Approaching the basis set limit for DFT calculations using an environment-adapted minimal basis with perturbation theory: Formulation, proof of concept, and a pilot implementation

Yuezhi Mao, Paul R. Horn, Narbe Mardirossian, Teresa Head-Gordon, Chris-Kriton Skylaris, and Martin Head-Gordon

Articles you may be interested in

CrossMark
click for updates

Approaching the basis set limit for DFT calculations using an environment-adapted minimal basis with perturbation theory: Formulation, proof of concept, and a pilot implementation

Yuezhi Mao,[1] Paul R. Horn,[1] Narbe Mardirossian,[1] Teresa Head-Gordon,[2,3,4,5] Chris-Kriton Skylaris,[6] and Martin Head-Gordon[1,5,a]

[1]*Kenneth S. Pitzer Center for Theoretical Chemistry, Department of Chemistry, University of California, Berkeley, California 94720, USA*
[2]*Department of Chemistry, University of California, Berkeley, California 94720, USA*
[3]*Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, USA*
[4]*Department of Bioengineering, University of California, Berkeley, California 94720, USA*
[5]*Chemical Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[6]*School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom*

Recently developed density functionals have good accuracy for both thermochemistry (TC) and non-covalent interactions (NC) if very large atomic orbital basis sets are used. To approach the basis set limit with potentially lower computational cost, a new self-consistent field (SCF) scheme is presented that employs minimal adaptive basis (MAB) functions. The MAB functions are optimized on each atomic site by minimizing a surrogate function. High accuracy is obtained by applying a perturbative correction (PC) to the MAB calculation, similar to dual basis approaches. Compared to exact SCF results, using this MAB-SCF (PC) approach with the same large target basis set produces <0.15 kcal/mol root-mean-square deviations for most of the tested TC datasets, and <0.1 kcal/mol for most of the NC datasets. The performance of density functionals near the basis set limit can be even better reproduced. With further improvement to its implementation, MAB-SCF (PC) is a promising lower-cost substitute for conventional large-basis calculations as a method to approach the basis set limit of modern density functionals. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4959125]

I. INTRODUCTION

Kohn-Sham density functional theory[1–3] (KS-DFT) has become the most widespread electronic-structure method because of its reasonable balance between accuracy and computational cost. Functionals using the generalized gradient approximation (GGA)[4,5] are usually regarded as the simplest that can give acceptable accuracy for chemistry. To overcome the plague of self-interaction error, new variables have been introduced, leading to meta-GGA,[6–8] global hybrid (GH),[9] and range-separated hybrid (RSH)[10,11] functionals. At the same time, a variety of models have been developed to account for van der Waals (vdW) interactions within DFT,[12,13] including the empirical DFT-D methods[14–16] and nonlocal correlation (NLC) functionals (e.g., vdw-DF-10,[17] VV10[18]). Most recently, two combinatorially designed functionals were developed: ωB97X-V[19] (RSH+VV10) and B97M-V[20] (meta-GGA+VV10), which demonstrated impressive accuracy on both thermochemistry (TC) and non-covalent interactions (NC), with an accessible complete basis set (CBS) limit, and low grid sensitivity.

With finite atomic orbital (AO) basis sets,[21] one prerequisite for attaining such accuracy is to approach the CBS limit. This issue has been carefully investigated,[22–25] but

is often neglected in practical applications, as exemplified by the prevalence of the "B3LYP/6-31G(d)" model chemistry. A basis set of *at least* triple-ζ and *preferably* quadruple-ζ quality is often required by hybrid functionals (e.g., B3LYP[9,26,27]) to obtain adequately converged thermochemistry results. Even for the semi-local B97M-V functional, the acceptable alternatives to aug-cc-pVQZ (aQZ)[28,29] (which almost represents the CBS limit) are still of least triple-ζ quality. Turning to the evaluation of NCs, a similar study on the A24[30] and S66[31,32] complexes indicates that augmented triple-ζ basis sets (e.g., aug-cc-pVTZ (aTZ), def2-TZVPD[33]) are in general required by B97M-V to properly converge the binding energies. Their double-ζ counterparts (e.g., aug-cc-pVDZ (aDZ), def2-SVPD) should only be carefully used with counterpoise (CP) corrections.[34] For ωB97X-V, the requirement on basis set qualities might be even higher due to the slower basis set convergence of functionals that contain exact exchange.

Figure 1 demonstrates the basis set convergence of several modern density functionals in terms of their root-mean-square deviations (RMSDs) for the G2 set[35] (atomization energies of 148 neutral molecules, whose reference values are taken from Ref. 36). Apart from B97M-V, using aQZ instead of aTZ for the other four functionals reduces their RMSDs by 0.8–1.4 kcal/mol, including the semi-local B97-D[15] functional (GGA). Using aDZ yields very poor accuracy (usually over 10 kcal/mol RMSDs) for all these functionals, which defeats

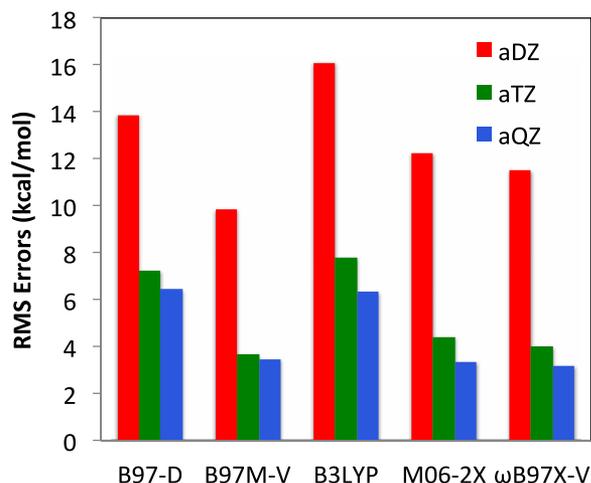a)Electronic mail: mhg@cchem.berkeley.edu

FIG. 1. Basis set dependence of the root-mean-square (RMS) errors for G2 atomization energies against the reference values from Ref. 36. Data for five functionals are collected: B97-D (GGA), B97M-V (meta-GGA), B3LYP (GH), M06-2X (GH),[39] and ωB97X-V (RSH). Dunning's augmented correlation-consistent basis sets (D, T, Q) are used to systematically approach the CBS limit.

the purpose of using state-of-the-art density functionals. One way to tackle this problem is by directly training a functional in a small basis, such as the EDF1 functional,[37] which was parameterized at the 6-31+G(d) level. By relying on cancellation between functional error and basis set error, the transferability of these methods are often limited, and further empirical corrections seem necessary to achieve adequate accuracy for relative energies.[38]

Each self-consistent field (SCF) cycle of a KS-DFT calculation involves two computationally demanding steps: (1) Fock matrix construction with a given density and (2) Fock matrix diagonalization to update the density. For fixed system size, the computational cost of the Fock build scales as $O(n^4)$ with respect to the basis size (n) when conventional AO algorithms are used, and the cost of the diagonalization step scales as $O(n^3)$. This steep cost increase inhibits large basis sets (e.g., those of quadruple-ζ size) from being routinely employed in DFT calculations. It should be noted that the scaling of cost vs. basis size is largely independent of the development of linear scaling (with system size) Fock matrix build algorithms[40–47] and many diagonalization replacements.[48,49] Moreover, near-complete basis sets are not favored by linear-scaling algorithms, especially when diffuse functions are included, since matrix element sparsity is diminished and the overlap matrix starts to be ill-conditioned (note that optimization of Gaussian basis sets up to triple-ζ quality with much reduced condition numbers while maintaining condensed phase accuracy has been reported[50]), which in turn destroys the sparsity of the density matrix.[51,52]

One successful strategy to make large basis KS-DFT calculations more tractable is to compute the full Coulomb (**J**) and exchange (**K**) matrices more efficiently by approximating two-electron repulsion integrals (ERIs) with the aid of auxiliary basis functions or grid points. The resolution-of-the-identity (RI) method[53–55] expands the product of AO function pairs with a preoptimized auxiliary basis. RI algorithms do not improve the system-size scaling unless local fit regions

are applied,[56,57] but they reduce the basis set size scaling from $O(n^4)$ to $O(n^3)$. Therefore, state-of-the-art RI algorithms (e.g., MARI-J,[58] PARI-K,[59] and occ-RI-K[60]) can speed up Fock matrix constructions in large basis sets significantly for small- to medium-sized systems, while retaining numerical accuracy. The diagonalization step is unaffected.

A second successful approach to accelerating large basis calculations is to perform the iterative SCF procedures in a primary (small) basis and then approximate the secondary (target) basis results by utilizing perturbation theory. This idea was introduced for post-SCF methods (e.g., MP2),[61,62] and was then developed for SCF methods, including the dual-basis SCF (DB-SCF) method developed by Head-Gordon and co-workers,[63–67] and the "Hartree-Fock/Density Functional Perturbation Corrections" (HFPC/DFPC) scheme proposed by Deng and Gill.[68–70] With a careful choice of primary/secondary basis set pairing, these methods can provide satisfactory accuracy for both TC[63,69,70] and NC[67] with significantly reduced computational costs (roughly 10 times faster), although system-size scaling remains unchanged. One limitation is the need to develop and validate the basis set pairings,[67] which determines accuracy and speedup. Furthermore, as the secondary basis approaches the CBS limit, the size of the primary basis needed to achieve a given accuracy also increases: for instance the optimized primary basis for cc-pVQZ is roughly of cc-pVTZ size.[64]

A related approach is the use of small adaptive basis sets. The idea of encoding chemical environment information into atomic/quasiatomic basis functions to understand chemical bonding dates back to early tools,[71–76] as well as some more contemporary methods.[77–82] Apart from interpretive purposes, the merits of utilizing small adaptive bases in KS-DFT calculations have been recognized with the development of fast (especially linear-scaling) SCF algorithms, leading to renewed interest in the concept of "polarized atomic orbitals" (PAOs), first put forward by Adams in the 1960s.[83,84] These adaptive sets usually have very tiny (often minimal) spans, which leads to vastly fewer variational degrees of freedom. In addition, an adaptive basis constructed with spatial confinement contributes to a well-conditioned overlap matrix, which is a property favored by $O(N)$ scaling methods.

The Adams PAO scheme treats atoms in a molecule as fragments and solves projected equations self-consistently on each of them, which is similar, in spirit, to projected SCF methods using fragment-localized, non-orthogonal molecular orbitals (MOs) to evaluate intermolecular interactions (SCF-MI).[85–87] In practice this scheme only works for weakly interacting atoms (e.g., rare-gas clusters) or ionic compounds (e.g., LiH, NaCl).[88] Later, the PAO approach was recast to form a minimal atom-centered adaptive basis as an atom-blocked contraction of the secondary basis functions on each atom.[89] The molecular energy is minimized simultaneously with respect to the atom-blocked contraction coefficients and the density matrix in the adaptive basis.[89] The PAO-SCF energy can be improved using perturbation theory,[90] similar to the dual basis approaches discussed above. The minimal rank of the PAO basis and its atomic locality makes it promising for linear-scaling algorithms,[52] but the "double" optimization problem is challenging and often causes convergence problems.

Significant progress on tractable adaptive basis schemes for KS-DFT has been made in the condensed matter physics community. Similar to the aforementioned PAO approach, Ozaki and Kino[91,92] and others[93,94] used numerical solutions to the atomic Kohn-Sham equations as the secondary basis, and a scheme resembling geometry optimization to obtain the adaptive basis. The CONQUEST program[95] forms local "support functions" (an adaptive basis) from either functions akin to plane waves[96] or pseudo-atomic orbitals.[97] The ONETEP package[98,99] forms non-orthogonal generalized Wannier functions (NGWFs)[100] as the environment-adapted basis, which is a linear combination of periodic sinc functions confined in an atom-centered sphere of fixed radius. The NGWFs are efficiently optimized via a preconditioned conjugate-gradient algorithm.[101]

Recently, adaptive basis schemes that do not require the global Hamiltonian or density matrix have been presented. The localized filter diagonalization (LFD) method builds an adaptive basis on-the-fly by contracting the atomic Gaussian functions within a local region, with contraction coefficients determined by diagonalizing a block of the Hamiltonian matrix corresponding to that region.[102,103] This algorithm has also been used to construct multisite local support functions,[104] and the general philosophy has been extended by Lin *et al.*,[105] including another model with more rigorous optimization.[106] While clearly promising, to our knowledge, the accuracy and performance of these methods on chemical systems have not been systematically assessed yet.

In the present work, we propose an inexpensive version of the PAO method (Sec. II). Instead of energy-optimizing the adaptive basis and density simultaneously, an inexpensive

FIG. 2. Illustration of the overall MAB-SCF (PC) procedure, which includes four steps in total.

converged SCF solution (density matrix) computed in a projected reference basis (PRB) is utilized as a reference (Sec. II A). Based on this reference, an atom-centered minimal adaptive basis (MAB) is found by minimizing a judiciously chosen surrogate function (Sec. II B), which only involves computationally inexpensive steps. The converged MAB is then used as the basis set for another SCF calculation, which requires small computational effort as well while providing comparable accuracy to PAO-SCF. Perturbation corrections (PCs) can be applied to the MAB-SCF energy for obtaining the desired accuracy (Sec. II E). The overall MAB-SCF (PC) procedure is illustrated in Figure 2. Details about the pilot implementation of this scheme and proof-of-concept calculations are summarized in Sec. III. As an approximate SCF method, its accuracy is assessed on a broad range of TC and NC datasets that is presented in Secs. IV and V.

II. THEORY

The notation used throughout this paper is as follows: $|\omega\rangle$: generic atomic basis functions; $|\psi\rangle$: generic molecular orbitals; capital Roman indices X, Y, ...: atomic centers; lowercase Greek letters μ, ν, λ, ...: secondary (large) AO basis indices; α, β, γ, ...: primary (PRB or MAB) AO basis indices; lowercase Romans i, j, k, ...: occupied MO indices; a, b, c, ...: virtual MO indices; p, q, r, ...: generic MO indices. For introducing the MAB optimization scheme, i, j, ... are also employed to denote the vectors retained in the MAB subspace, a, b, ... for the vectors in MAB's complementary subspace, and p, q, ... for the generic ones, which is analogous to the partitioning of MO space in SCF. The different basis sets that are involved in this work and the relationships between them are summarized in Table I.

Unless otherwise specified, matrices in the secondary AO basis are denoted by bold Roman letters (e.g., $\mathbf{F}$, $\mathbf{P}$), while those in the primary basis are by bold calligraphic Roman letters (e.g., $\mathcal{F}$, $\mathcal{P}$). To concisely show the character of quantities within a nonorthogonal basis, tensorial notation will be used in the derivation, i.e., covariant (subscript) and contravariant (superscript) indices are distinguished, following Ref. 107 and the appendix of Ref. 89. For instance, a matrix element denoted by $B_{X\alpha}^{X\mu}$ indicates that matrix $\mathbf{B}$ has rows corresponding to contravariant secondary basis functions and columns corresponding to covariant primary basis functions, and these basis functions belong to the same atomic center X. Einstein summation convention is applied for contractions between contravariant and covariant indices, except for summations over different atomic centers, which will be written out explicitly.

A. SCF in the projected reference basis (PRB)

The search for the MAB described in Sec. II B requires an inexpensively calculated reference density matrix in the secondary (target) basis. A converged SCF solution in a small PRB serves this purpose. The PRB is constructed by projecting the reference basis functions, $\{|\omega_{A\bar{\alpha}}\rangle\}$, into the space spanned by the secondary basis on each atom:[66]
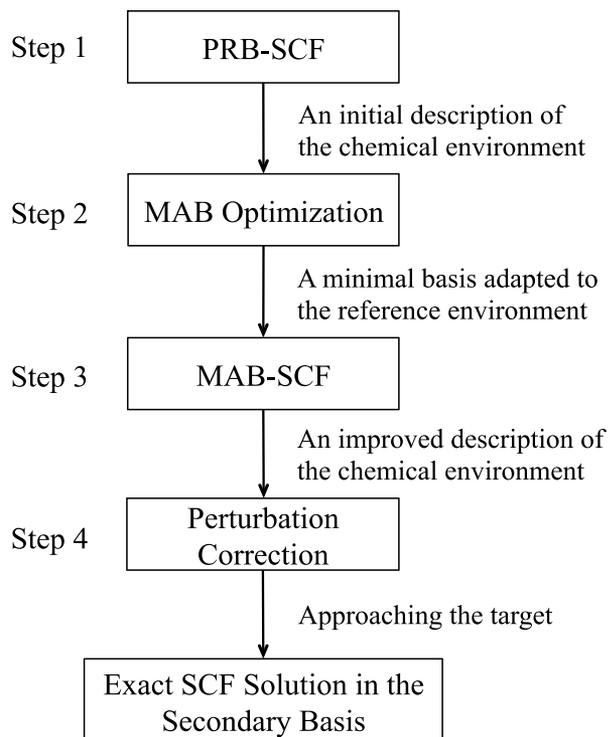
TABLE I. Summary of different level of atomic basis sets involved in the procedure of MAB-SCF.

	Name	Expression	Definition (origin)
Standard basis	Reference basis (RB)	$\lvert\omega_{A\tilde{\alpha}}\rangle$	Standard double-ζ basis sets (e.g., 6-31+G(d))
	Secondary basis (the target)	$\lvert\omega_{A\mu}\rangle$	Standard basis sets that are close to the CBS limit (e.g., aQZ)
Primary basis	Projected reference basis (PRB)	$\lvert\omega_{A\alpha}\rangle = \lvert\omega_{A\mu}\rangle (B_{ref})^{A\mu}_{A\alpha}$ (Eq. (1))	On-atom projection of RB into the secondary basis
	Minimal adaptive basis (MAB)	$\lvert\omega_{Ai}\rangle = \lvert\omega_{A\mu}\rangle B^{A\mu}_{Ai}$, $\mathbf{B} = \mathrm{argmin}\,E(\mathbf{B})^{a}$	Energetically optimized contraction of secondary basis functions on each atom

$^{a}E(\mathbf{B})$ is the surrogate energetic objective function used for MAB optimization (defined by Eq. (10)).

$$\lvert\omega_{A\alpha}\rangle = \lvert\omega_{A\mu}\rangle \left(S_A^{-1}\right)^{A\mu A\nu} (S_{12})_{A\nu A\alpha}$$
$$= \lvert\omega_{A\mu}\rangle (B_{ref})^{A\mu}_{A\alpha}. \tag{1}$$

Here, $(S_{12})_{A\nu A\alpha} = \langle\omega_{A\nu}\lvert\omega_{A\tilde{\alpha}}\rangle$ is the overlap between the (unprojected) reference basis and the secondary basis functions, while $\mathbf{S}_A$ is the overlap metric of secondary basis functions on atom A. Throughout this paper, $\mathbf{B}$ is used to denote atom-blocked matrices containing contraction coefficients of the secondary basis functions on each atomic site, which defines a primary basis.

Since the reference basis is small (to be specified later) while the secondary basis is close to the CBS limit, the block in $\mathbf{B}_{ref}$ will be very sparse since the contraction coefficients for the high angular momentum components of the secondary basis all vanish during the projection procedure. At this stage, an SCF calculation is performed in the PRB, by solving the following generalized eigenvalue equation:

$$\mathcal{F}C = \mathcal{S}C\epsilon, \tag{2}$$

where $\mathcal{F}$ and $\mathcal{S}$ have the dimension of the PRB, and they can be transformed from their counterparts in the secondary basis using the $\mathbf{B}_{ref}$ matrix,

$$\mathcal{F} = \mathbf{B}_{ref}^T \mathbf{F} \mathbf{B}_{ref}, \quad \mathcal{S} = \mathbf{B}_{ref}^T \mathbf{S} \mathbf{B}_{ref}. \tag{3}$$

In reverse, the PRB density matrix, $\mathcal{P}$, can be projected into the secondary basis via the following transformation:

$$\mathbf{P}_{ref} = \mathbf{B}_{ref} \mathcal{P} \mathbf{B}_{ref}^T. \tag{4}$$

Since the PRB is an exact subset of the secondary basis, no information in $\mathcal{P}$ is lost upon projection into the latter (Eq. (4)), i.e., $\mathbf{P}_{ref}$ and $\mathcal{P}$ contain the same information about the chemical environment (this is *not* true if $\mathcal{P}$ is optimized in the reference basis directly without doing the projection). We call this special property of $\mathbf{P}_{ref}$ "PRB-representability." The final PRB density matrix $\mathcal{P}$ therefore becomes the reference used in the search for the MAB.

B. Finding the minimal adaptive basis (MAB)

Sec. II A employs a basis defined by a fixed atom-blocked transformation (the PRB) and converges a density matrix in it. With the fixed $\mathbf{P}_{ref}$ in hand, our goal now is to optimize an energy-like function with respect to a variable $\mathbf{B}$ matrix that defines the MAB. (Note: In the following discussion $\mathbf{B}$ *exclusively* denotes the MAB coefficients.) Since a single

diagonalization minimizes $\mathrm{Tr}[\mathbf{P}\mathbf{F}]$ for a chosen number of electrons[64] when $\mathbf{F}$ is given, we shall, by analogy, minimize $\mathrm{Tr}[\tilde{\mathbf{P}}\mathbf{F}]$, where $\tilde{\mathbf{P}}$ is a "MAB-representable" density matrix in the secondary basis,

$$\tilde{\mathbf{P}} = \mathbf{B}\mathcal{D}\mathbf{B}^T. \tag{5}$$

$\mathcal{D}$ is a density matrix in the MAB that derives from the fixed $\mathbf{P}_{ref}$.

However, the MAB has smaller rank than the PRB, and the spaces spanned by them are rather different, so $\mathbf{P}_{ref}$ will not be exactly representable by the MAB. There exist many possible ways to construct $\tilde{\mathbf{P}}$. We choose to project $\mathbf{P}_{ref}$ into the space spanned by the MAB first, which gives $\mathcal{D}$, then transform it back into the secondary basis,

$$\mathcal{D} = (\sigma^{-1})\mathbf{B}^T \mathbf{S} \mathbf{P}_{ref} \mathbf{S} \mathbf{B}(\sigma^{-1}), \tag{6}$$
$$\tilde{\mathbf{P}} = \mathbf{B}(\sigma^{-1})\mathbf{B}^T \mathbf{S} \mathbf{P}_{ref} \mathbf{S} \mathbf{B}(\sigma^{-1})\mathbf{B}^T, \tag{7}$$

where σ refers to the overlap metric of the MAB,

$$\sigma = \mathbf{B}^T \mathbf{S} \mathbf{B}. \tag{8}$$

Recognizing that the projector into the MAB space is

$$\mathbf{R} = \mathbf{B}(\sigma^{-1})\mathbf{B}^T, \tag{9}$$

our surrogate energetic objective function becomes

$$E = \mathrm{Tr}\left[\mathbf{R}\mathbf{S}\mathbf{P}_{ref}\mathbf{S}\mathbf{R}\mathbf{F}\right]. \tag{10}$$

For brevity, in the following derivation we use $\mathbf{P}$ instead of $\mathbf{P}_{ref}$ for the *fixed* reference density matrix.

We note that the MAB-representable density matrix $\tilde{\mathbf{P}}$ usually does not contain exactly the right electron count. While the exact N_{elec} is given by $\mathrm{Tr}[\mathbf{P}\mathbf{S}]$, utilizing the idempotency of $\mathbf{R}$ (based on Eqs. (8) and (9), it is straightforward to derive $\mathbf{R}\mathbf{S}\mathbf{R} = \mathbf{R}$), we have

$$\mathrm{Tr}[\tilde{\mathbf{P}}\mathbf{S}] = \mathrm{Tr}[\mathbf{R}\mathbf{S}\mathbf{P}\mathbf{S}\mathbf{R}\mathbf{S}]$$
$$= \mathrm{Tr}[\mathbf{R}\mathbf{S}\mathbf{P}\mathbf{S}] \neq \mathrm{Tr}[\mathbf{P}\mathbf{S}]. \tag{11}$$

The inequality arises because the reference density matrix $\mathbf{P}$ is usually not MAB-representable.

Gradient-based optimization can locate the optimal $\mathbf{B}$ as the minimizer of Eq. (10). The initial guess for the MAB (and its orthogonal complement in the span of the secondary basis functions on each atomic site, which is denoted by $\mathbf{V}$) is obtained by diagonalizing atomic blocks of the reference density matrix, appropriately transformed[75]

as $\mathbf{P}'_A = \mathbf{X}_A^T \mathbf{S}_A \mathbf{P}_A \mathbf{S}_A \mathbf{X}_A$, where $\mathbf{S}_A$ is the overlap matrix of the secondary basis functions on A, and $\mathbf{X}_A$ is the canonical orthogonalizer for them. With $\mathbf{U}_A$ representing the eigenvectors of $\mathbf{P}'_A$, the initial $\mathbf{B}$ and $\mathbf{V}$ are set to

$$(B_A^{\text{init}})^{A\mu}_{Ai} = (X_A)^{A\mu}_{Ap}(U_A)^{Ap}_{Ai}, \tag{12a}$$

$$(V_A^{\text{init}})^{A\mu}_{Aa} = (X_A)^{A\mu}_{Ap}(U_A)^{Ap}_{Aa}, \tag{12b}$$

where $(U_A)^{Ap}_{Ai}$ denotes the eigenvectors corresponding to the m_A largest eigenvalues of $\mathbf{P}'_A$ (m_A is the rank of minimal basis for atom A), and $(U_A)^{Ap}_{Aa}$ denotes the remaining eigenvectors. This gives the initial partitioning of the Hilbert space that can be represented symbolically as

$$\mathbb{I}_\mathbf{A} = \mathbf{X_A} = \mathbf{B_A} \oplus \mathbf{V_A},$$
$$\mathbb{I} = \bigoplus_{A=1}^{N_A} \mathbb{I}_\mathbf{A}. \tag{13}$$

Since the MAB functions (and the complementary ones) are constructed by on-site contractions of the secondary basis, the variables that parameterize the MAB are intra-atomic orbital rotations. Akin to Ref. 52, a single on-block unitary transform is parameterized by the exponential of an antisymmetric matrix,[108] which ensures that the updated atomic orbitals stay on the same manifold,

$$[B(\theta)]^{X\mu}_{Xi} = C^{X\mu}_{Xr} \exp\left(\theta^{Xr}_{Xi}\right), \tag{14}$$

where $C^{X\mu}_{Xr}$ denotes the union of the MAB and the complementary functions on atom X. To enforce antisymmetry of $\boldsymbol{\theta}$, it is further parameterized by $\boldsymbol{\Delta}$ which contains all the independent variables,

$$\theta^{Xr}_{Xi} = \left[\Delta^{XrXs} - (\Delta^\dagger)^{XrXs}\right]\sigma_{XsXi}$$
$$= \left[\Delta^{XrXs} - \Delta^{XsXr}\right]\sigma_{XsXi}. \tag{15}$$

The desired gradient, evaluated at $\boldsymbol{\Delta} = \mathbf{0}$, is

$$\frac{\partial E}{\partial \Delta^{ZpZq}} = \sum_{X,Y} \frac{\partial E}{\partial R^{X\mu Y\nu}} \frac{\partial R^{X\mu Y\nu}}{\partial \Delta^{ZpZq}}$$
$$= \sum_{X,Y} (SPSRF + FRSPS)_{Y\nu X\mu} \frac{\partial R^{X\mu Y\nu}}{\partial \Delta^{ZpZq}}. \tag{16}$$

E is invariant with respect to orbital rotations within the MAB space ($p = i, q = j$), or within the space of complementary excluded vectors ($p = a, q = b$), as these rotations leave $\mathbf{R}$ unchanged. Therefore, the non-zero gradient comes only from variations of Δ^{ZiZa}. Using the identities

$$\frac{\partial B^{X\mu}_{Xj}}{\partial \Delta^{ZiZa}} = -V^{X\mu}_{Xa}\delta^X_Z \sigma_{XiXj} \tag{17}$$

and

$$\frac{\partial(\boldsymbol{\sigma}^{-1})}{\partial\boldsymbol{\Delta}} = -(\boldsymbol{\sigma}^{-1})\frac{\partial\boldsymbol{\sigma}}{\partial\boldsymbol{\Delta}}(\boldsymbol{\sigma}^{-1}), \tag{18}$$

the desired gradient expression is given by

$$\frac{\partial E}{\partial \Delta^{ZiZa}} = -2\sigma_{ZiZj}\left[(\boldsymbol{\sigma}^{-1})B^T G(I - RS)\right]^{Zj}_{Z\mu} V^{Z\mu}_{Za}, \tag{19}$$

where, for brevity, $\mathbf{G} = \partial E/\partial\mathbf{R}$ as defined in Eq. (16). More details about the derivation of Eq. (19) is provided in Appendix A.
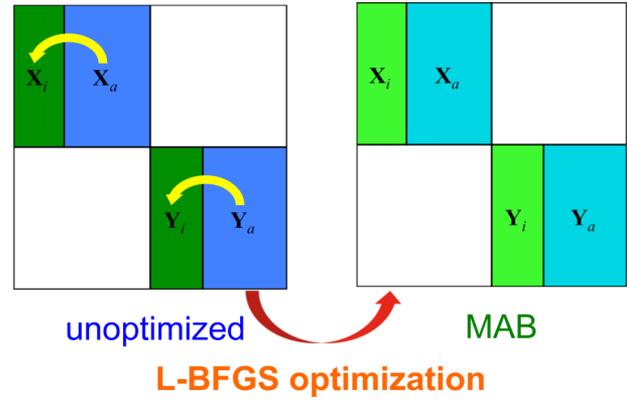

FIG. 3. Illustration of the MAB optimization procedure. Atom centers X and Y correspond to two distinct diagonal blocks of the $\mathbf{B}$ matrix. X_i, Y_i represent the MAB functions on these sites, and X_a, Y_a represent the excluded basis functions.

Once the gradient at the current position is computed, the optimization algorithm will generate a new step ($\boldsymbol{\Delta}$) based on it (and the previous gradients and steps). The equations for the exponential transformation were derived in Ref. 108. The update for the MAB can be represented as

$$\mathbf{B}^{(n)} = \mathbf{B}^{(n-1)}\begin{pmatrix} \mathbf{U}\cos\boldsymbol{p}^{1/2}\mathbf{U^T} \\ -\boldsymbol{\Delta}^\dagger\mathbf{U}\boldsymbol{p}^{1/2}\sin\boldsymbol{p}^{1/2}\mathbf{U^T} \end{pmatrix}. \tag{20}$$

$\mathbf{U}$ and $\boldsymbol{p}$ stand for eigenvectors and eigenvalues of the matrix quantity $\boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger$, respectively, and note that the unitary transformations are atom-blocked operations. When the iterative optimization converges, $\mathbf{B}$ represents a minimal basis energetically adapted to the chemical environment described by the reference density matrix (from PRB-SCF). Figure 3 illustrates the MAB optimization procedure. Finally, we note that for unrestricted cases, the MABs for α and β electrons are optimized separately (they are completely decoupled), using the same objective function form.

With the MAB defined, a converged SCF solution can be obtained in this basis. The SCF energy in the fixed MAB will be an approximation to the energy evaluated by PAO-SCF, which directly minimizes the SCF energy with respect to the generators of the MAB as well as the variables defining the density matrix. These two approaches will be compared in Sec. IV.

C. Modified definition of the minimal adaptive basis

The size of a minimal basis only depends on the principal quantum number of the atom's valence shell, since a complete set of angular momentum functions are needed to fulfill the requirement of spatial isotropy. This definition usually works very well, but there are two types of exceptions. First, in some cases, the standard rank of the minimal basis includes redundant functions. For example, the minimal basis of lithium ($n = 2$) consists of 5 functions, although only two of them are required to describe its $1s^2 2s^1$ configuration. The same applies to many other electron-deficient species like cations and radicals. The presence of redundant functions causes difficulties in converging the MAB optimization procedure.

ALGORITHM 1. Algorithm that flexibly adjusts (truncates or expands) the size of the MAB for special systems, based on quantities already computed when generating the initial guess for the MAB optimization.

---

for *atom* $A = 1, N_A$ **do**

    Compute $\mathbf{P}'_A = \mathbf{X}_A^T \mathbf{S}_A \mathbf{P}_A \mathbf{S}_A \mathbf{X}_A$

    Diagonalize $\mathbf{P}'_A$, get its eigenvalues E_A

    $i = 1, N_{Sig} = 0$

    while $E_A(i) \geq thresh1$ **do**

        $i = i + 1, N_{Sig} = N_{Sig} + 1$

    $m_A = N_{min}^A$

    if $N_{Sig} < N_{min}^A$ **then**

        $m_A = N_{Sig}$ (truncated)

    else

        $i = N_{min}^A + 1$

        if $E_A(N_{min}^A) > 0.25$ **then**

            $thresh2/ = 2$

            (considering large absolute values)

        while $E_A(i)/E_A(N_{min}^A) \geq thresh2$ **do**

            $m_A = m_A + 1$ (augmented)

            $i = i + 1$

            if $i > dim(E_A)$ **then**

                break

    m_A becomes the rank of MAB on atom A

Second, in some cases, the standard rank of the minimal basis is too small to accurately describe the bonding. Examples include some hypervalent molecules (e.g., SO_3 and ClF_3), and, occasionally, molecular anions. In such cases, the standard rank of the MAB will lead to larger errors in the resulting molecular energies, which can be greatly reduced if a certain number of additional MAB functions are judiciously added to the appropriate atomic centers.

In both cases, we can adjust the rank of the MAB appropriately based on information that is *already available* from the initial PRB-SCF calculation. The resulting procedure, shown in Algorithm 1, can either truncate or augment the MAB dimension on each atom. The number of significant eigenvalues (N_{Sig}) for each atom is set to the number of eigenvalues of $\mathbf{P}'_A$ that are above a first threshold (*thresh*1, whose default value is 0.01). The MAB dimension will be reduced to N_{Sig} if that is smaller than a minimal basis (N_{min}). On the other hand, when $N_{min} < N_{Sig}$, the algorithm expands the MAB dimension by the number of eigenvalues beyond N_{min} that satisfy $E_A(i)/E_A(N_{min}) > thresh2$ (*thresh*2 has a default value of 0.02, i.e., eigenvalues that are larger than $E_A(N_{min})/50$ will be included), which will allow a lower optimized MAB-SCF energy. The default values of these two thresholds are empirically determined based on the performance on the hypervalent molecules in the G2 set (see Sec. IV).

D. Modified MAB objective function

The objective function given by Eq. (10) can be rewritten as follows:

$$E = \text{Tr}\left[\mathbf{RSC}_o \mathbf{C}_o^T \mathbf{SRF}\right]$$
$$= \text{Tr}\left[\mathbf{C}_o^T \mathbf{SRFRSC}_o\right]$$
$$= \text{Tr}[\tilde{\mathbf{C}}_o^T \mathbf{F} \tilde{\mathbf{C}}_o], \tag{21}$$

where $\tilde{\mathbf{C}}_o = \mathbf{RSC}_o$ represents the PRB-optimized occupied MOs after being projected into the MAB space. For stable species, the energies of occupied MOs should all be negative, and thus minimizing E corresponds to retaining as many of the bound electrons as possible.

A disaster occurs in the MAB optimization if an occupied MO has a positive energy, because minimization will result in loss of those electrons. With inexact functionals, this occasionally happens for anions. For example, the energy of the three $2p$ orbitals in F^- is 0.001 Eh with the B3LYP functional (hybrid), and 0.056 Eh with B97-D (pure). With these functionals, the resulting value of $\text{Tr}[\mathbf{RSPS}]$ (number of electrons captured by $\tilde{\mathbf{P}}$) is close to 2 when the MAB is optimized, which indicates that the six $2p$ electrons are missing! This is completely unphysical, and causes the SCF energy computed with the MAB to be qualitatively incorrect.

Such difficulties can be avoided by modifying the eigenvalue structure of $\mathbf{F}$ to ensure that all occupied levels are negative. This can be done by applying a uniform shift to all the eigenvalues,

$$\mathbf{F}' = \mathbf{F} - \lambda \mathbf{S} \tag{22}$$

using $\mathbf{C}^T \mathbf{SC} = \mathbf{I}$. The shift, λ, is set to be

$$\lambda = \varepsilon(\text{HOMO}) + \alpha\left[\varepsilon(\text{LUMO}) - \varepsilon(\text{HOMO})\right], \tag{23}$$

where $0 < \alpha \leq 1$ so that the zero energy lies between the HOMO and LUMO calculated by PRB-SCF (note: such a shift is applied *only* when $\epsilon(\text{HOMO})$ is detected to be positive). The "mixing" parameter $\alpha = 0.75$ is empirically selected based on the performance of our method on small (monoatomic and diatomic) anions (for more details, see Sec. 1 of the supplementary material[109]). Replacing $\mathbf{F}$ with $\mathbf{F}'$ in Eq. (10) gives a modified objective function for the MAB,

$$E = \text{Tr}\left[\mathbf{RSPSR}(\mathbf{F} - \lambda\mathbf{S})\right]$$
$$= \text{Tr}[\mathbf{RSPSRF}] - \lambda\text{Tr}[\mathbf{RSPS}]. \tag{24}$$

When $\lambda > 0$, the new term resembles a penalty for losing electrons, which can be made explicit by adding an additional constant, λN_{elec}, to the RHS,

$$E = \text{Tr}[\mathbf{RSPSRF}] + \lambda\left(N_{\text{elec}} - \text{Tr}[\mathbf{RSPS}]\right). \tag{25}$$

E. Perturbation correction schemes

Based on the data presented in Refs. 89 and 90, a significant difference exists between PAO-SCF and exact SCF results. To reduce this gap, computationally inexpensive correction schemes based on perturbation theory are useful. Analogous to the dual-basis methods, the converged MAB-SCF solution serves as the primary basis reference, and the contribution of non-Brillouin singles to the second-order perturbative (PT2) energy correction is given by[63]

$$\delta E = \text{Tr}\left[\mathbf{F}_{OV}^{(1)} \mathbf{T}_{VO}^{(1)}\right]. \tag{26}$$

Here $\mathbf{F}$ denotes the Fock matrix built upon the MAB-SCF density projected into the secondary basis: $\mathbf{F} = \mathbf{F}(\tilde{\mathbf{P}})$. The first-order T-amplitude satisfies the following linear equation:

$$\mathbf{F}_{VO}^{(1)} + \mathbf{F}_{VV}^{(0)} \mathbf{T}_{VO}^{(1)} - \mathbf{T}_{VO}^{(1)} \mathbf{F}_{OO}^{(0)} = \mathbf{0}_{VO}^{(1)}. \tag{27}$$

In the pseudo-canonicalized MO basis (obtained by diagonalizing $\mathbf{F}_{OO}$ and $\mathbf{F}_{VV}$ separately, see Appendix C), Eq. (27) reduces to a simpler form,

$$T_{ai}^{(1)} = -F_{ai}^{(1)} / \left(\epsilon_a^{(0)} - \epsilon_i^{(0)} \right). \tag{28}$$

Correspondingly, the perturbative energy lowering becomes

$$\delta E = - \sum_{ia} |F_{ai}^{(1)}|^2 / \left(\epsilon_a^{(0)} - \epsilon_i^{(0)} \right), \tag{29}$$

which can be interpreted as an energy-weighted steepest descent (or an approximate Newton) step.[63,90] Alternatively, other corrections that involve a full diagonalization of the Fock matrix can be applied, such as the aforementioned DB-SCF (only slightly different from PT2) and DFPC methods. The latter performs a single update of the density matrix in the secondary basis (by diagonalizing $\mathbf{F}$), and then recomputes the full SCF energy based upon that density matrix (the result will thus be variational).

III. COMPUTATIONAL DETAILS

A pilot implementation of this new SCF scheme is accomplished in a development version of the Q-Chem 4.3 package.[110] A preconditioned limited-memory BFGS (L-BFGS) algorithm[111,112] is implemented for solving the MAB optimization problem efficiently. The inverted on-diagonal blocks of the Hessian matrix for the objective function (second derivatives of the function value with regard to orbital rotations on the same atomic site) are employed as the preconditioner of the L-BFGS algorithm. In most scenarios, this preconditioning strategy leads to convergence of the MAB optimization in a reasonable number of iterations (10^1–10^2), with only moderate additional cost to evaluate the preconditioner. More details about the preconditioned L-BFGS algorithm and the evaluation of the on-diagonal blocks of the Hessian are provided in Appendix B.

In the current implementation, all the density matrix updates are computed by diagonalizing $\mathcal{F}$ (Fock matrix in the dimension of PRB or MAB), and the only diagonalization in the secondary basis dimension ($N_v \times N_v$ to be exact) is required by the perturbation correction. However, to obtain $\mathcal{F}$ (for the time being) still requires contracting the ERI tensor with a PRB- or MAB-representable density matrix in the secondary basis (using restricted KS Fock matrix as an example),

$$F_{\mu\nu} = h_{\mu\nu} + 2 (\mu\nu|\lambda\sigma) \tilde{P}^{\lambda\sigma} - \kappa (\mu\sigma|\lambda\nu) \tilde{P}^{\lambda\sigma} + (V_{xc})_{\mu\nu}, \tag{30}$$

where κ is the proportion of exact exchange in the employed functional. Then $\mathbf{F}$ is transformed back into the primary basis through Eq. (3). This choice is actually less efficient, because quantities in the primary basis, like $\mathcal{P}^{\alpha\beta}$, C_i^α, can be directly utilized to construct $\mathcal{F}$, which will significantly reduce the dimension of the contraction. Therefore, with our preliminary implementation, we will focus on validating the accuracy of MAB-SCF (PC) in this work, and the potential efficient implementation of this method will be briefly discussed in Sec. VI.

All the results for TC and NC datasets are generated with the KS-DFT routines in Q-Chem 4.3 as well. A $(75, 302)$ grid

(75 radial shells with 302 Lebedev points in each) is used for all employed exchange-correlation (XC) functionals, and the SG-1 grid[113] is used for the VV10[18] NLC functional. Unless otherwise noted, 6-31+G(d) is used as the reference basis in PRB-SCF. Smaller reference bases could alternatively be employed, at the cost of diminished accuracy (see the results provided in Sec. 2 of the supplementary material[109]). The optimization of the MAB converges to 10^{-6} a.u., while all the SCF calculations converge to 10^{-8} a.u. (RMS of the gradient). To determine the appropriate dimension of the MAB, the default values of *thresh*1 and *thresh*2 adopted by Algorithm 1 are set to 0.01 and 0.02, respectively. We note that, in this work, the "adding vector" strategy is by default turned *off* and only utilized for specified hypervalent molecules.

IV. PRELIMINARY TESTS ON G2 THERMOCHEMISTRY

A. Comparison with PAO-SCF

We start investigating the accuracy of our method by performing a series of preliminary tests on the G2 thermochemistry set.[35] To test the quality of the optimized MAB, the performance of MAB-SCF on the G2 set is compared to PAO-SCF, since the latter gives the limiting behavior of an atom-centered minimal basis. For the reasons discussed in Sec. II C, the minimal basis models (including MAB and PAO) are not sufficient for describing hypervalent molecules. Thus, we designate molecules containing Al, Si, P, S, Cl centers that are coordinated by highly electronegative atoms (e.g., O, F, Cl) as hypervalent, including SO, ClO, SO_2, AlF_3, $AlCl_3$, SiF_4, $SiCl_4$, PF_3, ClF_3, and $(CH_3)_2SO$, and exclude them from the test set preliminarily.

The results for this "pruned" G2 set (138 molecules) computed with three functionals are collected in Table II (aQZ is the target secondary basis). To make it a fair comparison, molecules that fail to converge their PAOs (listed in the table

TABLE II. Summary of the errors of PAO-SCF and MAB-SCF (without and with PT2 correction) for the "pruned" G2 set against the conventional SCF results. aQZ is employed as the secondary (target) basis. Maximum errors (MAX), root-mean-square deviations (RMSD), and mean-signed errors (MSE) are reported in kcal/mol. Molecules where PAO failed to converge (listed in the footnotes) are excluded for both PAO-SCF and MAB-SCF when evaluating the statistical errors.

	B97-D[a]		B97M-V[b]		B3LYP[c]	
SCF energies						
	MAB	PAO	MAB	PAO	MAB	PAO
MAX	25.11	25.48	26.70	26.96	24.39	24.85
RMSD	7.07	7.27	8.06	8.07	7.05	7.18
MSE	5.59	5.72	6.54	6.49	5.62	5.66
With PT2 correction						
	MAB	PAO	MAB	PAO	MAB	PAO
MAX	−0.26	−0.14	−0.14	−0.08	0.84	0.91
RMSD	0.06	0.03	0.03	0.02	0.18	0.20
MSE	−0.02	0.00	−0.02	−0.01	0.14	0.15

[a]Convergence failures: ·CCH.
[b]Convergence failures: SO_2, ClF_3, ·SH.
[c]Convergence failures: NaCl.

footnotes) are also excluded when reporting the statistical errors. At the SCF level (i.e., without PT2 correction), PAO-SCF and MAB-SCF show similar accuracy with respect to the exact SCF results for all three tested functionals. Surprisingly, the MAB-SCF results are slightly better as a result of error cancellation (PAO-SCF is exact for atomic energies). Applying the PT2 correction significantly reduces the errors of both schemes. For the two pure functionals (B97-D and B97M-V), the RMSDs of MAB+PT2 are smaller than 0.1 kcal/mol ($\sim$0.05 kcal/mol), and they are close to those of PAO+PT2. For B3LYP, the RMSDs of both schemes noticeably increase, which suggests diminished effectiveness of PT2 when hybrid functionals are used. Nevertheless, we notice that the performance of MAB+PT2 is rather similar to that of PAO+PT2.

To better compute the energies of the hypervalent molecules, we increase the dimensions of their MAB based on Algorithm 1. The modified MAB+PT2 results are compared with those using the standard minimal basis dimensions in Table III. For a majority of these molecules (ClO, SiF$_4$, SiCl$_4$, PF$_3$, (CH$_3$)$_2$SO, and presumably SO$_2$ and ClF$_3$), the errors are reduced by over 10 times by using the "adding vector" strategy. The degree of inadequacy of the conventional minimal basis dimension is perhaps a measure of molecular hypervalency. Indeed, species like SiF$_4$, PF$_3$ do not formally violate the "octet" rule, which indicates that molecular hypervalency may exist beyond its usual definition. On the other hand, AlCl$_3$ and AlF$_3$ do not seem to be typical hypervalent species, because the use of standard minimal basis dimensions does not result in errors as large as those for the other molecules listed in Table III.

Combining these specially treated hypervalent molecules with the other 138 molecules computed with the standard MAB model, the overall RMSD of MAB+PT2 against conventional SCF results for the G2 set is 0.033 kcal/mol.

TABLE III. Errors (in kcal/mol) for the atomization energies of the hypervalent molecules in the G2 set computed with B97M-V. aQZ is used as the secondary basis. Results of unmodified MAB+PT2, MAB+PT2 with the "adding vector" strategy, and PAO+PT2 are compared against exact SCF results. The corresponding statistical errors for the full G2 set (including these molecules) are also reported.

	MAB+PT2 (normal)	MAB+PT2 (add_vec)	PAO+PT2
SO	0.049	0.022	0.085
ClO	−0.182	0.010	−0.120
SO$_2$	−0.990	0.001	N/A
AlF$_3$	−0.030	0.002	0.002
AlCl$_3$	0.050	−0.016	−0.017
SiF$_4$	−0.214	−0.007	−0.268
SiCl$_4$	−0.445	−0.009	−0.544
PF$_3$	−0.720	0.009	−0.544
ClF$_3$	−1.857	−0.020	N/A
SO(CH$_3$)$_2$	−1.173	−0.026	−1.061
G2 statistics (all molecules)			
MAX	−1.857	−0.141	−1.061
RMSD	0.238	0.033	0.114
MSE	−0.058	−0.017	−0.024

TABLE IV. Modifications of the MAB dimensions on the central atoms of the hypervalent molecules in the G2 set after the "adding vector" strategy is applied. The number of MAB functions for the coordinating atoms also occasionally changes (not shown).

Molecule	Central atom	*dim* (MAB) (original)	*dim* (MAB) (add_vec)
AlF$_3$	Al	9	16
AlCl$_3$	Al	9	13
SiF$_4$	Si	9	15
SiCl$_4$	Si	9	15
PF$_3$	P	9	14
SO	S	9	13
SO$_2$	S	9	13
(CH$_3$)$_2$SO	S	9	12
ClO	Cl	9	12
ClF$_3$	Cl	9	14

This result is only minimally different from the RMSD for the "pruned" G2 set (with B97M-V), by contrast with the poor results for standard MAB+PT2 and PAO+PT2 when hypervalent molecules are included. Therefore in the later G2 tests, unless otherwise specified, statistical errors evaluated including with *all* the molecules will be reported with the hypervalent ones *separately treated* via Algorithm 1.

Table IV shows how the dimensions of the MAB are increased on the central atoms of the hypervalent molecules after applying Algorithm 1. According to the rightmost column, the modified MAB function counts are usually close to those of a minimal basis plus one set of *d* (polarization) functions. Although AlF$_3$ and AlCl$_3$ do not show strong hypervalent character according to Table III, the Al atom nonetheless gains additional MAB basis functions.

In general, the MAB optimization problem is considerably easier to converge than the aforementioned "double" PAO-SCF optimization. In contrast to PAO-SCF that encounters several convergence problems (mentioned in Table II), no MAB convergence failure is detected for the entire G2 set with all three tested functionals. Furthermore, in contrast to PAO-SCF, the MAB optimization is decoupled from density matrix optimization, and thus MAB iteration counts will not directly affect the required number of SCF cycles. This is extremely important because of the much more significant cost per iteration for the latter. Table V lists molecules in the G2 set that require over 100 iterations to converge their PAOs (using B97M-V/aQZ). Apart from several convergence failures, the PAO-SCF scheme requires an enormous number of Fock matrix constructions for some molecules, such as NaCl. The MAB scheme, on the other hand, attains the optimized adaptive basis in fewer iterations for most of these molecules. Even in cases like COS and C$_2$H$_4$O (oxirane) where the iteration counts for optimizing the adaptive basis are similar, MAB-SCF is still far more efficient because many fewer Fock builds are required. Therefore, MAB-SCF appears to be a more feasible adaptive basis SCF scheme than PAO-SCF, with comparable accuracy.

TABLE V. Molecules in the G2 set that require over 100 iterations to converge their PAOs. The counts of iterative optimization steps and Fock builds required to converge PAO-SCF and MAB-SCF are compared. For several radicals (CH, ·SH, CH$_3$CH$_2$O·), considerably larger number of iterations are required by the PRB-SCF step because the geometric direct minimization (GDM) algorithm[114] is utilized to circumvent unstable SCF solutions.

Molecules	Num of opt steps		Num of Fock builds		
	PAO	MAB	PAO	PRB	MAB
CH	277	93	277	27	6
Na$_2$	1219	23	1219	7	3
Si$_2$	145	20	145	16	22
NaCl	3208	47	3208	9	5
SO$_2$	N/A	212	N/A	11	8
COS	148	146	148	10	7
ClF$_3$	N/A	83	N/A	12	8
C$_2$Cl$_4$	614	80	614	9	5
C$_4$H$_6$ (2-butyne)	490	149	490	10	7
C$_2$H$_4$O	109	111	109	9	6
SH	N/A	60	N/A	48	8
CH$_3$CH$_2$O	450	156	450	44	9

B. Functional dependence: pure vs hybrids

Table II already suggests that the performance of MAB+PT2 is not completely functional-independent. A clear difference exists between using pure and hybrid functionals. Therefore, we must investigate the performance of this method when different flavors of density functionals are employed. Using the G2 set with aug-cc-pVQZ as the secondary basis, Table VI explores the performance of 13 density functionals. The first seven functionals do not contain exact exchange, including three GGAs (B97-D, BLYP,[26,27] PBE[5]) and four meta-GGAs (TPSS,[8] MGGA_MS1,[115] M06-L,[116] B97M-V). For these functionals, the MAB+PT2 scheme demonstrates good accuracy (M06-L is the largest outlier), while the RMSDs computed by MAB+DFPC are roughly twice as large. Thus PT2 appears preferable.

TABLE VI. RMSDs (in kcal/mol) of MAB-SCF with two different perturbation correction schemes (PT2 and DFPC) for the G2 set. Different pure (B97-D, BLYP, PBE, TPSS, MGGA_MS1, M06-L, B97M-V) and hybrid (TPSSh, B3LYP, PBE0, M06-2X, ωB97X-D, ωB97X-V) functionals are investigated. aQZ is employed as the secondary basis for all the calculations.

Functionals	MAB+PT2	MAB+DFPC
B97-D	0.053	0.115
BLYP	0.074	0.148
PBE	0.065	0.132
TPSS	0.055	0.119
MGGA_MS1	0.042	0.108
M06-L	0.144	0.156
B97M-V	0.033	0.079
TPSSh	0.099	0.086
B3LYP	0.181	0.083
PBE0	0.220	0.066
M06-2X	0.458	0.076
ωB97X-D	0.645	0.119
ωB97X-V	0.668	0.133

The other six functionals in Table VI are hybrid functionals, including TPSSh (10%),[117] B3LYP (20%), PBE0 (25%),[118] M06-2X (54%), ωB97X-D (RSH),[119] and ωB97X-V (RSH) ("%" denotes the proportion of exact exchange). They show results that contrast with those of the pure functionals. For these hybrids, PT2 undershoots the exact SCF energy, and the size of the errors grows roughly with the amount of exact exchange. This leads to unsatisfactory accuracy (RMSD > 0.2 kcal/mol) for functionals that contain more exact exchange than B3LYP. The MAB+DFPC approach, on the other hand, shows comparatively better performance across the hybrids (RMSDs are around 0.1 kcal/mol), than PT2. For the two RSH functionals, the results of MAB+DFPC are 5–6 times more accurate than MAB+PT2.

We conclude that DFPC should be used as the correction to MAB-SCF when hybrid functionals are employed, at the expense of one more Fock build in the secondary basis. On the other hand, MAB+PT2 is less expensive and more accurate for pure functionals.

C. Basis set convergence

Different basis sets are employed for KS-DFT calculations, based on considerations such as accuracy, efficiency, and user experience. With the B97M-V functional, we assess the performance of MAB+PT2 with several widely used basis sets, including those in Dunning's correlation-consistent series (aTZ, QZ, aQZ),[28,29] Jensen's polarization-consistent series (apc-2, pc-3, apc-3),[120–122] and the Karlsruhe "def2" series (TZVPPD, QZVPP, QZVPPD).[33] The popular "large Pople" basis set 6-311++G(3df,3pd)[123,124] is also included. These basis sets are augmented triple-ζ quality or higher, because our goal is to approach CBS limit results.

The RMSDs for the G2 set using different target secondary basis sets are displayed in Figure 4. The errors are typically below 0.1 kcal/mol, which indicates excellent transferability. The best performance is achieved by two quadruple-ζ basis sets with diffuse functions, aQZ and QZVPPD. For their unaugmented counterparts, QZ and QZVPP, the RMSDs are slightly larger, although still satisfactory. Since diffuse functions typically have no major impact on the accuracy of evaluated energetics for bonded interactions, the compatibility of MAB+PT2 with these unaugmented basis sets is helpful. The largest RMS errors are produced by Jensen's pc-3 and aug-pc-3 basis sets. This is mostly due to the inaccurate atomic energy for Li: the RMSDs are reduced to 0.063 (pc-3) and 0.059 (apc-3) kcal/mol if we exclude three Li-containing molecules (Li$_2$, LiF, and LiH) from the G2 set. As these are high-quality basis sets, the outlier might be due to poor compatibility of pc-3/apc-3 with the employed reference basis, 6-31+G(d), for the Li atom. Computational cost aside, any of the largest basis sets approach the true CBS limit and also the top accuracy of our method.

Ultimately, MAB-SCF with a perturbation correction scheme (MAB-SCF (PC)) might replace conventional SCF in KS-DFT calculations to approach the CBS limit. Therefore, we want to see how MAB-SCF (PC) approaches the CBS
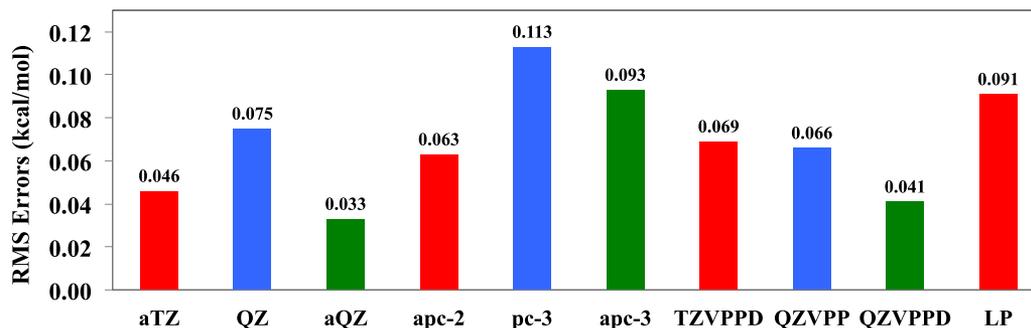
FIG. 4. Assessment of the performance of MAB+PT2 with 10 secondary basis sets. RMSDs (vs. exact SCF) for the G2 set calculated with the B97M-V functional are reported. Different colors are used to indicate distinct basis set categories: augmented triple-ζ (red), unaugmented quadruple-ζ (blue), and augmented quadruple-ζ (dark green).
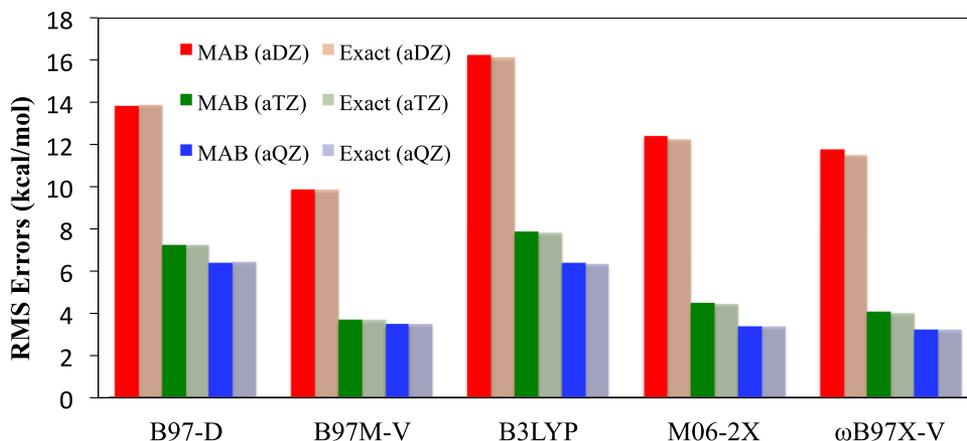

FIG. 5. Basis set convergence of the functional RMSDs for the G2 set evaluated by MAB-based methods (solid) and conventional SCF (transparent). The reference values are taken from Ref. 36. Dunning's augmented correlation-consistent basis sets are utilized to approach the CBS limit (D: red, T: green, Q: blue). PT2 and DFPC correction schemes are applied on top of the MAB-SCF results of local (B97-D, B97M-V) and hybrid (B3LYP, M06-2X, ωB97X-V) functionals, respectively.

limit with the increasing size of the secondary basis. For that purpose, we extend Figure 5 (which motivates this work) with the functional RMSDs evaluated by MAB+PT2 (for B97-D, B97M-V) and MAB+DFPC (for B3LYP, M06-2X, ωB97X-V), as shown in Figure 5. These results are quite encouraging: the conventional SCF convergence towards the CBS limit of each functional is closely reproduced by MAB-SCF (PC). At the aQZ level, the largest difference between MAB-SCF (PC) and exact SCF results is only about 0.03 kcal/mol (for B97-D and B3LYP), which is below 1% of the intrinsic (CBS) error of the functional itself, and thus negligible. Larger differences between these two sets of results exist at the aDZ level (the largest gap is 0.26 kcal/mol for ωB97X-V/aDZ), but this is not relevant to our target of the CBS limit.

V. ADDITIONAL ACCURACY TESTS

A. Thermochemistry

We will assess how the performance of MAB-SCF (PC) transfers to other thermochemistry (TC) datasets. Three density functionals (B97-D, B97M-V, B3LYP) will be employed to examine the accuracy of MAB-SCF (PC), using aQZ as the secondary basis, and the perturbation correction schemes are applied in the same way as in Figure 5. First we consider the W4-11 dataset,[125] which includes 99 bond dissociation energies (BDE99), 707 heavy-atom transfer reaction energies (HAT707), 20 isomerization energies

(ISO20), 13 nucleophilic substitution reaction energies (SN13), and 140 total atomization energies (TAE140). Note that the multi-reference (MR) species in W4-11 are included in this test. Thirteen species are separately treated as hypervalent molecules: AlF_3, $AlCl_3$, SiF_4, P_4, SO, SO_2, SO_3, S_2O, S_2, S_3, S_4, ClO, and OClO (see the previous discussion for the G2 set).

Table VII contains the RMSDs of the MAB-SCF (PC) approach (against exact SCF) for W4-11, where different TC categories have been separated. The overall performance is similar to that for the G2 set, which shows encouraging transferability. Taking B97M-V as an example, the smallest and largest RMSDs of MAB+PT2 (vs. exact SCF) are obtained on SN13 (0.02 kcal/mol) and HAT707 (0.06 kcal/mol), respectively, while the corresponding functional RMSDs (vs. W4 reference) are 1.39 kcal/mol and 3.90 kcal/mol. Therefore, the errors caused by replacing conventional SCF with MAB-SCF (PC) are usually smaller than intrinsic functional errors by one or two orders of magnitude. A more straightforward comparison is provided by Figure 6: for B97M-V, the functional RMSDs computed via the MAB+PT2 approach show almost no difference compared to those by normal SCF method.

We also examined the performance of MAB-SCF (PC) on other TC datasets, including adiabatic ionization potentials and electron affinities (21 for each: G21IP and G21EA),[126,127] 38 non-hydrogen transfer and 38 hydrogen transfer barrier heights (NHTBH38[128] and HTBH38[129]), and 14 alkane isomerization energies (Pentane14[130]). The computational

TABLE VII. RMSDs (in kcal/mol) of MAB-SCF (PC) for the W4-11 dataset (vs. exact SCF) evaluated with using three different functionals. PT2 is applied for B97-D and B97M-V while DFPC is used for B3LYP. aQZ is used as the secondary basis.

	B97-D	B97M-V	B3LYP
BDE99	0.061	0.038	0.070
HAT707	0.101	0.061	0.100
ISO20	0.068	0.048	0.085
SN13	0.034	0.018	0.053
TAE140	0.071	0.046	0.096
Overall	0.093	0.057	0.096

TABLE VIII. RMSDs (in kcal/mol) of MAB-SCF (PC) for TC datasets other than W4-11: G21IP, G21EA, HTBH38, NHTBH38, Pentane14. The numbers in parentheses are recalculated RMSDs for G21EA after applying the "adding vector" strategy for four diatomic anions: NO^-, PO^-, O_2^-, and S_2^-.

	B97-D	B97M-V	B3LYP
G21IP	0.050	0.036	0.040
G21EA	0.598	0.412	0.455
	(0.101)	(0.027)	(0.078)
HTBH38	0.016	0.035	0.031
NHTBH38	0.099	0.071	0.118
Pentane14	0.005	0.007	0.002

details are identical to those for W4-11, except for the anions in G21EA, HTBH38 and NHTBH38, where the modified MAB objective function (introduced in Sec. II D) is automatically applied to avoid the emergence of unphysical results. The RMSDs of MAB-SCF (PC) are presented in Table VIII. We see that HTBH38 and Pentane14 are relatively easier cases for MAB-SCF (PC) to approximate the exact SCF result; and for G21IP and NHTBH38, the size of the RMSDs is similar to that for those W4-11 subsets (e.g., BDE99 and SN13).

The largest RMSD appears on the G21EA dataset. Although none of the results are qualitatively incorrect by applying the modified MAB objective function, there are several molecular anions whose absolute energies evaluated by MAB-SCF (PC) are rather unsatisfactory: NO^-, PO^-, O_2^-, and S_2^-. Based on the discussions in Sec. II C, the "adding vector" strategy may also be applied for these electron-abundant species. As a result, the absolute energies of these anions are significantly improved, as shown in Table IX. With these 4 molecular anions specially treated, the RMSDs for the G21EA dataset are recalculated and the results are also presented in Table VIII, which turn out to be more comparable to the RMSDs for other TC test sets.
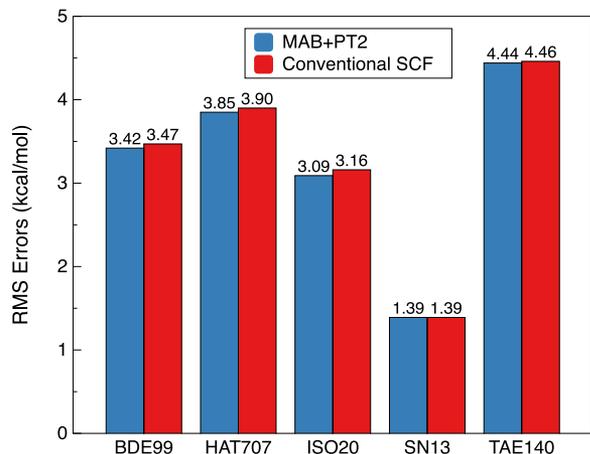
B. Non-covalent interactions

One of the key improvements in modern density functionals is for non-covalent interactions (NC). Therefore, we assess the performance of MAB-SCF (PC) on several NC datasets, including A24 (24 small NC complexes),[30] S22 (22 diverse small- to medium-sized NC complexes at the equilibrium geometries),[131,132] HB15 (15 ionic hydrogen bond interactions),[133] H2O6Bind8 (binding energies of eight configurations of water hexamers),[134,135] and FmH2O10 (binding energies of 10 configurations of $F^-(H_2O)_{10}$[134,135]). Since many NC interactions have smaller magnitudes than TC energy differences, and some modern density functionals are able to achieve very small errors for them (e.g., B97M-V's unweighted RMSD for 1458 non-covalent interactions is 0.22 kcal/mol[20]), higher accuracy is needed for the MAB approach to match the exact SCF results.

As before, three density functionals (B97-D, B97M-V, B3LYP-D3(0)) are employed to assess the performance of MAB-SCF (PC) on these NC datasets. To avoid counterpoise (CP) corrections, we choose def2-QZVPPD as the secondary basis, which has fewer functions than aQZ but generates even smaller BSSEs.[136] Table X contains the resulting RMSDs (vs. exact SCF) for these NC datasets. Very small MAB-SCF (PC) errors are found for dimer binding energies (A24, S22, and HB15), with all three functionals. This is very encouraging for treating the most common non-covalent interactions by MAB-SCF (PC) instead of conventional SCF. Larger errors appear for the cluster binding energies (H2O6Bind8, FmH2O10), due to the larger magnitude of these interactions (H2O6Bind8: −40 to −50 kcal/mol; FmH2O10:

FIG. 6. B97M-V's RMSDs for W4-11 (vs. reference values) evaluated by MAB+PT2 (blue) and conventional SCF (red). Very similar accuracy is obtained by these two SCF schemes at the B97M-V/aQZ level of theory.

TABLE IX. Errors in kcal/mol (vs. exact SCF) for the absolute energies of four diatomic anions (NO^-, PO^-, O_2^-, and S_2^-) evaluated with standard MAB-SCF (PC) ("Standard") and with the "adding vector" strategy turned on ("Add_Vec"). Other computational details are the same as in Table VIII.

	Standard			Add_Vec		
	B97-D	B97M-V	B3LYP	B97-D	B97M-V	B3LYP
NO^-	−2.36	−1.63	1.60	−0.13	−0.06	0.10
PO^-	−0.52	−0.55	0.92	−0.02	−0.02	0.17
O_2^-	−2.04	−1.25	1.66	−0.01	−0.02	0.07
S_2^-	−0.41	−0.16	0.23	−0.01	0.02	0.02

TABLE X. RMSDs (in kcal/mol) of MAB-SCF (PC) for five NC datasets: A24, S22, HB15, H2O6Bind8, and FmH2O10. Data points computed by conventional SCF using the def2-QZVPPD basis set (noCP) provide the reference values.

	B97-D	B97M-V	B3LYP-D3(0)
A24	0.009	0.005	0.009
S22	0.015	0.023	0.025
HB15	0.021	0.027	0.023
H2O6Bind8	0.061	0.085	0.118
FmH2O10	0.078	0.026	0.350

about −170 kcal/mol) and the uniformity of their interaction types (systematic errors accumulate on one single direction through all the data points). The largest outlier appears on FmH2O10 when the B3LYP-D3(0) functional is employed, mostly due to the less accurate MAB-DFPC result for F⁻ (the error is 0.3 kcal/mol vs. exact SCF) with that functional.

The accurate description of these non-covalent interactions by B97M-V can also be reproduced using MAB-SCF (PC), and the resulting RMS errors (vs. reference data) are compared with the exact SCF results in Figure 7. Only minimal differences exist between two sets of RMSDs for the dimer binding energies, while the monotonic deviations of the MAB-SCF (PC) results for each single data point contribute to more pronounced differences for the clusters. Nevertheless, even for H2O6Bind8 where the largest deviation occurs, the RMSD vs. exact SCF results (0.085 kcal/mol) is only 0.1%–0.2% of the magnitude of the corresponding binding energies. In terms of the evaluation of relative energies, fairly insignificant errors are caused by using MAB-SCF (PC).

Finally we test the accuracy of MAB-SCF (PC) on large non-covalent complexes, using the seven dispersion-bound systems of the recently proposed L7 dataset,[137] including (the abbreviations simply follow Ref. 137): stacked circumcoronene-adenine dimer (C3A), stacked circumcoronene with a Watson-Crick G-C base pair (C3GC), parallel displaced coronene dimer (C2C2PD), stacked Watson-Crick G-C base pairs (GCGC), stacked guanine trimer (GGG),

parallel stacked octadecane dimer (CBH), and phenylalanine residue trimer (PHE). Due to the tremendous computational effort required for these systems, our calculations were performed with two pure functionals (B97-D, B97M-V) using aug-cc-pVTZ as the secondary basis. Binding energies of these complexes evaluated by conventional SCF and MAB-SCF (PC) are compared in Table XI. For most of them, the differences between the results given by two SCF schemes are much smaller than 1% of the magnitude of their binding energies, which indicates the satisfactory accuracy of MAB-SCF (PC) for these large-scale non-covalent interacting systems. The only exception is the guanine trimer (GGG), which is primarily due to its very weak binding energy (−2.33 kcal/mol by B97M-V/aTZ). Due to the lack of reliable reference values for the time being, we cannot meaningfully assess the intrinsic functional errors for L7.

VI. DISCUSSION AND FUTURE WORK

The tests performed on a broad range of systems demonstrate that the overall accuracy of this new approximate SCF method seems to be encouraging. As the CBS limit is approached, the functional RMSDs for several representative TC and NC datasets (G2, W4-11, S22) evaluated by MAB-SCF (PC) show very small differences compared to the exact SCF results, as demonstrated in Figures 5, 6, and 7. In particular, "B97M-V/QZVPPD/MAB+PT2" (or "B97M-V/aQZ/MAB+PT2") turns out to be a promising model chemistry, because of B97M-V's very good accuracy for both TC and NC, its moderate computational cost as a semi-local functional, and its compatibility with the MAB+PT2 approach.

The MAB-SCF (PC) procedure is similar to that of dual-basis methods (including DB-SCF and DFPC), since they both first compute an SCF solution in a primary basis (preferably a subset of the secondary basis) then correct it using perturbation theory. It is encouraging that MAB-SCF (PC) is able to achieve comparable accuracy to DB-SCF with a primary basis adaptively prepared on the fly, whose size is also much smaller than those well-trained DB-SCF basis subsets[67] when the secondary basis approaches the CBS limit. Some comparisons of these two approaches and related discussion are provided in Sec. 4 of the supplementary material.[109]

At least three challenges remain in successfully applying the MAB-SCF (PC) method. First is the challenge of hybrid density functionals. According to Table II, there is no degradation in terms of the quality of the MAB when a hybrid functional is used: B3LYP in fact has the smallest RMSD for MAB-SCF among the three tested functionals. The reason for the poor performance of MAB+PT2 for hybrid functionals should therefore reside in the PT2 correction represented by Eq. (29). Assuming that the MABs calculated by two functionals are of similar quality, the numerators in Eq. (29) should also have fairly similar sizes. Thus, functional-dependent differences in the magnitude of the PT2 correction will be largely determined by the denominators (orbital energy differences) in Eq. (29). It turns out that the smaller gaps calculated by pure functionals help reduce the errors of MAB-SCF more effectively, while hybrid functionals that usually give larger orbital energy gaps undercorrect. This numerical
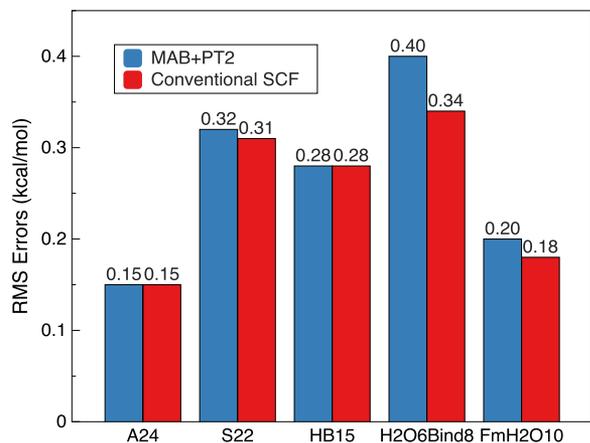

FIG. 7. B97M-V's RMSDs for 5 NC datasets (vs. reference values) evaluated by MAB-SCF (PC) (blue) and conventional SCF (red), respectively. def2-QZVPPD is employed as the secondary basis set without using counterpoise corrections.

TABLE XI. Binding energies (in kcal/mol) of the L7 complexes evaluated by conventional SCF and MAB-SCF (PC), using two pure density functionals with dispersion corrections (B97-D, B97M-V) and aug-cc-pVTZ as the secondary basis. Unsigned percentage error of the MAB-SCF (PC) results are provided.

Complex	B97-D			B97M-V		
	E_{bind} (exact)	E_{bind} (MAB)	Error (%)	E_{bind} (exact)	E_{bind} (MAB)	Error (%)
C3A	−18.514	−18.505	0.05	−17.466	−17.488	0.13
C3GC	−31.120	−31.113	0.02	−31.050	−31.130	0.26
C2C2PD	−22.355	−22.343	0.05	−22.316	−22.329	0.06
GCGC	−15.302	−15.295	0.05	−15.460	−15.514	0.35
GGG	−2.485	−2.530	1.81	−2.330	−2.392	2.66
CBH	−15.335	−15.330	0.03	−12.396	−12.394	0.02
PBH	−23.311	−23.290	0.09	−25.867	−25.872	0.02

result is intriguing since the gaps computed by the hybrids are usually deemed to be physically more correct. Although the accuracy of MAB+DFPC is satisfactory in most tested cases, an extra Fock build using the density matrix in the secondary basis (*not* MAB-representable) is required, which would be preferable to avoid (the reason will be elucidated below). Additionally, based on Table VI, the accuracy of MAB+DFPC slightly degrades for RSH functionals as well despite its clear advantage over MAB+PT2.

A second limitation is that hypervalent molecules are treated separately in this work by modifying the MAB dimensions in a semi-automated way (choosing to use the "adding vector" approach is user-specified). This is because our algorithm and its related thresholds were chosen to attain the desired accuracy for hypervalent molecules. This option increases the size of the MAB (and thus the computational cost of MAB-SCF) and potentially degrades the convergence of MAB optimization (see the discussion in Sec. 3 of the supplementary material[109]). Further work on wisely adjusting the algorithm parameters, or possibly refining the present scheme, is desirable.

The third general challenge is efficient computational implementation. Although this paper is mainly about the formulation MAB-SCF (PC) and the validation of its accuracy, the ultimate goal is to serve as an inexpensive substitute for conventional SCF as the CBS limit is approached. Compared to normal SCF, our method involves far fewer degrees of freedom. The cost of a single density matrix update, if achieved by diagonalizing the Fock matrix, can be reduced by up to a factor of $(n/m)^3$, where n is the size of the secondary basis, and m is the size of the PRB or MAB. For example, given a model system $(CH_2)_n$, there will be 129 basis functions per –CH_2– unit if def2-QZVPPD is employed as the secondary basis, which is roughly 6 times as large as the PRB (22 functions per unit), and about 18 times as the MAB (7 functions/unit). Therefore, the prefactor of the computational cost of this cubic scaling step is significantly reduced in our scheme. A linear equation solve in the secondary basis dimension is still required to obtain the PT2 correction, while that is far less expensive than diagonalization.

Alternatively, $O(N)$ scaling electronic structure methods can be potentially applied to the MAB-SCF step, since the overlap matrix of the MAB is extremely well-conditioned, as demonstrated by Table XII. The diagonalization-free density

matrix update algorithms introduced by Ref. 52 could be used for the MAB-SCF. However, the feasibility of these methods for PRB-SCF is not so clear because one set of diffuse functions is contained in 6-31+G(d).

In practice, however, even for systems as large as the L7 complexes, the Fock matrix construction step still dominates the computational cost of each SCF cycle due to its large prefactor when high accuracy is sought. This is despite the fact that asymptotically that step scales quadratically[138] (or even linearly with special algorithms)[40–47] with respect to system size. In our pilot implementation, the Fock matrix is built from PRB- or MAB-representable density matrices that are of secondary basis dimensions, which costs almost the same as a Fock build in a conventional SCF calculation within the secondary basis. Therefore, to speed up these Fock build steps by taking advantage of the properties of the PRB and MAB becomes the most urgent task for MAB-SCF (PC) to outperform conventional SCF in terms of computational efficiency, especially for medium-sized systems. PRB-SCF can be reformulated as a conventional SCF calculation within a basis set whose size and shell structure are identical to 6-31+G(d), due to the elimination of high angular momentum functions during the basis set projection procedure, and the cost of the involved Fock builds thereby can be significantly reduced without requiring more sophisticated techniques. For MAB-SCF, instead of using Eq. (30), the MAB Fock matrix can be constructed by forming the ERI tensor in the primary basis first then contracting it with the MAB density matrix, as

$$\mathcal{F}_{\alpha\beta} = h_{\alpha\beta} + 2\,(\alpha\beta|\gamma\delta)\,\mathcal{P}^{\gamma\delta} - \kappa\,(\alpha\gamma|\beta\delta)\,\mathcal{P}^{\gamma\delta} + (V_{xc})_{\alpha\beta}. \quad (31)$$

TABLE XII. Comparison of the sizes (N) and overlap matrix condition numbers (λ) of the MAB and the target aug-cc-pVTZ basis set on the L7 complexes.

	N (MAB)	N (aTZ)	λ (MAB)	λ (aTZ)
C3A	343	3473	12.17	5.07×10^{12}
C3GC	393	4002	12.57	7.93×10^{12}
C2C2PD	264	2760	13.50	4.32×10^{11}
GCGC	210	2208	9.02	1.39×10^{8}
GGG	180	1863	8.55	1.25×10^{8}
CBH	256	3404	15.63	1.45×10^{9}
PHE	267	3036	13.87	6.24×10^{8}

Furthermore, quantities like $(\alpha\beta|\gamma\delta)$ can be efficiently computed (and stored) using the RI approximation, due to the highly compact and atom-blocked structure of the MAB. We intend to present an optimized implementation of MAB-SCF (PC) and the resulting timings vs. conventional SCF as soon as these challenges are adequately addressed.

VII. CONCLUSION

In this work, we proposed a new minimal adaptive basis (MAB) SCF method that can be used in KS-DFT calculations. Its objective is to permit approach to the complete basis set (CBS) limit without explicitly performing the calculation in very large AO basis sets. The key aspects of this paper can be summarized as follows:

- An MAB is obtained as a molecule-adapted, atom-blocked transformation from the secondary AO basis. We have developed a viable optimization method to obtain an MAB using an inexpensive energy-like surrogate function based on a reference SCF calculation in a projected basis of moderate size (the PRB). Compared to exact energy optimization with respect to the transformation (the PAO method), our MAB yields similar total energies with far fewer convergence issues.
- A preconditioned L-BFGS algorithm that requires the gradient and the on-diagonal blocks of the Hessian of the objective function is implemented to solve the MAB optimization problem. In addition, an approach that modifies the MAB dimension based on chemical environment is proposed, and demonstrated to be necessary for hypervalent molecules. These ideas can potentially be used in PAO calculations and other adaptive basis models as well.
- Perturbation corrections (PC) are applied to MAB-SCF to approach the desired accuracy. This resembles DB-SCF without the need to select or develop the paired basis subset. The resulting accuracy is assessed on numerous TC and NC datasets. Measured against exact SCF results, MAB-SCF (PC) generates <0.15 kcal/mol RMSDs for most of the tested TC datasets, and even smaller errors (usually <0.1 kcal/mol) for the NCs. Encouragingly, as the CBS limit is approached, the MAB-SCF (PC) method deviates from full SCF one or two orders of magnitude less than the inherent errors in today's best functionals.
- Future work includes further refining the MAB-SCF (PC) model, and developing an efficient implementation, as discussed in Sec. VI. We note that hybrid functionals require a different PC to achieve adequate accuracy, and that the strategy to treat hypervalent species is not yet fully automated. Nonetheless, based on the accuracy demonstrated here, and the potential computational advantages of using an atom-centered minimal basis, we believe that this approach merits further development. We hope to report further progress in due course.

APPENDIX A: MORE DETAILS ABOUT THE GRADIENT OF THE MAB OBJECTIVE FUNCTION

Using the same notations as in Sec. II B, the most general form of the gradient of the MAB objective function is given by

$$\frac{\partial E}{\partial \Delta^{ZpZq}} = \sum_{X,Y} \frac{\partial R^{X\mu Y\nu}}{\partial \Delta^{ZpZq}} G_{Y\nu X\mu}, \tag{A1}$$

where $\mathbf{G} = \mathbf{SPSRF} + \mathbf{FRSPS}$. With the parameterization of $\mathbf{B}$ by Equations (14) and (15), we have

$$\frac{\partial R^{X\mu Y\nu}}{\partial \Delta^{ZpZq}} = \left[C_{Xp}^{X\mu}\delta_Z^X \sigma_{XqXi} - C_{Xq}^{X\mu}\delta_Z^X \sigma_{XpXi} \right] (\sigma^{-1})^{XiYj}(B^T)_{Yj}^{Y\nu}$$
$$+ B_{Xi}^{X\mu}(\sigma^{-1})^{XiYj}\left[C_{Yp}^{Y\nu}\delta_Z^Y \sigma_{YqYj} - C_{Yq}^{Y\nu}\delta_Z^Y \sigma_{YpYj} \right]$$
$$- \sum_W (RS)_{W\sigma}^{X\mu}\left[C_{Wp}^{W\sigma}\delta_Z^W \sigma_{WqWl} - C_{Wq}^{W\sigma}\delta_Z^W \sigma_{WpWl} \right] (\sigma^{-1})^{WlYj}(B^T)_{Yj}^{Y\nu}$$
$$- \sum_W B_{Xi}^{X\mu}(\sigma^{-1})^{XiWl}\left[C_{Wp}^{W\sigma}\delta_Z^W \sigma_{WqWl} - C_{Wq}^{W\sigma}\delta_Z^W \sigma_{WpWl} \right] (SR)_{W\sigma}^{Y\nu} \tag{A2}$$

and correspondingly the gradient becomes

$$\frac{\partial E}{\partial \Delta^{ZpZq}} = -2\sigma_{ZpZj}\left[(\sigma^{-1})B^T G(I-RS)C\right]_{Zq}^{Zj}$$
$$+ 2\sigma_{ZqZj}\left[(\sigma^{-1})B^T G(I-RS)C\right]_{Zp}^{Zj}. \tag{A3}$$

When $\mathbf{\Delta}$ stands for the orbital rotations within the MAB space, i.e., $p = i$, $q = j$, matrix $\mathbf{C}$ reduces to $\mathbf{B}$, and the gradient vanishes because $(\mathbf{I} - \mathbf{RS})\mathbf{B} = \mathbf{0}$. On the other hand, orbital rotations within the complementary space of the MAB ($p = a$, $q = b$) have no effect on the objective function value either, simply due to the enforced on-atom orthogonality ($\sigma_{ZaZj} = 0$). Therefore, the only non-zero block of this gradient is resulted from the rotations between these two subspaces. If we set $p = i$, $q = a$, only the first term in Eq. (A3) remains based on the arguments above, which immediately leads to the gradient represented by Eq. (19).

APPENDIX B: THE PRECONDITIONED L-BFGS ALGORITHM

The basic idea of L-BFGS is to construct the approximate Hessian (inverse Hessian in practice) for the current iteration with gradients and displacements computed in the most recent m steps, where m is the user-specified subspace size (number of "memorized" steps). If we define the gradient and displacement at kth iteration as $\mathbf{g_k}$ and $\mathbf{s_k}$, and $\mathbf{y_k} = \mathbf{g_{k+1}} - \mathbf{g_k}$, the kth approximate inverse Hessian can be evaluated as[111]

$$
\begin{aligned}
\mathbf{H_k} = {} & \left(\mathbf{V_{k-1}^T} \cdots \mathbf{V_{k-m}^T}\right) \mathbf{H_k^0} \left(\mathbf{V_{k-m}} \cdots \mathbf{V_{k-1}}\right) + \rho_{k-m}\left(\mathbf{V_{k-1}^T} \cdots \mathbf{V_{k-m+1}^T}\right) \mathbf{s_{k-m}}\mathbf{s_{k-m}^T}\left(\mathbf{V_{k-m+1}} \cdots \mathbf{V_{k-1}}\right) \\
& + \rho_{k-m+1}\left(\mathbf{V_{k-1}^T} \cdots \mathbf{V_{k-m+2}^T}\right) \mathbf{s_{k-m+1}}\mathbf{s_{k-m+1}^T}\left(\mathbf{V_{k-m+2}} \cdots \mathbf{V_{k-1}}\right) + \cdots + \rho_{k-1}\mathbf{s_{k-1}}\mathbf{s_{k-1}^T},
\end{aligned} \tag{B1}
$$

where

$$
\rho_\mathbf{k} = \frac{1}{\mathbf{y_k^T s_k}}, \quad \mathbf{V_k} = \mathbf{I} - \rho_\mathbf{k}\mathbf{y_k}\mathbf{s_k^T}. \tag{B2}
$$

In practice, a "two-loop" algorithm which only requires evaluating vector-vector products is implemented to compute $\mathbf{H_k}$ acting on $\mathbf{g_k}$. $\mathbf{H_k^0}$ is the preconditioner for the approximate inverse Hessian. By default, a constant scaling factor is used

for $\mathbf{H_k^0}$, which is considered as the unpreconditioned case here. Once $\mathbf{H_k^0}$ contains more information about the true inverse Hessian, the step generated by L-BFGS becomes closer to a Newton step, which can presumably accelerate the convergence.

If we still denote $\mathbf{G} = \mathbf{SPSRF} + \mathbf{FRSPS}$, the Hessian of the MAB objective function (Eq. (10)) can be formally represented as follows (only on-block mixings are allowed):

$$
\begin{aligned}
H_{X_iX_a,Y_jY_b} = {} & \left.\frac{\partial^2 E}{\partial \Delta^{X_iX_a}\partial \Delta^{Y_jY_b}}\right|_{\Delta=0} \\
= {} & \mathbf{R^\Delta SPSR^\Delta F} + \mathbf{R^{\Delta\Delta}G} \\
= {} & 2(P_{ov})_{X_iY_b}(F_{ov})_{Y_jX_a} + 2(F_{ov})_{X_iY_b}(P_{ov})_{Y_jX_a} + 2(P_{oo})_{X_iY_j}(F_{vv})_{Y_bX_a} + 2(F_{oo})_{X_iY_j}(P_{vv})_{Y_bX_a} \\
& - 2(S_{ov})_{X_iY_b}(G_{ov})_{Y_jX_a} - 2(G_{ov})_{X_iY_b}(S_{ov})_{Y_jX_a} + 2(S_{oo})_{X_iY_j}(G_{vv})_{Y_bX_a} - 2(G_{oo})_{X_iY_j}(S_{vv})_{Y_bX_a},
\end{aligned} \tag{B3}
$$

where $\mathbf{R^\Delta}$ and $\mathbf{R^{\Delta\Delta}}$ stands for first- and second-order derivatives of $\mathbf{R}$ with respect to $\mathbf{\Delta}$. The first four terms on the RHS of Eq. (B3) come from the "$\mathbf{R^\Delta R^\Delta}$" term, while the rest from the "$\mathbf{R^{\Delta\Delta}}$" term. The explicit forms of the involved matrix elements are

$$
\begin{aligned}
(P_{oo})_{X_iY_j} &= \sigma_{X_iX_k}\left[\sigma^{-1}B^TSPSB\sigma^{-1}\right]^{X_kY_l}\sigma_{Y_lY_j}, \\
(P_{ov})_{X_iY_a} &= \sigma_{X_iX_j}\left[\sigma^{-1}B^TSPS(I-RS)\right]_{Y_\nu}^{X_j}V_{Y_a}^{Y_\nu}, \\
(P_{vv})_{X_aY_b} &= (V^T)_{X_a}^{X_\mu}\left[(I-SR)SPS(I-RS)\right]_{X_\mu Y_\nu}V_{Y_b}^{Y_\nu},
\end{aligned} \tag{B4}
$$

$$
\begin{aligned}
(F_{oo})_{X_iY_j} &= \sigma_{X_iX_k}\left[\sigma^{-1}B^TFB\sigma^{-1}\right]^{X_kY_l}\sigma_{Y_lY_j}, \\
(F_{ov})_{X_iY_a} &= \sigma_{X_iX_j}\left[\sigma^{-1}B^TF(I-RS)\right]_{Y_\nu}^{X_j}V_{Y_a}^{Y_\nu}, \\
(F_{vv})_{X_aY_b} &= (V^T)_{X_a}^{X_\mu}\left[(I-SR)F(I-RS)\right]_{X_\mu Y_\nu}V_{Y_b}^{Y_\nu},
\end{aligned} \tag{B5}
$$

$$
\begin{aligned}
(G_{oo})_{X_iY_j} &= \sigma_{X_iX_k}\left[\sigma^{-1}B^TGB\sigma^{-1}\right]^{X_kY_l}\sigma_{Y_lY_j}, \\
(G_{ov})_{X_iY_a} &= \sigma_{X_iX_j}\left[\sigma^{-1}B^TG(I-RS)\right]_{Y_\nu}^{X_j}V_{Y_a}^{Y_\nu}, \\
(G_{vv})_{X_aY_b} &= (V^T)_{X_a}^{X_\mu}\left[(I-SR)G(I-RS)\right]_{X_\mu Y_\nu}V_{Y_b}^{Y_\nu},
\end{aligned} \tag{B6}
$$

and

$$
\begin{aligned}
(S_{oo})_{X_iY_j} &= \sigma_{X_iX_k}(\sigma^{-1})^{X_kY_l}\sigma_{Y_lY_j}, \\
(S_{vv})_{X_aY_b} &= (V^T)_{X_a}^{X_\mu}\left[S(I-RS)\right]_{X_\mu Y_\nu}V_{Y_b}^{Y_\nu}, \\
(S_{ov})_{X_iY_a} &= \sigma_{X_iX_j}\left[(\sigma^{-1})B^TS\right]_{Y_\nu}^{X_j}V_{Y_a}^{Y_\nu}.
\end{aligned} \tag{B7}
$$

In practice, we also have $\sigma_{X_iX_j} = \delta_{ij}$ since on-block orthogonality is enforced. More details about the Hessian derivation can be found in Ref. 139, which carefully derived the orbital Hessian for SCF-MI (a rather similar optimization problem).

The preconditioner we apply to the L-BFGS algorithm is the inverted on-diagonal blocks of the Hessian, i.e., the inverse of $\mathbf{H_{XX}}$ for all the different atom blocks (X). Within the "two-loop" implementation, $\mathbf{H_k^0}$ acts on vector $\mathbf{v} = \mathbf{V_{k-m}V_{k-m+1}} \cdots \mathbf{V_{k-1}g_k}$, which can be divided into contributions from each atom block. Therefore, the application of the preconditioner is equivalent to solving the following linear equation on each atom block:

$$
H_{X_iX_a,X_jX_b}u^{X_jX_b} = v_{X_iX_a}, \tag{B8}
$$

where $\mathbf{u}$ is the preconditioned vector: $\mathbf{u} = \mathbf{H_k^0 v}$. Based on the property of the Hessian matrix (symmetric positive-definite), a preconditioned conjugate-gradient (CG) algorithm is implemented to solve Eq. (B8) iteratively on each atom block. The implemented preconditioner for CG is actually the inverse of the on-diagonal part ($X = Y$) of the last two terms in Eq. (B3).

APPENDIX C: CONSTRUCTION OF PSEUDO-CANONICALIZED MOs UPON MAB-SCF SOLUTION

Once MAB-SCF converges, a Fock matrix in the secondary basis can be built upon the MAB density matrix projected into the secondary basis,

$$\mathbf{F} = \mathbf{F}(\tilde{\mathbf{P}}), \quad \tilde{\mathbf{P}} = \mathbf{B}\mathcal{P}\mathbf{B}^T. \tag{C1}$$

In the current implementation, the PT2 correction is evaluated based on pseudo-canonicalized occupied and virtual MOs, which can be obtained by diagonalizing $\mathbf{F}_{OO}$ and $\mathbf{F}_{VV}$ separately. In fact, the occupied ones are already available in this case, since we can simply project the occupied MOs optimized by MAB-SCF into the secondary basis: $(C_o)^\mu_i = B^\mu_\alpha (C_o)^\alpha_i$. Obviously, $\mathbf{C}_o$ diagonalizes $\mathbf{F}$,

$$(C_o^T)^\mu_i F_{\mu\nu}(C_o)^\nu_j = (C_o^T)^\alpha_i F_{\alpha\beta}(C_o)^\beta_j = \epsilon_i \delta_{ij}. \tag{C2}$$

To obtain the eigenvalues and eigenvectors of $\mathbf{F}_{VV}$, we first form an orthonormal basis that spans the virtual space. If the full but non-redundant span of the secondary basis is represented by $\mathbf{X}$ ($\mathbf{X}^T\mathbf{S}\mathbf{X} = \mathbf{I}$), the demanded vectors can be generated by projecting out the space spanned by occupied MOs,

$$\mathbf{V} = (\mathbf{I} - \tilde{\mathbf{P}}\mathbf{S})\mathbf{X}. \tag{C3}$$

The vectors in $\mathbf{V}$ can be orthonormalized again by performing a canonical orthogonalization (diagonalizing $\mathbf{V}^T\mathbf{S}\mathbf{V}$ will be required). Also, after doing this, the linear dependency of vectors in $\mathbf{V}$ will be eliminated and its column dimension reduces to N_v. We denote the resulting orthonormal basis as $\mathbf{V}'$. Solving the following standard eigenvalue problem

$$(\mathbf{V}'^T \mathbf{F} \mathbf{V}')\mathbf{C}'_v = \boldsymbol{\epsilon}_v \mathbf{C}'_v, \tag{C4}$$

the energies of the pseudo-canonicalized virtual orbitals (ϵ_a's in Eq. (29)) are given by $\boldsymbol{\epsilon}_v$, and their coefficients $\mathbf{C}_v = \mathbf{V}'\mathbf{C}'_v$. The Fock matrix elements coupling between occupied and virtual pseudo-canonicalized MOs ($\mathbf{F}_{OV}$) can be evaluated as

$$F^{(1)}_{ia} = (C_o^T)^\alpha_i (B^T)^\mu_\alpha F_{\mu\nu}(C_v)^\nu_a. \tag{C5}$$

[1]P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2]W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965).
[3]W. Kohn, A. D. Becke, and R. G. Parr, J. Phys. Chem. **100**, 12974 (1996).
[4]A. D. Becke, J. Chem. Phys. **84**, 4524 (1986).
[5]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[6]A. D. Becke, J. Chem. Phys. **109**, 2092 (1998).
[7]J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha, Phys. Rev. Lett. **82**, 2544 (1999).
[8]J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).
[9]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[10]T. Leininger, H. Stoll, H.-J. Werner, and A. Savin, Chem. Phys. Lett. **275**, 151 (1997).
[11]H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540 (2001).
[12]J. Klimeš and A. Michaelides, J. Chem. Phys. **137**, 120901 (2012).
[13]S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, Chem. Rev. **116**, 5105 (2016).
[14]S. Grimme, J. Comput. Chem. **25**, 1463 (2004).
[15]S. Grimme, J. Comput. Chem. **27**, 1787 (2006).
[16]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
[17]K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, Phys. Rev. B **82**, 081101 (2010).
[18]O. A. Vydrov and T. Van Voorhis, J. Chem. Phys. **133**, 244103 (2010).
[19]N. Mardirossian and M. Head-Gordon, Phys. Chem. Chem. Phys. **16**, 9904 (2014).
[20]N. Mardirossian and M. Head-Gordon, J. Chem. Phys. **142**, 074111 (2015).
[21]J. A. Pople, P. M. Gill, and B. G. Johnson, Chem. Phys. Lett. **199**, 557 (1992).
[22]K. S. Raymond and R. A. Wheeler, J. Comput. Chem. **20**, 207 (1999).
[23]A. D. Boese, J. M. L. Martin, and N. C. Handy, J. Chem. Phys. **119**, 3005 (2003).
[24]N. X. Wang and A. K. Wilson, J. Chem. Phys. **121**, 7632 (2004).
[25]N. X. Wang and A. K. Wilson, Mol. Phys. **103**, 345 (2005).
[26]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[27]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[28]T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[29]D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **98**, 1358 (1993).
[30]J. Řezáč and P. Hobza, J. Chem. Theory Comput. **9**, 2151 (2013).
[31]J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 2427 (2011).
[32]J. Řezáč, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 3466 (2012).
[33]D. Rappoport and F. Furche, J. Chem. Phys. **133**, 134105 (2010).
[34]F. B. Van Duijneveldt, J. G. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, Chem. Rev. **94**, 1873 (1994).
[35]L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **106**, 1063 (1997).
[36]R. Haunschild and W. Klopper, J. Chem. Phys. **136**, 164102 (2012).
[37]R. D. Adamson, P. M. Gill, and J. A. Pople, Chem. Phys. Lett. **284**, 6 (1998).
[38]S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen, J. Chem. Phys. **143**, 054107 (2015).
[39]Y. Zhao and D. G. Truhlar, Theor. Chem. Acc. **120**, 215 (2008).
[40]C. A. White, B. G. Johnson, P. M. Gill, and M. Head-Gordon, Chem. Phys. Lett. **230**, 8 (1994).
[41]C. A. White, B. G. Johnson, P. M. Gill, and M. Head-Gordon, Chem. Phys. Lett. **253**, 268 (1996).
[42]C. A. White and M. Head-Gordon, J. Chem. Phys. **104**, 2620 (1996).
[43]M. C. Strain, G. E. Scuseria, and M. J. Frisch, Science **271**, 51 (1996).
[44]M. Challacombe and E. Schwegler, J. Chem. Phys. **106**, 5526 (1997).
[45]C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[46]G. Lippert, J. Hutter, and M. Parrinello, Theor. Chem. Acc. **103**, 124 (1999).
[47]J. Dziedzic, Q. Hill, and C.-K. Skylaris, J. Chem. Phys. **139**, 214103 (2013).
[48]S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).
[49]D. Bowler and T. Miyazaki, Rep. Prog. Phys. **75**, 036503 (2012).
[50]J. VandeVondele and J. Hutter, J. Chem. Phys. **127**, 114105 (2007).
[51]P. E. Maslen, C. Ochsenfeld, C. A. White, M. S. Lee, and M. Head-Gordon, J. Phys. Chem. A **102**, 2215 (1998).
[52]G. Berghold, M. Parrinello, and J. Hutter, J. Chem. Phys. **116**, 1800 (2002).
[53]K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, Chem. Phys. Lett. **240**, 283 (1995).
[54]H. A. Früchtl, R. A. Kendall, R. J. Harrison, and K. G. Dyall, Int. J. Quantum Chem. **64**, 63 (1997).
[55]F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).
[56]A. Sodt, J. E. Subotnik, and M. Head-Gordon, J. Chem. Phys. **125**, 194109 (2006).
[57]A. Sodt and M. Head-Gordon, J. Chem. Phys. **128**, 104106 (2008).
[58]M. Sierka, A. Hogekamp, and R. Ahlrichs, J. Chem. Phys. **118**, 9136 (2003).
[59]S. F. Manzer, E. Epifanovsky, and M. Head-Gordon, J. Chem. Theory Comput. **11**, 518 (2015).
[60]S. Manzer, P. R. Horn, N. Mardirossian, and M. Head-Gordon, J. Chem. Phys. **143**, 024113 (2015).
[61]R. Jurgens-Lutovsky and J. Almlöf, Chem. Phys. Lett. **178**, 451 (1991).
[62]K. Wolinski and P. Pulay, J. Chem. Phys. **118**, 9497 (2003).
[63]W. Liang and M. Head-Gordon, J. Phys. Chem. A **108**, 3206 (2004).
[64]R. P. Steele, R. A. DiStasio, Jr., Y. Shao, J. Kong, and M. Head-Gordon, J. Chem. Phys. **125**, 074108 (2006).
[65]R. P. Steele, Y. Shao, R. A. DiStasio, and M. Head-Gordon, J. Phys. Chem. A **110**, 13915 (2006).
[66]R. P. Steele and M. Head-Gordon, Mol. Phys. **105**, 2455 (2007).
[67]R. P. Steele, R. A. DiStasio, Jr., and M. Head-Gordon, J. Chem. Theory Comput. **5**, 1560 (2009).
[68]J. Deng, A. T. Gilbert, and P. M. Gill, J. Chem. Phys. **130**, 231101 (2009).
[69]J. Deng, A. T. Gilbert, and P. M. Gill, J. Chem. Phys. **133**, 044116 (2010).
[70]J. Deng, A. T. Gilbert, and P. M. Gill, Phys. Chem. Chem. Phys. **12**, 10759 (2010).

[71]R. Mulliken, J. Chem. Phys. **36**, 3428 (1962).

[72]E. R. Davidson, J. Chem. Phys. **46**, 3320 (1967).

[73]K. R. Roby, Mol. Phys. **27**, 81 (1974).

[74]R. Heinzmann and R. Ahlrichs, Theor. Chim. Acta. **42**, 33 (1976).

[75]C. Ehrhardt and R. Ahlrichs, Theor. Chim. Acta. **68**, 231 (1985).

[76]A. E. Reed, L. A. Curtiss, and F. Weinhold, Chem. Rev. **88**, 899 (1988).

[77]M. S. Lee and M. Head-Gordon, Int. J. Quantum Chem. **76**, 169 (2000).

[78]I. Mayer, Can. J. Chem. **74**, 939 (1996).

[79]J. Cioslowski and A. Liashenko, J. Chem. Phys. **108**, 4405 (1998).

[80]W. Lu, C. Wang, M. Schmidt, L. Bytautas, K. Ho, and K. Ruedenberg, J. Chem. Phys. **120**, 2629 (2004).

[81]D. N. Laikov, Int. J. Quantum Chem. **111**, 2851 (2011).

[82]G. Knizia, J. Chem. Theory Comput. **9**, 4834 (2013).

[83]W. H. Adams, J. Chem. Phys. **34**, 89 (1961).

[84]W. H. Adams, J. Chem. Phys. **37**, 2009 (1962).

[85]H. Stoll, G. Wagenblast, and H. Preuβ, Theor. Chim. Acta. **57**, 169 (1980).

[86]R. Z. Khaliullin, M. Head-Gordon, and A. T. Bell, J. Chem. Phys. **124**, 204105 (2006).

[87]P. R. Horn, E. J. Sundstrom, T. A. Baker, and M. Head-Gordon, J. Chem. Phys. **138**, 134119 (2013).

[88]W. Adams, Chem. Phys. Lett. **12**, 295 (1971).

[89]M. S. Lee and M. Head-Gordon, J. Chem. Phys. **107**, 9085 (1997).

[90]M. S. Lee and M. Head-Gordon, Comput. Chem. **24**, 295 (2000).

[91]T. Ozaki, Phys. Rev. B **67**, 155108 (2003).

[92]T. Ozaki and H. Kino, Phys. Rev. B **69**, 195113 (2004).

[93]J. Junquera, Ó. Paz, D. Sánchez-Portal, and E. Artacho, Phys. Rev. B **64**, 235111 (2001).

[94]M. Basanta, Y. Dappe, P. Jelínek, and J. Ortega, Comput. Mater. Sci. **39**, 759 (2007).

[95]D. Bowler and T. Miyazaki, J. Phys.: Condens. Matter **22**, 074207 (2010).

[96]D. Bowler, T. Miyazaki, and M. Gillan, J. Phys.: Condens. Matter **14**, 2781 (2002).

[97]A. Torralba, M. Todorović, V. Brázdová, R. Choudhury, T. Miyazaki, M. Gillan, and D. Bowler, J. Phys.: Condens. Matter **20**, 294206 (2008).

[98]C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Chem. Phys. **122**, 084119 (2005).

[99]C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, J. Phys.: Condens. Matter **20**, 064209 (2008).

[100]C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, O. Diéguez, and M. C. Payne, Phys. Rev. B **66**, 035119 (2002).

[101]A. A. Mostofi, P. D. Haynes, C.-K. Skylaris, and M. C. Payne, J. Chem. Phys. **119**, 8842 (2003).

[102]M. Rayson and P. Briddon, Phys. Rev. B **80**, 205104 (2009).

[103]M. Rayson, Comput. Phys. Commun. **181**, 1051 (2010).

[104]A. Nakata, D. R. Bowler, and T. Miyazaki, J. Chem. Theory Comput. **10**, 4813 (2014).

[105]L. Lin, J. Lu, L. Ying, and E. Weinan, J. Comput. Phys. **231**, 2140 (2012).

[106]L. Lin, J. Lu, L. Ying, and E. Weinan, J. Comput. Phys. **231**, 4515 (2012).

[107]M. Head-Gordon, P. E. Maslen, and C. A. White, J. Chem. Phys. **108**, 616 (1998).

[108]J. Hutter, M. Parrinello, and S. Vogel, J. Chem. Phys. **101**, 3862 (1994).

[109]See supplementary material at http://dx.doi.org/10.1063/1.4959125 for more information about how the "mixing" parameter α (on page 6) is chosen, the effect of choosing different reference bases, the consequence of using the fully automated "adding vector" algorithm for all systems, and the comparison of MAB-SCF(PC) with dual-basis SCF method.

[110]Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng *et al.*, Mol. Phys. **113**, 184 (2015).

[111]J. Nocedal and S. Wright, *Numerical Optimization* (Springer Science & Business Media, 2006).

[112]D. C. Liu and J. Nocedal, Math. Program. **45**, 503 (1989).

[113]P. M. Gill, B. G. Johnson, and J. A. Pople, Chem. Phys. Lett. **209**, 506 (1993).

[114]T. Van Voorhis and M. Head-Gordon, Mol. Phys. **100**, 1713 (2002).

[115]J. Sun, R. Haunschild, B. Xiao, I. W. Bulik, G. E. Scuseria, and J. P. Perdew, J. Chem. Phys. **138**, 044113 (2013).

[116]Y. Zhao and D. G. Truhlar, J. Chem. Phys. **125**, 194101 (2006).

[117]V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, J. Chem. Phys. **119**, 12129 (2003).

[118]C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).

[119]J.-D. Chai and M. Head-Gordon, Phys. Chem. Chem. Phys. **10**, 6615 (2008).

[120]F. Jensen, J. Chem. Phys. **115**, 9113 (2001).

[121]F. Jensen, J. Chem. Phys. **116**, 7372 (2002).

[122]F. Jensen, J. Chem. Phys. **117**, 9234 (2002).

[123]R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).

[124]M. J. Frisch, J. A. Pople, and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984).

[125]A. Karton, S. Daon, and J. M. L. Martin, Chem. Phys. Lett. **510**, 165 (2011).

[126]L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).

[127]L. Goerigk and S. Grimme, J. Chem. Theory Comput. **7**, 291 (2011).

[128]Y. Zhao, N. González-García, and D. G. Truhlar, J. Phys. Chem. A **109**, 2012 (2005).

[129]Y. Zhao, B. J. Lynch, and D. G. Truhlar, Phys. Chem. Chem. Phys. **7**, 43 (2005).

[130]J. M. L. Martin, J. Phys. Chem. A **117**, 3118 (2013).

[131]P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).

[132]M. S. Marshall, L. A. Burns, and C. D. Sherrill, J. Chem. Phys. **135**, 194102 (2011).

[133]J. Řezáč and P. Hobza, J. Chem. Theory Comput. **8**, 141 (2012).

[134]K. U. Lao and J. M. Herbert, J. Chem. Phys. **139**, 034107 (2013).

[135]K. U. Lao and J. M. Herbert, J. Phys. Chem. A **119**, 235 (2015).

[136]J. Witte, J. B. Neaton, and M. Head-Gordon, J. Chem. Phys. **144**, 194306 (2016).

[137]R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay, and P. Hobza, J. Chem. Theory Comput. **9**, 3364 (2013).

[138]D. L. Strout and G. E. Scuseria, J. Chem. Phys. **102**, 8448 (1995).

[139]P. R. Horn and M. Head-Gordon, J. Chem. Phys. **143**, 114111 (2015).