Star-by-star simulations of the Milky Way accelerated by AI surrogate modeling

平島敬也 – Keiya Hirashima

SPDR fellow, RIKEN Center for iTHEMS
Guest Researcher, Flatiron Institute/Simons Foundation

2026/02/18

The 13th "FugakuNEXT" Application Seminar

“Colors of Liberty” project, **Eduardo Kobra**

Contents

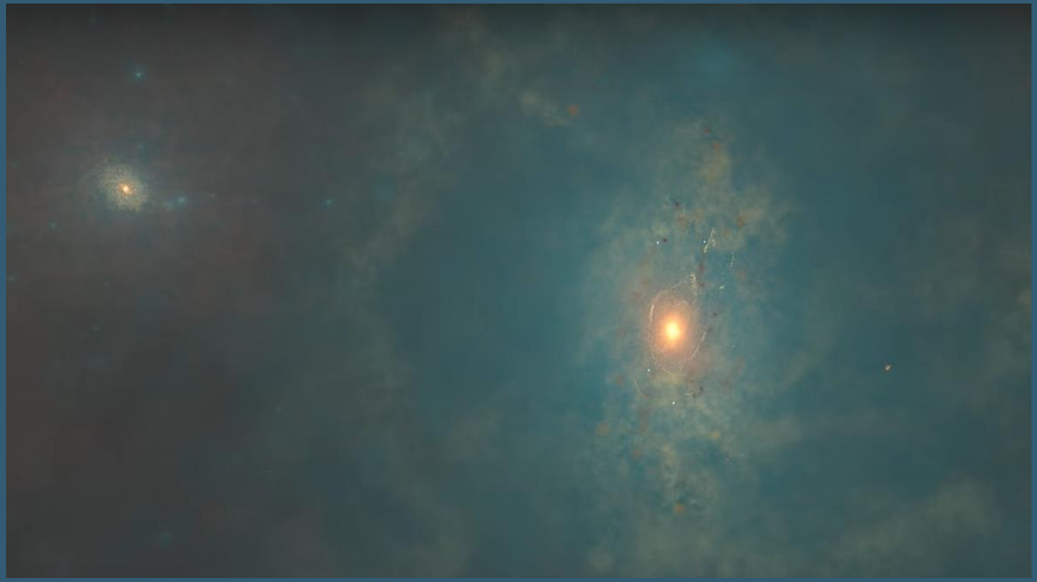
What are Galaxy Simulations?

- The system with dark matter, stars, and gas.
- Many particle system (10^{11} stars in the Milky Way)
- Multi-timescale ($\sim 10^8$ yr for galactic rotation v.s. 10^2 yr for supernovae)
- Track the time evolution of kinematics and thermodynamics by integrating ordinary equations for each pair of particles (Poisson equations and Euler equations).

e.g., self-gravity

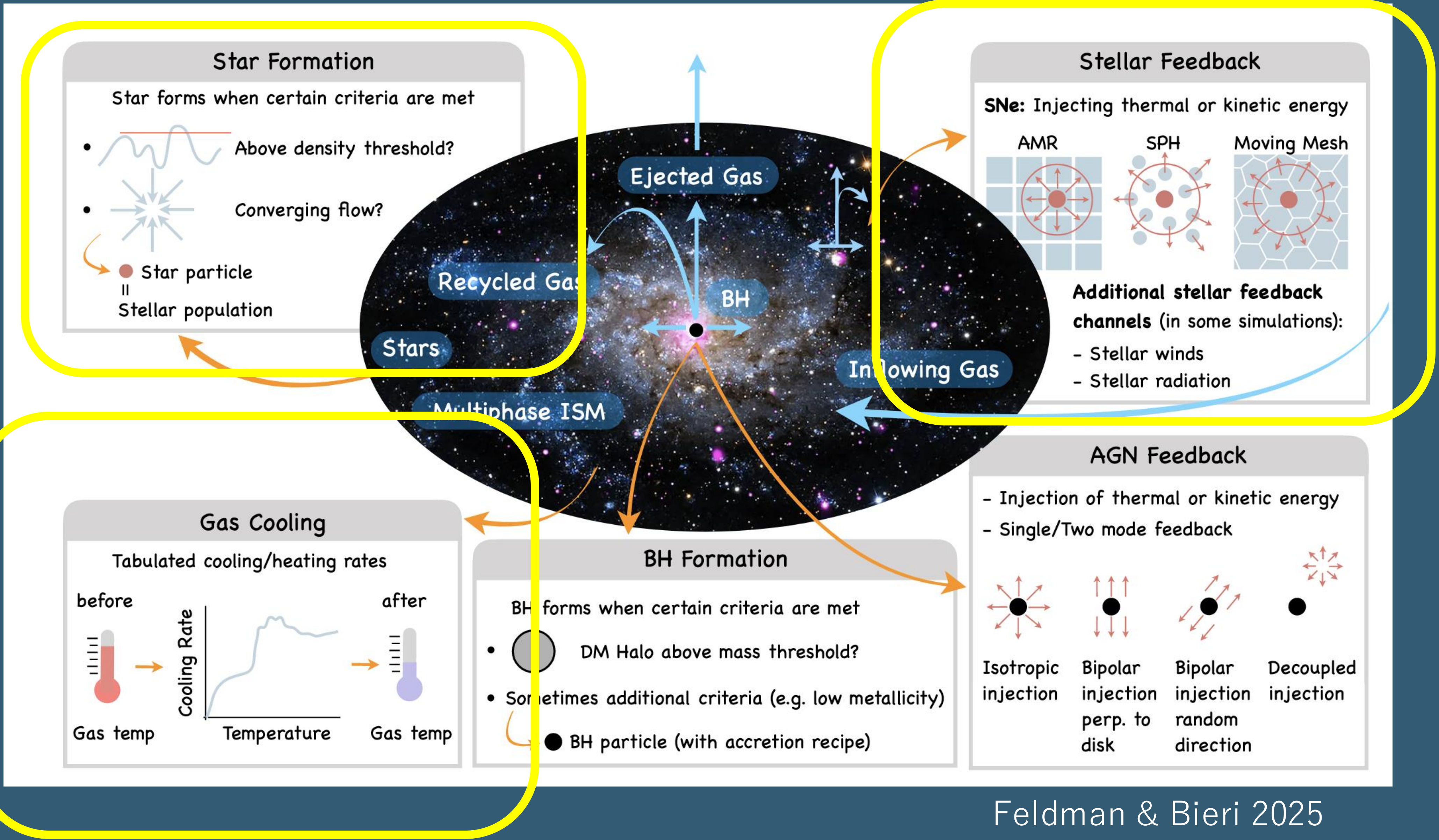
$$\frac{d^2 \boldsymbol{x}_i}{dt^2} = \sum_{j \neq i, 1 \leq j \leq N} G m_j \frac{\boldsymbol{x}_j - \boldsymbol{x}_i}{|\boldsymbol{x}_j - \boldsymbol{x}_i|^3}$$

- Other physics
  - Supernova, radiation, star-formation…



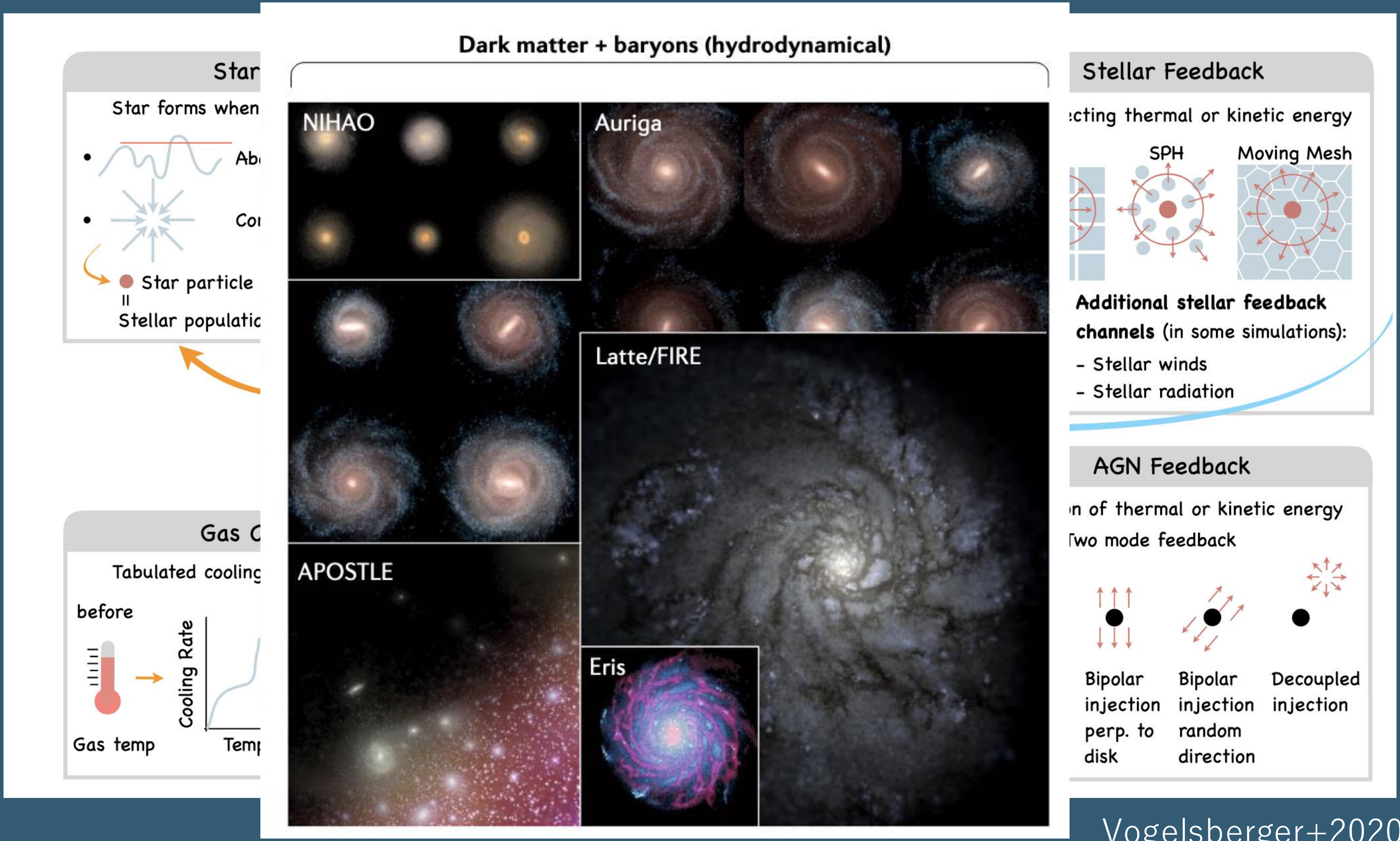
©Takayuki Saitoh

Models for Galaxy (Formation) Sims.



Feldman & Bieri 2025

Models for Galaxy (Formation) Sims.

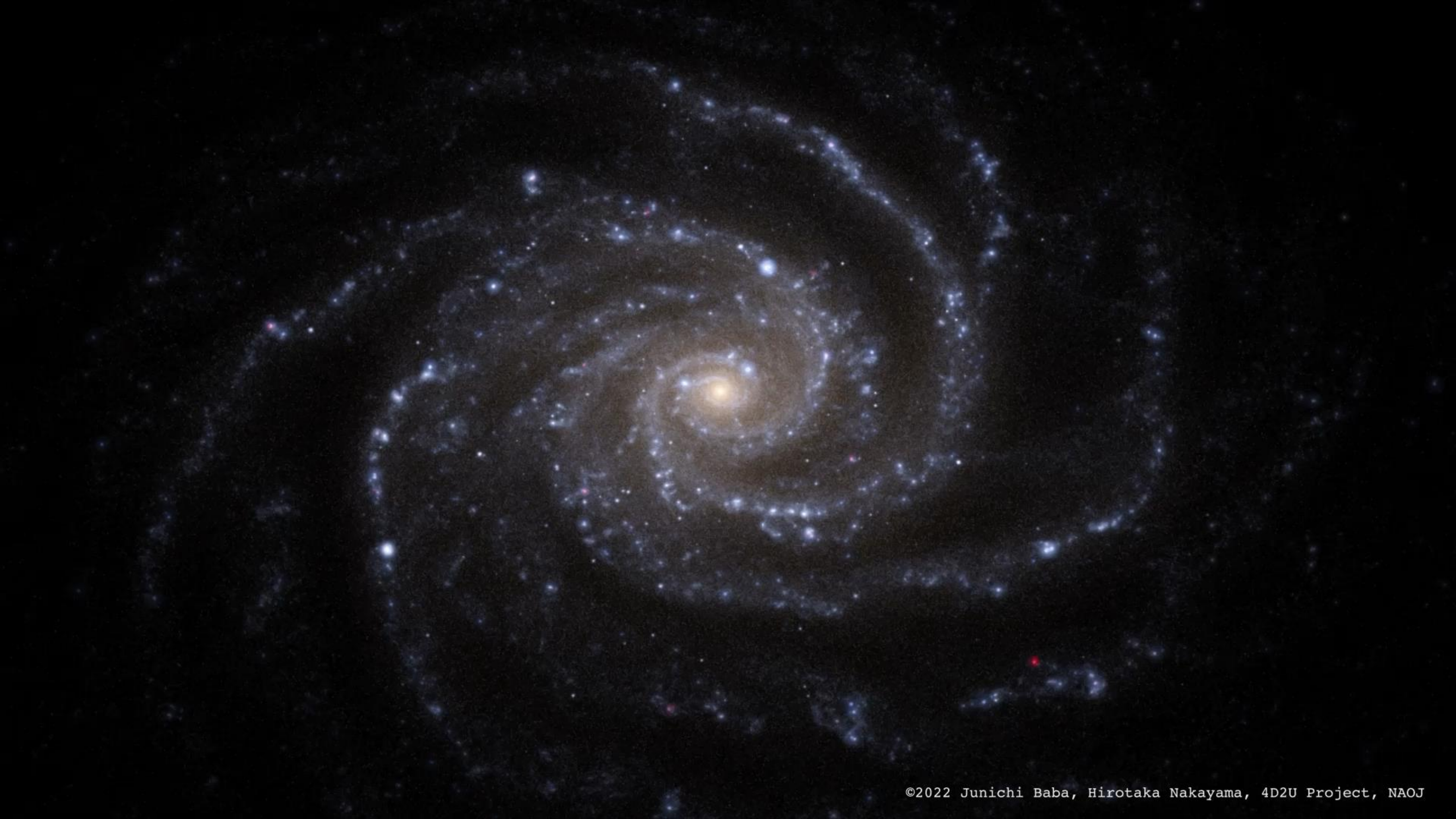


Vogelsberger+2020

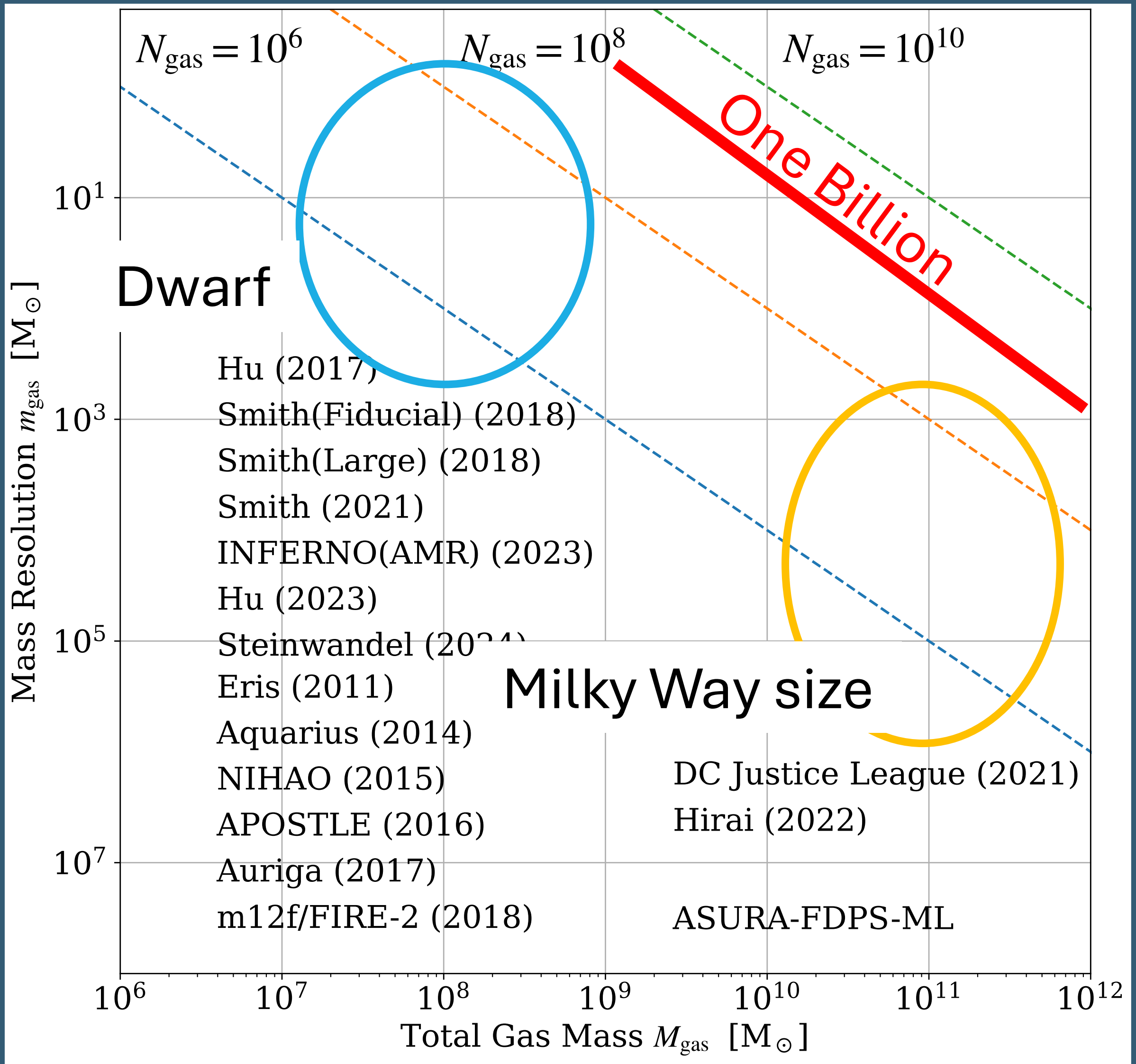
Idealized Galaxy Model

- Three components
  - Gas: hydro + gravity (4 Msun)
  - Dark matter (DM): gravity (4k Msun)
  - Stars: gravity (down to 4 Msun)

DM halo is 20—100 times more massive than the disk



Progress in Resolution of Galactic Disk Simulations



State-of-the-art simulations

1. Star-by-star simulations of dwarf galaxies.
2. Low-resolution Milky Way-sized galaxy simulations.

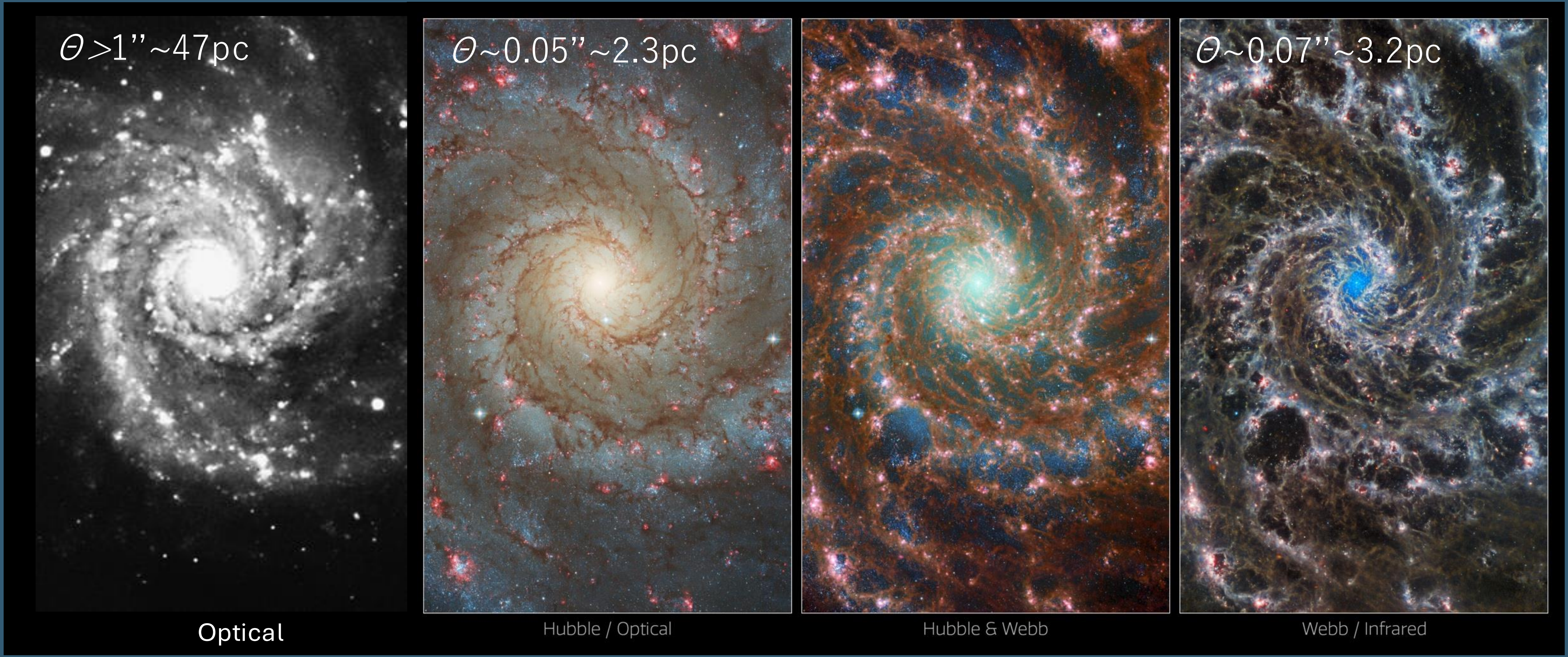
- Higher resolution further increases the cost of resolving small-scale physics.

Hirashima+25b, SC25

Why do we need to pursuit high-resolutions?

Observation of M74

1975 | 2003~2005 | 2022



Star-forming regions
Spiral arms

HII regions
Star clusters

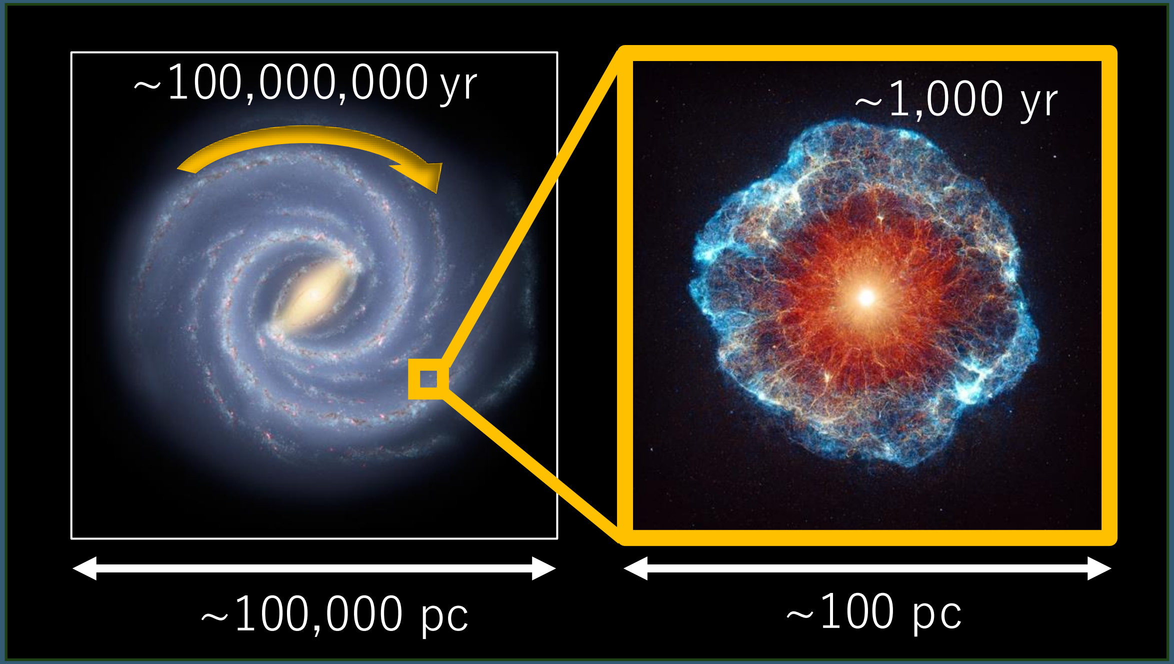
Filaments
Shock waves

KPNO 4-meter Mayall telescope, 1975
Credit: NOIRLab/NSF/AURA

ESA/Webb, NASA & CSA, J. Lee and the PHANGS-JWST Team; ESA/Hubble & NASA, R. Chandar Acknowledgement: J. Schmidt

Higher Resolution Demands Greater Costs

- High-resolution simulation domains demand short timesteps due to high density and temperature (e.g., SNe)



NASA/JPL-Caltech/ESO/R. Hurt

- Estimating a typical timestep
  - Size of SPH particles

$$\lambda = \left(\frac{3}{4\pi} \frac{m_{\rm SPH}}{\rho} \right)^{1/3}$$

Mass resolutions: $m_{\rm SPH}$
Density: ρ

  - Internal energy per particle

$$U_{\rm SN} := \frac{E_{\rm SN}}{N_{\rm NB}\ m_{\rm SPH}} \simeq \frac{5.0 \times 10^{17}}{N_{\rm NB}} \left(\frac{1\ M_\odot}{m_{\rm SPH}} \right) \ [{\rm erg\ g^{-1}}]$$

Neighbor number $N_{\rm NB} \sim 100$

  - Sound speed

$$c_{\rm SN} := \sqrt{\gamma(\gamma-1)U_{\rm SN}} \simeq \frac{7.5 \times 10^3}{N_{\rm NB}^{1/2}} \left(\frac{1\ M_\odot}{m_{\rm SPH}} \right)^{1/2} \ [{\rm km\ s^{-1}}].$$

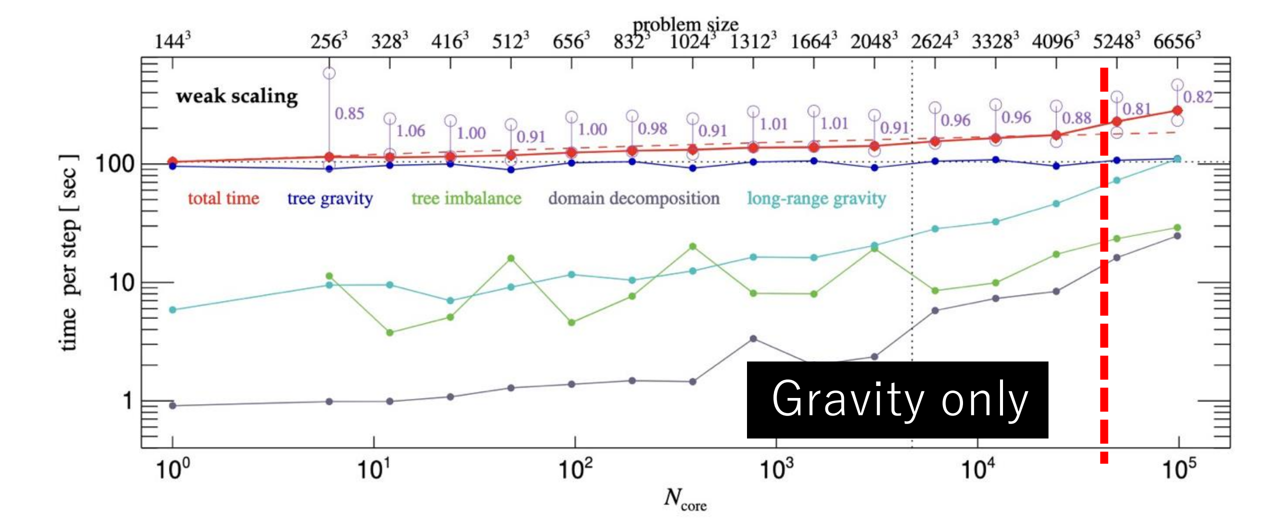
- Timescale of SNe at certain resolution and density

$$t_{\rm SN} \simeq 1.3 \times 10^3 \left(\frac{m_{\rm SPH}}{1\ M_\odot} \right)^{5/6} \left(\frac{100\ [{\rm cm^{-3}}]}{n_{\rm H}} \right)^{1/3} \ [{\rm yr}].$$

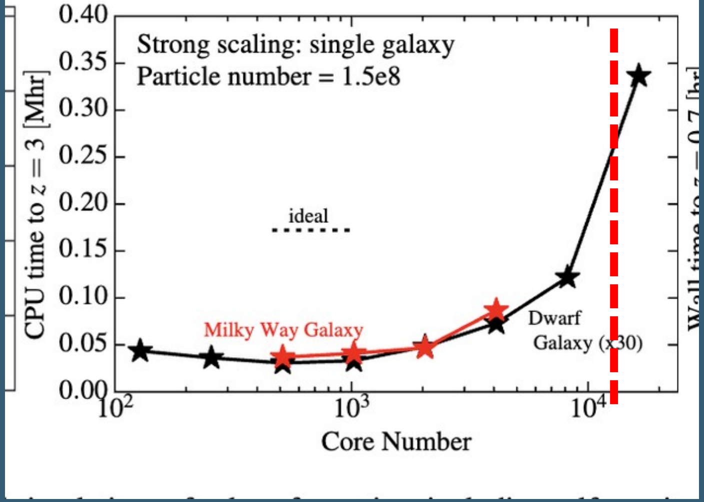
- Typical timestep: $0.1 t_{\rm SN} \sim 10^2$ yr

Parallelization Efficiency in Galaxy Simulations

Gadget4 (Springel et al. 2021)



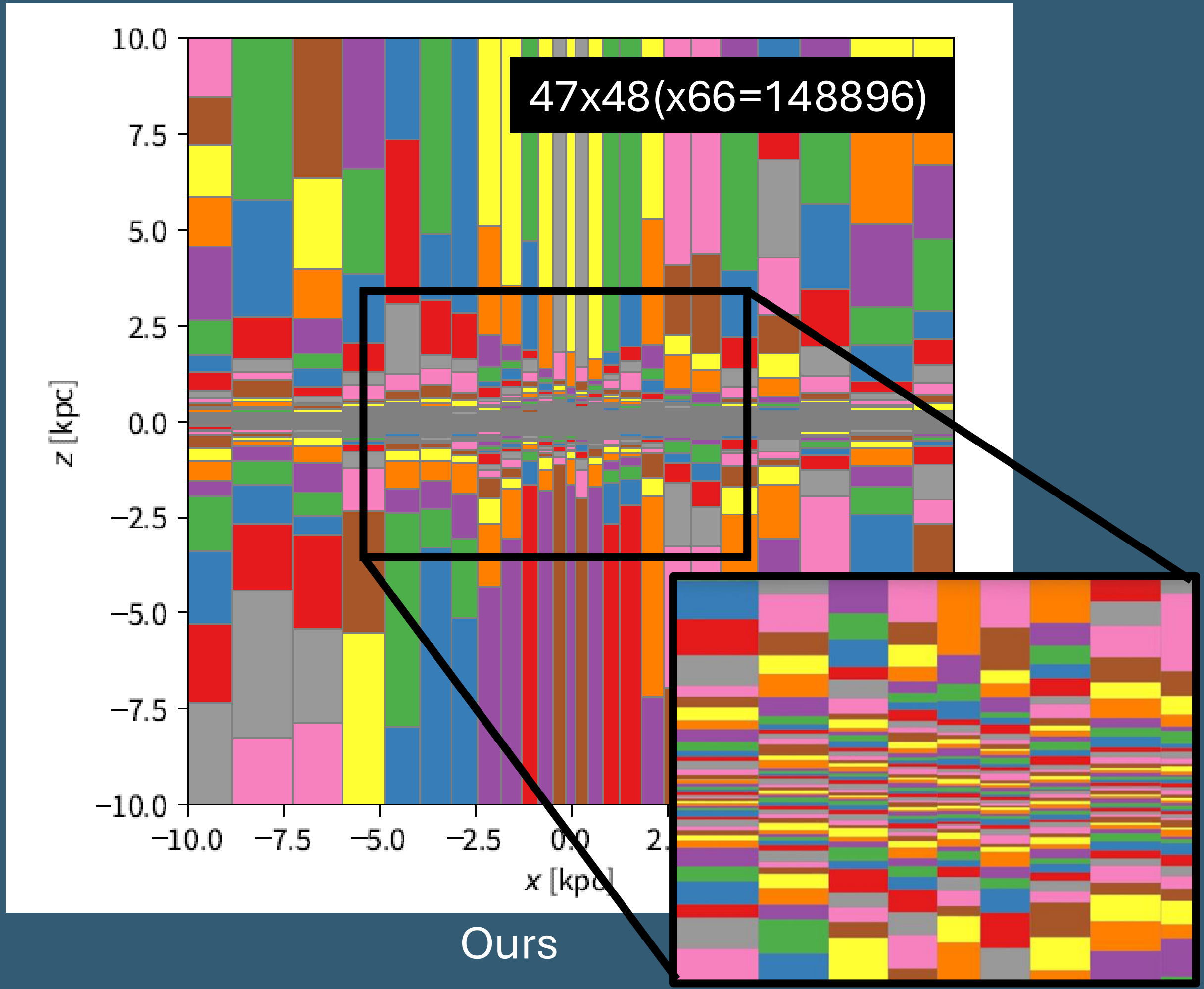
GIZMO (Hopkins et al. 2018)



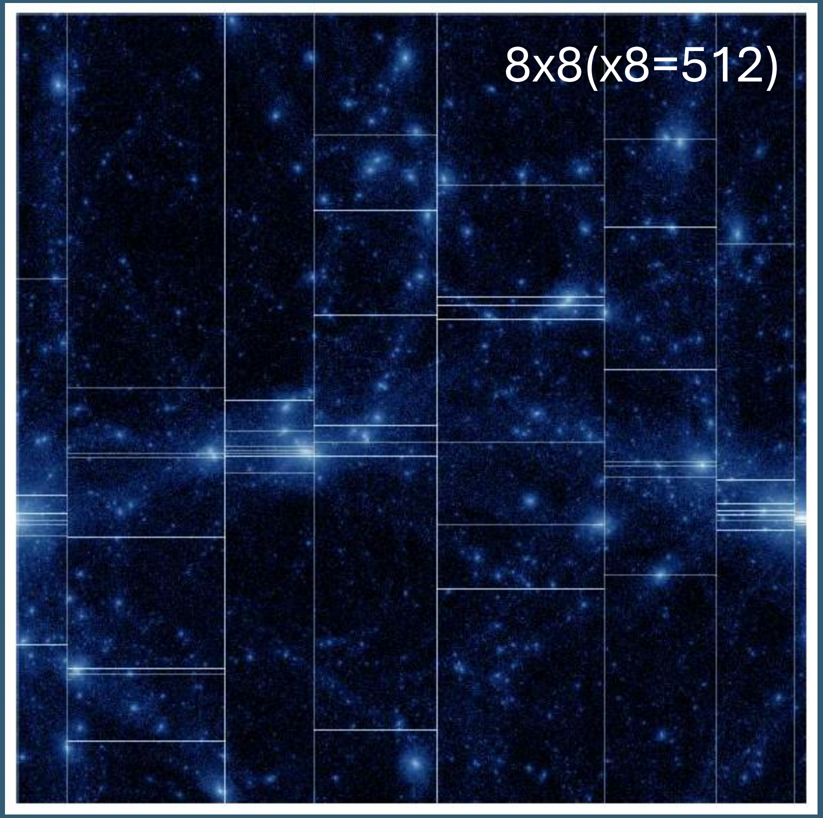
- The communication overhead increases at ~1k nodes.
- Given a certain mass, the number of required steps is proportional to the cubic root of #particles, $N^{1/3}$.
- Small-scall physics cause load-imbalance.

Challenges in Load Balance: Domain Decomposition

In a galaxy, the density increases toward the center, following a power-law profile, which could lead to severe load imbalance across nodes after domain decomposition.



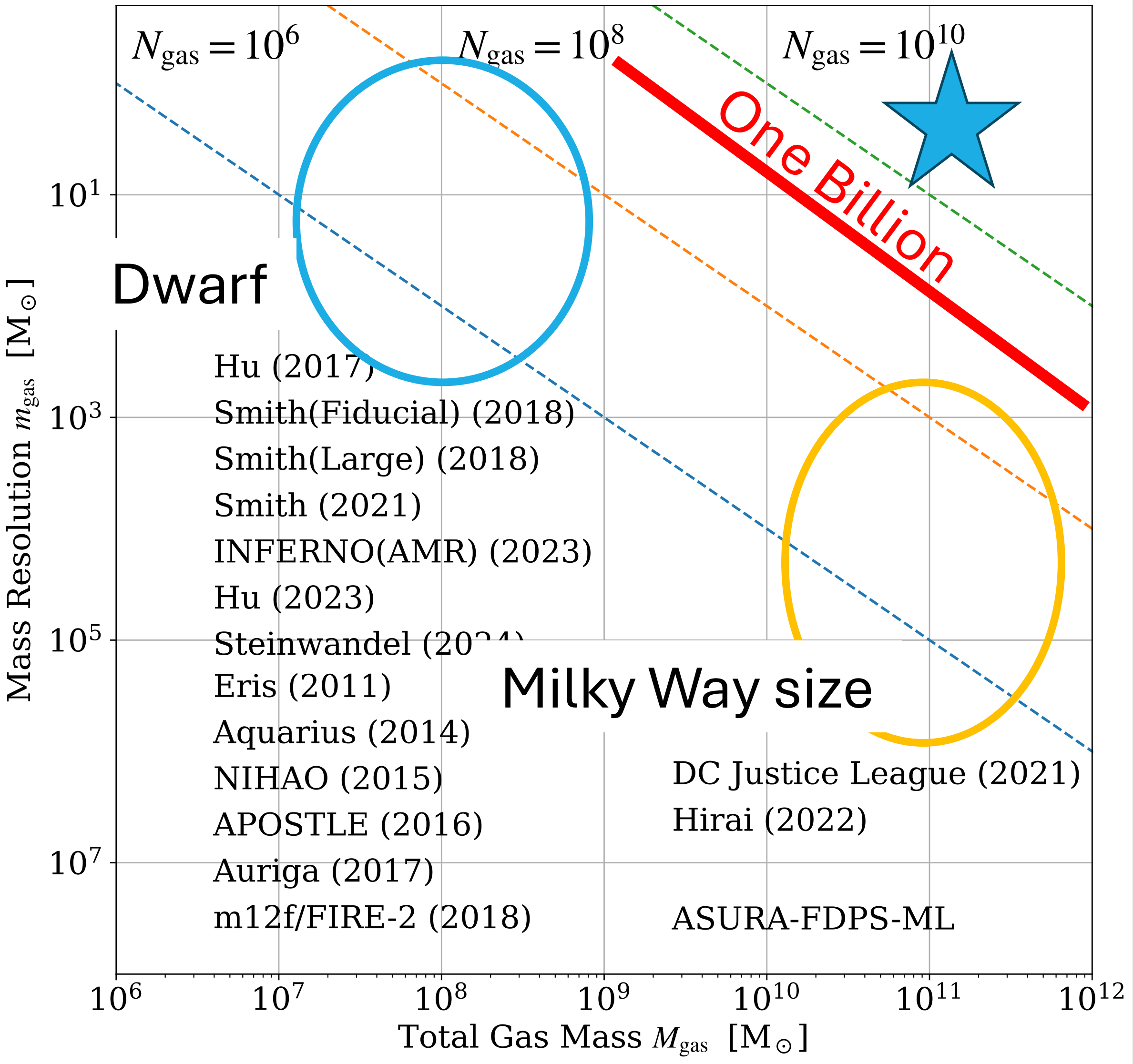
Ours



Ishiyama et al. (2012), Gordon-Bell prize

Contents

Current Status of Galactic Disk Simulations



State-of-the-art simulations

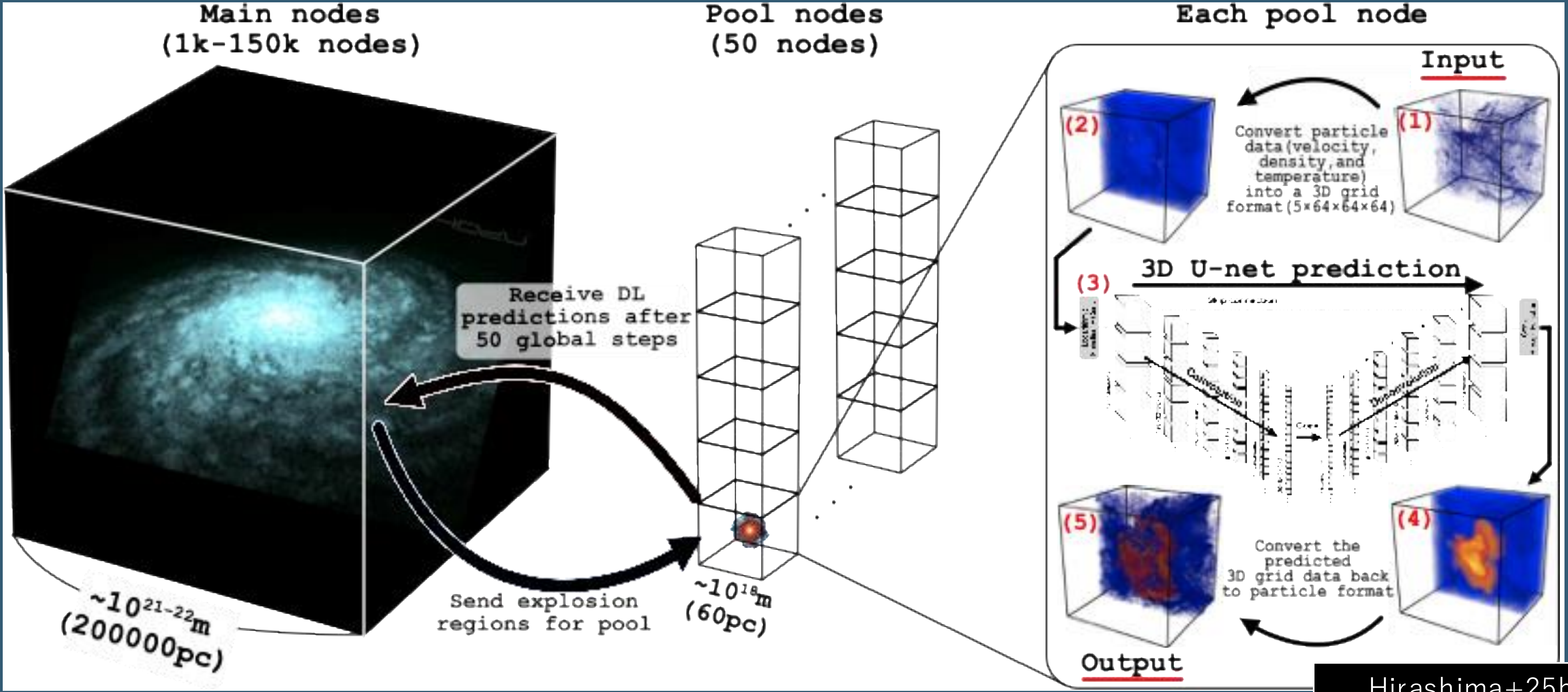
1. Star-by-star simulations of dwarf galaxies.
2. Low-resolution Milky Way-sized galaxy simulations.

Aim 10^{11} particles for star-by-star Milky Way-sized simulations

- Higher resolution further increases the cost of resolving small-scale physics.

Our Framework with AI: ASURA-FDPS-ML

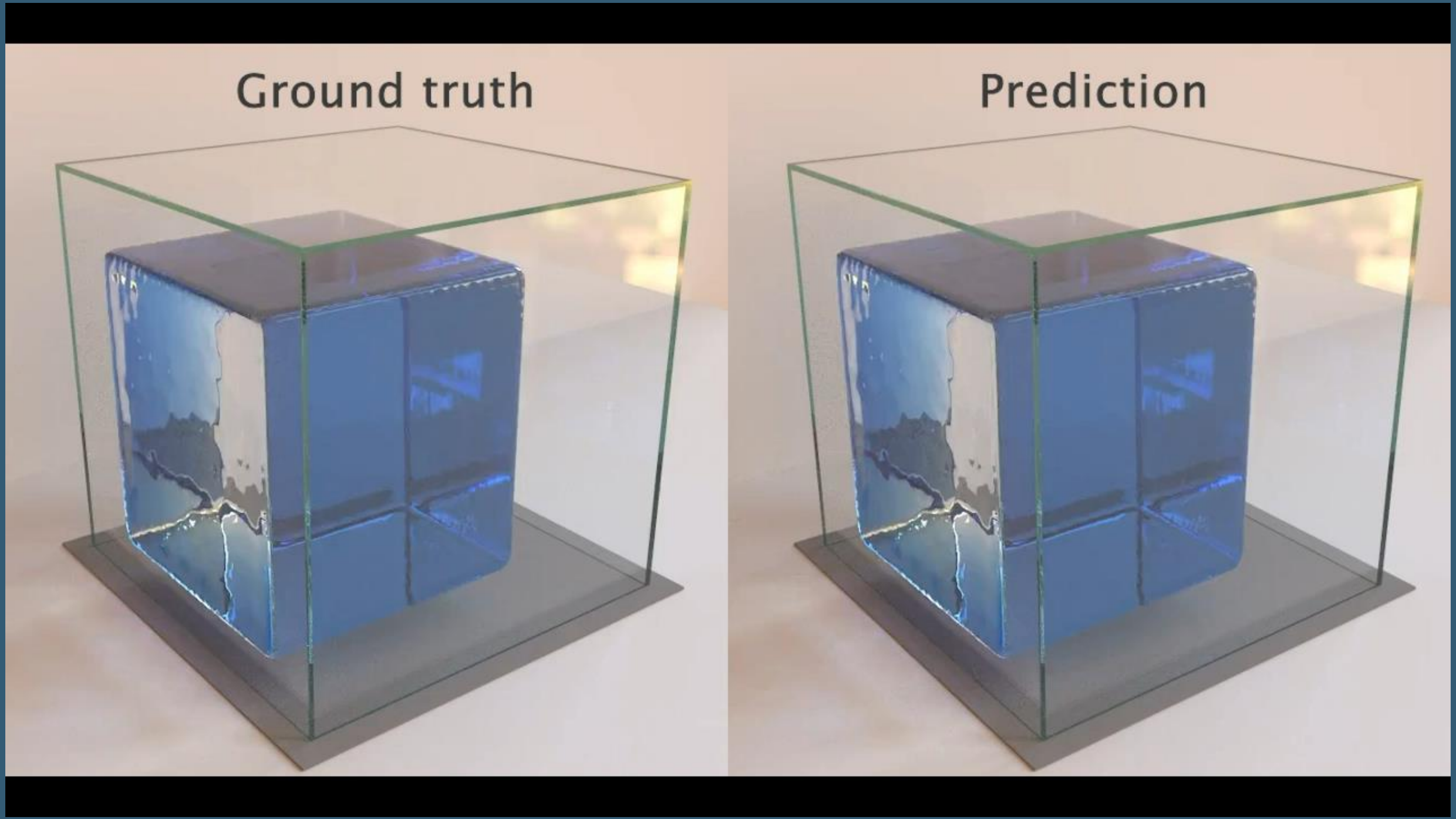
- ASURA (Physical model for galaxy simulations)
- FDPS (Framework for Developing Particle Simulator in parallel computing)
- PIKG (an automatic Particle-particle Interaction Kernel Generator)
- **Surrogate Modeling (Deep learning for fast computation of supernovae)**



Hirashima+25b, SC25

Surrogate Model

- Surrogating (replacing) expensive simulations with mathematical models (incl. machine learning)
- surrogate model, emulator, forward modeling are the same in the context of "AI for Science".



Sanchez-Gonzalez+2020

SPH interaction estimation using Graph Neural Network

Training Data (3D Cartesian Grids)

Training dataset: 300 simulations of a SN in turbulent clouds.

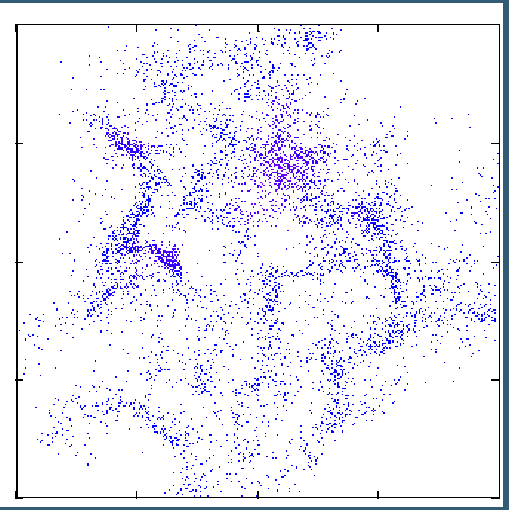
Temperature	10 [K]
Mean ambient density	4 ~ 1000 [cm^{-3}]
Input energy	10^{51} [erg]
Total mass	10^{6} [$M_{\odot}$]
Mass of a gas particle	1 [$M_{\odot}$]
Softening parameter	0.5 [pc]

The dataset is publicly available at **The Well** (https://github.com/PolymathicAI/the_well).

TheWell

Polymathic

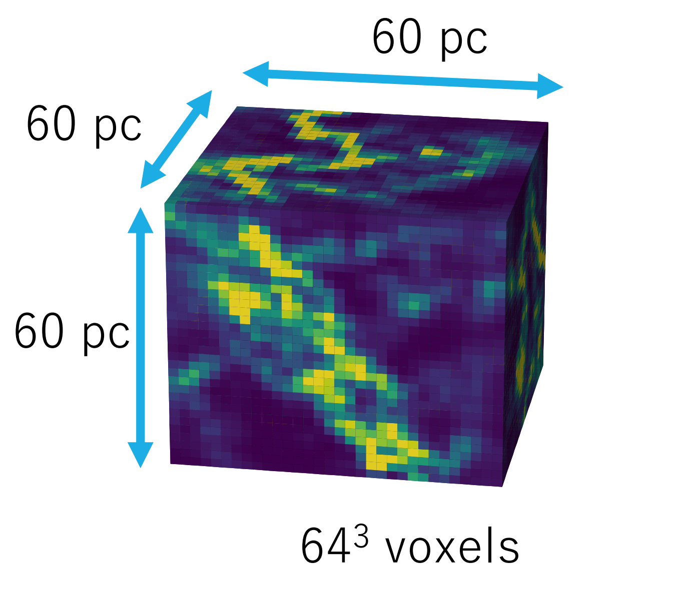
Ohana+(incl. KH), 2024, NeurIPS



Interpolation using

SPH summation

$$\rho_i = \sum_j m_j W(r_{ij}, h_j)$$



Convolutional Neural Network

Input

0	0	0	0
0	0	1	2
0	3	4	5
0	6	7	8

$*$

Kernel
(Learnable)

0	1
3	4

$=$

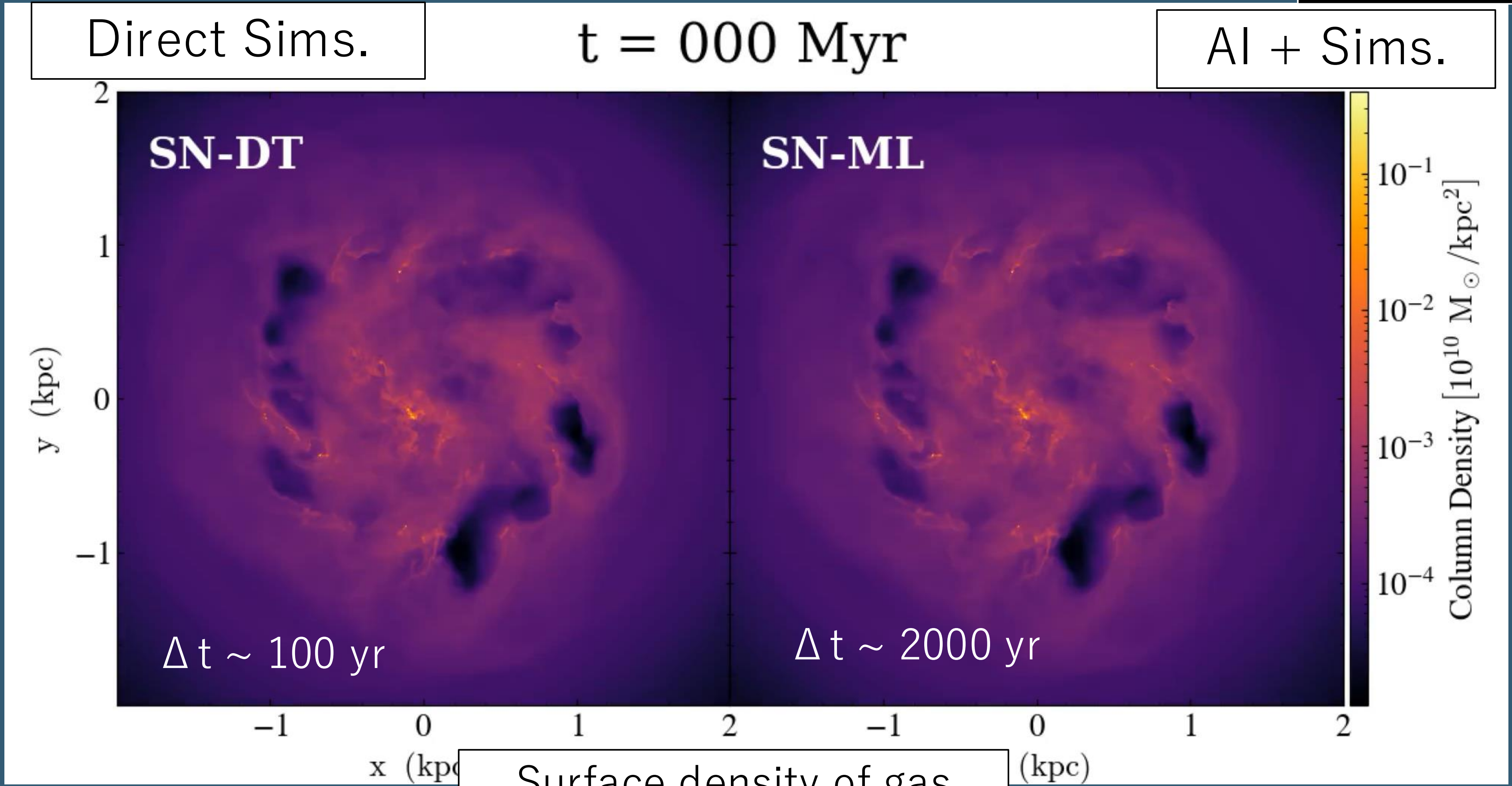
Output

0	4	11
12	26	34
27	50	58

Galaxy Simulations with Surrogate Model

- **Our surrogate model is used for SNe at** > 1 cm^{-3}
- Tested on a dwarf galaxy (M_{vir} ~ 10^{10} $M_{\odot}$ & m_{baryon} ~ 4 $M_{\odot}$)
- x4 faster and 6 months can be saved!

Hirashima+25a, ApJ

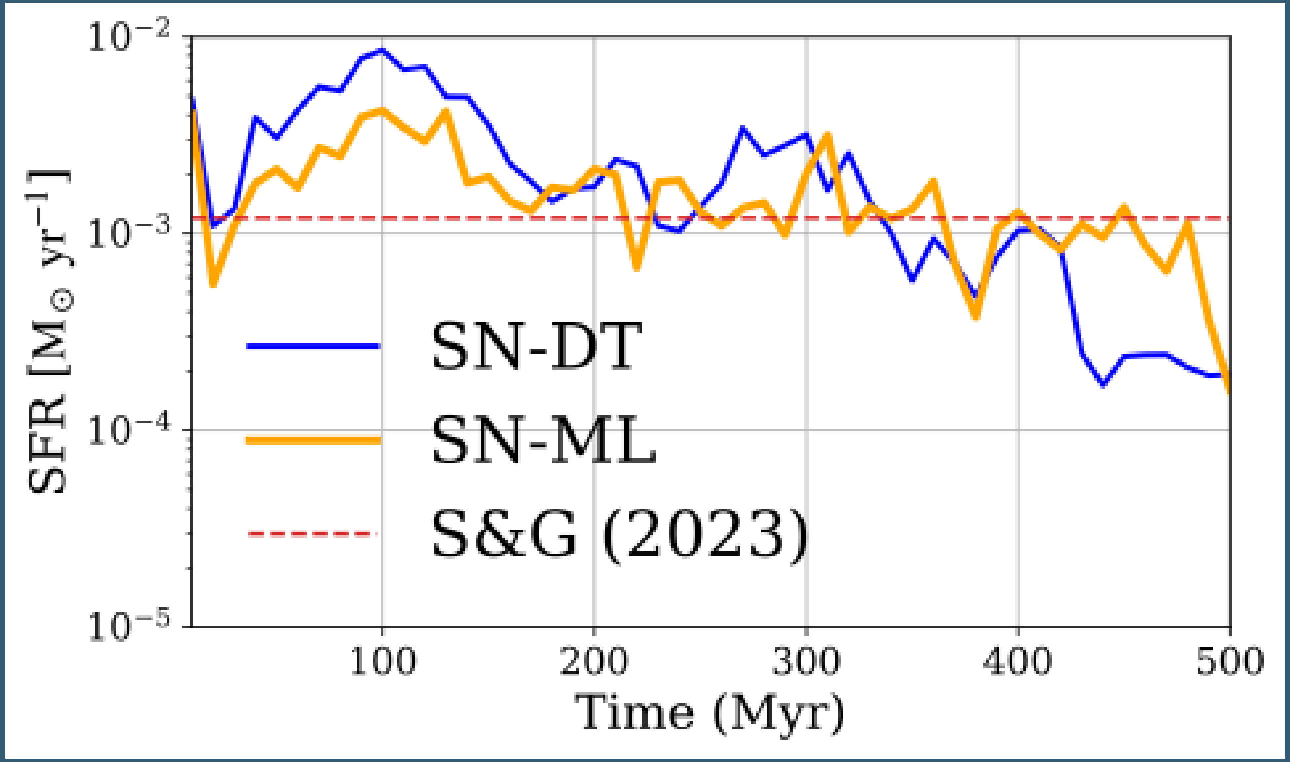


Surface density of gas

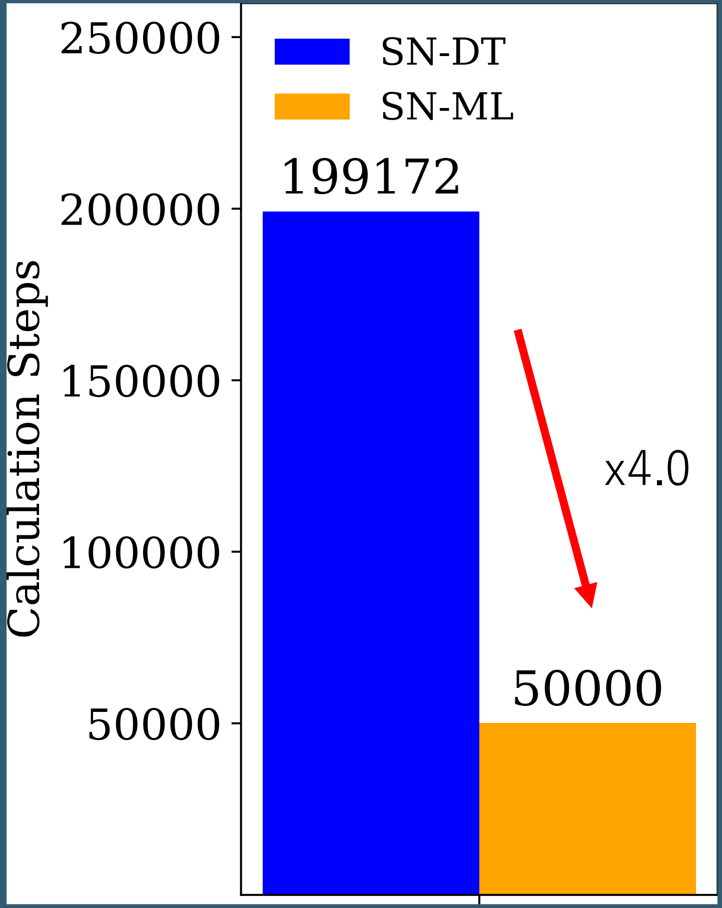
Video's Here!

Validation

Star Formation Efficiency



X4 speed-up



8 months to 2 months

Outflows

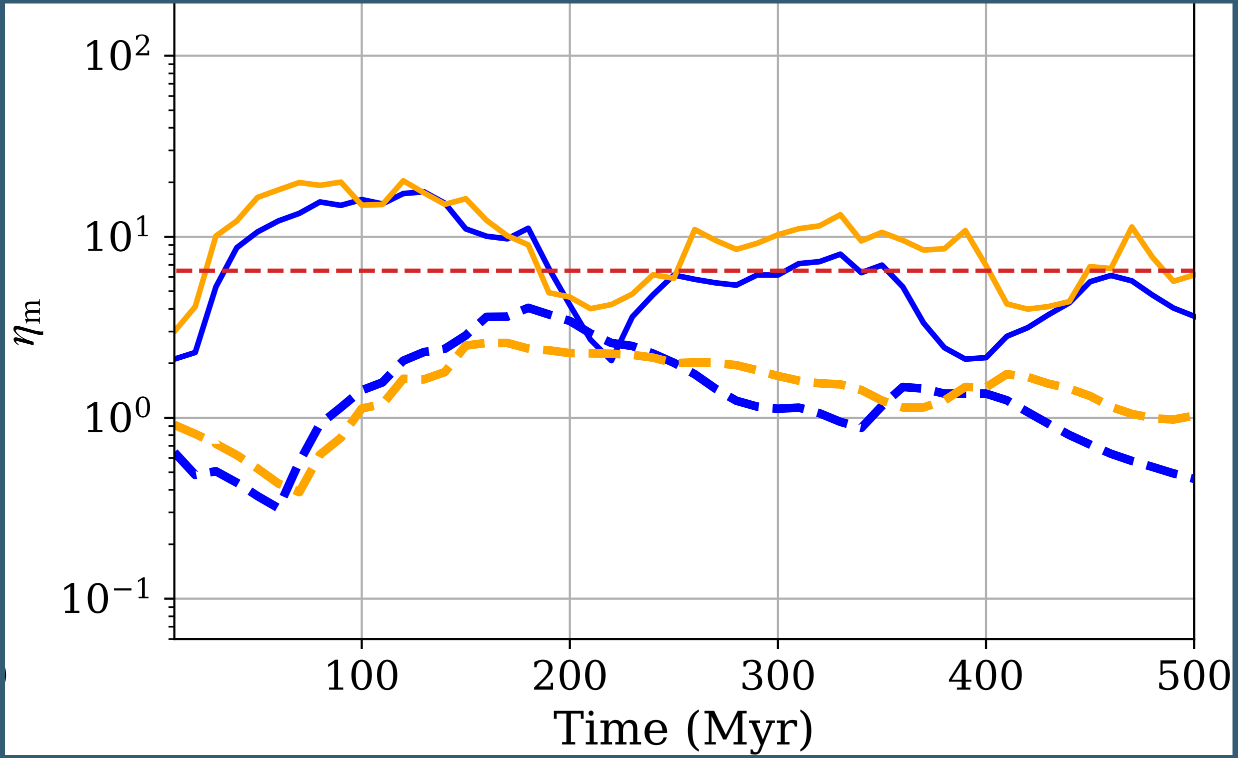
- Similar results between conventional and novel framework with AI.
- Only hot supersonic winds spout of the disk（Kim+20, Steinwandel+24）。

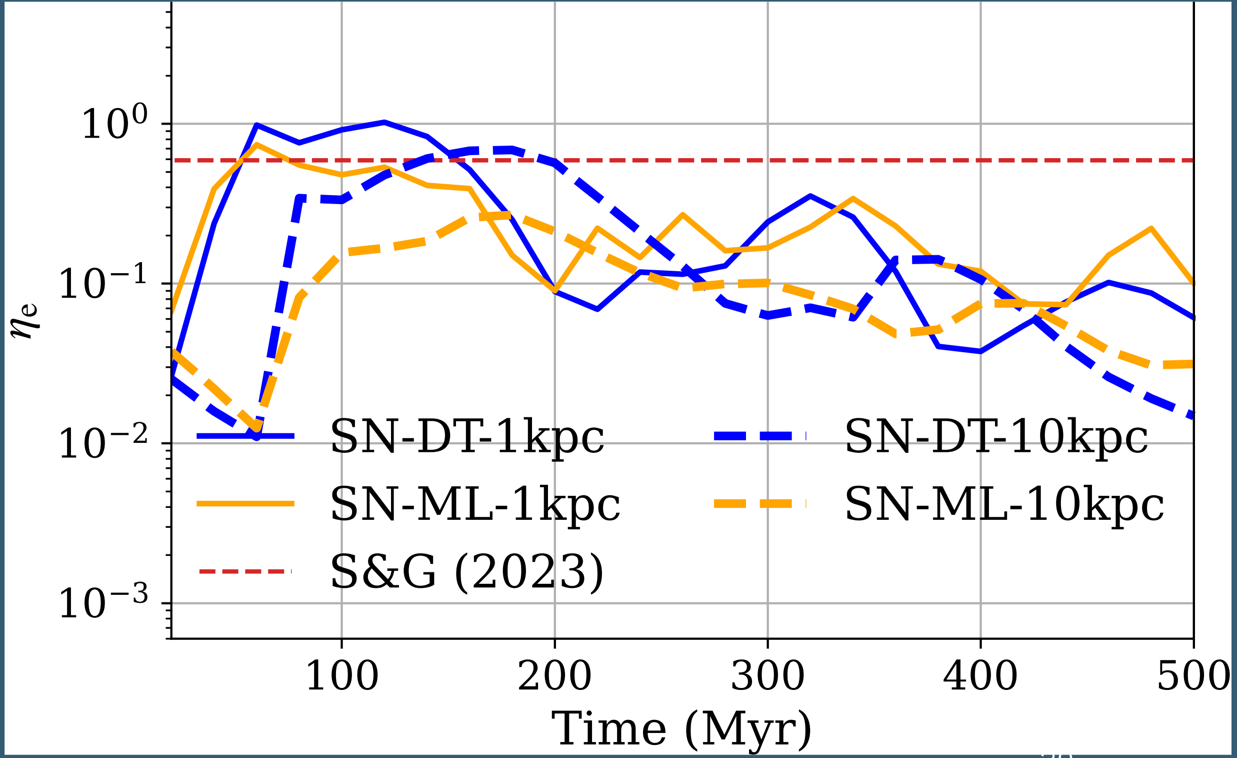
Mass loading factor

$$\eta_m^{\mathrm{out}} = \dot{M}_{\mathrm{out}} / \overline{\mathrm{SFR}}$$

Energy loading factor

$$\eta_e^{\mathrm{out}} = \dot{E}_{\mathrm{out}} / \left(E_{\mathrm{SN}} \overline{R_{\mathrm{SN}}}\right)$$





理化学研究所

RIKENメルマガ

検索

2025年11月21日

← 前の記事 ↑ 一覧へ戻る → 次の記事

理化学研究所
神戸大学
筑波大学

3,000億粒子の天の川銀河シミュレーションをAI×富岳で実現

ー星一つ一つを再現する高解像度モデルで銀河進化に迫るー

English Page

The First Star-by-star N-body/Hydrodynamics Simulation of Our Galaxy Coupling with a Surrogate Model

Keiya Hirashima
Center for Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS)
RIKEN
Wako, Saitama, Japan
keiya.hirashima@riken.jp

Michiko S Fujii
Department of Astronomy
The University of Tokyo
Tokyo, Tokyo, Japan
fujii@astron.s.u-tokyo.ac.jp

Takayuki R Saitoh
Department of Planetology
Kobe University
Kobe, Hyogo, Japan
saitoh@people.kobe-u.ac.jp

Naoto Harada
Department of Astronomy
The University of Tokyo
Tokyo, Tokyo, Japan
hrdnot-3997@g.ecc.u-tokyo.ac.jp

Kentaro Nomura
Preferred Networks Inc.
Tokyo, Tokyo, Japan
subarutaro@preferred.jp

Kohji Yoshikawa
Center for Computational Sciences
University of Tsukuba
Tsukuba, Ibaraki, Japan
kohji@ccs.tsukuba.ac.jp

Yutaka Hirai
Department of Community Service and Science
Tohoku University of Community Service and Science
Sakata, Yamagata, Japan
yutaka.hirai@koeki-u.ac.jp

Tetsuro Asano
Institut de Ci'encies del Cosmos
Universitat de Barcelona
Barcelona, Catalonia, Spain
asano@fqa.ub.edu

Kana Moriwaki
Research Center for the Early Universe
The University of Tokyo
Tokyo, Tokyo, Japan
kana.moriwaki@phys.s.u-tokyo.ac.jp

Masaki Iwasawa
Matsue College
National Institute of Technology
Matsue, Shimane, Japan
masaki@matsue-ct.ac.jp

Takashi Okamoto
Faculty of Science
Hokkaido University
Sapporo, Hokkaido, Japan
t.t.okamoto@gmail.com

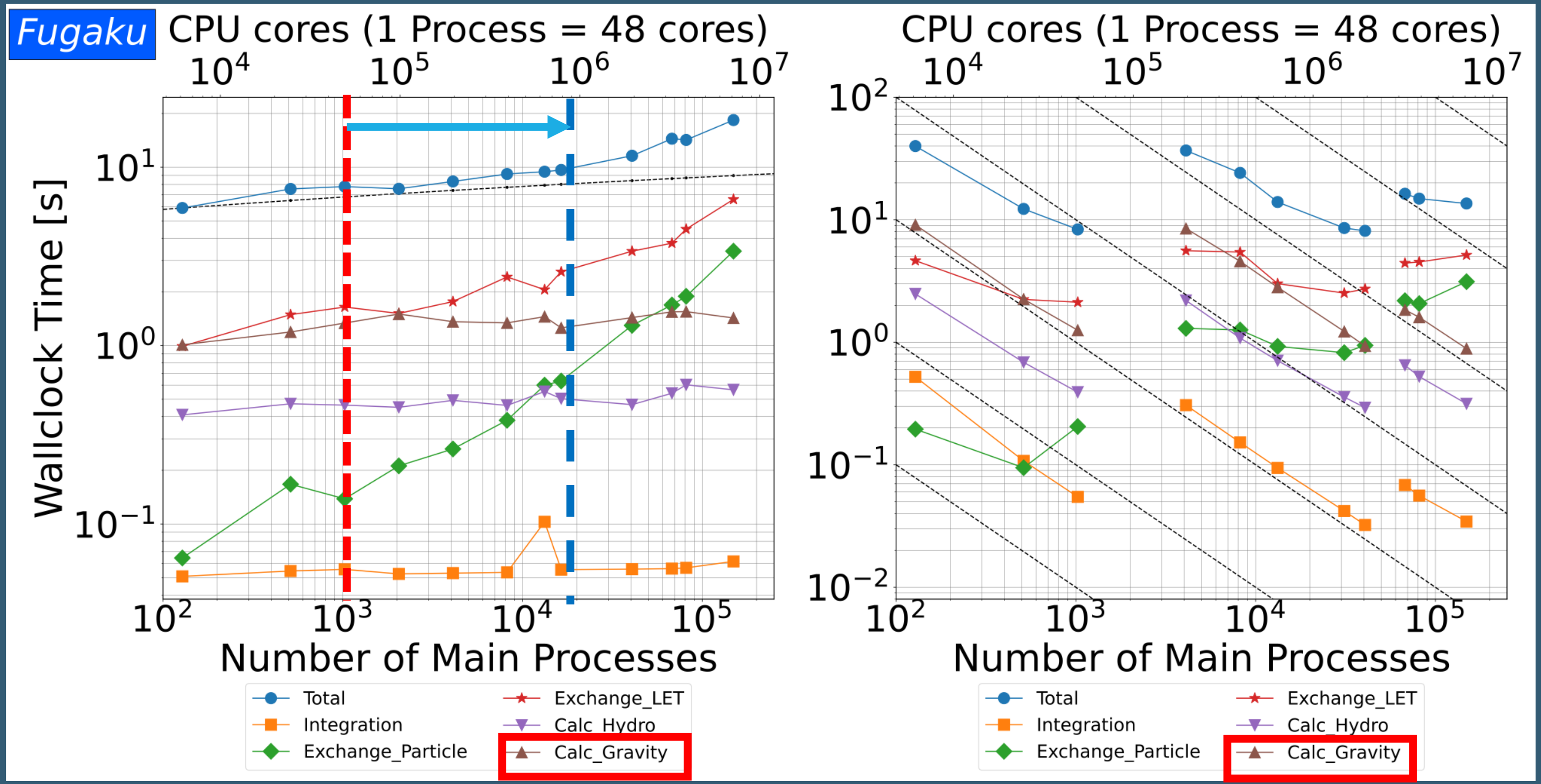
Junichiro Makino
Department of Planetology
Kobe University
Kobe, Hyogo, Japan
Preferred Networks Inc.
Tokyo, Tokyo, Japan
makino@mail.jmlab.jp

- 7M CPU cores （Fugaku）
- High performance across Arm & x86 & CUDA
  - **Efficient on most commercial architectures and NVIDIA GPUs**
- Resolving the Milky Way down to 1 Msun.
- 36 yr-> 3 months for 1 Gyr

Hirashima+25b, SC25

Scaling Performance across multi-architectures

	# Nodes	Processor	Architecture	PFLOPS	Efficiency	#
Fugaku	148,900	A64FX	Armv8.2-A SVE	90.2	9.9%	10^{11}
Rusty	193	Genoa	AVX-512	0.863	35.5%	10^{11}
Miyabi	1024	GH200	CUDA	5.60	8.1%	10^{10}



- The full simulation could finish within half a year if each step takes less than 10 seconds.
- When using ~20k nodes, the communication overhead is significant.
- As the number of particles per process increases, the relative impact of communication decreases.
- Gravity kernel is heavy but scalable.

Hirashima et al. SC25