

HBMC Ordering for SIMD Vectorization of IC Preconditioning

Takeshi Iwashita, Senxi Li, Takeshi Fukaya
(Hokkaido University)
20 min.

Research topics of HU group

- ▶ Mixed precision computing
 - GMRES and its preconditioned version
 - Invited talk by Dr. Fukaya
 - H-matrix
 - [R. Ooi, T. Iwashita, T. Fukaya, A. Ida and R. Yokota, Effect of Mixed Precision Computing on H-matrix Vector Multiplication in BEM Analysis, ACM/IPSJ HPC Asia 2020, pp. 92-101.](#)
- ▶ SIMD friendly algorithm for iterative linear solver
 - IC and ILU preconditioning

Colleagues & Publication

- ▶ Takeshi Fukaya (Hokkaido Univ.)
- ▶ Li Senxi (Hokkaido Univ. > The University of Tokyo)
- ▶ Takeshi Iwashita, Senxi Li, Takeshi Fukaya, “Hierarchical block multi-color ordering: a new parallel ordering method for vectorization and parallelization of the sparse triangular solver in the ICCG method”, CCF Transactions on High Performance Computing (Springer), volume 2, pages 84–97, (2020).
- ▶ <https://doi.org/10.1007/s42514-020-00030-z> (Open Access)

Background

- ▶ **A sparse triangular solver** is a main component of the Gauss-Seidel (GS) smoother, SOR method and IC/ILU preconditioning, which are used as building blocks in various computational science or engineering analyses.
- ▶ However, it is well known that the sparse triangular solver, which consists of **forward and backward substitutions**, cannot be straightforwardly parallelized.
- ▶ **Parallel ordering** is one of the most popular methods for parallelization of a sparse triangular solver.
 - But, it entails **a trade-off problem between convergence and number of synchronization points**.
- ▶ One of the solutions for the trade-off problem is **block multi-coloring**.

Background & Research Target

- ▶ Block multi-coloring (BMC)
 - Multi-color ordering is applied to blocks of unknowns.
- ▶ In the context of parallel IC/ILU preconditioning, the block coloring was investigated in a finite difference method (Iwashita et al. SISC 2005). And, the method was enhanced for a general sparse linear system (Iwashita et al. IPDPS2012).
- ▶ This technique has been used in various applications because of its advantages in terms of convergence, data locality, and the number of synchronizations (Semba et al. 2013; Tsuburaya 2015; Ruiz et al. 2018 etc.)
- ▶ However, the block multi-coloring method has a drawback in its implementation using SIMD vectorization.
- ▶ **In this research, we aimed to develop a new parallel ordering that has the same convergence rate as BMC and makes the SIMD vectorization of substitutions possible.**

Problem

- ▶ We consider an n -dimensional linear system of equation: $Ax = b$.
- ▶ We focus on parallelization of the sparse triangular solver:
 - $y = L^{-1}r, z = U^{-1}y$
 - L and U are lower and upper triangular matrices, respectively. They have the same nonzero element pattern as A . (Gauss-Seidel, SOR, IC/ILU preconditioning cases)
 - The sparse triangular solver (forward and backward substitutions) are not straightforwardly parallelized.
- ▶ **Reordering (parallel ordering)**
 - One of the most popular techniques for parallelization of the sparse triangular solver.
 - **Reordering**: permutation of the elements of index set I
 - $I = \{1, 2, \dots, n\}$ that corresponds to the index of each unknown.
 - i -th unknown is moved to $\pi(i)$ -th unknown in the new system.
 - New (reordered) linear system: $\bar{A}\bar{x} = \bar{b}$, $\bar{x} = P_{\pi}^T x$, $\bar{A} = P_{\pi} A P_{\pi}^T$, $\bar{b} = P_{\pi} b$
 - \bar{A} has a suitable form for parallel processing.

Equivalence of convergence

- ▶ When $\bar{x}^{(j)} = P_{\pi}x^{(j)}$ holds at every j -th step under initial setting $\bar{x}^{(0)} = P_{\pi}x^{(0)}$, we say the iterative solver has equivalence of convergence for two original and reordered linear systems. (Upper subscript means the iteration count.)
- ▶ Jacobi method and most of non-preconditioned Krylov subspace methods have equivalence of convergence.
 - Jacobi method and most of Krylov subspace methods are not affected by the ordering.
- ▶ Ordering usually affects the convergence of the iterative solver involving (S)GS, (S)SOR, or IC(0)/ILU(0) preconditioning parts.
 - However, when the condition shown in the next slide, which is called “Equivalent Reordering (ER) Condition”, is satisfied, the equivalence of convergence holds for a solver involving these preconditioning methods.

Equivalent Reordering Condition

ER condition -

$\forall i_1, i_2 \in I$ such that $a_{i_1, i_2} \neq 0$ or $a_{i_2, i_1} \neq 0$, $\text{sgn}(i_1 - i_2) = \text{sgn}(\pi(i_1) - \pi(i_2))$.

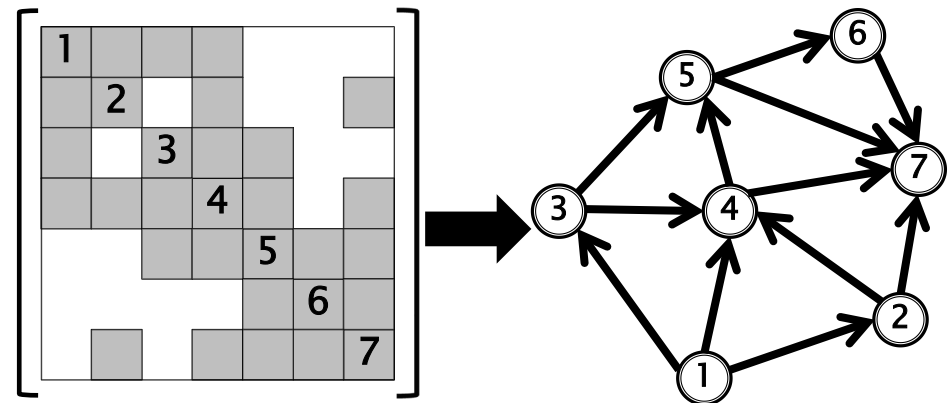
a_{i_1, i_2} : i_1 -th row i_2 -th column element of original coefficient matrix

In other words

Two orderings have an identical ordering graph.

Doi, S., and Lichniewsky, A., Res. Report No. 1452. INRIA, France, (1991)

The sketch of the proof is given in the appendix of our paper.



Coefficient matrix
(Colored elements: nonzero)

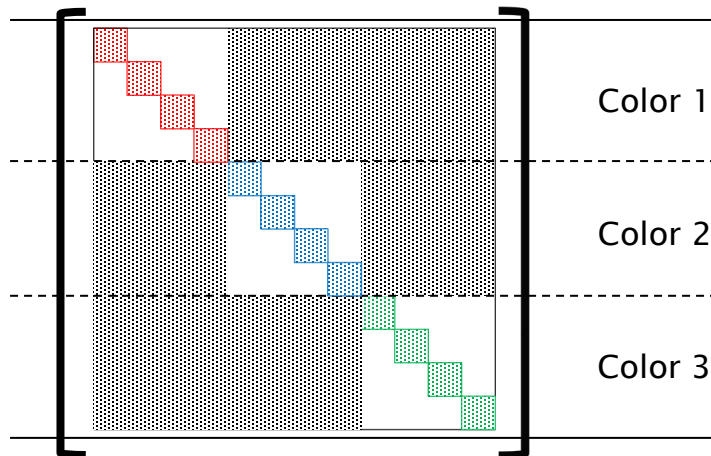
Ordering graph

Hierarchical block multi-color ordering

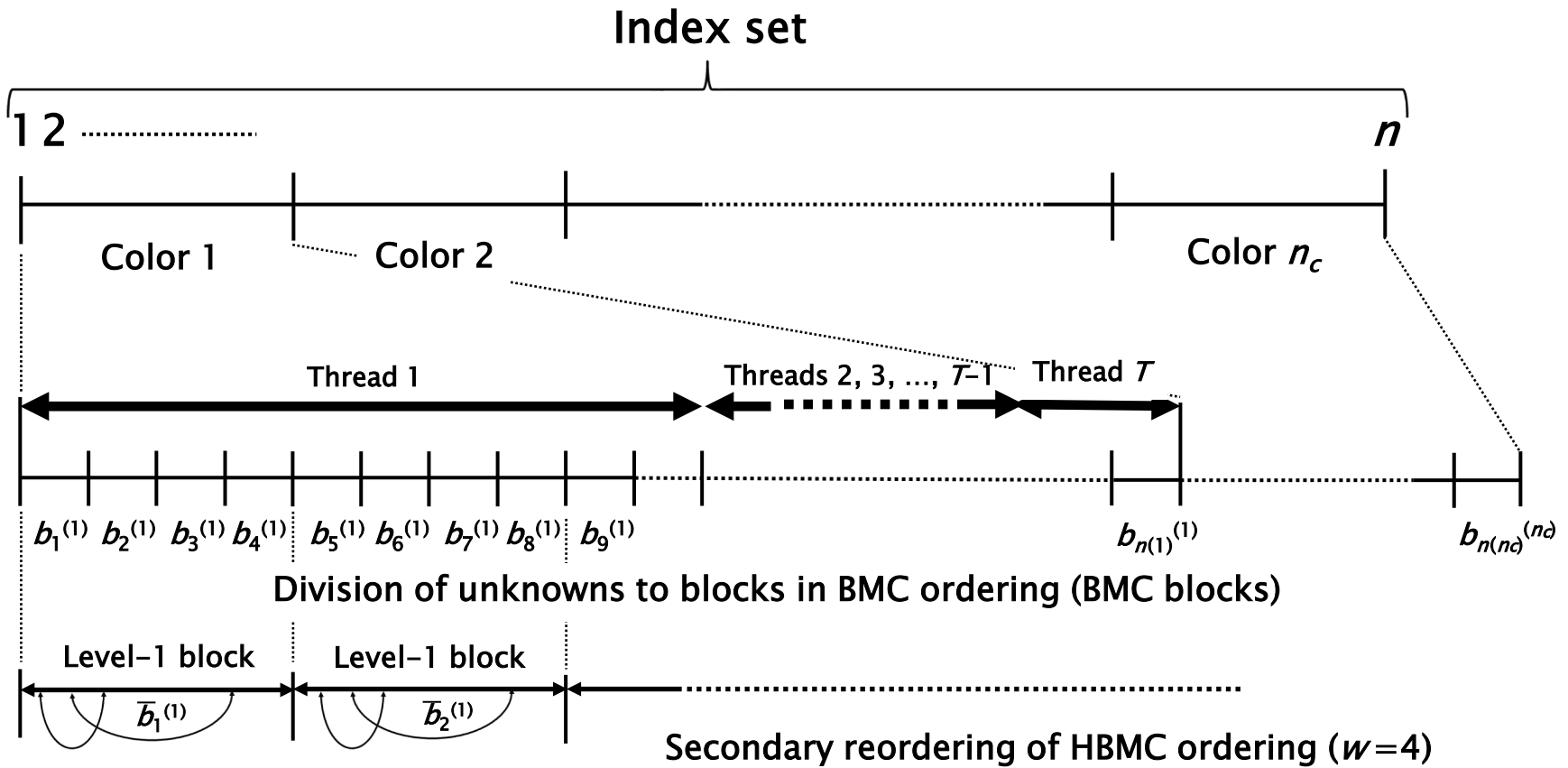
- ▶ Design points for new ordering
 - The same convergence rate of block multi-color ordering
 - The same number of synchronization points for multi-threads
 - Availability of SIMD vectorization for forward-backward substitutions

We first apply the (algebraic) block multi-color ordering to the linear system.

While keeping the ordering graph, we reorder it again (secondary reordering).



(Nonzero entries exist only in dotted parts)

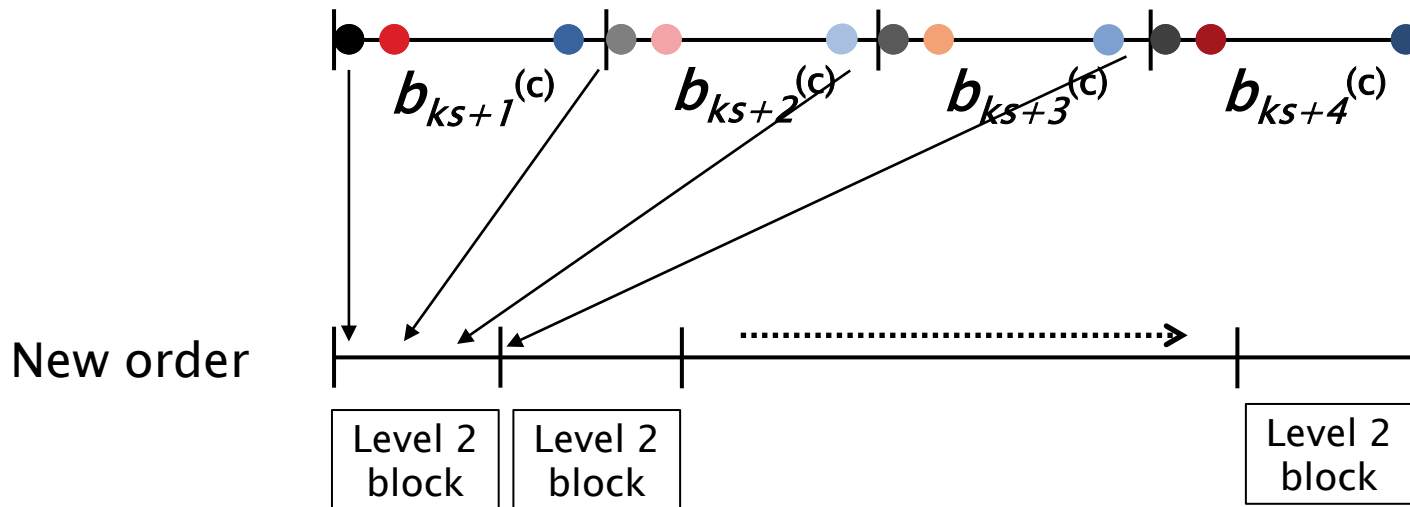


We generate **level-1 blocks** each of which consists of w **BMC blocks** in each color. (w : SIMD length)

We reorder the unknowns in each level-1 block. It does not affect the ordering graph between the unknowns belonging to different level-1 blocks.

Reordering inside level-1 block

Level 1 block which consists of 4 BMC blocks ($w=4$)



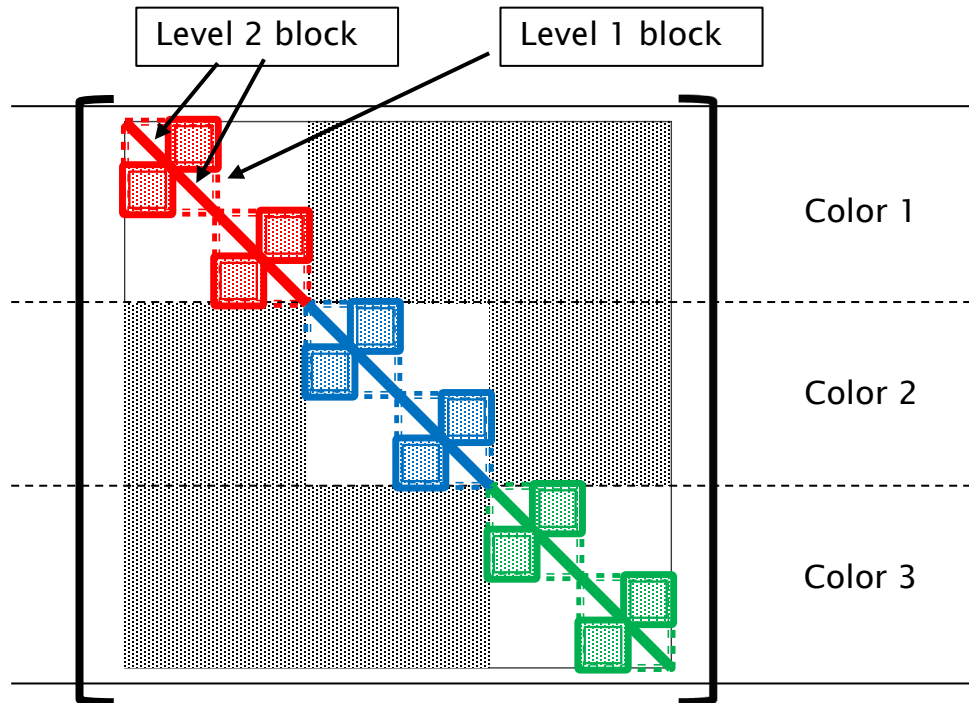
The unknowns in the same level-2 block has no data relationship.

→ SIMD vectorization for the substitutions is possible. (SIMD length: w)

The above reordering process does not change the ordering graph that corresponds to the unknowns in the level 1 block.

The new ordering is an equivalent ordering to BMC.

Hierarchical block multi-color ordering



The coefficient matrix arising from HBMC ordering.

Multithreaded and vectorized implementation of forward substitution using intrinsic functions

- ✓ The parallelism of level-1 blocks is exploited by multiple threads.
- ✓ The parallelism among unknowns in a level-2 block is exploited by SIMD instructions.

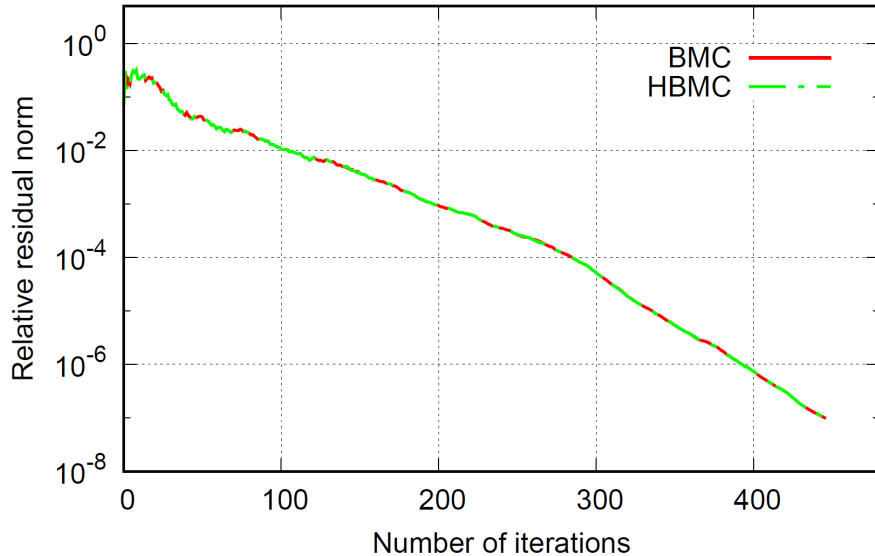
```
for (c = 0; c < nc; c++){
    #pragma omp for private(p, j, num, t, index, \
        mtmp, mval, pos, mb, mdiag)
    for (k = lev1b[c]; k < lev1b[c+1]; k++){
        num = k * 8 * bs;
        index = mat.offset[k];
        j = lev2b[k];
        for (p = j; p < j + bs; p++){
            mtmp = _mm512_load_pd( &r[num] );
            for (t = 0; t < mat.slen.lev2b[p]; t++){
                mval = _mm512_load_pd( &val[index] );
                pos = _mm512_load_epi32( &col[index] );
                mb = _mm512_i32logather_pd( pos, z, 8 );
                mtmp = _mm512_sub_pd( mtmp, \
                    _mm512_mul_pd(mval,mb) );
                index += 8;
            }
            mdiag = _mm512_load_pd( &diaginv[num] );
            mtmp = _mm512_mul_pd( mtmp, mdiag );
            _mm512_store_pd( &z[num], mtmp );
            num = num + 8;
        }
    }
}
```

Numerical tests

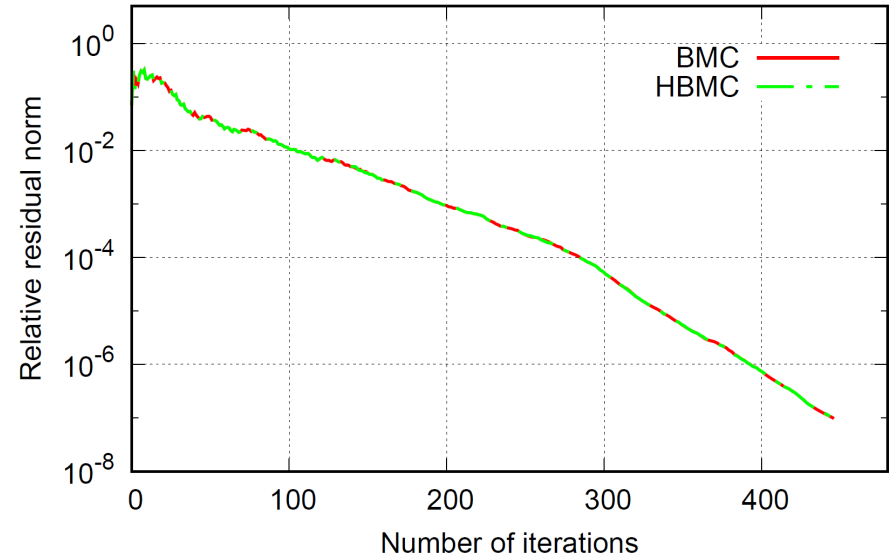
- ▶ The program is written in C.
- ▶ We used three types of nodes.
 - A node of Cray XC40 (Xeon Phi KNL processor)
 - A node of Cray CS400 (2 Xeon Broadwell processors)
 - A node of Fujitsu CX2550 (2 Xeon Skylake processors)
- ▶ Three multi-threaded IC(0)-CG solvers based on MC, BMC, and HBMC orderings were tested.
- ▶ Storage format
 - Coefficient matrix: CSR
 - Preconditioner: CSR(MC, BMC), SELL(HBMC)
- ▶ The convergence criterion: relative residual norm less than 10^{-7}
- ▶ Test problems: 7 matrices from SuiteSparse Matrix collection and a linear system arising in finite edge-element eddy current analysis

Equivalence of convergence

- ▶ We checked the convergence behaviors of BMC and HBMC.



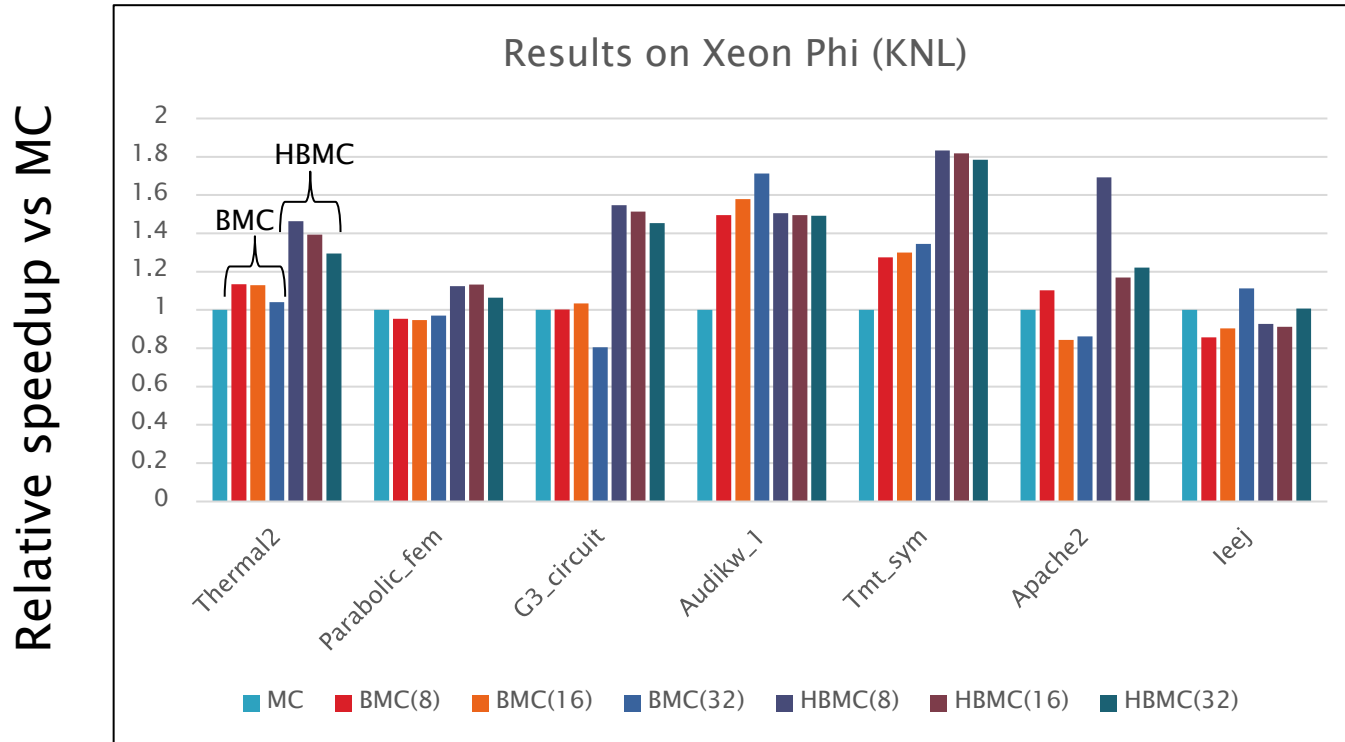
G3_circuit



leej

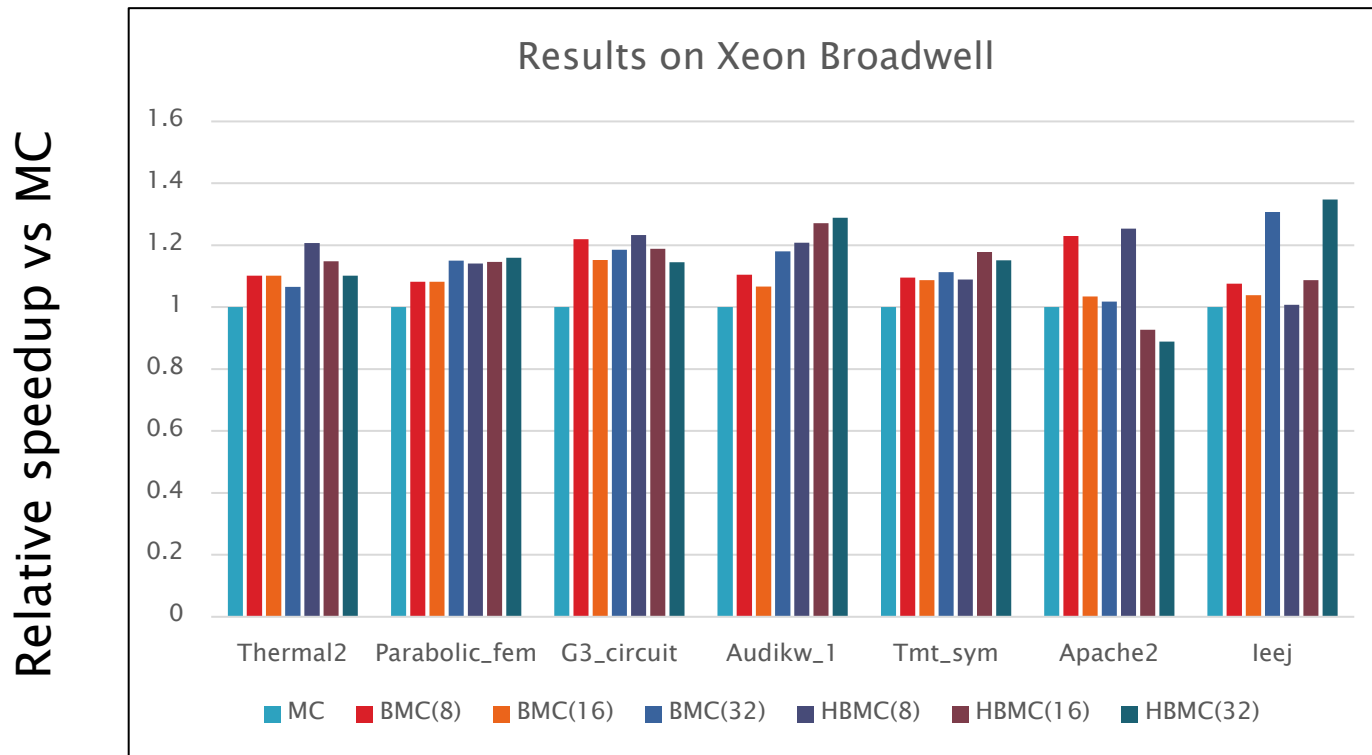
- ▶ The two lines of the relative residual norms for BMC and HBMC overlap, which indicates that the solvers had an equivalent solution process.
- ▶ The equivalence of convergence was also confirmed in all test cases.

Numerical results



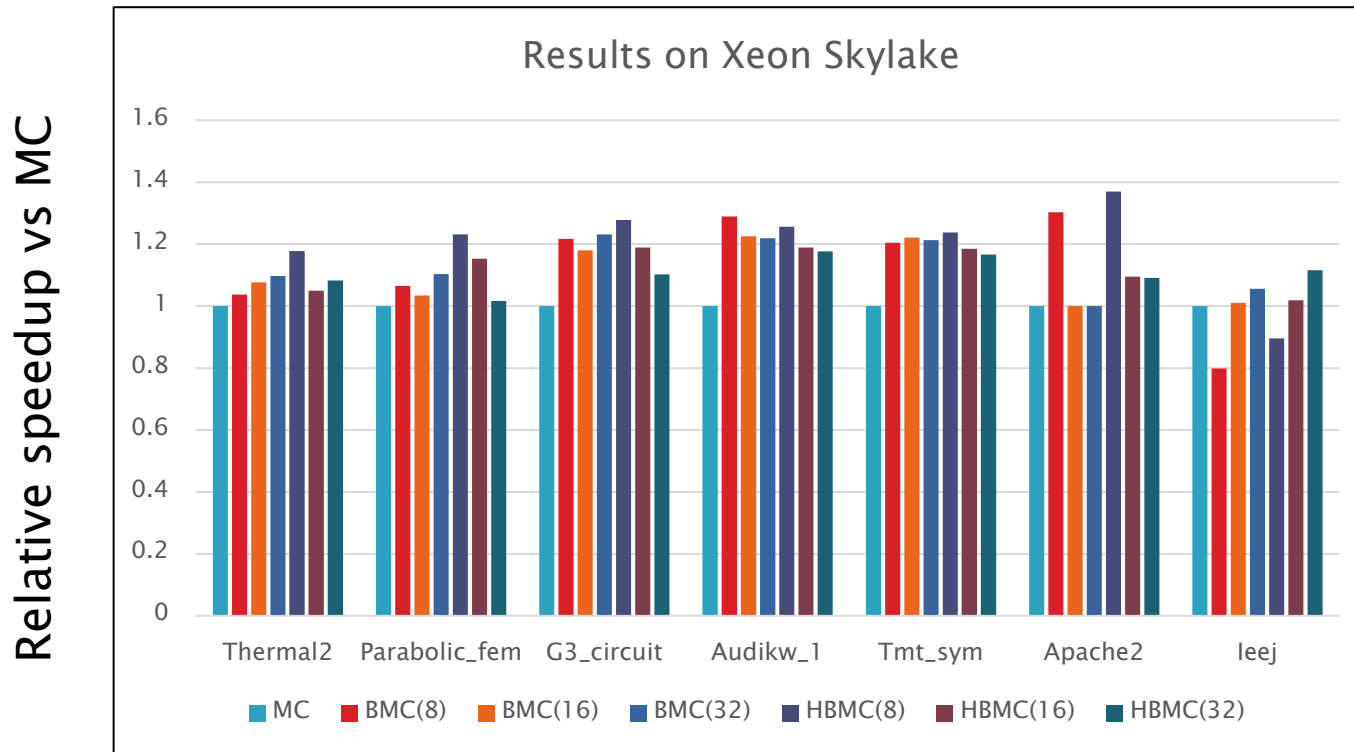
- ▶ BMC and HBMC are superior to MC in all tests.
- ▶ In 5 out of 7 datasets, HBMC outperforms BMC.

Numerical results



- ▶ BMC and HBMC are superior to MC with an appropriate block size in all tests.
- ▶ HBMC exhibits better performance than BMC.

Numerical results



- ▶ BMC and HBMC are superior to MC with an appropriate block size in all tests.
- ▶ HBMC exhibits better performance than BMC except for Audikw_1 test.
(When SIMD width is set to be 8, the number of padding elements for SELL format is large in the Audikw_1 test.)

Conclusions

- ▶ A new parallel ordering technique, **hierarchical block multi-color ordering (HBMC)**, was proposed for vectorizing and multithreading the sparse triangular solver.
- ▶ HBMC was designed to maintain the advantages of the block multi-color ordering (BMC) in terms of convergence and the number of synchronizations. In the method, the coefficient matrix was transformed into the matrix with hierarchical block structures.
 - **The level-1 blocks** were mutually independent in each color, which was exploited in **multithreading**.
 - Corresponding to **the level-2 block**, the substitution was converted into w (= SIMD width) independent steps, which were efficiently processed by **SIMD instructions**.
- ▶ Numerical tests were conducted to examine the proposed method using seven datasets on three types of computational nodes. The numerical results confirmed **the equivalence of the convergence of HBMC and BMC**. Moreover, numerical tests indicated that **HBMC outperformed BMC in 18 out of 21 test cases** (seven datasets \times three systems), which confirmed the effectiveness of the proposed method.
- ▶ Future work: we will examine our technique for other application problems, particularly a large-scale multigrid application.

Thank you

We will welcome your visit to Hokkaido Univ.

