

Local second order Møller-Plesset theory with a single threshold using orthogonal virtual orbitals: Theory, implementation and assessment

Zhenling Wang,^{†,‡} Abdulrahman Aldossary,[†] Tianyi Shi,[¶] Yang Liu,[¶] Xiaoye S. Li,[¶] and Martin Head-Gordon^{*,†,‡}

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

[‡]*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

[¶]*Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

E-mail: mhg@cchem.berkeley.edu

Abstract

It has long been clear that electron correlation methods exhibit unphysical compute scalings with molecular size, which has motivated the development of local correlation methods to discard effectively zero contributions in a controlled way to yield an approximate correlation energy. The ideal local correlation method should have a single numerical threshold that controls the dropping of terms, with the ability to have that threshold set small enough that the correlation energy is reproduced to enough significant figures that the result is chemically identical. This work reports such a method for

second order Møller-Plesset (MP2) theory. The theory, implementation, and testing of this local MP2 theory are reported. Thresholds ranging from 10^{-5} to 10^{-8} , and basis sets ranging from split valence plus polarization through to quadruple zeta are assessed for local MP2 calculations on a range of molecules, including linear chains and molecules with 2 and 3-dimensional character. The implementation is shared memory parallel, via OpenMP, and yields roughly 50% parallel efficiency with 16 cores for a large job. Considerable efforts were made to minimize memory demands, which increase as thresholds are tightened. A variety of relative energy calculations are presented as a function of threshold to provide some guidance to users on how to obtain adequate precision at low compute cost. It is particularly clear that derivative properties require tighter thresholds in order to achieve adequate precision.

1 Introduction

Solution of the Schrödinger equation is exponentially hard on a classical computer, notwithstanding recent advances in performing full configuration interaction with selection of only numerically significant amplitudes.¹⁻⁷ As a result, approximations beyond the introduction of a one-particle atomic orbital (AO) basis set^{8,9} continue to be mandatory. The most important concept associated with such approximations is that they should constitute a well-defined (“black-box”) theoretical model chemistry,¹⁰ as first clearly advocated by Pople.^{11,12} Such models require no user input apart from the choice of model, a molecular geometry, net charge and spin-state, and a finite one-particle basis set. They should also yield continuous potential energy surfaces, be reasonably efficient to compute, and be acceptably accurate. There are many ways to violate the ideas of a good model chemistry. For instance, specifying a geometry-specific active space, such as only the 2 occupied π and 2 valence π^* orbitals in planar C_4H_6 is not a model chemistry. Of more relevance to our present topic, neither is correlating only electrons in occupied orbitals that are less than some small distance (e.g. 4

Å) apart from each other.

Whilst Kohn-Sham density functional theory (DFT)¹³ is the most widely used class of model chemistries,¹⁴ great interest remains in the development of wavefunction theory (WFT) model chemistries because they are systematically improvable. Wavefunction models begin with the Hartree-Fock (HF) mean field reference determinant, $|\Phi_0\rangle$, and its energy, E_0 . The electron correlation effect of other determinants is then commonly modeled by low-order Møller-Plesset (MP) perturbation theory (PT),¹⁵ such as MP2 theory, or by infinite order coupled cluster (CC) theory,^{16,17} truncated at a relatively low level of excitation from $|\Phi_0\rangle$, such as CCSD¹⁸ or CCSD(T).¹⁹ Using KS orbitals, PT2 is also a component of double hybrid density functional theory (DFT).²⁰⁻²² The principal reason for the predominance of DFT over WFT is the fact that the compute costs of DFT (or HF) rise asymptotically as $\mathcal{O}(A^3)$ with number of atoms, A , for a fixed choice of AO basis, while exact implementation of MP2 scales $\mathcal{O}(A^5)$, CCSD scales $\mathcal{O}(A^6)$, and CCSD(T) scales $\mathcal{O}(A^7)$. As we briefly review next, such scalings are unphysical and can be circumvented both in principle and in practice.

Partial localization of electrons is a distinctive feature of molecules with significant HOMO-LUMO gaps, ΔE_{HL} . The one-particle density matrix (1PDM) in real-space, $\mathbf{P}(\mathbf{r}, \mathbf{r}')$ exhibits exponential decay with $|\mathbf{r} - \mathbf{r}'|$, with decay rate $\Gamma \propto \sqrt{\Delta E_{\text{HL}}}$,²³ which in turn means that occupied (and virtual) orbitals can be localized using a range of localization criteria.²⁴ A set of localized virtual orbitals²⁵ can then be obtained as projected atomic orbitals (PAOs) from each AO, μ : $\omega(\mathbf{r})_{a=\mu}^{\text{virt}} = (\mathbf{1} - \mathbf{P})\omega(\mathbf{r})_{\mu}$. For molecules much larger than the size of AOs, the PAOs will be exponentially localized,²⁶ although they are not only non-orthogonal, but also linearly dependent (i.e., for N AOs, with n occupied, the rank of the PAO space is $N - n$, although there are N PAOs). Nevertheless, since correlation amplitudes depend on matrix elements over occupied (i, j) and virtual (a, b) orbitals, such as $(ia|jb)$, it is clear that there are only a linear number of significant ia products. Since each such product has zero charge due to strong orthogonality, the magnitude of $(ia|jb)$ behaves at long range like

a dipole-dipole interaction:²⁷⁻²⁹ i.e., $(ia|jb) \propto |\mathbf{r} - \mathbf{r}'|^{-3}$. So with an appropriate set of localized occupied and virtual orbitals, the number of numerically significant $(ia|jb)$ exhibits only linear growth, $\mathcal{O}(A^1)$, for a large enough molecule. This is the basis of local correlation methods, which can likewise achieve $\mathcal{O}(A^1)$ scaling for sufficiently large systems.

The genesis of modern local correlation methods was the seminal research of Pulay.^{25,30} These early local correlation models took advantage of localized occupied orbitals and PAO virtuals as a basis to define “domains” of orbitals,³¹ using a type of overlap cutoff of roughly 0.2, a set of virtuals is assigned as being in the domain of each occupied orbital. The retained double substitutions from a given pair of occupied orbitals are only to virtuals that lie in the union of their two domains. In terms of model chemistry criteria, this approach only becomes a good model in the limit where all numerically significant correlation amplitudes are retained. Otherwise, the potential energy surface could not be smooth. In the early days of local correlation, compute limits made this impossible, and therefore other models were designed to guarantee smooth surfaces.^{32,33} However, having model chemistries for local MP2 that are similar but not identical to MP2 (or any other target correlation theory) is also undesirable. Therefore the most successful local correlation directions have been to pursue methods capable of near-quantitative reproduction of the exact MP2 result, in pursuit of linear scaling.^{34,35} Use of resolution-of-the-identity or density fitting approaches³⁶ in local MP2 was an important step forwards.³⁷ A particularly important development was the use of pair natural orbitals (PNOs)^{38,39} to define a compressed representation of correlation amplitudes for a given pair of occupied orbitals.⁴⁰ At the MP2 level, perhaps the most advanced implementations of PNO local MP2 methods are the PNO-LMP2 method,⁴¹ and the DLPNO-MP2 method.⁴² These approaches achieve full linear scaling and, as reported, deviations from conventional MP2 in energy differences that can be 1 kJ/mol or less.

Other highly effective local MP2 methods have also been reported. As an alternative to iterative solution of linear equations for the first order correlation amplitudes, the Laplace

transform approach replaces the solver with a fixed number of quadrature points.^{43,44} With use of localized orbitals, this becomes a basis for competitive fast local MP2 methods.^{35,45–48} Another family of fast methods is the divide and conquer paradigm^{49–52} as well as the divide, expand, consolidate approach.^{53,54} Other fragmentation ideas have also led to high efficient fast local MP2 methods.^{55–60} Tensor hypercontraction ideas have also been applied to MP2 to lower the formal scaling^{61–64} Combinations of these ideas have also proven effective.^{65–67} Implementation has been extended to include GPUs,⁶⁸ as well as massively parallel computing.^{52,69} Novel ideas continue to be put forwards, including an algorithm for the opposite spin MP2 energy^{70,71} that avoids two-electron integrals,⁷² and another that employs Slater orbitals.⁷³ This diversity of approaches reflects the fact that the problem of developing an optimal fast MP2 framework cannot be viewed as fully solved despite all the progress we have briefly reviewed.

The goal of the research we report here is to develop a new approach to the goal of approaching the exact MP2 limit, with the particular target of specifying a single numerical threshold to control the precision. The thresholds of most interest are those which can ensure that the results are identical to conventional MP2 such that one need not label the results as being from a local correlation calculation. We introduce the theory and algorithms necessary to implement this strategy, which follows the prototyping that we have recently reported.⁷⁴ The first design decision is a representation of the virtual space. Rather than the PAO/PNO strategy reviewed above, we take the potentially simpler approach of using a set of non-redundant orthogonal virtual orbitals, partitioned into valence virtuals (VV) and hard virtuals (HV),⁷⁵ which are preconditioned by an atomic pseudo-canonicalization of the HV space. Resolution-of-the-identity (RI) together with local fit domains are employed to accelerate two-electron integral construction whilst retaining asymptotic linear scaling. Together with use of the “ragged list” sparsity⁷⁴ that is defined by the full list of virtuals $\{a, b\}$ significant to a given pair of occupied orbitals $\{i, j\}$, this yields asymptotically linear scaling

memory requirements, and also asymptotic linear scaling of solver compute effort. All steps are OpenMP parallel. These developments are described in the Theory and Implementation Sections.

With a working implementation in hand, we then turn to the question of characterizing the performance of the local MP2 algorithm in terms of memory and compute requirements as a function of the three key variables. First is the nature of the molecule: close-packed molecules with higher than one-dimensional connectivity pose the greatest challenge for local correlation methods in the sense of yielding the least advantage versus a non-local exact solution.^{76,77} Second is the choice of AO basis set size: larger basis sets likewise pose greater challenges for local correlation,⁷⁸ yet are crucial for enabling useful applications of the code. Third is the target threshold for accuracy. With a sufficiently tight threshold, the local correlation results are indistinguishable from conventional MP2. Therefore we seek the loosest thresholds that can yield such results for energy differences. On the one hand, looser thresholds will yield (much) faster time-to-solution, but will only be reproducible running *exactly* the same algorithm. On the other hand, significantly tighter thresholds will simply eliminate the possibility of savings in memory requirements and compute effort versus exact MP2. We think a suitable target is $10^{-1} - 10^{-3}$ kJ/mol. A detailed set of calculations exploring these factors is reported in the Results and Discussions Sections.

2 Theory

In this section, we will summarize the framework of our local MP2 theory, and the implementation details will be introduced in the next section. This paper only considers spin-restricted closed-shell systems with frozen core orbitals for simplicity; for the energy the frozen orbitals are simply excluded. Table 1 lists all the index notations that we are using. All equations are given in terms of spatial molecular orbitals.

Table 1: Notations for different orbitals and basis functions.

orbitals / basis functions	indices	total number
occupied orbitals	i, j, k, \dots	n_{occ}
virtual orbitals	a, b, c, \dots	n_{virt}
basis functions	μ, ν, \dots	n_{basis}
auxiliary basis functions	P, Q, \dots	n_{aux}

The MP2 amplitude equation is

$$\sum_{i'j'a'b'} \Delta_{aba'b'}^{ijj'} t_{ab}^{ij} = 2J_{ab}^{ij} - K_{ab}^{ij}, \quad (1)$$

where $J_{ab}^{ij} = K_{ba}^{ij} = (ia|jb)$ are two-electron integrals and Δ is (if using orthogonal orbitals)

$$\Delta_{aba'b'}^{ijj'} = \delta_{ii'} \delta_{aa'} \delta_{jj'} F_{bb'} - \delta_{ii'} \delta_{aa'} F_{jj'} \delta_{bb'} + \delta_{ii'} F_{aa'} \delta_{jj'} \delta_{bb'} - F_{ii'} \delta_{aa'} \delta_{jj'} \delta_{bb'}; \quad (2)$$

where F is the Fock matrix. In a gapped system, Δ is negative definite, and its smallest eigenvalue (in absolute value) is $2\Delta E_{\text{HL}}$, so a unique solution is guaranteed. The MP2 correlation energy is then obtained as the dot product of the amplitude tensor with J :

$$E_{\text{corr}} = - \sum_{iajb} J_{ab}^{ij} t_{ab}^{ij}. \quad (3)$$

Our local MP2 theory has three components which will all be discussed in detail in the following subsections:

1. A choice of orbital localization technique – we employ the common choice of Boys occupied and the not-so-common choice VV-HV virtual orbitals as described below in Sec. 2.1;

2. A way to efficiently solve the linear system, Eq. 1 – we use a sparse “ragged list” representation of the right-hand side, and a “fixed sparsity pattern” approach to the solver, discussed in Sec. 2.2;
3. A way to build the significant elements of the J tensor efficiently – we use robust local density fitting via algorithms given in Sec. 2.3

2.1 One Particle Representations - Local Occupied and Virtual Orbitals

There are multiple ways to localize orbitals,^{24,79} for example, occupied orbitals can be localized by the well-known Boys, Pipek-Mezey, or Edmiston-Ruedenberg methods.^{80–82} Widely used sets of localized virtual orbitals include projected atomic orbitals (PAOs), orbital specific virtuals (OSVs),⁸³ pair natural orbitals (PNOs),^{38,39} and many others. Our previous work⁷⁴ compared a handful of such possibilities and concluded that of those, the optimal choice for local MP2 is to pick Boys occupied orbitals and valence virtual – hard virtual (VV-HV) orbitals.⁷⁵ This is a single set of localized orthogonal orbitals for the whole chemical system, which is simple both in concept and in implementation. Also, relative to use of PNOs or OSVs, the memory requirement to hold the J tensor is minimized.

The Boys occupied orbitals are a set of orthogonal orbitals that minimizes the following spread functional

$$f_B = \sum_i \langle \phi_i | (\mathbf{r} - \bar{\mathbf{r}}_i)^2 | \phi_i \rangle, \quad (4)$$

where $\bar{\mathbf{r}}_i = \langle \phi_i | \mathbf{r} | \phi_i \rangle$. In other words, Boys orbitals minimize the sum of orbital second moments. The optimization can be carried out either by Jacobi sweeps or other techniques.^{84–86} While the Boys orbitals typically localize well (subject to the magnitude of ΔE_{HL}) it should be noted that they are not necessarily unique for a given problem.

To obtain VV-HV orbitals,⁷⁵ we need to split our basis set space (\mathcal{B}) into the direct sum

of an adaptive minimal basis space (\mathcal{E}) and the remaining hard virtual space (\mathcal{H})

$$\mathcal{B} = \mathcal{E} \oplus \mathcal{H}. \tag{5}$$

Following ref. 75, our adaptive minimal basis is composed from the union of the span of occupied orbitals, and the span of the virtuals defined by projecting STO-3G into the working basis, and removing the occupied space. The hard virtual space contains many high angular momentum basis functions which are essentially atom-centered although they are not orthogonal (similar to PAOs). We then split \mathcal{E} into the occupied space (\mathcal{O}) and the valence virtual space (\mathcal{L})

$$\mathcal{B} = \mathcal{O} \oplus \mathcal{L} \oplus \mathcal{H}. \tag{6}$$

Here we assume that the minimal basis is able to describe the occupied space, which is usually true. The \mathcal{L} space is small, and can be easily localized by the Boys procedure; the resulting valence virtuals look like antibonding orbitals.⁸⁷ Our virtual space (\mathcal{V}) is therefore

$$\mathcal{V} = \mathcal{L} \oplus \mathcal{H}, \tag{7}$$

So we collect all localized valence virtuals and hard virtuals, and then perform a locality-preserving orthogonalization of the HVs. Our implementation basically follows Ref. 75, but with two differences. The first one is that we do a simple weighted symmetrical orthogonalization of the hard virtuals based on the inverse of their spatial spread instead of the “important class” technique in the original paper. The second difference is we further pseudo-canonicalize the hard virtuals atom by atom. As we will see later in Sec. 4.4, this pseudo-canonicalization is very beneficial when we want to use larger basis sets (Sec. 4.4). For calculations using extended AO basis sets, particularly those containing diffuse functions, and for molecules with close-packed 3-d structures, ill-conditioning in the AO overlap

matrix associated with near-linear dependence will occur.^{77,78} To manage this challenge, we employ the same linear-dependence threshold ($\sim 10^{-7} - 10^{-8}$ for double precision computer arithmetic) that is used to drop eigenvectors whose eigenvalues are below the threshold via canonical orthogonalization.⁸⁸ However, to accommodate the VV-HV scheme, we drop individual atomic orbitals one by one based on selecting the AO with largest amplitude in the eigenvector belonging to the smallest eigenvalue, and repeating until the smallest eigenvalue of the overlap matrix built with the remaining basis functions is above the threshold. To ensure the MP2 correction is well-defined, the same orthogonalization procedure is also used for the preceding SCF calculation.

2.2 Ragged List Representation and Fixed Sparsity Pattern Approximation

A local representation of the J tensor is central to local MP2 theory. As already pointed out in the introduction, there are multiple ways to represent the J tensor in an $\mathcal{O}(A)$ way, including the PAO/PNO approach and the PAO/OSV approach. In our prototyping paper,⁷⁴ we have shown that the “ragged list” representation actually is the most memory efficient one. The ragged list representation exploits all the numerical sparsity of J tensor, meaning that we store every $(ia|jb)$ element in J if

$$|J_{ab}^{ij}| > \epsilon \quad \text{or} \quad |K_{ab}^{ij}| > \epsilon, \quad (8)$$

where ϵ is the threshold. In practice, this is done by storing all significant a, b index pairs and their J_{ab}^{ij} values for every i, j index pair.

To solve the linear equations, Eq. 1, of MP2, we introduce an approximation called “fixed sparsity pattern”, which means that we require the solution t tensor to have exactly the same sparsity pattern as the J tensor. In other words, we remove the rows and columns

in Δ which corresponds to every (numerically) zero element in $2J - K$ (Figure 1). The error from the fixed sparsity pattern has been proven highly controllable in a wide range of molecules.⁷⁴ Because we have only $\mathcal{O}(A)$ elements on the right-hand side and the truncated Δ matrix still has some sparsity, it is possible to solve the MP2 equation with linear scaling computational effort. Conjugate gradient (CG, with a diagonal preconditioner) can be used since the truncated Δ remains negative definite.

$$\begin{array}{ccc}
 \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \begin{pmatrix} \bullet \\ \bullet \\ \bullet \\ \bullet \end{pmatrix} = \begin{pmatrix} * \\ 0 \\ * \\ 0 \end{pmatrix} & & \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \begin{pmatrix} \bullet \\ 0 \\ \bullet \\ 0 \end{pmatrix} = \begin{pmatrix} * \\ 0 \\ * \\ 0 \end{pmatrix} \\
 \text{Full System} & & \text{Fixed Sparsity Pattern}
 \end{array}$$

Figure 1: The “fixed sparsity pattern” approximation. Stars stand for known elements and dots for unknowns.

After the solution t is converged, we define t' such that

$$t_{ab}^{ij} = 2t_{ab}^{\prime ij} - t_{ba}^{\prime ij}, \quad (9)$$

The correlation energy is then calculated as

$$E_{\text{corr}} = -t' \cdot (2J - K) - t' \cdot r, \quad (10)$$

where r is the residual vector at the end of the conjugate gradient algorithm; i.e.,

$$\Delta t = (2J - K) - r. \quad (11)$$

This expression corresponds to the Hylleraas functional, so the energy error is quadratic in terms of r , meaning that a loose convergence threshold can be used to solve the MP2 equation. The seemingly redundant symmetric truncation in Eq. 8 is needed because other-

wise, the Hylleraas functional provides a different correlation energy under the fixed sparsity pattern.

2.3 Building the J Tensor

The remaining question is how to build the J tensor efficiently. For clarity, we adopt the language of “sparse maps” introduced in Ref. 42. A sparse map

$$A \rightarrow B \tag{12}$$

means each element of A has a list of corresponding B . For example, an $i \rightarrow a$ sparse map stores different lists of virtual orbitals linked to each occupied orbital. Sparse maps can be inverted: a $B \rightarrow A$ sparse map can be constructed from an $A \rightarrow B$ sparse map. Multiple sparse maps can also be concatenated, meaning that putting two sparse maps $A \rightarrow B$ and $B \rightarrow C$ together we can define $A \rightarrow C = A \rightarrow B \rightarrow C$.

Looking at the definition of $(ia|jb)$,

$$(ia|jb) = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{i(\mathbf{r}_1)a(\mathbf{r}_1)j(\mathbf{r}_2)b(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}, \tag{13}$$

we know that $(ia|jb)$ will be small if i is far away from j , or i is far away from a , or b is far away from j . So, two sparse maps, $i \rightarrow j$ and $i \rightarrow a$, are necessary. Notice that i, a are linear combinations of basis functions μ , so we need $i \rightarrow \mu$ and $a \rightarrow \mu$. Finally, *local* density fitting must be used in order to generate all required J elements in $\mathcal{O}(A)$ time, so $i \rightarrow P$ is also needed. In total, we need five basic sparse maps, and there could be multiple ways to construct each of them. We will describe our choice later in the implementation section.

The robust local density fitting uses the local fit coefficients

$$C_{ia}^P = \sum_{i \rightarrow Q} V_{PQ} (ia|Q); \quad (14)$$

here V is the inverse of $(P|Q)$ submatrix where $i \rightarrow P$, $i \rightarrow Q$. The expression for the J tensor elements is

$$(ia|jb) = - \sum_{i \rightarrow P, j \rightarrow Q} C_{ia}^P (P|Q) C_{jb}^Q + \sum_{i \rightarrow P} C_{ia}^P (jb|P) + \sum_{j \rightarrow Q} C_{jb}^Q (ia|Q), \quad (15)$$

so we need to compute and store:

1. all $(ia|P)$ if i and P can be connected through j : $i \rightarrow \rightarrow P = i \rightarrow j \rightarrow P$;
2. all C_{ia}^P but only when i and P are directed connected: $i \rightarrow P$.

Sparse maps with " $\rightarrow \rightarrow$ " are called extended sparse maps. Notice that in this way, we will use the same list of P in both $(ia|P)$ and C_{ia}^P for all virtual orbitals connected to i . Then, for each i, j pair, we can construct elements in the J^{ij} block, filter, and store the significant indices and integral values in the ragged list representation. This formalism is robust in the sense that the error in $(ia|jb)$ is quadratic in terms of the incompleteness of the fitting domain $i \rightarrow P$.

The entire integral assembly part can be done linear-scalingly. Asymptotically, only $\mathcal{O}(A)$ number of J^{ij} blocks need to be computed (by $i \rightarrow j$), and in each block, we only compute a constant number of $(ia|jb)$ integrals (by $i \rightarrow a$). Every integral takes asymptotically constant time to evaluate because the length of $i \rightarrow P$ is asymptotically constant. For the same reason, obtaining C_{ia}^P should take $\mathcal{O}(A)$ time. Computing all required $(ia|P)$ also scales linearly because the number of the required $(\mu\nu|P)$ is constant for every $(ia|P)$ integral (by $i \rightarrow \mu$ and $a \rightarrow \mu$).

We have a couple of different thresholds in our local MP2 formalism: the main threshold ϵ for J , the CG threshold, and the thresholds in the five basic sparse maps. In the next section, we will link all other thresholds to ϵ , such that the accuracy of our algorithm is easily controllable by a single parameter ϵ supplied by the user.

3 Implementation

In this section, we present the implementation details on how we perform the conjugate gradient iterations and how we build J by local density fitting.

3.1 Conjugate Gradient Iterations

We use the standard conjugate gradient algorithm with a diagonal preconditioner (i.e., Jacobi preconditioner) to solve the linear system. In the code, we store J in the following way:

1. (main part) for each i, j pair, we store the integral $(ia|jb)$ and the $an_{virt} + b$ value of each J_{ab}^{ij} element kept;
2. (transposed part) for each a, b pair, we store the $in_{occ} + j$ value and the position of $(ia|jb)$ in the vectorized J tensor of each J_{ab}^{ij} element kept.

The tensor is vectorized by putting elements with the same i, j together. Since the J tensor is symmetric, we only need to store and manipulate the upper half of it, saving us about 50% memory requirements and computation time.

In our framework, the slowest part of the CG iteration is the matrix-vector multiplication algorithm – the way to calculate Δt . The pseudo-code is shown in Algorithm 1. It can be easily OpenMP parallelized by adding relevant compiler directives (`#pragma omp parallel for schedule(dynamic)`) to the loops in the 2nd and 9th lines. The actual calculation of Δt is

in lines 6 and 13, and the number of FLOPs is

$$2 \sum_{i,j} \left[\sum_{a_0} |v_a|^2 + \sum_{b_0} |v_b|^2 \right] + 2 \sum_{a,b} \left[\sum_{i_0} |v_i|^2 + \sum_{j_0} |v_j|^2 \right], \quad (16)$$

where $|\cdot|$ means the length of a list. The first sum comes from the for loop in line 2, and the second sum comes from the for loop in line 9. In the first sum, as the number of significant j for each i and the cardinal number of a, b list for each significant i, j pair both become constant when the system size is large, the FLOP count is linear-scaling. A similar argument holds for the second sum, so overall this algorithm scales linearly with respect to system size.

Since preconditioned conjugate gradient needs 5 copies of the vector, together with 3 index tensors and 1 integral tensor, the memory requirement is 9 copies of the J tensor. With the help of the Hylleraas functional, we can converge our CG to a relatively loose threshold, namely $\|r\| < 2.5\sqrt{\epsilon}$, where r is the residual vector and ϵ is the threshold for J .

3.2 Local Density Fitting

As already introduced in Sec. 2.3, we need five basic sparse maps: $i \rightarrow j$, $i \rightarrow a$, $i \rightarrow P$, $i \rightarrow \mu$, and $a \rightarrow \nu$. There are multiple choices for each of these sparse maps, and we will briefly explain our choice below. A detailed example analysis of choosing all the thresholds can be found in the SI. Notice that our main threshold for the J tensor is ϵ .

1. $i \rightarrow j$: determined by pair uFERF interaction energies with threshold $10^{-2}\epsilon$.

This is used to determine whether or not i, j are far enough away from each other such that their pair correlation energy can be neglected. This could be estimated by the dipole-dipole interaction in the multipole expansion. We choose to use dipole uFERF orbitals to estimate the pair correlation energy, because the compressed form of dispersion interactions look like uFERF orbitals.^{29,89} Because the uFERF estimation

Algorithm 1 A linear-scaling algorithm to compute Δt

```
1: initialize the output vector  $(\Delta t)_{out} = 0$ 
2: // first and third terms in Eq. 2
3: for each  $i, j$  pair whose  $an_{virt} + b$  list is not empty do
4:   break the  $an_{virt} + b$  list into  $a$  list and  $b$  list
5:   // for example, if  $n_{virt} = 3$ ,  $an_{virt} + b$  list  $\{0,2,3,4\}$  will become a list  $\{0,0,1,1\}$  and
    $b$  list  $\{0,2,0,1\}$ 
6:   for each unique element  $a_0$  in  $a$  list do
7:     // the first term
8:     find the positions in the vectorized tensor of all elements equal to  $a_0$  in  $a$  list and
     store them in a vector  $v_a$ 
9:     find the corresponding  $b$  of those positions and store them in  $v_b$ 
10:     $(\Delta t)_{out}(v_a) += F_{virt}(v_b, v_b)t(v_a)$ 
11:   end for
12:   do lines 6-11 with  $a, b$  swapped // the third term
13: end for
14: // second and fourth terms in Eq. 2
15: for each  $a, b$  pair whose  $in_{occ} + j$  list is not empty do
16:   break the  $in_{occ} + j$  list into  $i$  list and  $j$  list
17:   for each unique element  $i_0$  in  $i$  list do
18:     // the second term
19:     find the positions in the vectorized tensor of all elements equal to  $i_0$  in  $i$  list and
     store them in a vector  $v_i$ 
20:     find the corresponding  $j$  of those positions and store them in  $v_j$ 
21:      $(\Delta t)_{out}(v_i) -= F_{occ}(v_j, v_j)t(v_i)$ 
22:   end for
23:   do lines 17-22 with  $i, j$  swapped // the fourth term
24: end for
```

undershoots the real pair correlation energy (Figure S3 in the SI) and the error is also accumulative, we choose a tight threshold of $10^{-2}\epsilon$ here.

2. $i \rightarrow a$: determined by the 1s-Slater-orbital-approximation of the integral $\int i^2 a^2 d\mathbf{r}$ with threshold 10^{-3} .

This map determines significant virtual orbitals for a certain occupied orbital. The value of the differential overlap integral $\int i^2 a^2 d\mathbf{r}$ has been shown to be of the same magnitude as the Schwarz integral $(ia|ia)$.⁴² The latter can be used to control $(ia|jb)$ effectively:

$$(ia|jb) \leq \sqrt{(ia|ia)(jb|jb)}. \quad (17)$$

A numerical integration scheme is required to accurately calculate $\int i^2 a^2 d\mathbf{r}$, but here we apply a very crude approximation to estimate the differential overlap integral: based on the orbital variance $\langle \phi_i | (\mathbf{r} - \bar{\mathbf{r}}_i)^2 | \phi_i \rangle$, we reduce i, a to 1s Slater orbitals $e^{-\lambda|r-\bar{r}|}$, and then calculate $\int i^2 a^2 d\mathbf{r}$. Unfortunately, this estimator is not effective when i, a are far away, so we choose a rather loose fixed threshold 10^{-3} , and use the extended sparse map $i \rightarrow j \rightarrow a = i \rightarrow \rightarrow a$ instead of $i \rightarrow a$ to determine the significant virtual orbitals for i .

3. $i \rightarrow P$: determined by the 1s-Slater-orbital-approximation of integral $\int i^2 P^2 d\mathbf{r}$ with threshold $10^2 \epsilon^2$.

This map determines the local density fitting domain of every occupied orbital. Similar to the previous map, we use the crude approximator to estimate the differential overlap integral. A tight threshold is needed for an accurate result, and now we pick $10^2 \epsilon^2$ such that the error in E_{corr} caused by local density fitting and the fixed sparsity pattern is similar. Other possibilities could be Mulliken population, center-to-center distance, and others.

4. $i \rightarrow \mu, a \rightarrow \nu$: determined by the coefficient matrix with threshold 10^{-3} .

These two sparse maps are used when we compute $(\mu\nu|P)$ and then transform to $(ia|P)$. A fixed value can be chosen since this does not heavily affect the efficiency and the quality of the result (Figure S6 in the SI). Notice that a truncation threshold of 10^{-3} means that in the localization step, we do not need to converge the Boys and the valence part of the VV-HV orbitals very tightly. Actually, as most quantum chemistry packages process integrals in terms of shells instead of single basis functions, we use slightly different $i \rightarrow P, i \rightarrow \mu, a \rightarrow \nu$ sparse maps, which use lists of shells instead of lists of single basis functions: $i \rightarrow P_{sh}, i \rightarrow \mu_{sh}, a \rightarrow \nu_{sh}$.

Our local density fitting algorithm can be decomposed into three steps: building $(ia|P)$, computing C_{ia}^P , and assembling $(ia|jb)$. To maximally utilize BLAS-3 subroutines, we use matrices to store $(ia|P)$ and C_{ia}^P – for each i , the row dimension of the matrix is the length of $i \rightarrow \rightarrow P$ list (for $(ia|P)$) or $i \rightarrow P$ list (for C_{ia}^P) and column dimension is the length of the $i \rightarrow \rightarrow a$ list. Also, all P_{sh} of the same atom are batched together when computing $(ia|P)$.

Eq. 15 in Sec. 3.2 suggests that we need to compute $(\mu\nu|P)$ if

$$P \rightarrow P_{sh} \rightarrow \rightarrow i \rightarrow \mu_{sh} \rightarrow \mu \quad \text{and} \quad P \rightarrow P_{sh} \rightarrow \rightarrow i \rightarrow \rightarrow a \rightarrow \nu_{sh} \rightarrow \nu. \quad (18)$$

Note the "double extension" in the second list, which is necessary to get the correct answer. Alg. 2 shows the pseudo-code for building the J tensor, where $C_{occ}(A \rightarrow P_{sh} \rightarrow \rightarrow i, A \rightarrow P_{sh} \rightarrow \rightarrow i \rightarrow \mu_{sh} \rightarrow \mu)$ means we pick out the submatrix of C_{occ} with rows in the $A \rightarrow P_{sh} \rightarrow \rightarrow i$ list and columns in the $A \rightarrow P_{sh} \rightarrow \rightarrow i \rightarrow \mu_{sh} \rightarrow \mu$ list. Alg. 2 can be OpenMP parallelized by adding OpenMP directives to lines 1, 8, and 13.

Algorithm 2 Building J tensor using local density fitting

- 1: **for** every atom A **do**
 - 2: Compute integrals $(\mu\nu|P)$ where μ_{sh} that are connected by $A \rightarrow P_{sh} \rightarrow i \rightarrow \mu_{sh}$ and ν_{sh} that are connected by $A \rightarrow P_{sh} \rightarrow i \rightarrow a \rightarrow \nu_{sh}$
 - 3: Pick submatrix of the coefficient matrix: $C_{occ}(A \rightarrow P_{sh} \rightarrow i, A \rightarrow P_{sh} \rightarrow i \rightarrow \mu_{sh} \rightarrow \mu)$ and $C_{virt}(A \rightarrow P_{sh} \rightarrow i \rightarrow a, A \rightarrow P_{sh} \rightarrow i \rightarrow a \rightarrow \nu_{sh} \rightarrow \nu)$
 - 4: **for** every P in $A \rightarrow P_{sh} \rightarrow P$ **do**
 - 5: Transform $(\mu\nu|P)$ into $(ia|P)$ by two dense matrix multiplications
 - 6: select (we have made more than we needed) and store the integrals
 - 7: **end for**
 - 8: **end for**
 - 9: *// computing C_{ia}^P*
 - 10: **for** every i **do**
 - 11: Compute the inverse of $(P|Q)$ submatrix V where $i \rightarrow P, i \rightarrow Q$
 - 12: Find the position of all P of $i \rightarrow P$ in $i \rightarrow P$
 - 13: $C_{ia}^P = \sum_{i \rightarrow Q} V_{PQ} (ia|Q)$
 - 14: **end for**
 - 15: *// assemble to get the J tensor*
 - 16: **for** all i, j pair **do**
 - 17: Find the positions of all P of $i \rightarrow P$ in the Q list of $j \rightarrow Q$ and the positions of all Q of $j \rightarrow Q$ in the P list of $i \rightarrow P$
 - 18: $(ia|jb) = - \sum_{i \rightarrow P, j \rightarrow Q} C_{ia}^P (P|Q) C_{jb}^Q + \sum_{i \rightarrow P} C_{ia}^P (jb|P) + \sum_{j \rightarrow Q} C_{jb}^Q (ia|Q)$
 - 19: Filter elements and store the numerically significant ones in the "ragged list" data structure introduced in Sec. 2.1
 - 20: **end for**
-

3.3 Summary of the Algorithm

A brief summary of our local MP2 scheme is in Algorithm 3. Although we have seven different thresholds in our algorithm (the main threshold ϵ , the CG threshold, and five thresholds for five basic sparse maps), we have either fixed or linked all other six thresholds to the main threshold ϵ , such that the accuracy of our algorithm can be easily controlled by a single input parameter.

Algorithm 3 Local MP2

- 1: Localize molecular orbitals (Boys occupied orbitals and VV-HV virtual orbitals)
 - 2: Construct five basic sparse maps $i \rightarrow j, i \rightarrow a, i \rightarrow P, i \rightarrow \mu, a \rightarrow \nu$ and all other required sparse maps
 - 3: Build the J tensor with local density fitting (Algorithm 2)
 - 4: Use CG to solve the local MP2 equation (with mat-vec function in Algorithm 1)
 - 5: Evaluate the Hylleraas functional Eq. 10 to obtain E_{corr}
-

4 Results and Discussion

In this section, we will characterize our local MP2 scheme in a few different ways. We first discuss how the threshold ϵ affects accuracy (the error in correlation energy and the error of energy difference between molecules), timing, and memory requirement. We then briefly show our OpenMP parallel efficiency, the performance of this algorithm with different basis sets, and the smoothness of the potential energy surface. All jobs are done on a local workstation with two 8-core Intel Xeon Silver 4110 CPUs and 64 GB memory. A development version of Q-Chem 6 is used,⁹⁰ and all calculations are done with the def2-TZVP basis set, frozen core, and 16 threads OpenMP, unless otherwise specified. The geometries of all molecules can be found in the SI.

4.1 Threshold and Accuracy

In our prototyping paper,⁷⁴ we showed that the error caused by the fixed sparsity pattern is linear with respect to ϵ , both in linear systems and 3-D molecules. This is still true when we also include errors associated with local density fitting. Figure 2 plots the absolute correlation energy error of linear alkane chains.

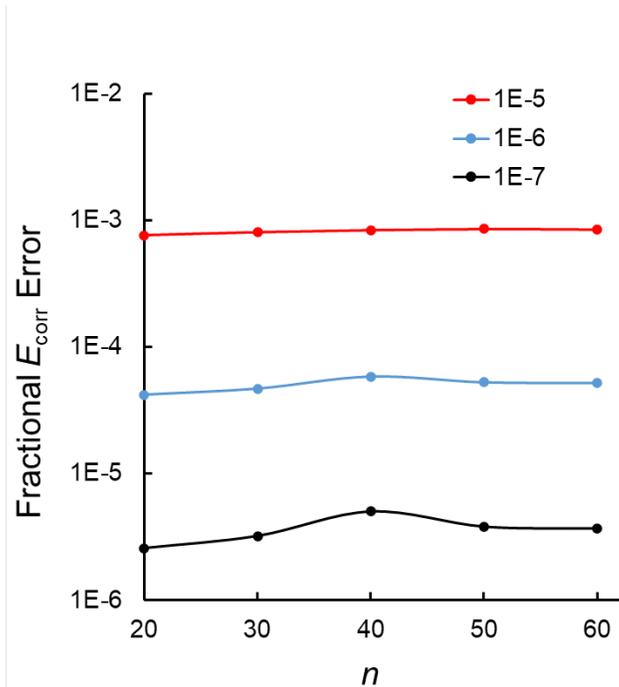


Figure 2: The fractional error in the def2-TZVP basis correlation energy of the local MP2 scheme compared with canonical MP2 for linear alkane chains $\text{C}_n\text{H}_{2n+2}$. Three different thresholds ϵ are used for the local correlation calculations.

It can be seen that the error of our local MP2 scheme is highly controllable by the threshold ϵ : $\epsilon = 10^{-5}$ can recover 99.9% of canonical MP2 correlation energy, and $\epsilon = 10^{-7}$ can recover more than 99.999%. To understand the performance of our algorithm for relative energies, i.e., chemical reactions, we computed the conformational energy of 21 conformers of $\text{C}_{20}\text{H}_{42}$ relative to the all-trans conformer (0 in Figure 3), using the ACONF20 molecule set by Ehlert *et al.*⁹¹ Figure 3 briefly summarizes the molecule set, and Table 2 presents the statistics of the error in conformational energy of our method under a different threshold,

compared with the canonical MP2 numbers.

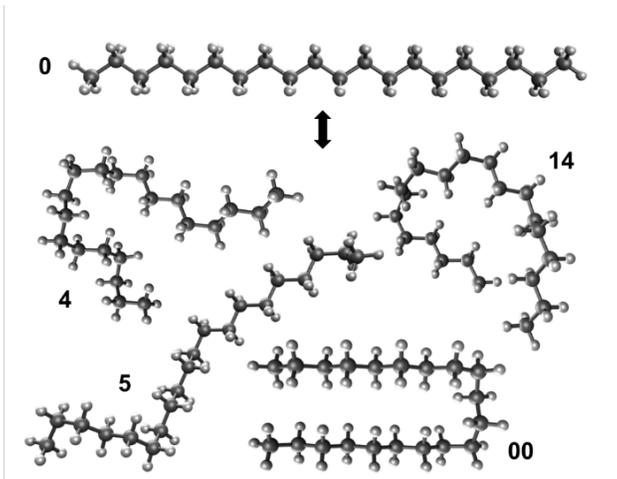


Figure 3: Some representative molecules from the 21 $C_{20}H_{42}$ conformers contained in the ACONF20 data set. The 00 conformer has the lowest energy. We calculate the conformational energy relative to the all-trans conformer 0.

Table 2: Error statistics (in kJ/mol) for the relative conformational energies of the ACONF20 molecule set between local and canonical MP2 under different thresholds, ϵ . Statistics include maximum deviation (MAX), mean absolute (signed) error (MAE), and root mean square deviation (RMSD).

Threshold	10^{-5}	$10^{-5.5}$	10^{-6}	$10^{-6.5}$	10^{-7}
MAX	5.94	1.88	0.49	0.18	0.05
MAE	2.66	0.83	0.23	0.06	0.02
RMSD	3.09	0.97	0.26	0.08	0.02

In Table 2, it can be seen that the goal of 0.1 kJ/mol relative energy goal is safely reached with a 10^{-7} threshold. Even $10^{-6.5}$ is sufficient most of the time. The accuracy is, again, generally linear with respect to ϵ . The conformational energy differences in the ACONF20 set range between roughly 5 and 20 kJ/mol. Therefore a 3 kJ/mol error, which is the RMSD with $\epsilon = 10^{-5}$ is very large. We see that in this case, local MP2 with $\epsilon = 10^{-5}$ would not be useful at all, even though it recovers about 99.9% of the correlation energy.

It is worth recalling that all errors in the absolute correlation energy are one-sided – our local MP2 always underestimates the correlation energy (destabilizing the molecule) and that all conformational energy errors in ACONF20 set are also one-sided – our local MP2 destabilizes compact conformers more. The former is the nature of the fixed-sparsity pattern: we ignore some i, a, j, b combinations, which have negative energy contribution most of the time. The latter could be explained by truncating more $(ia|jb)$ elements that are close to the threshold when the system is more compact. We will see this effect again in the next subsection, when we find that our local MP2 performs less ideally on very compact systems such as fullerene. Therefore, when using this algorithm to compute relative energies, *relative local correlation error* will be smallest for a given cutoff when there are no drastic conformational changes. Otherwise, as this ACONF20 data demonstrates, the user must employ a tighter threshold to ensure adequate precision in the energy differences.

4.2 Timing and Memory Requirement

To see the crossover between our local MP2 and RIMP2 at different ϵ , we plot the timing and memory requirement of linear alkane chains in Figure 4. The memory requirement of RIMP2 and $\epsilon = 10^{-7}$ local MP2 exceeds the total memory of the computer (64 GB) at $C_{70}H_{142}$. It is clear in Figure 4 that the timing and memory requirement is linear with respect to the system size after $C_{30}H_{62}$, and that even with the tightest threshold $\epsilon = 10^{-7}$, we see a crossover in timing between $C_{50}H_{102}$ and $C_{60}H_{122}$.

Table 3 presents a more detailed timing analysis for $C_{40}H_{82}$ for different ϵ . The memory requirement is always controlled by the solver, since CG requires 5 copies of the ragged list J tensor. When ϵ is made 10 times smaller, the memory requirement becomes 2.5 times bigger. The timing, however, is only dominated by the solver part when $\epsilon \leq 10^{-6.5}$. More than 80% of the time in local DF is used to build the $(ia|P)$ tensor, while the mat-vec function (Algorithm 1) takes most of the wall time in the solver. When ϵ is made 10 times

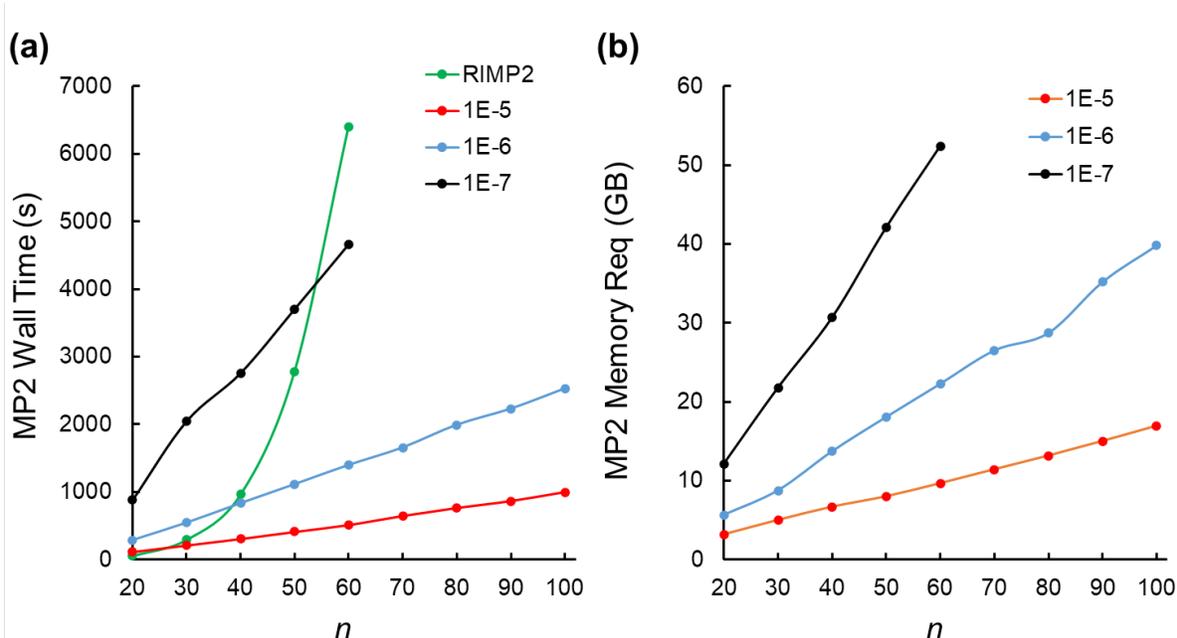


Figure 4: The MP2 (a) wall time and (b) memory requirement for def2-TZVP linear alkane chains C_nH_{2n+2} under different local MP2 threshold ϵ .

Table 3: The detailed wall time breakdown, memory requirement, percentage of significant occupied orbital pairs, and fractional correlation energy error for def2-TZVP linear $C_{40}H_{82}$ using different local MP2 thresholds, ϵ . Time is in seconds, and memory requirement is in gigabytes (GB).

Threshold	RIMP2	$10^{-4.5}$	10^{-5}	$10^{-5.5}$	10^{-6}	$10^{-6.5}$	10^{-7}	$10^{-7.5}$	
Significant i, j Pair (%)	100	23.4	30.5	36.0	42.7	49.6	58.2	67.3	
Fractional E_{corr} Error ($\%$)	0	3.335	0.839	0.211	0.059	0.016	0.005	0.001	
Wall Time	Total	964	215	309	482	834	1636	2757	5203
	LDF		158	181	202	226	248	271	296
	ERI		21	42	74	129	212	339	542
	Solver		32	82	201	474	1171	2142	4359
Local DF Mem Req			5.9	6.7	7.4	9.3	12.1	15.9	22.5
Solver Mem Req			5.9	6.7	8.6	13.7	19.7	30.7	46.3

smaller, the total wall time roughly triples but the solver time increases by 400%, which can be attributed to the locally quadratic nature seen in Eq. 16.

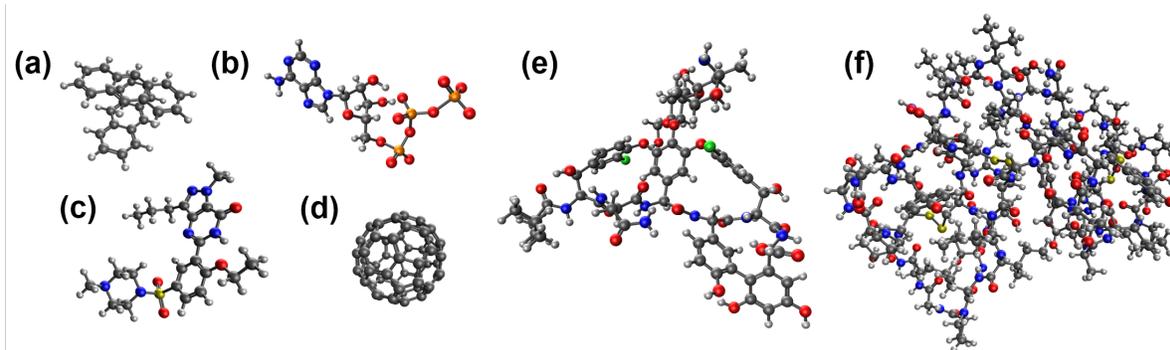


Figure 5: Testing 3-D molecules: (a) anthracene dimer, (b) ATP^{4-} , (c) sildenafil, (d) fullerene, (e) vancomycin, and (f) crambin.

We have also tested our methods on several 3-D molecules (Figure 5) and the results are shown in Table 4. We use the def-SV(P) basis set for fullerene, vancomycin, and crambin, and those molecules were run on a NERSC Perlmutter node with two 64-core AMD EPYC 7763 CPUs and 512 GB RAM. Since the molecules are not large (except for vancomycin and crambin, and we do not have enough RAM to run RIMP2 for crambin), local MP2 takes more time than RIMP2. The error in correlation energy is still about 0.1% when $\epsilon = 10^{-5}$, consistent with the findings in Sec. 4.1. However, in the anthracene dimer and fullerene, which are small and compact systems, the error is larger. In 3-D systems, when ϵ decreases 10-fold, we see a 3-fold to 4-fold increase in the memory requirement, higher than the sparser extended linear alkane chain system. The ratio of computing time between $\epsilon = 10^{-6}$ and $\epsilon = 10^{-5}$ is therefore also higher, at around 4 to 5.

4.3 OpenMP Parallel Efficiency

We have conducted a strong scaling analysis and a weak scaling analysis to understand the efficiency of our OpenMP parallel implementation. All jobs in this section were performed using a single 128 core node (2 AMD 7763 chips) on the NERSC Perlmutter machine.

Table 4: The detailed wall time breakdown, memory requirement, and fractional correlation energy error for some 3-D molecules using thresholds $\epsilon = 10^{-5}, 10^{-6}$. Time is in seconds, and memory requirement is in gigabytes (GB). The def2-TZVP basis set is used except for fullerene, vancomycin, and crambin, where def2-SV(P) is used. Fullerene, vancomycin and crambin were run 32-threaded on a NERSC Perlmutter node, while others were run 16-threaded on a workstation.

	Anthracenes		ATP ⁴⁻		Sildenafil		Fullerene		Vancomycin		Crambin
Threshold	10 ⁻⁵	10 ⁻⁶	10 ⁻⁵	10 ⁻⁶	10 ⁻⁵	10 ⁻⁶	10 ⁻⁵	10 ⁻⁶	10 ⁻⁵	10 ⁻⁶	10 ⁻⁵
Fractional E_{corr} Error (%)	2.91	0.18	1.31	0.08	1.58	0.10	3.87	0.25	0.92	0.07	–
RIMP2 Wall Time	120		187		300		20		514		–
Total	589	2712	413	1612	586	2328	102	355	115	467	2195
LMP2 LDF	175	180	142	147	211	228	15	16	44	52	1152
ERI	197	295	134	301	201	458	55	90	55	295	957
Solver	215	2234	137	1133	165	1527	33	249	15	119	83
Solver Mem Req	7.1	25.0	6.4	19.3	7.2	22.5	13.4	51.2	13.5	40.6	54.7

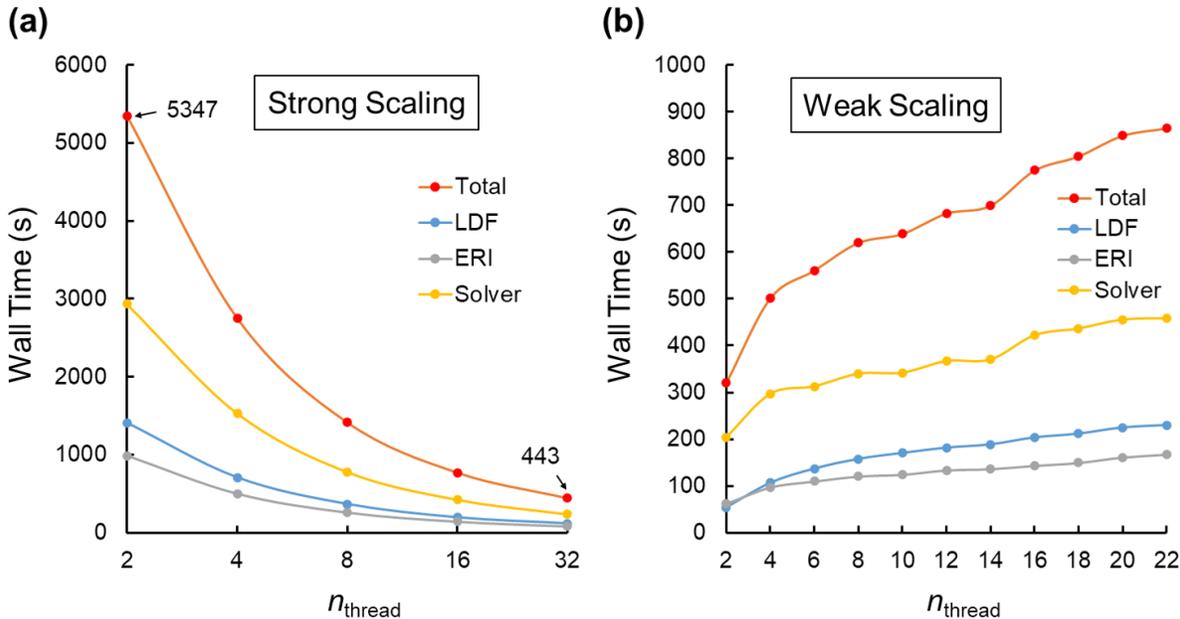


Figure 6: (a) Strong scaling and (b) weak scaling local MP2 wall time analysis of OpenMP parallel efficiency. Systems used are $\text{C}_{160}\text{H}_{322}$ for strong scaling and for weak scaling, $\text{C}_{20}\text{H}_{42}$, $\text{C}_{40}\text{H}_{82}$, \dots , to $\text{C}_{220}\text{H}_{442}$ using 2 to 22 threads, respectively. Threshold $\epsilon = 10^{-6}$, and basis set is def2-TZVP. These jobs are run on a NERSC Perlmutter node.

$C_{160}H_{322}$ was used for the strong scaling analysis, and we varied the number of OpenMP threads from 2 to 32. The sequence of molecules, $C_{20}H_{42}$, $C_{40}H_{82}$, \dots , to $C_{220}H_{442}$ were used for weak scaling analysis, with 2 to 22 threads, respectively. We encountered SCF issues for larger jobs. The def2-TZVP basis was used for all jobs, and a threshold of $\epsilon = 10^{-6}$ was employed for local MP2. Ideally, in the strong scaling analysis, the timing should be inversely proportional to the number of threads, while in the weak scaling analysis, the wall time should ideally be the same in each case.

Figure 6 plots the result. In the strong scaling analysis, we see that by using 32 threads, the wall time decreases by a factor of 12 compared with a 2-thread run. This indicates a satisfactory overall parallel efficiency of roughly 75% for this range of threads. In the weak scaling test, we do not see a flat line. The jump in the solver time from $C_{140}H_{282}$ to $C_{160}H_{322}$ is because the number of iterations increases by one. Other than that, we see generally flat lines in all three parts of the algorithm, resulting in a mildly increasing total time curve. $C_{20}H_{42}$ is an outlier because it is not in the linear region for $\epsilon = 10^{-6}$, and the LDF curve stabilizes later than the other two components as the doubly extended map (Eq. 18) becomes linear-scaling later than the other parts of the algorithm. Also, some small non-linear parts of the algorithm (for example, the building of all sparse maps and F_{occ}, F_{virt}) can contribute to the slight increase in the weak scaling time. Overall, the parallel efficiency is satisfactory, and helps enable the evaluation of large jobs.

4.4 Basis Set Transferability

In this subsection, we discuss the transferability of our results to other basis sets. Figure 7 plots the memory requirement and the wall time for $C_{30}H_{62}$ using several different Karlsruhe def2- series basis set with $\epsilon = 10^{-6}$. Naturally, a bigger basis set requires more time and memory. Fortunately adding diffuse functions does not seem to affect the performance of our algorithm. The fractional error of all calculations (Table 5) is under 0.01%, which is

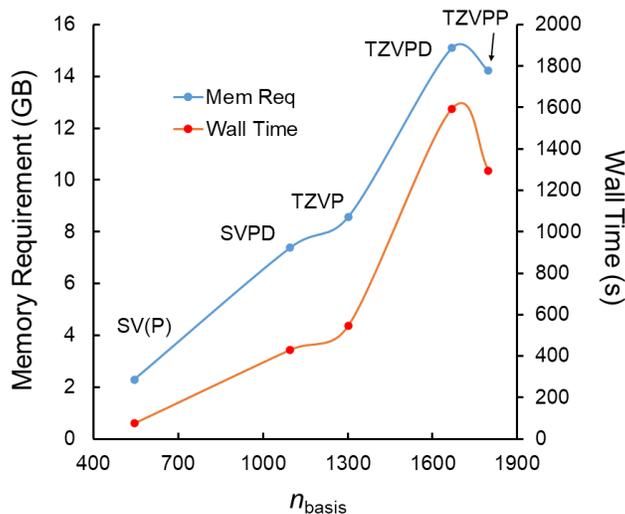


Figure 7: The timing and memory requirement for local MP2 of $\text{C}_{30}\text{H}_{62}$ with different basis sets. Threshold $\epsilon = 10^{-6}$. The def2- basis sets are used; i.e., “SV(P)” means the def2-SV(P) basis set.

consistent, but a larger basis set tends to have a larger error. Table 6 shows the sparsity of $(ia|P)$, C_{ia}^P , and $(ia|jb)$ tensors of vancomycin using different basis sets with $\epsilon = 10^{-5}$. Encouragingly, we keep about the same fraction of the three-index tensors, regardless of whether we add diffuse functions to our basis or not. The fraction of significant elements in $(ia|jb)$ decreases when the basis is bigger, suggesting that in correlation methods which must store the full amplitude tensor (e.g., CCSD), the local version will have a better memory footprint in bigger basis sets well before the linear scaling regime.

Table 5: The fractional correlation energy error between local and canonical MP2 and the number of local MP2 CG iterations of $\text{C}_{30}\text{H}_{62}$ with different basis sets. Threshold $\epsilon = 10^{-6}$. The def2- sets of basis functions are used; i.e., a “SV(P)” code means def2-SV(P) basis set. The original VV-HV set corresponds to the definition introduced in Ref. 75, while the canonicalized VV-HV set has hard virtuals on each atom pseudo-canonicalized.

Basis		SV(P)	SVPD	TZVP	TZVPD	TZVPP
Fractional E_{corr} Error (%)		0.020	0.032	0.047	0.066	0.075
# of Iterations	Original VV-HV	5	7	15	17	17
	Canonicalized VV-HV	5	6	7	10	9

Table 6: The fractional of significant i, j pair, $(ia|P)$, C_{ia}^P , and $(ia|jb)$ elements in vancomycin with different basis sets. Threshold $\epsilon = 10^{-5}$. The def2- sets of basis functions are used; i.e., a “SV(P)” code means def2-SV(P) basis set.

Basis		SV(P)	SVPD	TZVP	TZVPD
Fraction	i, j Pair	39.1	40.9	40.3	40.4
	$(ia P)$	36.4	37.9	37.4	37.2
Significant (%)	C_{ia}^P	8.4	10.7	8.5	10.2
	$(ia jb)$	0.169	0.128	0.076	0.078

The number of iterations, which is crucial to the efficiency of the solver, gradually increases when a larger basis set is used. The reason might be that to meet the threshold requirement of the residual r , a longer vectorized J tensor requires smaller element-wise error, which in turn requires more iterations. Also, in the Fock matrix, large off-diagonal values make the linear system deviate further from a diagonal system, meaning that our diagonal preconditioner will be less effective and therefore more iterations may be needed. If we use the original VV-HV orbitals as in Ref. 75, hard-virtuals on the same atom tend to have large off-diagonal values (Figure 8(b)), so the number of iterations needed goes much higher for the TZ basis set. After an atom-wise pseudo-canonicalization of the hard-virtuals, we make the Fock matrix more diagonal (Figure 8(a)) without losing much sparsity, and therefore, the number of iterations goes down by a large fraction. The fraction of $(ia|jb)$ elements kept and the fractional correlation energy errors are almost the same when using these two versions of the VV-HV orbitals.

4.5 Smoothness of the Potential Energy Surface

One of the drawbacks of many local correlation methods is that the potential energy surface ceases to be smooth and differentiable due to all the non-negligible truncations in the algorithm. Therefore, properties that depend on the (numerical) gradient will be much less

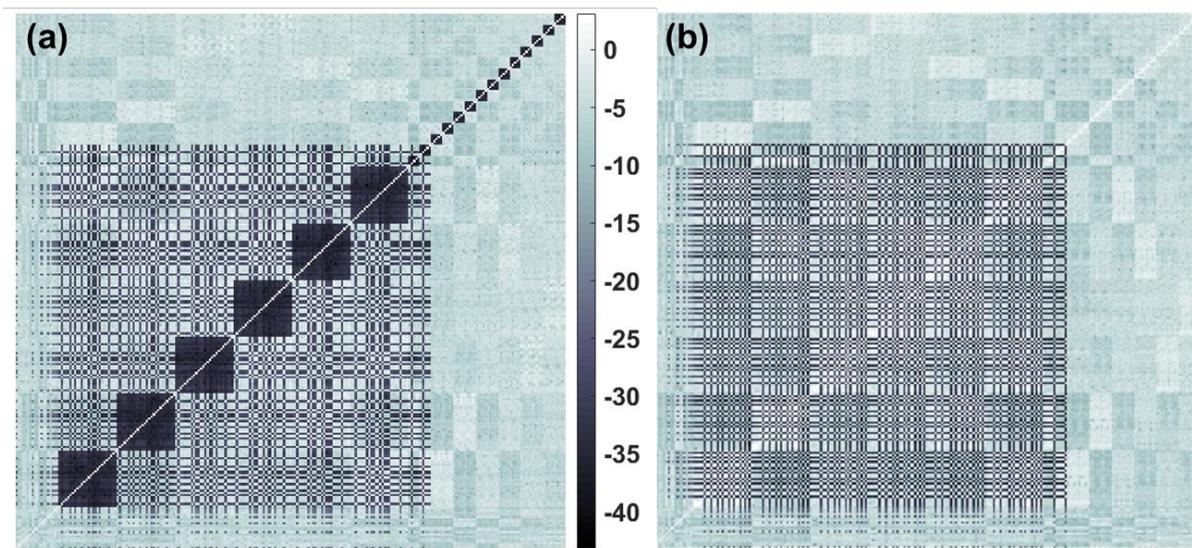


Figure 8: The magnitude of all elements in the Fock matrix of def2-TZVP C_6H_{14} with (a) pseudo-canonicalized VV-HV and (b) original VV-HV. Original VV-HV set corresponds to the orbital in Ref. 75, and canonicalized VV-HV set have hard-virtuals on each atom pseudo-canonicalized.

Table 7: The dipole vector and the polarizability tensor calculated from numerical differentiation for def2-TZVP ATP^{4-} . Step length 0.001 a.u.

Threshold	RIMP2	10^{-5}
Dipole	(29.3099, -2.0215, 2.6170)	(29.3169, -2.0305, 2.6319)
Polarizability	$\begin{pmatrix} 347.084 & -0.656 & 43.598 \\ -0.656 & 190.929 & 8.311 \\ 43.598 & 8.311 & 292.062 \end{pmatrix}$	$\begin{pmatrix} 347.005 & -2.988 & 47.810 \\ -2.988 & 173.941 & 18.930 \\ 47.810 & 18.930 & 289.684 \end{pmatrix}$
Threshold	10^{-6}	10^{-7}
Dipole	(29.3117, -2.0203, 2.6201)	(29.3103, -2.0213, 2.6175)
Polarizability	$\begin{pmatrix} 345.256 & 0.231 & 44.969 \\ 0.231 & 189.988 & 9.152 \\ 44.969 & 9.152 & 289.050 \end{pmatrix}$	$\begin{pmatrix} 347.582 & 0.409 & 44.448 \\ 0.409 & 191.037 & 8.782 \\ 44.448 & 8.782 & 291.331 \end{pmatrix}$

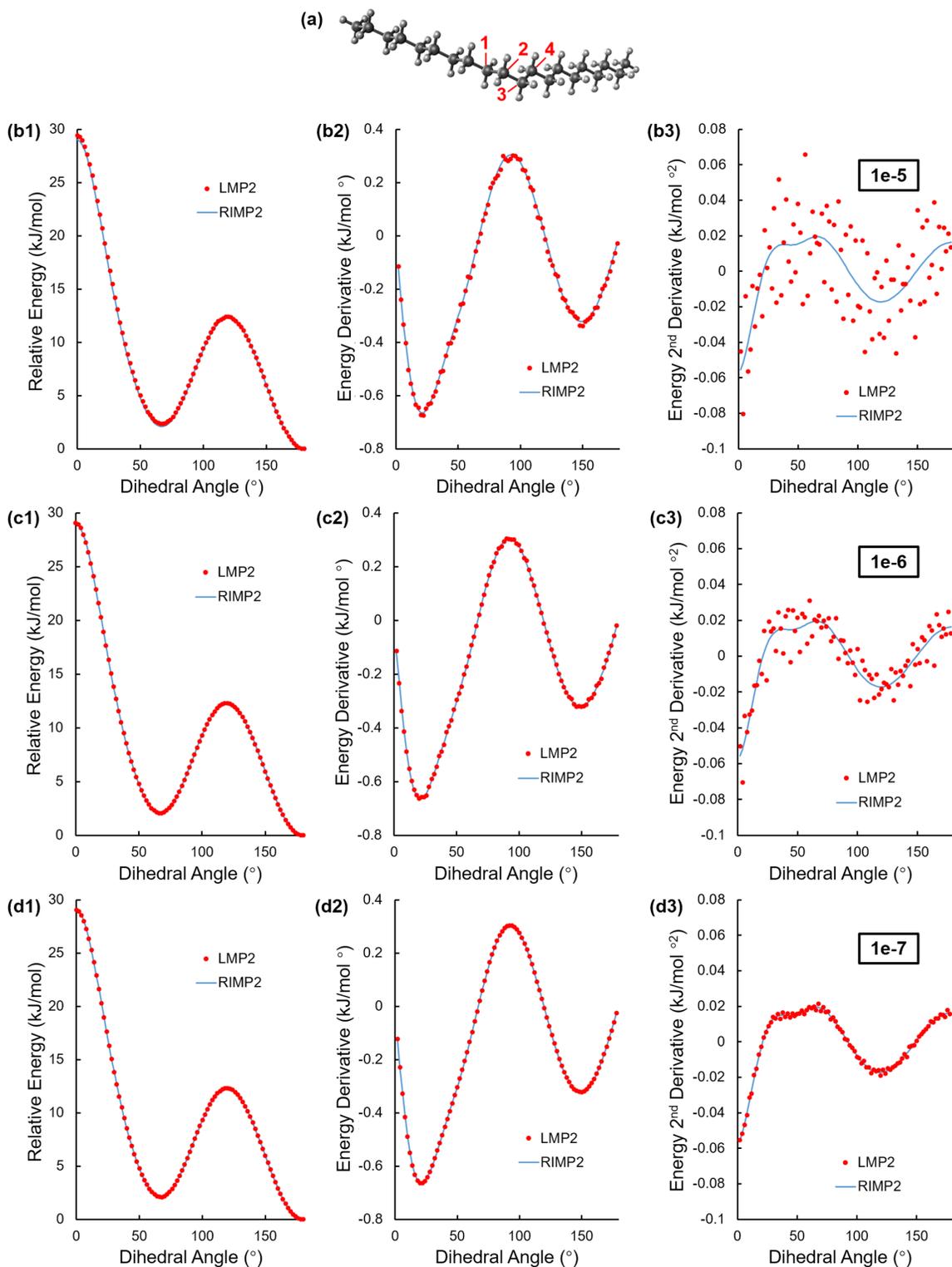


Figure 9: The potential energy surface and numerical energy derivative and second derivative of def2-TZVP $C_{20}H_{42}$ rotating the (a) C1-C2-C3-C4 dihedral angle under different threshold $\epsilon = 10^{-5}, 10^{-6}, 10^{-7}$. Energies are relative to the all-trans conformer, where the dihedral angle is 180° . Step length 2 degrees.

accurate, limiting the value of the local method for solving chemical problems. The numerical precision issue may be even more acute for second derivative properties such as force constants. The appropriate choice of truncation threshold should therefore be tested on some chemically relevant examples.

For this purpose, we tested the smoothness of our algorithm for a torsional potential energy surface of $C_{20}H_{42}$ with $\epsilon = 10^{-5}, 10^{-6}, 10^{-7}$. We varied the central C1-C2-C3-C4 dihedral angle as shown in Figure 9)(a). Panels (b1,c1,d1) plots the potential energy surface relative to the all-trans conformer (180° dihedral angle), which has the lowest energy. Since our local MP2 destabilizes molecules, we typically overestimate the conformational energy difference. The error is within 0.5 kJ/mol, 0.06 kJ/mol, and 0.007 kJ/mol for $\epsilon = 10^{-5}, 10^{-6}$ and 10^{-7} , respectively. The energy error is better than the performance in ACONF20 since the conformational differences between members of the ACONF20 set are more drastic and tend to create more errors. For the numerical energy derivative (steplength 2°), the surface in Figure 9(b1) is close to the RIMP2 reference line but visually rough, and the result improves gradually when ϵ decreases. The average (absolute) error goes from 0.1 kJ/(mol $^\circ$) when $\epsilon = 10^{-5}$ to 0.03 kJ/(mol $^\circ$) for $\epsilon = 10^{-6}$, and then to 0.005 kJ/(mol $^\circ$) for $\epsilon = 10^{-7}$. The speed of improvement in the energy derivative is slower than the speed of improvement in energy. Panels (b3,c3,d3) plot the performance in the energy second derivative. The $\epsilon = 10^{-5}$ result is far away from the RIMP2 reference, improving by roughly a factor of 3 when $\epsilon = 10^{-6}$, and another factor of 7 when $\epsilon = 10^{-7}$.

As a second test of another class of derivative properties, we also calculated the dipole vector and the polarizability tensor for ATP^{4-} (Table 7). We evaluated the first and second electric field derivatives numerically, using a step length of 0.001 a.u.^{92,93} The $\epsilon = 10^{-5}$ dipole moment is correct to about 0.1 a.u., and an extra significant figure can be gained by tightening ϵ to 10^{-6} and 10^{-7} . The $\epsilon = 10^{-5}$ polarizability has a very large error - the largest error is about 17 a.u. in α_{yy} . A significant improvement can be seen when using the 10^{-6}

threshold, as all elements are within 2 a.u., another 10 times smaller ϵ gives a maximum matrix element error of 0.7 a.u.

Summarizing this section, a loose threshold $\epsilon = 10^{-5}$ gives a visually good potential energy surface and marginally usable numerical energy derivatives, but unacceptable numerical second derivatives. $\epsilon = 10^{-6}$ nicely cleans up the first derivative properties but is still not adequate for second derivatives. Only when a strict threshold $\epsilon = 10^{-7}$ is used can we obtain a useful numerical second derivative.

5 Conclusions

To sum up, using Boys occupied orbitals and local orthogonal virtual orbitals (Boys valence virtuals and atomically pseudo-canonicalized hard virtuals), we designed and implemented a local MP2 algorithm based on fixed sparsity pattern and local density fitting. As a key feature, our framework only requires a single input parameter, the J tensor truncation threshold, ϵ . The error of our algorithm varies linearly with respect to ϵ : our local MP2 is generally able to capture 99.9% of the canonical RIMP2 correlation energy at $\epsilon = 10^{-5}$ and 99.999% of correlation energy at $\epsilon = 10^{-7}$. We tested our method on the ACONF20 benchmark set, and the difference between the LMP2 and RIMP2 conformational energy can also be easily controlled by ϵ : $\epsilon = 10^{-5}$ gives a discrepancy of about 3 kJ/mol, and $\epsilon = 10^{-6.5}$ is able to keep the error below 0.1 kJ/mol most of the time. Our algorithm is linear-scaling with respect to system size both in the wall time and in the memory requirement (although there are some small non-linear scaling steps). The shared memory OpenMP parallel efficiency is good but not perfect. The precision considerations and parallel performance can be transferred to other bigger or smaller basis sets.

For a torsional potential energy scan, the surface is visually smooth, and the numerical derivative may be acceptable when $\epsilon = 10^{-5}$, and is significantly improved with $\epsilon = 10^{-6}$.

A numerical second derivative calculation needs a tighter threshold $\epsilon = 10^{-7}$ to produce meaningful results. With care, our algorithm can be successfully deployed with a threshold that produces enough significant figures that it is a direct replacement for conventional MP2 in practice. We emphasize that the appropriate choice of ϵ is critical to achieving useful precision (i.e. adequately approximating MP2) and good performance (i.e. much lower compute cost than MP2) at the same time. Our examples provide some guidance in this tradeoff, as well as clearly revealing that the optimal ϵ is application-dependent. As computer resources improve, it should be possible to choose ϵ more conservatively to ensure too much precision rather than too little.

Beyond its deployment for large-molecule chemical applications, there is considerable scope for further development of the ideas presented here. One important and non-trivial next step is to develop a distributed memory implementation of our algorithm. A strong motivation is that sufficient memory is a requirement to run our algorithm, and that requirement increases steeply with tighter thresholds to easily exceed the single-node memory limit. Distributed computing also opens the door to larger molecules where the speedups relative to exact MP2 will still be larger. Another very desirable extension is to adapt our local correlation algorithm to implement local MP3 and CCD/CCSD. While considerable additional development is needed, our amplitude solver code can be reused without many modifications, and the same is true for much of the integral assembly algorithm. In addition, we can more readily implement other promising variations of MP2, such as regularized MP2,^{94,95} for applications to large molecules. Additionally, a closer look at the numerical gradient of our local MP2 energy may also be useful, as the analytical gradient should be viable to implement given that its accuracy can be assured.

Conflicts of Interest

MHG is a part-owner of Q-Chem Inc, which is the software platform used to implement the algorithms described here.

Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing, and Office of Basic Energy Sciences, via the Scientific Discovery through Advanced Computing (SciDAC) program. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231.

References

- (1) Cleland, D.; Booth, G. H.; Alavi, A. Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum Monte Carlo. *J. Chem. Phys.* **2010**, *132*, 041103.
- (2) Li Manni, G.; Smart, S. D.; Alavi, A. Combining the complete active space self-consistent field method and the full configuration interaction quantum Monte Carlo within a super-CI framework, with application to challenging metal-porphyrins. *J. Chem. Theory Comput.* **2016**, *12*, 1245–1258.
- (3) Tubman, N. M.; Lee, J.; Takeshita, T. Y.; Head-Gordon, M.; Whaley, K. B. A deterministic alternative to the full configuration interaction quantum Monte Carlo method. *J. Chem. Phys.* **2016**, *145*, 044112.

- (4) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic heat-bath configuration interaction method: Selected configuration interaction with semistochastic perturbation theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595–1604.
- (5) Zimmerman, P. M. Incremental full configuration interaction. *J. Chem. Phys.* **2017**, *146*, 104102.
- (6) Tubman, N. M.; Freeman, C. D.; Levine, D. S.; Hait, D.; Head-Gordon, M.; Whaley, K. B. Modern approaches to exact diagonalization and selected configuration interaction with the adaptive sampling CI method. *J. Chem. Theory Comput.* **2020**, *16*, 2139–2159.
- (7) Cotton, S. J. A truncated Davidson method for the efficient “chemically accurate” calculation of full configuration interaction wavefunctions without any large matrix diagonalization. *J. Chem. Phys.* **2022**, *157*, 224105.
- (8) Jensen, F. Atomic orbital basis sets. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 273–295.
- (9) Nagy, B.; Jensen, F. Basis sets in quantum chemistry. *Rev. Comput. Chem.* **2017**, *30*, 93–149.
- (10) Head-Gordon, M. Quantum chemistry and molecular processes. *J. Phys. Chem.* **1996**, *100*, 13213–13225.
- (11) Pople, J. A. In *Energy, Structure, and Reactivity: Proceedings of the 1972 Boulder Summer Research Conference on Theoretical Chemistry*; Smith, D. W., McRae, W. B., Eds.; John Wiley & Sons: New York, 1973; pp 51–61.
- (12) Pople, J. A. Nobel lecture: Quantum chemical models. *Rev. Mod. Phys.* **1999**, *71*, 1267.

- (13) Kohn, W.; Becke, A. D.; Parr, R. G. Density functional theory of electronic structure. *J. Phys. Chem.* **1996**, *100*, 12974.
- (14) 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.
- (15) Cremer, D. Moller-Plesset perturbation theory: from small molecule methods to methods for thousands of atoms. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 509–530.
- (16) Bartlett, R. J.; Musial, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291–352.
- (17) Bartlett, R. J. Coupled-cluster theory and its equation-of-motion extensions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 126–138.
- (18) Purvis, G. D.; Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910.
- (19) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479.
- (20) Goerigk, L.; Grimme, S. Double-hybrid density functionals. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 576–600.
- (21) Mardirossian, N.; Head-Gordon, M. Survival of the most transferable at the top of Jacob’s ladder: Defining and testing the ω B97M(2) double hybrid density functional. *J. Chem. Phys.* **2018**, *148*, 241736.

- (22) Martin, J. M. L.; Santra, G. Empirical double-hybrid density functional theory: A ‘third way’ in between WFT and DFT. *Isr. J. Chem.* **2020**, *60*, 787–804.
- (23) Baer, R.; Head-Gordon, M. Sparsity of the density matrix in kohn-sham density functional theory and an assessment of linear system-size scaling methods. *Phys. Rev. Lett.* **1997**, *79*, 3962–3965.
- (24) Høyvik, I.-M.; Jørgensen, P. Characterization and generation of local occupied and virtual Hartree–Fock orbitals. *Chem. Rev.* **2016**, *116*, 3306–3327.
- (25) Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151–154.
- (26) Maslen, P. E.; Ochsenfeld, C.; White, C. A.; Lee, M. S.; Head-Gordon, M. Locality and sparsity of Ab initio one-particle density matrices and localized orbitals. *J. Phys. Chem. A* **1998**, *102*, 2215–2222.
- (27) Hetzer, G.; Pulay, P.; Werner, H.-J. Multipole approximation of distant pair energies in local MP2 calculations. *Chem. Phys. Lett.* **1998**, *290*, 143–149.
- (28) Werner, H.-J. Communication: Multipole approximations of distant pair energies in local correlation methods with pair natural orbitals. *J. Chem. Phys.* **2016**, *145*, 201101.
- (29) Mackie, C. J.; Gonthier, J. F.; Head-Gordon, M. Compressed intramolecular dispersion interactions. *J. Chem. Phys.* **2020**, *152*, 024112.
- (30) Saebo, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213–236.
- (31) Boughton, J. W.; Pulay, P. Comparison of the boys and Pipek–Mezey localizations in the local correlation approach and automatic virtual basis selection. *J. Comput. Chem.* **1993**, *14*, 736–740.

- (32) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. Closely approximating second-order Møller–Plesset perturbation theory with a local triatomics in molecules model. *J. Chem. Phys.* **2000**, *112*, 3592–3601.
- (33) DiStasio, R. A.; Jung, Y.; Head-Gordon, M. A resolution-of-the-identity implementation of the local triatomics-in-molecules model for second-order Møller–Plesset perturbation theory with application to alanine tetrapeptide conformational energies. *J. Chem. Theory Comput.* **2005**, *1*, 862–876.
- (34) Schütz, M.; Hetzer, G.; Werner, H. J. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *J. Chem. Phys.* **1999**, *111*, 5691–5705.
- (35) Ayala, P. Y.; Scuseria, G. E. Linear scaling second-order Møller–Plesset theory in the atomic orbital basis for large molecular systems. *J. Chem. Phys.* **1999**, *110*, 3660–3671.
- (36) Dunlap, B. I. Robust and variational fitting. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2113–2116.
- (37) Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast linear scaling second-order Møller–Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* **2003**, *118*, 8149–8160.
- (38) Edmiston, C.; Krauss, M. Pseudonatural Orbitals as a Basis for the Superposition of Configurations. I. He²⁺. *J. Chem. Phys.* **1966**, *45*, 1833–1839.
- (39) Meyer, W. PNO–CI Studies of electron correlation effects. I. Configuration expansion by means of nonorthogonal orbitals, and application to the ground state and ionized states of methane. *The Journal of Chemical Physics* **1973**, *58*, 1017–1035.
- (40) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to

- coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *J. Chem. Phys.* **2009**, *130*, 114108.
- (41) Werner, H.-J.; Knizia, G.; Krause, C.; Schwilk, M.; Dornbach, M. Scalable Electron Correlation Methods I.: PNO-LMP2 with Linear Scaling in the Molecular Size and Near-Inverse-Linear Scaling in the Number of Processors. *J. Chem. Theory Comput.* **2015**, *11*, 484–507.
- (42) Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F. Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals. *J. Chem. Phys.* **2015**, *143*, 034108.
- (43) Almlöf, J. Elimination of energy denominators in Møller–Plesset perturbation theory by a Laplace transform approach. *Chem. Phys. Lett.* **1991**, *181*, 319–320.
- (44) Häser, M. Møller-Plesset (MP2) perturbation theory for large molecules. *Theor. Chem. Acc.* **1993**, *87*, 147–173.
- (45) Doser, B.; Lambrecht, D. S.; Kussmann, J.; Ochsenfeld, C. Linear-scaling atomic orbital-based second-order Møller–Plesset perturbation theory by rigorous integral screening criteria. *J. Chem. Phys.* **2009**, *130*, 064107.
- (46) Zienau, J.; Clin, L.; Doser, B.; Ochsenfeld, C. Cholesky-decomposed densities in Laplace-based second-order Møller–Plesset perturbation theory. *J. Chem. Phys.* **2009**, *130*, 204112.
- (47) Maurer, S. A.; Clin, L.; Ochsenfeld, C. Cholesky-decomposed density MP2 with density fitting: Accurate MP2 and double-hybrid DFT energies for large systems. *The Journal of chemical physics* **2014**, *140*.

- (48) Demel, O.; Lecours, M. J.; Habrovský, R.; Nooijen, M. Toward Laplace MP2 method using range separated Coulomb potential and orbital selective virtuals. *J. Chem. Phys.* **2021**, *155*, 154104.
- (49) Kobayashi, M.; Imamura, Y.; Nakai, H. Alternative linear-scaling methodology for the second-order Møller-Plesset perturbation calculation based on the divide-and-conquer method. *J. Chem. Phys.* **2007**, *127*, 074103.
- (50) Yoshikawa, T.; Kobayashi, M.; Nakai, H. Linear-scaling divide-and-conquer second-order Møller-Plesset perturbation calculation for open-shell systems: implementation and application. *Theor. Chem. Acc.* **2011**, *130*, 411–417.
- (51) Fujimori, T.; Kobayashi, M.; Taketsugu, T. Energy-based automatic determination of buffer region in the divide-and-conquer second-order Møller-Plesset perturbation theory. *J. Comput. Chem.* **2021**, *42*, 620–629.
- (52) Nakai, H.; Kobayashi, M.; Yoshikawa, T.; Seino, J.; Iwabata, Y.; Nishimura, Y. Divide-and-Conquer Linear-Scaling Quantum Chemical Computations. *J. Phys. Chem. A* **2023**, *127*, 589–618.
- (53) Kristensen, K.; Høyvik, I.-M.; Jansik, B.; Jørgensen, P.; Kjærgaard, T.; Reine, S.; Jakowski, J. MP2 energy and density for large molecular systems with internal error control using the Divide-Expand-Consolidate scheme. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15706–15714.
- (54) Baudin, P.; Ettenhuber, P.; Reine, S.; Kristensen, K.; Kjærgaard, T. Efficient linear-scaling second-order Møller-Plesset perturbation theory: The divide-expand-consolidate RI-MP2 model. *J. Chem. Phys.* **2016**, *144*, 054102.
- (55) Li, W.; Guo, Y.; Li, S. A refined cluster-in-molecule local correlation approach for

- predicting the relative energies of large systems. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7854–7862.
- (56) Li, W.; Ni, Z.; Li, S. Cluster-in-molecule local correlation method for post-Hartree–Fock calculations of large systems. *Mol. Phys.* **2016**, *114*, 1447–1460.
- (57) Nagy, P. R.; Samu, G.; Kállay, M. An integral-direct linear-scaling second-order Møller–Plesset approach. *J. Chem. Theory Comput.* **2016**, *12*, 4897–4914.
- (58) Ni, Z.; Li, W.; Li, S. Fully optimized implementation of the cluster-in-molecule local correlation approach for electron correlation calculations of large systems. *J. Comput. Chem.* **2019**, *40*, 1130–1140.
- (59) Szabó, P. B.; Csóka, J.; Kállay, M.; Nagy, P. R. Linear-scaling open-shell MP2 approach: Algorithm, benchmarks, and large-scale applications. *J. Chem. Theory Comput.* **2021**, *17*, 2886–2905.
- (60) Ni, Z.; Guo, Y.; Neese, F.; Li, W.; Li, S. Cluster-in-molecule local correlation method with an accurate distant pair correction for large systems. *J. Chem. Theory Comput.* **2021**, *17*, 756–766.
- (61) Hohenstein, E. G.; Parrish, R. M.; Martínez, T. J. Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller–Plesset perturbation theory. *J. Chem. Phys.* **2012**, *137*, 044103.
- (62) Kokkila Schumacher, S. I.; Hohenstein, E. G.; Parrish, R. M.; Wang, L.-P.; Martínez, T. J. Tensor hypercontraction second-order Møller–Plesset perturbation theory: Grid optimization and reaction energies. *J. Chem. Theory Comput.* **2015**, *11*, 3042–3052.

- (63) Lee, J.; Lin, L.; Head-Gordon, M. Systematically improvable tensor hypercontraction: Interpolative separable density-fitting for molecules applied to exact exchange, second- and third-order Møller–Plesset perturbation theory. *J. Chem. Theory Comput.* **2019**, *16*, 243–263.
- (64) Zhao, T.; Simons, M.; Matthews, D. A. Open-Shell Tensor Hypercontraction. *J. Chem. Theory Comput.* **2023**, *19*, 3996–4010.
- (65) Kjærgaard, T. The Laplace transformed divide-expand-consolidate resolution of the identity second-order Møller–Plesset perturbation (DEC-LT-RIMP2) theory method. *J. Chem. Phys.* **2017**, *146*, 044103.
- (66) Bangerter, F. H.; Glasbrenner, M.; Ochsenfeld, C. Low-scaling tensor hypercontraction in the Cholesky molecular orbital basis applied to second-order Møller–Plesset perturbation theory. *J. Chem. Theory Comput.* **2020**, *17*, 211–221.
- (67) Glasbrenner, M.; Graf, D.; Ochsenfeld, C. Efficient Reduced-Scaling Second-Order Møller–Plesset Perturbation Theory with Cholesky-Decomposed Densities and an Attenuated Coulomb Metric. *J. Chem. Theory Comput.* **2020**, *16*, 6856–6868.
- (68) Bykov, D.; Kjaergaard, T. The GPU-enabled divide-expand-consolidate RI-MP2 method (DEC-RI-MP2). *J. Comput. Chem.* **2017**, *38*, 228–237.
- (69) Kjærgaard, T.; Baudin, P.; Bykov, D.; Eriksen, J. J.; Ettenhuber, P.; Kristensen, K.; Larkin, J.; Liakh, D.; Pawłowski, F.; Vose, A., et al. Massively parallel and linear-scaling algorithm for second-order Møller–Plesset perturbation theory applied to the study of supramolecular wires. *Comput. Phys. Commun.* **2017**, *212*, 152–160.
- (70) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. Scaled opposite-spin second order Møller–Plesset correlation energy: an economical electronic structure method. *J. Chem. Phys.* **2004**, *121*, 9793.

- (71) Jung, Y.; Shao, Y.; Head-Gordon, M. Fast evaluation of scaled opposite spin second-order Møller–Plesset correlation energies using auxiliary basis expansions and exploiting sparsity. *J. Comput. Chem.* **2007**, *28*, 1953–1964.
- (72) Barca, G. M.; McKenzie, S. C.; Bloomfield, N. J.; Gilbert, A. T.; Gill, P. M. Q-mp2-os: Møller–plesset correlation energy by quadrature. *J. Chem. Theory Comput.* **2020**, *16*, 1568–1577.
- (73) Förster, A.; Franchini, M.; van Lenthe, E.; Visscher, L. A quadratic pair atomic resolution of the identity based SOS-AO-MP2 algorithm using slater type orbitals. *J. Chem. Theory Comput.* **2020**, *16*, 875–891.
- (74) Wang, Z.; Aldossary, A.; Head-Gordon, M. Sparsity of the electron repulsion integral tensor using different localized virtual orbital representations in local second-order Møller–Plesset theory. *J. Chem. Phys.* **2023**, *158*, 064105.
- (75) Subotnik, J. E.; Dutoi, A. D.; Head-Gordon, M. Fast localized orthonormal virtual orbitals which depend smoothly on nuclear coordinates. *J. Chem. Phys.* **2005**, *123*, 114108.
- (76) Krause, C.; Werner, H.-J. Comparison of explicitly correlated local coupled-cluster methods with various choices of virtual orbitals. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7591–7604.
- (77) Hansen, A. S.; Baardsen, G.; Rebolini, E.; Maschio, L.; Pedersen, T. B. Representation of the virtual space in extended systems—a correlation energy convergence study. *Mol. Phys.* **2020**, *118*, e1733118.
- (78) Høyvik, I.-M. The spectrum of the atomic orbital overlap matrix and the locality of the virtual electronic density matrix. *Mol. Phys.* **2020**, *118*, e1765034.

- (79) Folkestad, S. D.; Matveeva, R.; Høyvik, I.-M.; Koch, H. Implementation of occupied and virtual Edmiston–Ruedenberg orbitals using Cholesky decomposed integrals. *J. Chem. Theory Comput.* **2022**, *18*, 4733–4744.
- (80) Boys, S. F. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.* **1960**, *32*, 296.
- (81) Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* **1989**, *90*, 4916–4926.
- (82) Edmiston, C.; Ruedenberg, K. Localized atomic and molecular orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457.
- (83) Yang, J.; Kurashige, Y.; Manby, F. R.; Chan, G. K.-L. L. Tensor factorizations of local second-order Møller–Plesset theory. *J. Chem. Phys.* **2011**, *134*, 044123.
- (84) Subotnik, J. E.; Shao, Y.; Liang, W.; Head-Gordon, M. An efficient method for calculating maxima of homogeneous functions of orthogonal matrices: Applications to localized occupied orbitals. *J. Chem. Phys.* **2004**, *121*, 9220.
- (85) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. Localized orbital theory and ammonia triborane. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5522–5530.
- (86) Wang, Z.; Head-Gordon, M. A Linear Surrogate for Optimizing Functions of an Orthogonal Matrix with Applications in Wave Function Theory. *Mol. Phys.* **2023**, *121*, e2118185.
- (87) Aldossary, A.; Head-Gordon, M. Non-iterative method for constructing valence antibonding molecular orbitals and a molecule-adapted minimum basis. *J. Chem. Phys.* **2022**, *157*, 094102.

- (88) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover, 1996.
- (89) Gonthier, J. F.; Head-Gordon, M. Assessing Electronic Structure Methods for Long-Range Three-Body Dispersion Interactions: Analysis and Calculations on Well-Separated Metal Atom Trimers. *J. Chem. Theory Comput.* **2019**, *15*, 4351–4361.
- (90) Epifanovsky, E.; Gilbert, A. T.; Feng, X.; Lee, J.; Mao, Y.; Mardirossian, N.; Pokhilko, P.; White, A. F.; Coons, M. P.; Dempwolff, A. L., 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.
- (91) Ehlert, S.; Grimme, S.; Hansen, A. Conformational energy benchmark for longer n-alkane chains. *J. Phys. Chem. A* **2022**, *126*, 3521–3535.
- (92) Hait, D.; Head-Gordon, M. How accurate is density functional theory at predicting dipole moments? An assessment using a new database of 200 benchmark values. *J. Chem. Theory Comput.* **2018**, *14*, 1969–1981.
- (93) Hait, D.; Head-Gordon, M. How accurate are static polarizability predictions from density functional theory? An assessment over 132 species at equilibrium geometry. *Phys. Chem. Chem. Phys.* **2018**, *20*, 19800–19810.
- (94) Shee, J.; Loipersberger, M.; Rettig, A.; Lee, J.; Head-Gordon, M. Regularized second-order Møller–Plesset theory: A more accurate alternative to conventional MP2 for non-covalent interactions and transition metal thermochemistry for the same computational cost. *J. Phys. Chem. Lett.* **2021**, *12*, 12084–12097.
- (95) Carter-Fenk, K.; Head-Gordon, M. Repartitioned Brillouin-Wigner perturbation theory with a size-consistent second-order correlation energy. *J. Chem. Phys.* **2023**, *158*, 234108.

TOC Graphic

