pFoF

A highly scalable halo-finder for large cosmological data sets

Fabrice Roy, Vincent Bouillot, Yann Rasera

Brief history

Situation in 2008:

- Serial fof developped severial years earlier by Edouard Audit
- 9 large RAMSES n-body simulations planned (1024^3 particles)
- Discussion with Patrick Hennebelle who suggested the idea behind pFoF (parallel Friends-of-Friends halo finder)
- Development of the first version of pFoF (Roy, Bouillot, Rasera, 2014)
- New version in 2015

What is this idea behind pFoF?

- Divide the simulation box
- Perform a serial halo detection in each subdivision
- Merge the halos that extend across several subdivisions

Merge = give each halo particle the same halo ID

From this idea, 2 versions of pFoF

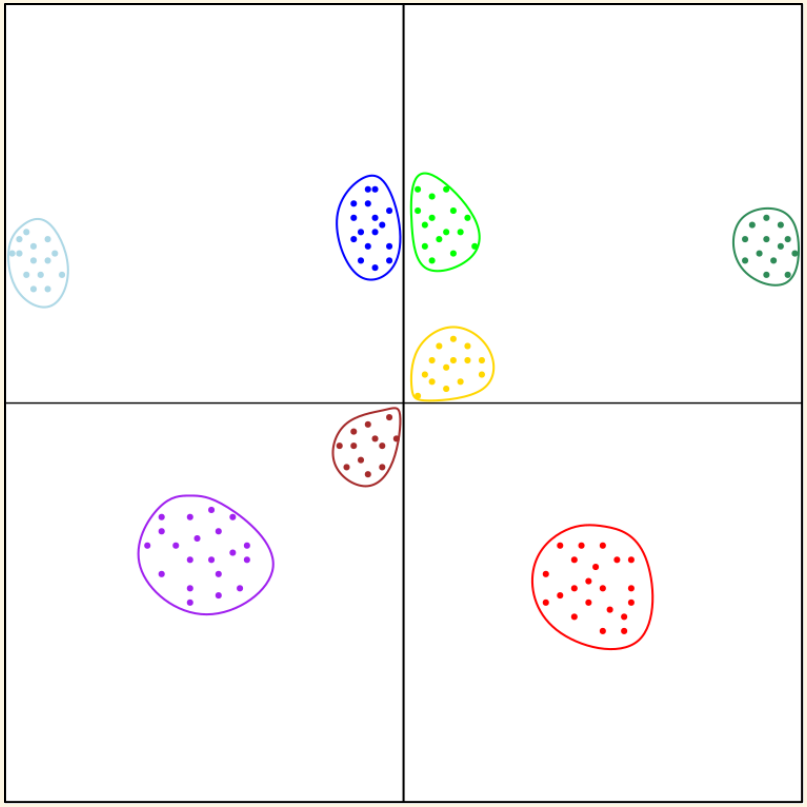
pfof_snap for RAMSES snapshots

- Analyses a RAMSES particles snapshot
- Assumes the boundary conditions are periodic
- Can write the particles distributed in cubic subdivisions ("cube files")
- Can read RAMSES binary files or previously written cube files
- Writes 2 kinds of files: halo properties and list of particles per halo

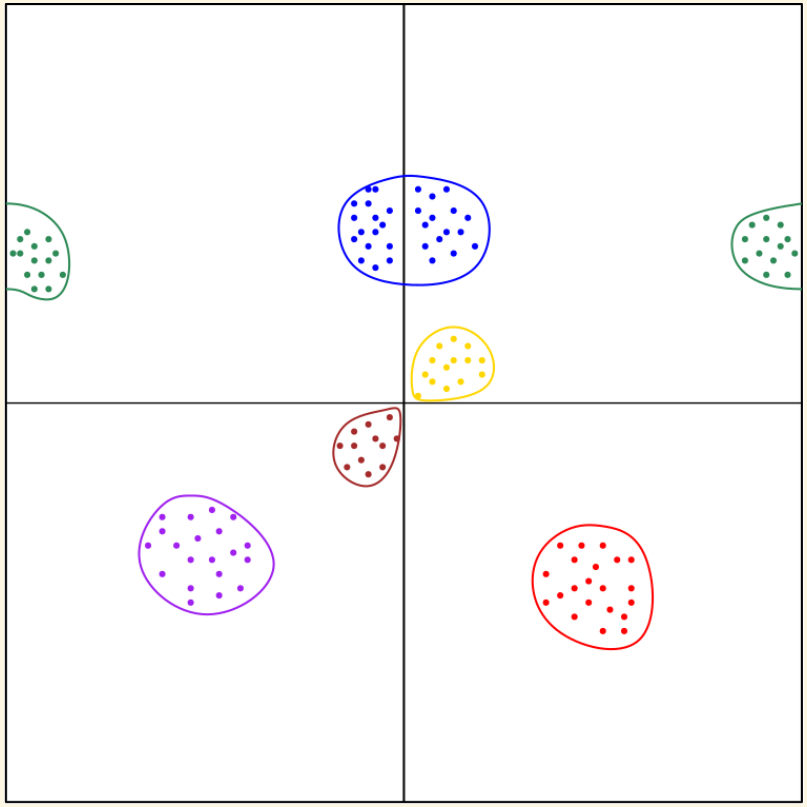
pfof_cone for RAMSES lightcones

- Analyzes a RAMSES particles lightcone
- First version by Vincent Bouillot for DEUS-FUR lightcones (>100 billions particles) in 2012
- New optimized version in 2015
- pfof_cone processes are "mapping" the cone
- No boundary conditions
- Only reads hdf5 shells produced by a tool (conepartcreator)
- Writes 2 kinds of files: halo properties and list of particles per halo

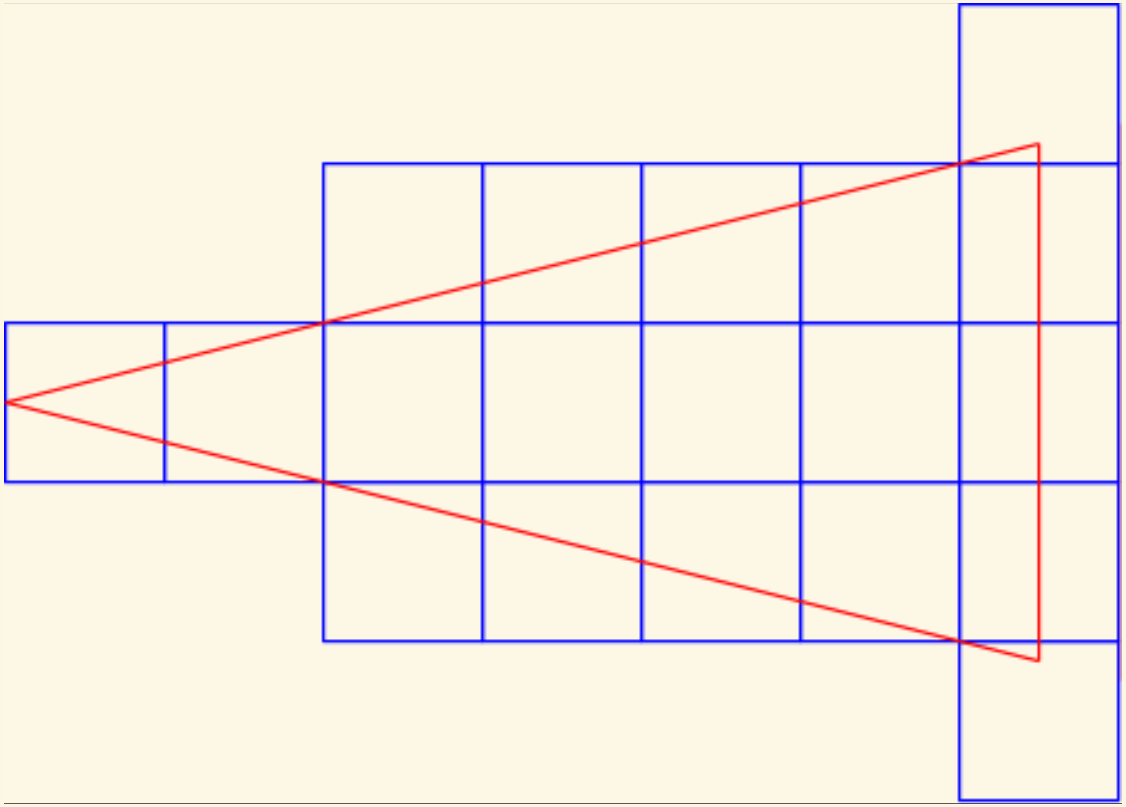
What does it look like?



What does it look like?



How is the lightcone mapped?



Halo properties file

HDF5 file, with some metadata (RAMSES parameters, pfof parameters, etc...)

- Only 1 file, parallel write
- Contains a list of halos with:
  - their number of particles
  - the position and velocity of their center of mass
  - the "radius" of the halo

Halo particles files

HDF5 files, with some metadata (RAMSES parameters, pfof parameters, etc...)

- Several files, 1 per process or 1 per group of processes (parallel write in this case)
- One group per halo
- In each group, the properties of each particle

Pros/Cons of these files

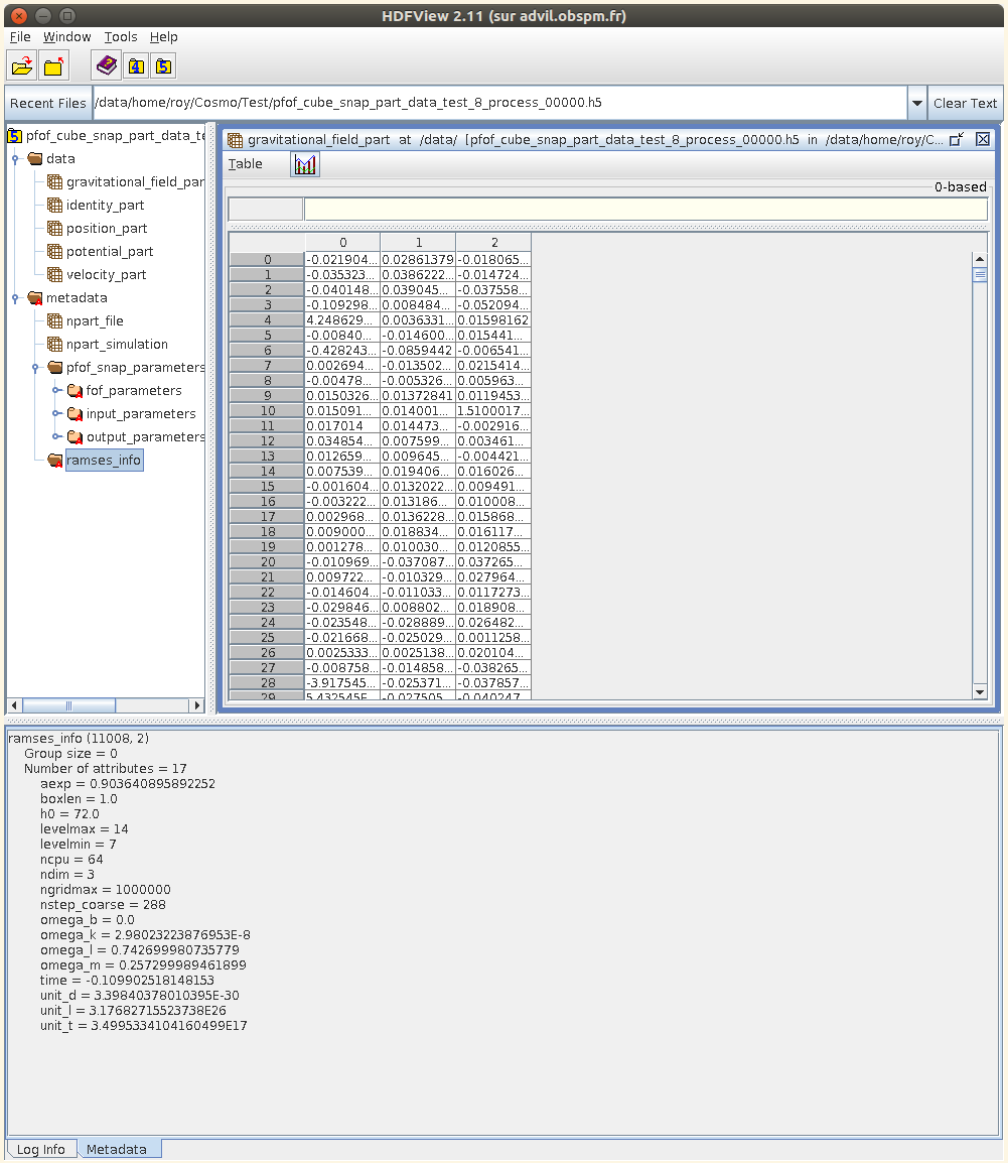
Pros:

- Output files allow easy further analysis (with pFoF, other halo finder, other codes)
- You can perform a FoF with a lower b from halo particles files
- Cubes are easier to handle than the RAMSES binary files
- Parallel writes reduce the number of files

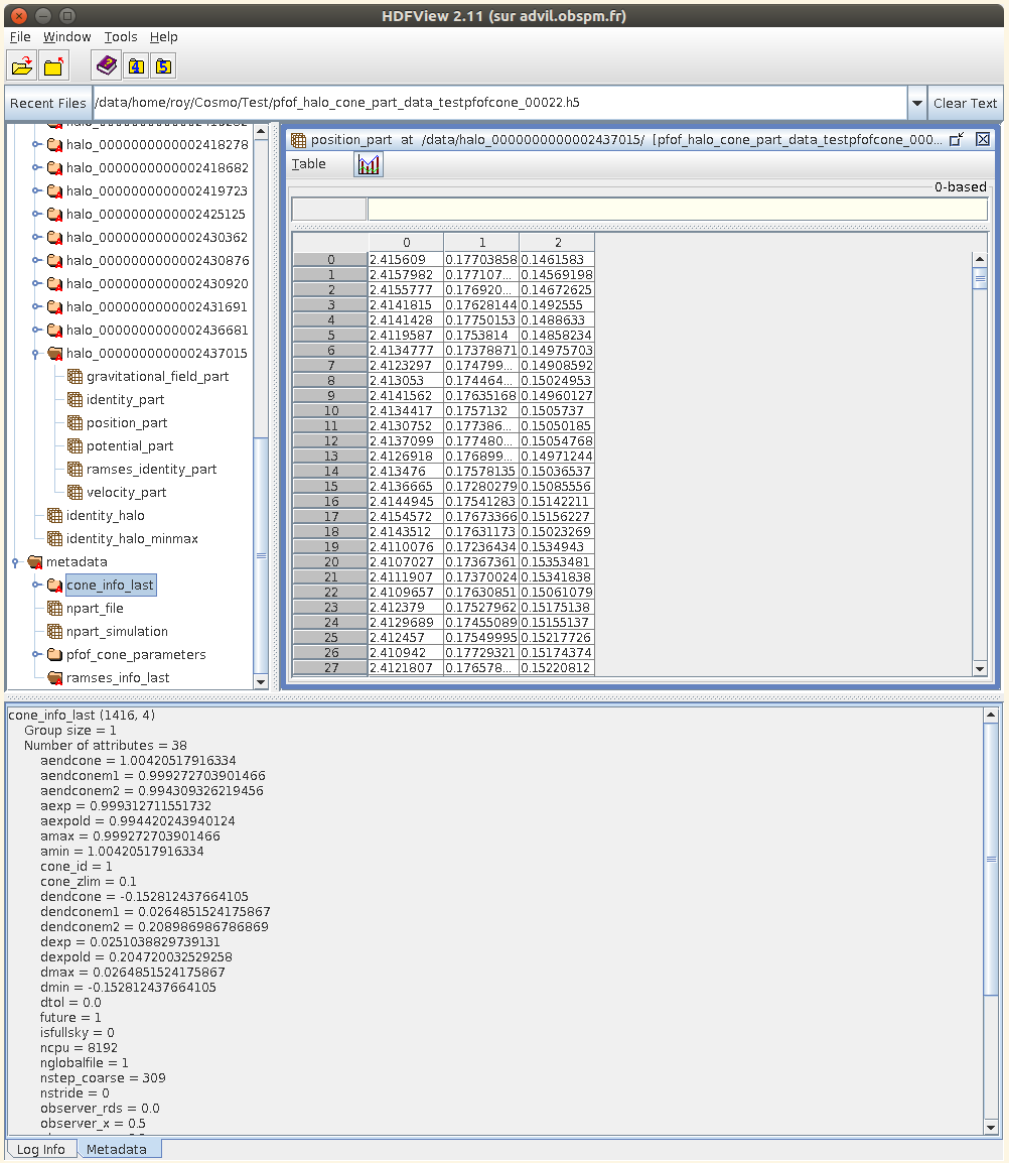
Cons:

- Very costly: lots of communications to gather the particles
- See performance for an example

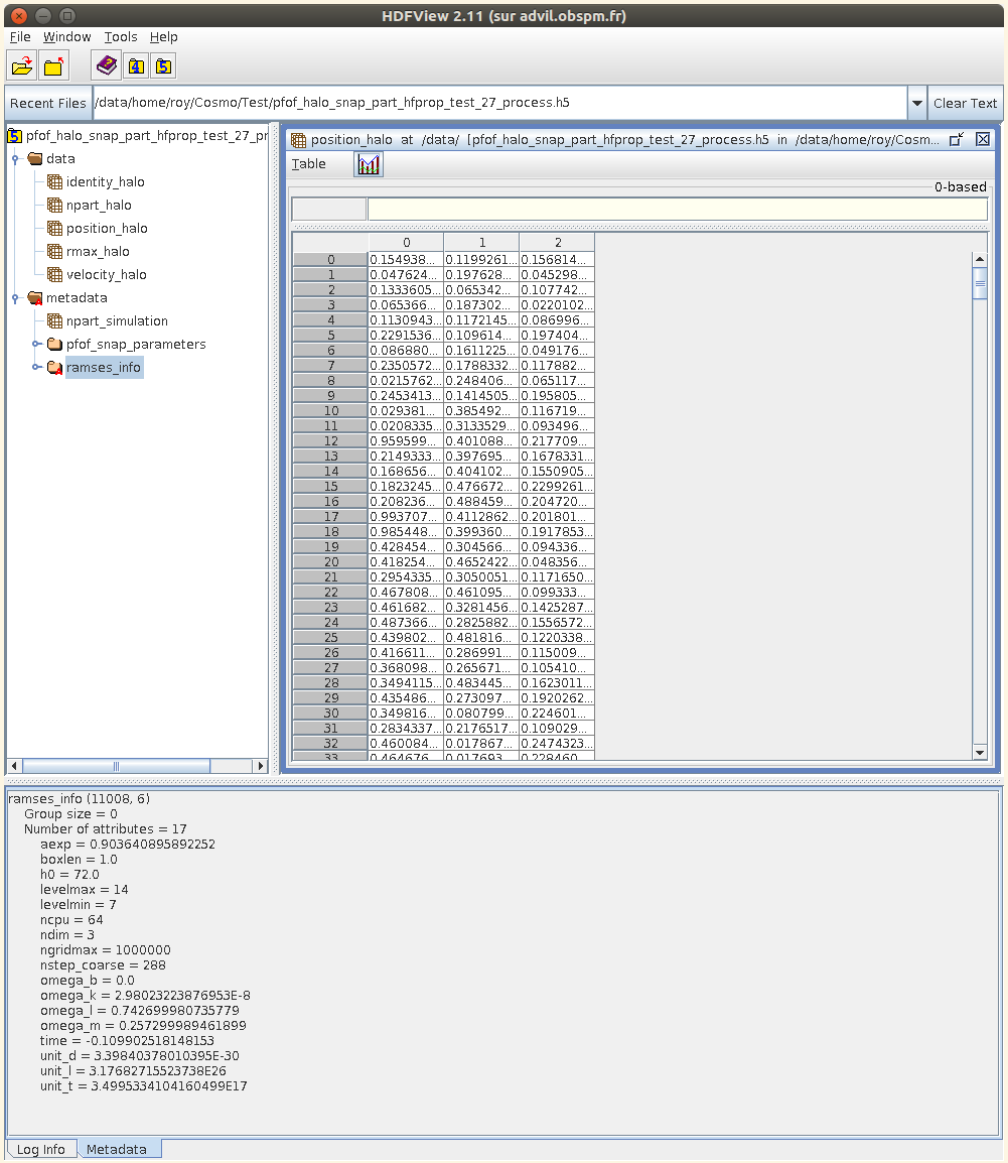
HDF5 structure of the files



HDF5 structure of the files



HDF5 structure of the files



Performances (1) - Weak scaling

ΛCDM, a=0.3, b=0.2, 1024^3, 64 proc., on Curie (TGCC): 483 s.

- input: 57 s.
- local fof + merging: 43 s.
- halo properties + output: 232 s.

ΛCDM, a=0.3, b=0.2, 4096^3, 4096 proc., on Curie: 1923 s.

- input: 605 s. (bad)
- local fof + merging: 47 s. (good)
- halo properties + output: 1015 s. (bad)

Bad comm. scheme + poor I/O scaling on Curie with >512 processes (Alimi et al. 2012)

Performance (2)

Some examples of pfof_cone execution on Curie

- Narrow cone, ΛCDM, 648 Mpc/h, 2048^3, ~7.9×10^9 part., b=0.2, 1332 proc.: 190s.
- A large part of the costly I/O is done by conepartcreator: 1h30 on 16 proc.
- Fullsky, ΛCDM, 2592 Mpc/h, 2048^3, ~9.2×10^9 part., b=0.2, 1472 proc.: 206s.
- conepartcreator: 3h. on 16 proc.

Performance (3)

Largest snapshot analyzed with pFoF (previous version, 2012):

- ΛCDM, 21Gpc/h, a=1, b=0.2
- 8192^3 particles, 32k proc., on Curie
- Walltime = approx. 2h

Future developments: some ideas

- improve the input and halo properties/output phases
- develop an hybrid MPI-OpenMP version
- add subhalo detection
- add unbinding
- adapt pFoF for zoom and hydro simulations
- ...

If you want to use pFoF

- No public git or subversion yet, it will come soon!
- You can send me an email: fabrice.roy@obspm.fr
- I would be glad to help you install/use/adapt pFoF
- People have already used it at Strasbourg Observatory and KASI in Daejeon (Korea)

Thank you!

l'Observatoire de Paris | LUTH
Laboratoire de l'Univers et de ses Théories