

T2K@京大と プログラム高度化共同研究

京都大学
学術情報メディアセンター
中島 浩



目次

- T2K@京大の現況
 - システム構成
 - 運用・売れ行き・顧客満足度
- プログラム高度化共同研究
 - スパコン関連共同研究
 - プログラム高度化の概要
 - 事例紹介
- まとめ



T2K @ 京大 システム構成

FUJITSU HX600 クラスタ

ノード数 = 416
コア数 = $16 \times 416 = 6656$
ピーク性能 = 61.2 TFlops
Linpack 性能 = 50.5 TFlops(#34→#95)
メモリ容量 = 13 TB

FUJITSU SPARC Enterprise M9000 fat node サブシステム

ノード数 = 7
コア数 = $128 \times 7 = 896$
ピーク性能 = 8.96 TFlops
メモリ容量 = $1\text{TB} \times 7 = 7\text{TB}$



FUJITSU ETERNUS 2000 ストレージシステム

容量 = 883TB
転送性能 = 16GB/s

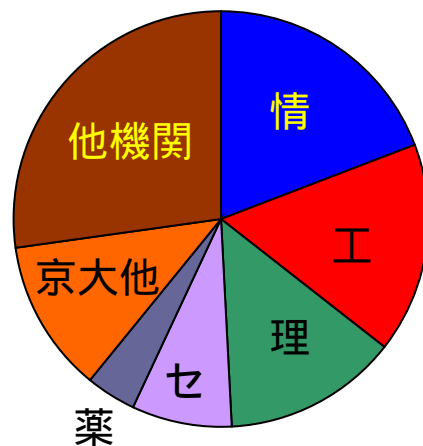
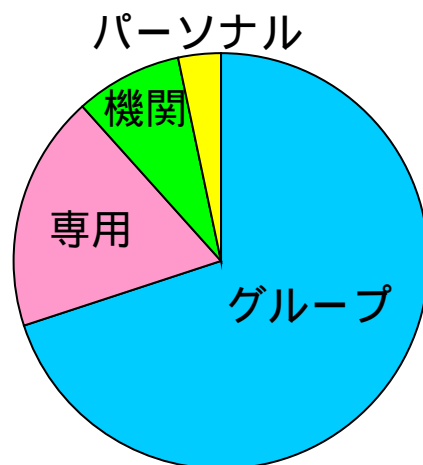
超高速 Infiniband 結合網

通信性能 = 3.3TB/s

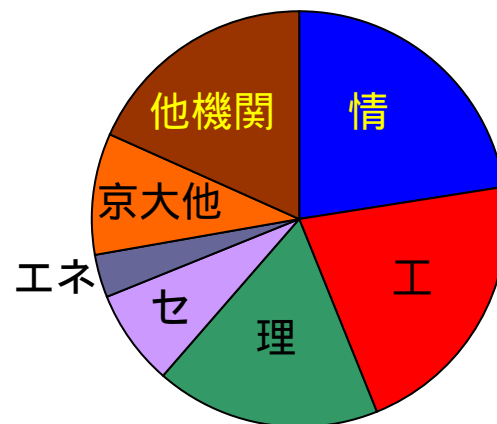
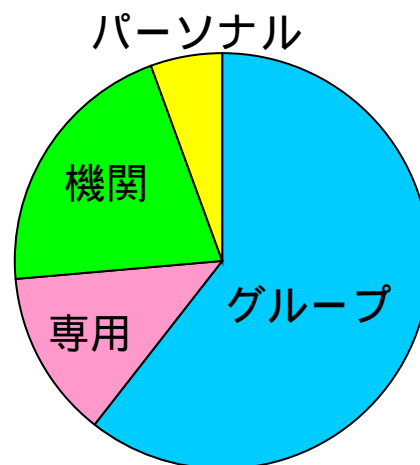
- 課金方針 = 資源買取制
 - 100万円/年 = 8nodes + 8TB + 24users
 - グループメンバーが利用可能な資源量で調整
 - ≠ 物理的なノード割当
 - 通常稼動時 (incl. 繁忙期) は 8node を (ほぼ) 保証
 - 非通常時 (大規模ジョブ実行時) も 4node を保証
 - 60万円/年 = 8node (保証2node) もH21に新設
- 利点
 - 費用対効果 / 計算進捗が読める
 - 「全員が納得する」 scheduling を実現
 - 機関(& 部局)定額利用の実現：4大学 + 2部局

売れ行き (¥70M/year)

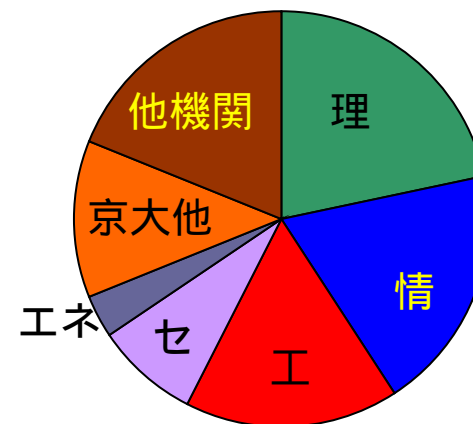
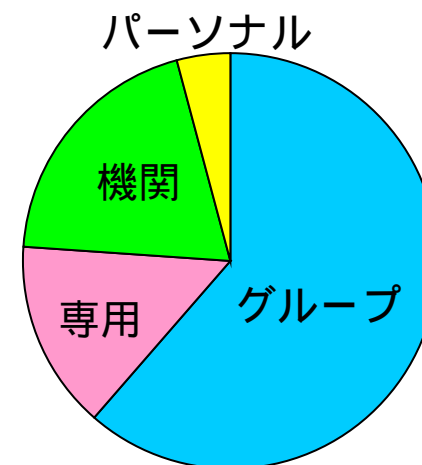
2008



2009



2010



顧客満足度調査@2009.3

■ システム性能

- 計算ノード：満足 = 80%
- ログインノード：満足 = 60%
- ファイルシステム：満足 = 60%

■ 運用制度

- 利用負担金額：満足 = 85%
- 制度設計：満足 = 85%

■ Q/A & Web

- Q/A：満足 = 90%
- Web：満足 = 50~80%



スパコン関連共同研究 (1)

計算科学研究グループ

素粒子物理



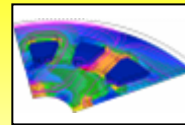
宇宙物理



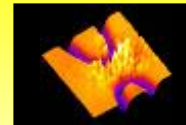
宇宙工学



電磁場解析



ナノ物性



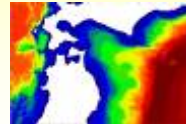
情報検索



気象学



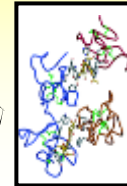
海洋物理



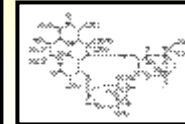
構造解析



分子動力学



創薬科学



医学



グループ / パーソナル
無償利用 (若手)

大規模ジョブコース
無償利用

共同研究
公募

プログラム高度化
無償委託

共用

外部委託

学術情報メディアセンター



スパコン関連共同研究 (2)

- 若手奨励 (< 40歳)
 - パーソナルコース (10万円/年) を無償提供
 - H20 = 20件; H21 = 13件; H22 = 10 ~ 15件
- 大規模計算
 - 128 ノード × 2週間 (相当) を無償/格安提供
 - H20 = 6件; H21 = 4件
 - H22 : 拠点 = ~ 5件; ユーザ = ~ 5件
- 共用促進 (民間利用)
 - 8 ノード × 1年間を無償/半額提供
 - H20 = 4件; H21 = 4件; H22 = 3~5件
- プログラム高度化





プログラム高度化：概要 (1)

- ユーザプログラムを高度化・高性能化
 - プログラム / データ構造改良チューニング
 - 大規模化のための並列化方式改良
(共有メモリ→分散メモリ)
 - 新たな並列アルゴリズムの適用・開発
- 対象課題をユーザから公募
- H20~21 実績 / H22 計画
 - 経費：H20=¥30M; H21=¥80M; H22=¥60M
 - 課題：H20=8件, H21=7件, H22=5~10件



プログラム高度化：概要 (2)

チューニング	複合プラズマ(x15) 分子動力学(x8) 海洋シミュレーション(x2) コロイド物理(x2)
分散メモリ並列化	CO2地中貯留(x20) コロイド物理(x?)
並列アルゴリズム	宇宙プラズマA(x30) 宇宙プラズマB(x13) 宇宙プラズマC(x?) 複合プラズマ(x?) 巨大地震(x7~)

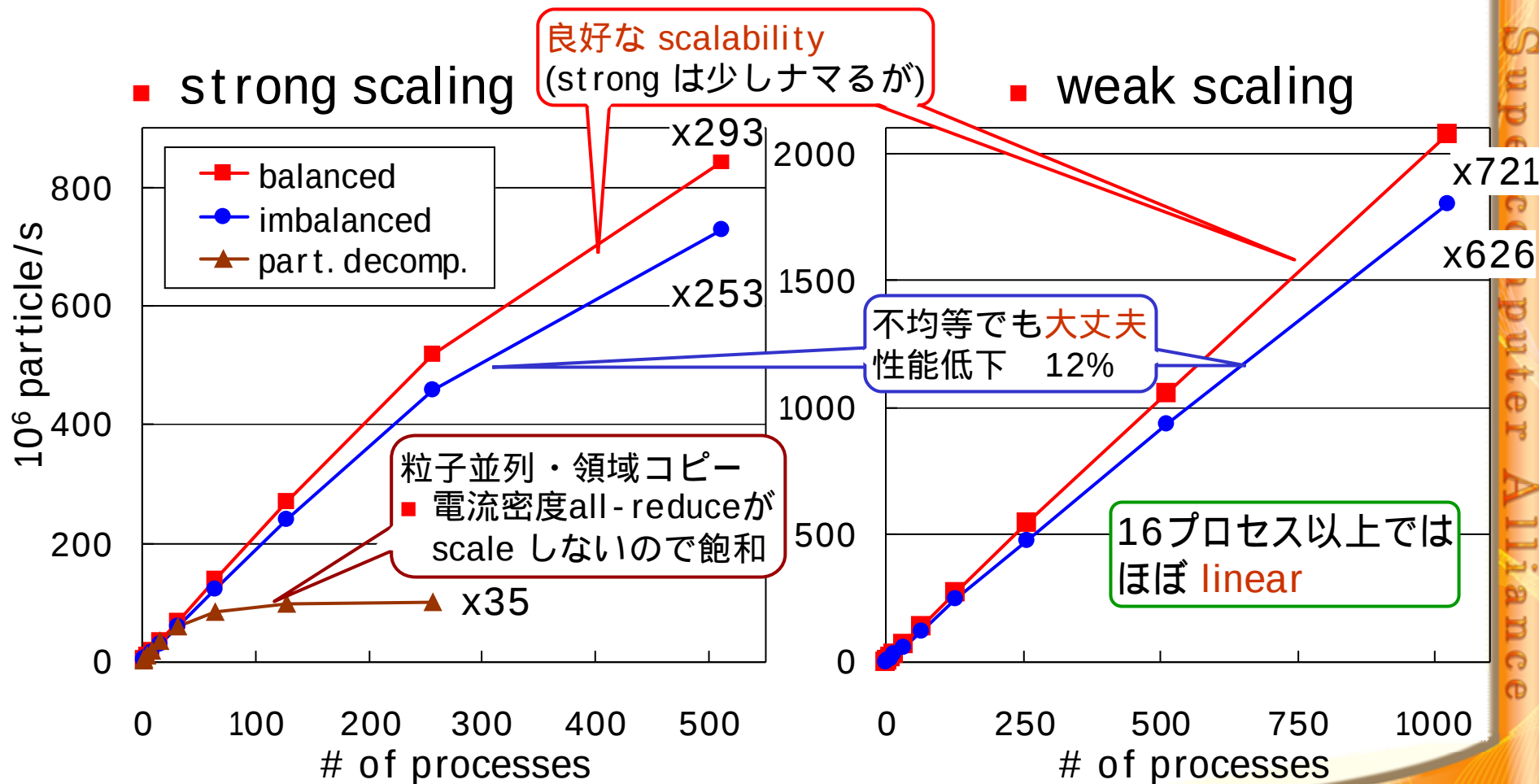




プログラム高度化 プラズマシミュレーション (1)

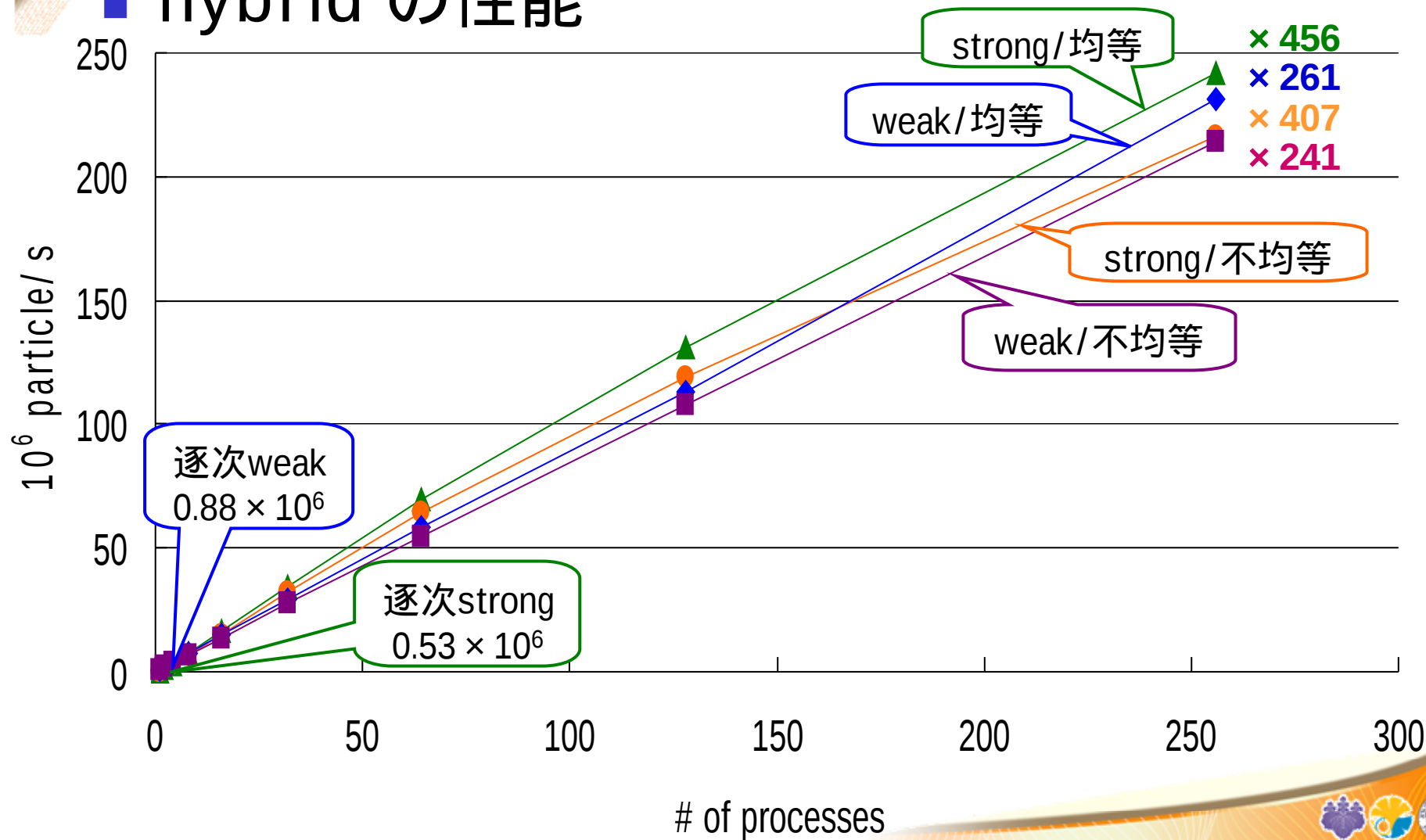
- 負荷分散アルゴリズム **OhHelp** を開発
 - 動的負荷均衡化
 - プロセス間粒子移送
 - 電磁場・電荷/電流密度データの通信
- ➔ ライブラリとして整備
- H20: full-particle simulation
- H21
 - 粒子・流体ハイブリッド
 - 粒子・流体ハイブリッド + Poisson Solver
 - 複合プラズマ (進行中)

■ full-particle の性能



プラズマシミュレーション (3)

■ hybrid の性能





プログラム高度化 CO₂地中貯留

- 格子ボルツマンによる岩石中の水・CO₂の流体シミュレーション
- オリジナルコード = OpenMP
 - ある問題の実行時間 = 40日@16core
→ OpenMPチューニング = 25日@16core
- 高度化版 = OpenMP + MPI
 - ある問題の実行時間 > 2日@16thr × 16proc
 - 負荷バランスの改良でさらに向上？



プログラム高度化 巨大地震サイクル予測

- 32000×32000 の巨大密行列ベクトル積
 - MPI のチューニング→性能改善 = 3%
 - 絶対値が小さい要素を無視した疎行列化
→起きるはずの地震が起きない
- 行列圧縮技法 Hmatrix の導入
 - 行列サイズを **1/40** に圧縮
 - 逐次性能 オリジナルの64proc性能
 - 簡単な並列化→64proc性能 **× 7**
 - 本格的な並列化を実施中



まとめ

- T2K@京大
 - 順調に稼働中
 - 売れ行き & 顧客満足度もOK
 - しかしそろそろ次のマシンを考えねば
- プログラム高度化
 - 1桁ぐらいいは大体上がる
 - 神戸 & 京大次世代機のために必須の投資
 - 今後の課題 = 成果のライブラリ化