

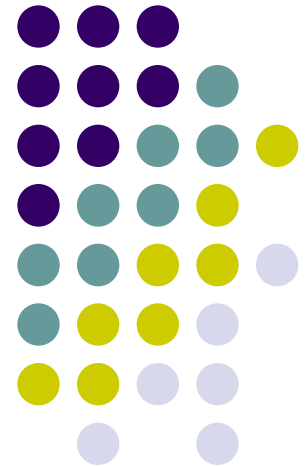
2013年度 実践的並列コンピューティング 第1回

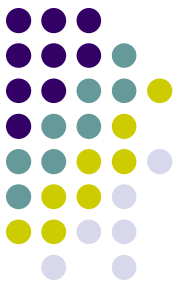
授業の概要と
TSUBAMEスパコンガイダンス

遠藤 敏夫

endo@is.titech.ac.jp

2013年4月7日



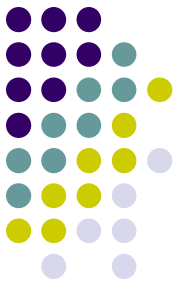


本授業のねらい

- 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)
 - (予定) Hadoop (Map-Reduceプログラミング)

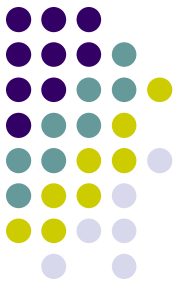
お知らせがあるときは下記webページにて

<http://www.el.gsic.titech.ac.jp/~endo/>



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
 - 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
 - コンパイル、makeコマンド



半年の流れと単位認定

基礎編(4~5月)

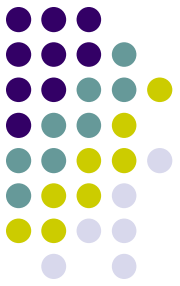
- OpenMP: 共有メモリ並列プログラミング
- MPI: 分散メモリ並列プログラミング

応用編(6月~8月)

- GPUプログラミング
- Map-Reduceプログラミング(予定)
- 応用アルゴリズム(予定)

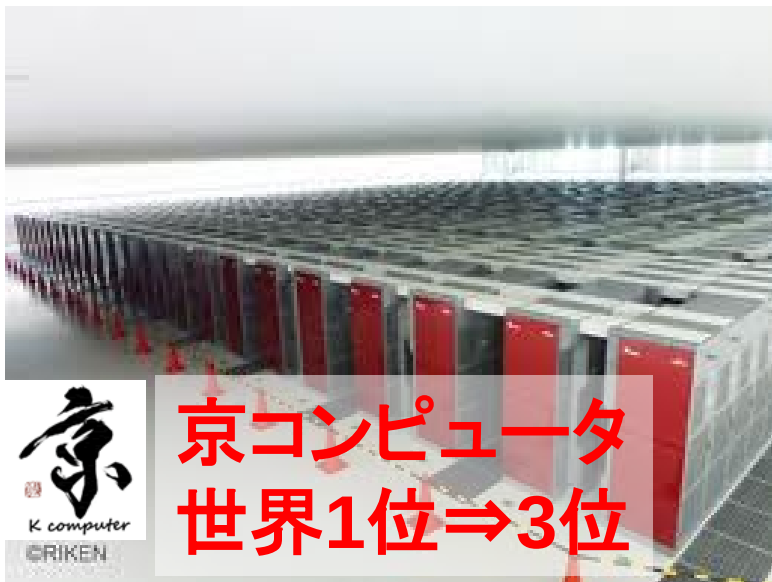
下記により採点・単位認定

- 基礎編の課題(2 or 3問)から一つ選択しレポート提出 (必須)
 - 締めきりは6月ごろ。二つ以上提出も可
- 応用編の課題(2 or 3問)から一つ選択しレポート提出 (必須)
 - 締めきりは8月上旬ごろ。二つ以上提出も可
- 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⇒3位



東工大TSUBAME2
世界4位⇒17位



BlueGene/P



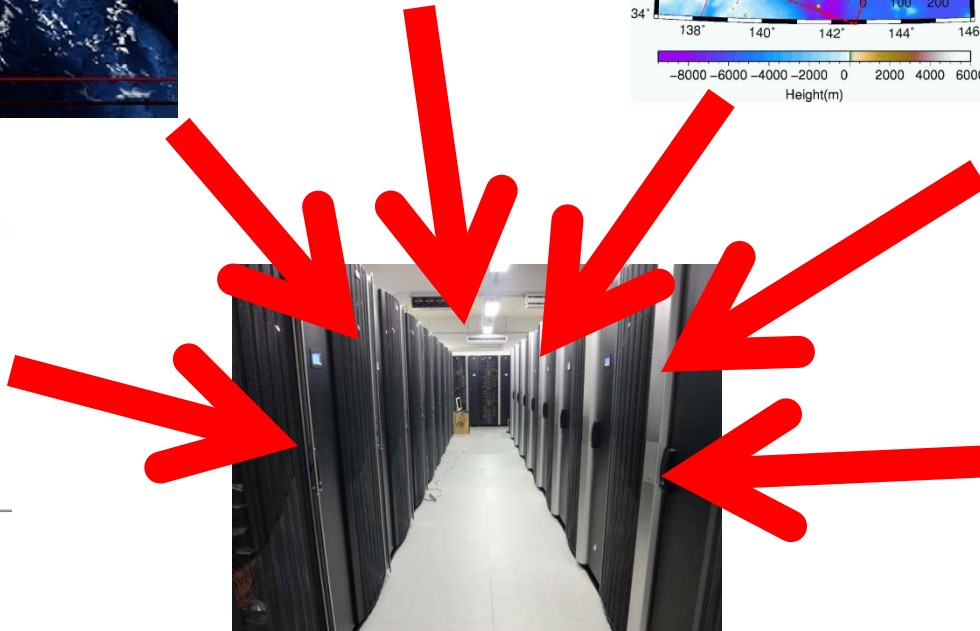
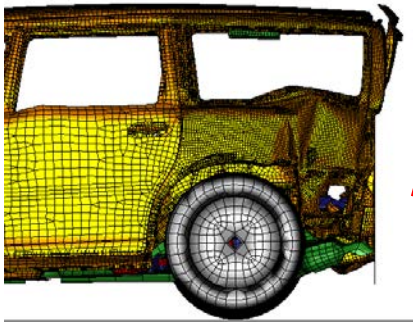
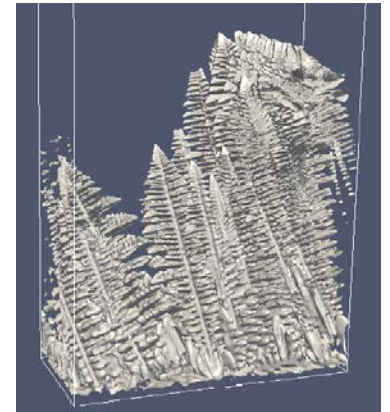
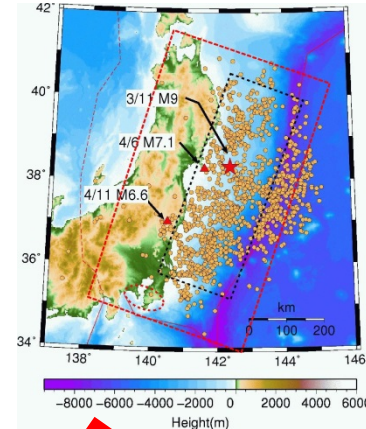
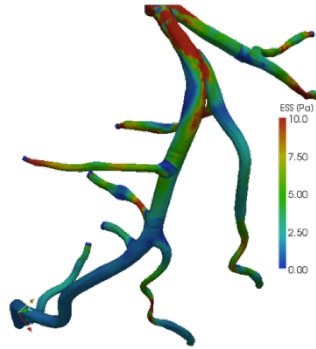
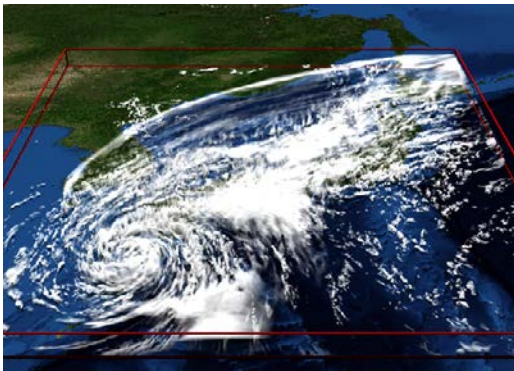
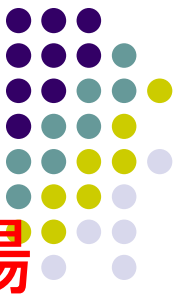
Jaguar



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**

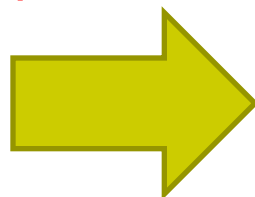


普通のコンピュータとの違い

- スパコンは、普通のコンピュータよりも**ずっと計算速度が速い**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20GFlops (一秒で200億回)

2.4PFlops (一秒で2400兆回)



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



Linpack演算速度 (Gflops)

Jun-93	CM-5/1024	50.7	LANL	US
Nov-93	Numerical Wind Tunnel	143.4	SNL	US
Jun-94	XP/S140	170	NAL	Japan
Nov-94	Numerical Wind Tunnel	170	NAL	Japan
Jun-95	Numerical Wind Tunnel	170	NAL	Japan
Nov-95	Numerical Wind Tunnel	170	NAL	Japan
Jun-96	SR2201/1024	2121	SNL	US
Nov-96	CP-PACS/2048	2379	SNL	US
Jun-97	ASCI Red	2379	SNL	US
Nov-97	ASCI Red	2379	SNL	US
Jun-98	ASCI Red	2379	SNL	US
Nov-98	ASCI Red	2379	SNL	US
Jun-99	ASCI Red	2379	SNL	US
Nov-99	ASCI Red	2379	SNL	US
Jun-00	ASCI Red	2379	SNL	US
Nov-00	ASCI White	4938	LLNL	US
Jun-01	ASCI White	7226	LLNL	US
Nov-01	ASCI White	7226	LLNL	US
Jun-02	Earth-Simulator	35860	ES Center	Japan
Nov-02	Earth-Simulator	35860	ES Center	Japan

60GFlops

19年間で性能300,000倍
アメリカと日本の一騎打ち
⇒中国の台頭

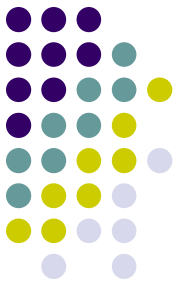
Jun-03	Earth-Simulator	35860	ES Center	Japan
Nov-03	Earth-Simulator	35860	ES Center	Japan
Jun-04	Earth-Simulator	35860	ES Center	Japan
Nov-04	BlueGene/L beta	70720	IBM/DOE	US
Jun-05	BlueGene/L	136800	DOE/NNSA/LLNL	US
Nov-05	BlueGene/L	136800	DOE/NNSA/LLNL	US
Jun-06	BlueGene/L	136800	DOE/NNSA/LLNL	US
Nov-06	BlueGene/L	136800	DOE/NNSA/LLNL	US
Jun-07	BlueGene/L	136800	DOE/NNSA/LLNL	US
Nov-07	BlueGene/L	136800	DOE/NNSA/LLNL	US
Jun-08	BlueGene/L	136800	DOE/NNSA/LLNL	US
Nov-08	BlueGene/L	136800	DOE/NNSA/LLNL	US
Jun-09	BlueGene/L	136800	DOE/NNSA/LLNL	US
Nov-09	BlueGene/L	136800	DOE/NNSA/LLNL	US
Jun-10	BlueGene/L	136800	DOE/NNSA/LLNL	US
Nov-10	BlueGene/L	136800	DOE/NNSA/LLNL	US
Jun-11	BlueGene/L	136800	DOE/NNSA/LLNL	US
Nov-11	BlueGene/L	136800	DOE/NNSA/LLNL	US
Jun-12	Sequoia	17590000	DOE/NNSA/LLNL	US
Nov-12	Titan	17590000	DOE/SC/ORNL	US

17.6PFlops

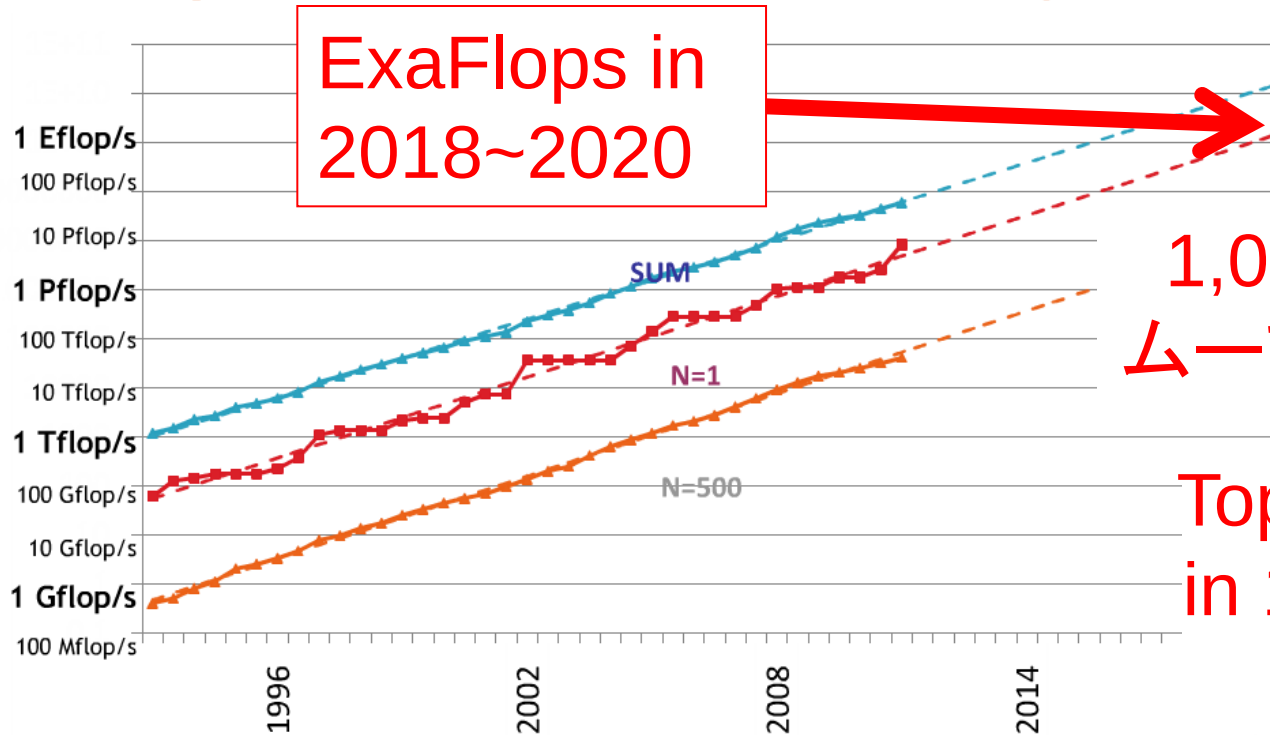
参考: www.top500.org

Development of Supercomputers

(from www.top500.org)



Projected Performance Development

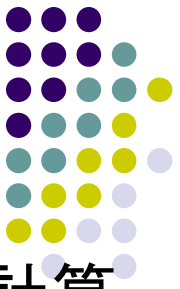


1,000x in 10 years
ムーアの法則より速い

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



なぜ計算速度が重要？

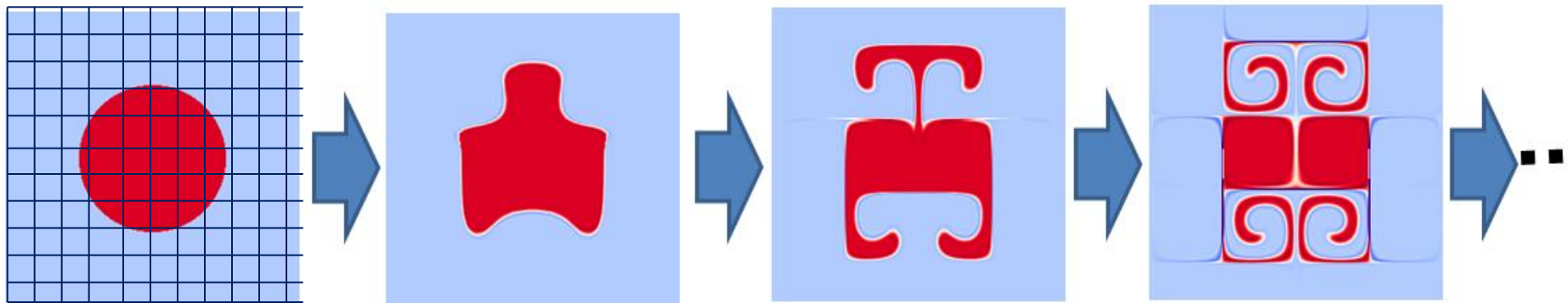


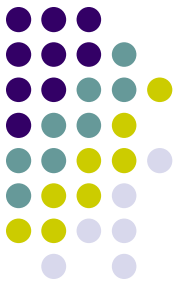
- 仮想実験(シミュレーション)のためには、**ばく大な量**の計算を、**タイムリー**にこなさなければならないから
⇒明日の天気計算に、一年かかっては意味がない！

天気予報はどうやって計算？

空間を、細かいマス目に分割 ⇒ 1点ずつ計算が必要
ある瞬間の計算が終了したら、次の瞬間の計算へ

- 因果律：過去⇒現在⇒未来⇒もっと未来
- パラパラ漫画のように計算を続ける

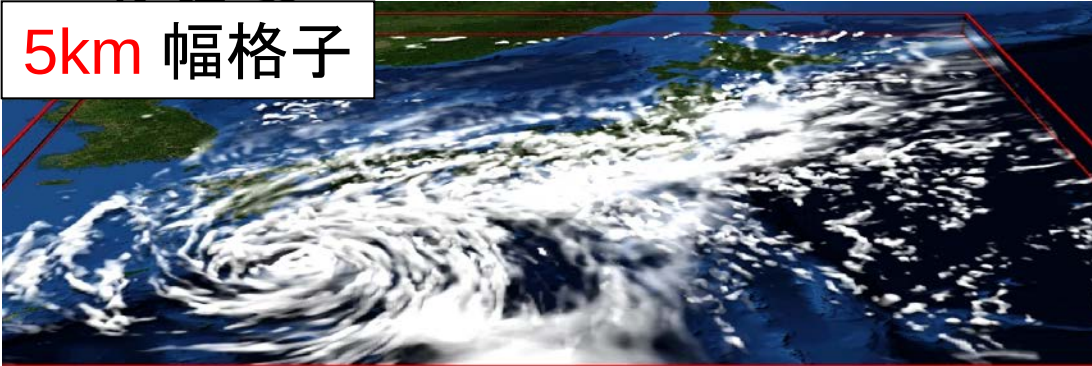




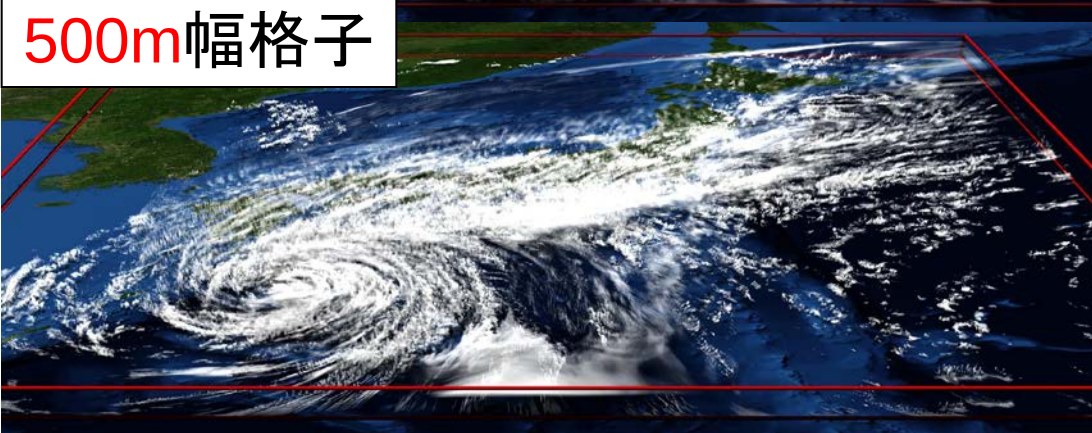
なぜ計算速度が向上し続ける必要？

- より厳密なシミュレーションを行うには、細かい解像度が必要

5km 幅格子



500m 幅格子



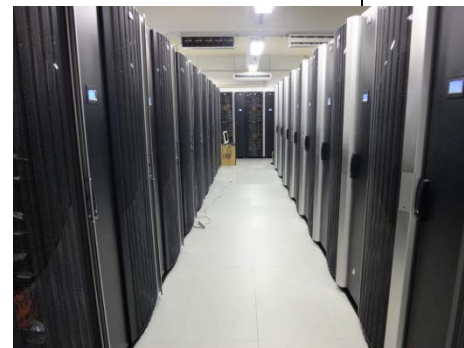
より細かい雲の
挙動が把握可能に！

10倍の高解像度のためには

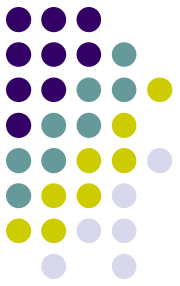
X方向10倍 × Y方向10倍 × Z方向10倍 × 時間方向10倍 = **10000倍**
の計算をこなす必要！

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？



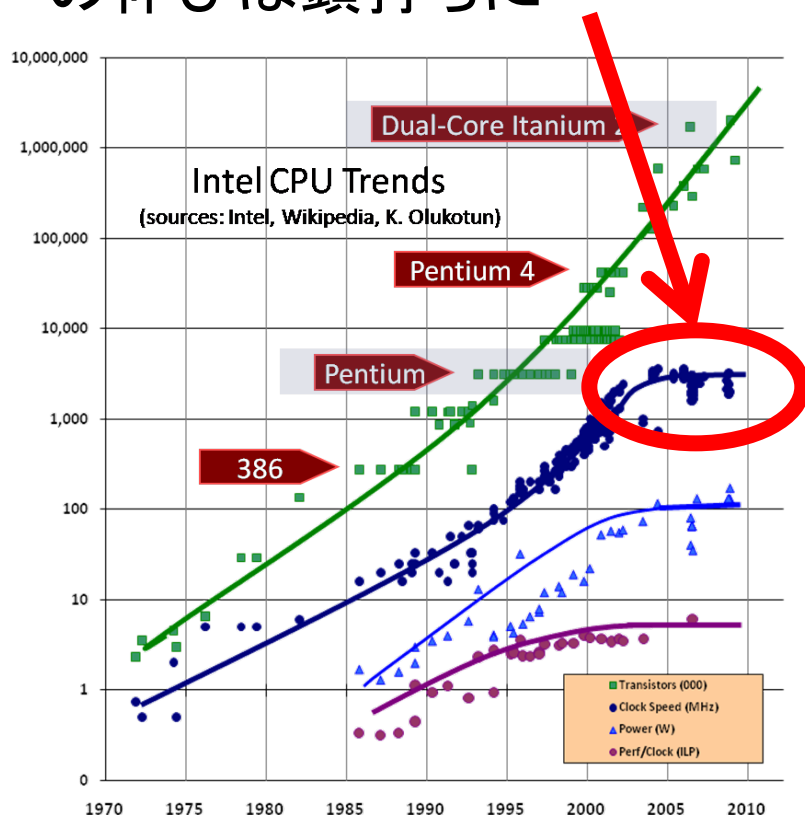
理論ピーク演算性能:

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

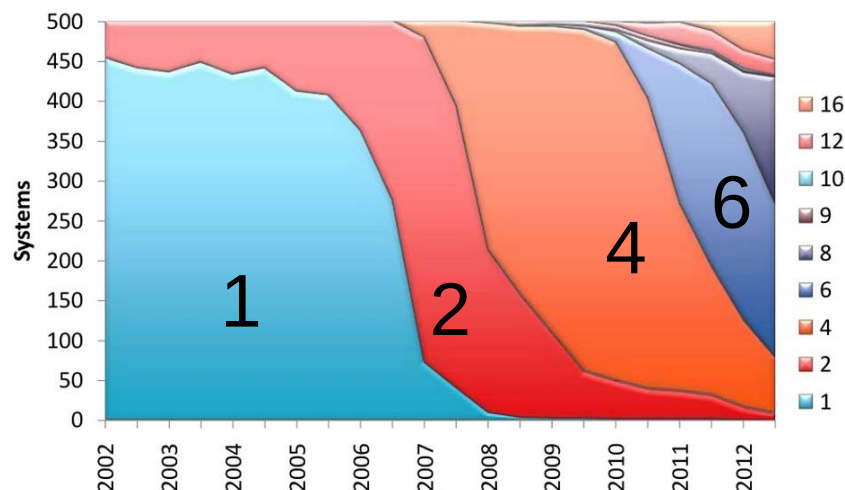
近年のスパコンの動向

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

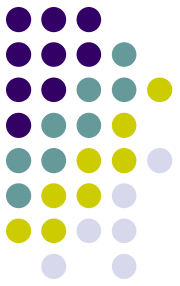
$$2.93 \text{ GHz} \times 4 \text{ Flop} \times 6 \text{ core} = 70.4 \text{ GFlops}$$



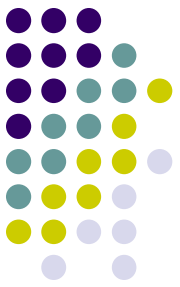
NVIDIA Tesla M2050

$$1.15 \text{ GHz} \times 32 \text{ Flop} \times 14 \text{ SM} = 515 \text{ GFlops}$$

- TSUBAME2.0, Titan(2012 No.1), Tianhe-1A(2010 No.1)などが採用
- NVIDIA GPU, AMD/ATI GPU, Intel Xeon Phiなど



- ここで、「TSUBAME2.0利用ガイドンス」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



- ノートPCを持参のこと / Please bring your laptop PC
 - 無線ネットワーク(Wifi)が使えること
 - 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
- 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 ⇒ TSUBAME2.0利用ポータルより
 - 作成後、遠藤へ下記をメール連絡

Subject: TSUBAME2 ppcomp account

To: endo@is.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>

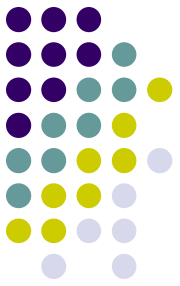
- マニュアル
- 統計情報(混雑具合など)
- メンテナンス予定

- スパコン相談ML:

- soudan@cc.titech.ac.jp

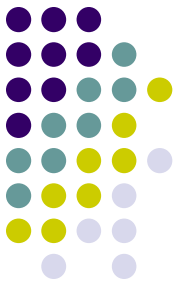
授業に関することは遠藤など
教員に！

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 High-Performance Computing at Tokyo Tech". Navigation tabs are present for "Top", "初めての方へ", "利用について", "TSUBAME2システム構成", and "お問い合わせ". A left sidebar menu lists various links including "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", and "TSUBAME2ソフトウェア構成". The main content area features a "重要なお知らせ" (Important Notice) section with several announcements dated from 2011.06.27 to 2011.10.02, and a "最新のお知らせ" (Latest Notice) section with announcements from 2011.9.02 to 2011.9.27.



連絡先/Contact

- 授業全体に関すること
 - 遠藤 endo@is.titech.ac.jp
- ノートPCの準備が困難な場合も、早急宛先に上記へ相談のこと
- 遠藤のページ
 - <http://www.el.gsic.titech.ac.jp/~endo/>



次回/Next Lecture

- April 15 (Mon)
 - 並列プログラミングモデル
 - OpenMPによる共有メモリプログラミング (1)