授業，演習の概要
HPCの概要

2007年4月11日

中島 研吾

並列計算プログラミング（616-2057）・先端計算機演習I（616-4009）

自己紹介

- 略歴
  - 工学部航空学科出身，博士（工学）
  - 株式会社三菱総合研究所等
  - 2004年よりCOE特任教員
- 専門
  - 数値流体力学
  - 並列プログラミングモデル，大規模数値解法
- 地球惑星科学とのかかわり
  - 平成10～14年度：文部科学省科学技術振興調整費「高精度の地球変動予測のための並列ソフトウェア開発に関する研究」
    - 固体地球シミュレーションプラットフォーム「GeoFEM」
  - 平成17年度～：科学技術振興機構「観測・計算を融合した階層連結地震・津波災害予測システム」（研究代表：松浦充宏教授（地球惑星科学専攻））（次回少し話す）

まずは質問から

- 所属，学年
  - 本郷，地震研，柏，海洋研，駒場・・・
- 並列計算経験
  - 実行，プログラミング
- プログラミング言語
  - FORTRAN，C，JAVA・・・
- 数値解析手法
  - SOR，CG（Conjugate Gradient，共役勾配法）
- ECCアカウント
- 地球物理数値解析 受講予定者

概　要

- 本授業・演習の概要
  - 目的，方針
  - 概要
  - スケジュール
- HPCの概要
  - 計算機ハードウェアの発達
  - 並列プログラミング言語

本授業の理念（HPより）（1/4）

- 近年，マイクロプロセッサの処理速度の上昇，並列計算機の発達によって，計算機システムの処理能力は飛躍的に増加している。世界の並列計算機の動向をまとめた「TOP500リスト」によると，世界最高速の計算機の速度は1993年6月から2006年11月の間に4,500倍にもなっている。2006年11月現在で世界最高速（約280 TFLOPS，TFLOPSとは「TERA Floating Point Operations Per Second」の略，1秒間に1兆回の浮動小数点演算）のIBM BlueGene/Lは約130,000台のプロセッサ，世界第14位（日本で2位，約36 TFLOPS）の「地球シミュレータ」は約5,000台のプロセッサから構成されている。

- これら，国家プロジェクトによる超大型並列計算機の他，PCを連結したPCクラスタも普及している。世界第9位（日本で1位，約47 TFLOPS）の「TSUBAME（東京工業大学）」は約11,000コアのOpteronプロセッサを使用したクラスタである。ネットワーク技術の発展により，広範囲に分布した大規模な計算機資源を効率的に利用する「グリッドコンピューティング」も実現されつつある。並列計算機の使用によって，より大規模で詳細なシミュレーションを高速に実施することが可能になり，新しい科学の開拓が期待される･･･しかしながら，いざ，自作のプログラムを並列計算機で動かそうとすると中々容易ではない。

本授業の理念（HPより）（2/4）

- 参考になる文献も少なく，英語のものが多い。これまで，計算機を専門としない学生に対して科学技術シミュレーションのための並列プログラミング技術を体系的に教える授業は，日本では皆無であった。多圏地球COEの一環として平成16年度から開講された「並列計算プログラミング」，「先端計算機演習I・II」は，そうした試みの日本における最初のもののうちの一つである。

- FORTRAN，C言語などで記述されたプログラムを並列計算機上で並列化するための手段の代表的なものとして MPI（Message Passing Interface）というプロセッサ間通信のための共通規格がある。MPIには400以上の関数があるが，科学技術シミュレーションにおいて必要になるのは10程度である。本授業では，科学技術シミュレーション手法を，局所的手法（差分法，有限要素法等）と大域的手法（境界要素法，スペクトル法等）に分類し，それぞれを並列化するために必要な最小限のMPI関数（1対1通信，グループ通信）について教え，あとはできるだけ実習によって経験を積んでもらうこととする。

本授業の理念（HPより）（3/4）

- アメリカのメリーランド大学（University of Maryland）で Applied Mathematics and Scientific Computation Programというコンピュータサイエンスと科学技術シミュレーションのジョイントプログラムを主宰している David Levermore教授によると，科学技術シミュレーションの真髄は「**SMASH**」，すなわち，
  - **S**cience
  - **M**odeling
  - **A**lgorithm
  - **S**oftware
  - **H**ardware

  であるという。本授業でカバーするのは，Algorithmの一部とSoftware全般，Hardwareの一部である。

本授業の理念（HPより）（4/4）

- 本授業の中で強調したいことは以下の4点である:
  - 並列計算プログラミングは決して難しくない。
  - 最も重要なのはscience, modeling, algorithmである。
  - 計算機に使われてはいけない。
  - 良い並列プログラムは，良いシリアルプログラム（serial program，単独CPUのためのプログラム）から生まれる。
- 授業，実習で学んだことが受講者がシミュレーションによる研究を本格的に実施する際に少しでも助けになればと願っている。もちろん授業内容が直接役に立つにこしたことはないが，それよりも「並列計算は難しいものではない」という意識を，本授業，実習を通して持ってくれることが最も重要である。

本授業・演習の背景

- 計算科学: 第三の科学
  - **と言われて久しいが，現実には実験，観測の後追いに過ぎない。それでも実験より手軽に実施可能である。**
  - **第一原理的手法を使用すれば，実験不可能なことでも，シミュレーションによって解明される，であろうことが明らかになりつつある。**
    - **バイオ，ナノテクノロジー**
    - **現在の計算機リソースでは不可能なものも多い：例えば，**AMD Opteron×1024**クラスタの**10,000**倍規模の計算機必要など**
- **異分野の研究者の協調の必要性**
  - **物理，応用数学，計算機科学**
  - **掛け声だけは**10**年以上前から聞かれるが，なかなか進んでいないのが現状**

本授業・演習の背景（続き）

- 地球惑星科学
  - **観測，実験とそれに基づく理論の重要性**
  - **シミュレーション技術が比較的発達しているのは，観測データが豊富な大気・海洋分野**
  - **観測が難しい分野においてはシミュレーションはより一層重要なはずである・・・**
    - **実際問題として検証ができないとシミュレーションも進歩しない**
  - **大規模な解析空間 ⇒ 大規模計算の必要性**
    - **全地球，連成現象：**Multiscale，Multiphysics
- **地球シミュレータ**
- **地球惑星科学 多圏**COE
  - **大規模シミュレーション技術に関する体系的教育の必要性**

本授業・演習の目的

- 大規模な数値シミュレーションに必須の技術である，並列計算プログラミング技法の習得
  - **地球シミュレータ，PCクラスタ**
  - MPI，OpenMP
  - **情報の探し方の習得：これは重要**
- **並列計算技術のあり方を考える端緒としたい**
  - **自分の研究，分野にとって必要な計算機，計算技術とは ？**
  - **計算機に使われるのではない，使う立場で考える**
- **最終的には，「第三の科学」を開拓するための術（すべ）となれば幸いである。**

担当者，時間割，講義室

- 担当教員
  - **中島研吾（COE特任准教授）**
    - **理**1-716，ex：28329
    - e-mail：nakajima@eps.s.u-tokyo.ac.jp
- **授業時間**
  - **水**：1300-1430 **並列計算プログラミング**
  - **水**：1445-1615 **先端計算機実習-I**
  - **質問等は随時**
- **講義室**
  - **4月11日，4月18日：理学部3号館320号室**
  - **4月25日以降：情報基盤センター5F 大演習室2**

授業内容

- High-Performance Computingの現状と動向
- 並列プログラミングモデルの概要
- スカラープロセッサ，ベクトルプロセッサの特徴と最適化
- PCクラスタ等を使用した実習
- SIMD（Single-Instruction Multiple-Data）方式の並列パラダイムを対象とする：大規模問題
- 前提となる知識
  - UNIXに関する基本的な知識，経験
  - FORTRAN，Cによるプログラミングの経験
  - 数値解析に関する基本的な知識，経験（SOR法等）

履修について

- 夏学期
  - 並列計算プログラミング
  - 先端計算機演習I
    - PCクラスタ利用
    - 演習の役割
      - 授業（並列計算プログラミング）の補足
      - 基本的なアプリケーション設計，開発
      - とは言え，明確に両者を区別しているわけではない
    - できるだけ両者を履修することが望ましい
      - 「演習」の時間にそのまま授業をやることも多い
    - 「並列計算プログラミング」のみの履修については応相談
- 冬学期
  - 先端計算機演習II
    - オーダリング，多重格子，OpenMP
    - 個別研究に応じた指導

評価（案）

- 「並列計算プログラミング（以下C）」，「演習I（以下P）」の評価は別々に実施する
  - 実習課題レポート（C：70%，P：90%）
  - 学期末レポート（C：20%）
  - ログ（C：10%，P：10%）
    - この授業（及び演習）のために毎日どのくらい時間を割いたか，簡単に記述したものを毎週水曜日正午までにメールで中島まで送付のこと

負担を把握するための参考にするものである。
勉強時間が多ければ点数が良いというものではない。
他のことで忙しければ白紙で報告してください。

5 月19日（水）
1000-1200　MPI調査：W.Gropp「Using MPI」
1300-1615　授業
5月20日（木）
1300-1400　クラスタにログイン，環境整備
5月22日（土）
1500-1600　課題C1プログラム作成
5月24日（月）
1800-2100　課題C1デバッグ
5月25日（火）
1600-1700　グループにて討論
1700-1900　課題C1レポート作成

課 題

- MPI例題（その1）（グローバル通信）（課題S1）：C
- MPI例題（その2）（1対1通信）（課題S2）：C
- MPI例題（その3）（共役勾配法）（課題S3）：C
- 熱伝導解析コードによる並列計算実習（課題P1）：P
  - 出題から2－3週間後に解説を実施，模範解答を公開する。

- 提出期限：9月19日（水）17：00

課題・評価についての考え方

- 色々なレベルがある
  - 全部自分でプログラムを作る
  - 解説を聞いてから自分でプログラムを作る
    - 解説に従ってやる
    - 敢えて解説に従わないでやる
  - 模範解答のプログラムを使って計算する

- 「模範解答のプログラムで計算するだけでは意味がない」と思うかも知れないがそんなことはない。何らかの形で全ての課題をこなすことを心がけてほしい。

課題・評価についての考え方（続き）

- プログラミングの習得のためには正しいやり方で実習によって経験を積むしかない。そのための課題である。
  - しっかりやれば，自分の研究に応用することも容易にできるに違いない。これがベスト。
  - 資料を見れば必要になったときにいくらでも習得できる，「並列計算プログラミング」はその程度のものである，という考え方も決して悪くは無い（そのくらいのマテリアルは用意するつもり）。積極的に勧めるものではないが。

- とにかく将来，何らかの形で受講者の今後の研究，仕事の役に立つような課題を設定していくつもりである。

ホームページ等

- http://www-solid.eps.s.u-tokyo.ac.jp/~nakajima/class/
  - 毎週火曜日正午までに翌日のマテリアルを公開する。
  - クラスタ使用法等を含めて，最新の情報を随時追加する。
  - 来週から資料の印刷は各自でお願いします。

- 履修用アンケート（24日1700までに中島宛送付）

氏名（学籍番号）：
専攻，大講座，指導教官：
学年（修士，博士）：
居室：
内線：
e-mail：

クラスタの希望ログイン名（第3希望まで）：

使用言語：
並列計算，並列プログラミングの経験：（使用計算機も含め，具体的に）
研究テーマの概要：（並列計算の必要性，可能性ついても記述する）

履修科目：並列計算プログラミング，先端計算機演習I
（履修しない科目は消す）

http://www-solid.eps.s.u-tokyo.ac.jp/~nakajima/07s/registration.html

参考文献

- いろいろな書籍が出版されているが，これ，と言ったものはない。
  - 全てに目を通しているわけでもない。
  - 特に和書には中々適切なものが無い
    - 計算機科学の専門家向けのものが多い：この分野の研究者が多いこともある
- とりあえず，特に推薦できる書籍を以下に示す。これらは全て中島のところにある。

分類	著者	書名	内容
HPC，並列計算法，MPI	奥田，中島	「並列有限要素解析〔I〕」培風館，2004.	手前味噌ですみません(..=)-。CD-ROM付き。「GeoFEM」プロジェクトの成果。
HPC，並列計算法，MPI	樫山他	「並列計算法入門」日本計算工学会編，計算力学レクチャーシリーズ③，丸善，2003.	CD-ROM付き，MPI中心。有限要素法，差分法，境界要素法の幅広い分野をカバーしている。巻末の付録（MPI関連）は見やすい。
HPC，並列計算法，MPI	三好他	「スーパーコンピューティング」，培風館，2001.	比較的新しい情報。
HPC，並列計算法，MPI	C.Douglas他	「A Tutorial on Elliptic PDE Solvers and Their Parallelization」SIAM,2003.	並列計算について線形ソルバなど数学的な側面からのアプローチ。
HPC，並列計算法，MPI	B.Wilkinson他	「並列プログラミング：ネットワーク結合UNIXマシンによる並列処理」丸善，2000.（原著1999）	計算機科学的アプローチ，基本的なことが一通り簡単に書いてある。
HPC，並列計算法，MPI	P.Pacheco	「MPI並列プログラミング」，培風館，2001（原著1997）	MPIプログラミングに関する初の日本語文献（訳書ではあるが）。例題豊富，自習書としては良い。C言語中心であるがFORTRANのソースもダウンロード可能。
HPC，並列計算法，MPI	W.Gropp他	「Using MPI second edition」，MIT Press, 1999.	同じ著者による「Using MPI2（これは日本語訳有り）」とは違う本，MPICH（フリーのMPIライブラリの決定版）を開発したアルゴンヌ国立研究所のグループが執筆している。アプリケーションを念頭においた使用例も多い。必携の書である。

分類	著者	書名	内容
HPC，並列計算法，MPI	W.Gropp他	「MPI：The Complete Reference Vol.I, II」，MIT Press, 1998.	こちらはどちらかというと辞書的に使用する本。
HPC，並列計算法，MPI	R.Chandra他	「Parallel Programming in OpenMP」，Morgan Kaufmann, 2001.	「地球シミュレータ」など共有メモリユニット用の並列ディレクティヴ「OpenMP」について書かれた世界でおそらく唯一の文献（・・・ではない）。
HPC，並列計算法，MPI	K.Dowd	「ハイパフォーマンスコンピューティング」，オーム社，1994（原著1993）	RISC計算機におけるキャッシュの利用法について解説した文献。
HPC，並列計算法，MPI	S.Goedecker他	「Performance Optimization of Numerically Intensive Codes (Software, Environments, Tools)」，SIAM，2001.	最適化に関する解説書
数値解析手法全般	登坂他	「偏微分方程式の数値シミュレーション」第2版，東京大学出版会，2003.	「並列計算」とは直接関係ないが，数値解析法に関する入門書として適切。
数値解析手法全般	高橋他	「差分法」，培風館，1991.	差分法を中心に数値計算，シミュレーション技術全般に関して記述。なかなか味わいのある本である。並列計算に関する記述も多少あり。
数値解析手法全般	J.J.Dongarra他	「Templates for the Solution of linear Systems: Building Blocks for Iterative Methods」, SIAM, 1994.（邦訳：長谷川他「反復法Templates」朝倉書店，1995）	線形ソルバに関する解説書。簡潔であるが非常に分かりやすい。英語版はタダで ダウンロード できる。例題も豊富。

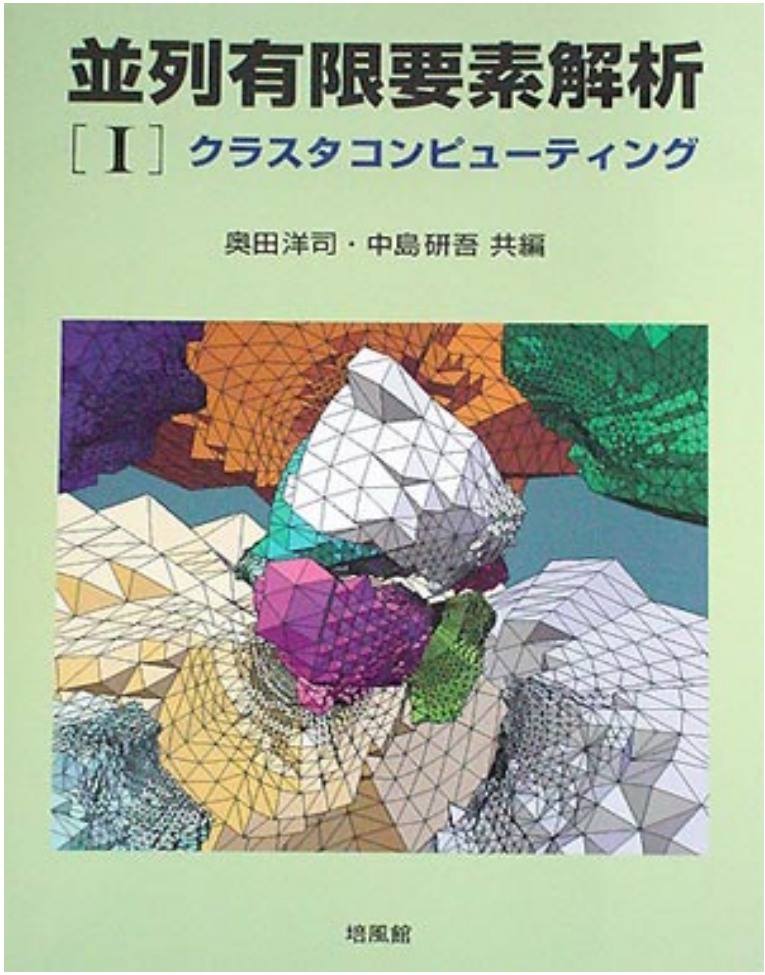
分類	著者	書名	内容
数値解析手法全般	J.J.Dongarra他	「Numerical linear Algebra for High-Performance Computing」, SIAM, 1998.	線形ソルバに関する代表的な解説書。
ハードウェア関連	長島他	「スーパーコンピュータ」, オーム社, 1992.	日立製作所の技術者が執筆。ベクトル計算機のハードウェアの仕組みについて詳説されている。最終章「スーパーコンピュータの将来展望」は今読むと興味深いものがある。
グリッドコンピューティング	日本IBM	「グリッドコンピューティングとは何か」, ソフトバンク, 2004.	前半が「Grid」の概要, 後半はGlobusを使ったGridの構築法について説明してある。
グリッドコンピューティング	I.Foster, K.Kesselmann	「The Grid 2: Blueprint for a New Computing Infrastructure」, Elsevier, 2003.	「Grid」のバイブルと言われている本。

奥田，中島編「並列有限要素解析〔I〕クラスタコンピューティング」

培風館，2004.

- 「GeoFEM」の成果のまとめ
  - **http://geofem.tokyo.rist.or.jp**
- 「地球シミュレータ」上での最適化，シミュレーション結果を紹介

- 初心者向けでは無い
- 高い・・・
  - 若干残部があるので希望者には貸し出します。



スケジュール①

日付	時間	教室	番号	内容
4月11日（水）	1300－1430	理3-320	CS-01	授業および演習の概要，High-Performance Computing（HPC）の概要
	1445－1615			
4月18日（水）	1300－1430	理3-320	CS-02	Grid Computingについて
	1445－1615			数値解析手法の基礎（Gauss-Seidel，SOR，CG法）
4月25日（水）	1300－1430	情基セ・大演習室-2	CS-03	MPIによるプログラミング概要（1）
	1445－1615			
5月2日（水）	1300－1430	情基セ・大演習室-2	CS-04	MPIによるプログラミング概要（2），課題S1出題
	1445－1615			
5月9日（水）	1300－1430	情基セ・大演習室-2	CS-05	MPIによるプログラミング概要（3），課題S2出題
	1445－1615			
5月16日（水）	1300－1430	情基セ・大演習室-2	CS-06	線形ソルバー，課題S3出題
	1445－1615			
5月23日（水）	1300－1430	-		（休講：地球惑星科学連合大会）
	1445－1615			
5月30日（水）	1300－1430	-		（休講：中島海外出張）
	1445－1615			
6月6日（水）	1300－1430	情基セ・大演習室-2	PS-01	課題S1解説
	1445－1615		PS-02	課題S2解説

スケジュール②

日付	時間	教室	番号	内容
6月13日（水）	1300－1430	情基セ・大演習室-2	CS-07	可視化
	1445－1615			チューニング
6月20日（水）	1300－1430	情基セ・大演習室-2	CS-08	並列アプリケーション開発法入門（I）有限体積法
	1445－1615			
6月27日（水）	1300－1430	-		（休講：中島海外出張）
	1445－1615			
7月4日（水）	1300－1430	情基セ・大演習室-2	PS-03	課題S3解説
	1445－1615		CS-09	並列アプリケーション開発法入門（II）有限体積法：並列データ構造，領域分割
7月11日（水）	1300－1430	情基セ・大演習室-2	CS-10	並列アプリケーション開発法入門（III）有限体積法：並列可視化，課題P1出題
	1445－1615			
7月18日（水）	1300－1430	情基セ・大演習室-2	CS-11	並列アプリケーション開発法入門（IV）粒子間熱伝導解析コード並列化
	1445－1615			

その他：演習

- 情報基盤センター演習室の端末を使用する
  - 4月25日（水）以降
  - それまでに，アカウントを取得しておくこと
    - 講習会を受講する必要あり
  - **http://www.ecc.u-tokyo.ac.jp/**

- わからないことがあったら，なるべく質問に来てください。
  - 「在室」の場合はいつでもOKです。本郷以外の方はメールで予約してください。
  - 必要な場合は出張補講にも応じます。

その他：言語

- FORTRAN90（F90）を基本とする
  - 中島はFORTRANユーザーである
  - 幸いなことに地球惑星専攻はFORTRAN利用者が多い
- 授業中の解説もFORTRANプログラムで行うが，比較的分かりやすく書いてあるので，Cユーザーでも容易に理解できると思う。
- C言語による模範解答も準備する。

教育用PCクラスタ

- cenju.eps.s.u-tokyo.ac.jp
  - 千手観音にちなむ
- 理1号館741号室
- 仕様
  - AMD Opteron 1.8GHz x 16ノード（32 PE）
  - 2GB RAM/ノード，1MBキャッシュ/PE
  - 1.28 TB Storage
  - Gigabit Ethernet
  - *PE：Processing Element，CPUのこと

本授業・演習の実情，問題点

- バックグラウンド
  - 数値解析
  - シミュレーション
  - 専攻
  - 各自の興味
    - 感想文を書いてもらうと色々な意見がある
- ボリューム
  - 本年度はできるだけ簡単に，量も少なめに
- 副読本
  - これはあると良いという意見が多かった

- 本授業・演習の概要
  - 目的，方針
  - 概要
  - スケジュール
- HPCの概要
  - 計算機ハードウェアの発達
  - 並列プログラミング言語

計算機ハードウェアの歴史

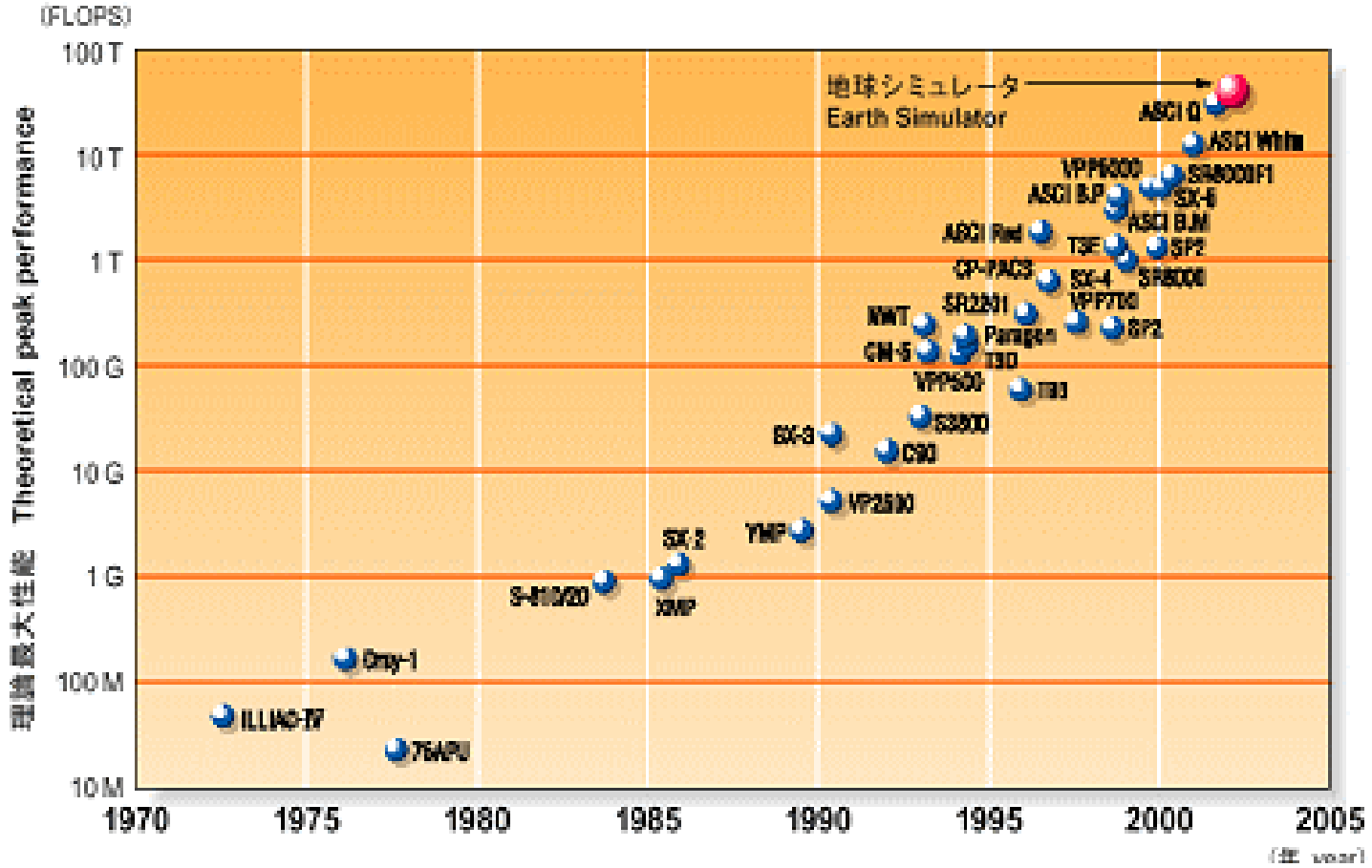
- プロセッサは1.5年に2倍の割合で処理速度が増加している
  - Moore’s Law：集積度が18ヶ月から24ヶ月で2倍
- 並列計算機の発達
- 1983年：1 GFLOPS，1996年：1 TFLOPS，2002年：36 TFLOPS，2005年：280TFLOPS
  - MFLOPS: Millions of FLoating Point OPerations per Second.（1秒間に10^6回の浮動小数点処理）
  - GFLOPS：10^9回，TFLOPS：10^{12}回，PFLOPS：10^{15}回
  - 2010年頃にはPFLOPS（Peta FLOPS）マシンが登場すると言われている。
- 「地球シミュレータ」はピーク性能 40 TFLOPS

次世代スーパーコンピュータ

- 理化学研究所
  - http://www.nsc.riken.jp/index_j.html
- 京速計算機
  - 「京」＝「兆」の10,000倍＝10×10^{15}＝10 Peta FLOPS

- 神戸に設置
- バイオ，ナノシミュレーションが中心といわれているが・・・

計算機ハードウェア発達の歴史



http://www.es.jamstec.go.jp/

アメリカにおけるHPC（High-Performance Computing）

- National Coordination Office for Networking and Information Technology Research and Development（NITRD）(http://www.nitrd.gov/）
  - 何回か名称は変わっている。
  - アメリカの科学技術計算に関する政策。
- Grand Challenge Applications
  - 科学的，経済的，政治的見地から解かなければならないアプリケーションを設定し，その分野に重点的に予算を投下する。
  - 例えば，地震シミュレーション，温暖化予測などもその例である。
  - 近年は，単一のアプリケーションから，ネットワークを利用したCollaborative Project などにシフトしつつある。
    - 複合型，横断型研究でないと予算がとれない：NSF

ASCI

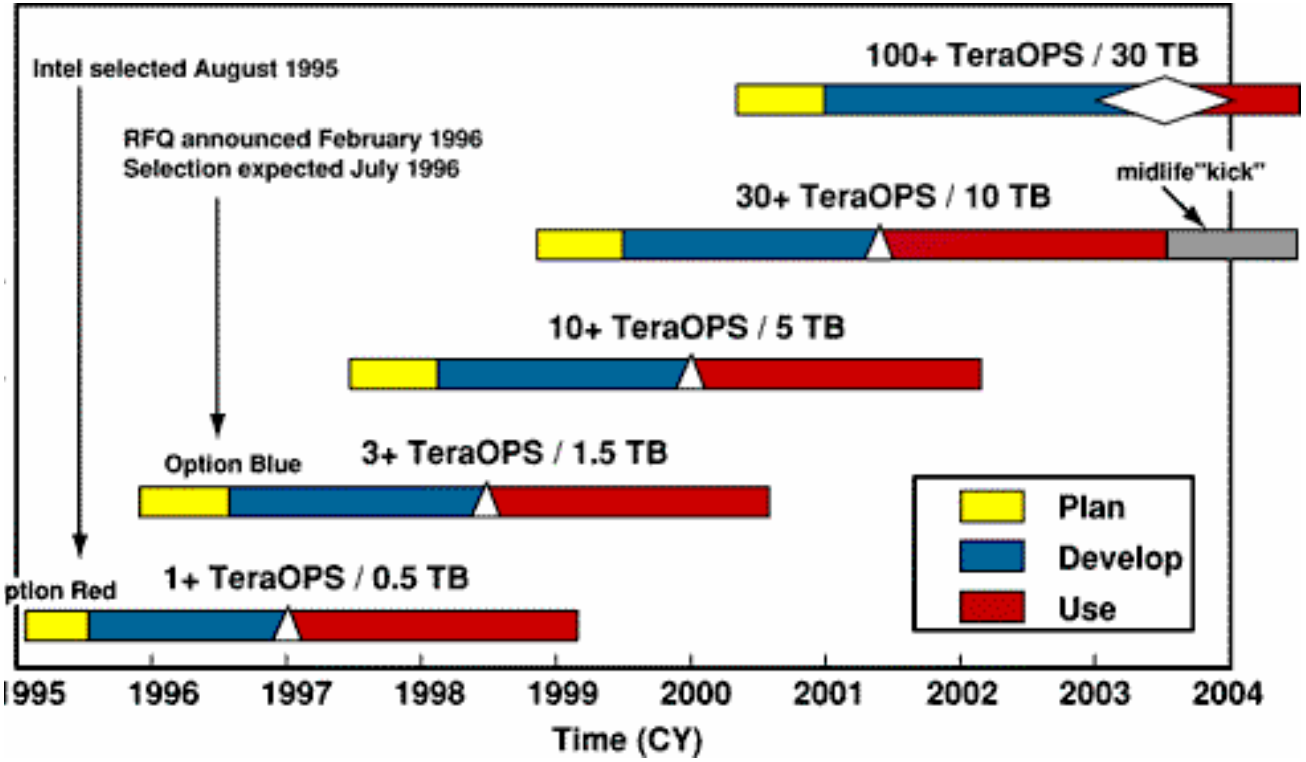
Accelerated Strategic Computing Initiative

http://www.llnl.gov/asc/

（今はASC: Advanced Simulation and Computing）

- もともとは核実験のシミュレーションによる代替
  - そのために必要なハードウェア，ソフトウェア，周辺技術を開発
- 1995年から10年計画で，1 TFLOPSから100 TFLOPSまでの並列計算機を開発：90年代のHW開発をリード
- COTS戦略（Commercial off-the-shelf）
  - 「地球シミュレータ」のように特殊なハードウェアを開発するのではなく，PCなど民生用のプロセッサを使用する。
    - IBM BG/Lは家電機器制御用のプロセッサを使用している。
  - 種類が多いので最良のものを選択できる。
    - たとえばスーパーマーケットで牛乳を買うとき・・・

ASCI Platforms



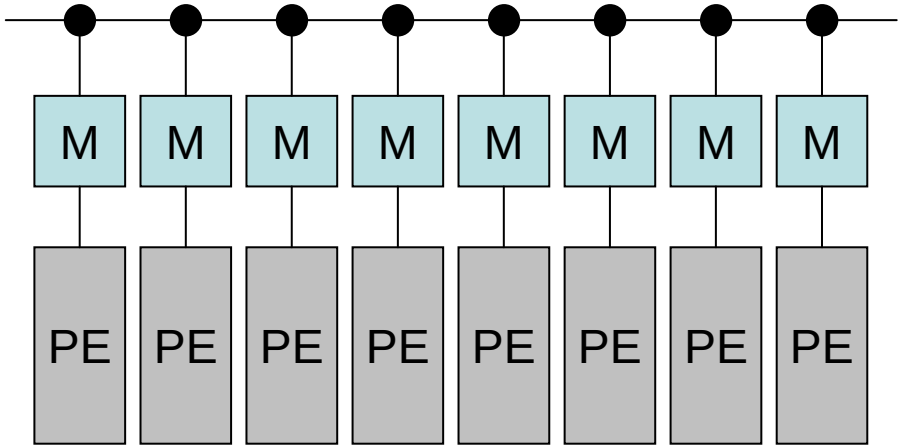
最終ハードウェア（ASCI Purple）は2005年完成
当初の予定とはかなり変わった

http://www.llnl.gov/asci/

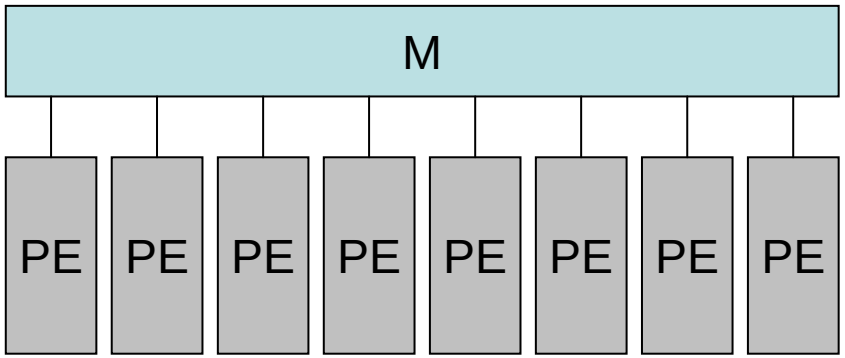
計算機ハードウェアの分類

- 個々のプロセッサ
  - ベクトルプロセッサ
  - スカラープロセッサ：Pentium, Power, Alpha, Itanium, Opteron
- 並列計算機の種類
  - PCクラスタ
    - 最近のPCクラスタは「PCをつなぎ合わせたもの」とは言えないが・・・
  - 専用並列計算機
  - 専用機・・・GRAPE等
- 並列計算機のアーキテクチュア（architecture）
  - 分散メモリ型並列計算機
  - SMP（Symmetrical Multi Processor）：共有メモリ型並列計算機
  - SMPクラスタ型並列計算機（「Constellation」とも言う）

並列計算機の分類



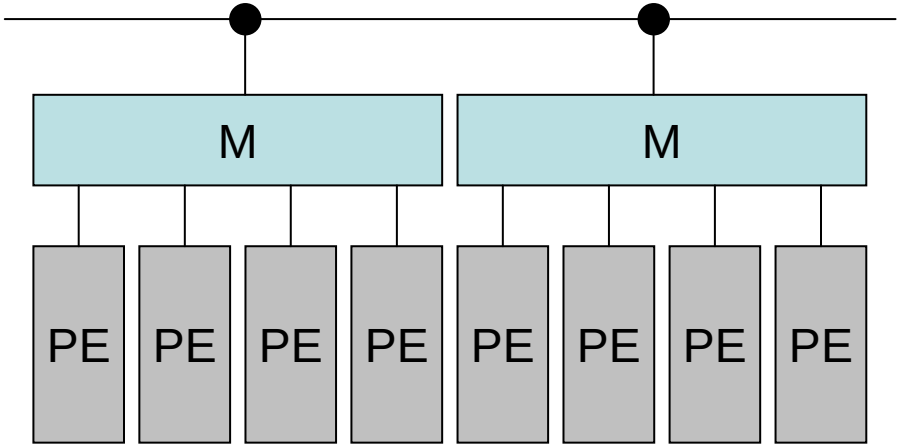
分散メモリ型並列計算機（Hitachi SR2201, Cray T3E）
現在はこのアーキテクチュアのマシンはほとんどない



SMP：共有メモリ型並列計算機（SGI Altix ）

それぞれのアーキテクチュアに応じた
並列プログラミングスタイルがあるのだが，
ここでは話が複雑になりすぎるので触れ
ない。

*PE: Processing Element



SMPクラスタ（地球シミュレータ，IBM-SP，PCクラスタ
（dual processor, dual core））

TOP 500 List

http://www.top500.org/

- 年2回更新
- LINPACKと言われるベンチマークテストを実施する。
  - 密行列を係数とする連立一次方程式を解く
  - ベクトル機でもスカラー機でも性能が出やすい
  - 例えば地球シミュレータはピーク性能40TFLOPS（設計値）に対して35 TFLOPS以上の性能が出ている。
- 実際のアプリケーションではこれほどの性能は出ない
  - 差分法，スペクトル法系の手法：ピーク性能の60%程度
    - AFES on the Earth Simulator： 26 TFLOPS（ピーク性能の65%）
  - 有限要素法
    - GeoFEM on the Earth Simulator（512ノード）：10 TFLOPS（30%）
    - スカラー機ではこれほど出ない：5%～10%

Rank	Site	Computer	Processors	Year	R_{max}	R_{peak}
1	DOE/NNSA/LLNL United States	BlueGene/L - eServer Blue Gene Solution, IBM	131072	2005	280600	367000
2	NNSA/Sandia National Laboratories United States	Red Storm - Sandia/ Cray Red Storm, Opteron 2.4 GHz dual core, Cray Inc.	26544	2006	101400	127411
3	IBM Thomas J. Watson Research Center, United States	BGW - eServer Blue Gene Solution,IBM	40960	2005	91290	114688
4	DOE/NNSA/LLNL, United States	ASC Purple - eServer pSeries p5 575 1.9 GHz, IBM	12208	2006	75760	92781
5	Barcelona Supercomputing Center Spain	MareNostrum - BladeCenter JS21 Cluster, PPC 970, 2.3 GHz, Myrinet, IBM	10240	2006	62630	94208
6	NNSA/Sandia National Laboratories United States	Thunderbird - PowerEdge 1850, 3.6 GHz, Infiniband, Dell	9024	2006	53000	64972.8
7	Commissariat a l'Energie Atomique (CEA), France	Tera-10 - NovaScale 5160, Itanium2 1.6 GHz, Quadrics, Bull SA	9968	2006	52840	63795.2
8	NASA/Ames Research Center/NAS United States	Columbia - SGI Altix 1.5 GHz, Voltaire Infiniband, SGI	10160	2004	51870	60960
9	GSIC Center, Tokyo Institute of Technology, Japan	TSUBAME Grid Cluster, NEC/Sun	11088	2006	47380	82124.8
10	Oak Ridge National Laboratory United States	Jaguar - Cray XT3, 2.6 GHz dual Core, Cray Inc.	10424	2006	43480	54204.8

R_{max}：実効性能（GFLOPS）

R_{peak}：ピーク性能（GFLOPS）

http://www.top500.org/

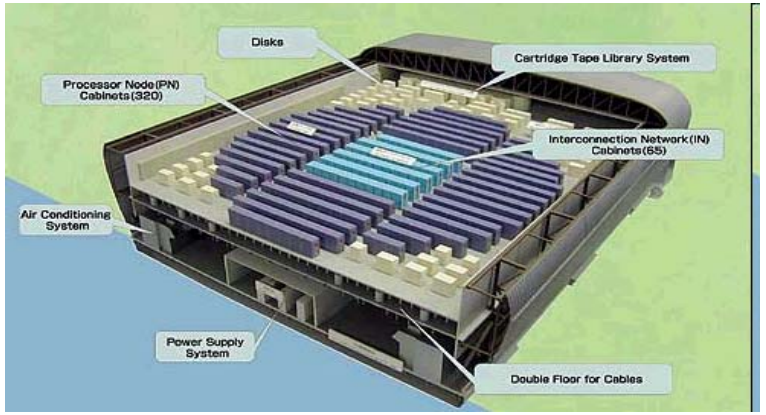
Rank	Site	Computer	Processors	Year	R_{max}	R_{peak}
491	Government Germany	Xeon 3.4 GHz, Myrinet, HP	512	2005	2738.6	3481.6
492	Government Germany	Xeon 3.4 GHz, Myrinet, HP	512	2005	2738.6	3481.6
493	Government United States	Xeon 3.4 GHz, Myrinet, HP	512	2005	2738.6	3481.6
494	Sandia National Laboratories United States	Xeon 3.4 GHz, Myrinet, HP	512	2005	2738.6	3481.6
495	Petroleum Company (G) Saudia Arabia	Xeon 3.06 GHz, GigEthernet, HP	800	2005	2736.9	4896
496	Petroleum Company (G) Saudia Arabia	Xeon 3.06 GHz, GigEthernet, HP	800	2005	2736.9	4896
497	Petroleum Company (G) Saudia Arabia	Xeon 3.06 GHz, GigEthernet, HP	800	2005	2736.9	4896
498	Telecommunication Company United States	Xeon 3.06 GHz, Gig-Ethernet, HP	800	2005	2736.9	4896
499	Telecommunication Company United States	Xeon 3.06 GHz, Gig-Ethernet, HP	800	2005	2736.9	4896
500	Telecommunication Company United States	Xeon 3.06 GHz, Gig-Ethernet, HP	800	2005	2736.9	4896

http://www.top500.org/

Earth Simulator (ES)

http://www.es.jamstec.go.jp/

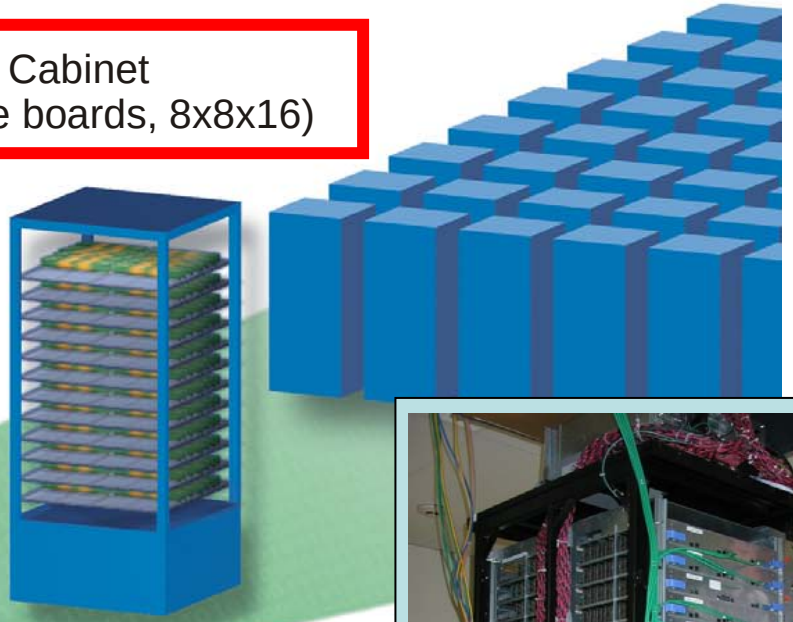
- 640×8= 5,120 Vector Processors
  - **SMP Cluster-Type Architecture**
  - 8 GFLOPS/PE
  - 64 GFLOPS/Node
  - 40 TFLOPS/ES
- 16 GB Memory/Node, 10 TB/ES
- 640×640 Crossbar Network
  - 16 GB/sec×2
- Memory BWTH with 32 GB/sec.
- **35.6 TFLOPS for LINPACK (2002-March)**
- **26 TFLOPS for AFES (Climate Simulation)**



BlueGene/L

System
(64 cabinets, 64x32x32)

Cabinet
(32 Node boards, 8x8x16)



Node Board
(32 chips, 4x4x2)
16 Compute Cards

Compute Card
(2 chips, 2x1x1)

Chip
(2 processors)

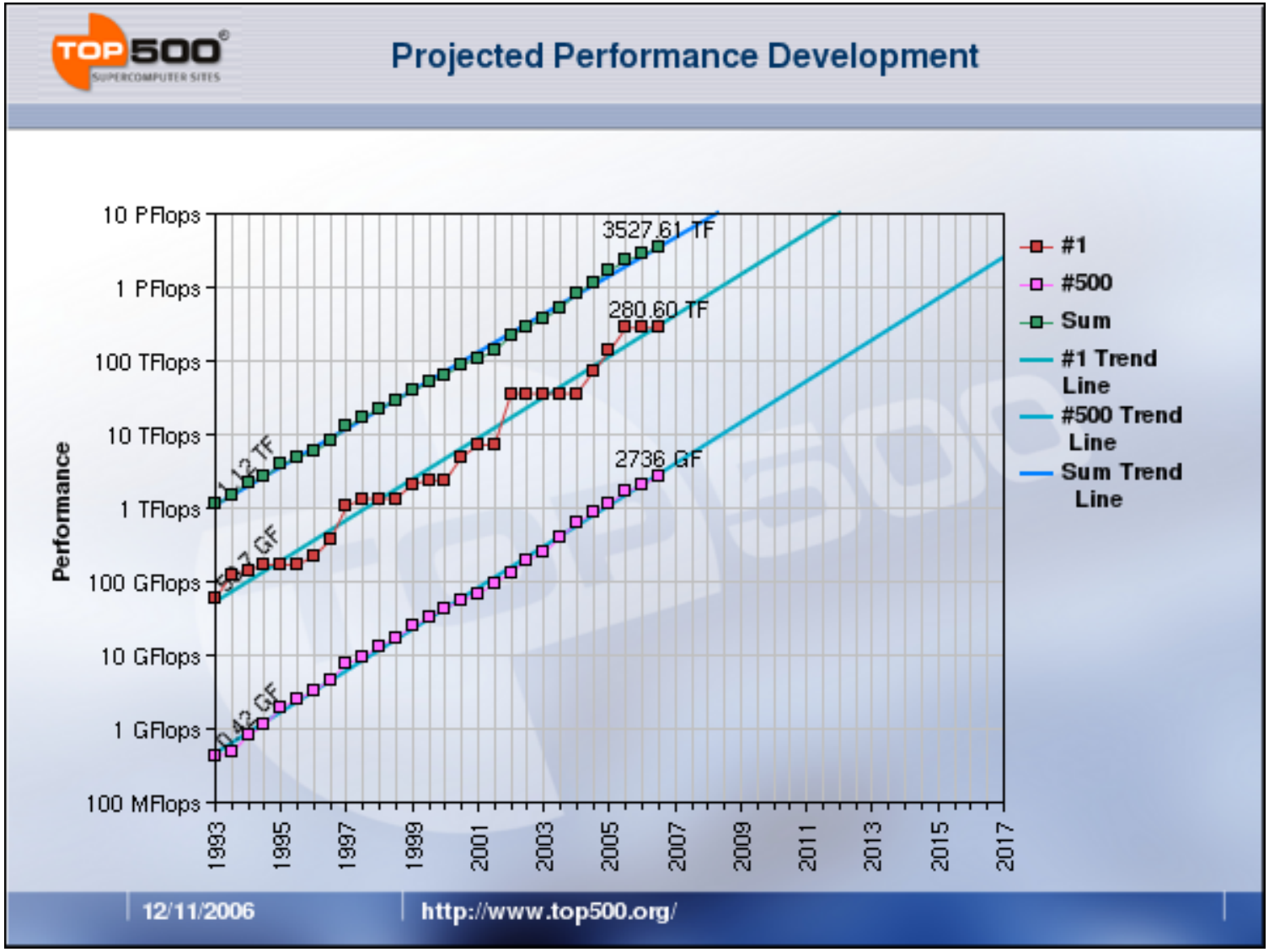
2.8/5.6 GF/s
4 MB

5.6/11.2 GF/s
0.5 GB DDR

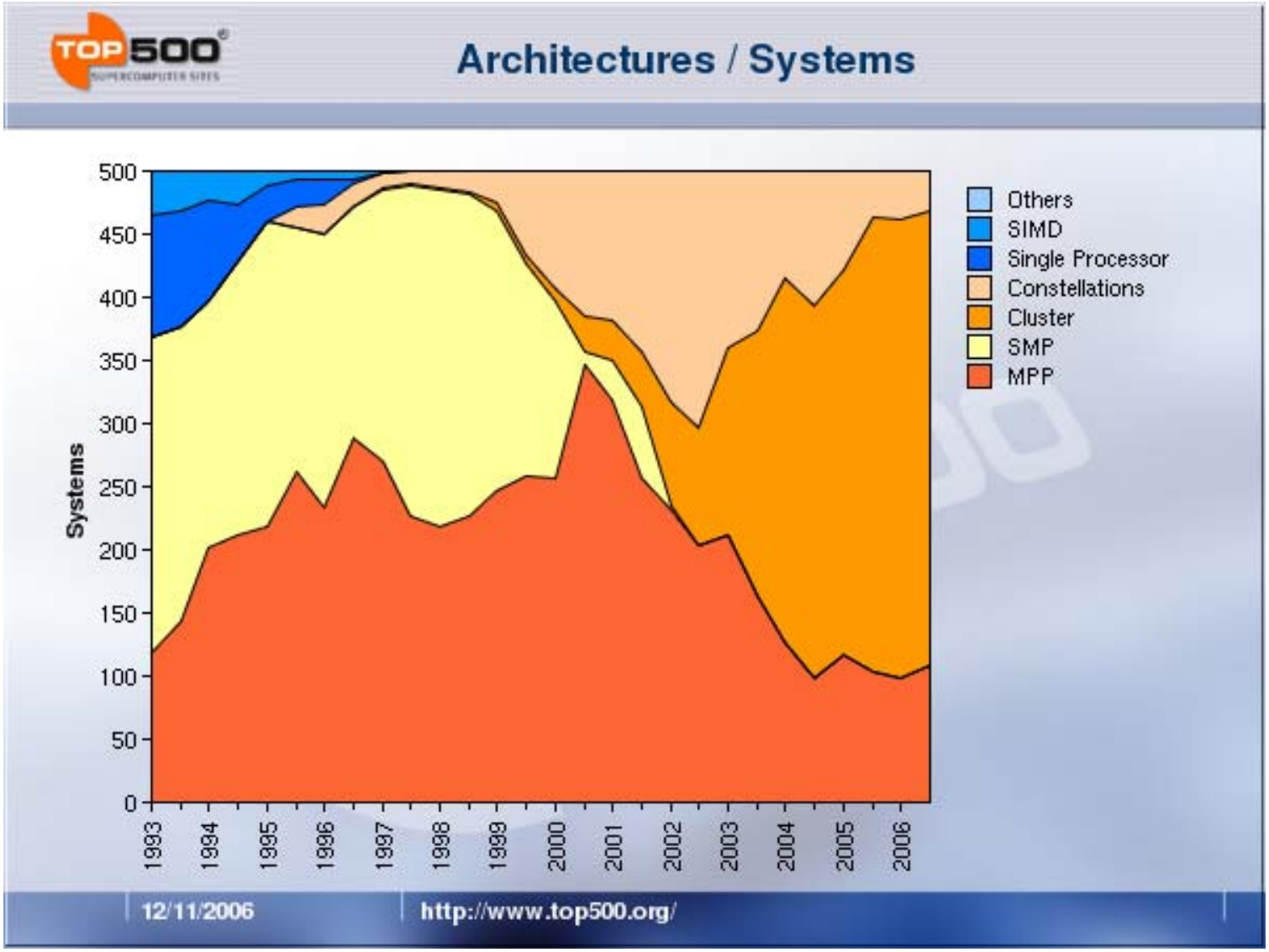
90/180 GF/s
8 GB DD

2.9/5.7 TF/s
256 GB DDR

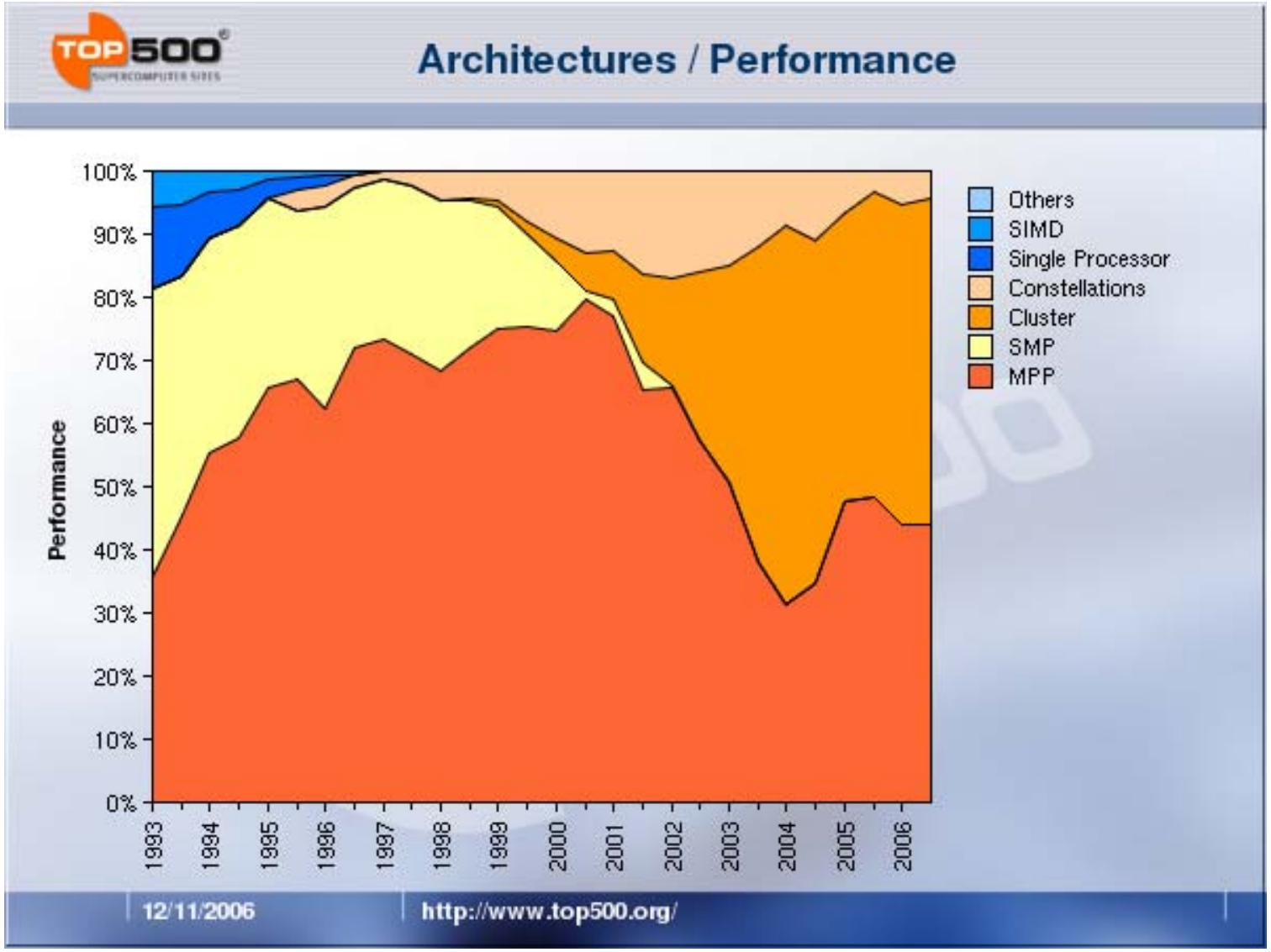
October 2003
BG/L half rack prototype
500 Mhz
512 nodes/1024 proc.
2 TFlop/s peak
1.4 Tflop/s sustained



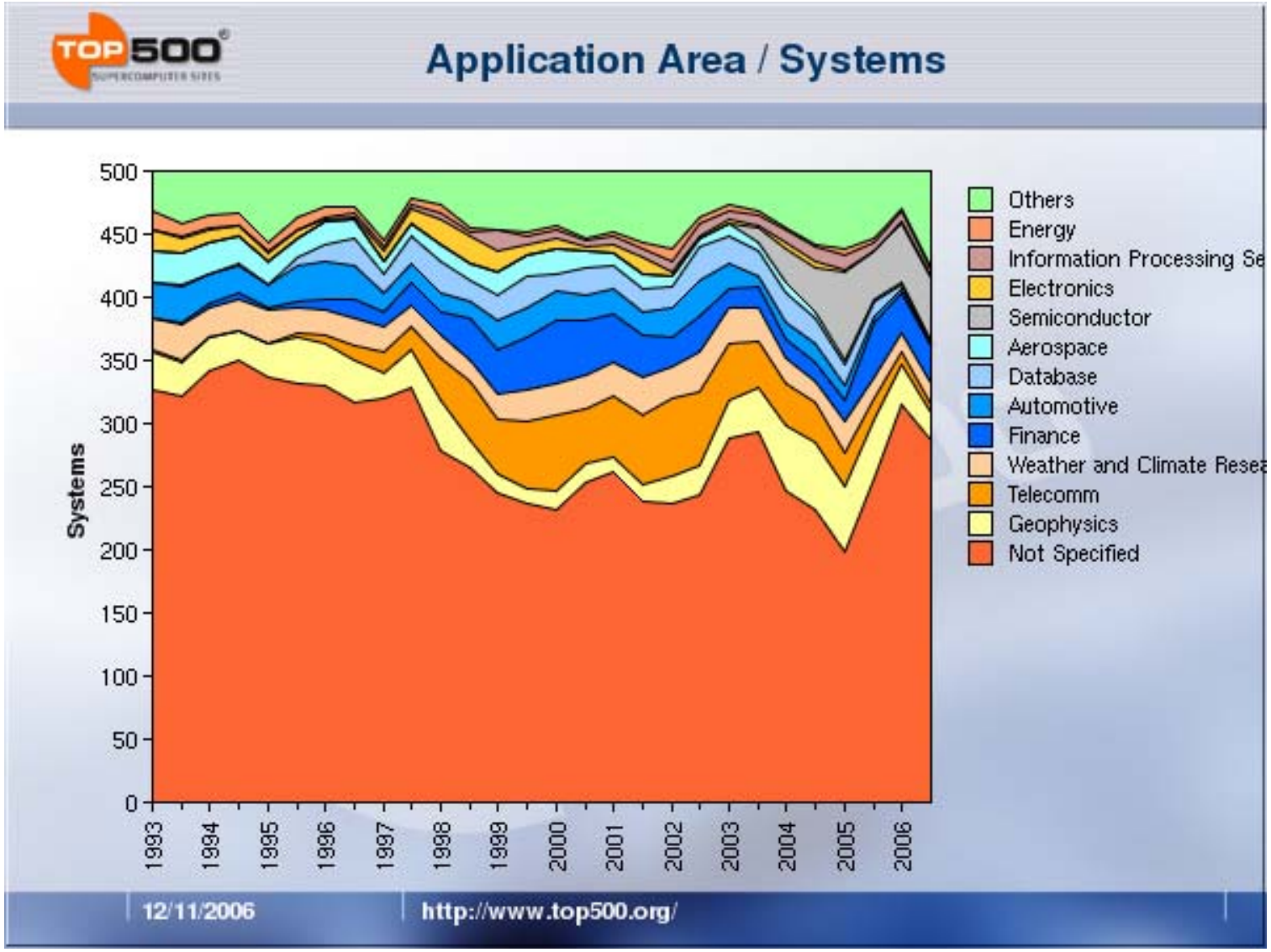
http://www.top500.org/



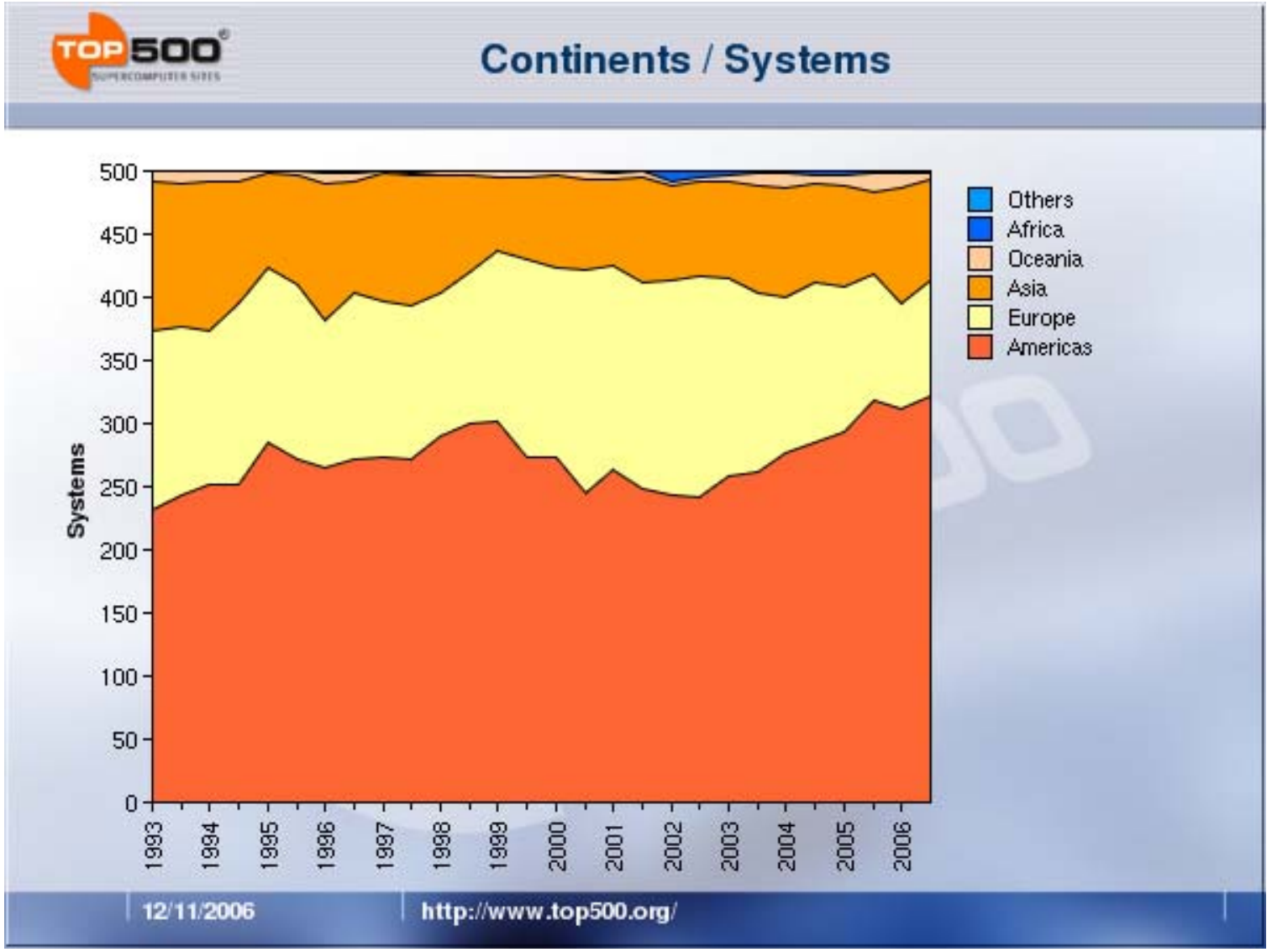
http://www.top500.org/



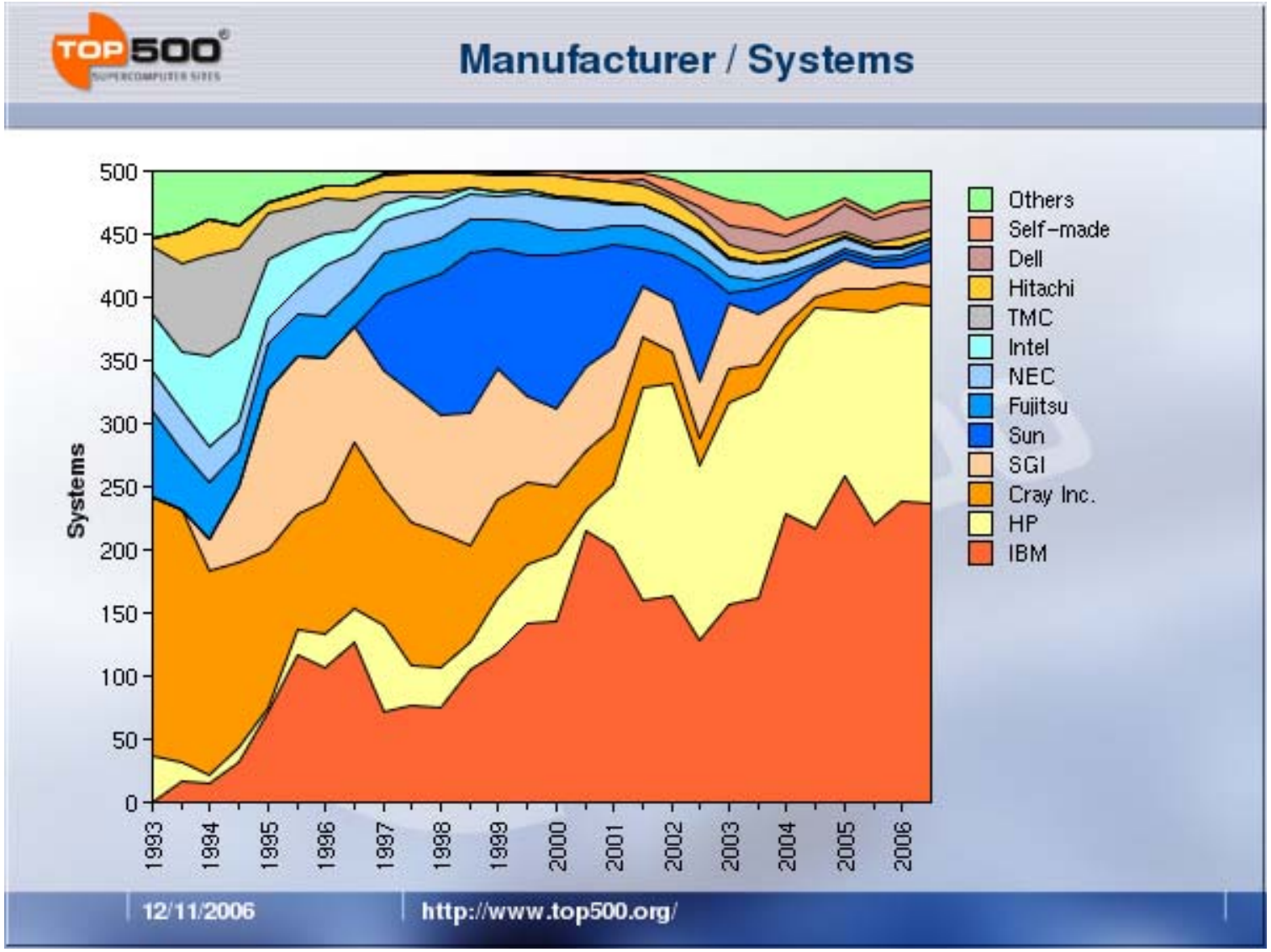
http://www.top500.org/



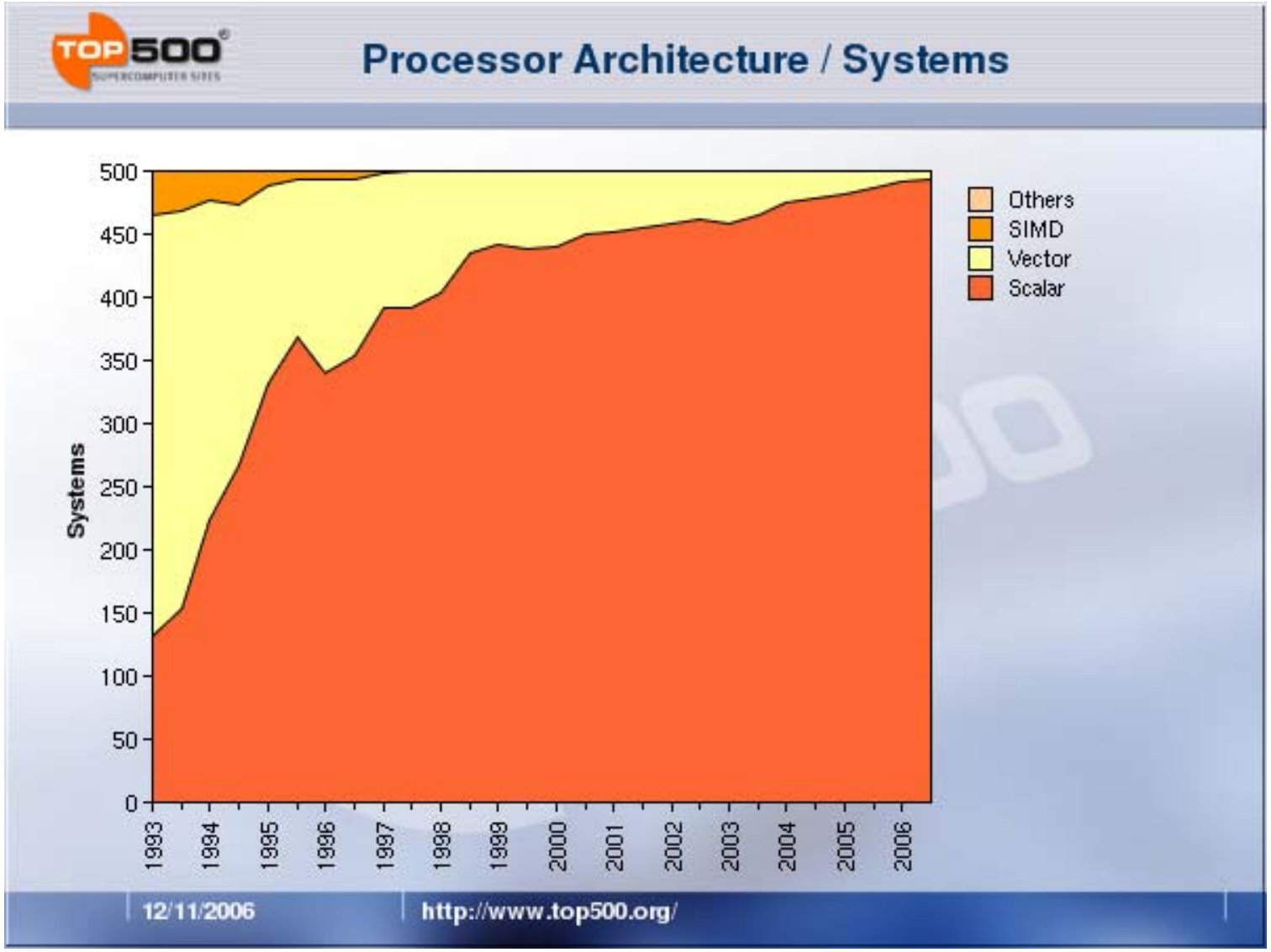
http://www.top500.org/



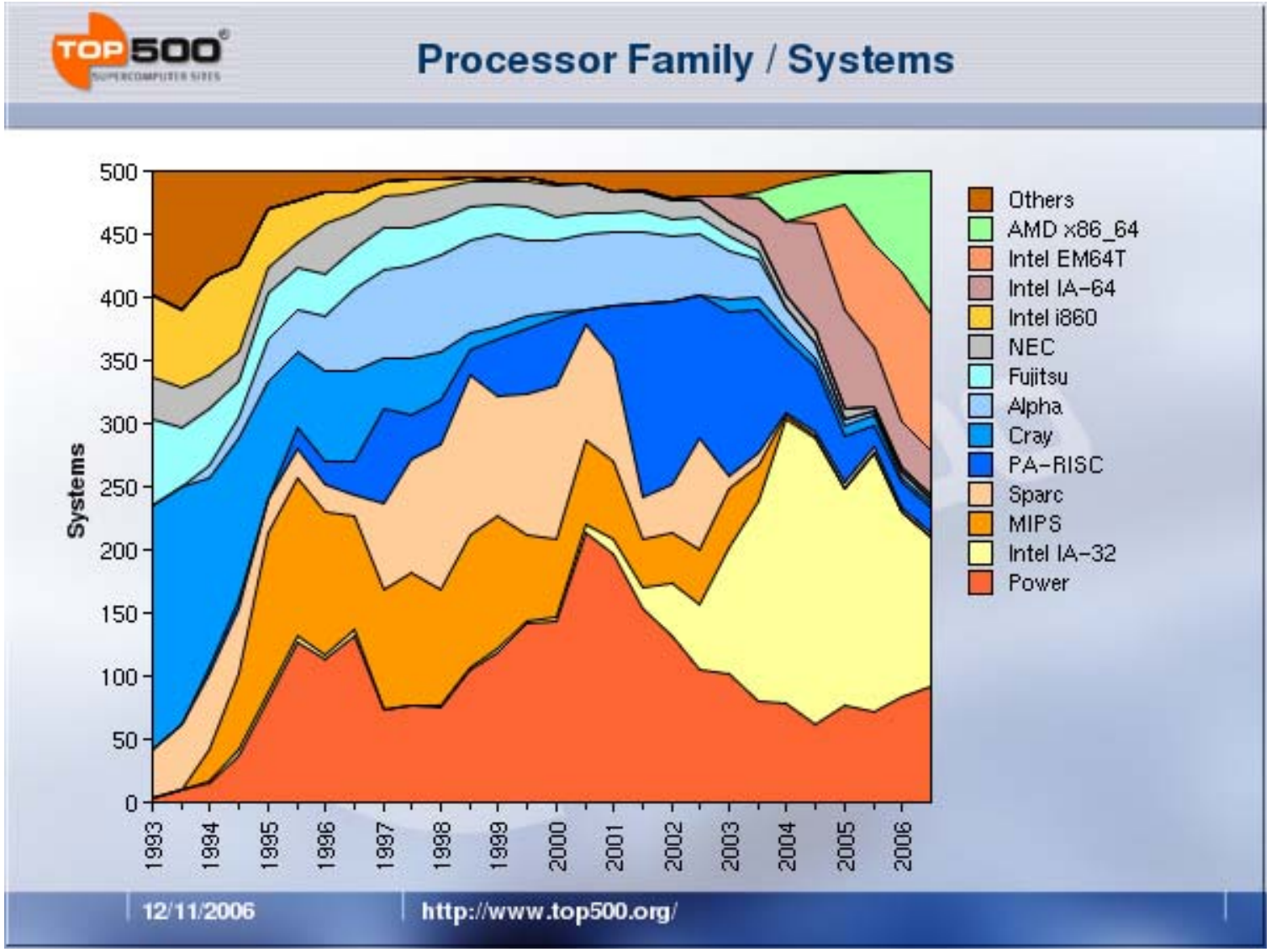
http://www.top500.org/



http://www.top500.org/

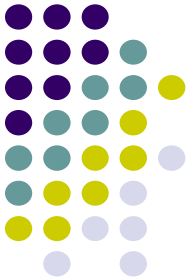


http://www.top500.org/



http://www.top500.org/

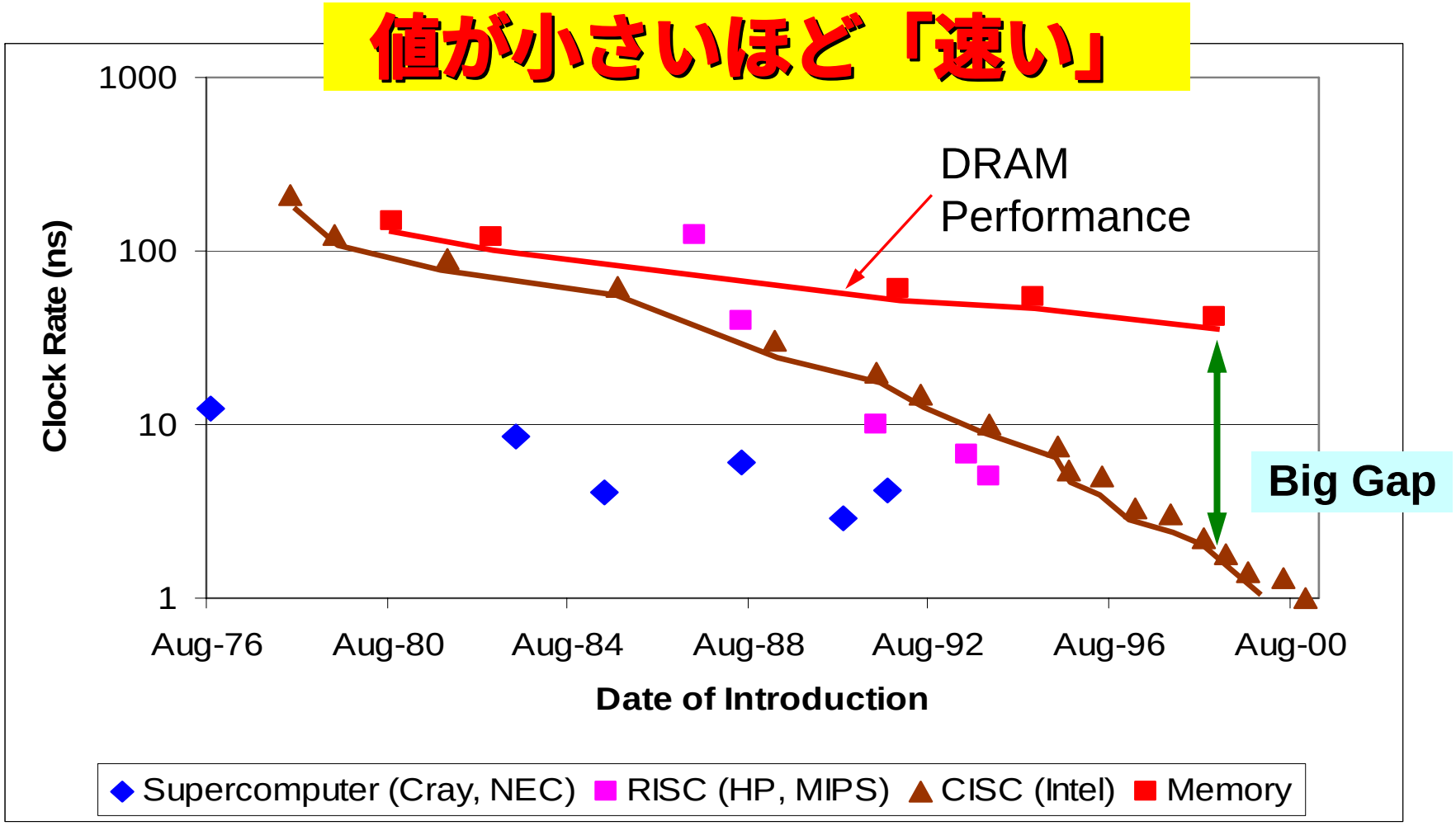
プロセッサの動向：スカラーとベクトル



- スカラープロセッサ
  - クロック数とメモリバンド幅のギャップ
  - 低い対ピーク性能比
    - 例：IBM Power-3，Power-4，FEM型アプリケーション → 5-8 %

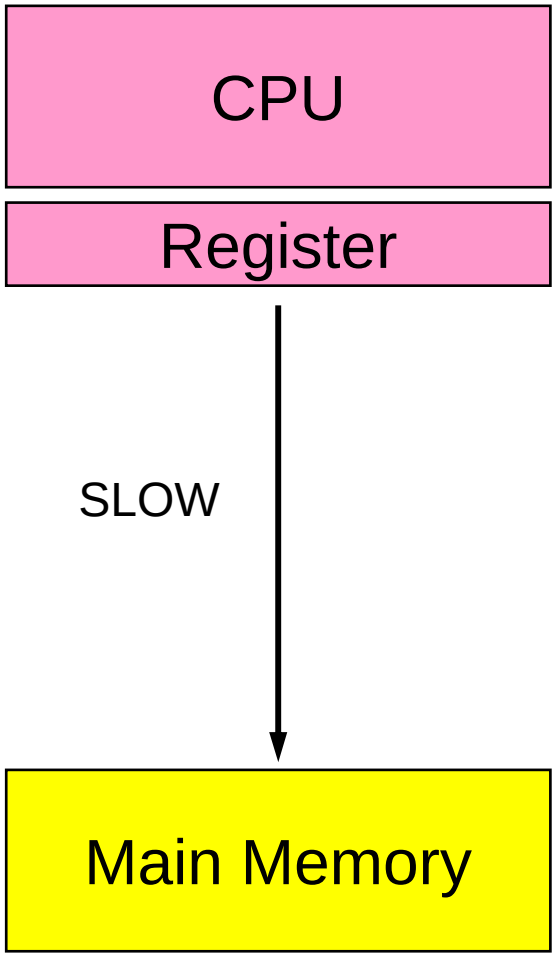
CPU and Memory Performance

Bill Gropp (ANL) "Algorithm and Architecture"
SIAM CSE03, February 2003, San Diego, CA.



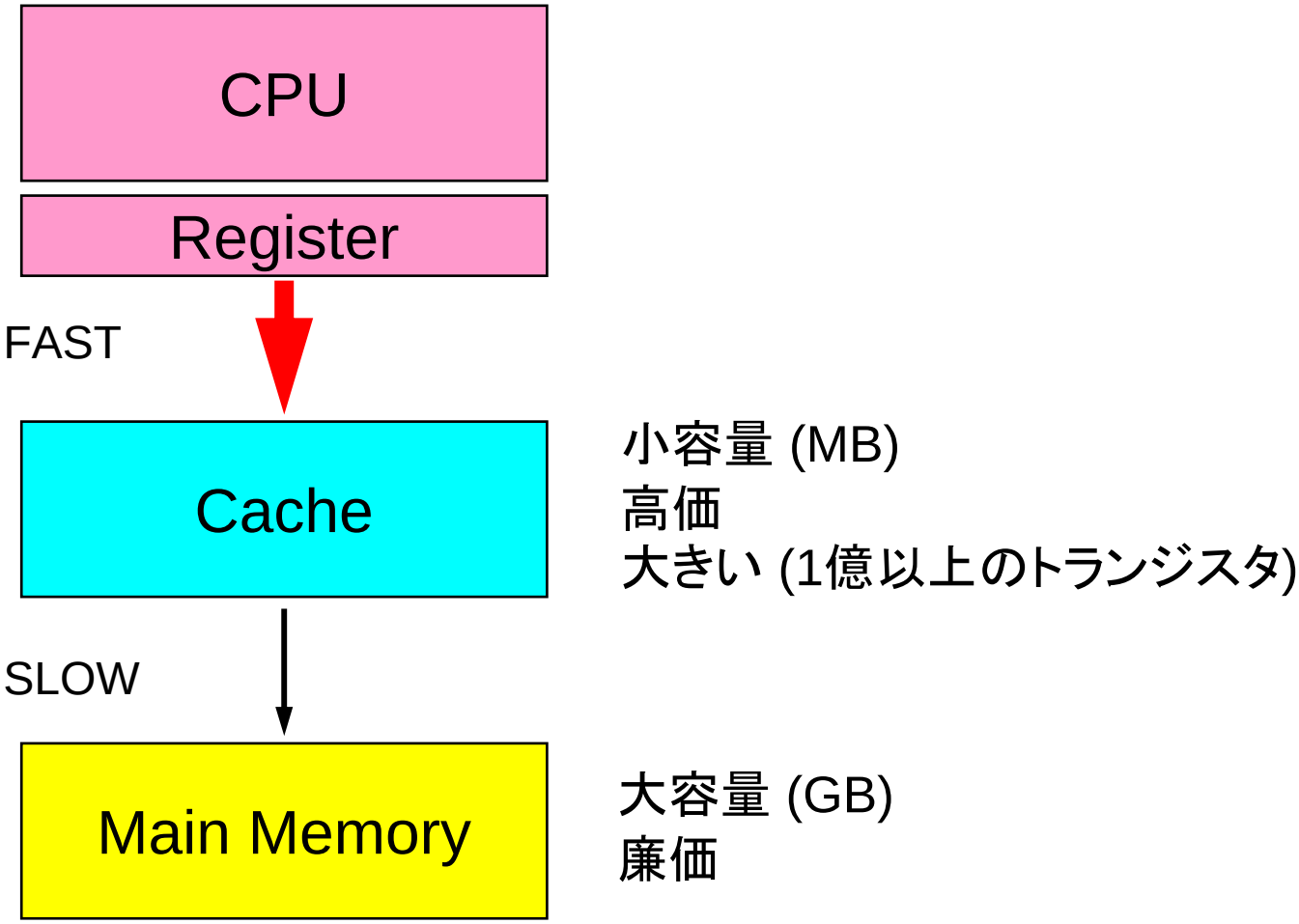
スカラープロセッサ

メモリへ直接アクセスするのは実際的でない

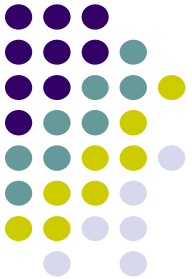


スカラープロセッサ

CPU–キャッシュ–メモリの階層構造

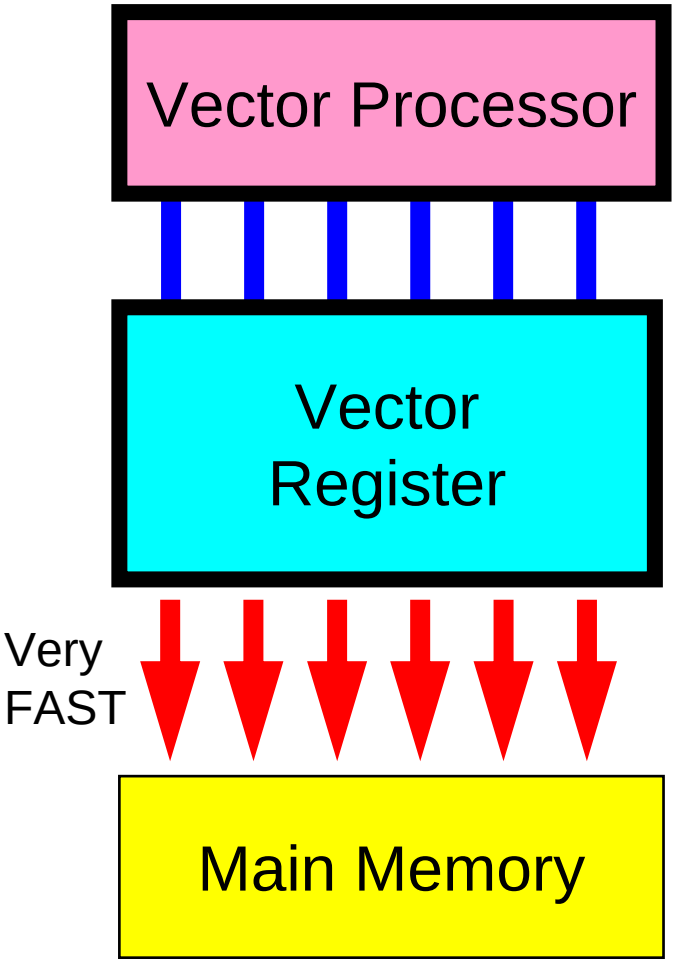


プロセッサの動向：スカラーとベクトル



- スカラープロセッサ
  - クロック数とメモリバンド幅のギャップ
  - 低い対ピーク性能比
    - 例：IBM Power-3, Power-4, FEM型アプリケーション → 5-8 %
- ベクトルプロセッサ
  - 高い対ピーク性能比
    - 例：地球シミュレータ, FEM型アプリケーション → >35 %
  - そのためには・・・
    - ベクトルプロセッサ用チューニング
    - 充分長いベクトル長（問題サイズ）
  - 比較的単純な問題に適している

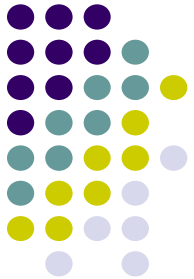
ベクトルプロセッサ
ベクトルレジスタと高速メモリ



- 単純構造のDOループの並列処理
- 単純，大規模な演算に適している

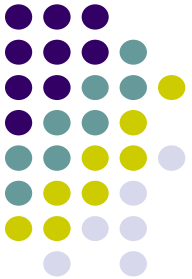
```
do i= 1, N
  A(i)= B(i) + C(i)
enddo
```

各種ハードウェアの比較



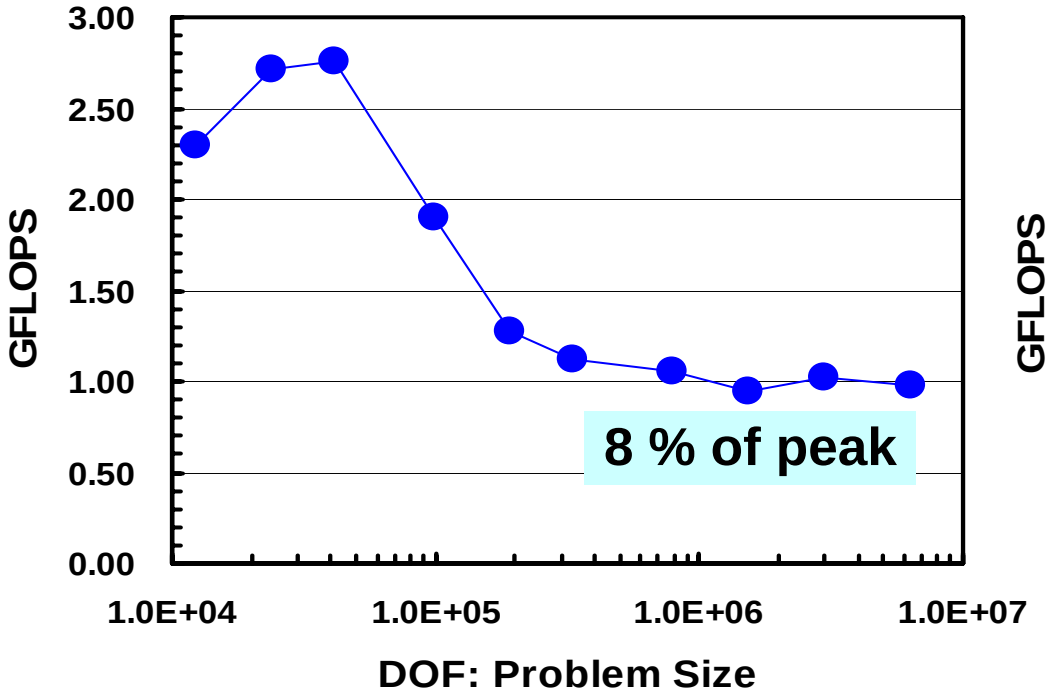
	Earth Simulator	Hitachi SR8000 (U.Tokyo)	IBM-SP3 (LBNL)	IBM p5-575 (LBNL)	IBM BG/L-proto (Prototype)
PE#/node	8	8	16	8	2
Clock rate (MHz)	500	450	375	1,900	500
Peak Performance (GFLOPS/PE)	8.00	1.80	1.50	7.60	1.00 (w/singe FPU)
Memory Size (GB/node)	16	16	16~64	32	0.256
Peak Memory BW (GB/sec/node)	256	32	16	100	3.4
Network Topology	single stage crossbar	3D crossbar	Switch	Switch	3D Torus
Network BW (GB/sec/node)	12.3	1.6	1.0	32.0	1.32
MPI Latency (μsec)	5.6-7.7	6-20	16.3	3.0	6.0

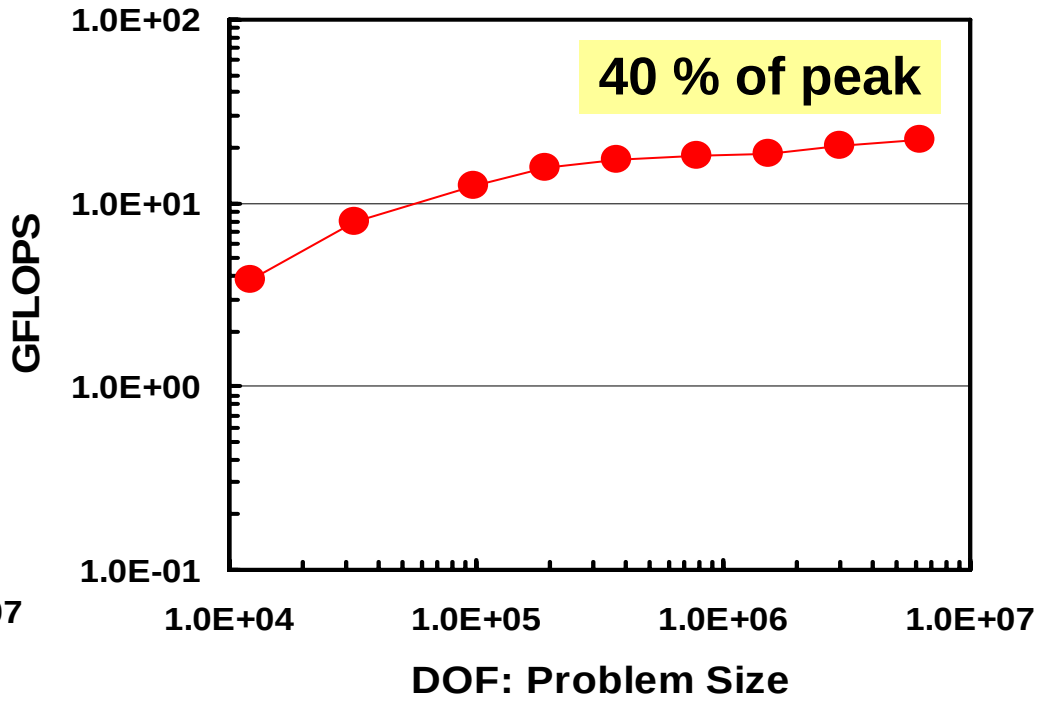
プロセッサの動向：スカラーとベクトル



- スカラープロセッサ
  - クロック数とメモリバンド幅のギャップ
  - 低い対ピーク性能比
    - 例：IBM Power-3, Power-4, FEM型アプリケーション → 5-8 %
- ベクトルプロセッサ
  - 高い対ピーク性能比
    - 例：地球シミュレータ, FEM型アプリケーション → >35 %
  - そのためには・・・
    - ベクトルプロセッサ用チューニング
    - 充分長いベクトル長（問題サイズ）
  - 比較的単純な問題に適している

典型的な挙動

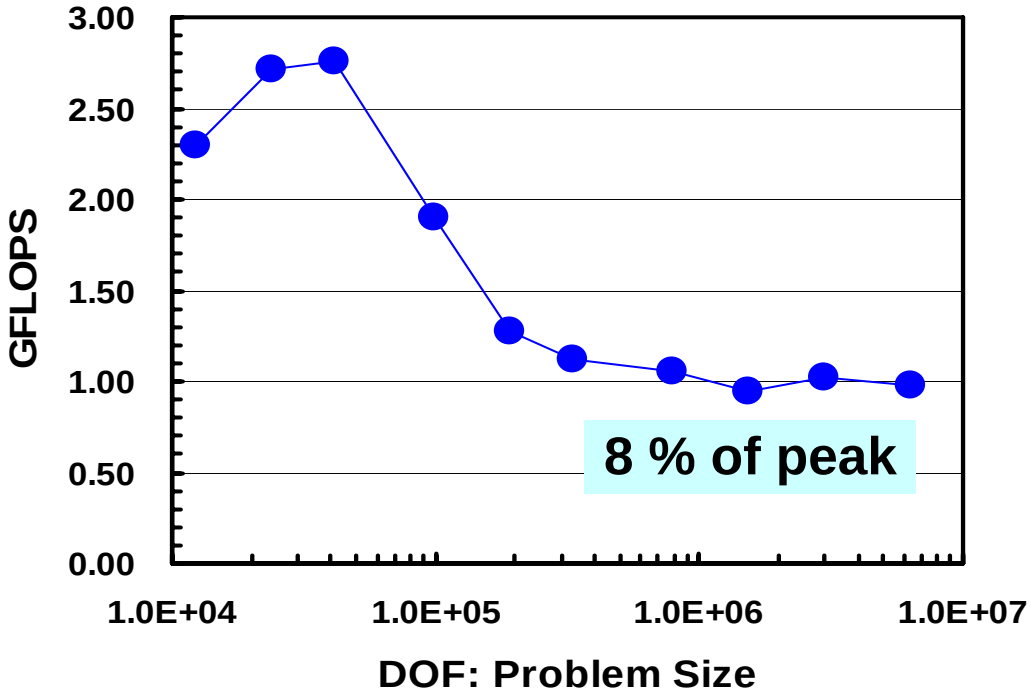


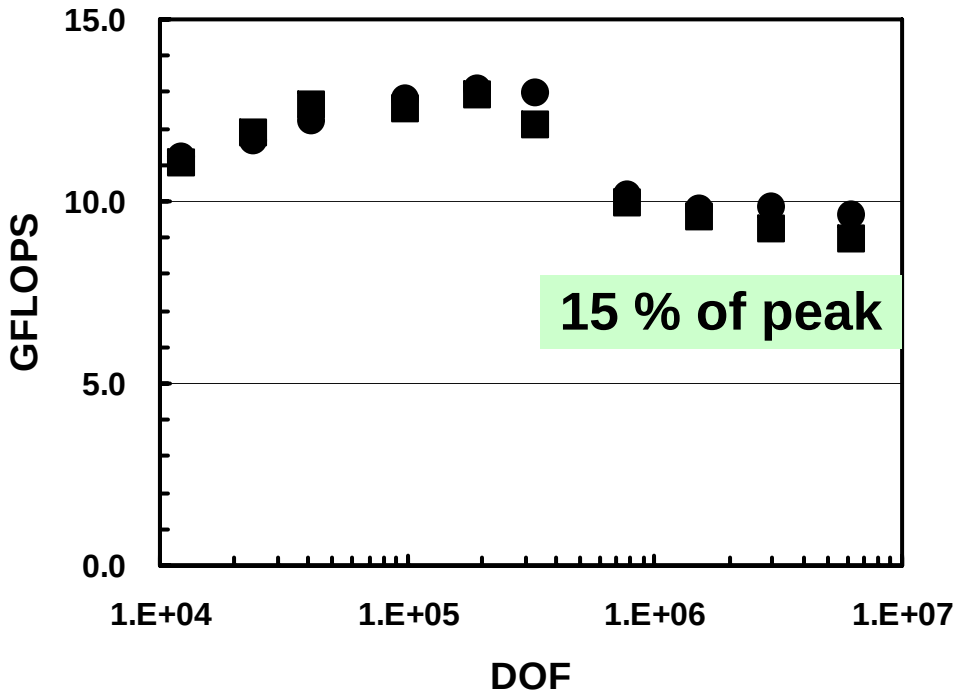


IBM-SP3:
問題サイズが小さい場合はキャッシュの影響のため性能が良い

Earth Simulator:
大規模な問題ほどベクトル長が長くなり，性能が高い

典型的な挙動

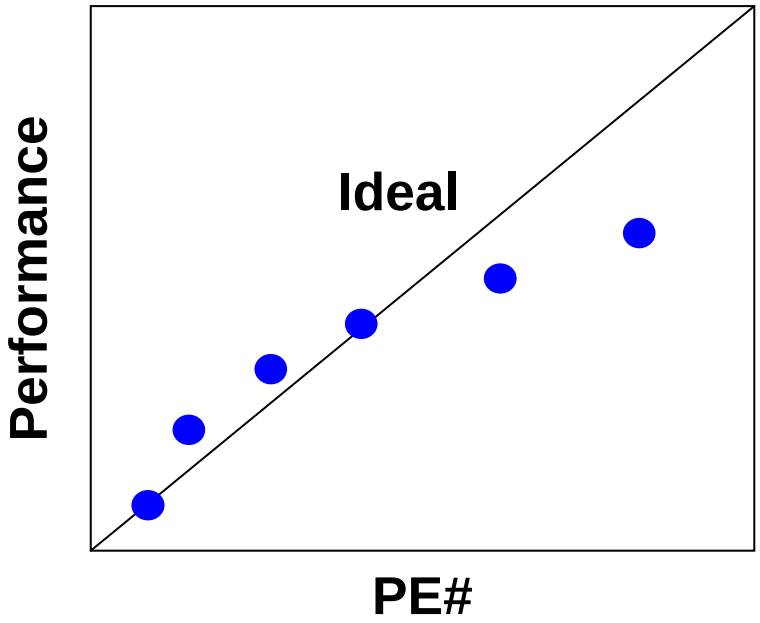


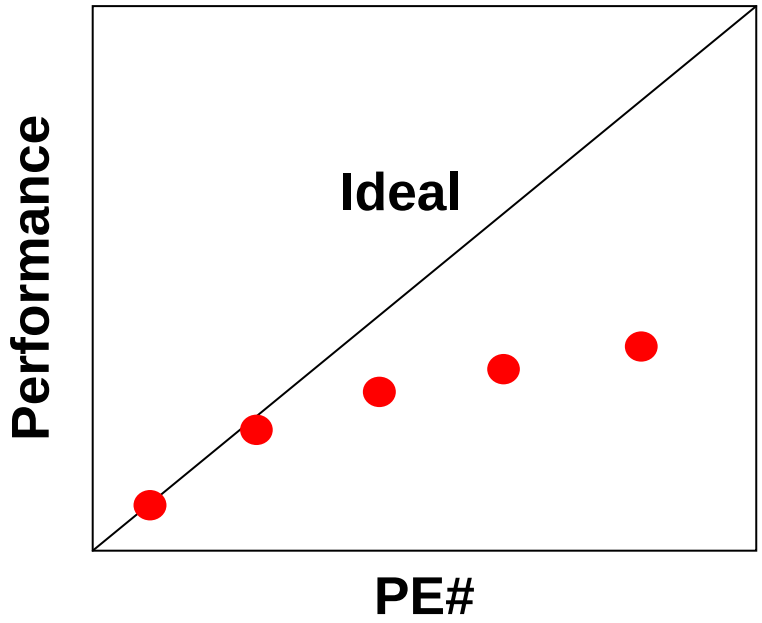


IBM-SP3:
問題サイズが小さい場合はキャッシュの影響のため性能が良い

IBM-p5-575:
キャッシュの影響はあるが，SP3と比べて改善（メモリバンド幅，メモリレイテンシ，キャッシュ容量）

並列計算
Strong Scaling（全体問題規模固定）

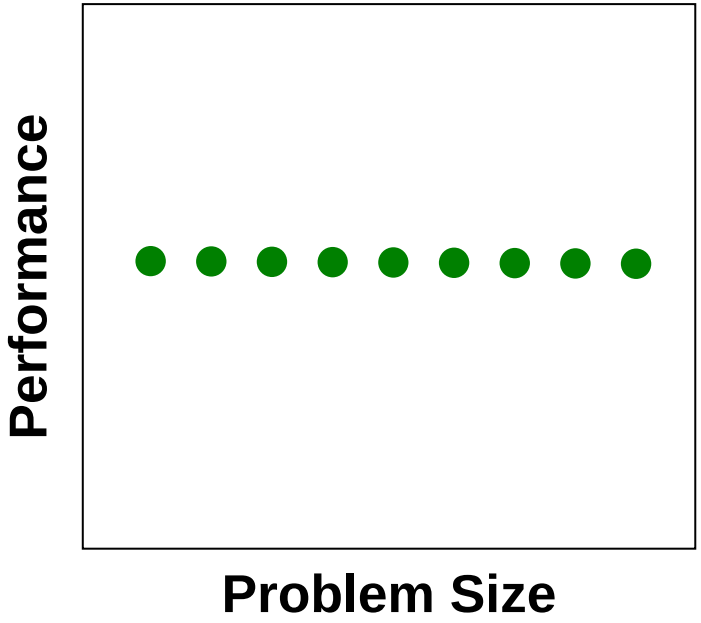


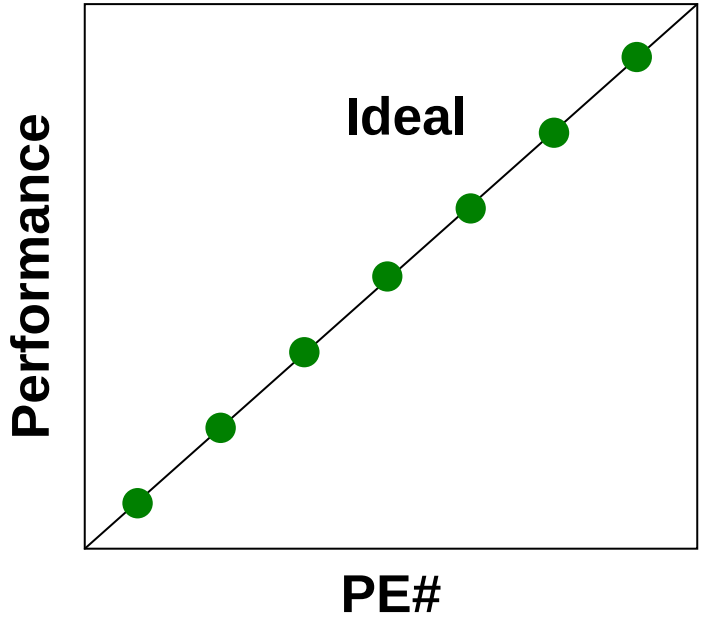


IBM-SP3:
PE（Processing Element）数が少ない場合はいわゆるスーパースカラー。
PE数が増加すると通信オーバーヘッドのため性能低下。

Earth Simulator:
PE数が増加すると，通信オーバーヘッドに加え，PEあたりの問題規模が小さくなるため性能低下。

理想の並列計算機システム





広い範囲の問題規模（もちろんアプリケーションの種類）にわたって一定の（高い）性能。

通信のオーバーヘッド少ない

まとめ

- ハードウェアの進展
  - Mooreの法則，ASCI，TOP500
  - 地球シミュレータ，京速計算機
- 並列計算機のアーキテクチュア
  - 分散メモリ
  - SMP，SMPクラスタ
- プロセッサの動向
  - ベクトル，スカラー
    - 性能を出すためにはチューニング必要（詳細は6月14日授業）
    - 両者に得意，不得意な計算がある（これも6月14日）
  - Cell

補 足

- 専用計算機
  - GRAPE（GRAvity PipE ）
    - 宇宙物理学におけるN対N問題用専用ハードウェア：MD，境界要素法等
    - コストパフォーマンス
    - http://grape.astron.s.u-tokyo.ac.jp/grape/
  - MDGRAPE
    - MD（Molecular Dynamics）専用のGRAPE
    - MDGRAPEを通常のクラスタのAcceleratorのように使用することもできる。
    - http://www.riken.go.jp/r-world/info/release/press/2007/070327/index.html
- 省電力
  - PCクラスタも意外に電気を食う ⇒ 空調の問題
  - ハードウェア，アルゴリズム，オペレーションに関する研究
  - dual core, quad core

並列プログラミングで重要なこと

- 最も重要なことは，まず，良い単体CPU用のプログラムを開発することである。そして，中で起こっていること全てに精通していること。
- 精度，安定性の検証
  - これができていないと，並列にすると答えが変わったりする。
- 単体チューニング
  - まず，単体CPU単位で充分な性能が出ている必要がある。
- 高いモジュラリティ
  - プログラムそのものの読み易さ。
  - 並列，非並列部分の区別：サブルーチン単位
  - 通信部，非通信部の区別：データ

科学技術計算の真髄：SMASH

- Prof. David Levermore
  - Applied Mathematics and Scientific Computation Program University of Maryland
- **SMASH**
  - **Science，Modeling，Algorithm，Software，Hardware**
- 多岐にわたる知識が必要ということもできるし，この順番に重要ということもできる。

- **本講義，演習ではSoftware全般と，AlgorithmとHardwareの一部をカバーする。**
  - **Science, ModelingとAlgorithmの大半は諸兄の仕事である。**

並列プログラミング言語

- メッセージパッシングライブラリ
  - MPI, PVM
  - 複雑と言われているが，柔軟な処理可能，移植性も大。
- 並列化コンパイラ
  - HPF（High Performance FORTRAN）
  - プログラミングは簡単だが（配列の分割），融通が効かない。複雑な問題に対しては効率も出にくい。移植性も悪い・・・多分実際に使っているのは一部の「地球シミュレータ」ユーザーのみ。
- SMP用ディレクティブ
  - OpenMPなど
  - 共有メモリユニット用の並列化用
  - SMPクラスタではMPIと組み合わせた「ハイブリッド」プログラミング

メッセージパッシングは難しいか？

- そんなことはない。
- 基本的にMPIの関数のうち10個程度を知っておけば，充分なことができる。
- 本授業・演習Iでは基本的にMPIを使います。
  - 演習IIで，OpenMPを少し扱う。

連絡事項

- 「履修用アンケート」できるだけ早く送付してください。
- 「教育用計算機」アカウントの取得も忘れずに。

- 次回
  - 理3-320
  - 教材は各自で印刷のこと