The “Planck Process Coordinator” Workflow Engine

ThePlanck Surveyor Satellite Project

The Planck Surveyor is an ESA satellite to be launched in the near future. The satellite will hover in the second Lagrange point of the Sun-Earth system, from where it will observe the cosmic microwave background (CMB) radiation, a remnant of the Big Bang. By analyzing the CMB radiation, in particular its so-called “autocorrelation function”, astrophysicists may derive the value of important cosmological parameters such as the Hubble constant, which measures the speed of the expansion of our Universe, or the fraction of cold dark matter (CDM), which apparently dominates the mass-energy content of the Universe.

Fig. 1: The Planck Surveyor satellite in an ESA assembly hall.

By exploiting the so-called Sunyayev-Zeldovich effect, the Planck satellite will also detect many distant galaxy clusters. Planck will also observe the indirect effects of the magnetic field in our own Galaxy, and of the enigmatic gravitational lenses.

The Planck Process Coordinator

The Planck Process Coordinator (ProC) is a scientific workflow engine, which has originally been designed for use within the Planck Surveyor project. The ProC executes modules (“jobs”) which can be combined to sub-workflows. Modules and sub-workflows in turn can be combined into executable workflows.

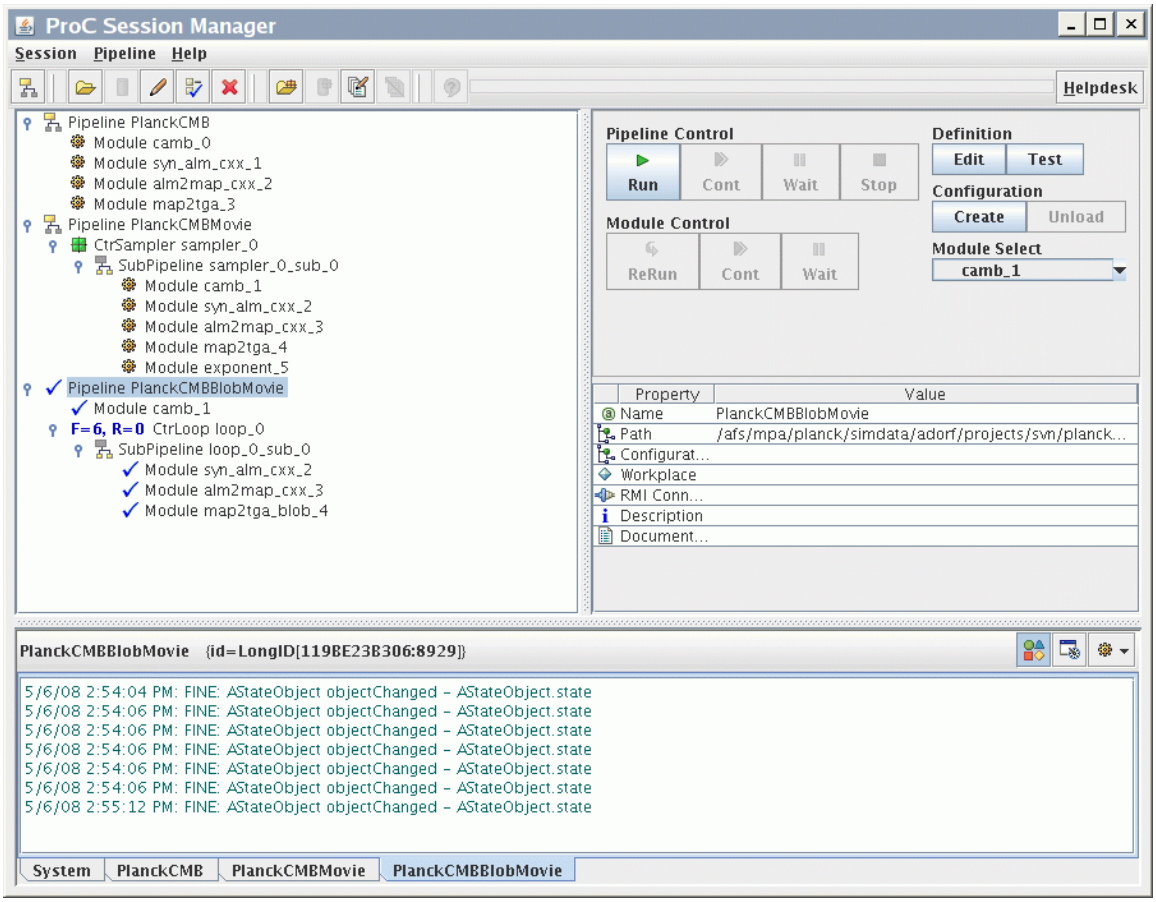

Fig. 2: The GUI of the Planck Process Coordinator (ProC) workflow engine. The ProC is mainly used within the Planck Surveyor project, but may be used for organizing the execution of other scientific tasks, including simulations and data analysis.

The ProC is a multithreaded application implemented in Java and XML. It operates in a data-driven, forward-chaining mode, with the capability to execute many jobs in parallel. To this end the ProC offers a standard parallel control element and its variant, the so-called sampler control element.

The ProC offers a sophisticated mechanism of logging workflow-specific information. It may be operated in a file-based standalone mode, or in conjunction with the Planck Data Management Component.

The Planck Data Management Component

The Planck Data Management Component (DMC) is a software product accompanying the ProC. on behalf of which it manages data and metadata. The DMC is implemented in Java and XML, and uses a Java Data Objects (JDO) layer for persisting the data /metadata into a local or remote database. Code written in C/C++ or Fortran may use the DMC through the Java Native Interface (JNI). The DMC has successfully been operated in conjunction with the Oracle, PostgrSQL, Apache Derby, and HSQLDB database management systems.

The benefit of using the DMC become obvious when many different simulations or data analysis tasks have to be carried out. Since the parameters for each run are stored in the database, the standard database query mechanisms may be used to acquire information about the work accomplished, and to locate the

data sets pertaining to certain parameter sets.

The Planck LevelS simulation package

The Planck LevelS simulation package is currently being developed by the Planck consortium in preparation of the launch of the Planck Surveyor satellite. Simulations are needed in order to predict the capabilities of the science instruments on-board the Planck satellite.

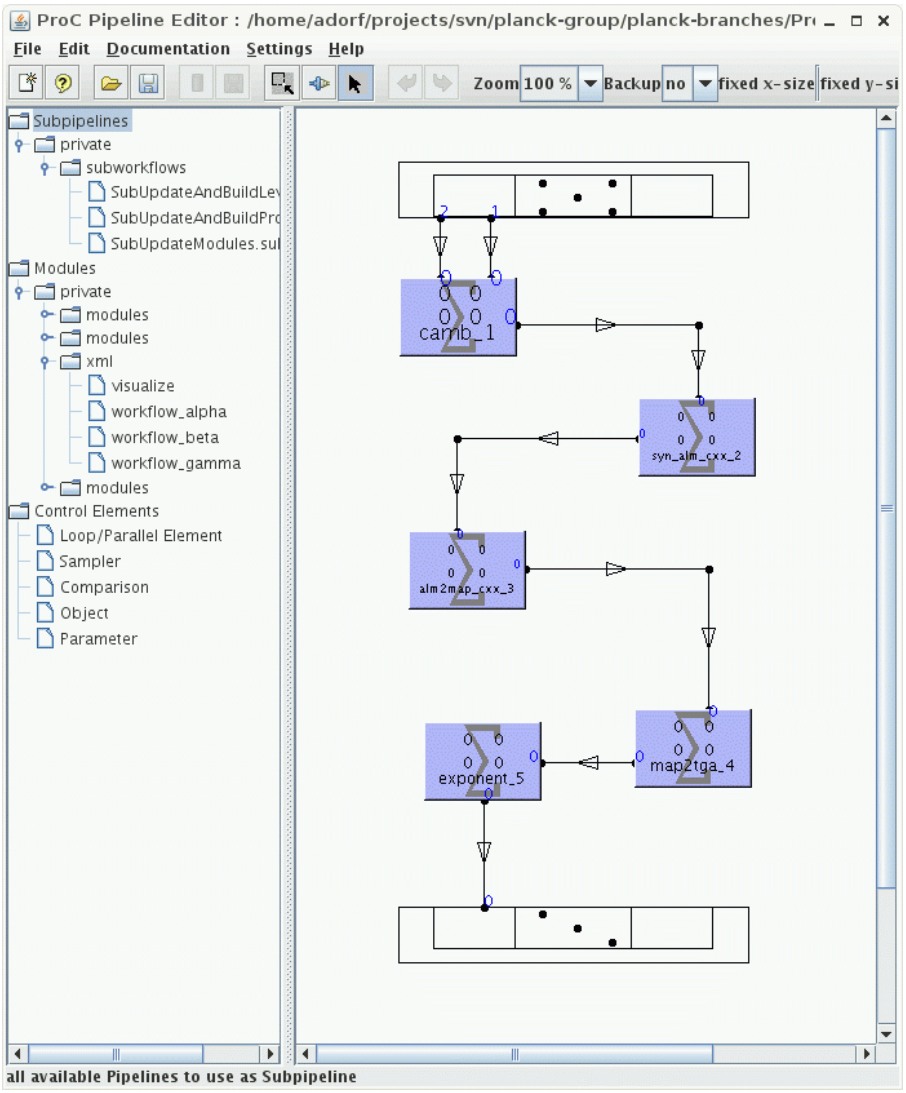

Fig. 3: The Planck CMB Sky video production workflow, as presented in the ProC editor GUI. At the top and the bottom the workflow is enclosed by the two halves of the *sampler control element*. Many such workflows can be executed in parallel. The individual workflow tasks are modules from the Planck LevelS simulation package.

ProC and DMC on the Grid

'ProC and DMC on the Grid' is a Grid use case developed within the AstroGrid-D project. The use case was designed and largely implemented at the Max Planck Institute for Astrophysics, Garching. The ProC was interfaced to the Grid Application Toolkit (GAT), which allows the execution of jobs on the submission host, on clusters via the PBS and SGE GAT adapters, and on the Grid, using the Globus Toolkit 2 and 4 GAT adapters for process-to-process communication.

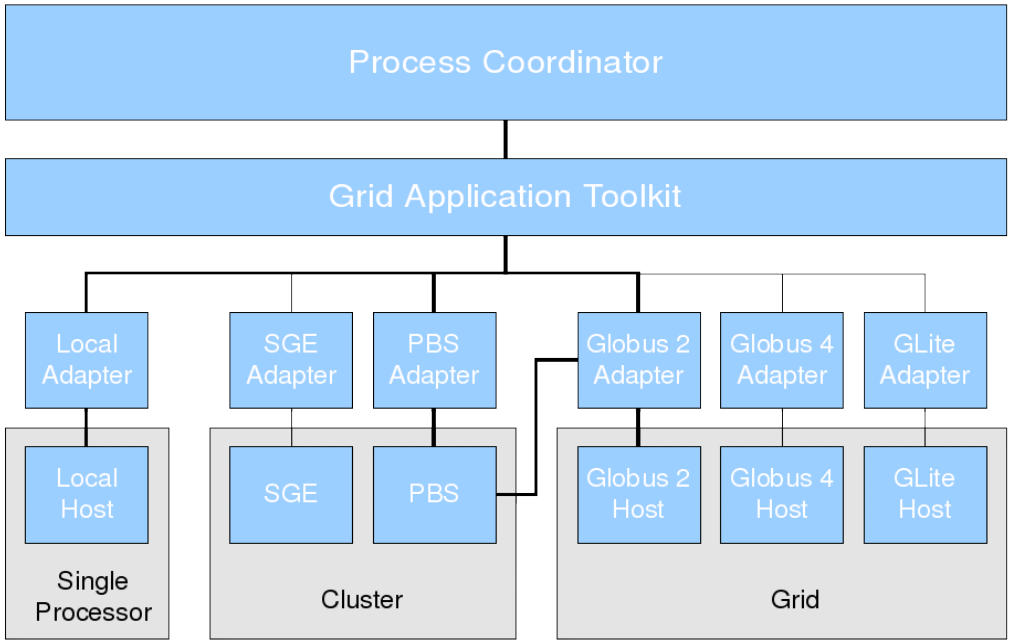

Fig. 4: The ProC interfaces to the Grid Application Toolkit (GAT) which via its set of adapters offers job execution on the local host, on worker nodes of a local cluster, and on remote Grid hosts.

The Cosmic Microwave Background Sky Demo

The CMB Sky demo is a technology showcase demonstrating the interplay between different technologies. In one scenario a path through the multi-dimensional cosmological parameter space is prescribed via a set of way points. A sampling point generator follows the path stopping a number of times on its way. At each stop the ProC starts a workflow with a CMB-simulation, which eventually produces an image of the CMB-sky as it will be observed by the Planck satellite. These images may be combined into a movie illustrating in an intuitive fashion how different cosmologies (characterized by different cosmological parameter sets) would produce different CMB-skies. By comparing the actual observation with a number of simulations, astrophysicists may find the best fitting set of cosmological parameters.

Even in the simplest cosmological setting and when excluding foreground objects, each simulation takes of the order of half a minute. Since the individual simulation workflows are independent, they may all be

executed in parallel. The power of the ProC + GAT was shown in several trials kindly supported by the TU Dortmund, where up to 200 cores were working in parallel each generating an image of the CMB-sky.

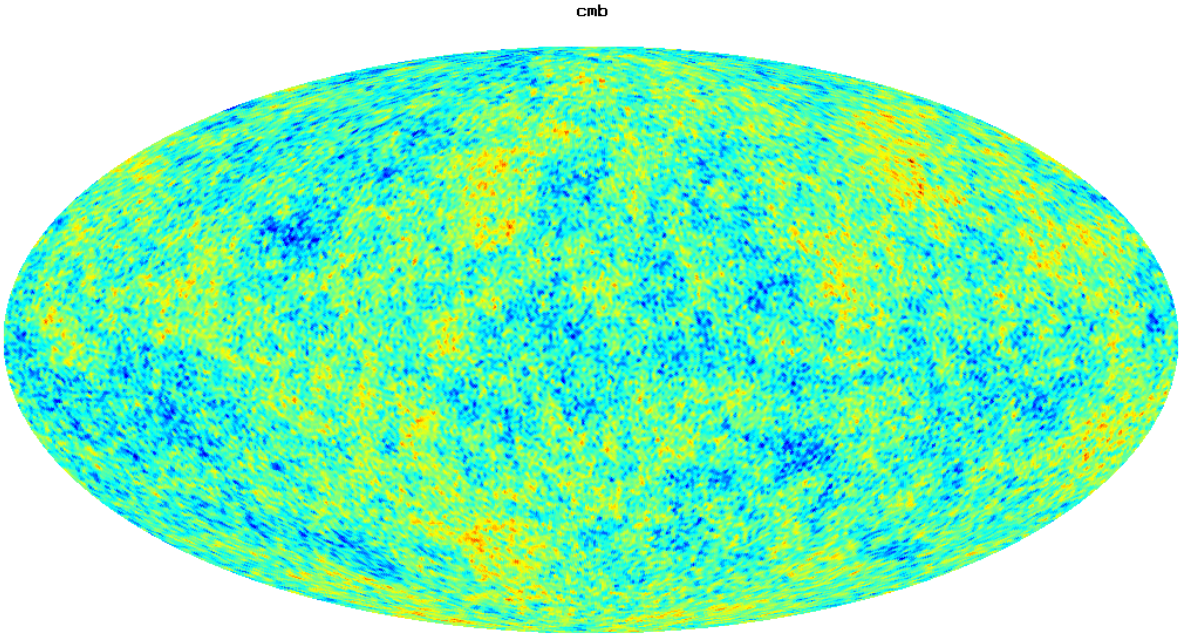

Fig. 5: An image of the Cosmic Microwave Background sky as it will be observed by the future ESA Planck Surveyor satellite. This image was produced during a CMB Sky demo run, using the Planck LevelS simulation package. The CMB has a temperature of 2.7 Kelvin, and the image displays temperature variations on the level of 10^-5.

Contact

If you are interested to learn more about the ProC and DMC on the Grid, have any suggestions or want to provide feedback, please contact MPA's Planck group headed by Torsten Ensslin.