

大島聡史（東京大学 情報基盤センター 助教／並列計算分科会主査）

最新の並列計算事情とCAE

アウトライン

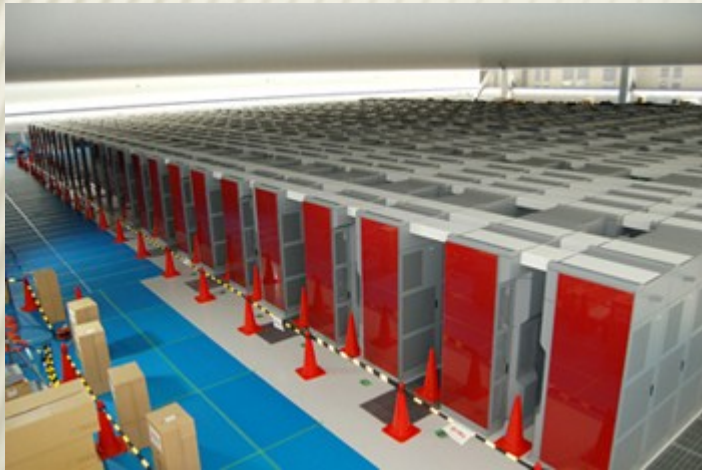
- × 最新の**並列計算機**事情とCAE
 - + 世界一の性能を達成した「京」について
 - + マルチコア・メニーコア・GPU・クラスタ……
- × 最新の**並列計算**事情とCAE
 - + MPI、OpenMP、CUDA、OpenCL、etc.

※「京」については、仕分けやら予算やら計画やらの面で問題視する意見もあるかと思いますが、なるべく技術的な視点でお話しさせていただきますのでご勘弁願います

最新の並列計算機事情

× 「京」

- + 先日発表されたTOP500ランキングにて一位を達成
- + 世界一のスーパーコンピュータ なのか？



「京」実物の写真
(理研のwebページから)

No.1の証書を渡された渡辺氏と佐相氏
(ドイツ、ハンブルクのISC2011にて撮影)



TOP500

- × スーパーコンピュータのランキング、半年ごとに更新
 - + 6月のISCと11月のSCにあわせて発表される
 - × 旧地球シミュレータが連続5期も君臨したためにアメリカが本気になったことは有名
- × 指標はLINPACKベンチマークの性能
 - + 密行列計算、主に行列積の性能が影響する
 - + 主にCPU性能（浮動小数点演算性能）が反映される
- × すなわち……

TOP500で一位（上位）にランキングされても
どんな計算でも高速にできるわけではない

- + CPUが強力であることはわかる
- + システムが安定して動くことがわかる（こともある）
- + ストレージが重要な場合は？
- + 整数演算性能は？
- + 疎行列系（キャッシュが利かない）は

「京」のNo.1は喜ばしいが、
それが全てではないことを
知っておいて欲しい

TOP500 List - June 2011 (1-10)

Rank	Site	Computer/Year Vendor	Cores	R _{max}	R _{peak}	Power
1	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect / 2011 Fujitsu	548352	8162.00	8773.63	9898.56
2	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C / 2010 NUDT	186368	2566.00	4701.00	4040.00
3	DOE/SC/Oak Ridge National Laboratory United States	Jaguar - Cray XT5-HE Opteron 6-core 2.6 GHz / 2009 Cray Inc.	224162	1759.00	2331.00	6950.60
4	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade, Intel X5650, NVIDIA Tesla C2050 GPU / 2010 Dawning	120640	1271.00	2984.30	2580.00
5	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows / 2010 NEC/HP	73278	1192.00	2287.63	1398.61
6	DOE/NNSA/LANL/SNL United States	Cielo - Cray XE6 8-core 2.4 GHz / 2011 Cray Inc.	142272	1110.00	1365.81	3980.00
7	NASA/Ames Research Center/NAS United States	Pleiades - SGI Altix ICE 8200EX/8400EX, Xeon HT QC 3.0/Xeon 5570/5670 2.93 Ghz, Infiniband / 2011 SGI	111104	1088.00	1315.33	4102.00
8	DOE/SC/LBNL/NERSC United States	Hopper - Cray XE6 12-core 2.1 GHz / 2010 Cray Inc.	153408	1054.00	1288.63	2910.00
9	Commissariat a l'Energie Atomique (CEA) France	Tera-100 - Bull bullx super-node S6010/S6030 / 2010 Bull SA	138368	1050.00	1254.55	4590.00
10	DOE/NNSA/LANL United States	Roadrunner - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband / 2009 IBM	122400	1042.00	1375.78	2345.50

「京」の概要

<http://www.nsc.riken.jp/project/spec.html> に基づき作成

× 全体構成

- + CPU数：8万以上
- + CPUコア数：64万以上
- + ピーク演算性能：10PFLOPS以上
 - × 全てのCPUを同時に動かして得られる理論上の最大性能
- + メモリ総容量：1PB以上（ノードあたり16GB）

× プロセッサ構成

- + 8コア SPARC64VIIIfx
 - × 富士通がカスタマイズしたSPARCプロセッサ
 - × $16\text{GFLOPS} \times 8\text{コア} = 128\text{GFLOPS}$

× プログラミング環境

- + Fortran, XPFortran, C/C++（少なくとも）
- + 主な想定はノード内自動並列化＋ノード間MPI（のはず）

疑問

- × 私たちも「京」を使えますか？
 - + （使えるルートはありますが）難しいでしょう
 - + 主なユーザは戦略5分野のメンバーとなります
 - × 分野1 予測する生命科学・医療および創薬基盤
 - × 分野2 新物質・エネルギー創成
 - × 分野3 防災・減災に資する地球変動予測
 - × 分野4 次世代ものづくり
 - × 分野5 物質と宇宙の起源と構造
- × もし「京」を使えたら手持ちのアプリケーションは速く動きますか？
 - + 一部の例外はありますが、困難です
 - + もちろん、単体プロセッサ（単体ノード）性能分の差は出ますが、数千・数万・数十万コアを活用するにはプログラムの作り直しが必要になるでしょう
 - × 並列計算の知識と技術が必要！

（「京」に特化した話はここまでにして、
ここからはもう少し広い話や身近な話をします）

最近のスーパーコンピュータ事情

(まずはハードウェア側の事情から)

× キーワード

+ ~~ベクトル~~計算機

+ PCクラスタ

+ マルチコアCPU、マルチソケットCPU

+ アクセラレータ

+ メニーコア

× 特にマルチコアCPUとアクセラレータについては 個人所有のPCレベルでも容易に利用可能

+ むしろ、積極的に活用する必要がある

スライド一枚でわかる (かもしれない) スパコンのトレンド

× 10年（15年？）前

- + 専用設計のベクトル計算機、強力なベクトルプロセッサを高速バスで接続、非常に「密」なシステム

× 近年

- + 汎用のPCに近いシステム（ノード）を大量にネットワーク接続した「疎」なシステム、PCクラスタ
 - × Xeon、Opteron、DDR SDRAM、HDD/SSD、InfiniBand
 - × 少ノードシステムの販売
 - ★ 流石に、マザーボードや筐体は専用品……？
- + Cell B.E.やGPUといった演算加速器「アクセラレータ」を搭載したヘテロジニアス（不均質）構成
 - × Roadrunner、Nebulae、Tianhe、Tsubame
- + スーパーコンピュータと汎用PCの距離は（ある意味）小さい
 - × 規模が圧倒的に違うのがポイント

マルチコアCPUとマルチソケット

× マルチコアCPU

+ 複数の「CPUコア」を搭載したCPU

× デュアルコアCPU、クアッドコアCPU、etc.

+ 一般に売られているPCの多くはマルチコアCPUを搭載

× ネットブックなど一部の激安PCの一部は除く



× マルチソケット

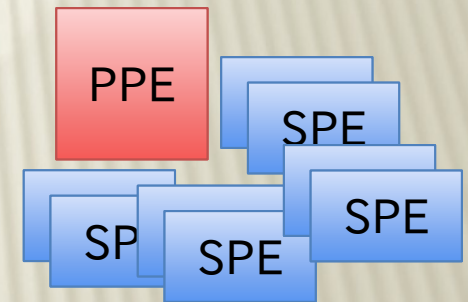
+ 複数個のCPUを1ノードに搭載

+ 例：T2Kスパコン東大版（1ノードあたり4つの4コアCPU＝1ノードあたり16コア）

+ 一般に売られているPCはシングルソケット

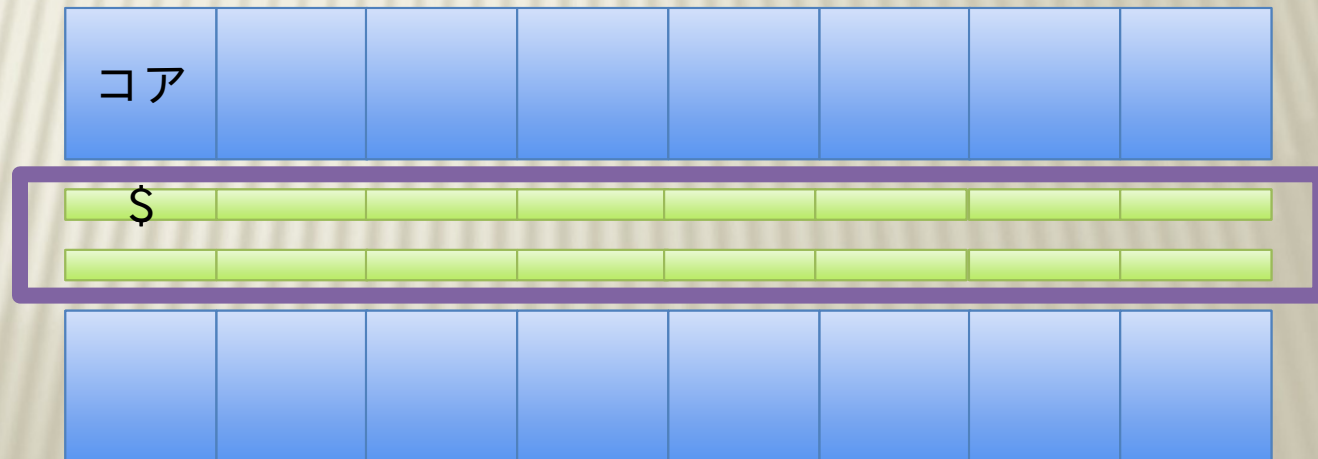
アクセラレータ (Cell B.E.とGPU)

- × 既存のCPUとは異なるアーキテクチャを持ち、特に高い並列計算性能を目指すハードウェア
- × Cell B.E.
 - + PS3に搭載されているプロセッサ（もしくはそれをカスタマイズしたもの）
 - + PowerアーキテクチャのマルチコアCPU + 複数のシンプルな計算コア
- × GPU (Graphics Processing Unit)
 - + 画像処理用のハードウェア、いわゆるビデオカード
 - + 多数のシンプルな計算コアのグループを多数搭載
 - + PCショップで売られているGPUの多くが利用可能
 - × PCショップに1TFLOPSの演算器が並んで売られている時代



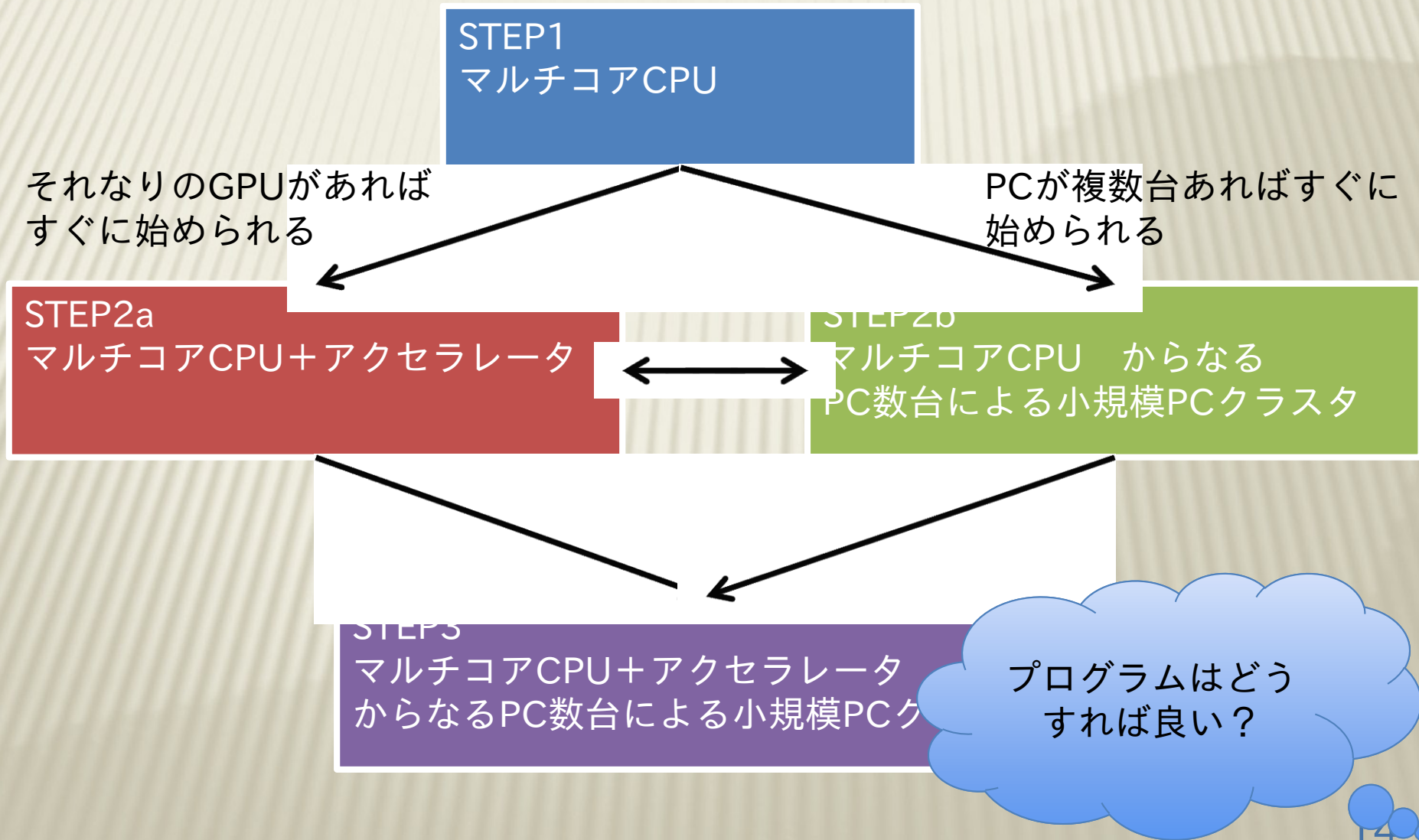
メニーコア

- × 多数のシンプルなコアを搭載
 - + Intel の Many Integrated Core (MIC)
 - × 躓いたLarrabee……
 - + まだ製品化されていない
 - + Pentium程度のコア*を多数搭載
 - + 2012年に製品化？



* Cell B.E.のSPEやGPUの計算コアよりも複雑、各コアでOSが動くレベル

今すぐにでも始められる並列計算



最新の並列計算事情（ややC言語よりの視点から）

- × 並列計算を行うためのプログラムを作る方法
- × 従来の並列化プログラミング環境
 - + 自動並列化
 - + OpenMP（ノード内専用）
 - + MPI（ノード内・ノード間兼用）
- × 最近のプログラミング環境
 - + GPU向け：CUDA、OpenCL
 - + CPU向け：High Performance Fortran (HPF)、Threading Building Block (TBB)、Unified Parallel C (UPC)、X10、Chapel、XcalableMP (XMP)、etc.
- × スーパーコンピュータ
 - + 従来：MPI
 - + 最近：MPI+OpenMP、「最近のプログラミング環境」
 - × コア数が増えすぎてMPIのみでは無理が生じてきた

CAEのためには何を学ぶ（使う）べきか 1/2

- × 実は（プロダクト、オープンソース等の種別を問わず）多くのソフトウェア・ライブラリは並列化されている（されつつある）
 - + 最近はGPUに対応したソフトウェアも増加中
 - + 既製品で事足りるならそれで良い
 - × 本当に進めたいサイエンスを進められることはとても重要
 - × サイエンスに集中できる vs 選択肢が増える
- × 何故並列化を学ぶ必要があるのか
 - + 対象問題と実行環境とを適合させるため（学習する）
 - × どの並列化済みソフトウェアを使うのがベストか
 - + 既製品では不足があるため（自前で実装する）
 - × 並列化済みソフトウェアに与えるデータの作成に時間がかかる
 - × 自分のプログラム向けにアルゴリズムを改良してある

CAEのためには何を学ぶ（使う）べきか 2/2

× 並列化が必要な場合は何を学ぶべきか（一例）

+ OpenMPとMPI

- × C/C++でもFortranでも使える（言語自体から学ぶ必要がない）
- × 分散メモリと共有メモリを学べる
- × 資料が非常に多い

+ GPUを使いたい場合にはCUDAかOpenCL

- × どちらか覚えればもう一方も使いやすい
- × OpenMPとMPIを知っていた方が理解しやすい

+ C/C++やFortran以外の言語を使っているユーザは??

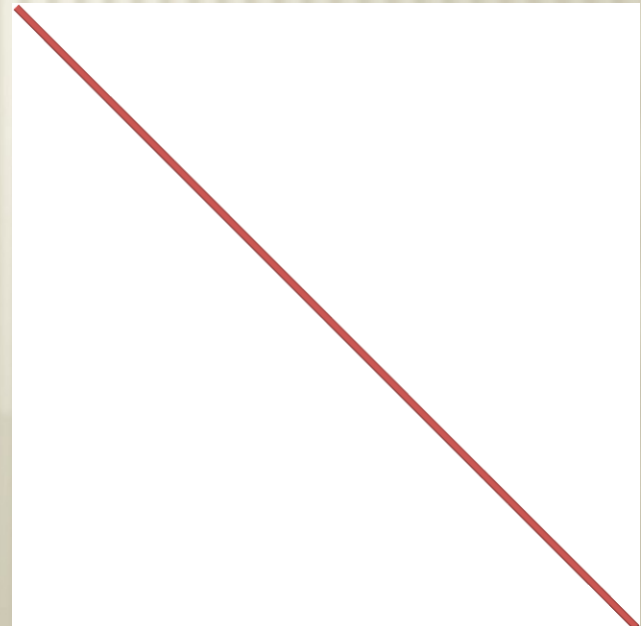
- × スクリプト・インタプリタからCUDAを利用する環境など

自前で実装する例の紹介

- × 有限要素法におけるCGソルバーのGPU実装
 - + 有限要素法：偏微分方程式の数値解法の1つ、連続体力学の広い範囲で使用
 - + CGソルバー：反復計算、疎行列ベクトル積・ベクトルの積や和や内積
 - + 疎行列ベクトル積(SpMV)が実行時間の大部分を占めるため、これを高速化することが重要
 - + GPUを用いたSpMVに関するライブラリは既にある（公開されている）

自前で実装する意義と内容

- × CPU向けに最適化された有限要素法プログラムが存在
 - + 行列を3x3ブロック化しており、対角・上三角・下三角に分けて保持している
- × CPU向けプログラムで行っている最適化を活用するには自前での実装が必要



性能評価

× 実行環境と設定

- + Xeon W3520 (2.67GHz Nehalem 4コア)
- + gcc v4.4 (-O3オプション)
- + 512,000要素、3x3ブロックLU前処理、double型

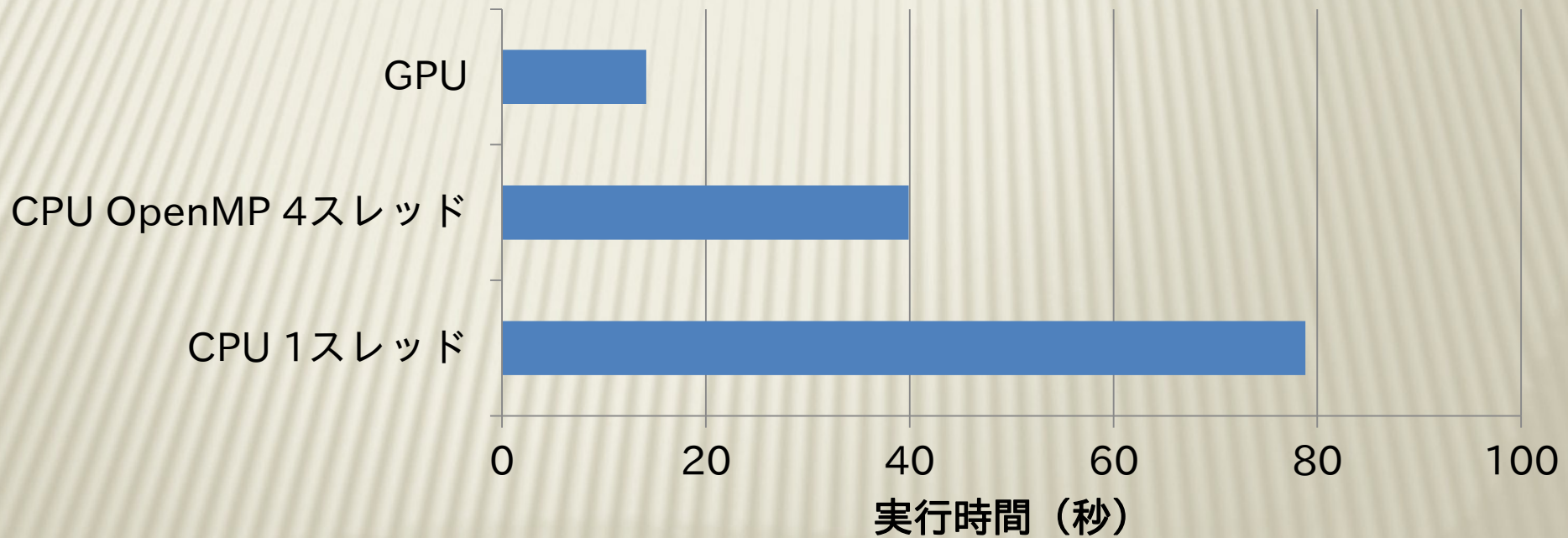
× 測定結果

- + SpMV on CPU(OpenMP 4スレッド)
 - × ブロック化有り 71 msec / iter (3.30GFLOPS)
 - × ブロック化無し 96 msec / iter (2.24GFLOPS)
- + SpMV on GPU
 - × ブロック化有り 20 msec / iter (11.73GFLOPS)
 - × ブロック化無し 52 msec / iter (4.51GFLOPS)

(↑既存のライブラリと同程度の性能)

成果

× CG法の実行時間比較



まとめ

- × 最新の並列計算機事情・並列計算事情をCAEと関連づけて紹介した
- × 並列計算機・並列計算について知ることはCAEの活用に役立つ（**と思います**）
 - + もちろん、全てを知ることとは不可能であり不要である
 - + 興味と需要と時勢にあわせて選択すればよい

連絡先（質問その他お気軽にどうぞ）
ohshima@cc.u-tokyo.ac.jp