GRAPE-DR:
2-Pflops massively-parallel computer with 512-core, 512-Gflops processor chips for scientific computing

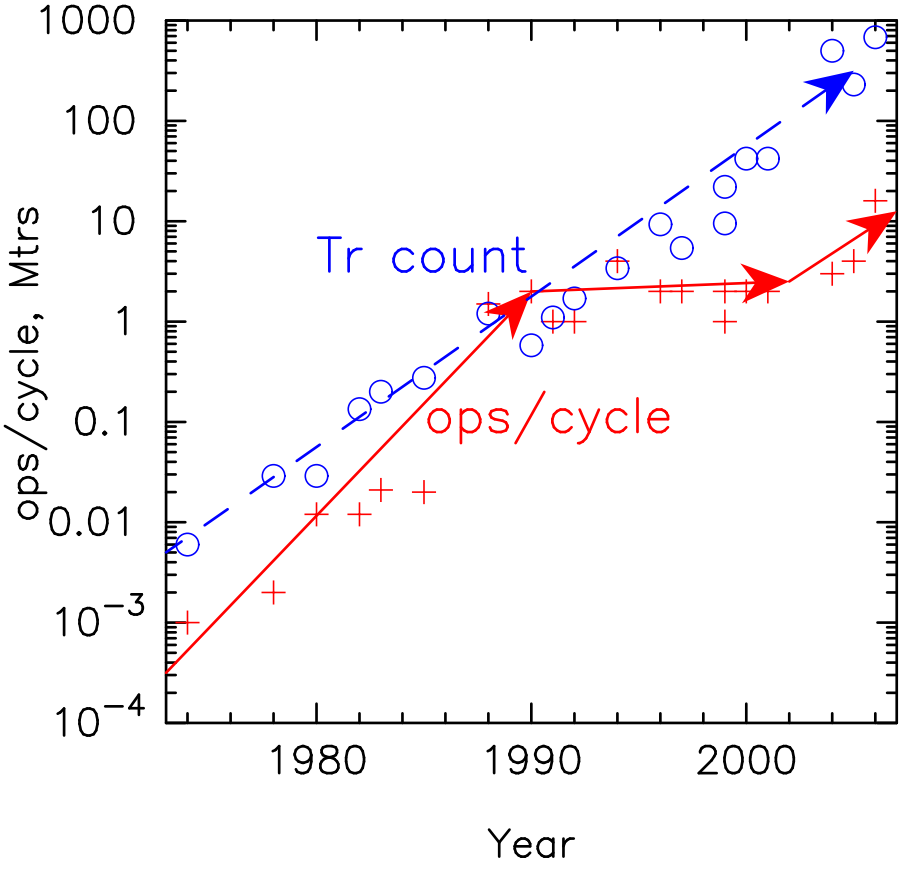
Jun Makino
Center for Computational Astrophysics
National Astronomical Observatory of Japan
and
Kei Hiraki and Mary Inaba
University of Tokyo

Talk structure

- Problem with microprocessors
- Possible alternatives
  - FPGA, GPGPU, Specialized hardware
  - One-chip massive SIMD architecture
- GRAPE-DR concept
- GRAPE-DR project status
- Comparison with other approaches
- Summary

Problem with microprocessors

Evolution of Microprocessors



- **Transistors: 1000 times in last 15 years**
- **FPUs: 8 times more in the same period**
- **a factor of 100 "lost"**

Why FPUs have not increased?

- Memory Wall
  - The bandwidth of off-chip main memory limits the performance. Added transistors need to be used for cache memory and other logic
- Amdahl's Law
  - Parallel speedup is limited by non-parallel part of the program
  - Certainly true for many business applications on Windows PC.

Is Memory Wall really a problem?

No, for many important applications.

- **Any application which requires dense matrix computation**
- **Many particle-based simulations (astronomy, molecular dynamics, gridless hydro etc)**
- **Quantum Chemistry application with local basis or $O(N)$ method**

Yes, for some other applications.

- **Anything which requires large-scale FFT**
- **CFD with explicit time integration**

Approaches which ignore MW

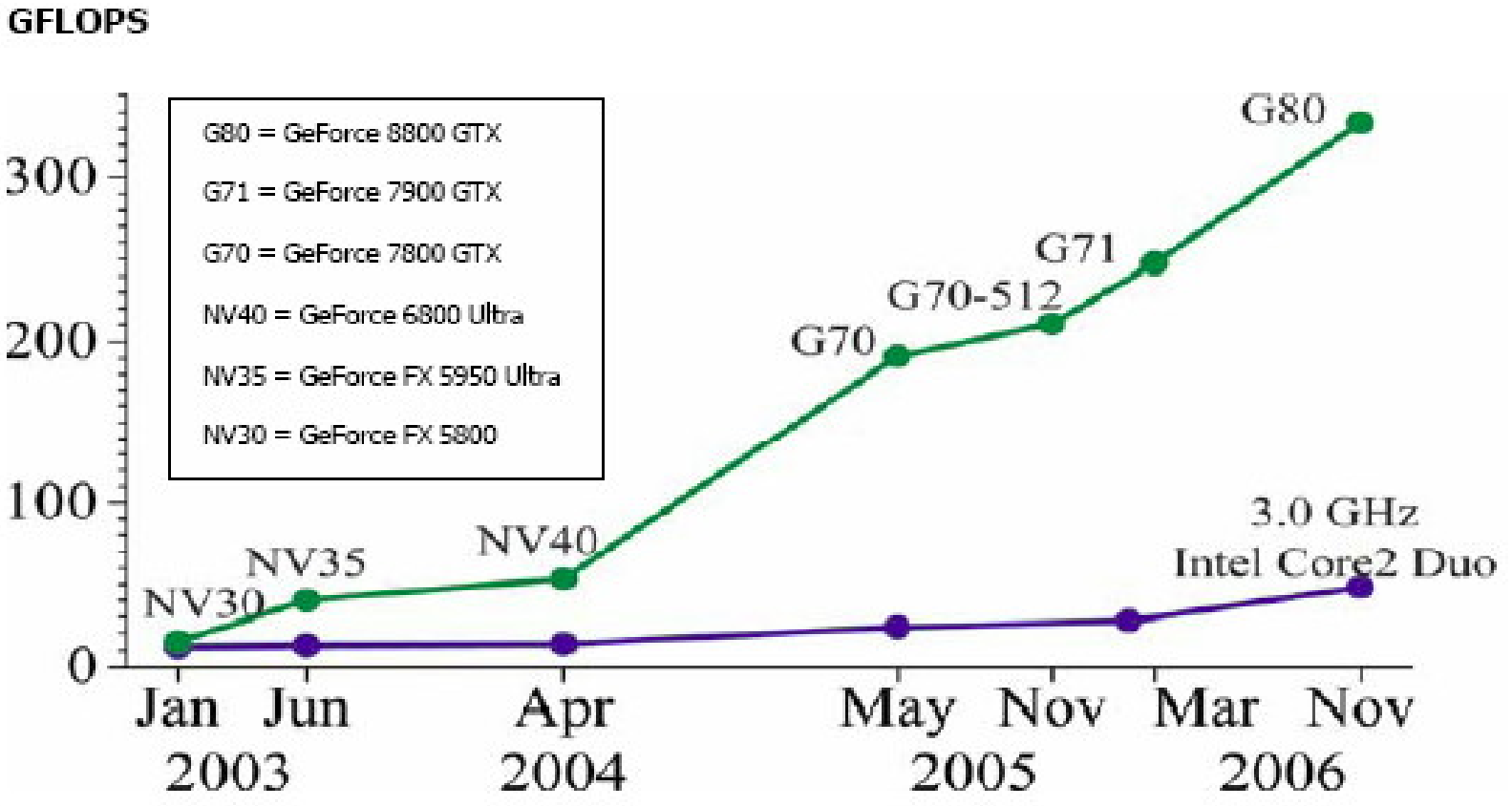
- Reconfigurable computing
- GPGPU
- Special-purpose systems
- SIMD chips

Reconfigurable computing

- Commercial FPGAs cannot do FP operations faster than microprocessors.
  - Has been so in last decade
  - Will remain so for foreseeable future
- Custom reconfigurable processors cannot compete with commercial FPGAs in price-performance ratio.
  - Longer development cycle
  - Far smaller quantity

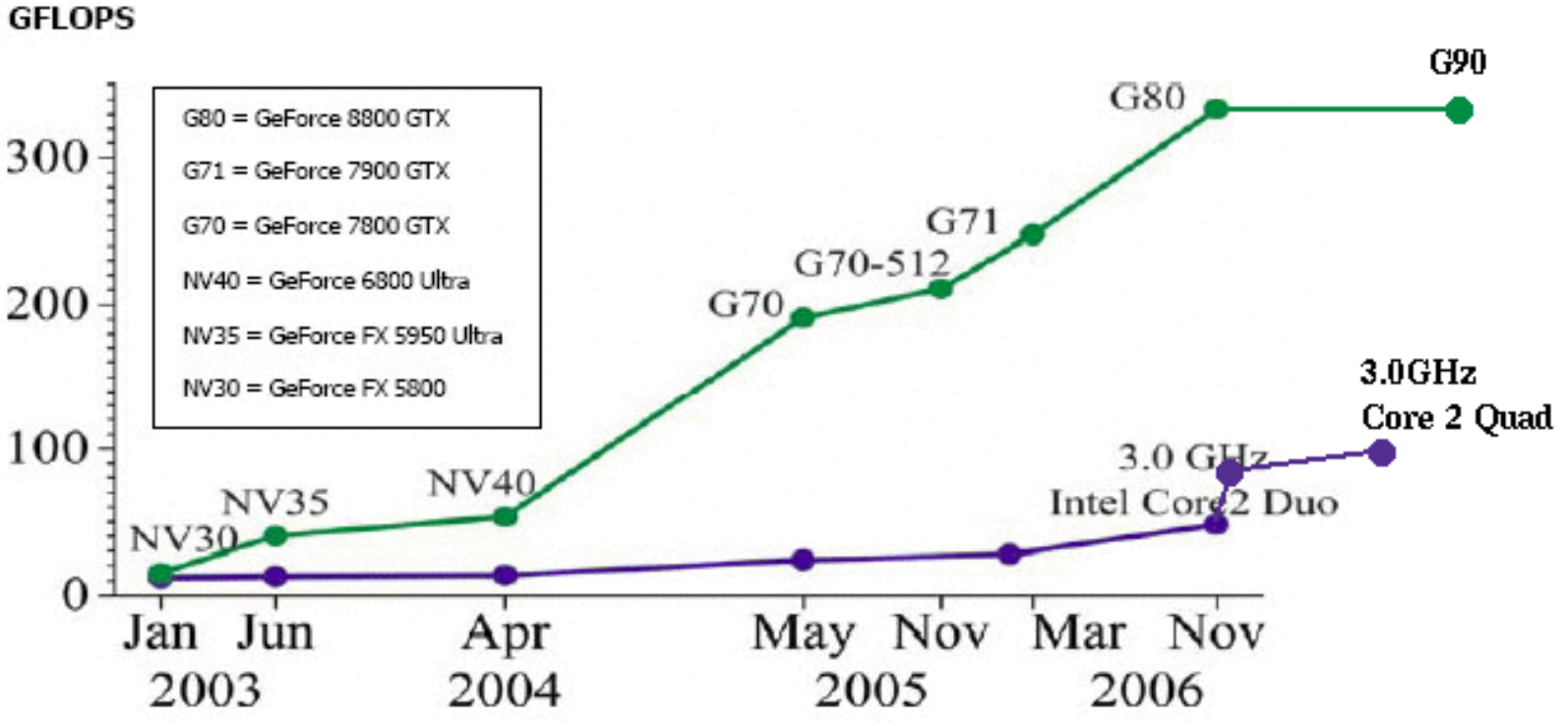
Good for applications with short-wordlength

GPGPUs —What manufacturers show:



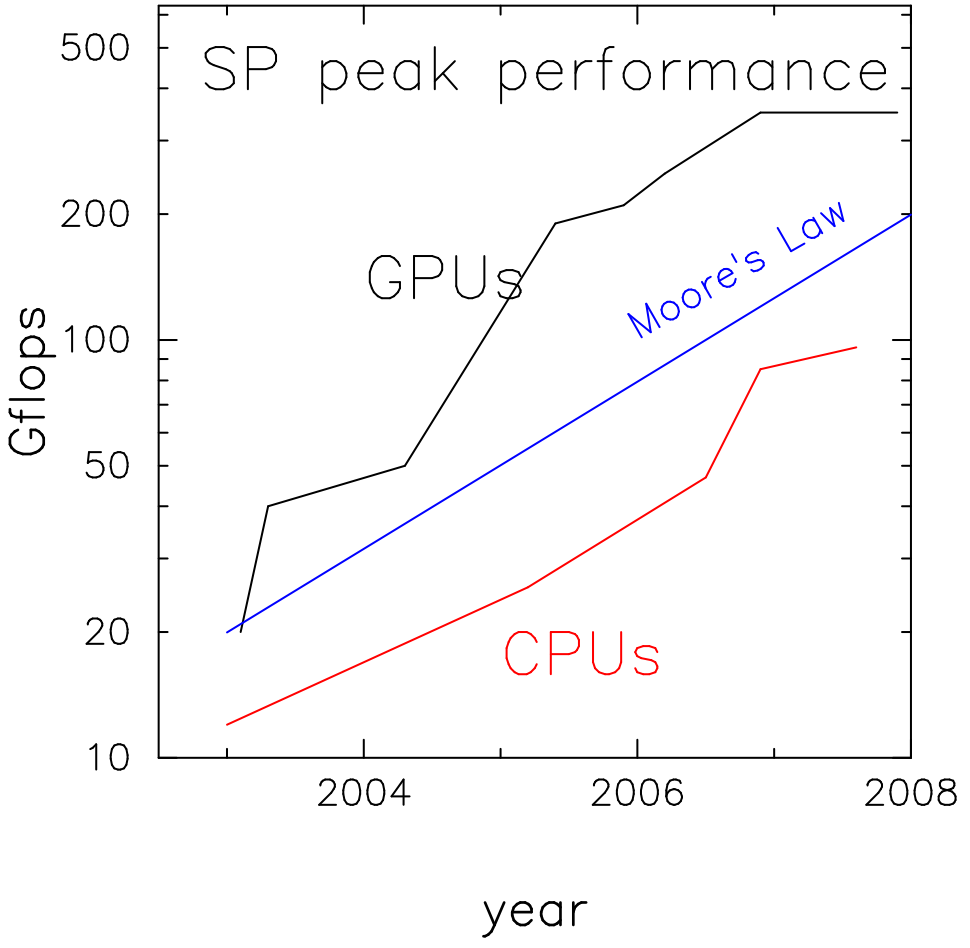
"GPUs beat Moore's Law!"

GPGPUs —Today



Hmm...

GPGPUs —Same data in log plot

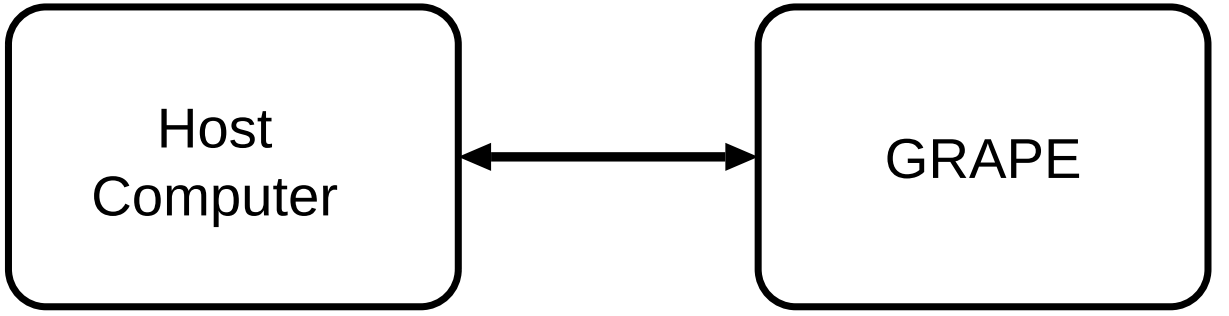


- **Faster-than-Moore period ended in 2005**
- **Microprocessors are catching up**
- **DP performance?**
- **Design limit with memory bandwidth**

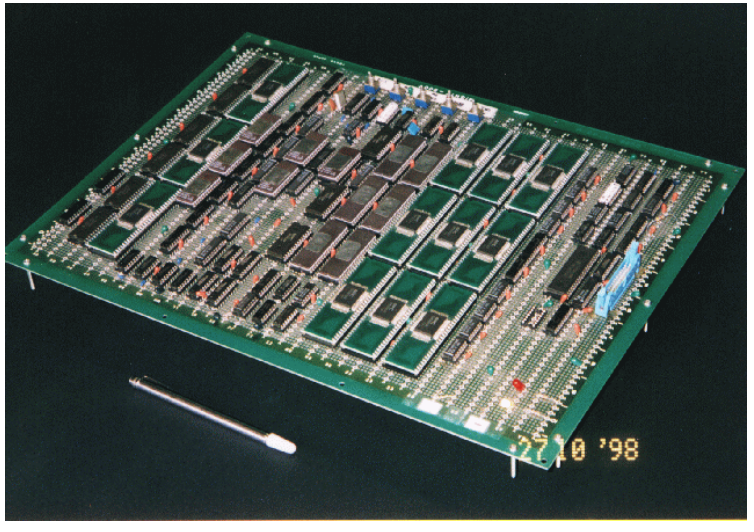
Special-purpose systems

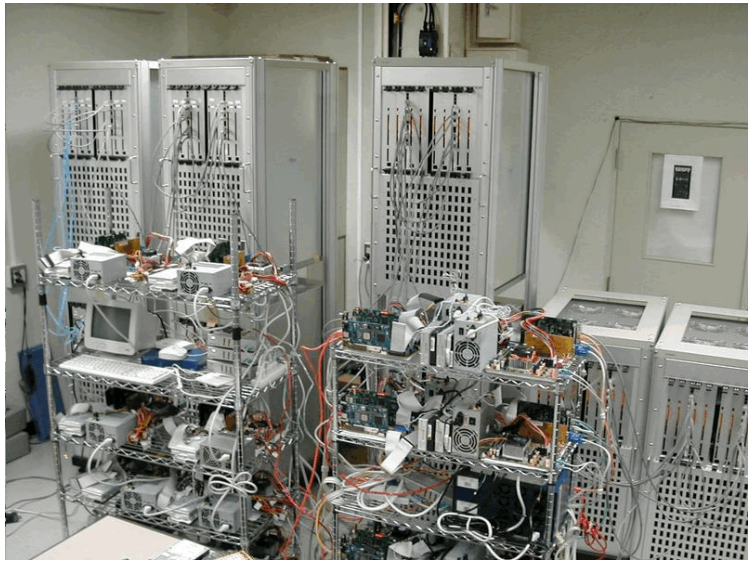
GRAPE and MDGRAPE

- **Specialized hardware for the calculation of interaction between particles**
- **Other operations are done in general-purpose PCs**



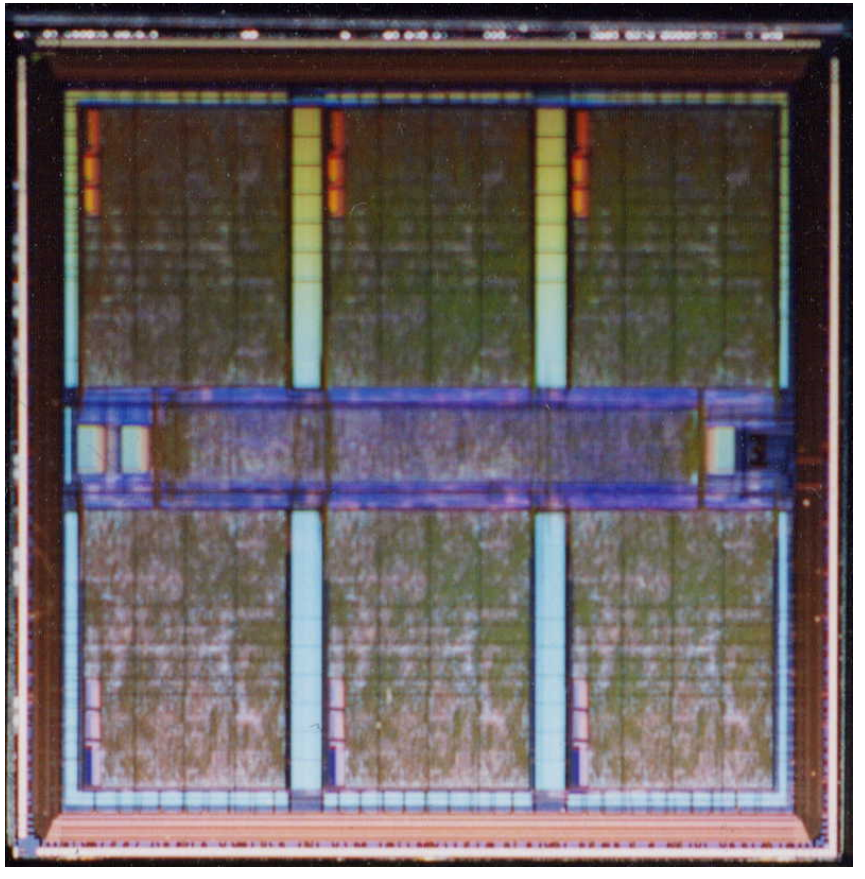
GRAPE-1 to GRAPE-6





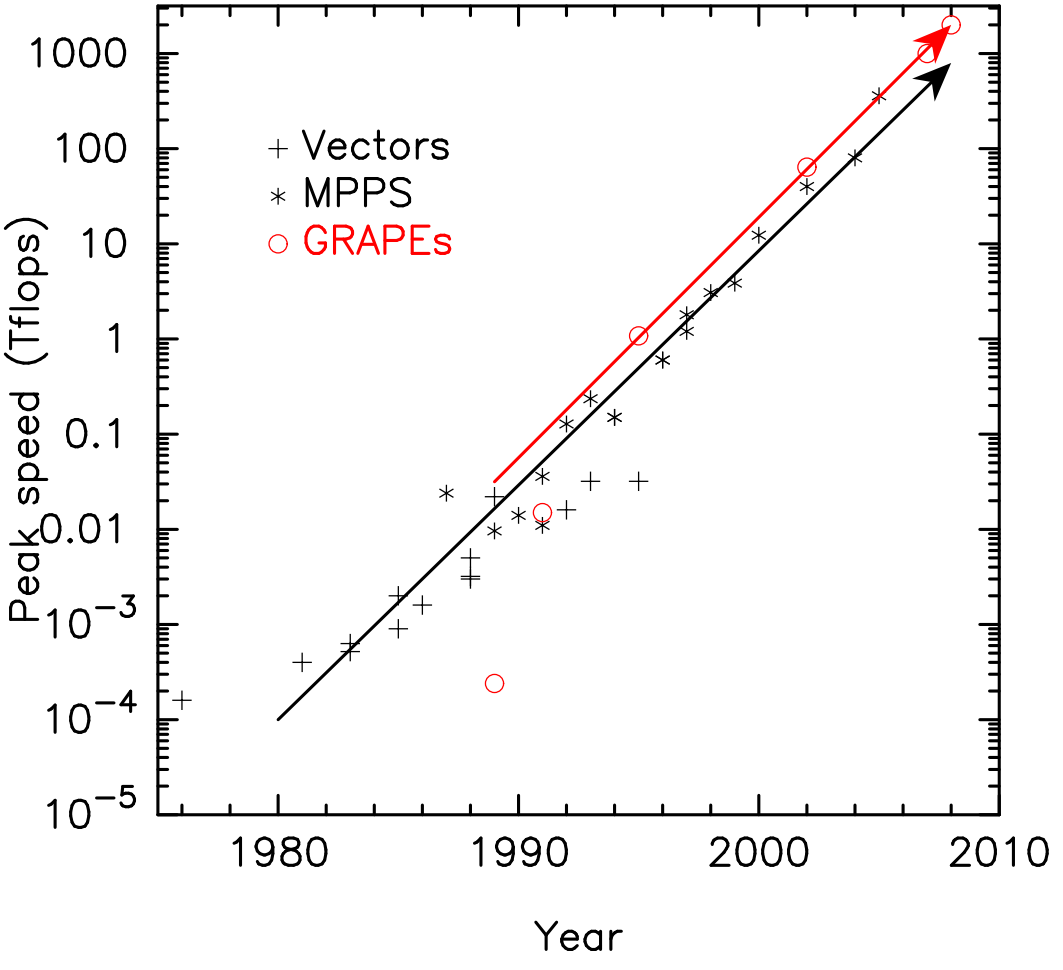
GRAPE-1: 1989, 308Mflops
GRAPE-4: 1995, 1.08Tflops
GRAPE-6: 2002, 64Tflops

Processor LSI



- 0.25 μm design rule (Toshiba TC-240, 1.8M gates)
- 90 MHz Clock
- 6 pipeline processors
- 32.4 Gflops / chip

Performance history



Since 1995 (GRAPE-4), GRAPE has been faster than general-purpose computers.

Development cost was around 1/100.

Comparison with a recent Intel processor

	GRAPE-6	Intel Xeon 5365
Year	1999	2006
Design rule	250nm	65nm
Clock	90MHz	3GHz
Peak speed	32.4Gflops	48Gflops
Power	10W	120 W
Perf/W	3.24Gflops	0.4 Gflops

"Problem" with GRAPE approach

- Chip development cost becomes too high.

Year	Machine	Chip initial cost	process
1992	GRAPE-4	200K$	1μm
1997	GRAPE-6	1M$	250nm
2004	GRAPE-DR	4M$	90nm
2008?	GDR2?	$\sim$ 10M$	65nm?

Initial cost should be 1/4 or less of the total budget.
How we can continue?

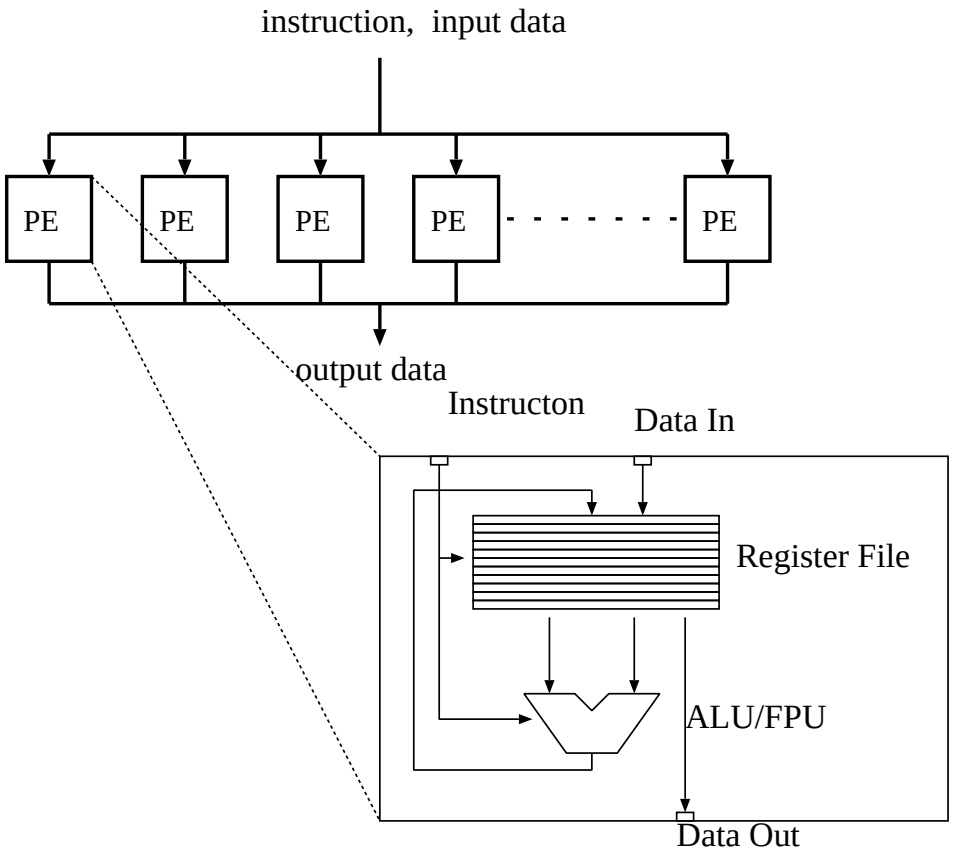
In short...

- **FPGAs: not enough FPUs because of reconfigurability**
- **GPUs: not enough FPUs because of the application requirements**
- **Special-purpose: We can't pay the initial cost**

We need something with

- **Very large number of FPUs**
- **reasonable amount of programmability**

SIMD processor without local memory

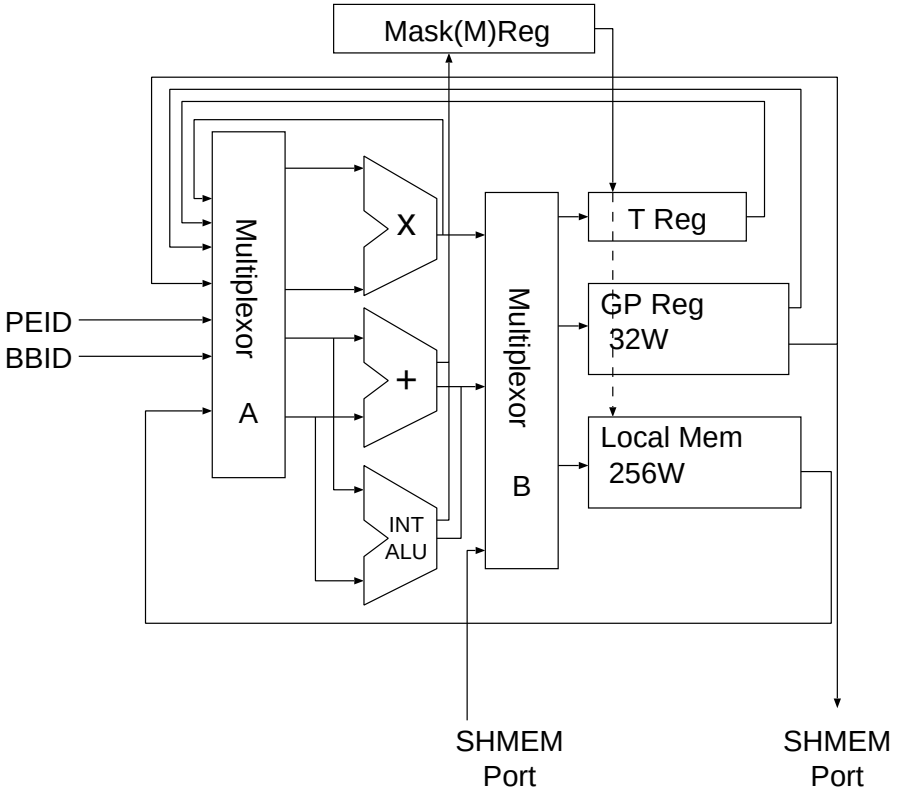


- Without external memory
- Without communication network
- Can do almost everything GRAPE or FPGA can do
- Can have much larger number of FPUs than FPGA or GPU

GRAPE-DR

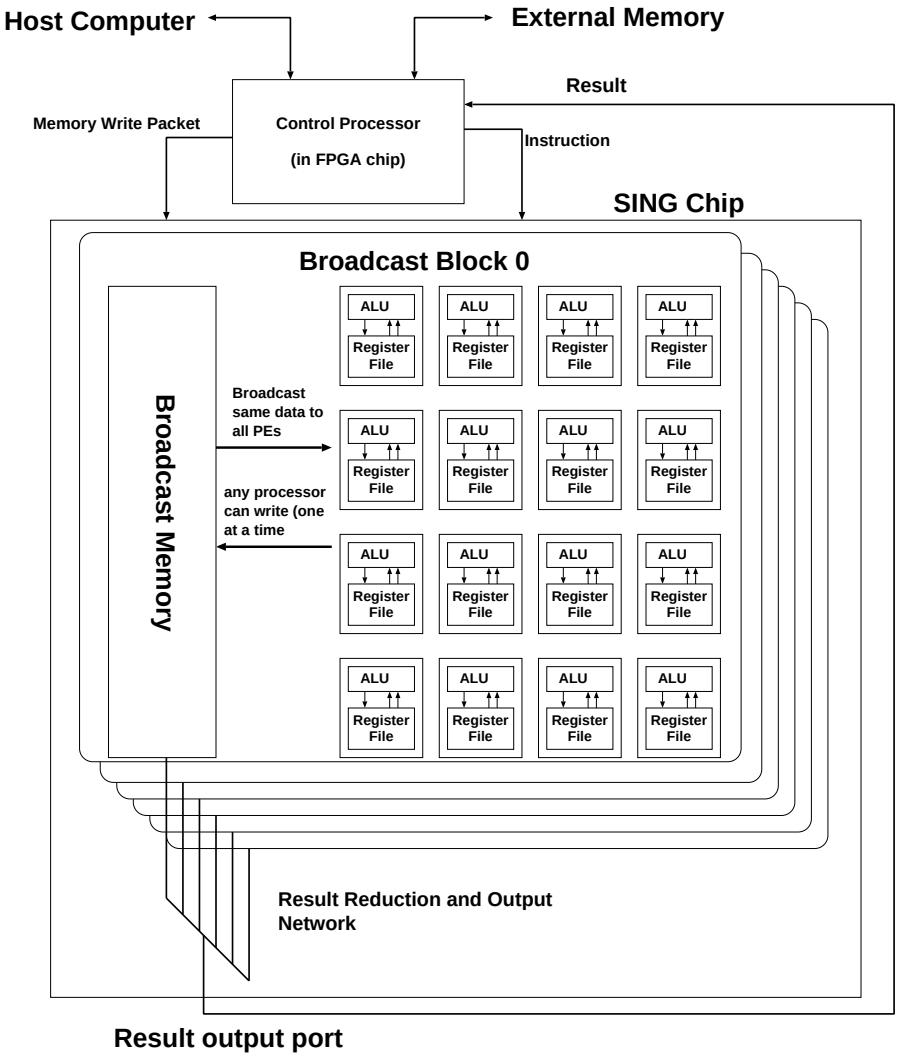
- Planned peak speed: 2 Pflops
- New architecture — wider application range than previous GRAPEs
- No force pipeline. SIMD programmable processor
- Planned completion year: FY 2008 (early 2009)
- "Greatly Reduced Array of Processor Elements with Data Reduction"

Processor architecture



- Float Mult
- Float add/sub
- Integer ALU
- 32-word registers
- 256-word memory
- communication port

Chip structure



Collection of small processors.

512 processors on one chip
500MHz clock

Peak speed of one chip: 0.5 Tflops (20 times faster than GRAPE-6).

Comparison with FPGA

- much better silicon usage (ALUs in custom circuit, no programmable switching network)
- (possibly) higher clock speed (no programmable switching network on chip)
- easier to program (no VHDL necessary; assembly language and compiler instead)

Comparison with GPGPU

Pros:

- **Significantly better silicon usage (512PEs with 90nm)**
- **Designed for scientific applications reduction, small communication overhead, etc**

Cons:

- **Higher cost per silicon area... (small production quantity)**
- **Longer product cycle... 5 years vs 1 year**

Good implementations of N-body code on GPGPU are coming (Hamada, Nitadori, Portegies Zwart, Harris, ...)

Comparison with GPGPU(2)

	GRAPE-DR	nV G92	AMD FS9170
Design rule	90	65	55
Clock(GHz)	0.5	1.5	0.8
# FPUs	512	112	320
SP peak(GF)	512	336	512
DP peak(GF)	256	—	?
Power(W)	65	70?	150?

How do you use it?

- Particle simulations: The necessary software is now ready. Essentially the same as GRAPE-6.
- Matrix etc ... Libraries will be provided
- New applications:
  - Primitive Compiler available
  - For high performance, you need to write the kernel code in assembly language (for now)

Primitive compiler (NAOJ/RIKEN)

(Nakasato 2006)

```
/VARI  xi, yi, zi, e2;
/VARJ  xj, yj, zj, mj;
/VARF  fx, fy, fz;
dx = xi - xj;
dy = yi - yj;
dz = zi - zj;
r2 = dx*dx + dy*dy + dz*dz + e2;
r3i= powm32(r2);
ff = mj*r3i;
fx += ff*dx;
fy += ff*dy;
fz += ff*dz;
```

- **Assembly code**
- **Interface/driver functions**
- **SIMD parallel data distribution**
- **Data reduction**

are generated from this "high-level description".
(Can be ported to GPUs and FPGAs)

Interface functions

```
struct SING_hlt_struct0{
  double xi;
  double yi;
  double zi;
  double e2;
};
int SING_send_i_particle(struct SING_hlt_struct0 *ip,
                         int n);
...

int SING_send_elt_data0(struct SING_elt_struct0 *ip,
                        int index_in_EM);
...
int SING_get_result(struct SING_result_struct *rp);

int SING_grape_run(int n);
```

A few more thoughts on software

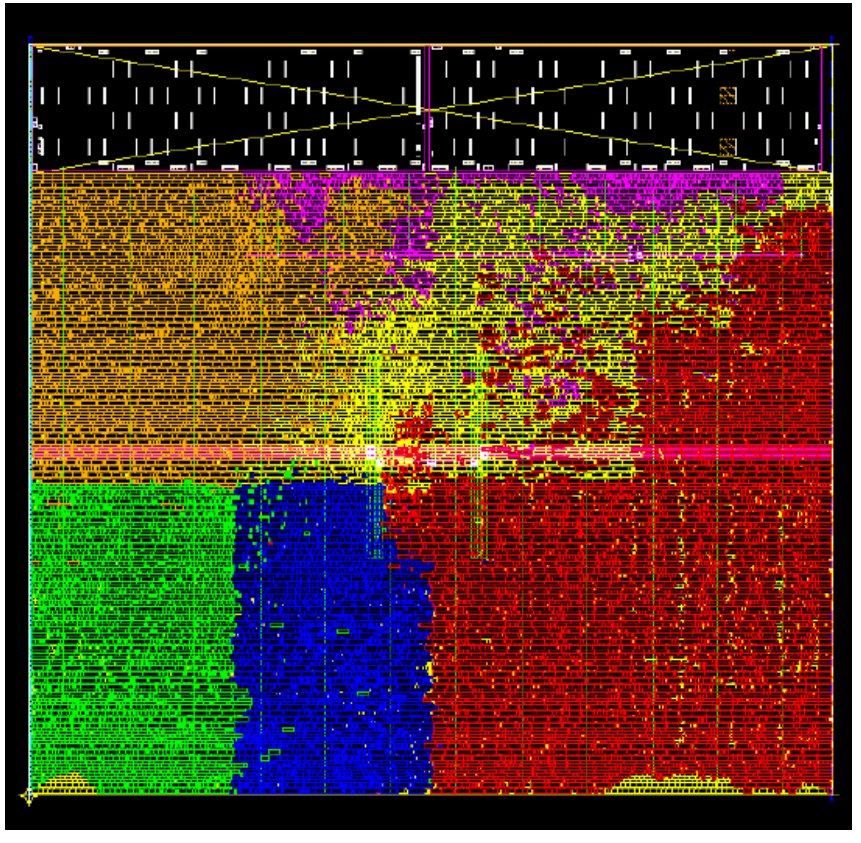
- The right way to separate the task between host CPU and (GRAPE, GRAPE-DR, GPU, FPGA) is the same
- The right way to make efficient use of large number of processors on (GRAPE, GRAPE-DR, GPU, FPGA, CPU) is the same

We should develop a common software platform for different hardwares

Development status

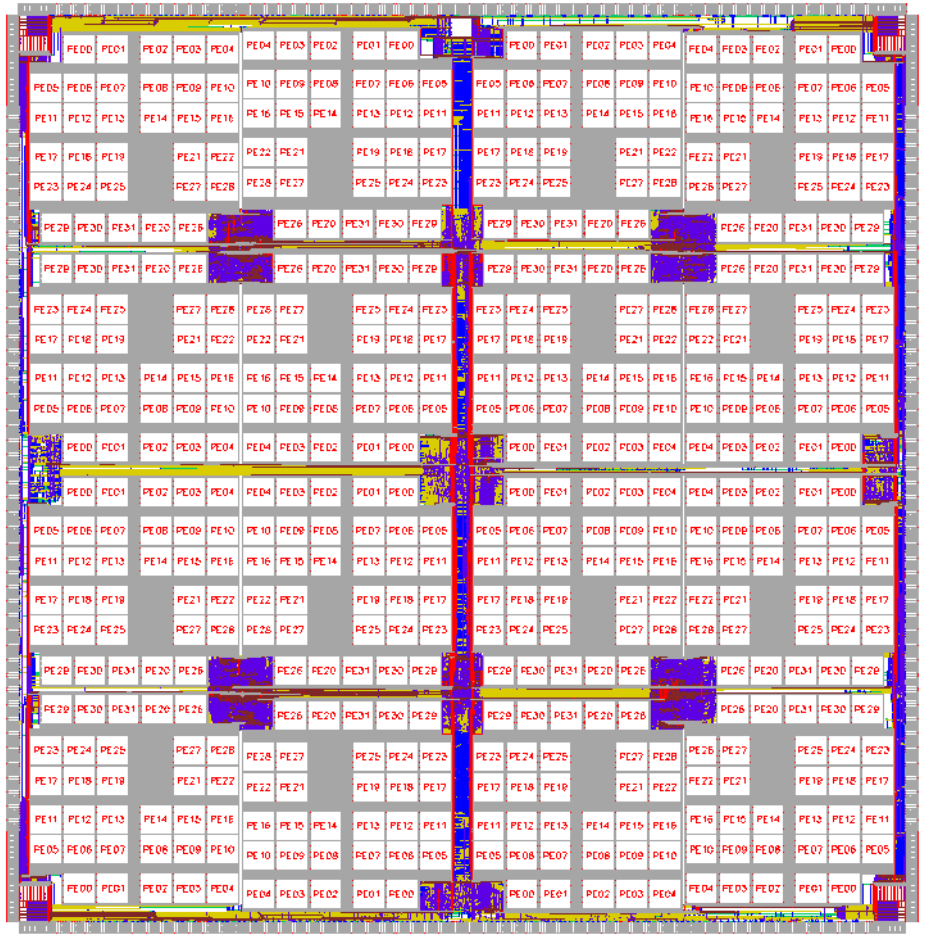
Sample chip delivered May 2006
Chip and board at Booth # 2133

PE Layout



0.7mm by 0.7mm
Black: Local Memory
Red: Reg. File
Orange: FMUL
Green: FADD
Blue: IALU

Chip layout



- 32PEs in 16 groups
- 18mm by 18mm

Prototype board

2nd prototype. (Designed by Toshi Fukushige)
Single-chip board
PCI-Express x8 interface
On-board DRAM
Designed to run real applications
(Mass-production version will have 4 chips)

Prototype board performance

- Measured on first board with PCI-X interface (communication limited)
- Assembly code not fully optimized yet
- 50 Gflops for $N = 1024$
- asymptotic speed 170Gflops

Preliminary data for first commercial version

- Prototype board working
- 1 Chip on a board (0.5Tflops peak)
- PCI-Express x4 interface
- 80W ...
- $\sim$ 5K USD ...

GDR-2?

- With 65nm, it is not difficult to achieve
  - 768 DP Gflops/chip
  - 1.5 SP Tflops/chip
  - On-chip memory (16-32MB)
- Could reach 10Pflops with 13,000 chips, 2-3MW
- With 45nm the performance more than doubles

Summary

- **GRAPE-DR is a massively parallel SIMD processor chip with 512 PEs in a chip.**
- **Can be applied to a fairly wide range of applications.**
- **The processor chip is completed and is working as designed.**
- **Peak speed of a card with 4 chips will be 2 Tflops.**
- **We plan to complete a 1 Pflops (DP) system by the end of FY 2008.**
- **Chip and board at Booth # 2133**
  — Please visit us!