

「京」からエクサスケールへ—スーパー コンピューター開発と国家プロジェクト

牧野淳一郎

東京工業大学理工学研究科

理学研究流動機構

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

ソフトウェアシンポジウム 2012/6/14

話の構成

- 自己紹介
- 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、エクサは？

というつもりだったのですが

というつもりだったのですが
—昨日、5年前と3年前の「京」関係
会議資料が一挙公開になったので

というつもりだったのですが
—昨日、5年前と3年前の「京」関係
会議資料が一挙公開になったので
だいぶ話を変えます

公開になった理由

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。

今日の話の構成

- 自己紹介
- 「京」とはなんだったか？
 - 「京」の概要
 - 開発の経緯 — 昨日までの話
 - 昨日公開された資料から
- 日本のスパコンと国家プロジェクト
- 計算機と半導体技術
- エクサは何を目指すのか

自己紹介

- 職歴

- 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)
- 2011/4- 東工大理学研究流動機構 教授
- 2006/6- 国立天文台理論研究部 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
- 1994/4- 東京大学教養学部 助教授
- 1990/4- 東京大学教養学部 助手

天文 (天体物理) 学者？

- 学歴

- 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
- 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
- 1985/3 東京大学教養学部基礎科学科第二卒業

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
 - － 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

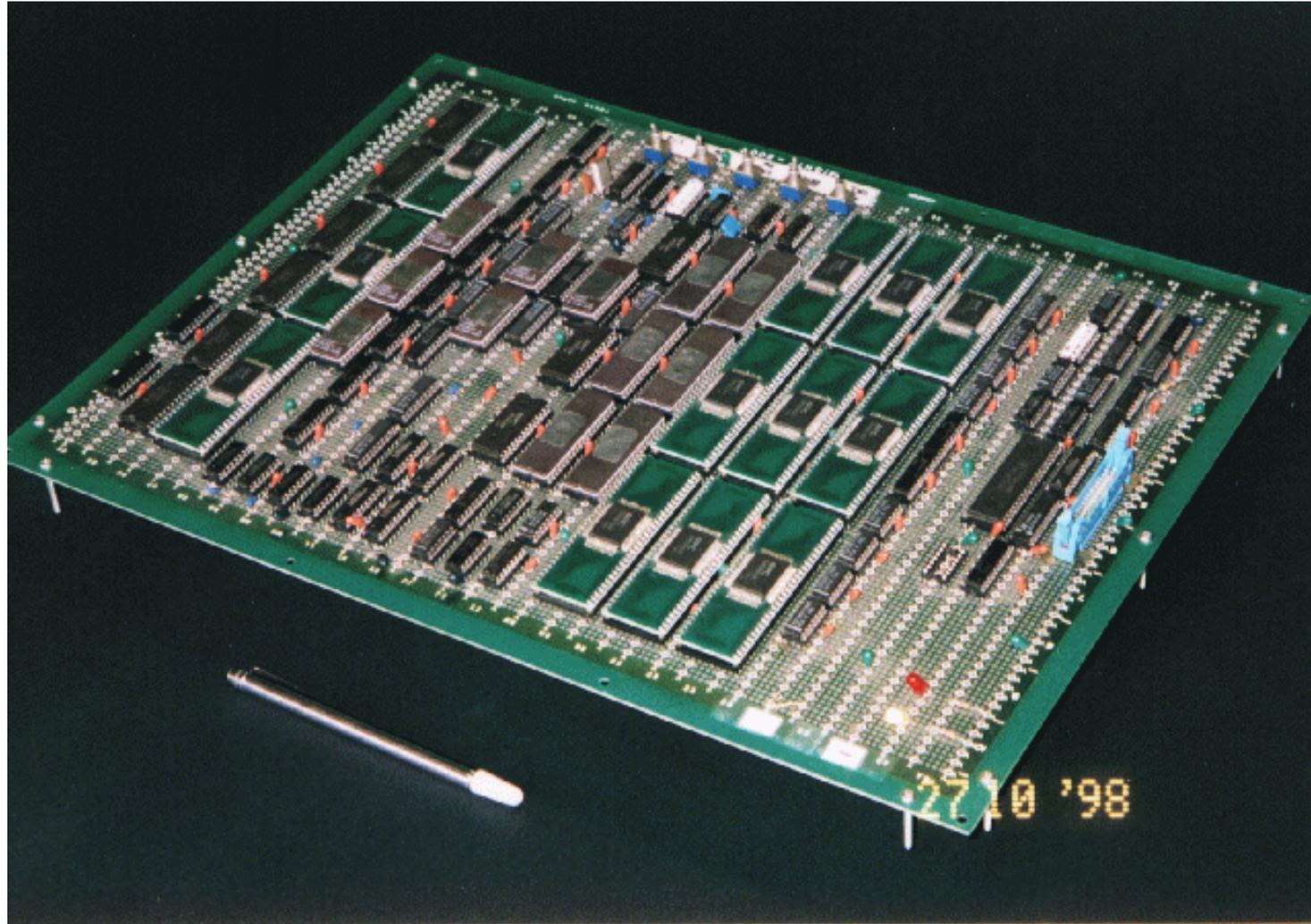
あんまり普通の天文学者というわけでもない。

計算機開発

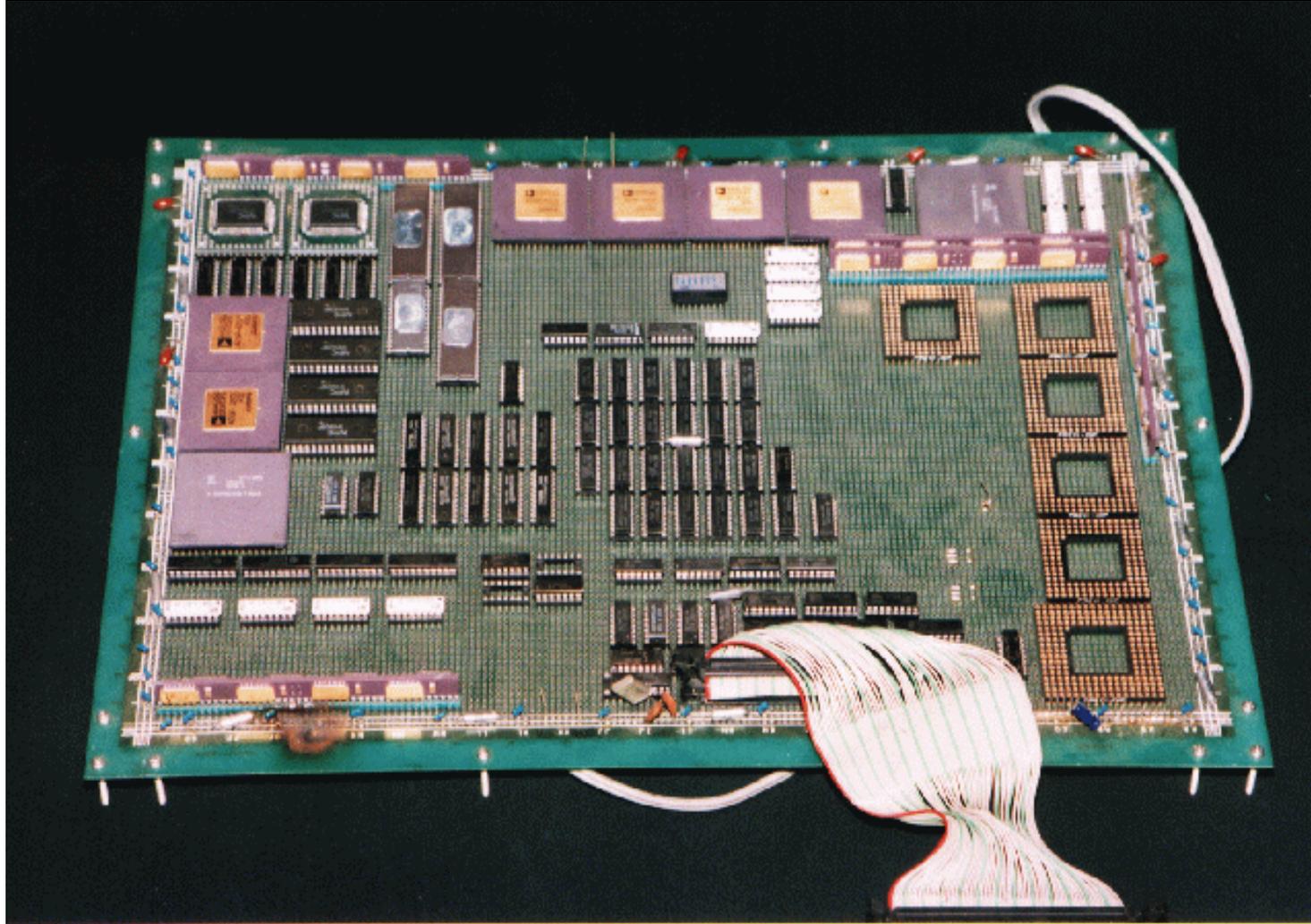
20年くらいだらだらやっている。

- 1989- GRAPE (Gravity PipE):
重力多体問題専用計算機
- 2004- GRAPE-DR (Greatly Reduced Array of Processor Element with Data Reduction):
「汎用」アクセラレータ

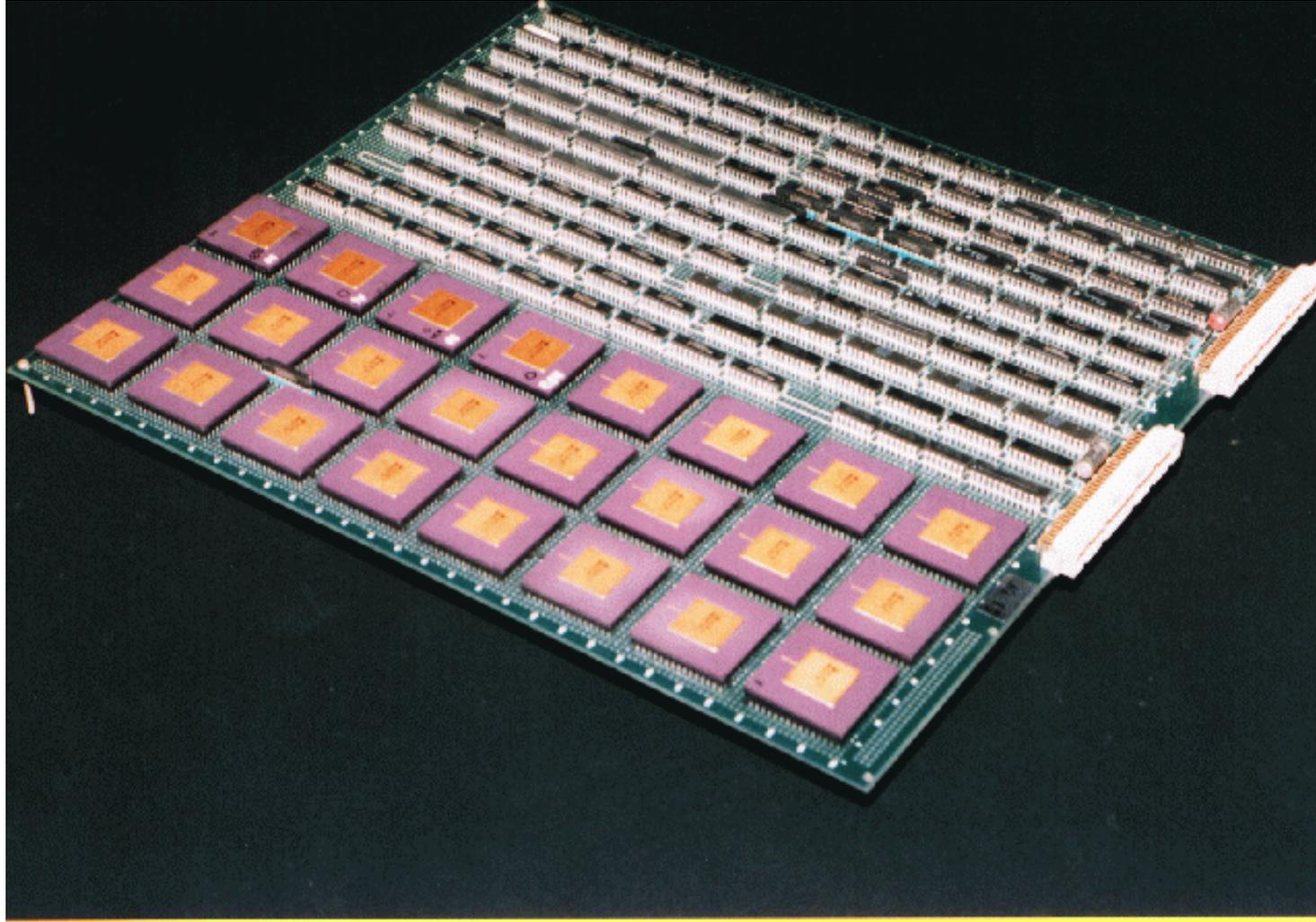
GRAPE-1(1989)



GRAPE-2(1990)



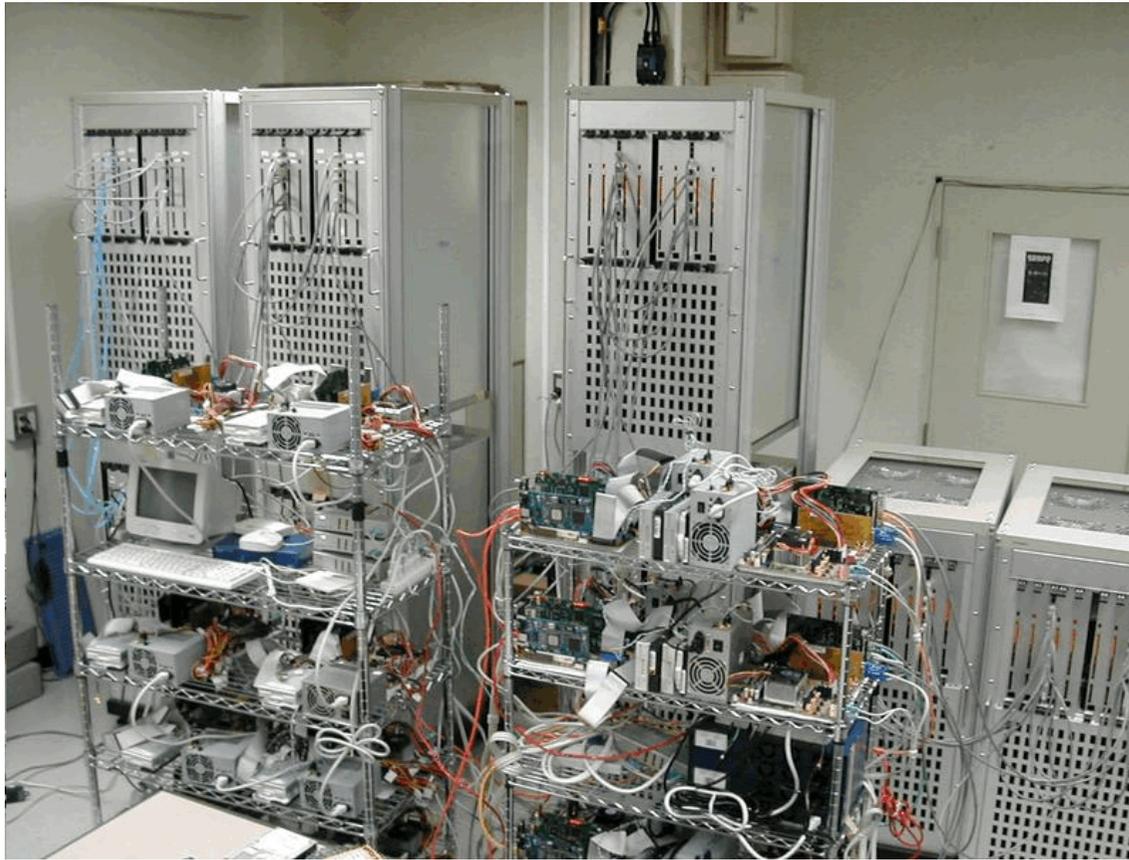
GRAPE-3(1991)



GRAPE-4(1995)



GRAPE-6 system



2002年の 64 Tflops システム

4 ブロック

16 ホスト

64 プロセッサボード

自己紹介(続き2)

昨年度

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

「京」の概要

- 2006-2012 の7年間で開発。開発費 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

開発の経緯

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
- 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)

時系列続き

- 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
- 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。

「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

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」
2. 結局、「何のため」かははっきりしないままスタート
3. 計画スタートまでに目標性能が二転三転した。
4. アーキテクチャも二転三転した。
5. 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

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
- 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(但しパイロットラインで量産用ではない)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。

- 根本問題は、「暗黙の大前提」
- 見かけ上は、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

もうちょっと詳しく アーキテクチャについて

—昨日公開された資料からいくつか

アーキテクチャ案の概要 (汎用システム) 【月末時点】

- 基本仕様 (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

アーキテクチャ案の概要 (アクセラレータ) 【月末時点】

■ 基本要件仕様

- 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		<div style="border: 1px dashed blue; padding: 2px; display: inline-block;">-10 MW</div> <small>※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。</small>	<div style="border: 1px dashed blue; padding: 2px; display: inline-block;">-10 MW</div> <small>※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。</small>

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

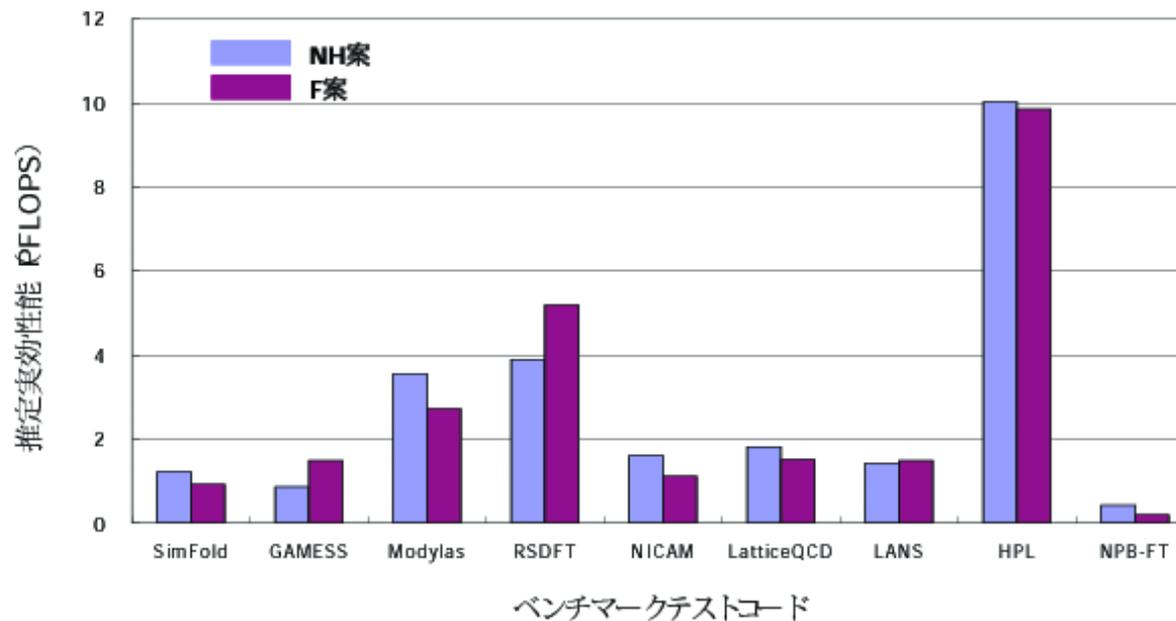
「消費電力については、10MW以下、10-20MW、20-30MWの範囲でまとめたもの。提案はホスト部を除いて1.7MWであった」

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。

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.

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
RSDFE	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った。
- 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
- という以上にもっと不透明かつ不可思議なことが起こっていた模様

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

- 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がない気がする

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る

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

根本的な問題

- スパコン開発が計算機産業、半導体産業に何か貢献するという幻想から抜け出す必要がある
- 但し、この幻想に基づいた官学複合体はできている(産は足を抜きたがってる気も?)
- 元は幻想でもシステムは実体
- 半導体産業、計算機産業に貢献しないならそもそも何のために? というのをちゃんと考え直す必要あり

まとめ

- 「京」の開発プロセス、できたものは色々おかしい
- その根本的な理由は、「国家によるスパコン開発が産業発展に寄与する」という幻想と、それが実体化した官学複合体
- エクサをやるなら幻想から一度脱却するべき

おしまい