

スーパーコンピューターの歴史 —— 「京」 コンピュータまで、その先

牧野淳一郎

理化学研究所 計算科学研究機構

エクサスケールコンピューティング開発プロジェクト

コデザイン推進チーム チームリーダー

話の構成

- 自己紹介
- スーパーコンピューターの歴史
- スーパーコンピューターの進歩を生むもの、阻むものは？
- 国策スパコン
 - 「京」を振り返る
 - 歴史的には？
- もうちょっと色々な科学技術政策
- 何故そうなるのか？
- で、「京」の次は？

自己紹介

職歴

- 1990/4- 東京大学教養学部 情報図形科学教室 助手
- 1994/4- 東京大学教養学部 情報図形科学教室 助教授
- 1999/4- 東京大学大学院理学系研究科天文学専攻 助教授
- 2006/6- 国立天文台理論研究部 教授
- 2011/4- 東工大理学研究流動機構 教授
- 2012/4- 理化学研究所 計算科学研究機構 粒子系シミュレータ研究チーム
チームリーダー (併任)
- 2013/1- 東工大地球生命研究所 主任研究者
- 2014/4- 理化学研究所 計算科学研究機構
エクサスケールコンピューティング開発プロジェクト
コデザイン推進チーム チームリーダー

天文 (天体物理) 学者？ 計算機科学者？

自己紹介

職歴

- 1990/4- 東京大学教養学部 情報図形科学教室 助手
- 1999/4- 東京大学大学院理学系研究科天文学専攻 助教授
- 1994/4- 東京大学教養学部 情報図形科学教室 助教授
- 2006/6- 国立天文台理論研究部 教授
- 2011/4- 東工大理学研究流動機構 教授
- 2012/4- 理化学研究所 計算科学研究機構 粒子系シミュレータ研究チーム
チームリーダー (併任)
- 2013/1- 東工大地球生命研究所 主任研究者
- 2014/4- 理化学研究所 計算科学研究機構
エクサスケールコンピューティング開発プロジェクト
コデザイン推進チーム チームリーダー

天文 (天体物理) 学者？ 計算機科学者？

なんだかよくわからない。

自己紹介続き

学歴

1985/3 東京大学教養学部基礎科学科第二卒業

1987/3 東京大学大学院総合文化研究科広域科学専攻修士課程修了

1990/3 東京大学大学院総合文化研究科広域科学専攻博士課程修了

自己紹介続き

学歴

1985/3 東京大学教養学部基礎科学科第二卒業

1987/3 東京大学大学院総合文化研究科広域科学専攻修士課程修了

1990/3 東京大学大学院総合文化研究科広域科学専攻博士課程修了

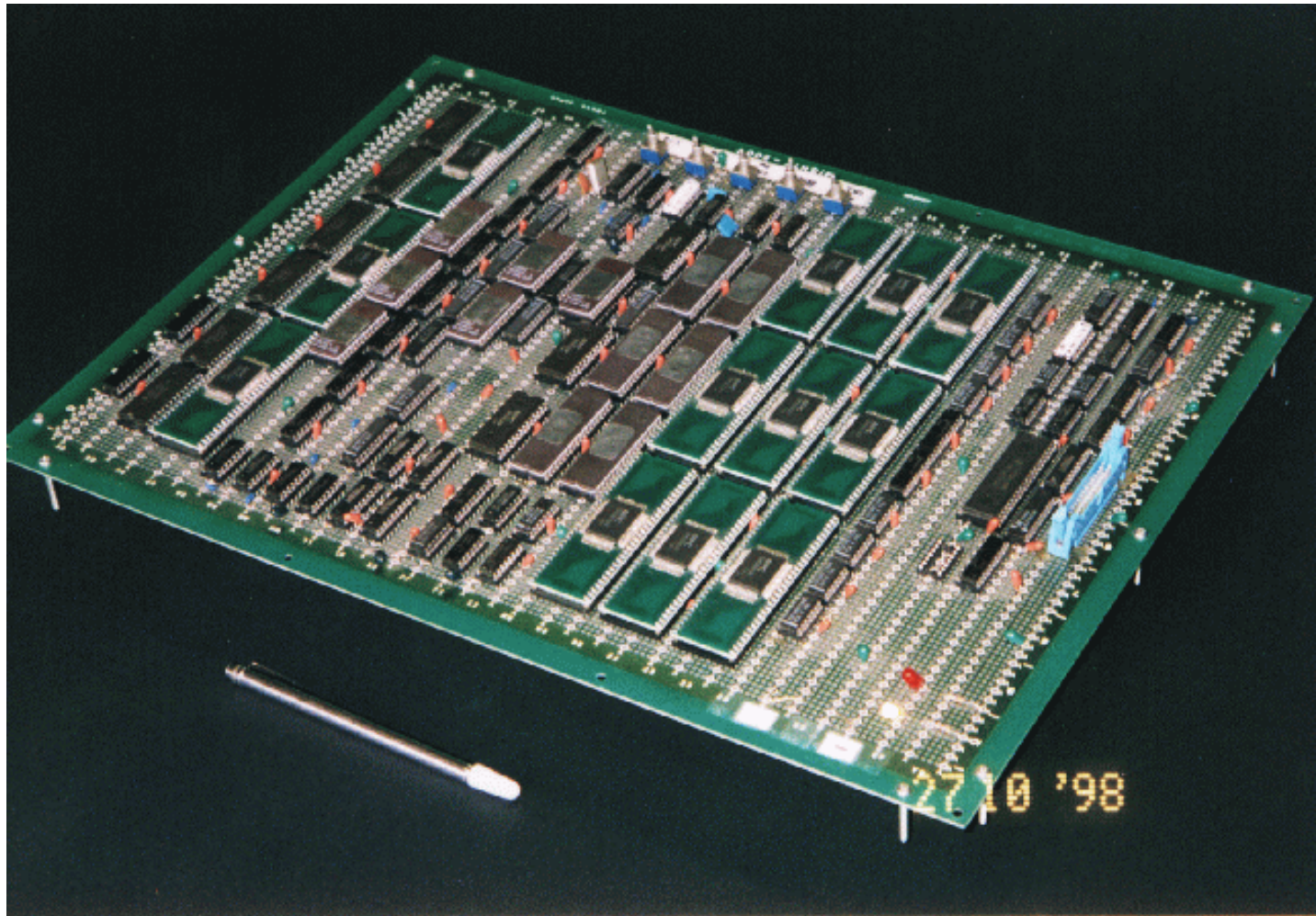
やっぱりよくわからない。

自己紹介(続き)

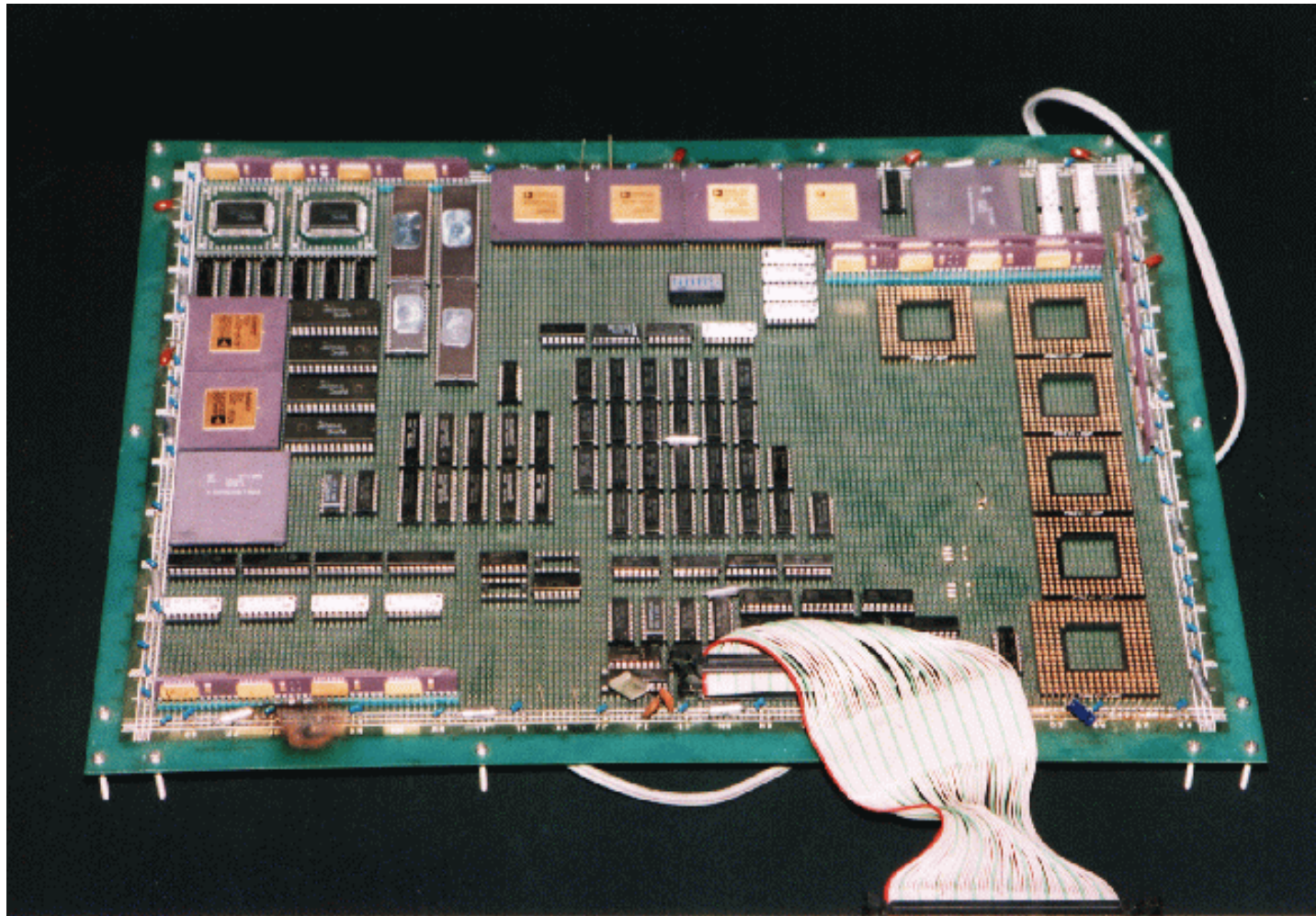
研究はどんなことをやってきたか

- 主に理論・シミュレーションによる天体形成・進化の研究
 - 宇宙の大規模構造の形成・銀河形成
 - 銀河中心・球状星団の力学進化
 - 惑星形成
- シミュレーションのための計算アルゴリズムの研究
- シミュレーションのための計算機の開発

GRAPE-1(1989)



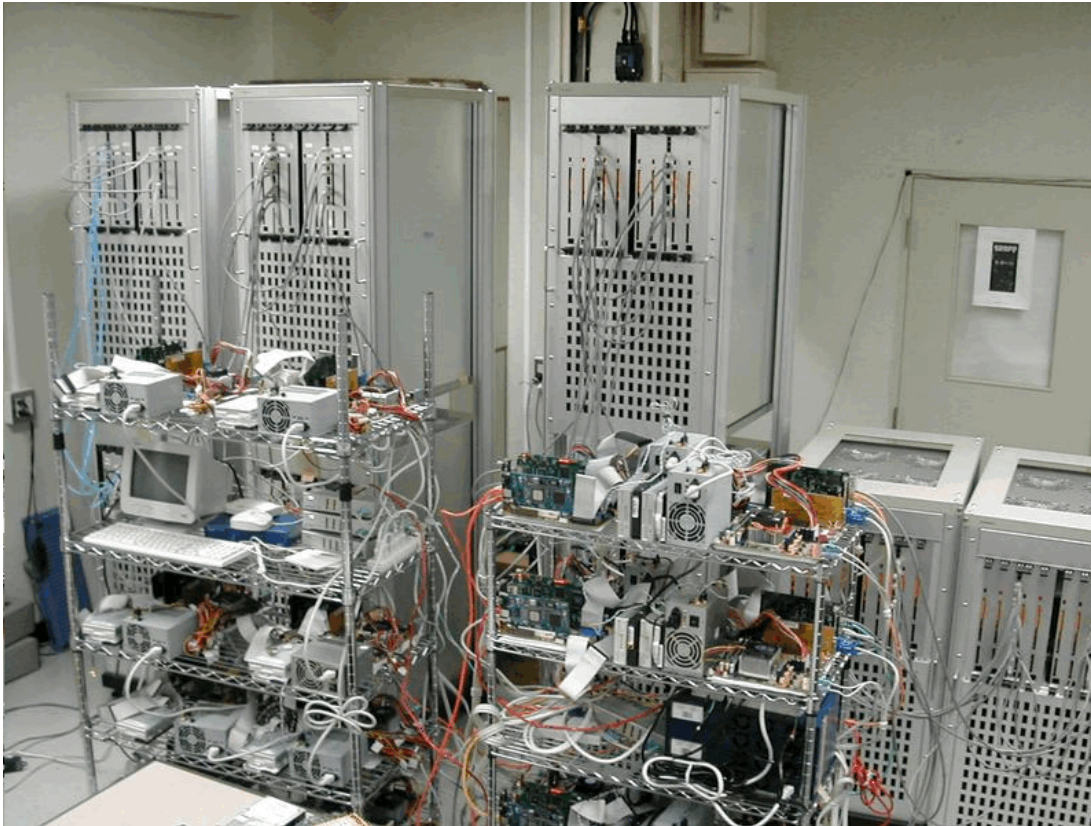
GRAPE-2(1990)



GRAPE-4(1995)



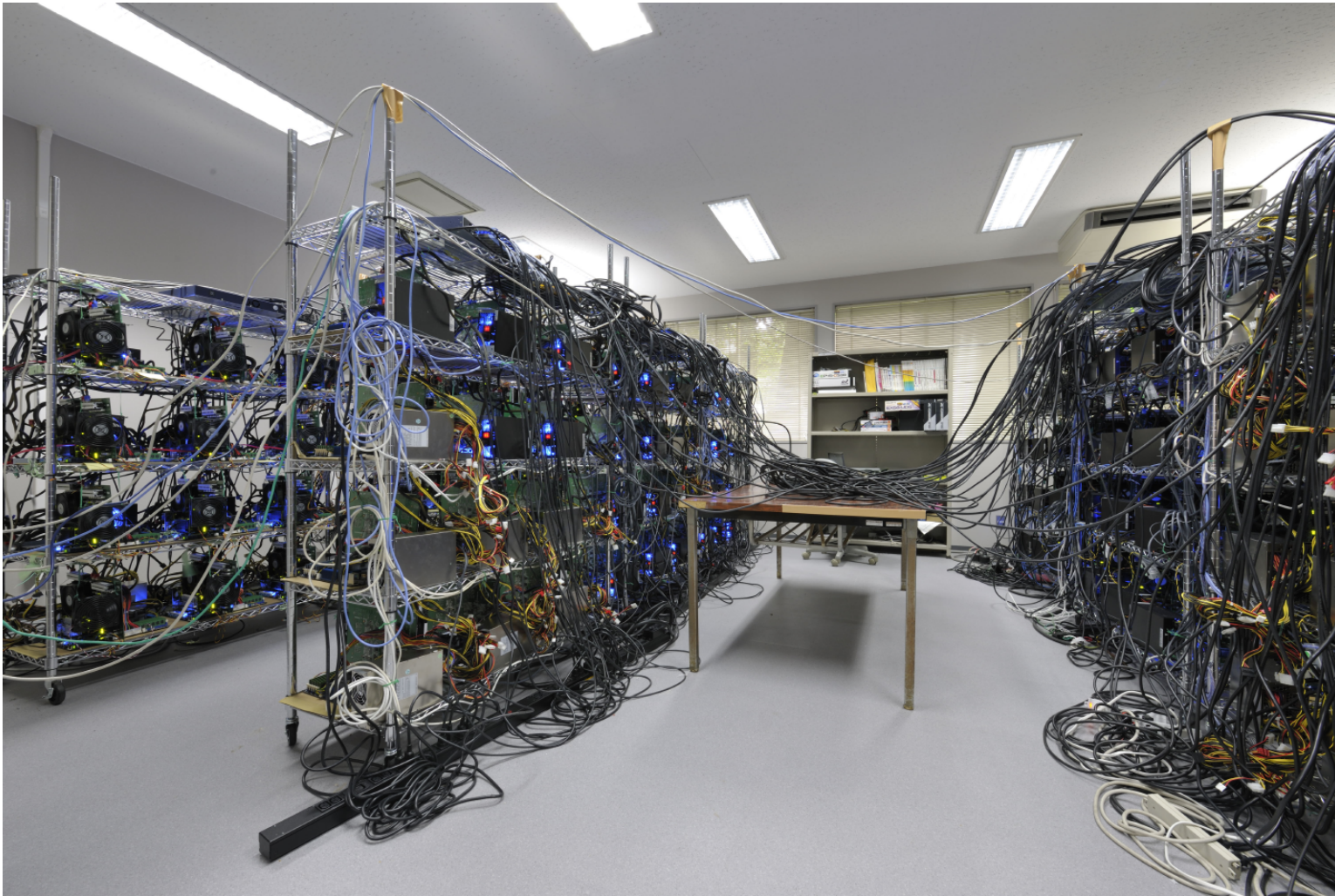
GRAPE-6(2002)



2002年の 64 Tflops
システム

初代地球シミュレー
タの 1.5 倍の性能
を 1/100 のコスト
で

GRAPE-DR(2009)



自己紹介 (続き 2)

2011 年度

文部科学省 研究振興局 情報課 が作った

HPCI 推進委員会 が作った

今後のハイパフォーマンス・コンピューティング（HPC）技術の研究開発
の検討ワーキンググループ
の答申を踏まえてできた

今後の HPC 技術の研究開発を検討する作業部会 (コンピュータアーキテクチャ、コンパイラ・システムソフトウェア、アプリケーションの 3 個)
のうちアプリケーション作業部会の取りまとめ作業を理研 (神戸の AICS)
の富田さんとやった

自己紹介 (続き 3)

2012-2013 年度

- 「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」
委員
- 「将来の HPCI システムのあり方の調査研究」にも参加

2014 年度

- 何とかチームリーダー。次期システムの開発に参加 (しているはず)

「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」

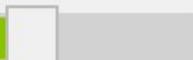


“京”の100倍の性能、
新型スパコン開発へ

新型スパコン

1,000億円ほどの予算で開発予定

NHK



00:48 / 01:20



国策スパコン

国家プロジェクトとしての計算機開発

- 通産省系

- 科学技術用高速計算機システム (1981-89)
- 第五世代 (1982-91)
- RWCP (1992-2001)

- 文部省系

- 数値風洞 (?-1993)
- CP-PACS (1992-96、国家プロジェクトというほどではない?)
- 地球シミュレータ (-2002)
- 「京」 (2006-2012)

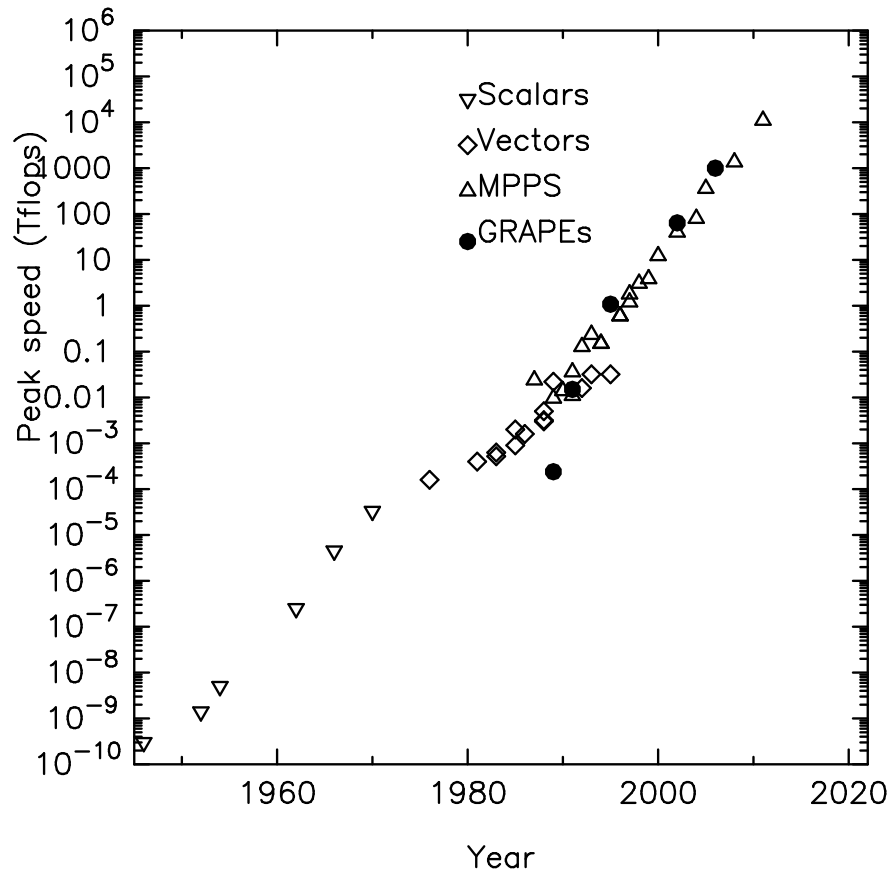
通産はこの前に: 「超高性能電子計算機の開発」 (1966-?)

とかいう話の前に

そもそもスパコンって何？パソコンとどう違うの？何故違うの？

という話を少し。

スパコンとは何か



- 単純には、「その時代で最高速クラスの計算機」
- 70年間で 14桁速くなった＝ほぼ10年で100倍
- 何故これほど進歩したか？は「スパコンとは何か」そのもの

1940年代の「スパコン」



ENIAC

ENIAC

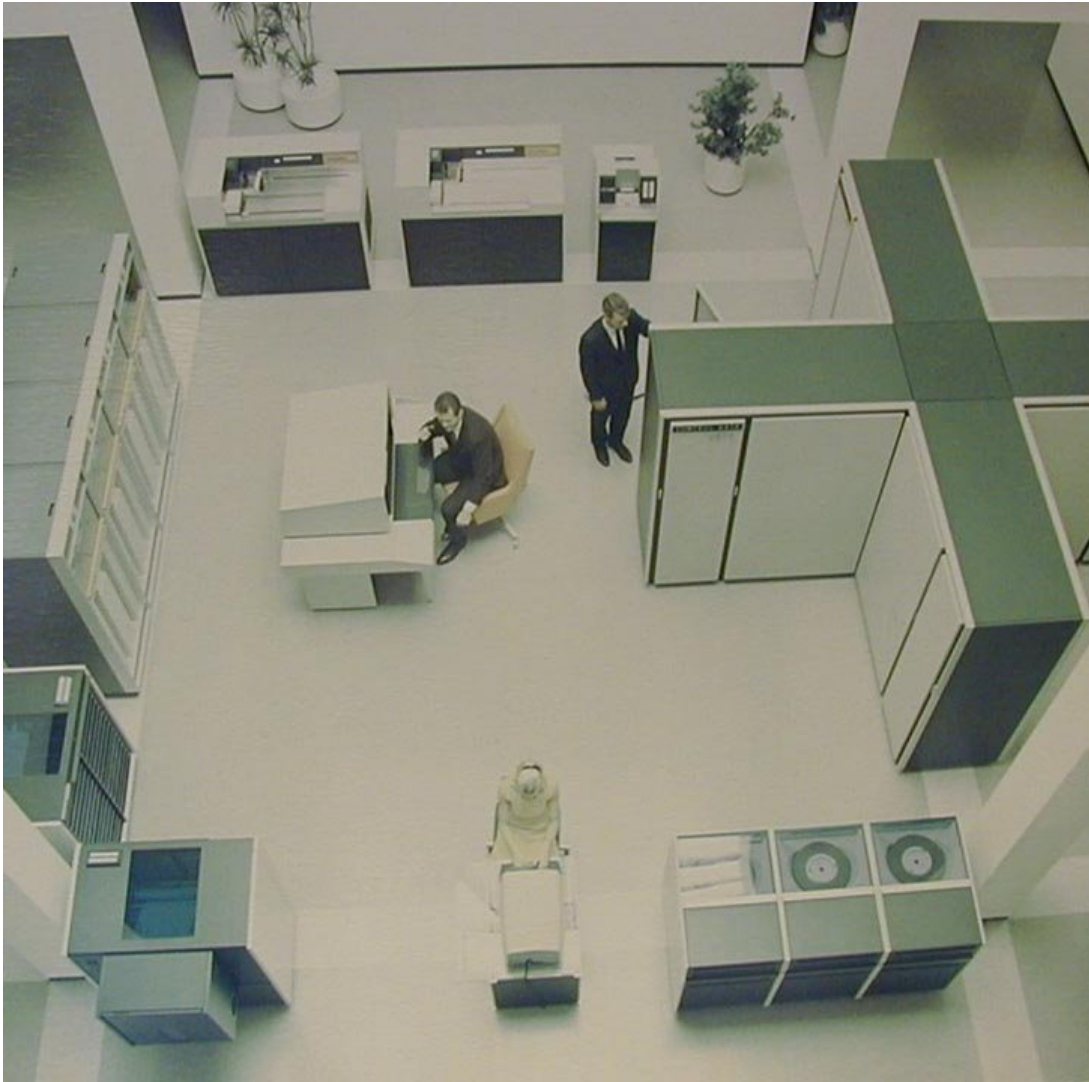
- 第二時世界大戦中に、アメリカ陸軍が弾道計算のために開発。完成は1946年
- 18,000本の真空管、消費電力 150KW、10進法10桁の数が基本
- 「プログラム」はスイッチ、ケーブルの変更で行う
- 加減算が1秒5000回、乗算が400回くらい(らしい)

1950年代の「スパコン」



IBM 709 — IBM 最後の真空管計算機

1960年代の「スパコン」



CDC 6600 1Mflops

CDC6600

- 1964 年から出荷。シリコントランジスタを利用。
- 10MHz クロック、1Mflops 程度の性能
- 近代的計算機アーキテクチャの原型。
 - ロード・ストアアーキテクチャ
 - 複数の演算器の「アウト・オブ・オーダー」実行
- シーモア・クレイが設計

1970年代の「スパコン」



CRI Cray-1 160Mflops

Cray-1

- 1976 年から出荷。IC (ECL 素子による小規模IC) を使用
- 80MHz クロック、160Mflops
- 最初に成功したベクトル計算機
 - ベクトルレジスタ
 - 半導体メモリ
- シーモア・クレイが設計

ベクトル計算機って？

- 「スカラー計算機」と「ベクトル計算機」と分ける
- 「ベクトル処理」をするかどうか
- スカラー計算機 (あるいは普通の計算機の基本命令): 命令1つで演算1つ。例えば掛け算1つとか。
- ベクトル計算機: 1命令で複数の演算を処理。ベクトルの、要素同士の四則演算が基本。それだけでは足りないので色々追加機能がある。
- 必ずしも並列処理するわけではない。Cray-1 では「パイプライン処理」。長さ n のベクトル命令の実行に 定数 $+ n$ クロックかかる。

ベクトル計算機の利点

- 演算器の有効利用ができる。クロックサイクル毎に結果を得ることが可能。
- 普通のスカラー計算機だと、演算以外の色々な処理が必要なので、複雑な「スーパースカラー」実行をする仕掛けが必要。ベクトル計算機は単純なハードウェアで演算器を有効利用できる。

欠点:

- メモリとベクトルレジスタの間のデータ移動が高速であることが前提: 高速なメモリがないと性能がでない。

1980年代の「スパコン」



NEC SX-2 1.3Gflops

NEC SX-2

- 1985 年から出荷。CML 論理素子による LSI を使用。
1000 ゲート程度 (Cray-1 の IC の 100 倍程度)
- 6ns クロック、掛け算、加減算パイプラインをそれぞれ 4 本
- 富士通、日立が同様なマシンを出荷、NEC は最後
- Cray は共有メモリマルチプロセッサに (Cray XMP, YMP)

1980年代にはこんなのも



PAX-128 4Mflops。筑波大学、星野ら

1990年代の「スパコン」

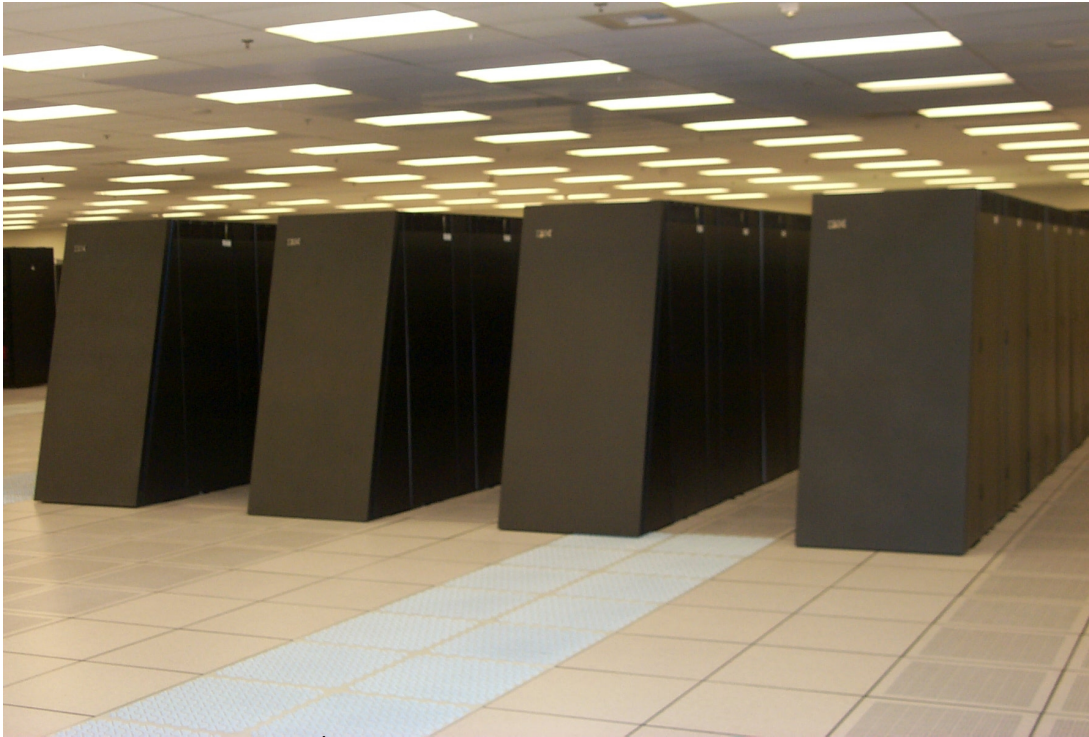


ASCI-Red, Intel Pentium Pro, 1.8TF

ASCI-Red

- 1996年に完成。200MHz Pentium Pro プロセッサ 9216台。
- 3次元メッシュネットワーク (ほぼ2次元、、、)
- 核爆弾のシミュレーション用

2000年代の「スパコン」



IBM BG/L 360TF

IBM BG/L

- 2004年完成、カスタムプロセッサを最大 131072 個使用。
3次元トーラスネットワーク
- ピーク性能 360TF (もっと大きい構成もあった模様)
- 後継の BG/P, アーキテクチャを一新した BG/Q の開発後、プロジェクト解散

2010年代の「スパコン」



神戸の「京」コンピュータ、10PF

「京」

- 2012 年完成、カスタムプロセッサを 8 万個使用。6 次元トーラスネットワーク
- ピーク性能 11PF
- 商用版富士通 FX10, 「ポスト FX10」が開発された。その後継が次の国家プロジェクトで開発される予定

現在 (2000 年以降) の普通のスパコン

- Intel のプロセッサを使用
- アプリケーションにあてれば NVIDIA の GPU を使用
- 沢山並べる
- 高速ネットワークを使う、あるいは Cray 社のシステムを
買う
- 独自プロセッサとかはあまり使われない。
- あんまり芸がないので写真は省略

過去のスパコンの進化

何の話をしたかったかというと： 何故計算機はどんどん速くなったのか？
基本的な理由：

- 使うスイッチ素子が高速になった
- 使うスイッチ素子が小型、低消費電力になって、沢山使えるようになった
- 使うスイッチ素子が安くなって、沢山使えるようになった
(スパコンの物理的大きさは70年代が最小。そこまで段々小さくなって、そこからまた大きくなった)

素子の高速化

- といっても、真空管でもそれなりに速かった。
- スイッチング速度が重要でないわけではないが、配線を信号が伝搬する速度のほうが昔から重要。
- 昔は信号はほぼ光の速さでつたわった。
- 最近の LSI 上の配線は非常に細く (抵抗が大きく)、キャパシタンスを充電しないといけないことによる RC 遅延のため、信号が伝わる速度は光速度よりはるかに低い。
- 太い配線に大電流を流せば速いが、大量の電力消費になる

素子の小型化

- 真空管 → トランジスタ → IC という進化は 1970 年代までは重要
- サイズだけでなく、消費電力が下がることが重要
- 80 年代から重要になったのは CMOS LSI の微細化。10 年でサイズが $1/10$ になる
- CMOS 素子では (2000 年くらいまでは) 微細化すると電圧を下げることができ、消費電力が下がり、速度は向上した。
- 2000 年頃からは電圧が下がらないので、電力はちょっと下がるが速度は上がらなくなった
- 80 年頃から現在までで、集積度 100 万倍、速度 1000 倍。但し、2000 年くらいで速度向上は止まるので、トランジスタの速度は 100 倍くらい

素子の性能向上、サイズ低下とスパコンの性能向上

「スパコンの進歩」は4つの時期に分けられる

- I スパコンが完全パイプライン化した乗算器をもたない時期
(1969年まで)
- II スパコンは1つ以上の演算器をもつが、1チップにはまだ演算器1つが入らない時代 (CMOS では1989年まで)
- III 1チップに複数の演算器がはいるが、微細化すると動作クロックが上がり、電力が下がった時代 (2001年頃まで)
- IV チップに多数の演算器がはいるが、微細化してもクロックが上がり、電力がへらない時代 (2001-)

このそれぞれで、素子の性能向上がどう使われたかは違う。

I期: 1 演算器未満の時代

- **この時期には、メモリアクセスのほうが演算より速い**
- **演算器の性能向上が最重要課題**
- **CDC 6600 くらいまで。**
- **計算機の構成方法もまだ手探り。**
- **CDC 7600 で1 演算器はいるようになる**

II期: 1 演算器以上の時代

- 1つ以上の演算器を有効に使うことが課題
- パイプライン化 (ベクトル命令)、複数の演算ユニットの SIMD and/or MIMD 並列実行が必要になる
- 演算器の数が増えると、メモリとどうつながかが課題になる。
- 演算器 16-32 個で破綻する (1992 年頃): メモリと演算器の間のデータ移動回路が大規模・複雑になり過ぎるため
- 末期の計算機: Cray T-90, 日立 S-3800
- 富士通 VPP500 では、分散メモリにすることで当座しのぎとした:ある程度の成功
- 他の方向: 1チップマイクロプロセッサでの分散メモリ。プロセッサ間の通信は細い線でいいことにする。これが90年代以降の主流になった

III期: 1チップ高速化の時代

- 1チップマイクロプロセッサを分散メモリで多数つなぐの
は II 期末期から
- 1チップに1演算器以上がはいるが、ありあまるトランジ
スタを演算器を増やすのにはつかわないで動作クロックの
向上、キャッシュメモリの大型化に使った時代
- 1990 年代。 牧野の意見としては「失われた10年」
- 計算機全体としては II 期に起こった問題がチップレベル
でも起こるのを先送りにした
- メモリとの接続は速度が不足になった (memory wall): キ
ャッシュでごまかす方針
- 迷走したプロセッサ開発も多い: 各社の複数チップ共有メ
モリプロセッサ。特定目的には有用だが、スパコンのビル
ブロックには本来なるべきではない。(失敗プロジェクト:
ASCI Red 以外の ASCI マシン)
- 末期の計算機: Intel Pentium 4, DEC Alpha 21264

IV 期: 1 チップ多演算器化の時代

- 引き続き、1 チップマイクロプロセッサ、分散メモリ。
- デザインルール 130 nm あたりから、動作電圧低下、速度向上に限界
- マルチコア化、コア内 SIMD 化を同時に進めている
- 80 年代のベクトルスパコンの辿った道を追いかけている
- ということは、16-32 コアで破綻がくるはず
- 破綻がくることの予見: Intel Xeon Phi (60 コア)
- GPGPU はちょっと別。

まとめると

- 昔の Cray-1 みたいな「スパコン」の進歩は、プロセッサコア 32 個くらいの並列度で限界、破綻した
- 現在のマイクロプロセッサは、それくらいのコア数に達してきており、破綻が近い
- 破綻をどうするか、という展望が必要。

ということで、国家プロジェクトとしてのスパコン開発はそういう展望を与えるものであったかどうかを振り返って(または予見して)みよう。

国策スパコン(再掲)

国家プロジェクトとしての計算機開発

- 通産省系

- 科学技術用高速計算機システム (1981-89)
- 第五世代 (1982-91)
- RWCP (1992-2001)

- 文部省系

- 数値風洞 (?-1993)
- CP-PACS (1992-96、国家プロジェクトというほどではない?)
- 地球シミュレータ (-2002)
- 「京」 (2006-2012)

- 通産はこの前に: 「超高性能電子計算機の開発」 (1966-?)

- 通産のプロジェクトは今回みなかったことにする

数値風洞 (1993)

- II期からIII期への移行の一方の象徴。CMOS 技術ではなく、まだ1チッププロセッサではないが、分散メモリに移行した
- 分散メモリに移行したため、従来のプログラムは基本的に全て走らなくなった。「VPP-Fortran」という特殊な方言で書くことが必須
- 普通はこんなの絶対メーカーが作らないが、、
- ネットワークは完全クロスバーとした。高コストだが、共有メモリベクトルプロセッサのメモリインターフェースに比べるとたいしたことはなかった。
- 商業的にも成功だったといえる。同時期の共有メモリベクトルプロセッサの10倍のピーク性能を実現したことが大きい。
- 他の計算機の10倍速くなるならプログラムの書換えをいとわない人は結構いるということ

CP-PACS (1996)

- 1チップマイクロプロセッサによる分散メモリ並列計算機
- III 期中期の標準的設計。革新的とはちょっといいがたい
- この時期の計算機としては高メモリバンド幅。1チップの速度はそこそこ、という意味では若干時代に逆行。
- ネットワークは豪華。同軸ケーブルの山 (3次元ハイパークロスバーというもの)
- ハイパークロスバーというとクロスバーよりいいみたいだが、そうではない。
- 商業的には微妙

地球シミュレータ

- 1 チップベクトルプロセッサ、8チップで共有メモリ、640ノード構成。
- ノード数を抑えることで、数値風洞と同様のクロスバー構造を維持
- 共有メモリとしたことで、メモリ、メモリインターフェースは高コスト、大消費電力に
- 数値風洞からアーキテクチャが逆行した(部分的に複数プロセッサ共有メモリにもどった)
- つまり、設計としては IV 期になって現れた II 期の亡霊的存在
- 商業的には微妙

「京」

- 8コアマイクロプロセッサ。2SIMD、2スーパースカラー
実行で4演算器
- IV 期中期のある意味標準的設計
- 細かくみると、8コアで L2 キャッシュを共有するのは II
期末期的。標準的な分散 L2 から見ると時代に逆行
- メモリバンド幅が高いのもどちらかというと逆行的
- 多次元トーラスは時期的には微妙
- 商業的には成功とはいいがたい。「京」以外の最大のシス
テム:東大基盤センターの 4800 ノード、60 億円。全部合
わせても 100-150 億円？

革新と成功

国家プロジェクト（あるいはそれに近い巨大プロジェクト）でのスパコン開発を見ると

- 革新的でなければ商業的成功に結びつかない
- 革新的＝他の計算機より圧倒的に高い性能

「京」を例に、「どうしてこうなった」をみていこう

「京」の概要

- 2006-2012の7年間で開発。プロジェクト総経費1200億円
- ピーク性能10PF強、LINPACK性能10PF以上
- 約8万ノード、3次元(本当はもうちょっとややこしい)トーラスネットワーク
- 1ノードは1ソケット、SPARC-64 VIIIfx (Venus) プロセッサ
- 8コア、2way SIMDユニット2つ、2GHz、128Gflops
ピーク
- プロセッサ消費電力は58W(typical)、システム全体は10MW
以上

大体の時系列

- 2004 年くらいから文科省の下の情報科学技術委員会、計算科学技術推進ワーキンググループで議論
- 2005 年の「中間報告」
 - － 8 分野からの要求をまとめた (ことになっている)
 - － ベクトル 2PF + スカラー 4PF + 「特定処理計算加速機」 20PF
- 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」実質的にはデザインコンペ。NFH の他、東大 (+ 天文台)、九大、筑波大等も参加

時系列続き

- 2005 年から 2006 年にかけて CSTP 評価委員会からボロクソに言われる。目標、アーキテクチャを文句がでないように色々変更。
- 2006 年初め (だったと思う) 開発実施本部設置、ヘッド:渡辺 (NEC OB)
- 2007 年夏: ベクトル (?PF) + スカラー (?PF) 合計 10PF 以上、と決定
- 2009 年春: ベクトル担当の N が撤退
- 2009 年 11 月 13 日 (金) 仕分け。「2 位じゃいけないんですか？」
- 2010 年 10 月。プロトタイプ機が Green500 4 位にランクイン
- 2011 年 6 月。8 割完成で Top500 1 位
- 2011 年 11 月。全ノード動作で Top500 1 位。ゴードンベル賞獲得

時系列続き

- 2012年11月。2度目のゴードンベル賞獲得(石山、似鳥、牧野)

「京」の問題点

- 値段が高い
- 2012年のマシンとして特色がない(競争力がない)

何故そうなったか？が問題

「京」の進み方の問題点と思われるもの

- **何をするために作るのかを「ワーキンググループで議論」**
- **結局、「何のため」かははっきりしないままスタート**
- **計画スタートまでに目標性能が二転三転した。**
- **アーキテクチャも二転三転した。**
- **開始してからメーカーが降りるという前代未聞の事態にいたった。**

どうしてそんなことになったのか？(1)

何をするために作るのかを「ワーキンググループで議論」

- 「スパコン開発は重要」が先にある議論＝「何故重要か」を説明するための議論。
- 成功したプロジェクトではこんなステップはない
- 色々沢山の分野の人に話をきいて全部入れる、という感じのレポートになる
- 必然的に「革新」は排除される。全ての人が賛成するようなものは(特に使う側の人賛成するものは)革新的ではないから

どうしてそんなことになったのか？ (1a)

「スパコン開発は重要」の「大前提」

- この「大前提」がどこからくるのかは語られることがないので良くわからないが、多分こういうのかなと想像してみる
- 多分、背景にあるのは過去の「成功体験」
- 1960-80年代の垂直統合型の計算機＋半導体産業モデル
- 当時のIBMに「追いつけ追い越せ」

60-80年代的モデル

- 半導体技術は「産業のコメ」
- 大規模・微細プロセスの半導体製造を牽引するのはハイエンドの超大型計算機
- 超大型計算機に使われた半導体技術が普及型の計算機やそれ以外のデジタル技術に波及

1980年代までは上手くいっているように見えていた

でもって

老人の学者さんとか、政策決定に関与する人の頭には依然これがある。で、このモデルによってお金が流れる構造はできている。

「産業のコメ」の嘘

- 今から振り返ると明らかなこと:このモデルは遅くとも1980年頃には破綻していた
 - マイクロプロセッサは初めから MOSであったように、民生技術は CMOSに 70年代に移行。スパコンが CMOS化したのは 90年代 (ETA 除く)。20年遅れた。
 - 古いアーキテクチャで CMOS化したスパコンはマイクロプロセッサベースの並列システムに駆逐された
 - つまり II期から III期への移行
- しかし、1980年代にはそうは見えなかった

1980年代

- 日米半導体摩擦の時代
- スパコン摩擦というのもあった。
- これらが一体となって「日本株式会社」の優位性みたいな
- どちらも今は見るかげもなく、、、

日本の半導体産業の凋落と崩壊

- LSIチップそのものを作ることを半導体産業の主要部分とするなら、日本からは事実上半導体産業は消滅した
- これ結構知らない人が偉い人には多くてびっくりすることがある
- 28nm の論理 LSI を自社製造する国内企業はない
- 日立は 130、富士通は 65、NEC は 55(?) で終了
- つまり、日本の「半導体企業」は全部ファブレス化した
- 元々ファブレスで始めた会社と競争できるわけがない

GRAPEでの経験

何故日本の論理LSIは駄目だったかについての個人的感想

- 実はGRAPEでは国内の半導体メーカーを使ったことがない
- 一応GRAPE-6は東芝。但しこれはIBMとライブラリ共通化して、IBMで設計、東芝で製造。
- それ以外はNS、LSI Logic、TSMC

日本メーカーは大規模LSIの少量生産に対応したがない(できない)

＝設計能力がない(設計が高コスト)

何故設計能力がないのか

- 私にはよくわからない
- 但し、国内某社と海外某社を比べると、設計の方法論レベルで5-10年のギャップあり
- 単純に変化の速度に適応できていないだけのようにも見える

「京」の話に戻ると

- Fが開発を受けるにあたっては自社の45nmプロセスで、
という条件があったらしいという噂
- 噂が事実かどうかはともかく、Fは「京」のプロセッサは
自社のライン(但し古くスケールの量産用ではない)で製造、
商品版のFX10はTSMCの40nm
- どういうわけか45nmから40nmになっただけで8コアから
16コアに。
- 結果的には、Fの45nmは「するべきではなかった投資」

目標設定における問題のまとめ

- 実は大型計算機開発への国家投資が、計算機だけでなく半導体産業を通して経済発展全体に貢献するはずという30年前に破綻したモデルが暗黙の大前提としてある。
- 結果的に、生き残れなくなった半導体製造産業に公共投資
- 半導体各社が今後どうするか?というのが根本問題にある。明らかなことはファブレスで生き残ってはいけないということ。

どうしてそんなことになったのか？(2)

計画スタートまでに目標性能が二転三転した。

- 要するに、初期の目標が予算の割に低過ぎて国際競争力がないものだったため(今でも低い)
- そうなった直接の理由(ここは推測): メーカーの見積りをそのままあげたから
- メーカー見積りが高い理由: アーキテクチャ・開発モデルが時代遅れになっていたから

つまり:

開発することにしたものも、開発のしかたも間違っていた

結局、開発自体が無理、ということでメーカー撤退。

どうしてそんなことになったのか？(3)

アーキテクチャも二転三転した。

これも、根本には「暗黙の大前提」がある。スパコンを開発することが目的で、何のために何ができるスパコンを作るか、が欠落している

ターゲットアプリケーション

<http://www.nsc.riken.jp/target-application/target-application.html>

名前がでたのは 2007/4 だが選定は 2006/4 までになされていた:

- 1 Cavitation 有限差分法によるキャビテーション流れの非定常計算
- 2 COCO 超高解像度海洋大循環モデル
- 3 FrontFlow/Blue Large Eddy Simulation (LES) に基づく非定常流体解析
- 4 FrontSTR 有限要素法による構造計算
- 5 GAMESS FMO分子軌道法計算
- 6 GNISC 遺伝子発現実験データからの遺伝子ネットワークの推定
- 7 LANS 航空・宇宙機解析における圧縮性流体計算
- 8 LatticeQCD 格子QCDシミュレーションによる素粒子・原子核研究
- 9 MC-Bflow 血流解析シミュレーション
- 10 MLTest オーダーメイド医療実現のための統計的有意差の検証

ターゲットアプリケーション(2)

- 11 Modylas 高並列汎用分子動力学計算ソフトウェア
- 12 NICAM 全球雲解像大気大循環モデル
- 13 NINJA/ASURA 天体の起源を探る超大規模重力多体シミュレーション
- 14 Octa 粗視化分子動力学計算
- 15 PHASE 平面波展開第一原理分子動力学解析
- 16 ProteinDF 巨大タンパク質系の第一原理分子動力学計算
- 17 RISM/3D-RISM 溶液内タンパク質の電子状態の 3D-RISM/FMO 法による解析
- 18 RSDFT 実空間第一原理分子動力学計算
- 19 Seism3D 地震波伝播・強震動シミュレーション
- 20 sievgene/myPresto タンパク質・薬物ドッキングシミュレーション
- 21 SimFold タンパク質立体構造の予測

但し

実際に性能評価に使われたのは HPL, FFT と以下の7個

- SimFold, Modylas: 古典MD
- GAMESS, RSDFT: 量子化学
- LANS, NICAM: 流体
- LQCD: 4次元格子でのCG反復

この時点で、

- したい計算
- 高効率でできる計算
- 高効率でできるがオーバースペックな計算

の乖離

大雑把なアプリケーションの特性

- 古典MD、量子化学: メモリバンド幅もネットワークもあまりいらない
- 流体: メモリバンド欲しい。ネットワークはそこそこ
- QCD: メモリ量いらない。メモリバンド幅もネットワークも欲しい

これからわかること:

- 一台で全て満たそうとすると帯に短し襷に長しになる
- 低レイテンシ要求があるアプリケーションははいってない

「京」のデザインポイント

- CPU:128Gflops, 64GB/s, 16GB
- ICC: リンク 5GB/s x 2 x 10, CPU 20GB/s, レイテンシ:論文に書いてない
- 最近の計算機にしては B/F 重視
- 最近の計算機にしてはネットワーク重視
- CPU, ICC 別チップ

BG/Q との比較

	「京」	BG/Q	比率
ピーク速度 (GF)	128	204.8	—
メモリバンド幅 (GB/s)	64	42.6	2.4
ネットワーク (GB/s)	5	2	4
電力あたり性能 (GF/W)	0.85	2.1	(1/2.5)

演算性能あたりで、「京」は BG/Q の 2.4 倍のメモリバンド幅、4 倍のネットワークバンド幅を持ち、2.5 倍電気を食う。

どうしてこうなったか？

形式的な理由：要求仕様にそう書いてあった

- HPL 10PF
- 電力 30MW 以下
- アクセラレータ付けるならやはり 10PF
- この範囲でアプリケーションの性能を上げること

アプリケーションの側からは

欲しいもの:

俺のコードが速く走る計算機をよこせ!

但し: 「俺のコード」には温度差あり

- 必要ならアセンブラでも
- コンパイラの出力見るのは基本だよね?
- プロファイラ使ったことある
- OpenMP という言葉聞いたことがある
- 並列化ってコンパイラがしてくれるんでしょ?

最近の問題

アセンブラで書くとかくらいではどうにもならないことがある

- メモリ階層への対応
- 分散メモリ並列化

アルゴリズム、データ構造どちらも根本から見直す必要あり

1 ステップもどって考える

何故分散メモリ＋深いメモリ階層なのか？

チップ内の処理能力の増加に比べてチップ間データ転送能力が上がらないから

これは、言い換えると、現在の典型的マイクロプロセッサの設計では

- 演算能力はデータ転送能力に比べて相対的に安い
- これは製造コストだけでなく電力コストも

ということ

さらに言い換えると

- 「京」より演算性能を大幅に上げるのは、メモリバンド幅やネットワークバンド幅をあげなくてもいいなら電力・コストを大きくは増やさなくてもできたはず
- 逆に、HPL 10PF だけならもっと安価にもできたかもしれない
- メモリバンド幅を増やすなら全く別の作りかたもあったかも

もうちょっと詳しくアーキテクチャについて

2012年に公開された資料からいくつか

第179回国会 決算行政監視委員会 第 4号平成二十三年十二月八日

- スーパーコンピュータ「京」については、当初のスカラー・ベクトル混合型の技術選択がなぜ途中で方針転換されたのか、また、その変更が予算執行にどのように反映されたのかを政府は国民に明確に説明する義務がある。開示されていない会議の資料、議事録を公開するとともに、技術選択の過程、ベクトル型スーパーコンピュータとの連携など今後の方針が明確に説明される必要がある。
- また、スーパーコンピュータに関しては、最速の一台の能力だけでなく、国内における必要な総計算能力、地域分散の必要性、民間のニーズなどについてのデータを政府は明らかにすべきであり、「京」完成後のスーパーコンピュータの開発については、その戦略を早急に検討して公表するとともに、費用を精査することによりコストの縮減を図る必要がある。

(つづき)

- 「京」の利用に当たっては、その能力を有効に活用するため、コンソーシアム体制に依存することなく、ニーズの高い利用者が透明・公平な手続で選定されるような枠組みを構築して早期に供用を開始するとともに、純粋な科学、自然大災害予測など重要な国家的要請に基づく研究利用と、対価を得られる民生技術開発とを区別した利用のためのルールを策定すべきであり、後者については利用料金等を徴収するなどして、運用経費負担を圧縮すべきである。

ということで、大量の会議資料が公開された

公開になったもの

- 情報科学技術委員会 次世代スーパーコンピュータ概念設計評価作業部会 議事要旨・議事録・配付資料 (2007/3-6)
 - － ベクトル＋スカラーと決定した委員会
- 次世代スーパーコンピュータプロジェクト中間評価作業部会 (2009/4-7)
 - － ベクトルなくなっても大丈夫と結論した委員会

キーポイント：概念構築のためのなんとか

- 概念構築といいつつすでに枠が決まっていた
 - HPL 10PF
 - 電力 30MW 以下
 - アクセラレータ付けるならやはり 10PF
 - この範囲でアプリケーションの性能を上げること
- その結果、各社の提案が収斂した

以下、2012/6 に公開になった当時の会議資料から

(2012/6 公開)



回収資料

資料2-2

次世代スーパーコンピュータの概念設計 について 続き)

平成19年3月27日

理化学研究所
次世代スーパーコンピュータ開発実施本部

(2012/6 公開)

アーキテクチャ案の概要 (汎用システム) 【月末時点】

- 基本要件仕様 性能評価の基準とするシステム構成)
 - 理論ピーク性能 :10PFLOPS
 - 総メモリ容量 :2.5ペタバイト

アーキテクチャ案	NEC	日立	富士通	筑波大学
コア数 (コア: 1演算プロセッサ)	中並列 10万以下	高並列 10～50万	超並列 50～100万	
計算ノード数	～5万		10～15万	
高速演算機構	ベクトル		SIMD	
消費電力 (本体のみ)	20-30 MW	10-20 MW	20-30 MW	10-20 MW
設置面積	3000 m ² 以上	1000-2000 m ²	3000 m ² 以上	1000-2000 m ²

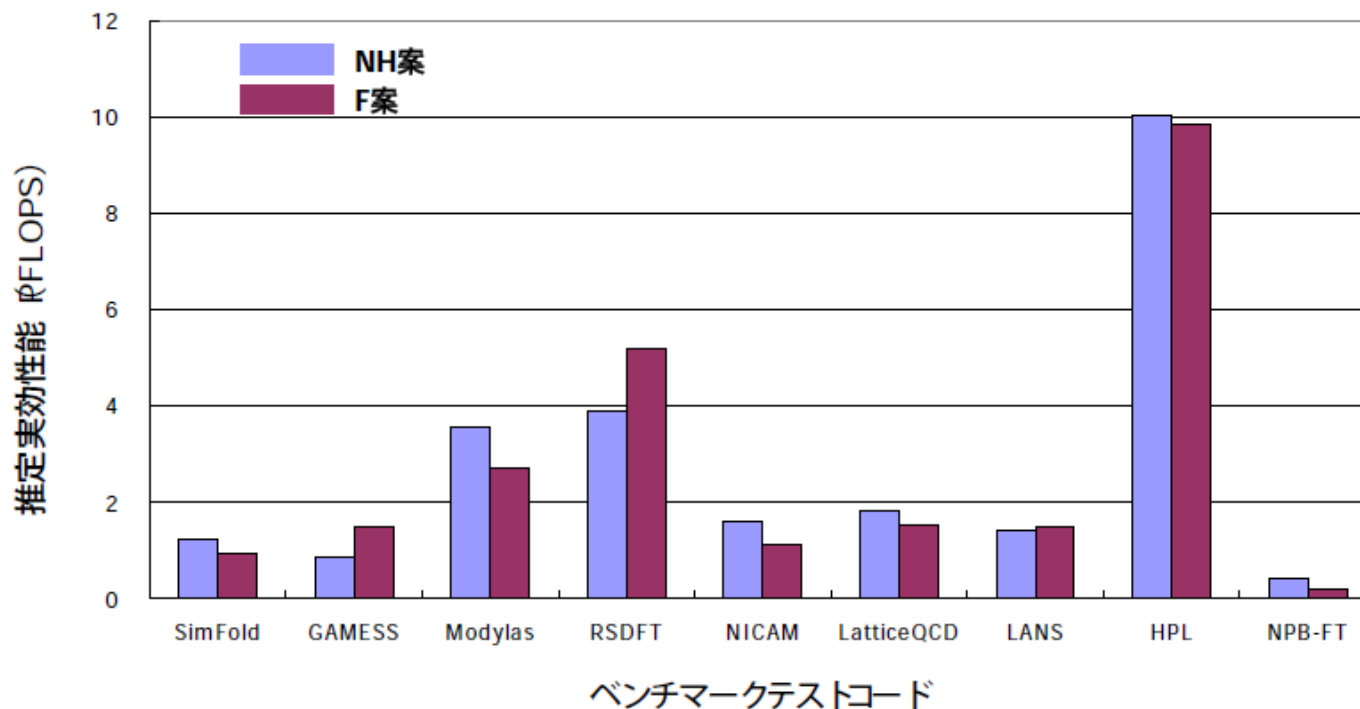
(2012/6 公開)

提案システムの演算部性能の比較

			NH案	F案
演算コア	動作周波数 (GHz)		2	
	演算性能 (GFLOPS)		64	16
	演算加速機構 演算器数)		ベクトル型 (6: 2FMA x 8VPP)	SIMD型 (4FMA)
	レジスタファイル		ベクトルレジスタ 256要素×64本	スカラーレジスタ 128本
CPUチップ 計算 ノード	演算性能 (GFLOPS)		256	128
	演算コア数		4	8
	メモリバンド幅 (Byte/Flop)		0.5	
	L2 キャッシュ	容量 (MB)	8	6
		Byte/Flop	4	2
		特殊機構	選択的登録機構	ライン・ロック機構

(2012/6 公開)

ベンチマーク・テストによる性能予測 (詳細9本)



- ターゲット・アプリケーションから7本のベンチマーク・テスト、及びHPL, NPB-FTについて、実効性能を推定。
- いずれのベンチマーク・テストもほぼ同等の性能。

NH/F の比較

- ベクトル・スカラーで違うはずだったが、アーキテクチャパラメータはほとんど同じものに収斂
- さらに消費電力等も収斂
- さらにベンチマーク性能も収斂

アクセラレータについては？

アーキテクチャ案の概要 (アクセラレータ) 【月末時点】

■ 基本仕様

- アクセラレータ部の理論ピーク性能：10PFLOPS
- 汎用サーバ (ホスト) のI/Oインターフェースに接続
- ホストより指定された演算処理をアクセラレータで実行し、結果をホストに格納

アーキテクチャ案		国立天文台	東京大学
アクセラレータ	アーキテクチャ	SIMD型 プロセッサアレイ	
	プロセッサチップ数	約 15,000	約 20,000
	ポート数	4,000	2,500
ホストサーバ数		2,000	2,500
アクセラレータ部 消費電力		-10 MW ※平成24年6月公開時の注意書き 消費電力については、10MW以下、10-20MW、20-30MWの範囲でまとめたもの。提案は、ホスト部を除いて、1.7MWであった。	-10 MW ※平成24年6月公開時の注意書き 消費電力については、10MW以下、10-20MW、20-30MWの範囲でまとめたもの。提案は、ホスト部を除いて、0.68MW (案1)、0.88MW (案2)であった。

4案の要点

	NH	F	NAOJ	UT
ピーク性能 (TF)	10.48	10.61	10	10
電力 (公開資料)	23	22.8	10	10
電力 (報告書)	23	22.8	1.7	0.88

注意書き拡大

「消費電力については、10MW 以下、10-20MW、20-30MW の範囲でまとめたもの。提案はホスト部を除いて1.7MWであった」

えええ、、、

1.7や0.88が10に？

- 公開時の説明:「消費電力については、10MW 以下、10-20MW、20-30MW の範囲でまとめたもの。提案はホスト部を除いて1.7MWであった」
- 但し、当時の議事録を見るとそんな説明はしていない。「10MW 以下」と議事録には書いてある。
- まあ、資料と説明をみたらこれは10MW くらいだろうと思ったであろう。
- 「誤解」したほうが悪いということで。

性能推定

ベンチマーク	NH	F	NAOJ
SimFold	1.3	1.1	2.5
GAMESS	1.0	1.7	2
Modylas	3.6	2.8	2.9
RSDFT	3.9	5.2	4.3
NICAM	1.8	1.4	2.5
LQCD	1.9	1.7	2

性能は全てペタフロップス単位
まあ私の見積りは楽観的かもしれないが、汎用でもアクセラ
レータでも大して変わらない。

では何故アクセラレータは不採用？

公式の説明

2者のシステム構成により、目標性能達成の見込みが確認できたため、アクセラレータの採用は考慮しない

言い換えると：

アクセラレータのほうが安くできるとか性能高いとかあるかもしれないけど汎用で目標は達成できるのでそっちは考えない

なんだそれ？

「京」まとめ

- プロジェクトとしてみると

- 数値目標が低すぎる
- アプローチが時代遅れなため低い目標も実現困難

という問題がある。

- そうなった理由は、「国産スパコンを作るプロジェクトをやる」という目標ありきで、作ったものに意味があるかどうかをきちんと検討してなかったから
- なにかもうちょっと謎な力が働いているようにも見えるが、よくわからない

国策スパコンは特殊か？という話

話を計算機ハードウェア開発に限っても、「×」なものはいくらでもある
先にあげた例:

- 科学技術用高速計算機システム
- 第五世代
- RWCP

もうちょっと違うところの例

旧科技庁系で目立つもの

- 原子力

- 高速増殖炉、新型転換炉
- 再処理工場
- というか、成功した大規模計画がない気がする

- 航空・宇宙開発

- J-1 ロケット
- JAXA(特に旧NASDA)の色々、ISAS では Lunar-A, Astro-G とか
- 航空関係で成功した大規模計画？

何故上手くいかないのか？

もう一度「京」の場合:

- 目標が適切でない
- アプローチも適切でない

まあ、そんなに細かく分析するまでもなく、この2つは大きな問題を起こしたプロジェクトに共通な特徴。

何故そうなるのか、もおそらくあまり変わらない

目標が適切でないのは何故か？

- 「京」の場合：要するに時代遅れ。30年前ならよかったかもしれない
- 原子力、宇宙、航空、、、

時代遅れな目標が設定される理由：多分2つ

- プロジェクトの時間スケールが長すぎる
- 実質上のプロジェクト立案者の知識が古い

日本におけるプロジェクト立案

- 官僚＋「専門家」の委員会、が基本的構造
- 但し、「専門家」を選ぶのは官僚
- どのような専門家を選べがどのような結論がでるか、は選ぶ側にはわかっていない

つまり、「専門家」にはあまり意味はなくて官僚が決める。但し、官僚はそんなに中身に詳しくない(ひとつの部門に 2-3 年しかいない)

このため、知識の不足に基づく判断をする。そうすると、ほぼ**自動的に時代遅れな判断**になる。

アプローチの決定

これも目標自体と同じ。知識が不足すれば時代遅れになる。

もうちょっとましな方法はあるか？

「専門家」に任せれば上手くいくか？

これはまあ、学術会議は日本に科学政策についてなにかまともな提案ができるか？というような話。各関係学会の希望を集めているようでは×なのはさっきした話と同じ。

政治主導なら？

素人では官僚と同じかもっと悪い。

What I learned from Steve Jobs

— Guy Kawasaki

1. Experts are clueless
2. Customers cannot tell you what they need
3. Jump to the next curve
4. The biggest challenges beget best work
5. Design counts
6. Changing your mind is a sign of intelligence
7. "Value" is different from "price"
8. A players hire A+ players

(いくつか省略)

エクサスケールはどうなってるか？

- 現在までのところ「京」と同様な進みかた
- 2011 年 7 月、アーキテクチャ、システムソフトウェア、アプリケーションの 3WG 発足
- 各分野の要求聞いてアーキテクチャまとめるみたいなのり
- 牧野はアプリケーション WG のメンバー
- アプリケーションを必要な「バンド幅」と「メモリ量」の 2 パラメータで見ることにした
- アーキテクチャとしてレファレンス+3 タイプを提案

1+3 アーキテクチャ

タイプ	B/F	M/F
レファレンス	0.1	0.01-0.1
オンチップメモリ	4	10^{-5}
アクセラレータ型	$10^{-3\sim 2}$	0.001-0.01
ベクター型	1	1

現在

- ほぼ、「京」の 2006-2007 年の段階
- 概ね同じように進んでいる。
- まずベクトルが脱落した (2013 夏)
- 次にアクセラレータが脱落した (公式には 2014 夏)

「京」とエクサ (「フラッグシップ 2020」) の関係
＝ 数値風洞と地球シミュレータの関係

違い:数値風洞は革新的だったが「京」はそうでない

まとめ

- 過去の日本における計算機開発には、大成功から大失敗まで色々ある
- 成功と、「成功ということになっているもの」は違う
- 成功は革新から生まれている。(もちろん、革新的な失敗は一杯あった)
- が、国家プロジェクトで革新は難しい
- なので、大抵の国家プロジェクトは上手くいかない。スパコンも同じ

おしまい