

# PCCCP

Physical Chemistry Chemical Physics

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: Z. Song and K. U. Lao, *Phys. Chem. Chem. Phys.*, 2026, DOI: 10.1039/D6CP01718E.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Cite this: DOI: 00.0000/xxxxxxxxxx

# Accelerating geometry optimization via Grassmann-DIIS extrapolation

Zihui Song and Ka Un Lao\*

Received Date  
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Geometry optimization in electronic structure theory requires repeated self-consistent field (SCF) calculations at every optimization step, making SCF convergence one of the primary computational bottlenecks in quantum chemistry simulations. Consequently, generating superior SCF initial guesses that reduce the number of required SCF iterations provides an effective strategy for significantly accelerating geometry optimization calculations. In this work, we develop a Grassmann extrapolation framework combined with direct inversion in the iterative subspace (G-Ext-DIIS) to generate improved density matrix initial guesses throughout geometry optimization. The proposed approach exploits the geometric structure of density matrices on the nonlinear Grassmann manifold, enabling physically consistent extrapolation while rigorously preserving the fundamental physical constraints of the density matrix. The method was systematically evaluated through calibration studies and benchmark calculations on diverse covalently bonded organic molecules and flexible noncovalent molecular clusters. Using either the overlap matrix or the core Hamiltonian matrix as the molecular descriptor, G-Ext-DIIS consistently reduces the total number of SCF iterations during geometry optimization. The performance of G-Ext-DIIS was found to be largely insensitive to the choice of density functional and basis set. For flexible cluster systems, reductions of approximately 20–30% in SCF iterations are routinely achieved, including a reduction of nearly 2000 SCF cycles for a hydroquinone clathrate cluster containing 129 atoms. Importantly, the computational overhead associated with the Grassmann extrapolation procedure is negligible compared with that of a single SCF iteration. These results demonstrate that G-Ext-DIIS provides an efficient, transferable, black-box, and computationally inexpensive strategy for accelerating geometry optimization calculations in large-scale quantum chemical simulations, making it a promising default SCF initialization scheme for future quantum chemistry software implementations.

## 1 Introduction

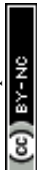
Geometry optimization is a fundamental and ubiquitous task in computational chemistry, serving as a cornerstone for modeling complex systems across chemistry, materials science, and chemical biology.<sup>1</sup> It is indispensable for characterizing molecular structures, assessing thermodynamic stability, and deciphering chemical reactivity. Beyond identifying equilibrium geometries, these routines are essential for locating transition states and mapping the intricate reaction pathways that govern chemical transformations.<sup>2</sup>

At the heart of this process is the minimization of the total electronic energy on the potential energy surface (PES) with respect to nuclear coordinates. Due to its favorable balance between accuracy and computational cost, density functional theory (DFT) has become the standard method for such evaluations. However, this optimization is inherently iterative; each

step requires a full electronic structure calculation to achieve self-consistent field (SCF) convergence, followed by the evaluation of energy gradients—and occasionally the Hessian—to guide the structural update. While robust quasi-Newton algorithms like Broyden–Fletcher–Goldfarb–Shanno (BFGS)<sup>3</sup> are ubiquitous in modern quantum chemistry packages, the primary computational bottleneck remains the sheer number of required DFT evaluations. Since each evaluation is computationally demanding, particularly for large-scale or condensed-phase systems, the total wall-clock time is directly governed by the number of iterative cycles. Therefore, developing advanced techniques to minimize these expensive DFT evaluations is essential for accelerating the discovery of stationary points on the PES.

To accelerate SCF convergence, Pulay's direct inversion in the iterative subspace (DIIS) method<sup>4,5</sup> remains the most prevalent technique and serves as the default in many modern quantum chemistry packages. Beyond acceleration algorithms, the quality of the initial guess is equally critical because it determines whether the SCF procedure converges to the desired electronic state and dictates the overall time-to-solution by influencing the

Department of Chemistry, Virginia Commonwealth University, Richmond, VA, USA. E-mail: laoku@vcu.edu



number of iterations. Various strategies have been developed to generate high-quality initial guesses, ranging from standard approaches, such as the superposition of atomic densities (SAD)<sup>6</sup> and basis set projection (BSP), to sophisticated methods, including molecular fragmentation<sup>7-10</sup> and, more recently, machine learning-based models.<sup>11-13</sup>

In the context of PES scans, geometry optimizations, and *ab initio* molecular dynamics (AIMD) simulations, efficiency is further enhanced by utilizing the converged density matrix or molecular orbital coefficients from the preceding configuration as the initial guess. More advanced schemes extrapolate the density matrix from multiple prior steps to provide a superior estimate for the new geometry,<sup>14</sup> effectively bridging the gap between successive electronic structure calculations. However, direct density matrix extrapolation is challenging. Since the density matrix in an orthonormal basis must satisfy the physical constraints of Hermiticity, particle number conservation, and idempotency, it is constrained to a nonlinear Grassmann manifold rather than a linear vector space. Consequently, as a linear combination of valid density matrices generally fails to preserve these physical constraints,<sup>15</sup> direct extrapolation is both physically and mathematically ill-defined.

To overcome these limitations, an indirect approach has been proposed for generating accurate density matrices by mapping them onto the tangent space of a Grassmann manifold.<sup>15-23</sup> Because this tangent space functions as a vector space, it is closed under both addition and scalar multiplication, which allows linear operations to be performed straightforwardly. The results are subsequently mapped back to the Grassmann manifold, ensuring the resulting density matrices remain valid and adhere to all physical constraints. These Grassmann interpolation and extrapolation schemes not only provide high quality initial guesses for SCF iterations but also facilitate the direct evaluation of expectation values for one-particle operators, such as atomic charges.<sup>21</sup>

Recently, our group introduced an approach combining a Grassmann extrapolation method with DIIS (G-Ext-DIIS) to accurately and efficiently extrapolate density matrices in electronic structure calculations.<sup>23</sup> Unlike Tikhonov-regularized G-Ext,<sup>16,18-20</sup> G-Ext-DIIS requires no tuning of regularization parameters. Moreover, its DIIS subspace is compact, numerically stable, and independent of descriptor dimensionality, system size, or basis set, ensuring both robustness and computational efficiency. This technique has been applied to electronic structure calculations during PES scans, where it significantly reduces the number of SCF iterations required for convergence or enables accurate energy predictions without the need for further SCF cycles, achieving sub-millihartree accuracy across extrapolation ranges that exceed typical geometry optimization steps.<sup>23</sup> Finally, G-Ext-DIIS demonstrates superior accuracy, variational consistency, and reliability across various basis sets compared to direct extrapolation methods and Löwdin extrapolation from nearby geometries.<sup>23</sup>

Given that computational chemistry frequently involves repeated quantum mechanical calculations on the same system, such as in geometry optimizations and AIMD simulations, we extend this robust, efficient, and transferable G-Ext-DIIS approach to geometry optimization in the present work. Specifically, ac-

curate density matrices generated by G-Ext-DIIS are employed to initialize the SCF iterations at each optimization step. This initialization strategy significantly accelerates convergence, thereby decreasing the total number of costly SCF cycles and reducing the overall wall clock time required for geometry optimization.

## 2 Computational Details

Geometry optimization aims to locate a local minimum on the PES with respect to the nuclear coordinates of a molecular system. This iterative process generates a sequence of geometries  $\{R_k\}$  such that:

$$R_{\min} \approx \arg \min_{R \in \mathbb{R}^{3N}} E(R) \quad (1)$$

where  $R_{\min}$  represents the nuclear configuration at the local minimum,  $E(R)$  denotes the total electronic energy at configuration  $R$ , and  $3N$  represents the degrees of freedom for a system containing  $N$  atoms. The optimization concludes when a set of pre-defined convergence criteria are satisfied. These thresholds typically include the maximum gradient component, the maximum atomic displacement, and the energy change between successive optimization cycles.

At each step of the geometry optimization procedure, an electronic structure calculation is performed to determine the total energy and its analytical derivatives. For mean-field methods such as DFT, this requires solving the nonlinear SCF equations, typically initialized using an initial guess of either the molecular orbital (MO) coefficients or the density matrix ( $\mathbf{P}$ ). In an orthonormal basis, the density matrix  $\bar{\mathbf{P}}$  is derived from the occupied MO coefficient  $\mathbf{C}^{\text{occ}}$  as follows:

$$\bar{\mathbf{P}} = \mathbf{S}^{1/2} \mathbf{C}^{\text{occ}} (\mathbf{C}^{\text{occ}})^{\text{T}} \mathbf{S}^{1/2}, \quad (2)$$

where  $\mathbf{S}^{1/2}$  denotes the matrix square root of the overlap matrix  $\mathbf{S}$ . This matrix  $\bar{\mathbf{P}}$  satisfies the aforementioned symmetry, idempotency, and trace properties, which constrains it to lie on the Grassmann manifold  $\mathcal{M}_{\text{Gr}}$ . However, the Grassmann manifold  $\mathcal{M}_{\text{Gr}}$  does not satisfy the closure property under linear addition; consequently, a linear combination of such density matrices generally does not result in another element of the Grassmann manifold.<sup>15</sup>

Rather than performing operations directly on the Grassmann manifold, a more effective approach involves projecting known density matrices,  $\{\bar{\mathbf{P}}_k\}$ , onto the tangent space  $\mathcal{T}_{r_0} \mathcal{M}_{\text{Gr}}$  at a reference point  $r_0$ . By mapping these matrices to tangent vectors  $\{\Gamma_k\}$  via the Grassmann logarithm,<sup>24</sup> we transform complex geometric problems into simple vector space operations. Interpolation or extrapolation can then be executed to estimate the target vector  $\Gamma_{\text{target}}$  as a linear combination of the sampled vectors:

$$\Gamma_{\text{target}} = \sum_{k=1}^n c_k \Gamma_k. \quad (3)$$

Crucially, the accuracy of the resulting  $\Gamma_{\text{target}}$  is largely invariant to the choice of reference point.<sup>15</sup> The coefficients  $\{c_k\}$  can be determined using a variety of schemes, including Lagrange interpolation,<sup>15,16,21,22</sup> Tikhonov-regularized least-squares,<sup>17-20</sup> and our recently developed DIIS-based minimization framework.<sup>23</sup> The latter determines the extrapolation coefficients within a com-



compact residual subspace whose dimensionality depends only on the number of sampling points  $n$ , leading to a numerically stable and computationally efficient formulation. Specifically, the DIIS minimization leads to a compact system of  $n + 1$  linear equations,

$$\begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1n} & 1 \\ B_{21} & B_{22} & \cdots & B_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \\ -\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (4)$$

where  $\lambda$  is a Lagrange multiplier enforcing the normalization constraint  $\sum_{k=1}^n c_k = 1$ . The elements of the symmetric error matrix  $\mathbf{B}$  are defined as

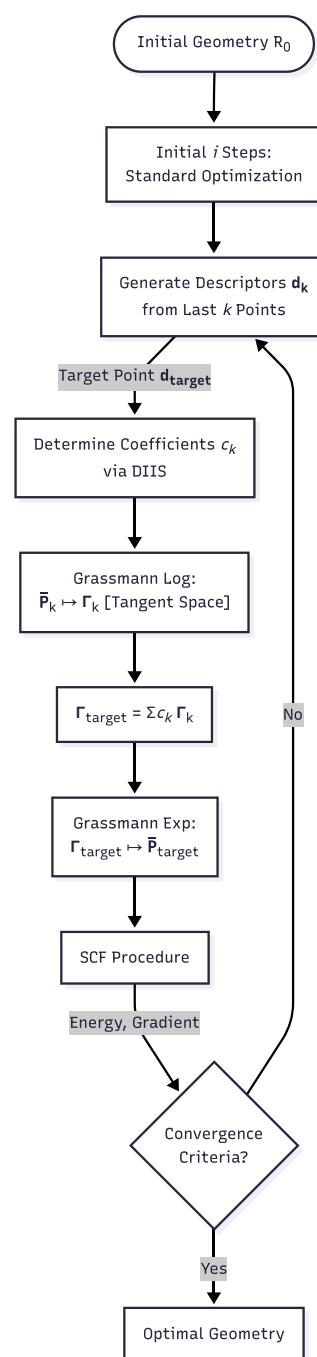
$$B_{ij} = B_{ji} = \text{Tr}((\mathbf{d}_i - \mathbf{d}_{\text{target}})^T (\mathbf{d}_j - \mathbf{d}_{\text{target}})), \quad (5)$$

where  $\mathbf{d}_i$ ,  $\mathbf{d}_j$ , and  $\mathbf{d}_{\text{target}}$  denote the molecular descriptors associated with sampling points  $i$ ,  $j$ , and the target geometry, respectively. The extrapolation coefficients  $\{c_k\}$  are obtained using singular value decomposition (SVD) together with the Moore–Penrose pseudoinverse and are subsequently employed to construct the target tangent vector  $\mathbf{\Gamma}_{\text{target}}$  according to Eq. (3). The resulting G-Ext-DIIS formulation retains the algebraic structure of conventional Pulay DIIS, but replaces the SCF residual vectors with descriptor residuals defined relative to the target geometry.

It is worth emphasizing that both the Tikhonov-regularized G-Ext formulation and the present DIIS-based minimization framework rely on the same central approximation: the extrapolation coefficients determined from the molecular descriptors are assumed to be transferable to the tangent-space representation. Consequently, the choice of molecular descriptor plays a critical role in determining the quality of the extrapolation coefficients, the accuracy of the resulting tangent vector, and ultimately the performance of the G-Ext procedure, as demonstrated in our previous study.<sup>23</sup>

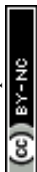
Subsequently, the resulting  $\mathbf{\Gamma}_{\text{target}}$  is mapped back to the Grassmann manifold via the Grassmann exponential,<sup>24</sup> recovering the orthonormal density matrix  $\bar{\mathbf{P}}_{\text{target}}$ . This geometric transformation ensures that all intrinsic physical constraints and geometric properties are rigorously preserved, as detailed in our previous G-Ext-DIIS study.<sup>23</sup> Finally, the orthonormal density matrix is transformed back to the nonorthogonal basis, yielding  $\mathbf{P}_{\text{target}}$ . This matrix can then be utilized either as a robust initial guess for subsequent SCF calculations or as a direct basis for predicting molecular properties.

Conventional software packages typically initialize each step of a molecular geometry optimization by reusing the converged electronic structure, such as MO coefficients, from the preceding geometry. To improve upon this practice, the recently developed Tikhonov-regularized G-Ext method predicts the current density matrix based on data from previous steps,<sup>20</sup> providing a more robust initial guess for the SCF process. However, solving the resulting least squares problem requires a regularization parameter that depends on the SCF convergence threshold,<sup>18</sup> as well as the specific molecular systems and descriptors used.<sup>20</sup> Re-



**Fig. 1** Flowchart of the geometry optimization procedure accelerated by the G-Ext-DIIS technique. The workflow initiates with  $i$  standard optimization steps for data accumulation, followed by the G-Ext-DIIS phase, which utilizes a sliding window of the last  $k$  density matrices to generate the SCF initial guess.

cently, we introduced a stable and computationally efficient DIIS-based technique to determine the interpolation/extrapolation coefficients  $c_k$ , eliminating the need for manual parameter tuning. In this work, G-Ext-DIIS is applied for the first time to generate SCF initial guesses throughout optimization cycles. By leveraging the geometry optimization history at marginal computational cost, this approach significantly reduces the number of SCF cycles per step and minimizes overall wall time.



The sequence of density matrices  $\{\mathbf{P}_k\}$  corresponding to the nuclear coordinates  $\{R_k\}$  is collected throughout the geometry optimization process. Subsequently, the G-Ext-DIIS technique is applied to generate an extrapolated density matrix  $\mathbf{P}_{k+1}$  based on this historical data, which serves as the SCF initial guess for the next geometry  $R_{k+1}$ . To determine the extrapolation coefficients  $c_k$  in Eq. (3), the overlap matrix (**S**), core Hamiltonian matrix (**H**), and Coulomb matrix (**CM**) are utilized as molecular descriptors. Notably, the **S** and **H** descriptors share the same symmetry properties as the density matrix, being invariant to translations and equivariant under rotations. In contrast, the **CM** descriptor is invariant to both translations and rotations. As demonstrated in our earlier study, descriptors that align with the symmetry of the density matrix produce more accurate coefficients, resulting in enhanced initial guesses.<sup>23</sup> In this work, all three descriptors are evaluated within the G-Ext-DIIS framework during geometry optimization. The integrated procedure for accelerating geometry optimization via G-Ext-DIIS is depicted in Figure 1. We systematically investigate the effects of the length of the initial standard optimization phase ( $i$ ) required to accumulate sufficient data for initiating G-Ext-DIIS, and the number of historical density matrices ( $k$ ) utilized in the extrapolation procedure on the overall accuracy, robustness, and stability of the G-Ext-DIIS scheme. Furthermore, we compare the performance of G-Ext-DIIS with the conventional strategy of using the converged MO coefficients from the previous optimization step as the initial SCF guess for the current geometry.

All methods were implemented in a locally modified version of Q-Chem.<sup>25</sup> In all calculations, the DIIS SCF algorithm was employed with an SCF convergence threshold of  $\tau_{\text{SCF}} = 10^{-8}$  a.u. and an integral screening threshold of  $\tau_{\text{ints}} = 10^{-14}$  a.u. Geometry optimizations were carried out using the default convergence criteria in Q-Chem, including a gradient threshold of  $300 \times 10^{-6}$  a.u., a displacement threshold of  $1200 \times 10^{-6}$  a.u., and an energy threshold of  $100 \times 10^{-8}$  a.u.

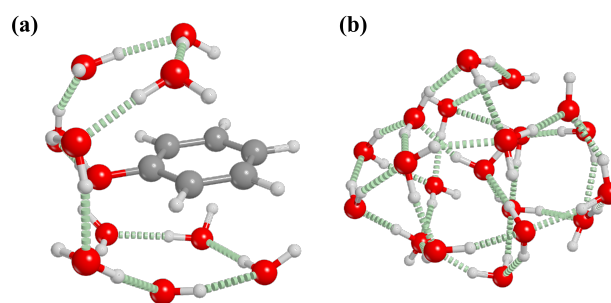
The geometry immediately preceding the target geometry is selected as the reference point for extrapolation in G-Ext-DIIS during geometry optimization. It was confirmed that the numbers of geometry optimization cycles are identical between the standard and G-Ext-DIIS-accelerated geometry optimization calculations for each system. All systems considered in this work are neutral closed-shell singlet systems.

### 3 Results and Discussion

In this section, we evaluate the performance of the proposed G-Ext-DIIS method when integrated into the standard geometry optimization procedure in Q-Chem. We first examine the impact of key computational settings and establish a consistent protocol, followed by a comprehensive benchmark across a diverse set of molecular systems.

#### 3.1 G-Ext-DIIS Settings for Geometry Optimization

To rigorously assess the effectiveness of the proposed G-Ext-DIIS method, calibration tests were first performed on two representative molecular systems, phenol( $\text{H}_2\text{O}$ )<sub>10</sub> and ( $\text{H}_2\text{O}$ )<sub>20</sub>, using ini-



**Fig. 2** Initial molecular structures of (a) phenol( $\text{H}_2\text{O}$ )<sub>10</sub> and (b) ( $\text{H}_2\text{O}$ )<sub>20</sub> used to investigate the optimal G-Ext-DIIS settings for geometry optimization.

tial structures generated by CREST at the GFN2-xTB level,<sup>26</sup> as shown in Fig. 2. The performance of the extrapolation strategy was then analyzed with respect to the choice of molecular descriptors, the length of the initial standard optimization phase ( $i$ ) required for data accumulation, and the number of historical density matrices ( $k$ ) included in the extrapolation procedure.

To determine the optimal parameters for each type of descriptor, we evaluate two strategies using these molecular systems. In the first strategy, G-Ext-DIIS is applied after  $i$  initial conventional geometry optimization steps, using a fixed number  $k$  of previous sampling points to construct the density matrix at the current geometry; if fewer than  $k$  sampling points are available, all previous points are used. This approach is denoted as ( $i, k$ ). In the second strategy, all available sampling points are employed in the G-Ext-DIIS extrapolation, denoted as ( $i, \text{all}$ ). The resulting reduction in SCF iterations at the B3LYP/def2-SVP level is compared to that obtained with the standard geometry optimization protocol, in which the MO coefficients from the preceding step are used as the initial guess for the current SCF calculation. The results, obtained using the molecular descriptor **S**, are summarized in Table 1.

In the first strategy, multiple sets of sampling parameters were examined. The number of initial optimized points was chosen as  $i = 10, 20,$  and  $25$ , while the number of historical points was varied as  $k = 2, 5, 8, 10, 15, 20,$  and  $25$ . Different molecular descriptors (**S**, **H**, and **CM**) were then applied to evaluate their impact on the total number of SCF iterations, which serves as the primary metric for assessing the performance of G-Ext-DIIS in accelerating geometry optimization (OPT). Table 1 compares the total SCF iterations for standard OPT, where previous MO coefficients are used as the initial guess, with those from G-Ext-DIIS-accelerated OPT employing the overlap matrix **S** as the descriptor. During the initial sampling stage, conventional OPT is carried out using the previous MO coefficients as the initial guess. Once G-Ext-DIIS is activated, the density matrix generated by G-Ext-DIIS is employed as the initial guess for subsequent SCF cycles. The total number of SCF cycles reported for G-Ext-DIIS in Table 1 includes those incurred during the initial sampling stage, thereby representing the full SCF cost required to complete the geometry optimization. Overall, G-Ext-DIIS consistently reduces the total number of SCF iterations across all tested parameter combinations, achieving reductions of 6–20% for phenol( $\text{H}_2\text{O}$ )<sub>10</sub> and 6–23% for ( $\text{H}_2\text{O}$ )<sub>20</sub>. These results indicate that greater savings in SCF cycles can be



**Table 1** Total number of SCF iterations for phenol( $\text{H}_2\text{O}$ )<sub>10</sub> and ( $\text{H}_2\text{O}$ )<sub>20</sub> at the B3LYP/def2-SVP level using standard geometry optimization (OPT) and G-Ext-DIIS-augmented OPT with the overlap matrix descriptor **S**. Results are reported for different combinations of the length of the initial standard optimization phase (*i*) required for data accumulation and the number of historical density matrices (*k*) used in the extrapolation. The corresponding percentage reduction (PR) in SCF iterations relative to the standard OPT procedure using molecular orbital (MO) coefficients from the preceding optimization step as the initial guess is also reported.

System	Standard	<i>i</i>	<i>k</i>	G-Ext-DIIS	SCF
	OPT			OPT	PR
Strategy-I					
Phenol( $\text{H}_2\text{O}$ ) <sub>10</sub>	2199	25	25	1750	-20.4%
		20	20	1754	-20.2%
		20	15	1781	-19.8%
		20	10	1829	-16.8%
		20	8	1859	-16.0%
		20	5	1924	-12.6%
		20	2	2063	-6.2%
		10	10	1824	-17.1%
		10	8	1847	-16.7%
		10	5	1911	-13.1%
10	2	2050	-6.8%		
( $\text{H}_2\text{O}$ ) <sub>20</sub>	2075	25	25	1606	-22.7%
		20	20	1603	-22.7%
		20	15	1618	-22.1%
		20	10	1658	-20.1%
		20	8	1701	-19.2%
		20	5	1769	-14.9%
		20	2	1954	-5.9%
		10	10	1666	-19.7%
		10	8	1656	-20.2%
		10	5	1761	-15.1%
10	2	1950	-6.0%		
Strategy-II					
Phenol( $\text{H}_2\text{O}$ ) <sub>10</sub>	2199	25	all	1628	-26.0%
		20	all	1617	-26.5%
		10	all	1603	-26.9%
		8	all	1605	-27.0%
		5	all	1606	-27.1%
		2	all	1605	-27.2%
( $\text{H}_2\text{O}$ ) <sub>20</sub>	2075	25	all	1478	-28.8%
		20	all	1464	-29.5%
		10	all	1448	-30.2%
		8	all	1447	-30.3%
		5	all	1443	-30.5%
		2	all	1437	-30.8%

obtained when a larger number of historical points is included in constructing the initial guess density.

Using the core Hamiltonian **H** as the descriptor yields comparable performance, with reductions of 6–20% for phenol( $\text{H}_2\text{O}$ )<sub>10</sub> and 6–23% for ( $\text{H}_2\text{O}$ )<sub>20</sub>, as reported in Table S1 of the Supporting Information (SI). This is consistent with our previous observation that **S** and **H** perform similarly when combined with G-Ext-DIIS for generating density matrices.<sup>23</sup> In contrast, the Coulomb matrix descriptor **CM** provides only marginal improvement (approximately 1%) and, in some cases, even increases the number of SCF iterations by about 1%.

In the second strategy, the number of initial optimized points is varied as *i* = 2, 5, 8, 10, 20, and 25, while all available historical points are used to construct the extrapolated density matrix for the current geometry. Incorporating the full set of historical points significantly enhances the performance of G-Ext-DIIS in geometry optimization, leading to reductions of 26–27% for

phenol( $\text{H}_2\text{O}$ )<sub>10</sub> and 29–31% for ( $\text{H}_2\text{O}$ )<sub>20</sub>. Notably, the best performance is achieved when the smallest number of initial sampling points (*i* = 2) is combined with the use of all historical points. This result indicates that the benefits of G-Ext-DIIS can be realized at a very early stage of the optimization, with improvements already evident from the third optimization step under the (2, all) setting.

A similar level of reduction is observed when using the core Hamiltonian **H** as the descriptor, yielding decreases of approximately 26–27% for phenol( $\text{H}_2\text{O}$ )<sub>10</sub> and 29–30% for ( $\text{H}_2\text{O}$ )<sub>20</sub>, as reported in Table S1. In contrast, the Coulomb matrix descriptor **CM** again shows no significant improvement in SCF convergence, with the maximum reduction limited to about 1%, and in some cases even leading to an increase of approximately 1% in the number of SCF cycles, as presented in Table S1. Thus, the largest reduction in SCF iterations is achieved when the first two optimization steps are performed conventionally, followed by the application of G-Ext-DIIS from the third step onward using all available historical points, with either **S** or **H** as the descriptor.

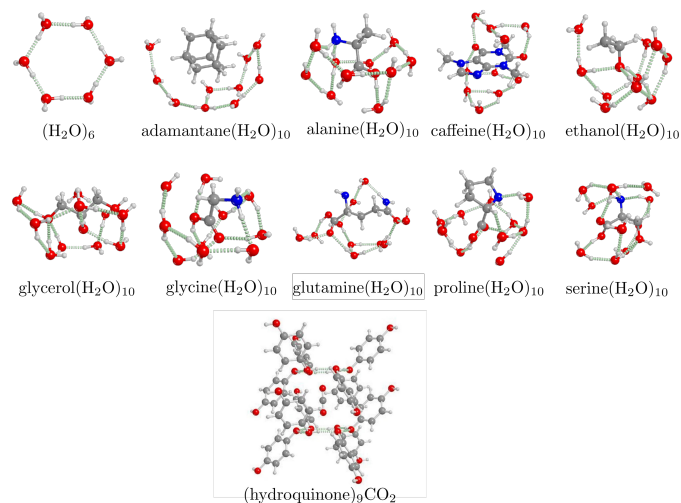
To further assess the robustness of this approach, we examined a different functional and a larger diffuse basis set. Specifically, G-Ext-DIIS was tested using various values of *i* combined with all available historical points (*k* = all), as summarized in Table S2 at the levels of PBE/def2-SVP and B3LYP/def2-TZVPPD. A similar trend in reducing the total number of SCF iterations is consistently observed. For PBE/def2-SVP, G-Ext-DIIS with **S** yields reductions of 24–25% for phenol( $\text{H}_2\text{O}$ )<sub>10</sub> and 26–28% for ( $\text{H}_2\text{O}$ )<sub>20</sub>, while using **H** results in comparable reductions of 23–24% and 26–27%, respectively. At the B3LYP/def2-TZVPPD level, G-Ext-DIIS with **S** achieves reductions of 23–24% for phenol( $\text{H}_2\text{O}$ )<sub>10</sub> and 24–28% for ( $\text{H}_2\text{O}$ )<sub>20</sub>, whereas **H** yields reductions of 23–25% and 20–28%, respectively. In contrast, the **CM** descriptor consistently increases the total number of SCF cycles across all tested cases at both levels of theory. Notably, for both PBE/def2-SVP and B3LYP/def2-TZVPPD, the largest reduction in SCF iterations is again obtained using the (2, all) setting, with improvements emerging from the third optimization step onward.

Overall, these results demonstrate that the effectiveness of G-Ext-DIIS in accelerating OPT is largely insensitive to the choice of functional and basis set. Moreover, initiating the procedure early and incorporating all available historical density matrices generally leads to more accurate and stable predictions, thereby reducing the total number of SCF iterations required for convergence.

### 3.2 Transferability of G-Ext-DIIS Across Molecular Systems

Following the calibration of the optimal parameters, the performance of the G-Ext-DIIS technique was further validated through benchmark tests on both covalently bonded organic molecules and noncovalently bonded clusters. Eight covalently bonded organic molecules containing up to 72 atoms and optimized at the B97-D/TZVP level were selected from the ISOL dataset, which was designed to include realistic organic systems commonly encountered in experimental studies.<sup>27</sup> Their structures are shown in Figure S1. In addition, the water hexamer and nine molecular



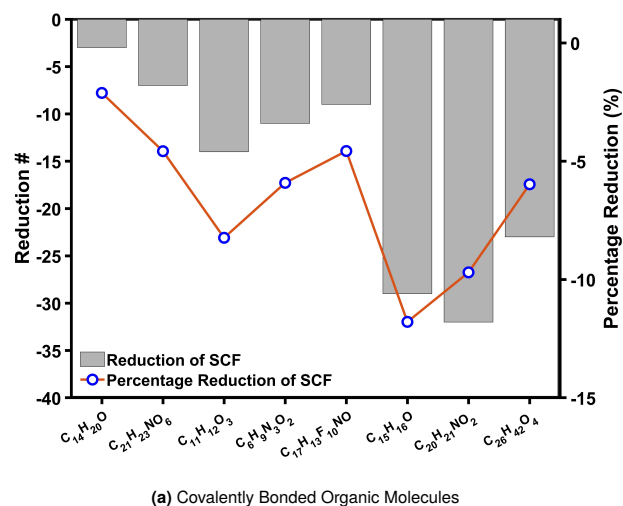


**Fig. 3** Initial molecular cluster structures used to assess the performance of G-Ext-DIIS in geometry optimization.

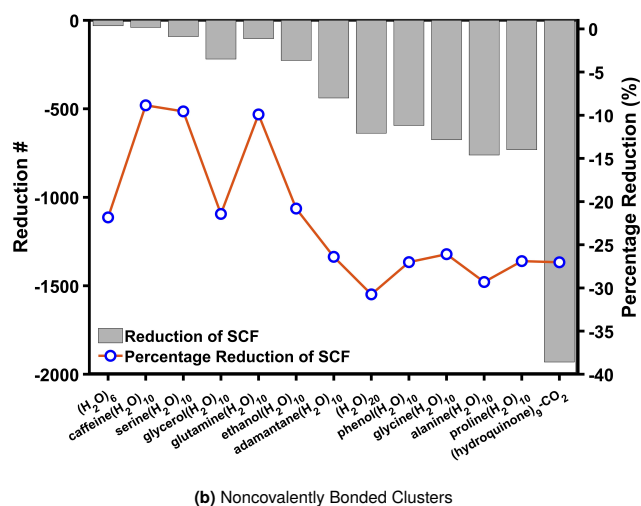
clusters were generated using CREST at the GFN2-xTB level. In these systems, adamantane, caffeine, ethanol, glycerol, and five amino acids (alanine, glycine, glutamine, proline, and serine) are solvated by ten surrounding water molecules. The largest system considered consists of a 129-atom CO<sub>2</sub>-adsorbed hydroquinone clathrate cluster constructed by extracting a molecular fragment containing nine hydroquinone molecules interacting with one CO<sub>2</sub> molecule from a solid-state PBE-D3/VTZP optimized structure.<sup>28–30</sup> The structures of all cluster systems are presented in Figure 3. Cartesian coordinates for all covalently bonded organic molecules and noncovalently bonded clusters are provided in the Supporting Information.

The performance of G-Ext-DIIS-accelerated geometry optimization using **S** as the descriptor with the (2, all) setting at the B3LYP/def2-SVP level, evaluated in terms of the reduction and percentage reduction in total SCF iterations relative to the standard optimization scheme employing the MO coefficients from the preceding step as the initial guess, is shown in Figure 4a for organic molecules and Figure 4b for molecular clusters, including phenol(H<sub>2</sub>O)<sub>10</sub> and (H<sub>2</sub>O)<sub>20</sub>. The corresponding results obtained using **H** as the descriptor are presented in Figure S2. Complete details of the SCF iteration counts and the percentage reduction in SCF cycles achieved by G-Ext-DIIS for all systems using the (2, all) setting are provided in Table S3.

For the eight covalently bonded organic molecules shown in Figure 4a, G-Ext-DIIS using **S** as the descriptor consistently lowers the total number of SCF iterations for all systems, with reductions ranging from 3 to 32 cycles, corresponding to 2–12%, and an average reduction of 6.6%. Similar performance is obtained when **H** is employed as the descriptor, yielding an average SCF reduction of 6.7%, as shown in Figure S2. In contrast, the use of **CM** leads to increased SCF iterations for every system examined, with increases ranging from 2 to 13% and an average increase of 5.9%, as reported in Table S3, further demonstrating that **CM** is not an appropriate descriptor for G-Ext-DIIS. The comparatively modest reductions obtained for these organic molecules, relative to the



(a) Covalently Bonded Organic Molecules



(b) Noncovalently Bonded Clusters

**Fig. 4** Reduction and percentage reduction (%) in total SCF iterations between standard and G-Ext-DIIS-accelerated geometry optimization calculations using the overlap matrix **S** as the descriptor with the (2, all) setting at the B3LYP/def2-SVP level for (a) eight covalently bonded organic molecules and (b) 13 noncovalently bonded clusters. The systems are arranged from left to right according to the increasing total numbers of SCF iterations in the standard geometry optimization calculations, as reported in Table S3.

approximately 30% reduction observed for phenol(H<sub>2</sub>O)<sub>10</sub> and (H<sub>2</sub>O)<sub>20</sub>, are likely attributable to two factors. First, these molecular geometries had already been optimized, albeit at a different level of theory. Second, the covalently bonded organic molecules considered here are generally less flexible and possess fewer degrees of freedom for structural rearrangement than the molecular clusters investigated in this work.

For the 13 noncovalently bonded clusters shown in Figure 4b, the performance improvement becomes substantially more pronounced. Using **S** as the descriptor, G-Ext-DIIS reduces the total number of SCF iterations for all systems by 31 to 1931 cycles, corresponding to reductions of 9–29% and an average reduction of 21.5%. The largest absolute reduction is observed for (hydroquinone)<sub>9</sub>CO<sub>2</sub>, where the total number of SCF iterations decreases from 7144 to 5213, corresponding to a reduction of



1931 cycles (27.3%) during geometry optimization. The largest percentage reduction is obtained for alanine( $\text{H}_2\text{O}$ )<sub>10</sub>, where the SCF iterations decrease from 2596 to 1835, yielding a reduction of 761 cycles (29.3%). Overall, the SCF savings achieved for molecular clusters are approximately three times larger than those observed for the covalently bonded organic molecules, primarily because these cluster systems require substantially more geometry optimization steps to reach the minimum-energy structure. Using **H** as the descriptor again provides comparable performance, with an average SCF reduction of 21.8%, as shown in Figure S2. The largest absolute reduction is again found for (hydroquinone)<sub>9</sub>CO<sub>2</sub> (1936 cycles), whereas the largest percentage reduction is observed for phenol( $\text{H}_2\text{O}$ )<sub>10</sub> (30.0%). In contrast, G-Ext-DIIS using **CM** performs significantly worse: it reduces SCF iterations for only four systems, with a maximum reduction of 6% for ( $\text{H}_2\text{O}$ )<sub>6</sub>, while increasing the number of SCF iterations for the remaining systems by as much as 10% for ethanol( $\text{H}_2\text{O}$ )<sub>10</sub>. Interestingly, although G-Ext-DIIS using **S** or **H** decreases the SCF iterations for (hydroquinone)<sub>9</sub>CO<sub>2</sub> by approximately 2000 cycles, the use of **CM** instead increases the total SCF iterations by about 200 cycles.

For completeness, Table S4 also compares the performance of G-Ext-DIIS against geometry optimization calculations that employ the superposition of atomic densities (SAD) initial guess at every optimization step. Relative to this SAD-based reference, the conventional strategy of reusing the converged MOs from the preceding optimization step reduces the total number of SCF iterations by approximately 17–28% for covalently bonded organic molecules and 11–35% for noncovalently bonded clusters. This reduction reflects the well-established benefit of orbital recycling during geometry optimization. The G-Ext-DIIS approach provides a further reduction beyond this conventional baseline. Relative to the SAD reference, the total number of SCF iterations is reduced by approximately 22–35% and 23–35% when using the **S** and **H** descriptors, respectively, for covalently bonded organic molecules, and by approximately 30–52% for noncovalently bonded clusters using either descriptor. These results demonstrate that the performance gains achieved by G-Ext-DIIS arise not merely from reusing information from previous optimization steps, but also from the descriptor-driven extrapolation of tangent vectors within the Grassmann manifold framework, which yields improved initial density matrix guesses for subsequent SCF calculations.

Overall, these results demonstrate that G-Ext-DIIS, using either **S** or **H** as the descriptor, consistently accelerates geometry optimization by reducing the total number of SCF iterations by approximately 7% on average for covalently bonded organic molecules and 22% on average for flexible noncovalent molecular clusters, relative to the default approach of reusing the converged MOs from the preceding optimization step. The results further establish that **S** and **H** serve as robust and transferable descriptors for G-Ext-DIIS, whereas **CM** is less suitable for practical geometry optimization applications. This trend is further illustrated in Figure 5, which presents the correlation map of total SCF iterations obtained from standard geometry optimization and G-Ext-DIIS-accelerated geometry optimization using the **S**, **H**, and **CM**

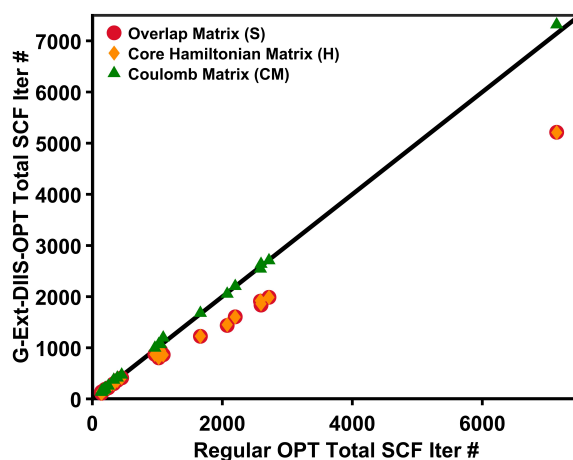


Fig. 5 Correlation map of the total number of SCF iterations between standard geometry optimization and G-Ext-DIIS-accelerated geometry optimization using the **S**, **H**, and **CM** descriptors for all 21 systems considered in this work.

descriptors for all 21 systems considered in this work. In particular, G-Ext-DIIS employing either **S** or **H** provides the largest reductions for systems requiring a greater number of geometry optimization steps to converge.

This behavior is consistent with our previous findings, which identified **S** and **H** as more effective descriptors for G-Ext-based extrapolation than **CM**.<sup>23</sup> The Coulomb matrix is constructed directly from the molecular geometry and therefore provides only an indirect representation of the underlying electronic structure. In contrast, the overlap matrix and core Hamiltonian arise directly from the electronic structure problem and consequently track changes in the self-consistent density matrix more closely throughout the optimization process. As a result, **S** and **H** provide descriptor spaces that are more strongly correlated with density matrix evolution, leading to more reliable extrapolation coefficients. Moreover, the overlap matrix, core Hamiltonian, and density matrix share the same symmetry properties, being invariant under translations and equivariant under rotations, whereas the Coulomb matrix does not possess the same transformation behavior. Since G-Ext-DIIS relies on the transferability of extrapolation coefficients determined from molecular descriptors to the tangent-space representation, descriptors whose symmetry properties and evolution more closely resemble those of the density matrix are expected to yield more transferable extrapolation coefficients. These factors collectively explain the superior performance and robustness of the **S** and **H** descriptors relative to **CM**.

It should be noted that the density matrix generated by G-Ext-DIIS is intended to provide a high-quality initial guess for the SCF procedure rather than the exact self-consistent solution at the target geometry. Although the extrapolated density matrix is generally very close to the converged solution, with typical DIIS errors of  $10^{-5}$ – $10^{-6}$  a.u. for covalently bonded organic molecules and  $10^{-6}$ – $10^{-7}$  a.u. for noncovalently bonded clusters, a small number of additional SCF iterations is still required to achieve full convergence. This is partly due to the stringent SCF convergence threshold of  $10^{-8}$  a.u. employed in the present work. Further-



more, the nonlinear nature of the DIIS-SCF procedure can give rise to residual oscillatory behavior, even when the initial density matrix is already very close to the converged solution. Consequently, several SCF iterations remain necessary to obtain the final self-consistent density matrix.

Finally, the computational cost associated with generating density matrices through G-Ext-DIIS is negligible compared with that of a single SCF iteration. For example, the extrapolation procedure requires only approximately  $10^{-4}$  s for covalently bonded organic molecules and  $10^{-2}$  s for noncovalent molecular clusters. In addition, the associated storage requirements remain modest, even for the largest system considered in this work (1320 basis functions) when all available historical density matrices are retained. The present study focuses primarily on reductions in SCF iteration counts, which provide a hardware-independent measure of the effectiveness of the proposed extrapolation procedure and facilitate direct comparisons across different molecular systems and computational environments. While direct wall-clock benchmarks would offer additional practical insight, particularly for larger systems and parallel computing environments where implementation-specific factors such as memory management, communication overhead, and I/O operations may become increasingly important, the negligible computational cost of the extrapolation step suggests that the overall efficiency gains are dominated by the reduction in SCF iterations. A systematic investigation of wall-clock performance and scaling behavior is beyond the scope of the present work and will be pursued in future studies.

Taken together, these results demonstrate that G-Ext-DIIS provides an efficient *ab initio* framework for substantially reducing the total number of SCF iterations during geometry optimization while introducing only minimal additional computational cost. Given the negligible overhead associated with the extrapolation procedure, the resulting computational savings are expected to be primarily determined by the reduction in SCF iterations. Consequently, G-Ext-DIIS represents a promising general strategy for generating initial density matrix guesses in geometry optimization calculations, and implementations for additional quantum chemistry software packages are currently under development in our group.

## 4 Conclusion

In this work, we developed and implemented G-Ext-DIIS, a Grassmann manifold-based extrapolation framework that incorporates DIIS to accelerate *ab initio* geometry optimization calculations through the generation of improved initial density matrix guesses for SCF calculations. By exploiting the geometric structure of density matrices on the Grassmann manifold, the proposed approach constructs physically meaningful extrapolated density matrices from information accumulated during previous optimization steps while rigorously preserving the fundamental constraints of the density matrix. The incorporation of the DIIS framework enables the stable and efficient determination of extrapolation coefficients within a compact residual subspace, eliminating the need for empirical regularization parameters and enhancing the robustness of the extrapolation procedure.

Comprehensive benchmark calculations demonstrate that G-Ext-DIIS can substantially reduce the computational effort associated with SCF convergence during geometry optimization across a diverse range of molecular systems. The optimal protocol was found to employ the first two optimization steps for data accumulation and to initiate G-Ext-DIIS extrapolation from the third step onward using all previously available density matrices. Using either the overlap matrix **S** or the core Hamiltonian matrix **H** as the molecular descriptor, the method consistently accelerates geometry optimization for both covalently bonded organic molecules and flexible noncovalent molecular clusters. The largest performance gains are observed for structurally flexible molecular clusters requiring many optimization steps to converge, where G-Ext-DIIS routinely reduces the total number of SCF iterations by approximately 20–30% relative to the conventional strategy of reusing the converged MOs from the preceding optimization step. For the largest system considered, (hydroquinone)<sub>9</sub>CO<sub>2</sub>, G-Ext-DIIS reduces the total SCF iteration count by nearly 2000 cycles, corresponding to a reduction of approximately 27%. The advantages of G-Ext-DIIS are even more evident when compared with the SAD-based reference, yielding reductions in total SCF iterations of up to 52% across the tested systems. In contrast, the Coulomb matrix descriptor **CM** was found to be significantly less effective, highlighting the importance of descriptors that accurately capture the evolution of the electronic structure while sharing symmetry properties with the density matrix in Grassmann-based extrapolation schemes. Finally, the performance of G-Ext-DIIS is largely insensitive to the choice of density functional and basis set, demonstrating the robustness, transferability, and broad applicability of the proposed framework.

Importantly, the computational overhead associated with G-Ext-DIIS is negligible compared with that of a single SCF iteration, making the method particularly attractive for large-scale quantum chemical calculations. Because the approach relies exclusively on information already generated during the geometry optimization procedure, it requires no additional electronic structure calculations while providing a simple, transferable, fully *ab initio*, and computationally efficient strategy for improving SCF convergence. Consequently, G-Ext-DIIS offers a practical and broadly applicable framework for reducing the overall wall time of geometry optimization calculations with essentially no additional computational cost.

Overall, these results establish G-Ext-DIIS as a practical and robust framework for accelerating geometry optimization in electronic structure theory. Owing to its black-box nature, broad transferability, and negligible computational overhead, the method is particularly well suited for large molecular clusters, biomolecular systems, and condensed-phase simulations, where repeated SCF calculations constitute a significant fraction of the overall computational cost. As a fully *ab initio* and computationally inexpensive approach, G-Ext-DIIS represents a promising default SCF initialization scheme for future quantum chemistry software implementations and large-scale electronic structure simulations.

While the present study focuses on neutral closed-shell molecular systems to establish and validate the G-Ext-DIIS framework,



the methodology itself is not inherently restricted to this class of problems. Future work will extend the benchmarking and application scope to electronically more challenging systems, including charged species, zwitterions, transition-state structures, systems with diffuse electronic densities, and open-shell molecules. Such systems often exhibit more complex electronic structure variations and more challenging SCF convergence behavior, providing a rigorous test of the robustness and transferability of the proposed approach. Beyond geometry optimization, future developments will focus on extending G-Ext-DIIS to AIMD simulations, periodic electronic structure calculations, and implementations in additional quantum chemistry software packages, thereby broadening the applicability and impact of the methodology.

### Author contributions

Zihui Song: conceptualization (contributing); data curation (lead); formal analysis (lead); investigation (lead); methodology (contributing); software (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal). Ka Un Lao: conceptualization (lead); formal analysis (contributing); funding acquisition (lead); methodology (lead); project administration (lead); supervision (lead); visualization (contributing); writing – review and editing (equal).

### Conflicts of interest

There are no conflicts to declare.

### Data availability

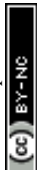
The Cartesian coordinates of all structures used in the geometry optimization calculations, along with additional details of the SCF iterations for each system, are provided in the Supporting Information (SI).

### Acknowledgements

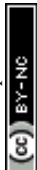
This work was supported by the National Science Foundation under CAREER Award No. CHE-2441101. High Performance Computing resources provided by the High Performance Research Computing (HPRC) Core Facility at Virginia Commonwealth University (<https://chipc.vcu.edu>) were used for conducting the research reported in this work.

### Notes and references

- H. B. Schlegel, *WIREs Comput. Mol. Sci.*, 2011, **1**, 790.
- R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Ltd, 2000.
- J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Nature, 2006.
- P. Pulay, *Chem. Phys. Lett.*, 1980, **73**, 393–398.
- P. Pulay, *J. Comput. Chem.*, 1982, **3**, 556–560.
- J. H. Van Lenthe, R. Zwaans, H. J. J. Van Dam and M. F. Guest, *J. Comput. Chem.*, 2006, **27**, 926–932.
- F. Ballesteros and K. U. Lao, *J. Chem. Theory Comput.*, 2022, **18**, 179.
- F. Ballesteros, J. A. Tan and K. U. Lao, *J. Chem. Phys.*, 2023, **159**, 074107.
- F. Ballesteros and K. U. Lao, *Phys. Chem. Chem. Phys.*, 2024, **26**, 4386.
- F. C. Y. Yu, C. Seidl, E. Palethorpe and G. M. J. Barca, *J. Chem. Theory Comput.*, 2025, **21**, 1230.
- S. Hazra, U. Patil and S. Sanvito, *J. Chem. Theory Comput.*, 2024, **20**, 4569.
- P. Stishenko, C. Qian, J. Westermayr, R. J. Maurer and A. J. Logsdail, *J. Chem. Phys.*, 2026, **164**, 064123.
- L. Zhang, P. Mazzeo, M. Nottoli, E. Cignoni, L. Cupellini and B. Stamm, *Digit. Discov.*, 2026, **5**, 1868.
- A. H. Mühlbach, A. C. Vaucher and M. Reiher, *J. Chem. Theory Comput.*, 2016, **12**, 1228–1235.
- J. A. Tan and K. U. Lao, *J. Chem. Phys.*, 2023, **158**, 051101.
- É. Polack, A. Mikhalev, G. Dusson, B. Stamm and F. Lipparini, *Mol. Phys.*, 2020, **118**, e1779834.
- É. Polack, G. Dusson, B. Stamm and F. Lipparini, *J. Chem. Theory Comput.*, 2021, **17**, 6965–6973.
- F. Pes, É. Polack, P. Mazzeo, G. Dusson, B. Stamm and F. Lipparini, *J. Phys. Chem. Lett.*, 2023, **14**, 9720–9726.
- P. Mazzeo, S. Hashem, F. Lipparini, L. Cupellini and B. Menucci, *J. Phys. Chem. Lett.*, 2023, **14**, 1222–1229.
- Z. Askarpour, M. Nottoli and B. Stamm, *J. Chem. Theory Comput.*, 2025, **21**, 1643–1651.
- J. A. Tan and K. U. Lao, *J. Chem. Phys.*, 2023, **158**, 214104.
- J. A. Tan and K. U. Lao, *Phys. Chem. Chem. Phys.*, 2024, **26**, 1436–1442.
- K. U. Lao, K. Wickramasinghe and J. A. Tan, *J. Chem. Phys.*, 2025, **163**, 144114.
- T. Bendokat, R. Zimmermann and P.-A. Absil, *Adv. Comput. Math.*, 2024, **50**, 6.
- E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, Á. Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyayev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer,



- S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, Á. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. DePrince, R. A. DiStasio, A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C.-P. Hsu, T.-C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert and A. I. Krylov, *J. Chem. Phys.*, 2021, **155**, 084801.
- 26 P. Pracht, S. Grimme, C. Bannwarth, F. Bohle, S. Ehlert, G. Feldmann, J. Gorges, M. Müller, T. Neudecker, C. Plett, S. Spicher, P. Steinbach, P. A. Wesolowski and F. Zeller, *J. Chem. Phys.*, 2024, **160**, 114110.
- 27 R. Huenerbein, B. Schirmer, J. Moellmann and S. Grimme, *Phys. Chem. Chem. Phys.*, 2010, **12**, 6940–6948.
- 28 J.-P. Torré, R. Coupan, M. Chabod, E. Pere, S. Labat, A. Khoukh, R. Brown, J.-M. Sotiropoulos and H. Gornitzka, *Cryst. Growth Des.*, 2016, **16**, 5330–5338.
- 29 W. Zhang, Z. Song, M. T. Ruggiero and D. M. Mittleman, *J. Infrared Millim. Terahertz Waves*, 2020, **41**, 1355–1365.
- 30 W. Zhang, Z. Song, M. T. Ruggiero and D. M. Mittleman, *Cryst. Growth Des.*, 2020, **20**, 5638–5643.



Data availability The Cartesian coordinates of all structures used in the geometry optimization calculations, along with additional details of the SCF iterations for each system, are provided in the Supporting Information (SI)

