Simulating self-interacting dark matter on GPUs

MPA Cosmology Seminar

Oliver Zier (CfA)

(together with Rüdiger Pakmor,
Xuejian Shen, Rongrong Liu,
Lucas Schleuss, Mark Vogelsberger, …)



What is SIDM?

CDM

COLD DARK MATTER

COLD
non-relativistic

DARK
no electromagnetic coupling

COLLISIONLESS
C[f] = 0

CONTINUOUS
f(x, v, t)

CONTINUUM DYNAMICS

Vlasov	$\partial f / \partial t + v \cdot \nabla_x f - \nabla_x \Phi \cdot \nabla_v f = 0$
Density	$\rho_{\mathrm{DM}}(x,t) = \int f(x,v,t)\, \mathrm{d}^3 v$
Poisson	$\nabla^2 \Phi = 4\pi G\, \rho_{\mathrm{DM}}$
Particles	Monte Carlo samples of f

What is SIDM?

CDM

COLD DARK MATTER

COLD
non-relativistic

DARK
no electromagnetic coupling

COLLISIONLESS
C[f] = 0

CONTINUOUS
f(x, v, t)

CONTINUUM DYNAMICS

Vlasov	$\partial f/\partial t + v\cdot\nabla_x f - \nabla_x \Phi \cdot \nabla_v f$	$= C[f; \mathrm{d}\sigma/\mathrm{d}\Omega]$
Density	$\rho_{\mathrm{DM}}(x,t) = \int f(x,v,t)\, \mathrm{d}^3 v$	
Poisson	$\nabla^2 \Phi = 4\pi G\, \rho_{\mathrm{DM}}$	
Particles	Monte Carlo samples of f	

SIDM

CDM + self-scattering

ELASTIC SCATTERING short range • pairwise • isotropic

STRENGTH: cross section σ

Impact of SIDM

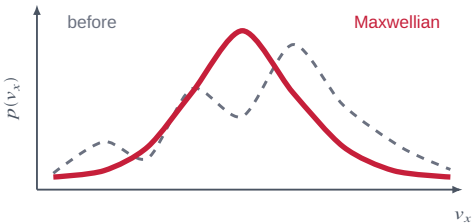
THERMALIZATION

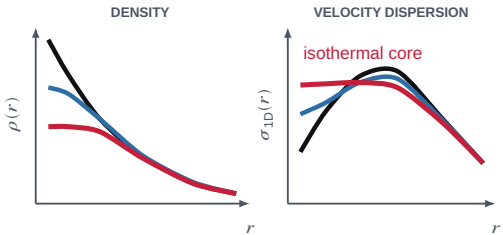
Repeated elastic scattering redistributes particle velocities toward a local Maxwellian.

THERMAL CONDUCTION

Scattering carries kinetic energy from the hotter outer halo into the initially cooler centre.

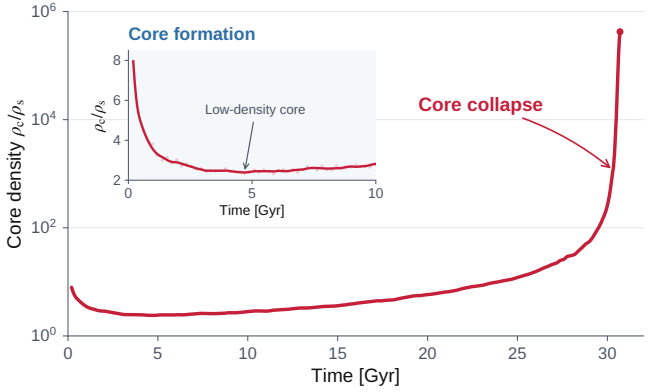
CDM small σ/m larger σ/m





Gravothermal collapse

Core formation is transient: a self-gravitating core has negative heat capacity.



NORI (smoothed) • $N_{200} = 10^7$ • $\sigma/m = 30\,\mathrm{cm^2\,g^{-1}}$

TRANSPORT REGIMES

Long mean-free-path: $\lambda/H \gg 1$
rare scatterings conduct heat

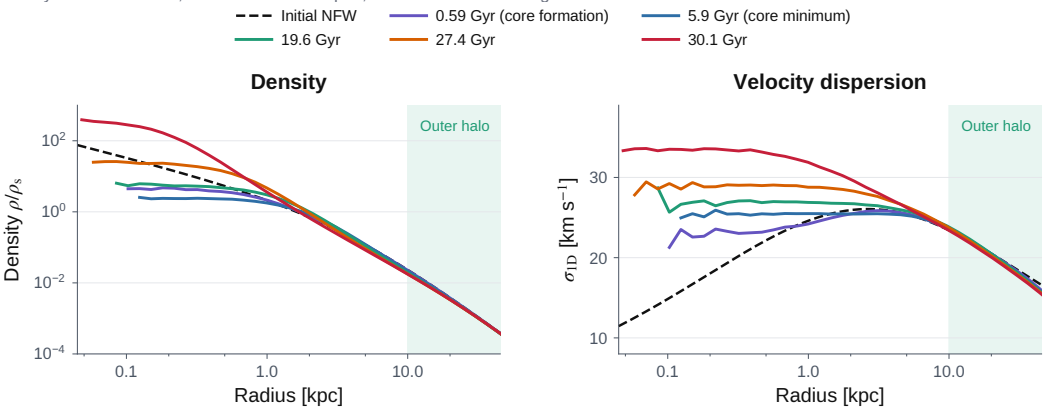
Fluid regime: $\lambda/H \ll 1$

OBSERVATIONAL INTEREST

- ultra-dense strong-lens perturbers
- compact stellar-stream perturbers
- possible seeds for high-z black holes

Only the halo centre collapses

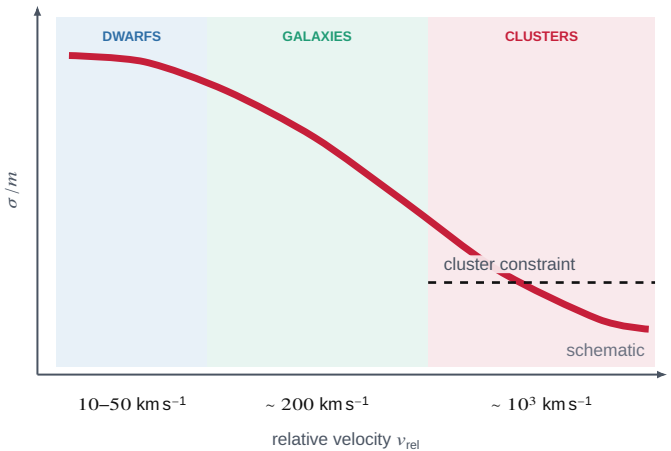
Early core formation, then central collapse; the outer halo changes little.



NORI • $N_{200} = 10^7$ • $\sigma/m = 30\,\mathrm{cm^2\,g^{-1}}$

Observational constraints on the cross section

Different systems probe different characteristic relative velocities.



MODEL SPACE

Cluster bounds disfavor a large constant cross section, while dwarfs can favor much higher scattering rates.

Velocity dependence appears naturally with a massive mediator.

Further directions

- angular dependence $\mathrm{d}\sigma/\mathrm{d}\Omega$
- anisotropic scattering
- dissipative / inelastic channels

Simulating SIDM

GRAVITY: LONG RANGE

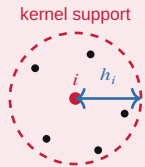
$$\dot{\mathbf{x}}_i = \mathbf{v}_i, \qquad \dot{\mathbf{v}}_i = -\nabla\Phi(\mathbf{x}_i)$$

$$\nabla^2\Phi = 4\pi G\rho, \qquad \rho = \sum_j m_j\,\delta(\mathbf{x}-\mathbf{x}_j)$$

Every particle contributes to every force.

SIDM: SHORT RANGE

$$P(i|j) = \tfrac{1}{2} m_j W(r_{ij}; h_i) \times \frac{\sigma(v_{\text{rel}})}{m} v_{\text{rel}} \Delta t$$



SPH kernels set local rates.

Candidates lie inside h_i.

Calculating gravity

DIRECT SUM

$$\mathbf{a}_i = -G \sum_{j \neq i} m_j \frac{\mathbf{r}_{ij}}{(r_{ij}^2 + \epsilon^2)^{3/2}}$$

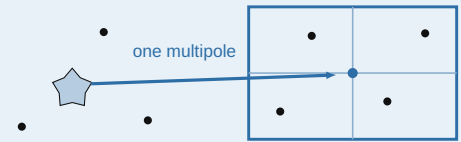
Reference solution with pairwise symmetry;

$\mathcal{O}(N^2)$ work.

HIERARCHICAL TREE

Group distant particles into a monopole or quadrupole. Open only cells that are too large or too close.

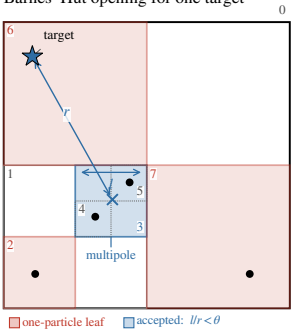
$$l/r < \theta \quad \text{or} \quad GMl^2/r^4 < \alpha |\mathbf{a}_{\text{old}}|$$



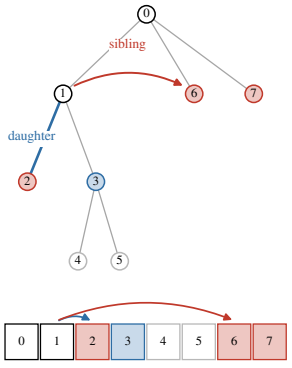
Adaptive resolution: $\mathcal{O}(N \log N)$ work.

Gravity tree walk

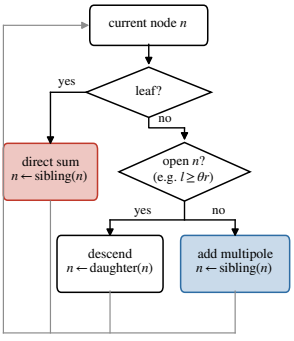
Barnes–Hut opening for one target



links in the flattened tree



binary opening decision

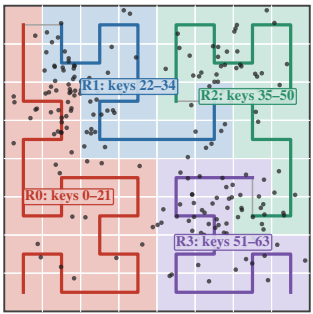


For every target: **direct sum at a leaf** | **descend** | **accept a multipole.**

Domain decomposition

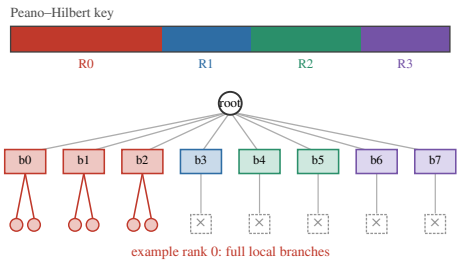
Cut a Peano–Hilbert curve into contiguous, work-balanced key ranges.

contiguous Peano–Hilbert key ranges



compact domains; unequal area can carry comparable work

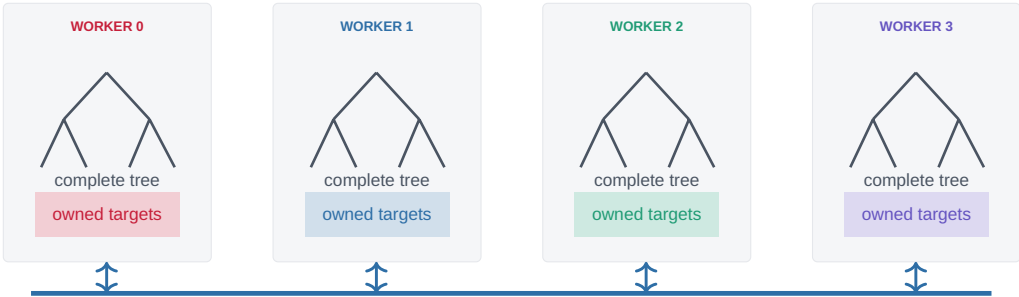
one rank builds a local tree plus pseudo-nodes



Contiguous key ranges preserve locality; unequal spatial volumes balance work and memory.

NORI: fully replicated tree

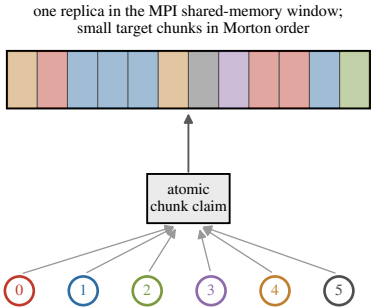
Replicate tree + tree construction, parallelize tree walk



Walk(i) is a pure function of target i and the frozen replica. Any worker can take the next target without changing its arithmetic.

NORI on CPUs

one shared CPU-node replica of the tree



SHARED-MEMORY ISLAND

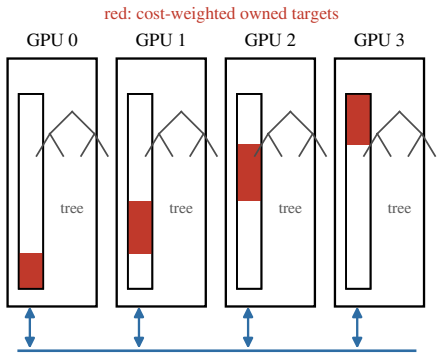
- one state and tree per node / NUMA island
- MPI ranks share one memory-window replica
- ranks claim small Morton-ordered target chunks
- dynamic scheduling balances irregular walks
- tree construction is parallel

Binary invariant: Result is independent of scheduling.

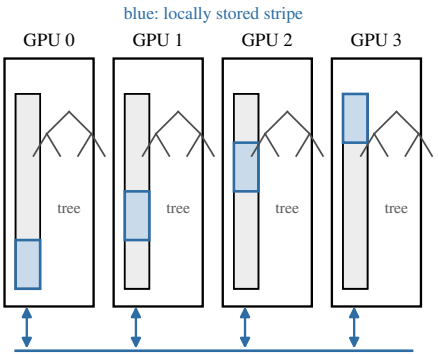
Limit: each island must hold the complete state.

NORI on GPUs

one replica per GPU



one striped replica per GPU island



SIDM in NORI

1 • NEIGHBOURS

Reuse the gravity tree

Adapt h_i; collect nearby partners.

2 • DECIDE

Independent pair trials

Frozen velocities; accept if $u < P(i|j)$.

3 • ORDER EVENTS

Gather and sort

Random priority; pair-index tie-breaks.

4 • APPLY KICKS

Replay pair kicks

Current velocities; isotropic directions.

Stateless random numbers

No RNG state to store or checkpoint.

COUNTER (i_i, i_j, n)
ordered Morton pair + step

KEY (p, s)
purpose + fixed realization salt

Philox4x32-10
stateless mapping

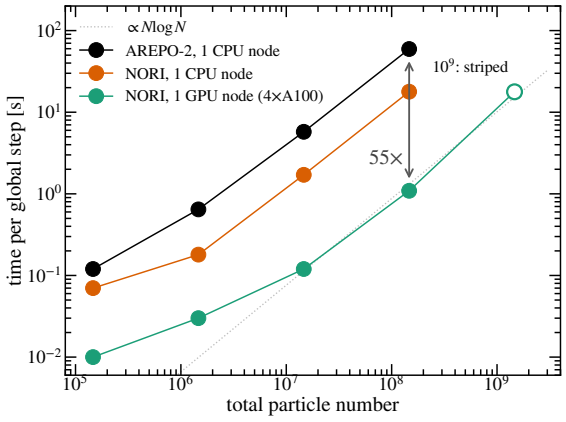
RANDOM DRAWS
Trial acceptance
Event priority
Scattering direction

Same inputs → same random draws on every CPU / GPU.

Pair momentum and kinetic energy conserved **to round-off**, even for repeated scattering.

NORI: problem size and performance

One node each: **GPU** 4×A100, 80 GB each • **CPU** 2×56-core Sapphire Rapids, 112 MPI ranks



GPU SPEED-UP

55×

one NORI GPU node
versus one AREPO-2 CPU node

at $N \simeq 1.5 \times 10^8$

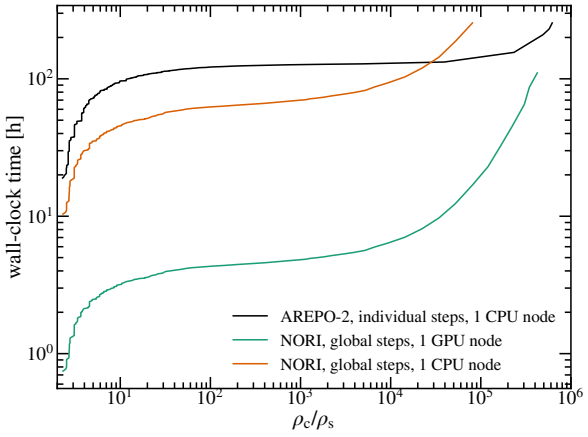
CPU SPEED-UP

$\simeq$ 3×

one NORI CPU node
versus one AREPO-2 CPU node

due to better load balancing

NORI: wall-clock time versus density



TIMESTEP STRATEGY

NORI: global steps

AREPO-2: individual steps

The collapsing core sets the step for **every NORI particle**.

GPU SPEED-UP AT FIXED DENSITY

$26\times \rightarrow 7\times$

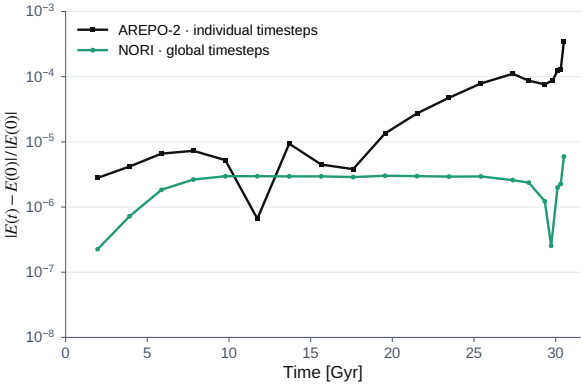
at $\rho_c/\rho_s = 10^3 \rightarrow 10^5$

Global timesteps become costly deep in collapse.

$N_{200} = 10^7$ • $\sigma/m = 30\,\mathrm{cm^2\,g^{-1}}$ • GPU node: 4 A100s • CPU node: 112 cores

NORI: energy errors

Total energy from direct-summation potential + snapshot kinetic energy.



NORI • GLOBAL STEPS

6×10^{-6}

final fractional energy error

AREPO-2 • INDIVIDUAL STEPS

3.5×10^{-4}

final fractional energy error

Same initial halo • $N_{200} = 10^7$ • $\eta = 0.005$ • final comparison at $t \approx 30.5$ Gyr

NORI: Idealized simulations

$\mathcal{O}(10^9)$ particles

GPU and CPU code for isolated, equal-mass SIDM systems.

RESTRICTIONS

- no cosmological integration
- no periodic boundaries
- no hydrodynamics
- no individual timesteps
- memory cost from replication

FEATURES

- cleaner long-term energy behavior
- independent, dynamically scheduled walks
- CPU/GPU execution from one algorithm
- deterministic event ordering

Ideal targets: gravothermal collapse • cluster collisions • interactions of satellites

Cosmological DM-only simulations

GADGET-4

- new SIDM implementation
- individual particle timesteps
- optimized for dark-matter dynamics
- fully cosmological integration
- custom optimizations: domain decomposition, parallelisation, etc.

TreePM on GPUs?

- short-range GPU tree walks
- long-range mesh with GPU FFTs
- periodic cosmological gravity
- keep particles and mesh on GPUs

WORK IN PROGRESS

AREPO-2?

AREPO-2: SIDM with hydrodynamics

JULY 2026 • arXiv:2607.05504

A Novel Implementation of Self-Interacting Dark Matter in AREPO

Oliver Zier, Xuejian Shen, Vinh Tran, Martin Rosenlyst, Mark Vogelsberger, and Rongrong Liu

SIDM + GALAXY FORMATION

- SMUGGLE
- RIGEL
- IllustrisTNG
- Auriga

LONG-TERM TARGET

Auriga “superstar”

SIDM in numerical clones of the MW and its environment.

NORI: summary

CPU + GPU EXECUTION

- Replicated tree; dynamically scheduled walks
- One algorithm for CPUs and GPUs
- $\mathcal{O}(10^9)$ particles

BINARY INVARIANCE: CPU vs GPU

- Bitwise-identical CPU/GPU results for the same numerical setup
- Independent of work assignment
- Stateless Philox; fixed event order

CONSERVATIVE PAIR UPDATES

- Elastic scattering
- Current velocities at every kick
- Pair momentum and kinetic energy conserved to round-off

LIMITATIONS

- Isolated equal-mass systems; no cosmology or hydrodynamics
- Global steps: collapsing core limits all particles
- Memory cost of full replication

What is NORI?

NORI (THE FOOD)

Japanese seaweed



SHEETS • FLAKES • SUSHI WRAPS

What is NORI?

NORI (THE FOOD)

Japanese seaweed



SHEETS • FLAKES • SUSHI WRAPS



NORI (THE CAT)