Proposal for Computing Time on the Supercomputer JUGENE

New project proposal

Period: 05.2011-05.2013

Magneticum Pathfinder II, towards the next generation of cosmological simulations

Principial investigator

Klaus Dolag [KD]
Sternwarte, Ludwig-Maximilians-Universität, Scheinerstraße 1, D-81679 München, Germany

Project contributors

Joseph Mohr [JM]
Sternwarte, Ludwig-Maximilians-Universität, Scheinerstraße 1, D-81679 München, Germany

Andreas Burkert [AB]
Sternwarte, Ludwig-Maximilians-Universität, Scheinerstraße 1, D-81679 München, Germany

Stefano Borgani [SB]
Astronomy Unit, University of Trieste, via Tiepolo 11, I-34143 Trieste, Italy

Project collaborators

Alexandro Saro [AS]
Sternwarte, Ludwig-Maximilians-Universität, München, Germany

Hans Böhringer [HB]
Max-Planck für Extraterestische Physik, Garching, Germany

Giuseppe Murante [GM]
INAF - Osservatorio di Torino, Torino, Italy

Matteo Viel [MV], **Luca Tornatore** [LT], **Gabriella DeLucia** [GL]
Astronomy Unit, University of Trieste, Trieste, Italy

Massimo Meneghetti [MM]
INAF-Astronomical Observatory of Bologna, Bologna, Italy

Contents

1. Project Description

1.1. Introduction

It is now well accepted that the observed structure of our universe is best reproduced in the presence of cold dark matter and dark energy, within the framework of LCDM cosmology, in which structures form in a hierarchical bottom up fashion. The increased size, range and completeness of observational data obtained using the latest generation of astronomical instruments recently opened the so-called era of precision cosmology, meaning that the basic parameters describing the standard cosmological model can be in principle determined with a precision of ten per cent or better. Therefore, we have now entered the period where we need to understand the formation of structures in the universe with high precision, e.g. investigating large volumes by following in detail the internal structures of observable tracers of the underlying matter distribution, like galaxies and galaxy clusters. In our hierarchical picture of structure formation, small objects collapse first and then merge in a complex manner to form larger and larger structures. To a first approximation one can study the formation of cosmic structures using N-body simulations, which basically follow the evolution of collisionless particles under gravity. Such simulations have been performed with high resolution for individual objects, like galaxies and galaxy clusters as well as for very large-scale structures. However, with the possible exception of gravitational lensing, observations mainly reflect the state of the ordinary (baryonic) matter. Therefore, their interpretation in the framework of cosmic evolution requires that we understand the complex, non-gravitational, physical processes, which determine evolution of the cosmic baryons. The evolution of each of the underlying building blocks – where the baryons fall into the potential well of the underlying dark matter distribution, cool, and finally condense to form stars – within the hierarchical formation scenario will contribute to the state and composition of the inter-galactic and intra-cluster media (IGM and ICM, respectively), and are responsible for energy and metal feedback, magnetic fields, and high-energy particles. Depending on their origin, these components will be blown out by jets, winds or ram pressure effects and finally mix with the surrounding IGM/ICM. Some of these effects will be naturally followed within hydrodynamic simulations (like ram pressure effects), others have to be included in simulations via effective models (like star formation and related feedback and chemical pollution by supernovae). Further components like magnetic fields and high-energy particles need additional modelling of their injection processes and evolution, and must also be self consistently coupled with the hydrodynamics. To fully exploit the potential of the upcoming large sky surveys for cosmology and for the study of the effects of Dark Energy, we need to improve the predictions of how the large-scale structure is traced by luminous matter. This requires a detailed description in simulation codes of the above-mentioned complex astrophysical processes and their impact on observational properties of cosmic structures. This project is aimed at providing an essential step forward for the overall ambitious programmes for precision cosmology.

Limitations of current cosmological simulations

Most current large-scale high-resolution cosmological simulations are based on pure gravitational physics (examples: Millennium, Springel et al. 2005; Millennium II, Boylan-Kolchin et al. 2009; Coyote Universe, Heitmann et al. 2008). These simulations are usually complemented by running the so-called Semi Analytic Models (SAMs) of galaxy formation. While SAMs provide a realistic description of the properties of galaxy populations, they provide at best indirect information on the properties of the IGM and, nor do they properly include the dynamical effects of the baryons on structure formation, which is highly relevant for the study of environmental effects. Attempts have been pursued to perform large-scale (e.g. 500 Mpc/h), hydrodynamical simulations, but they usually include only non-radiative physics (examples: The Marenostrum Universe, Gottlöber et al. 2006) or a very crude description of star formation (example: Millennium gas project, Gazzola & Pearce 2007). All such simulations additionally suffer from their poor resolution (typically 20 kpc/h). Simulations with higher spatial resolution and better treatment of cooling and star formation (examples: Borgani et al. 2004; CLEF simulation, Kay et al. 2004), can only be realized for relatively small simulation volumes (200 Mpc/h) and still are performed with moderate resolution (typically 10 kpc/h). Only relatively tiny volumes (25 Mpc/h) are so far explored at high resolution (2 kpc/h) and with moderate inclusion of physical processes (e.g., Tescari et al. 2009; OWLS, Schaye et al. 2009). However, a more complete and refined description of physical processes, larger volumes and higher resolution are needed to produce a theoretical counterpart to interpret data coming from current and forthcoming astronomical surveys and instruments.

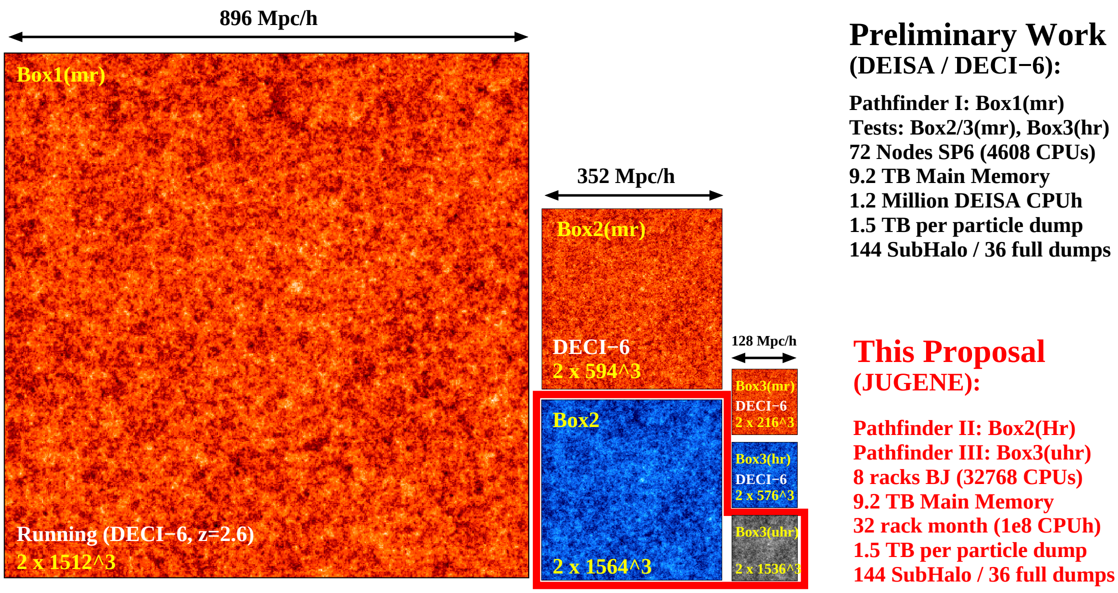

Fig. 1.— This collections illustrate the cosmological boxes (*Box1-3*) with the different resolutions (*mr*, *hr* and *uhr*) which are part of this project. Some of them where already performed using a previous, DECI-6 proposal. Note that due to the higher spatial resolution, the two simulations proposed for this project, are much more CPU demanding than the previous ones (see table 2 for the estimates of CPU time).

1.2. Preliminary work

Current state of hydrodynamical cosmological simulations

The most advanced simulations nowadays include the description of radiative cooling of the gas, a sub-resolution prescription to follow the formation and evolution of the stellar component and the release of energy and metals from Type II and Type Ia supernova and AGB stars (e.g., Tornatore et al. 2007a, Fabjan et al. 2008, Wiersma et al. 2009). A self-consistent treatment of these processes is necessary for a comprehensive description of the observational properties of the IGM/ICM and of galaxies (e.g. Saro et al 2006, Nuzza et al. 2010). Furthermore, Additional modelling of AGN feedback is required to obtain a more realistic description of the the X-ray properties of the ICM in central regions of galaxy clusters and to suppress low-redshift star formation within the most massive galaxies (e.g., Puchwein et al. 2008, Fabjan et al. 2010). Furthermore, transport processes, such as thermal conduction, also affect the structure of the ICM (e.g., Dolag et al. 2004). Including a self-consistent treatment of magnetic fields (Dolag & Stasyszyn 2009) allows one to extend comparison with observations towards radio wavelengths (Donnert et al 2010) as well as towards astronomy with Ultra-High Energy Cosmic Rays (UHECRs; Dolag et al 2005b). The underlying, hydrodynamical treatment is now advanced enough to even capture additional effects like turbulence within the ICM (Dolag et al 2005a), which should be detected by the next generation of X-ray telescopes. Finally, advanced tools for post-processing of large simulations allow one to reliably detect multi-component substructures (i.e., galaxies; Dolag et al. 2009a) as well as to distinguish different dynamical components within the formed structures (e.g. intra-cluster stars vs. stellar populations in galaxies; Murante et al. 2007, Dolag et al. 2010, Puchwein et al. 2010).

The numerical description of all such processes, and a study of their interplay, have been only realized so far in simulations of individual galaxy clusters and groups, with the purpose of describing their effect on the ICM and on other global cluster properties. However, they have never been brought together in a large-scale, cosmological simulation, that would follow not only the evolution of galaxy clusters, but also of the properties of galaxies in different environments and of the inter-galactic baryons permeating the cosmic web. To this purpose, a large volume (1 GPc/h) simulation with resolution at least comparable to that of the Millennium Run (10^9 M_o/h for the mass dark matter particles) would be needed, thus requiring 4096^3 DM particles and as many gas particles. Such a simulation is clearly out of range for current facilities, but in range for key projects on the next generation of European supercomputers.

1.3. Magneticum Pathfinder I

Therefore we have proposed (in the spirit of a pathfinder) several simulations of different cosmological boxes covering such large scales (with lower resolutions) down to much smaller scales (covering much higher resolution) allowing us to produce large area predictions for various upcoming surveys as well as studying the detailed evolution of cosmic structures with so far unreached level of detail using P-Gadget3(XXL). Several of the smaller simulations and one of the computational most demanding ones was already performed within a successfull application to the last DEISA DECI-6 call. Improvements in code performance, infrastructure and post processing, as well as further development of various physical modules which were already done to perform demanding simulations are listed in section 1.5. First highlights of the scientific results obtained from this first set of pathfinder simulations are shown in Figure 2. We want to stress that these are the first cosmological hydrodynamical simulation which start to allow studying multi wavelength ICM properties and photometric properties of the galaxy population at the same time. Therefore they build an unique theoretical counterpart to upcoming large volume and multiple wavelength astronomical surveys like Planck, SPT, Pan-STARRs, LOFAR, eROSITA, DES and many more. In combination with data sets from these surveys, our proposed simulation will enable forefront

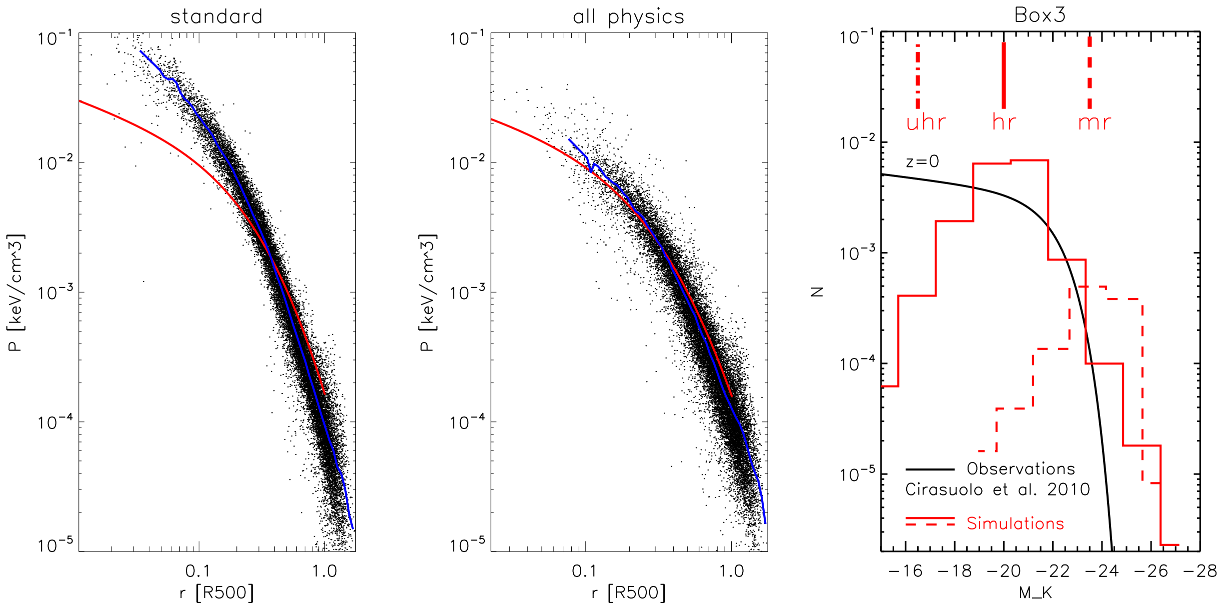

Fig. 2.— The two panels in the left illustrate results from the *Box3(mr)* simulation, comparing the obtained pressure profile (black points and blue lines) with the expected pressure profiles of galaxy clusters from a fit to observations (red line, Arnaud et al. 2008). The left of the two panels correspond to state of the art simulations, whereas the middle one reflects the results from our runs, including all the additional physics modules. The right pannels show the luminosity function of galaxies obtained directly from the hydrodynamical simulation (red) at different resolutions (*mr* and *hr*) compared to observations (black). The lines at the top mark where we expect the different resolutions to give convergent results, including the expectation for the *uhr* resolution as proposed in this proposal.

studies of the formation and evolution of galaxies and galaxy clusters as well as the study of dark energy and the origin of cosmic acceleration. A summary of all simulation setups, which in combination will allow us to obtain the maximum of scientific outcome are listed in table 1.

The PIs together with their collaborators are active members of these ongoing and upcoming observational efforts, especially in the working groups in charge to supply such experiments with theoretical predictions. The team of PIs and collaborators also cover well the expertise to run such demanding cosmological simulations and are well know for their work in computational cosmology. Successful application to computing time at world leading supercomputing facilities all over Europe allowed them to performed various, outstanding cosmological simulations during the last decade.

Simulation	BoxSize	Resolution	Softening	N_{DM}	Galaxies (hydro)	Galaxies (SAMs)	Cluster (hydro)
	[Mpc/h]	DM/gas [M_o/h]	[kpc/h]		M* [M_o/h]	M_{vir} [M_o/h]	M_{vir} [M_O/h]
Box1(mr)	896	$1.3\text{x}10^{10}$ / $2.6\text{x}10^{9}$	14	1512^3	$1.6\text{x}10^{10}$	$6\text{x}10^{11}$	$2\text{x}10^{14}$
Box2(mr)	352	$1.3\text{x}10^{10}$ / $2.6\text{x}10^{9}$	14	594^3	$1.6\text{x}10^{10}$	$6\text{x}10^{11}$	$2\text{x}10^{14}$
Box2(hr)	352	$6.9\text{x}10^{8}$ / $1.4\text{x}10^{8}$	5	1564^3	$8.6\text{x}10^{8}$	$3\text{x}10^{11}$	$1\text{x}10^{13}$
Box3(mr)	128	$1.3\text{x}10^{10}$ / $2.6\text{x}10^{9}$	14	216^3	$1.6\text{x}10^{10}$	$6\text{x}10^{11}$	$2\text{x}10^{14}$
Box3(hr)	128	$6.9\text{x}10^{8}$ / $1.4\text{x}10^{8}$	5	576^3	$8.6\text{x}10^{8}$	$3\text{x}10^{11}$	$1\text{x}10^{13}$
Box3(uhr)	128	$3.6\text{x}10^{7}$ / $7.3\text{x}10^{6}$	2	1536^3	$4.5\text{x}10^{7}$	$1.5\text{x}10^{10}$	$5\text{x}10^{11}$

Table 1: Parameters and resolution of the proposed simulations. In red marked are the two simulations foreseen for this proposal, which are by far the most computing time intensive among the list.

1.4. Project Details

1.4.1. Simulation Setup

Table 1 list the parameters foreseen for the various simulations and the mass down to which hydro-dynamical quantities of galaxy clusters (col. 8) and directly simulated galaxies (col. 6) will give convergent results as well as the mass limit for galaxies inferred from SAMs (col. 7). In combination, these simulations would:

- shed new light on and enable unique insights into many current astrophysical aspects of structure formation;
- help in the interpretation of observational results and the refinement of observational strategies for many of the current astronomical observatories;
- drive the development and exploration of new simulation and post-processing infrastructure.

1.4.2. Additinal Physics

The proposed simulations will for the first time combine the description of a number of physical processes, which so far have been developed and tested only separately. Specifically, we plan to include the following modules:

- low-viscosity SPH to allow the development of turbulence within the ICM (Dolag et al. 2005a);
- star formation and a detailed model of chemical enrichment (Tornatore et al. 2007a);
- AGN feedback (Springel et al. 2005, Fabjan et al. 2010);
- thermal conduction (Dolag et al. 2004);
- passive magnetic fields based on Euler potentials (Dolag & Stasyszyn 2009)
- non-ideal MHD effects, magnetic dissipation (Bonafede & Dolag, work in progress)

- **Magnetic fields within the large scale structures [KD]**
  These simulations will follow for the first time the evolution of magnetic fields within the large scale structures, filaments, galaxy clusters and groups simultaneously (Box1 + Box2 + Box3). They will provide a self-consistent charac-terization of the magnetic field structure within large representative volumes of the universe. This will enable statistically meaningful studies of the magnetic field amplification in cluster mergers (Box1 + Box2), the associated radio emission in the ICM (halos and relics; Box1 + Box2), and the propagation of UHECRs within the large-scale structures (Box2 + Box3).

- **Clusters as tracers of the large scale structures [JM,KD,SB]**
  These simulations will allow us to predict X-ray and Sunyaev-Zeldovich (SZ) signals from galaxy clusters with new levels of fidelity (Box2 + Box3) and for a very large set of simulated clusters (Box1). Individual clusters will be used as templates to improve extraction of cluster properties from joint X-ray, SZ and optical observations (Box1 + Box2 + Box3). Deep light-cones can be used to predict the performance of cluster finders applied to mock observations (Box1). Both will help to theoretically predict the significance with which cosmological parameters (especially dark energy parameters) can be extracted from current X-ray and SZ observations.

- **Cluster X-Ray and SZ scaling relations [SB,KD,JM]**
  Due to the large volume of Box1, X-Ray and SZ scaling relations can be extracted for a fair sample of massive clusters, which was not possible given current limitations of hydrodynamical simulations. Making use of the high resolution of Box2 + Box3, it will be possible to study the observed transition of the scaling relations towards groups and characterize especially the effect of AGN feedback, which was not taken into account in previous studies using large, cosmological volumes. Moreover, studying the biases affecting the lensing measurements will allow us to verify the degree to which lensing can be used as a powerful complement to X-ray and SZ observations to calibrate cluster masses.

- **Galaxy Formation and Evolution [AB, KD, SB]**
  The high-resolution achieved for the Box2 and Box3 simulations will allow us to investigate the formation and evolution of the different galaxy types and the history of star formation and chemical enrichment in great details. We will focus especially on the high-redshift (cosmological redshifts z=1-3) regime where new, pioneering observations e.g. with the ESO Very Large Telescopes have revealed young galaxies with structures that are very different compared with present-day galaxies. The simulations will provide insight into how these galaxies formed, what the origin of their surprising substructures might be and how high-redshift galaxies evolve into the present-day galaxy population. They will allow us to predict how metal-rich galactic winds and central supermassive black holes shape galaxies and limit their star formation and how the evolution of galaxies is linked to their intergalactic surroundings.

- **Properties of the Intergalactic Medium [SB, KD, MV]**
  The very high resolution achieved in Box3, along with the relatively large volume covered by this simulation, will allow us to study in details the evolution of the thermal and metal-enrichment properties of the IGM. For the first time, a single simulation will cover both the high-redshift regime, $z > 1.5$, where the IGM is observed in the optical band

through absorption features in the spectra of distant quasars, and the low-redshift regime ($z < 1$), where the future generation of high-sensitivity X–ray telescopes will reveal diffuse baryons in the form of the elusive Warm-Hot Intergalactic Medium (WHIM). Combined with the good statistics of clusters and groups in Box2 and Box3, our simulations will provide a unified picture of the interaction between galaxy formation and inter-galactic medium over unprecedented range of environments and cosmic times.

1.4.4. Secondary Scientific targets

These scientific targets will be evaluated within the team of PIs and the collaborators.

- Full-sky SZ maps to predict the diffuse SZ signal in the CMB power spectrum from PLANCK [KD]
- Full-sky X-Ray maps for eROSITA [KD,HB,JM]
- Appearence of massive clusters and protoclusters high redshift [HB,SB,KD,GD]
- Morphology of massive clusters and its evolution [HB,KD]
- Distribution of Luminous Red Galaxies and application to measure baryon acoustic oscillations (BAOs) [HB]
- Galaxy properties and mock light-cones from SAMs [GD]
- The origin of gas-rich, violently star forming galaxies at redshift 2-3 and their connection to present-day galaxies [AB,KD]
- The formation of elliptical galaxies [AB,KD]
- The role of supermassive black hole feedback and galactic winds in driving galaxy evolution [AB,KD]
- Properties of Lyman-alpha transmitted flux and their cosmological application [MV]
- WEAK and strong lensing signal and production of mock optical observations using SkyLens [MM]
- Detailed SZ maps of clusters and groups to study detectability with SPT [JM,KD]
- Morphology of groups [HB,KD]
- Galaxy dynamics in clusters and origin of the intra-cluster light [GM,KD,GD]
- Evolution of the galaxy population within clusters [GM,SB,KD]
- Effects of selection and ongoing formation on galaxy kinematic mass constraints on clusters of galaxies [AS,JM,KD]
- Galaxy properties in different environments [SB,GD,KD]
- Magnetic field properties of groups and forecast of visibility for LOFAR [KD]
- Galaxy properties and mock catalogs from SAMs [GD]

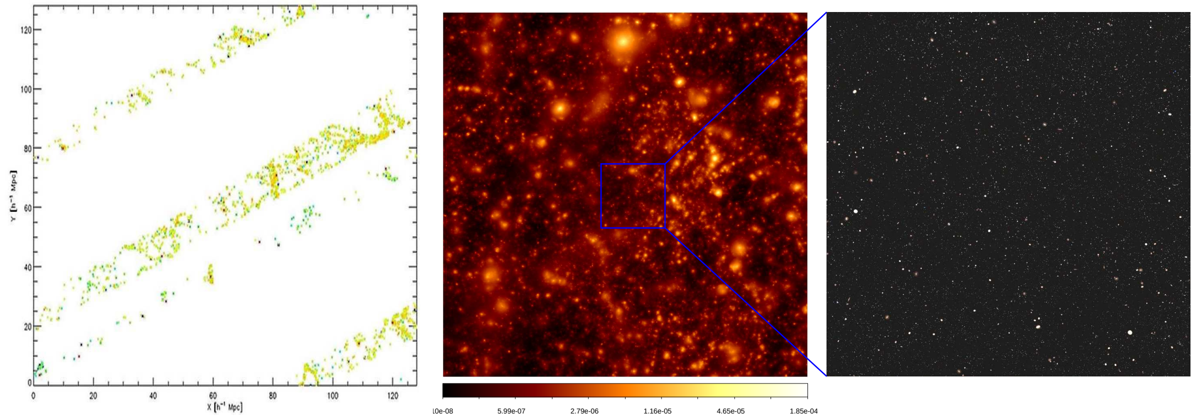

Fig. 3.— The left panel shows the galaxies (color coded by the V-K luminosities) within an abitrary lightcone through the Box3(hr) simulation. The middle panel shows the predicted SZ signal within such lightcone for Box3(mr). The right panel shows a synthetic SkyLens image for a virtual SUBARU observation of the central part.

1.5. Description of Methods and Algorithms

P-Gadget3(XXL) is a highly optimised, MPI/pthread/OpenMP parallelized N-Body TreePM MHD-SPH code. It has already been used to perform several of the worlds leading cosmological simulations (e.g. Millennium, MillenniumII, Aquarius and many more). It has been run on various platforms (IBM Power3,4,5,6 / Blue Gene, Linux based intel/amd clusters, SGI Altix, etc.). It shows a very good scaling with large numbers of CPUs, especially for cosmological boxes. From our code testing we expect less than 40% losses due to unbalanced workload effects, on the fly post processing and imperfect OpenMP paralelization, especially when using mixed MPI and OpenMP paralelization on 32768 CPUs.

However, to perform such world leading simulations, especially with such a large amount of particle data and with such rich amount of additional physics, further code optimizations are necessary. For performing the first set of pathfinder simulations we already modified and optimized the code in various places. This includes a complete re-organization of the internal data structure for the extra physics, which allows us already to save 50% of memory for the particle data. It also includes several optimizations developed in co-operation with the application performance department of IBM regarding OpenMP replacement to replace the pthread parallelization as well as several other optimizations developed within the core-developer team of Gadget. We also re-evaluated all physical modules which we want to use and optimized some of the implementations to ensure a proper behavior within the foreseen simulation setup. In total, we have run more than 100 small scale simulations to test and verify the different parts of the current code.

Such large simulations require not only a substantial effort in terms of development of the simulation code, but also require the development of appropriate post-processing and analysis tools. Data management and access can be foreseen to be already quite demanding, and therefore such pathfinder simulations are an ideal testbed for improving algorithms and infrastructure of the whole simulation pipeline. A challenge is that for such large simulations these tools also have to be parallelized and optimized. Some of the basic post-processing tools have already been developed for such large and complex simulations, others are currently under development as part of the preparation for the future, larger simulations. As part of the first pathfinder simulations, many of the most critical post-processing tools have been optimized and further developed:

- Visualization for static movies and time evolution. (developed and tested, fully MPI/OpenMP parallelized, Dolag et al. 2008)

- Map-making tools for flat maps, deep light cones and full sky maps. (developed and tested, parallelized based on producing and combining partial maps, adaptions for arbitrary lightcone angles through the simulations are in the testing phase, middle panel of figure 2 shows an example applied to Box3(mr), Dolag et al. 2005)

- Detection of multiple component halos and sub-halos.
  This code has been updated and extended within the on the fly version, halo properties have been extended to hold various, multi wavelength informations.
  (developed and tested, fully MPI parallelized, Springel et al. 2008, Dolag et al. 2010)

- Photometric code to assign optical/near-IR luminosities to galaxies identified in the simulations. Has now been extended and joined to the on the fly post processing. Right panel of figure 1 is using already this on the fly photometric galaxy results.
  (developed and tested, fully MPI parallelized, Saro et al. 2006, Nuzza et al. 2010)

- Merger-tree construction and application to SAMs, to produce mock galaxy surveys.
  (adapted and tested for hydro simulations, Saro et al. 2010)

- X-ray telescope simulator for mock observations.
  (Rasia et al. 2008, new, parallel version under development)

- Mock optical observations using SkyLens.
  Adapted to use the photometric properties available through the on the fly post processing and the light cone construction. Right panel of Figure 3 is already produced by this.
  (Meneghetti et al. 2010, currently being parallelized and ported to GPU)

- Cluster and group properties extractor.
  Has now been extended and joined to the on the fly post processing. (fully MPI parallelized)

- Novel sub-data access scheme based on Peano-Hilbert sorting and look-up table.
  Allows an extremely efficient read-out of particles belonging to a galaxy cluster from the full particle data. Figure 2 is made already using this new infrastructure. (developed and tested)

Besides the primary scientific targets, which will be reached by direct access of the team members to the data (we plan to have full copies of the simulation/postprocessing data at the major places involved), we also want to make simulation outcome (halo catalogues, maps etc.) available online. Prototype databases fed with small, test simulations are already available:

- http://www.g-vo.org/HydroClusters/

- http://www.g-vo.org/hydrosims/

1.6. Work Schedule

A large fraction of the preparation work for the code optimization has been already realized for the simulation within the last DEISA DECI-6 call. However, some further optimization will be needed. The time-line for the project is:

- OpenMP parallelization within additional physics modules (especially the density calculation needed in SubFind as well as the chemical enrichment part within the stellar evolution and feedback module). (**2 month**)
- Testing of the code performance by re-doing Box3(mr) and other tests with larger particle numbers. (**1 month**)
- Performing Box2(hr) simulation. Here, the simulation will be performed in several steps to allow monitoring and verification during the whole run. (**3 month**)
- Performing small scale test simulations at the resolution of Box3(uhr), to verify and calibrate the feedback parameters if needed. (**1 month**)
- Performing Box3(uhr) simulation. As before, it will be performed in several steps to allow monitoring and verification during the whole run. (**5 month**)

Scientific usage of the obtained simulation data will proceed as soon as the individual simulations are available and will be performed by the different working groups in parallel. They will form the bases of several PhD projects, and it can been foreseen that this data will be used far beyond the current grant period. It is also clear that these simulations still are in the spirit of pathfinder simulations and will be refined and complemented by further simulation activities within the next decade. They are a part of our continued theoretical effort to understand the state of the ordinary (baryonic) matter within the framework of cosmic evolution and a key ingredient to interpret data from current and forthcoming astronomical surveys and instruments.

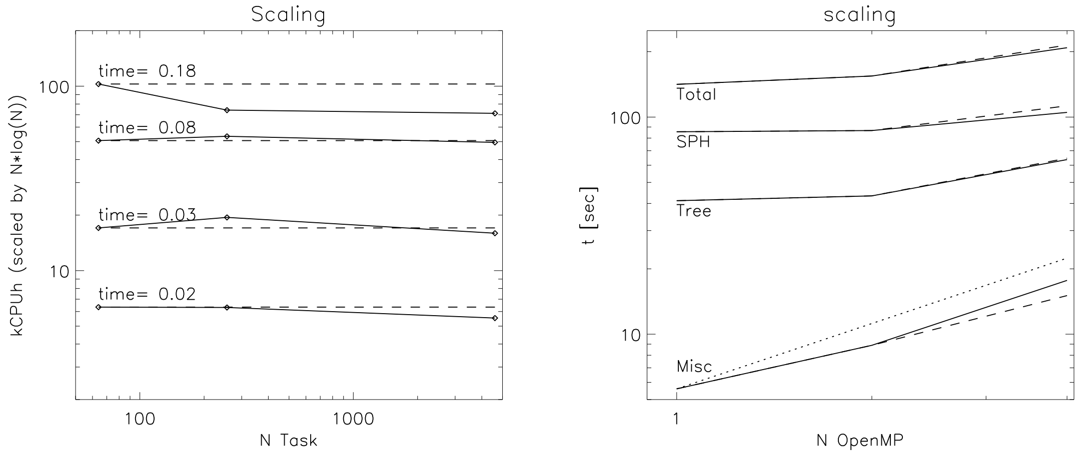

Fig. 4.— The left panel shows the scaling of CPU time for simulating *Box3(mr)*, *Box2(mr)* and *Box1(mr)* while increasing the number of MPI tasks from 64 to 256 to 4608. The different lines (from bottom to top) are with increasing time of the simulation. The actual CPU time is scaled by change of number of particles with the expected $N \times \log(N)$ scaling (as expected for an idealized Tree/SPH algorithm). A horizontal line (dashed) would indicate perfect scaling. Declining towards larger number of tasks indicates even better scaling and originates from minor parts of the code (like drift or kick of the particles), which naturally scale just with the number of particles N. The right panel shows the scaling of the code when going from pure MPI to mixed MPI/OpenMP while keeping the total number of CPUs used constant. The different line bundles correspond to different parts of the code. As before, a flat line would indicate perfect scaling. The slope of the dotted line as indicated for the "*Miscellaneous*" parts of the code would indicate no scaling at all. Dashed line represents the use of pthreads, the solid line reflects the new OpenMP implementation.

1.7. Code performance

P-Gadget3(XXL) is a highly optimised, MPI/pthread/OpenMP parallelized N-Body TreePM MHD-SPH code. It has already been used to perform several of the worlds leading cosmological simulations (e.g. Millennium, MillenniumII, Aquarius and many more). It has been run on various platforms (IBM Power3,4,5,6 / Blue Gene, Linux based intel/amd clusters, SGI Altix, etc.). It shows a very good scaling with large numbers of CPUs, especially for cosmological boxes when using MPI parallelization. Scaling of the code up to 4608 MPI tasks is shown in the left panel of figure 4. Especially on a Blue Gene architecture we expect a good performance of the MPI parallelization due to the underlying communication structure which in contrary to other platforms are not expected to result in loss of large amounts of memory for MPI buffers when going beyond several thousand MPI tasks. We plan to use the new OpenMP parallelization within each core (e.g. over the 4 CPUs). The most computationally intense parts (e.g. PM, Tree-force and hydro-force) are already OpenMP paralellized and also show acceptable scaling behaviour, other cpu intensive parts of the code (e.g. cooling and SubFind) will be also OpenMP parallelized within the preparation for this project. The right panel of figure 4 shows the scaling of the different parts of the code when switching from MPI to OpenMP parallelization for mixed jobs.

From our code testing we therefore expect overall less than 40% losses when running on 8 racks in a 8192 (MPI tasks) x 4 (OpenMP tasks) = 32768 CPUs.

Simulation	Box3(mr)	P6 to BG/p	resolution	Num. of Part	losses	PostProc	estimated time
Box2(hr)	4e5	x10	x(14/5)	x1.2	x1.3	x1.5	**2.6e7** core hours
Box3(uhr)	4e5	x10	x(14/2)	x1.1	x1.3	x1.7	**6.8e7** core hours

Table 2: Scaling of the CPU time (in core hours) measured for *Box3(mr)* to the proposed runs including change of architecture, increased number of time steps due to higher resolution, (slightly) larger number of particles, usage of mixed OpenMP and MPI and higher post processing costs due to the relative increase of resolved structures in such higher resolution runs. Therefore the two simulations sum up to **9.4e7 core hours**, e.g. **31 rack months**.

1.8. Estimated resources on JUGENE

The resources estimated on JUGENE come from basically two constraints. One is the memory requirements. Here the number of particles for *Box1(mr)*, *Box2(hr)* and *Box3(uhr)* are quasi identical, so we can safely extrapolate from our previous experience with *Box1(mr)*. To run our first pathfinder simulation (*Box1(mr)*) on 4068 CPUs we needed in total ca. 9TB of main memory. Including the increased number of MPI buffers and a slightly higher clumping (and therefore memory inbalance) when going to higher resolution gives us an estimate that the main memory of 8 racks (e.g. 16TB) should be enough to run the foreseen simulations.

For the CPU time estimate, we can also combine our experience with the previous simulations, including the time increase due to smaller timesteps when increasing the resolution and other cosmological factors like the shift of the cost of different parts of the code when changing box size and resolution. Table 2 lists the factors collected from various test runs. Including change of architecture, increased number of time steps due to higher resolution, (slightly) larger number of particles, usage of mixed OpenMP and MPI and higher post processing costs due to the relative increase of resolved structures in such higher resolution runs, the two proposed simulations sum up to **9.4e7 core hours**, e.g. **31 rack months**.

Rererences

Böhringer et al. 2010: Substructure of the galaxy clusters in the REXCESS sample: observed statistics and comparison to numerical simulations, A&A, 514

Borgani et al. 2004: "X-ray properties of galaxy clusters and groups from a cosmological hydrodynamical simulation", MNRAS 348

Boylan-Kolchin et al. 2009: "Resolving cosmic structure formation with the Millennium-II Simulation", MNRAS 398

Dolag & Stasyszyn 2009: "An MHD Gadget for cosmological simulations", MNRAS 398

Dolag et al. 2004: "The effects of Thermal Conduction in Simulated Galaxy Clusters", ApJ 606L

Dolag et al. 2005a: "Turbulent gas motion in galaxy cluster simulations: The role of SPH viscosity", MNRAS 364

Dolag et al. 2005b: "Constrained Simulations of the Magnetic Field in the Local Universe and the Propagation of UHECRs", JCAP 1

Dolag et al. 2008: "Splotch: Visualizing Cosmological Simulations", New J. Phys., 10

Dolag et al. 2009a: "Substructures in hydrodynamical cluster simulations", MNRAS 339

Dolag et al. 20010: "Dynamical difference between the cD galaxy and the stellar diffuse component in simulated galaxy clusters", MNRAS 405

Donnert et al 2010: "Radio Halos From Simulations And Hadronic Models I: The Coma cluster", MNRAS, 401

Fabjan et al. 2008: "Evolution of metal content with redshift in hydrodynamical simulations of galaxy clusters", MNRAS 386

Fabjan et al. 2010: "Simulating the effect of AGN feedback on the metal enrichment of galaxy clusters", MNRAS, 401

Gazzola & Pearce 2007: "A Heating Model for the Millennium Gas Run", 2007hvcg.conf..412G

Gottloeber et al. 2006: "The MareNostrum Universe", astro-ph 0608.289

Heitmann et al. 2008: "The Coyote Universe I: Precision Determination of the Nonlinear Matter Power Spectrum", astro-ph 0812.1052

Kay et al. 2005: "Clusters of galaxies: new results from the CLEF hydrodynamics simulation", 2005AdSpR..36..694K

Maio et al 2007: "Metal and molecule cooling in simulations of structure formation", MNRAS 379

Meneghetti et al. 2010: Weighing simulated galaxy clusters using lensing and X-ray, A&A 514

Murante et al. 2007: "The Importance of Mergers for the Origin of Intracluster Stars in Cosmological Simulations of Galaxy Clusters", MNRAS 377

Nuza et al. 2010: "Photometric and clustering properties of hydrodynamical galaxies in a cosmological volume: results at z=0", MNRAS, 407

Puchwein et al. 2008: "Simulations of AGN Feedback in Galaxy Clusters and Groups: Impact on Gas Fractions and the L_X-T Scaling Relation", ApJ 687L

Puchwein et al. 2010: "Intracluster stars in simulations with AGN feedback", MNRAS 406

Rasia et al. 2008: X-MAS2: Study Systematics on the ICM Metallicity Measurements, ApJ 674

Saro et al. 2006: "Properties of the galaxy population in hydrodynamical simulations of clusters", MNRAS 369

Saro et al. 2010: Gas cooling in semi-analytic models and SPH simulations: are results consistent?", MNRAS 406

Schaye et al. 2009: "The physics driving the cosmic star formation history", 2009MNRAS.tmp.1888S

Springel et al. 2005: "Simulations of the formation, evolution and clustering of galaxies and quasars", Nature 435

Springel et al. 2008: The Aquarius Project: the subhaloes of galactic haloes, MNRAS 391

Tescari et al. 2009: Damped Lyman systems in high-resolution hydrodynamical simulations, MNRAS 397

Tornatore et al. 2007a: "Population III stars: hidden or disappeared", MNRAS 382

Tornatore et al. 2007b: "Chemical enrichment of galaxy clusters from hydrodynamical simulations", MNRAS 382

Wiersma et al. 2009: Chemical enrichment in cosmological, smoothed particle hydrodynamics simulations, MNRAS 339