Top500, HPCGへの挑戦

計算科学研究機構　運用技術部門
ソフトウェア技術チーム　チームヘッド

南　一生

本日の趣旨

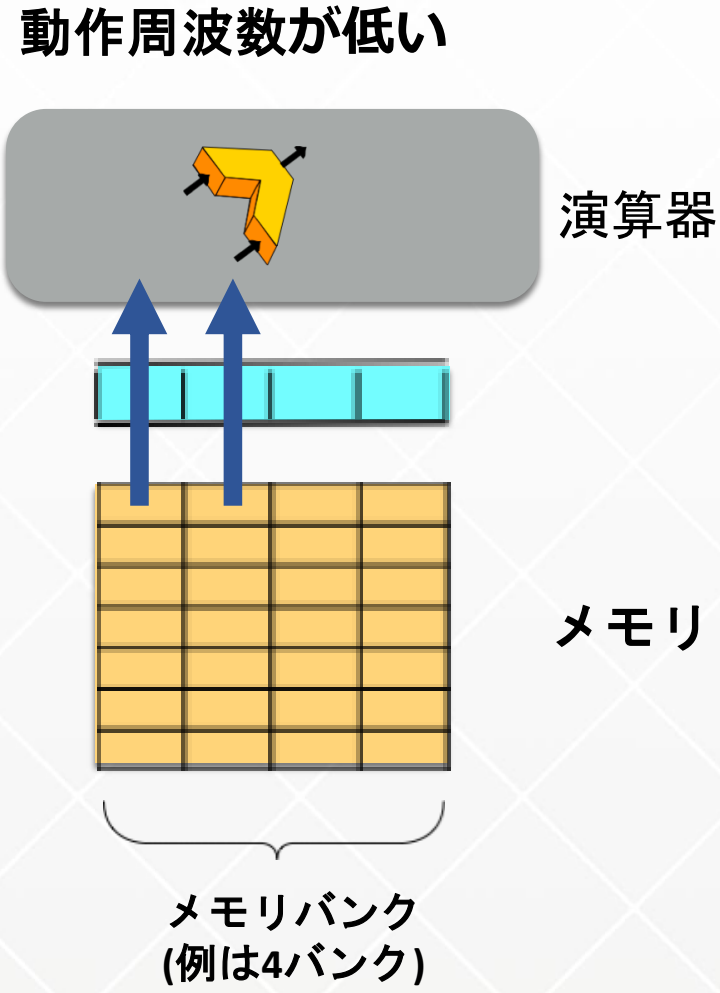
- 世界で最も高速なスパコンを評価するプロジェクトとしてTOP500がある.
- TOP500は，1993年に発足した.
- TOP500にはLINPACKというベンチマークプログラムが用いられている.
- LINPACKは，演算性能の評価に重きを置いているため，近年，実アプリケーションとの乖離が指摘されることもあった.
- そこで，実アプリケーションで使われる計算手法の性能評価に重きを置いたベンチマークプログラム:HPCGが提案されている.
- 「京」は，2011年にTOP500，1位，2014年にHPCG，2位を獲得した.
- 現代のスパコンの特徴から見たBMTの特性と，高い性能を引き出すベンチマークプログラムのチューニングへの挑戦について話す.

最初に
ーコンピュータの過去と現在ー

昔のコンピュータ

シングルプロセッサ

- メモリは一定時間を経過しないと同じメモリバンクにアクセスできない
- 昔は20サイクルくらい待っていた
- しかし動作周波数が低かったため演算器も遅く演算速度とメモリ転送性能は釣り合っていた



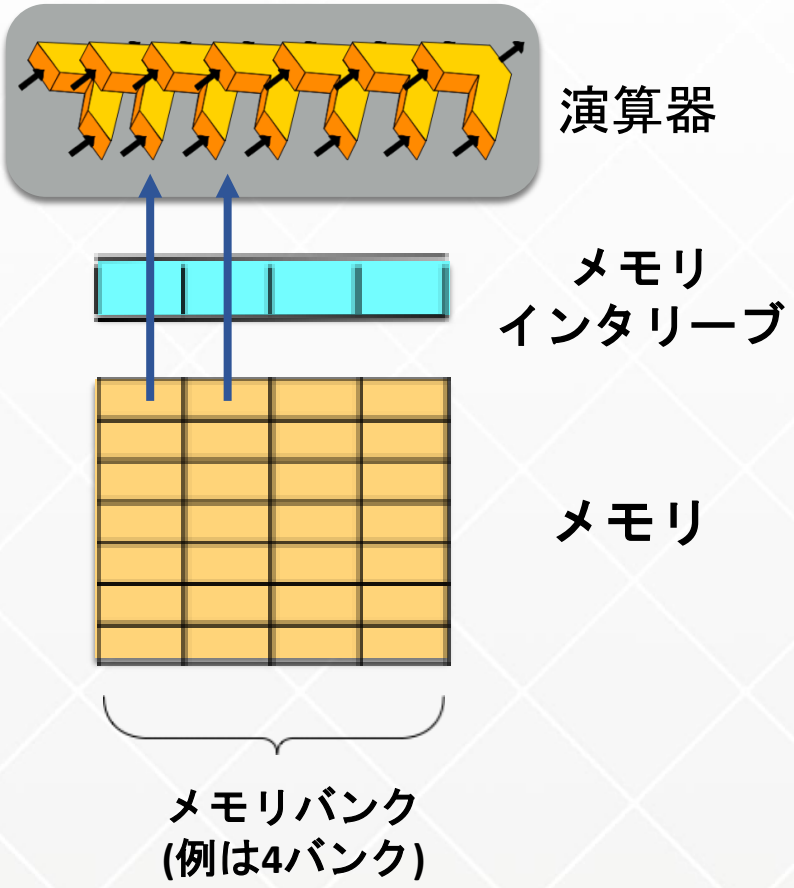
シングルプロセッサの問題

RIKEN AICS シミュレーションが未来をひらく

メモリウォール問題

- メモリは動作周波数が高くなってくるともっと待つ事となる(100–200サイクル)
- メモリに比べて演算器は動作周波数が高くなり高速になった
- さらに半導体プロセスの微細化により演算器はCPUにたくさん搭載可能となった
- 結果的に演算器に比べメモリのデータ転送能力が低くなった

動作周波数が高い場合(現在)

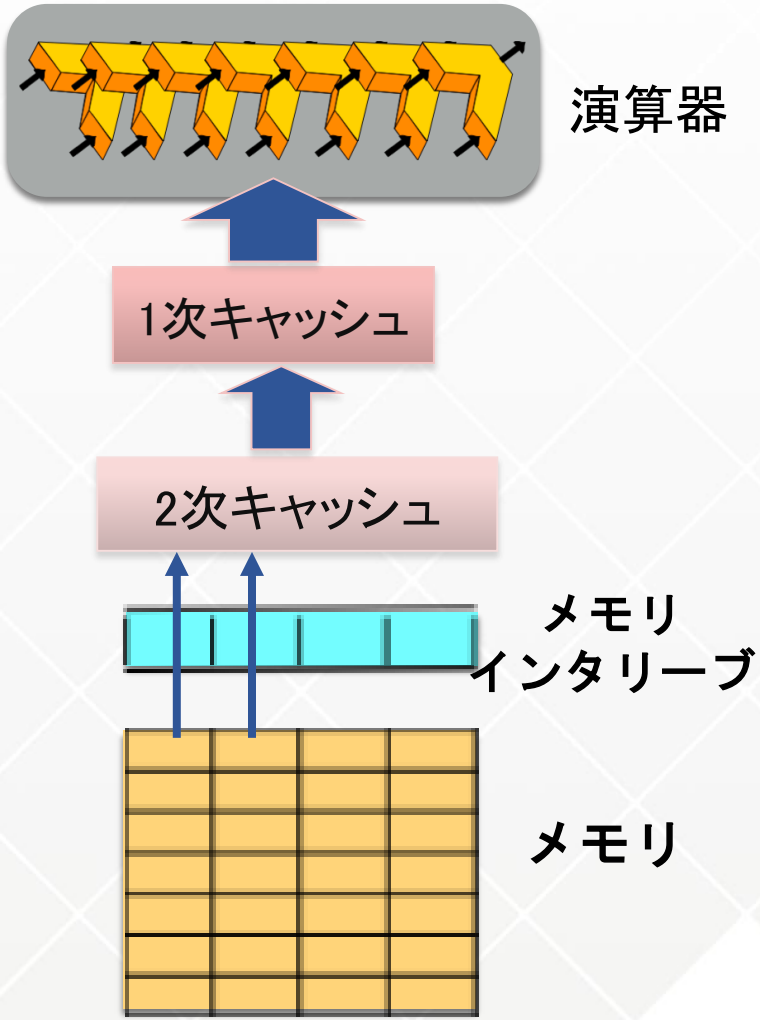


シングルプロセッサの問題

対策

- データ供給能力の高いキャッシュをメモリと演算器間に設ける
- データをなるべくキャッシュに置きデータを再利用する事でメモリのデータ供給能力の不足を補う
- こうすることで演算器の能力を使い切る
- 多くのスカラー機はこの方式を取っている

動作周波数が高い場合



シングルプロセッサの問題

一台のコンピュータの処理限界

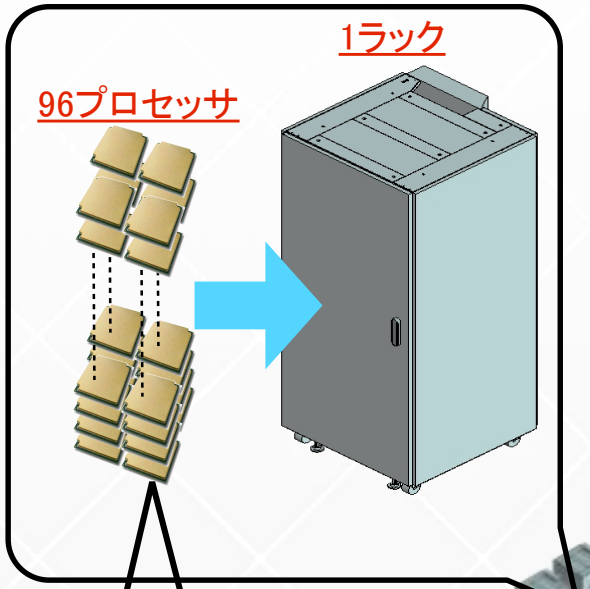
- 動作周波数を上げる事で電力が吹き上がる
- 動作周波数の限界を迎えつつある
- メモリウォール問題もあり一台のコンピュータの演算能力を上げてもメモリの能力が追いつかない

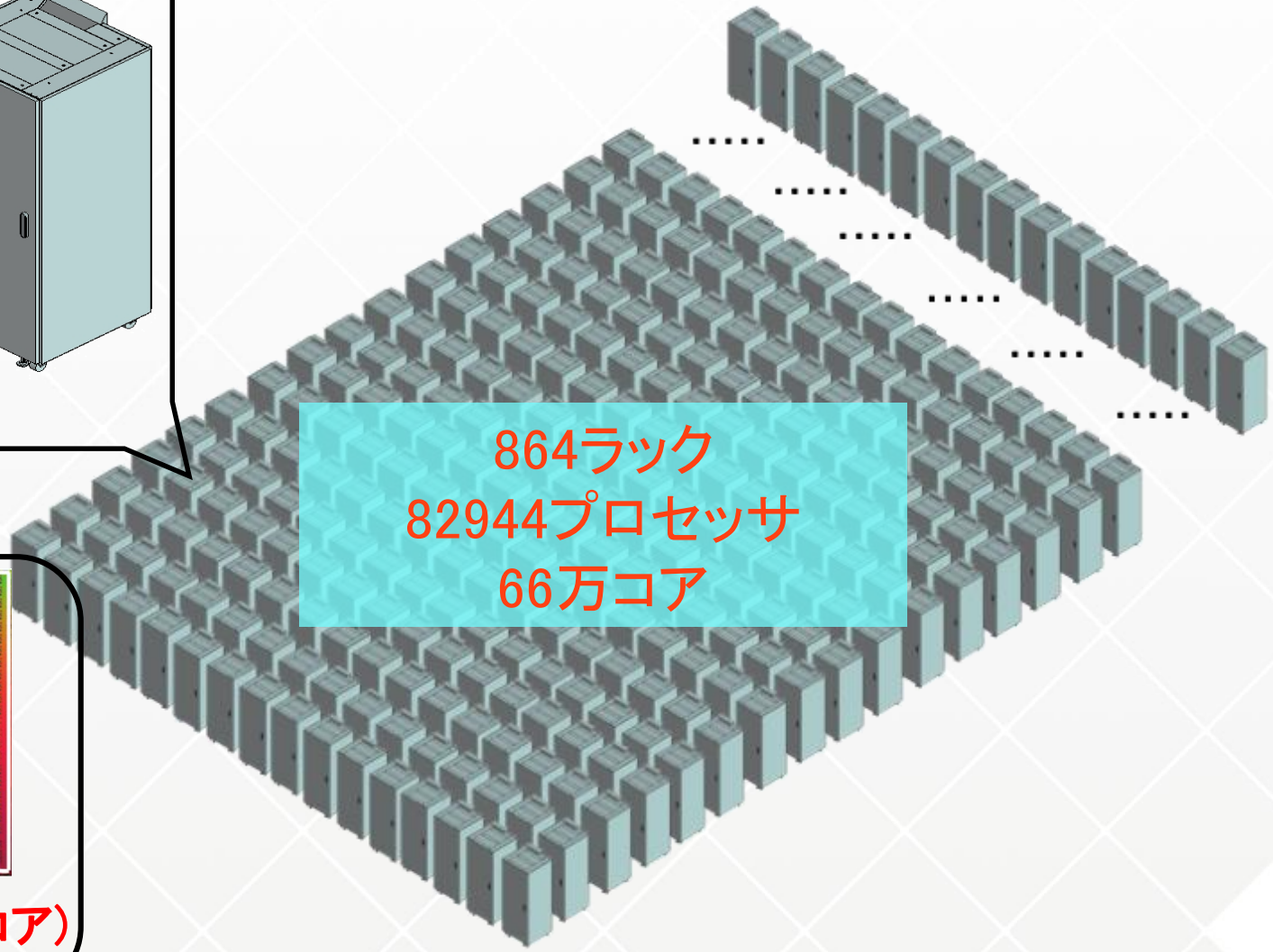
そこで！

そこで超並列処理！

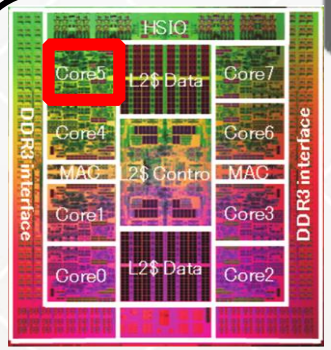
RIKEN AICS シミュレーションが未来をひらく

50m×60mの部屋に





864ラック
82944プロセッサ
66万コア



マルチコア(8コア)

次に
ー現代のスパコンについてー

現代のスパコン利用の難しさ

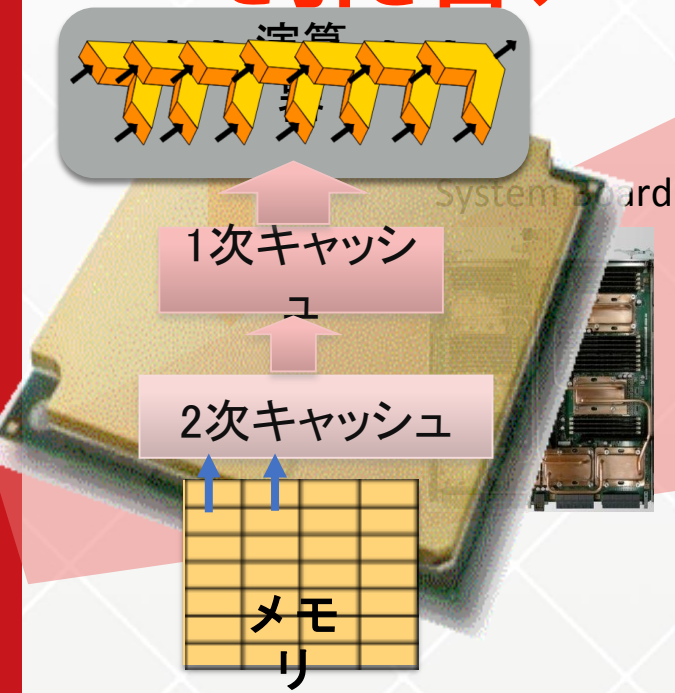
シミュレーションが未来をひらく

アプリケーションの超並列性を引き出す

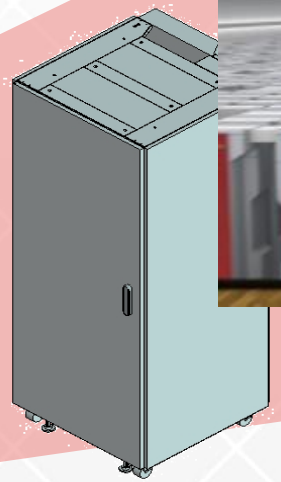
現代のスーパーコンピュータシステム

プロセッサの単体性能を引き出す

現代のプロセッサ



Rack

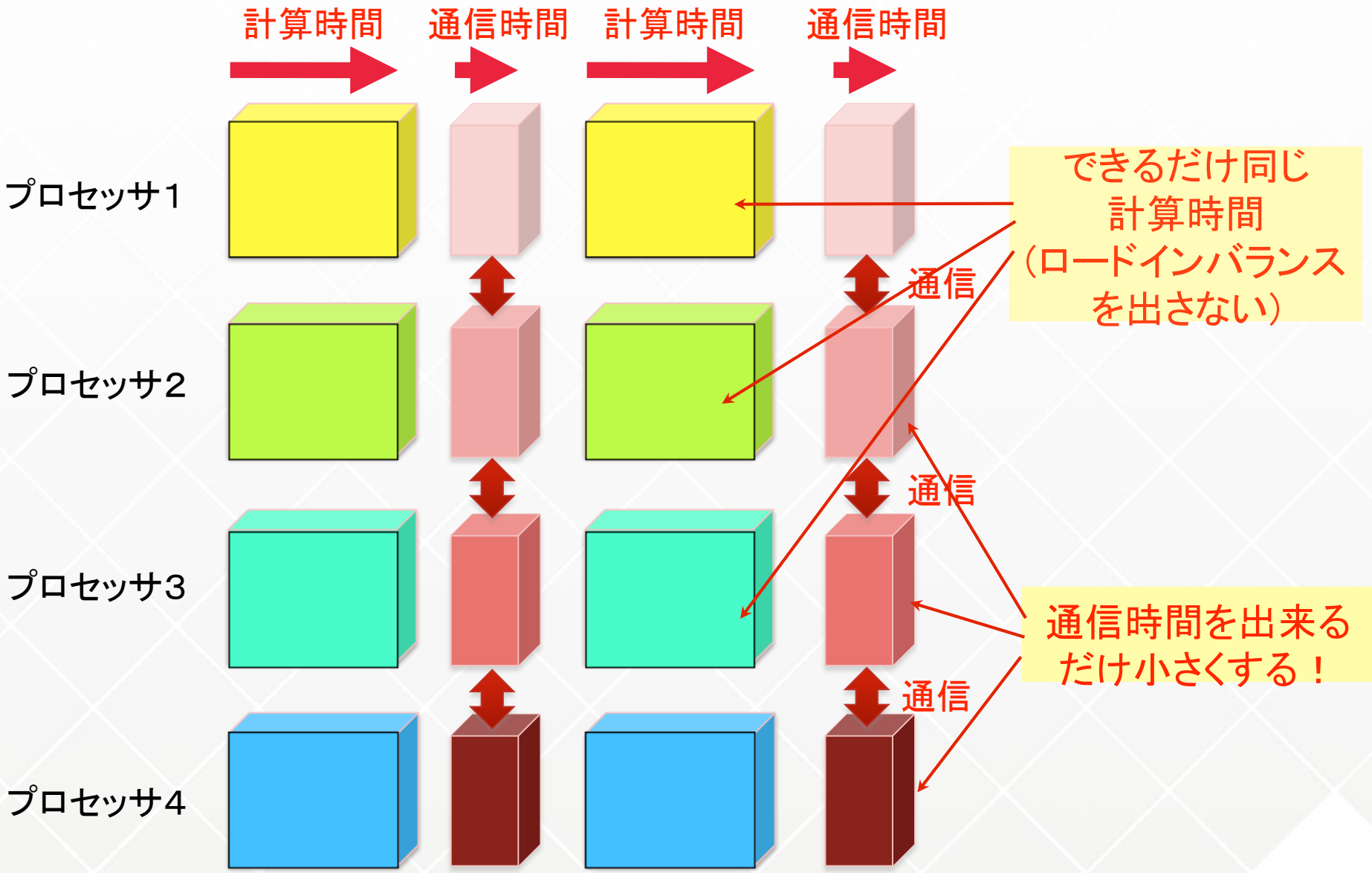


並列化とは？

逐次計算

並列計算

並列化とは？


計算時間
通信時間
計算時間
通信時間
プロセッサ1
プロセッサ2
プロセッサ3
プロセッサ4
通信
通信
通信
できるだけ同じ
計算時間
（ロードインバランス
を出さない）
通信時間を出来る
だけ小さくする！

並列化とは？

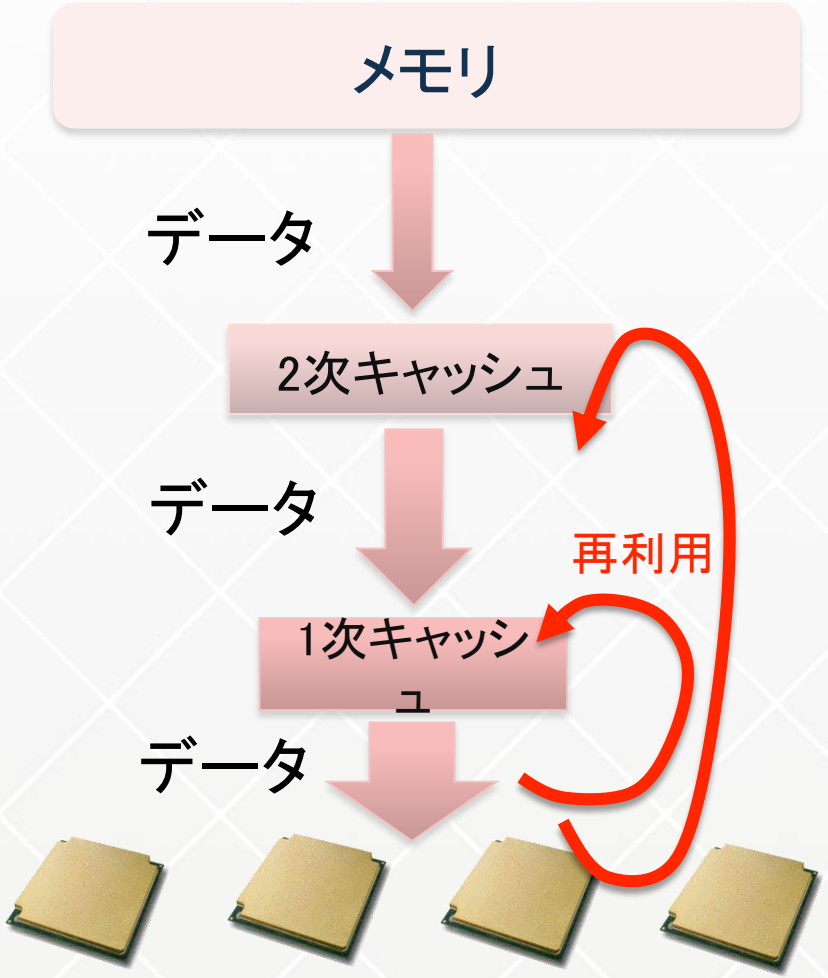
必要な並列度を確保する！

並列計算できない部分（非並列部）を限りなく小さくする！

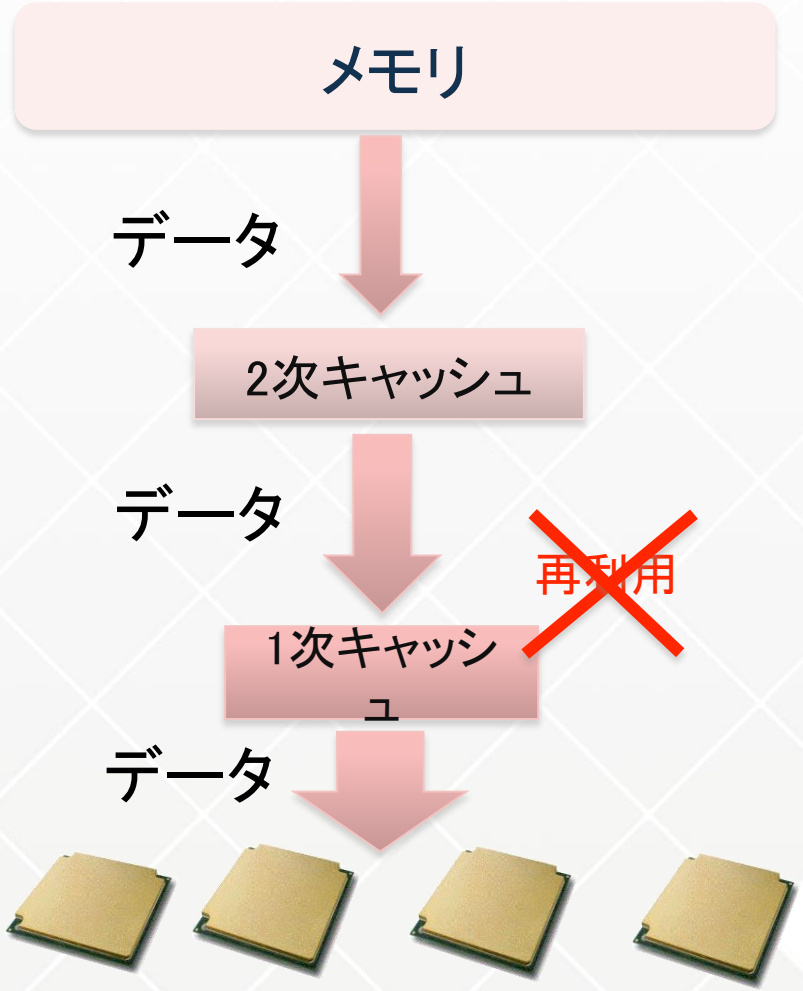
プロセッサの単体性能を引き出すとは？

マルチコア化 ＋ 以下を考慮したプログラミング

データの再利用可能なアプリ



データの再利用不可能なアプリ



LINPACKとHPCG

シミュレーションは・・・
連立1次方程式として解くことが多い

シミュレーションが未来をひらく

2元連立1次方程式

$$3x + 5y = 3$$
$$x + 4y = 8$$

書き換え

2元連立1次方程式の行列形式

$$\begin{pmatrix} 3 & 5 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 8 \end{pmatrix}$$

行列

解

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -4 \\ 3 \end{pmatrix}$$

$$\begin{pmatrix} a2 & a3 & 0 & 0 & 0 & 0 \\ b1 & b2 & b3 & 0 & 0 & 0 \\ 0 & c1 & c2 & c3 & 0 & 0 \\ 0 & 0 & d1 & d2 & d3 & 0 \\ 0 & 0 & 0 & e1 & e2 & e3 \\ 0 & 0 & 0 & 0 & f1 & f2 \end{pmatrix}\begin{pmatrix} u \\ v \\ w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} g1 \\ g2 \\ g3 \\ g4 \\ g5 \\ g6 \end{pmatrix}$$

行列

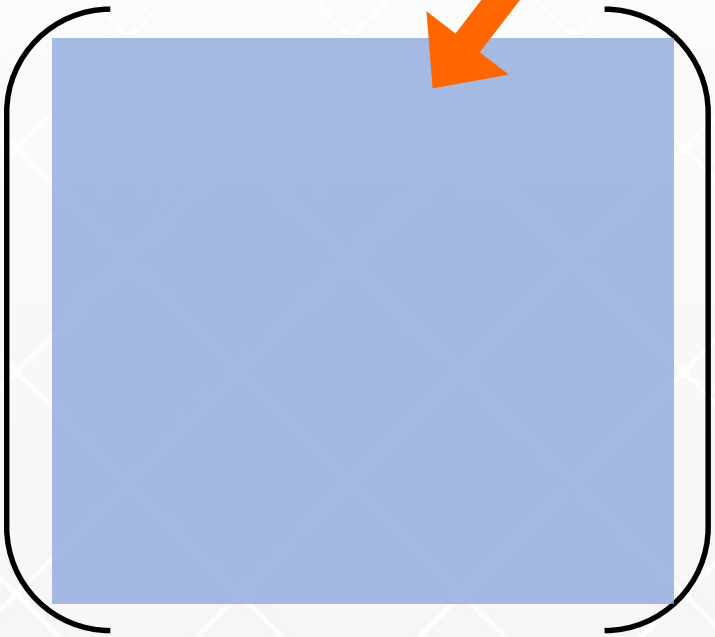
$$Ax = b$$

略してこう書く

LINPACKとHPCG

$$Ax = b$$

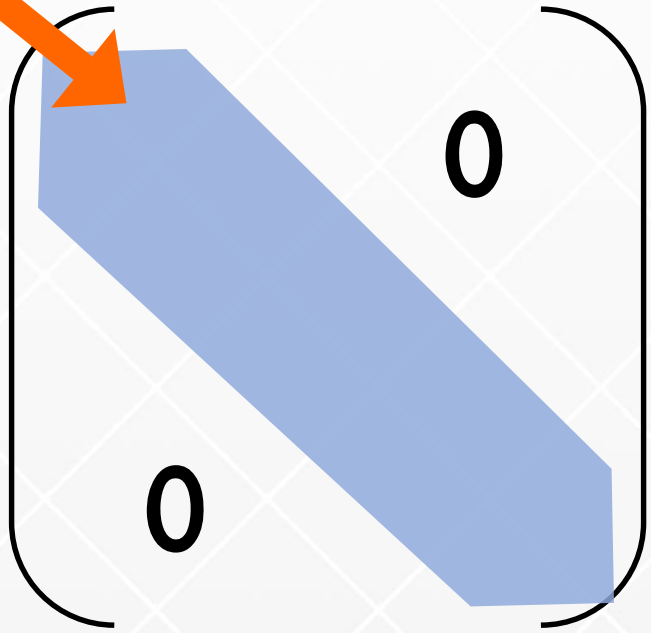
LINPACK
(密行列・直接解法)



精密な計算をする電子状態シミュレーション等の分野このタイプが多い

(*) 正確には連立一次方程式でなく固有値方程式となる

HPCG
(疎行列・反復解法)



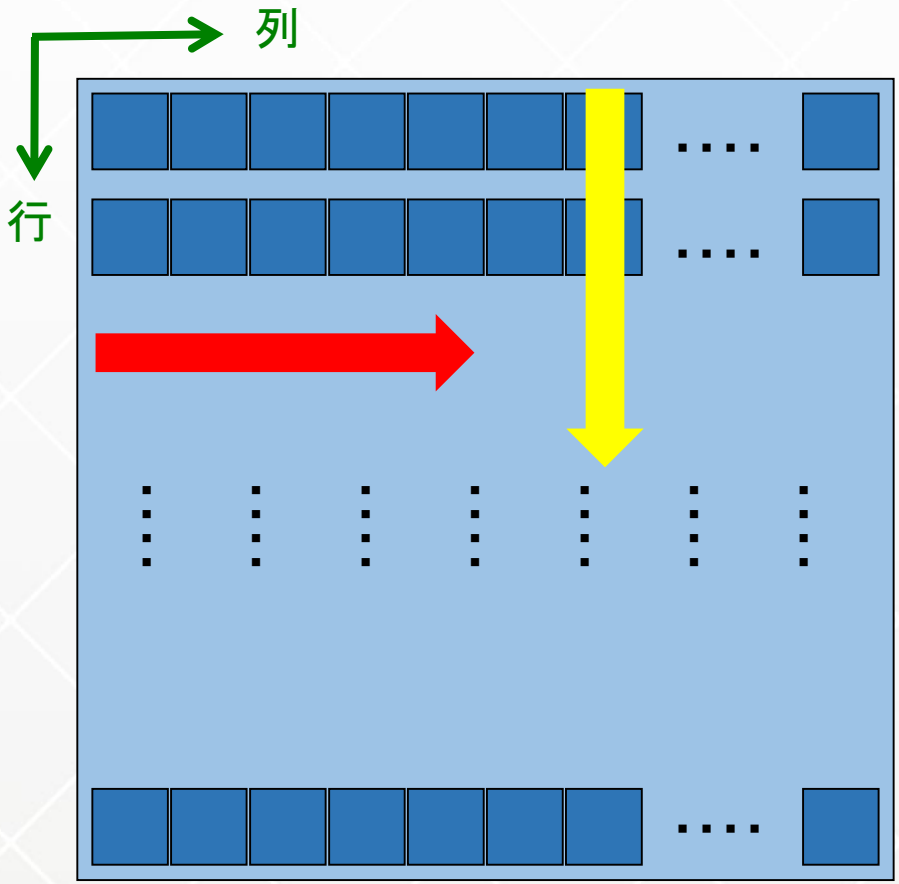
工学等の分野にこのタイプが多い

LINPACKとHPCGの難しい点

		LINPACK	HPCG
並列性能	・できるだけ同じ計算時間	○	○
	・通信時間を出来るだけ小さくする	×	○
	・必要な並列度を確保する	○	○
	・並列計算できない部分(非並列部)を限りなく小さくする	○	○
単体性能	・マルチコア化	○	×
	・データの再利用可能なアプリ	○	×

LINPACKのためのチャレンジ

- 行方向通信について
- 計算と通信が同時実行可. 計算と通信のオーバーラップ.
- 「京」独特の通信ハードウエア(Tofu)
- Tofuの特徴を生かした徹底的に速い通信ライブラリ開発



- 列方向通信について
- 計算と通信が同時に実行できない
- 行方向の通信を早くしたい
- Tofuはデータの配置を6次元に自由に配置可能
- 行方向通信が早くなるようデータ配置を工夫

（プロセッサ）

- プロセッサ内計算について
- 元々プロセッサの性能は出やすいが！
- プロセッサ性能限界まで高速化！

（全体）

- できるだけ大規模な問題を解いて性能を出す
- 大規模化して長時間実行となる
- 高い信頼性により長時間実行が可能

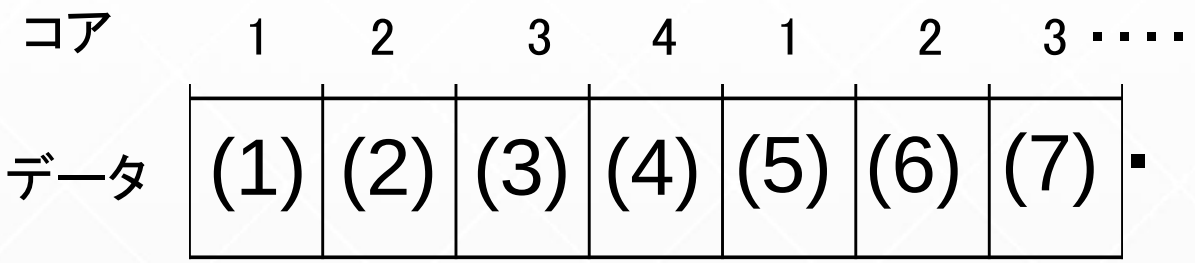
LINPACKの結果



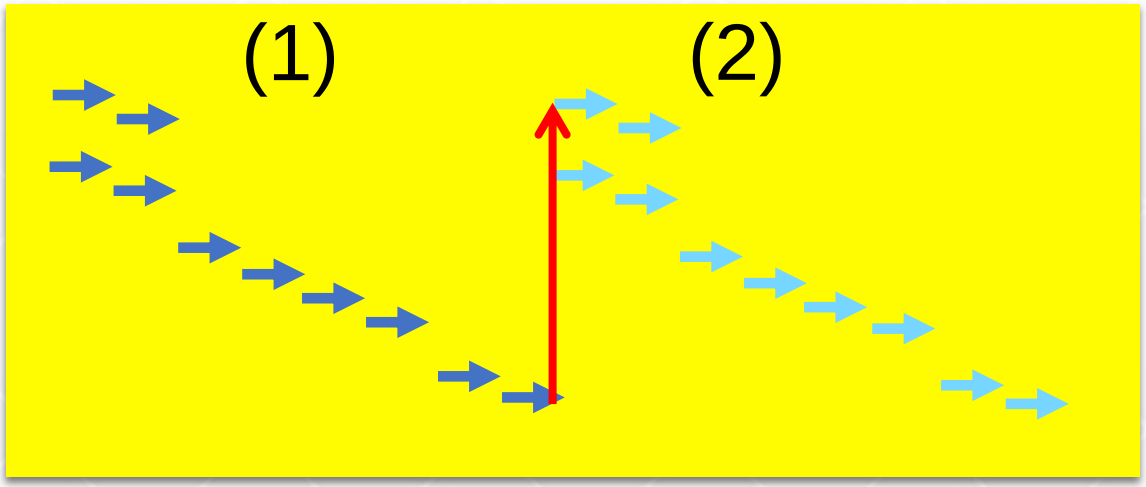

	Tianhe-2	Titan	Sequoia	K computer
latest TOP500 ranking (entry)	No.1 (Jun. 2013)	No.2 (Nov. 2012)	No.3 (Jun. 2012)	No.4 (Jun. 2011)
LINPACK Score (PFLOPS)	33.86	17.59	17.17	10.51
LINPACK efficiency (%)	61.67	64.8	85.3	93.2

HPCGのためのチャレンジ

マルチコア化




```
do i=1,100
      a(i)のロード
      b(i)のロード
      a(i)とb(i)の演算
      i番目の結果のストア
end do
```

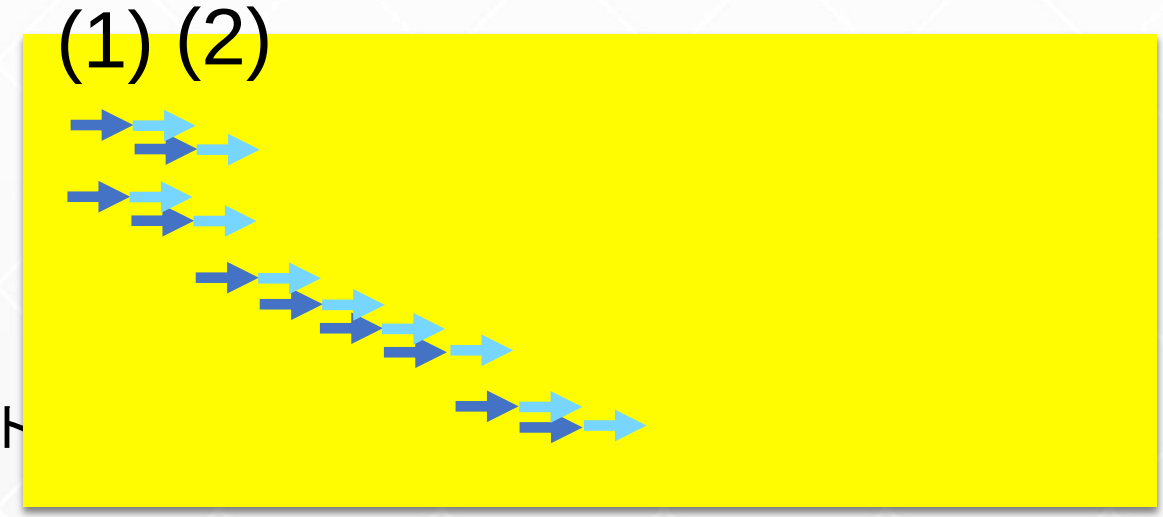


- マルチコア化は複数のコアを同時に使うこと
- HPCGのオリジナルプログラムは(1)の後に(2)の計算を行う必要あり
- したがって複数コアを同時に使えない

HPCGのためのチャレンジ

マルチコア化

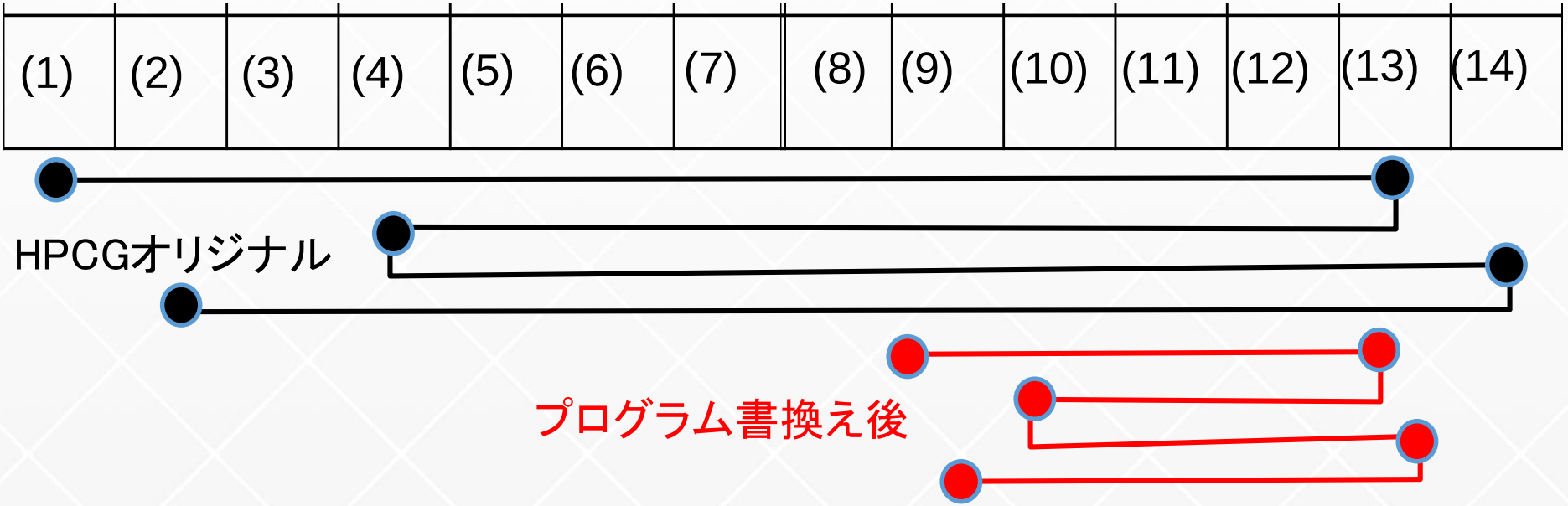
```
do i=1,100
      a(i)のロード
      b(i)のロード
      a(i)とb(i)の演算
      i番目の結果のスト
end do
```



- プログラムを書換え(1)と同時に(2)の計算をできるようにした
- 同時に複数コア使えるようにした
- 上図のように実行時間を大幅に短縮

HPCGのためのチャレンジ

データの再利用性が少ない場合の効率的なデータアクセスの実現



- データの再利用性が少ない
- しかも飛び飛びのアクセスとなっている
- このような飛び飛びのアクセスは非常に効率が悪い
- 飛び飛びの度合いを大幅に縮小するようにプログラムを書換え

HPCGの結果

Results: June 2015

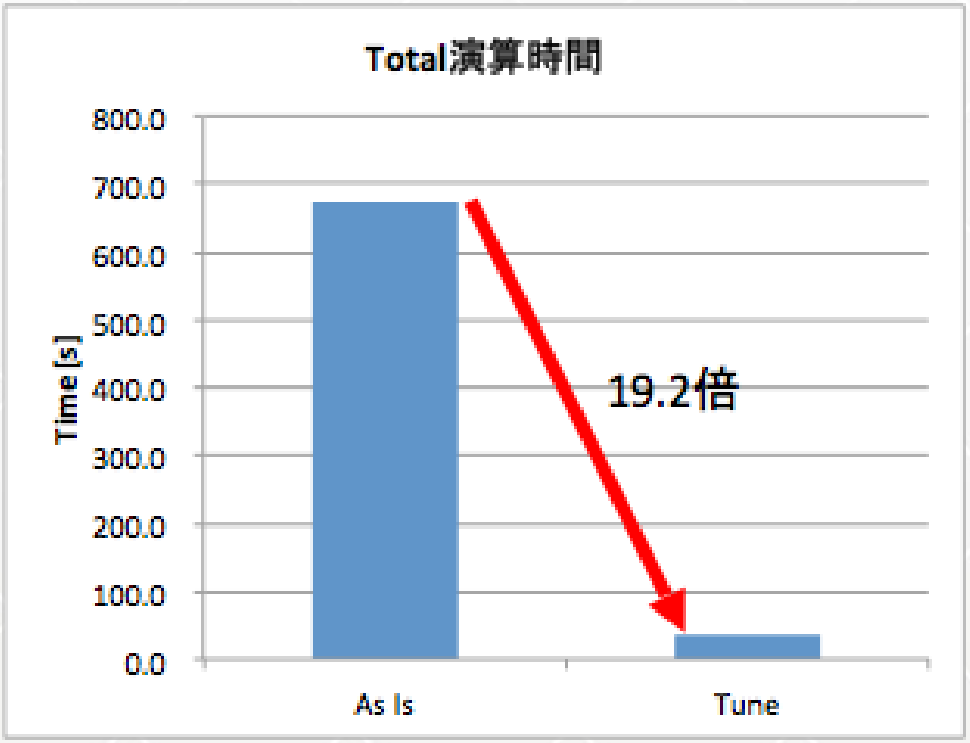
June 2015 HPCG Results

June 2015 HPCG Results

Site	Computer	Cores	HPL Rmax [Pflop/s]	TOP500 Rank	HPCG [Pflop/s]	HPCG/HPL
NSCC / Guangzhou	Tianhe-2 NUDT, Xeon 12C 2.2GHz + Intel Xeon Phi 57C + Custom	3120000	33.863	1	0.5800	1.7%
RIKEN Advanced Institute for Computational Science	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect	705024	10.510	4	0.4608	4.4%
DOE/SC/Oak Ridge Nat Lab	Titan - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x	560640	17.590	2	0.3223	1.8%
DOE/SC/Argonne National Laboratory	Mira - BlueGene/Q, Power BQC 16C 1.60GHz, Custom	786432	8.587	5	0.1670	1.9%

HPCGの結果

- HPCGオリジナルに比べて19.2倍の高速化
- １位の天河２はピーク性能で京の5倍程度
- しかしHPCGの性能差は1.3倍しかない
- HPCG/HPL性能比：4.1%は群を抜いている



ご清聴ありがとうございました