JD08 · Sydney, July 2003

Theoretical input to a Global Virtual Observatory

Simon D.M. White
Max Planck Institute for Astrophysics

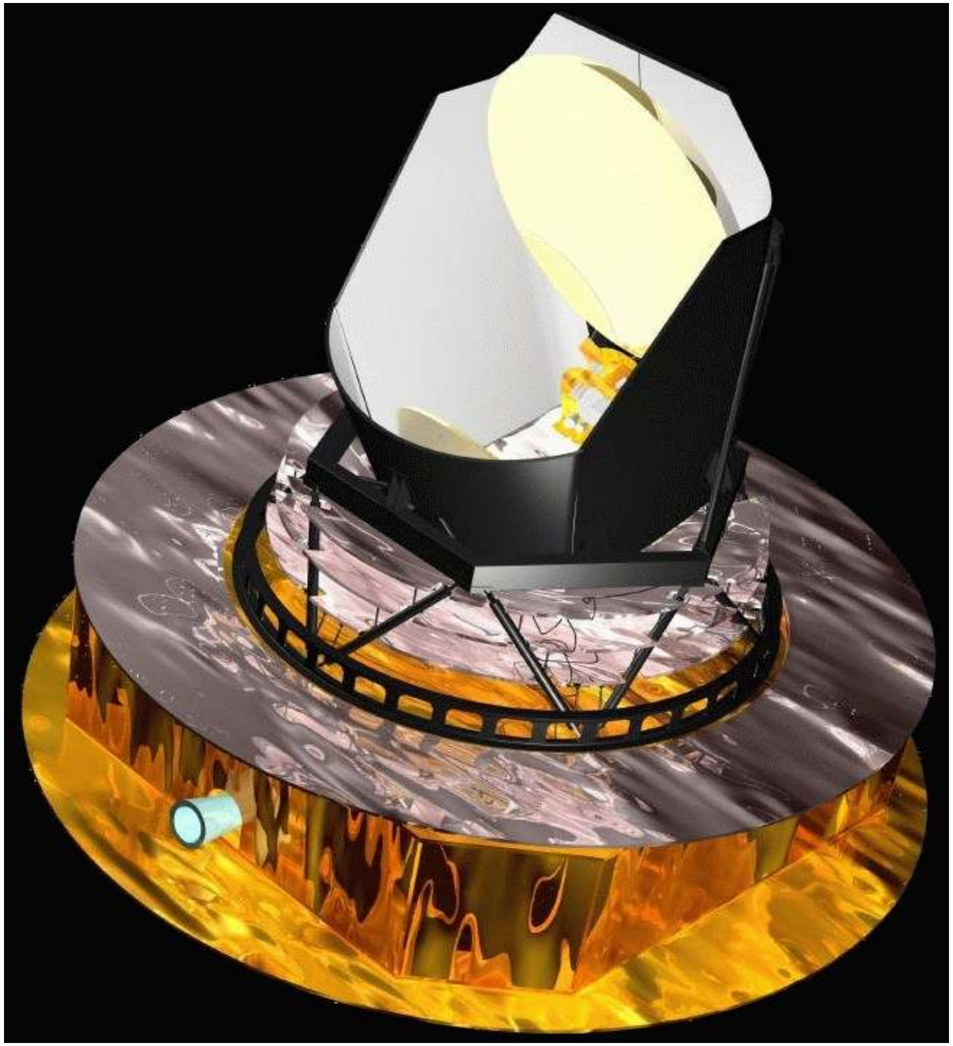
Functions of a Global Virtual Observatory

- Coordinated and transparent access
  - -- to data (MAST, CDS, Virgo)
  - -- to knowledge (astro-ph, ADS)
  - -- to tools (IRAF, Gissel, Cactus,...)
  - -- to computing resources (hardware, CPU cycles)
  - -- to expertise (web-tutorials)
- Communication
  - -- bandwidth
  - -- language (freedom from jargon, ease of use)
- Empowerment
  - -- capabilities independent of site and local resources
- Outreach
  - -- interfaces for non-professionals and non-academics

MPA Projects illustrating VO functionality

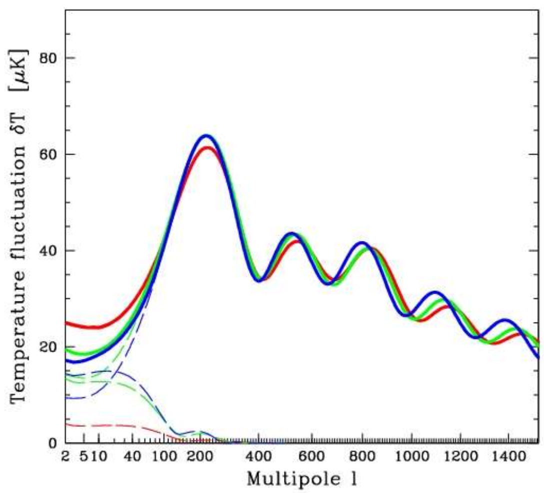
- Mission simulation pipeline for the Planck satellite
- Public archive for the Virgo Supercomputer Consortium
- Theoretical work within the German Astrophysical Virtual Observatory GAVO (see also poster 1433)
- GADGET distribution site
- Cosmic Cinema project

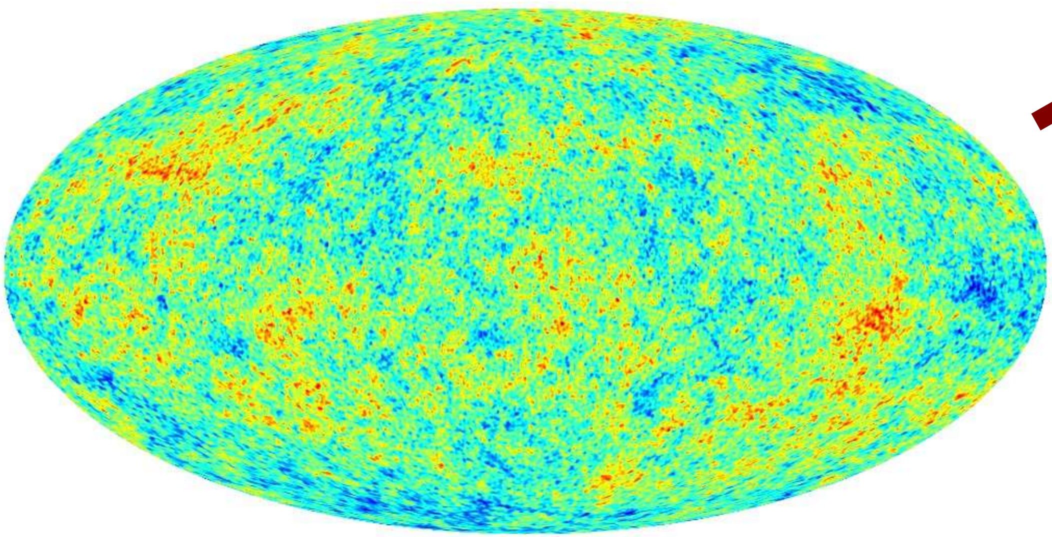
ESA's Planck Satellite

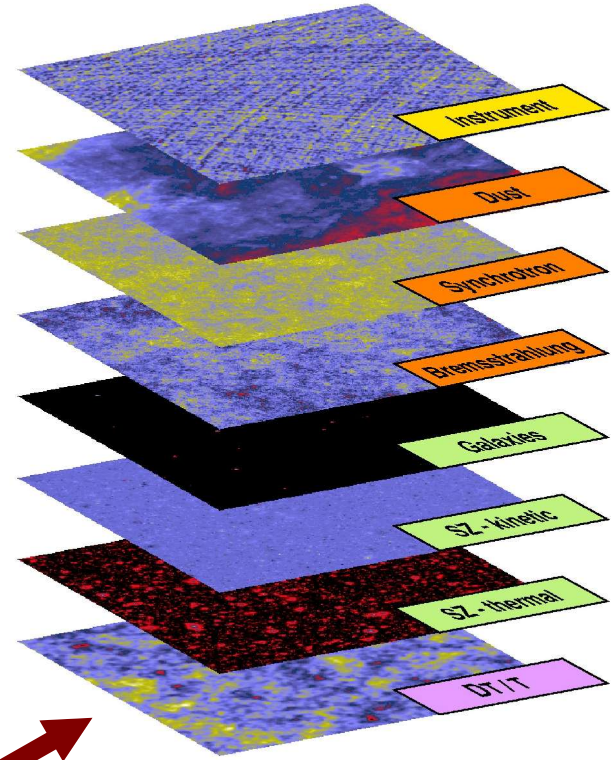


- Mission to map the Microwave background at 9 frequencies
- Launch in 2007
- Two broad instrument consortia and a telescope consortium
- Major contributions from ten European countries, USA and Canada
- Planned integrated analysis of all data in distributed Data Centres

Planck mission simulations







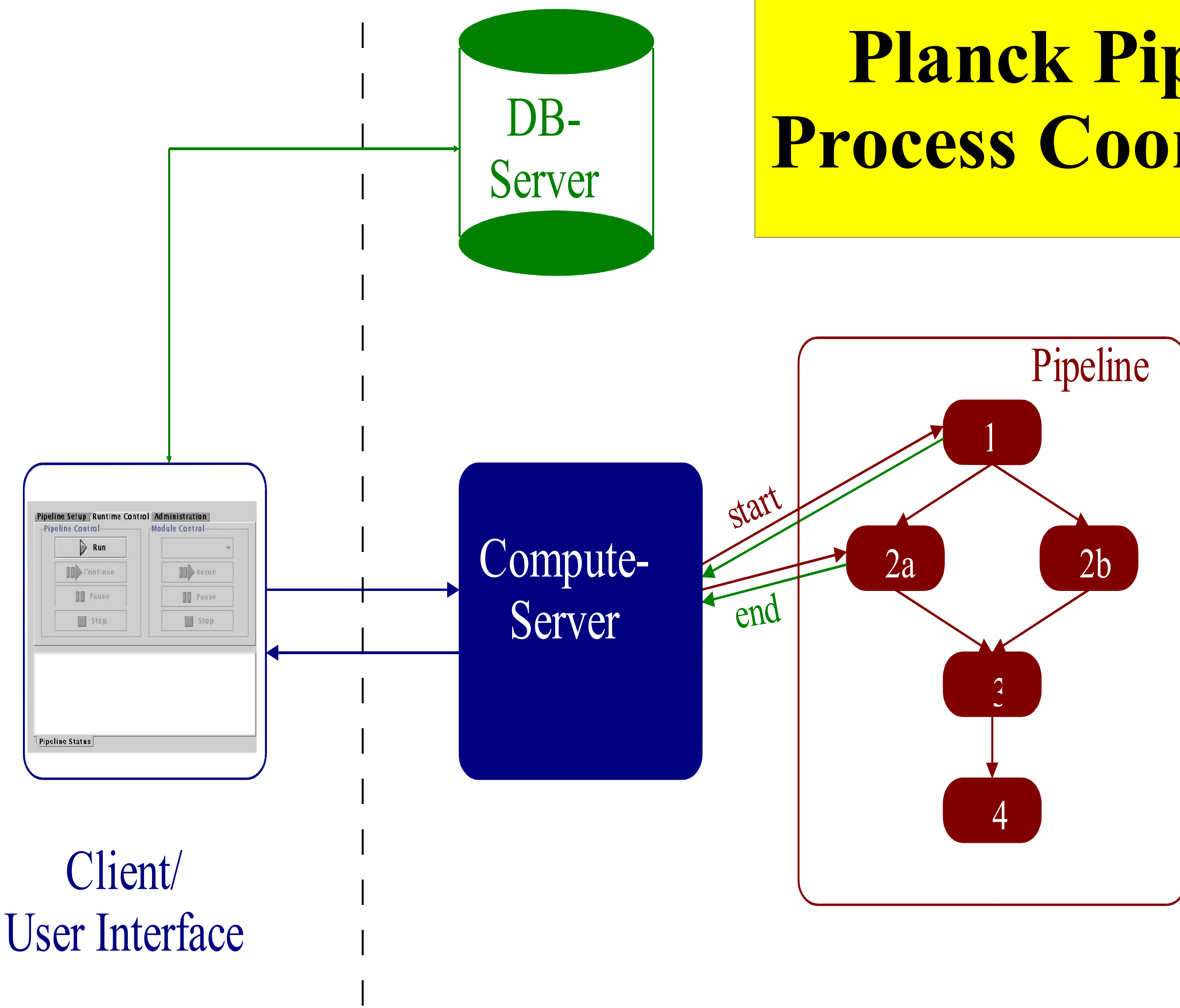
Time-line for each detector's signal

Features of simulation pipeline

- Programme modules user supplied in diverse languages
- Structure must be continually modified/updated
- Simulation process is distributed (non-simultaneous)
- Modular simulation products suitable for diverse groups

- Flexible but standard interfaces
- Object-oriented data model
- Distributed process coordinator

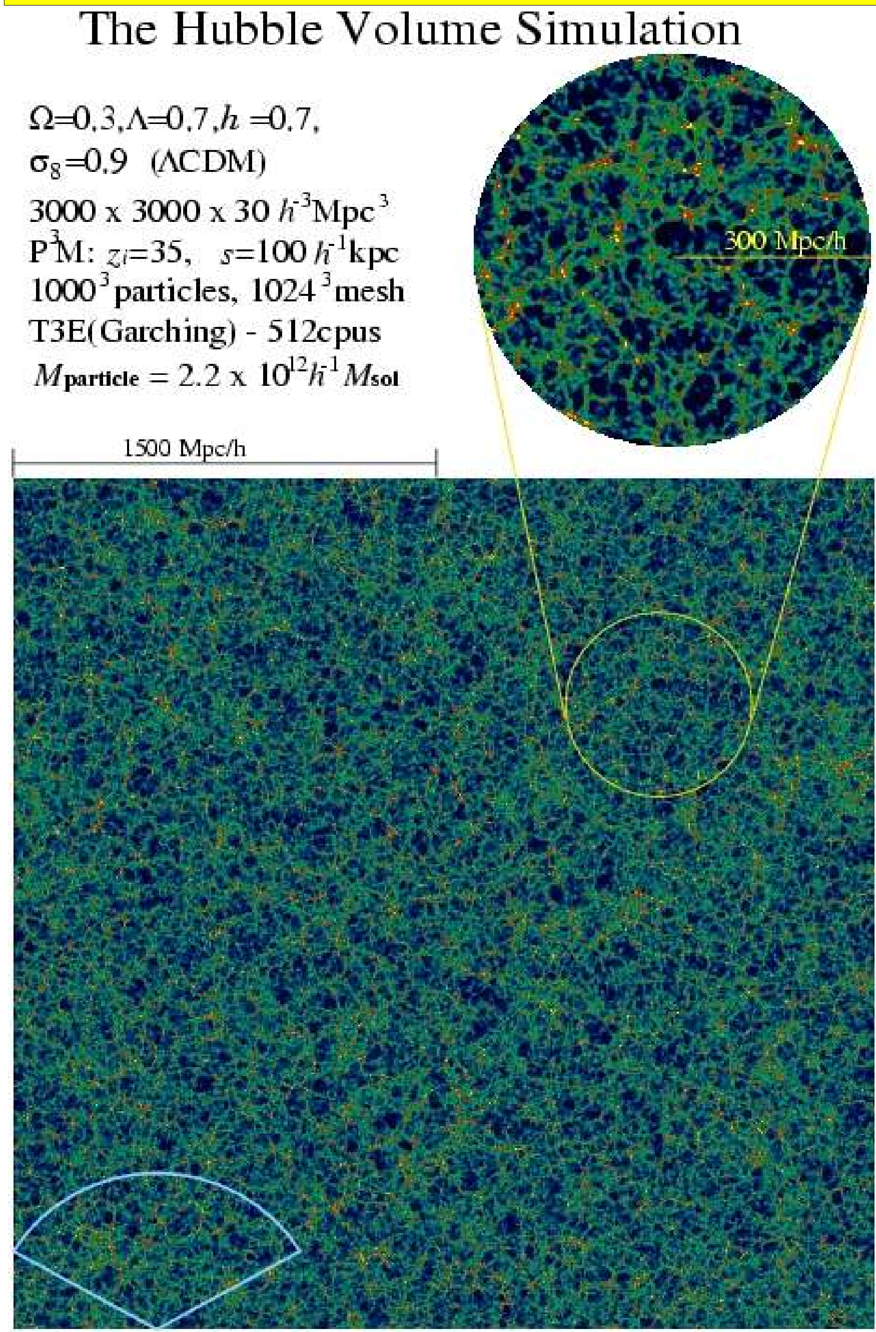
Planck Pipeline Process Coordinator

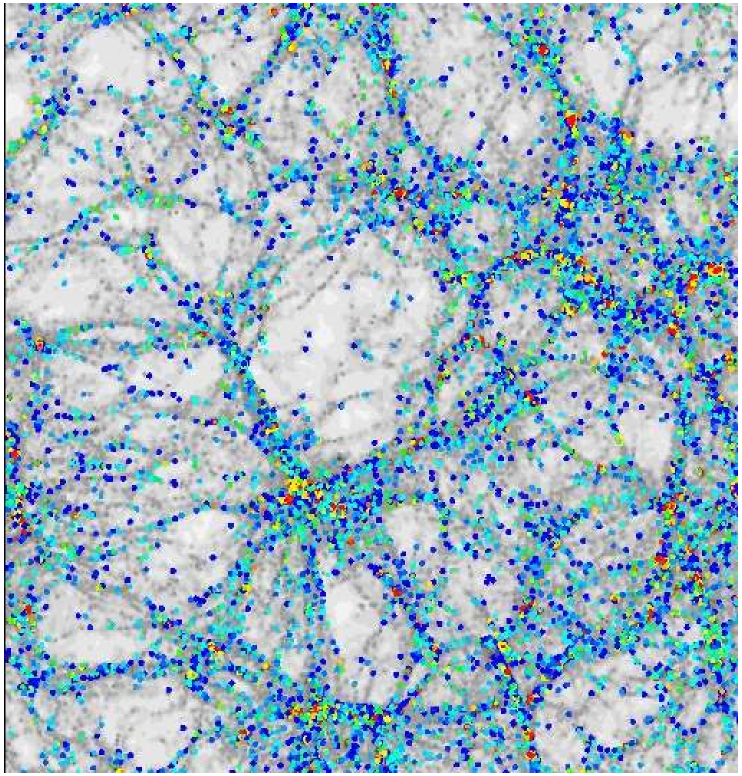


Simulating for a large distributed consortium

- Capability must match need
  - -- flexible, modular approach
  - -- users must be able to contribute elements
  - -- interfaces must work across a variety of platforms
  - -- CPU/memory/bandwidth resources must be sufficient
- Planning is needed to finish big tasks on time
- Documentation and communication are critical
  - -- well thought-out requirement/design documents
  - -- well defined development cycles
  - -- well defined interfaces

The Virgo Supercomputing Consortium

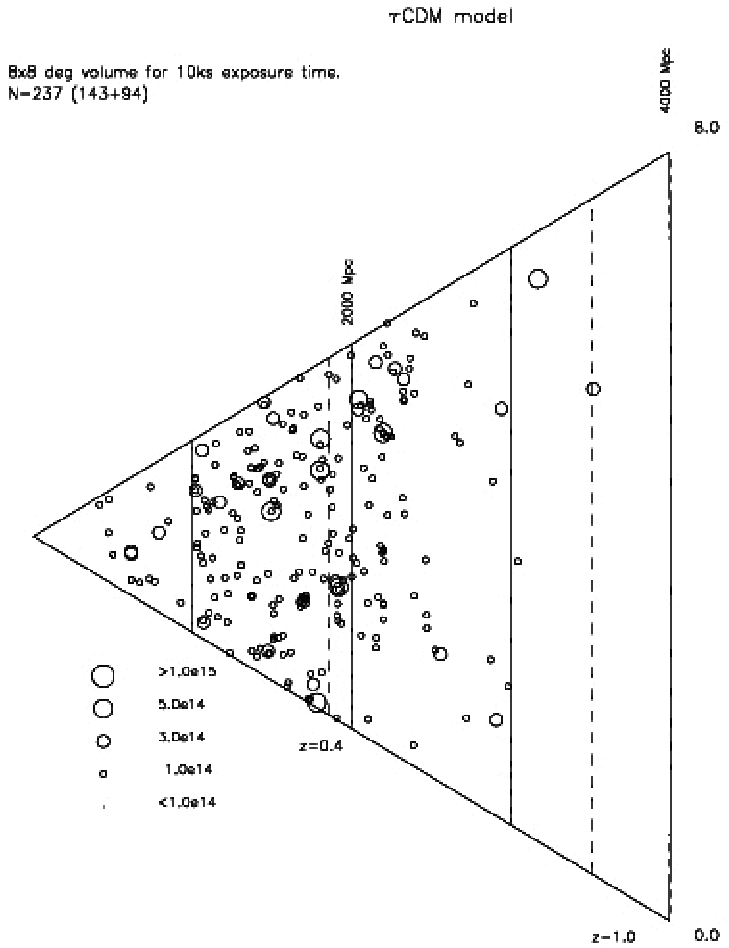


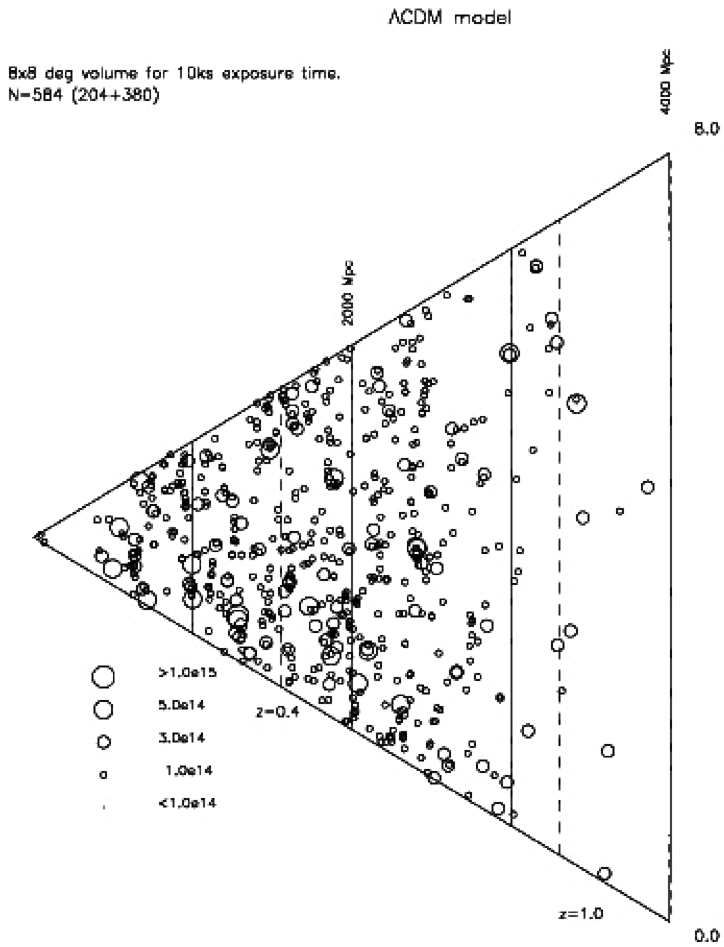


- Large consortium (UK, Germany, US, Canada)
- Aims to carry out largest feasible simulations of LSS/galaxy formation
- Goal to make simulation data public
- Archive/website at MPA

HV data use for a large XMM proposal

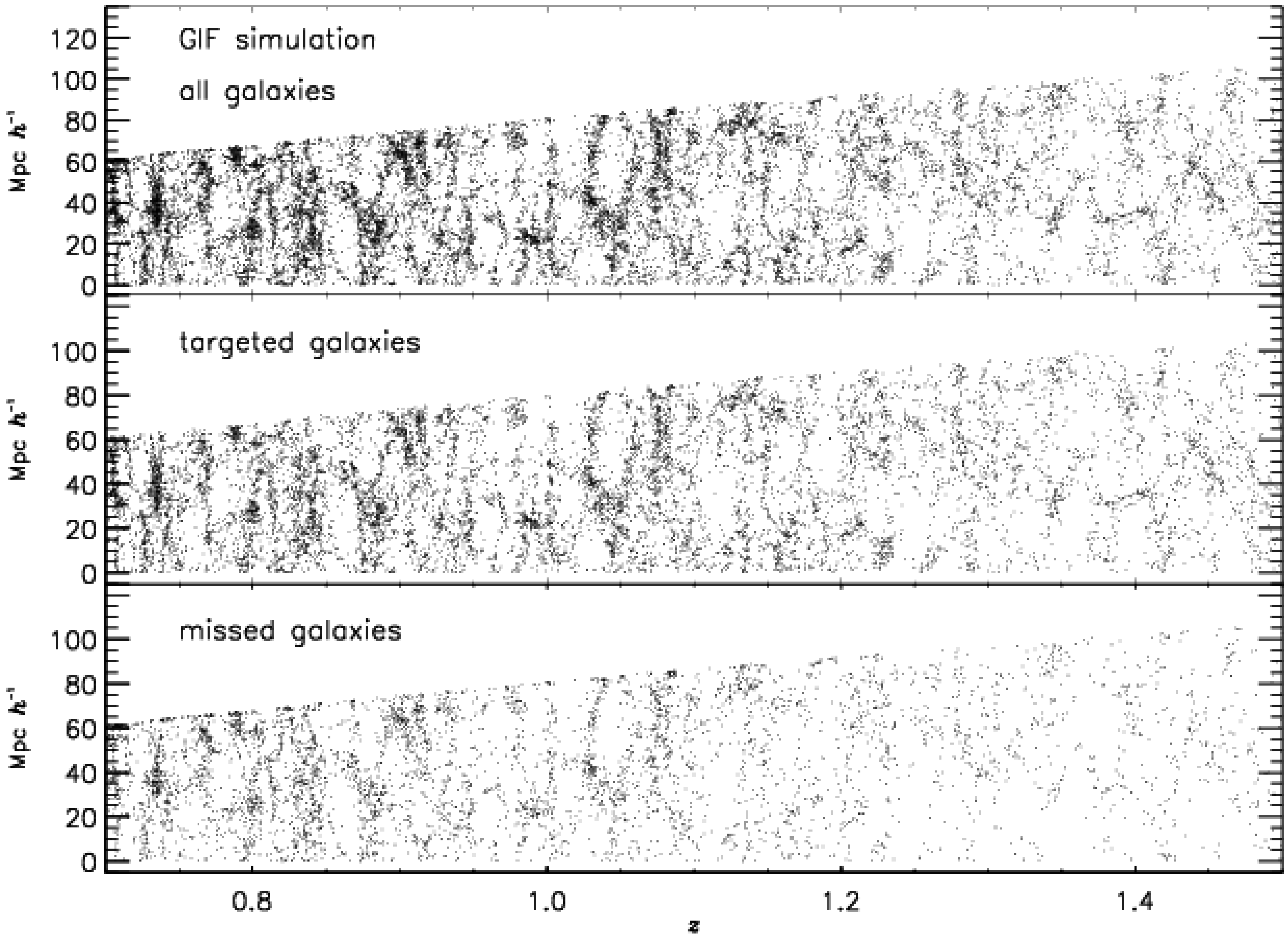
Pierre et al 2000





GIF data use for planning Keck/Deep surveys

Coil, Davis & Szapudi PASP (2001)



Evaluating infall models onto Perseus-Pisces

Hanski,Theureau, Ekholm & Teerikorpi A&A (2000)

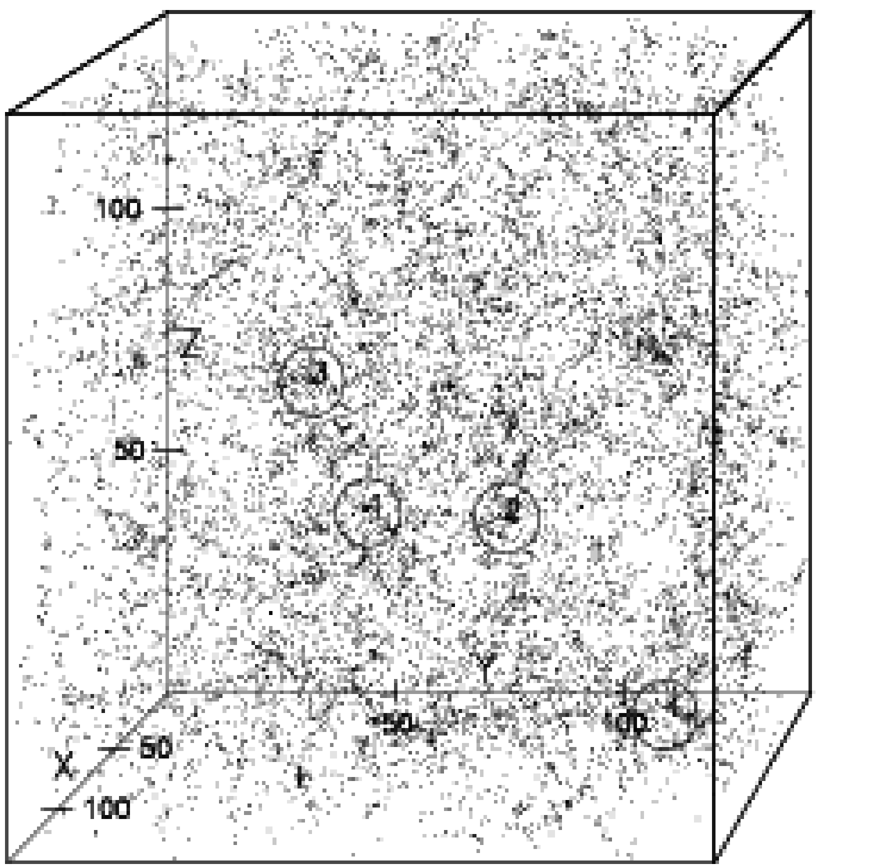


Fig. 5. The (141 h^{-1} Mpc)^3 cube of GIF N-body simulation galaxies. The simulation started with 256^3 dark halos, producing the 15000 galaxies seen in this figure. The four circles indicate the clusters for which we studied the applicability of the TB calculations.

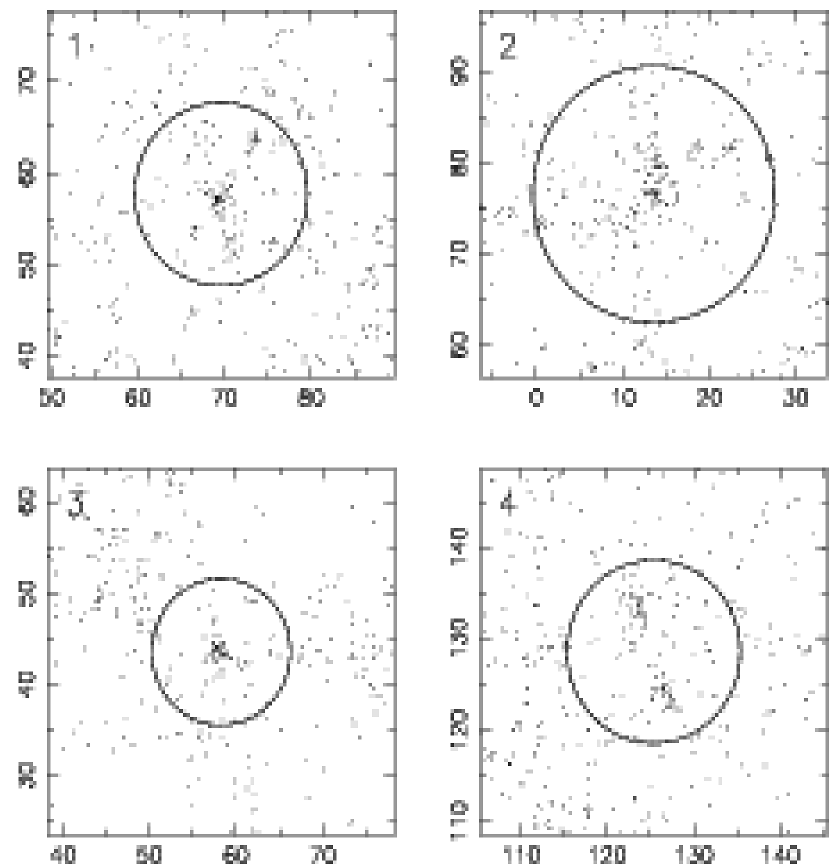


Fig. 6. The four GIF galaxy clusters extracted for further study. Figures show XY-planes from the previous figure, centered to the cluster centers. The width in Z direction of each plot is 30 Mpc. Circles with radii R_0 are drawn. It is notable that these complexes are not spherically symmetric, resembling the situation with PP.

Theoretical modelling of galaxy-galaxy lensing

Guzik & Seljak MNRAS (2001)

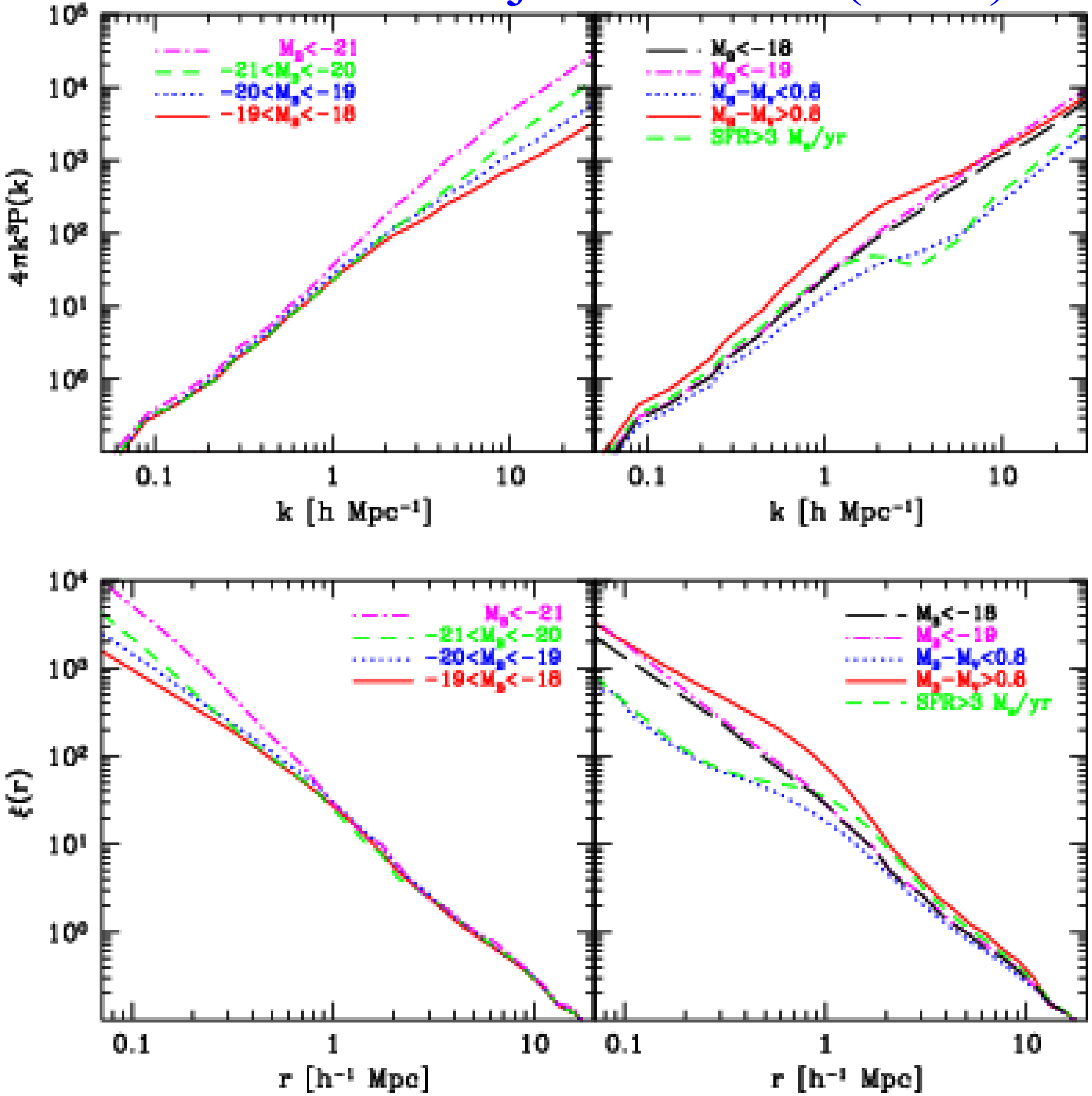


Figure 2. Galaxy-dark matter cross-power spectra and respective cross-correlation functions for luminosity bands (left panel) and colour and star formation rate (right panel). In the lower panels the cross-correlation functions are presented for the same samples as in the upper panel.

Galaxy halo masses from satellite motions

McKay + SDSS collaboration ApJ (2002)

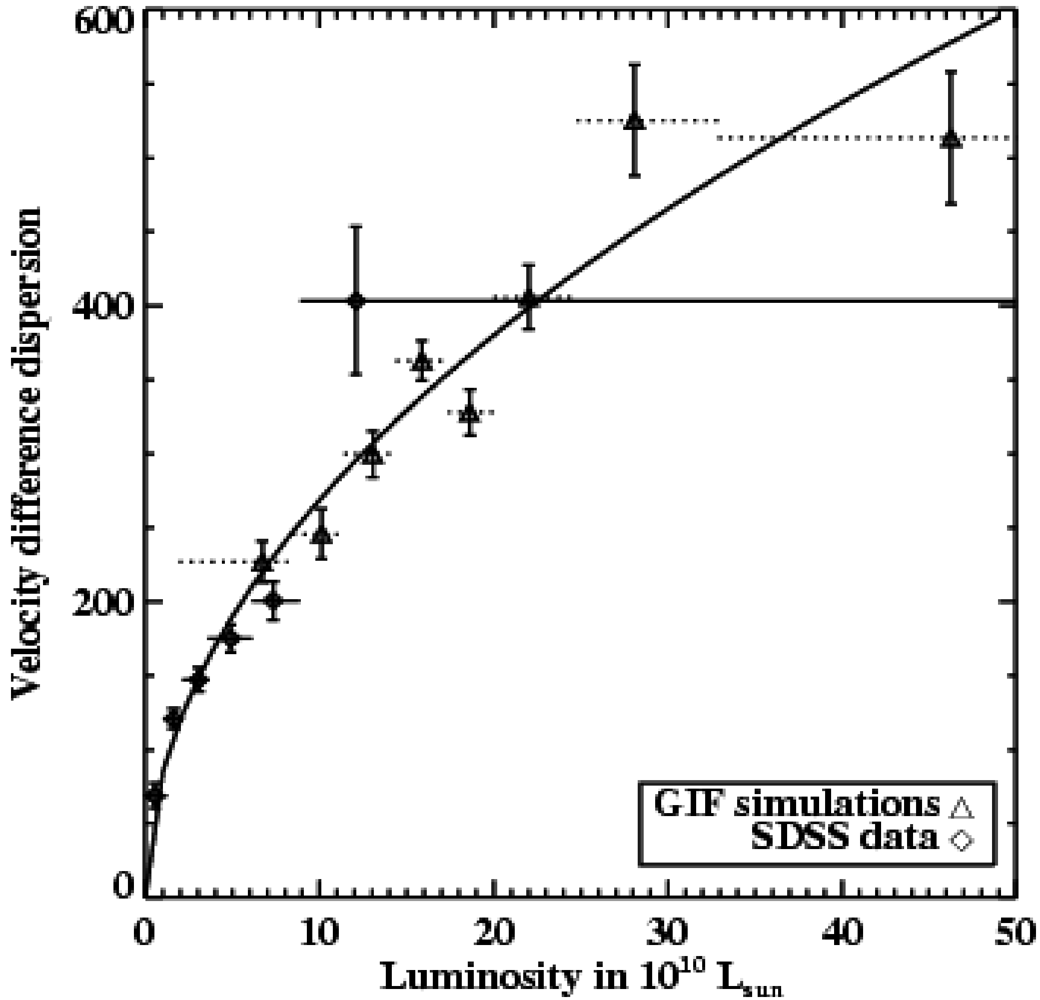


FIG. 2.— This figure compares measurements of satellite velocity dispersion vs. luminosity for real SDSS satellites (diamonds) and simulated GIF satellites (boxes). The two measures are well represented by a single relationship $\sigma_r \propto L^{0.5}$. Vertical error bars represent the uncertainties in determination of the satellite velocity dispersion. Horizontal error bars represent the range of host luminosities for each bin. The results from GIF simulations allow us to test the applicability of our mass estimator. The similarity of the simulations and observations at this observable level give some confidence in the comparison.

List of the available data

The following data except the cluster catalogues are very large in size and therefore it is practically impossible to download them via internet or ftp. By selecting data types and clicking "GO" you will get informations on the data such as number of pieces, size, magnetic tape format and also a detailed description of the data will be shown. Please read the description carefully and send the request as instructed.

Snapshot data

☐ tCDM at z=0.0 (16 GB)	☐ LCDM at z=0.0 (16 GB)

Lightcone outputs : DESCRIPTION(HTML/txt)

(* = not yet available)

☐ tCDM narrow wedge (0.5 GB)	☐ LCDM deep wedge (1.4 GB)
☐ tCDM octant A (6.3 GB)	☐ LCDM octant A (17.1 GB)
☐ tCDM octant B (6.3 GB)	☐ LCDM octant B (6.3 GB)
☐ tCDM sphere A (6.3 GB)	☐ LCDM sphere A (9.2 GB)
☐ tCDM sphere B (6.0 GB)	☐ LCDM sphere B (6.2 GB)
☐ tCDM planar map (*)	☐ LCDM planar map (*)

GO RESET

Cluster catalogues (downloadable)

- **Snapshot :** By FOF algorithm with linking parameter b=0.2 for tCDM and 0.164 for LCDM, ref 2.

tCDM snapshot z=0.0 (36 MB)	LCDM snapshot z=0.0 (42 MB)
tCDM snapshot z=0.5	LCDM snapshot z=0.5

- **Lightcone :** SO algorithm with density threshold 200, ref 1. See also DESCRIPTION(HTML/txt)

tCDM narrow (160 KB)	LCDM deep (430 KB)
tCDM octant A (10 MB)	LCDM octant A (22 MB)
tCDM octant B (10 MB)	LCDM octant B (11 MB)
tCDM sphere A (29 MB)	LCDM sphere A (40 MB)
tCDM sphere B (28 MB)	LCDM sphere B (40 MB)

Publically releasing simulation data

- Response is strong
  -- e.g. ~35 papers (published or astro-ph) using GIF simulations since Kauffmann et al (1999)
- Response is diverse
  -- >60% of GIF papers by authors unrelated to simulators
  -- theoretical, observational and "proposal" uses
  -- wide geographical distribution (Germany, France, UK, Finland, US, Canada, Israel, China, Japan...)
- Web download is preferred
- Support is needed
- Usage is difficult to monitor/control

German Astrophysical Virtual Observatory (GAVO)

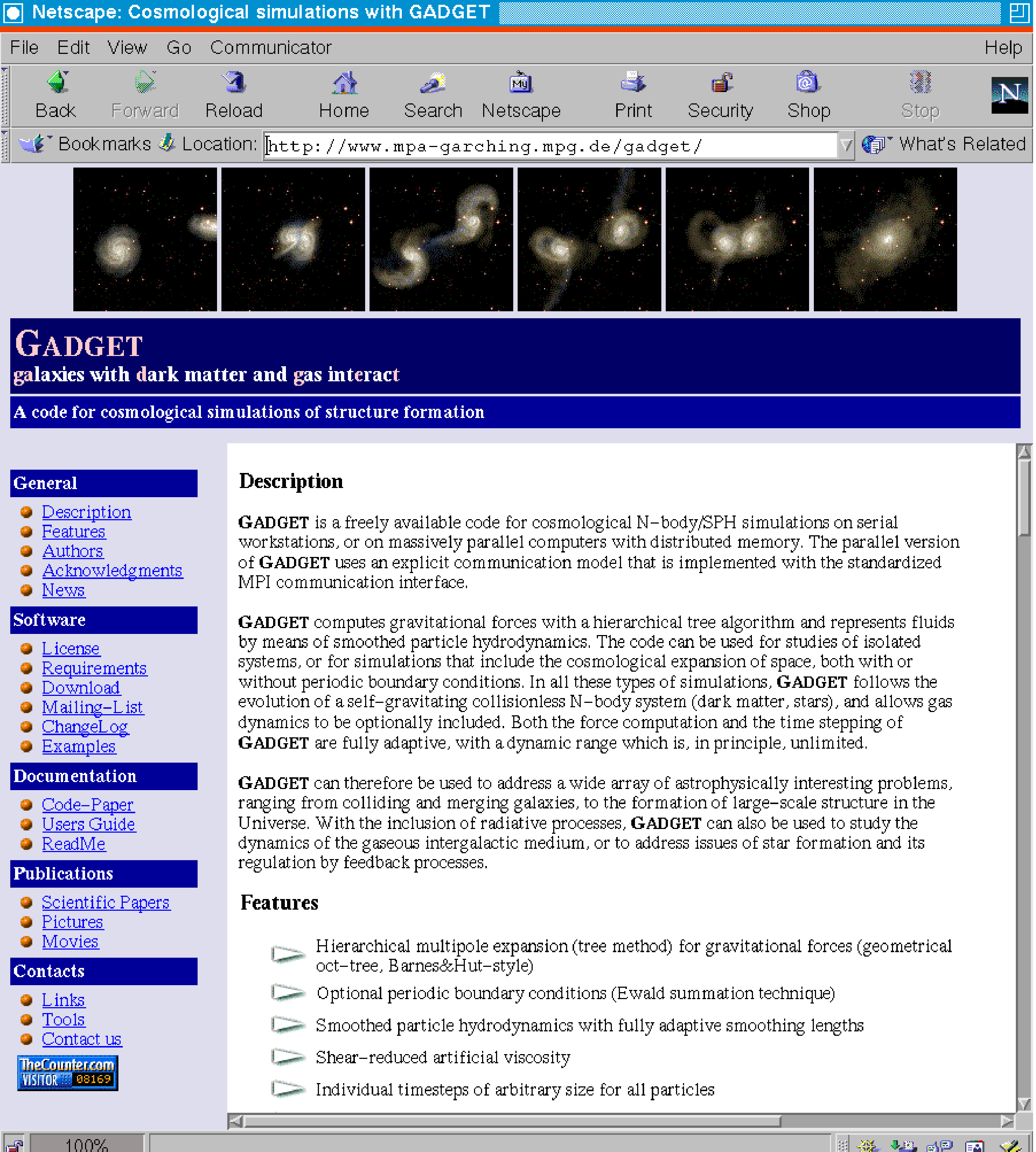
- **Collaboration of German institutes**:
  - Max-Planck-Institut für extraterrestische Physik, Garching
  - Astrophysikalisches Institut Potsdam
  - Max-Planck-Institut für Astrophysik, Garching
  - Hamburger Sternwarte
- **Goals:**
  - Publishing/federation of local catalogs (ROSAT, VIRGO, SDSS)
  - Develop advanced query and analysis algorithms (catalogue matcher, cluster extractor)
  - GRID solutions for federated database queries
  - GRID-enabled simulator
  - Develop the "Theoretical Virtual Observatory" (TVO)

Goals for Theory within the Virtual Observatory

- Public release of major simulation data
  - Products: raw data (position-velocity snapshots), halos, merger trees, mock galaxy catalogues, images, movies...
- Bridge theory/observation gap
  - Use compatible standards and interfaces
  - Develop common tools (e.g. Visualisation, joint queries)
- Develop standards for theoretical archives
  - Massive simulations need proper archiving with similar requirements to large observational datasets (history · metadata)
- Promote access to state of the art analysis algorithms
  - Online versions for friends-of-friends, N-point correlation functions, semi-analytical galaxy formation, telescope simulators
- Allow simulations on demand using GRID technology
- Allow a more detailed refereeing by clarifying the provenance of results.

Work Plan

- create a (*meta-)data model* describing simulation products. This model must be compatible with the IVOA data models for observations so that observational and simulation archives can be jointly queried
- define *protocols* for registering and maintaining metadata information about products available in simulation data centers.
- create *services* implementing these protocols for (simulation) data providers to register their products with GAVO and compatible VOs.
- create *tools/adapters/plug-ins* in standard computing languages for use by data centers/simulators, to ease publication/registration tasks.
- implement IVOA compatible *browsing and query services* on top of the simulation archives for astronomers to locate and retrieve desired products.
- expose advanced *data analysis algorithms* written by experts as standard, VO compatible (web) services, using the Globus GRID toolkit for optimal scheduling, security, accounting etc.
- use *GRID* techniques to allow users to bring their own query and analysis algorithms to where the data is stored.
- create the GRID-based GAVO Simulator allowing users to perform simulations on the fly, using observations as input for example to reproduce an observed galaxy merger.
- make all services available through a *web browser interface* (www.g-vo.org) as well as programmatically through XML/SOAP based *web services*.
- create a *prototype* using simulations available at MPA and AIP: large scale structure N-body, cluster hydrodynamics, supernovae, stellar evolution

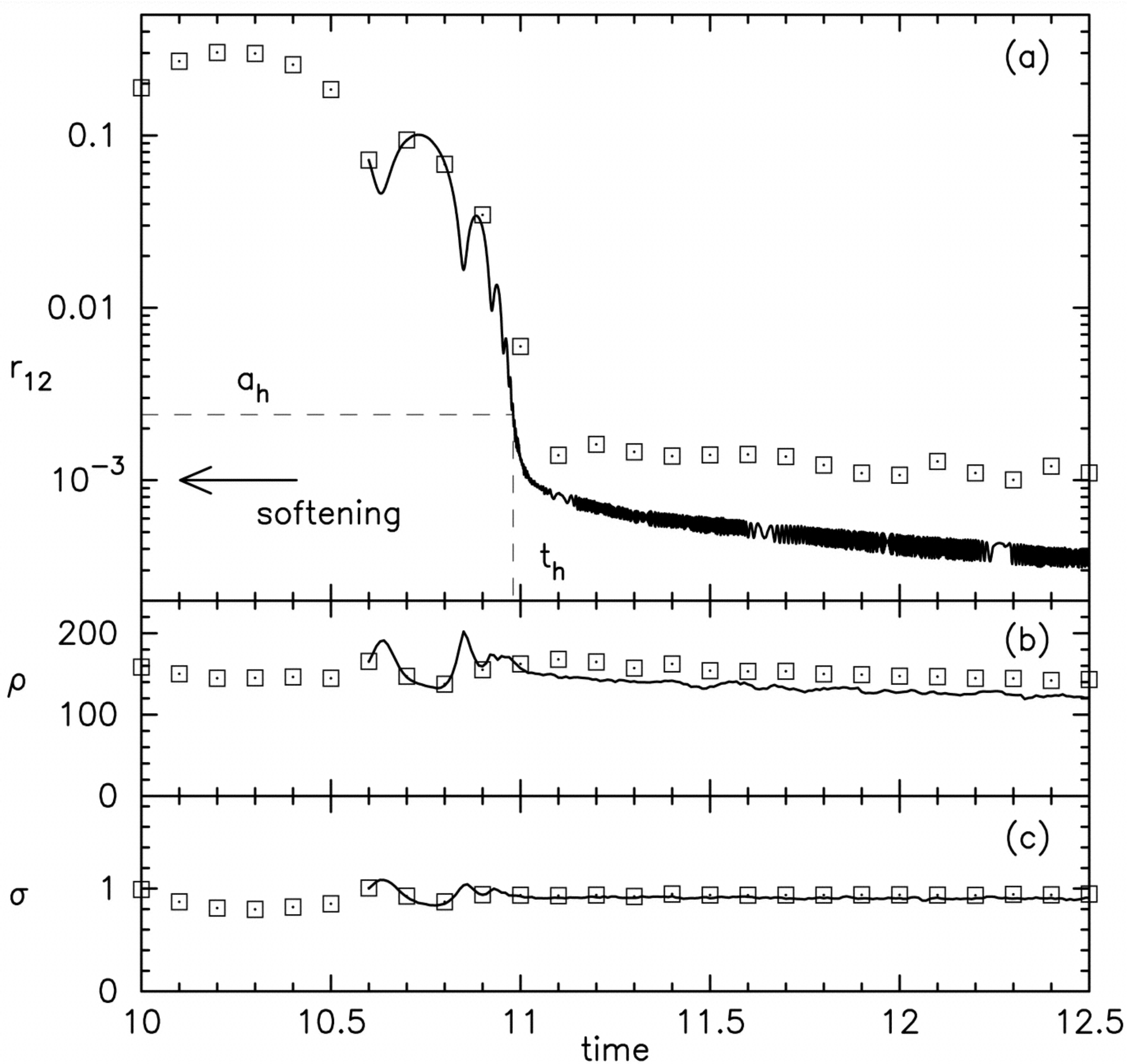


GADGET

- Public code for serial or massively parallel N-body/SPH simulation of galaxies and galaxy formation
- Single author
- Open source
- Web-downloadable
- Documentation and example problems

Merging of galaxies with central black holes

Milosavljevic & Merritt ApJ (2001)



Publically releasing simulation code

- Demand is high
- Continual support is needed
- Some comments lead to improvements, most not
- Impact of coding effort is substantially increased

MPA's Cosmic Cinema

- Web-based audio-visual presentation for the Deutsche Museum
- Bilingual
- Introduces research topics of the MPA to a lay audience
- Runs permanently and interactively on a PC
- Available also on cD and over the Net

Presenting theoretical work to a broad public

- Best combined with observational, instrumental and personal elements
- Visualisation through simulation movies or diagram animation combines well with non-technical narration and with music
- Presentations can be addressed to a broad range of audiences with relatively little modification
  - -- school and university students
  - -- senior or "club" gatherings
  - -- political levels of funding agencies
  - -- web browsers
  - -- museum visitors
- Helps establish the cultural importance of our science