A Fast Select Inversion Algorithm for Green's Function Calculation in Many-body Quantum Monte Carlo Simulation [1]

Chengming Jiang[1], Zhaojun Bai[1], and Richard Scalettar[2]

[1]Department of Computer Science/ [2]Department of Physics, University of California, Davis

Motivation

- **Hubbard Hamiltonian**[2]

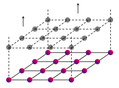
Electron interactions of quantum many-body system

- **DQMC simulation**[3]

Determinant Quantum Monte Carlo simulation of Hubbard Hamiltonian

- **QUEST**

- Quantum Electron Simulation Toolbox, state-of-the-art implementation of DQMC
- Computational kernel : selected entries of a large set of **Green's functions**



Lattice structure

http://quest.ucdavis.edu

Hubbard matrix

Hubbard matrix is of the block **p-cyclic** structure:

$$M = \begin{pmatrix} I & & & B_1 \\ -B_2 & I & & \\ & \ddots & \ddots & \\ & & -B_L & I \end{pmatrix}$$

Green's function matrix:

$$G = M^{-1} = (G_{kl})$$

Explicit expression of G_{kl}:

$$G_{kl} = W_{kk}^{-1} Z_{kl} \qquad (1)$$

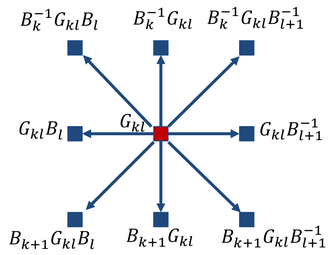
$$W_{kk} = I + B_k B_{k-1} \cdots B_1 B_L \cdots B_{k+1}$$

$$Z_{kl} = \begin{cases} -B_k \cdots B_1 B_L \cdots B_{k+1}, & k < l \\ I, & k = l \\ B_k B_{k-1} \cdots B_{l+1}, & k > l \end{cases}$$

** (1) can be extended to p-cyclic matrix in the normal form (all diagonals are arbitrary blocks).*

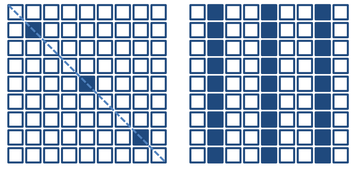
Adjacency relations

By the explicit expression (1), G_{kl}'s adjacent blocks can be easily computed.



Selected inversion

- A collection S of selected blocks of G
- Examples of selected inversion:



Block cyclic reduction

- A factor-of-c block cyclic reduction of M:

$$\widehat{M} = \begin{pmatrix} I & & & \hat{B}_1 \\ -\hat{B}_2 & I & & \\ & \ddots & \ddots & \\ & & -\hat{B}_b & I \end{pmatrix}$$

where b=L/c, and

$$\hat{B}_i = B_{j_0} B_{j_0-1} \cdots B_{j_0-c+1}$$

and $j_0 = ci - q, 0 \le q \le c-1$.

- $\hat{G} = \widehat{M}^{-1}$ is b^2 blocks of G:

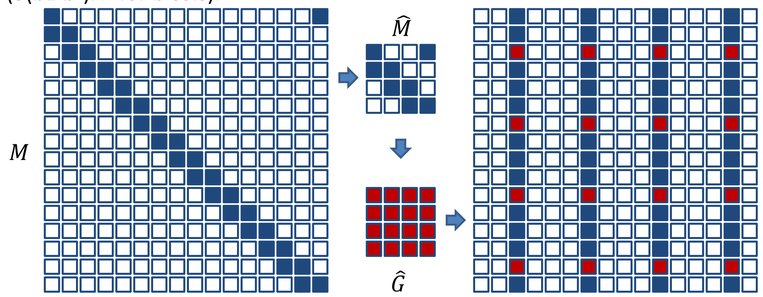
$$\hat{G}_{k_0,l_0} = G_{ck_0-q,cl_0-q}$$

where $1 \le k_0, l_0 \le b$.

Fast selected inversion (FSI)

Three steps :

1. CLS - apply the block cyclic reduction for a structure-preserving reduction ($2b(c-1)N^3$);
2. BSOFI [4] - compute the inverse of the reduced block p-cyclic matrix by block structured orthogonal factorization and inversion ($7b^2N^3$);
3. WRP - use $\hat{G}_{k_0,l_0}$ as seeds to form the selected inversion through adjacency relations ($3(bL-b^2)N^3$ for b cols).



The advantages of FSI:

- FSI is stable with moderate clustering size of c[5];
- FSI is **fast** (flops);
- FSI is embarrassingly **parallel**.

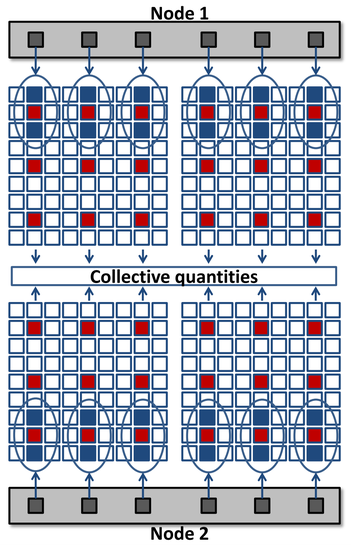
Sel. inv.	Explicit	FSI	Reduction
b diag.	$2b^2cN^3$	$[2(c-1)+7b]bN^3$	$O(b)$
b col.	$b^3c^2N^3$	$3b^2cN^3$	$O(bc)$

Hybrid MPI/OpenMP

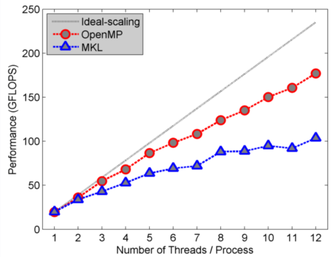
Input: $M_1, M_2, ..., M_m$ and c
Output: $\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_m$ and global_measurement_quantities
On MPI_root {
MPI_Init
m_per_MPI $= m/$num_MPI_process
MPI_Scatter (sbuff:{M_i},scount:m_per_MPI, ...)
}
On each MPI_process){
for $iter = 1, ..., m\_per\_task$ **do**
select $q \in \{0, ..., c-1\}$ randomly
!$omp parallel do
$\widehat{M} = \text{CLS}(M, c, q)$ by OpenMP multi-threads
!$omp end parallel do nowait
$\widehat{G} = \widehat{M}^{-1}$ by BSOFI
initialize $\mathcal{S} = \{\hat{G}_{k_0,\ell_0} | 1 \le k_0, \ell_0 \le b\}$
!$omp parallel do
execute WRP by OpenMP multi-threads
compute local_measurement_quantities
!$omp end parallel do nowait
end for
MPI_Reduce (sbuff:local_measurement_quantities,...)
}
On MPI_root{
MPI_Finalize
compute global_measurement_quantities
}

MPI Level: for each p-cyclic matrix in a large set

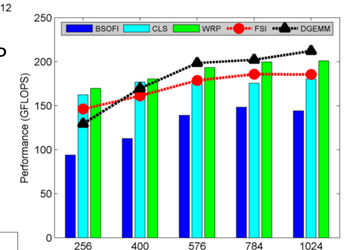
OpenMP Level: for each seed (red square) in one matrix



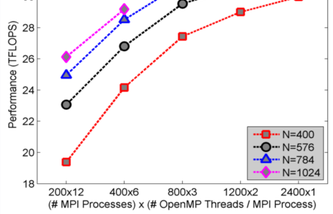
Performance



FSI scalability on 12-core Intel Ivy Bridge processor*: explicit OpenMP vs MKL (N=576, L=100, c=10)
80% improvement



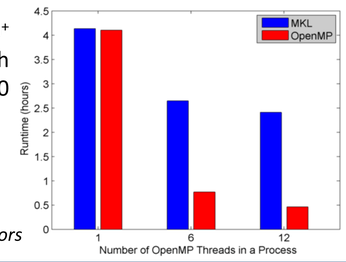
Performance rate* of OpenMP multi-threaded FSI and profile of three steps (L=100,c=10)
Performance rate of FSI close to the MKL BLAS-3 *dgemm*



Performance rate* of parallel application of FSI for multiple Green's functions with different numbers of MPI processes & OpenMP threads (L=100, c=10)

* *Edison, DOE NERSC's newest Cray X30 supercomputer*

FSI in QUEST. The CPU+ runtime of a full DQMC with 100 warmup loops and 200 measurement loops (N=400, L=100, c=10)
5x speedup



+ *Two 6-core Intel Westmere processors*

References

1. C. Jiang, Z. Bai, R. Scalettar, IPDPS submitted, 2015.
2. J. Hubbard, 1963.
3. R. Blankenbecler, D. Salapino, R. Sugar, 1981.
4. S. Gogolenko, Z. Bai, R. Scalettar, 2014.
5. Z. Bai, W. Chen, R. Scalettar, I. Yamazaki, 2012.