

Untangling Sources of Error in the Density-Functional Many-Body Expansion

Dustin R. Broderick and John M. Herbert*



Cite This: *J. Phys. Chem. Lett.* 2025, 16, 2793–2799



Read Online

ACCESS |



Metrics & More

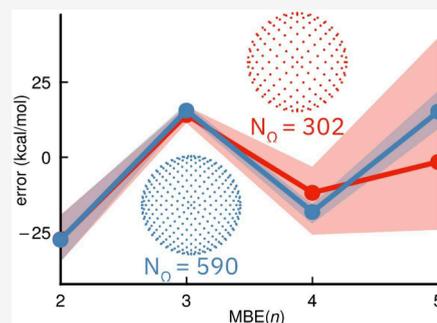


Article Recommendations



Supporting Information

ABSTRACT: The many-body expansion provides a framework for data-driven applications of electronic structure theory, including parametrization of classical force fields and machine learning. However, we demonstrate that its use significantly amplifies quadrature grid errors when modern density-functional approximations are employed. Standard grids that work well in conventional density-functional calculations result in runaway error accumulation when used with the many-body expansion. At the same time, delocalization error is also exacerbated, leading to exaggerated estimates of nonadditive n -body interactions. This is illustrated for anion–water clusters using the SCAN, r^2 SCAN, ω B97X-V and ω B97M-V functionals. By employing dense quadrature grids, the inherent self-interaction error is exposed, which can then be mitigated using a variety of other strategies.



Fragment-based approximations, anchored in the many-body expansion (MBE), are an appealing means to sidestep the steep nonlinear scaling of *ab initio* quantum chemistry.¹ There is significant interest in using the MBE to decompose interaction energies in molecular liquids,^{2–5} biomolecules,^{6–8} and other complex systems,^{9,10} as a systematic means to generate training data for machine learning applications. However, a growing body of work demonstrates that care must be taken to avoid error accumulation in MBE-based methods.^{11–15}

The enormity of the resource requirements for data-hungry machine learning applications places a premium on low-cost electronic structure models such as density functional theory (DFT), yet we have recently demonstrated that self-interaction error (SIE) causes catastrophic failure of the DFT-based MBE.¹⁶ In the present work, we illustrate how SIE is intermingled with quadrature grid errors, in a manner that is unique to the MBE and does not manifest in conventional DFT calculations. The functionals most strongly affected are meta-generalized gradient approximations (meta-GGAs), which represent many of the most accurate contemporary exchange-correlation functionals,¹⁷ although functionals such as ω B97X-V¹⁸ are also impacted. Using calculations on ion–water clusters, we demonstrate that grid-based error dominates higher-order MBE calculations that employ meta-GGAs and B97-based functionals. This obfuscates other sources of error, such as SIE, which must be addressed for accurate DFT-MBE calculations. Improving the grid quality brings SIE-based delocalization error into focus.

The MBE can be expressed as

$$E = \sum_{I=1}^N E_I + \sum_{I=1}^N \sum_{J>I} \Delta E_{IJ} + \sum_{I=1}^N \sum_{J>I} \sum_{K>J} \Delta E_{IJK} + \dots \quad (1)$$

where the E_I are single-fragment (monomer) energies,

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \quad (2)$$

is a two-body correction, etc.¹ If eq 1 is truncated at n -body interactions, then we call the resulting method MBE(n). In conjunction with high-quality basis sets and correlated wave function models, MBE(4) affords good accuracy for neat liquid water and monovalent ion–water interactions,^{13,19–21} yet requires electronic structure calculations on systems no larger than $(\text{H}_2\text{O})_4$ or $\text{X}^\pm(\text{H}_2\text{O})_3$. However, four-body calculations manifest a crippling $O(N^4)$ combinatorial prefactor, resulting in error accumulation for large systems.^{11–15}

Replacing wave function methods with DFT reduces the cost but SIE becomes catastrophic, with fluctuations as large as ± 200 kcal/mol in low-order MBE(n) calculations on $\text{F}^-(\text{H}_2\text{O})_{15}$ clusters.¹⁶ (Note that SIE is especially pernicious for anions.^{22–30}) This effect is only marginally reduced by a combination of standard hybrid functionals (such as PBE0 or B3LYP) and aggressive screening of the n -body subsystems.¹⁶ For functionals including SCAN³¹ and ω B97X-V,¹⁸ this strategy is only moderately effective. Because meta-GGAs

Received: December 18, 2024

Revised: February 27, 2025

Accepted: March 5, 2025

and B97-based functionals are known to have stringent grid requirements,^{32–41} we decided to revisit the quadrature grids used in DFT-based MBE calculations, even though previous MBE(4) calculations at the B3LYP/cc-pVDZ level suggested that the quality of the integration grid had a negligible effect on accuracy.¹²

As n increases, $\text{MBE}(n)$ should converge to same interaction energy (ΔE_{int}) as a supramolecular calculation at the same level of theory, because we employ single-monomer fragments so that no covalent bonds are severed. As such, it makes sense to define the error in the $\text{MBE}(n)$ approximation as

$$\text{error} = \Delta E_{\text{int}}^{\text{MBE}(n)} - \Delta E_{\text{int}}^{\text{supersystem}} \quad (3)$$

where ΔE_{int} is the ion–water interaction energy for a $\text{X}^-(\text{H}_2\text{O})_{15}$ cluster, and both calculations in eq 3 are performed using the same functional and basis set. As a control experiment, we demonstrate in Figure 1 that

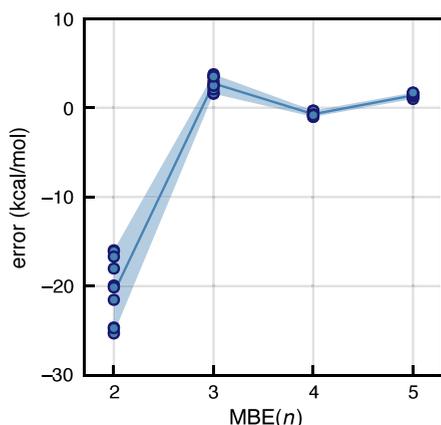


Figure 1. $\text{MBE}(n)$ errors in the ion–water interaction energy ΔE_{int} for 11 configurations of $\text{F}^-(\text{H}_2\text{O})_{15}$ computed at the HF/aug-cc-pVDZ level. The solid line connects mean errors at each value of n , and the shaded region highlights the range of the data.

convergence is indeed achieved for a set of $\text{F}^-(\text{H}_2\text{O})_{15}$ clusters, using calculations at the Hartree–Fock (HF)/aug-cc-pVDZ level. Residual errors spanning ≈ 2.5 kcal/mol, observed at the five-body level in Figure 1, have elsewhere been shown to be artifacts of basis-set superposition error (BSSE).^{19–21,42,43} For example, five-body terms for $\text{Cl}^-(\text{H}_2\text{O})_{15}$, computed at the HF/aug-cc-pVDZ level (Figure S1), are similar to those computed elsewhere for $\text{Cl}^-(\text{H}_2\text{O})_9$, using second-order Møller–Plesset perturbation theory (MP2) in the same basis set.²⁰ However, these five-body terms disappear when a BSSE correction is added or when the basis set is extended to aug-cc-pV5Z.²⁰

HF-MBE(n) calculations, for which there is no grid and no SIE, will serve as a baseline for the remaining discussion. Analogous PBE-MBE(n) data for $\text{F}^-(\text{H}_2\text{O})_{15}$ are shown in Figure 2, superimposed on the span of the corresponding HF-MBE(n) errors. As demonstrated in previous work,¹⁶ the PBE-MBE(n) errors diverge as n increases, with errors approaching ~ 150 kcal/mol at the five-body level. Data are provided for several different quadrature grids and errors change by more than 200 kcal/mol between $n = 4$ and $n = 5$, regardless of the choice of grid. We conclude that these oscillations are driven by SIE, as reported previously.¹⁶

The fluoride ion is a problematic case for many semilocal functionals in that its highest occupied molecular orbital

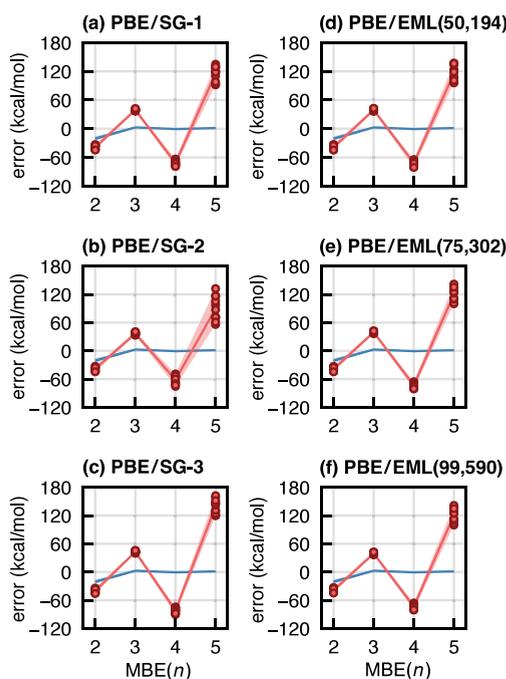


Figure 2. $\text{MBE}(n)$ errors in ΔE_{int} (circles), for 11 configurations of $\text{F}^-(\text{H}_2\text{O})_{15}$ computed at the PBE/aug-cc-pVDZ level using various quadrature grids: (a) SG-1, (b) SG-2, (c) SG-3, (d) EML(50,194), (e) EML(75,302), and (f) EML(99,590). The red shaded region and solid line connect the range of the $\text{MBE}(n)$ errors and their mean. In blue are the mean errors and their range computed at the HF/aug-cc-pVDZ level (from Figure 1). On this scale, the range of the HF-MBE(n) data is barely discernible.

(HOMO) is often unbound in large basis sets.^{22–25} Indeed, gas-phase F^- is unbound ($\epsilon_{\text{HOMO}} > 0$) at both the PBE/aug-cc-pVDZ and SCAN/aug-cc-pVDZ levels of theory, although it is bound by 3.0–3.4 eV in $\text{F}^-(\text{H}_2\text{O})_{15}$. Gas-phase F^- is bound in $\omega\text{B97X-V}$ and $\omega\text{B97M-V}$ calculations using aug-cc-pVDZ ($\epsilon_{\text{HOMO}} = -2.7$ eV), nevertheless these functionals still exhibit problems in the context of DFT-MBE(n) calculations, as documented below.

Meta-GGA functionals depend on the Laplacian of the electron density and/or the kinetic energy density, both of which are more oscillatory than the density gradients and thus more challenging for numerical integration, while the inhomogeneity factor in B97 makes $\omega\text{B97X-V}$ challenging to integrate as well.³⁴ For these functionals, the use of low-quality grids may cause oscillations in potential energy surfaces for noncovalent dimers.^{32–34} Before discussing grid quality, we first review the nomenclature for DFT quadrature grids.

Euler-Maclaurin-Lebedev (EML) quadrature grids have been used since the earliest days of molecular DFT,⁴⁴ and are indicated using the notation $\text{EML}(N_r, N_\Omega)$. Here, N_r is the number of radial shells (Euler-Maclaurin quadrature) on interval $[0, \infty)$, and N_Ω is the number of angular points per shell (Lebedev quadrature). We examine $(N_r, N_\Omega) = (50, 194)$, $(75, 302)$, and $(99, 590)$, representing standard choices for a low-quality, medium-quality, and high-quality grid, respectively.³⁴ The “standard grids” (SG- k) are pruned to reduce the number of Lebedev grid points near the nuclei (where the density is nearly spherically symmetric) and far away from the nuclei (where the it is slowly varying).^{34,45} The SG-1 grid,⁴⁵ for example, starts from $N_r = 50$ and $N_\Omega = 194$ and applies a pruning procedure; this grid is typically adequate for

GGAs.^{17,34} Importantly, SG-2 and SG-3 are pruned versions of (75,302) and (99,590) grids but they use a double-exponential quadrature for the radial integration.^{46,47} As such, they are not simply pruned versions of EML(75,302) and EML(99,590).³⁴

The SCAN functional³¹ is considered to be especially sensitive to grid choice, which has led to development of regularized versions including rSCAN and r²SCAN.^{35–37} However, much of the conventional wisdom regarding SCAN's grid sensitivity comes from the condensed-matter (periodic DFT) community, where grid quality is generally lower than what is typically used by default in electronic structure codes based on atom-centered basis functions. Prior to the development of SG-2 and SG-3, the EML(75,302) had been recommended for meta-GGA calculations,¹⁷ and subsequently SG-2 was found to be adequate for many meta-GGAs.³⁴ (Minnesota functionals are a categorical exception,³⁴ due to extreme oscillations in the exchange inhomogeneity factor.⁴⁸)

To demonstrate that SG-2 and SG-3 are adequate even for the SCAN functional, we computed ion–water interaction energies for F[−](H₂O)₁₅ and Cl[−](H₂O)₁₅ clusters using conventional, monolithic DFT calculations with a variety of grids, taking EML(250,974) as a benchmark.³⁴ Results in Tables S1–S2 and Figures S2–S3 indicate that interaction energies computed using SG-2 or SG-3 are converged to within $\lesssim 0.1$ kcal/mol. The interaction energies themselves are $\Delta E_{\text{int}} \sim 70$ kcal/mol for Cl[−](H₂O)₁₅ and $\Delta E_{\text{int}} \sim 120$ kcal/mol for F[−](H₂O)₁₅, so grid errors using SG-2 and SG-3 amount to $\lesssim 0.1\%$ of ΔE_{int} . As such, these standard grids are effectively converged for conventional DFT calculations, even for the SCAN functional.

For DFT-MBE(*n*), however, $\sim 10^4$ subsystem calculations (or more) might be required to evaluate a single-point energy. The robustness documented above for monolithic DFT calculations does not hold for MBE(*n*). Even for the regularized r²SCAN functional that was designed to mitigate grid sensitivity, calculations using SG-2 afford errors for F[−](H₂O)₁₅ that fluctuate over a range of 75 kcal/mol for MBE(5); see Figure 3. This is qualitatively different from the behavior of PBE-MBE(*n*) with the same grids. In particular, the PBE-MBE(*n*) errors have strictly alternating signs as a function of *n*, regardless of quadrature grid (Figure 2), which is not true for r²SCAN-MBE(*n*) calculations using SG-2 (Figure 3a). The difference lies in the fact that the PBE errors are relatively tightly clustered across the ensemble of structures, albeit with considerable spread for *n* = 5, whereas r²SCAN errors span a much larger range for *n* = 5, to the point that the error may be positive or negative depending on the cluster geometry.

When SG-2 is replaced by the EML(75,302) grid, the errors for r²SCAN-MBE(*n*) become much more tightly clustered and alternating signs are recovered; see Figure 3b. This behavior is not unique to r²SCAN. It is observed also for SCAN (Figure S4), ω B97X-V (Figure S5), and ω B97M-V (Figure S6). In this regard, the regularization used in r²SCAN is not effective in reducing the grid dependence of DFT-MBE(*n*) calculations. In fact, the SCAN-MBE(*n*) and r²SCAN-MBE(*n*) results are extremely similar, as documented in Tables S5–S6.

Alternating signs and relatively tight clustering of the errors are restored for SCAN- and r²SCAN-MBE(*n*) calculations using either SG-3 or EML(75,302). In contrast, MBE(5) errors obtained using SG-2 appear to be almost random across cluster geometries. The same is true for ω B97X-V (Figure S5) and

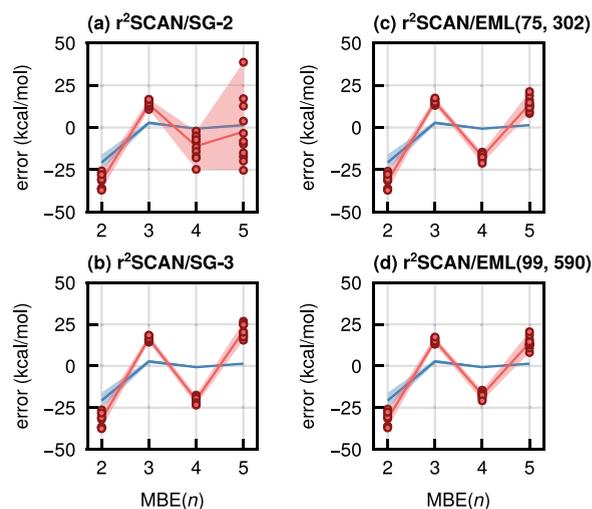


Figure 3. MBE(*n*) errors in ΔE_{int} (circles), for 11 configurations of F[−](H₂O)₁₅ computed at the r²SCAN/aug-cc-pVDZ level using various quadrature grids: (a) SG-1, (b) SG-2, (c) SG-3, (d) EML(50,194), (e) EML(75,302), and (f) EML(99,590). The red shaded region and solid line connect the range of the errors and their mean, for each value of *n*. In blue are the HF/aug-cc-pVDZ errors from Figure 1.

ω B97M-V (Figure S6), suggesting that DFT-MBE(*n*) calculations based on these hard-to-integrate functionals are dominated by grid error when SG-2 is used. For F[−](H₂O)₁₅, this problem seems to be eliminated by using SG-3 but it persists in Cl[−](H₂O)₁₅ (Figures S8–S11), with MBE(5) errors spanning a range of ~ 30 kcal/mol. For Cl[−](H₂O)₁₅, EML(75,302) is the smallest grid that affords tightly clustered errors across all geometries.

The influence of grid error on individual *n*-body terms is demonstrated in Figures S12–S13 by way of histograms for all subsystems derived from the F[−](H₂O)₁₅ clusters, considering both SCAN and r²SCAN. We define grid error (ϵ) with respect to an EML(250,974) benchmark and separate the error into individual terms $\Delta \epsilon_{IJ\dots}$ for each subsystem. For the SG-2 grid, and to a lesser extent the SG-3 grid, the magnitude of the grid error is independent of the many-body order (*n*), and all of the values $\Delta \epsilon_{IJ\dots}$ are contained between ± 0.1 kcal/mol. For the systems considered here, where all of the subsystems are relatively small, the grid-dominated errors are characterized by a small random error per subsystem, without much dependence on subsystem size.

This observation helps to explain why the error pattern for PBE (Figure 2) is qualitatively different from that of r²SCAN (Figure 3). SIE drives delocalization that stabilizes clusters relative to their constituent monomers, and stabilizes larger clusters relative to smaller ones. The lack of SIE in the smaller fragments is overrepresented in the MBE(*n*) calculations as system size and expansion order increase. This is more readily seen if the *n*-body corrections are written in closed form,¹²

$$E_{\text{MBE}(n)} = \sum_{k=1}^n (-1)^{n-k} \binom{N-k-1}{n-k} \sum_{\alpha=1}^{N C_k} E_{\alpha}^{(k)} \quad (4)$$

Here, α indexes the subsystems with *k* fragments each, whose individual energies are $E_{\alpha}^{(k)}$, and $N C_k$ is a binomial coefficient equal to the number of *k*-body subsystems. In this form, each fragment energy is scaled by a signed combinatorial coefficient.

As an example, these coefficients are listed in Table S3 for $N = 16$, as in $F^-(H_2O)_{15}$.

For $k < n$, the signs alternate with n for fixed k and increase in magnitude with n . Because the SIE inherent to $E_a^{(k)}$ is consistently stabilizing, the alternating signs in the lower-order terms will overwhelm the always-positive $n = k$ terms, leading to the pattern of errors that is observed in the PBE-MBE(n) calculations. Note that oscillatory signs in the MBE(n) corrections need not result in a divergent expansion, as demonstrated by the HF/aug-cc-pVDZ data in Figure 1. In that case, BSSE is strictly stabilizing for larger subsystems, nevertheless errors become smaller as n increases. This suggests that there exists some n -dependent threshold for the errors $\Delta E_{IJ\dots}$, related to the combinatorial coefficients in eq 4, beyond which MBE(n) calculations will diverge.

Contributions from individual n -body corrections $\Delta E_{IJ\dots}$ should decrease in magnitude with increasing n . The corresponding grid errors $\Delta \epsilon_{IJ\dots}$ arise from numerical integration artifacts and need not be strictly stabilizing or decrease with n , hence their pattern is more scattered. This is the case for r^2 SCAN-based MBE(5) calculations using SG-2 (Figure 3a). Only upon saturating the grid does SIE become the dominant source of error, leading to tightly clustered data points at each MBE(n) order. Errors then oscillate with n , as seen in Figure 3b. These oscillations can be quite large. For example, the error changes by almost 50 kcal/mol between $n = 4$ and $n = 5$, even when the EML(99,590) grid is used (Figure 3d). This is a manifestation of SIE.

Experience with various self-interaction corrections has demonstrated that SCAN is less sensitive to the underlying density as compared to PBE,^{49–51} which may suggest that SCAN has less SIE. If so, then this explains the reduction in SIE-driven oscillations that we observe for SCAN (Figure S4d) and for r^2 SCAN (Figure 3d), relative those for PBE (Figure 2f). In previous work,¹⁶ we tested several strategies to mitigate accumulation of SIE, including hybrid functionals, dielectric boundary conditions,⁵² and density-corrected (DC-)DFT.^{53–55} These procedures reduced the errors in DFT-MBE(n) calculations but none was sufficient to restore convergence, at least not for standard hybrid functionals with 20–25% exact exchange. Functionals with 50% exact exchange did afford convergent results, indicating that SIE plays a prominent role in the divergent behavior.¹⁶ However, calculations in ref 16 used the SG-2 quadrature grid, meaning that SIE-driven errors were intertwined with MBE-induced grid errors. We next revisit these calculations using EML(99,590). MBE(n) errors for $F^-(H_2O)_{15}$ clusters are plotted in Figure 4 for SCAN-based calculations with EML(99,590), using a variety of SIE remediation strategies that are described below.

Large-molecule GGA calculations are often difficult to converge due to “charge sloshing” in the self-consistent field (SCF) iterations,^{56–60} which can be mitigated by electrostatic stabilization of the molecular surface.^{60–62} This provides a driving force to localize charge that counteracts SIE’s tendency to delocalize charge, and we have shown that DFT-MBE(n) calculations on proteins require low-dielectric boundary conditions in order to obtain converged results in the presence of ionic residues.⁶² Following previous work,^{15,62,63} we implement dielectric boundaries using a polarizable continuum model (PCM) with a dielectric constant $\epsilon = 4$.^{64–66} This simple modification significantly reduces SCAN-MBE(n) errors, as shown in Figure 4b. For $n \geq 3$, the errors are

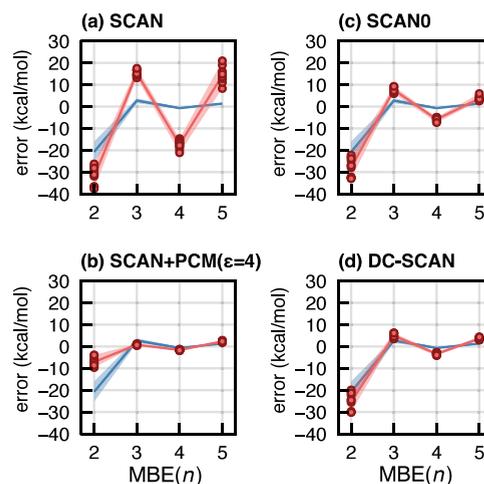


Figure 4. MBE(n) errors in ΔE_{int} (circles), for 11 configurations of $F^-(H_2O)_{15}$ computed using (a) SCAN, (b) SCAN + PCM($\epsilon = 4$), (c) SCAN0 with 25% exact exchange, and (d) DC-SCAN. All calculations used the aug-cc-pVDZ basis set and an unpruned EML(99,590) grid. Red lines connect the mean errors at each value of n , and the red shaded region highlights the span of the data. The blue line and shaded region represent the HF/aug-cc-pVDZ errors from Figure 1.

comparable to (and in some cases smaller than) the corresponding HF-MBE(n) errors.

The SCAN0 functional,⁶⁷ with 25% exact exchange, is an alternative that also results in convergent behavior for MBE(n), as shown in Figure 4c. This is not the case when SG-2 is used.¹⁶ For the latter grid, SCAN0-MBE(5) errors for this same set of $F^-(H_2O)_{15}$ clusters span a range of almost 250 kcal/mol and lack the telltale alternating signs that indicate suppression of grid errors.¹⁶ Nevertheless, oscillations in SCAN0-MBE(n) using the EML(99,590) grid remain larger than the corresponding HF-MBE(n) values. We attribute this to residual delocalization error in SCAN0, since HF-MBE(n) results should provide an estimate of errors due to finite-basis effects, specifically, mismatch between subsystem and super-system BSSE.^{13,21} The SCAN0-MBE(5) errors in Figure 4c are larger than any such effects.

In the DC-DFT approach,^{53–55} also known as the DFT@HF method,⁶⁸ an exchange-correlation functional is evaluated in a one-shot, non-self-consistent fashion using a converged HF density. This can significantly reduce SIE-driven errors,^{53–55,68–72} although it may be an overcorrection.⁶⁸ In particular, DC-SCAN has been suggested for development of MBE(n)-based force fields.^{71,72} Results in Figure 4d show errors on the order of 5 kcal/mol for MBE(4) and MBE(5) calculations based on DC-SCAN, which is a bit larger than the corresponding HF-MBE(n) errors but slightly smaller than SCAN0-MBE(n) errors. These residual errors are more surprising in the context of DC-SCAN, which uses a SIE-free density. Enhanced fluctuations in MBE(n) using DC-SCAN may arise from electron correlation effects and a more appropriate comparison for DC-DFT might be MP2, which is SIE-free but includes correlation. Indeed, MP2-MBE(n) fluctuations for $F^-(H_2O)_{15}$ clusters are larger than those observed for HF-MBE(n) calculations.²¹

Finally, we discuss how the choice of quadrature grid impacts the computational expense. The cost of the quadrature step is proportional to the number of grid points and

ultimately $O(N)$ with molecular size, although it can be a significant fraction of the computational time for semilocal functionals.³⁴ The cost of SCAN-MBE(4) calculations on $F^-(H_2O)_{15}$ is provided in Table 1 for various quadrature grids.

Table 1. Timings for MBE(4) Calculations on $F^-(H_2O)_{15}$ ^a

Grid	Time (sec) ^b
SG-2	59,176 ± 660
SG-3	96,282 ± 1205
EML(75,302)	99,744 ± 1450
EML(99,590)	202,371 ± 2282

^aSCAN/aug-cc-pVDZ level of theory. ^bAverage over 11 geometries. Uncertainties represent one standard deviation.

SG-2 is deficient but is included as a baseline, since it is a typical choice in atom-centered quantum chemistry codes.³⁴ The EML(75,302) grid is 1.7× more expensive although this cost can be significantly reduced by introducing a sensible screening protocol to cull unnecessary subsystems in DFT-MBE(*n*) calculations.²¹

In summary, the present work serves as a cautionary note that supramolecular tests are insufficient to reveal grid-based convergence problems in DFT-MBE(*n*) calculations. This is analogous to the manner in which MBE(*n*) is more sensitive to numerical thresholds as compared to supersystem calculations at the same level of theory.^{11–13} While modern density functionals such as SCAN, r²SCAN, ωB97X-V, and ωB97M-V have distinct advantages as compared to functionals from earlier generations of DFT, their more stringent grid requirements are significantly amplified by the MBE. In particular, pruned grids that are otherwise recommended for these functionals³⁴ can engender cumulative errors that mask other artifacts, notably SIE. To eliminate the grid errors, we recommend an unpruned grid such as EML(75,302).

This does not eliminate SIE-driven delocalization errors, and MBE(*n*) calculations using the aforementioned functionals remain divergent even if the order-by-order fluctuations are greatly reduced by means of high-quality integration grids. By saturating the grid and removing it as a source of error, other strategies for mitigating SIE can gain a foothold. These include the use of hybrid functionals such as SCAN0, low-dielectric boundary conditions (based on a PCM with ε = 4), and DC-DFT.

METHODS

Geometries for $F^-(H_2O)_{15}$ and $Cl^-(H_2O)_{15}$ were obtained from a molecular dynamics simulation of the ion in bulk water. All calculations were performed using the FRAGMENT code,⁷³ interfaced to Q-CHEM v. 6.2.⁷⁴ Timing data were obtained using dedicated compute nodes with two Intel Xeon CPU Max 9470 processors (52 cores each) with 128 Gb of memory. Each fragment calculation was provisioned to use 4 cores and 5 Gb of memory.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcllett.4c03619>.

Additional calculations and data (PDF)

Coordinates for $F^-(H_2O)_{15}$ and $Cl^-(H_2O)_{15}$ clusters (ZIP)

Transparent Peer Review report available (PDF)

AUTHOR INFORMATION

Corresponding Author

John M. Herbert – Department of Chemistry & Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States; orcid.org/0000-0002-1663-2278; Email: herbert@chemistry.ohio-state.edu

Author

Dustin R. Broderick – Department of Chemistry & Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpcllett.4c03619>

Notes

The authors declare the following competing financial interest(s): J.M.H. is part owner of Q-Chem Inc. and serves on its board of directors.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award No. DE-SC0008550. Calculations were performed at the Ohio Supercomputer Center.⁷⁵

REFERENCES

- Herbert, J. M. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.* **2019**, *151*, 170901.
- Riera, M.; Knight, C.; Bull-Vulpe, E. F.; Zhu, X.; Agnew, H.; Smith, D. G. A.; Simmonett, A. C.; Paesani, F. MBX: A many-body energy and force calculator for data-driven many-body simulations. *J. Chem. Phys.* **2023**, *159*, 054802.
- Yao, K.; Herr, J. E.; Parkhill, J. The many-body expansion combined with neural networks. *J. Chem. Phys.* **2017**, *146*, 014106.
- Bull-Vulpe, E. F.; Riera, M.; Götz, A. W.; Paesani, F. MB-Fit: Software infrastructure for data-driven many-body potential energy functions. *J. Chem. Phys.* **2021**, *155*, 124801.
- Maldonado, A. M.; Poltavsky, I.; Vassilev-Galindo, V.; Tkatchenko, A.; Keith, J. A. Modeling molecular ensembles with gradient-domain machine learning force fields. *Digital Discov* **2023**, *2*, 871–880.
- Li, W.; Ma, H.; Li, S.; Ma, J. Computational and data driven molecular material design assisted by low scaling quantum mechanics calculations and machine learning. *Chem. Sci.* **2021**, *12*, 14987–15006.
- Cheng, Z.; Du, J.; Zhang, L.; Ma, J.; Li, W.; Li, S. Building quantum mechanics quality force fields of proteins with the generalized energy-based fragmentation approach and machine learning. *Phys. Chem. Chem. Phys.* **2022**, *24*, 1326–1337.
- Liao, K.; Dong, S.; Cheng, Z.; Li, W.; Li, S. Combined fragment-based machine learning force field with classical force field and its application in the NMR calculations of macromolecules in solutions. *Phys. Chem. Chem. Phys.* **2022**, *24*, 18559–18567.
- Ricard, T. C.; Zhu, X.; Iyengar, S. S. Capturing weak interactions in surface adsorbate systems at coupled cluster accuracy: A graph-theoretic molecular fragmentation approach improved through machine learning. *J. Chem. Theory Comput.* **2023**, *19*, 8541–8556.
- Iyengar, S. S.; Ricard, T. C.; Zhu, X. Reformulation of all ONIOM-type molecular fragmentation approaches and many-body theories using graph-theory-based projection operators: Applications to dynamics, molecular potential surfaces, machine learning, and quantum computing. *J. Phys. Chem. A* **2024**, *128*, 466–478.

- (11) Richard, R. M.; Lao, K. U.; Herbert, J. M. Aiming for benchmark accuracy with the many-body expansion. *Acc. Chem. Res.* **2014**, *47*, 2828–2836.
- (12) Richard, R. M.; Lao, K. U.; Herbert, J. M. Understanding the many-body expansion for large systems. I. Precision considerations. *J. Chem. Phys.* **2014**, *141*, 014108.
- (13) Lao, K. U.; Liu, K.-Y.; Richard, R. M.; Herbert, J. M. Understanding the many-body expansion for large systems. II. Accuracy considerations. *J. Chem. Phys.* **2016**, *144*, 164105.
- (14) Liu, K.-Y.; Herbert, J. M. Understanding the many-body expansion for large systems. III. Critical role of four-body terms, counterpoise corrections, and cutoffs. *J. Chem. Phys.* **2017**, *147*, 161729.
- (15) Bowling, P. E.; Broderick, D. R.; Herbert, J. M. Convergent protocols for protein–ligand interaction energies using fragment-based quantum chemistry. *J. Chem. Theory Comput.* **2025**, *21*, 951–966.
- (16) Broderick, D. R.; Herbert, J. M. Delocalization error poisons the density-functional many-body expansion. *Chem. Sci.* **2024**, *15*, 19893–19906.
- (17) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315–2372.
- (18) Mardirossian, N.; Head-Gordon, M. ω B97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904–9924.
- (19) Heindel, J. P.; Xantheas, S. S. The many-body expansion for aqueous systems revisited: I. Water–water interactions. *J. Chem. Theory Comput.* **2020**, *16*, 6843–6855.
- (20) Heindel, J. P.; Xantheas, S. S. The many-body expansion for aqueous systems revisited: II. Alkali metal and halide ion–water interactions. *J. Chem. Theory Comput.* **2021**, *17*, 2200–2216.
- (21) Broderick, D. R.; Herbert, J. M. Scalable generalized screening for high-order terms in the many-body expansion: Algorithm, open-source implementation, and demonstration. *J. Chem. Phys.* **2023**, *159*, 174801.
- (22) Schwarz, K. Instability of stable negative ions in the $X\alpha$ method or other local density functional schemes. *Chem. Phys. Lett.* **1978**, *57*, 605–607.
- (23) Galbraith, J. M.; Schaefer, H. F. Concerning the applicability of density functional methods to atomic and molecular ions. *J. Chem. Phys.* **1996**, *105*, 862–864.
- (24) Rösch, N.; Trickey, S. B. Comment on “Concerning the applicability of density functional methods to atomic and molecular negative ions. *J. Chem. Phys.* **1997**, *106*, 8940–8941.
- (25) Jaręcki, A. A.; Davidson, E. R. Density functional theory calculations for F^- . *Chem. Phys. Lett.* **1999**, *300*, 44–52.
- (26) Lee, D.; Burke, K. Finding electron affinities with approximate density functionals. *Mol. Phys.* **2010**, *108*, 2687–2701.
- (27) Lee, D.; Furche, F.; Burke, K. Accuracy of electron affinities of atoms in approximate density functional theory. *J. Phys. Chem. Lett.* **2010**, *1*, 2124–2129.
- (28) Jensen, F. Describing anions by density functional theory: Fractional electron affinity. *J. Chem. Theory Comput.* **2010**, *6*, 2726–2735.
- (29) Herbert, J. M. The quantum chemistry of loosely-bound electrons. In *Reviews in Computational Chemistry*; Parill, A. L., Lipkowitz, K., Eds.; Wiley-VCH: Hoboken, NJ, 2015; Vol. 28, pp 391–517 DOI: 10.1002/9781118889886.ch8.
- (30) Peach, M. J. G.; Teale, A. M.; Helgaker, T.; Tozer, D. J. Fractional electron loss in approximate DFT and Hartree–Fock theory. *J. Chem. Theory Comput.* **2015**, *11*, 5262–5268.
- (31) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.
- (32) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. Application of 25 density functionals to dispersion-bound homomolecular dimers. *Chem. Phys. Lett.* **2004**, *394*, 334–338.
- (33) Johnson, E. R.; Becke, A. D.; Sherrill, C. D.; DiLabio, G. A. Oscillations in meta-generalized-gradient approximation potential energy surfaces for dispersion-bound complexes. *J. Chem. Phys.* **2009**, *131*, 034111.
- (34) Dasgupta, S.; Herbert, J. M. Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3. *J. Comput. Chem.* **2017**, *38*, 869–882.
- (35) Bartók, A. P.; Yates, J. R. Regularized SCAN functional. *J. Chem. Phys.* **2019**, *150*, 161101.
- (36) Mejía-Rodríguez, D.; Trickey, S. B. Comment on “Regularized SCAN functional”. *J. Chem. Phys.* **2019**, *151*, 207101.
- (37) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and numerically efficient r^2 SCAN meta-generalized gradient approximation. *J. Phys. Chem. Lett.* **2020**, *11*, 8208–8215.
- (38) Lehtola, S.; Marques, M. A. L. Many recent density functionals are numerically ill-behaved. *J. Chem. Phys.* **2022**, *157*, 174114.
- (39) Lehtola, S.; Marques, M. A. L. Reproducibility of density functional approximations: How new functionals should be reported. *J. Chem. Phys.* **2023**, *159*, 114116.
- (40) Sitkiewicz, S. P.; Zalesny, R.; Ramos-Cordoba, E.; Luis, J. M.; Matito, E. How reliable are modern density functional approximations to simulate vibrational spectroscopies? *J. Phys. Chem. Lett.* **2022**, *13*, 5963–5968.
- (41) Sitkiewicz, S. P.; Ferradás, R. R.; Ramos-Cordoba, E.; Zalesny, R.; Matito, E.; Luis, J. M. Spurious oscillations caused by density functional approximations: Who is to blame? Exchange or correlation? *J. Chem. Theory Comput.* **2024**, *20*, 3144–3153.
- (42) Richard, R. M.; Lao, K. U.; Herbert, J. M. Approaching the complete-basis limit with a truncated many-body expansion. *J. Chem. Phys.* **2013**, *139*, 224102.
- (43) Ouyang, J. F.; Cvitkovic, M. W.; Bettens, R. P. A. Trouble with the many-body expansion. *J. Chem. Theory Comput.* **2014**, *10*, 3699–3707.
- (44) Murray, C. W.; Handy, N. C.; Laming, G. J. Quadrature schemes for integrals of density functional theory. *Mol. Phys.* **1993**, *78*, 997–1014.
- (45) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. A standard grid for density-functional calculations. *Chem. Phys. Lett.* **1993**, *209*, 506–512.
- (46) Mitani, M. An application of double exponential formula to radial quadrature grid in density functional calculation. *Theor. Chem. Acc.* **2011**, *130*, 645–669.
- (47) Mitani, M.; Yoshioka, Y. Numerical integration of atomic electron density with double exponential formula for density functional calculation. *Theor. Chem. Acc.* **2012**, *131*, 1169.
- (48) Mardirossian, N.; Head-Gordon, M. Characterizing and understanding the remarkably slow basis set convergence of several Minnesota density functionals for intermolecular interaction energies. *J. Chem. Theory Comput.* **2013**, *9*, 4453–4461.
- (49) Sharkas, K.; Wagle, K.; Santra, B.; Akter, S.; Zope, R. R.; Baruah, T.; Jackson, K. A.; Perdew, J. P.; Peralta, J. E. Self-interaction error overbinds water clusters but cancels in structural energy differences. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 11283–11288.
- (50) Wagle, K.; Santra, B.; Bhattarai, P.; Shahi, C.; Pederson, M. R.; Jackson, K. A.; Perdew, J. P. Self-interaction correction in water–ion clusters. *J. Chem. Phys.* **2021**, *154*, 094302.
- (51) Hernandez, D. J.; Rettig, A.; Head-Gordon, M. A new view on density corrected DFT: Can one get a better answer for a good reason? *arXiv* **2023**, DOI: 10.48550/arXiv.2306.15016.
- (52) Herbert, J. M. Dielectric continuum methods for quantum chemistry. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, No. e1519.
- (53) Vuckovic, S.; Song, S.; Kozłowski, J.; Sim, E.; Burke, K. Density functional analysis: The theory of density-corrected DFT. *J. Chem. Theory Comput.* **2019**, *15*, 6636–6646.

- (54) Sim, E.; Song, S.; Vuckovic, S.; Burke, K. Improving results by improving densities: Density-corrected density functional theory. *J. Am. Chem. Soc.* **2022**, *144*, 6625–6639.
- (55) Song, S.; Vuckovic, S.; Sim, E.; Burke, K. Density-corrected DFT explained: Questions and answers. *J. Chem. Theory Comput.* **2022**, *18*, 817–827.
- (56) Rudberg, E.; Rubensson, E. H.; Salek, P. Kohn–Sham density functional theory electronic structure calculations with linearly scaling computational time and memory usage. *J. Chem. Theory Comput.* **2011**, *7*, 340–350.
- (57) Rudberg, E. Difficulties in applying pure Kohn–Sham density functional theory electronic structure methods to protein molecules. *J. Phys.: Condens. Matter* **2012**, *24*, 072202.
- (58) Slattery, S. A.; Surjuse, K. A.; Peterson, C.; Penchoff, D. A.; Valeev, E. Economical quasi-Newton unitary optimization of electronic orbitals. *Phys. Chem. Chem. Phys.* **2024**, *26*, 6557–6573.
- (59) Slattery, S. A.; Yon, J. C.; Valeev, E. F. Revisiting artifacts of Kohn–Sham density functionals for biosimulation. *J. Chem. Theory Comput.* **2024**, *20*, 6652–6660.
- (60) Lever, G.; Cole, J. D.; Hine, N. D. M.; Haynes, P. D.; Payne, M. C. Electrostatic considerations affecting the calculated HOMO–LUMO gap in protein molecules. *J. Phys.: Condens. Matter* **2013**, *25*, 152101.
- (61) Ren, F.; Liu, F. Impacts of polarizable continuum models on the SCF convergence and DFT delocalization error of large molecules. *J. Chem. Phys.* **2022**, *157*, 184106.
- (62) Bowling, P. E.; Broderick, D. R.; Herbert, J. M. Fragment-based calculations of enzymatic thermochemistry require dielectric boundary conditions. *J. Phys. Chem. Lett.* **2023**, *14*, 3826–3834.
- (63) Bowling, P. E.; Broderick, D. R.; Herbert, J. M. Quick-and-easy validation of protein–ligand binding models using fragment-based semiempirical quantum chemistry. *J. Chem. Inf. Model.* **2025**, *65*, 937–949.
- (64) Lange, A. W.; Herbert, J. M. Polarizable continuum reaction-field solvation models affording smooth potential energy surfaces. *J. Phys. Chem. Lett.* **2010**, *1*, 556–561.
- (65) Lange, A. W.; Herbert, J. M. A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: The switching/Gaussian approach. *J. Chem. Phys.* **2010**, *133*, 244111.
- (66) Herbert, J. M.; Lange, A. W. Polarizable continuum models for (bio)molecular electrostatics: Basic theory and recent developments for macromolecules and simulations. In *Many-Body Effects and Electrostatics in Biomolecules*; Cui, Q., Ren, P., Meuwly, M., Eds.; CRC Press: Boca Raton, FL, 2016; pp 363–416.
- (67) Hui, K.; Chai, J.-D. SCAN-based hybrid and double-hybrid density functionals from models without fitted parameters. *J. Chem. Phys.* **2016**, *144*, 044114.
- (68) Kaplan, A. D.; Shahi, C.; Sah, R. K.; Bhetwal, P.; Kanungo, B.; Gavini, V.; Perdew, J. P. How does HF-DFT achieve chemical accuracy for water clusters? *J. Chem. Theory Comput.* **2024**, *20*, 5517–5527.
- (69) Rana, B.; Coons, M. P.; Herbert, J. M. Detection and correction of delocalization errors for electron and hole polarons using density-corrected DFT. *J. Phys. Chem. Lett.* **2022**, *13*, 5275–5284.
- (70) Rana, B.; Beran, G. J. O.; Herbert, J. M. Correcting π -delocalisation errors in conformational energies using density-corrected DFT, with application to crystal polymorphs. *Mol. Phys.* **2023**, *121*, No. e2138789.
- (71) Dasgupta, S.; Lambros, E.; Perdew, J. P.; Paesani, F. Elevating density functional theory to chemical accuracy for water simulations through a density-corrected many-body formalism. *Nat. Commun.* **2021**, *12*, 6359.
- (72) Dasgupta, S.; Shahi, C.; Bhetwal, P.; Perdew, J. P.; Paesani, F. How good is the density-corrected SCAN functional for neutral and ionic aqueous systems, and what is so right about the Hartree–Fock density? *J. Chem. Theory Comput.* **2022**, *18*, 4745–4761.
- (73) Broderick, D. R.; Bowling, P. E.; Shockey, J.; Higley, J.; Dickerson, H.; Ahmed, S.; Herbert, J. M. “Fragme \cap t”, <https://gitlab.com/fragment-qc/fragment>, accessed 2025-02-14.
- (74) Epifanovsky, E.; et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155*, 084801.
- (75) Ohio Supercomputer Center, <https://osc.edu/ark:/19495/f5s1ph73> (accessed 2025-02-27).