Accelerated forward modelling of dark matter dynamics: ML-safety and perfect parallelism

ESI Vienna workshop “Putting the Cosmic Large-scale Structure on the Map: Theory Meets Numerics”

Florent Leclercq

www.florent-leclercq.eu

Institut d’Astrophysique de Paris
CNRS & Sorbonne Université

In collaboration with:
Mayeul Aubin (IAP), Deaglan Bartlett (IAP, Oxford), Marco Chiarenza (IAP, U. Milan), Ludvig Doeser (Stockholm University), Rémi Fahed (IAP), Tristan Hoellinger (IAP), Guilhem Lavaux (IAP)

and the Aquila Consortium

www.aquila-consortium.org

ANR-23-CE46-0006: INFOCW

22 September 2025

Let's define some concepts

- Machine-Learning safety: applying machine learning (ML) — in particular neural networks (NNs) — in a way that ensures the results are either correct by construction or, at worst, suboptimal.
  - Safe uses of ML eliminates the requirement for explainability.
  - Example: data compression, e.g. denoising autoencoders (DAE) to build summaries, information-maximising neural networks (IMNN) for implicit likelihood inference.

Charnock *et al.*, 1802.03537, Makinen *et al.*, 2107.07405, Makinen *et al.*, 2410.07548

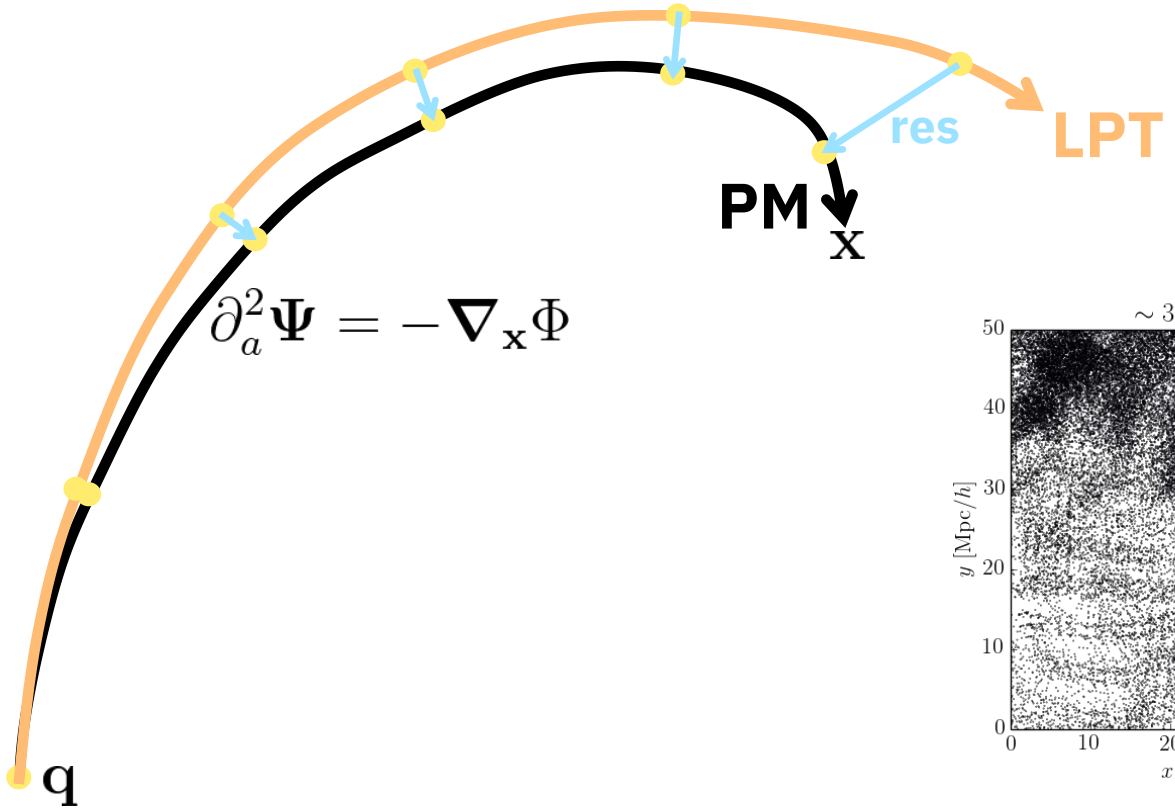
  - Counter-example: emulation of *N*-body simulations. There remains an emulation error [up to $\mathcal{O}(10\%)$] that we cannot ever correct for.

He *et al.*, 1811.06533, Lucie-Smith *et al.*, 1802.04271, Jamieson *et al.*, 2206.04594, Conceição *et al.*, 2304.06099, Doeser *et al.*, 2312.09271, Jamieson *et al.*, 2408.07699

- Perfect parallelism: dividing a computational task into independent sub-tasks with no communication between them, allowing for highly efficient parallel processing.
  - It is a.k.a. an "embarrassingly parallel workload" — but there's nothing to be embarrassed about, really.
  - Examples: computer simulations comparing many independent scenarios, ensemble calculation of i.i.d. numerical model predictions (e.g. for covariance matrix estimation).
  - Counter-examples: usual *N*-body simulation codes, recursive algorithms, Markov Chain Monte Carlo.

The tCOLA framework: (temporal) COmoving Lagrangian Acceleration

- Idea behind tCOLA: we can make use of the analytical solution at large scales and early times: Lagrangian perturbation theory (LPT).



- Write the displacement vector as:

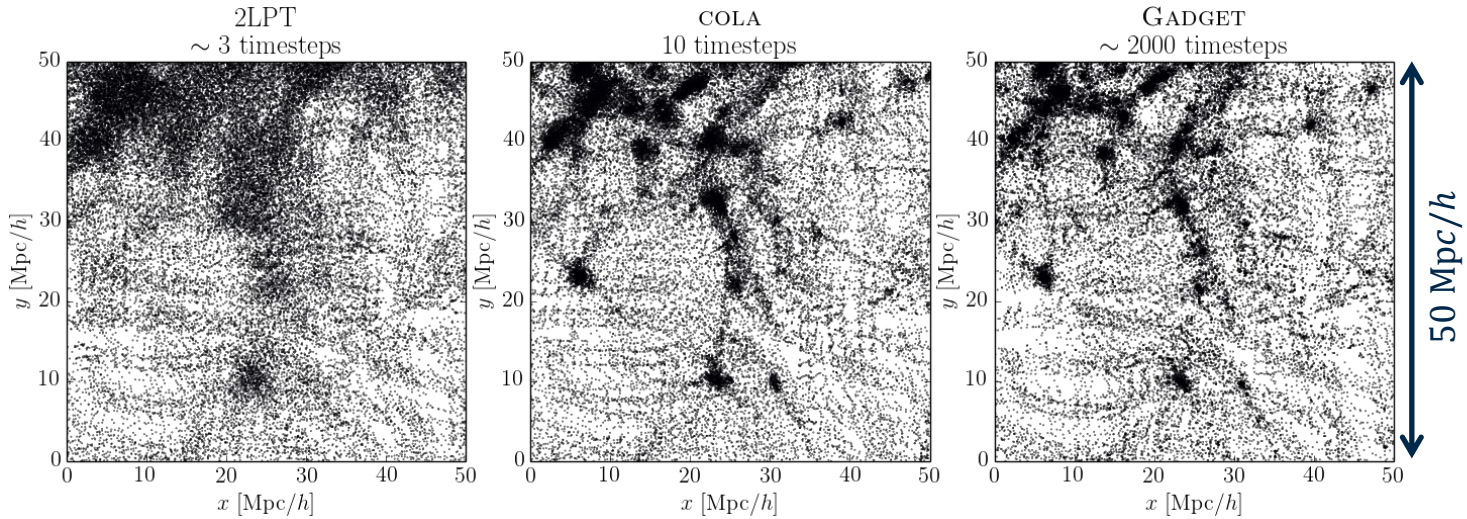
$$\Psi = \Psi_{\rm LPT} + \Psi^{\rm COLA}_{\rm res} \qquad (\mathbf{x} = \mathbf{q} + \Psi)$$

Tassev & Zaldarriaga, 1203.5785

- Equation of motion (omitted constants and Hubble expansion):

Analytical solutions!

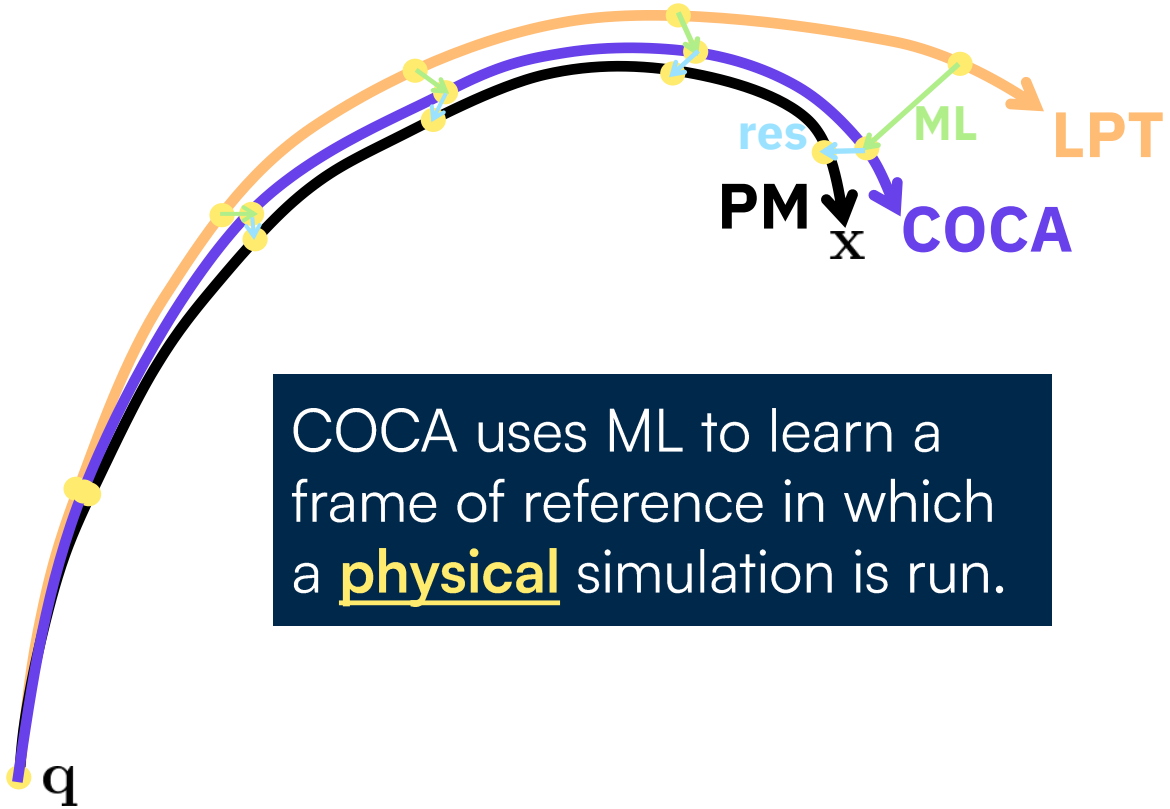
$$\partial_a^2 \Psi^{\rm COLA}_{\rm res} = \partial_a^2 (\Psi - \Psi_{\rm LPT}) = -\nabla_{\mathbf{x}} \Phi - \partial_a^2 \Psi_{\rm LPT}$$



Tassev, Zaldarriaga & Einsenstein, 1301.0322

The tCOCA framework: (temporal) COmoving Computer Acceleration

- Idea behind tCOCA: the easiest simulation to run is the one where nothing moves!



COCA uses ML to learn a frame of reference in which a **physical** simulation is run.

- Write the displacement vector as:

$$\boldsymbol{\Psi} = \boldsymbol{\Psi}_{\rm LPT} + \boldsymbol{\Psi}_{\rm ML} + \boldsymbol{\Psi}^{\rm COCA}_{\rm res} \qquad (\mathbf{x} = \mathbf{q} + \boldsymbol{\Psi})$$

- Equation of motion (omitted constants and Hubble expansion):

$$\partial_a^2 \boldsymbol{\Psi}^{\rm COCA}_{\rm res} = -\boldsymbol{\nabla}_{\mathbf{x}}\Phi - \partial_a^2 \boldsymbol{\Psi}_{\rm LPT} - \partial_a^2 \boldsymbol{\Psi}_{\rm ML}$$

$$\Leftrightarrow \quad \partial_a^2 \boldsymbol{\Psi} = -\boldsymbol{\nabla}_{\mathbf{x}}\Phi$$

- With COCA:
  - Any emulation error will be corrected by solving the *correct* physical equation of motion.
  - Any ML algorithm can do the job!
  - Building a data model is a safe use of ML.

Bartlett, Chiarenza, Doeser & FL, 2409.02154

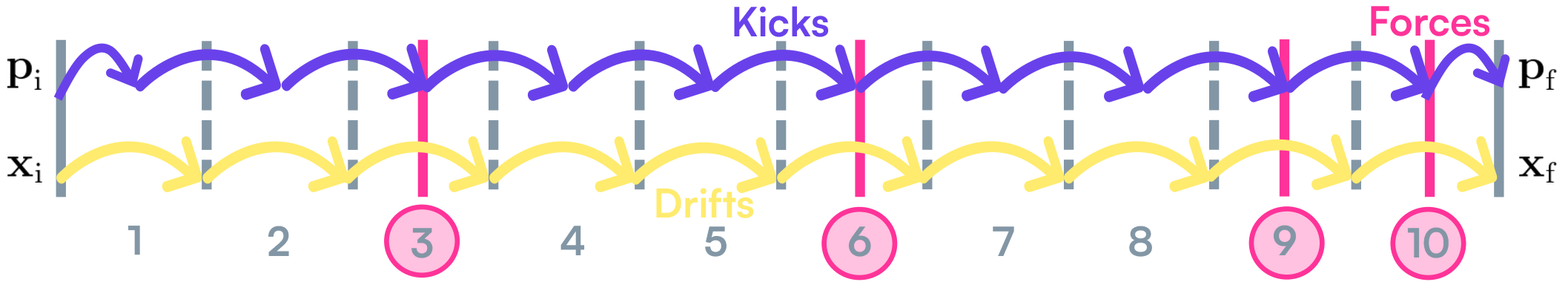
Time stepping and force calculations in COCA

- Our implementation of COCA in the Simbelmynë code uses the standard Kick-Drift-Kick (leapfrog) discretisation of the equation of motion.

https://simbelmyne.florent-leclercq.eu — Git:Aquila-Consortium/simbelmyne

- Learning the new frame of reference means emulating the COLA residual momenta at every time step: $\mathbf{p}_{\rm res}^{\rm COLA} = \mathbf{p} - \mathbf{p}_{\rm LPT}$.

- When the emulation error is small ($\mathbf{p}_{\rm ML} \approx \mathbf{p}_{\rm res}^{\rm COLA}$), particles are already at rest in the COCA frame of reference, so it is unnecessary to compute forces at every step.

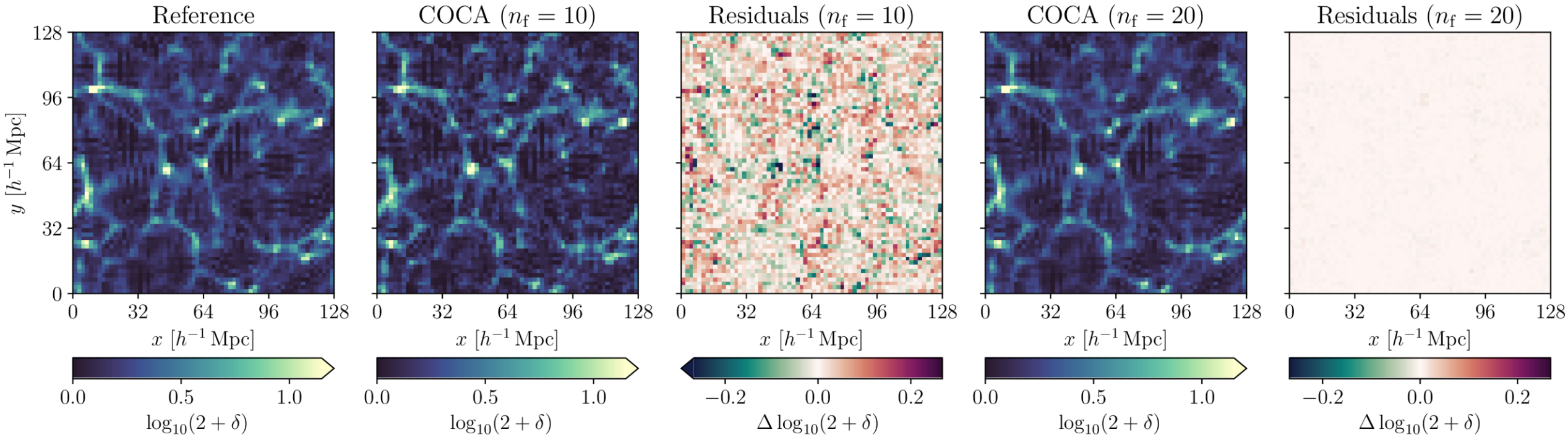


- A good frame-of-reference emulator therefore makes COCA cheaper than COLA.

Bartlett, Chiarenza, Doeser & FL, 2409.02154

Results: COCA density field

Deaglan Bartlett
(PDRA at IAP → Oxford)



We reach percent-level accuracy up to $k = 1\ h/\mathrm{Mpc}$ on standard correlations functions, using only 8 to 10 particle-mesh (PM) force evaluations (see the paper).

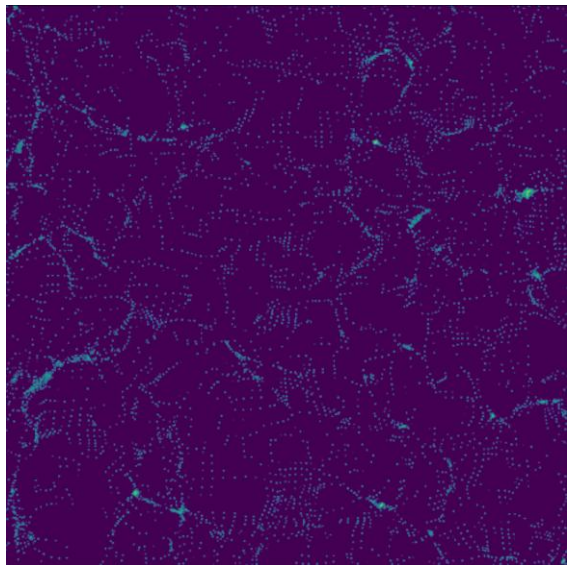
Bartlett, Chiarenza, Doeser & FL, 2409.02154

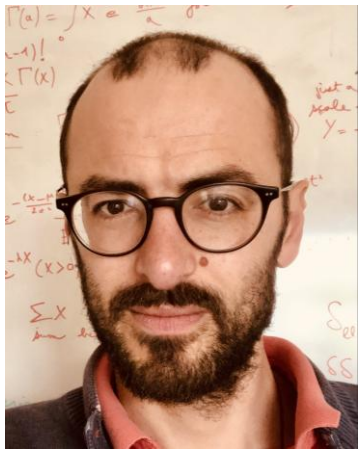
Force calculation and the small-scale accuracy of COLA/COCA

- A common misconception: COLA (or COCA) does not necessarily sacrifice small-scale accuracy for speed! (only implementations with PM forces usually do).
- Changing the frame of reference (to LPT or LPT+ML) can be done with *any* force calculation technique. Therefore, trajectories can be integrated to arbitrary accuracy.

Spectral sheet interpolation

(PRELIMINARY)





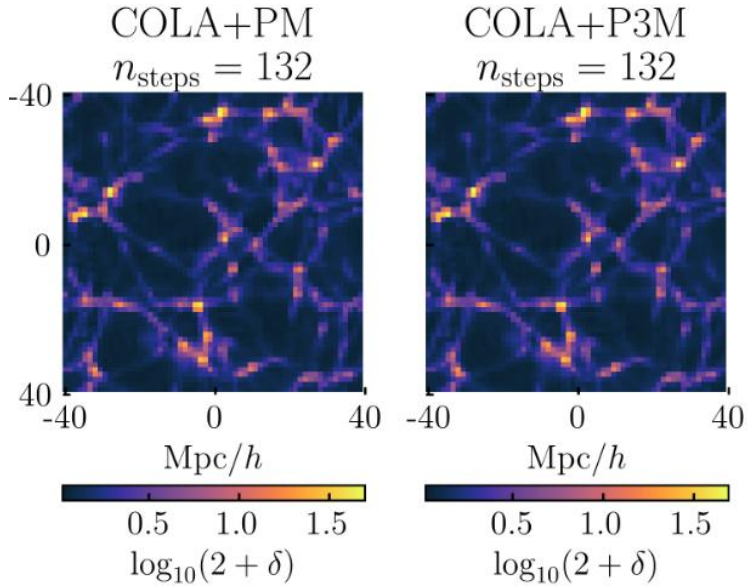
Rémi Fahed
(research engineer at IAP)

Implemented as in List *et al.*, 2309.10865

COLA particle-particle particle-mesh (P3M) dynamics

(PRELIMINARY)

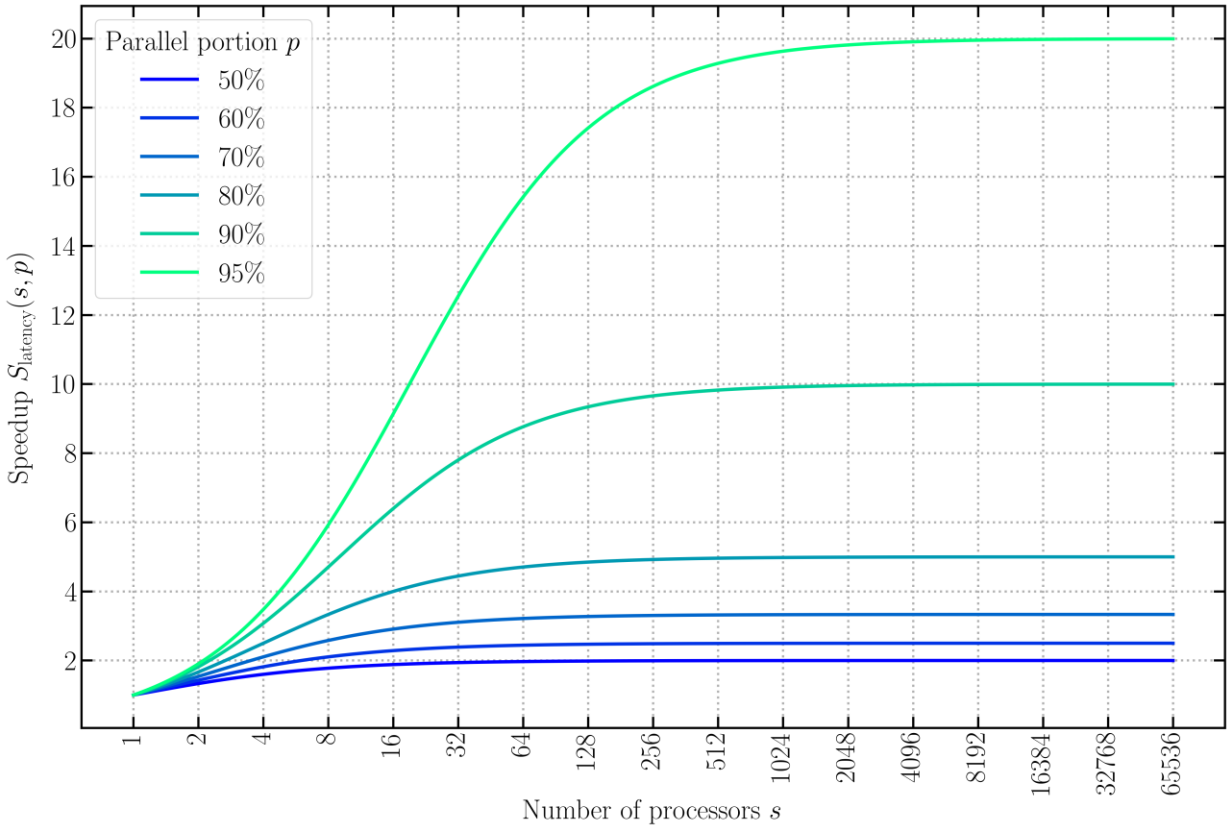
Tristan Hoellinger
(doctoral researcher at IAP)



Implemented as in Dakin *et al.*, 2112.01508

The real challenge of parallelisation: long-range gravity

- tCOLA/tCOCA provide a pleasant reduction in the number of timesteps…
- …but: the real challenge, preventing the easy parallelisation of *N*-body codes, is the long-range nature of gravitational interactions:
  - It is a communication-dominated problem.
  - It requires sophisticated memory management to maximise locality in the storage hierarchy (cache, RAM, disk I/O).
- Amdahl's law: latency kills the gains of parallelisation.
  - For example, Gadget-4 loses strong scaling before 1,000 cores due to domain decomposition and long-range gravity.



Are high-performance communications really needed to solve a quasi-linear problem at large scales? Shouldn't locality in memory come by construction?

Amdahl 1967, doi:10.1145/1465482.1465560, Springel *et al.*, 2010.03567 (Figure 63)

Perfectly parallel cosmological simulations using spatial comoving Lagrangian acceleration (sCOLA)

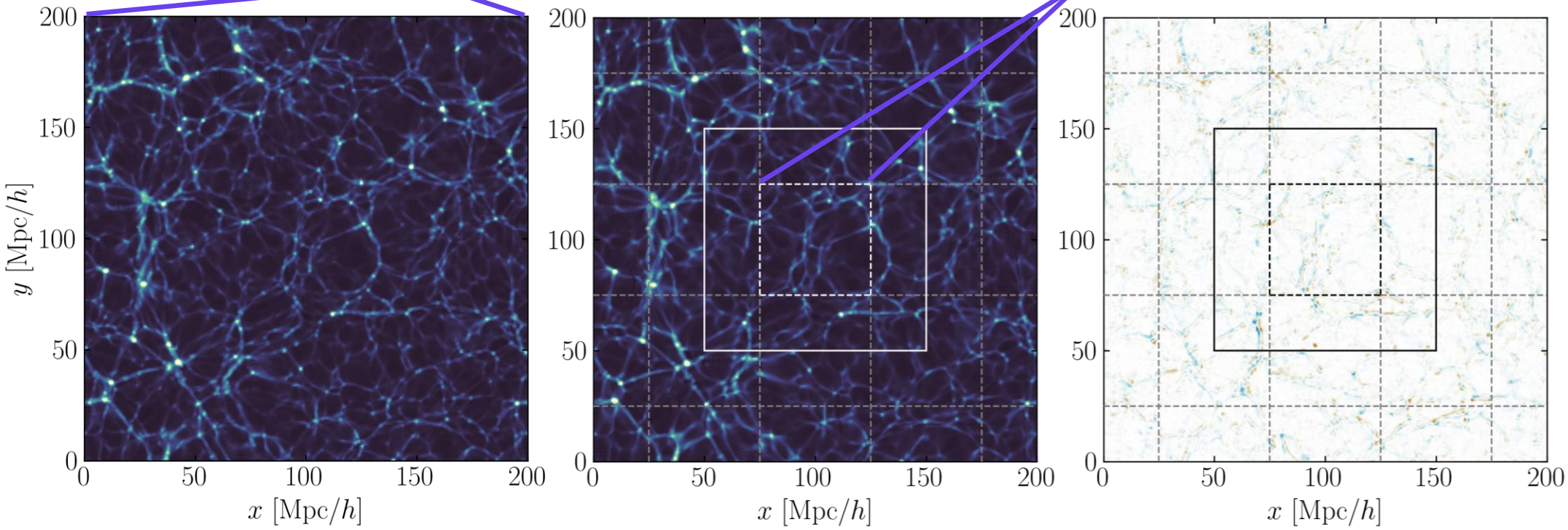
- Can we decouple sub-volumes by using the large-scale solution?

LPT so far (analytical solution) → sCOLA; soon ML solution → sCOCA

$$\partial_a^2 \boldsymbol{\Psi} = -\boldsymbol{\nabla}_{\mathbf{x}} \left[\Delta^{-1}\delta\right] \Leftrightarrow \partial_a^2(\boldsymbol{\Psi} - \boldsymbol{\Psi}_{\text{l.s.}}) = -\boldsymbol{\nabla}_{\mathbf{x}} \left[\Delta^{-1}(\delta - \delta_{\text{l.s.}})\right]$$

non-local

local



tCOLA (reference) | sCOLA | Difference

FL, Faure, Lavaux, Wandelt, Jaffe, Heavens, Percival & Noûs, 2003.04925

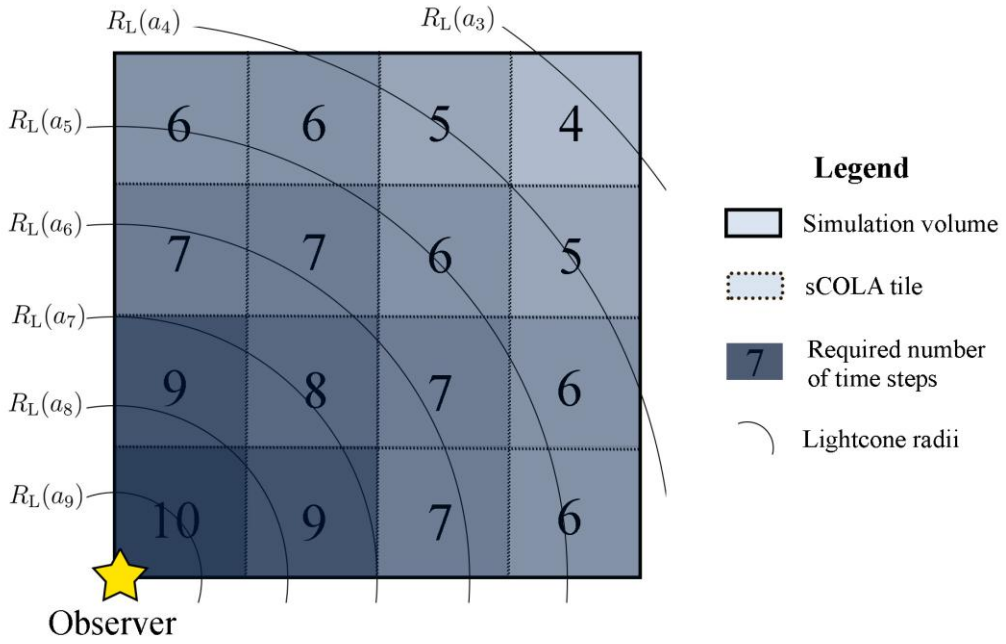
Lightcones and mock catalogues with sCOLA

- The workload in sCOLA is perfectly parallel, with a parallelisation potential factor:

$$p = s\left(\frac{L}{L_{\mathrm{sCOLA}}}\right)^3$$

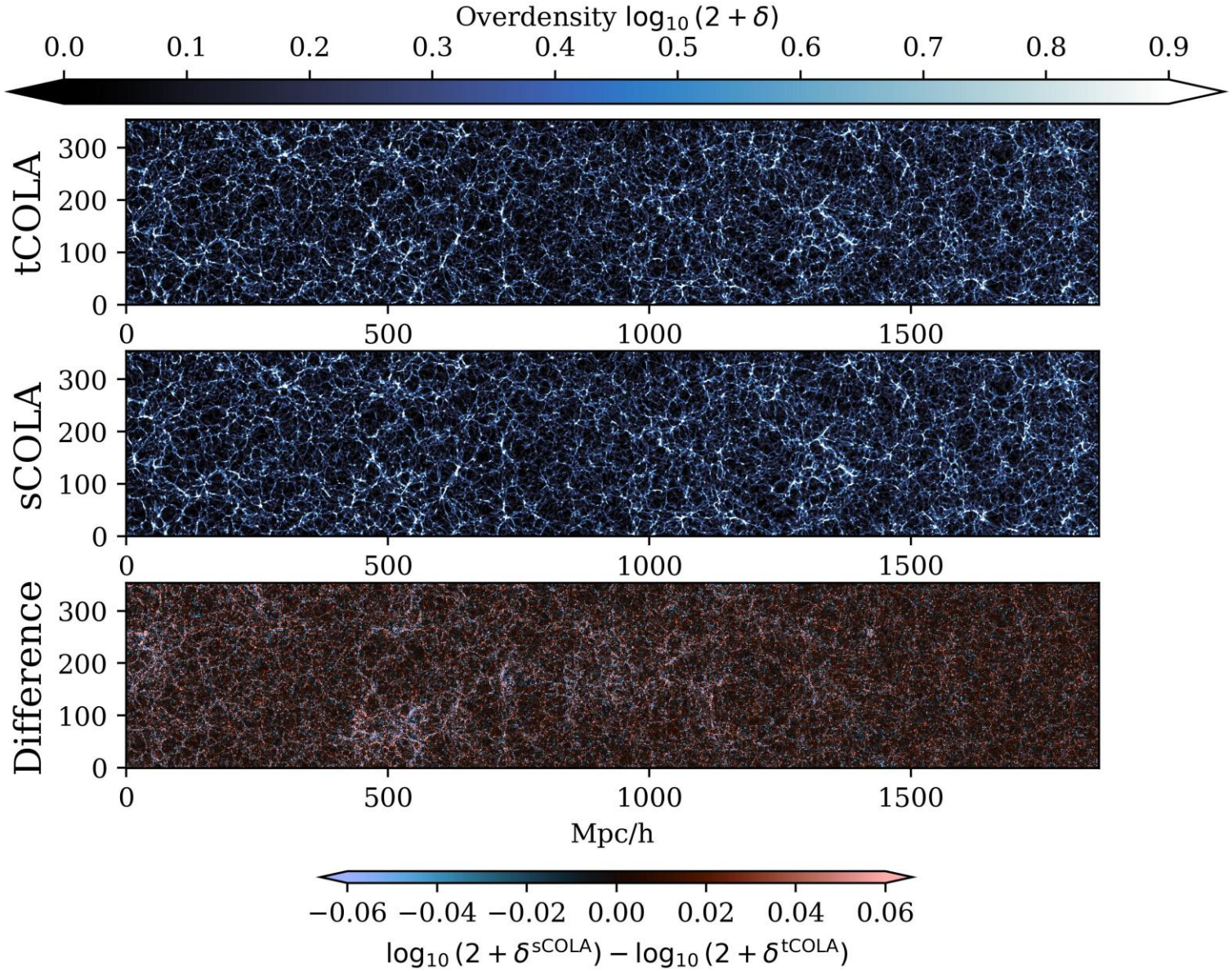
hardware "boost" factor due to small memory requirements

- Generation of lightcones and mock catalogues:
  - sCOLA boxes only need to run until they intersect the observer's past lightcone.
  - Most of the high-z volume will run faster than $z = 0$.
  - Many unobserved sCOLA boxes do not even have to run!
  - The wall-clock time limit is the time for running a single sCOLA box to $z = 0$ at the observer's position.



- Additional benefits:
  - Grid computing: the algorithm is suitable for inexpensive, strongly asynchronous networks
  - Robustness to node failure

tCOLA and sCOLA lightcone simulations

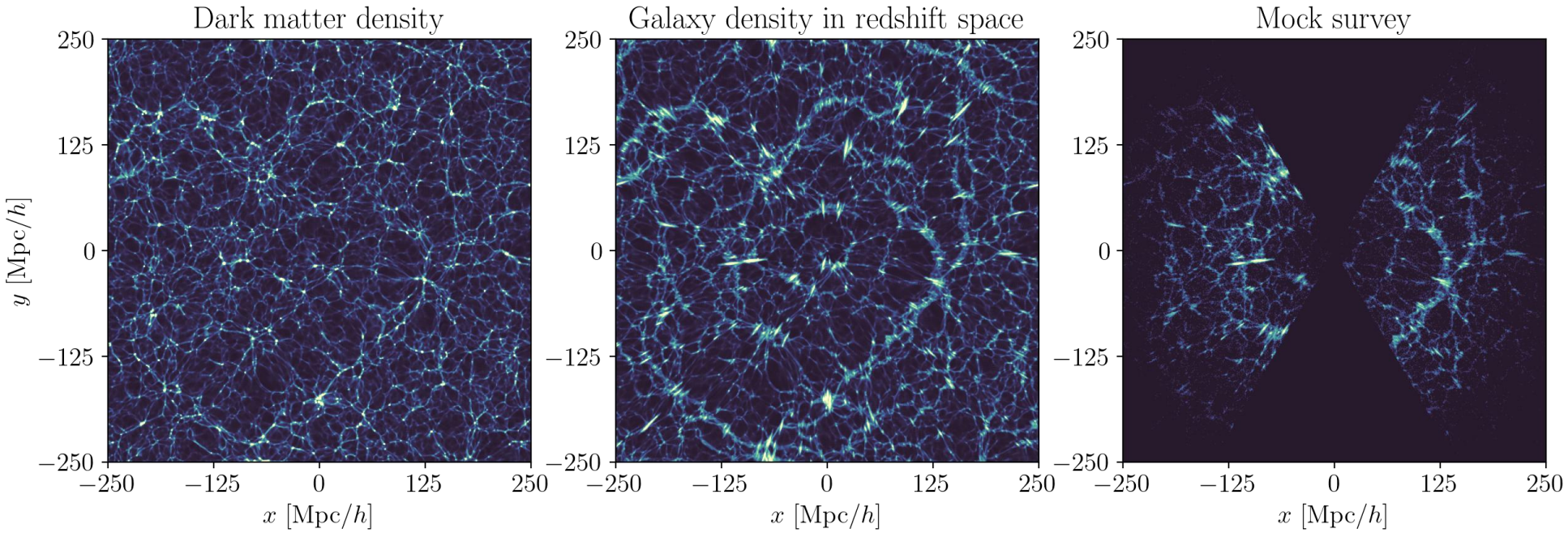


(PRELIMINARY)

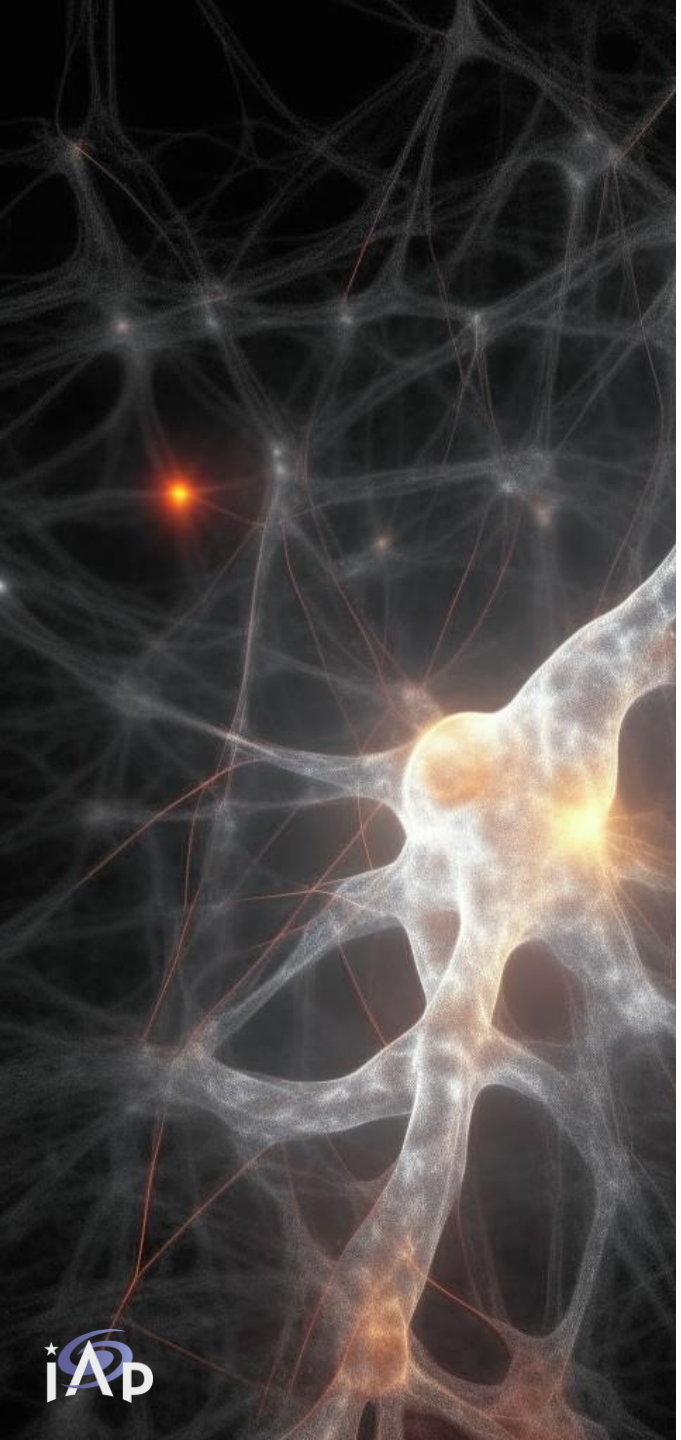
Mayeul Aubin
(doctoral researcher at IAP)

Simbelmynë: an end-to-end code for synthetic galaxy surveys

- Our code Simbelmynë is publicly available at https://simbelmyne.florent-leclercq.eu/.



"How fair are the bright eyes in the grass! Evermind they are called, simbelmynë in this land of Men, for they blossom in all the seasons of the year, and grow where dead men rest."



Conclusions – ML-safety and perfect parallelism

- tCOCA reimagines the use of neural networks for emulating *N*-body simulations:
  - It generalises the idea of tCOLA: running simulations in a new frame of reference. But it is not an emulator!
  - It solves the correct equations of motion, so it is a ML-safe use of neural networks. Explainability is not needed!
  - It makes simulations cheaper by skipping unnecessary force evaluations. But any force calculation technique (e.g. P3M) can be used.
- sCOLA uses the large-scale approximate solution to spatially split simulations in independent tiles:
  - It achieves perfect parallelism by fully removing the need for communications across the full computational volume.
  - It allows for fast lightcone and mock catalogue generation.
- Outlook: the large-scale ML solution can also be used to decouple sub-volumes, in the same spirit as sCOLA: the sCOCA framework!

Acknowledgements, credits, contacts

Slides at:
florent-leclercq.eu/talks.php

References:

- **Simbelmynë**: Leclercq, Jasche & Wandelt 2014, 1403.1260, *Bayesian analysis of the dynamic cosmic web in the SDSS galaxy survey* — https://simbelmyne.florent-leclercq.eu
- **sCOLA**: Leclercq *et al.* 2020, 2003.04925, *Perfectly parallel cosmological simulations using spatial comoving Lagrangian acceleration*
- **tCOCA**: Bartlett, Chiarenza, Doeser & Leclercq 2024, 2409.02154, *COmoving Computer Acceleration (COCA): N-body simulations in an emulated frame of reference*

www.florent-leclercq.eu

www.aquila-consortium.org

The author acknowledges the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-23-CE46-0006 (project INFOCW).
The author does not acknowledge any support from a famous American soda company.