Yukawa Institute for Theoretical Physics, Kyoto University
2-5 June 2025

Fundamentals of cosmological simulation

Qiao WANG

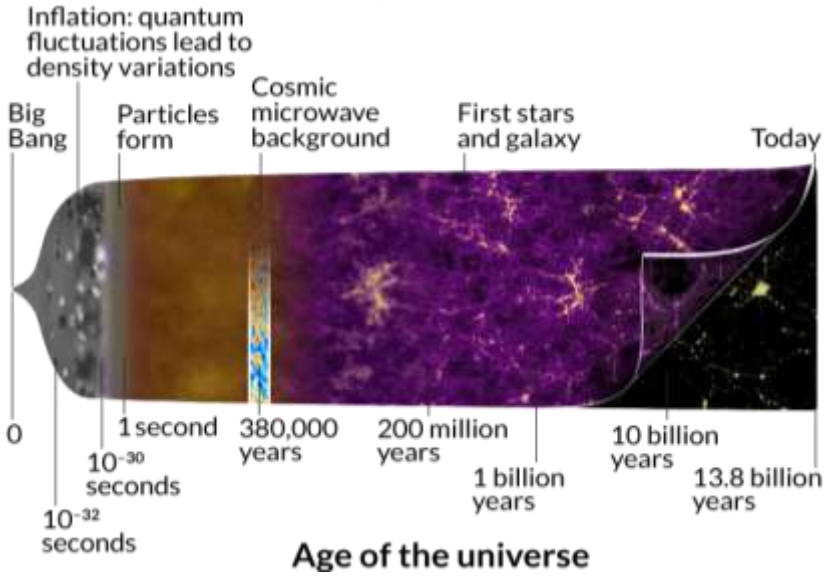
(National Astronomical Observatories, CAS)

NAOC 中国科学院国家天文台 NATIONAL ASTRONOMICAL OBSERVATORIES, CAS

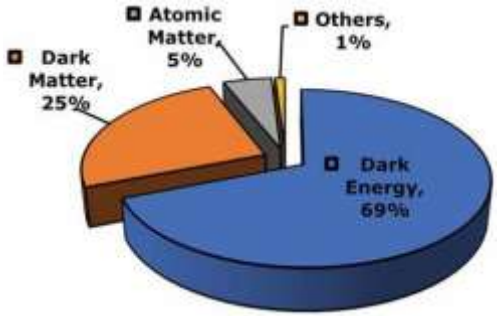
contents

- Background
- Methods of cosmological N-Body simulation
- PhotoNs code & Hyper-Millennium run
- Summary

Modern cosmology, Tensions & Nature of Dark sector

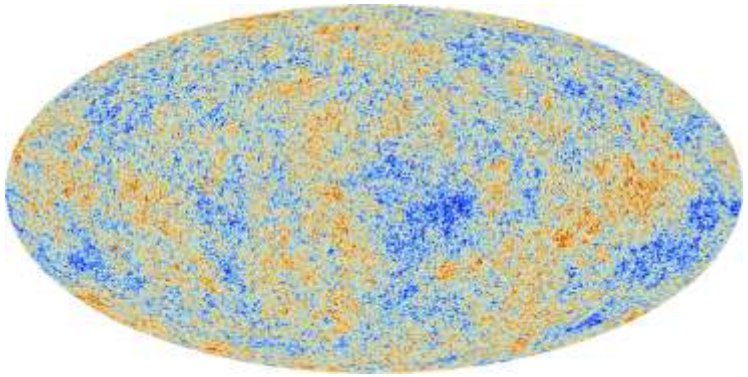


Timeline of the universe
Expansion history

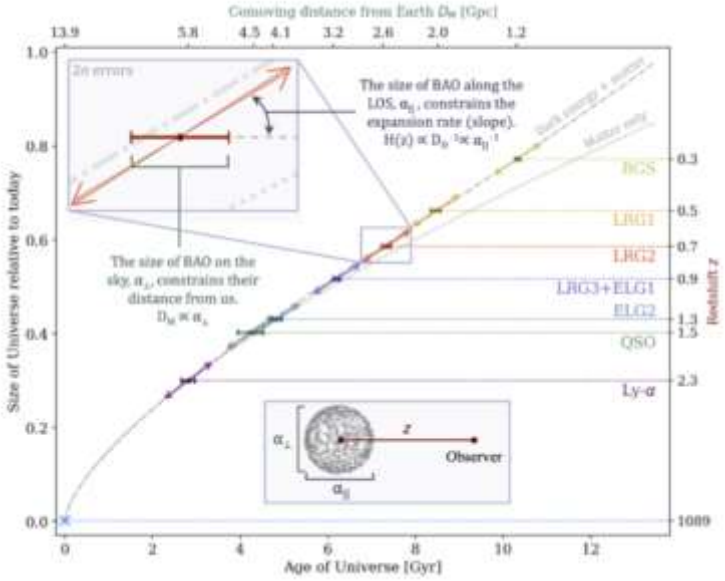


Mass-energy inventory
Standard model of ΛCDM

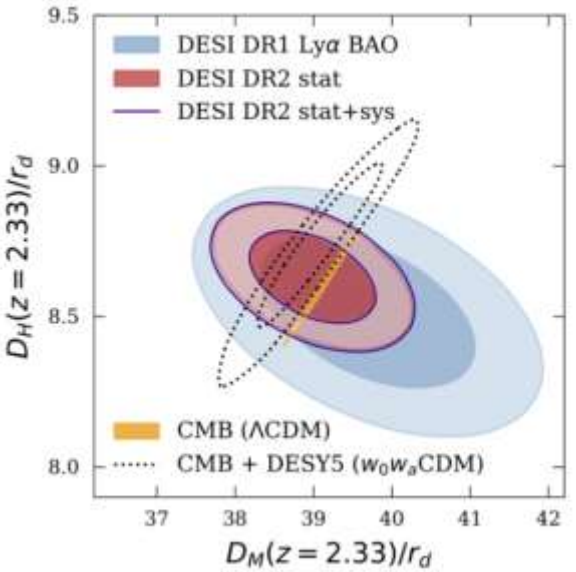
cosmic probes and parameter constraints



Planck ’18



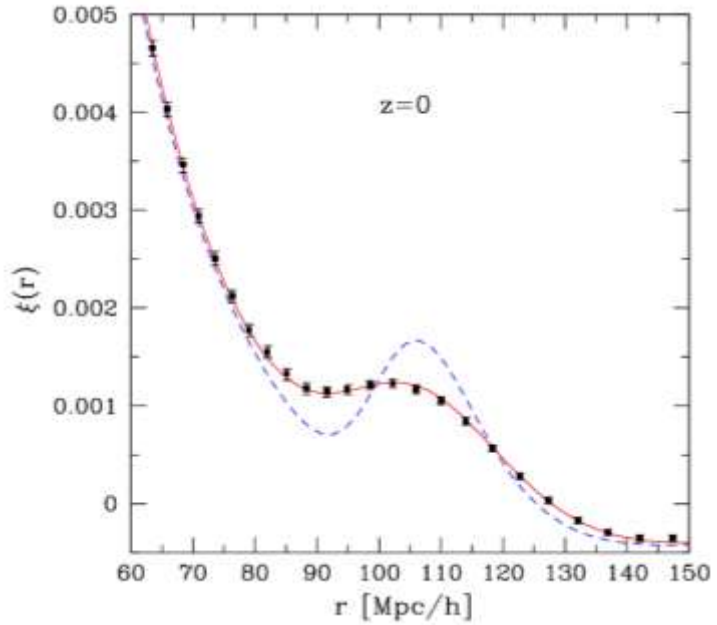
BAO (DESI DR2)



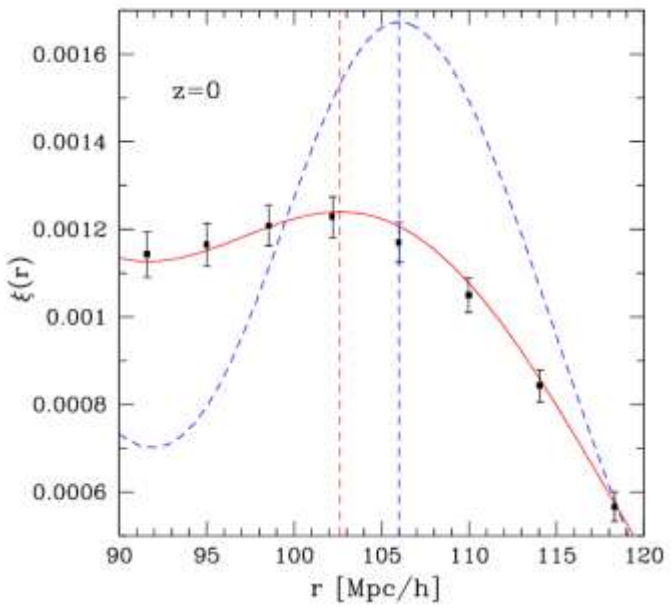
Joint constraint

See previous lectures

Nonlinear evolution



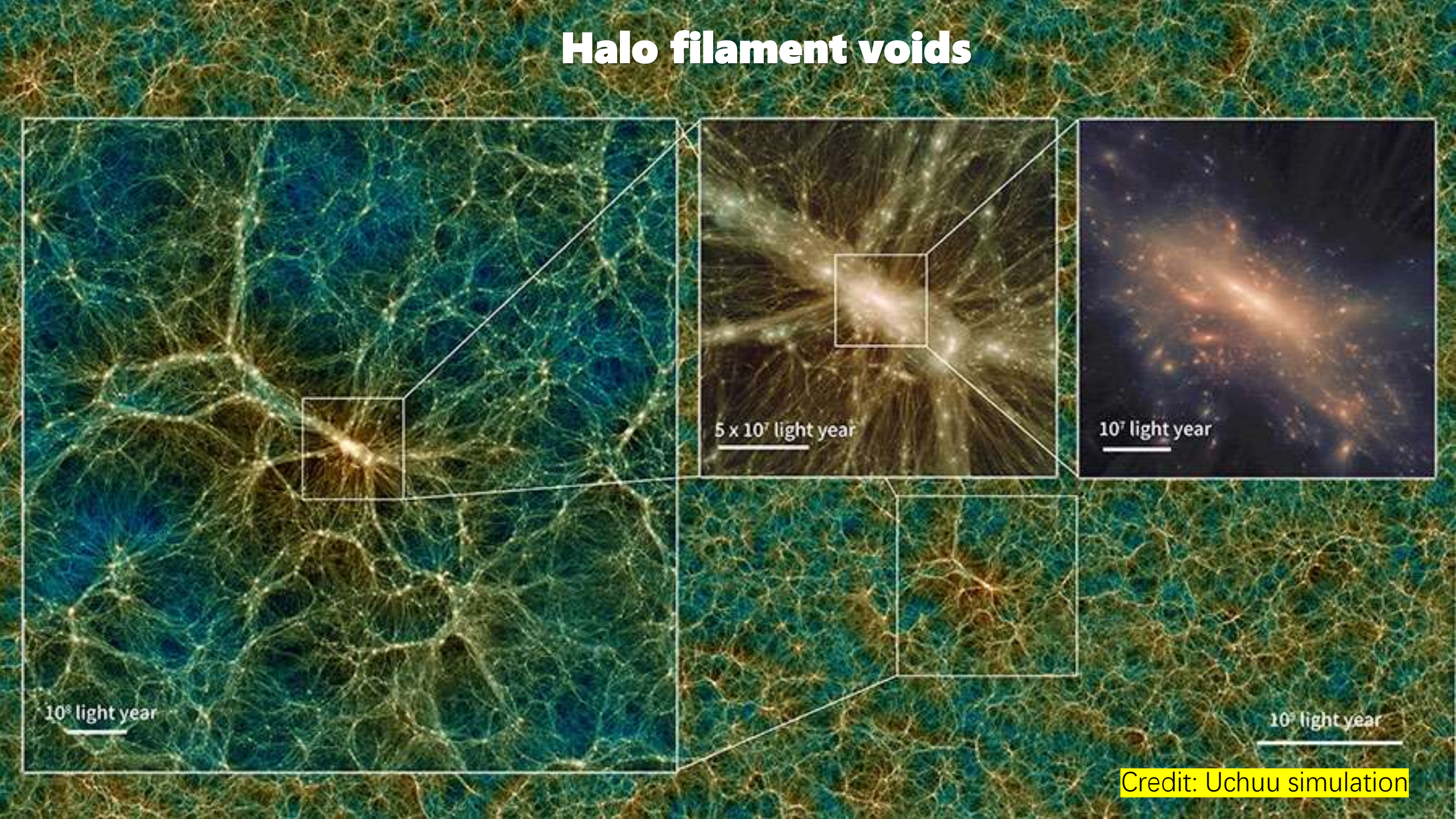
damping the amplitude



Peak shifted

Crocce & Scoccimarro 2007

Halo filament voids

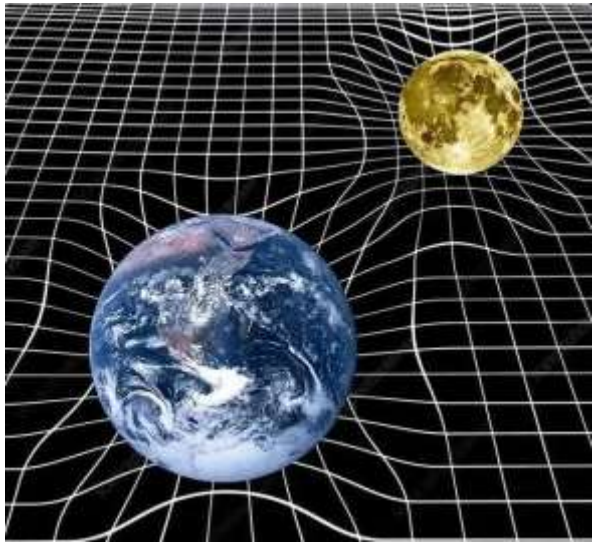


Credit: Uchuu simulation

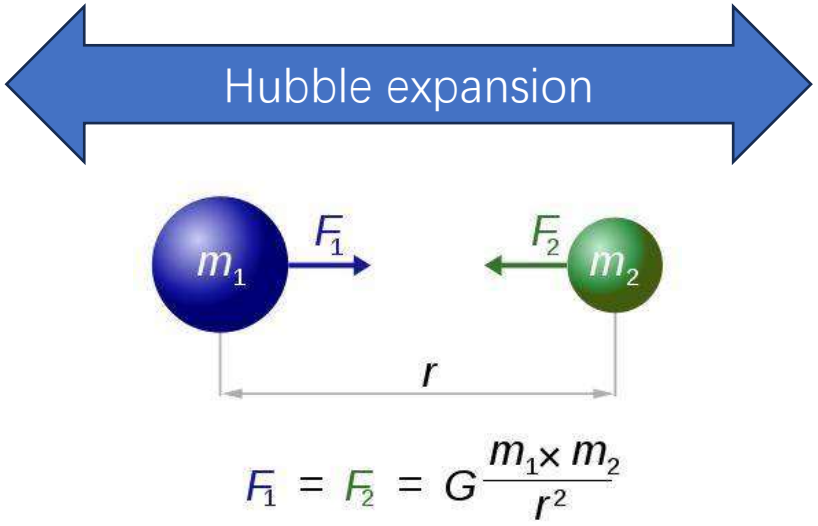
Motivation to cosmological simulation

- Nonlinear evolution of DM in expanding universe
- Solve the DM distribution on small-scale
- Platform for astrophysics, especially galaxy formation
- Mock galaxy catalogue by HOD, SHAM or SAM (previous lecture)
- Calibration of the systematic error of cosmic probes or various surveys or detections.

Framework of cosmological simulation



General Relativistic view



Newtonian gravity in expanding space

$$ds^2 = -a^2(1+2\psi)\,d\eta^2 + a^2(1+2\phi)\,\bar{g}_{\alpha\beta}dx^\alpha dx^\beta$$

N-Body method of cosmological simulation

$$\dot{\boldsymbol{r}} = \frac{\boldsymbol{p}}{a^2} \longrightarrow \rho(\boldsymbol{r}) = \sum_{\forall i} \frac{m_i}{a^3} \delta_{\mathrm{D}}(\boldsymbol{r} - \boldsymbol{r}_i)$$

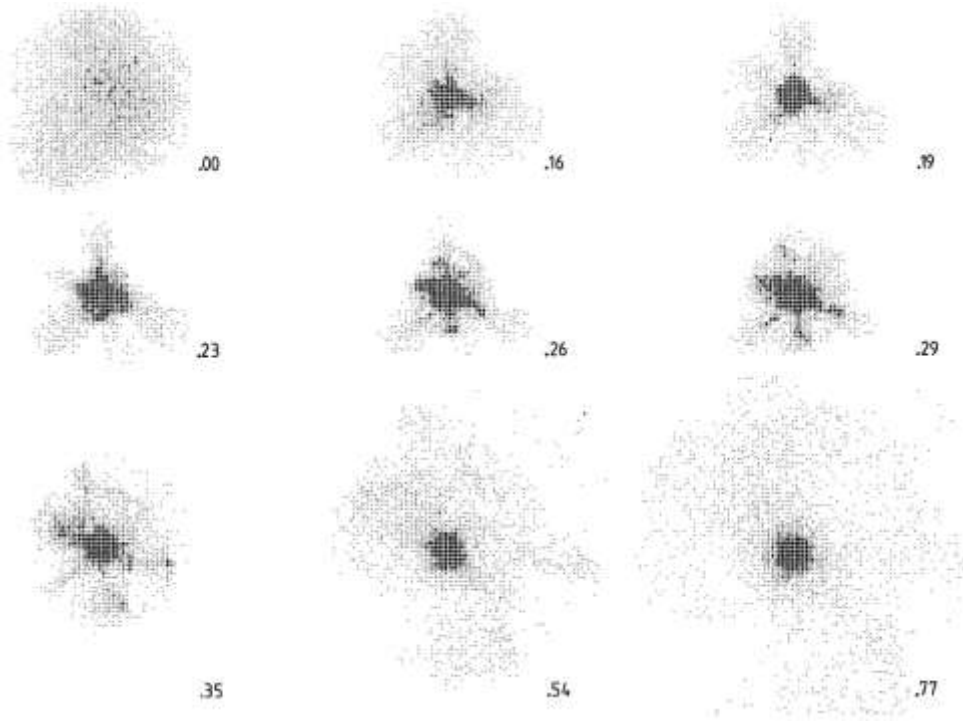
$$\dot{\boldsymbol{p}} = -\frac{\nabla\phi}{a} \longleftarrow \nabla^2\phi = 4\pi G a^3 \left[\rho_m(\boldsymbol{r}, t) - \bar{\rho}_m(t)\right]$$

- N-body method is lagrangian method (mass conversion with adaptive resolution)
- Gravity is determined by the density contrast
- The motion involves scale factor a(t) of the universe

- Lack of the inner detail of DM particle (mass points)
- Relativistic effect is linear order (~< extremely large scale)
- Retarded potential is neglectable, due to density contrast and average effect

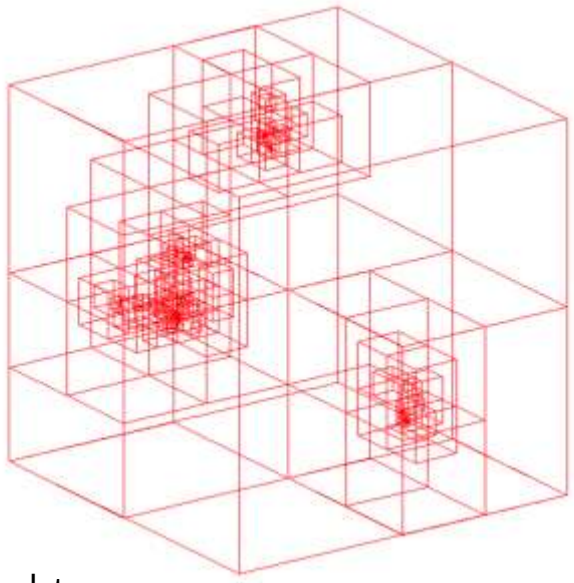
N-Body simulation

- The pairwise interaction of gravity is simply and precisely calculated.
- shortage: Direct N-Body summation is too expansive O(N^2)
- Development of approximate method is inevitable to reduce the computing amount of P2P
- Special-Purpose Computer (GRAPE)

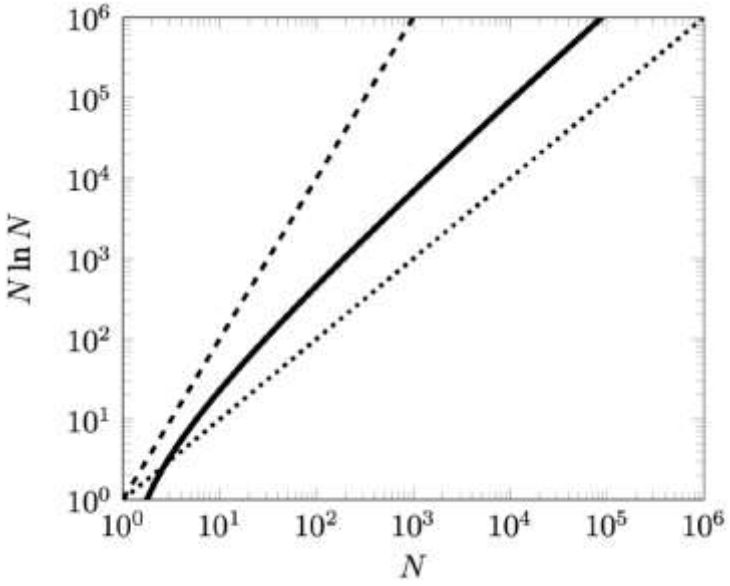


Tree code

- The particles are organized into a tree as leaves [Barnes & Hut 1986]
- The distant tree node is accepted as a mass point away from the target particle. Otherwise, the tree node will be open
- time complexity of tree code ~O(NlogN) (divide-and-conquer algorithm)

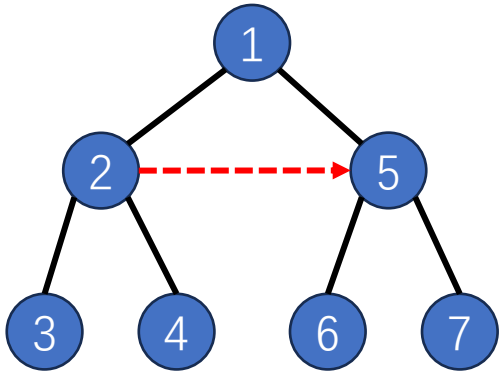


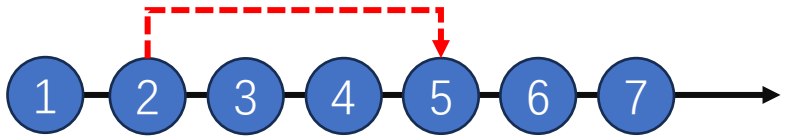
Octal tree



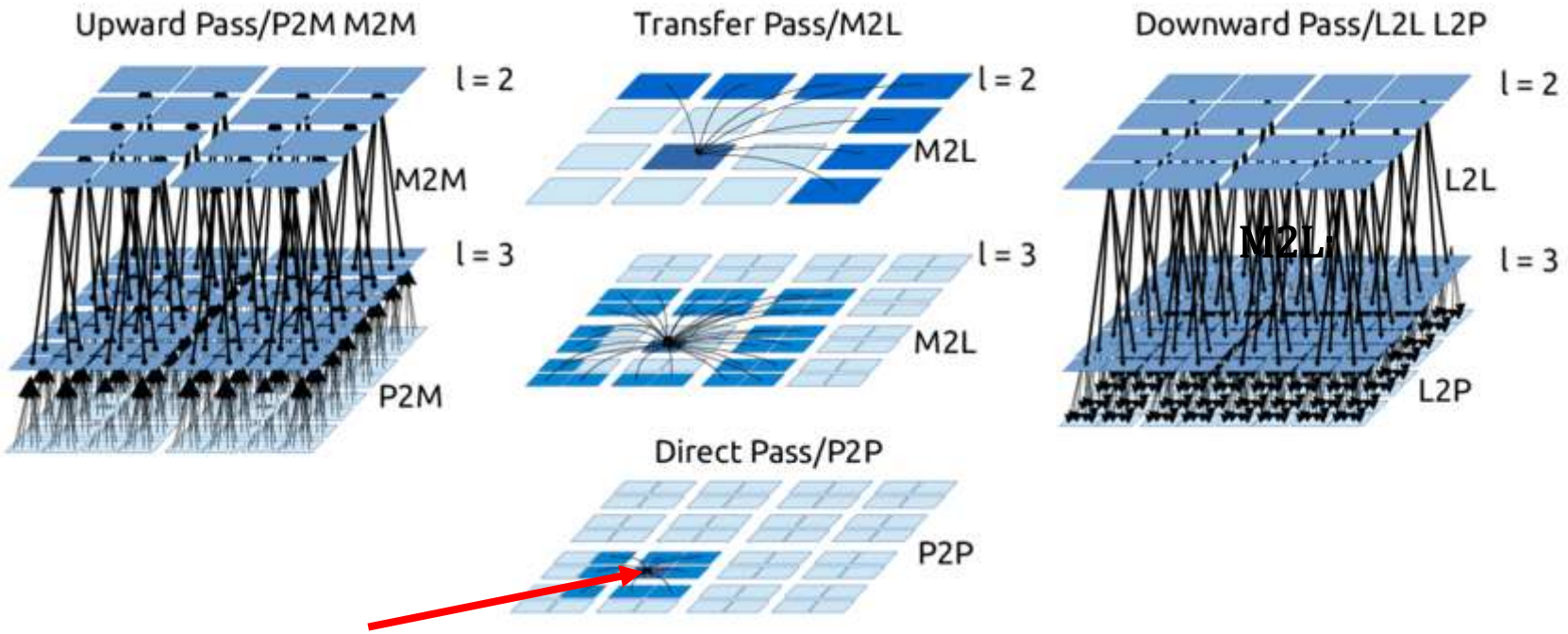
Tree code

- The particles are organized into a tree as leaves [Barnes & Hut 1986]
- The distant tree node is accepted as a mass point away from the target particle. Otherwise, the tree node will be open
- time complexity of tree code ~O(NlogN) (divide-and-conquer algorithm)
- Recursive traversal → iteration by preprocessing one time.





Fast Multipole Method (FMM)



$$\mathcal{L}_{\mathbf{n}}(\mathbf{r}_B) = \sum_{|\mathbf{m}|=0}^{p-|\mathbf{n}|} \mathcal{M}_{\mathbf{m}}(\mathbf{r}_A)\mathcal{D}_{\mathbf{n}+\mathbf{m}}(\mathbf{r}_B - \mathbf{r}_A),$$

- Pass up : P2M , M2M
- Pass down : L2L, L2P
- Treecode: M2P

Information transfer

- Greengard & Rokhlin, JCP, 1987 (2D) classic FMM
- Dehnen 2014 (stellar dynamics, 3D)
- Yokota & barba 2010 (GPU implementation)
- PKDGRAV code

Fast Multipole Method ~O(p*N)

Particle-mesh method

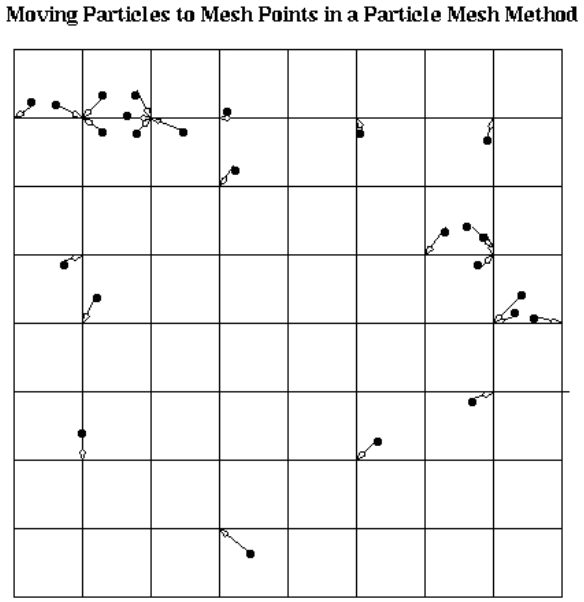
Poisson equation (Linear differential operator)

$$\nabla^2 \Phi = 4\pi G \rho$$

- Density field on a periodic regular mesh
- Convolution by using FFT ~O(NlogN)
- Interaction between a particle and potential mesh

Particle-mesh method

To construct density field from particle to a regular mesh



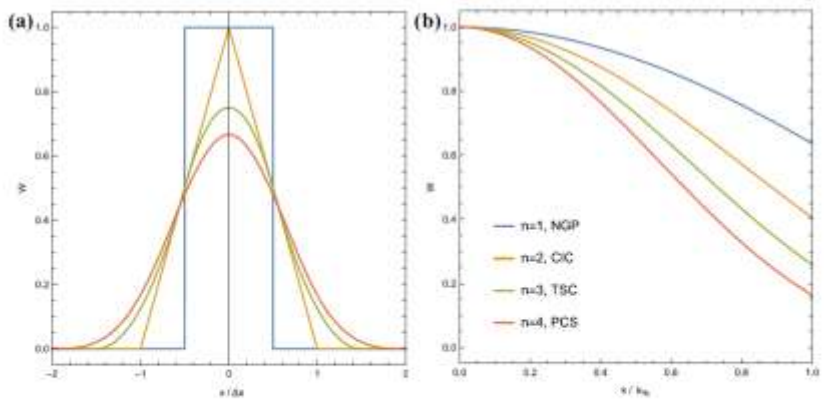


Fig. 6 Common particle-mesh mass assignment kernels in real space (panels a), and Fourier space (panels b) of increasing order: $n = 1$ NGP, $n = 2$ CIC, $n = 3$ TSC, $n = 4$ PCS. Note that the NGP kernel is not continuous, CIC is continuous but not differentiable, TSC is continuously differentiable, and PCS is twice differentiable. The support of the assignment functions is $n\Delta x$ per dimension, and they converge to a normal distribution for $n \to \infty$. Due to their increasing smoothness, they also act as increasingly stronger low-pass filters

$$W_{\rm NGP}(x) = \frac{1}{h}\begin{cases} 1 & \text{for } |x| \le \frac{\Delta x}{2} \\ 0 & \text{otherwise} \end{cases}$$

$$W_{\rm CIC}(x) = \frac{1}{h}\begin{cases} 1 - \frac{|x|}{\Delta x} & \text{for } |x| < \Delta x \\ 0 & \text{otherwise} \end{cases}$$

$$W_{\rm TSC}(x) = \frac{1}{\Delta x}\begin{cases} \frac{3}{4} - \left(\frac{x}{\Delta x}\right)^2 & \text{for } |x| \le \frac{\Delta x}{2} \\ \frac{1}{2}\left(\frac{3}{2} - \frac{|x|}{\Delta x}\right)^2 & \text{for } \frac{\Delta x}{2} \le |x| < \frac{3\Delta x}{2} \\ 0 & \text{otherwise} \end{cases}$$

Angulo & Hahn 2022

Fourier Transformation convolution

$$\mathscr{F}[f'(t)] = \int_{-\infty}^{+\infty} f'(t)e^{-j\omega t}dt$$
$$= f(t)e^{-j\omega t}\Big|_{-\infty}^{+\infty} + j\omega\int_{-\infty}^{+\infty} f(t)e^{-j\omega t}dt$$
$$= j\omega\mathscr{F}[f(t)]$$

differential operators can be transferred into “multiplication”

$$\mathscr{F}[f^{(n)}(t)] = (\mathrm{j}\omega)^n\mathscr{F}[f(t)]$$

$$\nabla^2\Phi = 4\pi G\rho$$

$$\partial \rightarrow ik \qquad \hat{\Phi} = -4\pi G\frac{\hat{\rho}}{k^2}$$

$$\tilde{\phi}_{\boldsymbol{n}} = \begin{cases} -\dfrac{4\pi G}{ak_0^2}\dfrac{\tilde{\rho}_{\boldsymbol{n}}}{\|\boldsymbol{n}\|^2} & \text{if} \quad \boldsymbol{n} \neq \boldsymbol{0} \\ 0 & \text{otherwise} \end{cases}$$

on regular mesh

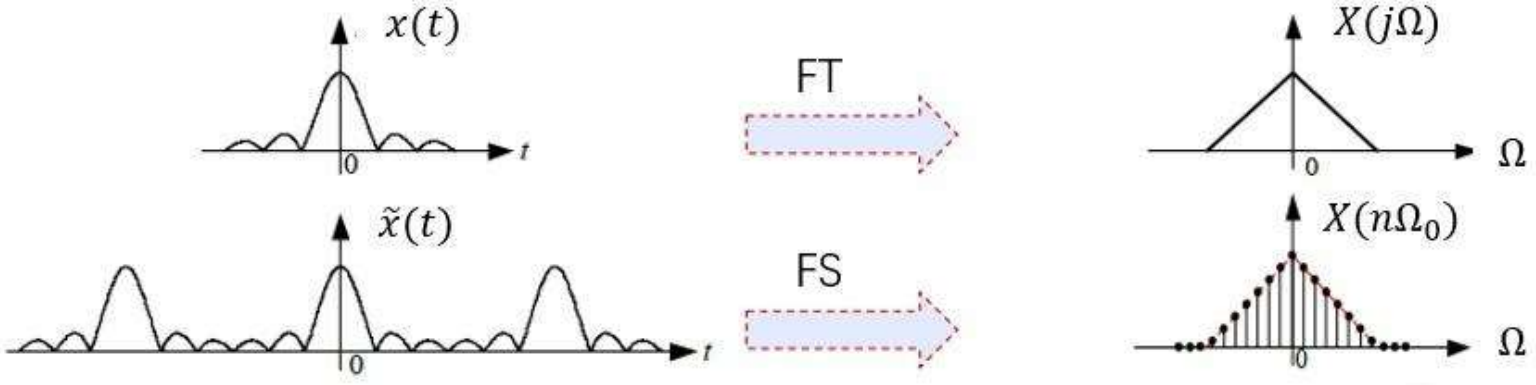
$$(ik)^2 \rightarrow -n^2$$

$$\tilde{\phi}_{\boldsymbol{n}}^{\mathrm{FD2}} = \begin{cases} -\dfrac{\pi G\Delta x^2}{a}\,\tilde{\rho}_{\boldsymbol{n}}\left(\sin^2\left[\dfrac{\pi n_x}{N_g}\right] + \sin^2\left[\dfrac{\pi n_y}{N_g}\right] + \sin^2\left[\dfrac{\pi n_z}{N_g}\right]\right)^{-1} \\ 0 \end{cases}$$

second order accurate finite difference Laplacian in Fourier space

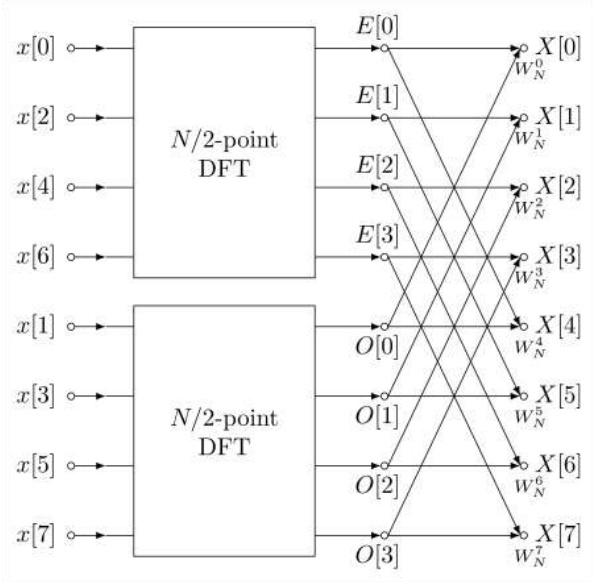
Discrete/Fast Fourier Transformation

discrete sampling



odd-even symmetry
O(N^2) => O(NlogN)

Cooley-Tukey algorithm



$$X_k = \underbrace{\sum_{m=0}^{N/2-1} x_{2m} e^{-\frac{2\pi i}{N/2} mk}}_{\text{DFT of even-indexed part of } x_n} + e^{-\frac{2\pi i}{N} k} \underbrace{\sum_{m=0}^{N/2-1} x_{2m+1} e^{-\frac{2\pi i}{N/2} mk}}_{\text{DFT of odd-indexed part of } x_n}$$

$$= E_k + e^{-\frac{2\pi i}{N} k} O_k \qquad \text{for } k = 0, \ldots, \frac{N}{2} - 1.$$

Butterfly for radix-2 FFT (from Wiki)

Scalability to the parallel of FFT

- Scalability issue on massively parallel implementation
- MPI communication will be dominant than computation
- A Subset corresponds to MPI-processes for FFT enlarge the particle number to trillion.

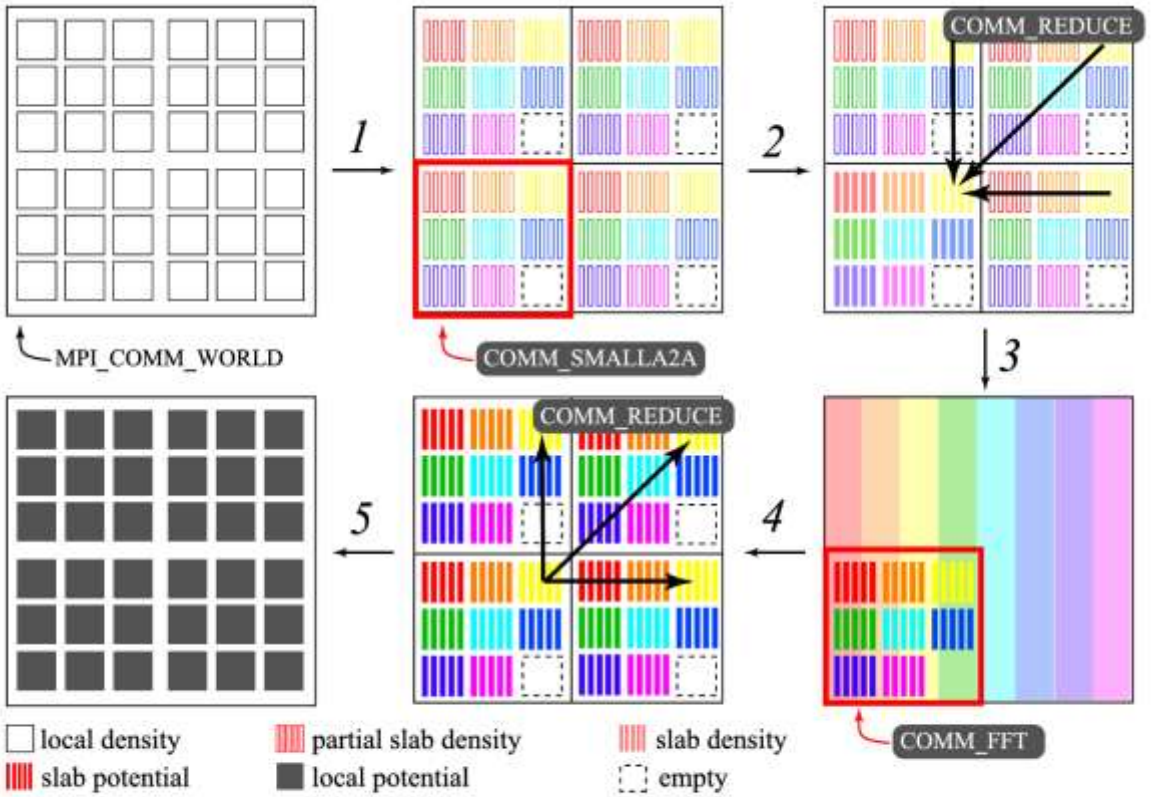
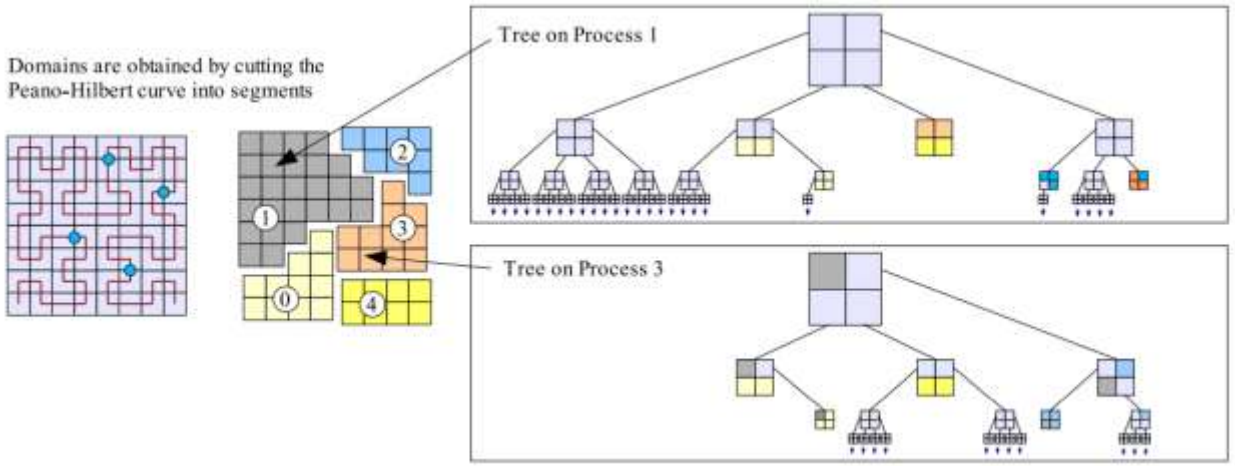


Fig. 5. Illustration of our communication algorithm, relay mesh method. There are 2-D decomposed 6×6 processes. The number of the PM mesh is $N_{\rm PM} = 8^3$, and that of processes perform FFT is eight. There are four groups that include nine processes. The detail explanation is written in the text. The background of bottom-right panel shows the physical regions that are corresponding to the slab density of each process in the root group.

Ishiyama+ 2014

Computing domain decomposition

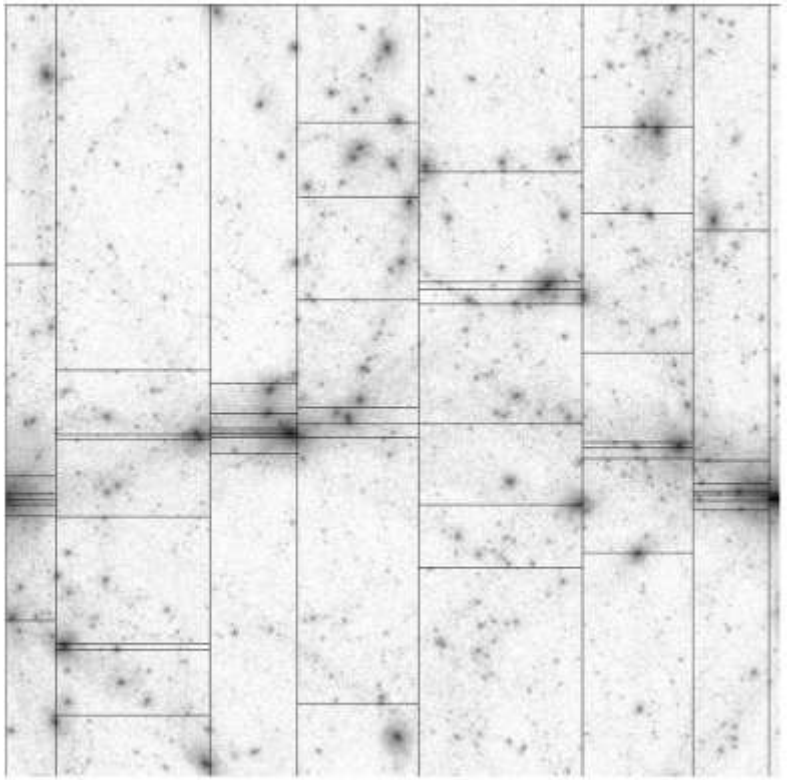
Space-filling curve



- Surface-area-to-volume ratio
- Space-filling Peano–Hilbert curve equally divides the particle distribution/work load.

adaptive straight boundary

- efficient MPI communication for massive parallel case, especially containing a FFT module

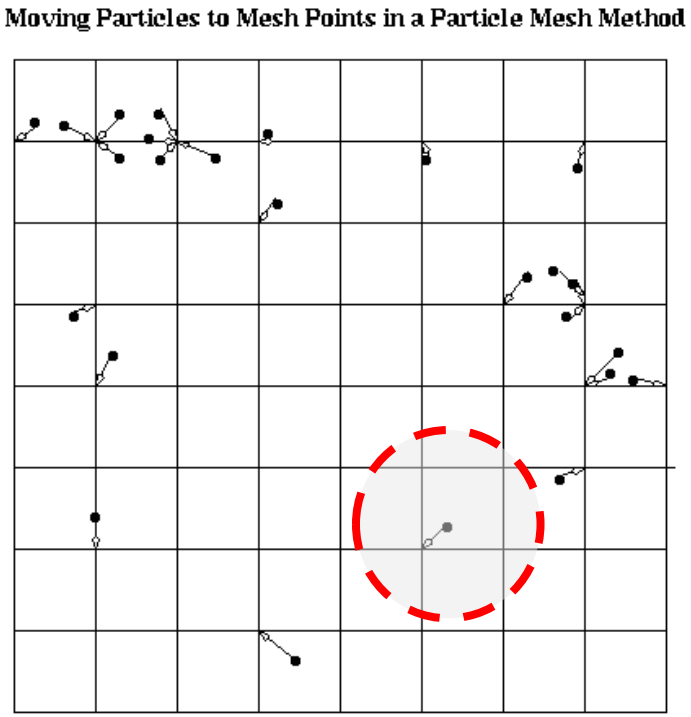


Particle-Particle + Particle-Mesh (P^3M) method

- the hybrid algorithm combined PM and P2P
- A compensated particle-particle interaction is necessary for short range interaction.

$$g_{\mathrm{P3M}}(R) = \begin{cases} 1 - \frac{1}{140}(224R^3 - 224R^5 + 70R^6 \\ +48R^7 - 21R^8) \quad \text{for} \quad 0 \le R < 1 \\ \\ 1 - \frac{1}{140}(12 - 224R^2 + 896R^3 - 840R^4 + 224R^5 \\ +70R^6 - 48R^7 + 7R^8) \quad \text{for} \quad 1 \le R < 2 \\ \\ 0 \quad \text{for} \quad 2 \le R. \end{cases} \tag{2}$$

- shortage: P2P calculation is still expansive O(N^2) once large number of particles in a single mesh.



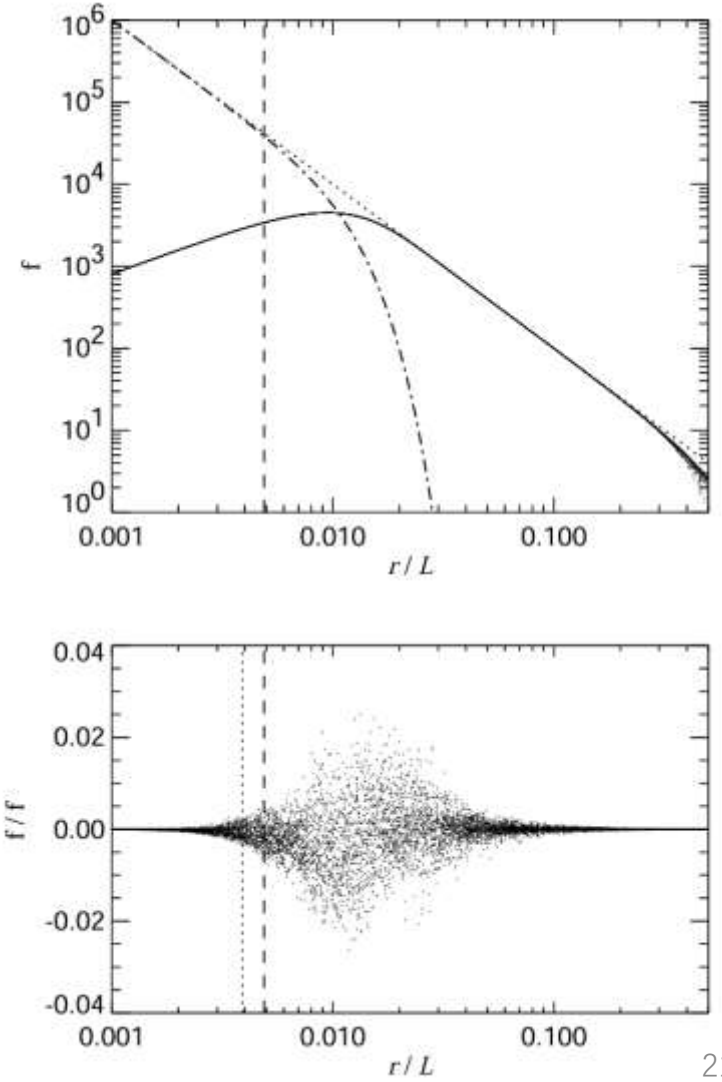
Tree-PM method

- improve the computing efficiency of short-range gravity, instead of Treecode.
- Error function (exponential) is alternative splitting formula (gird effect).

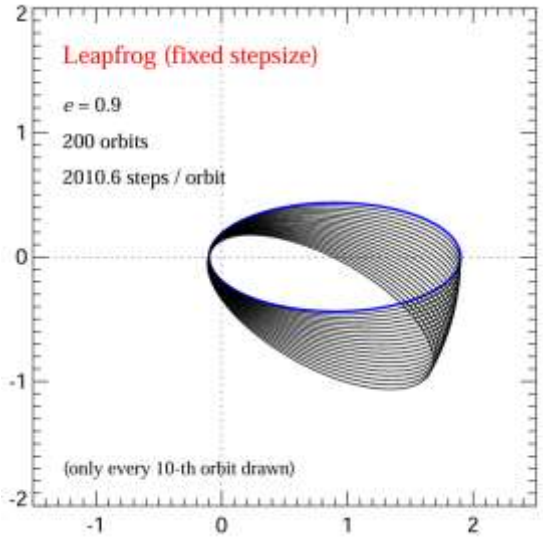
$$\phi^{\mathrm{short}}(\boldsymbol{x}) = -G\sum_i \frac{m_i}{r_i}\mathrm{erfc}\left(\frac{r_i}{2r_s}\right)$$

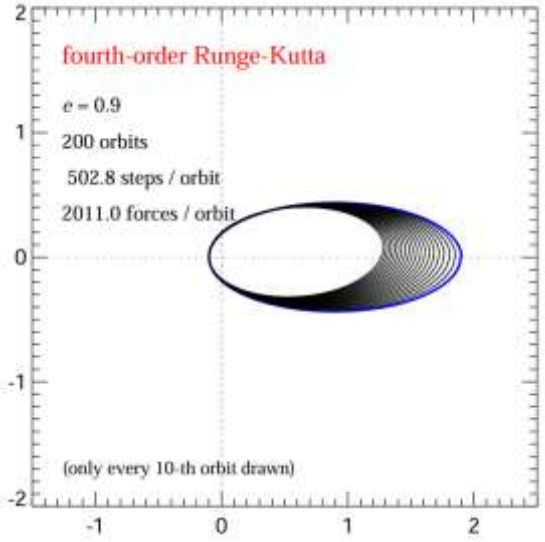
$$f_l = 1 - \mathrm{erfc}\left(\frac{r}{2\,r_s}\right) - \frac{r}{\sqrt{\pi}r_s}\exp\left(-\frac{r^2}{4\,r_s^2}\right)$$

GADGET 2-3 code
GREEM code



Sympletic time integrator





precession

dissipation

Quinn+1997

$$D_t(\Delta t): \begin{cases} \boldsymbol{p}_i & \mapsto & \boldsymbol{p}_i \\ \boldsymbol{x}_i & \mapsto & \boldsymbol{x}_i + \frac{\boldsymbol{p}_i}{m_i}\int_t^{t+\Delta t}\frac{\mathrm{d}t}{a^2} \end{cases}$$

$$K_t(\Delta t): \begin{cases} \boldsymbol{x}_i & \mapsto & \boldsymbol{x}_i \\ \boldsymbol{p}_i & \mapsto & \boldsymbol{p}_i + \boldsymbol{f}_i\int_t^{t+\Delta t}\frac{\mathrm{d}t}{a} \end{cases}$$

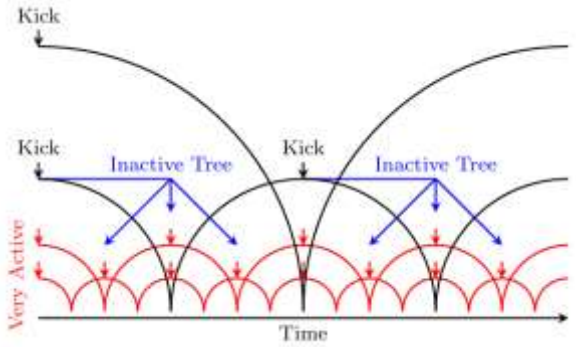
- Sympletic integrator in comoving coordinates

Sympletic time integrator

fixed step: K(Δt/2) D(Δt) K(Δt/2)

$$\Delta t_i \simeq \eta \sqrt{\frac{\varepsilon}{\|\boldsymbol{A}_i\|}}$$

- Softening scale
- Acceleration
- Control parameter



adaptive step:

$$\tilde{U}(\Delta t) = K_{\rm lr}\left(\frac{\Delta t}{2}\right)\left[K_{\rm sr}\left(\frac{\Delta t}{2m}\right) D\left(\frac{\Delta t}{m}\right) K_{\rm sr}\left(\frac{\Delta t}{2m}\right)\right]^m K_{\rm lr}\left(\frac{\Delta t}{2}\right)$$

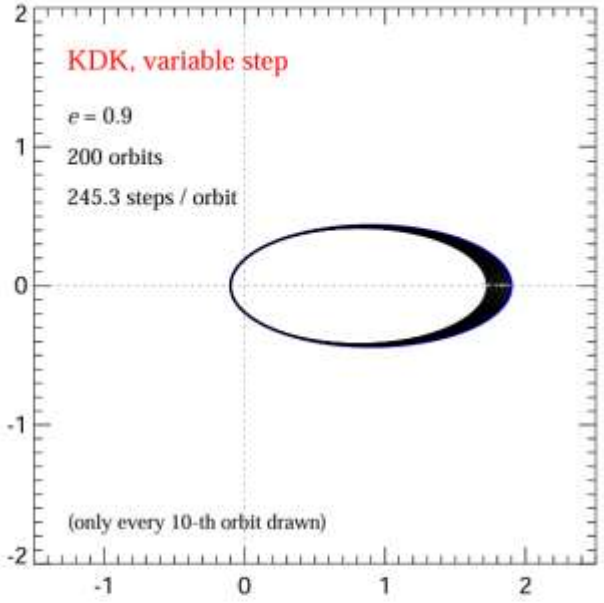
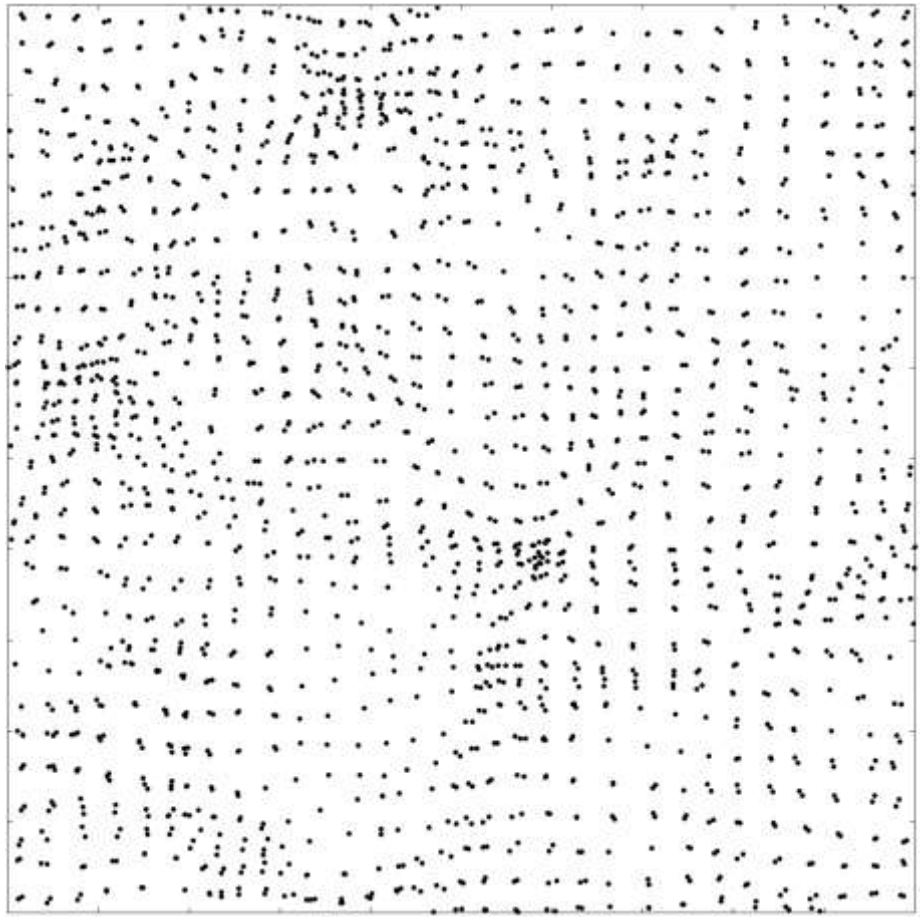


Figure 5. A Kepler problem of high eccentricity integrated with leapfrog schemes using a variable timestep from step to step, based on the $\Delta t \propto 1/\sqrt{|\boldsymbol{a}|}$ criterion commonly employed in cosmological simulations. As a result of the variable timesteps, the integration is no longer manifestly time reversible, and long term secular errors develop. Interestingly, the error in the KDK variant grows four times slower than in the DKD variant, despite being of equal computational cost.

Springel 2005

Initial Condition Generator



1. Input an initial power spectrum calculated by Boltzmann code, such as CAMB
2. Setup particles on a regular grid center
3. Zel'dovich shift

$$\mathbf{x}(\mathbf{q}, t_i) = \mathbf{q} + \Psi(\mathbf{q}, t_i), \quad \mathbf{v}(\mathbf{q}) = a\dot{\Psi}(\mathbf{q}, t_i).$$

$$\Psi(\mathbf{q}, t_i) = \mathcal{F}^{-1}\{i\mathbf{k}\hat{\delta}(\mathbf{k}, t_i)/k^2\}$$

4. 2nd order correction (optional)

   e.g., 2LPT Code

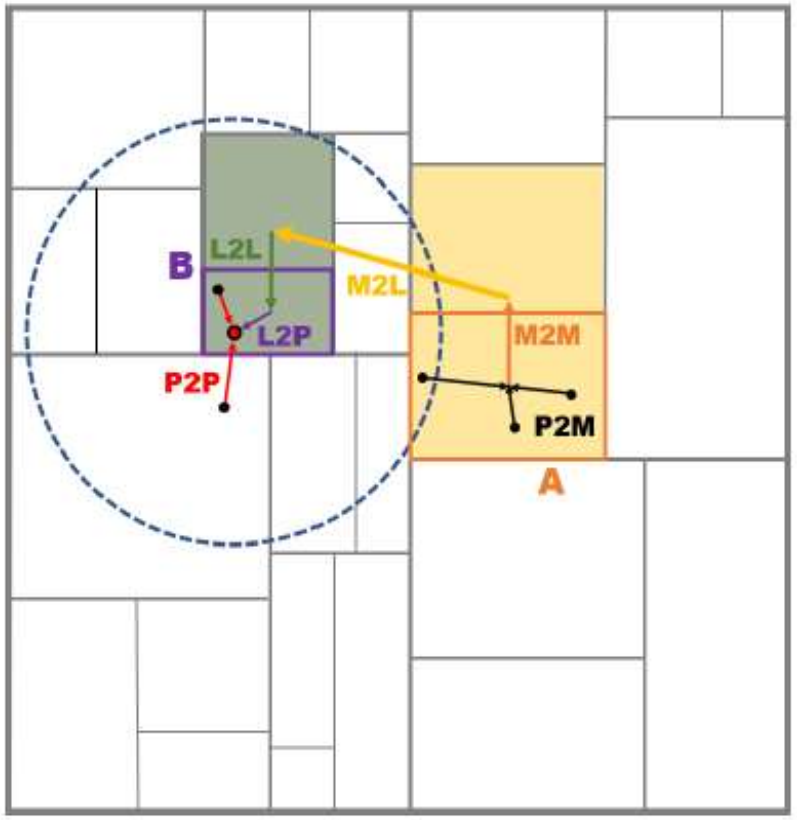
PhotoNs: cosmological simulation code

- PhotoNs-1: TreePM code aims at a heterogeneous accelerator (Tianhe2)
- PhotoNs-2: change gravity solver to PM-FMM （Taihu-light）
- PhotoNs-2-GPU: GPU implementation (CUDA/HIP)
- PhotoNs-3.7: optimize mpi communicator aim at the extreme simulation

  with post processing of halo + subfind (GPU-based supercomputer)
- Further version will involve more physics or module beyond N-Body simulation

Particle-Mesh Fast-Multipole-Method (PM-FMM)

FMM for short range
PM for long-range

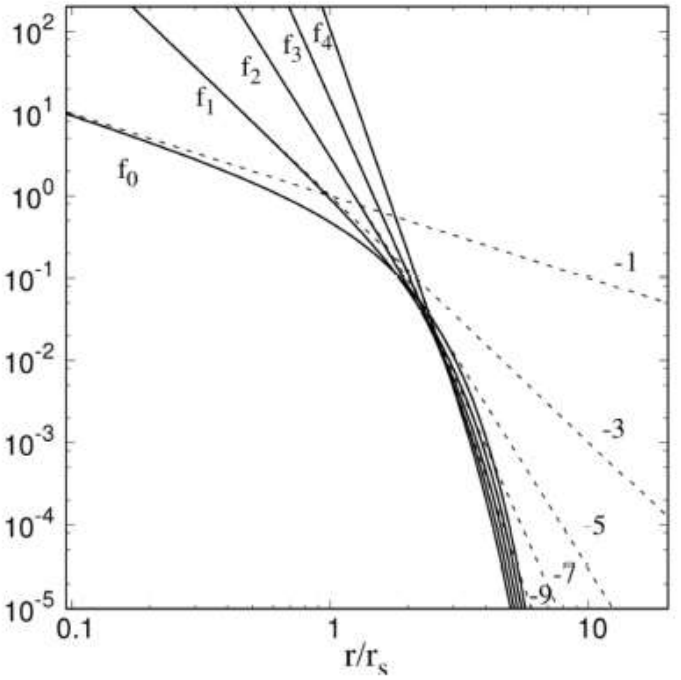


$$T \equiv \mathrm{erfc}\left(\frac{r}{2r_s}\right) + \frac{r}{r_s\sqrt{\pi}}\exp\left(-\frac{r^2}{4r_s^2}\right)$$

$$f_{(2)}(r) = \frac{3}{r^5}\mathrm{erfc}\left(\frac{r}{2r_s}\right) + \frac{1}{\sqrt{\pi}}\exp\left(-\frac{r^2}{4r_s^2}\right) \times \left[\frac{3}{r_s r^4} + \frac{1}{2r_s^3 r^2}\right],$$

$$f_{(3)}(r) = -\frac{15}{r^7}\mathrm{erfc}\left(\frac{r}{2r_s}\right) - \frac{1}{\sqrt{\pi}}\exp\left(-\frac{r^2}{4r_s^2}\right) \times \left[\frac{15}{r_s r^6} + \frac{5}{2r_s^3 r^4} + \frac{1}{4r_s^5 r^2}\right],$$

$$f_{(4)}(r) = \frac{105}{r^9}\mathrm{erfc}\left(\frac{r}{2r_s}\right) + \frac{1}{\sqrt{\pi}}\exp\left(-\frac{r^2}{4r_s^2}\right) \times \left[\frac{105}{r_s r^8} + \frac{35}{2r_s^3 r^6} + \frac{7}{4r_s^5 r^4} + \frac{1}{8r_s^7 r^2}\right].$$



Q. Wang, 2021

Particle-Mesh Fast-Multipole-Method (PM-FMM)

Multipole Acceptance Criteria

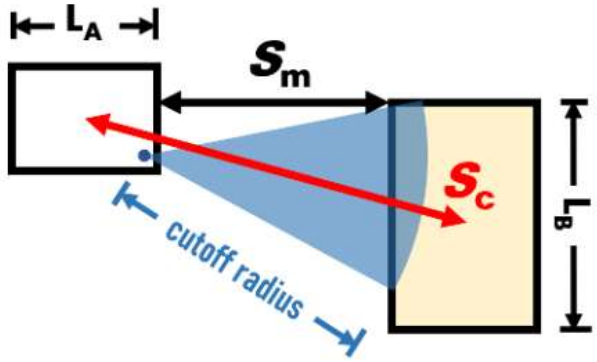


Figure 3: Multipole Acceptance Criteria. L_A is side length of sink (targeted) tree cell and L_B is the source one. The red arrow is the separation S_c between the centers of two nodes and S_m is minimum distance between tow boundaries of tree cells. Two boxes are still physically relevant despite S_c beyond the cutoff radius.

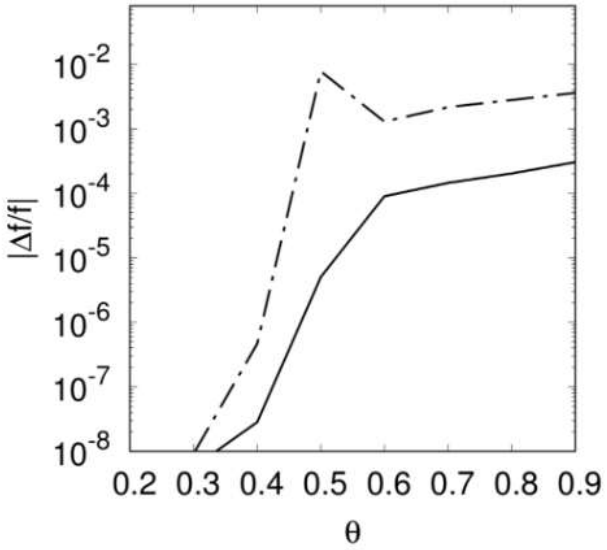
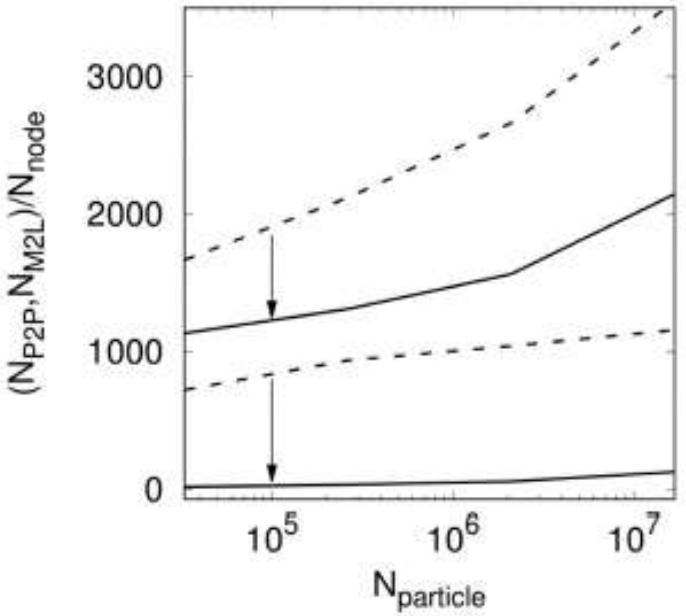


Figure 4: Relation of error to opening angle. Relative error of gravity or acceleration is calculated via truncated FMM upto hexpole. The solid curve denotes the rms error and dash-dotted curve denotes the maximum error of all particles.

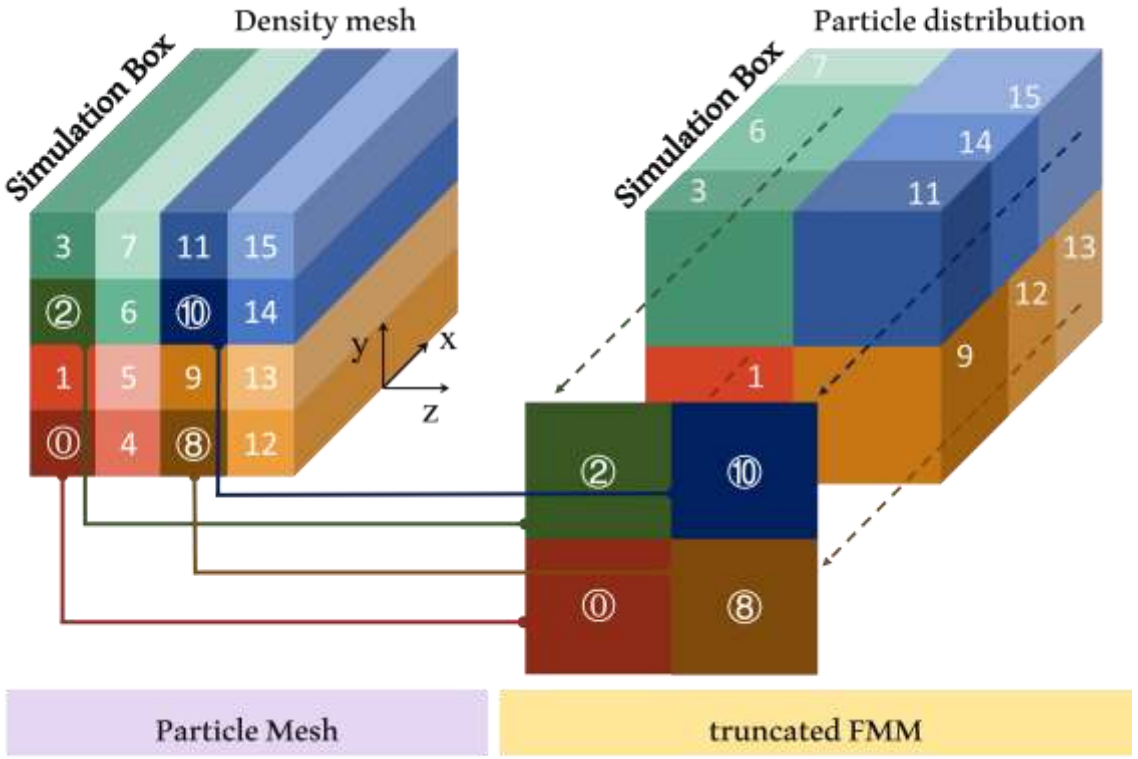


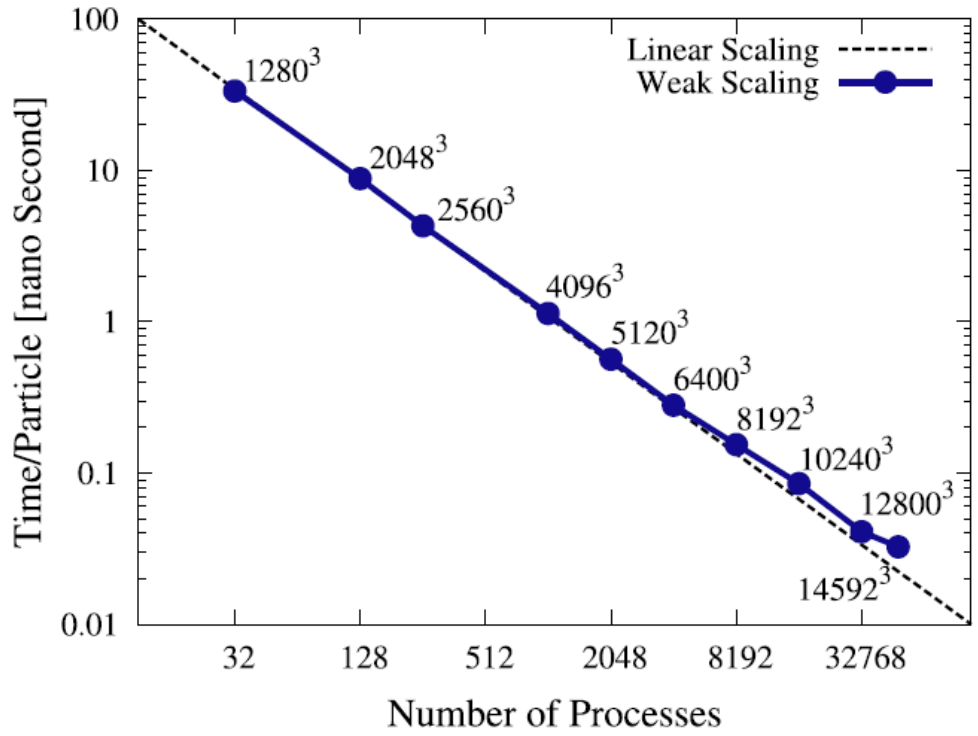
Q. Wang, 2021

$$O\ (\ N * p + n \log_2 n + m \log_2 m)$$

Dual tree traversal

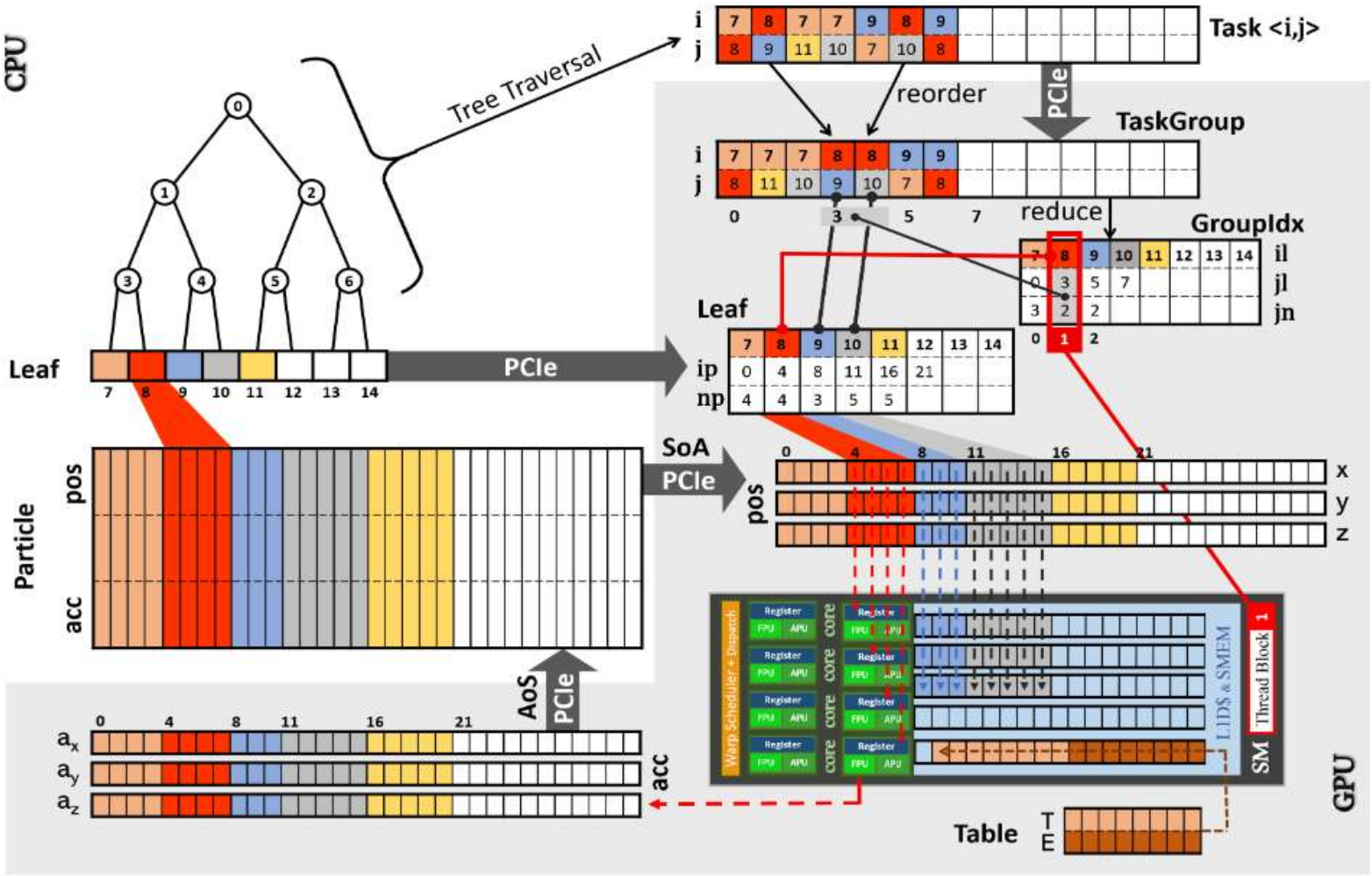
domain decomposition





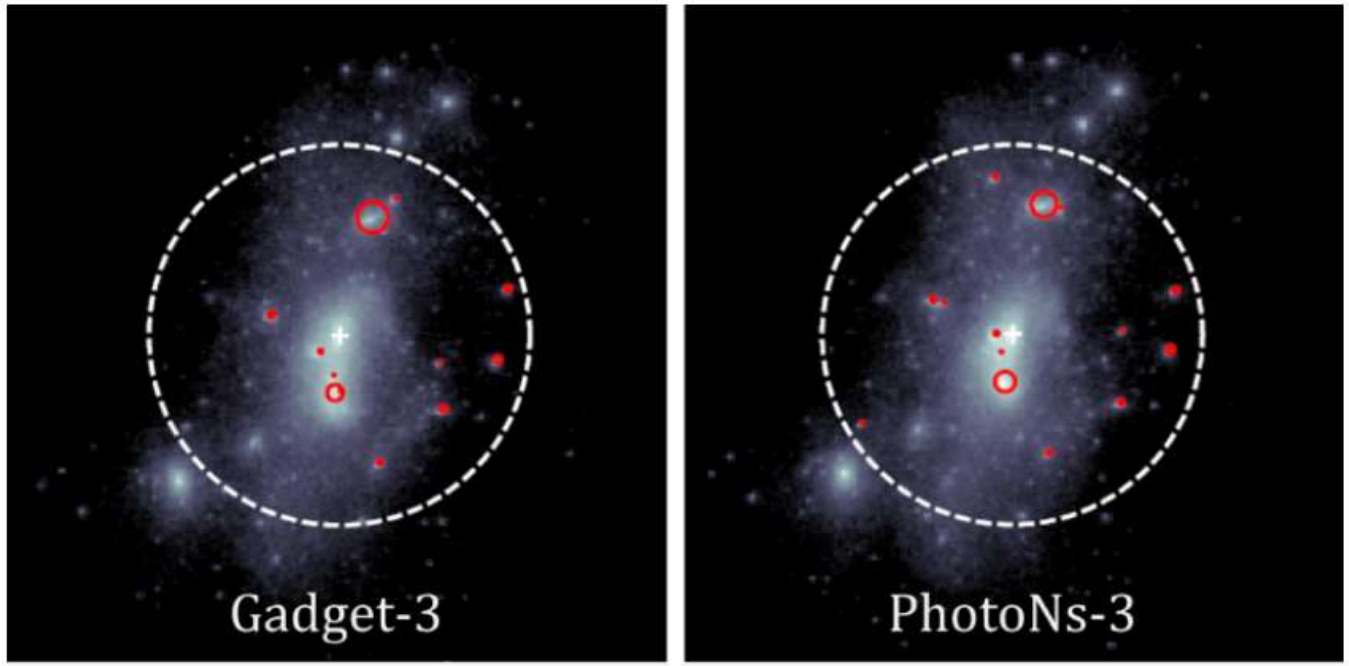
Wang+ 2022

GPU implementation



Wang+ 2021b

Comparison between PhotoNs-3 and Gadget-3



top 10 subhalos in the most massive dark halo

768^3 particles + 125Mpc/h

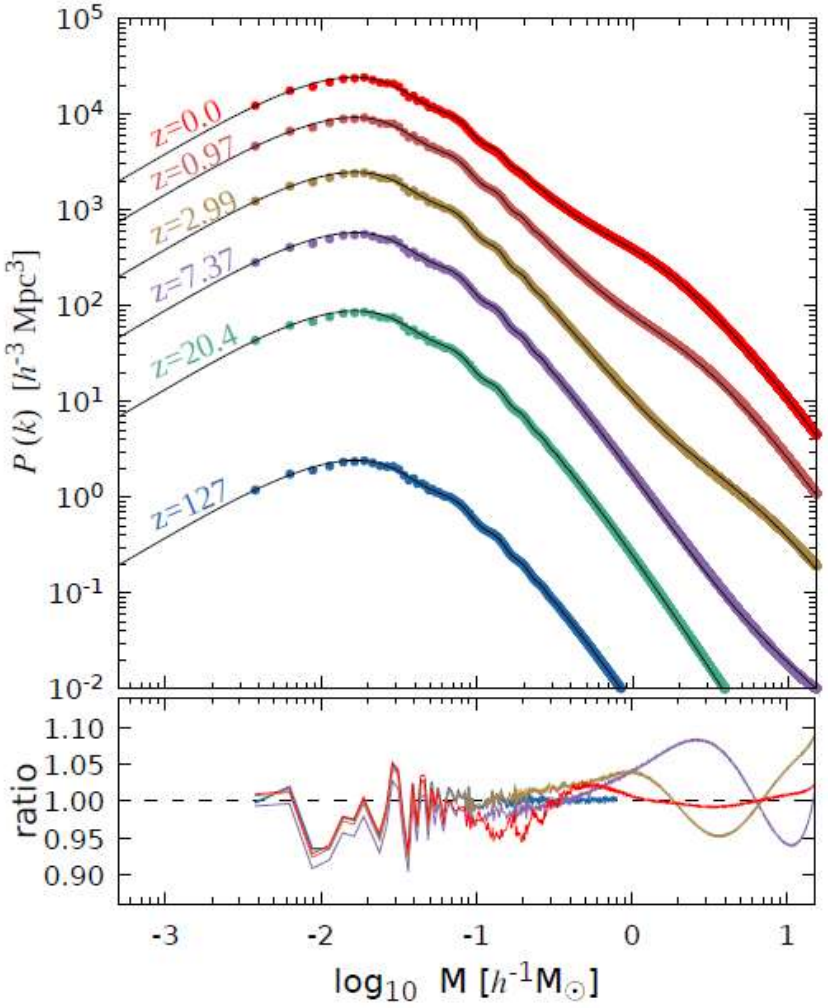
Hyper-Millennium simulation

- Large volume + high resolution
- 100 snapshots (z = 0.0 ~ 20.4)
- PhotoNs-3.7.20

Cosmology	Planck-2018
Volume	2.5^3 $[h^{-3}\mathrm{Gpc}^3]$
Particle No.	~4.2×10^{12}
Particle mass	3.21×10^8 $h^{-1}\mathrm{M}_\odot$

Ω_m	0.3111
Ω_Λ	0.6889
Ω_b	0.0490
h	0.6766
σ_8	0.8102
n_s	0.9665

ORISE Supercomputer



wang+ 2025 in prep

Hyper-Millennium simulation

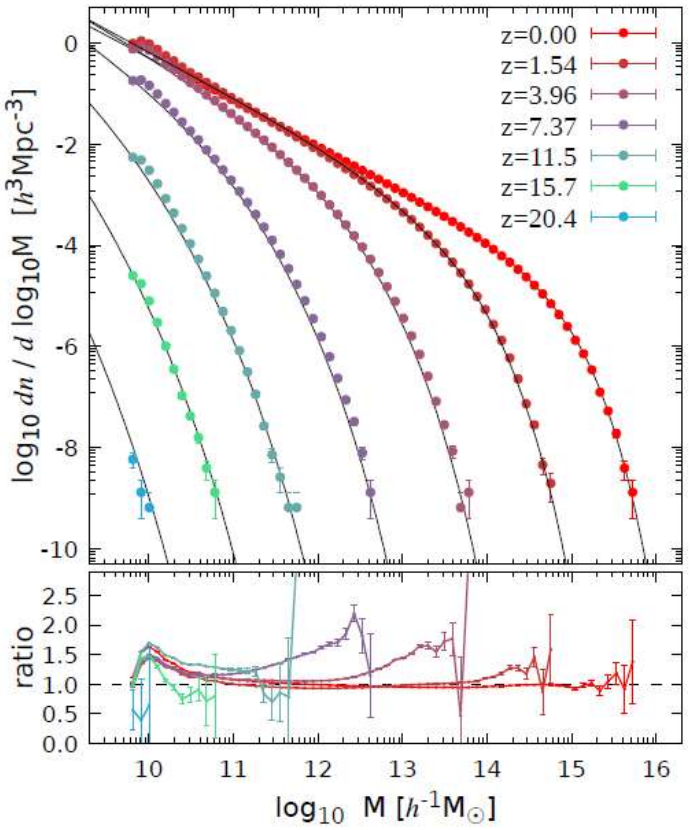
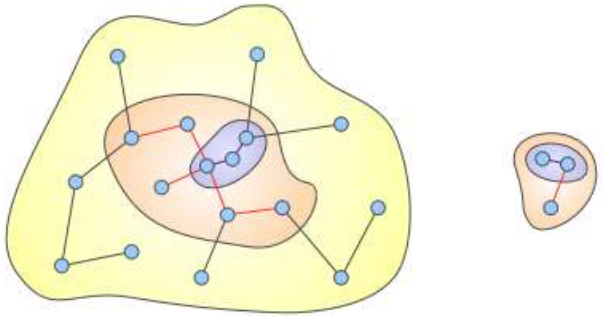


Figure 3. The FOF halo mass functions measured at seven redshifts, from $z = 20.4$ to $z = 0$. The minimum halo, containing 20 particles, corresponds to a mass of $\sim 6.5 \times 10^9\ h^{-1}\mathrm{M}_\odot$. The solid curves represent the predictions from the fitting model of Reed et al. (2007). The bottom panel shows the ratio as a function of halo mass, comparing the simulation results with the model predictions.

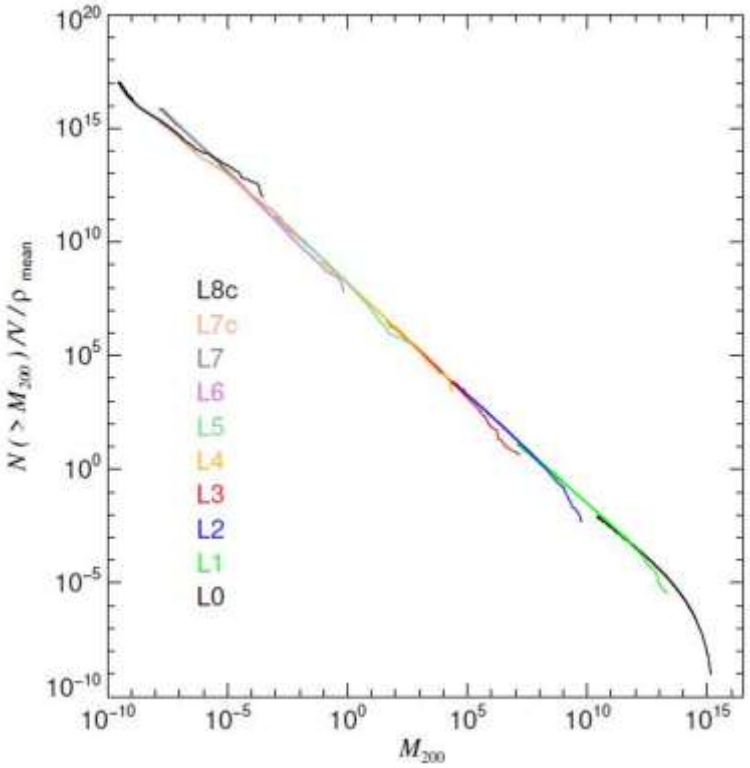
Raw data of snapshot → FoF + subfind + SAM



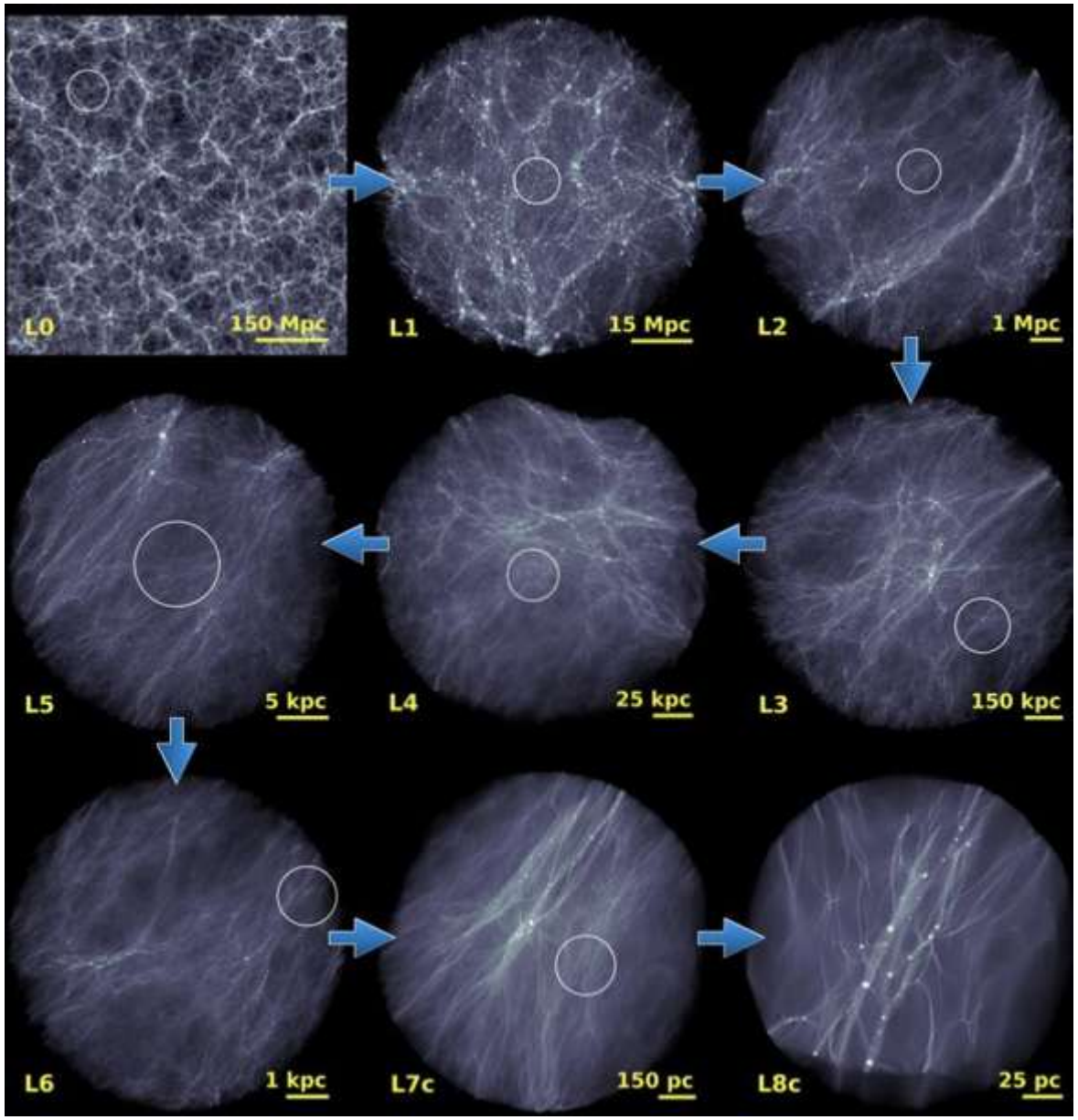
- Using the subhalo catalogue, merger trees were constructed to track subhalo evolution following the methods in Angulo et al. (2012) and Angulo et al. (2014)
- Recent version of the SAM code L-GALAXIES based on Henriques et al. (2015) and Pei et al. (2024)

zoom-in simulation

To improve the resolution



For example (VVV project) :
Dark halo property over 20 orders of magnitude in halo mass



J. Wang+ 2019

Summary

- N-Body simulation is essential and mature tool for studying the cosmology and galaxy formation
- efficient fast numerical method

Alg.	Computer	Code	Simulation
direct		J. Peebles	[Cosmology]
	HYDRA*	NBODY6++	DRAGON [cluster]
Tree	Titan*	2HOT	
	Piz Daint*	Bonsai	[Milky way]
P^3M	Tianhe-II	CUBEP3M	Tiannu[neutrino]
PMPM	π-2.0	CUBE	cosmo-π
		GADGET-2	Millennium(MS-I)
		GADGET-3	Millennium-II
	Tachyon-II	GOTPM	Horizon run-3
		GADGET-3	Millennium-XXL
Tree		GADGET-3	Pangu
+	Aterui/K	GREEM	ν^2GC
PM	Aterui-II/K	GREEM	(Shin) Uchuu
	MIRA*	HACC	Outer Rim
	MIRA*	HACC	Last Journey
	Summit*	HACC	Farpoint
		GADGET-3	Jiutian(JT)-1G
AMR	Pleiades	ART	Bolshoi
		ART	MultiDark
FMM	Piz Daint*	PKDGRAV3	EuclidFlapship
	Piz Daint*	PKDGRAV3	EuclidFlapship-v2
PM	ORISE*	PhotoNs-3.4	Ultramarine(EoR)
+	Cosma8	GADGET-4	Millennium-TNG
FMM		GADGET-4	Jiutian-300/2G
	ORISE*	PhotoNs-3.7	Hyper-Millennium

Summary

- N-Body simulation is essential and mature tool for studying the cosmology and galaxy formation
- efficient fast numerical method

- TNG surveys require the high mass resolution with a huge simulation box
- the extreme scale simulation is still challenging on modern heterogeneous supercomputer.

Thanks very much!