Exascale computers ?/ Can we believe SPH simulations of Giant Impact?

Jun Makino
ELSI, Tokyo Institute of Technology
RIKEN Advanced Institute for Computational Science

Science as Method and Methodology for Problems on the Earth and Life,
Sept 15-17, Nagoya

Structure of the talk

- Who am I?
- Exascale computer?
  - Advance of the Supercomputers: 1950-2010
  - Problems we see today
  - Solutions?
  - Japanese Exascale Project
- SPH simulations of Giant Impact
  - Problem with "standard" SPH — failure at contact discontinuity
  - New formulation
  - Preliminary results

Who am I?

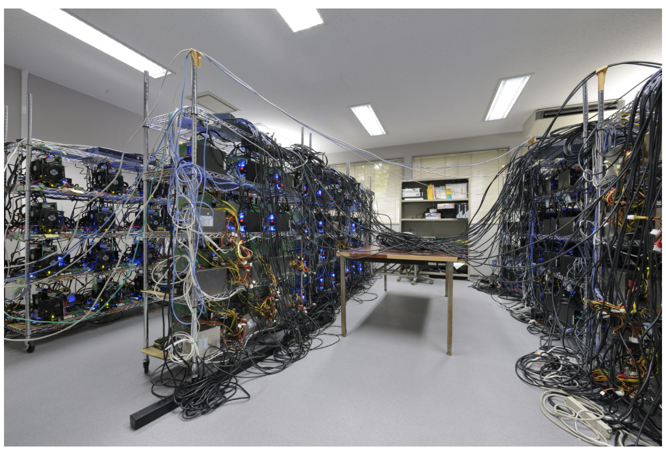
Current position (as of Apr 1st):

- PI, ELSI
- Team leader, Particle Simulator Research Team, Advanced Institute of Computational Science, RIKEN

What I have been doing for the last 20 years:

Develop GRAPE and similar hardware for astrophysical N-body simulations. Use them for:

Planetary formation, star cluster dynamics, galactic dynamics, cosmology



GRAPE-DR

K Computer

The way computational science "should" proceed

- **Four ingredients**
  - Scientific problem
  - Numerical simulations
  - Numerical methods (and software)
  - Computer hardware (and software)
- **Numerical methods and computer hardware should be "optimal" for the problem you want to solve**

In practice, "optimization" is either very poor or nonexistent

Why?

Because

- **Very few people are working on numerical methods**
- **Those who working on numerical methods knows very little about scientific problems**
- **Virtually nobody is working on computer hardware**
- **Those who working on computer hardware ....**

Examples

- **Evolution of supercomputers**
- **Smoothed Particle Hydrodynamics**

Advance of the Supercomputers

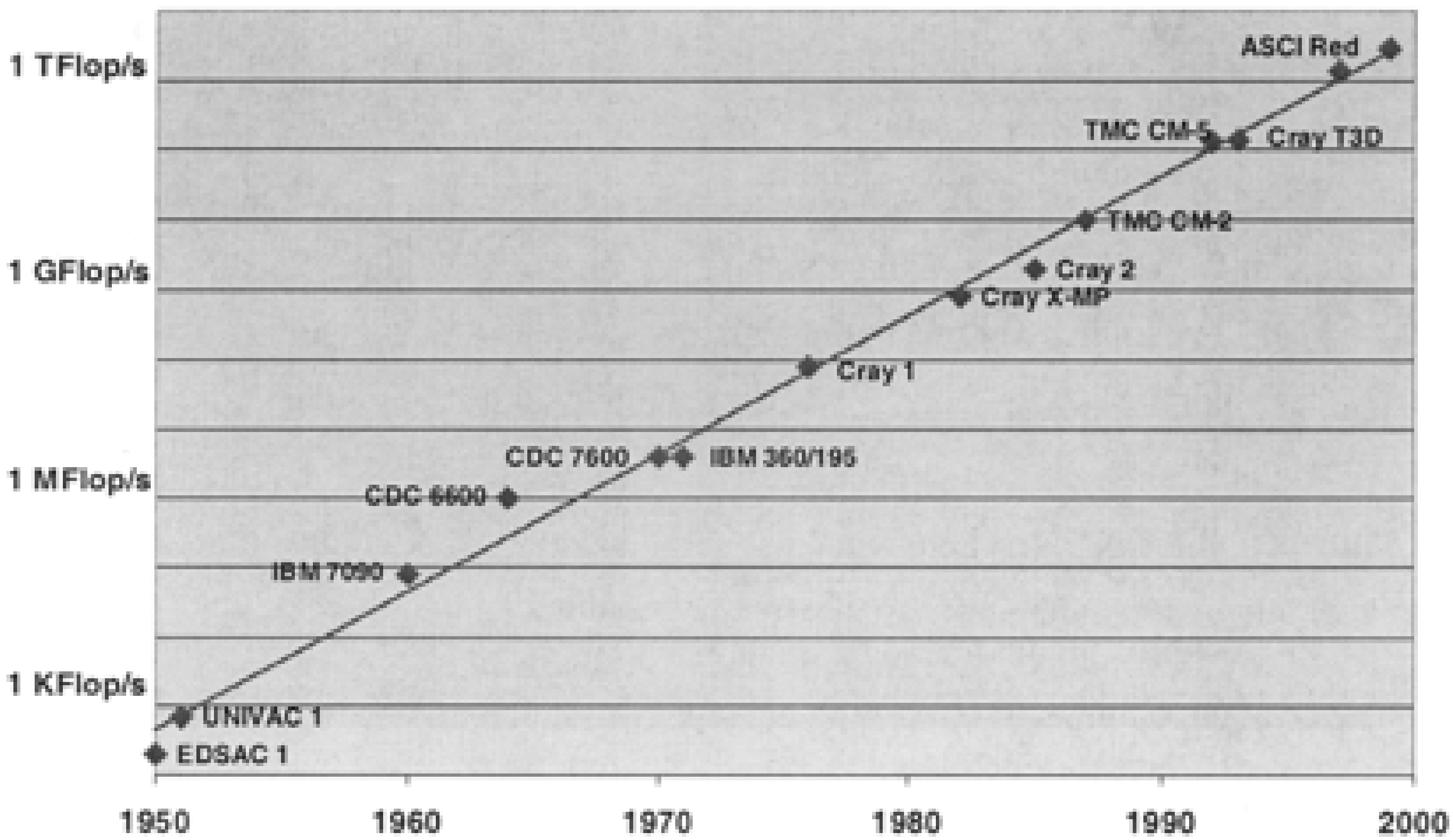
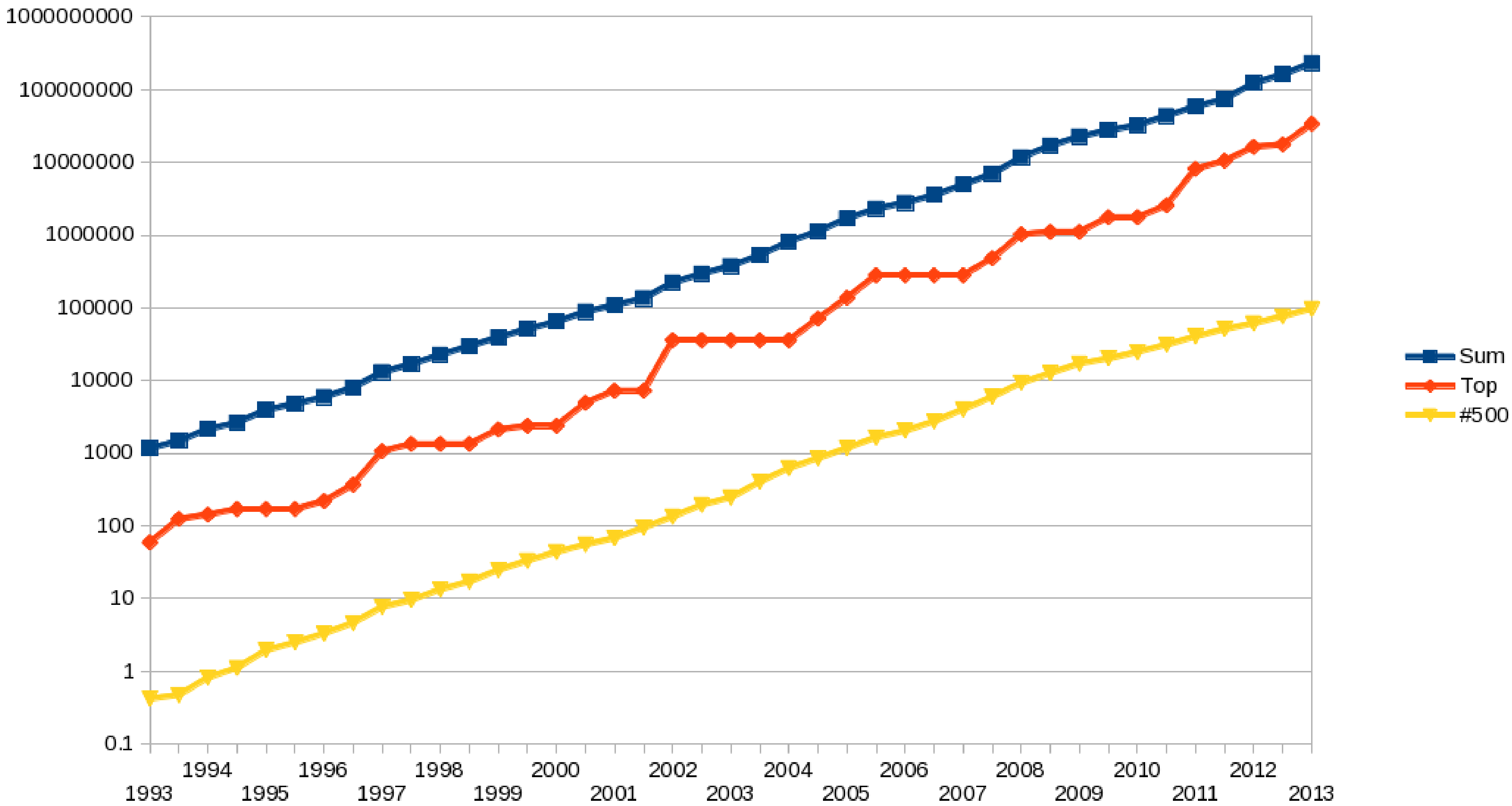


Figure 1. *Moore's law and peak performance of various "supercomputers" over time.*

1940-2000: 100 times per decade

Advance of the Supercomputers

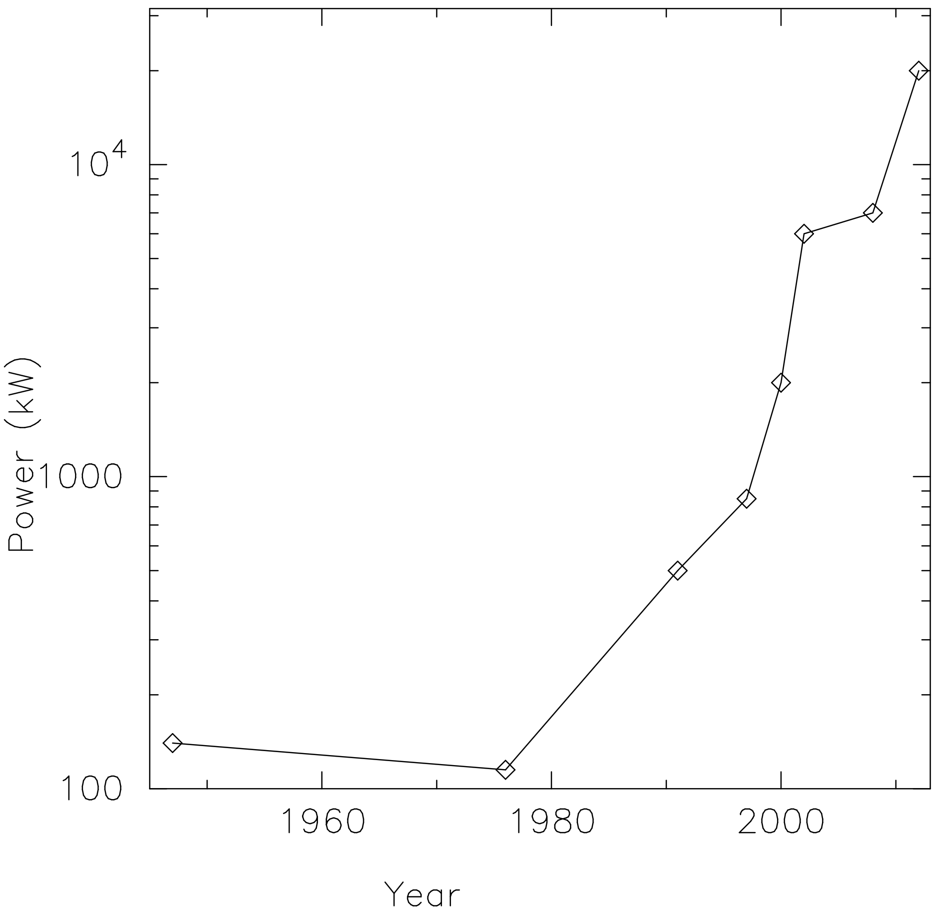


1993-2013: 500 times per decade(!?)

Problem 1: Power consumption

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**
K-computer	**2012**	**20MW**

If we plot data...



Little increase from ENIAC to Cray-1

Increase by a factor of 10 in 1975-95

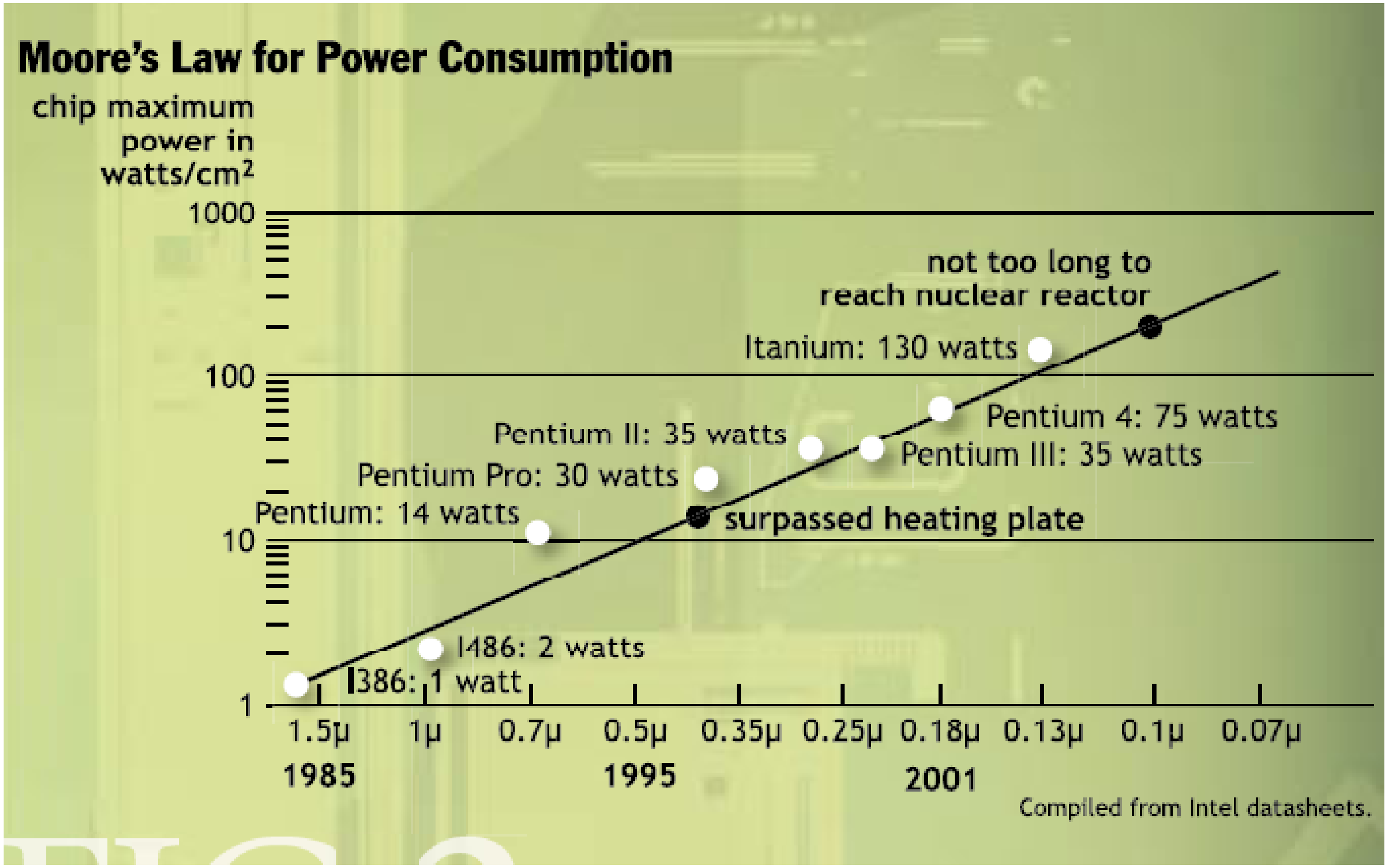
Factor of 30 in 1995-2012

Faster-than-exponential increase

Why?

- **Price increased: ASCI Red: $ 50M, K-computer: $ 1G**
- **Power consumption per chip (or per cm^2 of silicon) increased**
- **Price per chip (or per cm^2 of silicon) decreased**

Power consumption per cm^2 of silicon



(Not much increase since 2003. Practical limit of cooling reached)

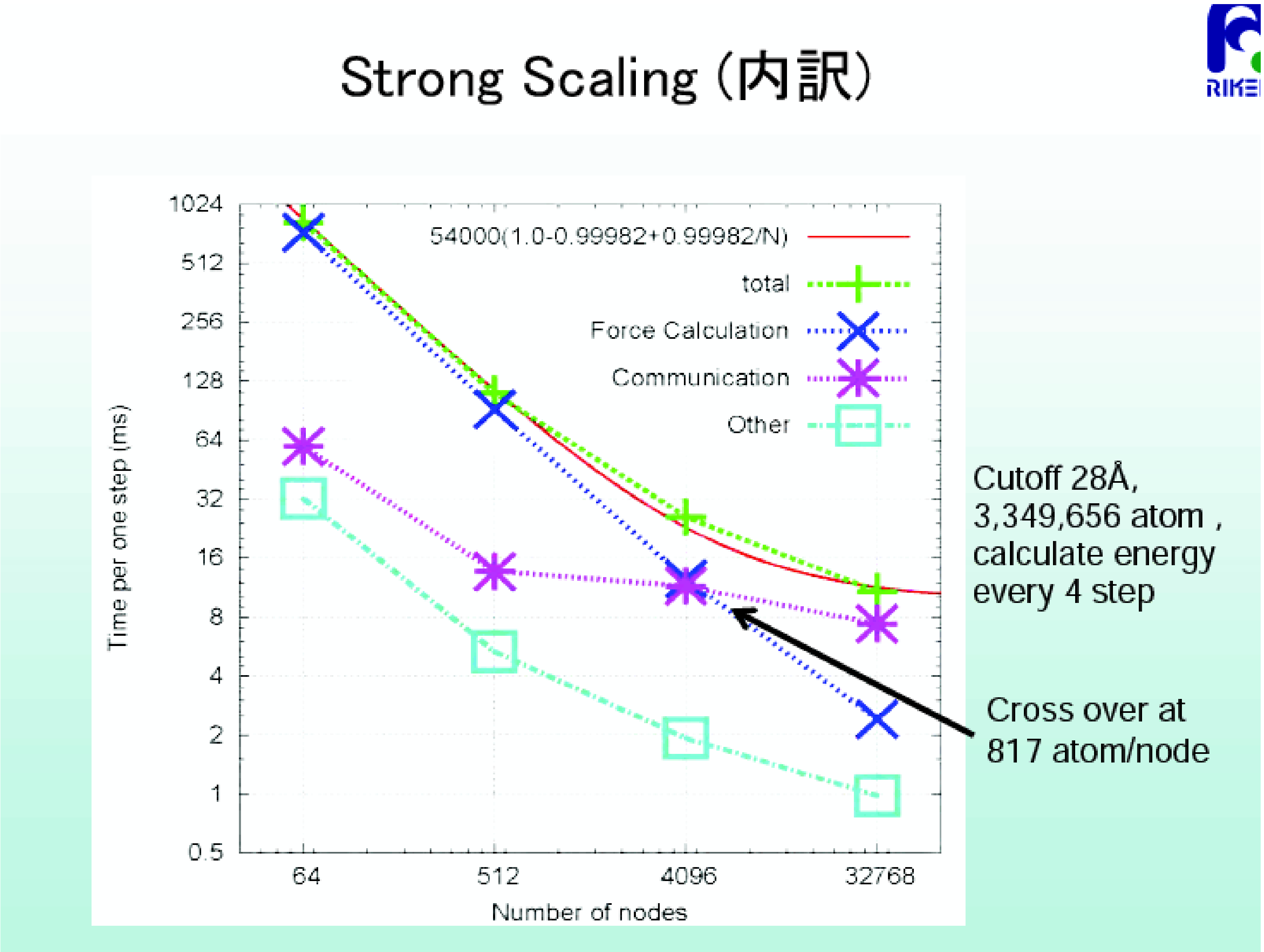
Problem 2: Parallelization overhead

Number of floating-point units (Multiply and add)

Cray-1	1976	1
Cray C90	1991	16
ASCI White	2000	16,384
ES	2002	40,960
K computer	2012	2,820,096

K computer is good for large problems (with small number of timesteps) but not so good for problems that require large number of timesteps.

Example of performance scaling



Molecular Dynamics on K computer

- One cannot go below 5ms/timestep
- Limitation: communication overhead

Is 5ms/step fast enough?

- No — for problems that require long simulation time
- Simulation is 10^{12} times slower than real time.
- Simulation of 1 ms would take a century...

Problem 3: How you develop/maintain codes???

- MPI
- OpenMP
- SIMD extensions
- Cache-friendly code
- Accelerators
- ...
- ...

Problem 3: How you develop/maintain codes???

- MPI
- OpenMP
- SIMD extensions
- Cache-friendly code
- Accelerators
- ...
- ...

(I'll not discuss this aspect much...)

What we want and what they give

We want a machine which

- can solve both small and large problems fast
- is power efficient
- is easy to program

What they give to us

- cannot solve small problems fast
- is power inefficient
- is nearly impossible to program

Solutions?

- **We need to reduce power consumption AND communication overhead.**
- **We do not need much memory (1TB would be enough to keep 10^{10} particles)**

Possible solution:

- **Processors with "small" on-chip memory (small means 256MB or more)**
- **Large number of cores, but in SIMD mode to reduce communication overhead**

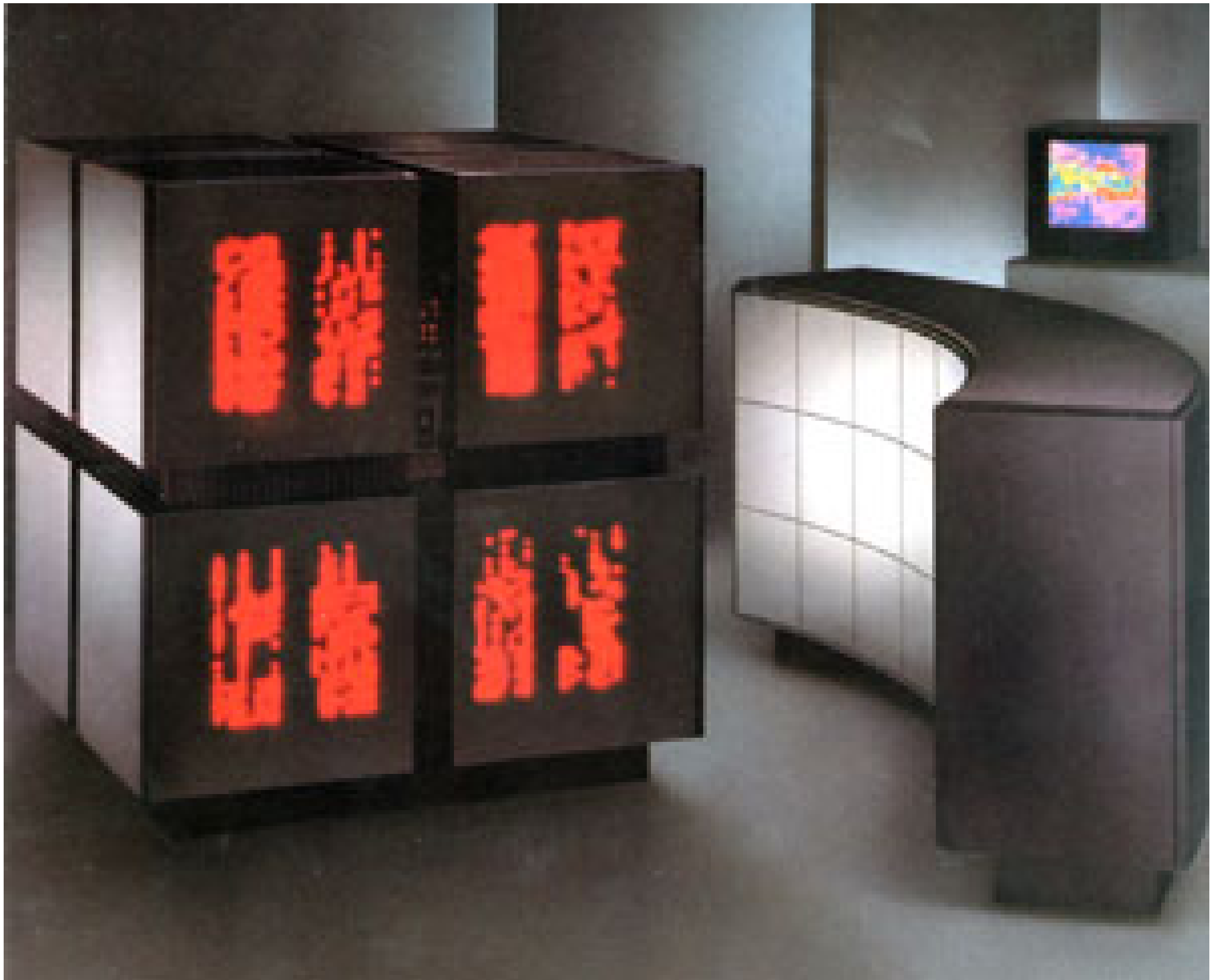
Massively-parallel SIMD machines

This is a “lost technology”

- **Goodyear MPP (1970s)**
- **ICL DAP (Late 1970s)**
- **Thinking Machines Connection Machine-1/2 (Late 1980s)**
- **Maspar MP-1/2 (Early 1990s)**

CM-2 was pretty successful

TMC CM-2



2048 floating point units in SIMD mode

TMC CM-2

- 64k 1-bit processors, each with 64k-bit memory
- 2048 floating-point units, each shared by 32 processors (Feynman's suggestion)
- 12-dimensional hypercube network between processor chips (16 processors in one chip)

With the present-day technology, we can integrate 4-8 CM-2s into one chip, for the peak performance of 10-20 Tflops at < 100W

How we reduce power and communication overhead

- Power:
  - Minimize data movement: Remove external memory and cache
  - Minimize instruction fetch and decode: Massive SIMD
- Communication overhead:
  - Minimize data movement: Remove external memory and cache, reduce the number of chips
  - Reduce the handshake overhead: Cores in SIMD operation do not need handshake, since they are executing the same instruction

Japanese Exascale Project

NHK TV news reporting: Japan to develop new supercomputer with 100x power of K-computer

I was there as a member of a working group organized by the ministry of education

Current rough plan

- **Follow-up of K-computer: would require 60-80 MW to reach exaflops in 2020**
- **Combine SIMD "accelerators" with MIMD general-purpose machine**
- **MIMD part: Fujitsu design**
- **SIMD part: Based on our design**
  - **reduce power consumption by 80%**
  - **reduce communication latency by at least a factor of 10, hopefully by 100.**

Summary for computer hardware

- Current big supercomputers are not ideal for long-term integration of "small" problems ("small" means 10^7 particles now and 10^9 particles in 2020)
- We need a new architecture (or revival of an old architecture...) to solve this problem: Massively-parallel SIMD
- If everything goes well, we will put this MP-SIMD system as part of Japanese Exascale project

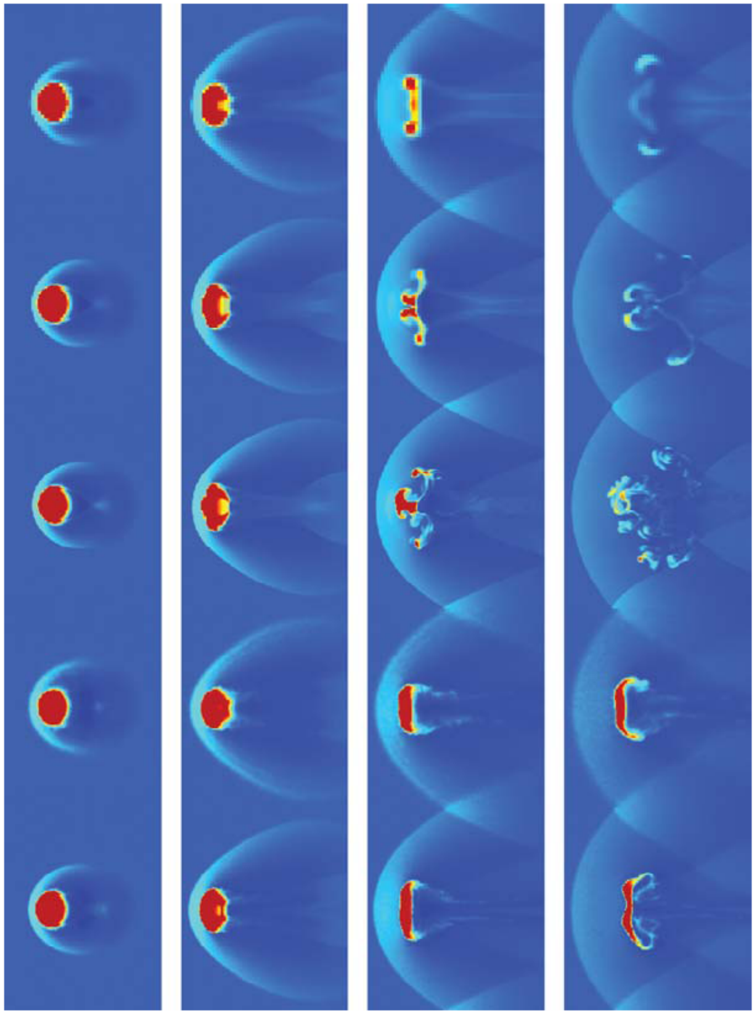
Can we believe SPH simulation of Giant Impact?

- SPH (smoothed particle hydrodynamics) is a scheme to express fluid equations by particles
- SPH has been used in many fields of science and engineering. One example is the Giant Impact
- Recently, its ability to handle contact discontinuity was put into question.

Agertz et al (MN 2007, 380, 963)

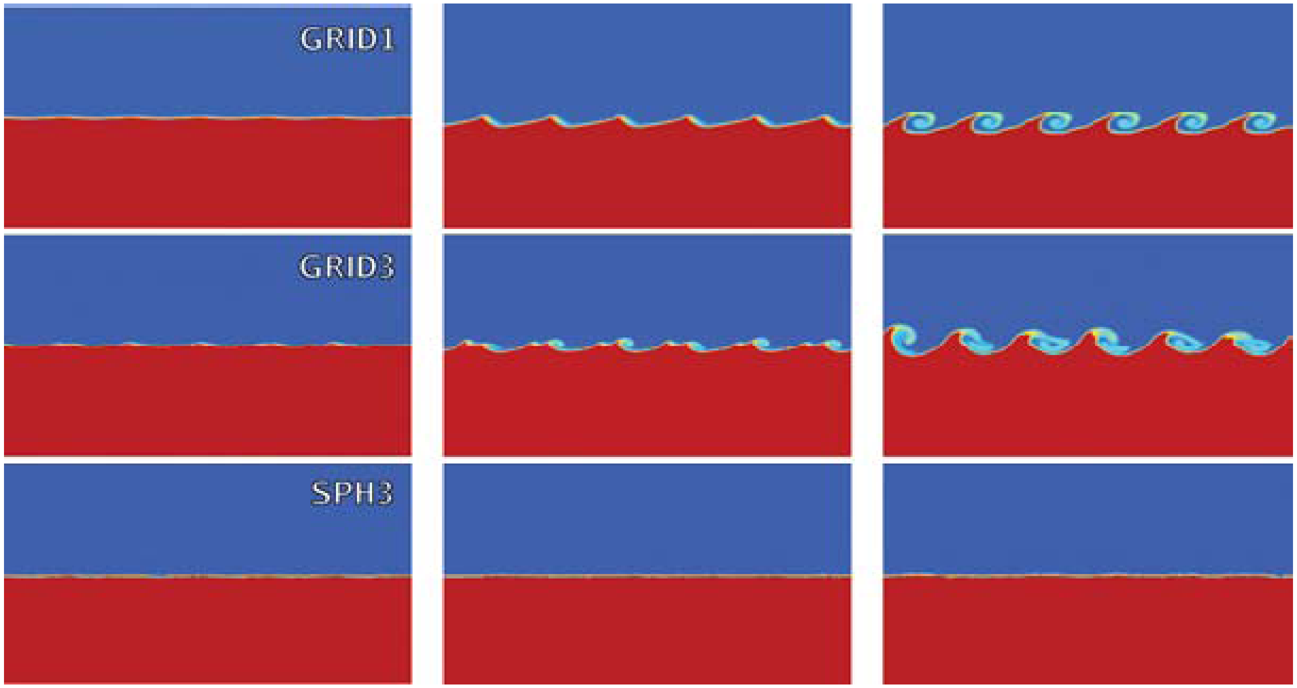
- The result of a simple "Blob test" quite different on SPH Grid
- Kelvin-Helmholtz Instability is not correctly handled with SPH
- Is SPH usable?

Difference (1)



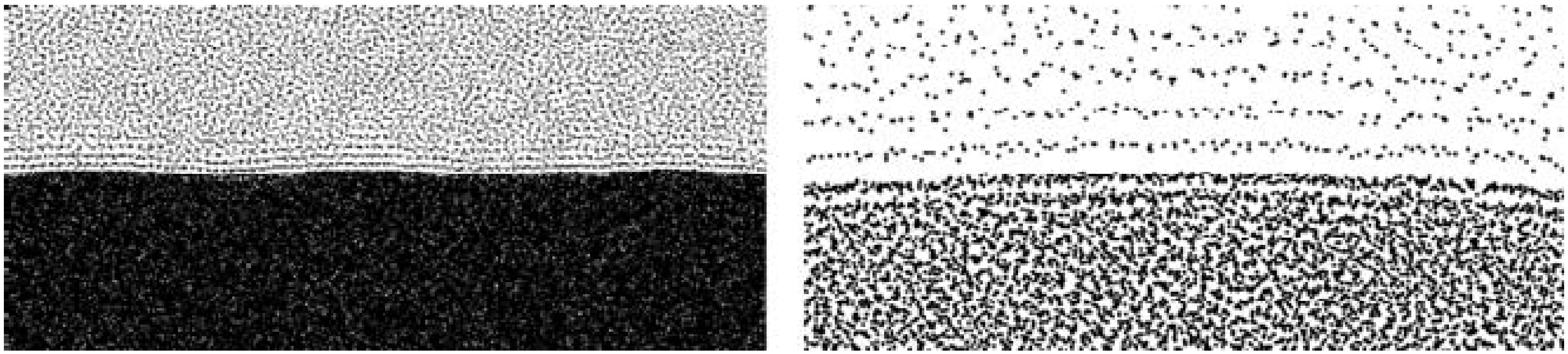
- Let a cold cloud (Temperature 1/10, density 10x) move with a supersonic velocity
- Upper three: Grid
- Lower two: SPH (1 and 10M particles)
- SPH suppresses the KHI at the fluid boundary

How different? (2)



SPH suppress KHI

How different? (3)



Strange-looking gap of particles at the two-fluid boundary.

Why does this happen?

Fundamental problem with SPH approximation

101 of SPH

Density estimate

$$\rho(x) = \sum_j m_j W(x - x_j), \tag{1}$$

Estimate of a quantity f

$$\langle f \rangle(x) = \int f(x') W(x - x') dx'. \tag{2}$$

101 of SPH continued(1)

grad of f: $\langle \nabla f \rangle = \nabla \langle f \rangle$ use the following identity

$$1 = \sum_j m_j \frac{1}{\rho(x)} W(x - x_j). \tag{3}$$

and with a bit more approximation we have

$$\langle \nabla f \rangle(x) \sim \sum_j m_j \frac{f(x_j)}{\rho(x_j)} \nabla W(x - x_j). \tag{4}$$

101 of SPH continued(2)

Equation of motion evaluates $-\frac{1}{\rho}\nabla P$. Use the identity

$$\frac{1}{\rho}\nabla P = \frac{P}{\rho^2}\nabla\rho + \nabla\frac{P}{\rho^2}. \tag{5}$$

and symmetrize. The we have

$$\dot{v}_i = -\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right)\frac{\partial}{\partial x_i}W(x_i - x_j), \tag{6}$$

Contact discontinuity

Standard SPH assumes the differentiability of ρ in the following two identities

$$1=\sum_j m_j\frac{1}{\rho(x)}W(x-x_j). \tag{7}$$

$$\frac{1}{\rho}\nabla P=\frac{P}{\rho^2}\nabla\rho+\nabla\frac{P}{\rho^2}. \tag{8}$$

Density estimated with SPH is smoothed

- Density in the low- (high-) density side (near CD) is over- (under-)estimated,
- Therefore, pressure and its derivatives have $O(1)$ errors, and particles are redistributed.

Solution?

"Fundamental" reason

ρ is smoothed, but u contains jump

We could solve the problem by smoothing u. Several proposals

- Use kernel-estimated u
- Let u diffuse (artificial conductivity)
- Use density which is continuous at CD.

Sort of working, but not a "true" solution.

Our proposal: Basic idea

At CD, there is not jump in the pressure or internal energy. Only the density jumps. Why SPH approximation breaks down?

Because we use density to calculate other quantities.

$$\langle f \rangle(x) = \sum_j \frac{m_j f(x_j)}{\rho(x_j)} W(x - x_j). \tag{9}$$

What we do here is to replace volume element dx by $m_j/\rho(x_j)$

In principle, **ANY** quantity should by okay as far as it gives correct estimate for the volume element, but there seems to be no other quantity used in the literature.

Our proposal: Principle

What should we use instead of the mass density?

An ideal gas is described by the equation of state $PV = nRT$. Here, mass density does not appear. The RHS is the thermal energy.

Can't we use the pressure itself, which is equivalent to the energy density?

Each SPH particle has energy (or entropy). So we can evaluate pressure distribution without using mass density.

Pressure is continuous at CD. So there can be no large error.

Formulation (1)

Define internal energy per particle as

$$U_j = m_j u_j, \tag{10}$$

(u is per unit mass). Define the energy density as

$$q = \sum_j U_j W(x - x_j). \tag{11}$$

Other quantities can be calculated as

$$\langle f \rangle(x) = \sum_j \frac{U_j f(x_j)}{q(x_j)} W(x - x_j), \tag{12}$$

Spacial derivatives are given by

$$\langle \nabla f \rangle(x) = \sum_j \frac{U_j f(x_j)}{q(x_j)} \nabla W(x - x_j). \tag{13}$$

Formulation (2)—Energy Equation

$$\frac{du}{dt} = -\frac{P}{\rho}\nabla \cdot v. \tag{14}$$

The divergence of the velocity is given by

$$\nabla \cdot v = \sum_j (v_i - v_j)\frac{U_j}{q_j}\nabla W(x - x_j). \tag{15}$$

P/ρ **is calculated as follows. Using EOS**

$$P_i = (\gamma - 1)q_i. \tag{16}$$

Formulation (3)—Energy Equation

The density appears since the LHS is per unit mass. To rewrite this to per-particle form, use

$$\boldsymbol{\rho_i = \frac{m_i q_i}{U_i}}. \tag{17}$$

Then we have

$$\boldsymbol{\dot{U}_i = \sum_j (\gamma - 1)\frac{U_i U_j}{q_j}(v_i - v_j)\nabla W(x_i - x_j)}. \tag{18}$$

Formulation (4)—Equation of Motion

From Energy equation we derive EoM using energy conservation. Energy change of two particles, due to the interaction between them are

$$\dot{U}_{ij} + \dot{U}_{ji} = (\gamma - 1) U_i U_j \left(\frac{1}{q_i} + \frac{1}{q_j} \right) (v_i - v_j) \nabla W (x_i - x_j). \quad (19)$$

This should be equal to the change of the kinetic energy

$$\frac{m_i m_j}{m_i + m_j} (v_i - v_j)(\dot{v}_i - \dot{v}_j). \quad (20)$$

Therefore, velocity change is

$$(\dot{v}_i - \dot{v}_j) = -(\gamma - 1) \frac{m_i + m_j}{m_i m_j} U_i U_j \left(\frac{1}{q_i} + \frac{1}{q_j} \right) \nabla W (x_i - x_j), \quad (21)$$

Formulation (5)—Equation of Motion

Using the conservation of the center of mass we have

$$\boldsymbol{m_i \dot{v}_i = -\sum_j (\gamma - 1) U_i U_j \left(\frac{1}{q_i} + \frac{1}{q_j}\right) \nabla W (x_i - x_j).} \quad (22)$$

- **RHS does not depend on mass**
- **This form is symmetric (between i and j particles)**

Examples

Standard SPH1
New SPH1
Standard SPH2
New SPH2

Non-ideal Equation of state

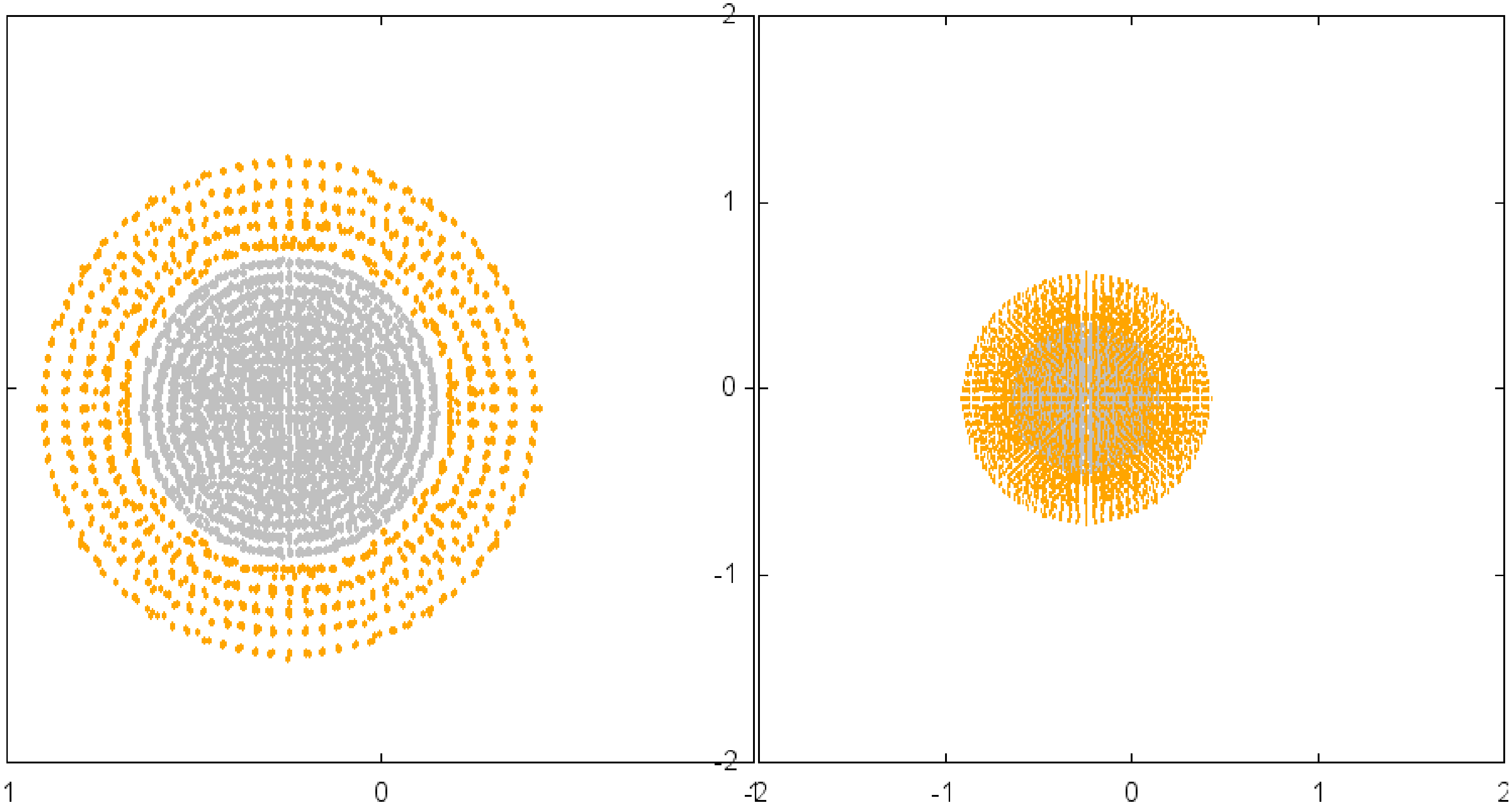
Hosono et al. 2013 (PASJ accepted)

- Integrate the internal energy and pressure separately
- Force them to be consistent at the end of each timestep

A possible different approach: integrate entropy only

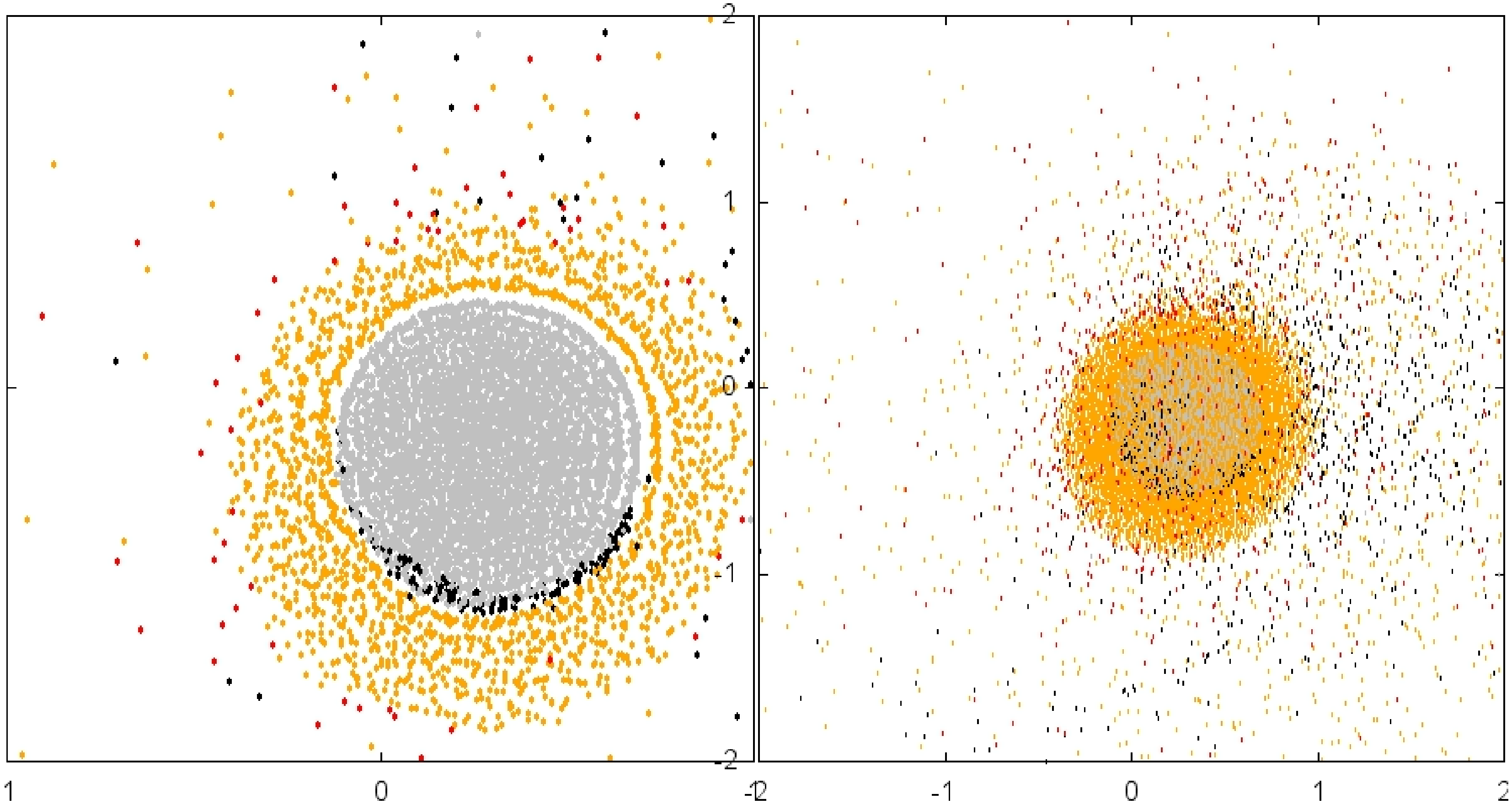
GI simulation

From "the most favorable" model of Canoop et al (2004?)

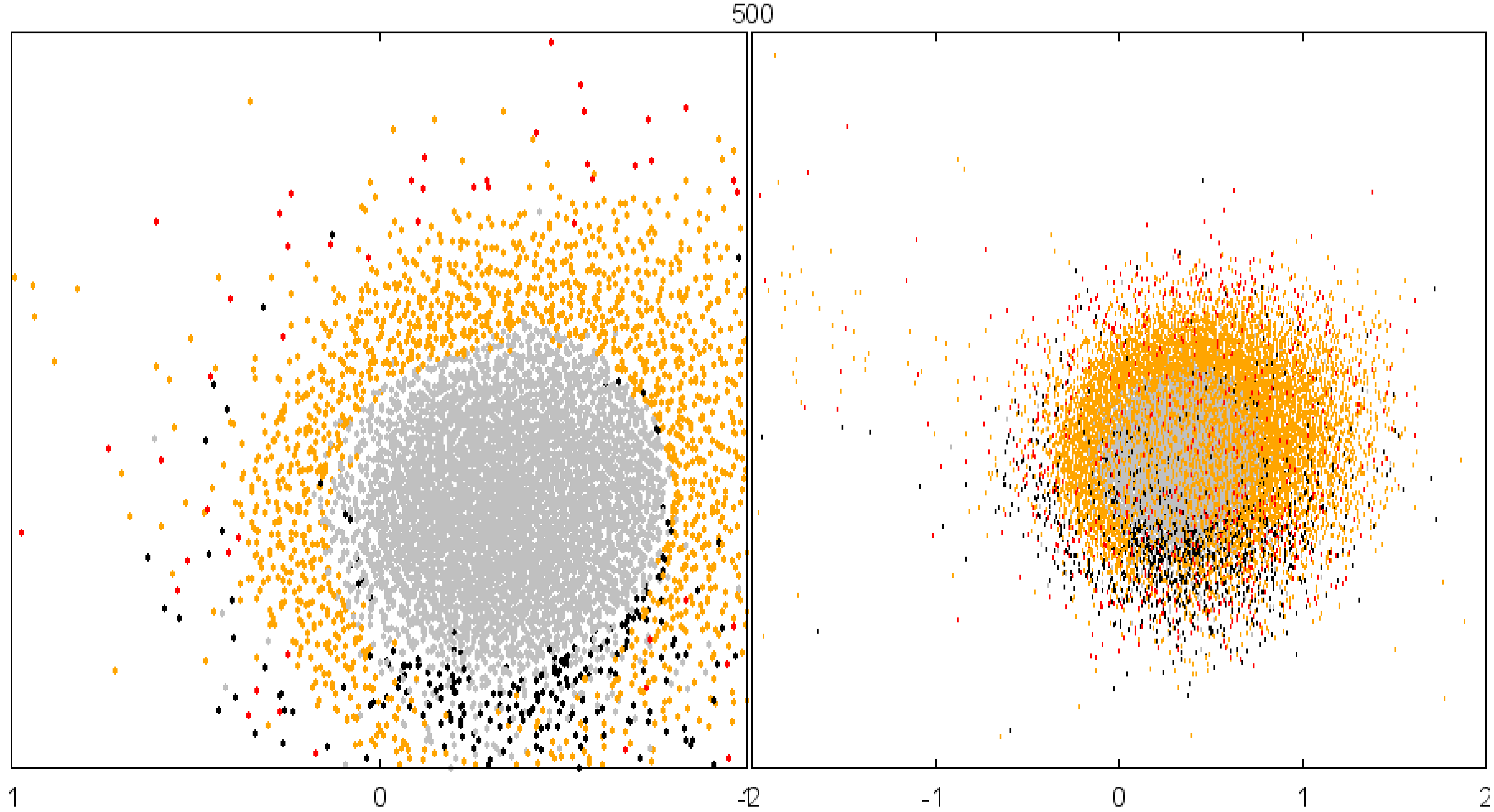


Standard SPH
New SPH

Standard SPH)



DISPH (Our method)



Results (preliminary)

- **Debris disk is much smaller with new SPH**
- **In previous calculations with standard SPH, both impactor core and mantle particles gain too much angular momenta**
- **Probably we need much higher specific angular momentum in impactor orbit**
- **Not clear if we can form moon or not**

Summary
(for the SPH part of the talk)

- We developed new SPH which can handle contact discontinuity
- It gives a result completely different from that of traditional SPH.
- We do not know if we can form moon through GI...