Large scale parallel multireference wave function methods: development and implementation in NWChem

Dongxia Ma,[1] Wibe Albert de Jong, [2] Giovanni Li Manni,[1] Jeppe Olsen[3], Laura Gagliardi,[1]
[1] Department of Chemistry and Minnesota Supercomputing Institute, UMN
[2] Scientific Computing Group, Lawrence Berkeley National Laboratory
[3] Department of Chemistry, Aarhus University, Denmark

SciDAC Scientific Discovery through Advanced Computing

NWCHEM HIGH-PERFORMANCE COMPUTATIONAL CHEMISTRY SOFTWARE

Introduction

Multi-reference methods are designed to treat systems with strong correlations. These calculations are often very expensive in both memory and time. We aim to provide users efficient computational chemistry tools from two different ways: to develop new theoretical methods to reduce the scale of the problem, and to take advantage of the available parallel computing resources.

Implementing GASSCF in NWChem

The MCSCF we have integrated into NWChem is based on General Active Space concept (GASSCF) which is highly flexible.

The parallel version has been successfully implemented in NWChem.

Direct CI is the most time consuming part in an MCSCF iteration.

$$\sigma(I_\alpha, I_\beta) = \sum_{J_\alpha J_\beta} \langle\beta(J_\beta)\alpha(J_\alpha)|\hat{H}|\alpha(I_\alpha)\beta(I_\beta)\rangle C(J_\alpha, J_\beta), \quad \text{(σ vector)}$$

$$|0\rangle = \sum_{I_\alpha I_\beta} C(I_\alpha, I_\beta)|\alpha(I_\alpha)\beta(I_\beta)\rangle, \quad \text{(Reference wf)}$$

$$\hat{H} = \sum_{kl} h_{kl}\hat{E}_{kl} + \frac{1}{2}\sum_{ij,kl}(ij|kl)(\hat{E}_{ij}\hat{E}_{kl} - \delta_{jk}\hat{E}_{il}), \quad \text{(Hamiltonian)}$$

$$\hat{E}_{kl} = a^\dagger_{k\alpha}a_{l\alpha} + a^\dagger_{k\beta}a_{l\beta}$$

Two e$^-$ contribution is split to 3 parts: 2 α excitations, 2 β excitations, and mixed excitations.

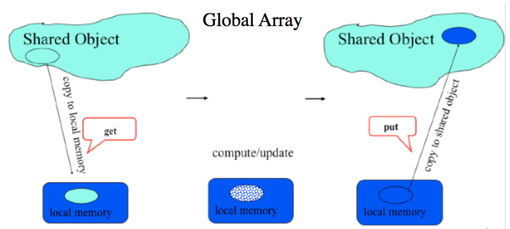
Orbital optimization uses true 2nd order Newton scheme.

Details on integrating serial LUCIA MCSCF/CI code into NWChem

- Overhaul of allocation to NWChem Memory Allocator infrastructure
- CI vector, σ vector, MO integrals are stored in global arrays that are blocked and spread over processors;
- Mutiple CI and σ vectors are stored to disk in parallel; Future algorithm to store as many in memory as possible
- Parallelization:
  - Distribute block of C to all processors
  - Each processor computes a partial contribution to the σ vector (where code spend 90% of the time)
  - Global reduction of σ vector

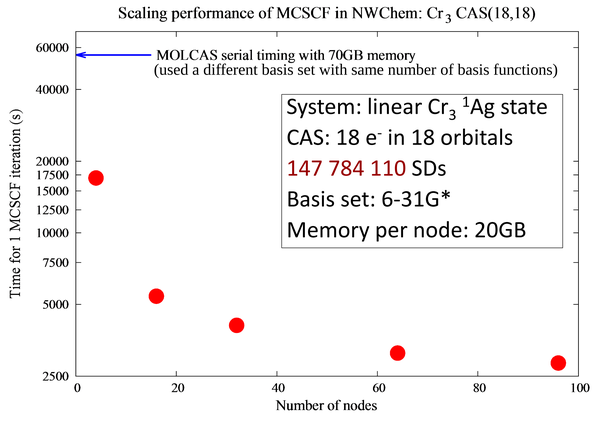
The parallel infrastructure of NWChem

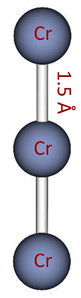
Built on the GA toolkits. To tackle the large-scale problem where the memory of a single node are not sufficient, data distribution is the only way. And it is desirable to be able to allocate the memory of the whole parallel computer to store large arrays.



(Picture taken from PNNL website)

Parallelization scaling performance of MCSCF in NWChem





Timing details for most time consuming segments

	4 cores	16 cores	32 cores	64 cores	96 cores
Building σ vector	13420	3477	2139	1350	1109
2-body density matrix	1497	479	306	218	195
Creation/annihilation mapping	450	114	60	32	24
Obtain SD blocks	309	43	56	37	44
Data movement (sigma/CI)	881	991	1218	1180	1157

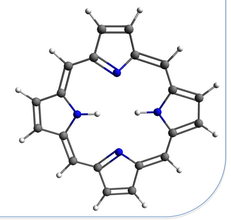
- Scalable routines to construct σ vector and density matrices
- Overhead of parallel implementation is data movement (broadcast) of CI vector and summation (reduction) of sigma vector

Challenging the limit: Free base porphin

Excited states in porphin and metal porphyrines are important for photochemistry and biochemistry. The excited states are difficult to deal with given the large conjugation system of 26 electrons in 24 orbitals. Among them 18 electrons in 20 orbitals are important.

There are 7052808360 SDs for the singlet symmetric state.

"PLACEHOLDER, RESULTS AVAILABLE ON FINAL POSTER"



Conclusion and future work

Currently

- We have implemented a parallel version of MCSCF based on general active space (GASSCF) in NWChem
- Good scaling for parallel build of σ vector; data movement is the next bottleneck to improve upon
- For the current CASSCF limit (18,18), each MCSCF iteration can be run in under an 1 hour, comparing to 15+ hours in serial
- Able to tackle bigger science problems not possible before

Next step

- Implement a scalable version of the SplitGAS approach within the NWCHEM GASSCF framework to approximate FCI

Reference

1. D. Ma, W. A. de Jong, G. Li Manni, J. Olsen, L. Gagliardi; *"Large-scale parallel Multiconfigurational Self Consistent Field method implementation in NWChem"*, in preparation.
2. G. Li Manni, D. Ma, F. Aquilante, J. Olsen and L. Gagliardi; *"SplitGAS Method for Strong Correlation and Challenging Case of Cr_2"*, J. Chem. Theo. Comp. **9**, 3375 (2013).
3. D. Ma, G. Li Manni, L. Gagliardi; *"The Generalized Active Spcace concept in Multiconfigurational Self-Consistent-Field methods"*, J. Chem. Phys., **135**, 044128 (2011).

Acknowledgement

This work is supported by DOE SciDAC program Grant No. DE-SC0008666.