

専用機の性能を持つ汎用超並列計算機へ

- GRAPE-DRプロジェクト -

東京大学
平木 敬

平成17年9月28日

東京大学

GRAPE-DR project

GRAPE- 6, 天体シミュレーション専用機 (牧野)

項目	サイズ	消費電力	価格	ピーク性能	実性能
GRAPE-6	4本	30 KVA	5億円	64Tflops	35.3Tflops
地球シミュレータ	770本	6 MVA	400億円	40Tflops	26.6Tflops
比率	192倍	200倍	80倍	0.63倍	0.75倍
備考	19インチラック換算、 プロセッサ、ネットワーク部分	空調用電力を除く	開発費用を含む 原価		GRAPE-6は天体シミュレーション 地球シミュレータは大気大循環全球モデル

- 汎用スーパーコンピュータと比較して非常に効率的
 - サイズ:実性能比 256倍
 - 消費電力:実性能比 267倍
 - 価格:実性能比 107倍



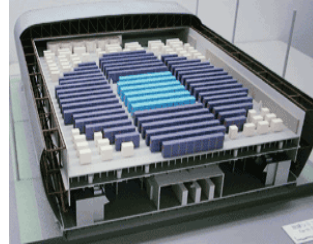
Grape-6

GRAPE-DR project

東京大学

Pflops実現への外挿

- サイズ(19インチ換算ラック数)
 - 地球シミュレータ 19,250本/Pflops
 - BlueGene/L 176本/Pflops
 - GRAPE-6 63本/Pflops
- 消費電力
 - 地球シミュレータ 150 MW/Pflops
 - BlueGene/L 4.2 MW/Pflops
 - GRAPE-6 0.5 MW/Pflops
- 価格
 - 地球シミュレータ 1兆円/Pflops
 - クラスタ(Opteron) 500億円/Pflops
 - BlueGene/L 280億円/Pflops
 - GRAPE-6 78億円/Pflops



地球シミュレータ



BlueGene/L

東京大学

GRAPE-DR project

Grape-6からGRAPE-DRへ

- Grape-6が高効率な理由
 - プログラムのハードウェア演算器による置き換え
 - 数倍のGain(多くの専用計算機の加速要因)
 - GRAPE-DRでは演算器直結によるGainは失う
 - 一般化した計算スキームレベルでのデータパス最適化
 - 並列化手法
 - スケジューリング
 - メモリ能力の最大利用、演算器の有効利用
 - GRAPE-DRでは、演算器結合を一般化することにより問題の適合範囲を増やす
 - 従来のベクトル・スカラーアプローチより高い効率

東京大学

GRAPE-DR project

専用計算機を出発点とした汎用スーパーコンピュータ

- Cosmic Cube Intel iPSC Intel Paragon
– CALTECH Intel
- QCD-PAX CP-PACS SR-2201, SR-8000
– 筑波大学 日立製作所
- QCDSF QCDOC BlueGene/L
– ブルックヘブン研究所 IBM

東京大学

GRAPE-DR project

GRAPE-DRシステムの概要

- 超並列プロセッサの実用化
 - PFLOPS以上の性能の安価な実現
 - 応用ソフトの高効率実行(超並列プロセッサの汎用化)
 - GRAPE-DRシステム全体の計算速度 2 PFLOPS
- 研究開発の概要
 - プロセッサシステム開発
 - 超並列チップ開発、PCB開発、制御プロセッサ開発
 - システムソフトウェア開発
 - コンパイラ、オペレーティング・システム、ネットワークソフト
 - アプリケーション整備
 - 天体シミュレーション、流体力学、線形計算
 - 分子動力学、探索問題

東京大学

GRAPE-DR project

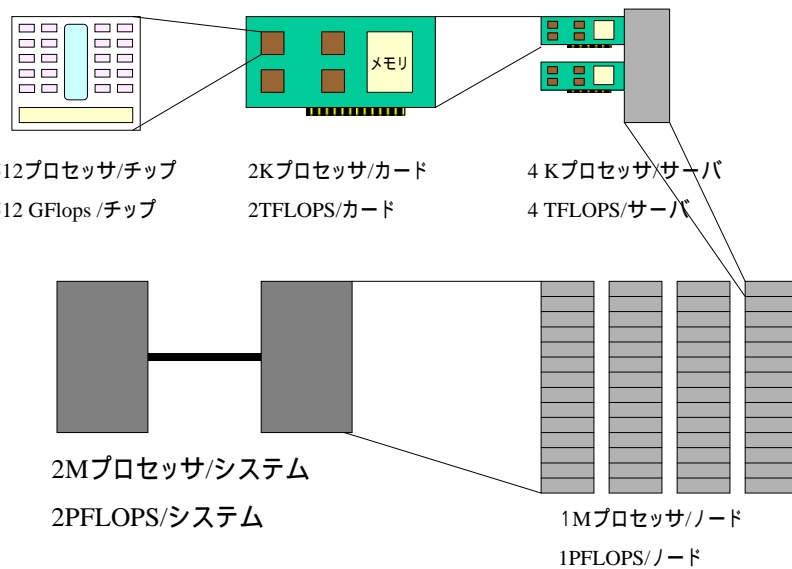
GRAPE-DRの効率性

- メモリバンド幅の有効利用
 - 放送型メモリアクセス
 - 専用Reduction演算ネットワーク
 - 階層化したデータ流と再利用性の確保
 - 実行するアルゴリズムに立脚したメモリ量の削減
- LSI上の資源有効利用
 - プロセッサ間相互結合の排除
 - プロセッサは演算器、ローカルメモリ、レジスタだけ
 - 512個のプロセッサを集積
 - 低クロック速度による消費電力低減
- 汎用部品利用によるコスト節減
 - 低価格サーバへの接続
 - Ethernetによる相互接続

東京大学

GRAPE-DR project

GRAPE-DRシステム構成

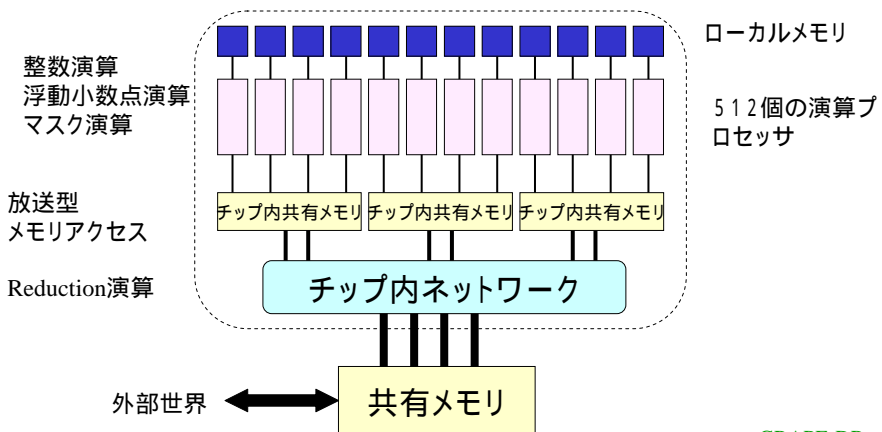


東京大学

GRAPE-DR project

GRAPE-DRプロセッサ

- 超並列プロセッサ技術
- 多数の演算器をパイプライン接続
- 共有メモリと、ネットワーク・パイプラインで接続
 - 演算器間の相互接続の排除
 - Reduction専用ネットワークの埋め込み

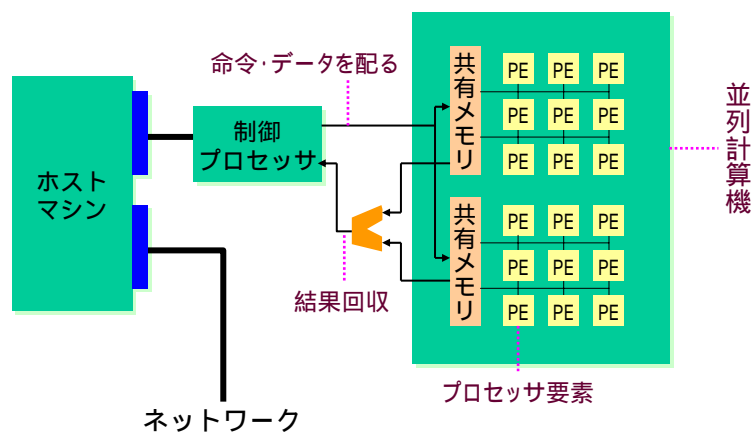


東京大学

GRAPE-DR project

基本アーキテクチャ

- GRAPE-DRが採用する計算機モデルの概要

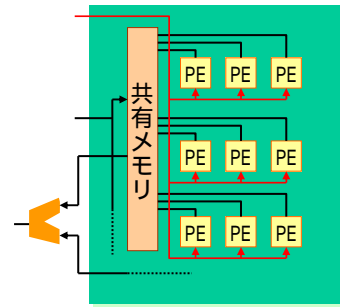


東京大学

GRAPE-DR project

プログラム実行の概要(1)

- プロセッサ要素間のネットワークを単純化
 - 通信は共有メモリを介してのみ
- プロセッサ要素の制御を単純化
 - すべてのプロセッサ要素で同一の命令流
 - 分岐をもたない、マスク演算
 - デコードされた命令を受け取る
 - 制御信号を受け取る

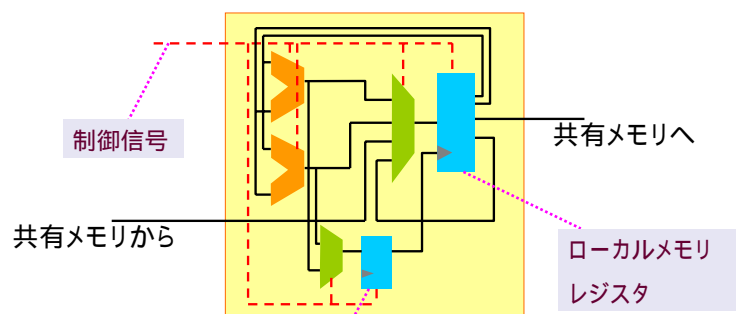


東京大学

GRAPE-DR project

要素プロセッサ

- 1チップに512個実装
 - PEの機能



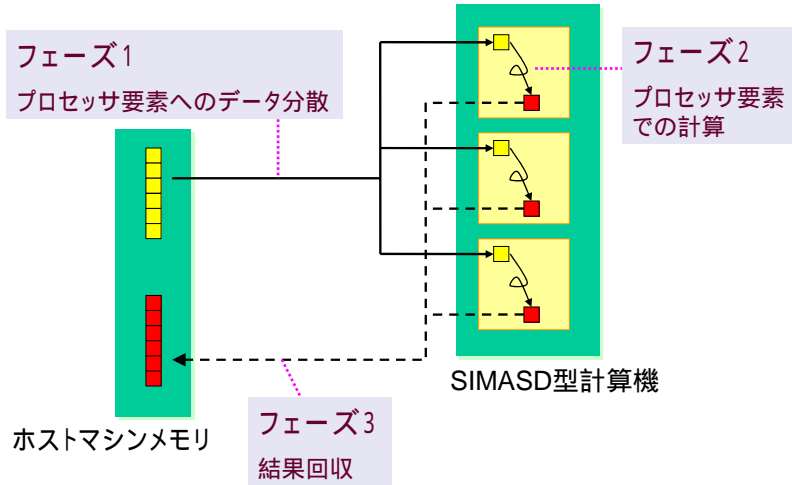
マスクレジスタ...記憶装置への書き込みを演算結果によって制御(PEごとに異なった動作を可能に)

東京大学

GRAPE-DR project

プログラム実行モデル

- 3フェーズ実行

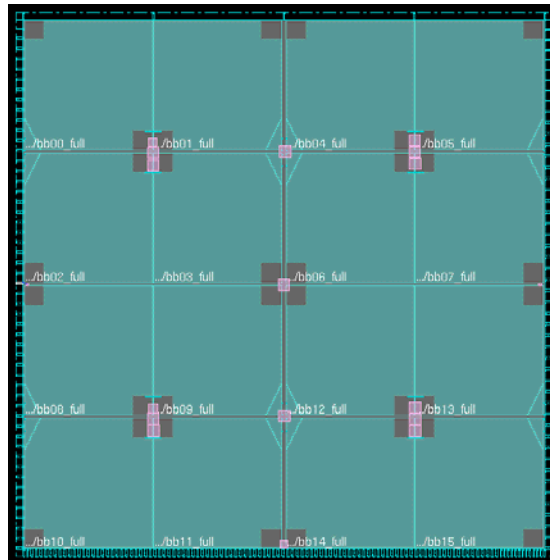


東京大学

GRAPE-DR project

GRAPE-DRチップのフロアプラン

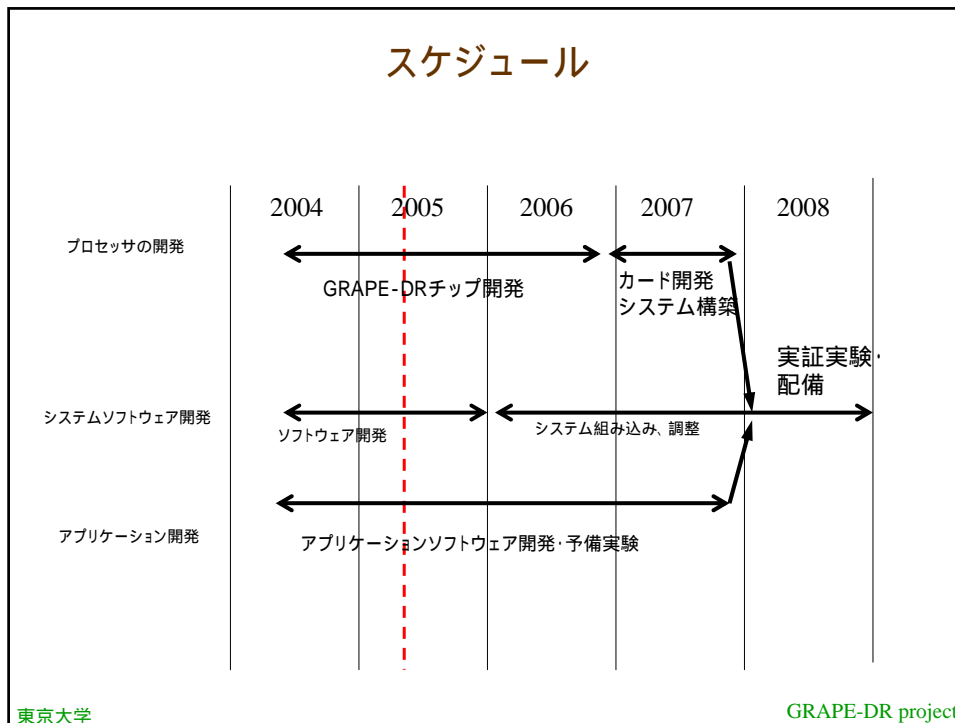
- 90nm CMOS
- 17mm x 17mm
- BGA 実装
- 2006年1月サンプル



東京大学

GRAPE-DR project

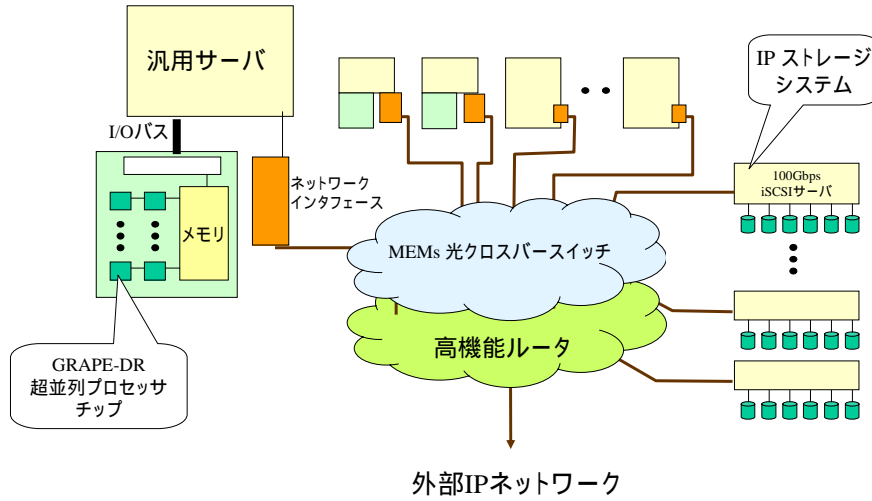
スケジュール



IPネットワークの内部利用

- 現在のプロセッサ内部ネットワークの問題点
 - 高コスト (利用範囲が狭いため)
 - きめの細かい制御の欠如 (バッファ制御、QoSなど)
 - 応用プログラムとの制御の乖離
- 本研究開発が解決するもの
 - MEMsスイッチなど準スタティックネットワークの効率的利用
 - ネットワーク資源(バンド幅、名前空間、メモリ、プロセッサ)のネットワークによる一元的管理
 - ユーザ層からの直接の高効率ネットワーク利用
- コモディティ・ネットワークによる性能:コスト比の向上
 - IT分野で多様されるネットワーク・コンポーネントの利用
 - 高機能ルータ、光クロスバ、光モジュール、PHYなど
 - ネットワークソフトウェア層の共通化(IP, TCP/IPなど)

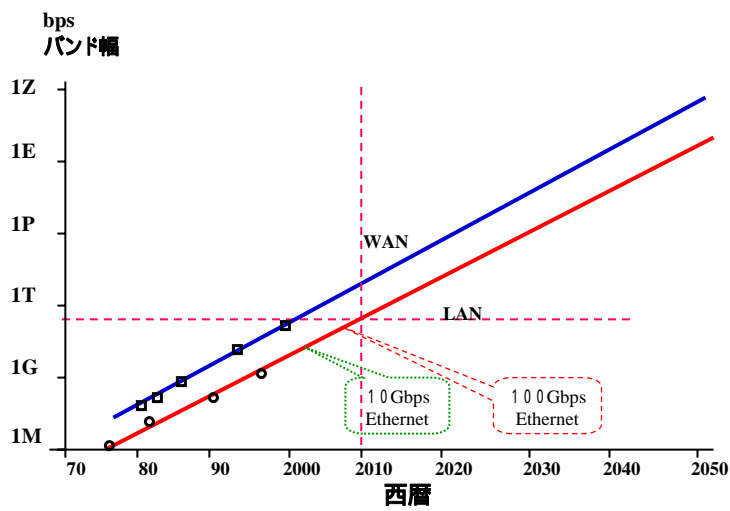
GRAPE-DR相互接続



東京大学

GRAPE-DR project

ネットワーク速度のロードマップ



東京大学

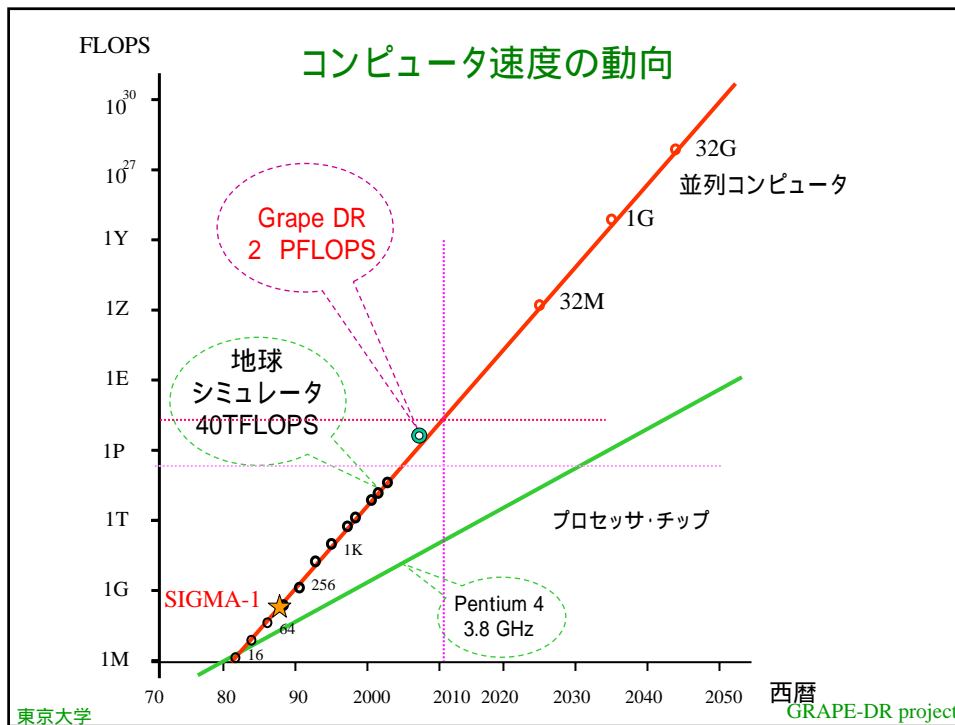
GRAPE-DR project

2010年システムを目指して

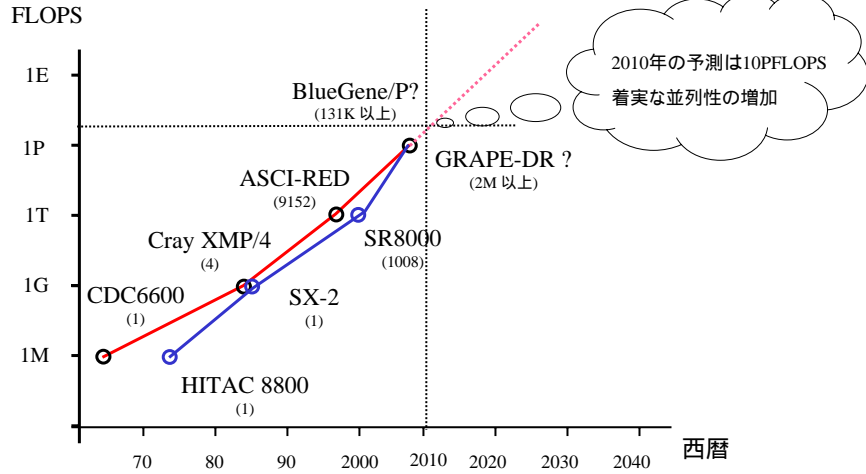
- 2010年スーパーコンピュータへの条件
 - 目標性能の達成 --- 10 Pflops Linpack
 - クラスタに勝るコスト: 性能比
 - 幅広いアルゴリズムがプログラミング可能
 - コンパイラによるプログラミング可能

東京大学

GRAPE-DR project



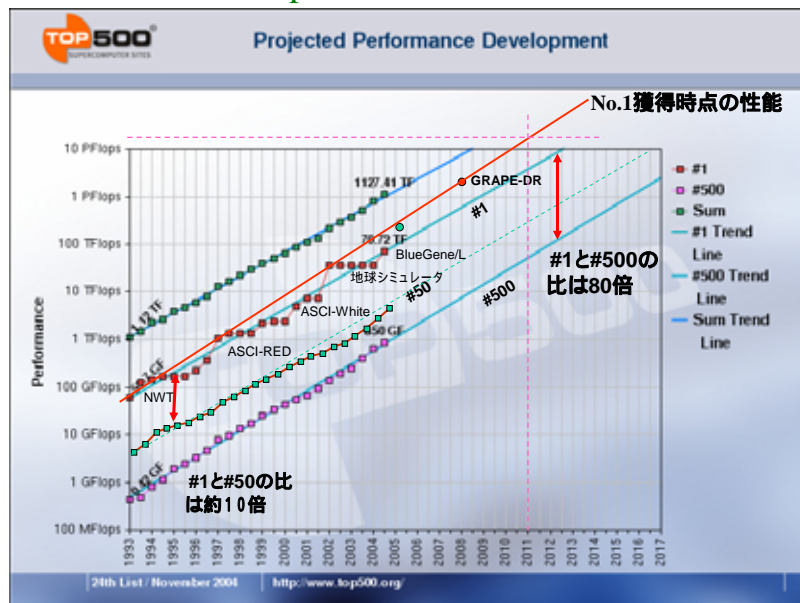
高速化の歴史(大台を越したシステム)



東京大学

GRAPE-DR project

Top500性能の推移



東京大学

www.top500.org による

GRAPE-DR project

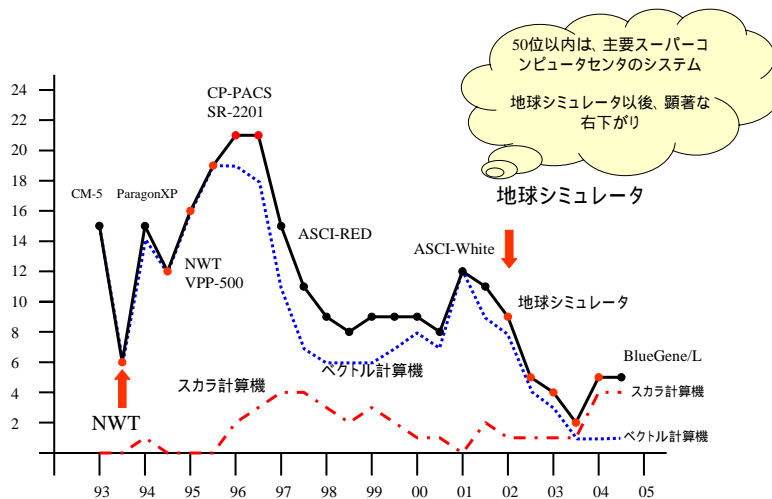
2010年目標性能の見積もり

- 2010年にLinpackでTop500のNo.1システムになるために必要な性能
- 外挿する資料により、10Pflopsから30Pflopsと見積もり
- Top500の50位システムは、1位システムの10%の性能
 - 50位は7大学計算センター + 主要研究機関の計算センター
- Top500の500位システムは、1位システムの1.25%の性能
- 10 Pflopsマシンが最先端のセンターへ設置
- 1 Pflopsマシンが主要大学、主要計算センターへ (50億円?)
- 125 Gflopsマシンが大学・研究所・企業へ (1億円?)

東京大学

GRAPE-DR project

日本の計算システム基盤:50位以内

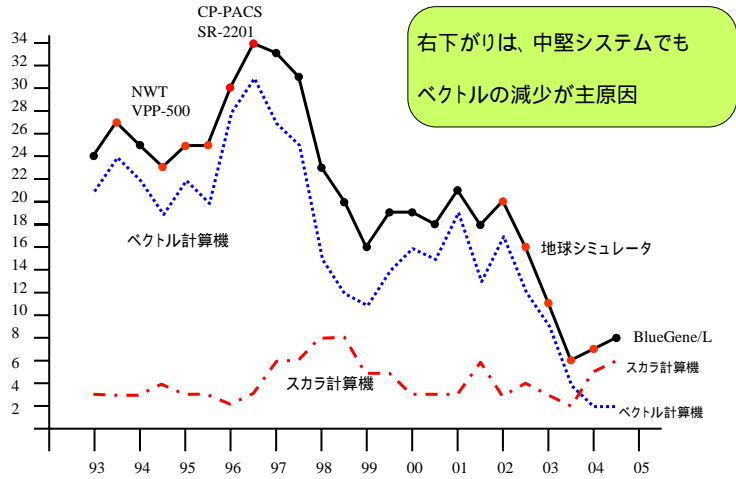


世界のTOP 50位に含まれる日本の計算システム台数の推移
(www.top500.org による)

東京大学

GRAPE-DR project

日本の計算システム基盤: 100位以内

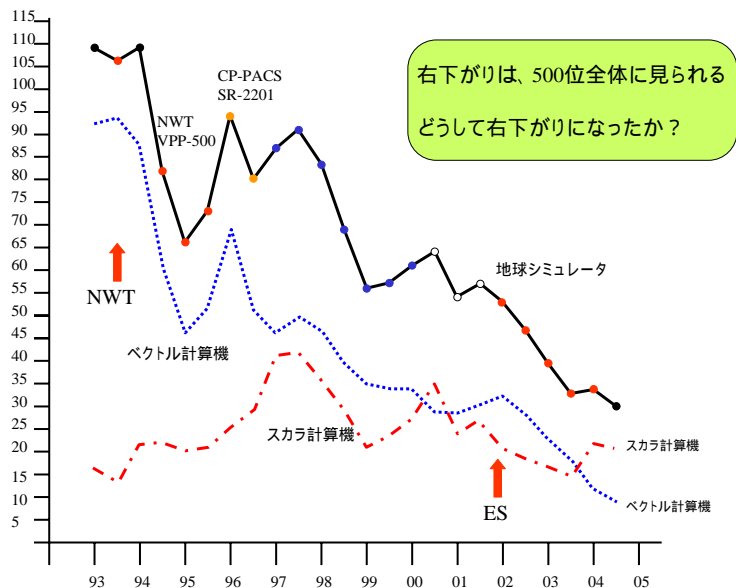


世界のTOP 100位に含まれる日本の計算システム台数の推移
(www.top500.org による)

東京大学

GRAPE-DR project

日本の計算システム基盤: 500位以内



世界のTOP500位に含まれる日本の計算システム台数の推移
(www.top500.org による)

東京大学

GRAPE-DR project

グラフが示す問題点

- Top500の50位以内、100位以内、500位以内とも急激な右下がり
- 国内でもクラスタシステムの急速な増加
 - 理研、産総研など
- 次期スーパーコンピュータは、この流れを変える原動力

東京大学

GRAPE-DR project

GRAPE-DRでのLinpackの性能実現へ

- LINPACKの実行形態
 - Linpack の本体はサーバで実行。DGEMMをGRAPE-DRプロセッサで実行
 - DGEMMは、サーバメモリ (ネットワーク) サーバメモリ
外部メモリ チップ内共有メモリ 演算器 の順で階層的に使う
 - F.Gustavsonのデータ分散法の利用を仮定
- 各部分のバンド幅、メモリ容量の配置は、設計であり、バンド幅とメモリ容量が相補的な関係にある。

東京大学

GRAPE-DR project

Linpackを高効率で実行するI/Oバンド幅の計算

汎用サーバを用いる場合の見積もり

- GRAPE-DRでは、2GB/sのI/Oバスで倍精度ピークで1Tflops
- 次期機種は、4チップで、倍精度ピークで4Tflops
- GDR-2で4Tflopsを実現するためのバス要求 8GB/s
 - 例えば、PCI-express x16 で5GHzクロック
- 20Pflops実現には5000個
- 1200ノードでは、2カード/ノード
- 1ノードに、16GB/s(双方向で32GB/s)が必要

- 安価な市販サーバを用いて20Pflops実現は可能

東京大学

GRAPE-DR project

サイズ・コスト・電力の実現性(10Pflops)

- サイズ(19インチ換算ラック数)

	ラック数	性能	導入年
- 地球シミュレータ	19,250本	(770本/40Tflops)	2002年
- BlueGene/L	1,756本	(64本/365Tflops)	2004年
- GRAPE-DR	200本	(40本/2Pflops)	2007年
- 消費電力

	消費電力	性能	半導体テクノロジー
- 地球シミュレータ	1500 MW	(6 MW/40Tflops)	150 nm
- BlueGene/L	42 MW	(1.5MW/365Tflops)	130 nm
- GRAPE-DR	3MW	(0.6MW/2Pflops)	90 nm
- 価格

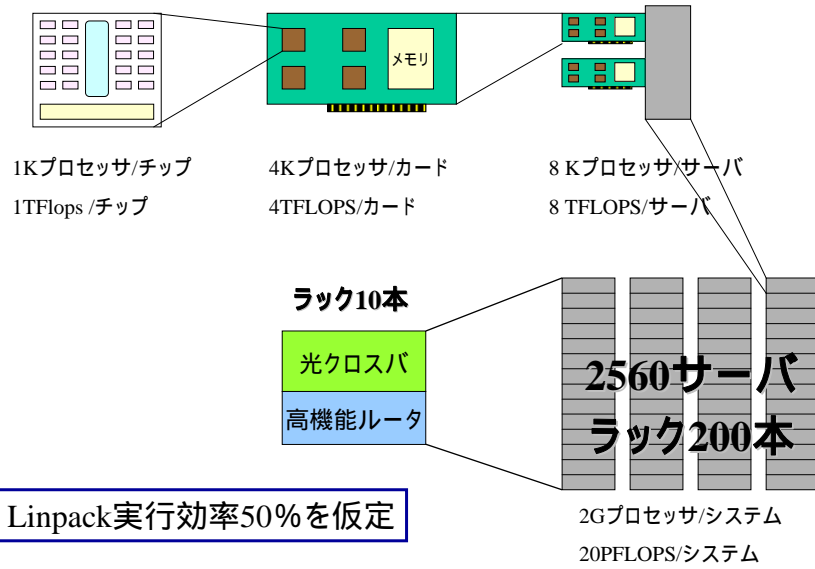
	価格	性能
- 地球シミュレータ	100,000億円	(400億円/400Tflops)
- クラスタ(Opteron)	5,000億円	
- BlueGene/L	2,800億円	(100億円/365Tflops)
- GRAPE-DR	75億円	(15億円/2Pflops)

開発時期によるテクノロジー差による効率差を含む

東京大学

GRAPE-DR project

10Pflopsシステム構成 (GDR-2)



東京大学

GRAPE-DR project

まとめ

• GRAPE-DRの特徴

- Grape-6の持つ高効率 + LSI技術の進歩 - 演算器接続の汎化
- 従来のベクトル計算機、SMP計算機、クラスタより高効率
 - システムのサイズ、消費電力、コスト
- 広い応用範囲

- 離散的探索問題
 - 自明並列問題
 - 多体シミュレーション(天体、MD等)
 - 陽解法シミュレーション(圧縮性流体など)
 - Linpack(DGEMM)
 - 陰解法シミュレーション(CG法等)
 - FFT
- } GRAPE-DR }
ベクトル
計算機

東京大学

GRAPE-DR project

2010年を目指し、何をするか

- GDR-2
 - 非専門ユーザが使えるソフトウェア環境
 - より広い応用分野への適合性(離散系、シミュレーション)
 - カード電力の低減
- スカラ計算機
 - 電力の低減 (3Gflops/Wレベル@2010年)
 - コストの低減 (500万円/Tflops@2010年)
 - CPU + メモリあたり1万円くらいが目標
 - メモリバンド幅の有効利用(SMPは困難)
 - 実行効率の向上
- ベクトル計算機(実行効率がスカラより数倍良いと仮定)
 - コストの低減 (2000万円/Tflops@2010年)
 - 電力の大幅低減
 - メモリコストの大幅低減
 - チップあたりの計算能力の大きな増加が必須

東京大学

GRAPE-DR project

平木 敬のプロフィール

東京大学理学部物理学卒業(1976)、東京大学理学系研究課物理学専門課程修了(1982)、理学博士以後、電子技術総合研究所研究員(1982-1991)、IBM社T.J.Watson研究センター客員研究員(1988-1990)、東京大学理学部助教授をへて、現在東京大学情報理工学系研究科教授
 コンピュータ・アーキテクチャ、並列分散計算、再構成可能デバイスを用いた計算、超高速ネットワーク、分散共有ファイルシステム研究に従事。主なコンピュータ構築などに関する業績は:



数式処理計算機FLATS
(理研、1982)



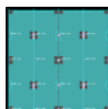
データフロー・スーパーコンピュータ
SIGMA-1(電総研、1988)



分散共有メモリ・並列コンピュータ
JUMP-1(7大学共同研究、1995)



データレザポワールシステム
(東大、2003)



GRAPE-DRシステム
(東大、2008(予定))



Internet2 Land Speed Record
(IPv4, Single and Multiple Stream, 2004)

東京大学

GRAPE-DR project

