Galaxies, Dark halos, and microhalos

Jun Makino
Center for Computational Astrophysics/
Division Theoretical Astronomy
National Astronomical Observatory of Japan

August 26, 2010

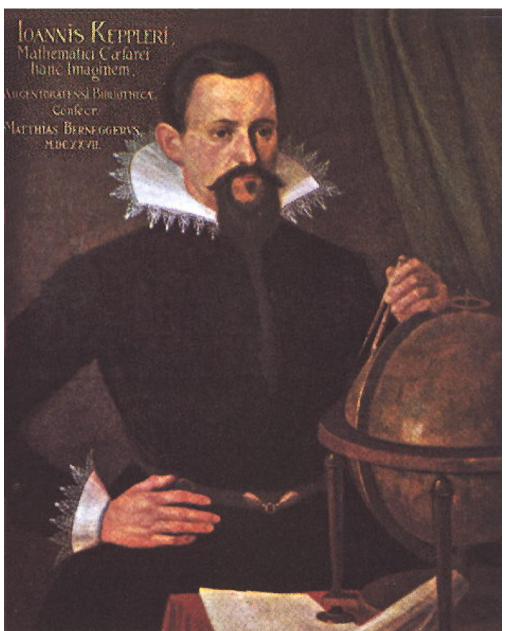
Center for Computational Astrophysics

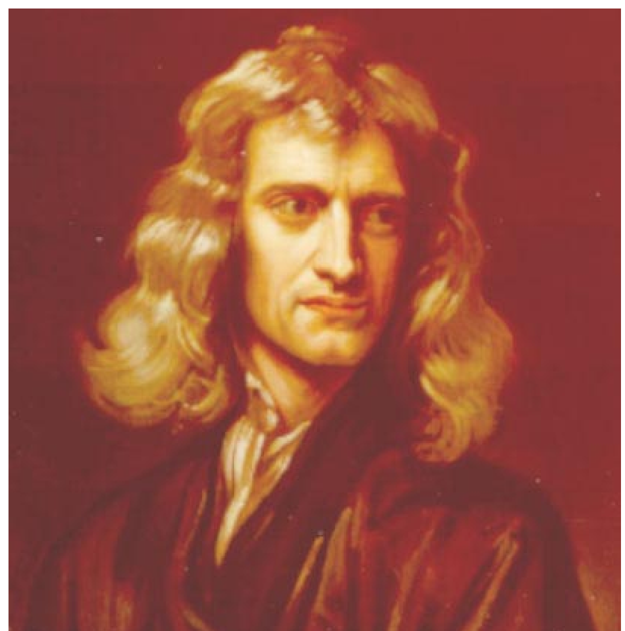
Structure of my talk

- Gravitational N-body problem
  - Solar system dynamics
- Stellar Dynamics
  - Some basics
  - Example of stellar systems
  - Galactic dynamics
  - Dark Matter Halos
- Numerical methods

First example of Theoretical Astronomy

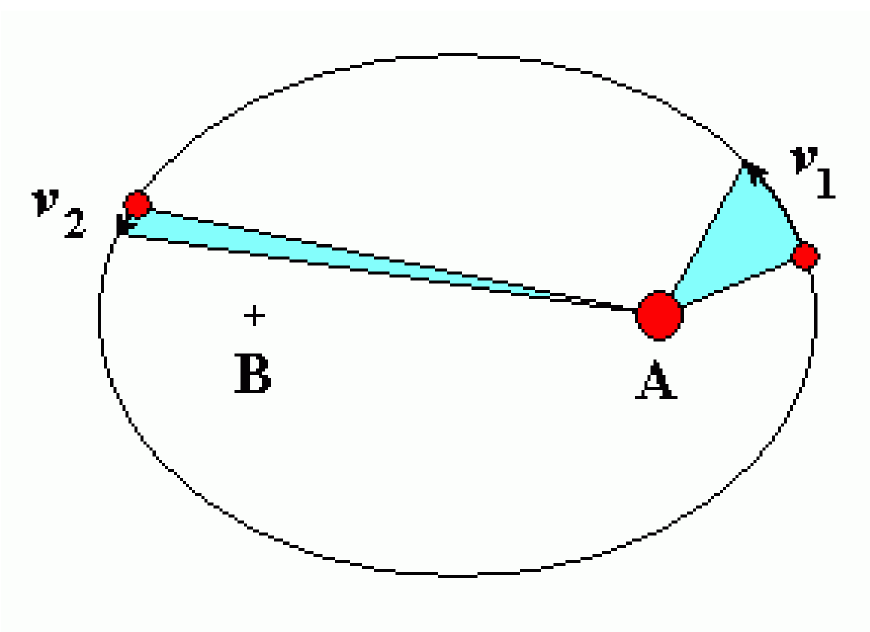
- Kepler formulated, from Tycho's observations, Kepler's three laws.
- Newton showed that Kepler's laws are derived from Newtonian mechanics and Newtonian gravity.





Kepler's laws

- The orbits of planets are ellipse with one focus at Sun.
- $dS/dt = \text{const.}$
- $T \propto a^{3/2}$



Newtonian equation of motion for planets: Two-body problem

$$\frac{d^2\mathbf{r}}{dt^2} = -GM\frac{\mathbf{r}}{|\mathbf{r}|^3},$$

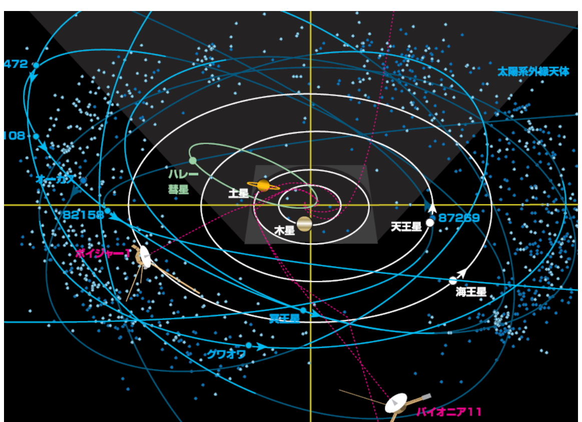
- Gravity from other planets were neglected.
- Simple closed elliptic orbits.

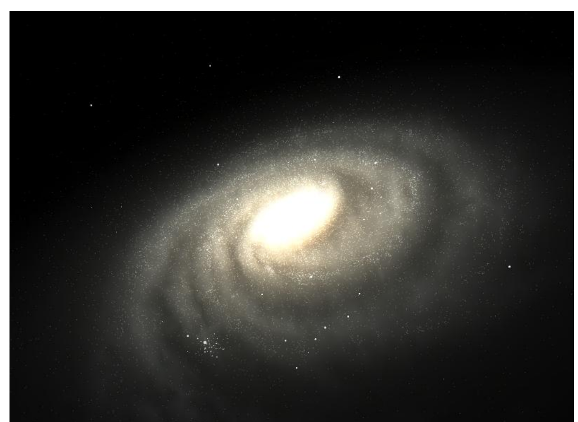
$N > 2$

- Celestial mechanics: What happens if we include planet-planet interaction?
- Stellar Dynamics: How stars themselves move?

Both are very natural "next steps" from the two-body problem.

In both fields, there are significant recent advances.





Two planets

- Simple example: Mars under the effect of Jupiter
  - Gravity from Mars to Jupiter is small
- More general case: Saturn and Jupiter
  - Saturn is not small

Perturbation technique

Basic idea:

- Start from unperturbed Kepler orbits
- Derive the equation for the change of orbital elements due to the gravity of Jupiter
- Expands it by the mass of the Jupiter
- Evaluate the first term (or first few terms...)
- (Usually assume that orbits are close to circular and close to coplanar)

Can be extended to general cases

Success of the perturbation technique

- Explained high-precision observations of the orbits of planets
- Unexplained motions led to
  - findings of new planets (Neptune)
  - Confirmation of general relativity (Mercury)

So, is everything OK?

— not quite.

One problem:

Long-term “stability” of the solar system.

Last 20 years of stability study

1987: Sussman and Wisdom
850Myrs numerical integration of outer five "planets"

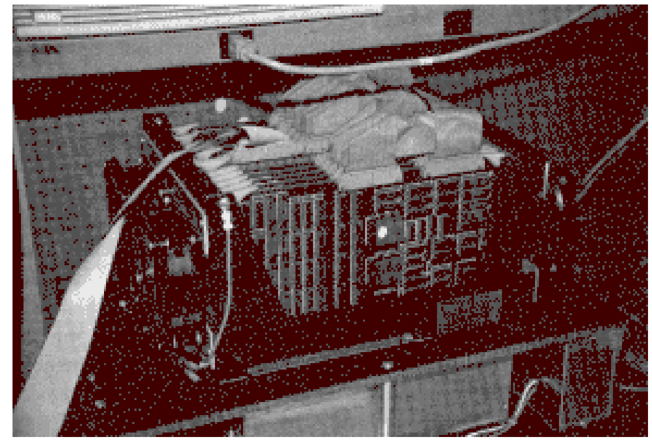
Lyapunov timescale: 20Myrs

Lyapunov timescale: (Roughly speaking) the distance between two (infinitesimally different) systems grows in this timescale

The Digital Orrery

Computer used by Sussman and Wisdom

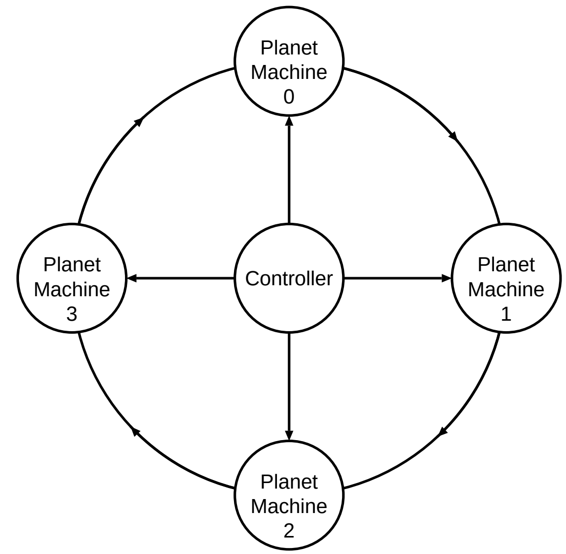
- A custom-built parallel computer for long-term integration of the Solar system



- Consists of 9 “planet computers” connected in a ring network
- 10 Mflops
- MIT AI lab + Planetary Science

Digital Orrery (2)

- SIMD (Single-Instruction Multiple-Data) parallel computer
- Programmable: Integration scheme etc can be changed
- Effective Quadruple-precision integration



One of (very few) examples of the successful development of special-purpose computer for numerical simulation

Naive question:

Lyapunov timescale $\ll$ Age of the solar system

Is solar system unstable? Why is it there?

Even longer numerical integration

- **Kinoshita and Nakai 1996 (4.5Gyrs)**
- **Ito and Tanikawa 2002 (45 Gyrs, 10 times the age of the solar system)**

Solar system seems to be “stable”

What do we mean by “stable”?

- Planets do not collide, exchange positions, escape from system, etc.
- not “linear stability”

Much simpler setup

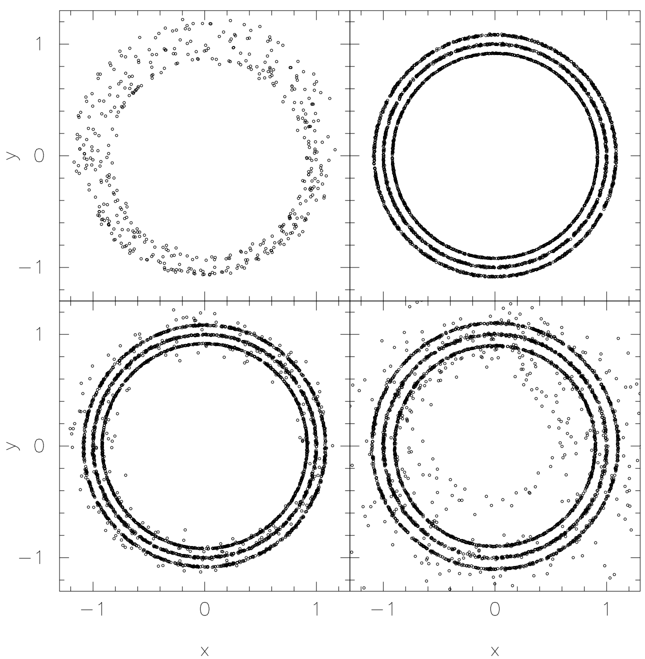
What is known:

- **Sun + two planets: STABLE (if two planets are well separated)**

What happens to the system of *three planets***?**

Simple experiment

- planet mass: 10^{-5} (Sun=1)
- planet separation: 0.06, 0.08, 0.1



Left top: 0.06, T=5000

Right top: 0.08, T=50000

Left bottom: 0.08, T=60000

Right bottom: 0.1, T=90000

"Suddenly" become unstable

Numerical experiments suggest:

- "Instability timescale" $\propto$ exp(separation)
- Weak dependence on the number of planets
- separation normalization: Hill radius $r_H = R(m/M)^{1/3}$
- Initial eccentricity reduces the timescale

Might imply:

- Planetary system (with more than three planets) is always unstable, if you wait long enough
- In the case of our solar system, instability timescale is longer than 10× its age.

Is the stability of our solar system such an important problem?

- Extrasolar planetary systems
- Trans-Neptunian Objects
- Formation theory for normal planets
- Earth's long-term climate change

Summary for planetary systems

N: number of planets

- $N = 1$ solved by Newton
- $N = 2$: stable if large separation
- $N \geq 3$: Everything becomes unstable?
  - Why does our solar system exist?
  - Wide variety of extrasolar planets

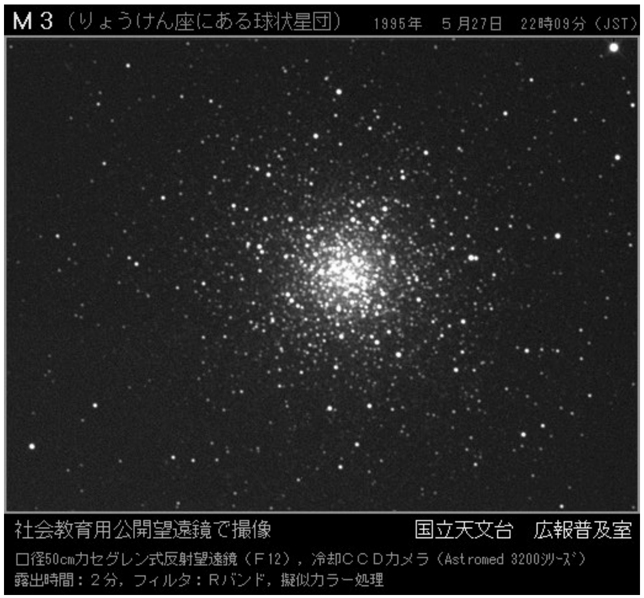
Stellar Systems

Planetary systems: Sun + “small” planets. Kepler orbit+perturbation.

Stellar systems: Consists of many stars

Examples of stellar systems

Globular clusters



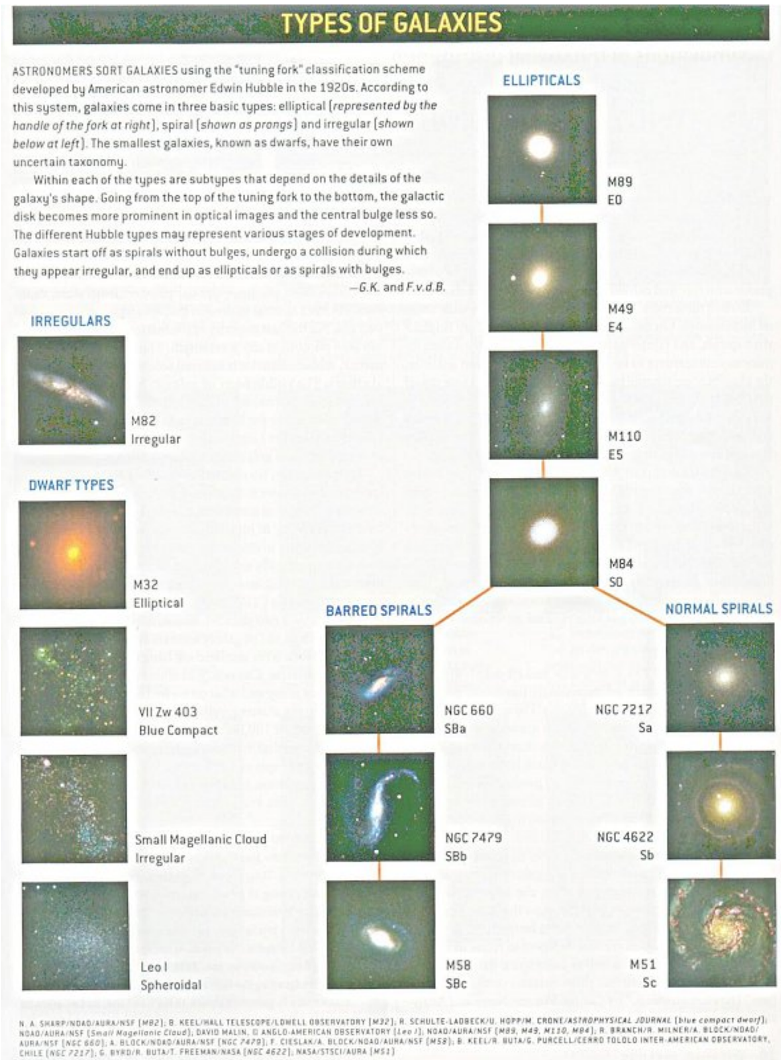
Galaxies

Globular clusters

- 10^5-10^7 stars
- Old stars ,> 10Gyrs (age of the Universe: 13.7Gyrs)
- Mostly spherical (some are a bit elliptical, rotating)
- Globular clusters all look alike
- "Clean" systems, no gas, star formation etc

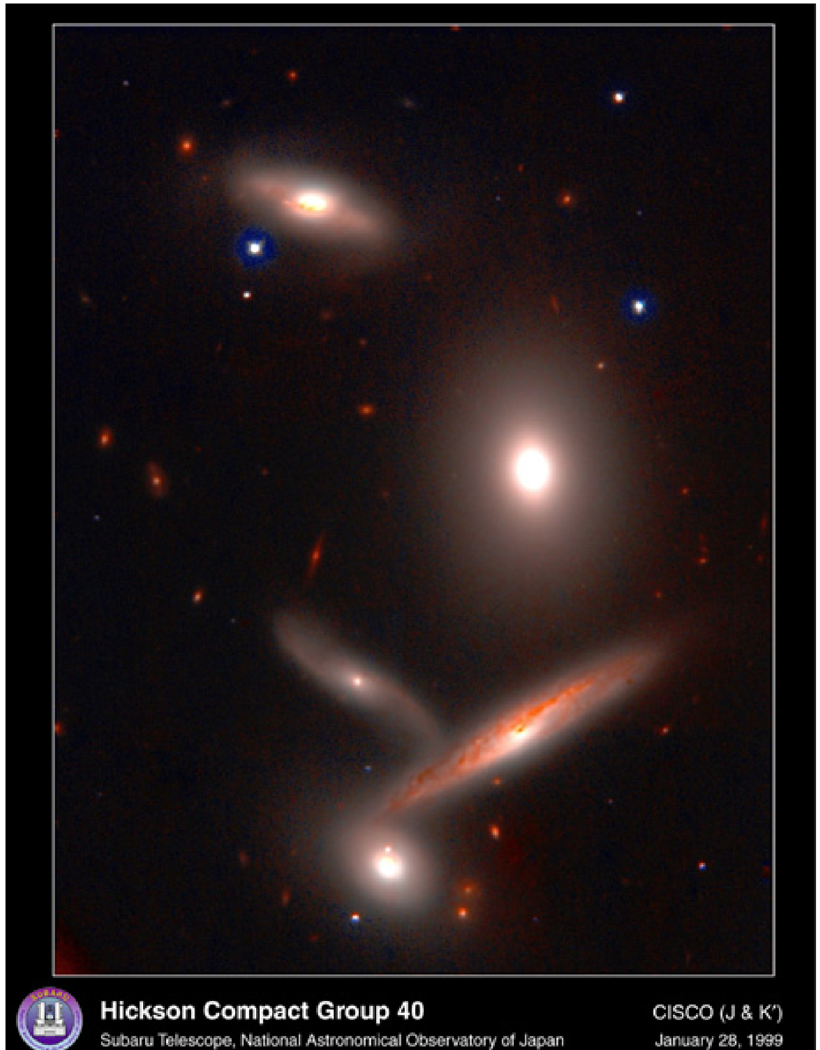
Natural lab for stellar dynamics

Galaxies

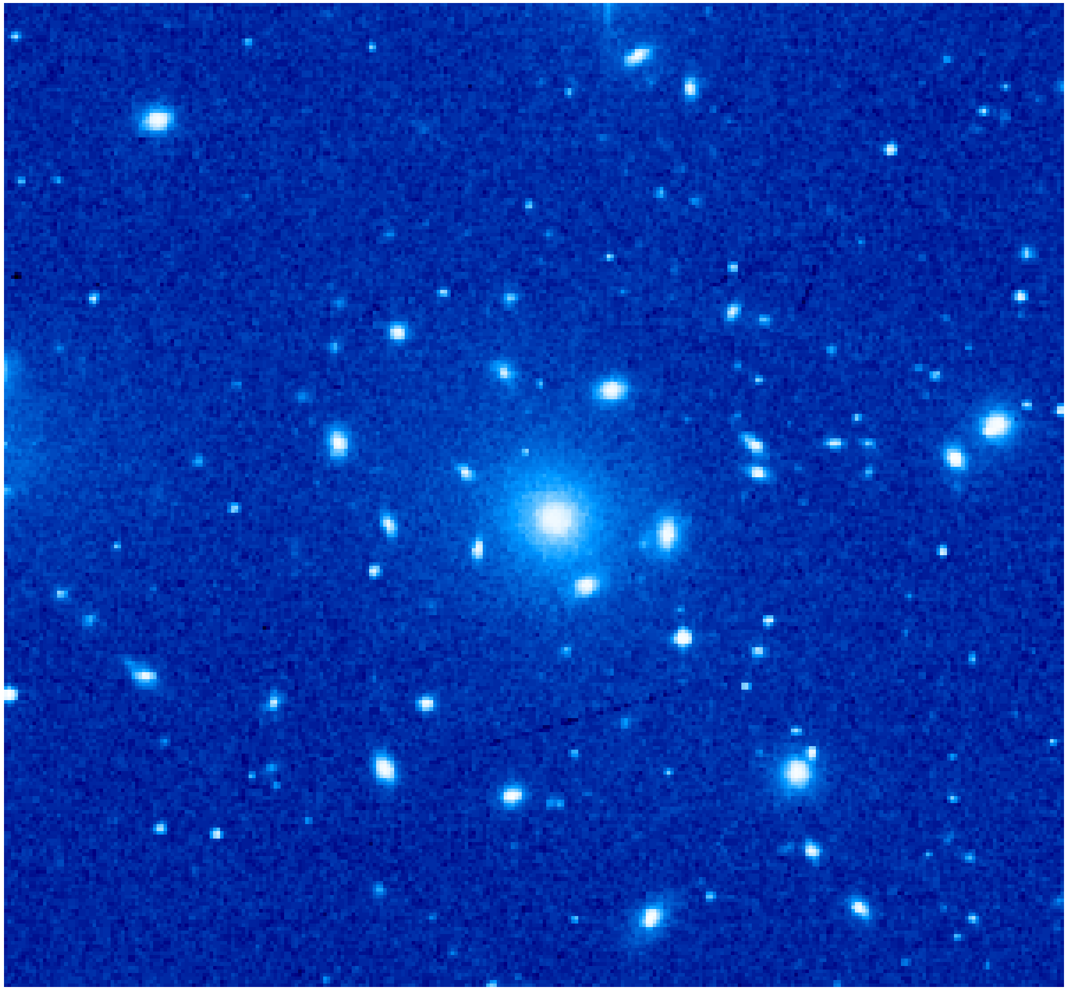


- $\sim 10^{11}$ stars (wide variety)
- Complex systems, gas, stars are forming
- Wide variety in morphology

Galaxy groups

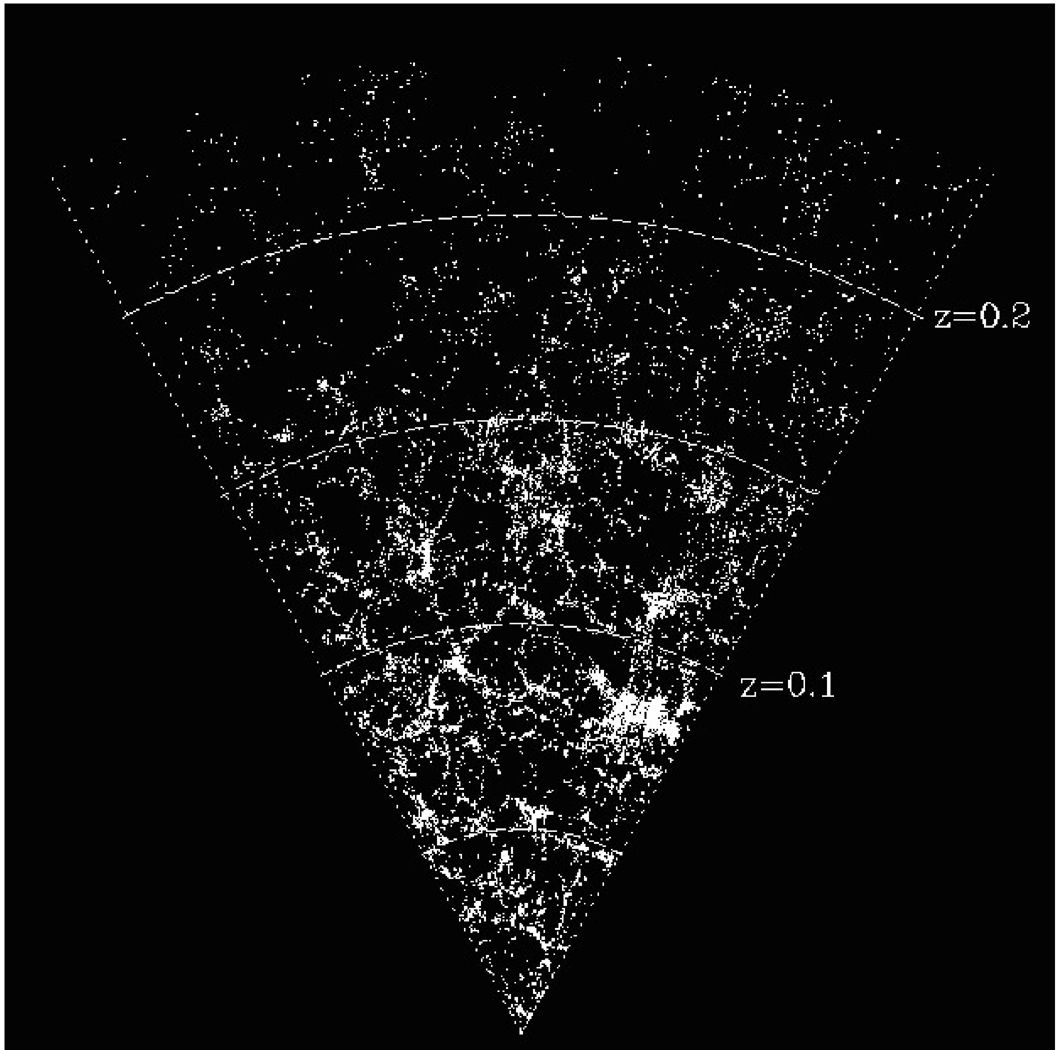


Clusters of Galaxies

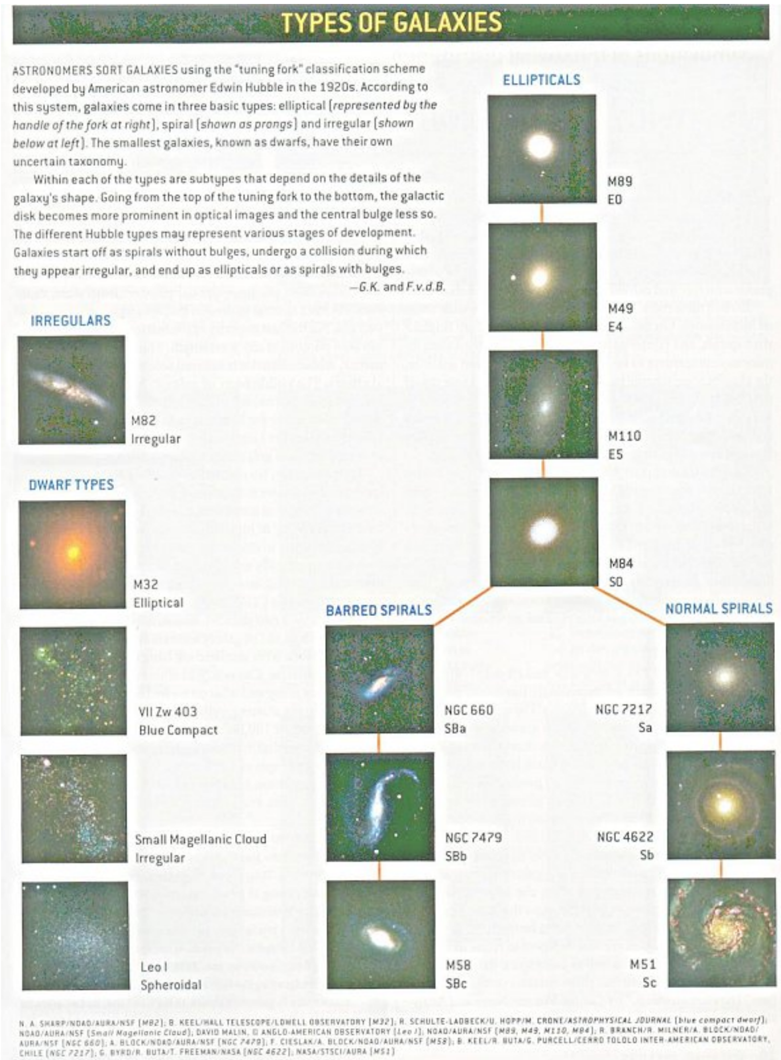


http://antwrp.gsfc.nasa.gov/apod/ap950917.html

Large-Scale Structure



Simulation of galaxy formation



TYPES OF GALAXIES

ASTRONOMERS SORT GALAXIES using the "tuning fork" classification scheme developed by American astronomer Edwin Hubble in the 1920s. According to this system, galaxies come in three basic types: elliptical (*represented by the handle of the fork at right*), spiral (*shown as prongs*) and irregular (*shown below at left*). The smallest galaxies, known as dwarfs, have their own uncertain taxonomy.

Within each of the types are subtypes that depend on the details of the galaxy's shape. Going from the top of the tuning fork to the bottom, the galactic disk becomes more prominent in optical images and the central bulge less so. The different Hubble types may represent various stages of development. Galaxies start off as spirals without bulges, undergo a collision during which they appear irregular, and end up as ellipticals or as spirals with bulges.

—*G.K.* and *F.v.d.B.*

IRREGULARS

M82
Irregular

DWARF TYPES

M32
Elliptical

VII Zw 403
Blue Compact

Small Magellanic Cloud
Irregular

Leo I
Spheroidal

ELLIPTICALS

M89
E0

M49
E4

M110
E5

M84
S0

BARRED SPIRALS

NGC 660
SBa

NGC 7479
SBb

M58
SBc

NORMAL SPIRALS

NGC 7217
Sa

NGC 4622
Sb

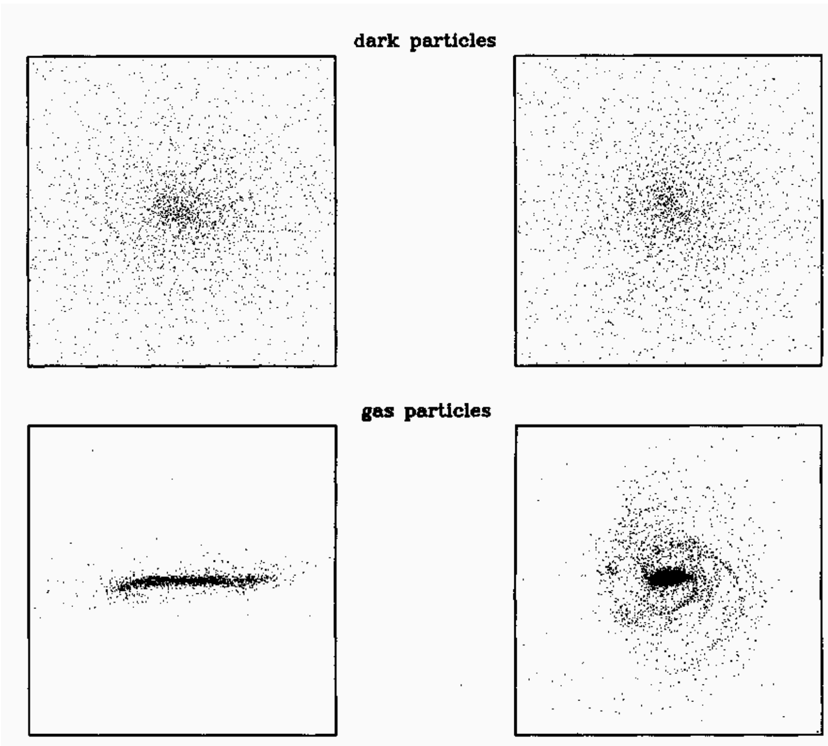
M51
Sc

N. A. SHARP/NOAO/AURA/NSF (*M82*); B. KEEL/HALL TELESCOPE/LOWELL OBSERVATORY (*M32*); R. SCHULTE-LADBECK/U. HOPP/M. CRONE/*ASTROPHYSICAL JOURNAL* (*blue compact dwarf*); NOAO/AURA/NSF (*Small Magellanic Cloud*); DAVID MALIN, © ANGLO-AMERICAN OBSERVATORY (*Leo I*); NOAO/AURA/NSF (*M89, M49, M110, M84*); R. BRANCH/R. MILNER/A. BLOCK/NOAO/AURA/NSF (*NGC 660*); A. BLOCK/NOAO/AURA/NSF (*NGC 7479*); F. CIESLAK/A. BLOCK/NOAO/AURA/NSF (*M58*); B. KEEL/R. BUTA/G. PURCELL/CERRO TOLOLO INTER-AMERICAN OBSERVATORY, CHILE (*NGC 7217*); G. BYRD/R. BUTA/T. FREEMAN/NASA (*NGC 4622*); NASA/STSCI/AURA (*M51*)

Basic Idea:

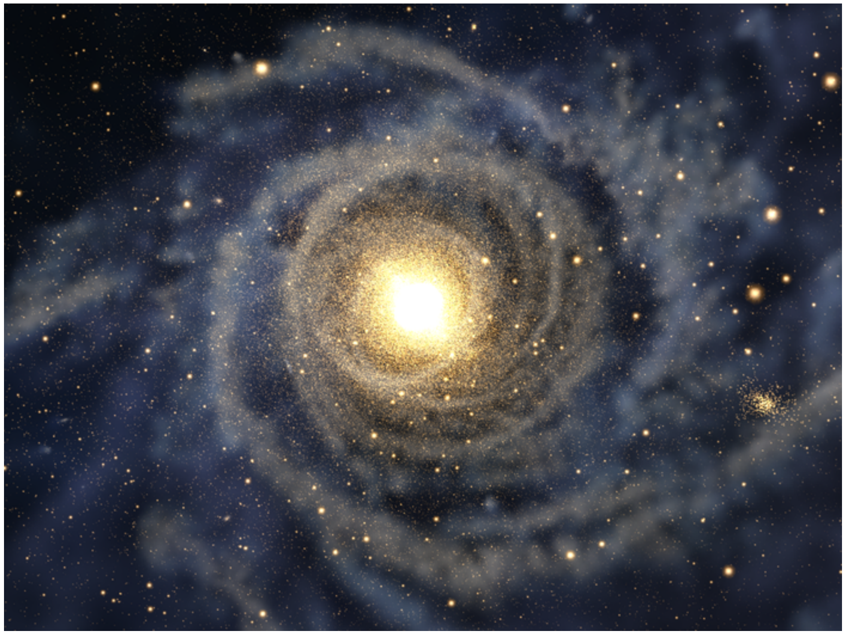
- "Holistic" simulation of galaxy, from initial density fluctuation
- To understand the origin of the variety of galaxies

Katz and Gunn 1992



- Dark Matter + gas + stars
- DM, star: particles gas :SPH particles
- 10^4 particles, Cray YMP 500-1000 hours
- mass resolution : 10^7 solar mass

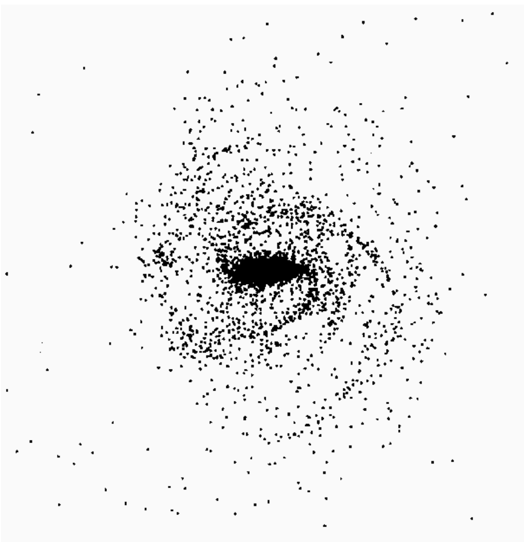
Saitoh et al. 2005

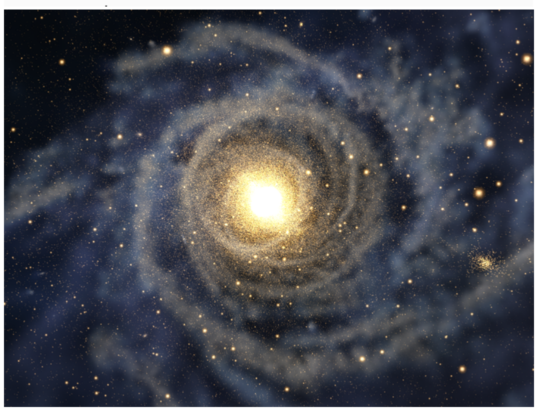


animation

- Dark Matter + gas + stars
- DM, stars: particles gas:SPH particles
- 2×10^6 particles, GRAPE-5 $\sim$ 1 year
- mass resolution : 10^4 solar mass

What gain from improved resolution?





- Not much?
- Important things: improved parametrization of “micro-physics”, such as star formation mechanism, energy input from supernovae.

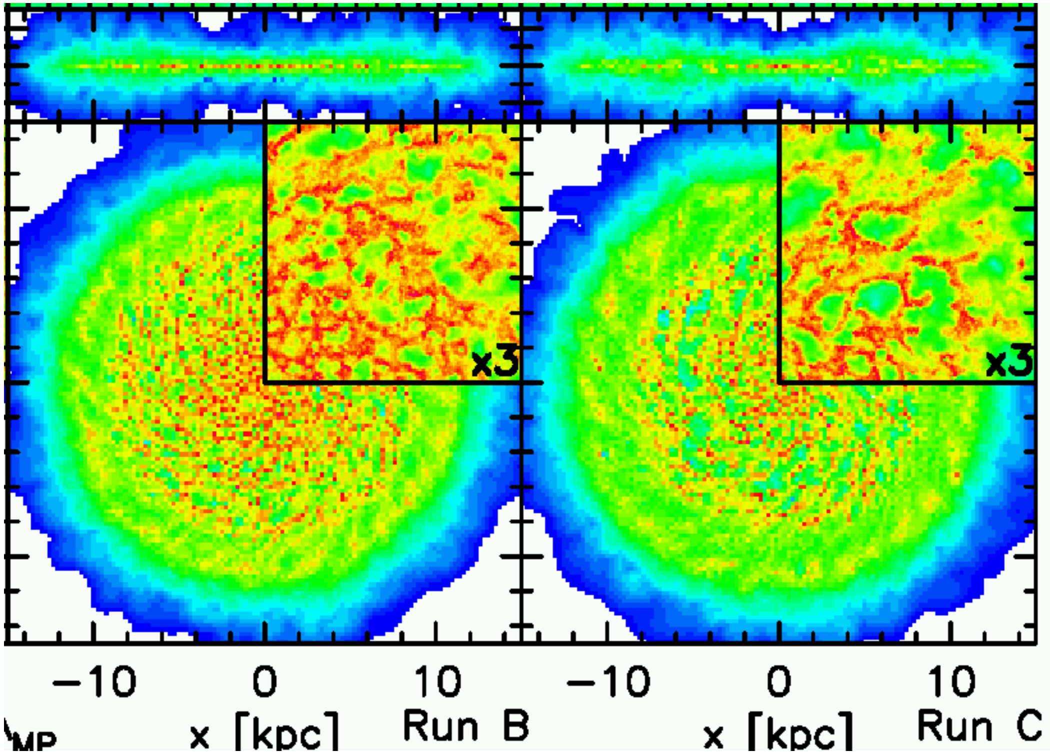
Modeling star formation

- **Minimum need for star formation modeling: : 10^{-4} solar mass**
- **What we can do now: : 10^3 solar mass (10^7 times too large)**
- **Need some way to form stars**
  - **Usual model: if interstellar gas is dense and cold enough, part of it will become stars in appropriate timescale.**
  - **three free parameters**
  - **The structure of the galaxy depends on these parameters**
- **Similar problem on supernovae.**

What resolution do we need?

- We will know when we reach there....
- If mass of SPH particles is more than that of GMCs, clearly we are not doing things right.
- Theoretically, if we have sufficient resolution, we can just change all mass to stars (that is what the nature does).
- We are approaching there.
- One or two orders of magnitude more?

Saitoh et al. 2007



Changed the star formation timescale by a factor of 15

little difference in the result

(In low-resolution calculation, the galaxy would have exploded.)

Some more animations

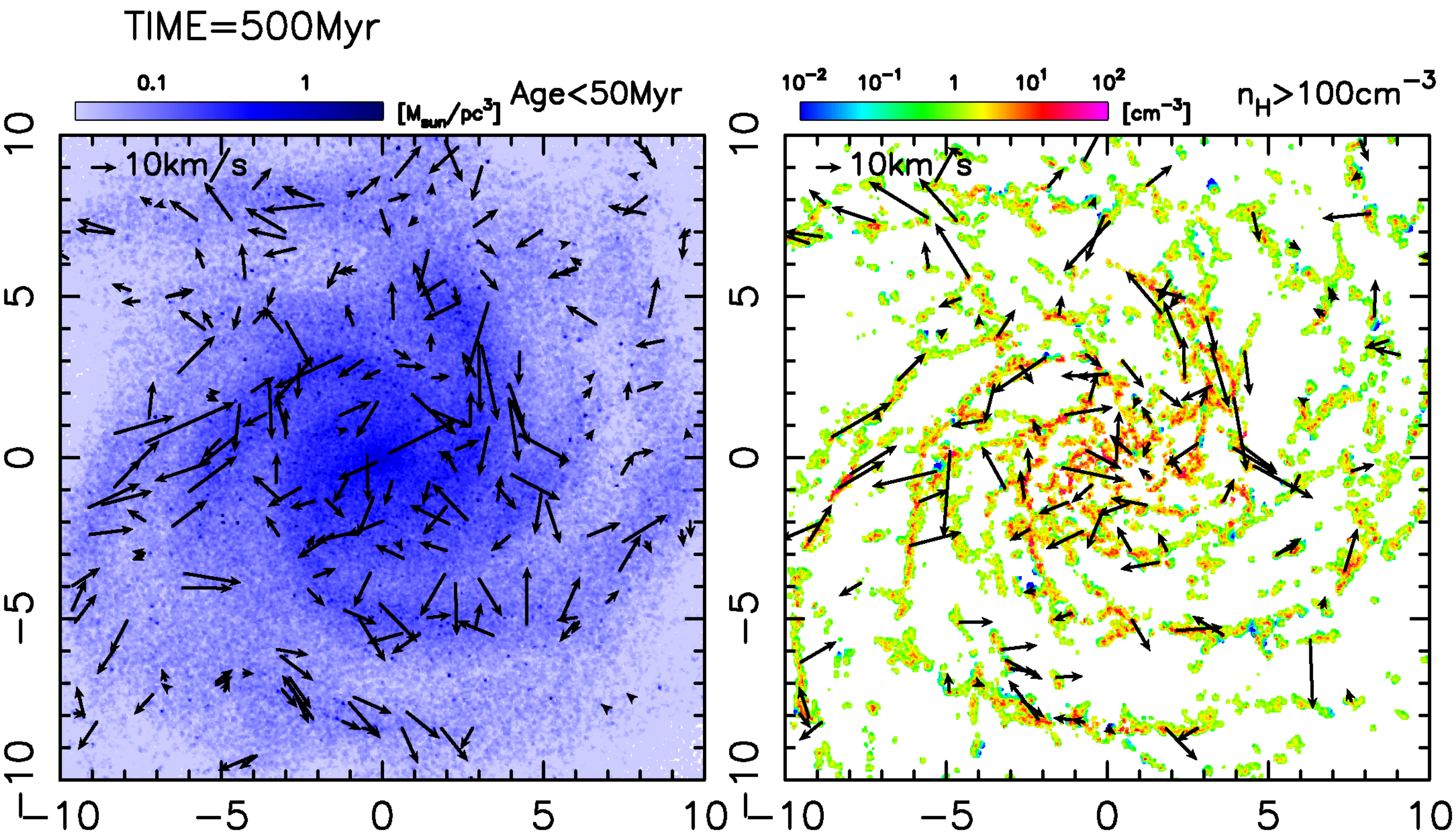
Star formation with SPH

Large scale structure formation with AMR

Galactic disk

animation (Baba et al 2009) 1 2

Spiral structure and deviation from the circular motion

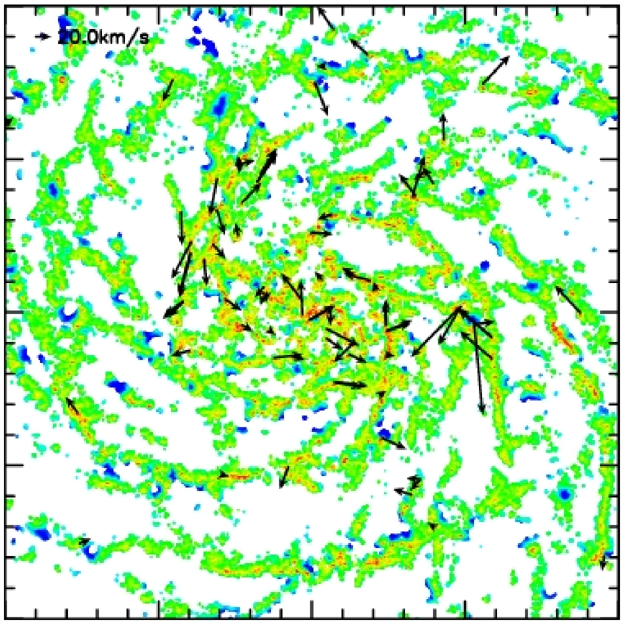


Left: distribution of stars Right: cold gas

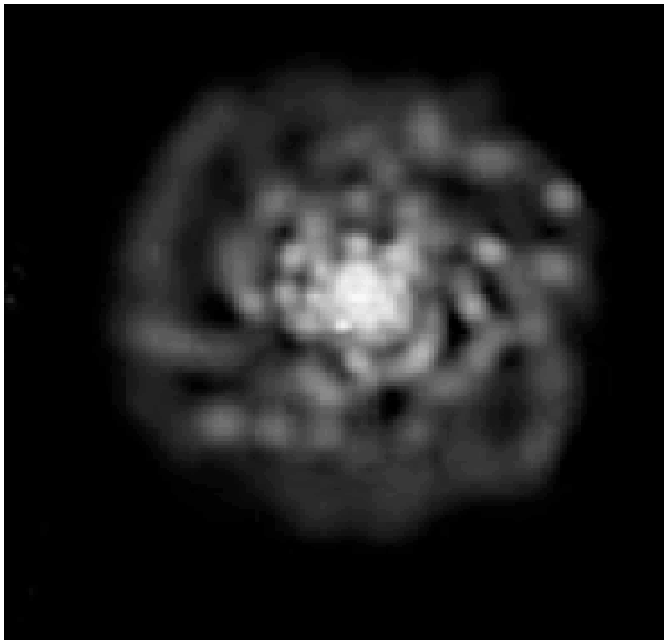
Simulation details

- **Allow gas to become cold and dense**
- **Requires large number of SPH particles**
- **Parallel SPH code ASURA (Saitoh) on Cray XT4@NAO**
- **10pc softening ($\leftarrow$ 500pc)**
- **Gas can cool down to 10k ($\leftarrow$ 10^4K)**
- **$3000M_\odot$ ($\leftarrow$ $10^5 M_\odot$)**

High-resolution model and observation



Low-resolution model and observation



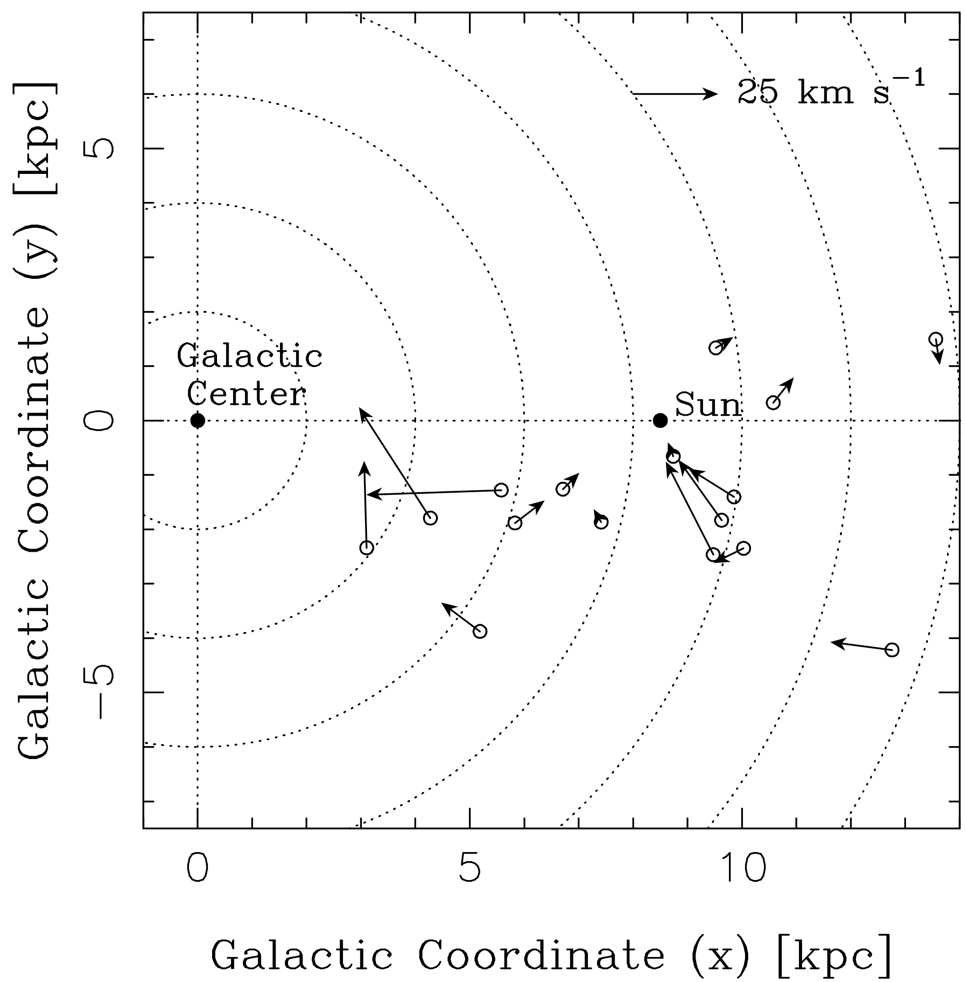
Results from high-resolution simulations

- Star-formation is regulated by large-scale dynamics.
- Observed (multi-arm) spirals can be explained by transient, but recurrent arms.
- These results are robust. Independent of assumption on microphysics such as star-formation timescale.

Observation of Milkyway spiral arms (VLBI)

- **2006: Xu et al, Science 311, 54**
- **Nov 2008: Burst of results from VLBA**
- **Several data from VERA**

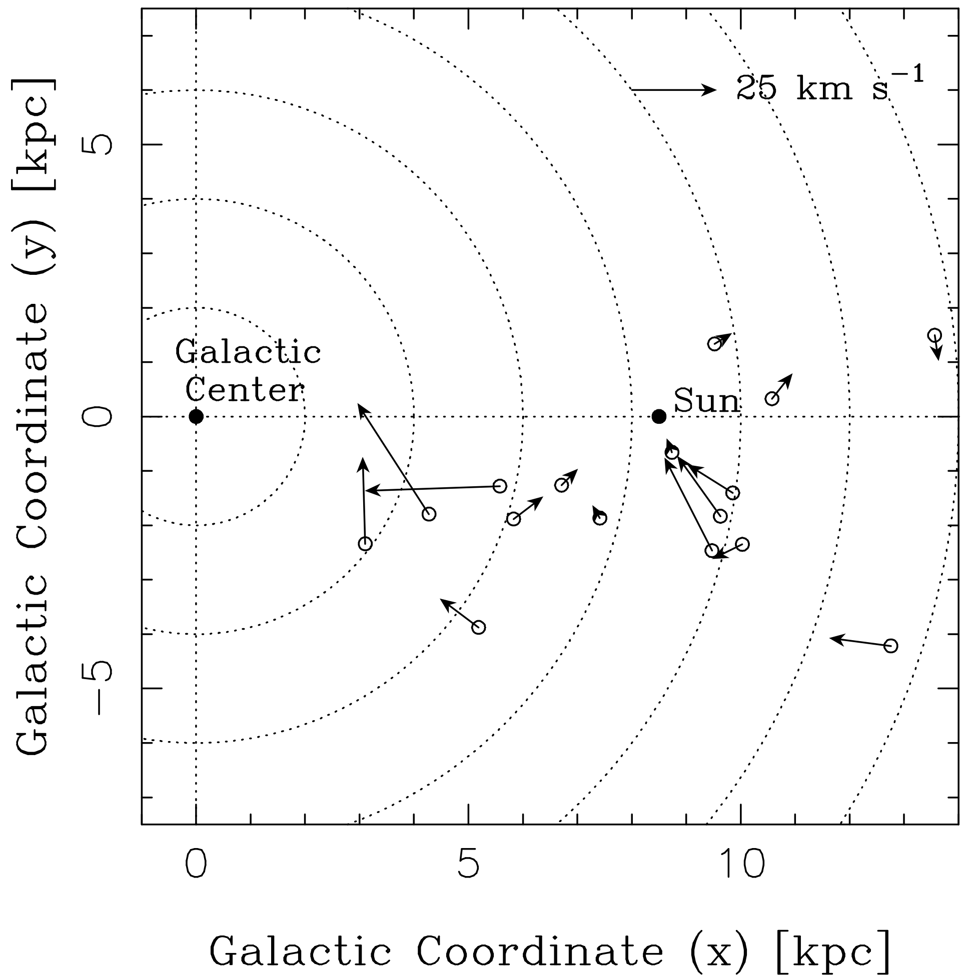
(Compiled by Dr. Asaki)



Observation of Milkyway spiral arms (VLBI)

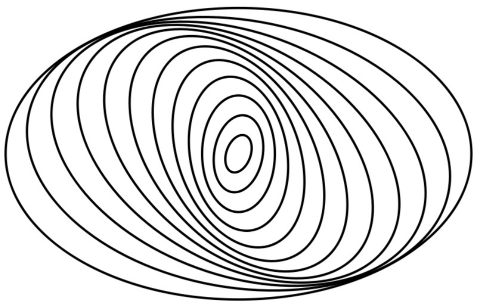
- **Large non-circular motion ($\sim$ 30km/s)**
- **Many data points shows inward motion and counter rotation**
- **Some signs of spacial correlation?**

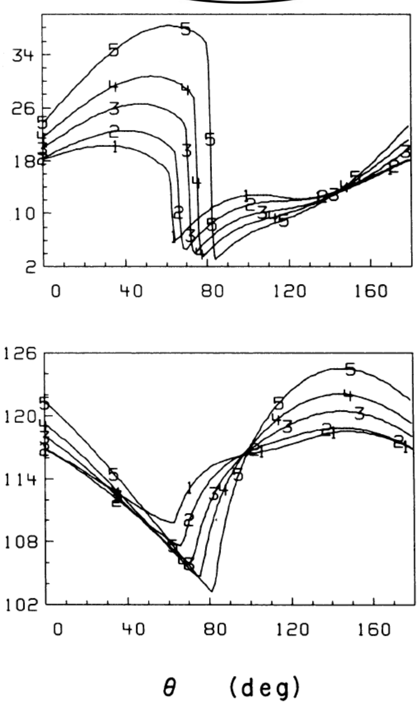
How these motions are induced?



What you learn from textbooks

Stationary density wave



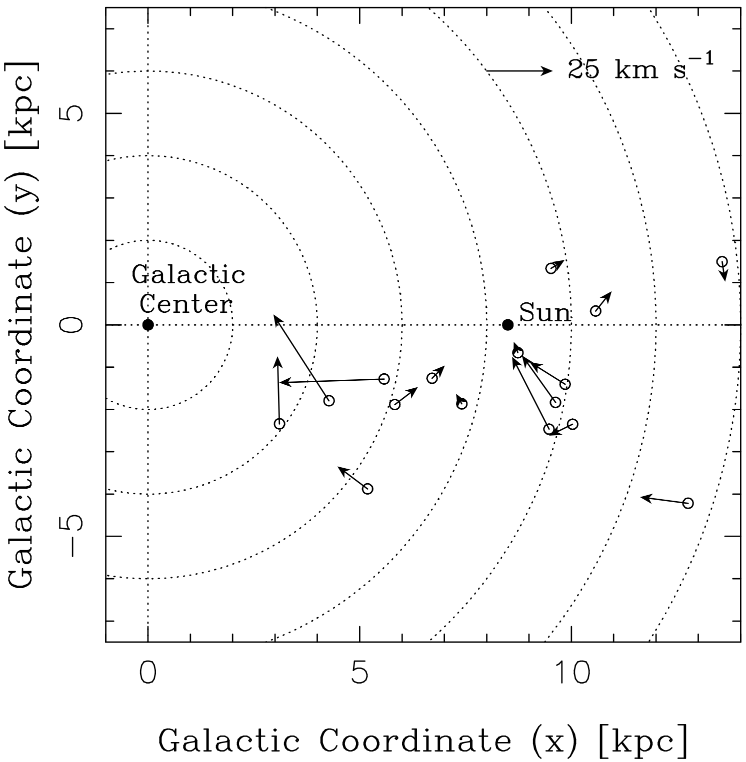


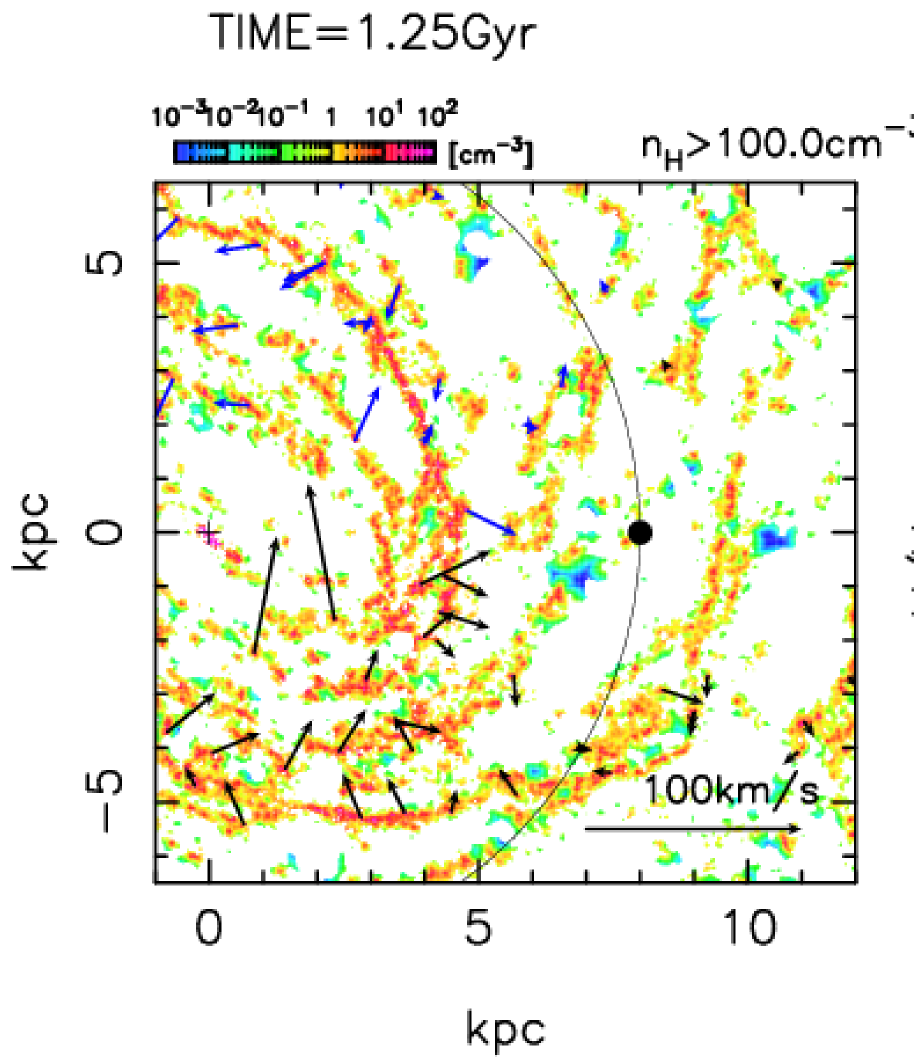
- Spiral arms are not material arms, but density waves
- gas is compressed when it passes through the botton of the potential well, and form stars there
- It is very difficult to generate non-circular velocity > 10km/s

Quite different from both observation and simulation

Comparison

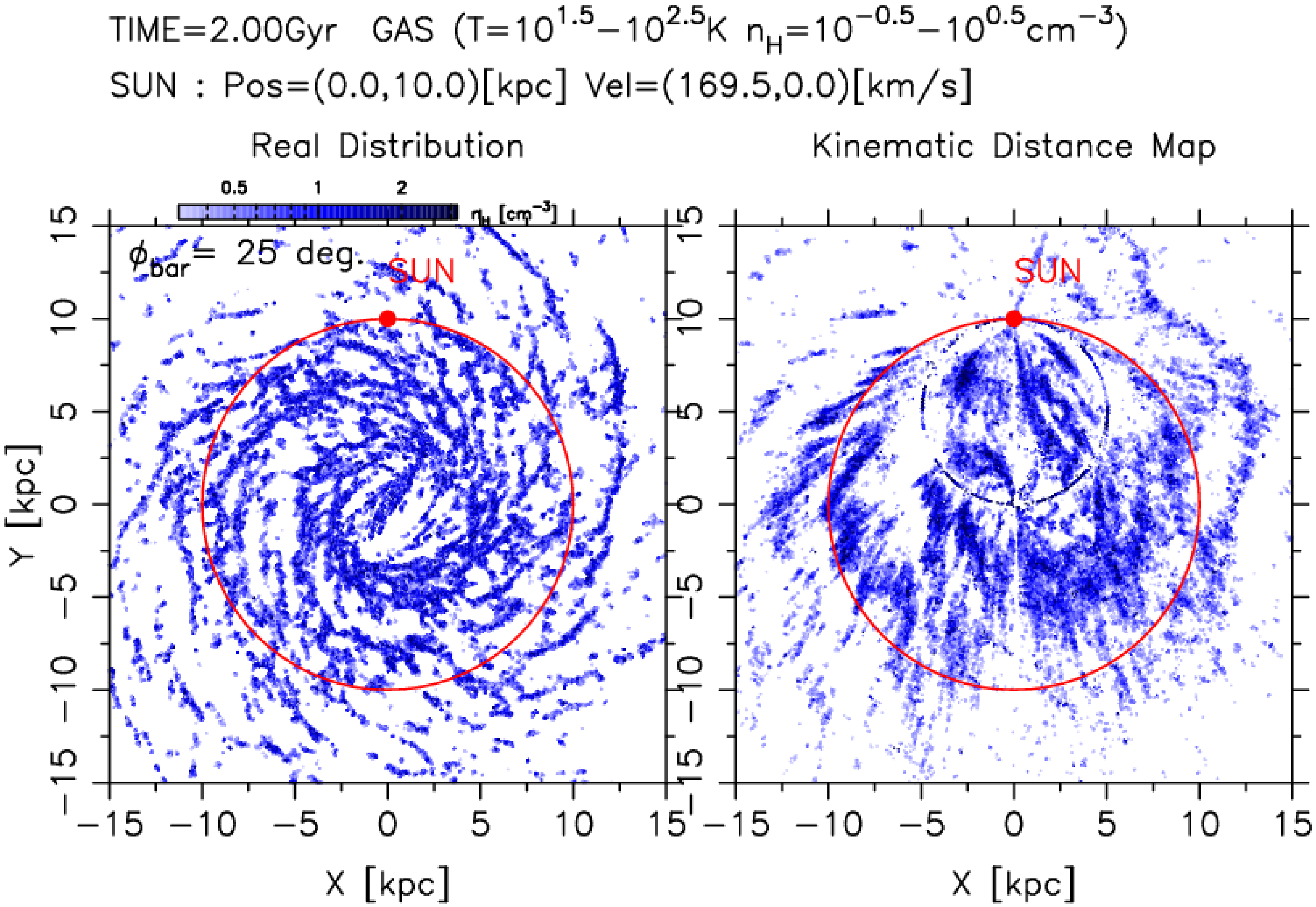
between observation and simulation





Look sort of similar?

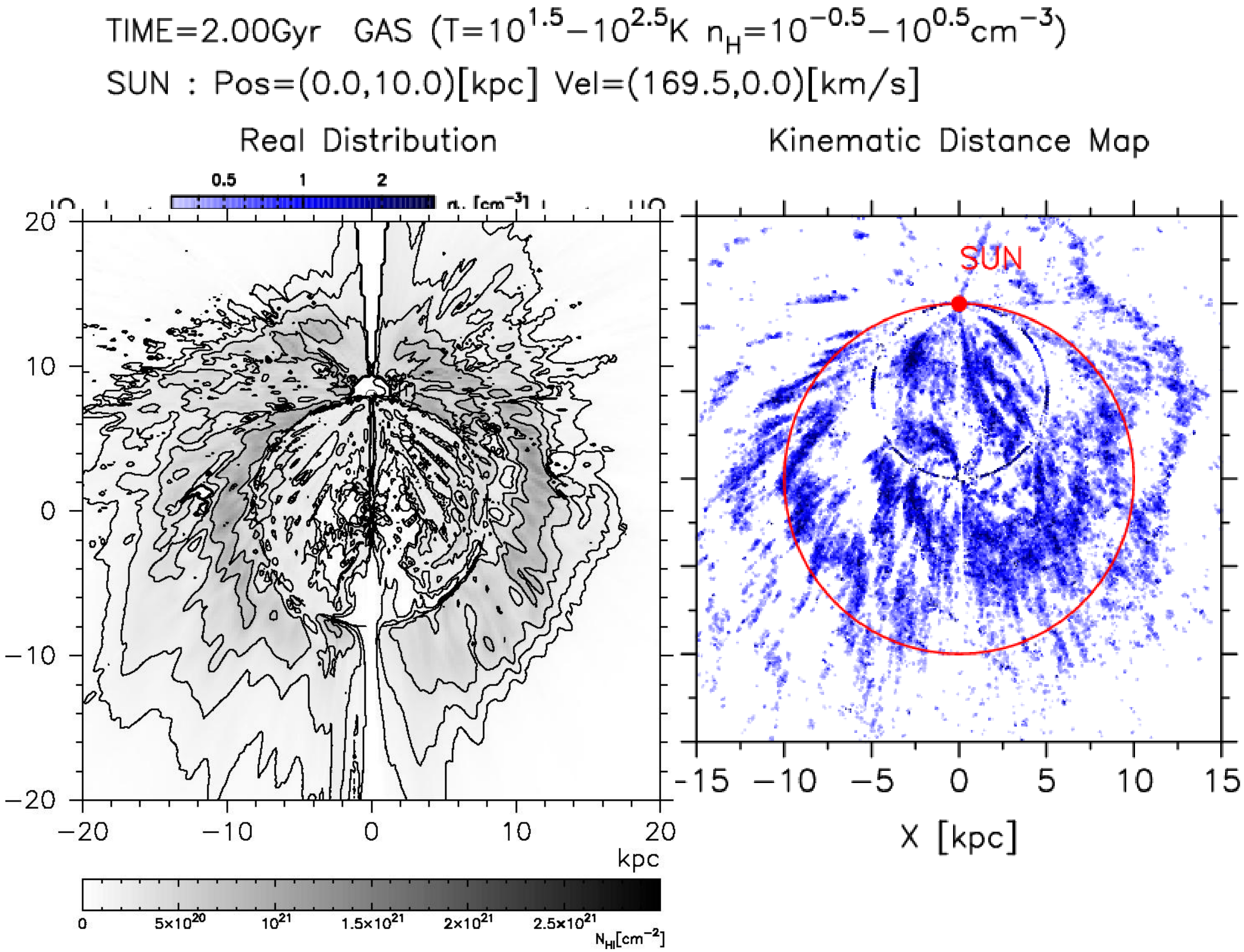
Kinematic distance



Left: Actual distribution Right: Kinematic distance

Quite different...

Kinematic distance

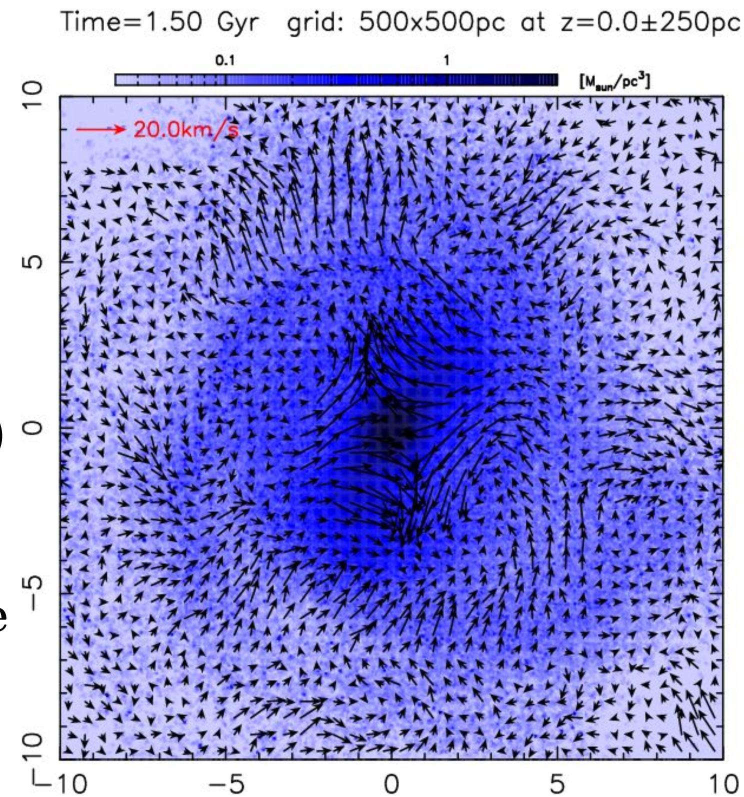


Left : HI observation (Nakanishi and Sofue 2003)

Lots of similar structures

Motion of stellar spiral

average non-circular motion of all stars



- **Spiral arms are real (material) arms, not waves**
  - **Old stars have fairly large non-circular motions**
  - **few-kpc scale structures**

Summary on SPH simulation of spiral arms

- In high-resolution SPH simulations, spiral arms naturally form
- Spiral arms are not stationary, but transient and recurrent
- "VLBI" and "HI" observations of simulation results look very similar to those of Milky way.

Simulation of pure stellar spiral

(Fujii et al. 2010)

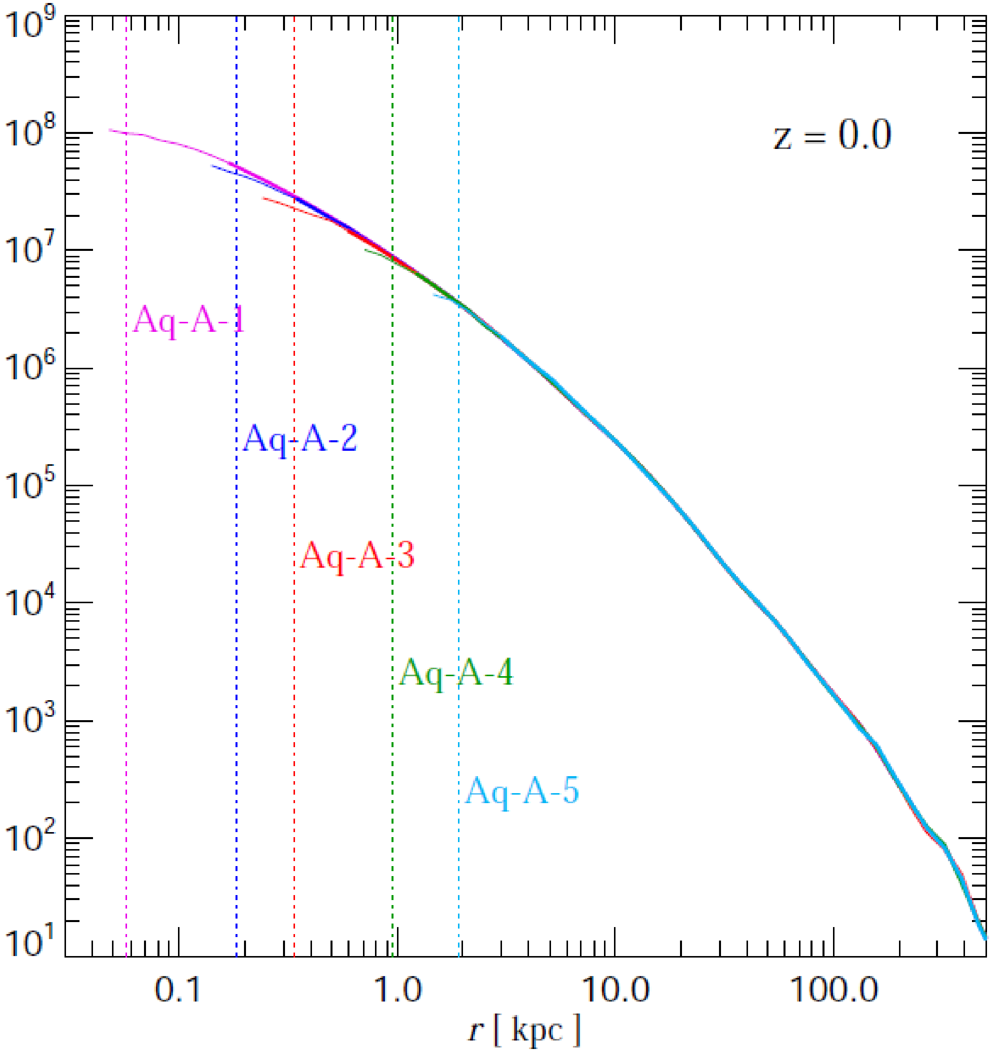
animation a1

animation a2

animation b1

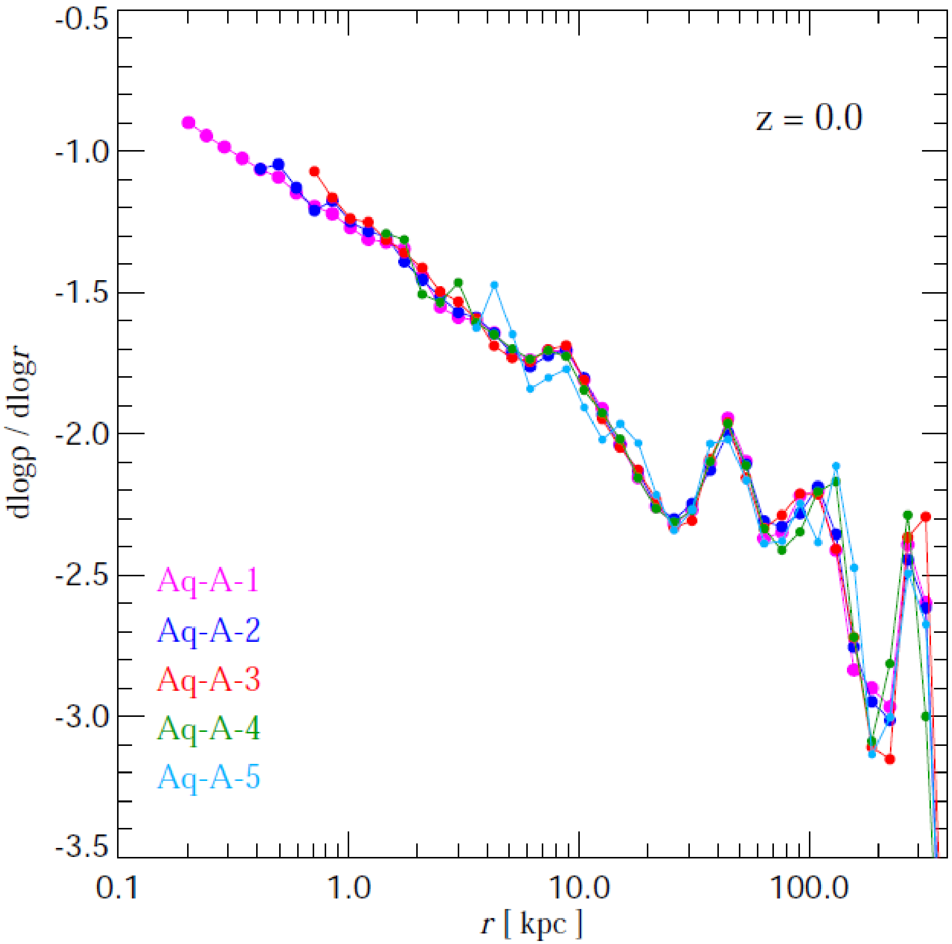
- **Stable against radial mode (a1, a2)**
- **Spiral arms form**
- **They seem to be maintained for very long time**

Structure of DM halo



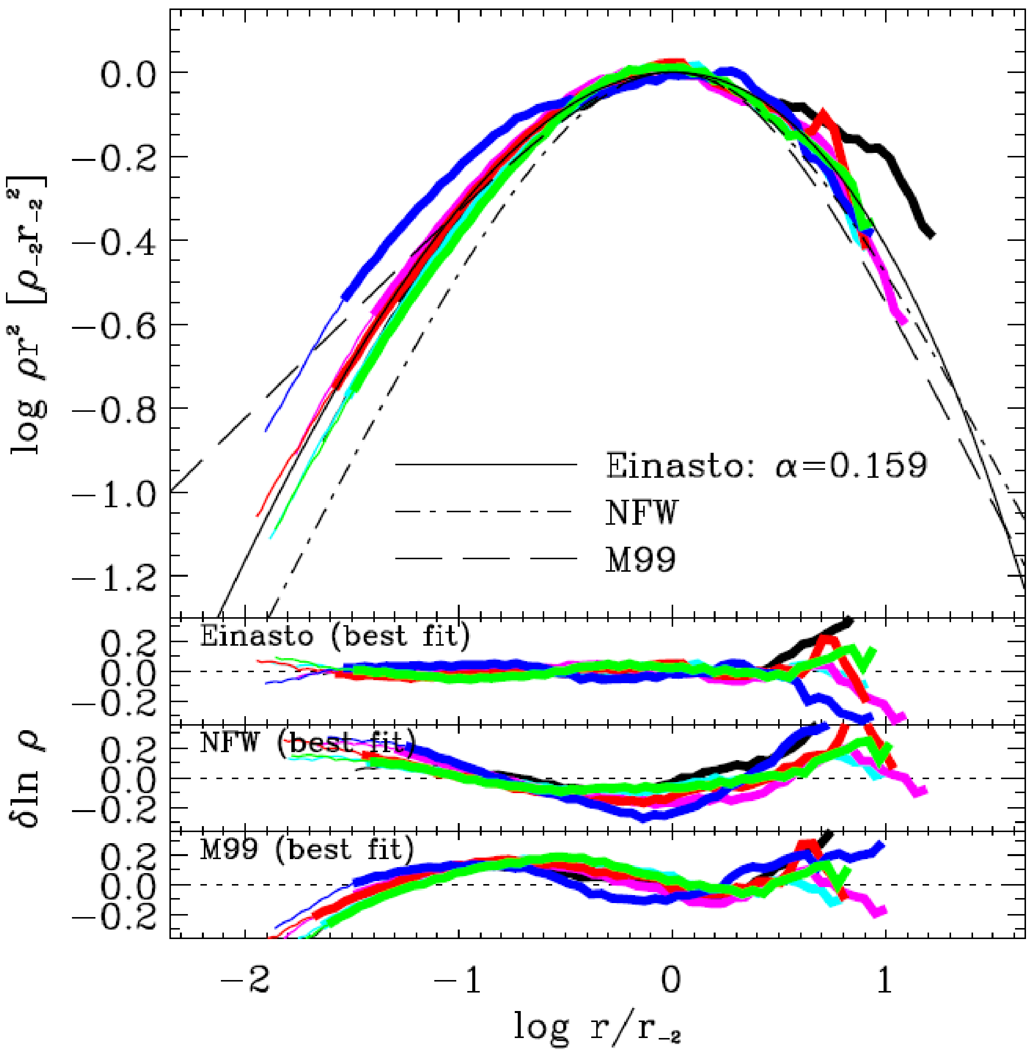
Highest-resolution run done so far : Springel et al (2008)
Mass resolution changed by three orders of magnitude
Sign of convergence?

Density Slope



Not a single power...

NFW and other profiles



Navarro et al. 2008 (not accepted yet?)

NFW: $\dfrac{1}{r(1+r)^2}$

Moore99:

$$\frac{1}{r^{1.5}(1+r^{1.5})}$$

Einasto:

$\exp[(-2/\alpha)(r^\alpha - 1)]$

Historical perspective

- **NFW (1996): Density structure of DM halo is expressed by the NFW profile, independent of initial power spectrum or cosmological parameters.**
  - **Simulation used 10-20k particles/halo**
  - **Two-body relaxation affected the central structure**
- **Fukushige and JM (1997) : Central slope is steeper for 1M particle simulation.**
- **Moore et al. (1999) Slope is steep with 3M particles, Moore99 profile**
- **Now: $>$ 1B particles: NFW fit is not good, but slope becomes shallower at the very center.**

To summarize...

- Central structure of DM halos is quite strange
- It is not a single power-law cusp. In simulations, slope becomes shallower as we improve resolution
- Numerical result is "reliable"

Naively:

- Initial density fluctuation is power-law
- No characteristic scale other than particle mass

So why not a single power?

Theoretical difficulties

- **We cannot understand numerical result**
  - **Well, though numerical result is "reliable", it is pretty hard to believe...**
  - **Result depends strongly on resolution**
- **Isn't there something wrong in our simulations?**

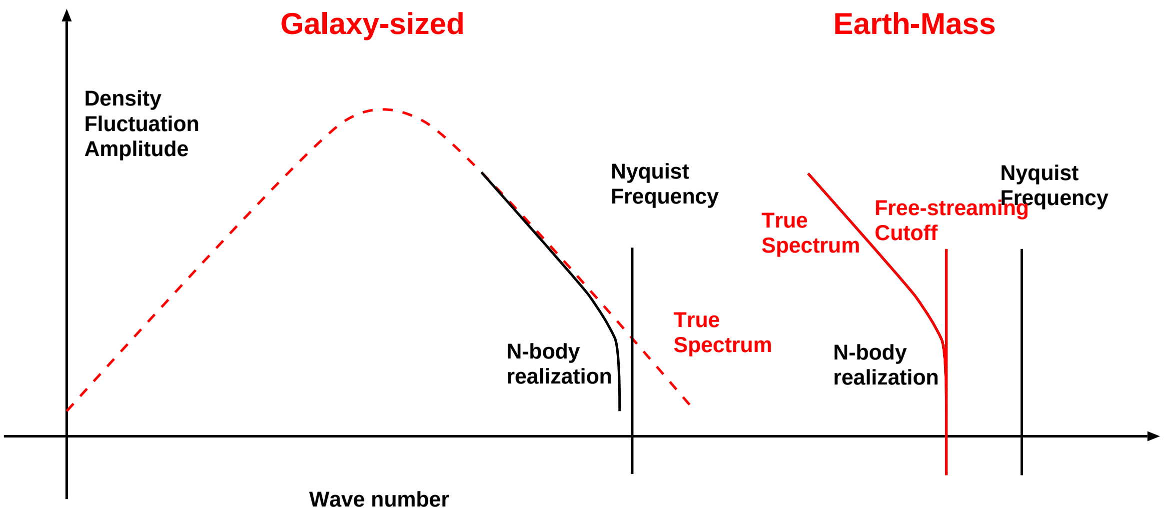
Fundamental problem?

Cosmological N-body simulations are not "correct" simulation of collisionless N-body systems.

Construction of initial condition:

1. **Place particles uniformly**
2. **Add random fluctuations to position and velocity of particles according to the power spectrum of density fluctuation (cutoff: order of interparticle distance)**

Smallest structures



Small halos form first in CDM = first "halos" contain ~ 10 particles

Common belief: Since the process of hierarchical merging determines the structure, the structure of halos in the smallest scale should not affect the result.

No one has confirmed this belief, though.

Can we confirm this belief?

In principle, it is easy to confirm.

- **Fix the cutoff scale of initial power spectrum**
- **vary the mass resolution**

Practical difficulties:

- **Requires huge amount of computing resources.**
- **Not clear if the result is of any scientific interest...**

Smallest-mass halos

Actually, it might have some scientific importance:

Cold dark matter has free-streaming cutoff

Typical mass scale (depending on the nature of DM particle): Earth mass

Smallest halo: Earth mass, size 100AU.

- **Structure of these halos?**
- **Do they survive in galaxies?**

Importance of the structure of smallest halos

Primary question: Have they survived?

Important for detection of CDM particles

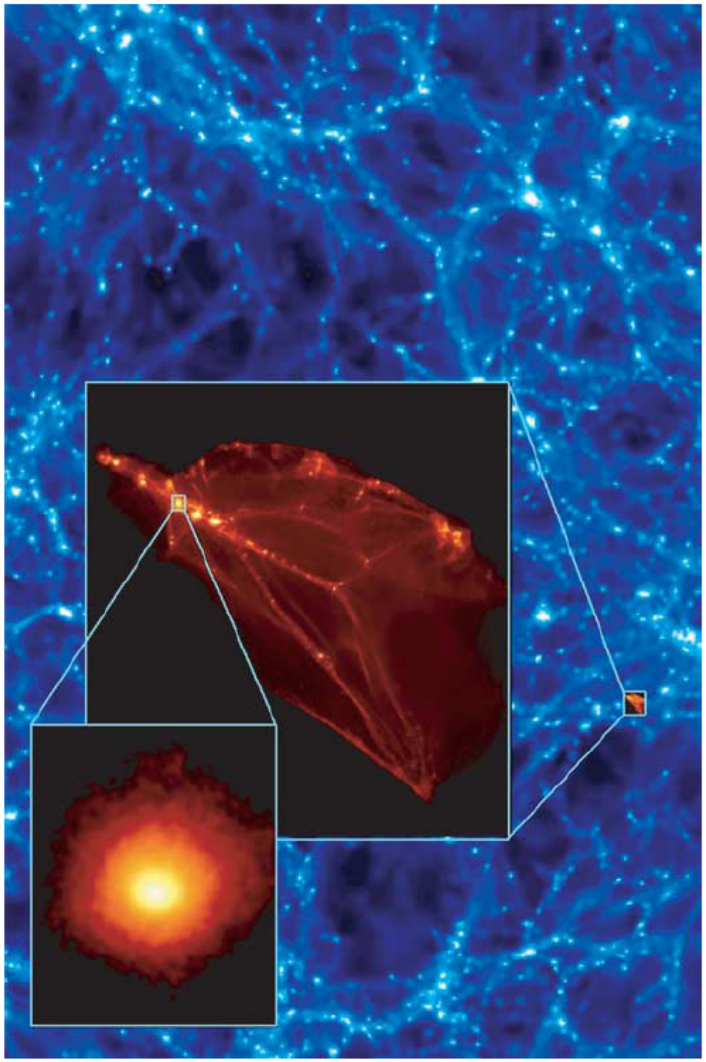
- **Direct detection: Large fluctuation in DM density**
- **Annihilation γ-ray: If survived, they dominate the flux.**

Processes which affect the survival

- Absorbed by larger halos
- Disrupted by potential of larger halo
- Disrupted by encounters with stars

Central structure is critical.

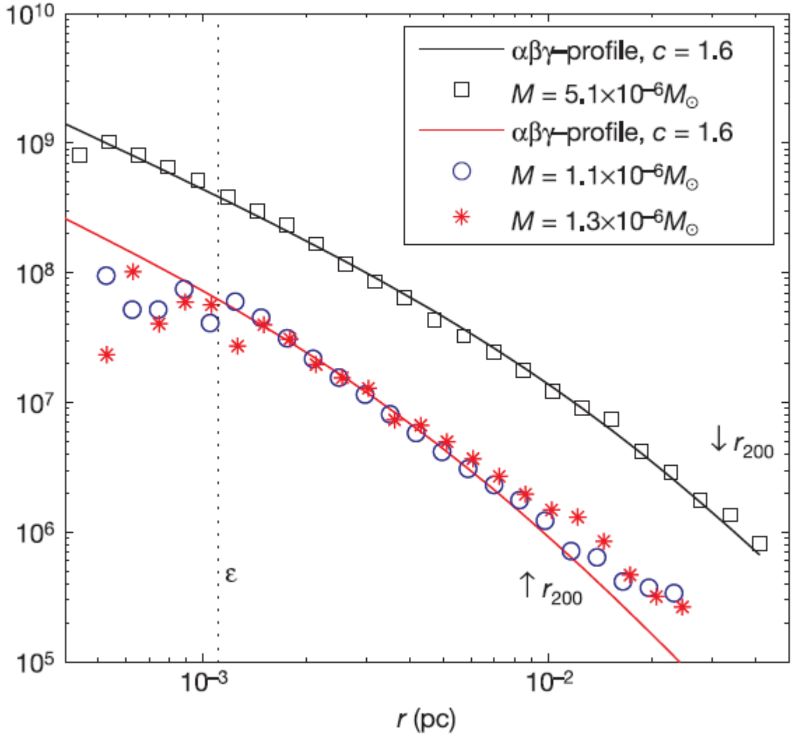
Previous work(s)



Diemand et al. 2005, Nature 433, 389

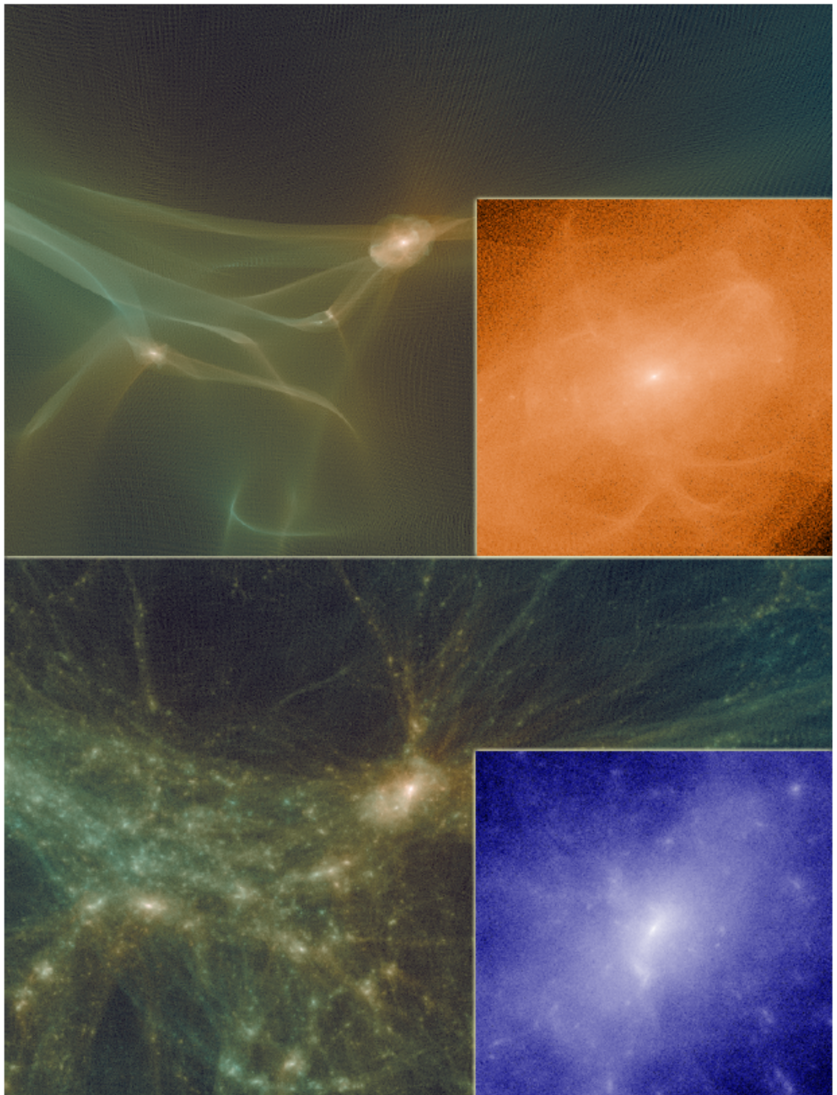
Usual Cosmological simulation
10^4 particles for Earth-mass halo

Density Profile



- **Essentially same result as NFW(1996)**
- **Quite natural because of low resolution**
- **Most likely completely wrong**

Ishiyama et al 2010



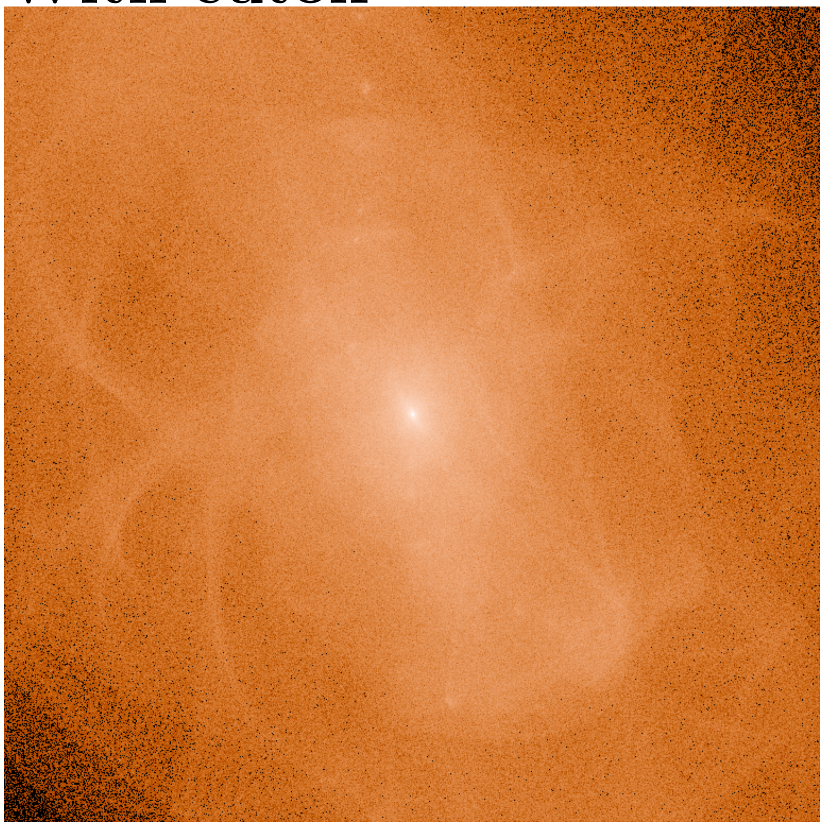
Ishiyama et al., 2010 (arXive1006.3392)

100 times more particles compared to Diemand et al.

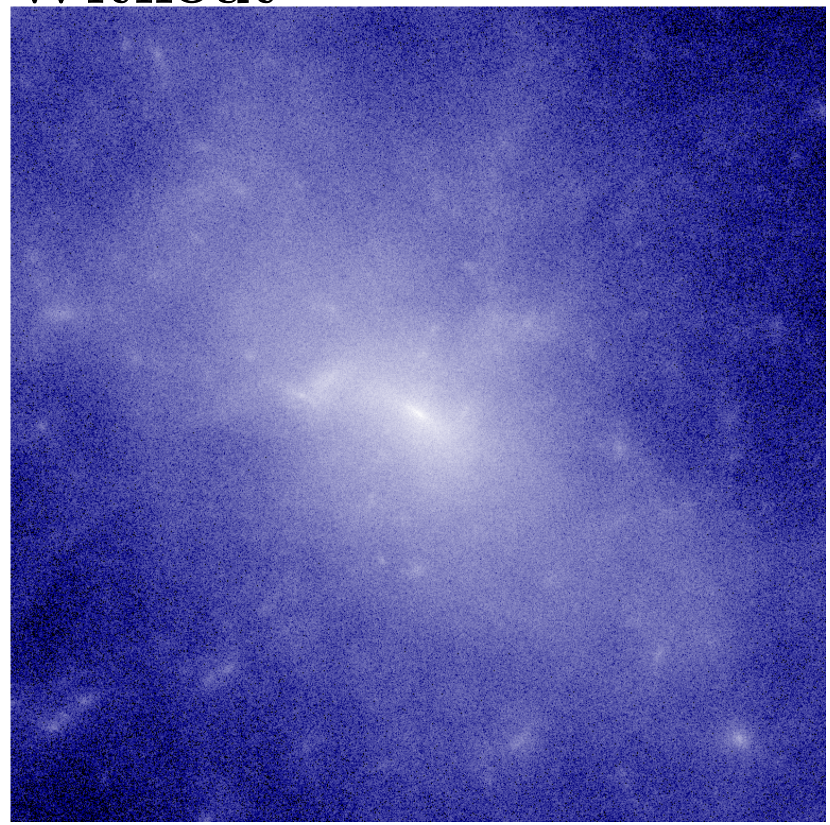
- Top: with cutoff
- Bottom: no cutoff

Halos

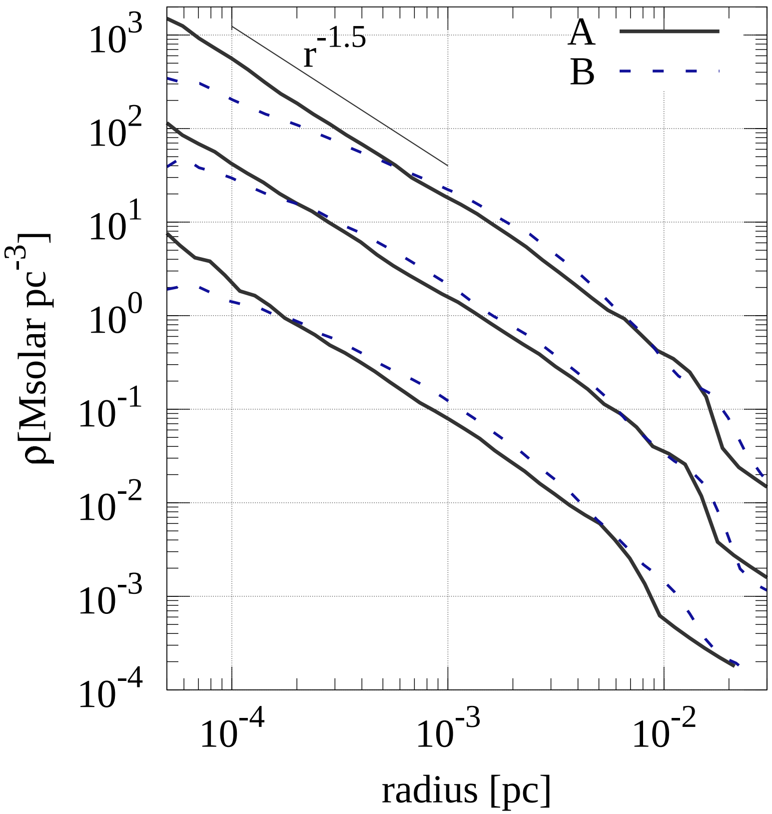
With cutoff



Without



Structure of microhalos



Solid: with cutoff. quite clear single power

Dashed: without cutoff. Similar to galaxy-sized halos.

Earth-mass microhalos have steep, $\rho \propto r^{-1.5}$ cusp

Meaning of -1.5

Annihilation γ-ray flux diverges as $r \to 0$.

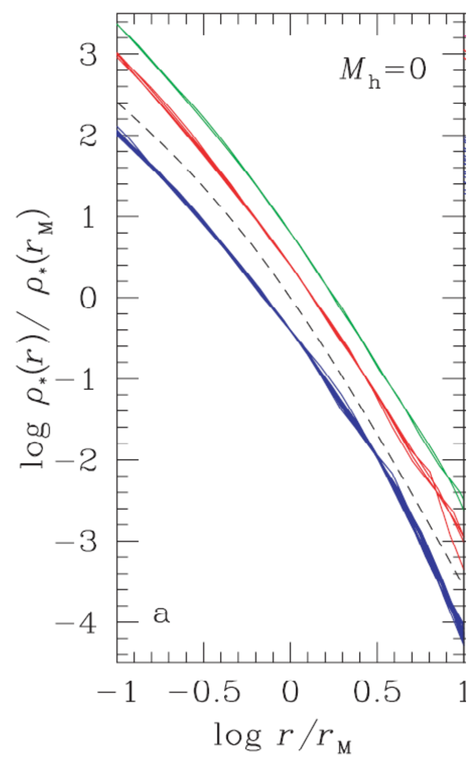
Two questions:

1. **Why -1.5?**
2. **Is there any limit radius?**

Why −1.5?

No real clue yet...

Resent cold-collapse simulations show the same −1.5 slope. (Nipoti et al 2006)



Single power is sort of natural

- **"Cold" initial condition: no limit in the central density**
- **No characteristic scale: result should be a power law?**

Is there any limit radius?

- **“Cold” dark matter still have finite temperature.**
- **Liouville’s theorem**
  — maximum phase space density is conserved (or does not increase): $\sim 10^{15} \mathrm{M}_{\odot} \mathrm{pc}^{-3} (\mathrm{km/s})^{-3}$.

$$\downarrow$$

- **Core radius: $r_{\mathrm{c}} \sim 10^{-5} \mathrm{pc}$**
- **Core density: $\rho_{\mathrm{c}} \sim 2 \times 10^{4} \mathrm{M}_{\odot} \mathrm{pc}^{-3}$.**

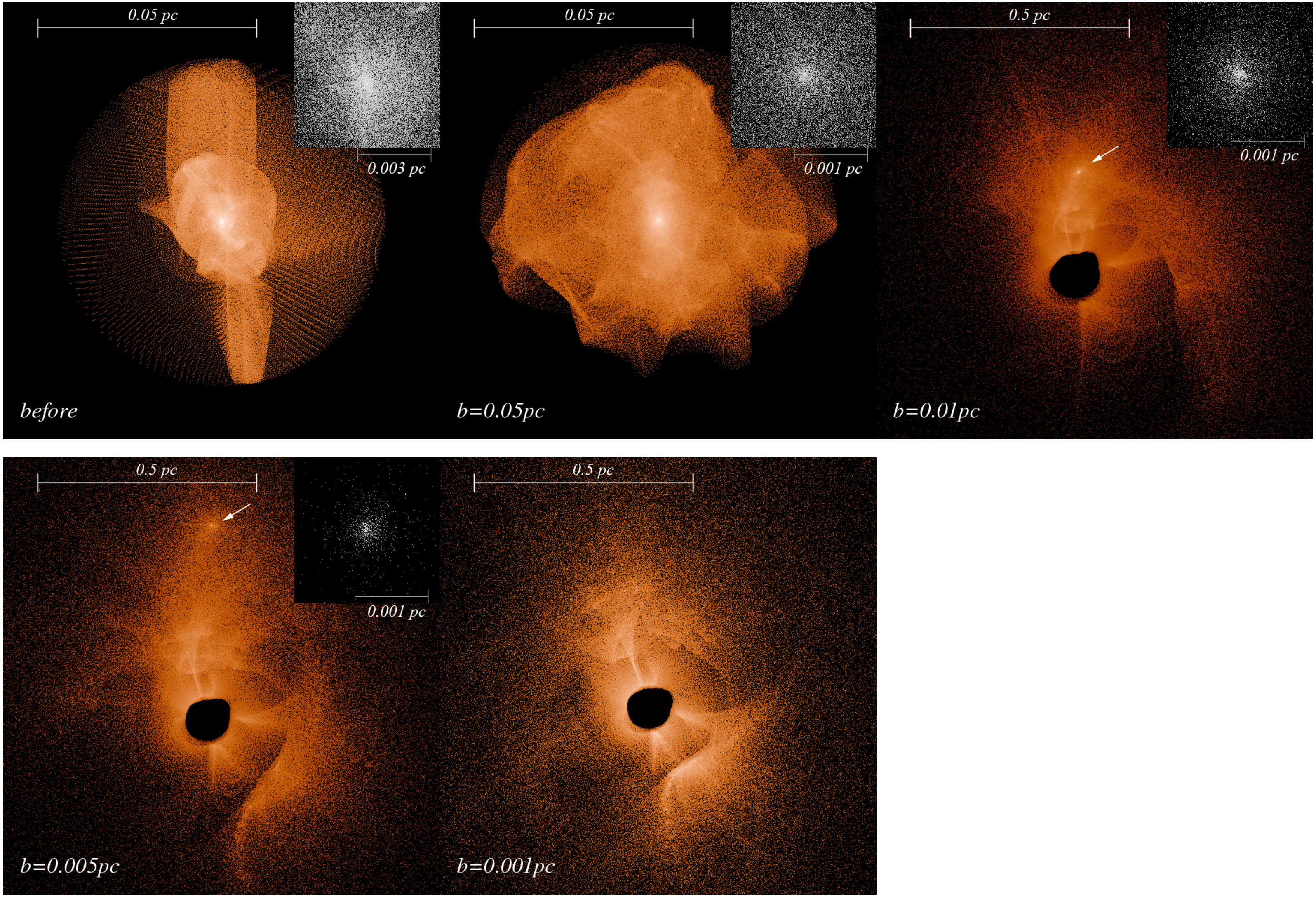
Disruption by tidal fields

In previous studies, microhalos were assumed to have shallow central slope (~ -1.2).

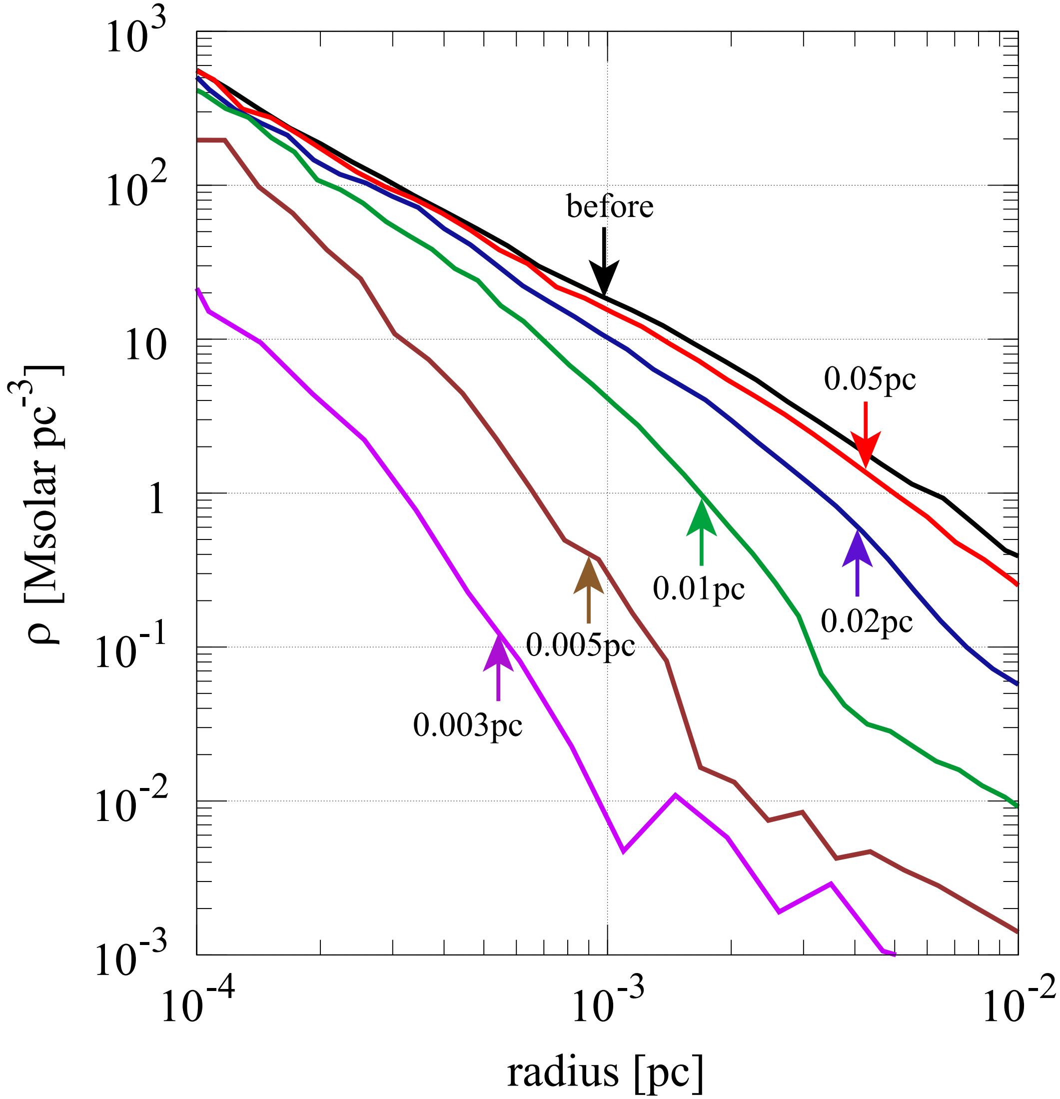
Our high-resolution simulation:

- **Central density is very high — difficult to disrupt**
- **γ-ray flux distribution logarithmic in radius — heavily stripped halos still retain most of luminosity**

Encounters with stars



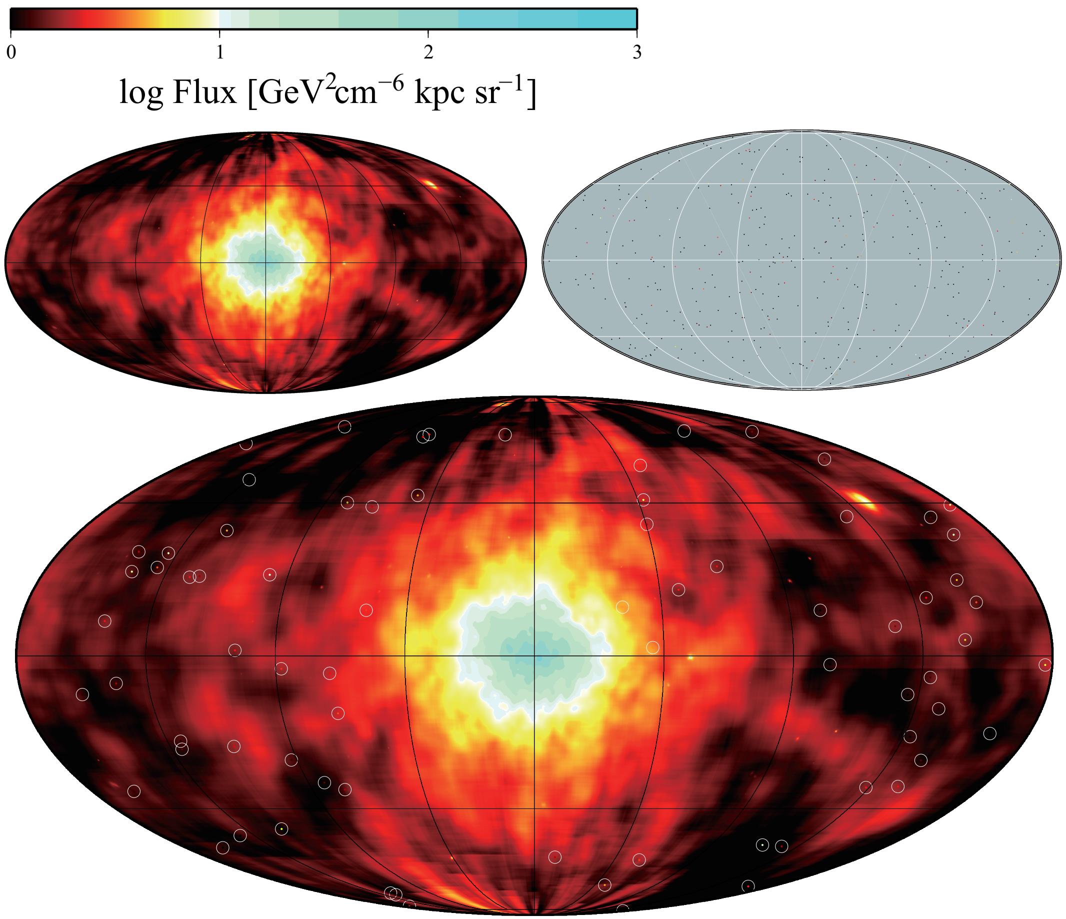
Structure after encounters



Central parts of Halos do survive very close encounters with stars.

Complete disruption requires impact parameter $b = 5 \times 10^{-5}$pc.

γ-ray all-sky map



Top left: Smooth component due to microhalos
Top right: resolvable flux from microhalos (within 1pc)

Theoretically, $r_{\rm tidal} \propto b^{8/11}$.

Nearby microhalos

- distance $\sim$ 0.2pc, core size $\sim 1AU \rightarrow$ image size $\sim$ 1 arcmin
- Proper motion: 300km/s, 0.2 pc $\rightarrow$ $\sim$0.2deg/y
- total flux: $\sim 10^{6}$ of the total galactic flux
- 10-100 times blighter than average background

Detectability by Pulsar timing

Encounter with Pulsars causes variation in the time of arrival.

$$\Delta T = 40 \left(\frac{R}{5000\mathrm{AU}}\right)^{-2} \left(\frac{M}{10^{-6} M_\odot}\right) \left(\frac{t}{10\mathrm{yr}}\right)^{2} \mathrm{ns}.$$

Change in the relative position should show up as the residual of TOA.

Current PPTA timing accuracy: 100ns

Many MSPs are in the direction of GC: High DM density.

PPTA might find microhalo in 10 years.

Summary on microhalos

- **Microhalos (mass $\sim$ earth mass) do survive to the present time.**
- **Their contribution dominates the annihilation γ-ray flux.**
- **Nearest halos might be observed as pointlike sources with extremely large proper motions**
- **Pulsar timing might also detect these halos.**

Summary

- **High-resolution simulation of galaxies are now beginning to reproduce observed spiral structures.**
- **There are lots of yet-to-be-answered questions, on relatively simple and well-defined questions such as the structure of pure DM halo.**

Gravitational Many-Body problem

Equation of Motion:

$$m_i \frac{d^2 x_i}{dt^2} = \sum_{j \neq i} f_{ij} \tag{1}$$

x_i, m_i: position and mass of particle i

f_{ij}: gravitational force from particle j to particle i

$$f_{ij} = G m_i m_j \frac{x_j - x_i}{|x_j - x_i|^3}, \tag{2}$$

G: gravitational constant.

This equation, however, does not tells much about the behavior of the system.

Why not?

- the equation does not have analytic solution
- there are special cases....
  - $N = 2$
  - $N = 3$ from special initial condition
  - Solar-like systems (well....)
  - $N \to \infty$, dynamical equilibrium

On the other hand, we can numerically integrate the equation of motion using computer. Isn't that enough?

Numerical integration

In principle, numerical calculation should be enough.

In practice, it is not.

Reason:

- **Computers are not fast enough**
- **Additional physics**
  - **gas dynamics**
  - **stellar evolution**
  - **...**

Computer power and calculation cost

A naive estimate:

If we have N stars, calculation cost per timestep is N^2.

A 10^8-body system would need a computer 10^8 times faster than a 10^4-body system needs.

Three ways to reduce calculation time

- Improve numerical methods to reduce operation count
- Buy faster computers
- Build faster computers

Improve numerical methods to reduce operation count

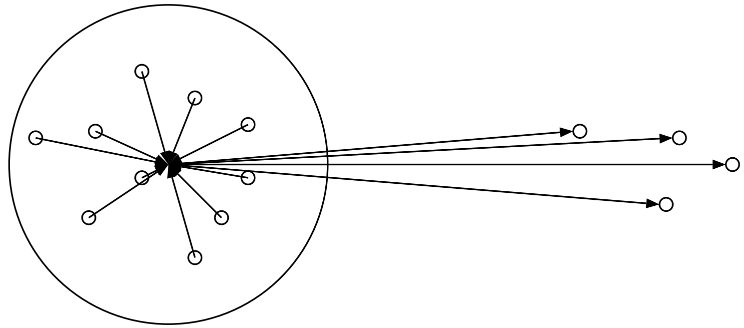
- Better methods for orbit integration
  - high-order integrator
  - symplectic/symmetric methods
  - hybrid
  - ...
- Fast and approximate methods for interaction calculation
  - Tree and FMM

Basic idea for tree method and FMM

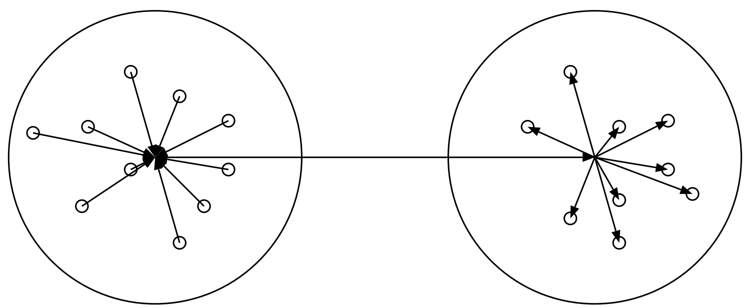
Force from distant particle: Weak

↓

Can't we evaluate many forces at once?



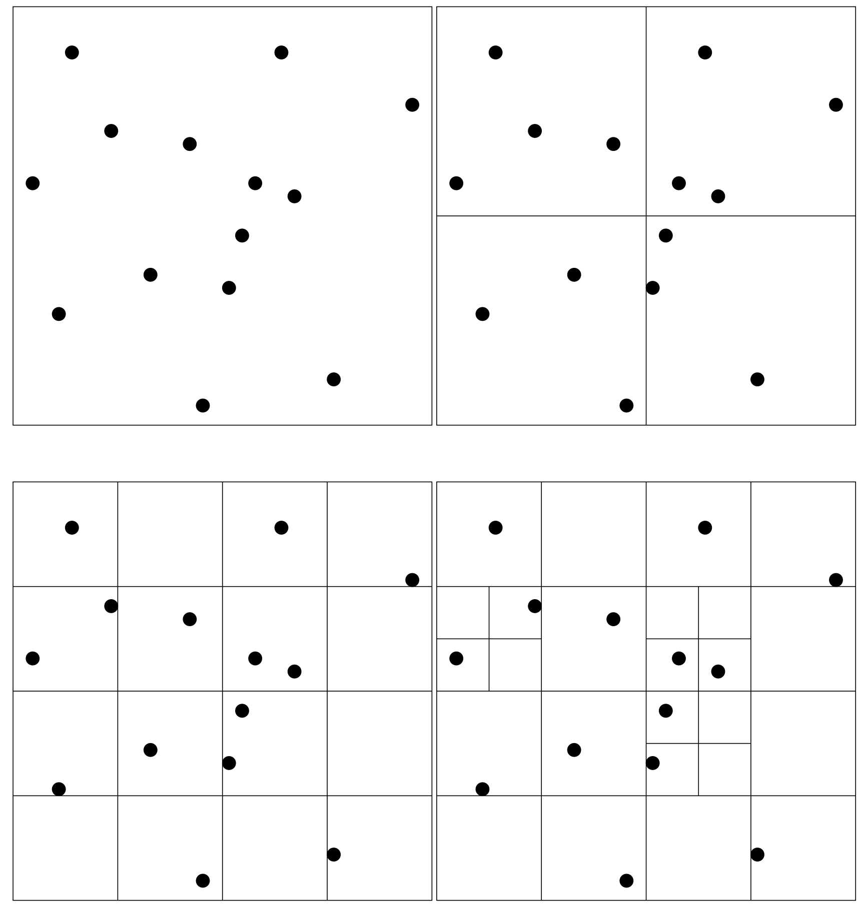
Tree



FMM

- **Tree: aggregate stars which exert the forces**
- **FMM: aggregate both side**

How do we aggregate — Barnes-Hut tree

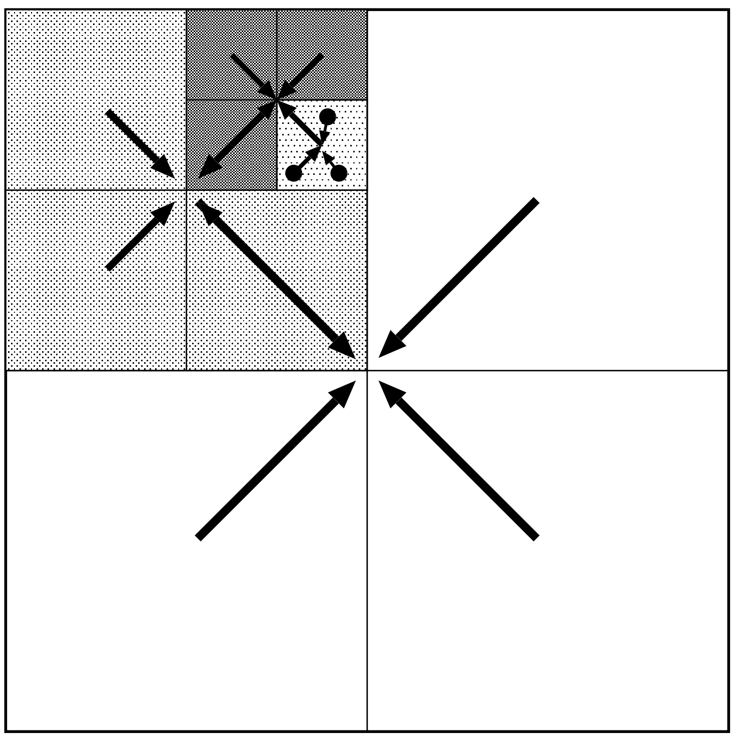


Use tree structure

- First make a cell with all stars in it
- Recursively subdivide the cells to 8 subcells
- Stop if there is small enough stars

Construction of the multipole expansion

Form the expansion for cells.

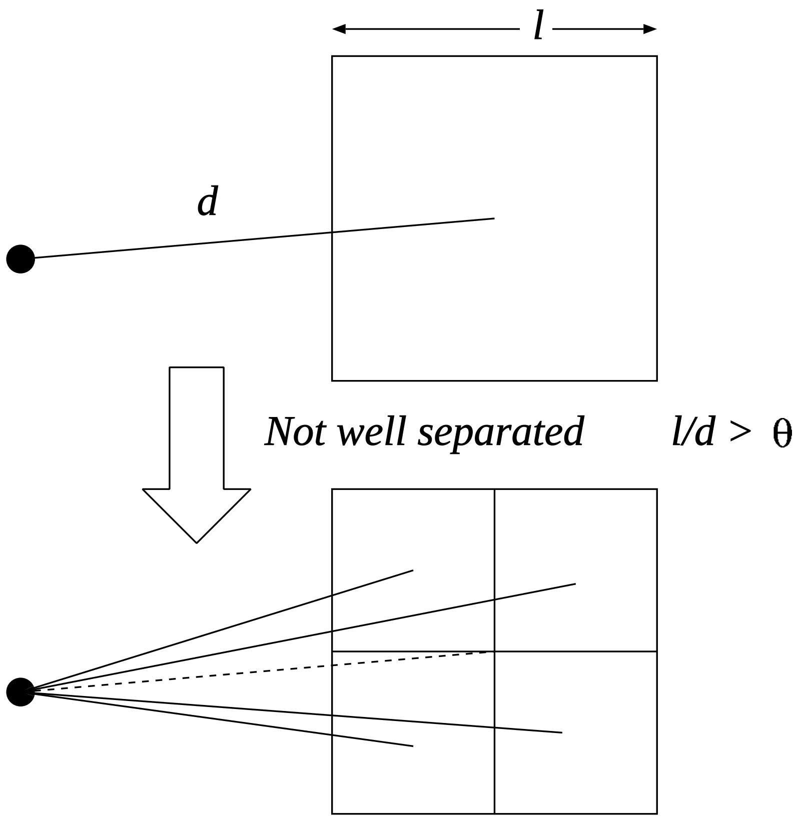


- **lowest-level cells: Directly calculate the expansions for stars in it.**
- **Higher-level cells: Shift and add the expansions for child cells.**

Calculate bottom-up.
Calculation cost: $O(Np^4)$ (p: expansion order)

Force calculation in tree method

Recursive expression:



- Well separated: apply the multipole expansion
- not: take summation of the forces from the child cells

Total force = force from the root cell

Second approach: Buy fast computer

We can do fast calculation by using fast computer.

... not that simple ...

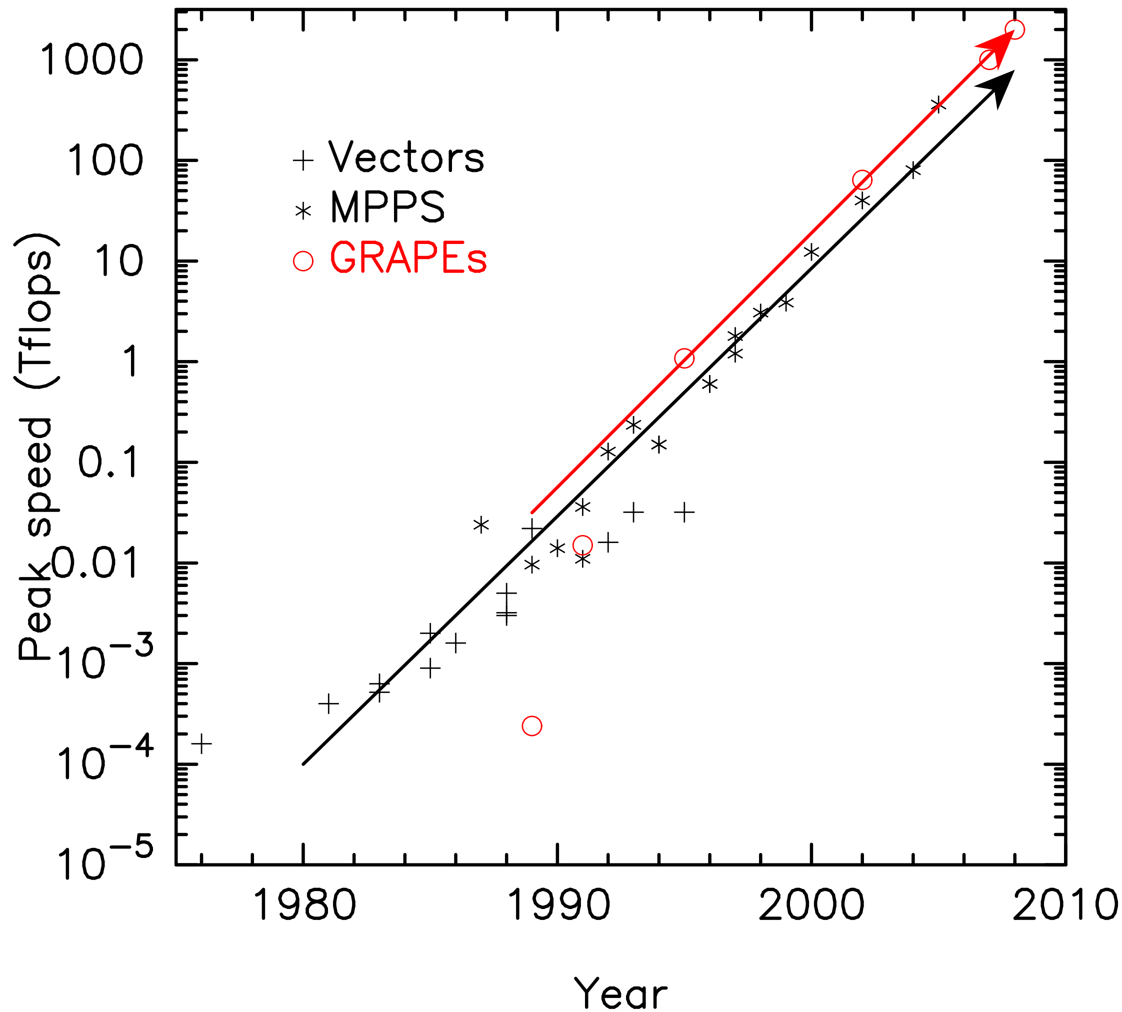
Basic reason:

The development of high-performance computers in the last 30 years made it more and more difficult to use them

Advance in computers

Speed improvement: 10^{10} in 50 years

Roughly exponential in time



How the exponential increase made possible?

1. Moore's law: Size of transistors halves every three years
   - 4 times more transistors
   - 2 times faster
2. Change in computer architecture
   Scalar $\longrightarrow$ Vector $\longrightarrow$ distributed parallel

We need parallel algorithm which is efficient on parallel machines with relatively slow network

(I'll not discuss it here...)

Third approach — build your own computer

Using fast computers is not easy...

- In 10 years, computer architecture might completely change, making your program totally useless. (Have changed in every 10 years)
- Using modern machines is hard:
  - Parallelization on distributed-memory machine
  - Cache reuse
  - Other complicated techniques

Isn't there a somewhat better way of life?

One approach: build your own computer

It's difficult to use the computer somebody else mad for some other purpose

Could be simpler to design the machine suited for your goal (special-purpose hardware).

Why consider special-purpose?

(Might be) faster and cheaper than general-purpose computers.

Why?

- Characteristics of the problem itself
- Technical aspects
- Historical, economical aspects

Characteristics of the problem itself

Stellar system :one star interacts with all other stars

- Large calculation cost (compared to memory requirement)
- Calculation is simple loop
- Communication pattern is simple

We do need some additional considerations for individual timestep and tree code.

Classification of the physical systems

Continuous:(Hydro etc): regular, near-neighbor communication, small calculation cost

Particles: regular $N \times N$ comm), high calculation cost

Others(discrete irregular systems)

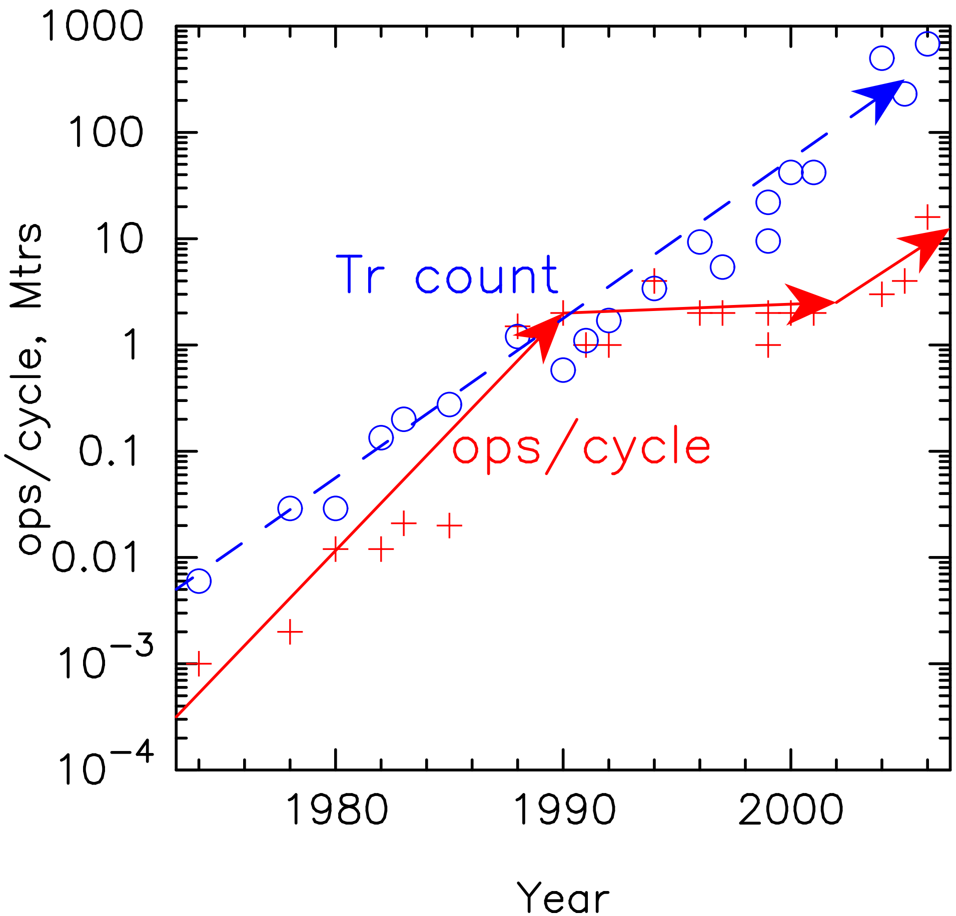
$\boxed{\textbf{Regular and costly = suited to special-purpose hardware}}$

Technical aspects

- **Advance in semiconductor technology: Large-scale circuits with large number of arithmetic units becomes technologically feasible**

- **Limit in design method = rapid decrease in transistor efficiency**

“Evolution” of microprocessors

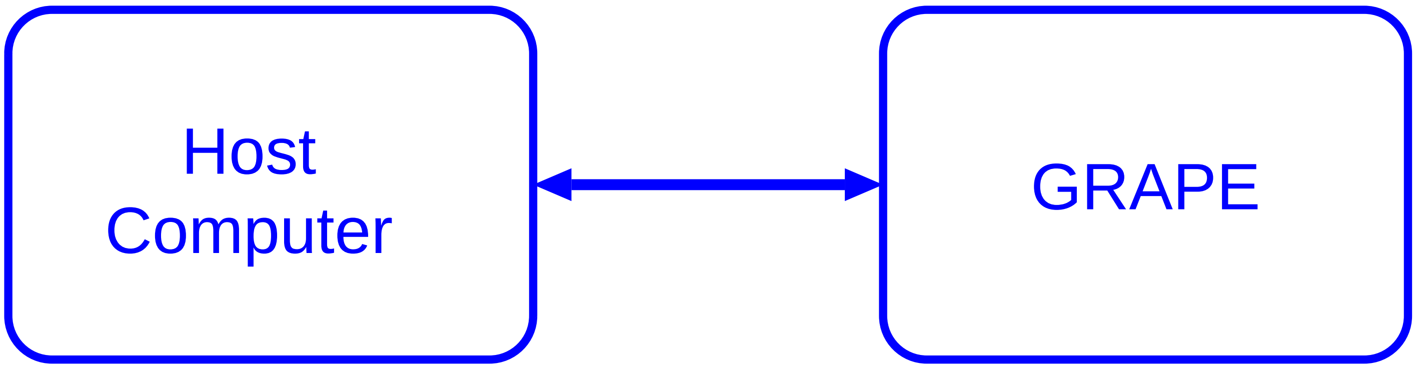


Number of transistors and Number of arithmetic operations per clock cycle

Transistor number increases exponentially
Operation count stuck at 1

Could be improved?

Basic idea of GRAPE



Time integration etc. Interaction calculation

Special hardware: interaction calculation
General-purpose host: everything else

Special-purpose hardware

- **Pipeline processor specialized for interaction calculation**
  - **Large number of FPUs**
  - **Small overhead**
  - **All FPUs always run in parallel**

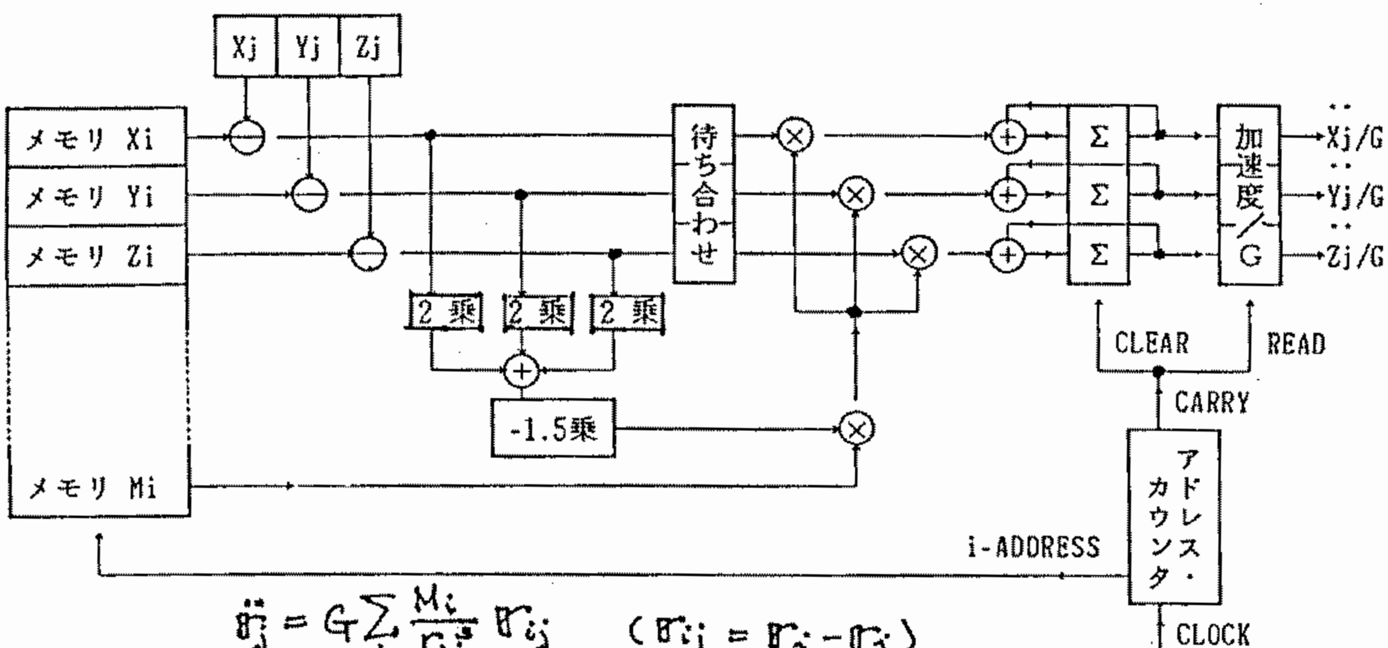
  $\rightarrow$ Very high performance

Important condition: low memory bandwidth requirement

General-purpose host computer

- “High-level” languages (Fortran, C, C++...)
- Existing programs with minimal changes
- Individual timestep, tree method, FMM

GRAPE Pipeline



$$\ddot{\mathbf{r}}_j = G\sum_i \frac{M_i}{r_{ij}^3}\mathbf{r}_{ij} \quad (\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j)$$

+, −, ×, 2乗は1 operation, -1.5乗は多項式近似でやるとして10operation 位に相当する.
総計24operation.
各operation の後にはレジスタがあって, 全体がpipelineになっているものとする.
「待ち合わせ」は2乗してMと掛け算する間の時間ズレを補正するためのFIFO(First-In First-Out memory).
「Σ」は足し込み用のレジスタ. N回足した後結果を右のレジスタに転送する.

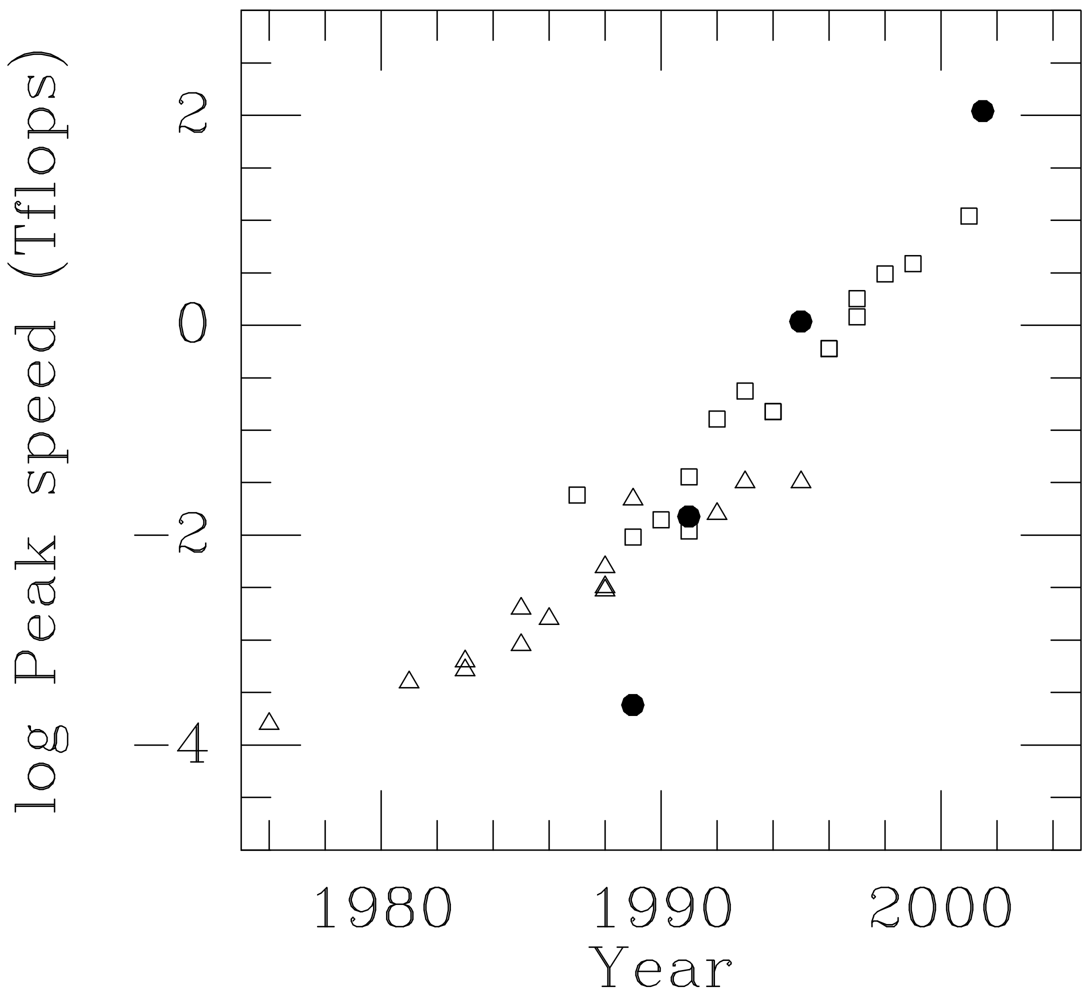
図2. N体問題のj-体に働く重力加速度を計算する回路の概念図.

(Chikada1988)

Evolution of GRAPEs

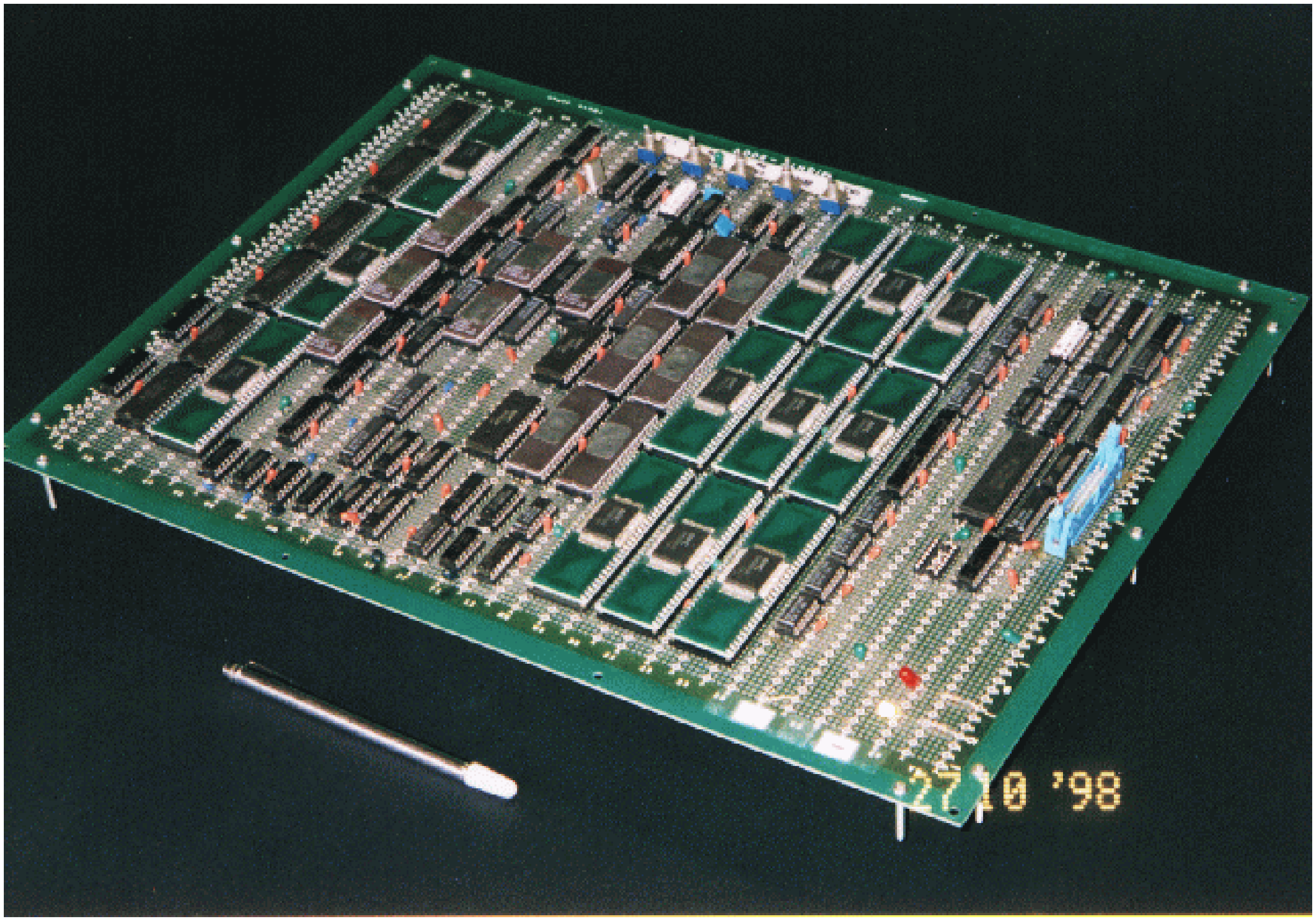
1989	**GRAPE-1**	**low acc, EPROM**
1990	**GRAPE-2**	**high acc, FPU chips**
1991	**GRAPE-3**	**low acc, Custom LSI**
1995	**GRAPE-4**	**high acc, custom LSI, Massively Parallel**
1998	**GRAPE-5**	**low acc, two pipes in a chip**
2001	**GRAPE-6**	**high acc,six pipes in a chip, MP**
2005	**GRAPE-7**	**low acc, 20 pipes in a chip**

Evolution of speed



filled circles: GRAPE

GRAPE-1



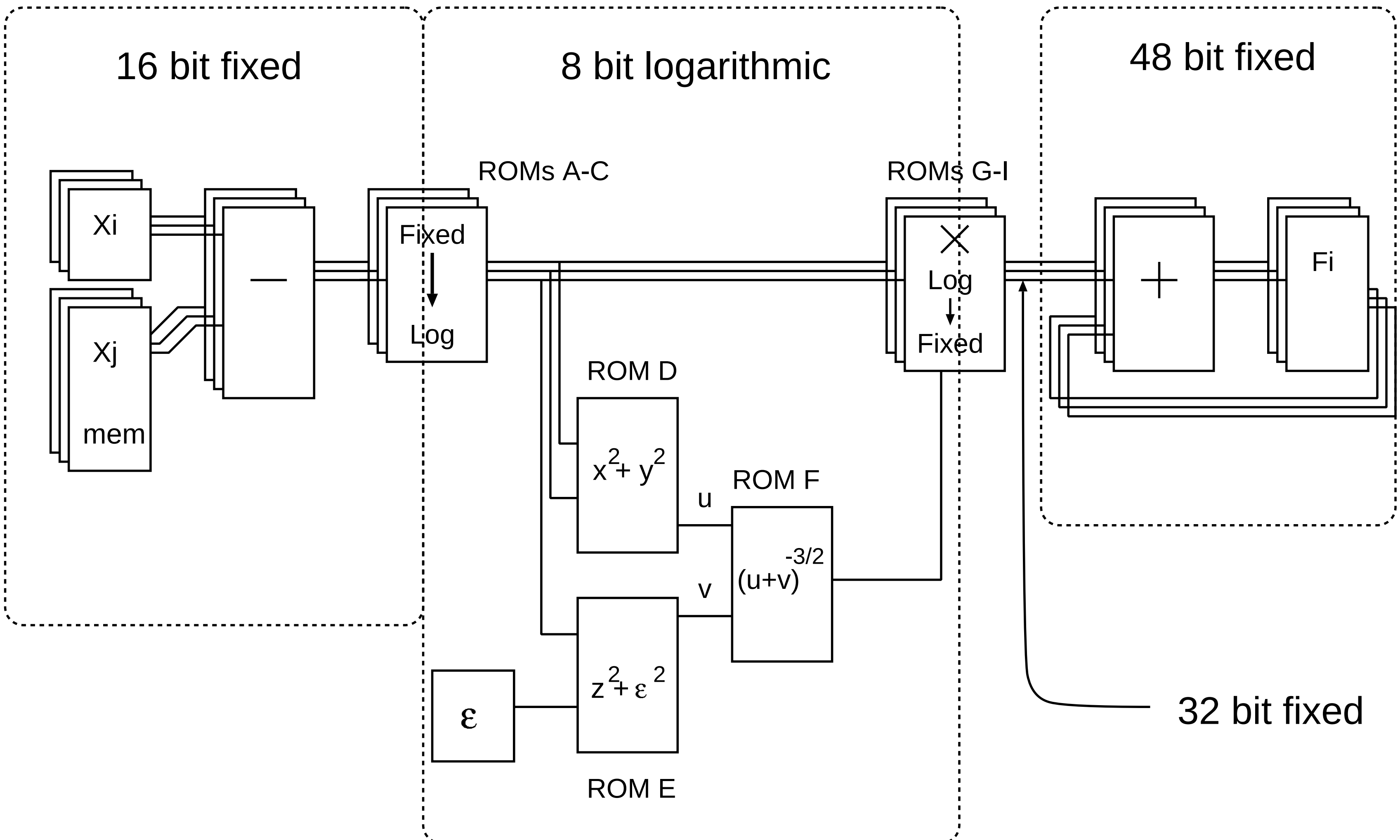
GRAPE-1 internals

“Digital Circuit for the beginners”

Initial goal

- **Make something like a force pipeline**
- **Connect to the host and evaluate performance**
- **Do not care much if it is useful for real calculation**

GRAPE-1 pipeline



Troubles during development

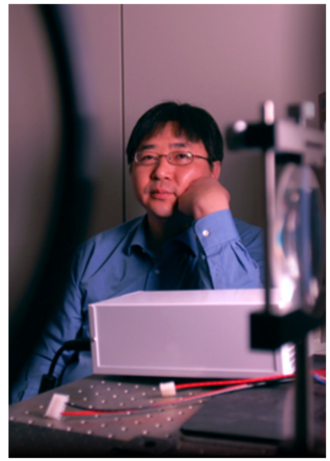
Hardware seemed to be completed without much problem (since Ito did the work, not me)

Performance problem:

Initially we used one MS-DOS PC (NEC PC-98). It was fine

We moved to a Unix workstation (Sony NEWS): Communication became very slow
We had to hack the operating system...

Tomoyoshi Ito

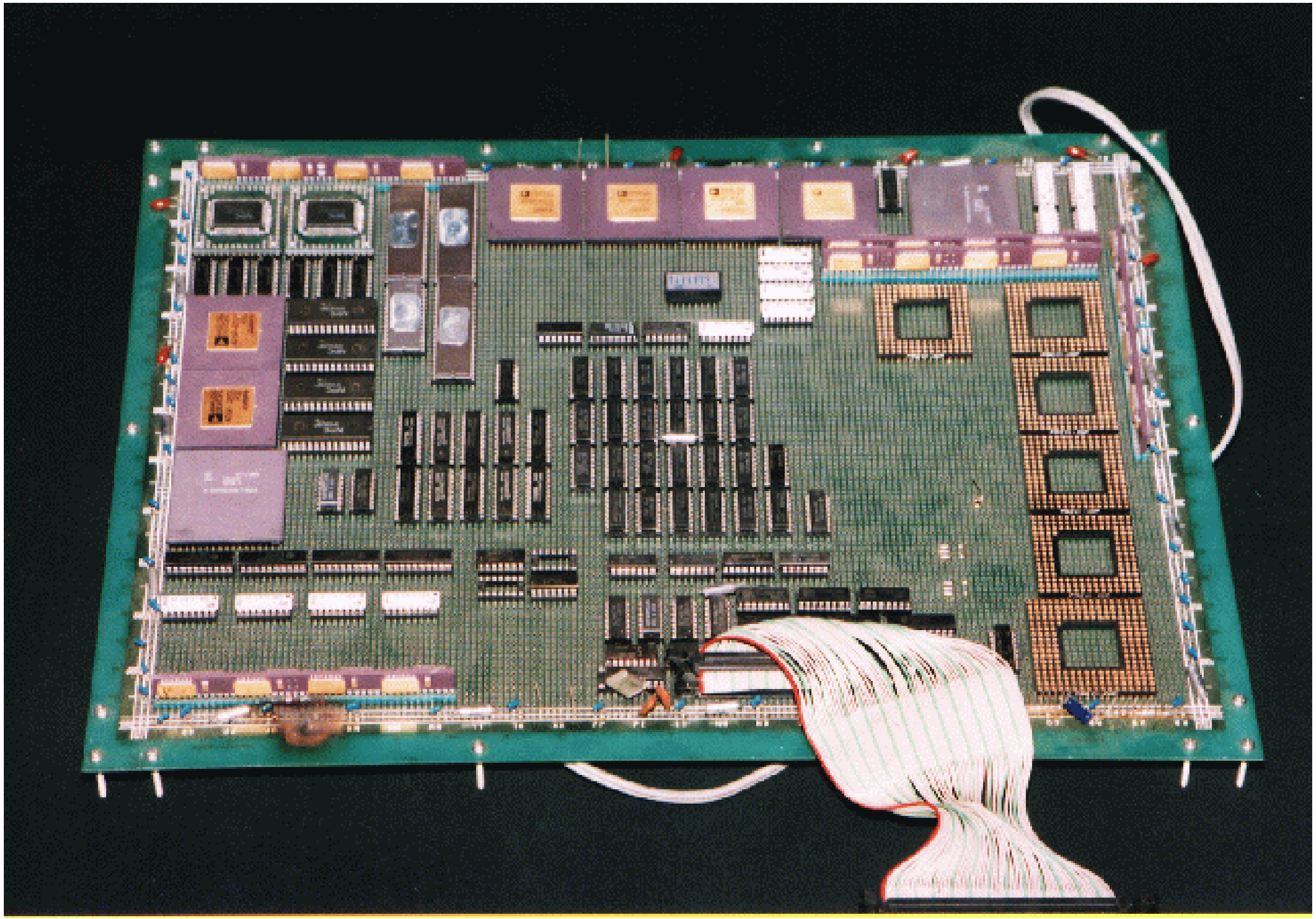


Might be better known as the author of the comic series "Eiko-naki tensai tachi" (Geniuses without fame)
Now professor of EE at Chiba Univ.

Lessons learned

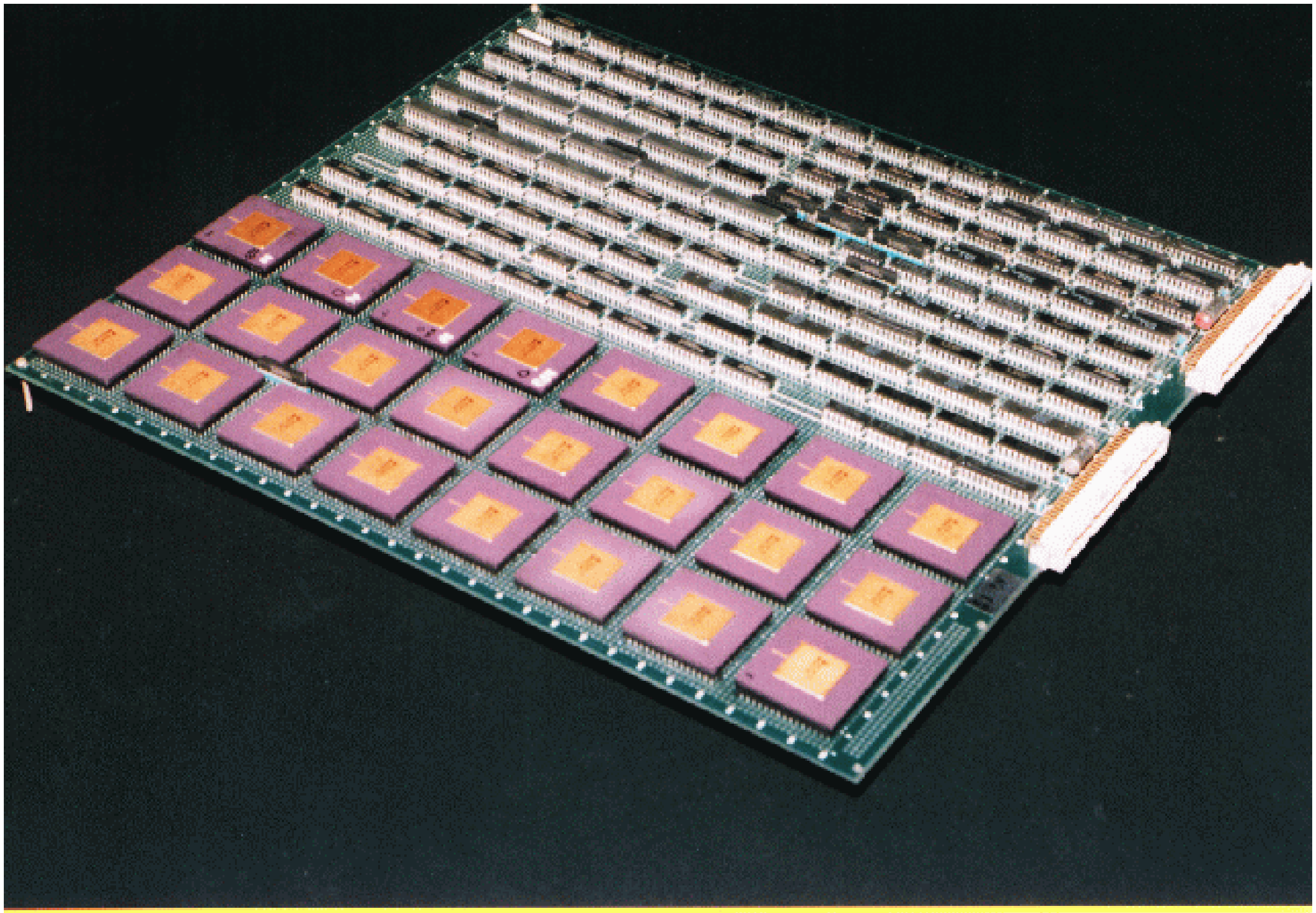
- Communication software is difficult
- "Recommended" or usual methods in textbook does not necessarily give good result
- Good result justifies whatever approach used

GRAPE-2

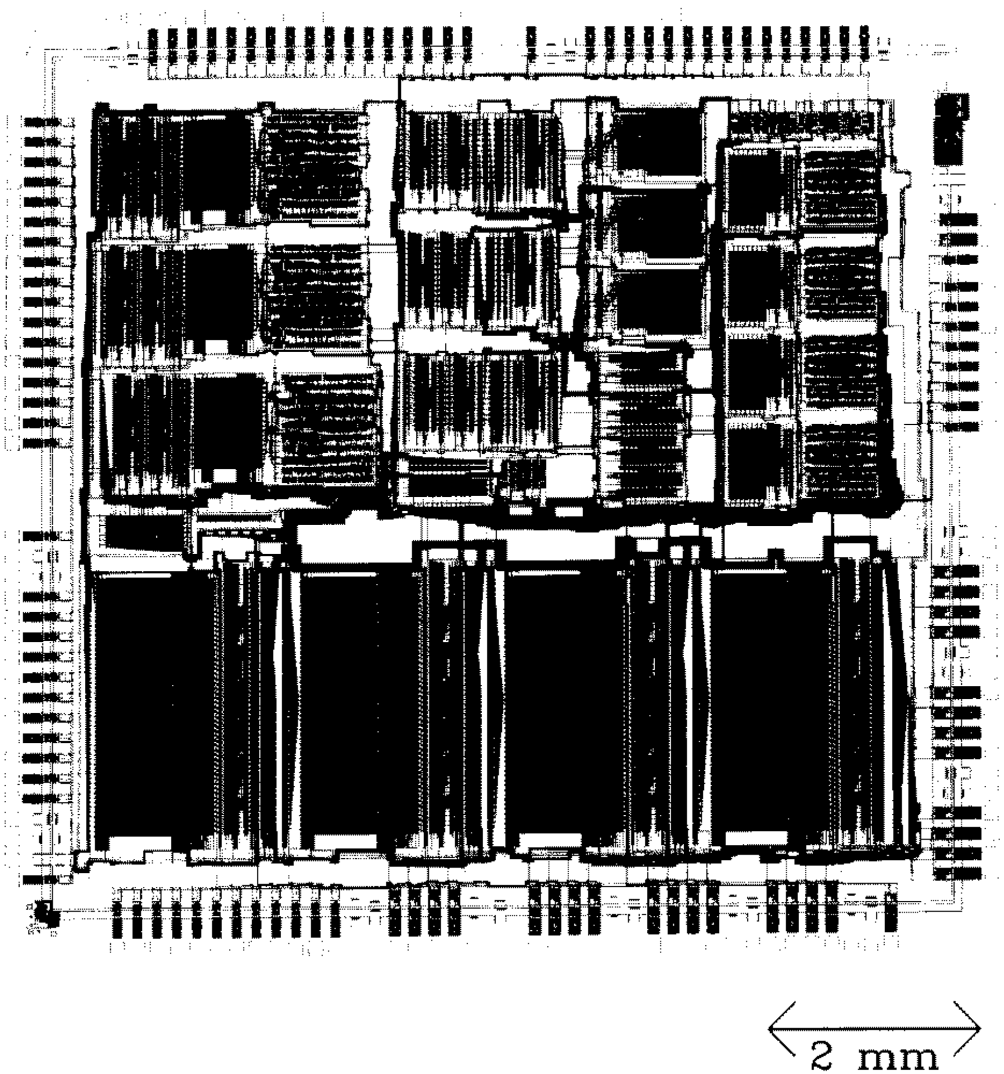


x

GRAPE-3



GRAPE-3 Custom LSI



GRAPE-3 chip design

- Specification, behavior simulation: JM
- Detailed logic design: Fuji-Xerox
- SCS Genesil design tool
- National Semiconductor. 1μm

How the chip-making affect your health?

We never had the budget for "respin", or redesign of the chip

Division of the responsibility

- If the test pattern did not get through, that is manufacturer's fault
- If other faults found, that's my fault...

In theory, if we can prepare perfect test pattern, there will be no problem.

In practice...

GRAPE-4

GRAPE-6

- Design principle
- Processor chip
- Overview

Design principle

Goal of the project (when we got budget)

— achieve the world's best performance

Our real goal:

To build a machine which can do real sciences.

Boundary condition

- **Budget: 500MJYE (Earth Simulator 50BJYE, ASCI Q 200MUSD**

Target performance: 200Tflops (5 times that of ES)

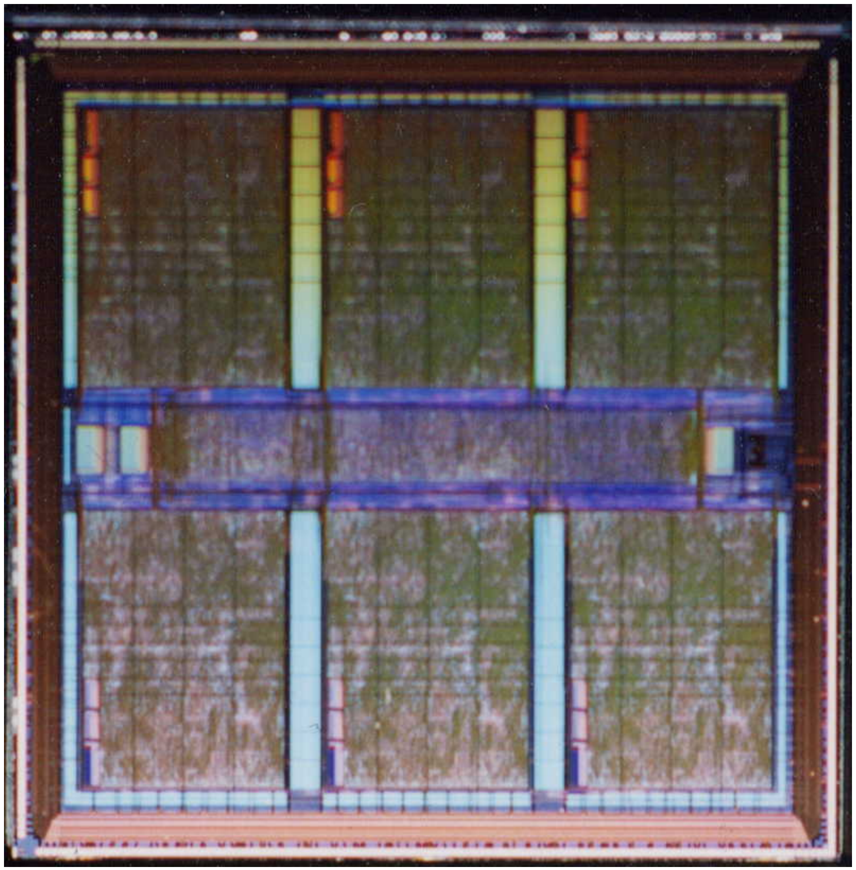
Performance prediction for GRAPE-6

Prediction: Extrapolation from GRAPE-4

	G4	G6 (pred)	G6 (real)
Design	1μm	0.25μm	0.25μm
Clock	32 MHz	125 MHz	90MHz
Pipelines	1/3	5-10	6
Performance	600Mflops	36-72 Gflops	31 Gflops
Initial Cost	25M	70M	More than 100M
Chip cost	8000K	10-20K	30K

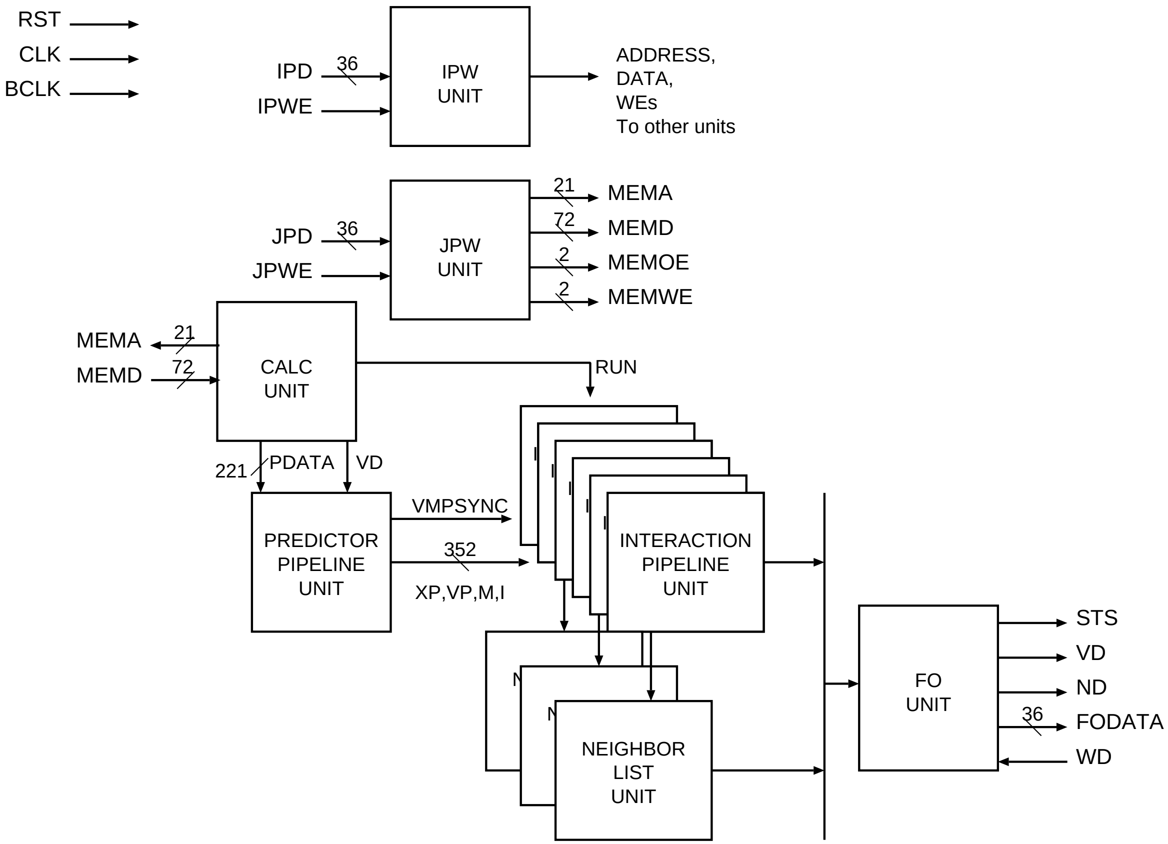
Accurate except for the cost estimate...

Pipeline chip



- 0.25 μm (Toshiba TC-240, 1.8M gates)
- 90 MHz Clock
- 6 pipelines
- 31 Gflops

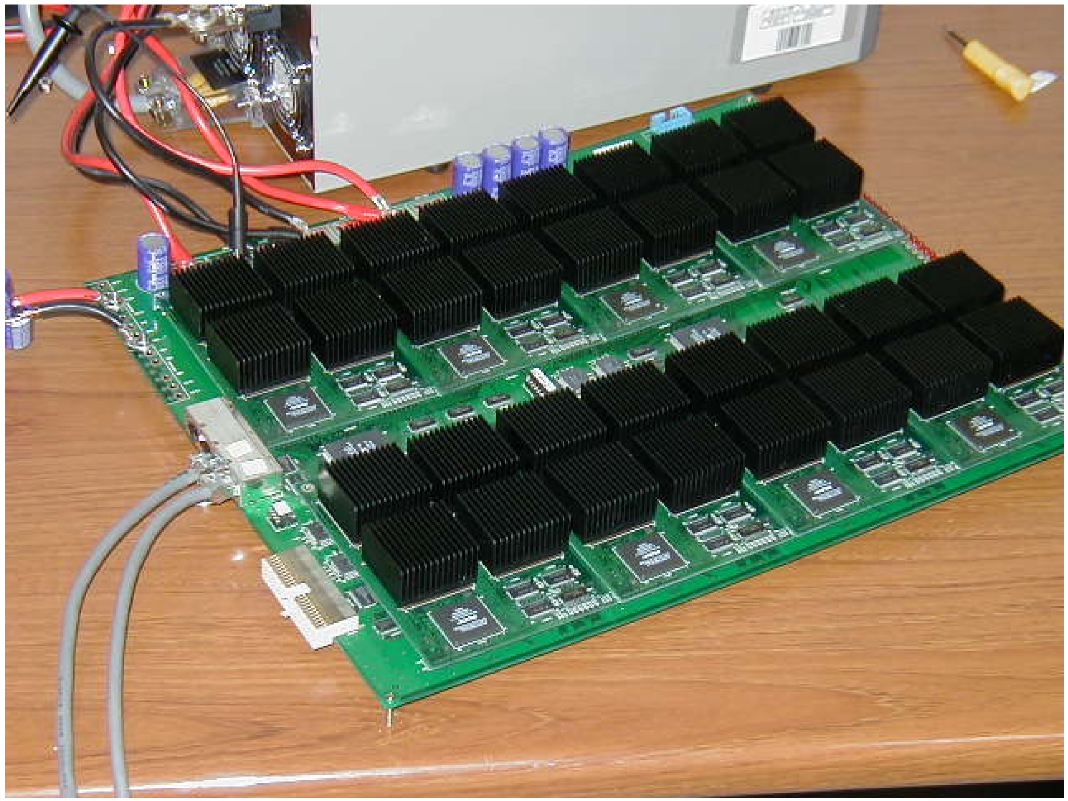
Details of LSI



Equivalent to single GRAPE-4 board

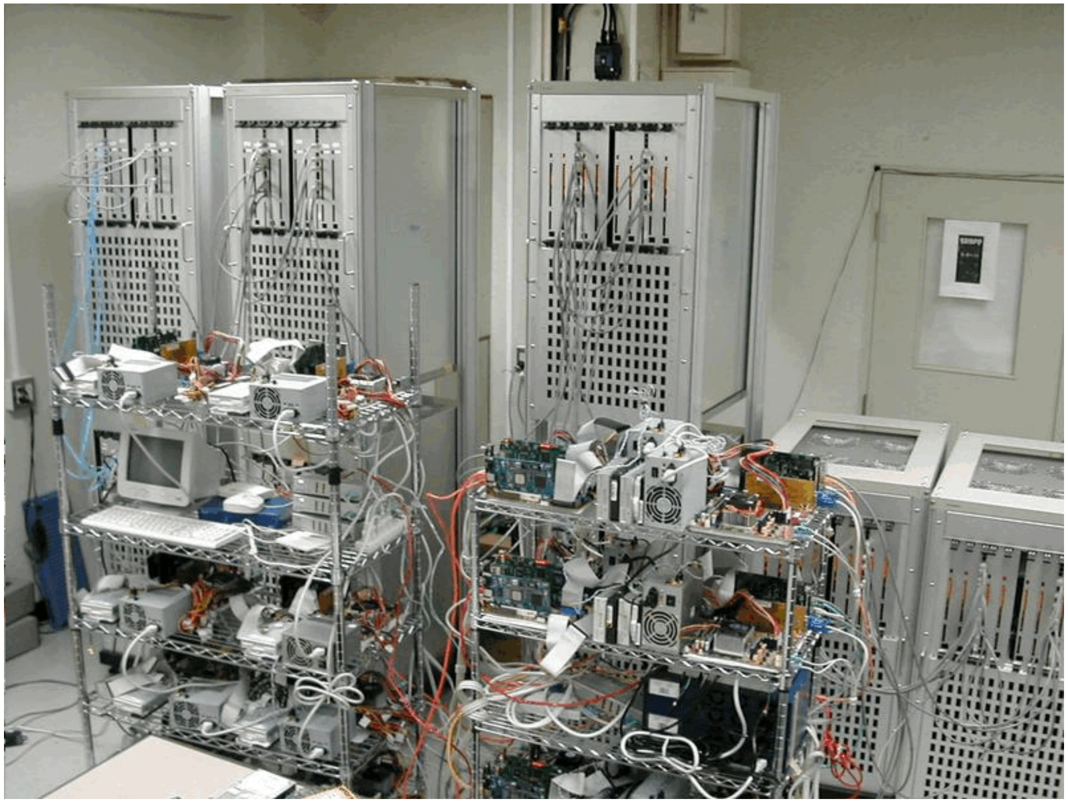
- Host IF
- Memory IF
- Pipelines
- Controls

GRAPE-6 processor board



- 32 chips/board
- LVDS interface(350MHz clock, 4 wires, about 1Gbps)

The 64-Tflops GRAPE-6 system

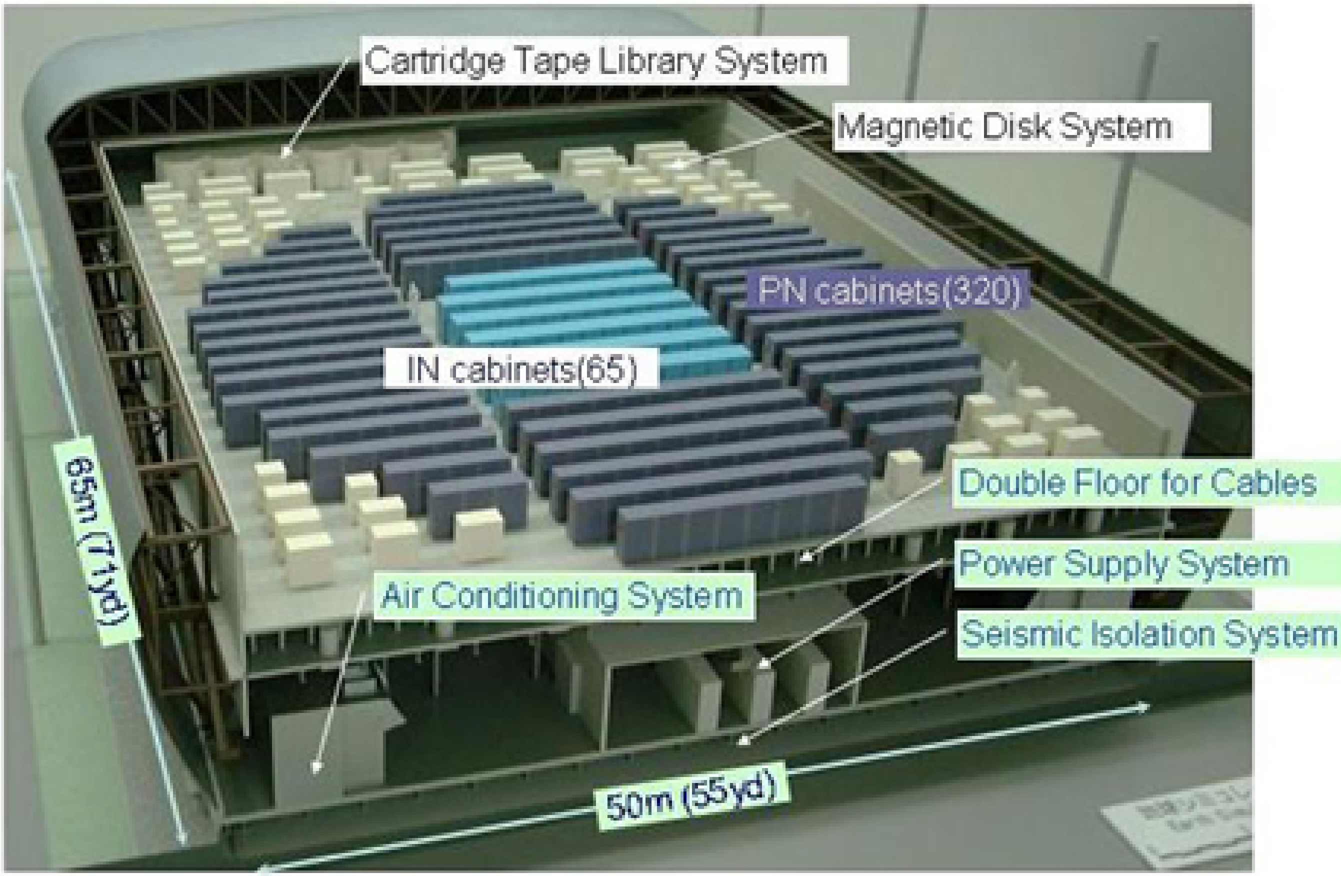


64-Tflops system.

4 blocks with 16 host computers.

In one room in Building 3, Asano-area of UT

The 40-Tflops Earth Simulator



Comparison with a recent Intel processor

	GRAPE-6	Intel Xeon X7460
Year	1999	2008
Design rule	250nm	45nm
Clock	90MHz	2.66GHz
Peak speed	32.4Gflops	64Gflops
Power	10W	130 W
Perf/W	3.24Gflops	0.49 Gflops

Even after 10 years...

"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?

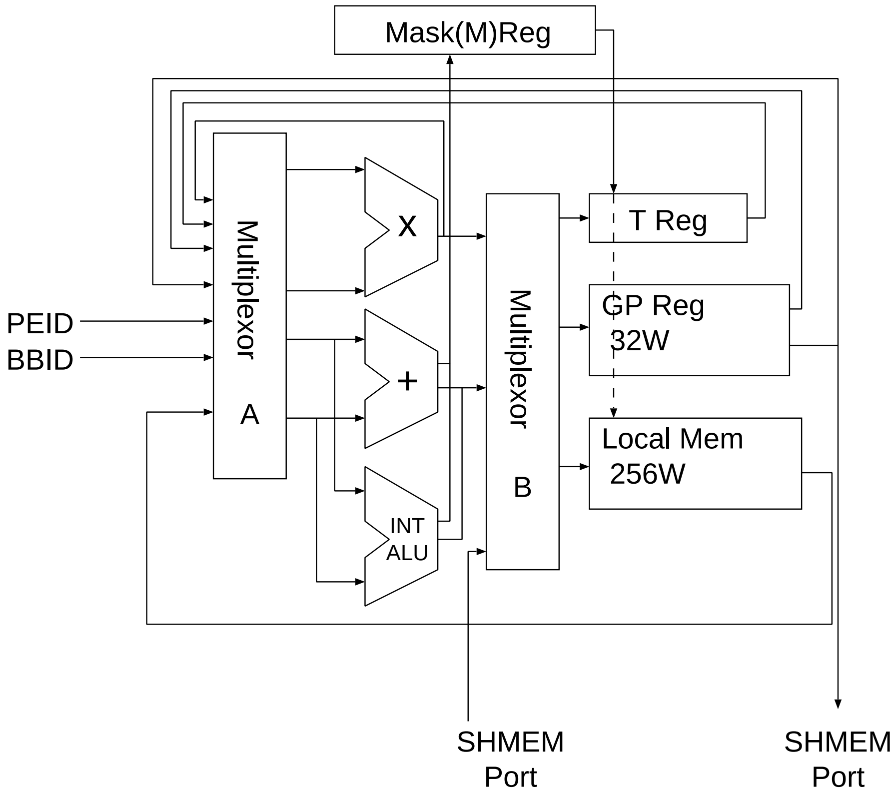
Initial cost should be 1/4 or less of the total budget.

How we can continue?

Next-Generation GRAPE — GRAPE-DR

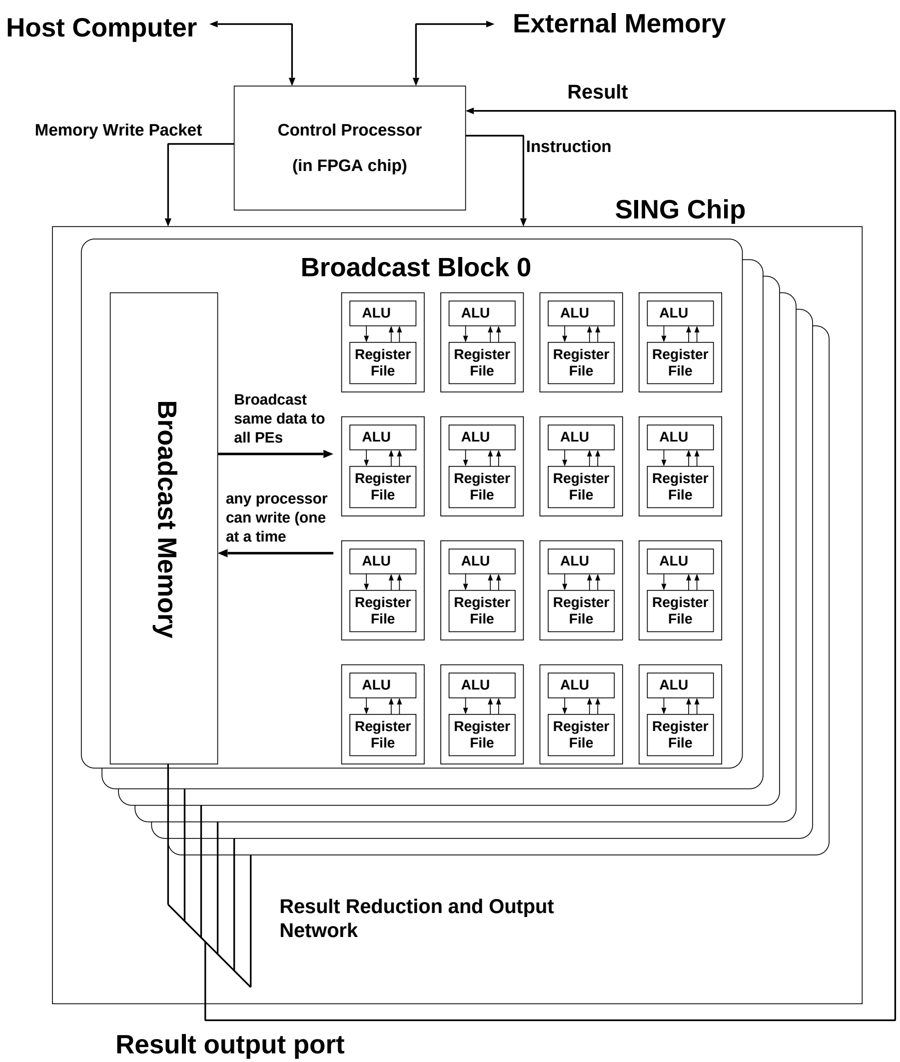
- Planned peak speed: 2 Pflops
- New architecture — wider application range than previous GRAPEs
- primarily to get funded
- No force pipeline. SIMD programmable processor
- Planned completion year: FY 2008 (early 2009)

Processor architecture



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

Chip structure



Collection of small processors.

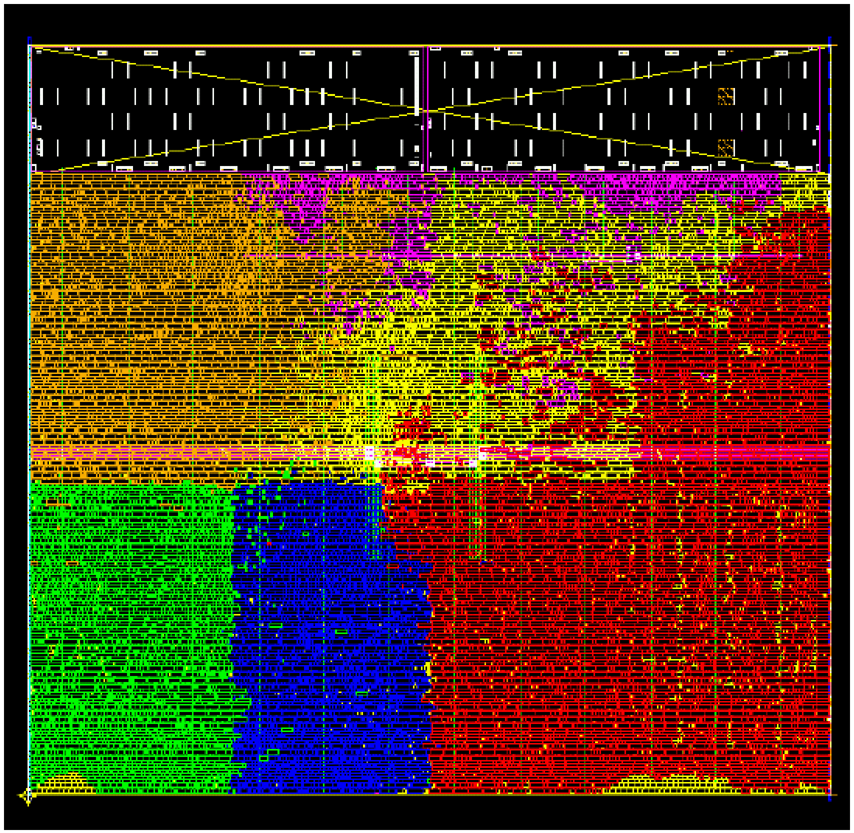
512 processors on one chip
500MHz clock

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

Why we changed the architecture?

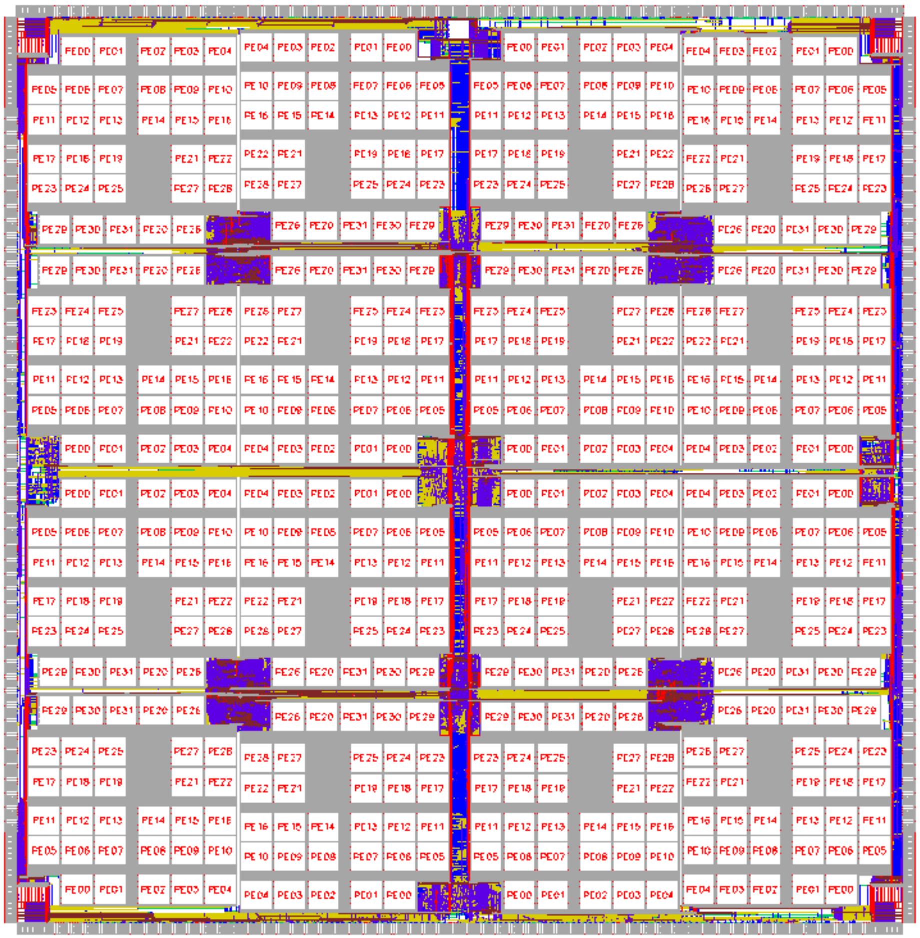
- To get budget (N-body problem is too narrow...)
- To allow a wider range of applications
  - Molecular Dynamics
  - Boundary Element method
  - Dense matrix computation
  - SPH
- To allow a wider range of algorithms
  - FMM
  - Ahmad-Cohen
  - ...

PE Layout



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

Chip layout

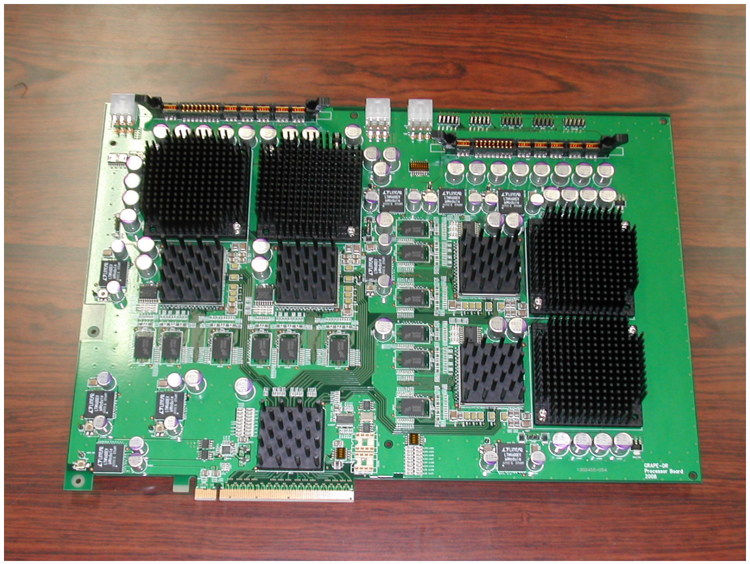


- **32PEs in 16 groups**
- **18mm by 18mm**

The processor chip

Sample chip delivered May 2006

Processor board

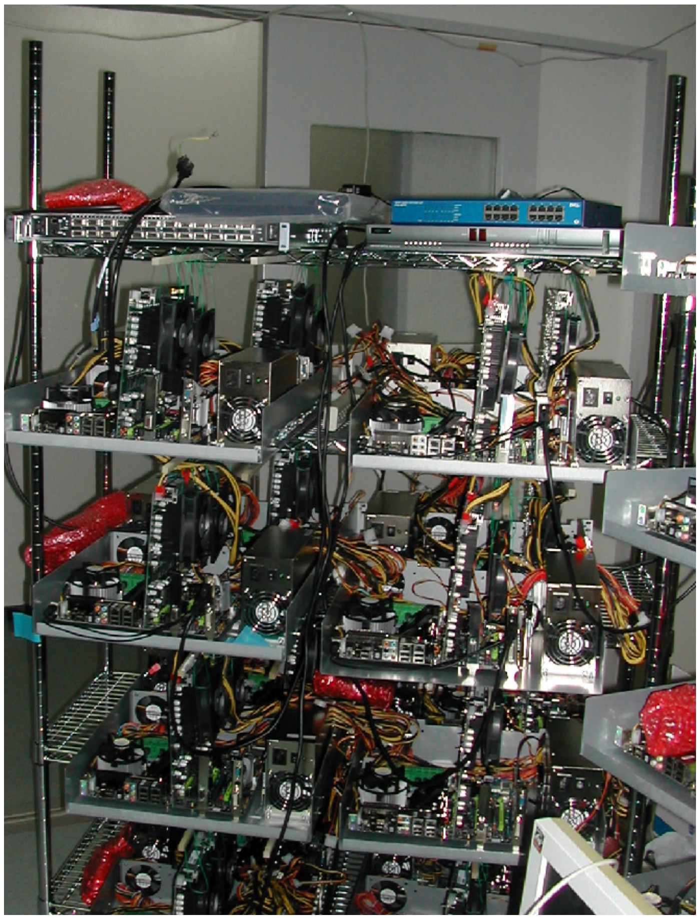


PCIe x16 (Gen 1) interface
Altera Arria GX as DRAM controller/communication interface

- Around 250W power consumption
- Not quite running at 500MHz yet... (FPGA design not optimized yet)
- 900Gflops DP peak (450MHz clock)
- Available from K&F Computing Research

GRAPE-DR cluster system

Just to show that the system exists...



Host computer: Intel Core 2 Quad Q6600 with nVidia 780i chipset

8GB memory

Network: IB (4x DDR)

HPC Linpack passed (not tuned yet....)

The system and (preliminary) performance numbers submitted to TOP500

Major concern: Effective host memory bandwidth

GDR cluster in 2009

- **Nehalem with 3way DDR3 memory should resolve bandwidth problem.**
- **IB network**
- **800T-1P DP peak range.**

Japan's Next-generation Supercomputer Project

- FY 2006-2012
- Total budget: 110 BJYE (about 80 times that of GRAPE-DR)
- Peak speed: 10Pflops (about 10 times that of GRAPE-DR)
- Vector (like ES) + Scalar (???) hybrid

Summary

- GRAPEs seems to be fairly successful
- However, we cannot continue...
- With GRAPE-DR, we moved to new architecture
- We'll see if this was the right move or not.