

Parallelized MP2 calculation

Olga Borodina², Artur Lyssenko¹, Mouza Almualla², and Hieu Dinh¹

¹Harvard Department of Chemistry and Chemical Biology

²Harvard Department of Astronomy

February 22, 2023

1 Background

Solving the Schrodinger equation for systems with more than one electron is a formidable task due to the electron-electron interaction. To obtain a rough estimation for the ground state energy, the electron-electron interaction term is often treated using a mean-field approach with a single electronic configuration (Hartree-Fock calculation). There are many approaches, such as CCSD, CI, and MP2, to calculate the remaining electron correlation starting from the mean-field solution. Each method has its own advantage, depending on the type of interactions (static vs dynamic). In computational chemistry, we want to calculate the electronic ground state energy of molecules with high accuracy to model the quantum mechanical behavior of molecules. This would aid experimental research in understanding reaction mechanisms and the design of molecules. Furthermore, the ground state energy calculation can serve as a metric/input to other data-driven methods (e.g: using machine learning to predict the properties of molecules).

In this project, we aim to parallelize the calculation for MP2 energy correction starting from the mean-field solution. MP2 (Møller–Plesset) energy is derived from perturbation theory in quantum mechanics. Given the orbital energy and the two-electron integrals in the Molecular Orbitals basis, we have the expression

$$E_{MP2} = \sum_{ij} \sum_{ab} \frac{(ia|jb) [2 (ia|jb) - (ib|ja)]}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (1)$$

where the electron correlation can be calculated using auxiliary basis functions to reduce the run time requirement to transform from Atomic Orbitals basis to Molecular Orbitals basis.

$$(ia|jb)^{RI} = \sum_P B_{ia}^P B_{jb}^P \quad (2)$$

2 Algorithms and Code Parallelization

The workflow for our calculation is presented in Figure 1. We will search for interesting molecules from public databases, such as CCDC. We then use QChem, a computational chemistry software, to perform the Hartree Fock calculation and get orbital energies (ϵ_p) with the corresponding B matrices. The implementation is given in Algo 1, which consists of multiple loops to do matrix multiplication. The parameters that will determine the scale of the problem is **number of occupied orbitals (N)**, **number of virtual orbitals (M)**, and **number of auxiliary basis (P)**. In a typical calculation routine, P is 4-5 times the total number of orbitals ($N + M$).

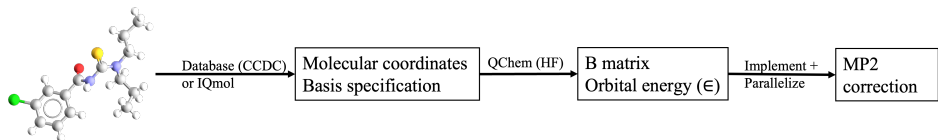


Figure 1: Workflow to obtain the MP2 energy

Code Parallelization

- We are planning on calculating the MP2 electron correlation energy on a large molecular system. For average-size system, the number of occupied orbitals is about 200, the number of virtual orbitals is about 1000, and the number of auxiliary basis is 6000. The memory requirement is $8 \times 200 \times 1000 \times 6000 \times \frac{1}{10^9} = 9.6$ Gb. For large-size systems, the basis is of dimensions 1000, 4000, and 20000. The memory requirement is $8 \times 1000 \times 4000 \times 20000 = 640$ Gb.
- We are considering doing both shared memory models (OpenMP) and distributed memory models (MPI) to parallelize the code. The NUMANode size on a node on the academic cluster is 64 Gb, and we have two per node, which is 128 Gb. Thus, for very big molecular system, which would require about 500 Gb RAM or more, we would have to request multiple nodes and parallelize with MPI. We will use the single instruction multiple data (SIMD) model for the calculation since the for loop can be divided into parallel tasks.
- We plan to do this calculation on multiple large molecules (from 5 to 10). We do not think that the problem would not depend on a high volume of I/O output. The success of this project depends on how much data we can generate and transfer to the academic cluster. Right now, we will have to put the code in C++ and test it on average size molecules first.

Validation, Verification

We will first do the calculation for small molecules (H₂O, CO₂, CH₄) using both sequential and parallel code. This would verify that the parallel code is working properly.

Algorithm 1: RI-MP2 Implementation

Data: Orbital energy(ϵ), $B_{ia}^P matrix$

Result: MP2 correlation energy

$E_{MP2} \leftarrow 0$;

for $i, j < N$ **do**

for $N \leq a, b < N+M$ **do**

for $P < AUX_BASIS$ **do**

$(ia|jb) \leftarrow (ia|jb) + B_{ia}^P B_{jb}^P$;

end

end

end

for $i, j < N$ **do**

for $N \leq a, b < N+M$ **do**

$\Delta_{ij}^{ab} = \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b$;

$E_{MP2} = E_{MP2} + \Delta_{ij}^{ab} \times [(ia|jb) [2 (ia|jb) - (ib|ja)]]$;

end

end
