バイオスーパーコンピューティングの更なる発展のために

バイオ・スーパーコンピューティング研究会
会長　姫野龍太郎

なぜこの研究会が必要か？

- Biosupercomputing: 生物学、医学、生理学、薬学、工学、情報処理、統計学、計算科学んどの複合領域
- 専門化されていない、それぞれの分野ではマイノリティー
- 情報を共有する事で、それぞれにメリット
- ライフ・グランドチャレンジ（京の開発プロジェクト）や戦略分野１予測する生命科学（HPCI)などのMEXTプロジェクトからの一定の距離を持つ研究者の自由な集まり
- 共通したメッセージを発信する事で政策に反映

この研究会では

- 研究会会員間:それぞれの領域間の問題:ミクロからメゾへ、メゾからマクロへ、それぞれの計算方法とコンピュータでの高速化・並列化、開発された方法の活用、使い方の共有
- 研究者から社会へ:現状や成果を社会に理解してもらう情報発信
- 研究者から政策へ:領域に共通するニーズを抽出、提言等を発信

本研究会の活動の方針

- 研究会会費は取らない
- 研究会主催のイベントで収益を上げて、会の運営を行う
  - ボランティアベース
  - 会誌なし

- 会員数を増やして、情報発信力を強化、会員のためのイベントを継続して企画

設立趣旨

目的

バイオスーパーコンピューティング研究会 (BSCRC)は、スーパーコンピュータを用いた生命科学分野でのシミュレーションや大規模データ解析を通して、生命体における現象を統合的に理解し予測 を行う新たな科学領域の進展を支え、育成・発展させて、わが国における学問・文化・産業の発展に寄与する。

この目的のため、個体、器官、組織、細胞、分子の様々なレベルにおける研究およびそれを統合する研究を対象として、 医学、細胞生物学、分子生物学、 生化学、生物物理学、薬学、化学、物理学、工学、情報科学等さまざまな分野の研究者が集い、産学官にわたる研究者間の交流を図る場とするとともに、国際的 な連携や情報発信を推進する。(研究会会則第2条より)

沿革と活動(1)

- 2008年12月から発足準備
- 2009年7月1日発足、初代会長・中村春木
- 2009年10月第一回総会
- 2009年11月行政刷新会議の事業仕分け（京の開発プロジェクト）に対して仕分け評価意見書提出
- 2010年4月2nd Biosupercomputing Symposium共催
- 2010年5月生命動態システム科学～生命の動的理解・予測・制御を目指して～を共催

沿革と活動(2)

- 2010年5月生命動態システム科学～生命の動的理解・予測・制御を目指して～を共催
- 2010年スーパーコンピューティング・セミナー（スパコン産業応用協議会主催）を後援
- 2010年10月第二回総会・講演会
- 2011年2月阪大タンパク研セミナー共催
- 2011年8月サマースクール2011主催
- 2011年10月第3回総会・講演会
- 2011年11月「次世代生命体統合シミュレーションソフトウェアの研究開発 成果報告会」(ISLiM成果報告会2011)に協賛

沿革と活動(3)

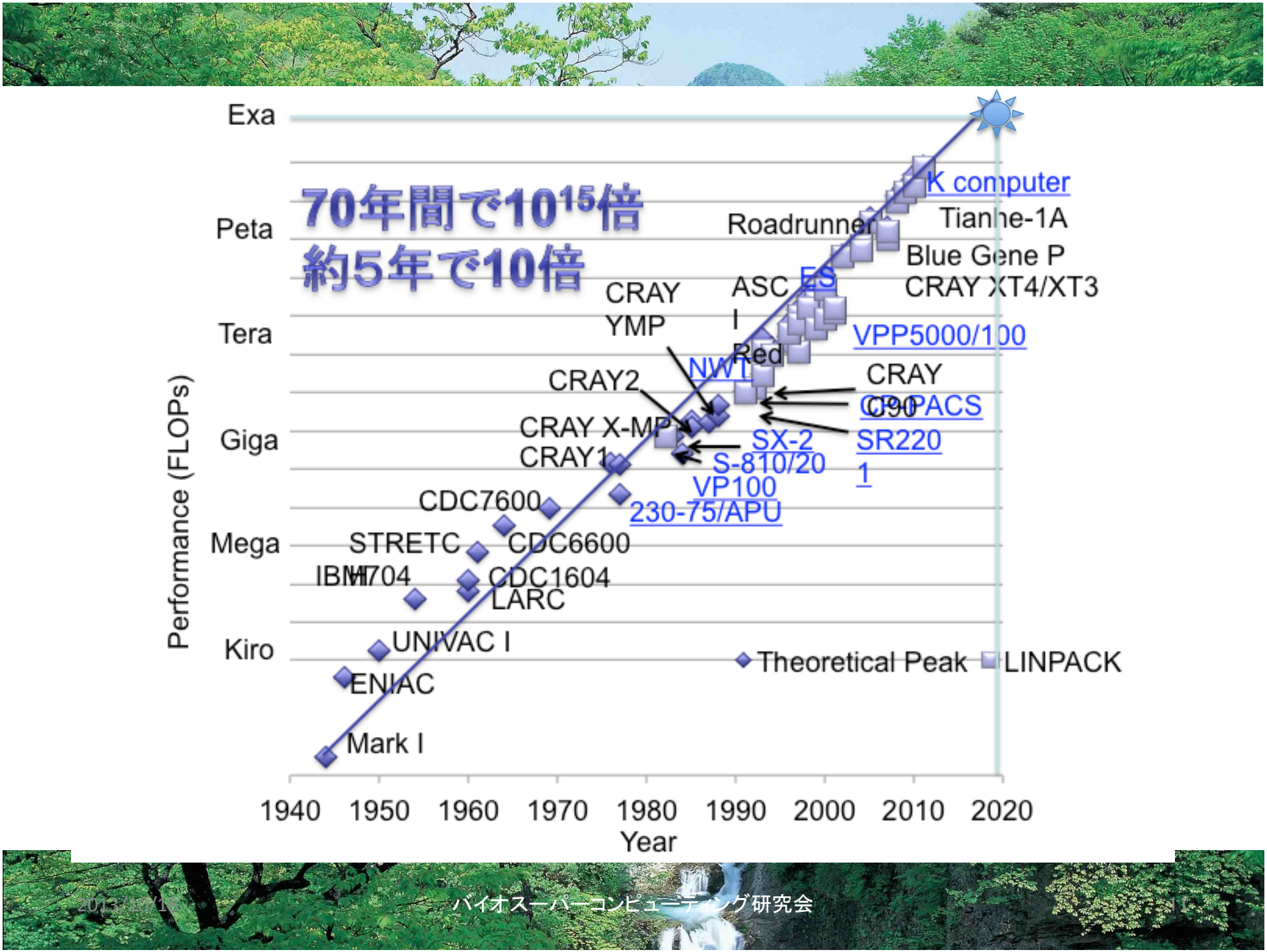
- 2012年２月ウィンター・スクール開催
- 2012年10月第４回総会・講演会
- 2011年11月「次世代生命体統合シミュレーションソフトウェアの研究開発 成果報告会」(ISLiM成果報告会2011)協賛
- 2013年10月第５回総会・講演会
- 2014年1月ウィンタースクール2014を開催
- 各種ソフトウェアの講習会を共催、「だれにでもわかる拡張サンプリングシミュレーション－長時間分子動力学シミュレーションへの挑戦－」、「やさしいcppmd (超並列高速MD計算コア)ソフトの紹介と利用方法の詳細」、「CafeMol」粗視化分子モデル計算ソフト講習会(実習つき)など

今日

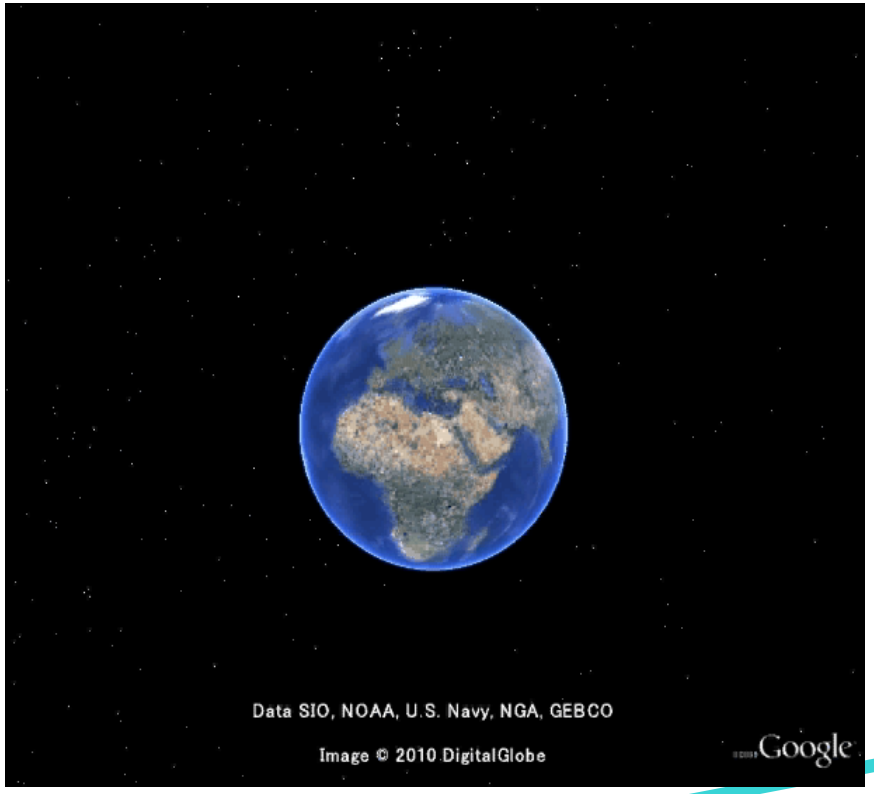
バイオスーパーコンピューティング東北2014
健康社会♫計算科学×(バイオ + 医療)
日時:6月2日(月) 13:30 - 18:00
会場:東北大学 流体科学研究所(片平キャンパス)
写真提供:仙台市観光交流課

研究会活動補足

- ISLiM(京の開発プロジェクトの中の生命科学グランドチャレンジ)の成果を活用したり、戦略分野1の状況を知るとともに開発中のものを活用する活動を含む
- 2020年を目指したExa級スーパーコンピュータ開発の状況を知るとともに、設計や利用方法などの提案も検討してゆく

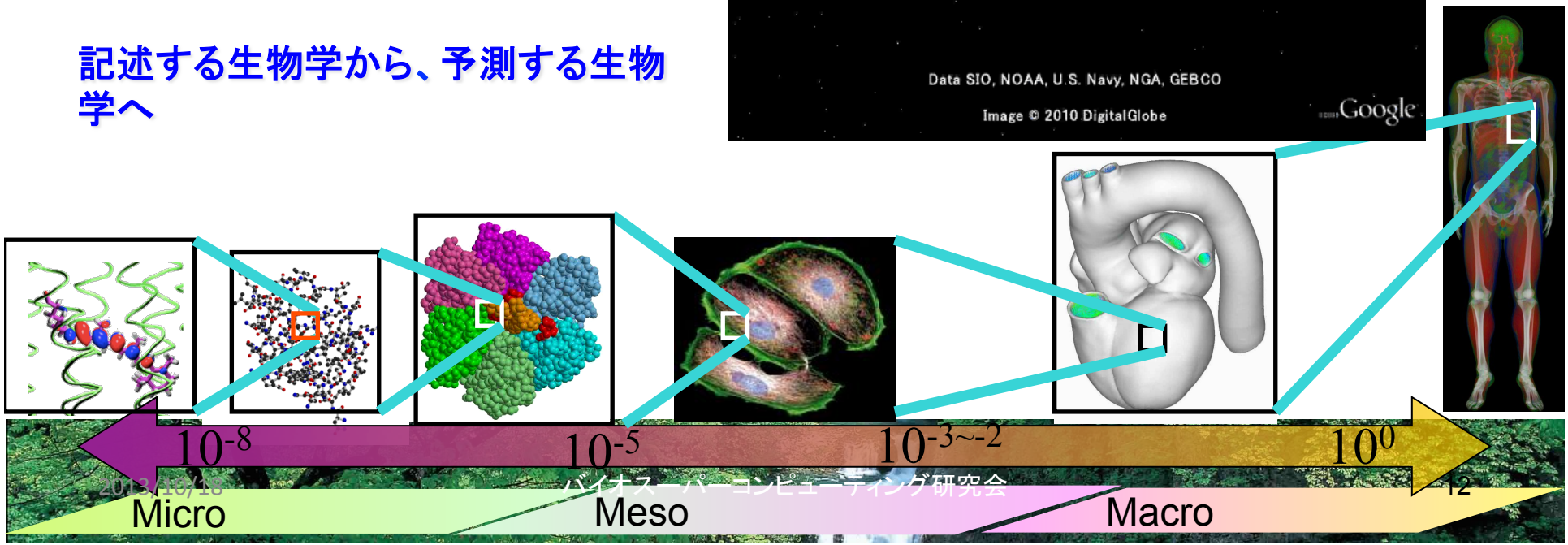

70年間で10¹⁵倍
約5年で10倍
Exa
Peta
Tera
Giga
Mega
Kiro
Performance (FLOPs)
K computer
Tianhe-1A
Roadrunner
Blue Gene P
CRAY XT4/XT3
ASC
ES
I
VPP5000/100
CRAY
YMP
Red
NWT
CRAY2
CRAY
CP-PACS
C90
CRAY X-MP
SX-2
SR220
1
CRAY1
S-810/20
VP100
230-75/APU
CDC7600
CDC6600
STRETC
IBM
704
CDC1604
LARC
UNIVAC I
ENIAC
Mark I
Theoretical Peak
LINPACK
1940
1950
1960
1970
1980
1990
2000
2010
2020
Year

生命現象はどこまで物理で説明できるか？



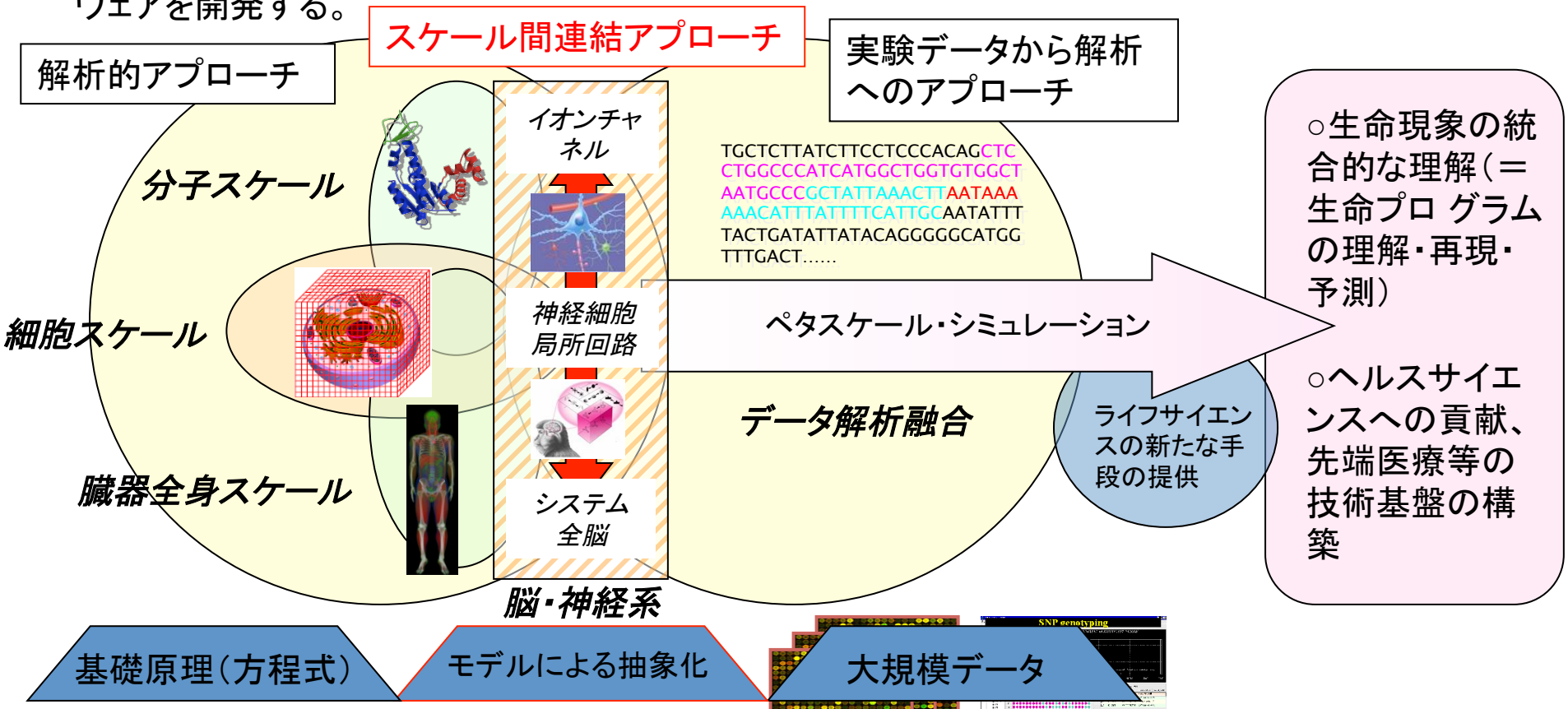
スーパーコンピュータを使って、この複雑な生命現象を解析

記述する生物学から、予測する生物学へ



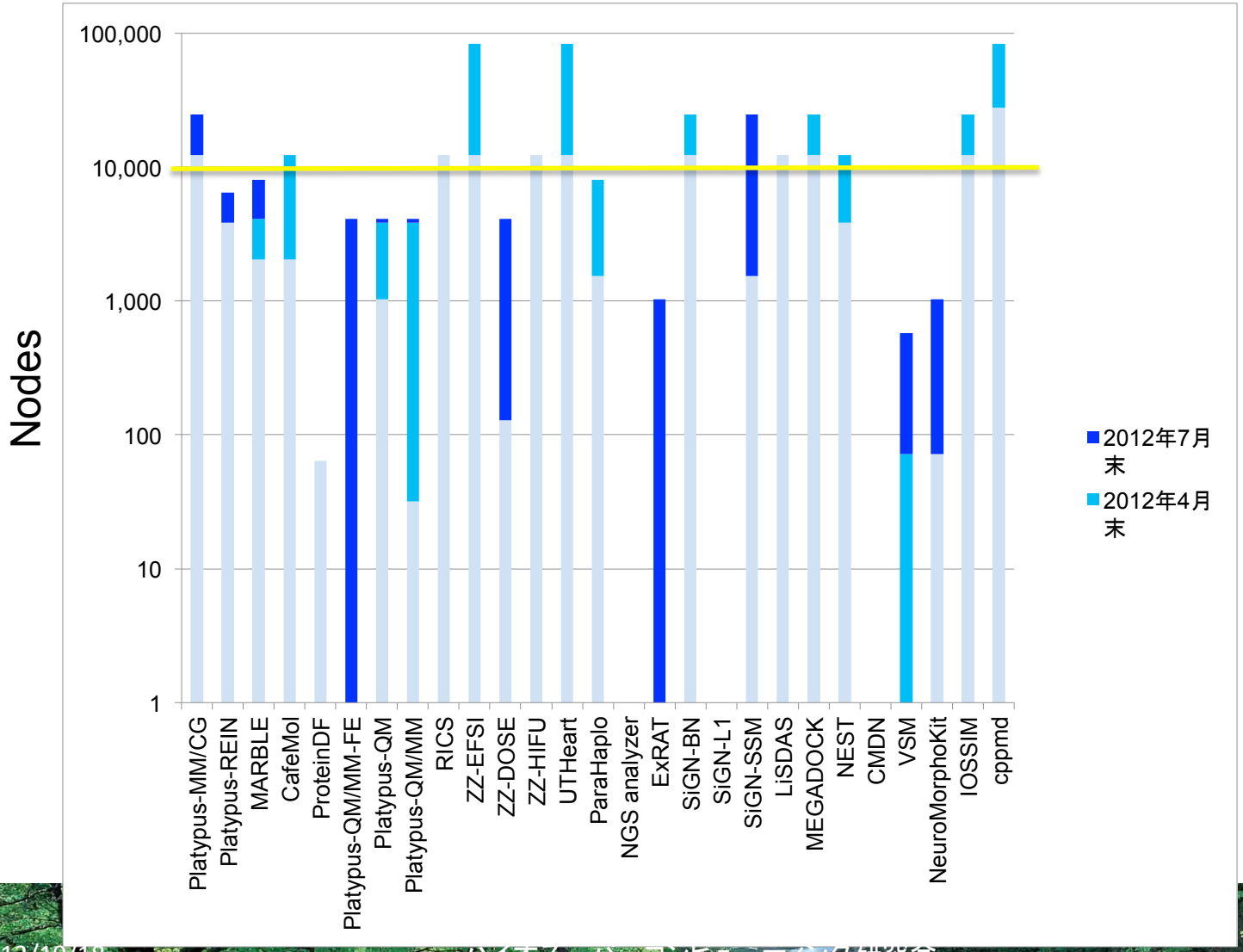
京での生命科学のグランドチャレンジ

基礎方程式に基づく解析的アプローチと、大量の実験データから未知の法則に迫る実験データから解析へのアプローチ、さらには多階層を連結するアプローチにより、異なるスケールの研究と実験データを統合的かつ有機的に結びつけ、ペタスケールという桁違いの性能を持つスーパーコンピュータの性能をフルに発揮し、生体で起こる種々の現象を理解し医療に貢献するためのソフトウェアを開発する。

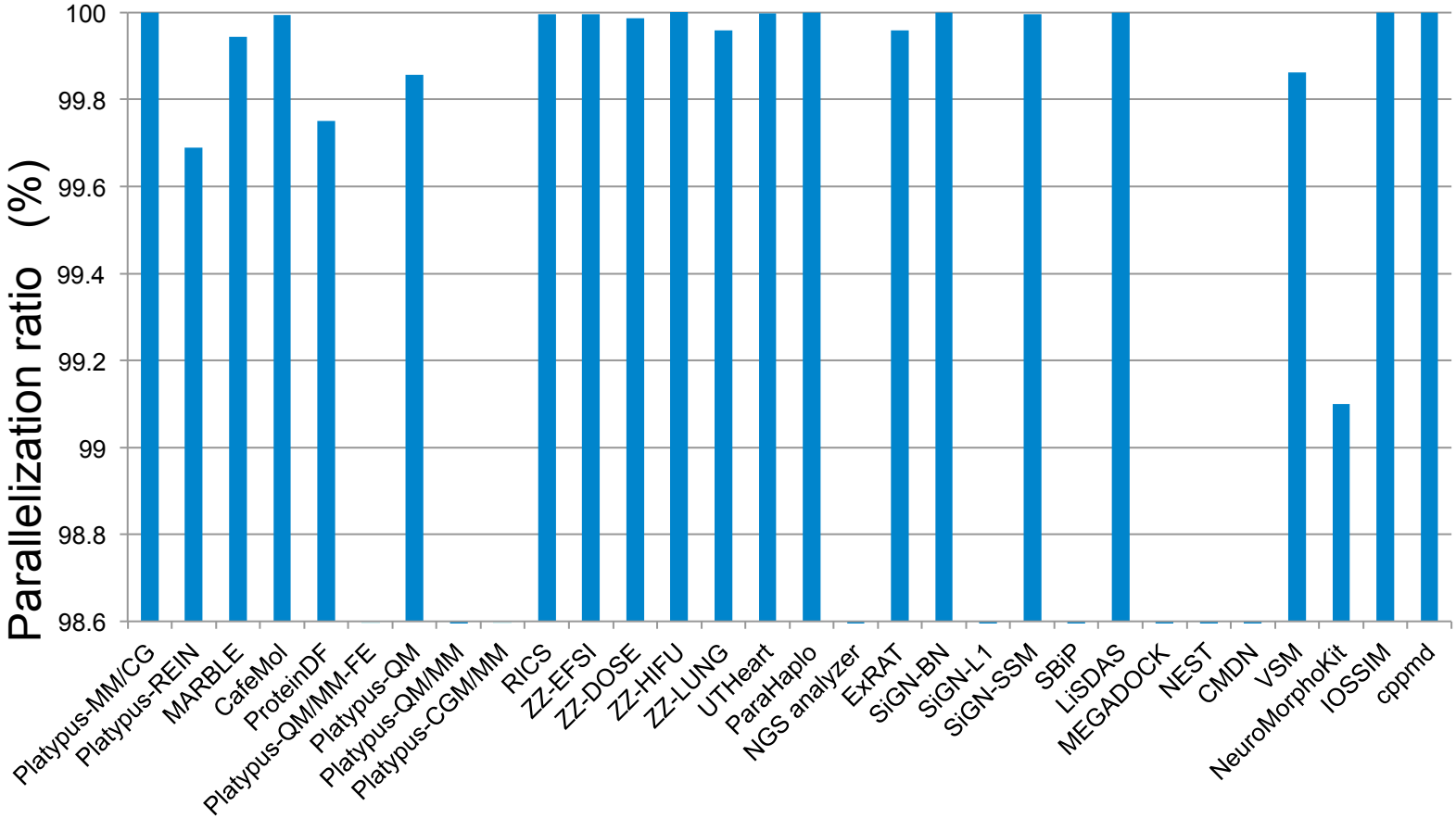


全アプリの京で使用した最大ノード数

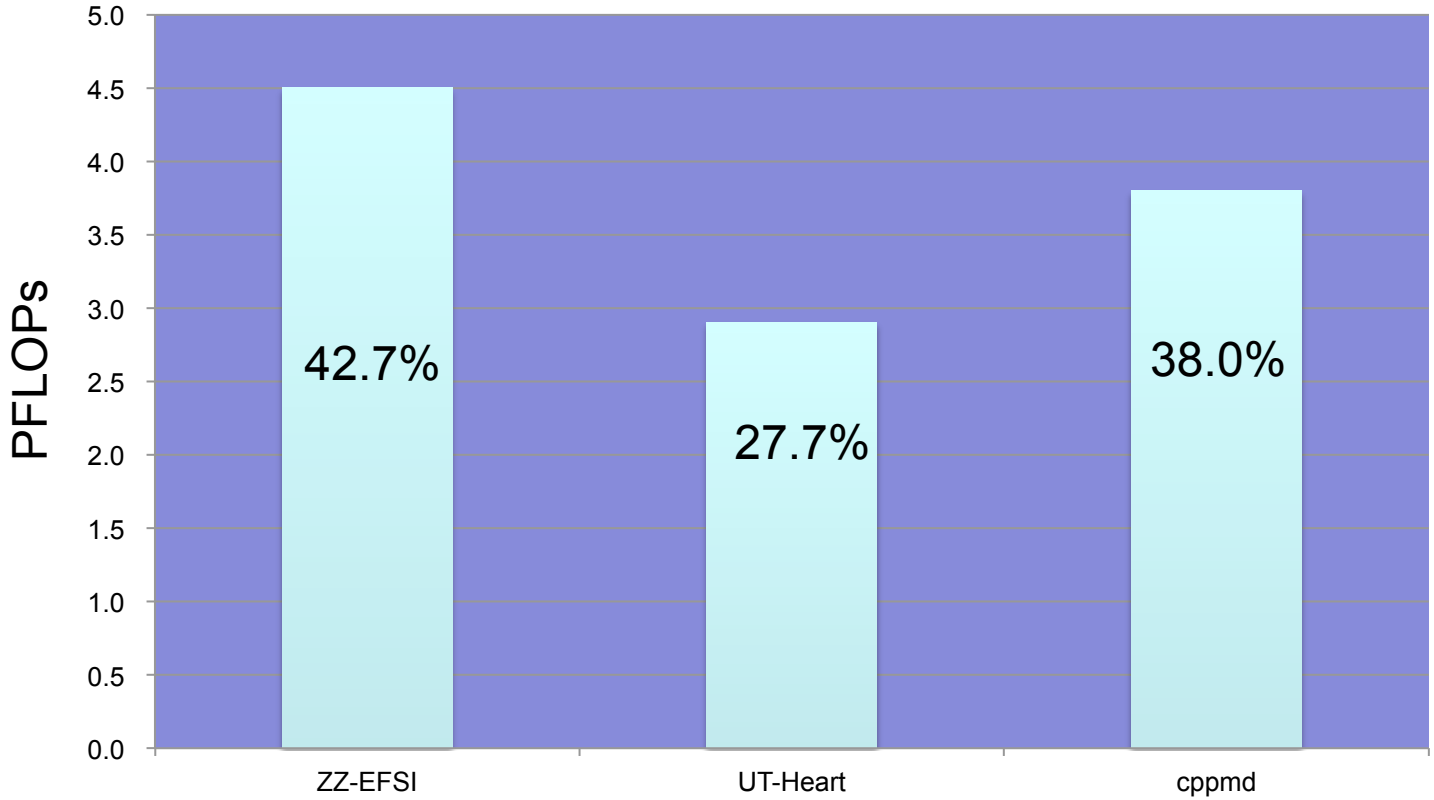
27本中13本が10,000ノード以上を達成



並列化率


Parallelization ratio (%)
100
99.8
99.6
99.4
99.2
99
98.8
98.6
Platypus-MM/CG
Platypus-REIN
MARBLE
CafeMol
ProteinDF
Platypus-QM/MM-FE
Platypus-QM
Platypus-QM/MM
Platypus-CGM/MM
RICS
ZZ-EFSI
ZZ-DOSE
ZZ-HIFU
ZZ-LUNG
UTHeart
ParaHaplo
NGS analyzer
ExRAT
SiGN-BN
SiGN-L1
SiGN-SSM
SBiP
LiSDAS
MEGADOCK
NEST
CMDN
VSM
NeuroMorphoKit
IOSSIM
cppmd

京全システムを使った性能測定結果（3本）



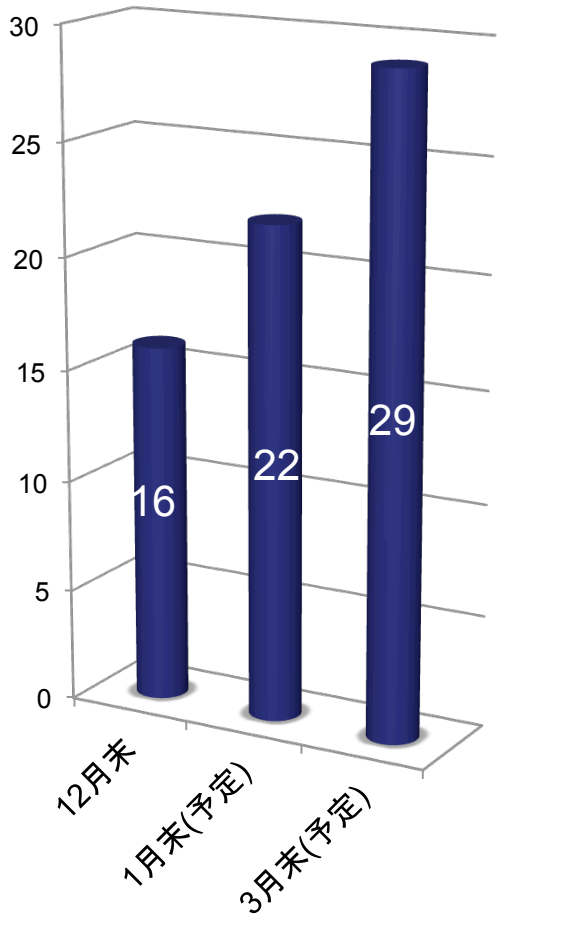
82,944 nodes

ISLiM研究開発ソフトウェアのソース公開

ビッグ・プロジェクトの成果を広く社会に提供し、発展に寄与する

ダウンロードサイト (*http://www.islim.org/islim-dl_j.html*) の一部

ソース公開ソフトウェア数



産業利用化に向けた活動

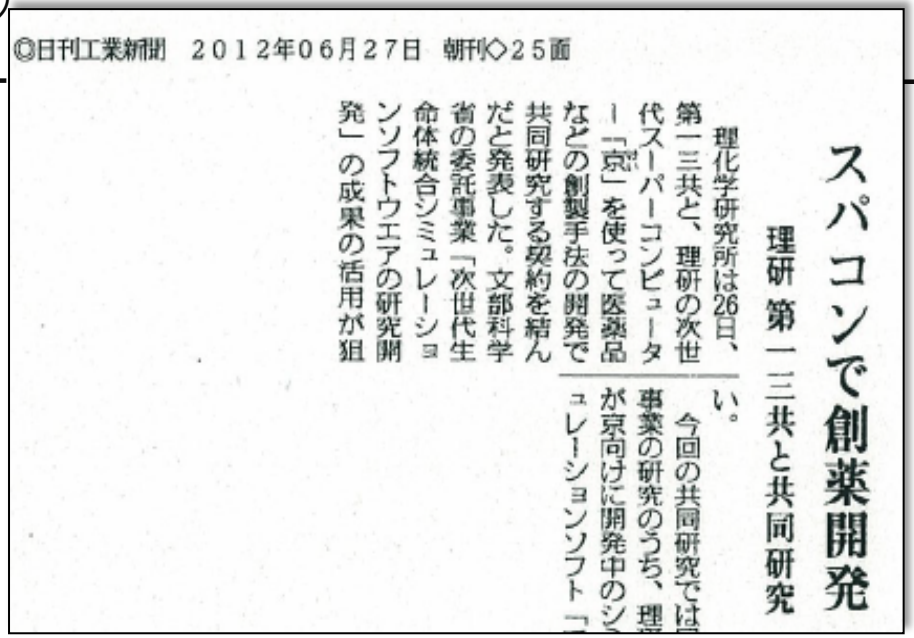
1. 産業利用化ワーキング・グループを発足し、製薬企業に対しISLiM研究開発コードの

認知度向上とニーズ把握を実施

- NPO法人バイオグリッドセンター関西との連携作業 (2011.8 ～)

2. 製薬企業を対象に、ISLiM研究開発コードを利用した共同作業の発掘 (～ 2012.3)

3. 第一


◎日刊工業新聞　2012年06月27日　朝刊◇25面

スパコンで創薬開発

理研 第一三共と共同研究

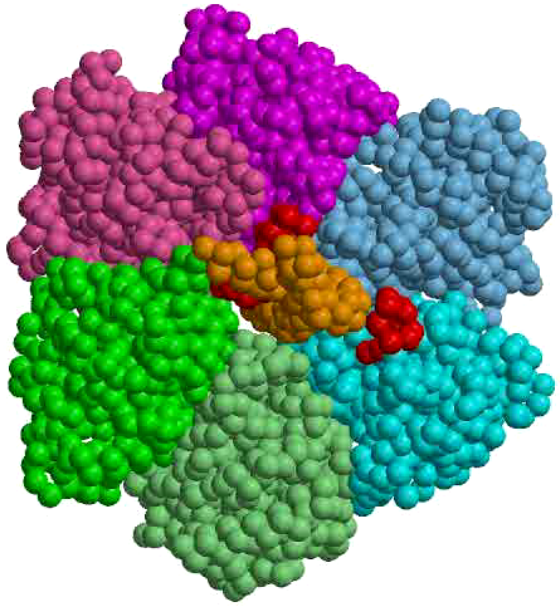
理化学研究所は26日、第一三共と、理研の次世代スーパーコンピューター「京」を使って医薬品などの創製手法の開発で共同研究する契約を結んだと発表した。文部科学省の委託事業「次世代生命体統合シミュレーションソフトウエアの研究開発」の成果の活用が狙い。

今回の共同研究では同事業の研究のうち、理研が京向けに開発中のシミュレーションソフト「

ース: 2012.6

ミクロスケールからメゾスケールへ

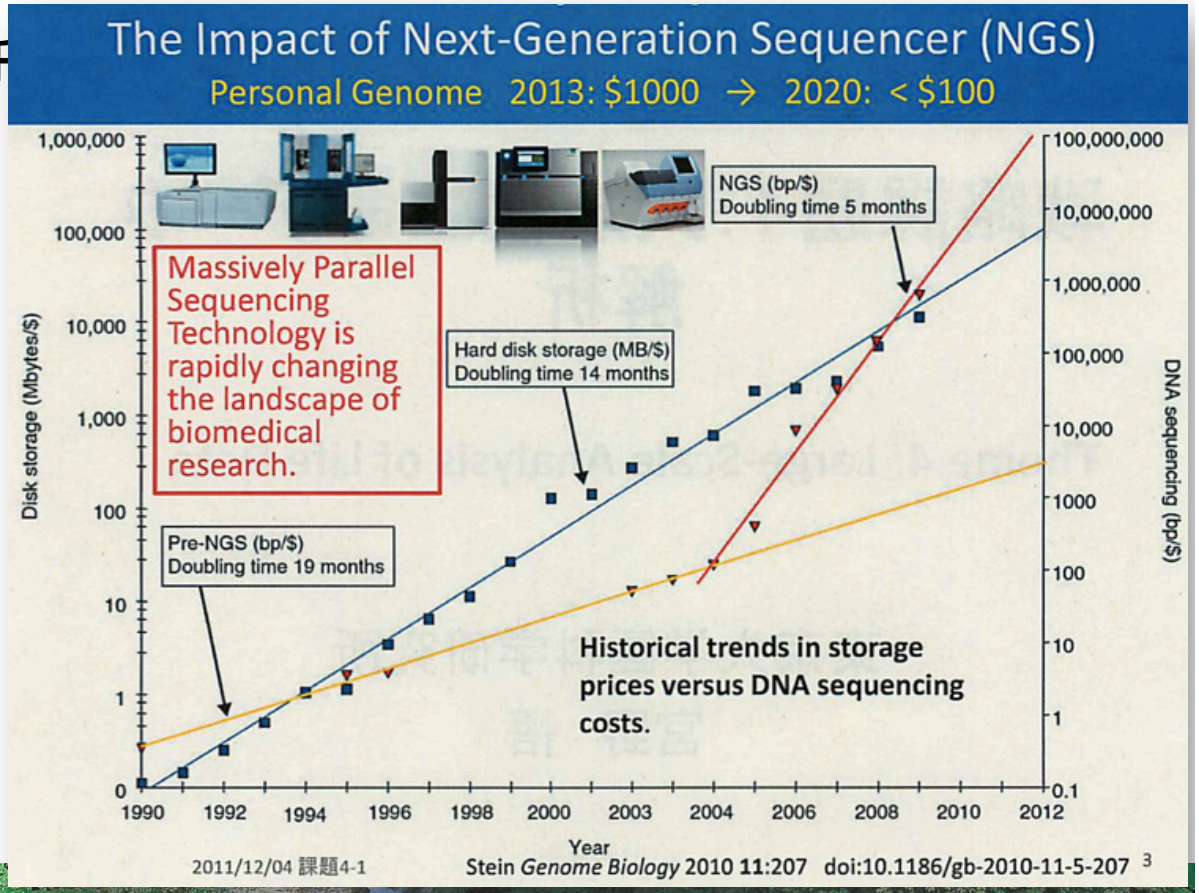
- 生命現象の要素部分はコンピュータ・シミュレーションで解明がすすむ
- 現状では
  - msec 以上の計算は困難
  - 電子状態の変化は計算が大変
  - 細胞内の混雑状況は分子数が多く、大規模な計算が必要
- 今後の計算機の発達に期待



F1ATPace, 高田先生（京大）

生命科学でのBig Data処理ニーズ

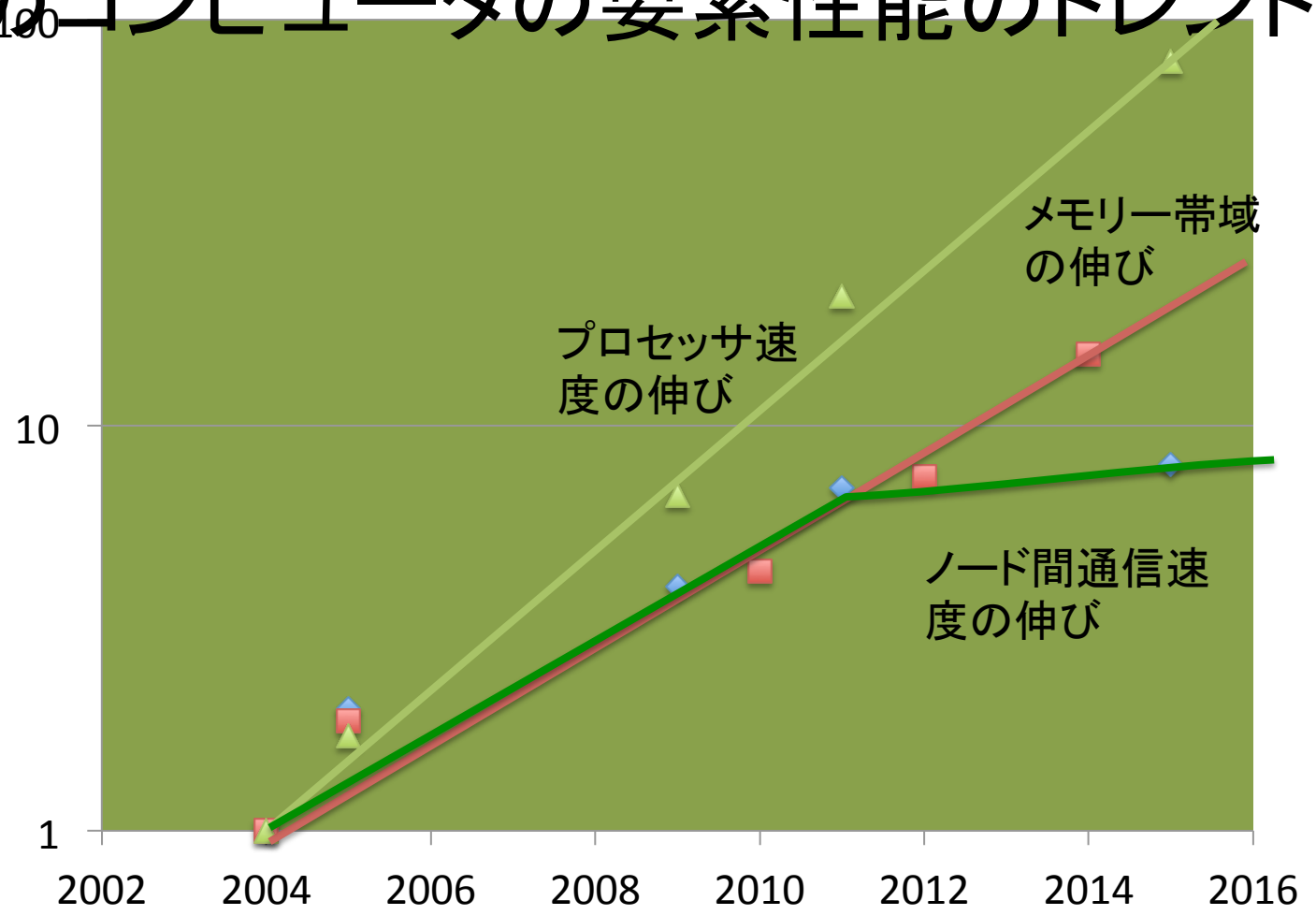
- 次世代シーケンサーの大量データ
  - パスウェー解析
  - ガンゲノム
  - SNP解析
  - ….



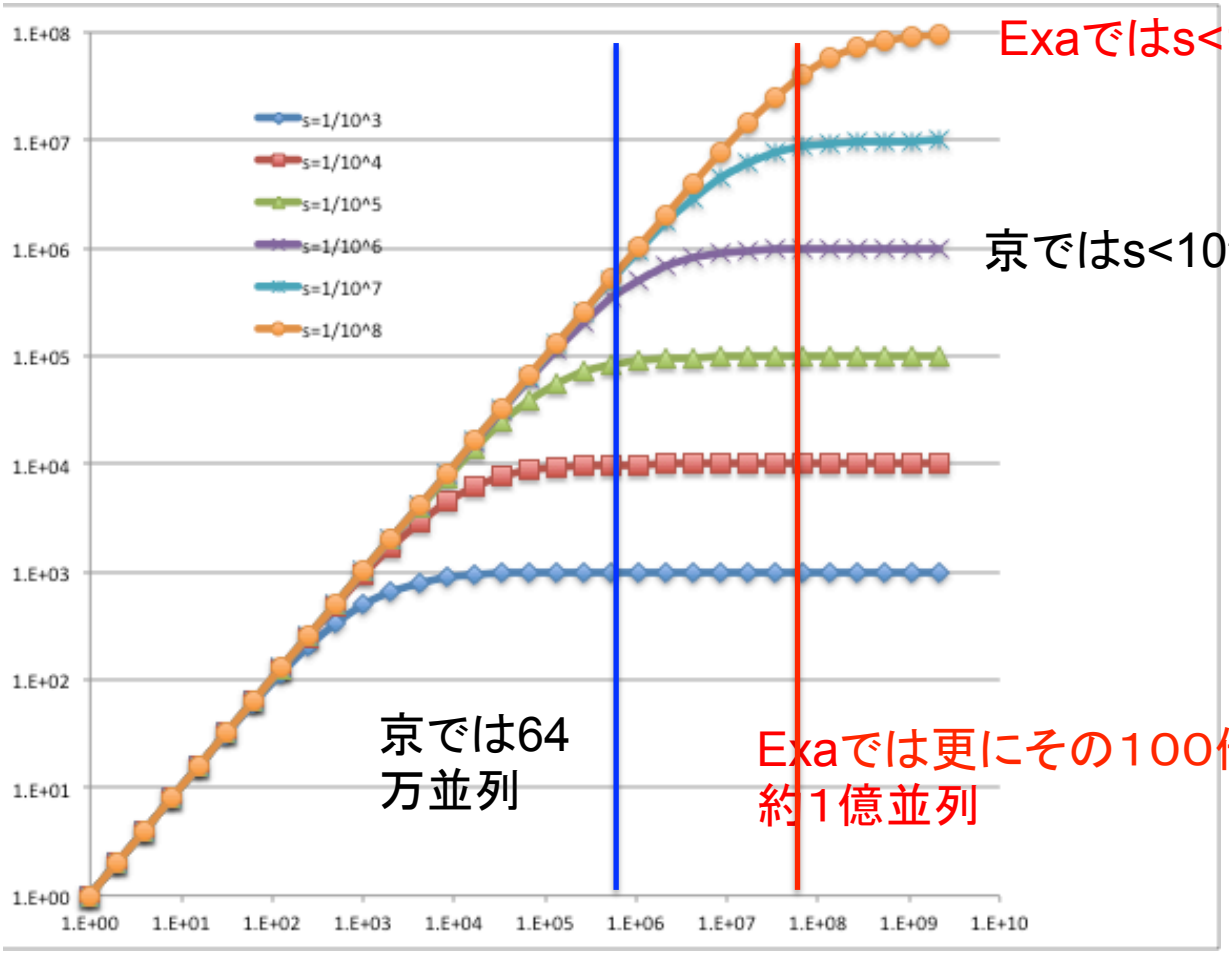
異なるコンピュータハードに対する要求

- コンピュータシミュレーション
  - 浮動小数点演算
  - 通信性能
- ビッグデータ解析
  - ディスクIO
  - 大きなメモリー
  - 整数演算

今後のコンピュータの要素性能のトレンド



アムダールの法則から見たExaの難しさ



Exaではs<10^{-8}

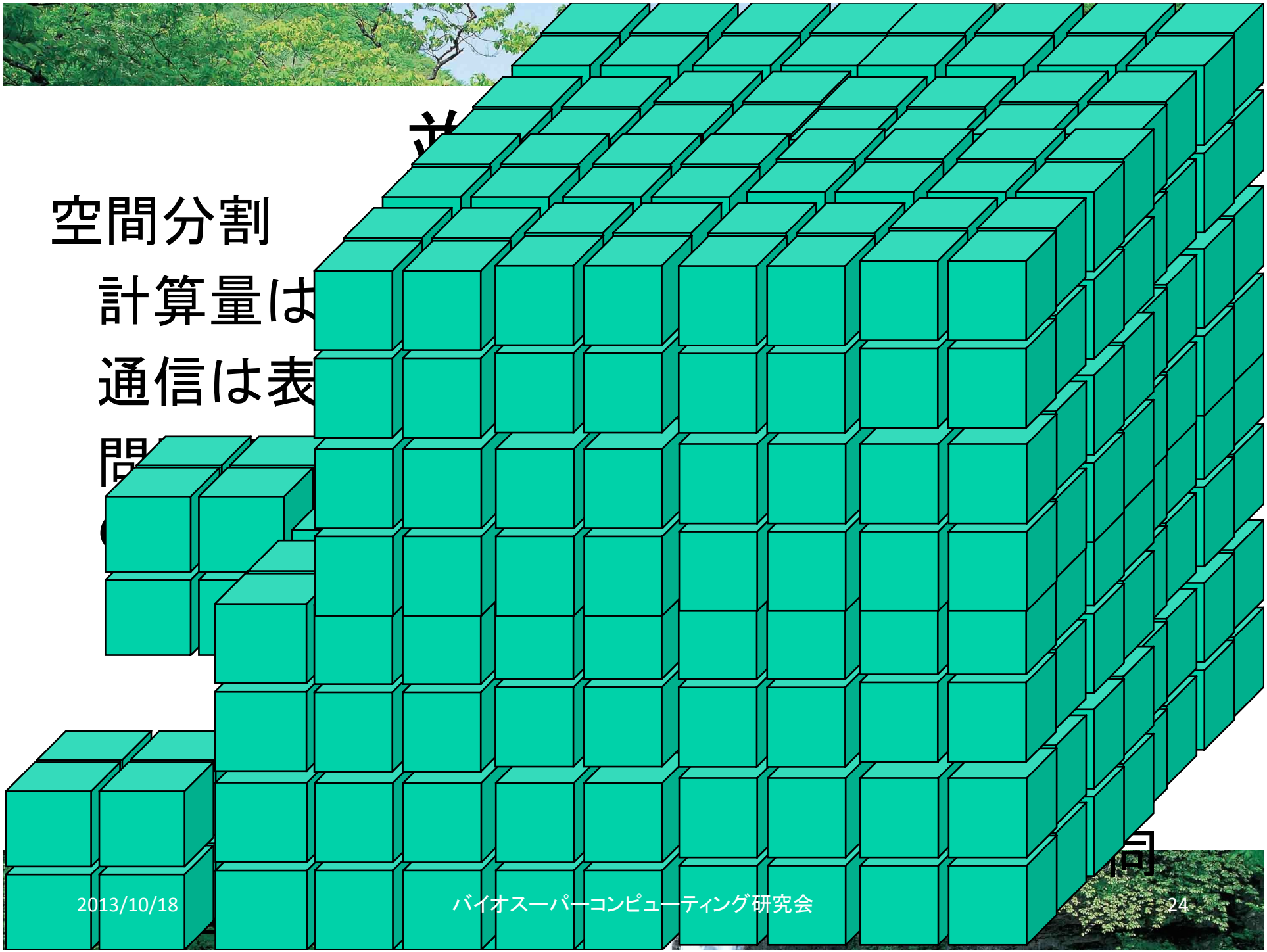
京ではs<10^{-6}

京では64
万並列

Exaでは更にその100倍、
約1億並列

- 並列化できない部分の比を s とすると
- 並列化できる部分は $(1 - s)$
- この部分は n 個のプロセッサで $1/n$ の時間で実行可能
- 並列化できない部分はいくつ使おうと同じ時間がかかる
- 実行時間は $s + (1 - s)/n$
- この逆数が速度向上比

$$\frac{1}{s + \frac{1 - s}{n}}$$



空間分割

計算量は

通信は表

問

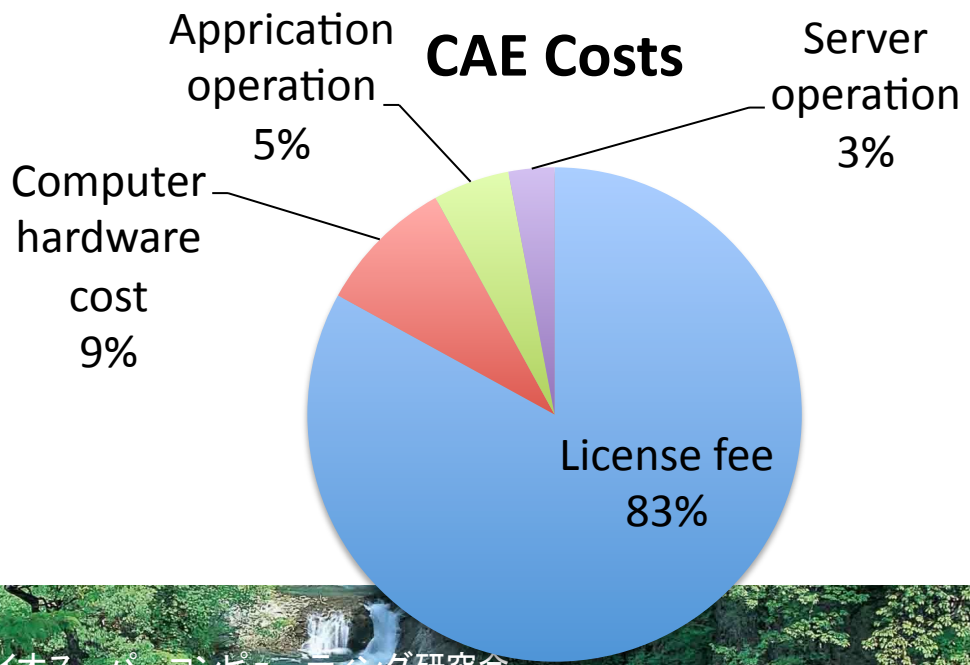
コンピュータハードウェアのトレンド

- 大規模ペタ・サブエクサマシン
  - 大規模な数値計算（1jobが全体を使う）
  - 通信が速くならないため、weak scalingしかしない
    - 計算規模を大きくすることで効率を稼ぐ
  - 京で初めてできるようになった事は何か
  - 多数のパラメータでの並列計算は軽んじられる
- 専用機
  - Antonの脅威
  - Anton2, MD-GRAPE4計画
  - Strong scaling, パラメータ最適化、新薬開発
    - 実時間での長時間のシミュレーションが可能

アプリの開発の仕組みで考えなければいけないこと

企業におけるコンピュータ活用に関する状況

- コンピュータを活用した開発は必須
  - コンピュータ速度向上で、精度良く予測が可能で設計開発のコストと期間の短縮が可能
  - 一方で、寡占化が進んだアプリのライセンス料がコンピュータ活用コストの大きな比率を占めるようになった
  - コンピュータの高性能化の恩恵は多くのアプリで受けられない
    - 京で動かない



市販ソフトの問題

- ライセンス料が高い
  - 使えば使うほど高くなる？
- 計算機の性能向上に追随していない
  - 計算機性能は上がるのに、同じ問題の計算時間は短縮されない
  - モデル規模を大きくした場合、使うノード数を増やしても計算時間が短縮されない（典型的には10ノード以上で速度向上は頭打ち）
- 敵対する関係←→互いに依存する共生関係

大学や国立研究機関発のソフト

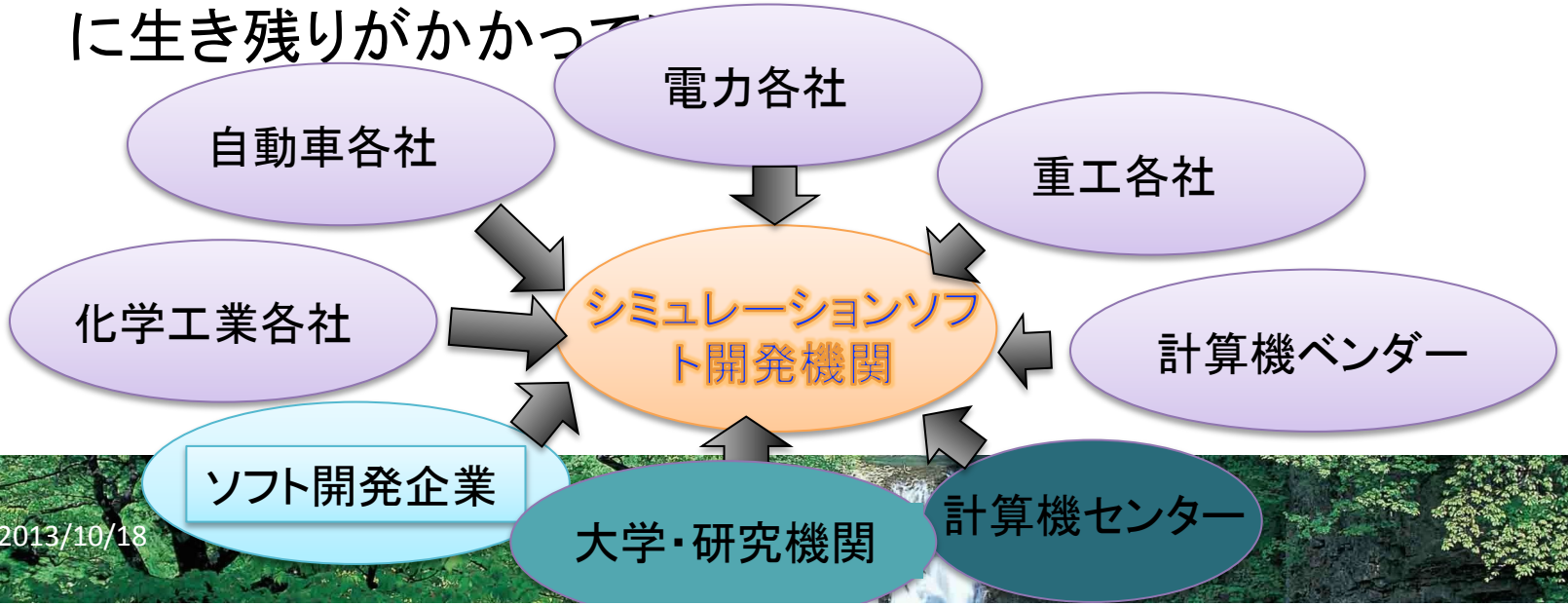
- 多くのソフトがこれまで開発された
  - 流体解析コード、大規模構造解析コード
  - 企業で使われているか？
- 使われていない理由
  - 設計：使われるように作っていない
    - 設計段階からの使う側の参加
  - 信頼性：使い続けられるか
    - 開発プロジェクトが終了すると開発者がいなくなる
    - 改良や機能追加ができるか
  - 計算機の進歩に追随した性能向上が可能か

フリーソフトの問題

- 信頼して使い続けることができるか
  - 機能追加、性能向上は可能か
    - 一般にはできる。でも、誰が責任を持つか。
    - 現実的には現在主流にはなっていない
    - 今後ライセンス料の状況が変わらないと、フリーソフトへの移行が起こる可能性もありえる

シミュレーションソフト開発機関を作り、育てる

- 自動車研究所や電力中央研究所、XX産業組合のように、産業界が出資して、実用的なシミュレーション・ソフトを開発する総合的な機関を作る。
- 民間主導で官学が協力：民間で必要とするものを作成。維持機能向上も担う。
- 製造・生産のソフト化の波（フレキシブルで高付加価値）：ここに生き残りがかかって



今後に向けて

- 異なる研究分野で一緒に議論する事で
  - 境界領域の分野の更なる発展
  - 共通する問題に共同して取り組む
  - 成果の普及、一般への啓蒙
  - 政策への反映

- まずは会員の増強!!