

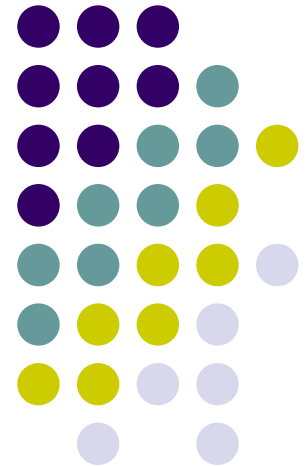
2016年度 実践的並列コンピューティング 第1回

授業の概要と
TSUBAMEスパコンガイダンス

遠藤 敏夫

数理・計算科学系 & 学術国際情報センター

endo@is.titech.ac.jp



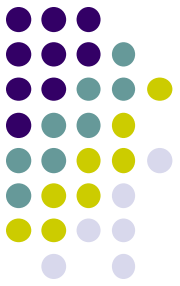
本授業のねらい



- 並列計算・高性能計算技術を実践的に学ぶ
 - 講義＋**実習**
 - この場で実際に東工大TSUBAME2スーパーコンピュータを使います
- 予定内容
 - HPCのためのライブラリや言語
 - OpenMP
 - MPI
 - CUDA (GPUのためのプログラミング)
 - その他追加の可能性あり

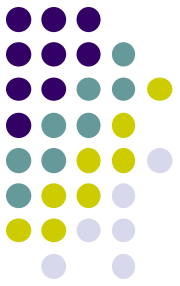
お知らせがあるときはOCW webページにて

- <http://ocw.titech.ac.jp> → 「実践的並列コンピューティング」で検索



必要となる基礎知識

- 本講義ではC言語をベースに
 - ポインタの概念、malloc/free
 - 配列とポインタの関係
 - Pthread, Javaスレッドは知っているが良いが必須ではない
- 基礎的なLinuxコマンド
 - ls, cp, mkdir...
 - コンパイル、makeコマンド

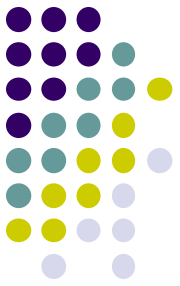


半年の流れと単位認定

- OpenMP: 共有メモリ並列プログラミング
- MPI: 分散メモリ並列プログラミング
- GPUプログラミング
- その他

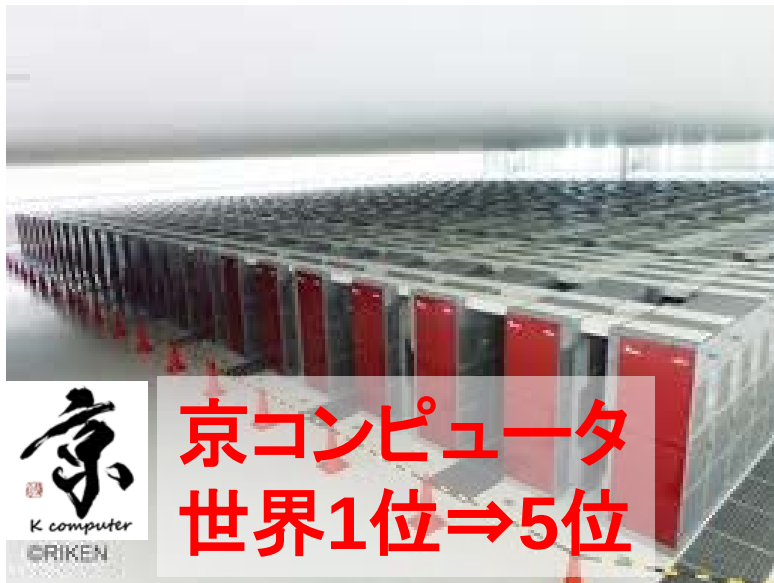
下記により採点・単位認定

- 各パートで課題を出す。2つ以上のパートのレポート提出を必須とする
 - ×切は、パート終了の2~3週間後
- 出席点



スーパーコンピュータとは

色々なスーパーコンピュータ



京コンピュータ
世界1位⇒5位



東工大TSUBAME2
世界4位⇒25位



BlueGene/P



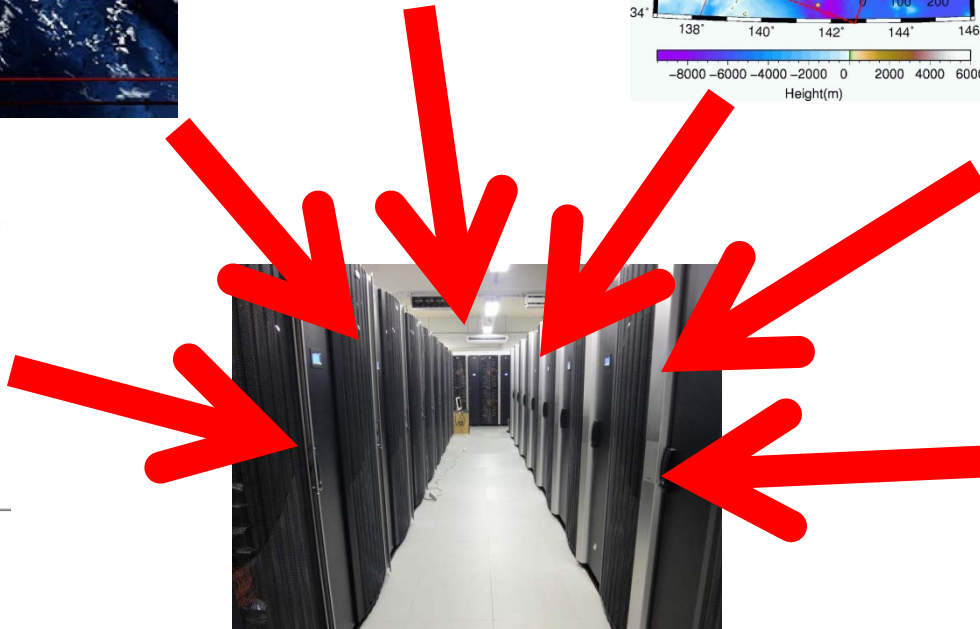
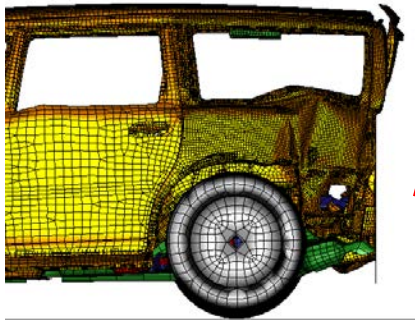
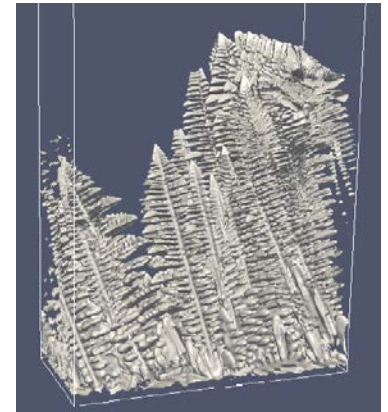
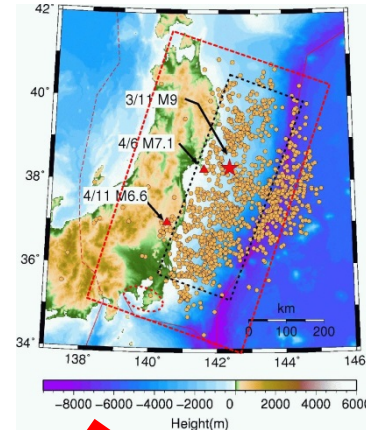
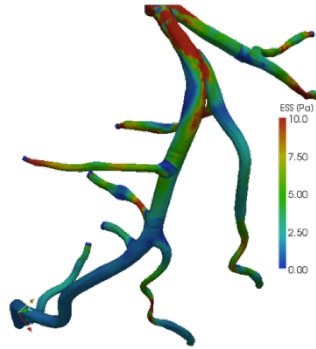
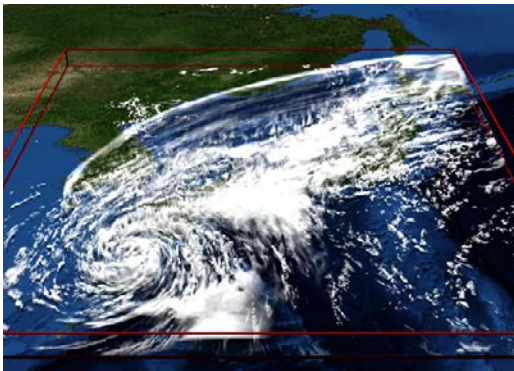
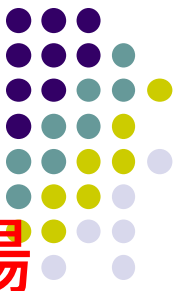
Jaguar

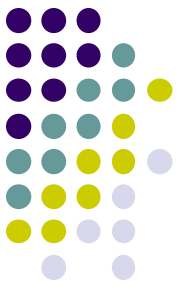


Cray-1 in 1976
160MFlops

スパコンは何に使われる？

スパコンはあらゆる科学分野の**仮想実験場**



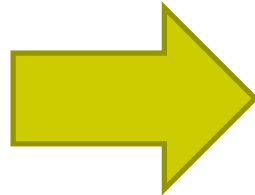


普通のコンピュータとの違い

- スパコンは、普通のコンピュータよりも**ずっと大規模な計算・高速な計算**が可能なコンピュータ
 - 速度の単位はFlops: 一秒間に加・減・乗を何回できるか
 - 倍精度浮動小数(double)の演算速度がポピュラー



約10万倍!



TSUBAME2

約50GFlops (一秒で500億回)

5.7PFlops (一秒で5700兆回)



1980年のパソコン

Cray-1 (1976)
160MFlops

現代の携帯電話

スパコン開発競争の激化—世界—スパコンの変遷



参考: www.top500.org

Linpack演算速度 (Gflops)

Jun-93	CM-5/1024	60	LANL	US
Nov-93	Numerical Wind Tunnel	60	AL	Japan
Jun-94	XP/S140	60	NL	US
Nov-94	Numerical Wind Tunnel	60	AL	Japan
Jun-95	↓	1		Japan
Nov-95	↓			
Jun-96	SR2201			
Nov-96	CP-PACS			
Jun-97	ASCI			
Nov-97	↓			
Jun-98	↓			
Nov-98	↓			
Jun-99	↓	2121	↓	US
Nov-99	↓	2379	↓	US
Jun-00	↓	2379	↓	US
Nov-00	ASCI White	4938	LLNL	US
Jun-01	↓	7226	↓	US
Nov-01	↓	7226	↓	US
Jun-02	Earth-Simulator	35860	ES Center	Japan
Nov-02	↓	35860	↓	Japan
Jun-03	↓	35860	↓	Japan

60GFlops

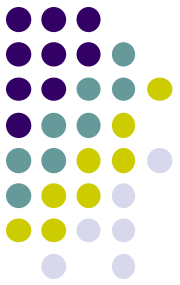
20年間で性能500,000倍
アメリカと日本の一騎打ち
⇒中国の台頭

Nov-03	↓	35860	↓	Japan
Jun-04	↓	35860	↓	Japan
Nov-04	BlueGene/L beta	70720	IBM/DOE	US
Jun-05	BlueGene/L	136800	DOE/NNSA/LLNL	US
Nov-05	↓	280600	↓	US
		280600	↓	US
		280600	↓	US
		280600	↓	US
		478200	↓	US
		1026000	↓	US
		1105000	↓	US
		1105000	↓	US
Nov-09	Jaguar	1759000	ORNL	US
Jun-10	↓	1759000	↓	US
	Tianhe-1A	2566000	NSC	China
	K computer	8162000	RIKEN AICS	Japan
Nov-11	↓	10510000	↓	Japan
Jun-12	Keuka	16324000	DOE/NNSA/LLNL	US
Nov-12	↓	33.9PFlops	DOE/SC/ORNL	US
Jun-13	Tianhe-2	33863000	NSC	China
Nov-13	↓	33863000	↓	China

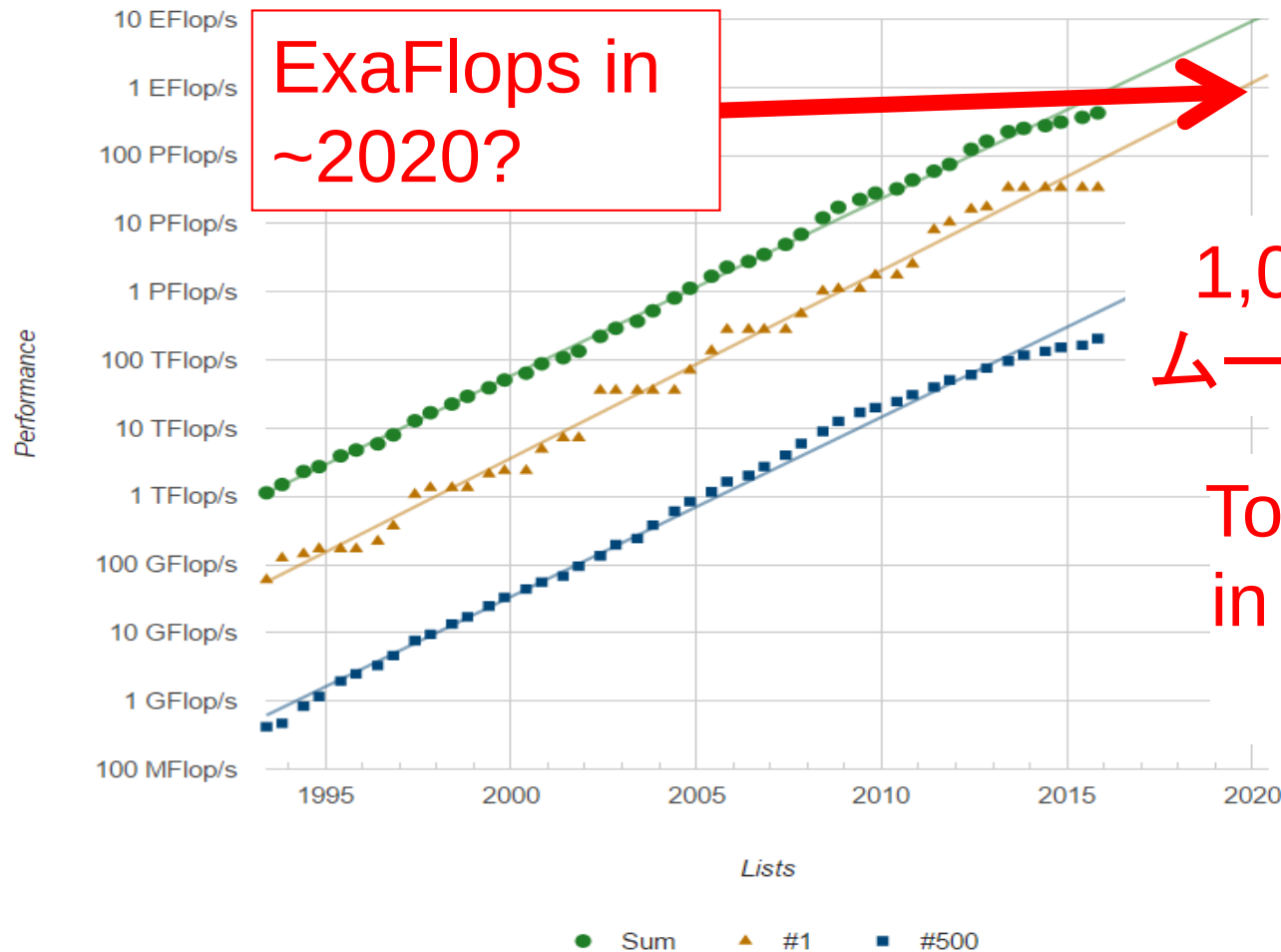
33.9PFlops

Development of Supercomputers

(from www.top500.org)



Projected Performance Development

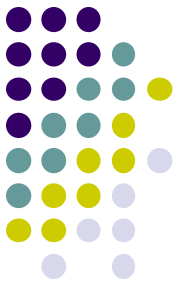


1,000x in 10 years
ムーアの法則より速い

Top Supercomputer
in 1993 (60GFlops)
≡ PC in 2011



なぜ計算速度 & 大容量が重要？



- 仮想実験(シミュレーション)のためには、**ばく大な量の計算**を、**タイムリーに**こなさなければならないから
 - ⇒明日の天気の計算に、一年かかっては意味がない
 - ⇒競争相手よりも先に薬を発売しなければ
- シミュレーションのためには、**大量のデータ**を利用・蓄積する必要
 - ⇒遺伝子データを大量に比較して発見をしたい
 - ⇒分子の動きをコマ送りで可視化したい

天気予報はどうやって計算されている？



大気・雲の動きは偏微分方程式で表現される

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{1}{\rho} \nabla p + \nu \Delta \mathbf{v} - g \hat{\mathbf{z}}$$

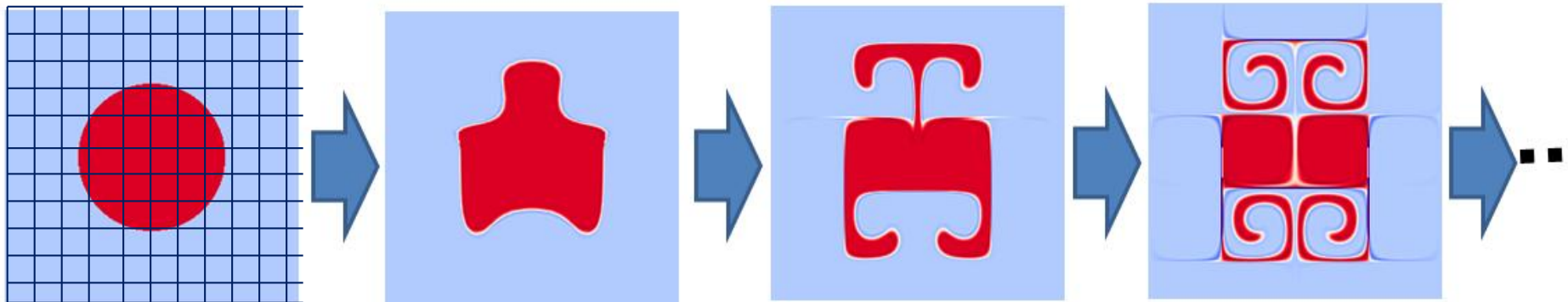
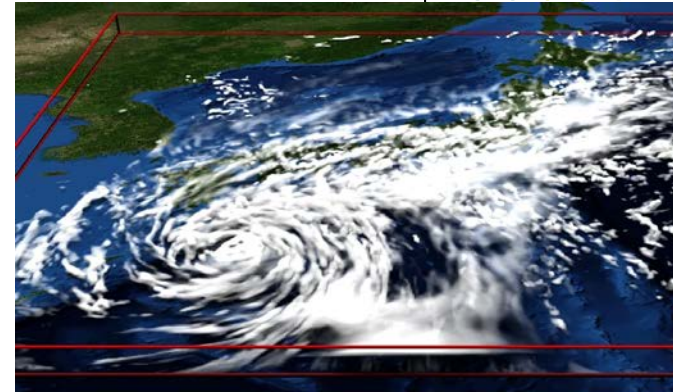
でも、解析解は存在しない(積分できない)
コンピュータは無限大や無限小は計算できない

⇒ ポイントは空間・時間の離散化

(1) 空間方向: 細かいマス目に分割

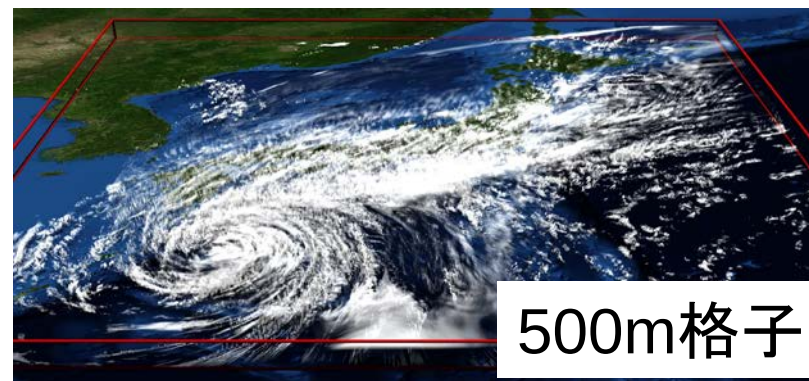
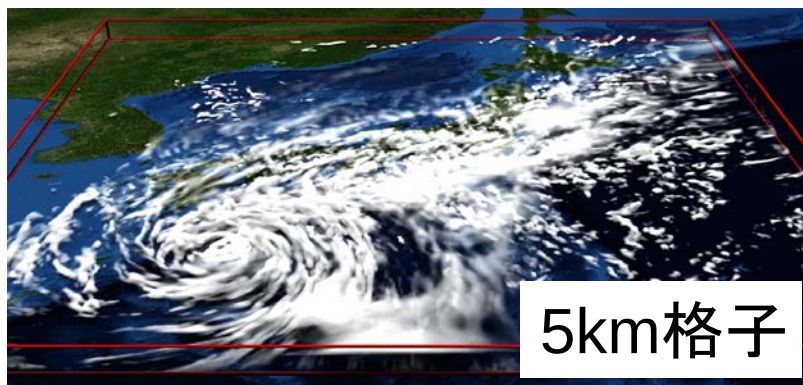
⇒ 1点ずつ計算が必要

(2) 時間方向: ある瞬間の計算が終了したら, 次の瞬間の計算へ (cf. パラパラ漫画)

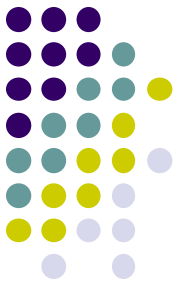


なぜ計算速度が重要？

- 各時間ステップごとに、各空間点を計算しなければならない
 - (経度方向 × 緯度方向 × 鉛直方向) × 時間方向
- ⇒ 莫大な計算。**明日の天気予報の計算に一週間かかっては意味がない！！**



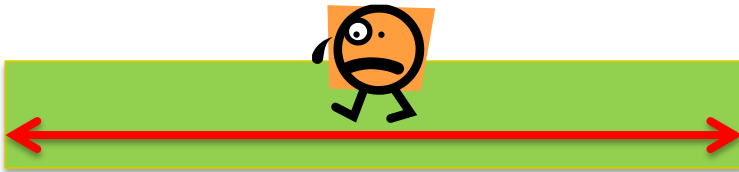
天気予報を当たりやすくするには、「細かく」する必要
解像度を10倍にするには、 $x \times y \times z \times t$ 方向で
10000倍の計算が必要になる！
→ さらに将来50m格子にしたい・・・



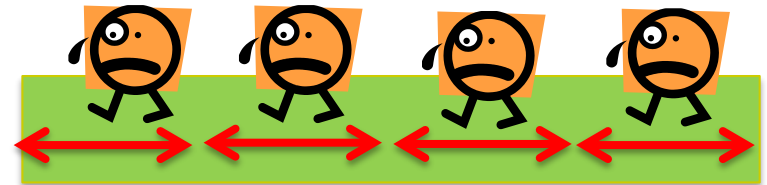
なぜスパコンは速い？

- 基本的な考え：一人(コンピュータ)だけで働くよりも、多くの人(コンピュータ)が協力して働くほうがすごい仕事ができる ⇒ 並列処理

一人の小人が大きな畑を耕す場合



複数の小人が分担して耕すと速く終わる



- スパコンは多数のコンピュータから成り立つ
- 各コンピュータは多数のCPUコア(頭脳)から成り立つ
⇒ クロック周波数が10万倍速いわけではない

一般的なスパコンの構造

- システム = 多数の計算ノード + 外部ストレージ
 - パーツ間はネットワークで接続
- 計算ノード = 1以上のプロセッサ + メモリ + ローカルストレージ
 - パーツ間はPCI-e, QPIなどの通信路で接続
- プロセッサ = 1以上のコア (+ L3 キャッシュ等)

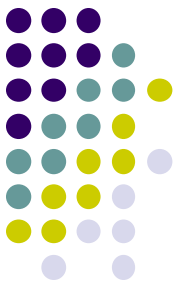


スパコンの計算性能は何で決まる？

理論ピーク演算性能:

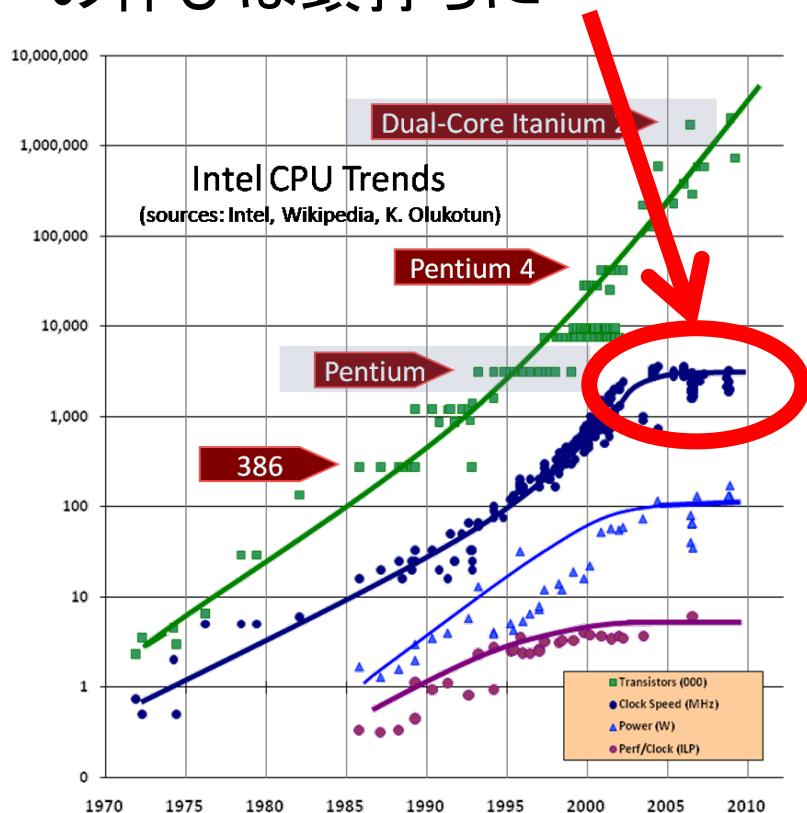
システムが仮に浮動小数演算のみを続けたときのFlops値。
実効演算性能とは区別が必要。以下の積となる:

- クロック周波数(Hz=1/sec)
 - 1~3GHz程度。2003ごろより頭打ち
- 1クロックあたりの同時計算数(flop)
 - TSUBAME2のCPUでは4, Sandy Bridge世代(AVX)で8
 - 単精度はその2倍
- プロセッサあたりのコア数
 - 4~16程度, これからも伸びる見込み
- 計算ノードあたりのプロセッサ数
 - 1~4程度
- 計算ノード数
 - TSUBAME2では1400, 京コンピュータでは88000



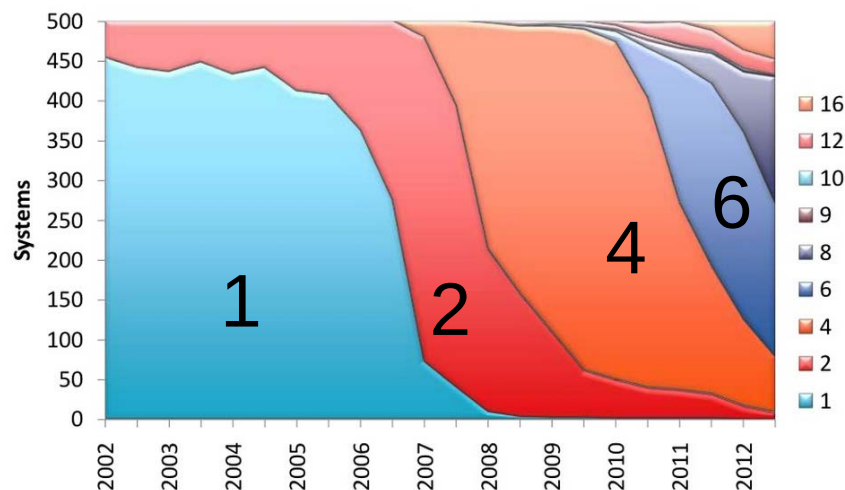
近年のスパコンの動向

2003年ごろからクロック周波数の伸びは頭打ちに



プロセッサあたりのコア数
を増やして性能を稼ぐ方向に

Cores per Socket



90年代までの常識:「5年待てば私のプログラムは速くなる」
今の常識:「並列プログラミング覚えないと！」

GPUやアクセラレータ



- 「周波数はあまり高くせず、並列度で稼ぐ」という方針をさらに推し進めたのが、GPUやアクセラレータ
- GPU(graphic processing unit)はもともと画像出力用専用プロセッサだが、演算へ転用(GPGPU)



Intel Xeon X5670

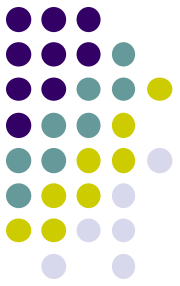
$$2.93 \text{ GHz} \times 4 \text{ Flop} \times 6 \text{ core} = 70.4 \text{ GFlops}$$



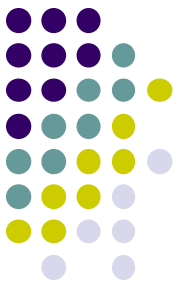
NVIDIA Tesla K20X

$$0.73 \text{ GHz} \times 128 \text{ Flop} \times 14 \text{ SM} = 1310 \text{ GFlops}$$

- TSUBAME2.5, Titan(2012 No.1), Tianhe-1A(2010 No.1)などが採用
- NVIDIA GPU, Intel Xeon Phi (現在世界一の天河2号)など

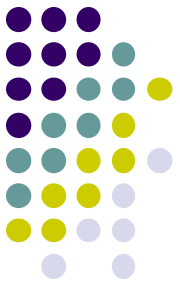


- ここで、「TSUBAME2.5利用ガイダンス」の説明



本授業では

- 本授業の受講者は t2g-ppcomp ユーザグループに入ってもらおう予定
 - t2subコマンドで “-W grouplist=t2g-ppcomp”
 - ちょっと大きなプログラムも動かせるように
- 授業に関する質問は担当教員あてに
 - 「レポートの×切はいつですか？」などをsoudan MLに投げないように



次回までに準備してください

- 可能ならノートPCを持参 / Please bring your laptop PC
 - titech-pubnetの使い方を調べてください
 - SSHの使える端末ソフトをインストールすること
 - WindowsノートPCなら、Putty, Teraterm+SSHなど
- 用意が難しい場合は、家などで各自演習
- TSUBAME2のアカウントを作ってください / Please make your account on TSUBAME2
 - 東工大ポータル ⇒ TSUBAME2.5利用ポータルより
 - 作成後、下記をメール連絡

件名(Subject): TSUBAME2 ppcomp account

宛先(To): ppcomp@el.gsic.titech.ac.jp

- 専攻・研究室
- 学年
- 氏名
- アカウント名

TSUBAME情報源



- <http://tsubame.gsic.titech.ac.jp>

- マニュアル
- 統計情報(混雑具合など)
- メンテナンス予定

- スパコン相談ML:

- soudan@cc.titech.ac.jp

↑のMLは利用法相談用
授業に関することは遠藤など
教員に！

The screenshot shows the homepage of the Tsubame High-Performance Computing service at Tokyo Tech. The browser address bar displays tsubame.gsic.titech.ac.jp. The page header includes the GSIC logo and the title "TSUBAME計算サービス" (TSUBAME Computing Service) with the subtitle "High-Performance Computing at Tokyo Tech". Navigation tabs include "Top", "初めての方へ" (For New Users), "利用について" (About Usage), "TSUBAME2システム構成" (TSUBAME2 System Configuration), and "お問い合わせ" (Contact Us). A left sidebar menu lists various links under "メニュー". The main content area features "重要なお知らせ" (Important Notices) and "最新のお知らせ" (Latest Notices), both containing recent updates and system status reports.

TSUBAME計算サービス
High-Performance Computing at Tokyo Tech

Top 初めての方へ 利用について TSUBAME2システム構成 お問い合わせ

メニュー

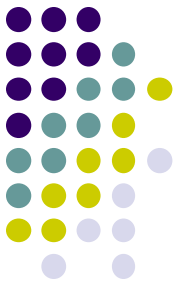
- ▶ Top
- ▶ お知らせ
- ▶ FAQ
- ▼ 利用について
 - ▶ TSUBAME2計算サービス種別
 - ▶ アカウント取得方法
 - ▶ 有料サービス利用方法
 - ▶ ログイン方法
 - ▶ TSUBAMEポータル
 - ▶ 各種利用の手引き
 - ▶ TSUBAME1利用者の方へ
- ▼ TSUBAME2システム構成
 - ▶ TSUBAME2ハードウェア構成
 - ▶ TSUBAME2ソフト

重要なお知らせ

- ▶ 2011.10.02
現在のTSUBAMEの運用状況はこちら(グランドチャレンジ実施中)
- ▶ 2011.08.15
平成23年度秋期グランドチャレンジ制度実施に伴うサービス休止のお知らせ(9/29更新)
- ▶ 2011.07.22
グループディスク課金本運用開始について
- ▶ 2011.06.27
7月以降の夏季のTSUBAME運用・停止について(9/29更新)

最新のお知らせ

- ▶ 2011.9.27 : 【障害報告】2011.9.16発生:/work0ストレージ障害
- ▶ 2011.9.22 : Hキュー予約システムメンテナンスのお知らせ(9/26実施)
- ▶ 2011.9.22 : Yキューの終了について
- ▶ 2011.9.13 : 【障害報告】2011.9.12発生:ジョブスケジューラ障害
- ▶ 2011.9.13 : 【障害報告】2011.9.9発生:ジョブスケジューラ障害
- ▶ 2011.9.06 : Materials Studio 5.5.3, Discovery Studio 3.1公開、学内配布開始について
- ▶ 2011.9.02 : 【障害報告】2011.9.2発生:ジョブスケジューラ障害



連絡先/Contact

- 授業全体に関すること
 - ppcomp@el.gsic.titech.ac.jp (遠藤あて)
- 東工大OCW
 - <http://ocw.titech.ac.jp> → 情報理工学院 → 実践的並列コンピューティング (平成28年度1Q)