited and un-excited determinants, which are zero in CIS with canonical orbitals due to Brillouin

theorem. However, these terms are non-zero in ESMF with rotated canonical orbitals. The last two terms arise from the contributions of purely excited determinants.

These expressions reveal that the most expensive part of the ESMF energy is the construction of two Fock matrices: $\mathbf{F}[\Theta^T\Theta]$ and $\mathbf{F}[\Theta^T\sigma\Gamma]$. Note that this is different from HF theory, in which only one Fock matrix, $\mathbf{F}[\Theta^T\Theta]$, is needed to evaluate the energy. In fact, we note that if we set $\sigma = \mathbf{0}$ and $\mathbf{U} = \mathbf{I}$ then the ESMF energy becomes,

$$E = 2Tr[\mathbf{PG}] + F[\mathbf{P}]\cdot\mathbf{P} \tag{13}$$

in which $P_{rs} = \sum_i C_{ri}C_{si}$. one would immediately realize that this is the HF energy expression.

3. Derivatives of Lagrangian-based Objective Function

Starting from the energy, we have derived the first derivatives $\partial E/\partial c_0$, $\partial E/\partial\sigma$, $\partial E/\partial\Theta$, and $\partial E/\partial\Gamma$, detailed expressions for which can be found in the Appendix. Note that these derivatives require the construction of one additional Fock matrix, $\mathbf{F}[\mathbf{A}]$, in addition to the two required for the energy itself. In this subsection, we use these components to find the analytic derivatives of the Lagrangian-based objective function of Equation 3. Once we have these derivatives in hand, we will see in the next subsection how they can be modified to produce analytic derivatives of the newer GVP-based objective function of Equation 5. Note that, as the derivation of the different variables' derivatives is very similar, we will often work in terms of derivatives with respect to a generic wave function variable x, which could be either a configuration coefficient from Equation 1 or an element of the X matrix from Equation 2.

We first expand the Lagrange multiplier dot product as

$$L_{\vec{\lambda}} = W + \mu_0\frac{\partial E}{\partial c_0} + \sum_{ia}\mu_{ia}\frac{\partial E}{\partial\sigma_{ia}} + \sum_{pq}M_{pq}\frac{\partial E}{\partial X_{pq}} \tag{14}$$

and note that the derivative of W with respect to any variable x is simple once the energy first derivatives have been evaluated.

$$\frac{\partial W}{\partial x} = -2(\omega - E)\frac{\partial E}{\partial x} \tag{15}$$

Noting that the remaining part of L can be written in terms of the auxiliary quantities L_1, L_2, and L_3 (defined in the Appendix), the objective function becomes

$$L = W + (\mu_0 L_1 - 2\mu_0 E c_0)/N_2$$
$$+ \left(L_2 - 4E\sum_{ia}\sigma_{ia}\mu_{ia}\right)/N_2 + L_3/N_2. \tag{16}$$

Inspecting the definition of L_3, we find that this approach requires two additional Fock builds: $\mathbf{F}[\mathbf{R}^T\Theta]$ and $\mathbf{F}[\mathbf{R}^T\sigma\Gamma]$, bringing us to five total Fock builds for this approach to constructing the energy, its derivatives, and L.

Although we again relegate the details to the Appendix, we find that evaluating the derivatives of L_1, L_2, and L_3 with respect to c_0, σ, Θ, and Γ requires another four Fock builds:

$\mathbf{F}[\Theta^T\mu\Gamma]$, $\mathbf{F}[\Theta^T\sigma\Phi]$, $\mathbf{F}[\Theta^T\mu\sigma^T\Theta]$, and $\mathbf{F}[\mathbf{B}]$. Thus, after constructing a total of nine Fock-like matrices, the evaluation of the derivatives of L with respect to all of the ESMF wave function variables amounts to an inexpensive (relative to the Fock builds) collection of operations on matrices whose dimensions are no worse than the number of orbitals.

We have so far derived derivatives with respect to the elements of $\mathbf{V}$, but the actual orbital rotation parameters are the elements of $\mathbf{X}$. To get all the way to derivatives with respect to $\mathbf{X}$, we first transform the derivatives with respect to $\mathbf{V}$ to the derivatives with respect to $\mathbf{U}$ through the following way,

$$\frac{\partial E}{\partial\mathbf{U}} = \mathbf{C}\frac{\partial E}{\partial\mathbf{V}^T} \tag{17}$$

In order to get derivatives with respect to $\mathbf{X}$, there are at least two possible routes. On the one hand, AD (e.g. via TensorFlow) can be used to complete the last step in a reverse-accumulation approach, in which the values for $\partial E/\partial U$ that we have calculated via our analytic expression are fed in to reverse accumulation through the matrix exponential function. On the other hand, if we wish to strictly avoid AD (although for this part of the evaluation there is not such a clear efficiency case for doing so) we can instead redefine our orbital rotation matrix as

$$U \to \tilde{U}\exp(-X) \tag{18}$$

where after each optimization step we reset X to zero by absorbing its rotation into $\tilde{U}$ and then exploiting the simple relationship $\exp(X) = 1 + X$. Either way, once we have evaluated the energy derivatives with respect to U, converting them to derivatives with respect to X is inexpensive compared to the Fock builds.

4. Derivatives of GVP-based Objective Function

We now turn to the derivatives of the GVP-based objective function in Equation 5 with respect to some wave function variable x.

$$\frac{\partial L_{\mu\chi}}{\partial x} = \chi\left(-2\mu(\omega - E)\frac{\partial E}{\partial x} + 2(1-\mu)\sum_y\frac{\partial E}{\partial y}\frac{\partial^2 E}{\partial x\partial y}\right)$$
$$+ (1-\chi)\frac{\partial E}{\partial x} \tag{19}$$

Now, compare this expression to the derivatives of $L_{\vec{\lambda}}$.

$$\frac{\partial L_{\vec{\lambda}}}{\partial x} = -2(\omega - W)\frac{\partial E}{\partial x} + \sum_y\lambda_y\frac{\partial^2 E}{\partial x\partial y} \tag{20}$$

If we first evaluate the energy derivatives, which requires three Fock builds, we can replace the Lagrange multipliers in Eq. (20) with these energy derivatives, at which point the approach discussed above for evaluating the derivatives of $L_{\vec{\lambda}}$ can be used for evaluating the term inside the large parentheses in Eq.

(19). Thus, as for the Lagrange multiplier objective function, the GVP objective function's derivatives can be evaluated at the cost of nine Fock builds (the three already done for the energy derivatives plus the remaining six).

C. Shell-Pair Screening

In total, our analytic evaluation of the derivatives needed for ESMF optimization requires building nine Fock matrices.

1. $\mathbf{F}[\Theta^T \Theta]$ 6. $\mathbf{F}[\Theta^T \mu \Gamma]$
2. $\mathbf{F}[\Theta^T \sigma \Gamma]$ 7. $\mathbf{F}[\Theta^T \sigma \Phi]$
3. $\mathbf{F}[\mathbf{A}]$ 8. $\mathbf{F}[\Theta^T \mu \sigma^T \Theta]$
4. $\mathbf{F}[\mathbf{R}^T \Theta]$ 9. $\mathbf{F}[\mathbf{B}]$
5. $\mathbf{F}[\mathbf{R}^T \sigma \Gamma]$

Although this ensures that ESMF has the same asymptotic cost scaling as HF theory, it is clearly going to suffer from a higher prefactor. How to mitigate this extra cost? In this study, we employ the shell-pair screening approach to Fock matrix construction, both because it exploits sparsity in the TEIs in larger systems and because it allows us to easily group Fock builds together in order to minimize the number of times the sparse TEIs must be accessed. The shell-pair screening algorithm can be summarized as follows. At the beginning of an ESMF optimization we loop over all four indices of the two-electron integrals (p,q,r,s), and for each index pair (p,q), find the (r,s) index pair that maximizes the absolute value of integral $|(pq|rs)|$. The shell pair (p,q) is discarded unless $\max_{r,s}|(pq|rs)|$ is above a certain threshold, which in this work we set to 10^{-9} a.u., small enough so that we achieve direct agreement with results from our older TensorFlow-based implementation. After this screening, Fock builds proceed according to Eq. (7) by looping over the retained shell pairs. Crucially, multiple Fock matrices can be constructed during a single loop through the TEIs. As the TEIs still take up a lot of memory even after screening, it is more cache efficient to evaluate multiple fock matrices at once. Thus, when using the Lagrange multiplier objective function, we evaluate all nine Fock matrices during a single loop over the shell pairs. For the GVP objective function derivatives, we require two loops through the shell pairs, the first to evaluate the first three matrices and the second to evaluate the other six, which in that case depend on the first three due to the Lagrange multiplier values having been set equal to the energy derivatives. Finally, note that we have implemented a simple approach to shared-memory parallelism by threading the loop over shell pairs.

III. RESULTS

A. Computational Details

All of the ESMF results are obtained via our own software, which extracts one- and two-electron integrals from PySCF.[45]

The DFT, TDDFT, CIS, ROHF, and CCSD results were obtained from QChem.[46] We use the VESTA[47] software to plot the density difference between ground and excited state. The molecular geometries can be found in the Supplemental Materials.
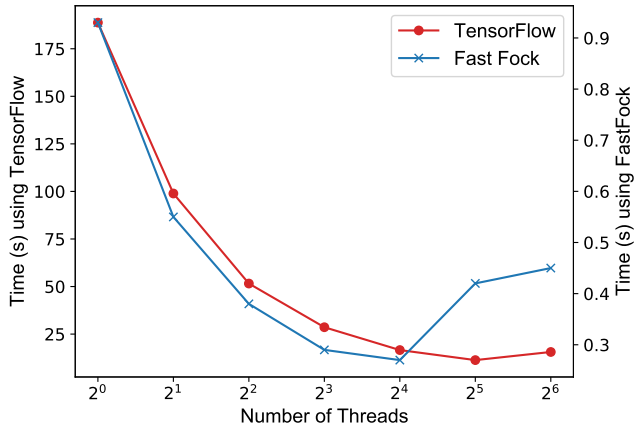
FIG. 1: Time taken to evaluate the 24 Lagrangian gradients required for inverting the Hessian during the first finite-difference NR iteration of the Cl^--H_2O optimization.
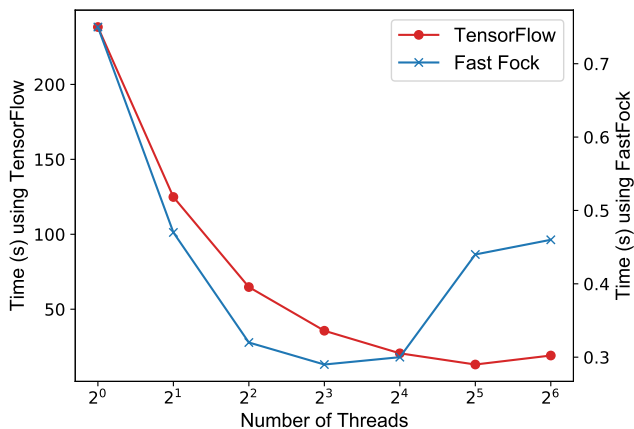
FIG. 2: Time taken to evaluate the 22 Lagrangian gradients required for inverting the Hessian during the first finite-difference NR iteration of the NH_3-F_2 optimization.

B. Efficiency Gain and Cost Scaling

Here we compare the cost of two example ESMF optimizations when performed with our group's previous AD-based implementation (denoted here as TensorFlow) with the new implementation based on analytic expressions and shell-pair screening (denoted here as FastFock). Working in the cc-pVDZ basis,[48] we optimize the lowest singlet charged transfer

excited states of the Cl^--H_2O dimer and the NH_3-F_2 dimer, both of which have been studied before.[15,49,50] The optimization uses the finite-difference NR approach,[33] in which the NR linear equation is solved via the generalized minimal residual method (GMRES), with the Hessian-vector multiplication evaluated via a simple finite difference formula involving gradients of the objective function. In Figures 1 and 2, we compare the amount of time that the two implementations took to solve the linear equation for each optimization's first NR step. Although we would clearly need to work with larger systems for our simple multi-threading approach to saturate the 32 cores on the processor, we see a roughly 100-fold increase in speed in both cases when running 16 threads (be careful to note the different left and right axes). When run in serial, the new implementation is 202 and 314 times faster than the TensorFlow implementation for the two different systems.

We now illustrate the origin of the efficiency of our new ESMF implementation by comparing the performance of the 5 different ESMF implementation options described in Table II. These five implementations differ in programming language/framework, how they conduct differentiation, and whether shell-pair screening is performed. Option 1 is our old TensorFlow ESMF implementation and option 5 is our new FastFock implementation. Option 2-4 connect option 1 with option 5 by changing one variable at a time. We compare the amount of time each implementation took to evaluate the same 22 Lagrangian gradients in Figure 2.

TABLE II: Comparison of how long five different ESMF implementations took to evaluate the Lagrangian gradients in Figure 2. The "Fock Build" column indicates which language the bottleneck Fock-build step is implemented in.

Option	Differentiation	Fock Build	Shell-Pair Screening	Time (s)
1	Automatic	TensorFlow	No	35.64
2	Analytic	TensorFlow	No	3.41
3	Analytic	NumPy	No	3.69
4	Analytic	C	No	0.33
5	Analytic	C	Yes	0.35

As shown in the Table, by merely changing automatic differentiation to analytic differentiation but maintaining the TensorFlow framework, we already see a 10-fold speed-up in terms of the gradient evaluation. Such an acceleration shows the benefits of deriving analytic derivatives and collecting all the Fock-like matrices, since doing so allows us to drastically reduce the number of times we must access the memory-intensive two-electron integrals. Another 10-fold speed-up is achieved by moving the Fock build implementation from Python to C. In the final row, we see that this system is too small to benefit from shell-pair screening, and so we now turn our attention to larger cases in which this benefit can be realized.

In order to show the cost scaling of our method, in Figure 3 we plot the time taken to evaluate one Lagrangian gradient as a function of the number of orbitals in a NH_3-$(H_2O)_n$ system. The structure is obtained by first putting water molecules around NH_3 randomly and then performing geome-

TABLE III: Excitation energies (eV) for the TensorFlow and FastFock implementations, demonstrating that the shell-pair screening within the latter has not affected the energetics.

System	TensorFlow	Fast Fock
Cl^--H_2O	4.7195	4.7195
NH_3-F_2	4.5367	4.5367

try optimizations in a 6-31G* basis with MP2. As more water molecules are added, we find that the implementation's cost scaling is roughly $O(N^3)$. This is expected, as although the shell-pair screening can greatly reduce the cost of the Fock builds, the remaining matrix-matrix multiplications involved in forming our derivatives should give an N^3 scaling.

Finally, in order to verify that the gain we obtained in efficiency does not come with a loss in accuracy from the shell-pair screening, we show the predicted excitation energies of Cl^--H_2O and NH_3-F_2 in Table III. We found that the predictions match to better than 10^{-4}eV between the two implementations, indicating that there was no meaningful accuracy loss at the screening threshold we employed.

C. Minimally-solvated charge transfer

With these efficiency improvements in hand, we are able to make an initial investigation of ESMF's predictions for how charge transfers in a minimally solvated environment. To this purpose, we study the system shown in Figure 4, in which a Li atom and a F atom are each surrounded by a small collection of water molecules. Although the geometry optimizations used a different basis (see figure), all subsequent calculations were performed in the cc-pVDZ orbital basis. When the two
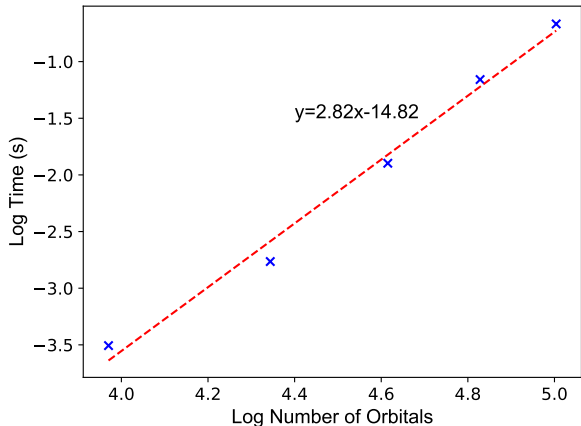

FIG. 3: A log-log plot of the time taken to evaluate one Lagrangian gradient as a function of the number of atomic orbitals for the NH_3-$(H_2O)_n$ system with varying numbers of water molecules. 8 OpenMP threads were used.

TABLE IV: Differences in atomic Mulliken charges for atoms in the lower cluster between the excited (lower cluster neutral) and ground (lower cluster negatively charged) states. For UKS, ROHF, and IP/EA-EOM-CCSD, we report the results for the corresponding charge states of the lower cluster in isolation (see text). DFT methods used the ωB97X functional.

Method	F	O	H	H	O	H	H	O	H	H	O	H	H	O	H	H	O	H	H
CIS	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
TDDFT	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
UKS	0.06	0.04	0.00	0.02	0.11	0.01	0.03	0.10	0.03	-0.01	0.38	0.06	0.05	0.03	-0.01	0.01	0.08	0.01	0.03
ROKS	0.03	0.14	0.02	0.03	0.09	0.00	0.02	0.07	0.03	-0.01	0.32	0.06	0.05	0.04	-0.01	0.01	0.06	0.00	0.02
ROHF	0.01	0.02	-0.01	0.02	0.02	-0.02	0.01	0.02	0.03	-0.03	0.72	0.10	0.09	0.02	-0.01	0.01	0.02	-0.02	0.01
ESMF	0.01	0.03	-0.01	0.02	0.02	-0.02	0.01	0.01	0.03	-0.04	0.71	0.11	0.09	0.03	-0.01	0.00	0.02	-0.02	0.01
CCSD	0.02	0.01	-0.01	0.01	0.01	-0.01	0.00	0.01	0.01	-0.02	0.78	0.09	0.08	0.01	-0.01	0.00	0.01	-0.01	0.00

clusters are simulated as a single system with ground state methods, the F and Li clusters carry negative and positive net charges, respectively. Our purpose here is to investigate the predictions that various excited state methods make for how the charge density changes in the (lowest-energy singlet) charge transfer excitation that returns both clusters to charge neutrality.

As seen in Figure 5, the charge density changes predicted by ESMF differ noticeably from those of CIS and ωB97X-based[52] TDDFT. Although all three methods (and all others tested, see below) agree that the transferred electron comes
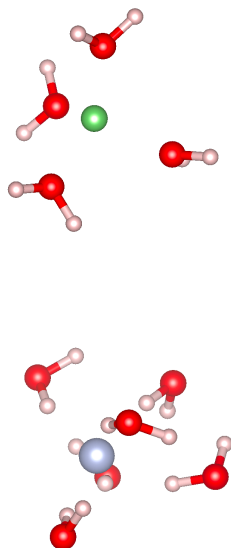

FIG. 4: The geometry of the minimally-solvated charge transfer system, in which a Li atom surrounded by four water molecules can be seen above a F atom surrounded by six water molecules. The geometry was determined by two MP2 optimizations in the 6-31G* basis,[51] one for the positively-charged Li cluster in isolation and one for the negatively-charged F cluster in isolation. The two clusters were then arranged one atop the other such that the Li and F atoms were 8.05$\mathring{A}$ apart. See Supplemental Materials for the atomic coordinates.

from a lone pair on the lower-left water molecule, ESMF also displays two types of significant orbital-relaxation effects that CIS and TDDFT fail to capture. First, electron density in the OH bonds on the affected water molecule is seen to shift towards the oxygen atom, showing a net charge depletion on the hydrogen atoms and some net charge accumulation close to the oxygen atom's center. Second, the other water molecules show a polarization of their electron densities in response to the newly created hole on the lower-left water. As seen in Table IV, these relaxation effects create changes in the Mulliken populations for ESMF that differ significantly from those shown by CIS and TDDFT, both of which predict only significant changes to the lower-left oxygen atom.

At this point, we take advantage of the fact that the two clusters are well separated in order to perform IP/EA-EOM-CCSD[53–55] on the separate clusters in isolation in order to produce a high-level benchmark on what types of charge-relaxation effects other methods should display. As seen in Table IV, the coupled cluster predictions for the Mulliken population changes on both the lower-left water and the other water molecules agree reasonably well with the ESMF predictions while disagreeing with the CIS and TDDFT predictions. In particular, CIS and TDDFT fail to predict the significant relaxation near the lower-left water's hydrogen atoms as well as the degree to which the charge densities on the other water molecules shift in response to the newly created hole. To verify that these errors are due to the additional approximations (namely a lack of secondary orbital relaxation) associated with CIS and TDDFT as excited state linear-response methods, we have also performed ground state restricted open-shell HF (ROHF) and ωB97X-based unrestricted Kohn-Sham

TABLE V: Comparison of excitation energies (eV) for the Li-F system's charge transfer excitation. See text for details of the the CCSD approach.

Method	Excitation Energy
CIS	8.82
TDDFT/ωB97X	6.12
ESMF	5.92
CCSD	6.64

(a) CIS   (b) TDDFT   (c) ESMF
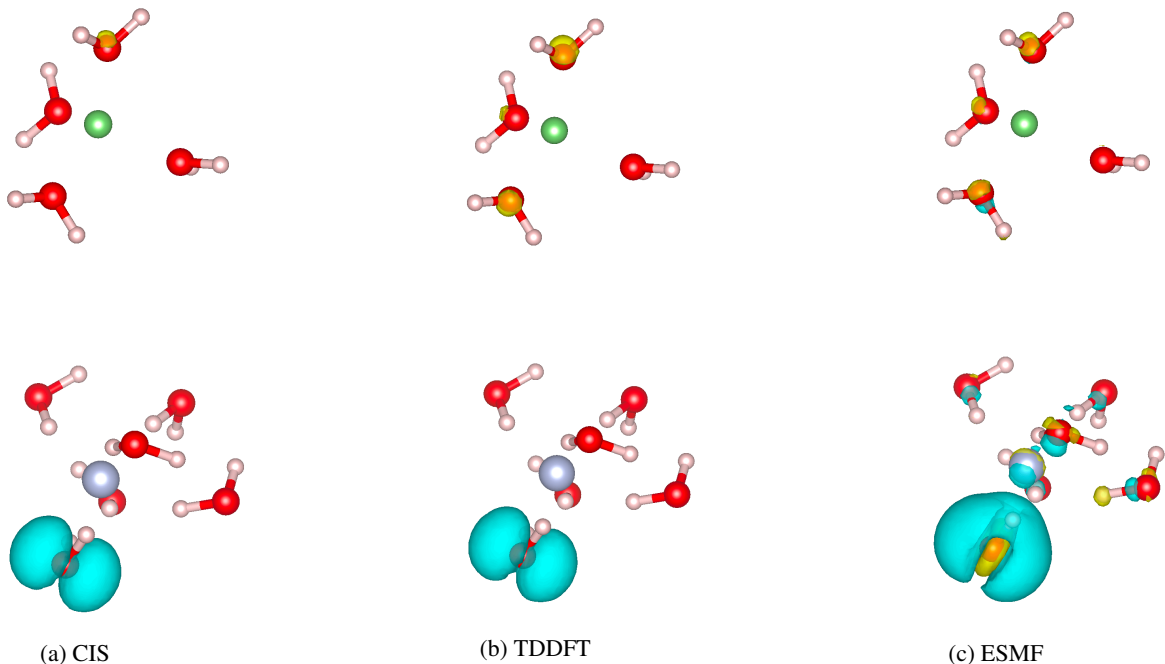
FIG. 5: Isosurface plots for the charge density changes following the charge transfer excitation that moves an electron from the lower F cluster to the upper Li cluster, with blue surfaces showing charge depletion relative to the ground state and yellow surfaces showing charge accumulation. We compare the results using CIS, TDDFT, and ESMF in Figure 5a, 5b, and 5c. TDDFT employed the ωB97X functional. Note that the primary orbital on the Li cluster that accepts the transferred electron is very diffuse, and so at the isosurface value that makes charge-relaxation effects visible in the lower cluster the charge accumulation on the Li cluster is not visible. See the Supplemental Materials for a plot with an isosurface value that makes the Li cluster charge accumulation more clear.

(a) ROHF   (b) UKS   (c) ESMF
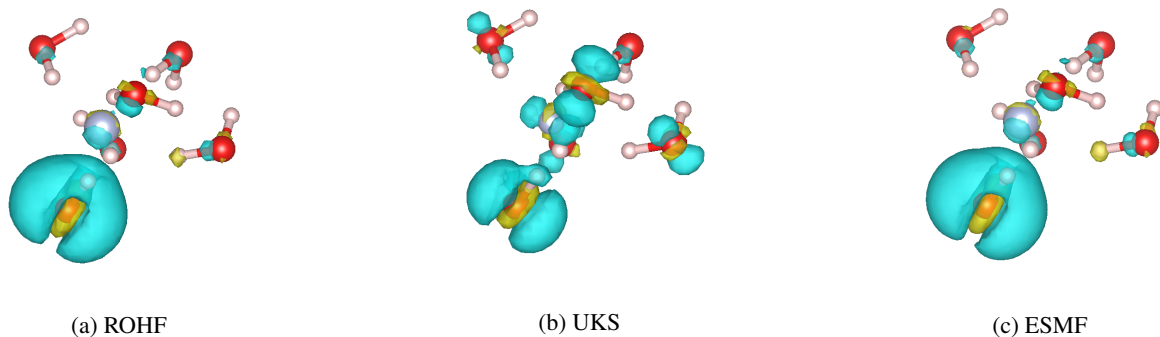
FIG. 6: As for Figure 5, but showing ROHF and UKS ground-state charge density differences between the neutral and negatively charged F cluster when modeled in isolation alongside ESMF's charge density differences when it treats the charge transfer state in the full two-cluster system. We compare the results using ROHF, UKS, and ESMF in Figure 6a, 6b, and 6c. UKS employed the ωB97X functional.

(UKS) calculations on the anionic and neutral lower cluster in isolation. As seen in Figure 6 and Table IV, these methods both display significant relaxation effects upon creation of the hole on the lower-left water, as should be expected of self-consistent field methods upon the removal of an electron. However, the ROHF and UKS predictions are quite different from each other, with the latter incorrectly delocalizing the hole across all of the water molecules in the lower clus-

ter in a clear display of the difficulties posed by DFT's self-interaction error, which is known to over-stabilize delocalized states.[56,57] ROHF, ESMF, and coupled cluster, in contrast, are by construction free of self-interaction errors and keep the hole localized on the lower-left water molecule, with the other water molecules showing much smaller net population changes while still correctly shifting their electron clouds to re-polarize after the creation of the hole. Note especially that

TABLE VI: Differences in atomic Mulliken populations for atoms between the excited and ground states of NaCl. DFT methods used the ωB97X functional.

Method	Na	Cl	total error
CIS	-0.93	0.93	0.41
TDDFT	-0.93	0.93	0.39
ROKS	-0.57	0.57	0.31
ESMF	-0.69	0.69	0.08
EOM-CCSD	-0.73	0.73	N/A

the ROHF and ESMF results are in close agreement, which is not surprising given that the spin-purity offered by ESMF has only a very small energetic effect in this type of long-range charge transfer excitation. The same logic explains why the ROKS results are very similar to those of UKS, displaying the same clear signs of self-interaction-induced over-delocalization.

As a final note in this system, we remind the reader that ESMF, like HF, should not be expected to deliver highly accurate energetics as it neglects weak correlation effects. For the cluster under study, excitation energy estimates for different methods are shown in Table V. Note that due to the large system size, we did not pursue a direct EOM-CCSD calculation in the full system. Instead, we created an estimate of the coupled cluster excitation energy via the expression

$$\Delta E = [\text{EA}]_{\text{upper}} - [\text{IP}]_{\text{lower}} + 1/R \qquad (21)$$

in which $[\text{EA}]_{\text{upper}}$ is the electron affinity of the upper cluster in isolation and $[\text{IP}]_{\text{lower}}$ is the ionization potential of the lower cluster in isolation, as predicted by IP/EA-EOM-CCSD. The $1/R$ term is used to estimate the Coulomb attraction of the two charged clusters in ground state, with R computed as the distance between Li and F. As we have seen in many other systems,[33] ESMF appears to underestimate the excitation energy. To reach quantitative energetics, a correlation correction such as the one provided by ESMP2[33] is clearly in order. While the current formulation and implementation of that theory is too expensive to use in a system of this size, work exploring more affordable formulations is ongoing.

D. More Charge Density Tests

In light of ESMF's success in predicting charge density changes in the previous section, we now turn our attention to how accurately the method predicts charge density changes in four additional charge transfer excitations. We examine charge transfer excitations in the NaCl molecule, the Cl$^-$-(H$_2$O)$_3$ system, the C$_2$H$_4$-C$_2$F$_4$ dimer, and the aminophenol molecule. In these excitations, the charge flow is from Cl to Na, from Cl$^-$ to H$_2$O, from C$_2$H$_4$ to C$_2$F$_4$, and from aminophenol's OH and NH$_2$ groups to its benzene ring. Note that we provide the molecular structures and Cartesian coordinates of these test systems in the Supplemental Materials.

Tables VI, VII, VIII, and IX show how well ESMF, TDDFT, CIS, ROKS, and CIS do at predicting charge density differences in comparison to EOM-CCSD. Note that we use standard EOM-CCSD here on the full systems, as they are small enough for this to be practical, and that in some cases we use ΔSCF instead of ROKS due to QChem's current limitation of ROKS to HOMO/LUMO excitations. As their charge densities should be the same, we use the ROKS and ΔSCF labels interchangeably in this section. We find that ESMF is consistently more accurate in these predictions than either CIS or TDDFT. When comparing ESMF and ROKS, we find that in three systems — Cl$^-$-(H$_2$O)$_3$, C$_2$H$_4$-C$_2$F$_4$, and aminophenol — these methods' accuracies are similar. However, for the charge density changes in NaCl and in the solvated charge transfer example of the previous section, ESMF is substantially more accurate than ROKS. The poorer performance of CIS and TDDFT can largely be explained by their inability to relax the orbitals not directly involved in the excitation. ROKS, on the other hand, does contain these types of relaxations through its ΔSCF approach, but in some cases self-interaction error causes it to significantly overestimate the delocalization of the hole. We emphasize that we have employed a range-separated hybrid functional (ωB97X) for TDDFT and ROKS in order to maximize their ability to deal successfully with charge transfer. Nonetheless, we find that ESMF is more reliable for predicting how the charge density changes in these charge transfer excitations.

IV. CONCLUSION

We have presented explicit analytic expressions for the derivatives required for ESMF wave function optimization. Although these expressions are somewhat long winded, the upshot is that the necessary derivatives can be evaluated through nine Fock matrix builds and a collection of (much less expensive) single-particle matrix operations. This formulation allows ESMF to immediately benefit from methods that improve the efficiency of Fock builds, a situation we have exploited via the shell-pair screening strategy in order to construct a thread-parallel ESMF implementation that is multiple orders of magnitude faster than the previous implementation, which relied on automatic differentiation. We then applied our method to study a Li-F charge transfer system in which these atoms were surrounded by minimal water solvation shells. By comparing against IP/EA-EOM-CCSD, we found that ESMF's predictions for how the charge transfer excitation changes the electron cloud were qualitatively more accurate than the predictions of CIS, TDDFT, UKS, or ROKS. In particular, ESMF correctly captured post-excitation orbital relaxation effects and, by virtue of being free of self-interaction errors, did not erroneously delocalize the hole.

Looking forward, these findings point to multiple priorities for further research. First, by giving ESMF access to larger systems, they make the development of an N^5-scaling post-ESMF perturbation theory (as opposed to the N^7-scaling of the existing[22] version) even more pressing. Happily, a path forward in this direction has been found and will be published

soon.[58] Second, we would expect the recently-introduced density functional extension to ESMF to show similar efficiency improvements upon development of Fock-build-based analytic gradients. Finally, in light of the efficiency offered by the successful generalization of the geometric direct minimization method to excited state orbital optimization,[59] it seems likely that employing our new analytic gradients in a custom-tailored quasi-Newton method would lead to further efficiency gains.

V. SUPPLEMENTAL MATERIAL

See supplementary material for the molecular geometries, Cartesian coordinates, and overlap integrals between ground and excited states.

ACKNOWLEDGMENTS

[1] G. D. Scholes, G. R. Fleming, A. Olaya-Castro, and R. van Grondelle, "Lessons from nature about solar light harvesting," Nat. Chem. **3**, 763–774 (2011).

[2] S. Refaely-Abramson, F. H. da Jornada, S. G. Louie, and J. B. Neaton, "Origins of singlet fission in solid pentacene from an ab initio green's function approach," Phys. Rev. Lett. **119**, 267401 (2017).

[3] J. K. Stolarczyk, S. Bhattacharyya, L. Polavarapu, and J. Feldmann, "Challenges and prospects in solar water splitting and CO_2 reduction with inorganic and hybrid nanostructures," ACS Catal. **8**, 3602–3635 (2018).

[4] Q. Wang and K. Domen, "Particulate photocatalysts for light-driven water splitting: Mechanisms, challenges, and design strategies," Chem. Rev. **Article ASAP**.

[5] K. J. Oosterbaan, A. F. White, and M. Head-Gordon, "Non-orthogonal configuration interaction with single substitutions for the calculation of core-excited states," J. Chem. Phys. **149**, 044116 (2018).

[6] M. L. Vidal, X. Feng, E. Epifanovsky, A. I. Krylov, and S. Coriani, "New and efficient equation-of-motion coupled-cluster framework for core-excited and core-ionized states," J. Chem. Theory Comput. **15**, 3117–3133 (2019).

[7] X. Zheng and L. Cheng, "Performance of delta-coupled-cluster methods for calculations of core-ionization energies of first-row elements," J. Chem. Theory Comput. **15**, 4945–4955 (2019).

[8] K. R. Siefermann, C. D. Pemmaraju, S. Neppl, A. Shavorskiy, A. A. Cordones, J. Vura-Weis, D. S. Slaughter, F. P. Sturm, F. Weise, H. Bluhm, *et al.*, "Atomic-scale perspective of ultrafast charge transfer at a dye–semiconductor interface," J. Phys. Chem. Lett **5**, 2753–2759 (2014).

[9] P. Chábera, K. S. Kjaer, O. Prakash, A. Honarfar, Y. Liu, L. A. Fredin, T. C. B. Harlang, S. Lidin, J. Uhlig, V. Sundström, R. Lomoth, P. Persson, and K. Wärnmark, "Feii hexa n-heterocyclic carbene complex with a 528 ps metal-to-ligand charge-transfer excited-state lifetime," J. Phys. Chem. Lett. **9**, 459–463 (2018).

[10] M. H. Beck, A. Jäckle, G. A. Worth, and H.-D. Meyer, "The multiconfiguration time-dependent Hartree (MCTDH) method: an efficient method for propagating wavepackets of several dimensions," Phys. Rep **324**, 1–105 (2000).

[11] J. C. Tully, "Perspective: Nonadiabatic dynamics theory," J. Chem. Phys. **137**, 22A301 (2012).

[12] T. Ziegler, M. Seth, M. Krykunov, J. Autschbach, and F. Wang, "On the relation between time-dependent and variational density functional theory approaches for the determination of excitation energies and transition moments," J. Chem. Phys. **130**, 154102 (2009).

[13] J. E. Subotnik, "Communication: Configuration interaction singles has a large systematic bias against charge-transfer states," J. Chem. Phys. **135**, 071104 (2011).

[14] Y. C. Park, M. Krykunov, and T. Ziegler, "On the relation between adiabatic time dependent density functional theory (TDDFT) and the Δ SCF-DFT method. introducing a numerically stable Δ SCF-DFT scheme for local functionals based on constricted variational DFT," Mol. Phys. **113**, 1636–1647 (2015).

[15] L. Zhao and E. Neuscamman, "Density functional extension to excited-state mean-field theory," J. Chem. Theory Comput. **16**, 164 (2020).

[16] A. Dreuw and M. Head-gordon, "Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules," Sciences-New York , 4009–4037 (2005).

[17] S. Hirata, M. Head-Gordon, and R. J. Bartlett, "Configuration interaction singles, time-dependent Hartree-Fock, and time-dependent density functional theory for the electronic excited states of extended systems," J. Chem. Phys. **111**, 10774–10786 (1999).

[18] K. Burke, J. Werschnik, and E. K. U. Gross, "Time-dependent density functional theory: Past, present, and future," J. Chem. Phys. **123**, 062206 (2005).

[19] A. I. Krylov, "Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: the Hitchhiker's guide to Fock space." Annu. Rev. Phys. Chem. **59**, 433–462 (2008).

[20] J. D. Watts, S. R. Gwaltney, and R. J. Bartlett, "Coupled-cluster calculations of the excitation energies of ethylene, butadiene, and cyclopentadiene," J. Chem. Phys. **105**, 6979–6988 (1996).

[21] L. Zhao and E. Neuscamman, "An Efficient Variational Principle for the Direct Optimization of Excited States," J. Chem. Theory Comput. **12**, 3436–3440 (2016), arXiv:1508.06683.

[22] J. A. R. Shea and E. Neuscamman, "Communication: A mean field platform for excited state quantum chemistry," J. Chem. Phys. **149** (2018).

[23] P. S. Bagus, "Self-consistent-field wave functions for hole states of some ne-like and ar-like ions," Phys. Rev. **139**, A619 (1965).

[24] H.-l. Hsu, E. R. Davidson, and R. M. Pitzer, "An scf method for hole states," J. Chem. Phys. **65**, 609–613 (1976).

[25] A. Naves de Brito, N. Correia, S. Svensson, and H. Ågren, "A theoretical study of x-ray photoelectron spectra of model molecules for polymethylmethacrylate," J. Chem. Phys. **95**, 2965–2974 (1991).

[26] N. A. Besley, A. T. Gilbert, and P. M. Gill, "Self-consistent-field calculations of core excited states," J. Chem. Phys. **130**, 124308 (2009).

[27] M. Filatov and S. Shaik, "A spin-restricted ensemble-referenced kohn-sham method and its application to diradicaloid situations," Chem. Phys. Lett. **304**, 429–437 (1999).

[28] T. Kowalczyk, T. Tsuchimochi, P. T. Chen, L. Top, and T. Van Voorhis, "Excitation energies and Stokes shifts from a restricted open-shell Kohn-Sham approach," J. Chem. Phys. **138** (2013).

[29] J. Olsen, B. O. Roos, P. Jo/rgensen, and H. J. A. Jensen, "Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces," J. Chem. Phys. **89**, 2185–2192 (1988).

[30] P. Å. Malmqvist, A. Rendell, and B. O. Roos, "The restricted active space self-consistent-field method, implemented with a split graph unitary group approach," J. Phys. Chem. **94**, 5477–5482 (1990).

[31] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications, Mineola, N.Y., 1996).

[32] T. Helgaker, P. Jøgensen, and J. Olsen, *Molecular Electronic Structure Theory* (John Wiley and Sons, Ltd, West Sussex, England, 2000) p. 162.

[33] J. A. R. Shea, E. Gwin, and E. Neuscamman, "A generalized variational principle with applications to excited state mean field theory," J. Chem. Theory Comput. **16**, 1526 (2020).

[34] H.-Z. Ye, M. Welborn, N. D. Ricke, and T. Van Voorhis, "σ-scf: A direct energy-targeting method to mean-field excited states," J. Chem. Phys **147**, 214104 (2017).

[35] H. Ye and T. V. Voorhis, "Half-projected σ self-consistent field for electronic excited states," J. Chem. Theory Comput. **15(5)**, 2954–2964 (2019).

[36] C. Umrigar, K. Wilson, and J. Wilkins, "Optimized trial wave functions for quantum monte carlo calculations," Phys. Rev. Lett. **60**, 1719 (1988).

[37]J. A. R. Shea and E. Neuscamman, "Size consistent excited states via algorithmic transformations between variational principles," J. Chem. Theory Comput. **13**, 6078–6088 (2017).

[38]S. D. Pineda Flores and E. Neuscamman, "Excited state specific multi-slater jastrow wave functions," J. Phys. Chem. A **123**, 1487–1497 (2019).

[39]M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," (2015), software available from tensorflow.org.

[40]D. S. Lambrecht, B. Doser, and C. Ochsenfeld, "Rigorous integral screening for electron correlation methods," J. Chem. Phys. **123**, 184102 (2005).

[41]A. Sodt, J. E. Subotnik, and M. Head-Gordon, "Linear scaling density fitting," J. Chem. Phys. **125**, 194109 (2006).

[42]E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, "Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory," J. Chem. Phys. **137**, 044103 (2012).

[43]R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, "Tensor hypercontraction. II. Least-squares renormalization," J. Chem. Phys. **137**, 224106 (2012).

[44]E. G. Hohenstein, R. M. Parrish, C. D. Sherrill, and T. J. Martínez, "Communication: Tensor hypercontraction. III. Least-squares tensor hypercontraction for the determination of correlated wavefunctions," J. Chem. Phys. **137**, 221101 (2012).

[45]Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. L. Chan, "PySCF: the Python-based simulations of chemistry framework," Wiley Interdiscip. Rev. Comput. Mol. Sci. **8** (2018).

[46]Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. W. III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. D. Jr., H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. Hanson-Heine, P. H. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. Chan, D. M. Chipman, C. J. Cramer, W. A. G. III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. S. III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. V. Voorhis, J. M. Herbert, A. I. Krylov, P. M. Gill, and M. Head-Gordon, "Advances in molecular quantum chemistry contained in the q-chem 4 program package," Mol. Phys. **113**, 184–215 (2015).

[47]K. Momma and F. Izumi, "Vesta 3 for three-dimensional visualization of crystal, volumetric and morphology data," J. Appl. Cryst. **44**, 1272–1276 (2011).

[48]T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. the atoms boron through neon and hydrogen," J. Chem. Phys. **90**, 1007 (1988).

[49]Y. Zhao and D. G. Truhlar, "Density functional for spectroscopy: No long-range self-interaction error, good performance for Rydberg and charge-transfer states, and better performance on average than B3LYP for ground states," J. Phys. Chem. A **110**, 13126–13130 (2006).

[50]D. Majumdar, J. Kim, and K. S. Kim, "Charge transfer to solvent (CTTS) energies of small $X^-(H_2O)_{n=1-4}(X = F, Cl, Br, I)$ clusters: Ab initio study," J. Chem. Phys. **112**, 101 (2000).

[51]G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley, and J. Mantzaris, "A complete basis set model chemistry. i. the total energies of closed-shell atoms and hydrides of the first-row atoms," J. Chem. Phys. **89**, 2193–218 (1988).

[52]J.-D. Chai and M. Head-Gordon, "Systematic optimization of long-range corrected hybrid density functionals," J. Chem. Phys. **128**, 084106 (2008).

[53]D. Sinha, S. K. Mukhopadhyay, R. Chaudhuri, and D. Mukherjee, "The eigenvalue-independent partitioning technique in fock space: An alternative route to open-shell coupled-cluster theory for incomplete model spaces," Chem. Phys. Lett. **164**, 544 (1988).

[54]J. F. Stanton and J. Gauss, "Analytic energy derivatives for ionized states described by the equation-of-motion coupled cluster method," J. Chem. Phys. **1010**, 8938 (1994).

[55]M. Nooijen and R. J. Barlett, "Equation of motion coupled cluster method for electron attachment," J. Chem. Phys. **102**, 3629 (1995).

[56]M. Lundberg and P. E. Siegbahn, "Quantifying the effects of the self-interaction error in DFT: When do the delocalized states appear?" J. Chem. Phys. **122**, 224103 (2005).

[57]A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Insights into Current Limitations of Density Functional Theory," Science **321**, 792–794 (2008).

[58]R. Clune, J. A. R. Shea, and E. Neuscamman, in preparation.

[59]D. Hait and M. Head-Gordon, "Excited state orbital optimization via minimizing the square of the gradient: General approach and application to singly and doubly excited states via density functional theory," arXiv:1911.04709 (2019).

TABLE VII: Differences in atomic Mulliken populations for atoms between the excited and ground states of $Cl^--(H_2O)_3$. DFT methods used the ωB97X functional.

Method	Cl	O	H	H	O	H	H	O	H	H	total error
CIS	0.86	0.04	-0.21	-0.09	0.07	-0.11	-0.24	0.04	-0.26	-0.09	0.32
TDDFT	0.88	0.01	-0.18	-0.13	0.04	-0.14	-0.19	0.02	-0.19	-0.13	0.25
ROKS	0.78	0.07	-0.21	-0.12	0.08	-0.12	-0.19	0.09	-0.24	-0.13	0.17
ESMF	0.85	0.06	-0.19	-0.14	0.07	-0.15	-0.21	0.08	-0.24	-0.15	0.16
EOM-CCSD	0.80	0.06	-0.18	-0.13	0.08	-0.14	-0.20	0.06	-0.21	-0.13	N/A

TABLE VIII: Differences in atomic Mulliken populations for atoms between the excited and ground states of C_2H_4-C_2F_4. DFT methods used the ωB97X functional.

Method	C	C	H	H	H	H	C	C	F	F	F	F	total error
CIS	0.49	0.49	0.00	0.00	0.00	0.00	-0.46	-0.46	-0.02	-0.02	-0.02	-0.02	1.30
TDDFT	0.49	0.49	0.00	0.00	0.00	0.00	-0.44	-0.44	-0.03	-0.03	-0.03	-0.03	1.19
ROKS	0.22	0.22	0.16	0.13	0.16	0.13	-0.28	-0.28	-0.10	-0.12	-0.10	-0.12	0.52
ESMF	0.24	0.24	0.15	0.12	0.15	0.12	-0.28	-0.28	-0.10	-0.12	-0.10	-0.12	0.45
EOM-CCSD	0.30	0.30	0.11	0.09	0.11	0.09	-0.33	-0.33	-0.08	-0.09	-0.08	-0.09	N/A

TABLE IX: Differences in atomic Mulliken populations for atoms between the excited and ground states of aminophenol. DFT methods used the ωB97X functional.

Method	H	C	C	H	C	O	C	H	C	H	C	N	H	H	H	total error
CIS	-0.00	-0.10	-0.14	-0.00	0.10	0.05	-0.13	0.00	-0.07	0.00	0.12	0.12	0.00	0.00	0.00	0.57
TDDFT	-0.00	-0.11	-0.14	-0.00	0.11	0.09	-0.13	0.00	-0.09	0.00	0.07	0.19	0.00	0.00	0.00	0.60
ROKS	-0.02	-0.04	-0.06	-0.02	0.05	0.04	-0.06	-0.02	-0.03	-0.02	0.00	0.11	0.01	0.03	0.03	0.13
ESMF	-0.01	-0.04	-0.07	-0.02	0.07	0.03	-0.07	-0.02	-0.03	-0.01	0.00	0.11	0.01	0.03	0.03	0.18
EOM-CCSD	-0.01	-0.05	-0.06	-0.01	0.03	0.05	-0.05	-0.01	-0.04	-0.01	0.00	0.12	0.01	0.02	0.02	N/A

Appendix

We provide the analytic expression of ESMF energy and target function derivatives in this section. Although the final results look tedious, the derivation is trivial as the ESMF energy and target function only involve matrix trace and products. Here we take the derivative of the first term in E_1 and E_2 with respect to Θ_{kp} for example,

$$
\begin{aligned}
&\frac{\partial}{\partial \Theta_{kp}} \left(2N_2 \mathrm{Tr}[\Theta \mathbf{G}\Theta^T]\right) \\
&= \frac{\partial}{\partial \Theta_{kp}} \left(2N_2 \sum_i [\Theta \mathbf{G}\Theta]_{ii}\right) \\
&= \frac{\partial}{\partial \Theta_{kp}} \left(2N_2 \sum_i \sum_{rs} \Theta_{ir} G_{rs} \Theta_{is}\right) \\
&= 2N_2 \sum_i \sum_{rs} \Theta_{ir} G_{rs} \delta_{ki}\delta_{ps} + 2N_2 \sum_i \sum_{rs} \delta_{ki}\delta_{rp} G_{rs} \Theta_{is} \\
&= 2N_2 \sum_r \Theta_{kr} G_{rp} + 2N_2 \sum_r G_{ps}\Theta_{ks} \\
&= 4N_2 (\Theta \mathbf{G})_{kp}
\end{aligned}
\tag{22}
$$

$$
\begin{aligned}
&\frac{\partial}{\partial \Theta_{kp}} \left(N_2 \mathbf{F}[\Theta^T\Theta]\cdot(\Theta^T\Theta)\right) \\
&= \frac{\partial}{\partial \Theta_{kp}} \left(N_2 \sum_{rs} \mathbf{F}[\Theta^T\Theta]_{rs} (\Theta^T\Theta)_{rs}\right) \\
&= N_2 \sum_{rs} \frac{\mathbf{F}[\Theta^T\Theta]_{rs}}{\Theta_{kp}} (\Theta^T\Theta)_{rs} + N_2 \sum_{rs} \mathbf{F}[\Theta^T\Theta]_{rs} \frac{(\Theta^T\Theta)_{rs}}{\Theta_{kp}}
\end{aligned}
\tag{23}
$$

The derivative of the Fock matrix is

$$
\begin{aligned}
&\frac{\partial \mathbf{F}[\Theta^T\Theta]_{rs}}{\partial \Theta_{kp}} \\
&= \frac{\partial}{\partial \Theta_{kp}} \left(2\mathbf{J}[\Theta^T\Theta]_{rs} - \mathbf{K}[\Theta^T\Theta]_{rs}\right) \\
&= 2\sum_v \Theta_{kv}(pv|rs) + 2\sum_u \Theta_{ku}(pu|rs) \\
&\quad - \sum_v \Theta_{kv}(rp|sv) - \sum_u \Theta_{ku}(ru|sp)
\end{aligned}
\tag{24}
$$

Plugging this into the first term of Equation 22, we obtained

$$
\begin{aligned}
&N_2 \sum_{rs} \frac{\mathbf{F}[\Theta^T\Theta]_{rs}}{\Theta_{kp}} (\Theta^T\Theta)_{rs} \\
&= N_2 \sum_{rs} (\Theta^T\Theta)_{rs} [2\sum_v \Theta_{kv}(pv|rs) + 2\sum_u \Theta_{ku}(pu|rs)] \\
&\quad - N_2 \sum_{rs} (\Theta^T\Theta)_{rs} [\sum_v \Theta_{kv}(rp|sv) + \sum_u \Theta_{ku}(ru|sp)] \\
&= N_2 \sum_v \Theta_{kv} \mathbf{F}[\Theta^T\Theta]_{pv} + N_2 \sum_u \Theta_{ku} \mathbf{F}[\Theta^T\Theta]_{pu} \\
&= 2N_2 (\Theta \mathbf{F}[\Theta^T\Theta])_{kp}
\end{aligned}
$$

The second term of Equation 22 is

$$
\begin{aligned}
&N_2 \sum_{rs} \mathbf{F}[\Theta^T\Theta]_{rs} \frac{(\Theta^T\Theta)_{rs}}{\Theta_{kp}} \\
&\sum_{rs} \mathbf{F}[\Theta^T\Theta]_{rs} \sum_i \delta_{ki}\delta_{pr}\Theta_{is} + \sum_{rs} \mathbf{F}[\Theta^T\Theta]_{rs} \sum_i \Theta_{ir}\delta_{ki}\delta_{ps} \\
&= 2N_2 \left(\Theta \mathbf{F}[\Theta^T\Theta]\right)_{kp}
\end{aligned}
\tag{25}
$$

Putting these together, we obtain

$$
\frac{\partial}{\partial \Theta_{kp}} \left(N_2 \mathbf{F}[\Theta^T\Theta]\cdot(\Theta^T\Theta)\right) = 4N_2 \left(\Theta \mathbf{F}[\Theta^T\Theta]\right)_{kp}
\tag{26}
$$

The other terms of the energy and objective function derivatives could be computed in similar ways, and their expression are as follows.

The derivative with respect to c_0 is,

$$
\begin{aligned}
\frac{\partial E}{\partial c_0} &= \left(4c_0 \mathrm{Tr}[\Theta\,\mathbf{G}\Theta^T] + 4\mathrm{Tr}[\Theta\mathbf{G}\Gamma^T\sigma^T] \right. \\
&\quad + 2c_0 \mathbf{F}[\Theta^T\Theta]\cdot(\Theta^T\Theta) + 4\mathbf{F}[\Theta^T\Theta]\cdot(\Theta^T\sigma\Gamma)\left.\right)/N_2 \\
&\quad - 2c_0 E/N_2
\end{aligned}
\tag{27}
$$

The derivative with respect to σ_{jb} is,

$$
\begin{aligned}
\frac{\partial E}{\partial \sigma_{jb}} &= \left(4c_0 \left(\Theta\mathbf{G}\Gamma^T\right)_{jb} + 4\left(\sigma\Gamma\mathbf{G}\Gamma^T\right)_{jb} \right. \\
&\quad - 4\left(\Theta\mathbf{G}\Theta^T\sigma\right)_{jb} + 8\mathrm{Tr}[\Theta\mathbf{G}\Theta^T]\sigma_{jb} \\
&\quad + 4c_0 \left(\Theta\mathbf{F}[\Theta^T\Theta]\Gamma^T\right)_{jb} + 4\left(\sigma\Gamma\mathbf{F}[\Theta^T\Theta]\Gamma^T\right)_{jb} \\
&\quad - 4\left(\Theta\mathbf{F}[\Theta^T\Theta]\Theta^T\sigma\right)_{jb} + 4\left(\Theta\mathbf{F}[\Theta^T\sigma\Gamma]\Gamma^T\right)_{jb} \\
&\quad + 4\sum_{pq} \left(\mathbf{F}[\Theta^T\Theta]\Theta^T\Theta\right)_{pq}\sigma_{jb}\left.\right)/N_2 \\
&\quad - 4E\sigma_{jb}/N_2
\end{aligned}
\tag{28}
$$

The derivative with respect to Θ_{kp} is,

$$
\begin{aligned}
\frac{\partial E}{\partial \Theta_{kp}} &= \left(4N_2 (\Theta\mathbf{G})_{kp} + 4c_0 (\sigma\Gamma\mathbf{G})_{kp} \right. \\
&\quad - 4\left(\sigma\sigma^T\Theta\mathbf{G}\right)_{kp} + 4N_2 \left(\Theta\mathbf{F}[\Theta^T\Theta]\right)_{kp} \\
&\quad + 4c_0 \left(\sigma\Gamma\mathbf{F}[\Theta^T\sigma\Gamma]\right)_{kp} \\
&\quad + 4c_0 \left(\Theta \left(\mathbf{F}[\Theta^T\sigma\Gamma] - F[\Theta^T\sigma\Gamma]^T\right)\right)_{kp} \\
&\quad + 2\left(\Theta \left(\mathbf{F}[\mathbf{A}] - \mathbf{F}[\mathbf{A}]^T\right)\right)_{kp} - 4\left(\sigma\sigma^T\Theta\mathbf{F}[\Theta^T\Theta]\right)_{kp} \\
&\quad + 4\left(\sigma\Gamma\mathbf{F}[\Theta^T\sigma\Gamma]\right)_{kp} /N_2 \left.\right)
\end{aligned}
\tag{29}
$$

The derivative with respect to Γ_{cp} is,

$$
\begin{aligned}
\frac{\partial E}{\partial \Gamma_{cp}} &= \left(4c_0 \left(\sigma^T\Theta\mathbf{G}\right)_{cp} + 4\left(\sigma^T\sigma\Gamma\mathbf{G}\right)_{cp} \right. \\
&\quad + 4c_0 \left(\sigma^T\Theta\mathbf{F}[\Theta^T\Theta]\right)_{cp} + 4\left(\sigma^T\sigma\Gamma\mathbf{F}[\Theta^T\Theta]\right)_{cp} \\
&\quad + 4\left(\sigma^T\Theta\mathbf{F}[\Theta^T\sigma\Gamma]\right)_{cp} \left.\right)
\end{aligned}
\tag{30}
$$

Let us now define different pieces of L,

The derivatives of L_1, L_2, and L_3 with respect to σ_{jb} are,

$$
\begin{aligned}
L_1 =\ & 4c_0\mathrm{Tr}[\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Theta}^T] + 4\mathrm{Tr}[\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Gamma}^T\boldsymbol{\sigma}^T] \\
& + 2c_0F[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\Theta}) + 4\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}) \\
L_2 =\ & 4c_0\mathrm{Tr}[\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Gamma}^T\boldsymbol{\mu}^T] + 4\mathrm{Tr}[\boldsymbol{\sigma}\boldsymbol{\Gamma}\mathbf{G}\boldsymbol{\Gamma}^T\boldsymbol{\mu}^T] \\
& - 2\mathrm{Tr}[\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\mu}^T + \boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Theta}^T\boldsymbol{\mu}\boldsymbol{\sigma}^T] + 8\boldsymbol{\sigma}\cdot\boldsymbol{\mu}\mathrm{Tr}[\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Theta}^T] \\
& + 4c_0\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\mu}\boldsymbol{\Gamma}) \\
& + 2\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\boldsymbol{\Gamma}^T\boldsymbol{\sigma}^T\boldsymbol{\mu}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}^T\boldsymbol{\mu}^T\boldsymbol{\sigma}\boldsymbol{\Gamma} - \boldsymbol{\Theta}^T\boldsymbol{\mu}\boldsymbol{\sigma}^T\boldsymbol{\Theta} - \boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\mu}^T\boldsymbol{\Theta}) \\
& + 4(\boldsymbol{\sigma}\cdot\boldsymbol{\mu})(\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\Theta})) + 4\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}) \\
L_3 =\ & 4N_2\mathrm{Tr}[\mathbf{R}\mathbf{G}\boldsymbol{\Theta}^T] + 4c_0\mathrm{Tr}[\mathbf{R}\mathbf{G}\boldsymbol{\Gamma}^T\boldsymbol{\sigma}^T] \\
& - 2\mathrm{Tr}[\mathbf{R}\mathbf{G}\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\sigma}^T + \boldsymbol{\Theta}\mathbf{G}\mathbf{R}^T\boldsymbol{\sigma}\boldsymbol{\sigma}^T] \\
& + 2N_2\left(\mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}] + \mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}]^T\right)\cdot(\boldsymbol{\Theta}^T\boldsymbol{\Theta}) \\
& + 4c_0\left(\mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}] + \mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}]^T\right)\cdot(\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}) \\
& + 4c_0\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\mathbf{R}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}) \\
& + 2\left(\mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}] + \mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}]^T\right)\cdot\mathbf{A} \\
& - 2\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot\left(\mathbf{R}^T\boldsymbol{\sigma}\boldsymbol{\sigma}^T\boldsymbol{\Theta} + (\mathbf{R}^T\boldsymbol{\sigma}\boldsymbol{\sigma}^T\boldsymbol{\Theta})^T\right) \\
& + 4\mathbf{F}[\mathbf{R}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}) \\
& + 4c_0\mathrm{Tr}[\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Phi}^T\boldsymbol{\sigma}^T] + 2\mathrm{Tr}[\boldsymbol{\sigma}\boldsymbol{\Gamma}\mathbf{G}\boldsymbol{\Phi}^T\boldsymbol{\sigma}^T + (\boldsymbol{\sigma}\boldsymbol{\Gamma}\mathbf{G}\boldsymbol{\Phi}^T\boldsymbol{\sigma}^T)^T] \\
& + 4c_0\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Phi}) \\
& + 2\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot\left(\boldsymbol{\Phi}^T\boldsymbol{\sigma}^T\boldsymbol{\sigma}\boldsymbol{\Gamma} + (\boldsymbol{\Phi}^T\boldsymbol{\sigma}^T\boldsymbol{\sigma}\boldsymbol{\Gamma})^T\right) \\
& + 4\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Phi})
\end{aligned}
$$

(31)

$$
\frac{\partial L_1}{\partial \sigma_{jb}} = 4\left(\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Gamma}^T\right)_{jb} + 4\left(\boldsymbol{\Theta}\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\boldsymbol{\Gamma}^T\right)_{jb}
$$

$$
\begin{aligned}
\frac{\partial L_2}{\partial \sigma_{jb}} =\ & 4\left(\boldsymbol{\Gamma}\mathbf{G}\boldsymbol{\Gamma}^T\boldsymbol{\mu}^T\right)^T_{jb} - 4\left(\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Theta}^T\boldsymbol{\mu}\right)_{jb} + 8\mathrm{Tr}[\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Theta}^T]\mu_{jb} \\
& + 2\left(\boldsymbol{\mu}\boldsymbol{\Gamma}\left(\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}] + \mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]^T\right)\boldsymbol{\Gamma}^T\right)_{jb} \\
& - 2\left(\boldsymbol{\Theta}\left(\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}] + \mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]^T\right)\boldsymbol{\Theta}^T\boldsymbol{\mu}\right)_{jb} \\
& + 4\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\Theta})\mu_{jb} + 4\left(\boldsymbol{\Theta}\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\mu}\boldsymbol{\Gamma}]\boldsymbol{\Gamma}^T\right)_{jb}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L_3}{\partial \sigma_{jb}} =\ & 4c_0\left(\mathbf{R}\mathbf{G}\boldsymbol{\Gamma}^T\right)_{jb} - 4\left(\mathbf{R}\mathbf{G}\boldsymbol{\Theta}^T\boldsymbol{\sigma}\right)_{jb} + 16\mathrm{Tr}[\mathbf{R}\mathbf{G}\boldsymbol{\Theta}^T]\sigma_{jb} \\
& + 4c_0\left(\boldsymbol{\Theta}\left(\mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}] + \mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}]^T\right)\boldsymbol{\Gamma}^T\right)_{jb} \\
& + 4\left(\boldsymbol{\sigma}\boldsymbol{\Gamma}\left(\mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}] + \mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}]^T\right)\boldsymbol{\Gamma}^T\right)_{jb} \\
& - 4\left(\boldsymbol{\Theta}\left(\mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}] + \mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}]^T\right)\boldsymbol{\Theta}^T\boldsymbol{\sigma}\right)_{jb} \\
& + 4c_0\left(\mathbf{R}\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\boldsymbol{\Gamma}^T\right)_{jb} \\
& - 2\left(\mathbf{R}\left(\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}] + \mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]^T\right)\boldsymbol{\Theta}^T\boldsymbol{\sigma}\right)_{jb} \\
& - 2\left(\boldsymbol{\Theta}\left(\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}] + \mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]^T\right)\mathbf{R}^T\boldsymbol{\sigma}\right)_{jb} \\
& + 4\left(\mathbf{R}\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}]\mathbf{U}^T\right)_{jb} + 4\left(\boldsymbol{\Theta}\mathbf{F}[\mathbf{R}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}]\boldsymbol{\Gamma}^T\right)_{jb} \\
& + 8F[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot\left(\mathbf{R}^T\boldsymbol{\Theta} + \theta\mathbf{R}^T\right)\sigma_{jb} \\
& + 4c_0\left(\boldsymbol{\Phi}G\boldsymbol{\Theta}^T\right)^T_{jb} + 4\left(\boldsymbol{\Phi}G\boldsymbol{\Gamma}^T\boldsymbol{\sigma}^T\right)^T_{jb} + 4\left(\boldsymbol{\Gamma}G\boldsymbol{\Phi}^T\boldsymbol{\sigma}^T\right)^T_{jb} \\
& + 4c_0\left(\boldsymbol{\Theta}F[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\boldsymbol{\Phi}^T\right)_{jb} \\
& + 2\left(\boldsymbol{\sigma}\boldsymbol{\Phi}\left(F[\boldsymbol{\Theta}^T\boldsymbol{\Theta}] + F[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]^T\right)\boldsymbol{\Gamma}^T\right)_{jb} \\
& + 2\left(\boldsymbol{\sigma}\boldsymbol{\Gamma}\left(F[\boldsymbol{\Theta}^T\boldsymbol{\Theta}] + F[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]^T\right)\boldsymbol{\Phi}^T\right)_{jb} \\
& + 4\left(\boldsymbol{\Theta}\left(F[\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}] + F[\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}]^T\right)\boldsymbol{\Phi}^T\right)_{jb} \\
& + 4\left(\boldsymbol{\Theta}\left(F[\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Phi}] + F[\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Phi}]^T\right)\boldsymbol{\Gamma}^T\right)_{jb}
\end{aligned}
$$

(33)

The derivatives of L_1, L_2, and L_3 with respect to c_0 are,

$$
\frac{\partial L_1}{\partial c_0} = 4\mathrm{Tr}[\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Theta}^T] + 2\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\Theta})
$$

$$
\frac{\partial L_2}{\partial c_0} = 4\mathrm{Tr}[\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Gamma}^T\boldsymbol{\mu}^T] + 4F[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\mu}\boldsymbol{\Gamma})
$$

$$
\begin{aligned}
\frac{\partial L_3}{\partial c_0} =\ & 8c_0\mathrm{Tr}[\mathbf{R}\mathbf{G}\boldsymbol{\Theta}^T] + 4\mathrm{Tr}[\mathbf{R}\mathbf{G}\boldsymbol{\Gamma}^T\boldsymbol{\sigma}^T] \\
& + 4c_0\left(\mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}] + \mathbf{F}[\mathbf{R}^T\boldsymbol{\Theta}]^T\right)\cdot(\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}) \\
& + 4\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\mathbf{R}^T\boldsymbol{\sigma}\boldsymbol{\Gamma}) \\
& + 4\mathrm{Tr}[\boldsymbol{\Theta}\mathbf{G}\boldsymbol{\Phi}^T\boldsymbol{\sigma}^T] + 4\mathbf{F}[\boldsymbol{\Theta}^T\boldsymbol{\Theta}]\cdot(\boldsymbol{\Theta}^T\boldsymbol{\sigma}\boldsymbol{\Phi})
\end{aligned}
$$

(32)

in which no new Fock build is needed.

The derivatives of L_1, L_2, and L_3 with respect to Θ_{kp} are,

$$\frac{\partial L_1}{\partial \Theta_{kp}} = 8c_0 \left(\Theta G\right)_{kp} + 4\left(\sigma \Gamma G\right)_{kp}$$
$$+ 8c_0 \left(\Theta F[\Theta^T \Theta]\right)_{kp} + 4\left(\sigma \Gamma F[\Theta^T \Theta]\right)_{kp}$$
$$+ 4\left(\Theta \left(F[\Theta^T \sigma \Gamma] + F[\Theta^T \sigma \Gamma]^T\right)\right)_{kp}$$

$$\frac{\partial L_2}{\partial \Theta_{kp}} = 4c_0 \left(\mu \Gamma G\right)_{kp} - 4\left(\mu \sigma^T \Theta G\right)_{kp} + 16\sigma \cdot \mu \left(\Theta G\right)_{kp}$$
$$+ 4c_0 \left(\mu \Gamma F[\Theta^T \Theta]^T\right)_{kp}$$
$$+ 4c_0 \left(\Theta \left(F[\Theta^T \mu \Gamma] + F[\Theta^T \mu \Gamma]^T\right)\right)_{kp}$$
$$- 4\left(\Theta \left(F[\Theta^T \mu \sigma^T \Theta] + F[\Theta^T \mu \sigma^T \Theta]^T\right)\right)_{kp}$$
$$- 2\left(\left(\mu \sigma^T \Theta + \sigma \mu^T \Theta\right)\left(F[\Theta^T \Theta] + F[\Theta^T \Theta]^T\right)\right)_{kp}$$
$$+ 8\sigma \cdot \mu \left(\Theta \left(F[\Theta^T \Theta] + F[\Theta^T \Theta]^T\right)\right)_{kp}$$
$$+ 4\left(\mu \Gamma F[\Theta^T \sigma \Gamma]^T\right)_{kp} + 4\left(\sigma \Gamma F[\Theta^T \mu \Gamma]^T\right)_{kp}$$

$$\frac{\partial L_3}{\partial \Theta_{kp}} = 4N_2 \left(RG\right)_{kp} - 4\left(\sigma \sigma^T RG\right)_{kp}$$
$$+ 2N_2 \left(R \left(F[\Theta^T \Theta] + F[\Theta^T \Theta]^T\right)\right)_{kp}$$
$$+ 4N_2 \left(\Theta \left(F[R^T \Theta] + F[R^T \Theta]^T\right)\right)_{kp}$$
$$+ 4c_0 \left(R \left(F[\Theta^T \sigma \Gamma] + F[\Theta^T \sigma \Gamma]^T\right)\right)_{kp}$$
$$+ 4c_0 \left(\sigma \Gamma \left(F[R^T \Theta] + F[R^T \Theta]^T\right)\right)_{kp}$$
$$+ 4c_0 \left(\Theta \left(F[R^T \sigma \Gamma] + F[R^T \sigma \Gamma]^T\right)\right)_{kp}$$
$$+ 4c_0 \left(\Theta \left(F[\Theta^T \sigma \Phi] + F[\Theta^T \sigma \Phi]^T\right)\right)_{kp}$$
$$+ 4c_0 \left(\sigma \Phi F[\Theta^T \Theta]\right)_{kp} + 2\left(R \left(F[A] + F[A]^T\right)\right)_{kp}$$
$$- 4\left(\sigma \sigma^T \Theta \left(F[R^T \Theta] + F[R^T \Theta]^T\right)\right)_{kp}$$
$$+ 4\left(\Theta \left(F[B] + F[B]^T\right)\right)_{kp}$$
$$- 2\left(\sigma \sigma^T R \left(F[\Theta^T \Theta] + F[\Theta^T \Theta]^T\right)\right)_{kp}$$
$$+ 4\left(\sigma \Gamma F[R^T \sigma \Gamma]^T\right)_{kp}$$
$$+ 4\left(\sigma \Phi F[\Theta^T \sigma \Gamma]^T\right)_{kp}$$
$$+ 4\left(\sigma \Gamma F[\Theta^T \sigma \Phi]^T\right)_{kp}$$

$$(34)$$

The derivatives of L_1, L_2, and L_3 with respect to Γ_{cp} are,

$$\frac{\partial L_1}{\partial \Gamma_{cp}} = 4 \left(\sigma^T \Theta G\right)_{cp} + 4 \left(\sigma^T \Theta F[\Theta^T \Theta]\right)_{cp}$$

$$\frac{\partial L_2}{\partial \Gamma_{cp}} = 4c_0 \left(\mu \Gamma G\right)_{cp} - 4 \left(\mu \sigma^T \Theta G\right)_{cp}$$
$$+ 4 \left(\sigma^T \mu \Gamma G\right)_{cp} + 4c_0 \left(\mu^T \Theta F[\Theta^T \Theta]\right)_{cp}$$
$$+ 2 \left(\sigma^T \mu \Gamma \left(F[\Theta^T \Theta] + F[\Theta^T \Theta]^T\right)\right)_{cp}$$
$$+ 2 \left(\mu^T \sigma \Gamma \left(F[\Theta^T \Theta] + F[\Theta^T \Theta]^T\right)\right)_{cp}$$
$$+ 4 \left(\sigma^T \Theta F[\Theta^T \mu \Gamma]\right)_{cp} + 4 \left(\mu^T \Theta F[\Theta^T \sigma \Gamma]\right)_{cp}$$

$$\frac{\partial L_3}{\partial \Gamma_{cp}} = 4c_0 \left(\sigma^T RG\right)_{cp} + 4 \left(\sigma^T \sigma \Phi G\right)_{cp} \qquad (35)$$
$$+ 4c_0 \left(\sigma^T \Theta \left(F[R^T \Theta] + F[R^T \Theta]^T\right)\right)_{cp}$$
$$+ 4 \left(\sigma^T \sigma \Gamma \left(F[R^T \Theta] + F[R^T \Theta]^T\right)\right)_{cp}$$
$$+ 4c_0 \left(\sigma^T RF[\Theta^T \Theta]\right)_{cp}$$
$$+ 4 \left(\sigma^T RF[\Theta^T \sigma \Gamma]\right)_{cp}$$
$$+ 4 \left(\sigma^T \Theta F[R^T \sigma \Gamma]\right)_{cp}$$
$$+ 2 \left(\sigma^T \sigma \Phi \left(F[\Theta^T \Theta] + F[\Theta^T \Theta]^T\right)\right)_{cp}$$
$$+ 4 \left(\sigma^T \Theta F[\Theta^T \sigma \Phi]\right)_{cp}$$

Supplemental Materials: Excited State Mean-Field Theory without Automatic Differentiation

Luning Zhao[1] and Eric Neuscamman[2, 3, a)]

[1)]*Department of Chemistry, University of Washington, Seattle, Washington 98195, USA*

[2)]*Department of Chemistry, University of California, Berkeley, California 94720, USA*

[3)]*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA*

(Dated: 2 April 2020)

I. MOLECULAR GEOMETRIES

The molecular structures of NaCl, Cl^--$(H_2O)_3$, C_2H_4-C_2F_4, and aminophenol are show in Figure 1, 2, 3, and 4. We also include the Cartesian coordinates of these molecules, and the bond length used for NaCl is 2.36 $\AA$.
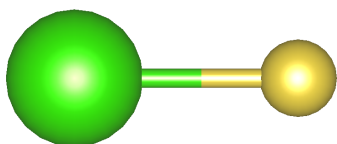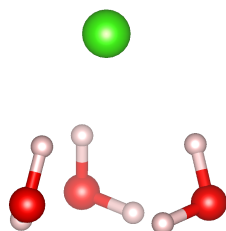

FIG. 1: Geometry of NaCl.

FIG. 2: Geometry of Cl^--$(H_2O)_3$.

FIG. 3: Geometry of C_2H_4-C_2F_4.

The geometry of Cl-H_2O is, (units are in $\AA$)

Cl	-0.848210	0.000015	0.074343
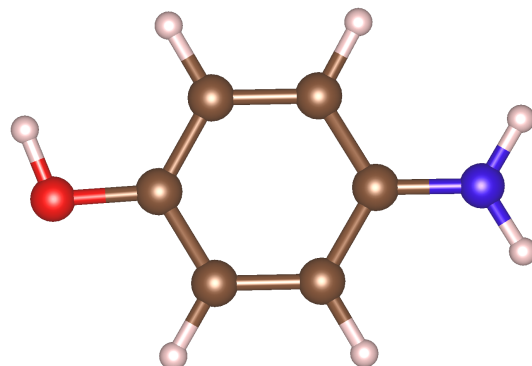
FIG. 4: Geometry of aminophenol.

O	2.337083	0.000017	0.186525
H	1.693454	0.725249	0.163751
H	1.693562	-0.725306	0.163748

The geometry of NH_3-F_2 is,

N	0.000000	0.000000	6.000000
H	0.000000	0.939731	5.611703
H	0.813831	-0.469865	5.611703
H	-0.813831	-0.469865	5.611703
F	0.000000	0.000000	12.000000
F	0.000000	0.000000	13.430000

The geometry of Li-F system is, (units are in $\AA$)

Li	1.796944	0.402653	7.714191
O	0.913870	1.932310	6.845200
O	0.377075	-0.964195	7.698416
O	2.515306	0.844431	9.484318
O	3.091745	-0.174192	6.348637
H	3.349025	0.540540	9.879509
H	2.133055	1.467664	10.123912
H	3.887425	-0.730479	6.379252
H	3.192501	0.388996	5.562935
H	0.429815	-1.696650	7.061571
H	-0.089280	-1.321908	8.472304
H	1.105006	2.884390	6.866262
H	-0.019955	1.852997	6.587612
F	0.243106	-0.058090	-0.170912
O	2.303862	1.382309	0.978656
O	-2.282249	-0.646317	-1.078369
O	0.137241	2.682825	-0.515020
O	2.626535	-0.051411	-1.623087
O	-1.783124	1.055466	1.206736
O	-1.154603	-1.907153	1.277859

H   1.641197   0.687837   0.777891
H   1.835652   2.163873   0.627908
H   -1.334445   -0.386089   -1.061835
H   -2.258393   -1.373147   -0.426111
H   -0.447054   -1.443480   0.773265
H   -1.562261   -1.153285   1.740562
H   2.944434   0.519587   -0.901516
H   1.727649   -0.252872   -1.297097
H   -0.614192   2.646830   0.104027
H   0.260635   1.727937   -0.704091
H   -0.914261   0.722919   0.879951
H   -2.352180   0.731683   0.478898

The geometry of $Cl^--(H_2O)_3$ is, (units are in Å)
Cl 0.050329   0.035472   -2.015391
O   -1.047843   -1.483881   0.617731
H   -0.907505   -1.210859   -0.310726
H   -0.146825   -1.432497   0.972306
O   -0.843184   1.553755   0.686802
H   -1.271587   0.728045   0.960088
H   -0.626924   1.361459   -0.247573
O   1.675737   -0.141942   0.774992
H   1.461043   -0.169950   -0.178449
H   1.151415   0.618930   1.069156

The geometry of $C_2H_4-C_2F_4$ is, (units are in Å)
C   0.000000   0.000000   0.000000
C   0.000000   1.336380   0.000000
H   0.923269   -0.570117   0.000000
H   -0.923269   -0.570117   0.000000
H   0.923269   1.906497   0.000000
H   -0.923269   1.906497   0.000000
C   7.000000   0.004179   0.000000
C   7.000000   1.332202   0.000000
F   8.113928   -0.719141   0.000000
F   5.886072   -0.719141   0.000000
F   8.113928   2.055521   0.000000

F   5.886072   2.055521   0.000000

The geometry of aminophenol is, (units are in Å)
H   1.247720   -2.140105   0.000005
C   0.694193   -1.202493   0.000004
C   -0.697351   -1.230213   0.000000
H   -1.235703   -2.174088   -0.000001
C   -1.423742   -0.037870   -0.000001
O   -2.800842   -0.135154   -0.000005
C   -0.742494   1.180062   0.000001
H   -1.295685   2.118976   -0.000000
C   0.653176   1.206486   0.000004
H   1.170235   2.164577   0.000006
C   1.395219   0.016120   0.000005
N   2.778587   0.038060   0.000013
H   -3.166110   0.766515   -0.000007
H   3.285260   0.907401   -0.000018
H   3.311516   -0.815438   -0.000017

II. ISOSURFACE PLOTS

III. OVERLAP BETWEEN GROUND AND EXCITED STATE WAVE FUNCTIONS

As ESMF does not enforce an orthogonality constraint, the optimized excited state wave functions are not necessarily orthogonal to ground state. In this section we report the overlap between ground and excited state wave functions: $\langle \Psi_{gs} | \Psi_{es} \rangle$. RHF determinant is used for ground state and ESMF wave function is used for excited state. In Table I we show the $\langle \Psi_{gs} | \Psi_{es} \rangle$ for all the systems studied in this paper. We find the values are at most on the order of 10^{-6} for all the systems tested. This shows that although the ESMF excited states are not rigorously orthogonal to RHF ground state, the amount of ground state contamination is negligible.

(a) Surface Value=0.003   (b) Surface Value=0.001   (c) Surface Value=0.0005
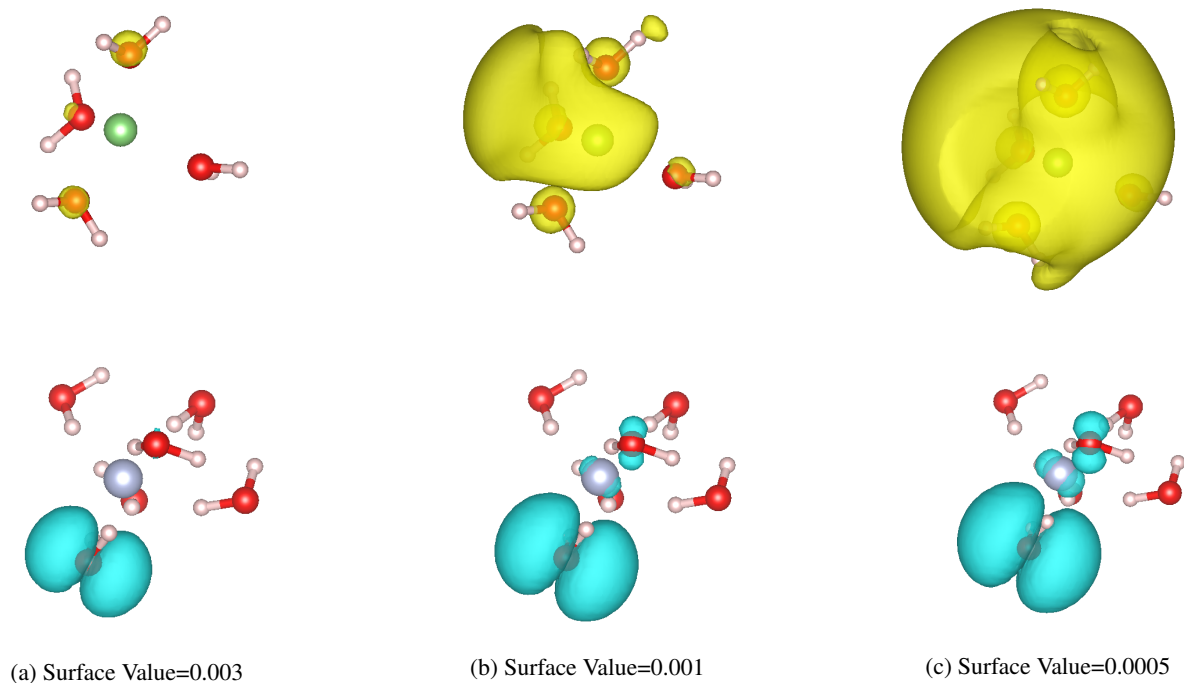
FIG. 5: CIS isosurface plots with different thresholds for the charge density changes following the charge transfer excitation that moves an electron from the lower F cluster to the upper Li cluster, with blue surfaces showing charge depletion relative to the ground state and yellow surfaces showing charge accumulation.

TABLE I: Overlap Between Ground and Excited State Wave Functions.

System	$\langle \Psi_{gs} | \Psi_{es} \rangle$
Cl^--H_2O	-2.90×10^{-6}
NH_3-F_2	1.59×10^{-7}
LiF-$(H_2O)_{10}$	2.27×10^{-6}
$NaCl$	1.17×10^{-6}
Cl^--$(H_2O)_3$	-2.20×10^{-6}
C_2H_4-C_2F_4	7.22×10^{-13}
aminophenol	8.80×10^{-8}