

# PetaからExaへ

神戸大学  
システム情報学研究科  
小柳義夫

# 神戸商業大学の経営機械化

- 1941年、我が国の大学で初めてPSCを導入した
    - 経営計算研究室→1944経営機械化研究所
    - 電動穿孔機、検孔機、分類機、統計機（IBMから無償貸与）
    - ナショナル銀行会計機2000号
  - 現在「経営機械化展示室」あり。貴重な機械や資料を保存
    - 情報処理学会の「コンピュータ博物館」でも触れられていない
- <http://www.rieb.kobe-u.ac.jp/riebcomp/history/gallery.html>



# 神戸商業大学の経営機械化

- 産業経済新聞1943/2/24(平井泰太郎)
  - 事務会計機械の国産化
  - 「人間の方が安い」「人間の方が機械より正確」「我が国では機械を作れない」との批判
  - 反論「竹槍で戦闘機は落とせない」
  - 兵器の機械化は理解しても、事務の機械化は理解せず
  - 日本の銀行では円天井の下に百人がペンを走らせているが、アメリカでは接客係と貸し付け担当者しかいない。サイノグラフで10階の信用係と電氣的に通信。
  - [http://www.lib.kobe-u.ac.jp/das/jsp/ja/ContentViewM.jsp?METAID=10030452&TYPE=IMAGE\\_FILE&POS=1](http://www.lib.kobe-u.ac.jp/das/jsp/ja/ContentViewM.jsp?METAID=10030452&TYPE=IMAGE_FILE&POS=1)

# HPCのmilestones

岩波講座『計算科学』第1巻「計算の科学」参照

- 1 MF: 1964
- 10 MF: 1969
- 100 MF: 1976
- 1 GF: 1984
- 10 GF: 1987
- 100 GF: 1993
- 1 TF: 1997
- 10 TF: 2002
- 100 TF: 2005
- 1 PF: 2008
- 10 PF: 2011
- 100 PF: ?
- 1 EF : ?

# 1 MFの壁

- ENIAC (1946, pk 0.7 kF)  
add 0.2 ms, mult 2.6 ms
- IBM System/360-65(1965)  
single-precision matrix mult. (100x100) : 33 kF
- CDC6600 (1964, pk 4 MF) 実効1 MF?  
10個のfunctional units. LLNLに納入  
clock 10 MHz, add 4 clocks, mult 10 clocks(x2)
- IBM2983 Array Processor(1969)  
I/O channelに接続する付加ベクトル・プロセッサ  
single-precision matrix mult. (100x100): 1.59 MF



Wikipedia

# 10 MFの壁

- 命令パイプライン
  - CDC7600 (1969, pk 36 MF) 実効 5 MF?
- ベクトル計算機
  - TI ASC (1972, pk 30 MF) 7機製作
  - CDC Star-100(1973, pk 50 MF) 4機製作
  - FACOM 230-75 APU (1977, pk 22 MF) 2機製作
    - NALで稼動
  - HITAC M200H IAP (1979, pk 48 MF)
- 並列計算機 (SIMD)
  - Burroughs BSP(1977発表、1980製作中止、4x4, pk 50 MF)



Wikipedia

# 100 MFの壁

- ベクトル計算機

- Cray-1 (1976, pk 160 MF)
- Cray X-MP/2 (1982, pk 630 MF)
- CDC Cyber 203 (1980, pk 200 MF),
- Cyber 205 (1981, pk 400MF)
- HITAC S-810 (1983, pk 630 MF)
- FACOM VP-200 (1983, pk 570 MF)

- 並列計算機

- ILLIAC IV (1973-6, pk 150 MF) 8x8 1機製作



Wikipedia

# 1 GFの壁

- ベクトル計算機

- Cray X-MP/4 (1984, pk 1.26 GF)
- Cray-2 (1985, pk 1.95 GF)
- VP-400 (1985, pk 1.14 GF)
- SX-2 (1986, pk 1.3 GF)

- ベンチマーク性能

- SX-2はLivermore Loop No. 7 (Equation of States)で**1.042** GFlopsを実測
- X-MP/4では、最大0.807 GFlops (No. 7) を達成。
- Cray-2は？



# アメリカの戦略

- 1985年、米国立研で日本品排斥
- 1989年、米スーパー301条の対日適用
- 1991年高性能コンピュータ通信法
  - HPCC 計画(1991-96)
- PITAC(大統領情報技術諮問委員会)
  - クリントン大統領が設置(1997-2001, 2003-05)
- 「21世紀に向けての情報技術IT2」(2000)
  - \$266Mの研究投資を追加
- ASCI計画(1995から)→ASC計画
  - 3倍ずつ高性能なコンピュータを戦略的に設置
- “Enabling Technologies for **Petaflops** Computing”  
(Pasadena, 1994/2)

# 10 GFの壁

- ベクトル並列計算機

- ETA-10 (1987, pk 10 GF) 液体窒素冷却  
ほとんど安定に作動せず。Princeton大学や東工大が購入
- PHI (1989, pk 10 GF)「スパコン大プロ」評価後解体
- SX-3/44R (1990, Lp 23.2 GF) NEC社内 (pk 25.6)
- C916/16256 (1992, Lp 13.7 GF) Cray社内
- HITAC S-3800 (1993, Lp 27.5 GF) 東大
  - 当時日立の田中輝雄、山本有作両氏ががんばったとか。

- スカラ並列計算機

- QCD-PAX (1989, single pk 13 GF) 筑波大学
- CM-5/1024 (1993, Lp 59.7 GF) LANL  
1993/6のTop500の首位4件独占

# 100 GFの壁

- ベクトル並列計算機
  - NWT (1993, Lp 124 GF) NAL  
1993/11, 94/11, 95/6, 95/12のTop500でトップ  
94/11以降はLp 170 GF  
この頃Petaflops計画始まる(1994)
- スカラ並列計算機
  - Paragon XP/S140 (1994, Lp 143.4 GF) SNL  
1994/6のTop500の首位
  - SR2201/1024 (1996, Lp 220.4 GF) 東大  
1996/6の首位
  - cp-pacs (1996, Lp 368.2 GF) 筑波大  
1996/11の首位

# 1 TFの呪い

- FPS T-series (1986年発表)
  - Transputer T414, Weitek FPU, hypercube接続、Video RAM
  - 16K nodesで1 TF実現と主張
  - 私も1986年8月にBeavertonまで見学に
  - 1992年破産してCrayの子会社、コンパイラはPGIへ
- TMC CM-5 (1991年発表)
  - Sparc, Weitek FPU, fat tree接続。8Kで1 TF実現と主張。
  - 1993/6の最初のTop500ではTop 4はCM-5 (5,6はSX-3)
  - 1994/8 破産。多くの技術者はSun Microsystemsへ
- NEC SX-4 (1995年発表)
  - CMOSベクトル, 2GF/node, 最大512 nodesで1 TF実現
  - 出荷は128nodesが最大。1998/6のTop500で13位 (Lp 244)
  - 会社は存続！

# 1 TFの壁

- ASCI Red (1997, Lp 1.068 TF → 2.379 TF)
  - SNL, Pentium Pro → Pentium 2
- ASCI Blue Mountain (1998, Lp 1.608 TF)
  - LANL, MIPS R10000
- ASCI Blue Pacific (1998, Lp 2.144 TF)
  - LLNL, PowerPC604e
- ASCI White (2000, Lp. 7.304)
  - LLNL, Power3
- **すべてスカラ並列**

# 10 TFの壁

- Earth Simulator (2002, Lp 35.86 TF)
  - JAMSTEC(横浜)
- ASCI Q (2002, Lp 13.88 TF)
  - LANL, Alpha
- Red Storm (2005, Lp 36.19 TF)
  - SNL, Opteron
- ASCI Purple (2006, Lp 75.78 TF)
  - LLNL, Power 5
- Tsubame (2006, Lp 38.18 TF)
  - 東京工業大学、Opteron+ClearSpeed
  - ESが首位を譲ってから約2年

# 100 TFの壁

- BlueGene/L (2005, Lp 136.8→478.2)
  - LLNL
  - 最初2004/11では70.72 TFで登場、ESを破る
  - このころExascale project始まる(後述)
- BlueGene/P (2007, Lp 167.3→825.5)
  - Jülich
- Earth Simulator 2 (2009, Lp 122.4 TF)
  - SX-9
- Kraken (2009, Lp 831.7 TF)
  - Tennessee大学

# Petaflopsへの道

- 1994/2, “Enabling Technologies for Petaflops Computing” (Pasadena)
  - 当時はNWTが100 GFを越えた頃。10000倍へ！
  - Seymour Crayが基調講演（当時Cray-4を発表）
- その後、7回のWS
- SC96でPetaflops Computingのパネル
  - 2010年頃を想定
  - 応用として、実時間3D心臓モデルが例示。その他、新材料、ナノ、VR、バイオ、プラズマなど
  - 要素技術としては、latency hiding, million threads, 消費電力（1～3 GWを想定する人も。2-3 MW/PFが主流）、費用（\$100M～1B）、信頼性、プログラムできるか、など。
  - 何となく今のExaの議論を彷彿させる

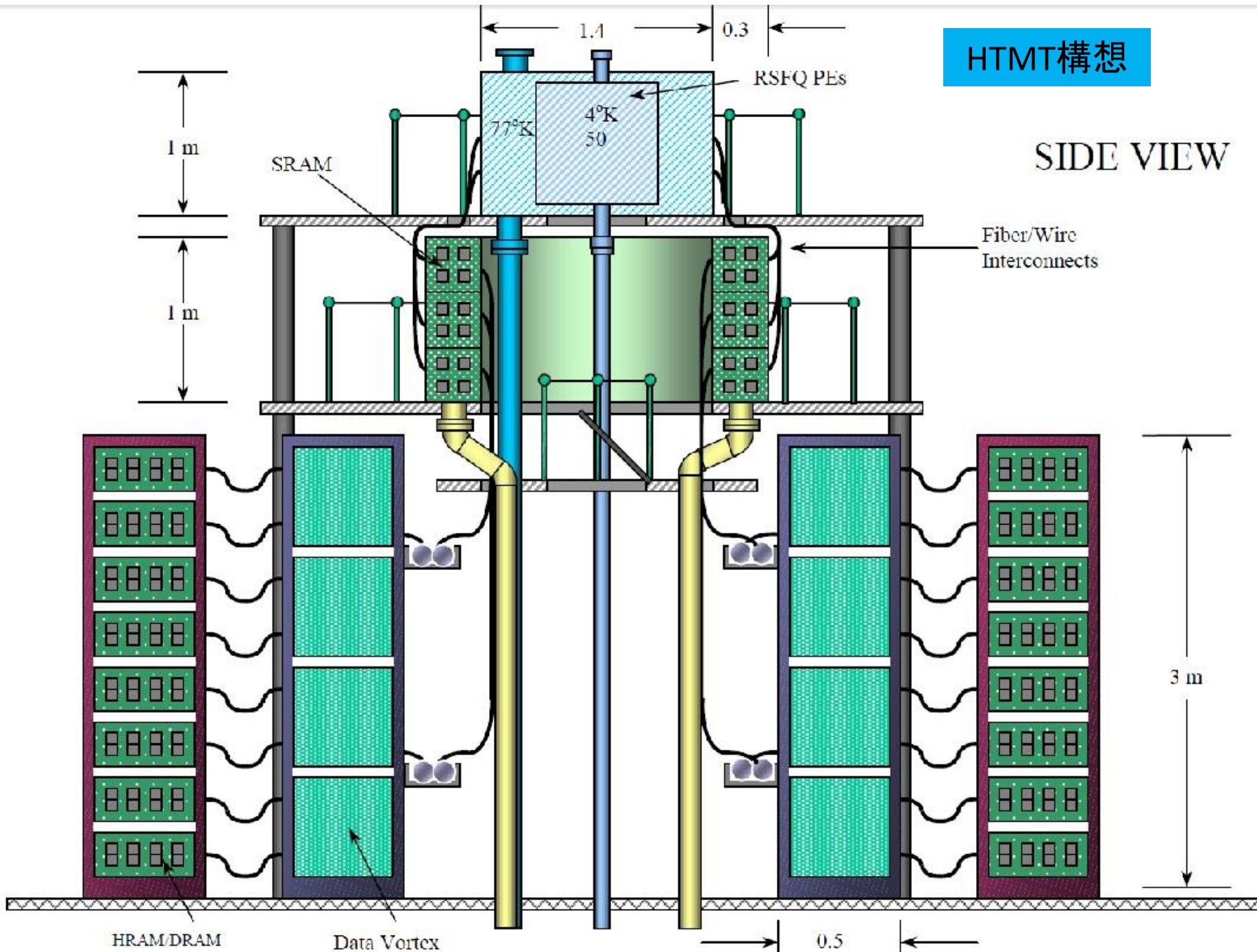


# 1999/2, “PETAFLOPS II” (Santa Barbara)

- Tilak Agerwala (IBM)
  - Powerの延長でPetaflopsができる
- Peter Kogge (Notre Dame)
  - PIM (Processor-in-Memory)ならlatencyは1/10, バンド幅は100倍
  - でもPIM間、PIM-nonPIM間の接続はどうする。
- Thomas Stirling (当時CalTech)
  - HTMT (超伝導、PIM、光接続、ホログラム記憶などを組み合わせる)
  - 超伝導なら、250nmで200GHzが実現できる

# HTMT構想

## SIDE VIEW



# アメリカの動き

- HEC Revitalization Task Force
  - 2003/3: 米国政府は国家科学技術会議 (NSTC) の下に特別プロジェクトとしてこのタスクフォースを編成 (ES対抗策か?)
- 2004/11: High-End Computing Revitalization Act of 2004 成立 (2005fy-\$50M, 2006fy-\$55M, 2007fy-60M for DoE)
- 2004/10: Zettaflops Project (実際は ExaFlops) が始まる
- 2005/3: High Performance Computing Revitalization Act of 2005 (H.R. 28) was approved by House Science Committee
  - 4月、House 通過
  - NSFを中心に、NASA, NOAA, DoE, NIST, EPA

# 日本の動き

- 1990年代、アメリカはPetaflopsに向けて動いていたが、日本の動きは緩慢
- 2002年に地球シミュレータが驚異的性能
- 2004/5: 文部科学省情報科学技術委員会にWGを設置し、11回の会合を行う
- 2005/8: 「計算科学技術の推進方策」
- 2005/6: 総合科学技術会議「科学技術基本政策策定の基本方針」に次世代スーパーコンピューティング技術を入れる

# 日本の動き

- 2005/7: スーパーコンピュータ推進**議員連盟**
  - 座長: 尾身幸次、副座長: 安倍晋三、事務局長: 後藤茂之、100人以上集まる
  - 8月、政府に勧告を提出
- 2005/7: 「最先端・高性能汎用スーパーコンピュータ開発利用」プロジェクトが正式決定
  - 2006～2012年度、総額1154億円、Lp 10 PF
  - 次世代スーパーコンピュータの開発
  - 応用ソフト(ナノ、バイオなど)の開発
  - 研究教育拠点の形成

# 日本の動き

- 2005年度～7年度、要素技術の研究開発
  - 総額40億円／年
  - システム相互接続(九大、富士通)
  - IPによる相互接続(東大、慶応等)
  - 低消費電力素子(日立、東大、筑波大)
  - CPUとメモリの光接続(NEC、東工大)
- 2005/8: 計算科学推進WG中間報告(>10 PF)
- 2005/8: 文部科学省が概算要求
- 2005/10: 理研を開発主体とする
- 2005/9, 10: 総合科学技術会議の評価検討会

# Requirements from these Applications

## 1. Large-scale processing part

- Although never explicitly stated, this part is believed to be a vector/pseudovector computer.
- 2 PF from disaster prevention  
1 PF from drug design  
0.2-0.6 PF from various fields

# Requirements from these Applications

## 2. Scalar computer part

- 4 PF from device simulation with electron correlation
- 0.3-0.5 PF from various fields

## 3. Special purpose computer

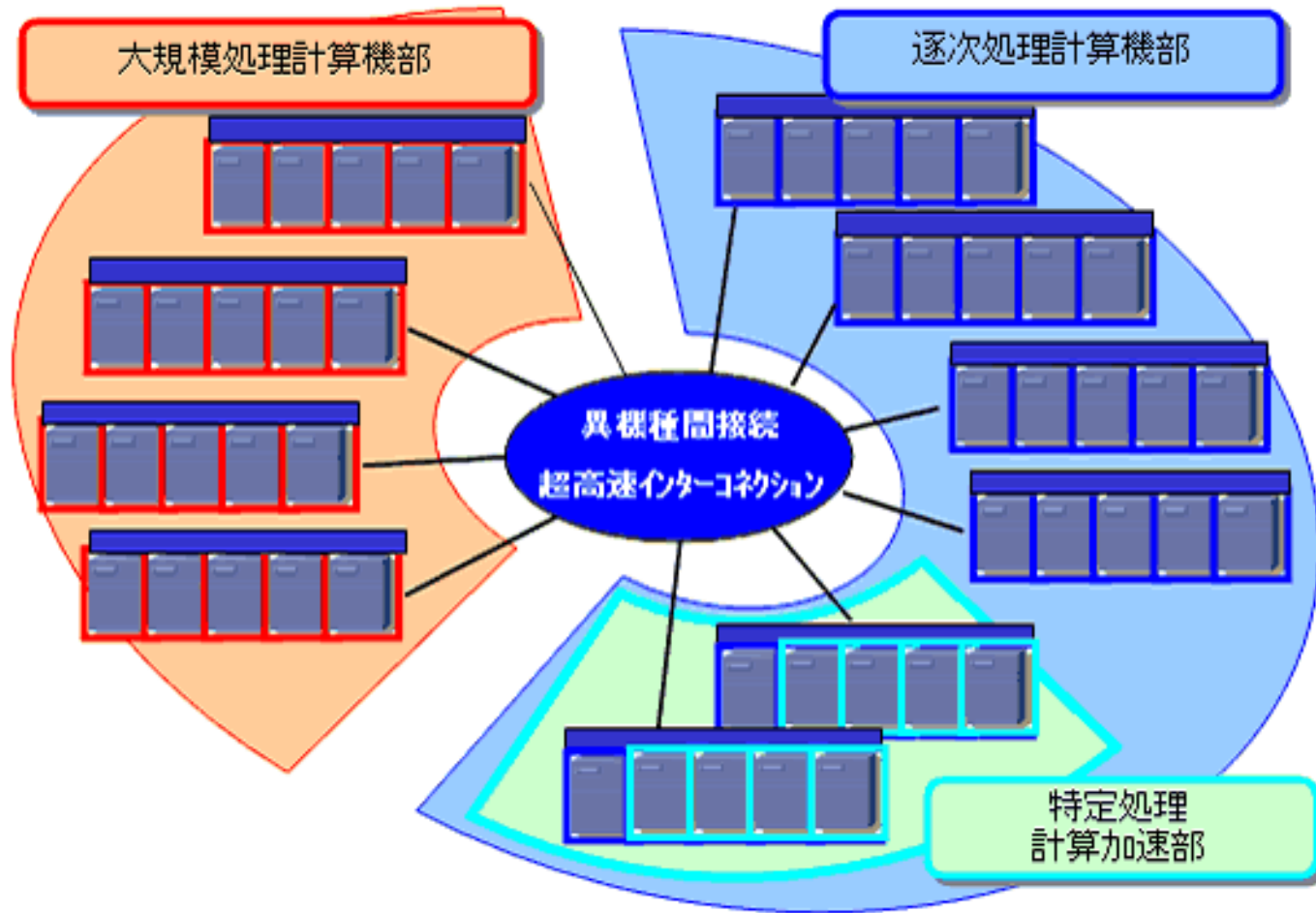
- 20 PF from drug design (MD)
- 20 PF from astrophysics



# Original Proposal

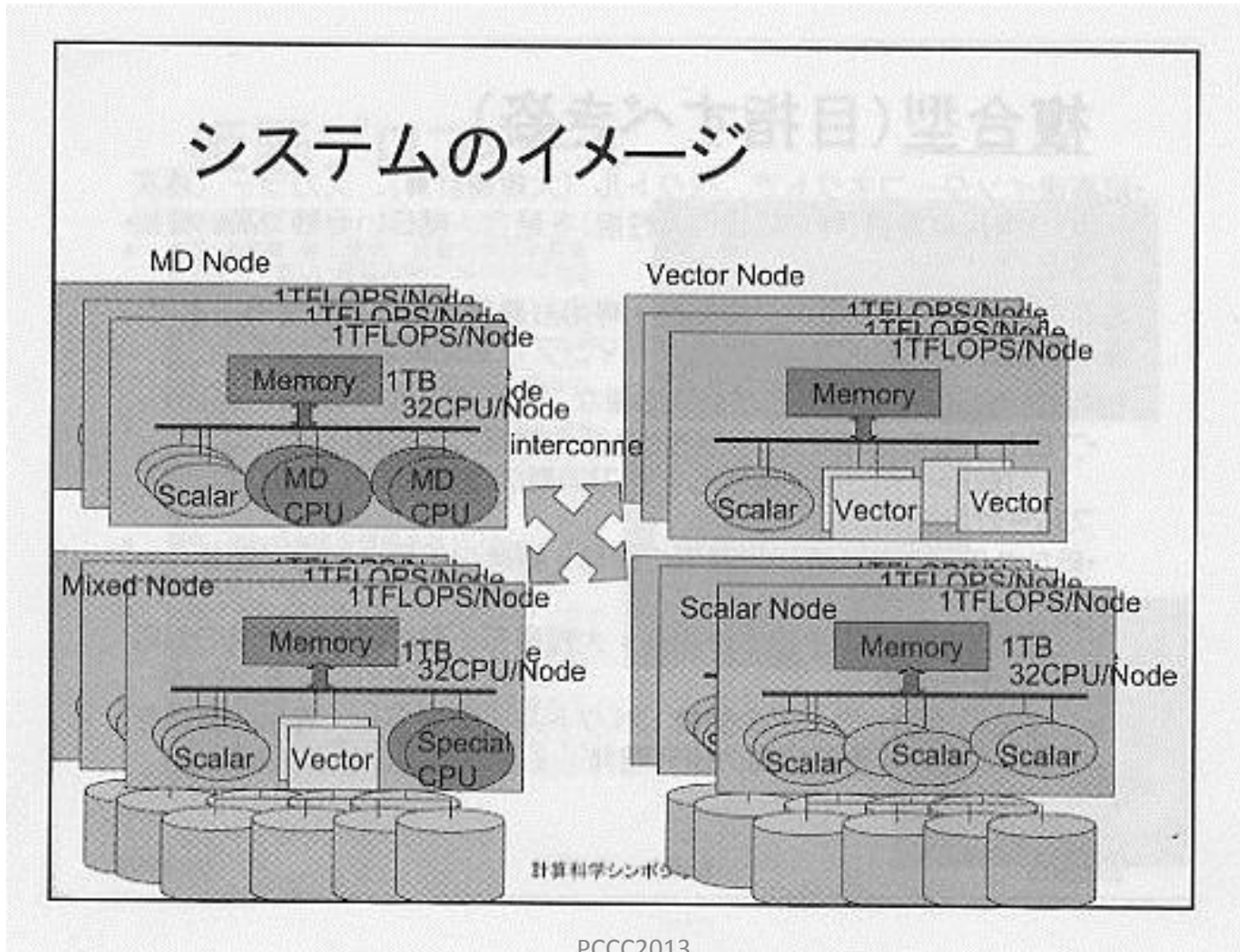
Large scale processing

Scalar computer

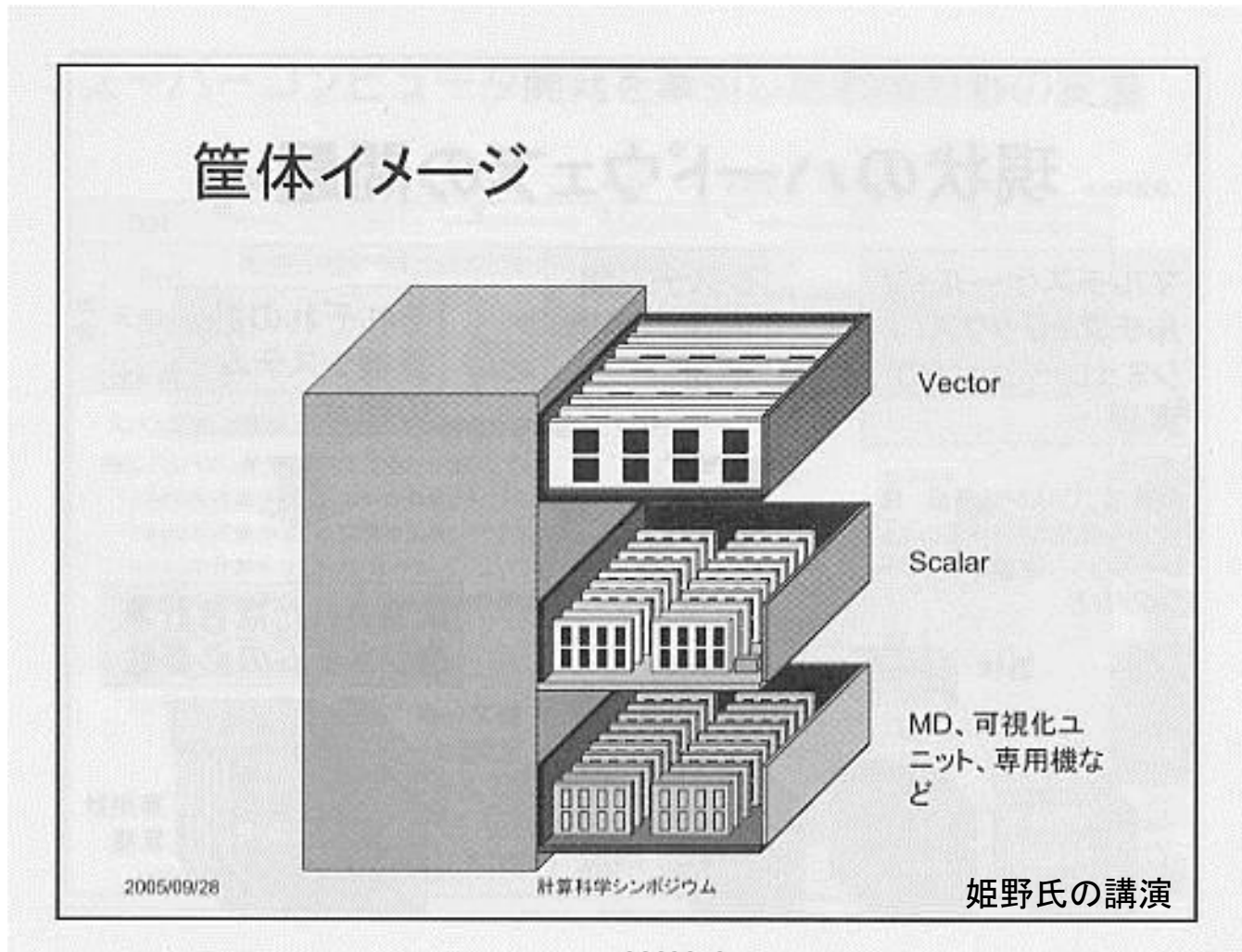


Special-purpose computer

# Proposed System Image



# Possible Rack Image



# 日本の動き

- 2005/9-10: 評価検討会
  - 「汎用性だけを目標としては使える計算機はできない」I先生や私が強調。
  - 「ぼろくそに言われた」(牧野)
  - ターゲットアプリをいくつか選定し、性能目標を設定し、それを実現するアーキ、基盤ソフト、ミドルウェアを設計する方法論が主張された
- 2005/11: 総合科学技術会議で「実施することが適当」と報告
- 2006/1: 文部科学省に推進本部
  - 21本のターゲットアプリを選定

# 日本の動き

- 2006/3: 第三期科学技術基本計画、閣議決定。スーパーコンピュータを国家基幹技術と位置付け
- 2006/7: 共用促進法を改正、立地委員会設置
- 2006年度: 「次世代スーパーコンピュータ概念構築に関する共同研究」(10以上の提案でコンペ)
- 2007/3: 神戸に決定
- 2007/4: 理研からシステム構成案
  - スカラ演算部(富士通担当)、ベクトル演算部(NEC、日立担当)から成る複合システム。合計>10 PF
- 2007/9から詳細設計、2009/4から中間評価

# 日本の動き

- 2009/5/13にNECが製造段階への不参加。スカラ部だけの構成とする
- 2009/7: 文科省、5戦略分野を決定
- 2009/11/13(金): 行政刷新会議の「事業仕分け」第3分科会
- 見送りに近い縮減→必要な改善を行いつつ推進

# 1 PFにどうやって到達したか

- Roadrunner (2008, Lp 1.026 PF), LANL
  - Cell processorは一種のPIMといえるか？
- Jaguar (2008, Lp 1.059→1.759) ORNL
  - Opteronのhomogeneous system。在来型の延長？
- Nebulea (2010, Lp 1.271 PF) Shenzhen深圳市
  - Intel+NVIDIA GPU
- 天河1A (2010, Lp 2.566 PF) 天津
  - Intel+NVIDIA GPU
- Tsubame2.0 (2010, Lp 1.192) 東工大
  - Intel+NVIDIA GPU

# 日本の動き

- 2010/2: NEC、ベクトルコンピュータの開発継続を表明
- 2010/7: 理研、愛称「京」を決定。AICS設立。
- 2010/11: 東工大、TSUBAME2.0: 1.192 PF
- 2011/3: 京の一部が稼働開始
- 2011/6: 京が8.162 PFでTop500の1位
- 2011/11: 京、10.51 PF。Gordon-Bell賞
- 2012/4: 登録機関、HPCIコンソーシアム
- 2012/5: 京の利用課題募集開始
- 2012/6: 京、正式に理研に引き渡し



# 10 PFの壁

- 京(2011, Lp 8.162→10.51) 理研
  - Sparc64 viiifx and Tofu interconnect
- Sequoia (2012, Lp 16.32→17.17 PF) LLNL
  - BG/Q
- Titan (2012, Lp 17.59 PF) ORNL
  - Opteron+Gemini interconnect
- 天河2号(2013, Lp 33.86) NUDT→広州
  - Intel Xeon+Phi
  - 2015年に100 PF (peak?)を目指す

# ExaScaleへの挑戦

- Linpack 1 EFは可能か？
  - メモリ100 PBとすると、 $n=10^8$ の行列が入る。これを1 EF ( $10^{18}$  flops)で実行すると、 $2/3 \times 10^{24} / 10^{18} = 7 \times 10^5$ 秒 = 7.7 日
  - 1週間安定に動作するとは思えない
  - もっと小さい $n$ でEFが出れば別であるが。
    - $n=5 \times 10^7$ なら1日

# Exascaleへの挑戦

- Zettaflops activity (SNL)
  - 1990's : Petaflops
  - 2004 : Zettaflops (extreme forward-looking focus)
- The Path to Extreme Supercomputing
  - 2004/10 Santa Fe, 30 people
  - Erik P. DeBenedictis, SNL, organizer
  - P. Jones, P. Kogge, W. Gropp, M. Frank, C. Lent
  - Panel: T. Stirling, H. Simon, T. Michalske, F. Johnson, W.Campなど
  - ちょうど地球シミュレータがトップを譲った頃

# Exascaleへの挑戦

- Horst Simonの2007/10の展望

<http://www.zettaflops.org/fec07/presentations/Monday-1330-Simon7Challenges.pdf>

1. No R&D program beyond Petaflops
2. Cannot afford EF until 2019 at current budget levels
3. Cannot afford the power requirements
4. Productivity for science not adequately addressed (data tsunami)
5. Application at 100K way parallel is hard
6. How to express parallelism
7. System software??

# Exascaleへの挑戦

- アメリカ: DARPA, NSF, DOE
- IESP (International Exascale Software Project)
  - SC08 (Austin)で発足、Santa Fe, Paris, つくば, Oxford, Maui, San Francisco, 神戸(2010/4)
- 2010/8: DARPA's Ubiquitous HPC Program selected --NVIDIA-Cray-ORNL troika, Intel, ... 終了？
- EESI (European Exascale Software Initiative):
  - 2010/6に発足、18ヶ月、4 WG's
  - 2010/11にAmsterdamで第1回会合
- 2011/1: Obama--State of the Union Address
  - Supercomputingに重点

# Exascaleへの挑戦

- 日本
  - 2008～：IESPへの参加
  - 2010/8～：SDHPC WS
  - 2011/7：WG（アプリ、システムソフト・アーキ）
    - 若手中心。老人はアドバイザーとして棚上げ
  - 2012/2：次世代HPCI-WG（2年の予定）
  - 2012/3：WG報告書
    - <http://www.open-supercomputer.org/workshop/sdhpc/>
  - 2012：FS（アプリ、3アーキ）
  - SC12：日米システムソフトウェア共同研究
  - 2013：日米科学技術協定

# Exascaleへの挑戦

- 日本(続き)
  - 2013/5: WG中間報告案、パブリックコメント
  - 2013/6/25: 正式な中間報告
    - 今後のHPCI計画推進の在り方について(中間報告)
  - 2013/7/2: 今後のHPCI計画推進のあり方に関する検討ワーキンググループ システム検討サブワーキンググループ
  - その他のワーキンググループ?
- 2013/8: 概算要求

[http://www.mext.go.jp/component/b\\_menu/other/\\_icsFiles/afieldfile/2013/08/30/1339148\\_5.pdf](http://www.mext.go.jp/component/b_menu/other/_icsFiles/afieldfile/2013/08/30/1339148_5.pdf)

# ポスト「京」(エクサスケール・スーパーコンピュータ)の開発

平成26年度要求・要望額 : 3,000百万円 (新規)  
うち優先課題推進枠要望額 : 3,000百万円

## 事業概要

スーパーコンピュータは科学技術の発展、産業競争力強化、安全安心の国づくりに不可欠な国家の基幹技術。我が国として、2020年頃までに「京」の約100倍の計算性能を有するエクサスケールのスーパーコンピュータの実現を目指し、ハードウェア等(ライブラリ等含む)の設計・開発に取り組むとともに、エクサ級スパコンに向けたアプリケーションの開発を行う。(平成26年度よりプロジェクト開始)なお、総事業費は1,200億円程度。

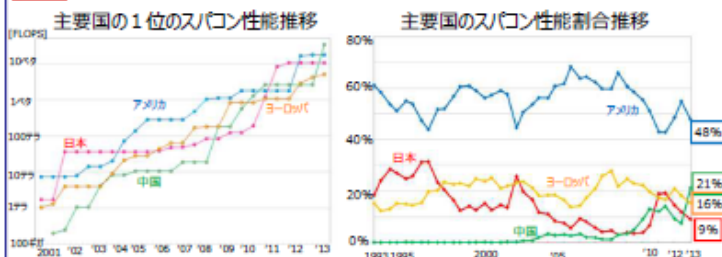
## 国際的な開発動向

スパコンは経済成長、国家安全保障、産業競争力・科学技術力強化に必須。国際的に開発・整備・導入が活発。

: 世界の計算性能の約半分, 2020年頃のエクサ級スパコン開発・稼働予定

: 日本を超える総計算能力, 2020年頃のエクサ級スパコンを整備予定

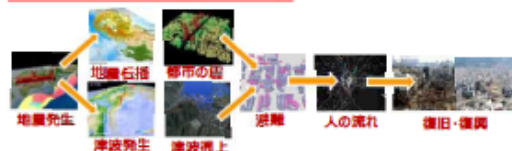
: 最新ランキングで一位獲得, CPUの自主開発を進めエクサ開発に着手



## エクサスケール実現の達成目標

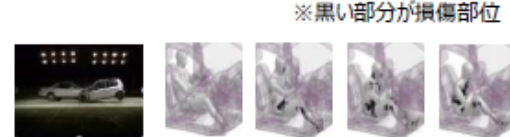
### <防災・減災対策>

- 「京」で1.5日の計算が1時間以内で可能となり、津波到達前の浸水被害予測が実現し、地震発生直後の住民避難誘導など初動対応に大きく貢献。
- 地震発生から避難予測までを統合した広域複合災害の被害予測によるきめ細やかな防災・減災・復興対策に貢献



### <ものづくり(自動車開発)>

- 「京」で40日の計算が10時間で可能となり、衝突試験コストの大幅削減等、自動車設計プロセスが大幅に効率化
- 自動車衝突時の影響を、車体だけでなく乗員の体への影響(骨や内臓等の損傷)も評価し、より安全性の高い車体の開発に貢献



## 実施内容

### 【平成26年実施内容】

(1)ハードウェア開発 : 2,188百万円

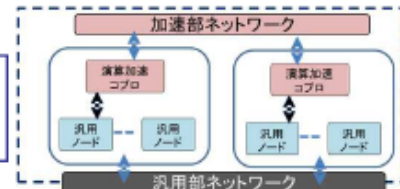
プロジェクト初年度となる平成26年度は汎用部及び加速部の基本設計を行い、ライブラリ等を含むハードウェア仕様の詳細について検討を行う。

(2)アプリケーション開発 : 812百万円

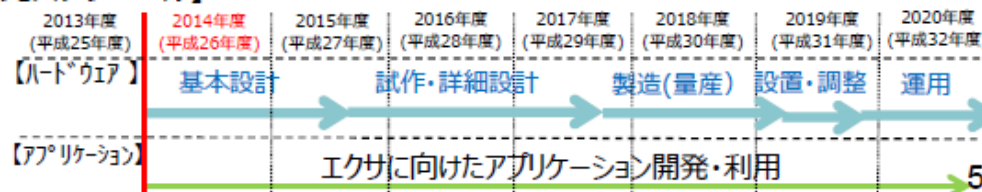
エクサスケールスパコンを最大限活用し、世界最高水準の研究成果を創出するため、必要なアプリケーションの開発、革新的なアルゴリズム開発に取り組む

### 【ハードウェアの仕様】

- アーキテクチャ : 汎用部 + 加速部
- 目標演算性能 : 1 I/Pコア級 (「京」の100倍)
- 消費電力 : 30~40MW (「京」は12.7MW)



### 【開発スケジュール】





# Exascaleへの挑戦

- Russia :
  - 2012/4: T-Platforms 6, IBM 17, HP 16, others 11
  - 2010: 自主開発への動き(\$37M, 2010)
  - National Exascale effort, \$1.5 b over several years  
人材育成、ミドルウェア開発、応用開発、相互接続、プロセッサなどをすべて含む
  - the Russian Academy of Sciences, T-Platforms, and RosAtom, the state controlled Russian nuclear regulatory body (応用は限定的)
  - 2014/15にIntel+ NVIDIAで10+ PFを狙う
  - 2017/18に自主開発プロセッサ(?)で100 PF
  - 2020頃のExascaleは自主プロセッサ

# まとめ

- 人類はExascaleに向かっている
  - Linpack 1 Exaflopsを意味する訳ではない
  - Exaでしか解決できない社会的・科学的課題
  - それを実現するためのコンピュータシステム
- 日本はどこまで自主開発すべきか
  - 基幹技術は何か(歴史的)
  - グローバル時代の国際協力
  - 技術安全保障
  - 納税者、政治家の理解