Direct Simulation of Dense Stellar Systems with GRAPE-6

Jun Makino (U. Tokyo)

Summary:

- **GRAPE-6 is currently under development.**
- **12-chip processor board (400 Gflops peak) is up and running.**
- **10-Tflops system will be ready by this summer.**
- **>100(?)-Tflops full system will be ready by 2001.**

Contents

- Introduction — Direct N-body simulation
- Basic Idea of GRAPE
- History of GRAPE Project
- GRAPE-4 and its achievement
- GRAPE-5
- GRAPE-6
- Current status of GRAPE-6
- Science on GRAPE-6

Direct N-body simulation of stellar systems

Most straightforward took to study the dynamics and evolution of stellar systems.

Fokker-Planck:

- good for spherical (or maybe axisymmetric) system in large-N limit.
- Difficult to incorporate tidal field, finite-N effect etc.

MC: Essentially the same as FP code.

HOWEVER,

N-body simulation had not been a really useful tool to study star cluster dynamics, until rather recently

Primary reason: N too limited: $O(100)$ in 1970s, $O(10^3)$ in 1980s.
Calculation cost $\sim N^{3.3}$

Recent advance in computer technology (and GRAPE) have changed the situation.

Some basics on N-body simulation of star clusters

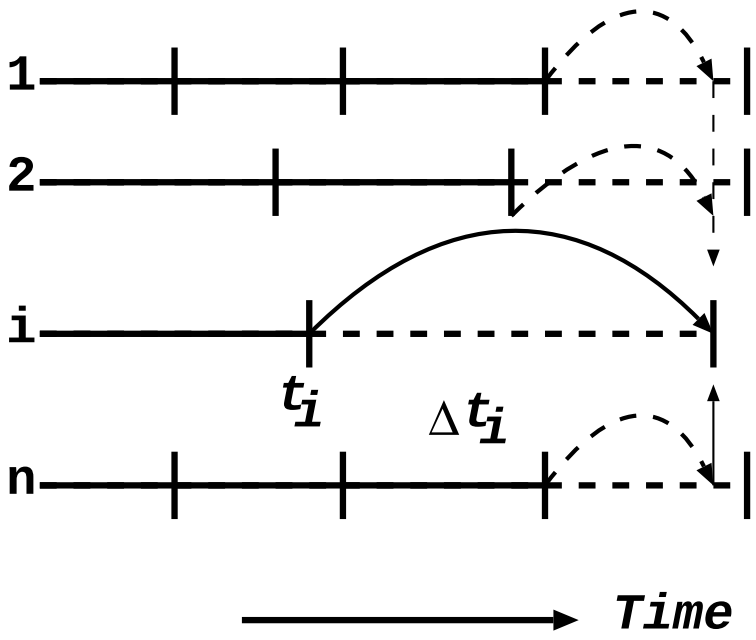
Two key algorithms:

- individual timestep
  - close encounter $\longrightarrow$ small timestep
  - high-density core $\longrightarrow$ small timestep

  Gain in speed compared to usual variable (shared) timestep: $O(N^{1/3} \sim N)$.
- special treatments for close encounters, binaries and triples
  - individual timestep alone is not enough to handle close encounters (roundoff error)
  - stable binaries (and other hierarchical systems)

  Immense effect on both accuracy and CPU time

Individual timestep



- Assign particles their own time and timesteps
- "Event driven" integration — the particle with minimum $t_i + \Delta t_i$ is updated

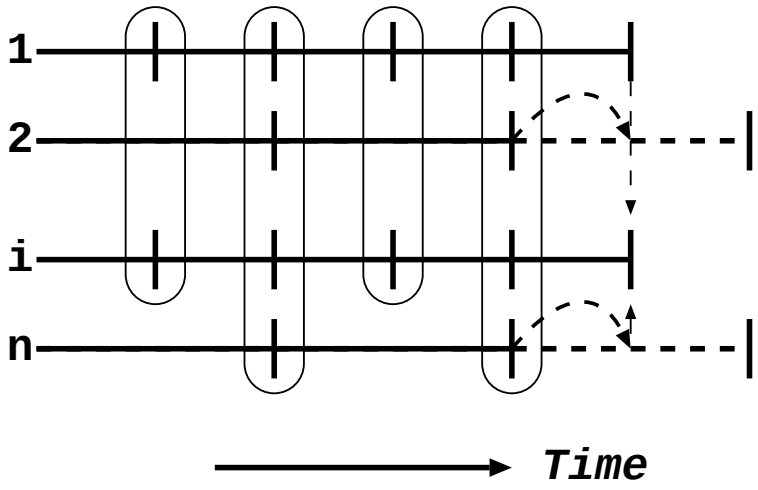
Drawback

- Serial algorithm (one particle at a time): Not much parallelism

The block-timestep algorithm

Modification for vector/parallel machines scheme (McMillan 1986)

- timesteps are limited to 2^{-k}
- particles with same $t_i + \Delta t_i$ are updated simultaneously



- $O(N_c^{2/3})$ particles are updated in parallel — not a very large number

Special treatment for binaries etc

Problems:

- round-off error: $|x_i| >> |x_i - x_j|$
- timesteps for stable binaries

Solutions — Mainly by Aarseth and his coworkers

Round-off → use relative coordinate (and regularization)

Timestep → Skip integration for "isolated" system, use perturbation treatment for weakly perturbed systems...

(Of course, always there are almost stable, but not quite isolated, systems....)

How can we speed up the calculation

Two basic options:

- develop better algorithm
- use faster computer (if necessary, adapt algorithms for new hardware)

Better algorithms?

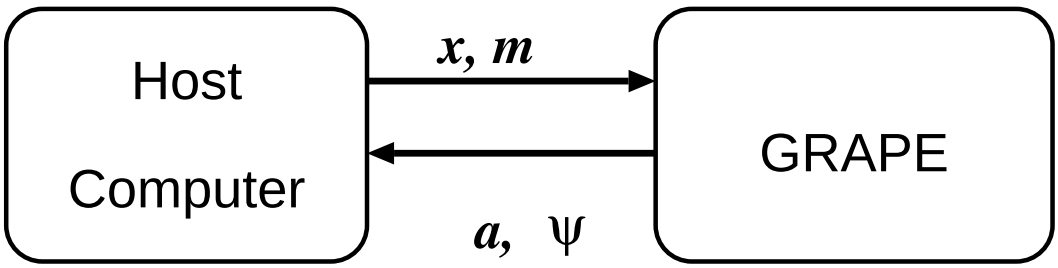
- the neighbour scheme — at the best $O(N^{1/4})$ gain
- Barnes-Hut treecode — ???

Faster computers?

- 1980s — Vector processors: easy to use, no problem
- 1990s and 2000s — MPPs and Beowulf clusters:
  - good performance for large N, shared-timestep code
  - fairly good performance is possible for simple blockstep code without the neighbour scheme
  - (As far as I know of) no successful implementation of high-accuracy treecode with high-order individual timestep scheme

Basic idea of GRAPE

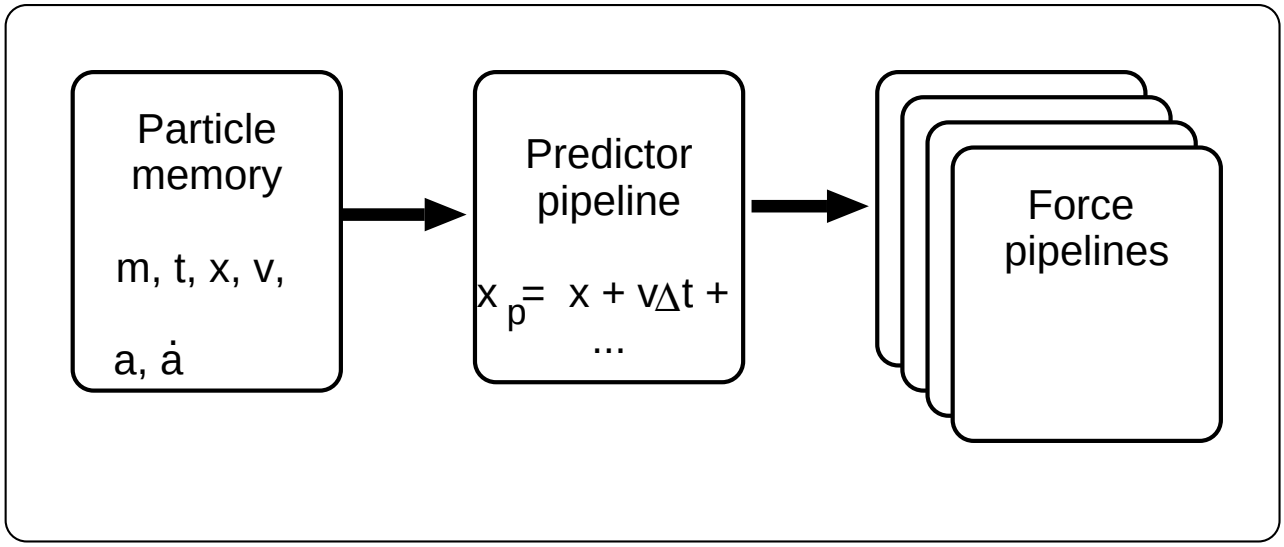
Special-purpose hardware for force calculation
General-purpose host for all other calculation



- Special-purpose hardware
  - Hardwired logic → No software needed
  - Pipelined processor → High performance
- General-purpose host
  - Programmable in high-level languages
  - existing programs can be used with small modifications
  - can be used with Barnes-Hut tree/Aarseth scheme

Individual timestep on GRAPE

- GRAPE hardware performs prediction and force calculation

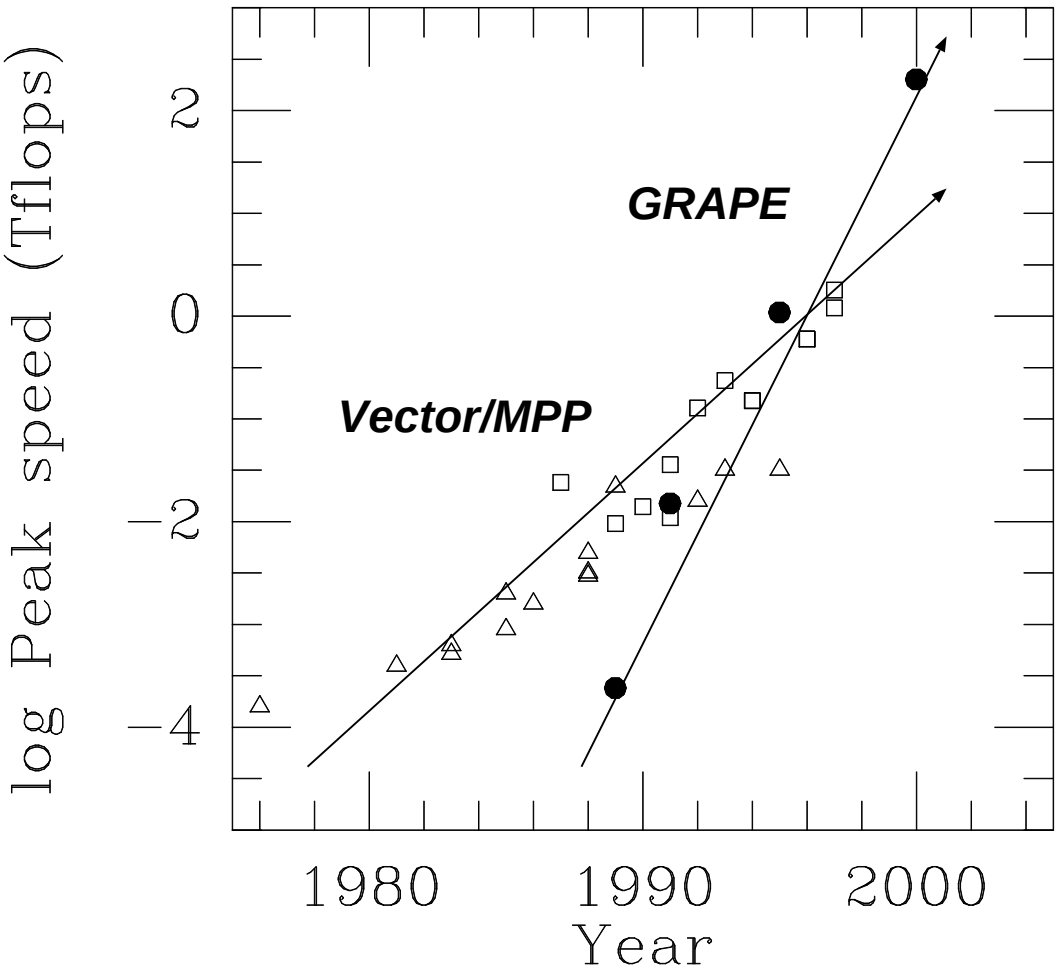


- Special treatments are all handled by the host. Calculation cost for this part is, well, not too high for most cases...

History of the project

Name	development period	speed	Accuracy	Note
GRAPE-1	89/4 — 89/10	120 M	L	
GRAPE-2	89/8 — 90/5	40 M	H	
GRAPE-1A	90/4 — 90/10	240 M	L	
GRAPE-3	90/9 — 91/9	14 G	L	custom chip
GRAPE-2A	91/7 — 92/5	180 M	H	
HARP-1	92/7 — 93/3	180 M	H	Hermite scheme
GRAPE-3A	92/1 — 93/7	6 G	L	∼100 boards made
GRAPE-4	92/7 — 95/7	1T	H	8 sites
MD-GRAPE	94/7 — 95/4	4G	H	programmable interaction
GRAPE-5	96/4 — 99/4	40G	L	
GRAPE-6	97/8 —	100T	H	under development

Evolution of GRAPE systems



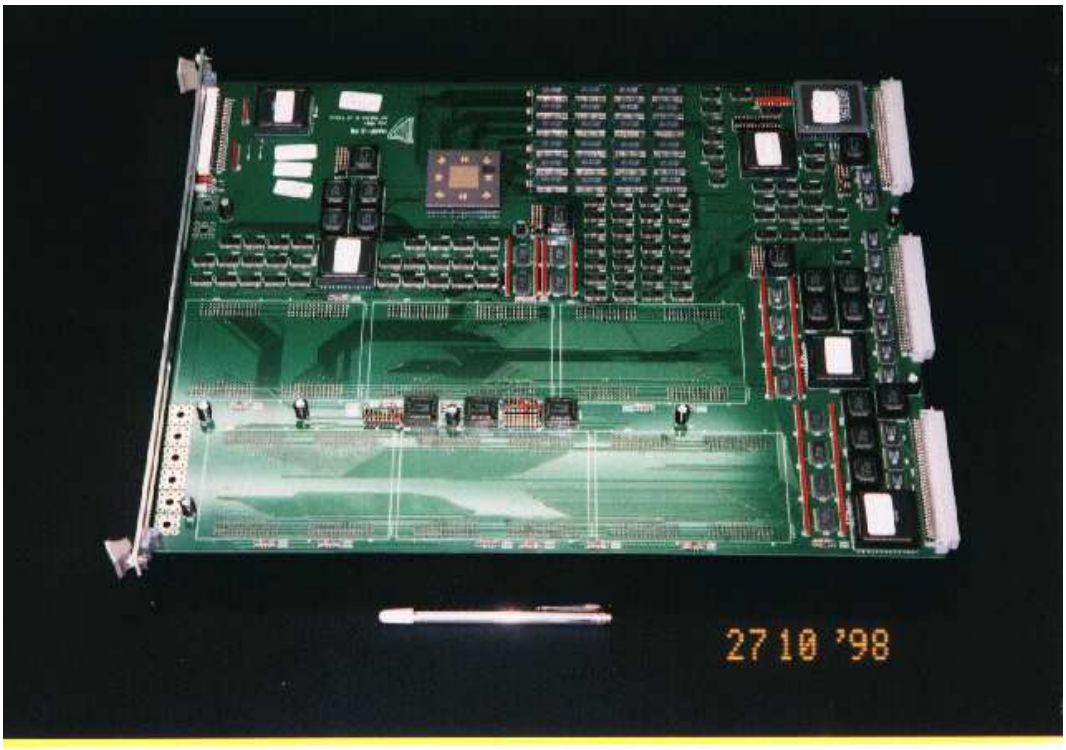
(Figure made in 1997. We were a bit too optimistic then)

GRAPE-4

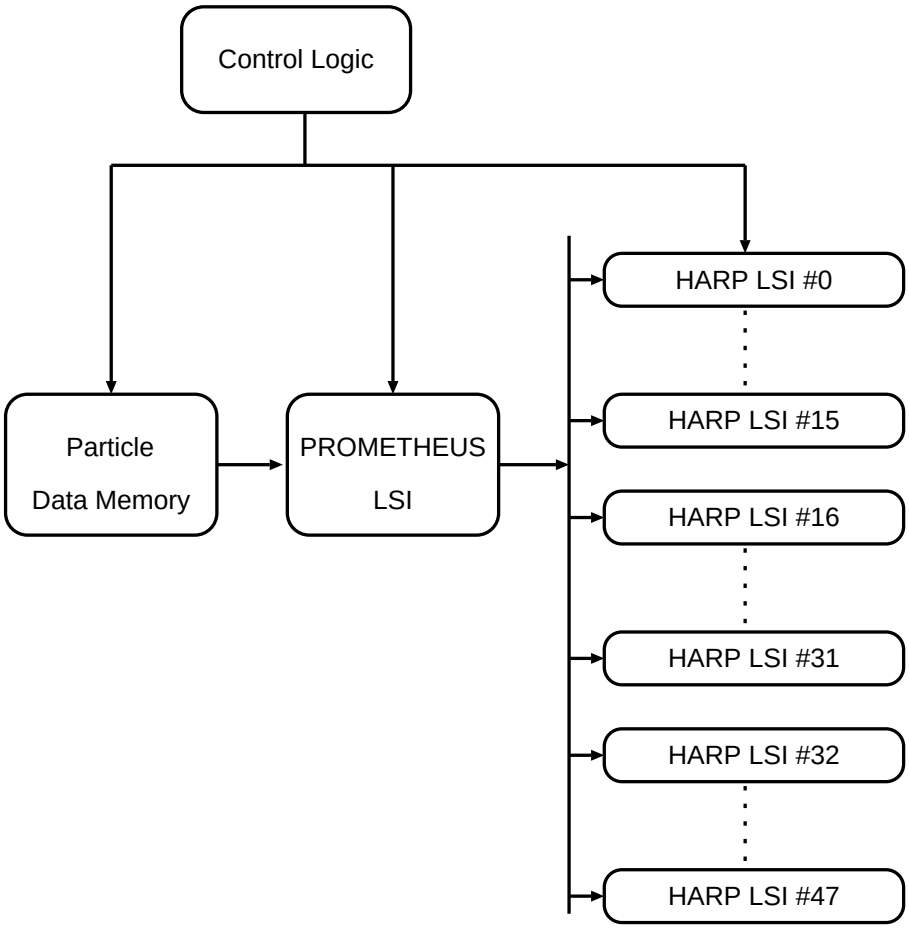
Features:

- Predictor pipeline in hardware
- Massively parallel system with 1692 pipeline chips
- calculates force on 94 particles

GRAPE-4 processor board



GRAPE-4 processor board



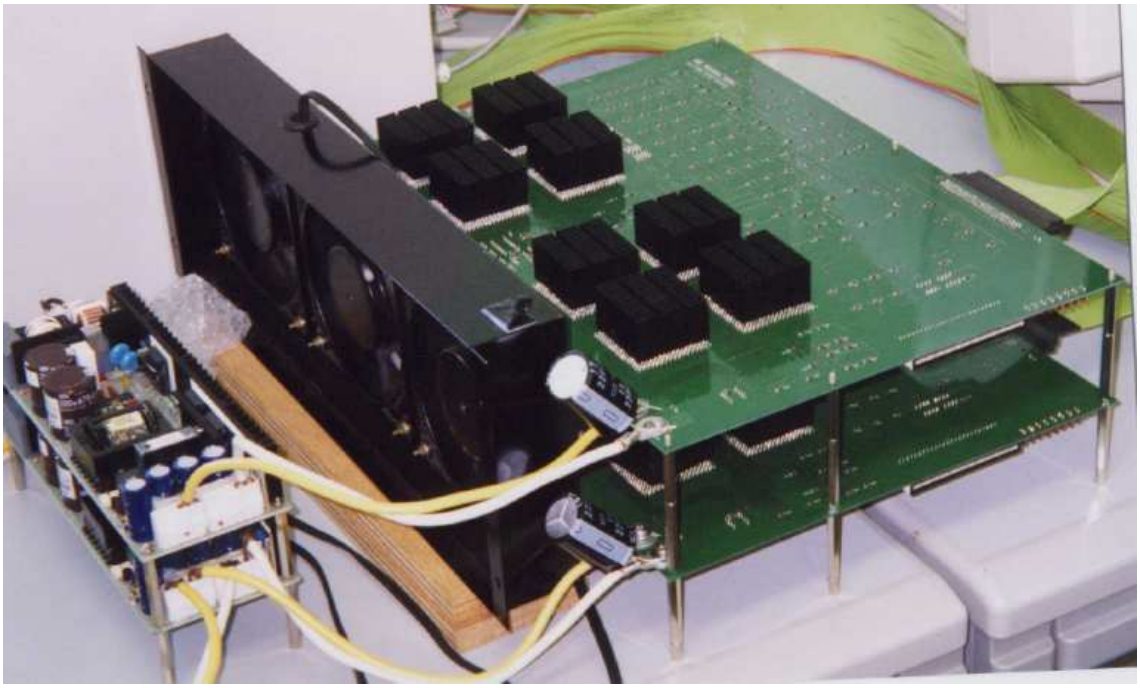
Many pipeline LSIs share one memory unit
→ Massively parallel system with small parts count (=low cost)

All processor boards calculate the force on same set of 94 particles, but from different subset of particles in the system.

The summation is done partly on GRAPE hardware, and partly on the host.

GRAPE-5

- Successor of GRAPE-3
- Improved calculation speed ($> 8\times$ per chip)
- Peak speed 40 Gflops /board
- Improved communication speed ($>$ 50MB/s)
- Support for periodic boundary condition



GRAPE-5 (production version)

GRAPE-5 Availability

- First delivery last summer
- Price/board $\sim$ 15k\$ (plus handling fee of trading company, plus ...)
- Host requirement: PCI interface.
- Recommended host: DEC (Compaq) Alpha with Digital UNIX.
- The software for Linux (RH) PC is now ready
- Check http://grape.astron.s.u-tokyo.ac.jp/grape/ for details

GRAPE-6

In 1995 we thought:
GRAPE-4 is a nice machine. However, the semiconductor technology is advancing rapidly

↓

A new machine to be completed by 2000 would be 10 ∼ 100 times faster for the same cost.

So we started to look for money, and somehow Sugimoto succeeded to get 500M JYE grant.

Project goals for GRAPE-6

- Peak speed > 100Tflops
- Improved communication speed (> 2GB/s)
- Minor fixes from GRAPE-4...

Design considerations

We cannot simply scale up GRAPE-4 design

- calculation speed becomes 100 times faster
- memory bandwidth remains almost the same

→ Too high parallelism. Not enough particles with blockstep algorithm.

New architecture

- each chip has its own memory
- use fast clock for memory
- pipelines on a chip share the memory
- different chips in a cluster calculate the force on the same particles

↓

- reasonable performance with blockstep scheme (48 virtual pipelines per chip)

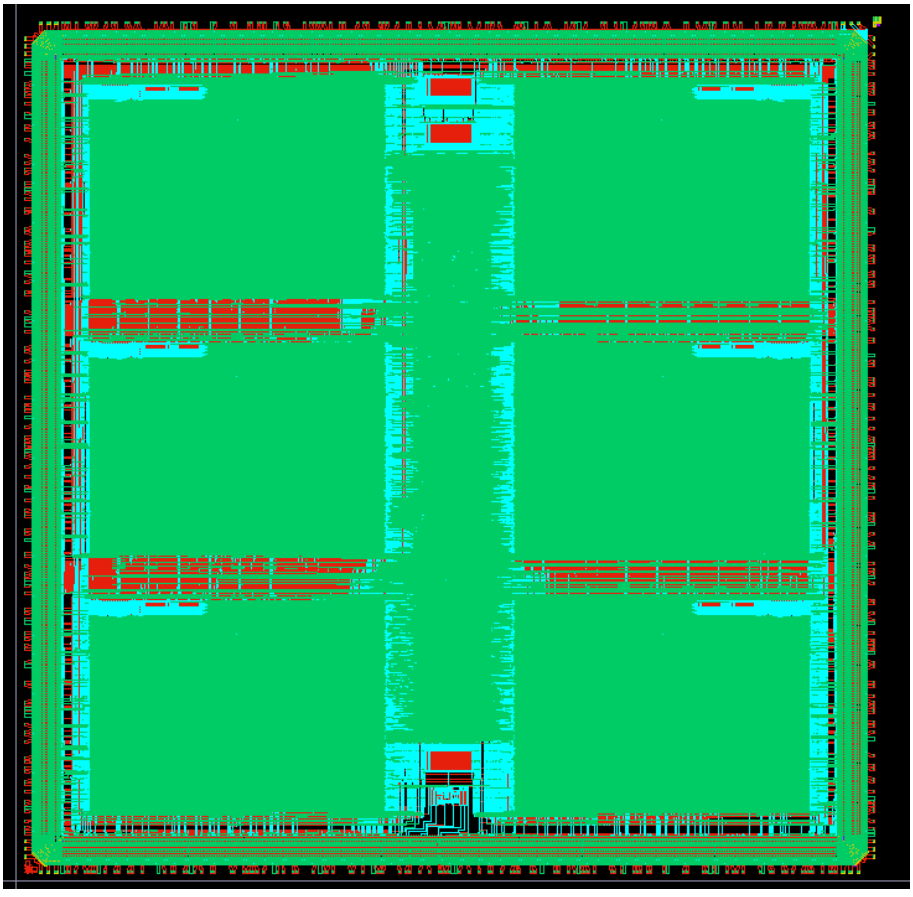
Overall structure

- 16 (or somewhat less...) clusters
- 16/8 processor boards per cluster
- 8 processor modules per board
- 2/4 chips per modules
- 6 pipelines per chip
- 60 arithmetic units per pipeline

In total,1.5×10^6 Arithmetic units operating at 100 MHz = 150 Tflops

GRAPE-6 Processor Chip

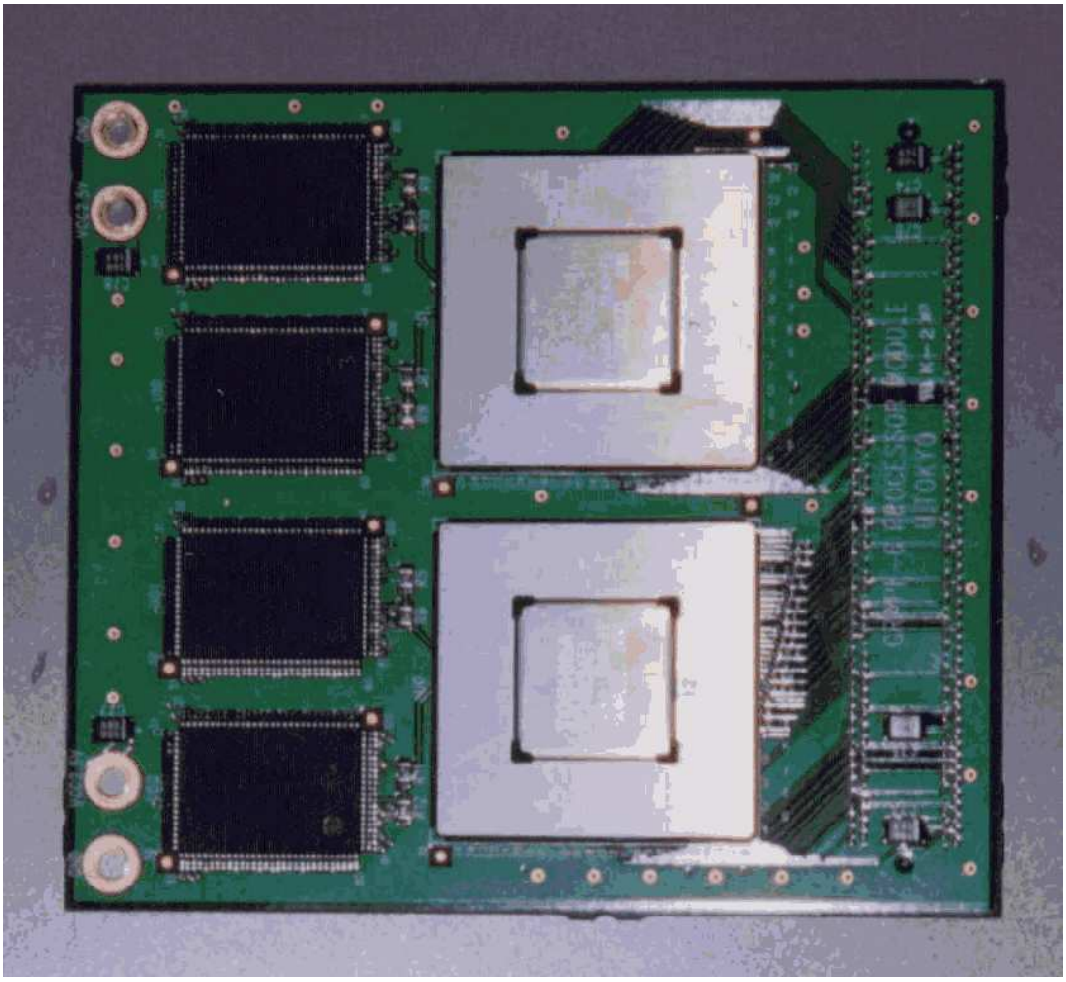
- 0.25 μm custom design (Toshiba TC240E, 1.8M gates)
- 2.5 V power supply, 100 MHz operation
- 6 pipelines in a chip
- 1 predictor pipeline
- communication and control logic integrated on chip
- $>$ 30 Gflops (equivalent) per chip
- system clock: 25 MHz
- Single chip equivalent to single GRAPE-4 board



Difference from GRAPE-4 chip

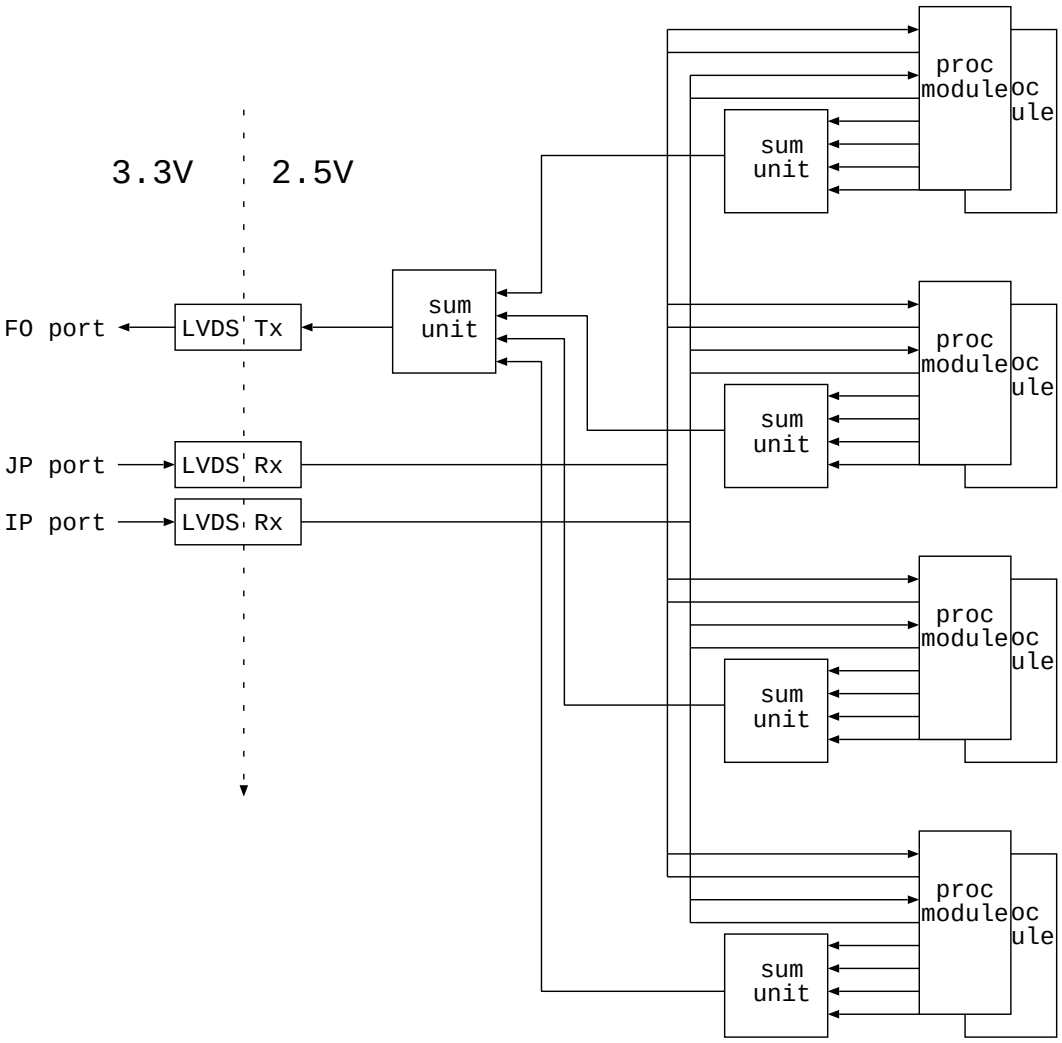
- Force is accumulated in fixed-point format (exponent is given by host) format — single-chip and multi-chip systems give exactly the same result
- Returns the index of the nearest neighbour — useful to recognize close encouner etc
- some other minor problems are fixed

GRAPE-6 processor module



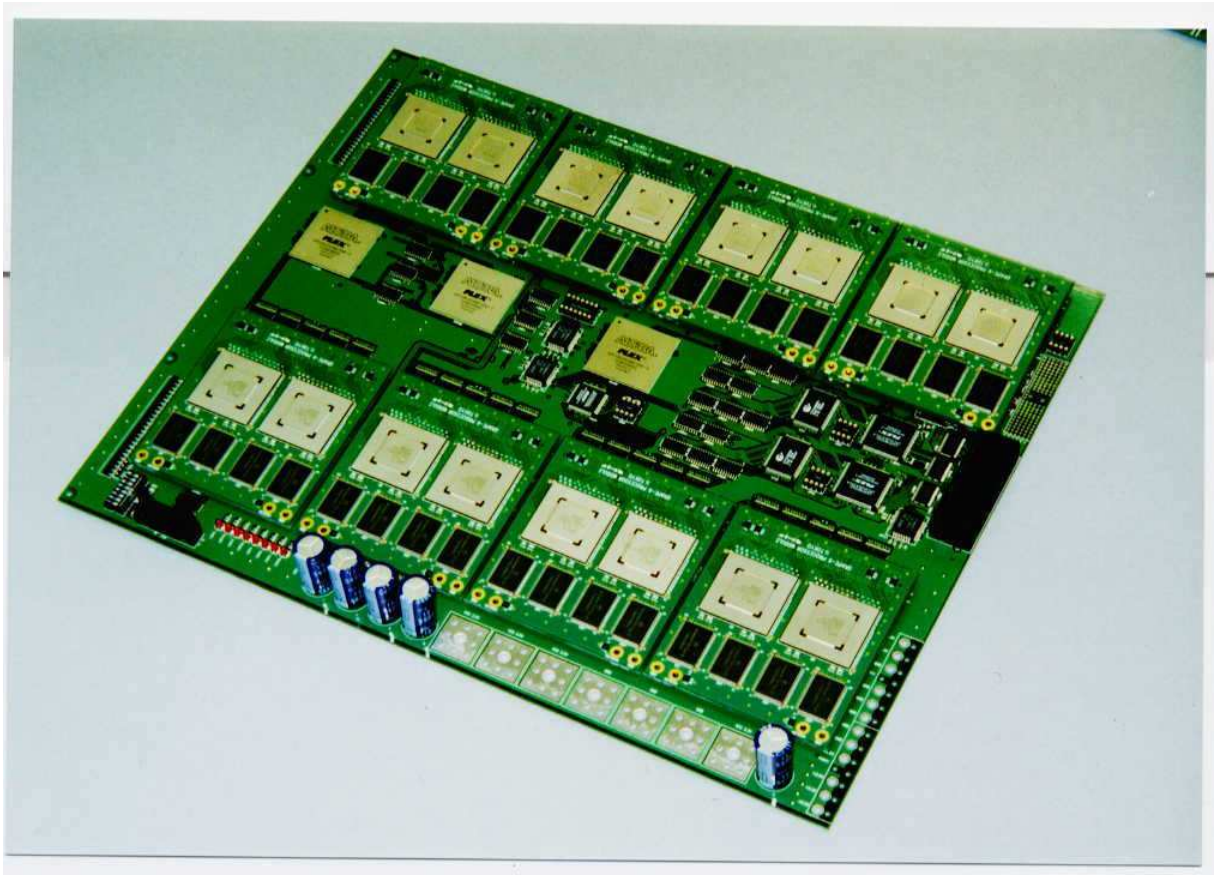
- Two processor chips + Memory chips.
- Integrates all functions of GRAPE-4 processor board.

The GRAPE-6 Processor board



- Single board with 16-chips
- Semi-serial (LVDS) interface (350MHz clock, 4 wires)
- Two input ports and one output port
- Result from 16 chips are summed by a reduction tree

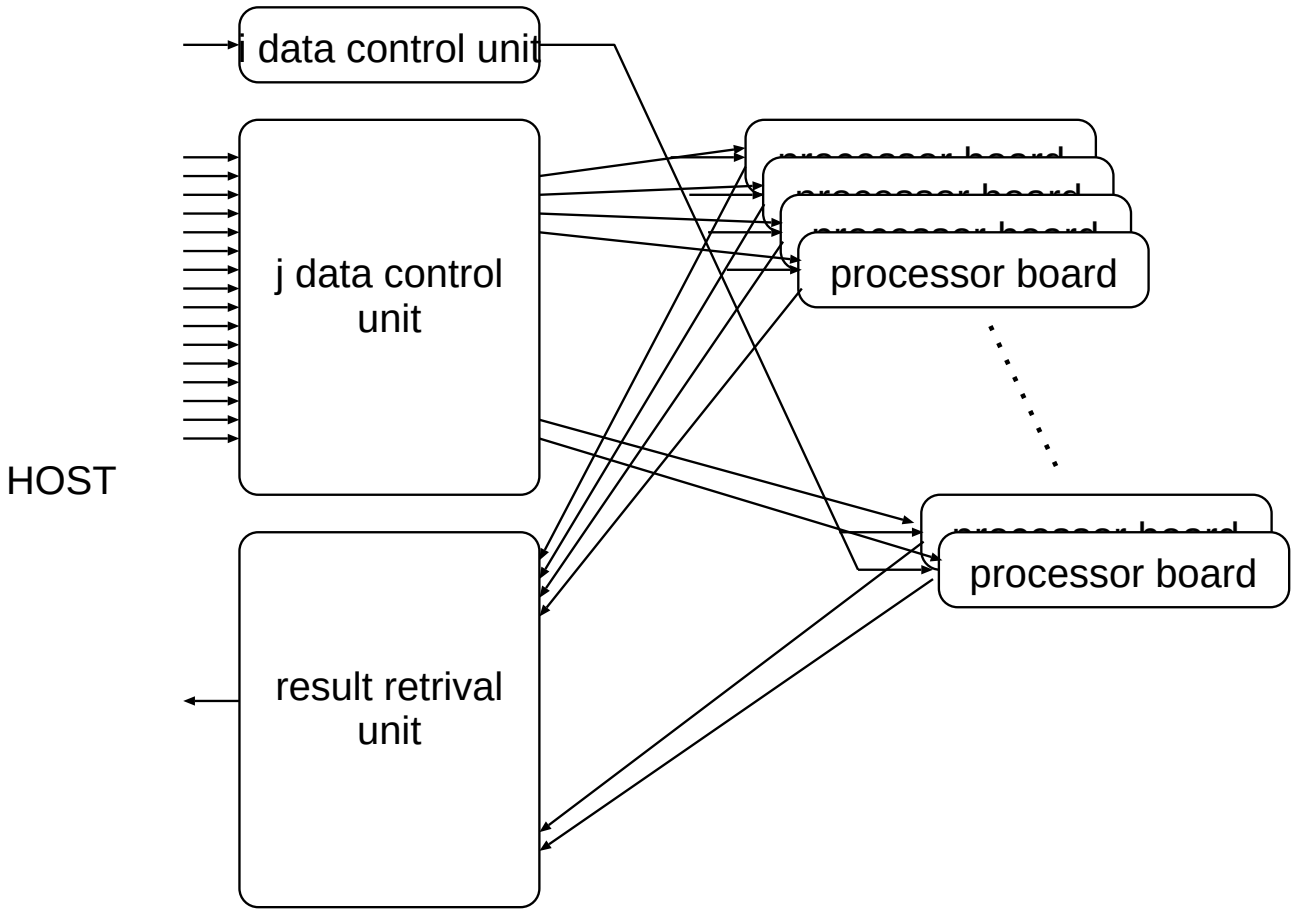
The GRAPE-6 processor board



Current status

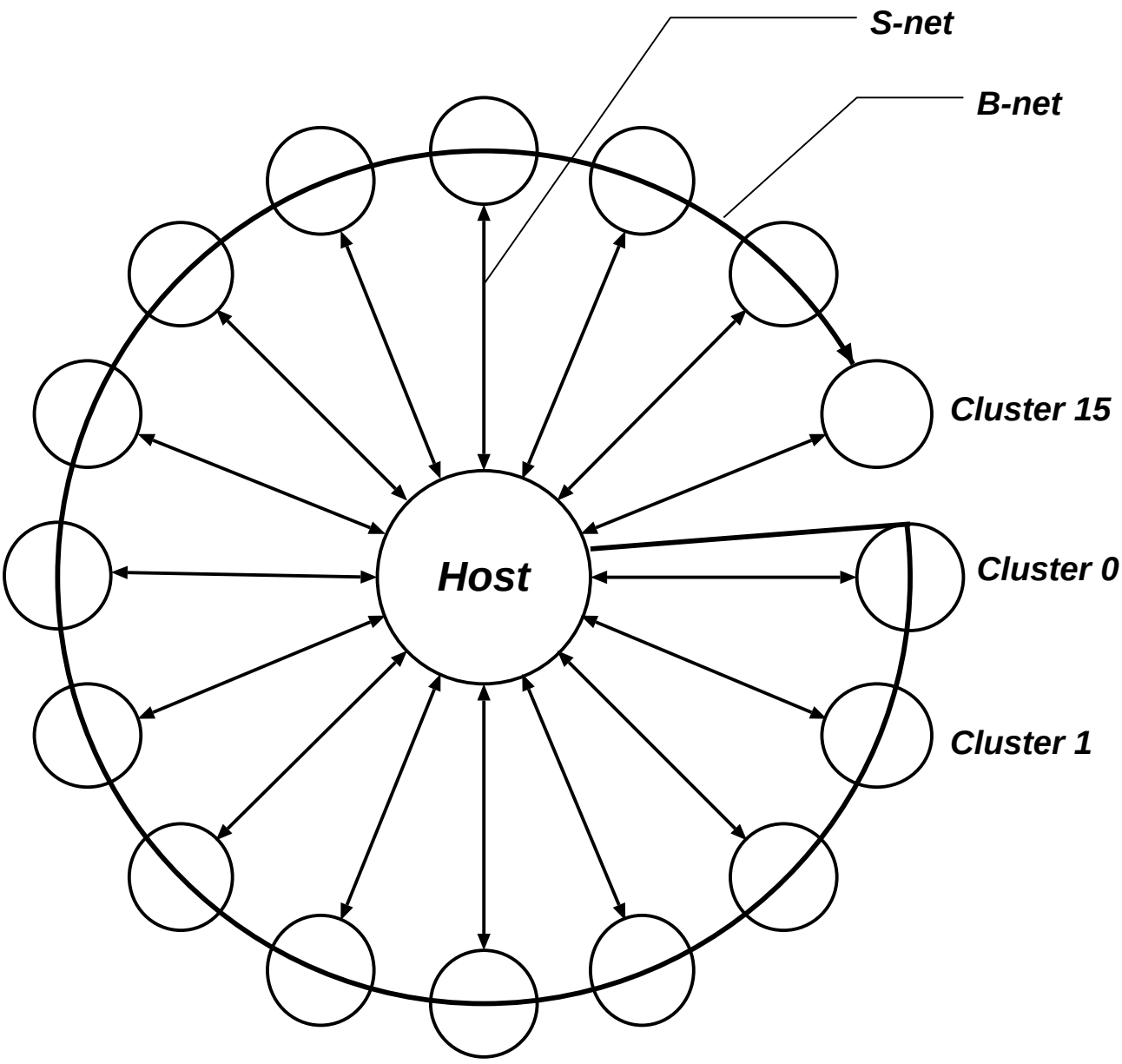
- First PCB board is up and running now (as of March 16)
- Current clock: 80 MHz
- Number of chips: 12
- First batch of production version should arrive in two weeks

Structure of a cluster



- 16 (or 8) processor boards
- 3 dedicated communication networks
- jdata port (particle memory data) has a multicast network to support multi-cluster calculation

Structure of the total system



- 16 clusters
- 256 chips/cluster
- S-net for register read/write
- B-net for particle memory write (multicast)

GRAPE-6 project schedule

Apr/May 2000:	First multi-board system
Jul/Aug 2000:	New version (4-chip module)
Dec 2000:	Half of the system (40-50 Tflops)
Dec 2001:	Full system (100 Tflops)

This is about 6 months behind our original schedule

Main reason: Deliverly of the fully working chip was about a year after the orignal schedule, due to

- physical design delay
- fabrication error in the first sample
- physical design error (insufficient power supply capacity)

Making a eight-million-transistor chip is not easy...

Science on GRAPE-6

GRAPE-6 (either full system or smaller system) offers about 100 times more raw computational speed compared to similar configulation of GRAPE-4.

Thermal (long-term) evolution

GRAPE-4: can do equal-mass 32k-particle globular cluster simulations in weeks
GRAPE-6: can handle 150k-particle simulations in similar time. For "realistic" calculation with tidal field, mass spectrum and stellar evolution, 300-500k is within reach

Dynamical (short-term) evolution of dense stellar system

GRAPE-4: up to about 1M
GRAPE-6: up to 10M in direct method

Galactic dynamics or cosmology with treecode

GRAPE-4: currently (host) memory limited. 10M is easy if the host has that much memory
GRAPE-6: Theoretically, 1G is within reach, if 100 GB of memory is available... Need to use clusters

Summary

- Development phase of GRAPE-6 is almost finished.
- At present, single processor board with 12 processor chips (peak speed 350 Gflops) is up and running.
- Multi-board system will be ready soon, and full system is scheduled to be ready by next year
- GRAPE-6 will offer 100-fold increase of the computing power over GRAPE-4, which is still more than competitive with any other machines for star cluster simulations.