北海道大学

INFORMATION INITIATIVE CENTER
Hokkaido University

北大スパコン Polaire

ー設計思想と活用研究ー

北海道大学 情報基盤センター

岩下　武史

北海道大学情報基盤センター

- ネットワーク型「学際大規模情報基盤共同利用・共同研究拠点」
  （**JHPCN**: 北大, 東大, 京大, 東北大, 東工大, 名大, 阪大, 九大）
  → 共同研究の公募（スパコン，クラウドの資源提供）
- **HPCI** (High Performance Computing Infrastructure) 資源提供機関
- 1962年：大型計算機センター発足（全国共同利用施設）
- 2003年：1979年に発足した情報処理教育センター（後に情報メディア教育研究総合センター）と統合 → 情報基盤センター発足
- スーパーコンピュータ，ネットワーク，クラウド，情報メディア教育，コンテンツ，システムザイン，サイバーセキュリティなどに関する研究開発を推進

INFORMATION INITIATIVE CENTER
Hokkaido University

現行スパコンシステムの設計思想

全国共同利用のスーパーコンピューティングサービスを展開し，共同研究拠点，HPCI第二階層資源に貢献するスパコンシステム

◆ 高い性能

- 省電力性能の重要性
  - ✓ HPC（高性能計算）のトレンド：消費電力がシステム性能を制限
  - ✓ 情報基盤センターのファシリティの制約
  - ✓ 負担金（電気代に充当）当たりの演算性能の拡大

◆ 使いやすさ

- オープンスーパーコンピュータ（T2K）の思想
  - ✓ コモディティデバイス，オープンソースソフトウェアを活用
  - ✓ 幅広いサイエンス分野に貢献
- 旧スパコンからのソフトウェア資産の継承
  - ✓ 旧スパコンの特徴： 高メモリ帯域，大容量共有メモリ，高い通信性能

現行システムの概要

- 最新のx86 CPUとLinux OSを搭載した二つの演算サブシステムとストレージシステム
- 旧スパコンの約20倍の理論演算性能
- オープンソースを含む多様なソフトウェアを整備
- ユーザニーズ，計算資源の効率的利用，省電力運転を同時に実現する北大センター独自のジョブスケジューラを設計・導入

現行スパコンシステム

2種類のサブシステム（計算ノード群）とストレージシステムで構成

高速ネットワーク（Intel Omni-Path）

Grand Chariot
（サブシステムA）
※仏語で“北斗七星”

使いやすさ重視の主力システム

サブシステムB

ストレージシステム
16PB

Polaire導入の狙い

- 限られた電力キャップの中で，性能を上げたい
  - 消費電力性能に優れたシステム
- HPCのトレンドに追従する
  - 大規模シミュレーションを実施するユーザ，将来のHPCシステムに向けた計算科学プログラム開発を行うユーザの開発環境の提供
    - 他機関の「大規模なシステム」の利用を想定
  - 主にJHPCN／HPCIでの利用を想定
- 2017年8月の選択
  - GPUかメニーコアか？
  - OFPの存在
  - 富岳に向けたプログラム開発
    - メニーコア，Wide SIMD

現行スパコンシステム

2種類のサブシステム（計算ノード群）とストレージシステムで構成

高速ネットワーク（Intel Omni-Path）

Grand Chariot
（サブシステムA）
※仏語で“北斗七星”

使いやすさ重視の主力システム

Polaire
（サブシステムB）
※仏語で“北極星”

HPCのトレンドを見据えた省電力システム

ストレージシステム
16PB

サブシステム（Grand Chariot / Polaire） の構成

Grand Chariot

ノードの構成

CPU	Intel Xeon Gold 6148 x 2個 （Skylake, 2.4GHz, 20コア）
メモリ	384GB
ストレージ	240GB SSD
OS	Cent OS

理論演算性能：3.07FLOPS/ノード

×

1,004ノード

（ノード間ネットワーク：Intel Omni-Path）

総演算性能：3.08PFLOPS

Polaire

ノードの構成

CPU	Intel Xeon Phi 7250 x 1個 （KNL, 1.4GHz, 68コア）
メモリ	96GB (+16GB 高速メモリ)
ストレージ	64GB SATA Flush
OS	Cent OS

理論演算性能：3.04FLOPS/ノード

×

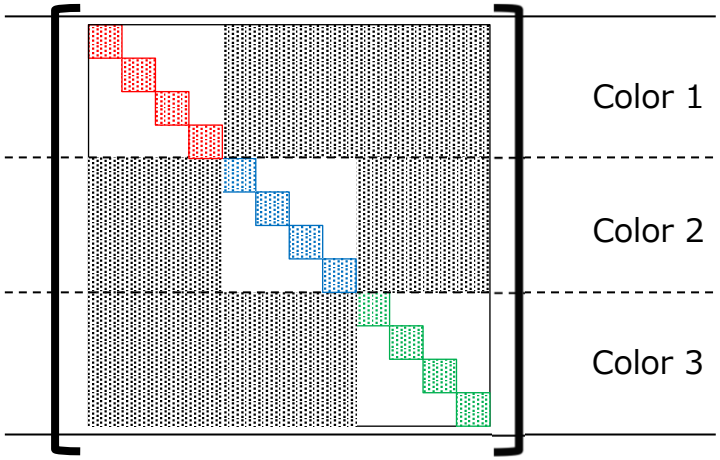
288ノード

（ノード間ネットワーク：Intel Omni-Path）

総演算性能：0.87PFLOPS

Polaireの活用研究

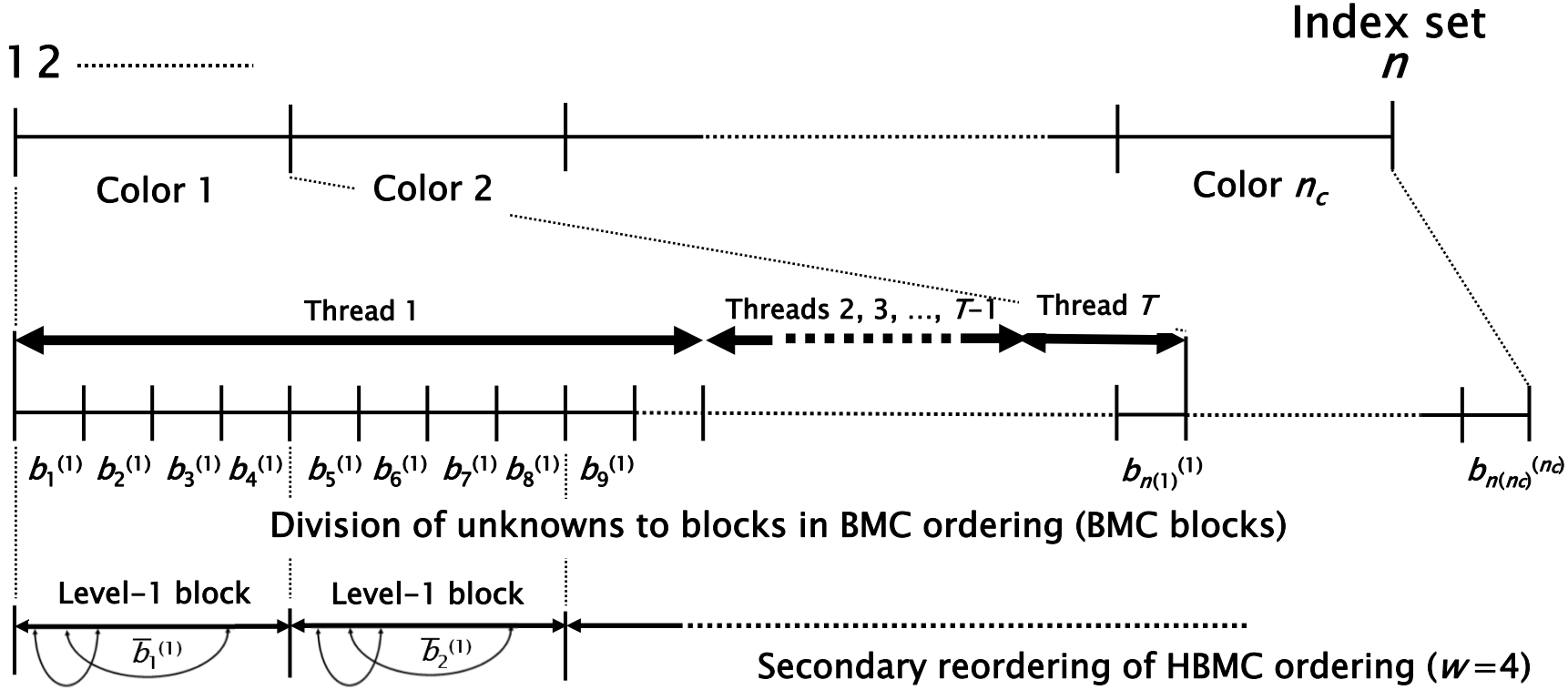
- SIMDをうまく使いたい
- ブロック化多色順序付け法
  - IC/ILU分解前処理，GSスムーザのマルチスレッド並列処理手法（Iwashita et al., IPDPS2012）
  - HPCGベンチマークでも利用される
  - 並列化された代入計算の最内ループがSIMD化できない



ブロック化多色順序付け法による係数行列

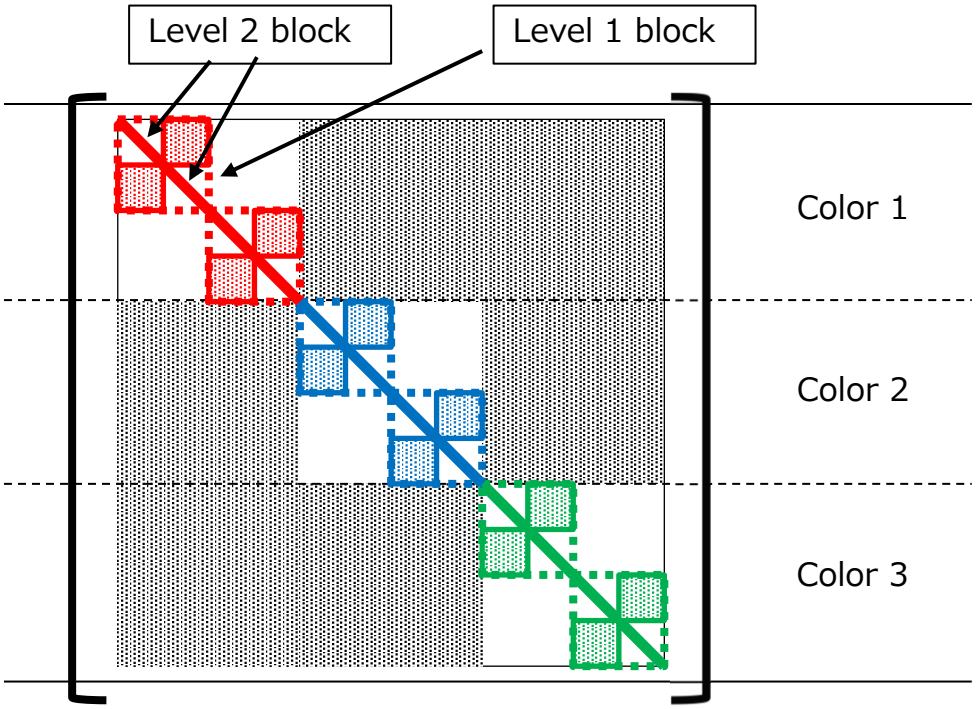
Polaireの活用研究

- 階層ブロック化多色順序付け法
  - ブロック化多色順序付け法と同じ収束性を保証
  - T. Iwashita et al., CCF Transactions on High Performance Computing (Springer), vol. 2, (2020), pp. 84-97で発表



Polaireの活用研究

- 階層ブロック化多色順序付け法
  - Level 1 ブロック：マルチスレッド並列処理
  - Level 2 ブロック：SIMD並列



階層ブロック化多色順序付け法による係数行列

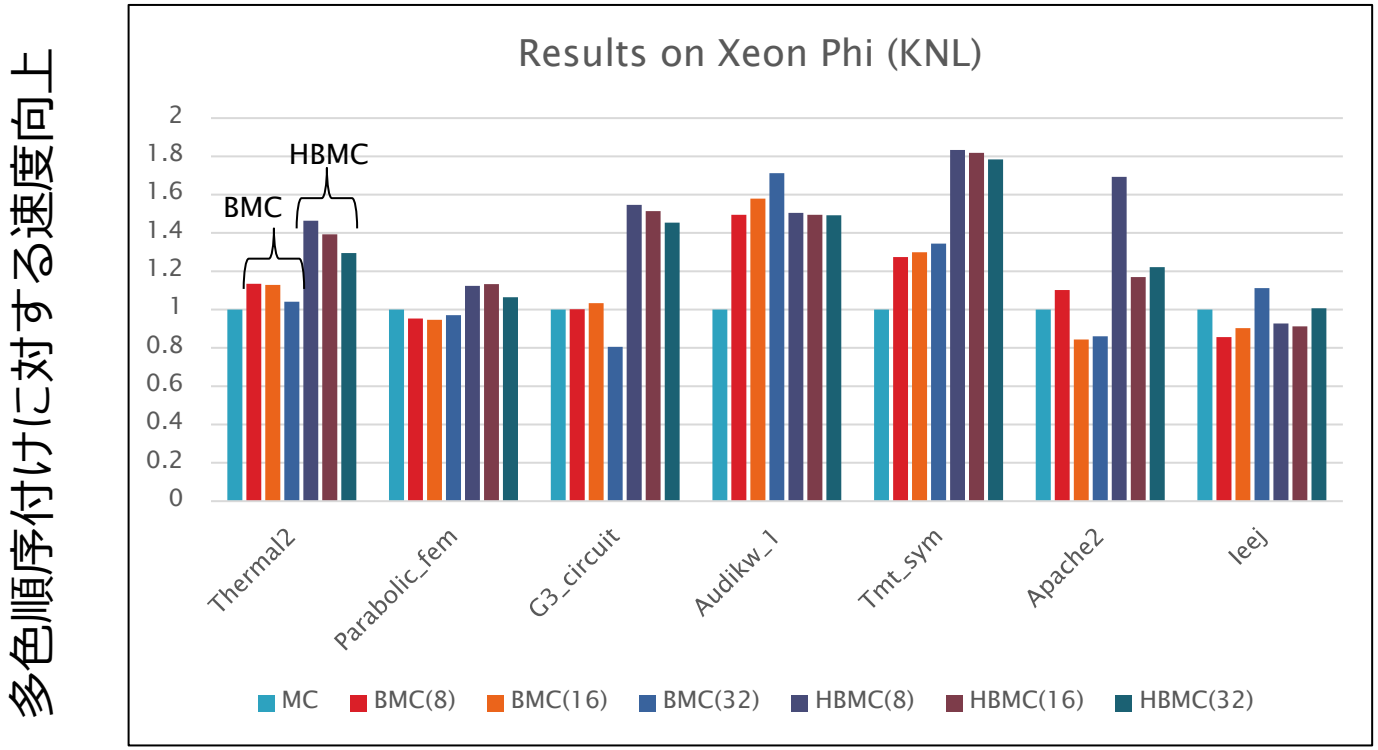
代入計算の実装例

```
for (c = 0; c < nc; c++){
  #pragma omp for priv
  for ( k = lev1b[c]
    num = k * 8 * b
    index = mat.offs
    j = lev2b[k];
    for ( p = j; p < j
      mtmp = _mm512_load
      for ( t = 0; t <                    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;
    }
  }
}
```

最内ループのSIMD化を可能に

Polaireの活用研究

- 数値実験
  - Suite Sparse Matrix Collectionを利用
  - 並列ICCGソルバの性能を比較



まとめ

- 北大では，副次的なシステムとして，省電力性能が高く，今後のHPCのトレンドにそったシステムの導入を検討
- OFPの存在，富岳の開発を踏まえ，Xeon Phiによるシステムの導入を選択
- OFPやPolaireを活用するためには，SIMDをうまく使うことを考える必要がある
  - 定量的な評価は単純ではないが，最内ループにおいて，メモリアクセスが連続であり，かつ計算が並行的に実施できるというのは，今のHPCシステムにはよい実装
    - アプリ側ではこのような実装を可能とする解法開発が必要