Quantum Chemistry for Molecules at Extreme Pressure on Graphical Processing Units: Implementation of Extreme Pressure Polarizable Continuum Model

Ariel Gale,[1] Eugen Hruska,[1] and Fang Liu[1, a)]

Emory University, Department of Chemistry, Atlanta, Georgia 30322,
USA

(Dated: 11 May 2021)

Pressure plays essential roles in chemistry by altering structures and controlling chemical reactions. The extreme-pressure polarizable continuum model (XP-PCM) is an emerging method with an efficient quantum mechanical description of small and medium-size molecules at high pressure (on the order of GPa). However, its application to large molecular systems was previously hampered by CPU computation bottleneck: the Pauli repulsion potential unique to XP-PCM requires the evaluation of a large number of electric field integrals, resulting in significant computational overhead compared to the gas-phase or standard-pressure polarizable continuum model calculations. Here, we exploit advances in Graphical Processing Units (GPUs) to accelerate the XP-PCM integral evaluations. This enables high-pressure quantum chemistry simulation of proteins that used to be computationally intractable. We benchmarked the performance using 18 small proteins in aqueous solutions. Using a single GPU, our method evaluates the XP-PCM free energy of a protein with over 500 atoms and 4000 basis functions within half an hour. The time taken by the XP-PCM-integral evaluation is typically 1% of the time taken for a gas-phase density functional theory (DFT) on the same system. The overall XP-PCM calculations require less computational effort than that for their gas-phase counterpart due to the improved convergence of self-consistent field iterations. Therefore, the description of the high-pressure effects with our GPU accelerated XP-PCM is feasible for any molecule tractable for gas-phase DFT calculation. We have also validated the accuracy of our method on small molecules whose properties under high pressure are known from experiments or previous theoretical studies.

[a)] Electronic mail: fang.liu@emory.edu

I. INTRODUCTION

Pressure plays important roles in chemistry by inducing phase transitions of molecule crystals,[1–3] altering chemical bonds,[4,5] controlling chemical reaction rates,[6–8] and tuning photochemical reactions.[9,10] Modeling the influence of pressure in quantum chemical calculations is of great importance to reveal the mechanism of exotic phenomena under high pressure, including pressure-induced α-helix stabilization,[11–13] formation mechanism of amino acids in deep space,[14,15] and metallic behavior of hydrogen under extreme pressure.[16–18] Highly-efficient quantum chemistry simulation under high pressure will also enable virtual high-throughput screening[19] using pressure as a tuning parameter for chemical discovery.

Periodic density functional theory (DFT) and Hartree-Fock have been used to investigate high-pressure effects in materials and molecule crystals in the past few decades.[20–23] However, pressure effects on large molecules, such as proteins, can hardly be simulated with periodic DFT approach due to the high computational cost for systems with hundreds or thousands of atoms per unit cell. Classical molecular dynamics, as a low-cost alternative, lacks the description of electronic structure changes under pressure. Its accuracy in describing high-pressure phenomena is highly dependent on the force field parameterization,[24] and could lead to results contradicting experimental findings.[25–27]

The extreme pressure polarizable continuum model (XP-PCM) by Cammi and co-workers[28–30] emerges as a computationally efficient approach to incorporate pressure effects (on the order of GPa[29]) into quantum chemistry calculations on single molecules. It has been applied to the study of small to medium size molecules from single atoms,[30–32] small organic molecules,[29,33] to crystals[34,35] and fullerenes.[36] While XP-PCM DFT calculations are much more efficient than their periodic DFT counterparts,[35] their applications in quantum chemistry calculation of large molecular and biomolecular systems are still hampered by the high computational overhead for evaluating large numbers of related electron integrals. Unlike the DFT calculations in the gas phase or in standard-pressure conductor-like screening models,[37,38] XP-PCM DFT calculations involve the unique Pauli repulsion potential, whose evaluation requires $O(MN^2)$ of electric field integrals. Here, N is the number of basis functions, and M is the number of XP-PCM cavity grid points typically at $O(10N)$ or $O(100N)$. This large number of electric field integrals can result in computational overhead many times greater than the total runtime of the corresponding gas-phase DFT calculation. Because of this prohibitively high computational overhead, no XP-PCM calculation has been reported for proteins. Graphical Processing Units (GPUs) are especially suitable for parallel computing involving massive data,[39] and have been successfully applied in accelerating various types of electron integrals by numerous groups.[40–43] Speedups of more than two orders of magnitude have been observed in the GPU-accelerated quantum chemistry methods at different levels of theory, including Hartree-Fock,[40,44] density functional theory,[45] second-order Moller-Plesset perturbation theory,[46] coupled-cluster theory,[47] and multireference methods.[48,49] Here we exploit the advances of GPUs to accelerate the electric field integrals uniquely required in XP-PCM calculations and enable quantum chemistry simulation of large molecular systems under high pressure.

II. THEORY

The XP-PCM theory is an extension of conductor-like screening models (COSMO,[37] C-PCM,[38] GCOSMO,[50] and IEF-PCM[51–53]), which are introduced to describe the free energy of solvated molecules. In these models, the solute molecule is embedded in a dielectric continuum with permittivity ε, forming a cavity with unit permittivity. The solute polarizes the continuum, whose electric field is described by a set of polarization charges on the cavity surface. Then, the free energy of a solvated system in C-PCM can be expressed as

$$G^{\text{C-PCM}} = E_0 + G_{\text{pol}}, \tag{1}$$

where E_0 is the energy of the solute, and G_{pol} is the electrostatic component of the solvation free energy represented by the interaction between the polarization charges and the solute, in addition to the self-energy of the surface charges. Numerous publications have described the detailed formalism of G_{pol}[37,38,50,54,55] and algorithms for large molecular systems,[43,56–59] so we will not elaborate on them in this work.

To describe molecules at extreme pressure, Cammi and coworkers[28–30] proposed the XP-PCM method, where the free energy of the system at the given pressure p is

$$G^{\text{XP-PCM}}(p) = E_0(p) + G_{\text{pol}}(p) + G_{\text{r}}(p) + G_{\text{cav}}(p). \tag{2}$$

Compared to the free energy formula of C-PCM [Eq. (1)], the XP-PCM free energy [Eq. (2)] introduces the Pauli repulsion contribution, G_{r}, and the cavitation energy term, G_{cav}.

The cavitation energy G_{cav} is the isotherm-isobar reversible work required for the formation of a void cavity to host the molecular solute in the pure solvent at the given pressure and temperature.[33,60,61] The G_{cav} term does not contribute to the electronic Hamiltonian of the solute, and is sometimes omitted in XP-PCM implementations.[30] The XP-PCM free energy excluding G_{cav} is denoted as

$$G_{\mathrm{er}} = E_0(p) + G_{\mathrm{pol}}(p) + G_{\mathrm{r}}(p). \tag{3}$$

In this work, we only focus on G_{er} without considering G_{cav}. The Pauli repulsion term, G_{r}, describes the exchange-repulsion term of the interaction energy of the solute electrons and the solvent electrons. Evaluation of G_{r} modifies the electronic Hamiltonian and is essential for implementing XP-PCM in self-consistent field (SCF) calculations.

In the following subsections, we describe the essential equations for XP-PCM, focusing on the evaluation of G_{r}.

A. Basic Formula of Pauli Repulsion Potential

In the XP-PCM model, the electronic Schrodinger equation for the solute molecule is given by

$$\left(\hat{H}_0 + \hat{V}_{\mathrm{pol}} + \hat{V}_{\mathrm{r}}\right)\Psi = E\Psi \tag{4}$$

where $\hat{H}_0$ is the Hamiltonian of the solute molecule in vacuum, Ψ is the solute wave function, $\hat{V}_{\mathrm{pol}}$ is the electrostatic solute-solvent interaction covered in C-PCM, $\hat{V}_{\mathrm{r}}$ is the Pauli repulsion operator, and E is the energy eigenvalue.

The Pauli repulsion operator corresponds to a repulsive potential located at the boundary of the solute cavity:

$$\hat{V}_{\mathrm{r}} = \sum_{i}^{N} \int \hat{\rho}(\mathbf{r})\Gamma(\mathbf{r})d\mathbf{r} \tag{5}$$

Here, $\hat{\rho}(\mathbf{r}) = \sum_i^N \delta(\mathbf{r} - \mathbf{r}_i)$ is the electron density operator over the N electrons of the solute molecule, and the repulsion potential $\Gamma(\mathbf{r})$ is a step barrier potential at the boundary of the cavity:

$$\begin{gathered}\Gamma(\mathbf{r}) = Z\Theta_C(\mathbf{r}),\\ \Theta_C(\mathbf{r}) = \begin{cases} 0, \mathbf{r} \subseteq \mathbb{D}_C \\ 1, \mathbf{r} \nsubseteq \mathbb{D}_C \end{cases}\end{gathered} \tag{6}$$

where $\mathbb{D}_C$ denotes the domain of the physical space inside the cavity, and the height of the step barrier, Z, is determined by the extent to which the cavity is compressed and can be expressed as a function of the cavity scaling parameter f,

$$Z(f) = Z_0 \left(\frac{V_{\mathrm{c}}(f)}{V_{\mathrm{c}}(f_0)}\right)^{\frac{-(3+\eta)}{3}}. \tag{7}$$

Here, f_0 = 1.2 is the value of f at the standard condition of pressure; $V_c(f)$ is the volume of the molecular cavity obtained with the cavity scaling parameter f; η is a semi-empirical parameter that gauges how strong the Pauli repulsive barrier of the external medium is; and Z_0 is the barrier at standard pressure calculated from the following equation [adapted from Eq. (13) of Ref. 62]:

$$Z_0 = (4\pi/\xi)\rho_{\mathrm{B}} n_{\mathrm{pair}}^{\mathrm{B}}. \tag{8}$$

Here, ρ_{B} is the number density of the solvent molecule B; $n_{\mathrm{pair}}^{\mathrm{B}}$ is the number of valence electron pairs of the solvent; and $\xi = 0.7$ is the exponent of the Gaussian representation of localized orbitals.[62] In practical implementations, an empirical scaling coefficient is applied,[62] given by

$$Z_0 = 0.063\rho_{\mathrm{B}} \frac{n_{\mathrm{val}}^{\mathrm{B}}}{M_{\mathrm{B}}}, \tag{9}$$

where ρ_{B} is redefined as the density of the solvent relative to the density of water at 298 K, $n_{\mathrm{val}}^{\mathrm{B}}$ is the number of valence electrons of the solvent, and M_{B} is the molecular weight of the solvent.

In HF/DFT, $|\Psi\rangle$ is the ground state Slater determinant, so we can apply the rules for the integral of Slater determinants with one-electron operators[63] and get

$$G_r = \left\langle \Psi \left| \hat{V}_r \right| \Psi \right\rangle = \sum_{\mu\nu} \sum_{\sigma=\alpha,\beta} P_{\mu\nu}^{\sigma} \left\langle \mu \left| \Gamma(r) \right| \nu \right\rangle, \tag{10}$$

where $\mathbf{P}^{\sigma}$ is the density matrix of the solute electrons with spin σ; μ and ν are atomic basis functions. The contribution of Pauli repulsion to the Fock matrix is

$$h_{\mu\nu}^{\mathrm{r}} = \frac{\partial G_r}{\partial P_{\mu\nu}^{\sigma}} = \left\langle \mu \left| \Gamma(r) \right| \nu \right\rangle, \tag{11}$$

Combining Eqs. (6), (7) and (11), $h_{\mu\nu}^{\mathrm{r}}$ can be rewritten as[62]

$$h_{\mu\nu}^{\mathrm{r}} = Z(f)\left(S_{\mu\nu} - S_{\mu\nu}^{(\mathrm{in})}\right), \tag{12}$$

where $S_{\mu\nu}$ is an element of the overlap matrix, and

$$S_{\mu\nu}^{(\mathrm{in})} = -\frac{1}{4\pi} \oiint_{S(C)} \left(\mathbf{E}_{\mu\nu} \cdot \hat{\mathbf{n}}\right) dS \tag{13}$$

is the electric flux of the electric field ($\mathbf{E}_{\mu\nu}$) contributed by the electron density $\mu(\mathbf{r})\nu(\mathbf{r})$ through the cavity surface $S(C)$. Here $\hat{\mathbf{n}}$ is the surface normal vector. Gauss' Law is used to convert the volume integral inside the cavity [Eqs. (5)-(6)] to the surface integral in Eqs. (12)-(13). A detailed derivation is provided in Supplementary Materials Text S1.

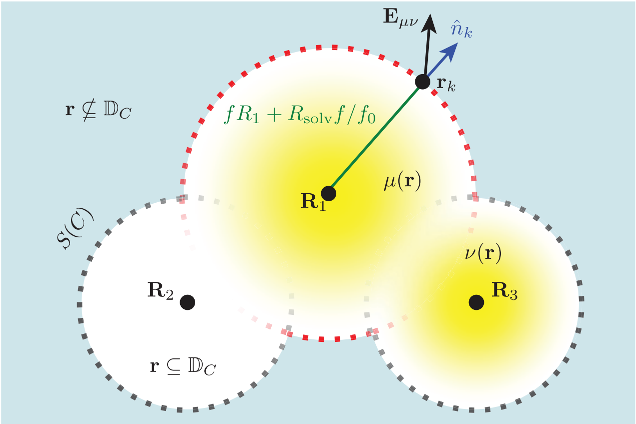

FIG. 1. Schematic presentation of the discretization of the SAS in the SWIG approach for a water molecule as an example. The atomic spheres have centers $\mathbf{R}_A$ and radii $\{f \cdot R_A^0 + R_{\mathrm{solv}} f/f_0\}$, where $\{R_A^0\}$ are the VWD radii, f is the cavity scaling factor, and R_{solv} is the solvent radius. The space inside the cavity (white background) and outside the cavity (blue background) are denoted by $\mathbf{r} \subseteq \mathbb{D}_C$ and $\mathbf{r} \nsubseteq \mathbb{D}_C$, respectively. Cavity surface grid points (located at $\{\mathbf{r}_k\}$ with norm vectors $\{\hat{\mathbf{n}}_k\}$) are presented by dots colored by the element of the center atom (red for O, gray for H). Transparency of the grid points indicates the value of the switching function (opaque for exposed points with $\mathscr{S}_k \approx 1$, transparent for partially "buried" points with $\mathscr{S}_k << 1$, $\mathscr{S}$ is defined in Eq. (16)). An example pair density $\mu(\mathbf{r})\nu(\mathbf{r})$ contributed by the atomic basis functions $\mu(\mathbf{r})$ (centered around $\mathbf{R}_1$) and $\nu(\mathbf{r})$ (centered around $\mathbf{R}_2$) is presented by yellow volume, with its electric field presented by $\mathbf{E}_{\mu\nu}$.

Therefore, in practical XP-PCM implementations, the essential task is to evaluate the Pauli repulsion matrix $\mathbf{h}^{\mathrm{r}}$ [Eq. (11)], and hence, the one-electron integrals $S_{\mu\nu}^{(\mathrm{in})}$. We will elaborate on the details of the numerical evaluation of $S_{\mu\nu}^{(\mathrm{in})}$ in Section II C after introducing the discretization scheme in Section II B.

B. Discretization of the molecular cavity surface in switching-Gaussian approach

In conductor-like screening models, the electrostatic interaction $\hat{V}_{\mathrm{pol}}$ is evaluated numerically by discretizing the cavity surface into "tesserae".[37] To evaluate the $S_{\mu\nu}^{(\mathrm{in})}$ integrals in XP-PCM, we use the readily built solvent accessible surface (SAS) discretized by the switching-Gaussian (SWIG) approach,[54] to be consistent with our implementation of $\hat{V}_{\mathrm{pol}}$ (Figure 1). In this discretization scheme, the molecular surface is formed from inter-locking Van der Waals (VDW) spheres centered around the composing atoms, and the surface of each sphere is discretized by Lebedev grid points[54]. Geometry of the solvent molecule is represented by expanding the VDW radii with a scaling factor f, and adding an optional effective solvent radius R_{solv}. The radius of each VDW sphere, R_J is then expressed as

$$R_J = fR_J^0 + R_{\text{solv}} f/f_0 \tag{14}$$

where R_J^0 is the Bondi radius[64,65] of atom J, and R_{solv} is the solvent radius, and the scaling factor f/f_0 for R_{solv} is adapted from the scaling of solvent radius in the solvent exclusion surface (SES).[29]

To avoid singularities in the evaluation of $\hat{V}_{\text{pol}}$, surface polarization charges are presented as spherical Gaussian functions centered at the grid point, and the Gaussian exponent for the kth point charge belonging to the Ith nucleus is given as

$$\zeta_k = \frac{\zeta}{R_I \sqrt{w_k}} \tag{15}$$

where ζ is an optimized exponent for the specific Lebedev quadrature level being used (as tabulated by York and Karplus),[54] w_k is the Lebedev quadrature[66] weight for the kth point, and R_I is the atomic radius of the Ith nucleus that the kth tessera belongs to.

To get a smooth change of cavity surface area during geometry optimization, the switching function is introduced to indicate how much a grid point is either buried inside the molecular surface or exposed and accessible to solvent molecules. For the improved SWIG (ISWIG) scheme[67] used by default in our implementation, the switching function is defined as

$$\mathscr{S}_k = \prod_{J, k \notin J}^{\text{atoms}} S_{\text{wf}}(\mathbf{r}_k, \mathbf{R}_J) \tag{16}$$

$$\begin{aligned} S_{\text{wf}}(\mathbf{r}_k, \mathbf{R}_J) = 1 - \frac{1}{2}\{&\text{erf}[\zeta_k(R_J - |\mathbf{r}_k - \mathbf{R}_J|)] \\ &+ \text{erf}[\zeta_k(R_J + |\mathbf{r}_k - \mathbf{R}_J|)]\} \end{aligned} \tag{17}$$

where $\mathbf{r}_k$ is the location of the kth Lebedev point, $\mathbf{R}_J$ is the location of the Jth nucleus with atomic radius R_J, and erf is the Gauss error function. The area of the kth tessara can then be calculated as

$$a_k = w_k R_I^2 \mathscr{S}_k. \tag{18}$$

C. Numerical Evaluation of Pauli Repulsion Integrals

With the discretization scheme of section II B (also see Figure 1), the $S_{\mu\nu}^{(\text{in})}$ integral [Eq. (13)] can be rewritten as

$$S_{\mu\nu}^{(\text{in})} = -\frac{1}{4\pi}\sum_{k}^{M}\left(\mathbf{E}_{\mu\nu}^{k}\cdot\hat{\mathbf{n}}_{k}\right)a_{k}. \tag{19}$$

Here, $\hat{\mathbf{n}}_k$ is the norm vector of the kth tessera pointing outward from the cavity, a_k is the area of the kth tessera [Eq (18)], and $\mathbf{E}_{\mu\nu}^{k}$ is the electric field caused by the electron distribution of basis function pair $\mu(\mathbf{r})\nu(\mathbf{r})$ at the kth tessera center

$$\mathbf{E}_{\mu\nu}^{k} = -\int \mu(\mathbf{r})\nu(\mathbf{r})\frac{\mathbf{r}_k-\mathbf{r}}{|\mathbf{r}_k-\mathbf{r}|^3}d\mathbf{r}, \tag{20}$$

where $\mathbf{r}_k$ is the position of the kth Lebedev grid point (Figure 1). It is worth noting that this type of electric field integral does not exist in C-PCM, where only the electric potential integrals are evaluated to obtain $\hat{V}_{\text{pol}}$. As usual, the atom-centered basis functions are contractions over a set of primitive atom-centered Gaussian functions

$$\mu(\mathbf{r}) = \sum_{i=1}^{l_\mu} c_{\mu i}\chi_i(\mathbf{r}) \tag{21}$$

Thus, the one-electron integral $S_{\mu\nu}^{(\text{in})}$ [Eq. (19)] can be expressed as

$$\begin{aligned} S_{\mu\nu}^{(\text{in})} &= \frac{1}{4\pi}\sum_{k}^{M}\left(\mu(\mathbf{r})\mu(\mathbf{r})\left|\frac{(\mathbf{r}_k-\mathbf{r})\cdot\hat{\mathbf{n}}_{\mathbf{k}}}{|\mathbf{r}_k-\mathbf{r}|^3}\right.\right) \\ &= \frac{1}{4\pi}\sum_{i=1}^{l_\mu}\sum_{j=1}^{l_\nu}\sum_{k}^{M}c_{\mu i}c_{\nu j}\left[\chi_i(\mathbf{r})\chi_j(\mathbf{r})\left|\frac{(\mathbf{r}_k-\mathbf{r})\cdot\hat{\mathbf{n}}_{\mathbf{k}}}{|\mathbf{r}_k-\mathbf{r}|^3}\right.\right] \end{aligned} \tag{22}$$

where we use brackets to denote one-electron integrals over primitive basis functions and parentheses to denote such integrals for contracted basis functions. In the following, we use the indices μ,ν for contracted basis functions, and the indices i, j for primitive Gaussian basis functions. We discuss the GPU algorithm for evaluating $S_{\mu\nu}^{(\text{in})}$ in Section III.

D. Numerical Calculation of Pressure

We use the numerical fitting approach proposed by Cammi and coworkers[33] to calculate the pressure p associated with each value of the cavity scaling factor f. For a given molecule with fixed structure, multiple XP-PCM calculations with different f values are performed to obtain a series of G_{er} values. These G_{er} values are fitted as a nonlinear function of the associated cavity volumes, V_{c}, based on the following expression[68]

$$G_{\mathrm{er}}(V_{\mathrm{c}}) = G_{\mathrm{er}}(V_{\mathrm{c}}^{0}) + aV_{\mathrm{c}}\left[\frac{1}{b-1}\left(\frac{V_{\mathrm{c}}^{0}}{V_{\mathrm{c}}}\right)^{b} + 1\right] + cV_{\mathrm{c}} \tag{23}$$

where a,b, and c are fitting parameters. Then the pressure p can be computed by differentiation (derivation available in Supplementary Materials Text S2)

$$p(f) = -\frac{\partial G_{\mathrm{er}}(V_{\mathrm{c}})}{\partial V_{\mathrm{c}}} = a\left[\left(\frac{V_{\mathrm{c}}^{0}}{V_{\mathrm{c}}}\right)^{b} - 1\right] - c. \tag{24}$$

III. IMPLEMENTATION ON GPUS

Unlike the $\hat{V}_{\mathrm{pol}}$ related integrals that need to be re-evaluated in each SCF iteration, $S_{\mu\nu}^{(\mathrm{in})}$ is evaluated only once before the SCF calculation starts and is directly added to the core Hamiltonian. However, this does not mean that the evaluation of $S_{\mu\nu}^{(\mathrm{in})}$ is computationally trivial. In Section V A, we will demonstrate that XP-PCM usually requires a significantly denser grid than regular C-PCM to ensure numerical integration accuracy, which increases computational cost significantly if no acceleration strategy is applied.

Building $S_{\mu\nu}^{(\mathrm{in})}$ requires one-electron integral evaluations and involves a significant amount of data parallelism, making it well suited for GPU acceleration. We elaborate the GPU-based acceleration strategies in the following subsections.

A. Fine-grained parallelism

Analogous to our GPU-based implementation of $\hat{V}_{\mathrm{pol}}$ related integrals in C-PCM,[43,56] we wrote six separate GPU kernels for evaluating $S_{\mu\nu}^{(\mathrm{in})}$ of the following angular momentum classes: *ss*, *sp*, *sd*, *pp*, *pd*, and *dd*. Each individual GPU thread calculates integrals corresponding to a batch of primitive pairs sharing the same set of pair quantities, similar to the One Thread ↔ One Batch mapping[42]originally proposed for the evaluation of Coulomb integrals. For instance, in the *ss* kernel, each GPU thread calculates a single integral, $[\chi_{s_1}\chi_{s_2}]$, in each loop, whereas in the *sp* kernel, each GPU thread calculates 3 primitive pairs, $[\chi_s\chi_p^x]$, $[\chi_s\chi_p^y]$, and $[\chi_s\chi_p^z]$.

The algorithm for evaluating $S_{\mu\nu}^{(\mathrm{in})}$ for *sp* pairs is shown schematically in Figure 2 for a system with one *s* shell and two *p* shells and a GPU block size of 1× 6 threads. The *s* shell contains

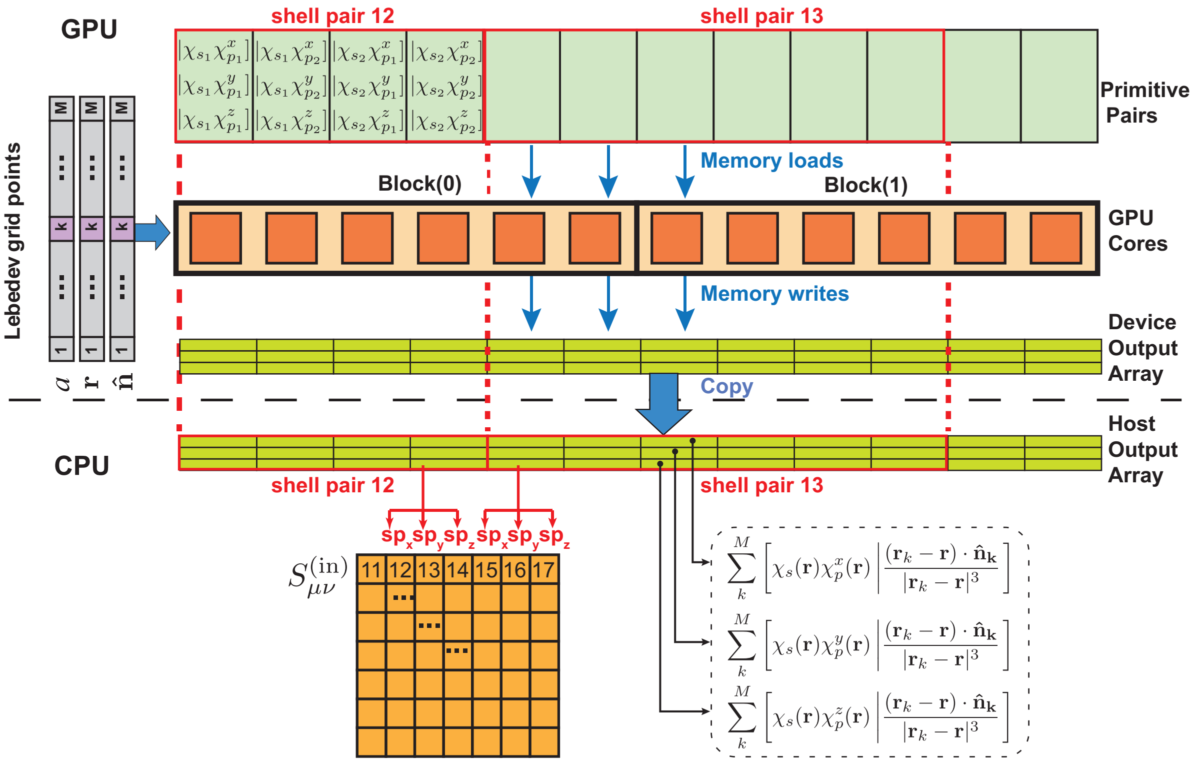

FIG. 2. Algorithm for calculating $\mathbf{S}^{\rm (in)}$ for *sp* integrals of a system composed of one *s* shell and two *p* shells (the *s* shell contains 2 primitive Gaussian function each; the first and second *p* shells have 2 and 3 primitive Gaussian functions, respectively). On top of the graph, the pale green array represents primitive pairs belonging to *sp* shell pairs. The GPU cores are represented by orange squares (threads) embedded in pale yellow rectangles (one-dimensional blocks with 6 threads/block). The 1 × 6 block is used for illustrative purposes only, and a 1 × 128 block is used in actual implementation. The output is a $3 \times N_{\rm threads}$ array where each GPU thread generates 3 integrals for primitive pairs $[\chi_s\chi_p^x]$, $[\chi_s\chi_p^y]$, and $[\chi_s\chi_p^z]$. Primitive pair integrals are finally added to the Fock matrix entry of the corresponding contracted function pair. All red lines and text indicate contracted Gaussian integrals. Blue arrows and text indicate memory operations.

2 primitive Gaussian functions; the first and second *p* shells have 2 and 3 primitive Gaussian functions, respectively. A block of size 1 × 6 is used for illustrative purposes. In practice, a 1 × 128 block is used for optimal occupancy and memory coalescence. Primitive pairs, $\chi_i\chi_j$, that make negligible contributions are not calculated, and these are determined by using a Schwartz-like bound[69] with a cutoff, $\varepsilon^{\rm screen} = 10^{-14}$ atomic units

$$[ij|_{\rm Schwartz} = \left[\chi_i\chi_j \,|\, \chi_i\chi_j\right]^{1/2} < \varepsilon^{\rm screen}. \tag{25}$$

Here we use a tighter threshold than the default 10^{-12} threshold for $\hat{V}_{\text{pol}}$[43] because G_{r} usually has a smaller magnitude than G_{pol} and thus is more sensitive to the integral threshold.

The surviving pair quantities are preloaded to the GPU global memory, and each GPU thread fetches a batch of 3 *sp* primitive pairs sharing the same set of pair quantities at the beginning of the integral kernel. Quantities related to each Lebedev grid point (area a_k, coordinates $\mathbf{r}_k$, and norm vector $\hat{n}_k$ pointing towards outside of cavity) are also preloaded in global memory. Each GPU thread loops over all Lebedev grid points to accumulate the electric flux of the electric field contributed by its primitive pair $[\chi_i\chi_j]$ through all tesserae of the cavity surface as follows

$$S_{ij}^{(\text{in})} = \frac{1}{4\pi}\sum_{k}^{M}\left[\chi_i(\mathbf{r})\chi_j(\mathbf{r})\left|\frac{(\mathbf{r}_k-\mathbf{r})\cdot\hat{\mathbf{n}}_{\mathbf{k}}}{|\mathbf{r}_k-\mathbf{r}|^3}\right.\right] \tag{26}$$

It worth noting that for the GPU kernel shown in Figure 2, which evaluates the *sp* angular momentum class, three integrals in the form of Eq. (26) are evaluated by each GPU thread for the primitive pairs $[\chi_s\chi_p^x]$, $[\chi_s\chi_p^y]$, and $[\chi_s\chi_p^z]$. Evaluation of the primitive integral of Eq. (26) is discussed in subsection III B.

The result is stored to an output array in GPU global memory, which is later copied to the CPU memory after the accumulation of $S_{ij}^{(\text{in})}$ is done. The last step is to form $S_{\mu\nu}^{(\text{in})}$

$$S_{\mu\nu}^{(in)} = \sum_{i=1}^{l_\mu}\sum_{j=1}^{l_\nu} c_{\mu i}c_{\nu j}S_{ij}^{(\text{in})} \tag{27}$$

on the CPU by adding each entry of the output array (primitive pair) to its corresponding atomic orbital pair entry.

All algorithms discussed above can be easily generalized to other angular momentum classes other than *sp*. The numbers of primitive pairs evaluated by each GPU thread in momentum classes *ss*, *sp*, *sd*, *pp*, *pd*, and *dd* are 1, 3, 6, 9, 18, and 36, respectively, since our implementation uses Cartesian format basis function and each *d* orbital has 6 components. These kernels are launched sequentially.

B. Evaluation of primitive integrals

The primitive electric field integral of Eq. (26) for different angular momentum classes are evaluated analytically based on the algorithm of McMurchie and Davidson.[70] For a pair of Cartesian

Gaussian basis functions

$$\chi_i = x_A^n y_A^l z_A^m \exp(-\alpha_A r_A^2) \tag{28}$$

$$\chi_j = x_B^{\bar{n}} y_B^{\bar{l}} z_B^{\bar{m}} \exp(-\alpha_B r_B^2) \tag{29}$$

centered at $\mathbf{A} = (x_A, y_A, z_A)$ and $\mathbf{B} = (x_B, y_B, z_B)$, the product $\chi_i \chi_j$ can be expanded as combinations of Hermite polynomial Gaussians

$$\begin{aligned}[\chi_i \chi_j| &= \sum_N^{n+\bar{n}} \sum_M^{m+\bar{m}} \sum_L^{l+\bar{l}} D_{NLM} \Lambda_N(x_P) \Lambda_L(y_P) \Lambda_M(z_P) \exp(-\alpha_P r_P^2) \\ &= \sum_{NLM} D_{NLM} [NLM| \end{aligned} \tag{30}$$

Here, the Hermite polynomial Gaussian Λ_j is related to the Hermite polynomial H_j by $\Gamma_j(x_P; \alpha_P) = \alpha_P^{j/2} H_j(\alpha_P^{1/2} x_P)$; $\mathbf{P} = (x_p, y_p, z_p)$ is the center of the product Gaussian function formed from the overlap of the two Gaussian functions in Eq (28) and (29), and α_P is the Gaussian exponent of the product Gaussian; indices (N, L, M) run over all possible combinations within the appropriate range.

The electric field integral of $\chi_i \chi_j$ can then be calculated as combinations of the electric field integral of Hermite polynomial Gaussians (N, M, L):

$$\begin{aligned} S_{ij}^{(\text{in})} = \frac{1}{4\pi} \sum_k^M \sum_{NLM} D_{NLM} \Big(&\left[NLM \left| x_k r_k^{-3} \right.\right] \hat{n}_{k,x} \\ &+ \left[NLM \left| y_k r_k^{-3} \right.\right] \hat{n}_{k,y} + \left[NLM \left| z_k r_k^{-3} \right.\right] \hat{n}_{k,z} \Big) \end{aligned} \tag{31}$$

where $\hat{n}_{k,x}$, $\hat{n}_{k,y}$, and $\hat{n}_{k,z}$ are the components of $\hat{\mathbf{n}}_k$ in x, y, and z directions, and the integrals of $[NML|$ are given by

$$\begin{aligned} \left[NLM | x_k r_k^{-3}\right] &= -(2\pi/\alpha_P) R_{N+1,L,M} \\ \left[NLM | y_k r_k^{-3}\right] &= -(2\pi/\alpha_P) R_{N,L+1,M}, \\ \left[NLM | z_k r_k^{-3}\right] &= -(2\pi/\alpha_P) R_{N,L,M+1} \end{aligned} \tag{32}$$

Here, the auxiliary functions R_{NLM} can be calculated from recursive relations to tabulated Boys functions.[71]

We wrote separate GPU kernel functions for evaluating the primitive integrals of each momentum class. Each kernel function generates expansions of primitive pairs [Eq. (30)] based on recursive relations of D_{NLM}, and then evaluates the electric field integrals of Eq. (32) in terms of Boys functions.

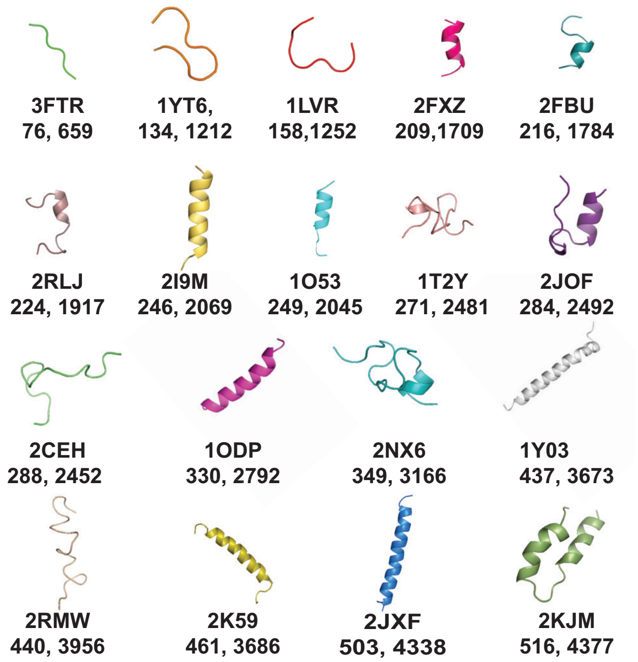

FIG. 3. Structures for the benchmark proteins for XP-PCM performance test. For each protein, the PDB ID, number of atoms, and the number of orbitals with 6-31G* basis set are listed.

IV. COMPUTATIONAL DETAILS

We have implemented a GPU-accelerated XP-PCM formulation in a development version of the TERACHEM[72,73] package. All XP-PCM calculations use parameters stated as follows unless otherwise specified. An ISWIG screening threshold of 10^{-8} is used, meaning that molecular surface (MS) points with a switching function value less than this threshold are ignored. The Pauli repulsion gauge parameter η is set to 6 as recommended by literature.[30] This choice of η is known to give a dependence of the computed pressure on the cavity volume V_c in reasonable agreement with the dependence of the experimental pressure on the molar volume in molecular solids.[33]

To study the computational performance of our GPU implementation on large biomolecules, we select a test set of 18 experimental protein structures[74] obtained with aqueous solution NMR where inclusion of a solvent environment was essential to find optimized structures in good agreement with experimental results (Figure 3). The proteins range in size from around 70 to 500 atoms, and their detailed properties (PDB ID, number of residues and atoms, charge, secondary structure) are summarized in the Supplementary Materials Table S1. For these test molecules, we conduct a number of XP-PCM single-point energy evaluations at B3LYP[75–77]/6-31G*[78] level of theory, with the cavity scaling factor f values ranging from 0.85 to 1.20 (pressure values on the order 100 GPa to 1 GPa, see Supplementary Materials Figure S1 and S2). The environment dielectric constant corresponds to aqueous solvation (ε = 78.39, $\rho_B = 1.0$, $n_{\mathrm{val}}^{\mathrm{B}} = 8$, M_B=18). The default revised Bondi radii[65] are used ($R_{\mathrm{H}}^0 = 1.1$ Å, $R_{\mathrm{C}}^0 = 1.7$ Å, $R_{\mathrm{N}}^0 = 1.55$ Å, $R_{\mathrm{O}}^0 = 1.52$ Å, and $R_{\mathrm{S}}^0 = 1.8$ Å). Similar calculations are carried out in the gas phase and in C-PCM implicit solvent to test the computational cost of adding in the XP-PCM high pressure environment. All timings have been obtained using a single core of the Intel Xeon Gold 6248 "Cascade Lake" CPU clocked at 2.50 GHz and one NVIDIA Tesla V100 GPU.

To validate our XP-PCM implementation, we carry out XP-PCM calculations on an argon (Ar) atom and an acetylene molecule. To simplify the comparison with previous XP-PCM works by Cammi *et al.* on these systems,[30] we have used the same simulation parameters wherever possible. Therefore, we fix the geometry at the equilibrium geometry optimized in the gas phase without XP-PCM. Due to the lack of f-type basis functions in TERACHEM, we are not able to use the aug-cc-pVTZ basis set in Cammi's work,[30] and the single point calculations are carried out at B3LYP/aug-cc-pVDZ level of theory. The environment dielectric constant corresponds to cyclohexane ($\varepsilon = 2.0165$, $\rho_B = 0.779$, $n_{\mathrm{val}}^{\mathrm{B}} = 36$, M_B=84.16). We use the same set of Bondi atomic radii ($R_{\mathrm{H}}^0 = 1.2$ Å, $R_{\mathrm{C}}^0 = 1.7$ Å, and $R_{\mathrm{Ar}}^0 = 1.88$ Å) as Cammi's work[30] to facilitate comparison of results. The hydrogen atom radius is slightly different from the default revised Bondi radius[65] used in TERACHEM ($R_{\mathrm{H}}^0 = 1.1$). The cavity uses an ISWIG[67] discretization density of 1202 Lebedev points/atom and cavity radii that are varied by applying a scaling factor f on the Bondi radii.[64]

V. RESULTS AND DISCUSSION

The GPU implementation of XP-PCM could be a computationally efficient approach to investigate the electronic structure of molecular and biomolecular systems. In this section, we investigate the efficiency and accuracy of our XP-PCM implementations. We first look for the appropriate discretization level to obtain numerically converged XP-PCM results for a small protein. With the optimal discretization parameter, we benchmark the performance of XP-PCM calculation of a set of proteins varying in sizes to estimate the time scaling and the applicability of the this to large biomolecules. Then we compare the computational performance of the XP-PCM, C-PCM, and the gas phase counterpart to evaluate the extra computational cost for describing the pressure and solvent effects. Finally, we assess the quality of our implementation for describing high pressure effects by applying the method on an Argon atom and an acetylene molecule.

A. Convergence with respect to discretization level

In this subsection, we look for the optimal XP-PCM cavity discretization level that balances the numerical accuracy and computational cost.

We examine the convergence of XP-PCM calculation with respect to the discretization level of the cavity surface, which is defined as the density of Lebedev grid points per sphere (Figure 4). For the tested peptide (PDB ID: 3FTR), both the Pauli repulsion energy (G_r) and the electrostatic solvation free energy (G_pol) values vary with the discretization levels, and hence determine the convergence behavior of the total free energy (G_er). At low, medium, high, and very high discretization levels (26-50, 110-302, 434-770, and 974-1202 points/atom), the relative error in G_pol is typically less than %5, %1, %0.3, and %0.04, whereas the relative error for G_r rapidly changes from %111 to %44, %7, and %2 (Figure 4 and Table I). Since G_pol and G_r have similar magnitudes and opposite signs, it is essential to reach sufficient numerical accuracy for both terms to ensure the accuracy of the total free energy.

Ideally, a very-high discretization level ($\geqslant$974 points/atom) is preferred, but the total runtime is twice the runtime of the medium discretization level (Figure 4). As the grid density increases, the number of effective (non-buried) Lebedev points increases linearly, and the number of XP-PCM related primitive integrals (for both $\mathbf{h}_\mathrm{r}$ and $\hat{V}_{pol}$) increases linearly (Supplementary Materials Table S2). Since $\hat{V}_{pol}$ is re-evaluated for each SCF iteration, the increase in computational cost with grid density is magnified, resulting in very high computational overhead (>50% of total runtime) at very high discretization level (Supplementary Materials Table S3). To balance accuracy and efficiency, a high discretization level of 434 points/atom is chosen for performance tests in sections V B and V C unless otherwise specified.

It is worth noting that this choice of discretization level for XP-PCM is significantly higher than the recommended discretization level for C-PCM,[54] which only has free energy contribution from G_pol. As shown in our test for 3FTR (Figure 4) as well as previous works in this field,[54,67] sufficient accuracy for G_pol (error $\approx$ 1 kcal/mol) can already be reached at a medium discretization level (ca. 110 points/atom). Apart from the convergence of free energies, the convergence of cavity volume is also crucial for XP-PCM because the volume is explicitly needed for deriving the pressure [Eq. (5)]. Our choice of 434 points/atom has less than 0.2% error in cavity volume for the tested

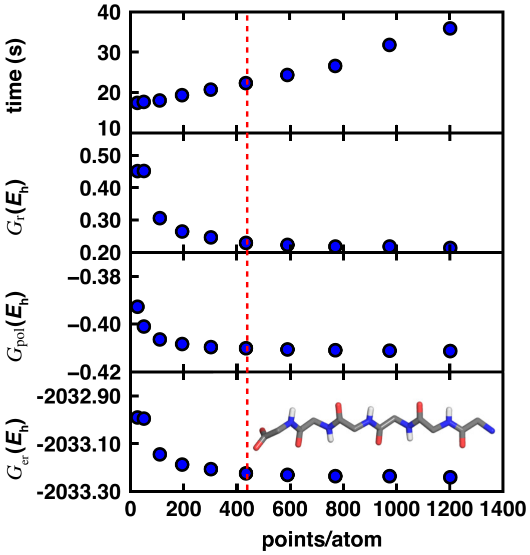

FIG. 4. Convergence of the free energies (G_r, G_{pol}, G_{er} in E_h) and runtime (in sec) with respect to discretization level (points/atom) for the XP-PCM B3LYP/6-31G* calculation of a peptide (PDB ID: 3FTR). A constant cavity scaling factor $f = 1.0$ is used for all calculations. The main chain of 3FTR is shown in the inset, with carbon, nitrogen, oxygen, and hydrogen colored grey, blue, red, and white, respectively. All XP-PCM calculations are conducted with TERACHEM using a single core of Intel Xeon Gold 6248 "Cascade Lake" CPU clocked at 2.50 GHz and one NVIDIA Tesla V100 GPU.

system, significantly lower than the 2% error at low discretization level (Supplementary Materials Table S2). Because of the higher discretization level required in XP-PCM, a GPU accelerated implementation is even more critical for XP-PCM than C-PCM to ensure its applicability to large molecules where the electric field integral evaluations may dominate the total runtime.

B. Performance for large molecules

Two primary concerns about applying XP-PCM to large molecules are the time scaling of the algorithm and the efficiency compared to its gas phase, or normal-pressure solution-phase counterparts. To test these, we collected the timings of XP-PCM calculation of a set of proteins at different pressures presented by different cavity scaling factors f (Figure 5). The observed empirical scaling of the evaluation of Pauli-Repulsion matrix $\mathbf{h}_r$ unique to XP-PCM is $O(N^{1.9})$ regardless of cavity scaling factor f. Here, the normal ($f = 1.2$), moderately compressed ($f = 1.0$), and highly compressed ($f = 0.85$) cavities correspond to pressure values on the order of 1 GPa,

TABLE I. Free energies (G_r and G_pol, in E_h) obtained at different discretization levels compared to the results obtained with the highest grid density (1202 points/atom) for the XP-PCM B3LYP/6-31G* calculation with cavity scaling factor $f = 1.0$ of a protein fibril (PDB ID: 3FTR).

discretization level	grid density (pts./atom)	$G_\mathrm{r}(E_\mathrm{h})$	$G_\mathrm{pol}(E_\mathrm{h})$	relative error	
				G_r	G_pol
low	26	0.451286	-0.392746	111.32%	4.51%
	50	0.451692	-0.401015	111.51%	2.50%
medium	110	0.305759	-0.406418	43.18%	1.18%
	194	0.264476	-0.408335	23.84%	0.72%
	302	0.246270	-0.409623	15.32%	0.40%
high	434	0.229174	-0.410099	7.31%	0.29%
	590	0.223057	-0.410603	4.45%	0.16%
	770	0.218007	-0.410924	2.08%	0.09%
very high	974	0.218098	-0.411119	2.13%	0.04%
	1202	0.213555	-0.411281	0.00%	0.00%

10 GPa, and 100 GPa, respectively (Supplementary Materials Figure S1 and S2). However, the prefactor of the scaling increases as f decreases, meaning that the XP-PCM calculations at higher pressure have higher computational costs for evaluating $\mathbf{h}_\mathrm{r}$. This is a natural result of the fact that the number of effective molecule surface grid points (not "buried" in the cavity) roughly increases linearly as f increases (Supplementary Materials Figure S3). As the scaling factor f decreases, the radii of the atom-centered spheres decreases, and there is less overlap between the spheres and more exposure of grid points.

The total runtime of XP-PCM follows a similar trend. The runtime has O(N$^{2.0}$) scaling at all f values, but larger prefactors are observed at lower f values (Figure 5). Here the total runtime includes the evaluation of $\mathbf{h}_\mathrm{r}$, the electrostatic solvent effect term $\hat{V}_\mathrm{pol}$, and other terms in regular gas phase SCF. It is worth noting that the $\hat{V}_\mathrm{pol}$ evaluation also needs more computational cost at lower f values because of the increase in grid points. For the largest protein (PDB ID: 2KJM) in the benchmark set, the time for evaluating $\hat{V}_\mathrm{pol}$ doubles as the cavity is compressed from $f = 1.2$ to $f = 0.8$, while the time for evaluating $\mathbf{h}_\mathrm{r}$ also increases by 1.75 times (Supplementary Materials Figure S4-S5).

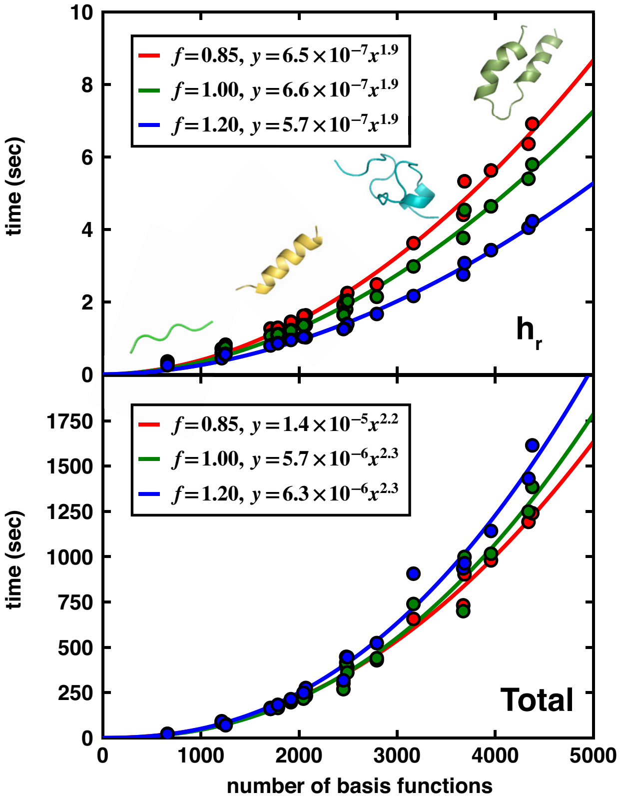

FIG. 5. Timings (in sec) for the XP-PCM B3LYP/6-31G* single point calculations for a set of 18 benchmark proteins in normal ($f = 1.2$, p at ca. 1 GPa), moderately compressed ($f = 1.0$, p at ca. 10 GPa), and highly compressed ($f = 0.85$, p at ca. 100 GPa) cavities. An ISWIG discretization scheme is used with 434 Lebedev points/atom. Timing data are presented with dots, with curves with respective colors showing the empirical scaling fitted by power function. Timings for the evaluation of the Pauli repulsion integral ($\mathbf{h}_\mathrm{r}$) (upper) and for the total single point calculation (lower), with representative proteins of different sizes shown in the inset structure. All XP-PCM calculations are conducted with TERACHEM using a single core of Intel Xeon Gold 6248 “Cascade Lake” CPU clocked at 2.50 GHz and one NVIDIA Tesla V100 GPU.

In summary, our XP-PCM implementation demonstrates a sub-quadratic scaling based on tests of molecules with up to 5000 basis functions, which is similar to the performance of our GPU-accelerated implementation of C-PCM.

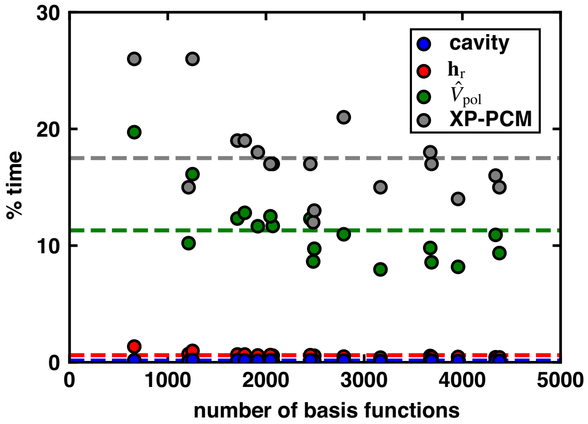

FIG. 6. Percentage of the runtime spent on different components of XP-PCM calculations of 18 small proteins using B3LYP/6-31G*. The constructions of molecular cavity, Pauli repulsion matrix $\mathbf{h}_\mathrm{r}$, and solvation electrostatic interactions $\hat{V}_\mathrm{pol}$, and their sum are presented by blue, red, and green dots, respectively. The sum of these XP-PCM-specific components is denoted with grey dots. For each component, the average value over the 18 protein set is indicated by a dotted line with the corresponding color. All XP-PCM calculations are conducted with TERACHEM using a single core of Intel Xeon Gold 6248 "Cascade Lake" CPU clocked at 2.50 GHz and one NVIDIA Tesla V100 GPU.

C. Performance comparison to gas phase

To obtain a comprehensive comparison between XP-PCM and its gas phase or normal-pressure solution-phase counterparts, we decompose the timings for XP-PCM at $f = 1.0$ (moderately compressed, p at about 10 GPa) into different contributions (Figure 6). The total runtime of XP-PCM can be partitioned into the XP-PCM-specific part, and other parts in common with gas phase SCF. The former includes three major components: building the cavity surface, constructing the Pauli-Repulsion matrix $\mathbf{h}_\mathrm{r}$, and evaluating electrostatic solvent effects related terms $\hat{V}_\mathrm{pol}$. Among these terms, $\hat{V}_\mathrm{pol}$ takes the majority of the time because the related integrals depend on solute electron density and are re-evaluated in each SCF iteration. The molecular cavity and the $\mathbf{h}_\mathrm{r}$ matrix only need to be built once before SCF starts and take less than 0.2% and 1.4% of the time for any tested protein. In total, the percentage of time taken by the XP-PCM specific terms fluctuate between 12%-26% with an average value of 17%. Based on these timings, and assuming that SCF converges similarly in different environments, we expect that the runtime of an XP-PCM calculation is about 1.2X of the gas phase runtime for the same system.

However, direct comparison of the runtime for XP-PCM, C-PCM, and gas phase calculations

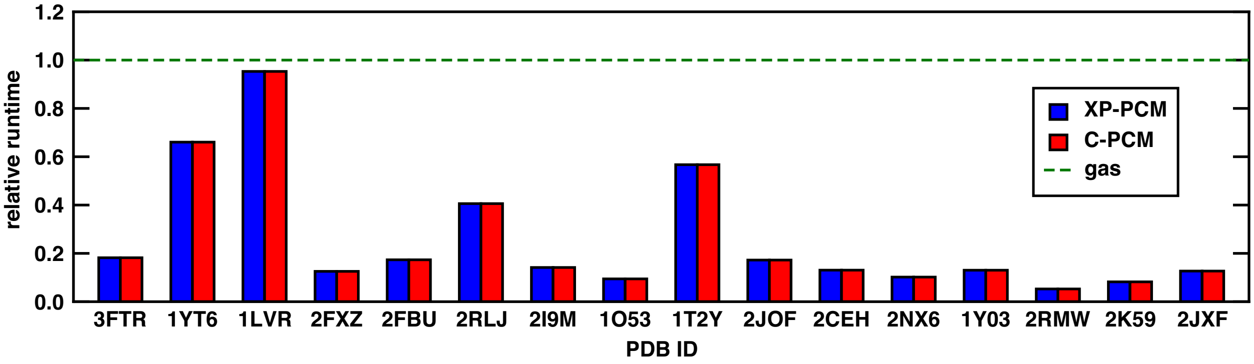

FIG. 7. Relative runtime (in seconds) taken by XP-PCM (f=1.0) and C-PCM B3LYP/6-31G* single point energy evaluations compared to their gas phase counterpart for benchmark proteins ordered by increasing size. Two proteins (PDB ID: 1ODP and 2KJM) in the benchmark set are not included because of unconverged gas phase calculations. An ISWIG discretization scheme is used with 434 Lebedev points/atom. Runtime for XP-PCM and C-PCM are visually the same, because the extra time for building $\mathbf{h}_r$ in XP-PCM is negligible due to our efficient implementation. Timings were obtained with TERACHEM single core of Intel Xeon Gold 6248 "Cascade Lake" CPU clocked at 2.50 GHz and one NVIDIA Tesla V100 GPU.

shows that XP-PCM and C-PCM require less runtime than the gas phase counterpart for all tested proteins (Figure 7).

This is caused by the fact that DFT calculations of large molecules like proteins tend to converge much faster in XP-PCM than in the gas phase, as was observed for the comparison between C-PCM and gas phase in our previous works.[43,56] The number of SCF iterations taken by the gas phase calculation is 1.4X to 22.8X of that taken by XP-PCM (Supplementary Materials Table S4). Two proteins (PDB IDs: 1ODP and 2KJM) that failed convergence in 2000 iterations in the gas also successfully converged in XP-PCM within 17 steps. Hence, we expect that in practical applications, XP-PCM is computationally feasible to any large molecular system that is computationally tractable in the gas phase.

D. Compressed argon atom

To assess the accuracy of our implementation for describing high-pressure effects, we compare our XP-PCM calculation of a compressed argon atom with previous theoretical and experimental results. The free energy G_{er} at different cavity volumes (V_{c}) is obtained with XP-PCM calculations at different cavity scaling factor values (f = 0.85 – 1.20, Figure 8 and Supplementary Materials

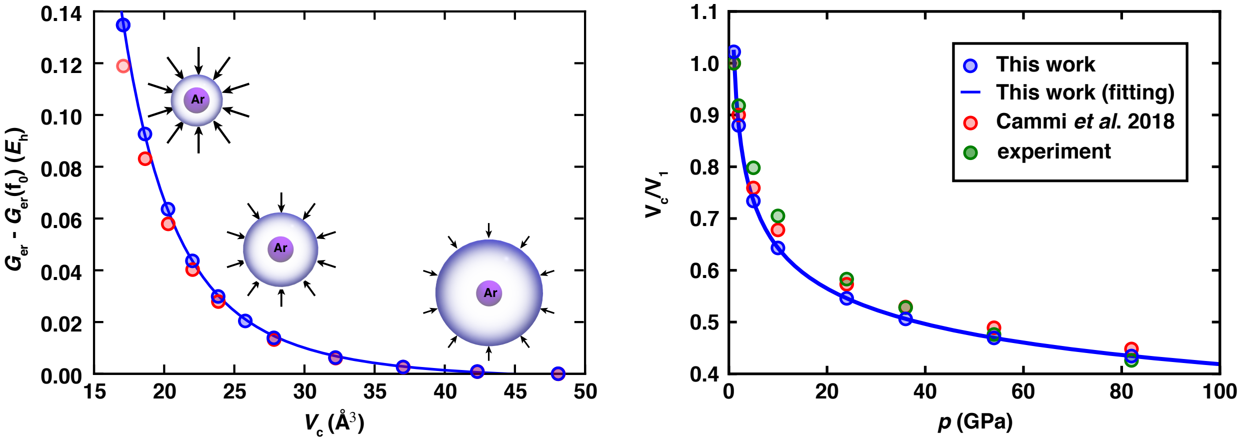

FIG. 8. (left) XP-PCM free energy of a compressed argon atom in cyclohexane solvent as a function of cavity volume (cavity scaling factor f = 0.85 – 1.20). G_{er} is calculated relative to its value obtained with f_0=1.2. (right) Comparison of the volume compression of argon (V_{c}/V_1) as a function of pressure obtained with XP-PCM calculation (this work and Cammi *et al.* work[30]) and experiments.[79] V_1 is the reference volume of the cavity corresponding to the pressure p = 1.1 GPa. Schematic illustration of the argon atom in a compressed cavity is shown in the insets. Length of arrows indicates the strength of pressure.

Table S5). To obtain the corresponding pressures, we used Equation (23) to fit G_{er} as a function of V_{c} (Figure 8 and Supplementary Materials Table S5). The fitting parameters for compressed argon are $a = 1.3048 \times 10^{-4}$ E_{h}/Å^3, b = 5.4057, and c = -1.7593 $\times 10^{-4}$ E_{h}/Å^3. Using these parameters with Equation (24), we determined the pressure as a function of the cavity volume (Figure 8 and Supplementary Materials Table S6). We see excellent agreement with Cammi's previous XP-PCM work for the free energy and the pressure functions, due to the usage of a slightly different basis set (see details in Section IV). We have also compared the volume compression as a function of pressure to the experimental values for solid argon compression and found very good agreement (Figure 8).

E. Compressed acetylene molecule

We further assessed the XP-PCM description of pressure effects in an acetylene molecule, where the cavity has an irregular shape in contrast to argon's spherical cavity. With a similar numerical fitting approach, we obtained the pressure as a function of the cavity volume (Figure 9 and Supplementary Materials Table S7). The fitting parameters for compressed acetylene are a = 1.5457 $\times 10^{-4}$ E_h/Å^3, b = 5.9367, and c = -1.8521$\times 10^{-4}$ E_h/Å^3. Although there are some differences in the free energy values, the $p - V_\mathrm{c}$ curve gives a good agreement with Cammi's results.

It is not surprising that the energies are not the same, as we have noticed some differences in the cavity definition. Although we obtained the same cavity volumes as Cammi's results for the argon atom, we noticed that our cavity volumes for acetylene are systematically lower at all f values, even when the same set of atomic radii are used (Supplementary Materials Table S8). The difference is likely caused by the fact that we use the switching Gaussian approach to smooth the cavity surface, which can influence the calculation of V_c. To facilitate direct comparison with Cammi's work, we added a small solvent radius (0.135 Å) to our cavity to expand the volumes to be more comparable to Cammi's results, and the final results are reported in Figure 9). However, we can still see subtle differences in volumes between this work and Cammi's at the same f value. This difference in cavity construction also influences $\mathbf{h}_\mathrm{r}$, leading to the difference in G_er (Figure 9).

VI. CONCLUSIONS

In this work, we demonstrated that by implementing the Pauli repulsion integrals of primitive basis functions with fine-grained parallelism, the free energy of the XP-PCM method can be efficiently evaluated on GPUs and can be applied to the simulation of large molecules under high pressure.

The performance was tested by calculating the XP-PCM free energy of 18 proteins with a size range of 70-500 atoms at high pressure (on the order of 1 to 100 GPa). The benchmark calculations are used to demonstrate the feasibility of applying the method on large molecules under high pressure with up to 5000 orbitals. We achieve the same scaling as the C-PCM method, showing that the evaluation of Pauli repulsion integrals in XP-PCM only introduces minimal increase of computational cost. For all tested proteins, the XP-PCM calculation took less runtime than its gas phase counterpart due to improved SCF convergence. We showed our XP-PCM implementation is feasible for any system that can be calculated in the gas phase. We also validated the accuracy of our implementation by comparing the XP-PCM calculated pressure-volume relationship with previous XP-PCM and experimental results. Very good agreement is obtained for an argon atom and an acetylene molecule.

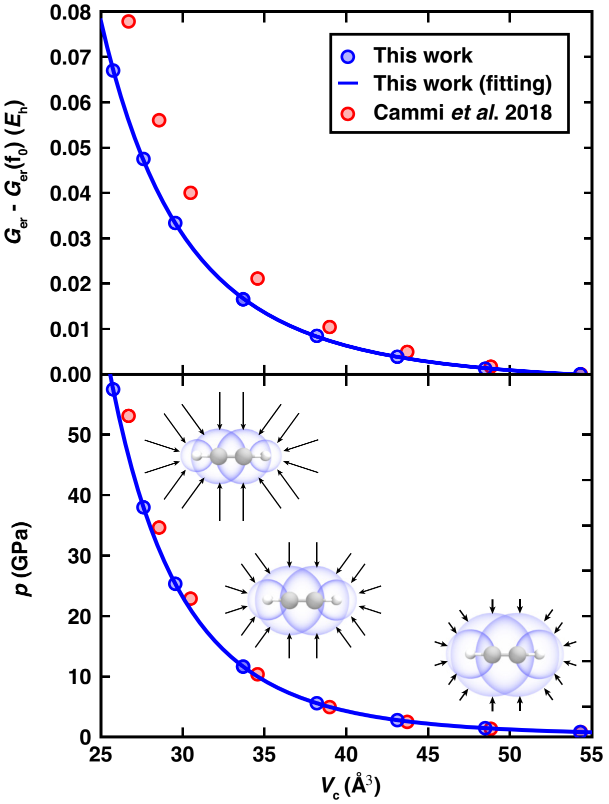

FIG. 9. Relative electronic energy G_{er} and pressure with respect to cavity volume from XP-PCM B3LYP/aug-cc-pVDZ single point calculations (f = 0.9 − 1.2) for a compressed acetylene molecule in cyclohexane. XP-PCM free energy G_{er} as a function of cavity volume for a compressed acetylene molecule in cyclohexane solvent (upper). G_{er} is calculated relative to its value obtained with cavity scaling factor f_0=1.2. Pressure-volume relationship derived from the numerical fitting approach of Eq. (23) (lower). Schematic illustration of the acetylene molecule in a compressed cavity is shown in the insets. Length of arrows indicates the strength of pressure.

In the future, we will extend our acceleration strategies to the evaluation of analytical energy gradients[29] and analytical pressure[30] of XP-PCM method. These efforts will enable efficient geometry optimization and *ab initio* molecular dynamics of large molecular systems under pressure.

SUPPLEMENTARY MATERIAL

See the supplementary material for characteristics of benchmark protein set; convergence of cavity volume as a function of grid density for 3FTR; detailed timings as a function of grid density for protein 3FTR; number of grid points, runtime, and percentage runtime as functions of cavity scaling factor f for protein 2KJM; comparison of XP-PCM and gas phase timings and SCF iterations for the benchmark protein data set; G_{er} and pressure as functions of cavity volume for argon atom and acetylene; calculated and experimental volume compression V_{c}/V_1 as a function of pressure for argon atom; and cavity volume as function of f for acetylene obtained with difference cavity radii. (PDF)

Geometries of all proteins in benchmark set; geometries of argon atom and acetylene molecule. (ZIP)

ACKNOWLEDGMENTS

This work was supported by start-up funds provided by Emory University. This work used the Extreme Science and Engineering Discovery Environment[80] (XSEDE) Bridges-2 at Pittsburgh Supercomputing Center through allocation CHE200099, , which is supported by National Science Foundation grant number ACI-1548562.

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

[1] R. J. Hemley, "Effects of high pressure on molecules," Annu. Rev. Phys. Chem. **51**, 763 (2000).

[2] O. Mishima and H. E. Stanley, "The relationship between liquid, supercooled and glassy water," Serb. Ac. B. **396**, 329 (1998).

[3] Y. Fujii, K. Hase, N. Hamaya, Y. Ohishi, A. Onodera, O. Shimomura, and K. Takemura, "Pressure-induced face-centered-cubic phase of monatomic metallic iodine," Phys. Rev. Lett. **58**, 796 (1987).

[4] L. Pauling, *The Nature of the Chemical Bond*, Vol. 260 (Cornell university press Ithaca, NY, 1960).

[5] S. Duwal, Y.-J. Ryu, M. Kim, C.-S. Yoo, S. Bang, K. Kim, and N. H. Hur, "Transformation of hydrazinium azide to molecular n8 at 40 gpa," J. Chem. Phys. **148**, 134310 (2018).

[6]T. Razzaq and C. O. Kappe, “Continuous flow organic synthesis under high-temperature/pressure conditions,” Chem-Asian. J. **5**, 1274 (2010).

[7]J. V. Badding, “High-pressure synthesis, characterization, and tuning of solid state materials,” Annu. Rev. Mater. Sci. **28**, 631 (1998).

[8]P. Stoltze and J. Nørskov, “Bridging the" pressure gap" between ultrahigh-vacuum surface physics and high-pressure catalysis,” Phys. Rev. Lett. **55**, 2502 (1985).

[9]M. Tsuda and T. Ebrey, “Effect of high pressure on the absorption spectrum and isomeric composition of bacteriorhodopsin,” Biophys. J. **30**, 149 (1980).

[10]J. Wang, A. Li, S. Xu, B. Li, C. Song, Y. Geng, N. Chu, J. He, and W. Xu, “Tunable luminescence of a novel organic co-crystal based on intermolecular charge transfer under pressure,” J. Mater. Chem. C **6**, 8958 (2018).

[11]S. Neumaier, M. Büttner, A. Bachmann, and T. Kiefhaber, “Transition state and ground state properties of the helix–coil transition in peptides deduced from high-pressure studies,” P. Natl. Acad. Sci. U.S.A. **110**, 20988 (2013).

[12]H. Imamura and M. Kato, “Effect of pressure on helix-coil transition of an alanine-based peptide: An ftir study,” Mol. Biol. Intell. Unit. **75**, 911 (2009).

[13]T. Takekiyo, A. Shimizu, M. Kato, and Y. Taniguchi, “Pressure-tuning ft-ir spectroscopic study on the helix–coil transition of ala-rich oligopeptide in aqueous solution,” Biochim. Biophys. Acta. Proteins Proteom. **1750**, 1 (2005).

[14]Y.-J. Kuan, S. B. Charnley, H.-C. Huang, W.-L. Tseng, and Z. Kisiel, “Interstellar glycine,” Astrophys. J. **593**, 848 (2003).

[15]K. Hadraoui, H. Cottin, S. Ivanovski, P. Zapf, K. Altwegg, Y. Benilan, N. Biver, V. Della Corte, N. Fray, J. Lasue, *et al.*, “Distributed glycine in comet 67p/churyumov-gerasimenko,” J. Phys. Conf. Ser. **630**, A32 (2019).

[16]C. F. Richardson and N. W. Ashcroft, “High temperature superconductivity in metallic hydrogen: Electron-electron enhancements,” Phys. Rev. Lett. **78**, 118 (1997).

[17]R. P. Dias and I. F. Silvera, “Observation of the wigner-huntington transition to metallic hydrogen,” Ieee S. Log. **355**, 715 (2017).

[18]P. M. Celliers, M. Millot, S. Brygoo, R. S. McWilliams, D. E. Fratanduono, J. R. Rygg, A. F. Goncharov, P. Loubeyre, J. H. Eggert, and J. L. Peterson, “Insulator-metal transition in dense fluid deuterium,” Ieee S. Log. **361**, 677 (2018).

[19] J. Bajorath, "Integration of virtual and high-throughput screening," Nat. Rev. Drug Discov. **1**, 882 (2002).

[20] J. E. Jaffe and A. C. Hess, "Hartree-fock study of phase changes in zno at high pressure," Phys. Rev. B **48**, 7903 (1993).

[21] B. Moses Abraham, J. Prathap Kumar, and G. Vaitheeswaran, "High-pressure studies of hydrogen-bonded energetic material 3,6-dihydrazino-s-tetrazine using dft," ACS Omega **3**, 9388 (2018).

[22] W. Zhu, X. Zhang, T. Wei, and H. Xiao, "Dft studies of pressure effects on structural and vibrational properties of crystalline octahydro-1, 3, 5, 7-tetranitro-1, 3, 5, 7-tetrazocine," Theor. Chem. Acc. **124**, 179 (2009).

[23] Q. Guo, Y. Zhao, C. Jiang, W. L. Mao, and Z. Wang, "Phase transformation in sm2o3 at high pressure: In situ synchrotron x-ray diffraction study and ab initio dft calculation," Solid State Commun. **145**, 250 (2008).

[24] R. B. Best, C. Miller, and J. Mittal, "Role of solvation in pressure-induced helix stabilization," J. Chem. Phys. **141**, 22D522 (2014).

[25] D. Paschek, S. Gnanakaran, and A. E. Garcia, "Simulations of the pressure and temperature unfolding of an α-helical peptide," P. Natl. Acad. Sci. U.S.A. **102**, 6765 (2005).

[26] H. W. Hatch, F. H. Stillinger, and P. G. Debenedetti, "Computational study of the stability of the miniprotein trp-cage, the gb1 β-hairpin, and the ak16 peptide, under negative pressure," J. Phys. Chem. B **118**, 7761 (2014).

[27] Y. Mori and H. Okumura, "Molecular dynamics of the structural changes of helical peptides induced by pressure," Mol. Biol. Intell. Unit. **82**, 2970 (2014).

[28] R. Cammi, V. Verdolino, B. Mennucci, and J. Tomasi, "Towards the elaboration of a qm method to describe molecular solutes under the effect of a very high pressure," Prog. T. Chem. **344**, 135 (2008).

[29] R. Cammi, C. Cappelli, B. Mennucci, and J. Tomasi, "Calculation and analysis of the harmonic vibrational frequencies in molecules at extreme pressure: Methodology and diborane as a test case," J. Chem. Phys. **137**, 154112 (2012).

[30] R. Cammi, B. Chen, and M. Rahm, "Analytical Calculation of Pressure for Confined Atomic and Molecular Systems Using the Extreme-Pressure Polarizable Continuum Model, volume = 39, year = 2018," J. Comput. Chem. , 2243.

[31] M. Rahm, M. Ångqvist, J. M. Rahm, P. Erhart, and R. Cammi, "Non-bonded radii of the atoms under compression," Chemphyschem **21**, 2441 (2020).

[32] M. Rahm, R. Cammi, N. W. Ashcroft, and R. Hoffmann, "Squeezing all elements in the periodic table: Electron configuration and electronegativity of the atoms under compression," J. Am. Chem. Soc. **141**, 10253 (2019).

[33] R. Cammi, "A new extension of the polarizable continuum model: Toward a quantum chemical description of chemical reactions at extreme high pressure," J. Comput. Chem. **36**, 2246 (2015).

[34] M. Pagliai, R. Cammi, G. Cardini, and V. Schettino, "Xp-pcm calculations of high pressure structural and vibrational properties of p4s3," J. Phys. Chem. A **120**, 5136 (2016).

[35] C. Caratelli, R. Cammi, R. Chelli, M. Pagliai, G. Cardini, and V. Schettino, "Insights on the realgar crystal under pressure from xp-pcm and periodic model calculations," J. Phys. Chem. A **121**, 8825 (2017).

[36] M. Pagliai, G. Cardini, and R. Cammi, "Vibrational frequencies of fullerenes c60and c70under pressure studied with a quantum chemical model including spatial confinement effects," J. Phys. Chem. A **118**, 5098 (2014).

[37] A. Klamt and G. Schuurmann, "Cosmo: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," J. Chem. Soc., Perkin Trans. 2 , 799 (1993).

[38] V. Barone and M. Cossi, "Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model," J. Phys. Chem. A **102**, 1995 (1998).

[39] U. J. Kapasi, S. Rixner, W. J. Dally, B. Khailany, J. H. Ahn, P. Mattson, and J. D. Owens, "Programmable stream processors," P. Ieee. Csfw. **36**, 54 (2003).

[40] A. Asadchev, V. Allada, J. Felder, B. M. Bode, M. S. Gordon, and T. L. Windus, "Uncontracted rys quadrature implementation of up to g functions on graphical processing units," J. Chem. Theory Comput. **6**, 696 (2010).

[41] K. Yasuda, "Two-electron integral evaluation on the graphics processor unit," J. Comput. Chem. **29**, 334 (2008).

[42] I. S. Ufimtsev and T. J. Martínez, "Quantum chemistry on graphical processing units. 1. strategies for two-electron integral evaluation," J. Chem. Theory Comput. **4**, 222 (2008).

[43] F. Liu, N. Luehr, H. J. Kulik, and T. J. Martínez, "Quantum chemistry for solvated molecules on graphical processing units using polarizable continuum models," J. Chem. Theory Comput. **11**, 3131 (2015).

[44] I. S. Ufimtsev and T. J. Martinez, "Quantum chemistry on graphical processing units. 2. direct self-consistent-field implementation," J. Chem. Theory Comput. **5**, 1004 (2009).

[45] X. Andrade and A. Aspuru-Guzik, "Real-space density functional theory on graphical processing units: Computational approach and comparison to gaussian basis set methods," J. Chem. Theory Comput. **9**, 4360 (2013).

[46] L. Vogt, R. Olivares-Amaya, S. Kermes, Y. Shao, C. Amador-Bedolla, and A. Aspuru-Guzik, "Accelerating resolution-of-the-identity second-order møller-plesset quantum chemistry calculations with graphical processing units," J. Phys. Chem. A **112**, 2049 (2008).

[47] A. E. DePrince III and J. R. Hammond, "Coupled cluster theory on graphics processing units i. the coupled cluster doubles method," J. Chem. Theory Comput. **7**, 1287 (2011).

[48] E. G. Hohenstein, N. Luehr, I. S. Ufimtsev, and T. J. Martínez, "An atomic orbital-based formulation of the complete active space self-consistent field method on graphical processing units," J. Chem. Phys. **142**, 224103 (2015).

[49] C. Song and T. J. Martínez, "Reduced scaling caspt2 using supporting subspaces and tensor hyper-contraction," J. Chem. Phys. **149**, 044108 (2018).

[50] T. N. Truong and E. V. Stefanovich, "A new method for incorporating solvent effect into the classical, ab initio molecular orbital and density functional theory frameworks for arbitrary shape cavity," Chem. Phys. Lett. **240**, 253 (1995).

[51] B. Mennucci, E. Cancès, and J. Tomasi, "Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method: Theoretical bases, computational implementation, and numerical applications," J. Phys. Chem. B **101**, 10506 (1997).

[52] E. Cancès, B. Mennucci, and J. Tomasi, "A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics," J. Chem. Phys. **107**, 3032 (1997).

[53] J. Tomasi, B. Mennucci, and E. Cancès, "The ief version of the pcm solvation method: an overview of a new method addressed to study molecular solutes at the qm ab initio level," J. Mol. Struct.: THEOCHEM **464**, 211 (1999).

[54] D. M. York and M. Karplus, "A smooth solvation potential based on the conductor-like screening model," J. Phys. Chem. A **103**, 11060 (1999), https://doi.org/10.1021/jp992097l.

[55] A. W. Lange and J. M. Herbert, "A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: The switching/gaussian approach," J. Chem. Phys. **133** (2010).

[56] F. Liu, D. M. Sanchez, H. J. Kulik, and T. J. Martínez, "Exploiting graphical processing units to enable quantum chemistry calculation of large solvated molecules with conductor-like polarizable continuum models," Int. J. Quantum Chem. **119**, e25760 (2019).

[57] E. Cancès, Y. Maday, and B. Stamm, "Domain decomposition for implicit solvation models," J. Chem. Phys. **139**, 054111 (2013).

[58] B. Stamm, E. Cancès, F. Lipparini, and Y. Maday, "A new discretization for the polarizable continuum model within the domain decomposition paradigm," J. Chem. Phys. **144**, 054101 (2016).

[59] F. Lipparini, B. Stamm, E. Cancès, Y. Maday, and B. Mennucci, "Fast domain decomposition algorithm for continuum solvation models: Energy and first derivatives," J. Chem. Theory Comput. **9**, 3637 (2013).

[60] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," Chem. Rev. **105**, 2999 (2005).

[61] J. Tomasi and M. Persico, "Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent," Chem. Rev. **94**, 2027 (1994).

[62] C. Amovilli and B. Mennucci, "Self-consistent-field calculation of pauli repulsion and dispersion contributions to the solvation free energy in the polarizable continuum model," J. Phys. Chem. B **101**, 1051 (1997), https://doi.org/10.1021/jp9621991.

[63] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Courier Corporation, 2012).

[64] A. Bondi, "Van der waals volumes + radii," J. Phys. Chem-Us. **68**, 441 (1964).

[65] M. Mantina, A. C. Chamberlin, R. Valero, C. J. Cramer, and D. G. Truhlar, "Consistent van der waals radii for the whole main group," J. Phys. Chem. A **113**, 5806 (2009).

[66] V. I. Lebedev, "Quadratures on a sphere," USSR Comput. Math. Math. Phys. **16**, 10 (1976).

[67] A. W. Lange and J. M. Herbert, "A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: The switching/gaussian approach," J. Chem. Phys. **133** (2010).

[68] R. Fukuda, M. Ehara, and R. Cammi, "Modeling molecular systems at extreme pressure by an extension of the polarizable continuum model (pcm) based on the symmetry-adapted cluster-configuration interaction (sac–ci) method: Confined electronic excited states of furan as a test case," J. Chem. Theory Comput. **11**, 2063 (2015).

[69] J. L. Whitten, "Coulombic potential energy integrals and approximations," J. Chem. Phys. **58**, 4496 (1973).

[70] L. E. McMurchie and E. R. Davidson, “One- and two-electron integrals over cartesian gaussian functions,” J. Comput. Phys. **26**, 218 (1978).

[71] S. F. Boys and A. C. Egerton, “Electronic wave functions - i. a general method of calculation for the stationary states of any molecular system,” P. Roy. Soc. Lond. A. Mat. **200**, 542 (1950).

[72] “Petachem,” `http://www.petachem.com/`, [Online; accessed 08-April-2021].

[73] S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, C. M. Isborn, S. I. L. Kokkila-Schumacher, X. Li, F. Liu, N. Luehr, J. W. Snyder Jr, C. Song, A. V. Titov, I. S. Ufimtsev, L.-P. Wang, and T. J. Martínez, “Terachem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics,” "Wiley Interdiscip. Rev. Comput. Mol. Sci." **11**, e1494 (2021).

[74] H. J. Kulik, N. Luehr, I. S. Ufimtsev, and T. J. Martinez, “Ab initio quantum chemistry for protein structures,” J. Phys. Chem. B **116**, 12501 (2012).

[75] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, “Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields,” J. Phys. Chem. A **98**, 11623 (1994).

[76] A. D. Becke, “Density-functional thermochemistry. iii. the role of exact exchange,” J. Chem. Phys. **98**, 5648 (1993).

[77] C. Lee, W. Yang, and R. G. Parr, “Development of the colle-salvetti correlation-energy formula into a functional of the electron density,” Phys. Rev. B **37**, 785 (1988).

[78] Harihara.Pc and J. A. Pople, “Influence of polarization functions on molecular-orbital hydrogenation energies,” Theor. Chim. Acta **28**, 213 (1973).

[79] D. A. Young, H. Cynn, P. Söderlind, and A. Landa, “Zero-kelvin compression isotherms of the elements $1 \leq z \leq 92$ to 100 gpa,” J. Phys. Chem. Ref. Data **45**, 043101 (2016), https://doi.org/10.1063/1.4963086.

[80] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, and G. D. Peterson, “Xsede: Accelerating scientific discovery,” Comput. Sci. Eng. **16**, 62 (2014).

[81] A. Asadchev and M. S. Gordon, “New multithreaded hybrid cpu/gpu approach to hartree–fock,” J. Chem. Theory Comput. **8**, 4166 (2012).