グリーンスパコンを作る

牧野淳一郎

国立天文台天文シミュレーションプロジェクト

グリーンスパコンを作る

牧野淳一郎

国立天文台天文シミュレーションプロジェクト (3/31 まで)

東工大理工研究科理学研究流動機構 (4/1 から)

話の概要

- 何故「グリーンスパコン？」
- ENIAC から「京」まで
- 地球シミュレータと GRAPE-6
- GRAPE-DR
- 今後

何故「グリーンスパコン」？

もちろん、

スパコンがどんどんグリーンでなくなってきたから

どれくらいグリーンでないか？

某社某氏との会話

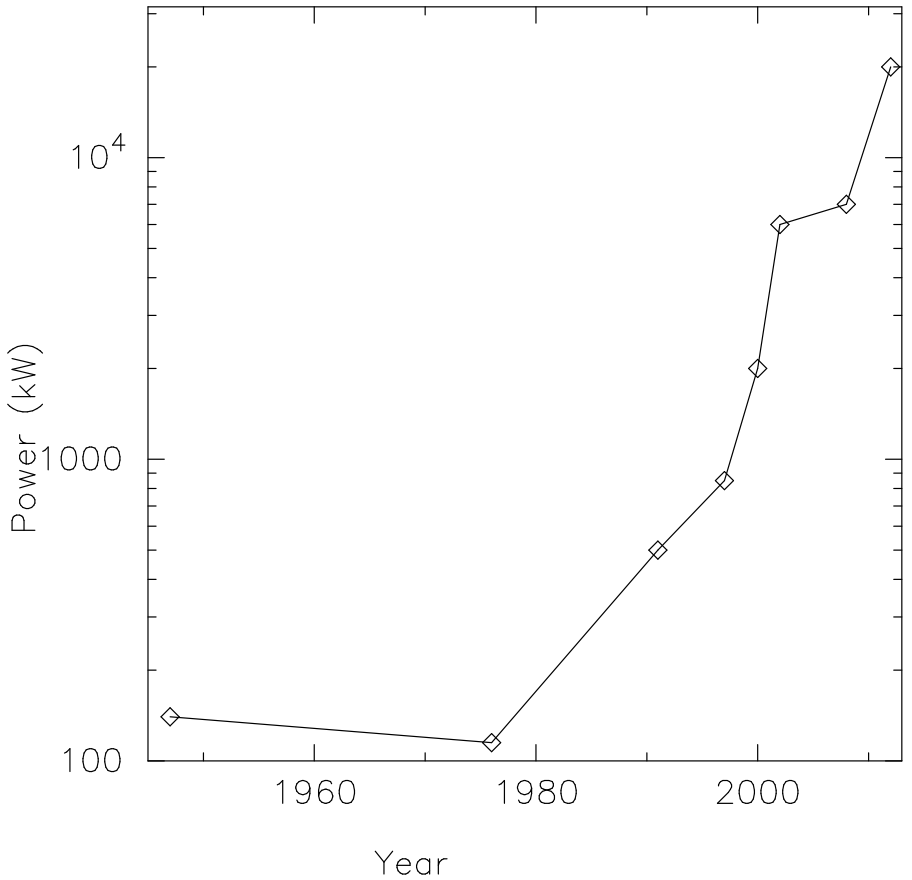
「天文台の次期システムはどうですか」
私「電力がですねえ、、、今 140 で、、、今のレンタル料でそちらのシステム入れると 600kW くらいにはなりますよね？」
「良い線ですね、空調もいれると×××」

国立天文台三鷹キャンパスの総電力量は 1.5MW しかない、、、

ENIAC から「京」まで

ENIAC	1947	140kW
Cray-1	1976	115kW
Cray C90	1991	500kW
ASCI Red	1997	850kW
ASCI White	2000	2MW
ES	2002	6MW
ORNL XT5	2008	7MW
「京」	2012	20MW

グラフにしてみると、、、



ENIAC から Cray-1 まであまり変わらない

そのあと20年間で10倍

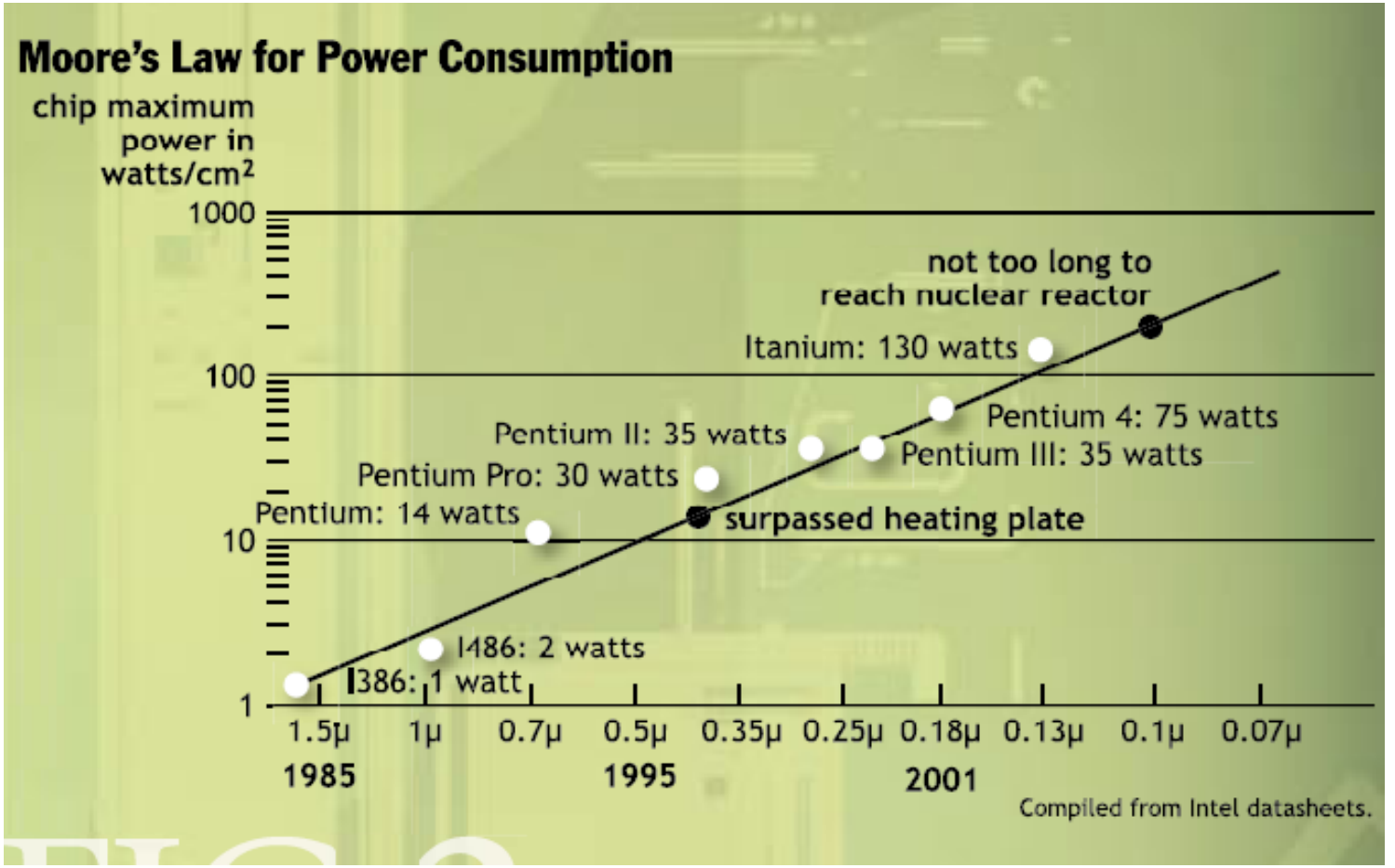
そのあと10年間でさらに10倍

何故こんなことに？

理由

- 計算機に使うお金が増えている。ASCI Red は $ 50M、「京」は、、、、
- プロセッサの面積当り消費電力が増えた
- プロセッサの面積当りの値段が下がった

面積当りの消費電力



Feng 2003 から、 実は 2003 年以降は増えてないけど、、、

$100\mathrm{W}/\mathrm{cm}^2$ の意味

- 普通にパッケージにいれて強制空冷で (まあ水冷でも同じ) 冷やせる限界
- クロックの上限を決める。この 10 年間 CPU のクロックは殆ど上がってない

古典的 CMOS スケーリングとの関係

古典的スケーリング
フィーチャーサイズ $1/2$ → 面積当りキャパシタンス 2 倍、
電圧 $1/2$、クロック 2 倍 → 消費電力不変

現実

- **電圧下げるのは限界**
- **リーク分の増大**

消費電力一定 → クロックほぼ一定

面積当りの値段

というか、プロセッサチップ当りのシステムコスト

- x86 とその周辺チップは量産効果で安い。
- 2000 年以降、 x86 とそれ以外で性能が逆転。

値段が下がった分、沢山買える → 電力増える

で、これは問題か？

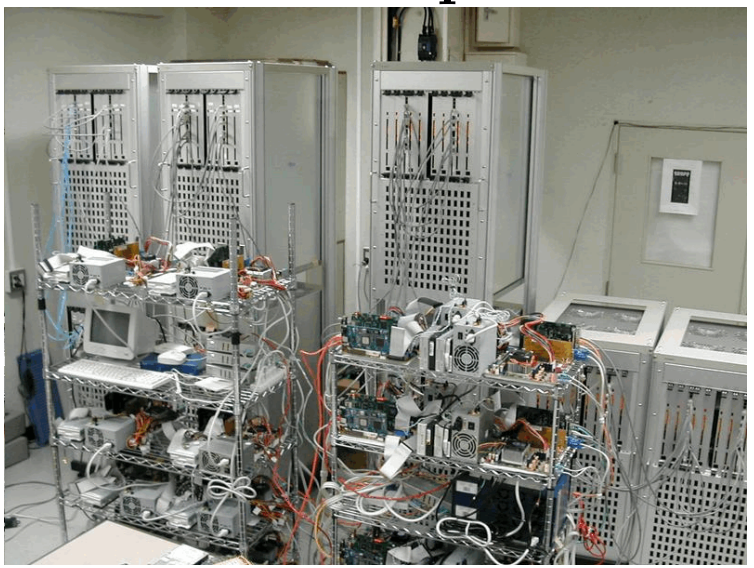
- 「そういうものだ」と思えば別に問題ではない？
- 10 年で 100 倍とかのペースで計算速度をあげたいと思うと問題
  2020 年くらいにエクサフロップスになっても、
  200MW 必要では困る (電気代のほうが高い)

性能を落とさずに電力を減らす？どうやって？

地球シミュレータと GRAPE-6

GRAPE-6 2002 250nm

50kW 64 Tflops



ES 2002 150nm

6MW 40 Tflops

電力当り性能は 100 倍以上違う

何が違うか?

プロセッサチップの比較

	地球シミュレータ	GRAPE-6
テクノロジ	150nm	250nm
面積	400mmsq	200mmsq
トランジスタ数	6000万	800万
クロック	500MHz	90MHz
消費電力	140W	15W
演算性能	8Gflops	30Gflops
演算器の数	16	～ 300
Gflops/W	0.06	2

要するに

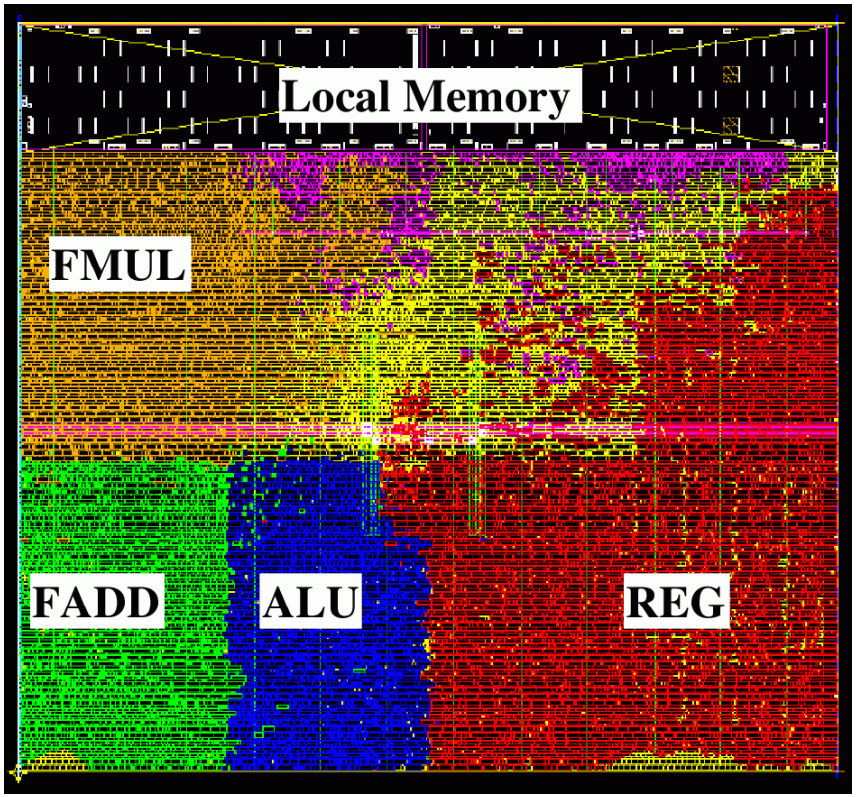
演算器当りのトランジスタ数が違う

GRAPE-3	4千
GRAPE-6	3万
GRAPE-DR	40万
Cray-1	40万
Intel 80860	60万
地球シミュレータ	400万
Fermi	300万
Sandy Bridge	4000万

Cray-1 の 40 万はベクトルレジスタ含んでなさそう

SB は大半のトランジスタがキャッシュメモリなのでちょっと不当な数えかた？でもキャッシュも電気は食っている。

GRAPE-DR PE のフロアプラン



0.7mm by 0.7mm

浮動小数点演算器の部分は
チップ面積の 1/5 以下
(1/3 以上くらいにはしたかった、、、)

比較からわかること

- 専用パイプラインで計算精度まで切り詰めれば演算器当り 1 万トランジスタ以下にできる
- 汎用プロセッサでは 30 万トランジスタくらいが限界
- Cray-1, Intel 80860, GRAPE-DR はその辺
- GPU はその 10 倍
- x86 はさらにその 10 倍

実際の数字は？

Little Green 500, June 2010

Green500 Rank	MFLOPS/W	Site*	Computer*	Total Power (kW)
1	815.43	National Astronomical Observatory of Japan	GRAPE-DR accelerator Cluster, Infiniband	28.67
2	773.38	Forschungszentrum Juelich (FZJ)	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	57.54
2	773.38	Universitaet Regensburg	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	57.54
2	773.38	Universitaet Wuppertal	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	57.54
5	536.24	Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw	BladeCenter QS22 Cluster, PowerXCell 8i 4.0 Ghz, Infiniband	34.63

#1: GRAPE-DR, #2: QPACE: German QCD machine

#9: NVIDIA Fermi

Green 500, Nov(Dec) 2010

Green500 Rank	MFLOPS/W	Site*	Computer*	Total Power (kW)
1	1684.20	IBM Thomas J. Watson Research Center	NNSA/SC Blue Gene/Q Prototype	38.80
2+	1448.03	National Astronomical Observatory of Japan	GRAPE-DR accelerator Cluster, Infiniband	24.59
2	958.35	GSIC Center, Tokyo Institute of Technology	HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows	1243.80
3	933.06	NCSA	Hybrid Cluster Core i3 2.93Ghz Dual Core, NVIDIA C2050, Infiniband	36.00
4	828.67	RIKEN Advanced Institute for Computational Science	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect	57.96
5	773.38	Universitaet Wuppertal	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	57.54

#1 BG/Q #2+: GRAPE-DR,

#2,3: NVIDIA Fermi, #4: K-computer

トランジスタ数を切り詰めることの効果

- 半導体技術では 3 世代遅れの GRAPE-DR でも Little Green500 でトップになれたりする。
- チップ単体の性能はもっと高い。 4Gflops/W くらい。
- ホストの電力が下がればこれに近くなる。来年くらい。

トランジスタ数を切り詰めることの効果2

- 少ないマンパワーで設計できる。
- そもそも設計しないといけないものが少ないため
- 開発コストも小さくなる。まあ、それでも 10 億くらい。

トランジスタ数を切り詰めることの問題点

- GRAPE-DR の場合、 メモリバンド幅を犠牲にしている。
- GRAPE-DR の場合、各 PE が外付メモリをランダムアクセスとかはできない。
- 但し、これは対象にしたアプリケーションがあんまりメモリバンド幅いらないものだったから。1 桁くらいなら増やせなくもない。

まとめ

- 演算当りのトランジスタ数を減らせればより「グリーン」にできる
- Cray-1 はとっても「グリーン」だった。GRAPE-DR と同じくらい。
- 設計を単純化すれば限られたマンパワーでも開発できる。
- メモリバンド幅を減らせばできることは制限される。但し、バンド幅は電力消費にそれほど本質的ではない。