Halo-matter cross-correlation in cosmological simulations

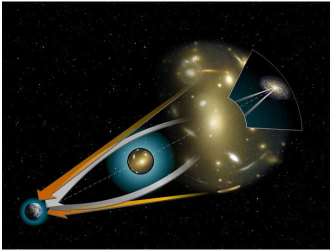
Relatore: Prof. Giuseppe Tormen
Correlatore: Prof. Ravi Sheth
Laureando: Brunetto Marco Ziosi

Dipartimento di Astronomia
Università di Padova

22 Luglio 2011

Outline

Large scale structure investigation



Compare theoretical predictions to:

- Numerical simulations
- Observations (e.g. gravitational lensing)

Halo model extension: sub-haloes structure and distribution

Brief history

- Quantum fluctuations amplified by inflation
- Perturbations grow and collapse due to gravitational instability
- Formation of the first structures: dark matter haloes
- Hierarchical aggregation to form larger and larger structures
- Galaxies form inside dark matter haloes

Halo model

Gathers all informations on dark matter structures to predict galaxy distribution and properties

Ingredients

- spatial distribution of dark matter haloes (correlation functions)
- mass functions of dark matter haloes (Press-Schechter)
- internal structure of dark matter haloes (NFW profile)
- spatial distributions of sub-haloes in haloes
- mass function of sub-haloes in host haloes
- internal structures of sub-haloes
- merging history and dynamical evolution of haloes and sub-haloes

Simulations

Simulations properties

Name	$\mathbf{l_{box}}$	$\mathbf{N_{part}}$	**Mass** $(\mathbf{M_\odot})$
GIF	141 Mpc/h	256^3	1.4×10^{10}
GIF2	110 Mpc/h	400^3	1.4×10^{10}
Milli-millennium	62.5 Mpc/h	270^3	1.7×10^{9}
Millennium	141 Mpc/h	2160^3	1.4×10^{10}
Millennium II	100 Mpc/h	2160^3	6.9×10^{6}

Spatial correlation of structures

Characterize spatial distributions of particles in simulations/galaxies and matter in the universe:

Two point correlation function: $\xi(r)$

- Measure the probability to find a pair of objects separated by a distance r compared to a random distribution
- $\mathrm{d}^2 P = n^2 \mathrm{d}V_1 \mathrm{d}V_2 [1 + \xi(r)]$ (discrete)

Estimators

- uses two sets of points: D simulations data, R random set of points
- count different set of pairs: DD (data-data), RR (random-random) DR (data-random)
- Build the Landy&Szalay estimator

$$\xi_{LS} = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)} \quad (1)$$

- Good behavior: $\Delta\hat{\xi}_{LS} = \frac{1+\xi}{\sqrt{DD}}$

Spatial correlation of structures

$\xi(r)$ is the Fourier anti-transform of the power spectrum of fluctuations

The Wiener-Khintchine theorem

- Density fluctuations field: $\delta(x) = \frac{\rho(x)-\rho_{bg}}{\rho_{bg}}$
- The power spectrum of density fluctuations is the Fourier transform of $\xi(r)$:

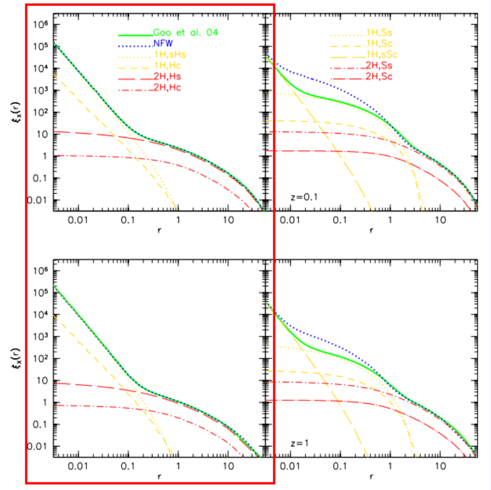
$$\xi(r) = \frac{1}{(2\pi)^3}\int P(k)\exp(i\mathbf{k}\cdot\mathbf{r})\mathrm{d}^3k$$

- For a continue field : $\xi(r) = \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle$
- Using different fields (matter, haloes, sub-haloes) one measures their cross-correlation, e.g.

$$\xi_{hm}(r) = \langle\delta_h(\mathbf{x})\delta_m(\mathbf{x}+\mathbf{r})\rangle$$

Halo-matter cross-correlation

Halo-matter, Subhalo-matter

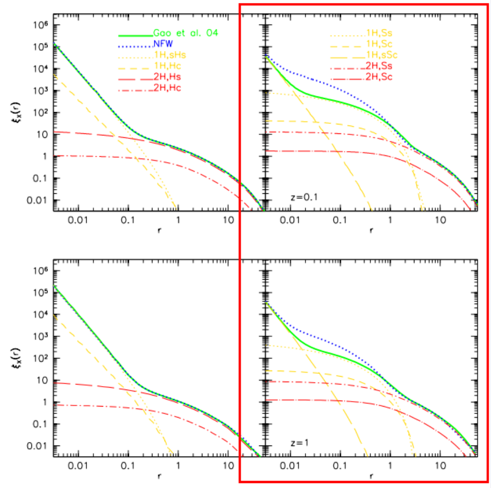


Two contributions:

- **One-halo term**: measure the internal density profile of a halo
- **Two-halo term**: measure the average matter profile outside the virial radius $\xi_{hm;2H}(r) \propto b\xi_{mm}(r)$

Subhalo-matter cross-correlation

Halo-matter, Subhalo-matter

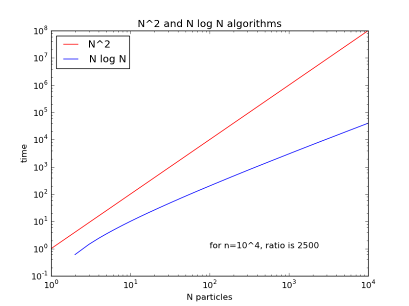


Two contributions:

- **One-halo term**:
  - Small scales: sub-halo density profile
  - Intermediate scale: mean matter distribution around sub-haloes
- **Two-halo term**: measure the average matter profile outside the virial radius of the host halo

Nearest-neighbors problem algorithms

Algorithm comparison



- Raw approach has $t \propto n^2$ in the number of particles
- For the GIF2 it means $4.09600 \times 10^{15}/??$ seconds on SP7: ??? years
- We need a fast and efficient algorithm
- We choose to build a Python tree code

Python

Advantages

- Flexibility
- Fast developing
- Math libraries
- Portability
- Parallelization

Disadvantages

- Not compiled
- Not so fast

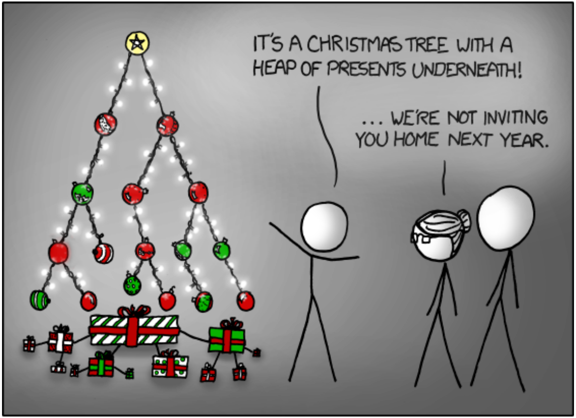
But:

- Optimization
- Use it as a glue
- With fast C/C++/Fortran core

Tree

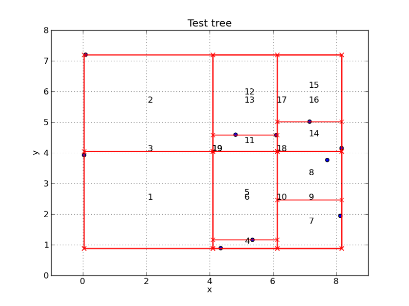
A *kd*-tree is

a hierarchical data structure in which each node represents a subset of the dataset.



Tree

KD-Tree



Properties

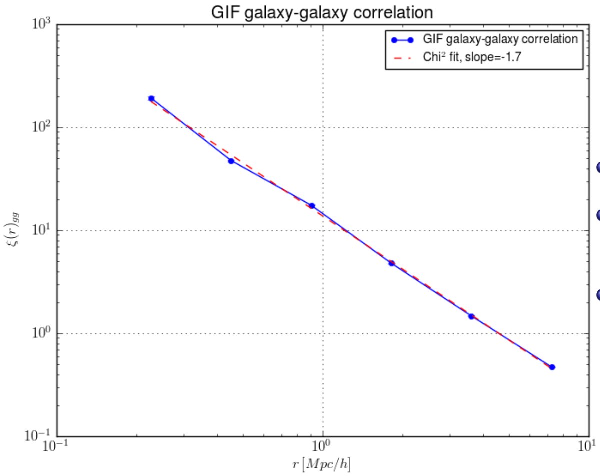
- Binary: each node has two subnodes
- Best *nearest neighbours problem* data structure
- Balanced when building it

Tree

Tree optimization

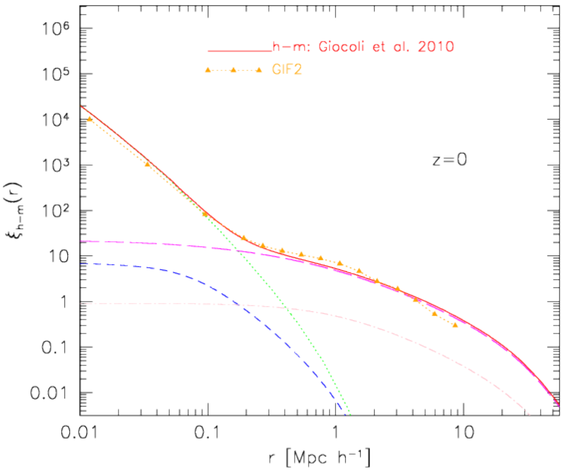
- Dual tree search
- Exclusion pruning (r_{min}, r_{max})
- Inclusion pruning
- Non redundant search
- Multiple radii
- Cached statistic
- Leaf opening strategies
- Leafsize optimization
- Possible persistence with smart node caching (PyTables)
- Time $\propto N \log(N)$ vs N^2

GIF galaxies-galaxies test, + grande l'immagine



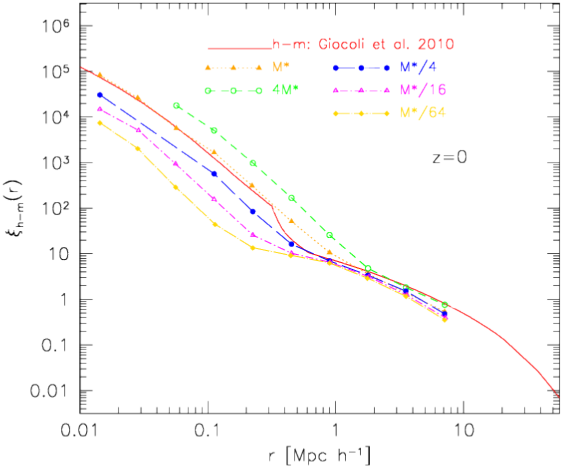
- First test of the code
- χ^2 slope of -1.7 vs -1.8 in literature
- Compare with Diaferio *et al.* 1999

GIF2 halo-matter



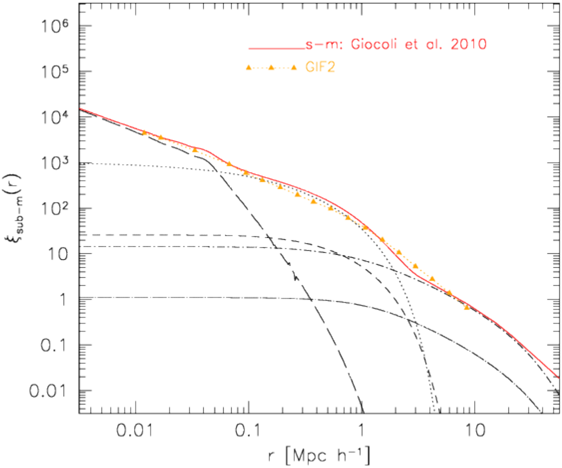
- Second, winning, test
- Small scales: density halo profile (NFW fit possible)
- Intermediate scales: transition between 1H and 2H terms
- Large scales: $\xi_{hm}(r) = b\xi_{mm}(r)$
- Compare with Giocoli *et al.* 2010

GIF2 halo-matter: different mass bins



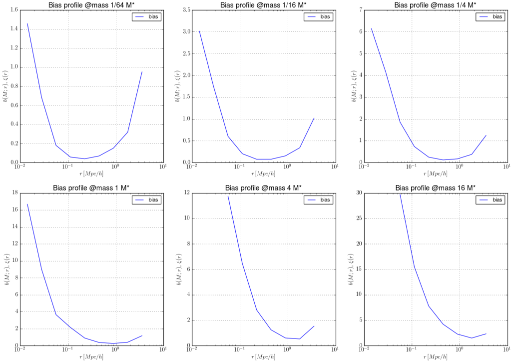
- Bin M/M^*; $M^* = 8.9 \times 10^{12} M_\odot$
- Different signals from haloes of different mass
- Very good agreement with theoretical predictions (for M^*) (Giocoli *et al.* 2010)
- Poor sampling for higher masses
- Intermediate scales feature: theoretical semplification
- Large scales disagreement: under work

GIF2 subhalo-matter



- 3 scales:
  - sub-haloes averaged profile
  - matter around sub-haloes
  - subhaloes-matter bias
- Compare with Giocoli *et al.* 2010

Bias



- $\xi_{hm}(r)/\xi_{mm}(r) \propto b$
- High resolution shows halo profiles
- Small scales: density profile
- Large scales: linear bias
- Larger scales (future work): convergence

Conclusions

Conclusions

- The original goal was on Millennium II
- Not every task is done but the code works very well
- The code is a embarrassing parallel Python tree code
- Tests were done on intermediate resolution simulations
- Very good agreement with literature and theoretical models
- Work is in progress to extend the measurement to the highest resolution Millennium II simulation
- da quanto è su la mill2???
- Problems: this work has been slowed down by a number of problems:
  - Code choices
  - Data retrieve
  - Data reading (Bad formats, few informations)
  - Hardware problems (failures and power off)
  - HW/SW compatibility

Further works

- Further refinement and optimization of the code
- More cached statistic
- Next generation: CPUs+GPUs?
- Modified persistence