Bulgarian GRID, Bulgarian Virtual observatory and some astronomical applications

Georgi Petrov, Momchil Dechev and Emanouil Atanasov

VII Bulgarian-Serbian Astronomical Conference - ASTROINFORMATICS
01 – 04 June 2010 – Chepelare, BULGARIA

1. Introduction
2. Grid portals, WEB services etc.
3. Middleware, Frameworks etc.
4. Software and standards
5. Modelling, simulations etc.
6. Astronomical applications and projects

BG GRID, BG VO and some astronomical applications

INTRODUCTION

1. A new age for the oldest science
2. The Free Lunch Is Over
3. The Grid: separating fact from fiction
4. Back to Basics - Why go parallel?
5. What makes parallel programming hard?
6. BOINC - *Kilo, mega, giga, tera, peta . . . exa?*

1. A new age for the oldest science

For millennia, astronomy meant looking at the night sky and sketching what you saw... But ***the human visual system remained an integral component, limiting data gathering*** to what could be seen and sketched by human observers in real time.

The advent of the photographic plate caused a revolution... But ***extracting data still involved human effort***, from developing the photographic plates to reducing data into a standard format.

In recent years, the photographic plates has been superseded by ***digital photography***. As digital detectors became larger and cheaper they gave birth to a ***new kind of astronomy***: Fewer ***surveys***, but ***on a much larger scale***, mapping vast areas of the sky at a time. ***Data reduction is automatic*** and ***ends up in query-able databases*** that astronomers worldwide can use.

How much data would surveying the whole sky generate? ***The atmospheric interference*** limits the resolution you can observe to ***about half an arcsecond***, and allowing 2-to-4 bytes per pixel puts ***the size of a whole-sky survey at roughly 20 terabytes***.

In the old days of photographic plates, producing 20 terabytes might take 60 years of observing time, and another ten years of digitization. ***Current digital sky surveys can produce 20 terabytes in a year***. *The newest generation of sky surveys will produce 20 terabytes every night for a decade.*

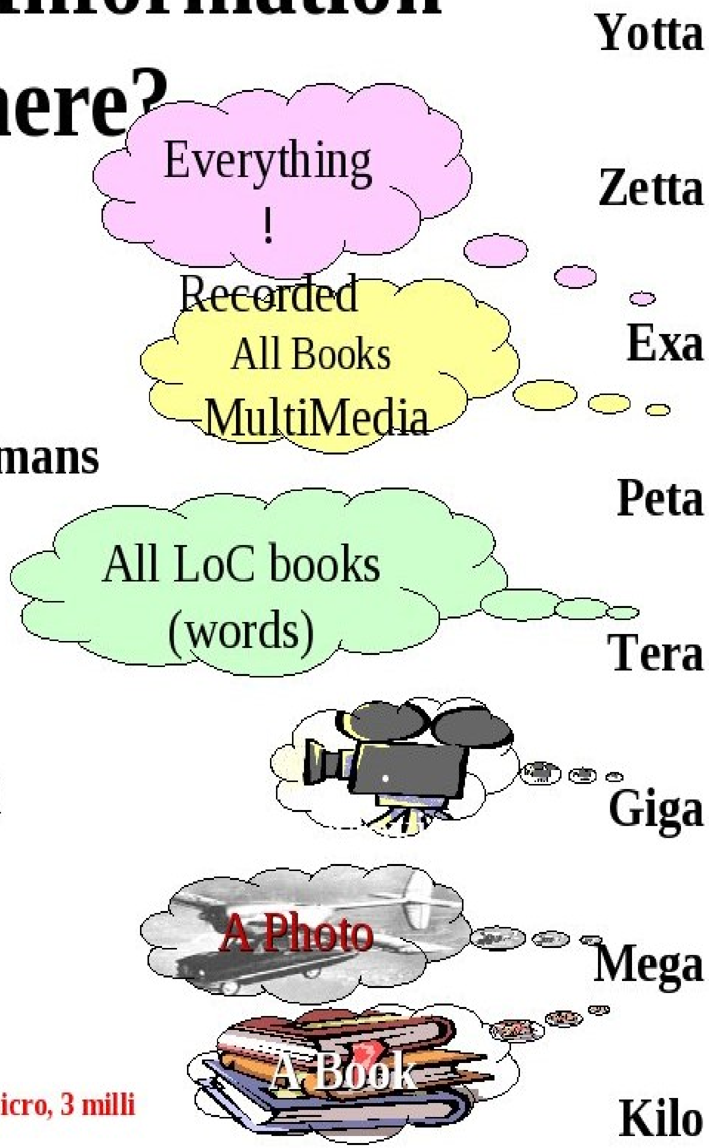
If a sky survey is to be able to issue an alert within one minute of detecting some, e.g. a transient effect, the ***data reduction system needs to achieve a data rate of about 2 terabytes per hour***.

Therefore, ***computational astronomy is an excellent sandbox for data-mining algorithms, and an effective way to teach both astronomy and computer science***. Consequently, the new generation of sky surveys will encourage the development of new computational techniques.

How Much Information Is there?

- **Soon everything can be recorded and indexed**
- **Most data never be seen by humans**
- **Precious Resource: Human attention**
  Auto-Summarization
  Auto-Search
  is key technology.
  www.lesk.com/mlesk/ksg97/ksg.html

24 Yecto, 21 zepto, 18 atto, 15 femto, 12 pico, 9 nano, 6 micro, 3 milli



2. The Free Lunch Is Over: A Fundamental Turn Toward ***Concurrency in Software*** (by Herb Sutter- 2005 - 2009)

The major processor manufacturers and architectures have run out of room with most of their traditional approaches to boosting CPU performance. Instead of driving clock speeds and straight-line instruction throughput ever higher, they are instead turning *en masse* to ***hyperthreading*** and ***multicore*** architectures.

No matter how fast processors get, software consistently finds new ways to eat up the extra speed.

Moore's Law and the Next Generation(s)

"There ain't no such thing as a free lunch." *—R. A. Heinlein, The Moon Is a Harsh Mistress*

Moore's Law predicts exponential growth, and clearly exponential growth can't continue forever before we reach hard physical limits; light isn't getting any faster.

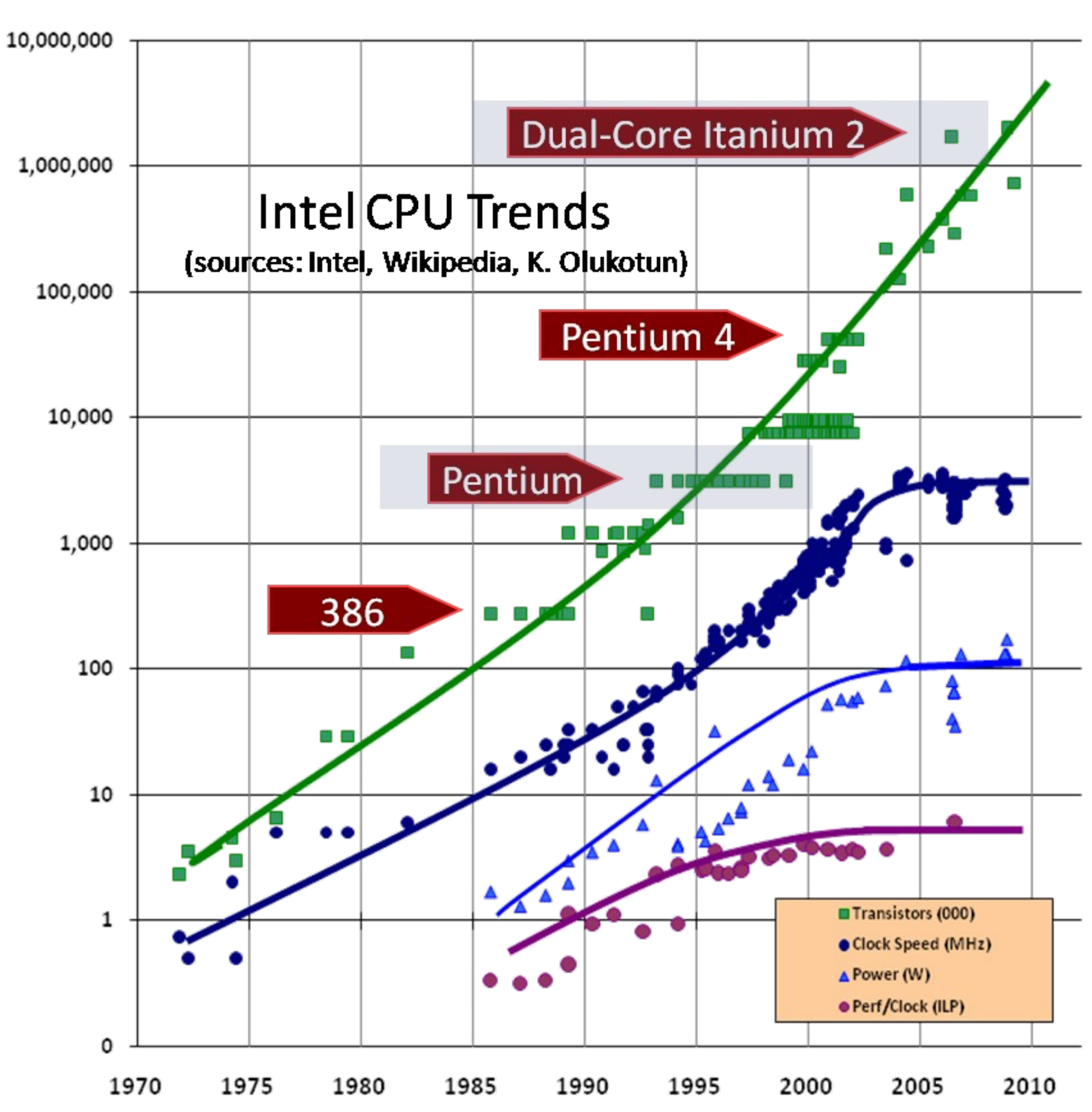
Moore's Law

- Performance/Price doubles every 18 months
- 100x per decade
- Progress in next 18 months
  = ALL previous progress
  - New storage = sum of all old storage (ever)
  - New processing = sum of all old processing.
- Rem.: E. coli double ever 20 minutes!

Over the past 30 years, CPU designers have achieved ***performance gains*** in three main areas:

- ***clock speed -*** *limited*
- ***execution optimisation –*** *quite dangerous*
- ***cache – most perspective***

CPU performance growth as we have known it hit a wall two years ago. Figure 2 graphs the history of Intel chip introductions by clock speed and number of transistors. The number of transistors continues to climb, at least for now. ***Clock speed***, however, ***is a different story.***



Myths and Realities: 2 x 3GHz < 6 GHz

So a dual-core CPU that combines two 3GHz cores practically offers 6GHz of processing power. Right?

Wrong. They should run faster than on a single-core CPU; the performance gain just isn't linear, but that's all. Today, a two- or four-processor machine isn't really two or four times as fast as a single CPU even for multi-threaded applications. If you're running a single-threaded application, then the application can only make use of one core.

For the near-term future, the ***performance gains*** in new chips will be fueled by three main approaches:

- ***hyperthreading***
- ***multicore***
- ***cache***

What This Means For Software: The Next Revolution

Starting today, the performance lunch isn't free any more. If you want ***your application*** to benefit from the throughput advances in new processors, it will ***need to be a well-written concurrent (usually multithreaded) application***. And ***that's easier said than done***, because ***not all problems are*** inherently ***parallelisable*** and because ***concurrent programming is hard***. A few rare classes of applications are naturally parallelizable, but most aren't.

3. The Grid: separating fact from fiction

Grid computing revolutionizes the way scientists share and analyse data. So ***what about grid computing is fact, and what is fiction?***

Fiction: The Grid will replace the Internet.

Fact: Grid computing, like the World Wide Web, is an application of the Internet.

Fiction: People will be able to download movies 10,000 times faster using the Grid.

Fact: First, in order to get such data-transfer rates, individuals would have to do what the large particle physics computing centres have done. Second, today's grid computing technologies and projects are geared toward research and businesses with highly specific needs... so free logging onto a computing grid may not be soon.

Fiction: The Grid was invented at CERN.

Fact: The first pioneering steps in grid computing were taken in the US.

4. Back to Basics - Why go parallel?

Parallel programs are no longer limited to the realm of high-end computing. C***omputers with multiple processors*** have been around for a long time, and people have been studying parallel programming techniques for just as long. However, only in the last few years have ***multi-core processors and parallel programming become truly mainstream***... If you want an application to get faster, you can no longer rely on processor clock speed increasing over time... ***The application must be written in parallel and it must be able to scale to the number of available cores.*** Parallelizing applications is not just important, but necessary!

Glossary of terms:

- **core:** the part of a processor responsible for executing a single series of instructions at a time.
- **processor:** the physical chip that plugs into a motherboard. A computer can have multiple processors, and each processor can have multiple cores
- **process:** a running instance of a program. A process's memory is usually protected from access by other processes.
- **thread:** a running instance of a process's code. A single process can have multiple threads, and multiple threads can be executing at the same on multiple cores. within a single process each thread represents an independent, concurrent path of execution. Threads within a process share memory.
- **parallel:** the ability to utilize more than one processor at a time to solve problems more quickly, usually by being multi-threaded. Running instance of a program. A process's memory is usually protected from access by other processes.

5. What makes parallel programming hard?

Dual and even quad-processor computers may be increasingly common, but ***software that is designed to run in parallel is another story*** entirely. ***What makes parallel programming different and why that makes it harder?***

We think of a function as a series of discrete steps. Let's say a function ***f*** is composed of steps ***f1, f2, f3... fn***, and another function ***g*** is composed of steps ***g1, g2, g3...gn***.

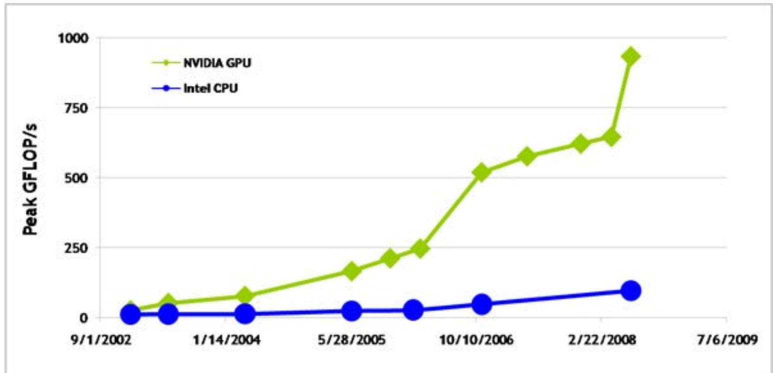
A ***single-threaded application*** can run *f* and *g* in only two ways: ***f(); g(); or g(); f();*** ==> That would lead to ***two possible orders for the steps***: ***f1, f2, f3... fn, g1, g2, g3... gn*** or ***g1, g2, g3... gn, f1, f2, f3... fn***. ***The order is determined by the order in which the application calls f() and g().***

What happens ***if f and g are called in parallel*** – that is to say, at the same time. Although ***f2*** will still execute before ***f3***, ***there is no control over the order in which f2 will be executed relative to g2.*** Thus, ***the order could be g1, g2, f1, g3, f2, f3*** .... or ***f1, g1, f2, f3, g2, g3***... or ***any other valid order***. ***Each time the application calls these functions in parallel, the steps could be executed in a different order.***

Summary: ***The order in which single threaded code executes is determined by the programmer***. This makes the execution of a single threaded application relatively predictable. ***When run in parallel, the order in which the code executes is not predictable***. This makes it much harder to guarantee that code will execute correctly.

6. BOINC or the Berkeley Open Infrastructure for Network Computing.

Remember your prefixes? ***Kilo, mega, giga, tera, peta . . . exa***? Each denoting a thousand times more than the one before? Today, the ***average personal computer*** can do a few ***GigaFLOPS*** . ***A modest cluster*** might do one thousand GigaFLOPS, or ***1 TeraFLOPS***. And for several years, one thousand TeraFLOPS, or ***one PetaFLOPS***, was the Holy Grail of ***high-performance computing (HPC)***. That one PetaFLOPS milestone was reached in the past year, first by Stanford's Folding@Home (a volunteer computing project), then by various volunteer computing projects using BOINC, a ***middleware system for volunteer computing*** developed by my research group at the University of California at Berkeley. More recently, that same milestone was attained by IBM's RoadRunner supercomputer.



Over the last few years, Graphic Processor Units (green) and Central Processor Units (blue) have increased exponentially in speed, but ***the doubling time for GPUs has been about 8 months, and 16 months for CPUs...***

BOINC has recently added support for GPU computing.

In summary: the combination of volunteer computing and GPUs can feasibly provide Exa-scale computing power for science in a remarkably short time frame, years ahead of other paradigms. Scientists wanting to share in this resource can do so by developing GPU-enabled applications and deploying them in BOINC-based volunteer computing projects. ***********

2. GRID PORTALS, WEB SERVICES etc...

1. Grid Projects

EGI - European Grid Initiative EMI - European Middleware Initiative gLite - European middleware distribution Open Science Grid - the U.S scientific Grid infrastructure Virtual Data Toolkit - provides middleware distribution	**National Grid projects** Dutch Grid GridPP INFN Grid LCG France NorduGrid WestGrid GÉANT - European academic and research network infrastructure	**Previous projects** EGEE European Data Grid Datatag Grid2003 GriPhyN iVDGL PPDG

2. AccessGrid

The ***Access Grid is an ensemble of resources*** including multimedia large-format displays, presentation and interactive environments, ***and interfaces to Grid middleware*** and to visualization environments. These resources are used to support group-to-group interactions across the Grid.

You can now view the ***locations of Access Grid nodes around the world*** on an interactive map. This convenient visualization provides understanding about the locations and distribution of AG nodes.

Access Grid 3.1 is now available for download. This release includes new functionality, including the latest VIC (video) and RAT (audio) tools from the Sumover project, support for hierarchical venue data storage, improved bridging support, and certificate-based Venue access controls.

3. AstroWISE - *Astronomical Wide-field Imaging System for Europe*

co-ordinated by OmegaCEN-NOVA – NL.

4. BalticGrid-II project

On 1 May 2008, the BalticGrid Second Phase project has started. It is designed to increase the impact, adoption and reach, and to further improve the support of services and users of the recently created e-Infrastructure in the Baltic States. This will be achieved by an ***extension of the BalticGrid infrastructure to Belarus***; interoperation of the greasy-based infrastructure with UNICOR and ARC based Grid resources in the region; Identifying and addressing the needs of new specialize in scientific communities such as nano-science and engineering sciences , and by Establishing new Grid services for linguistic research, Baltic Sea environmental research, data mining tools for communication modeling and bioinformatics.

Some Applications (extraction):

- ATOM *is a set of computer programs for theoretical studies of atoms.*
- Crystal06 *A quantum-chemistry package to model periodic systems. Limited to VU ITPA users.*
- Computational Fluid Dynamics (FEMTOOL) *Modelling of viscous incompressible free surface flows.*
- Density of Montreal *A molecular electronic structure program. Limited to VU ITPA users.*
- **ElectroCap** *Stellar Rates of Electron Capture. A set of computer codes produce nuclear physics input for **core-collapse supernova simulations**.*
- MATLAB Distributed computing server™
- Vilnius Parallel Shell model code *An implementation of nuclear spherical shell model approach.*

and many others...

5. BOINC - *Open-source software for volunteer computing and grid computing.*

What is volunteer computing? Volunteer computing is an arrangement in which people (volunteers) provide computing resources to projects, which use the resources to do distributed computing and/or storage.

- ***Volunteers*** are typically members of the general ***public who own Internet-connected PCs***. Organizations such as ***schools*** and ***businesses*** may also volunteer the use of their computers.
- ***Projects are typically academic*** (university-based) and do scientific research, but there are exceptions...

The first volunteer computing project was GIMPS (Great Internet Mersenne Prime Search), which started in 1995. ***Today there are over 50 active projects.***

Computing power - ***Top 100 volunteers*** · ***Statistics:***

Active: 320,115 volunteers, 588,156 computers. 24-hour average: 5,166.26 TeraFLOPS.

Compute with BOINC

- **Scientists**: use BOINC to create a volunteer computing project giving you the computing power of thousands of CPUs.
- **Universities**: use BOINC to create a Virtual Campus Supercomputing Center.
- **Companies**: use BOINC for desktop Grid computing.

BOINC is supported by the National Science Foundation.

6. COSMA

ICC Institute for Computational Cosmology	The Cosmology Machine (COSMA) was first switched on in July 2001. Now ***QUINTOR consists of a 256 SunFire V210s with a total of 512 UltraSparc IIIi 1 GHz processors and 576 GByte of RAM***

7. DEISA

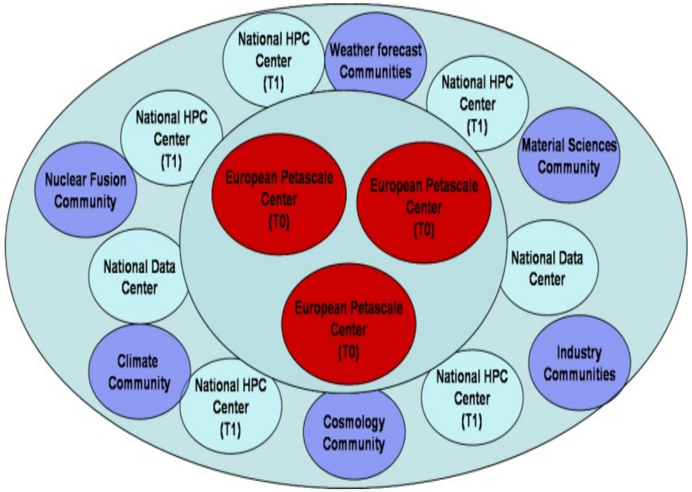
Distributed European Infrastructure for Supercomputing Applications	**DEISA Concept and Objectives**

DEISA2 support and develop the distributed ***high performance computing infrastructure*** in Europe. A ***major task*** is the ***federated operation of the powerful European HPC infrastructure built on top of national services***. This specific e-infrastructure will facilitate Europe's world-leading computational science research.

DEISA1, the previous project funded by the European Commission in the sixth Framework Programme, has proved its relevance for advancing computational sciences in leading scientific and industrial disciplines within Europe.

DEISA1 and now DEISA2, funded in FP7, are paving the way towards the deployment of a cooperative

European HPC ecosystem.



The existing infrastructure is based on the tight coupling of ***eleven leading national Supercomputing centres,*** using dedicated network interconnections of GÉANT2 and the National Research and Educational Network providers (NRENs). In DEISA2, activities and services concerning Applications Enabling, Operation, and Technologies are continued and enhanced, as these are indispensable for the effective support of computational sciences in the area of Supercomputing.

A two-fold strategy is applied:

- ***Consolidation of the existing DEISA infrastructure*** by guaranteeing the continuity of those activities and services that currently contribute to the effective support of world-leading computational science in Europe.
- ***Evolution of this European infrastructure towards a robust and persistent European HPC ecosystem***, by enhancing the existing services, by deploying new services including support for European Virtual Communities, and by cooperating and collaborating with new European initiatives, especially ***PRACE*** that will enable ***shared European PetaFlop/s supercomputer systems***.

8. EGEE - *Enabling Grids for E-sciencE*

Rem.: *The EGEE project* ***officially ended on April 30, 2010***

The distributed computing infrastructure is now supported by the ***European Grid Infrastructure***. This long-term organisation ***coordinates National Grid Initiatives***, which form the country-wide building blocks of the pan-European Grid.

Regional Web Sites

- EGEE South East Europe
- Spain EGEE-III site
- Romanian Grid site
- **Bulgarian Grid portal**
- Cyprus Grid site
- Hungarian Grid site
- Russian website
- Portuguese website
- Slovak website

9. FermiGRID

In order ***to better serve the entire program*** of Fermilab, the Computing Division has undertaken ***the strategy of placing all of its production resources in a Grid "meta-facility" infrastructure called FermiGrid.***

Among other things ***this strategy is designed to allow Fermilab*** to allow opportunistic use of the dedicated resources by various Virtual Organizations (VO's) that participate in the Fermilab experimental program and by certain VOs that use the Open Science Grid (OSG) and to make a coherent way of putting Fermilab on the Open Science Grid.

Open Science Grid Interfaces.

At Fermilab, compute resources are available in the context of the Open Science Grid (OSG) Compute Element (CE). The goal of the open science grid interfaces element effort is to enable the opportunistic use of Fermilab compute elements in a secure manner by external Virtual Organizations (VO's) through the use of Open Science Grid (OSG) interfaces .

Grid Projects at Fermilab

Fermilab is actively participating in the development and deployment of ***grid technology for high energy physics research***. Scientists are involved in a variety of grid projects, some involving CDF and D0 Run II data handling and other current research projects at the lab, others looking forward to and preparing for physics that will be coming from the LHC at CERN in a few years. These ***grid projects are collaborations of scientific and computer professionals*** from a number of participating labs, universities and other organizations throughout the U.S., Europe and Asia.

Scientific projects:

dCache | FermiGrid | Grid2003 | **GriPhyN** | interactions.org | iVDgL | OSG | PPDG

SAMGrid | **SDSS-GriPhyN** | SRM | USCMS S&C | VOX | VO Privilege | WAWG

Sloan Digital Sky Survey SDSS-GriPhyN Work Space ***Griphyn*** is a project to develop technologies around the ***concept of "virtual data" in which derived datasets can be recreated on-demand in a grid computing environment***. SDSS is applying these technologies to various analyses of the SDSS dataset, creating derived datasets such as galaxy cluster catalogs for use in studying phenomena such as dark energy.

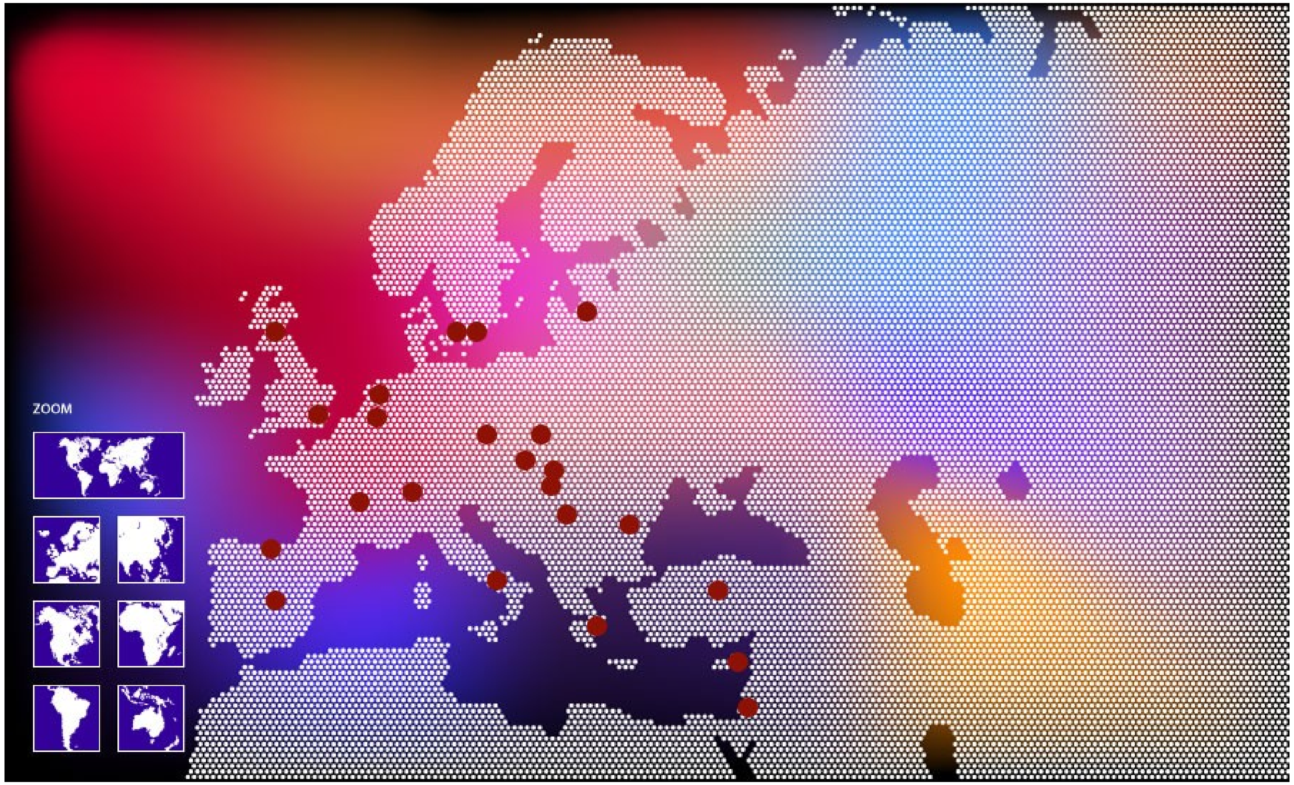
10. GridGuide

GridGuide is an innovative introduction to the sites — and sights — that contribute to ***global grid computing***, a technology that ***connects computers from around the world*** to create a powerful, shared resource for tackling complex scientific problems. The launch of GridGuide comes as part of the Enabling Grids for E-sciencE (EGEE) User Forum. While still a work-in-progress, the ***GridGuide website*** already ***allows visitors to explore an interactive map of the world***, visiting a sample of the thousands of scientific institutes involved in ***grid computing projects. Sites from 23 countries*** already ***appear on the GridGuide,*** offering insider snippets on everything from research goals and grid projects to the best place to eat lunch and the pros and cons of their jobs.

GridGuide is an EC-funded project, and most of the sites included so far are European.

GridGuide – Europa.

Rem.: ***BG is not presented***, bur Serbia and Romania are...



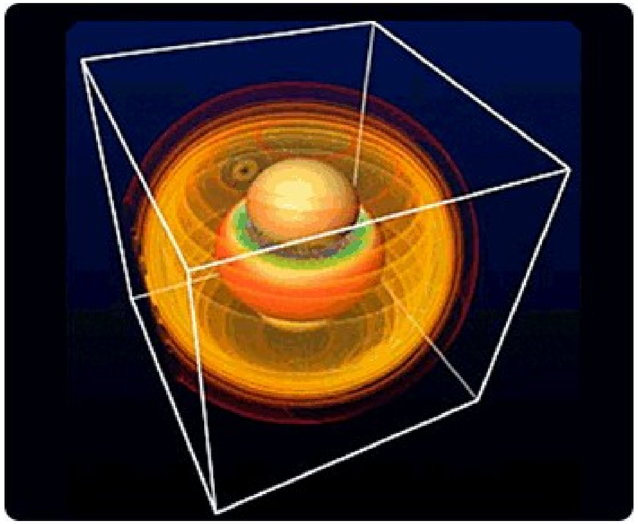
GridTALK coordinating grid reporting across Europe GridTalk brings the success stories of Europe's e-infrastructure to a wider audience. The project coordinates the dissemination outputs of EGEE and other European grid computing efforts, ensuring their results and influence are reported in print and online.

GridCAFE is an introduction to grid computing for the general public.

11. GLOBUS

Globus *is* ***open source Grid software*** that addresses the most challenging problems in distributed resource sharing.

The ***Globus Alliance*** is a ***community of organizations and individuals*** developing fundamental technologies behind the "Grid," which ***lets people share computing power, databases, instruments, and other on-line tools*** securely across corporate, institutional, and geographic boundaries without sacrificing local authonomy.



The ***Globus Toolkit*** is an ***open source software toolkit*** used for building Grid systems and applications. It is being ***developed by the Globus Alliance*** and many others all over the world.

Physicists used the Globus Toolkit and MPICH-G2 to harness ***the power of multiple supercomputers to simulate the gravitational effects of black hole collisions.***

12. EU India Grid

Europe and ***India*** are ***collaborating*** to exploit the vast potential of global eScience infrastructures through Grid computing in order to address ***global challenges such as climate warming and disease***. A series of European initiatives are involved in deploying and operating the European-wide e-Infrastructure. These initiatives cooperate with national programs at European and extra-European level. ***Eu-IndiaGrid supports and fosters collaboration between researchers from Europe and India in a wide range of scientific areas***.

GARUDA, the National Grid Initiative of India (http://www.garudaindia.in/) is a collaboration of scientific and technological researchers on a nationwide grid comprising of computational nodes, storage devices and scientific instruments. It aims to provide the technological advances required to enable data and compute intensive science for the 21st century. The establishment of Indian Grid Certification Authority (**IGCA**), (http://ca.garudaindia.in) for the first time in India by CDAC in November 2008 has allowed full access to worldwide grids for Indian Researchers and represented a landmark in this domain. This important milestone was achieved through strong cooperation with the FP6-EU-IndiaGrid.

Rem.: Included because of common BG - India astronomical project co. blazars variabilities...

13. GRID observatory

The ***Grid Observatory collects, publishes, and analyzes, data on the behaviour of the EGEE grid***.

The EGEE grid offers an unprecedented opportunity to observe, and start understanding the new computing practices of e-science. With ***more than 40000 CPUs*** and ***5PB of storage*** distributed worldwide, the management of ***100.000 concurrent jobs***, and the perspective of a sustainable development, the ***EGEE grid is one of the more exciting artificial complex systems to observe.***

The ultimate goal of the Grid Observatory is to ***integrate data collection, data analysis, and the development of models*** and of an ontology for the domain knowledge. The Grid Observatory is part of the EGEE-III project. Because grid data and models are equally relevant for computer science, middleware development and system administration, ***the Grid Observatory is an open project***.

14. NASA ADVANCED SUPERCOMPUTING (NAS) DIVISION

For nearly 25 years, the ***name "NAS" has been associated with leadership and innovation throughout the high-end computing (HEC) community***. They play a significant role in shaping HEC standards and paradigms, and have a leading part in the development of large-scale, single-system image computers. As part of the Exploration Technology Directorate at Ames Research Center, ***NAS division supplies some of the world's most powerful supercomputing resources to NASA and U.S. scientists***. NAS provides an integrated high-end computing environment ***to accelerate NASA missions and make revolutionary advances in science***. The ***245-teraflop Pleiades supercomputer***, will increase the computing capability 2.5 times over the current ***14,336-processor Columbia system*** — one of the fastest operational supercomputers in the world. In addition, the integrated environment includes smaller testbed and next-generation systems, ***high-fidelity modeling and simulation***, high-bandwidth local and wide area networking, ***parallel performance analysis and optimization***, distributed information infrastructure, and ***advanced data analysis and visualization***.

15. Open Science GRID

Open Science Grid	***A national, distributed computed grid for data-intensive research.***

- ***Einstein@Home***, an application that uses spare cycles on volunteers' computers, ***is now running on the OSG.***
- Grid technology enables students and researchers in the petroleum industry.
- **Hadoop** is an open-source ***data processing framework*** that includes a scalable, fault-tolerant distributed file system, HDFS. It is now included in the OSG Virtual Data Toolkit (VDT).
- Superlink-online helps geneologists perform compute-intensive analyses to discover disease-causing anamolies.
- ***The STAR experiment***, running compute-intensive analyses on the OSG to study nuclear decay, has successfully made use of cloud computing as well.April 2009

16. Sloan Digital Sky Sever

17. SEE-GRID

Contractors:

- Greek Research & Technology Networok (GRNET S.A.) - ***Grece***
- European Organization for Nuclear Research – ***CERN***
- **Institute for Parallel Processing – IPP,** Bulgaria
- ***Ruder Boskovic Institute (RBI), Serbia***
- Faculty of Electrical Engineering Banja Luka-UoBL, ***Bosnia and Hercegovina***
- Parallel and Distributed Systems Laboratory – SZTAKI, ***Hungary***
- Academy of Sciences-Institute of Informatics and Applied Mathematics – ASA-INIMA - ***Albania***
- National Institute for Research and Development in Informatics - ICI Bucharest, ***Romania***
- Ss Cyril and Methodius University in Skopje – UKIM, ***Macedonia***
- Research and Educational Netwoking Association of ***Moldova*** – RENAM
- University of ***Montenegro*** - UOM
- ***University of Belgrade – UOB, Serbia***
- The Scientific and Technological Research Council of ***Turkey***

18. Tier2 site in Prague

Collaborating scientists in Prague can now do their analysis at lightning speed, thanks to their new local Tier2 site. For the Prague collaborators to analyze data more efficiently, the datasets from BNL (Brookhaven National Laboratory) needed to be brought onsite at NPI ASCR. ***With the new Tier2 site, NPI ASCR now has 20 terabytes of space to store these datasets***, which can be rotated periodically depending on the researchers' demands and interests.

Researchers at NPI ASCR retrieve the data from BNL via a physical fiber cable running between the two countries that provides ***Ethernet connectivity at 1 Gigabit line***. The Tier2 data transfer framework allows the BNL datasets to be deposited into a ***"Disk Pool Manager,"*** developed by the ***LHC Grid Computing project***, where Prague collaborators can easily access them using tools developed by Open Science Grid.

19. TeraGrid

	TeraGrid is an open scientific discovery infrastructure combining leadership class resources at ***eleven partner sites*** to create an integrated, persistent computational resource.

Using ***high-performance network connections***, the ***TeraGrid integrates high-performance computers, data resources and tools***, and high-end experimental facilities around the country. Currently, ***TeraGrid resources*** include more than ***a petaflop of computing capability*** and more than ***30 petabytes of online and archival data storage***, with rapid access and retrieval over high-performance networks. Researchers can also access more than ***100 discipline-specific databases***. With this combination of resources, the ***TeraGrid is the world's largest, most comprehensive distributed cyberinfrastructure for open scientific research***.

TeraGrid is coordinated through the Grid Infrastructure Group (GIG) at the University of Chicago, working in partnership with the Resource Provider sites: Indiana University, the Louisiana Optical Network Initiative, National Center for Supercomputing Applications, the National Institute for Computational Sciences, Oak Ridge National Laboratory, Pittsburgh Supercomputing Center, Purdue University, San Diego Supercomputer Center, Texas Advanced Computing Center, and University of Chicago/Argonne National Laboratory, and the National Center for Atmospheric Research.

TeraGrid Wide Area Network

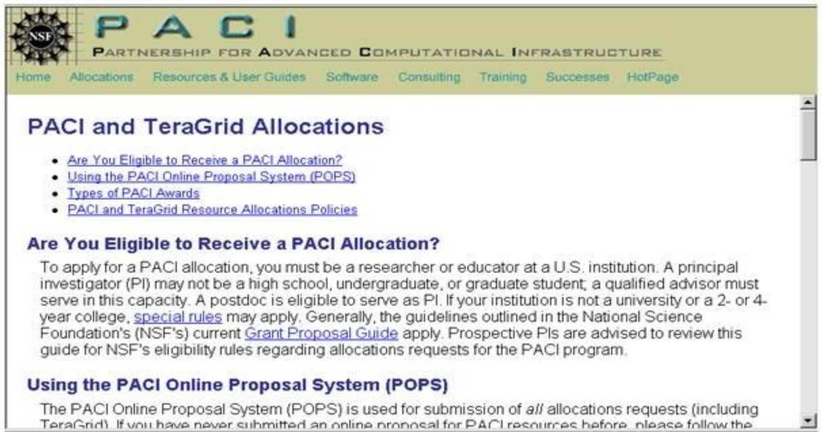
TeraGrid Components

- Compute hardware
  - Intel/Linux Clusters, Alpha SMP clusters, POWER4 cluster, ...
- Large-scale storage systems
  - hundreds of terabytes for secondary storage
- Very high-speed network backbone
  - bandwidth for rich interaction and tight coupling
- Grid middleware
  - Globus, data management, ...
- Next-generation applications

Distributed Parallel Processing

- Decompose application over geographically distributed resources
  - functional or domain decomposition fits well
  - take advantage of load balancing opportunities
  - think about latency impact
- Improved utilization of a many resources
- Flexible job management

Requesting a TeraGrid Allocation



PACI
PARTNERSHIP FOR ADVANCED COMPUTATIONAL INFRASTRUCTURE

Home Allocations Resources & User Guides Software Consulting Training Successes HotPage

PACI and TeraGrid Allocations

- Are You Eligible to Receive a PACI Allocation?
- Using the PACI Online Proposal System (POPS)
- Types of PACI Awards
- PACI and TeraGrid Resource Allocations Policies

Are You Eligible to Receive a PACI Allocation?

To apply for a PACI allocation, you must be a researcher or educator at a U.S. institution. A principal investigator (PI) may not be a high school, undergraduate, or graduate student; a qualified advisor must serve in this capacity. A postdoc is eligible to serve as PI. If your institution is not a university or a 2- or 4-year college, special rules may apply. Generally, the guidelines outlined in the National Science Foundation's (NSF's) current Grant Proposal Guide apply. Prospective PIs are advised to review this guide for NSF's eligibility rules regarding allocations requests for the PACI program.

Using the PACI Online Proposal System (POPS)

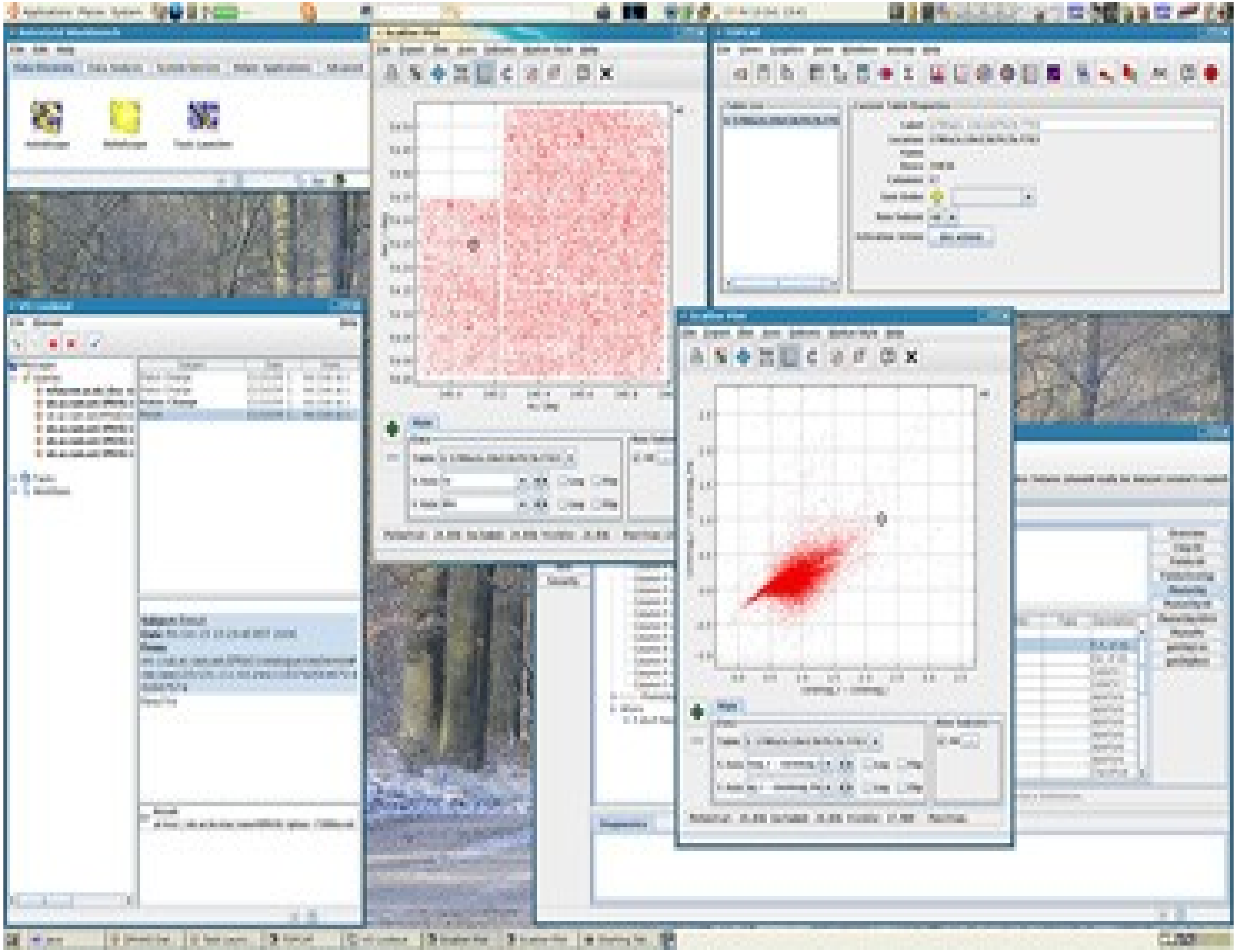
The PACI Online Proposal System (POPS) is used for submission of *all* allocations requests (including TeraGrid). If you have never submitted an online proposal for PACI resources before, please follow the

20. AstroGrid

AstroGrid is the doorway to the Virtual Observatory (VO). It provide a suite of ***desktop applications*** to enable ***astronomers to explore and bookmark resources from around the world, find data, store and share files in VOSpace, query databases, plot and manipulate tables, cross-match catalogues, and build and run scripts to automate sequences of tasks***. Tools from other Euro-VO projects inter-operate with AstroGrid software, so one can also view and analyse images and spectra located in the VO.

AstroGrid, a UK-government funded, open-source project, helps ***create universal access to observational astronomy data*** scattered around the globe. ***The AstroGrid consortium***, which ***consists of 11 UK university groups***, represents astronomy and computing groups with backgrounds in handling and publishing such data. The consortium worked with international partners to agree upon ***standards for published observational astronomy data***, so that all astronomers could interact with all data sets.

The AstroGrid workbench is the main user interface for astronomers accessing the virtual observatory. The global set of standards agreed upon by the consortium and its partners allows any astronomer to query the virtual observatory to ask for information on a certain area of the sky. Through AstroGrid, UK astronomers can also access workflows and applications for data analysis. AstroGrid has also created the “voSpace” program that allows astronomers to share their workflows.



21. AstroGRID-D

22. International Virtual Observatory Alliance

The International Virtual Observatory Alliance (IVOA) was formed in June 2002 with **a mission** to ***"facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory."***

The IVOA now comprises 17 VO projects from Armenia, Australia, Brazil, Canada, China, Europe, France, Germany, Hungary, India, Italy, Japan, Korea, Russia, Spain, the United Kingdom, and the United States.

Membership is open to other national and international projects according to the IVOA Guidelines for Participation.

23. Worldwide LHC Computing Grid

The Worldwide LHC Computing Grid (WLHCG) is ***a global collaboration of more than 170 computing centres in 34 countries***. The mission of the WLHCG project is to build and maintain a data storage and analysis infrastructure for the entire high energy physics community that will use the Large Hadron Collider at CERN. ***The LHC is the largest scientific instrument on the planet***. At full operation intensity, the ***LHC will produce roughly 15 Petabytes (15 million Gigabytes) of data annually***, which thousands of scientists around the world will access and analyse.

Today, the **WLCG** combines the computing resources of ***more than 100,000 processors from over 130 sites in 34 countries***, producing a massive distributed computing infrastructure that provides more than 8,000 physicists around the world with near real-time access to LHC data, and the power to process it.

Why Grid Computing? The answer is ***"money"***... In 1999, the "LHC Computing Grid" was merely a concept on the drawing board for a computing system to store, process and analyse data produced from the Large Hadron Collider at CERN. However when work began for LHC data analysis, it rapidly became clear that the required computing power was far beyond the funding capacity available at CERN.

Additional benefits of a Grid system

- ***Multiple copies of data can be kept in different sites***, ensuring access for all scientists involved, independent of geographical location.

- Allows ***optimum use of spare capacity for multiple computer centres***, making it more efficient.
- Having ***computer centres in multiple time zones*** eases round-the-clock ***monitoring*** and the availability of ***expert support***.
- No single points of failure.
- The cost of maintenance and upgrades is distributed, since ***individual institutes fund local computing resources*** and retain responsibility for these, while still contributing to the global goal.
- Independently managed resources have encouraged novel approaches to computing and analysis.
- So-called ***"brain drain",*** where researchers are forced to leave their country to access resources, ***is reduced when resources are available from their desktop***.
- The system can be ***easily reconfigured*** to face new challenges, making it able to dynamically evolve throughout the life of the LHC, growing in capacity to meet the rising demands as more data is collected each year.
- Provides ***considerable flexibility*** in deciding how and where to provide future computing resources.
- Allows community to take advantage of new technologies that may appear and that offer improved usability, cost effectiveness or energy efficiency. **********

Bulgarian GRID, Bulgarian Virtual observatory and some astronomical applications

Georgi Petrov, Momchil Dechev and Emanouil Atanasov

VII Bulgarian-Serbian Astronomical Conference - ASTROINFORMATICS
01 – 04 June 2010 – Chepelare, BULGARIA

1. Introduction
2. Grid portals, WEB services etc.
3. Middleware, Frameworks etc.
4. Software and standards
5. Modelling, simulations etc.
6. Astronomical applications and projects

MIDDLEWARE AND FRAMEWORKS

1. How to run a million jobs

- **FALKON,** a Fast and Light-weight tasK executiON framework for Clusters, Grids, and Supercomputers – ***AstroPortal***
- **GRACIE:** Grid Resource Virtualization and Customization Infrastructure
- **NIMROD -** A million questions or a few good answers?
- SWIFT

2. IMAGER: A Parallel Interface to Spectral Line Processing

3. CoG Kit

4. gLite ...

5. Globus Toolkit

6. Parallel-Processing Astronomical Image Analysis Tools for HST and SIRTF

7. Grid Observatory

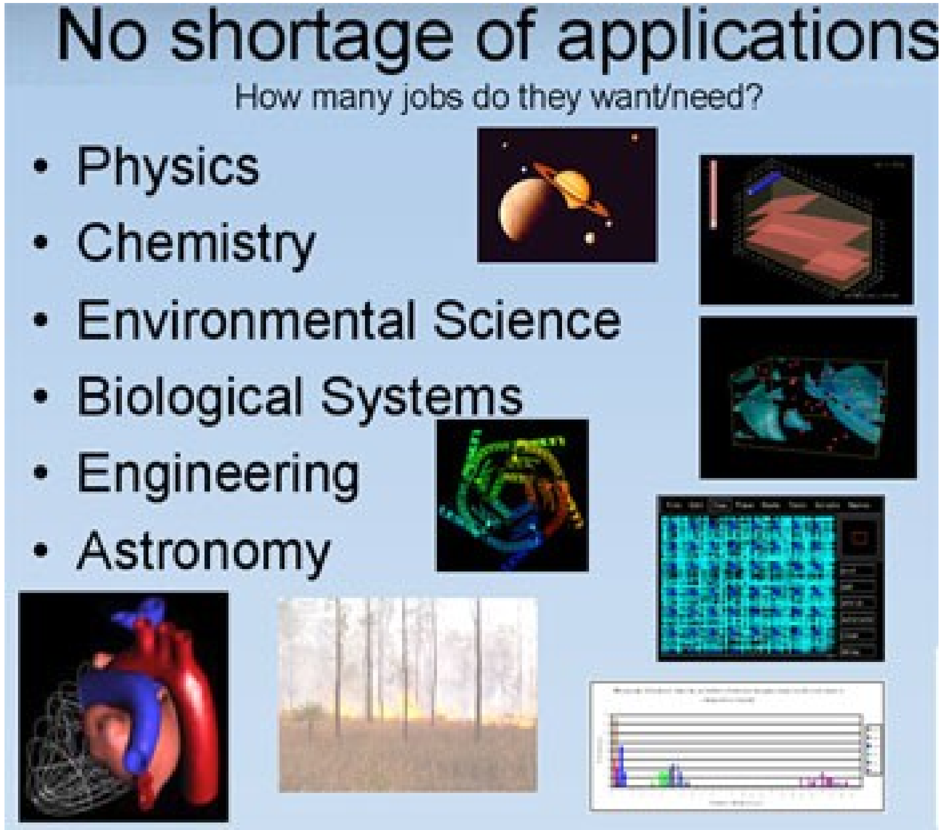
8. SEEGRID Infrastrucure

9. Stellaris

10. Virgo Consortium

11. ProC - The Planck Process Coordinator Workflow Engine

1. How to run a million jobs



As large systems surpass 200,000 processors, more scientists are running "megajobs", thousands to millions of identical or very similar, but independent, jobs executed on separate processors. Some older, well-established job management systems are extremely feature-rich, but their high overhead in scheduling and persistency makes them inefficient for executing many short jobs on many processors. The newer systems work well up to many thousands of jobs, short or long. Still newer ones, like ***Falkon*** and ***Gracie,*** which aim to scale even higher, have yet to achieve wide-scale deployment.

There is a class of applications called ***Many Tasks Computing (MTC)***. An MTC application is composed of many tasks, both independent and dependent, that are (in Foster's words) "***communication-intensive but not***

naturally expressed in Message Passing Interface," referring to a standard for setting up communications between parallel jobs. In contrast to ***h***igh ***t***hroughput ***c***omputing (HTC), MTC uses many computing resources over short periods of time to accomplish many computational tasks. ***Megajobs*** naturally fit in both the HTC and MTC class of applications.

- **FALKON, a Fast and Light-weight tasK executiON framework for Clusters, Grids, and Supercomputers.** Amongst the projects is**...*AstroPortal***

The astronomy community has an abundance of imaging datasets at its disposal which are essentially the "crown jewels" for the astronomy community; ***however the terabytes of data makes the traditional analysis of these datasets a very difficult process.*** Large ***astronomy datasets are generally terabytes in size and contain hundreds of millions of objects separated into millions of files.*** The ***solution is*** to use ***grid computing*** as the main mechanism to enable the dynamic analysis of large astronomy datasets on the ***TeraGrid*** spanning many physical resources.

The key question is: "How can the analysis of large astronomy datasets be made a reality for the astronomy community using Grid resources?" ***The answer*** is: the ***"AstroPortal"***, a ***science gateway to grid resources that is specifically tailored for the astronomy community***. The prototype was implemented as a ***web service*** using the ***Globus Toolkit 4*** (GT4) and it is deployed on the ***TeraGrid***. The astronomy dataset using is the Sloan Digital Sky

Survey ***(SDSS), DR4***, which is comprised of about ***300 million objects*** dispersed ***over 1.3 million files*** adding up to ***3 terabytes of compressed data***.

- ***GRACIE: Grid Resource Virtualization and Customization Infrastructure***

Gracie is a lightweight execution ***framework for efficiently executing massive independent tasks in parallel on distributed computational resources.***

Three optimization strategies have been devised to improve the performance of Grid system.

- Pack up to ***thousands of tasks into one request***.
- Share the effort in resource discovery and allocation among requests by ***separating resource allocations from request submissions***.
- Pack ***variable numbers of tasks into different requests***, where the task number is a function of the destination resource's computability.

Gracie is a computational grid software platform developed by ***Peking University***.

- **NIMROD** - A million questions or a few good answers?

The ***tool set***, called ***Nimrod***, automates the process of ***finding good solutions to demanding computational experiments***. Nimrod includes tools that perform a complete parameter sweep across all possible combinations :

Tool **Purpose**

- Nimrod/G provides two services: Parameter sweeps and ***grid***/cloud execution tools including scheduling across multiple compute resources.
- Nimrod/O provides an ***optimisation*** framework for optimising a target output value of an application. Used with Nimrod/G, it can exploit parallelism in the search algorithm.
- Nimrod/OI provides an ***interactive*** interface for Nimrod/O. In some applications, it might require someone to decide which output is better. Those results are fed back into Nimrod/O to produce more suggestions.
- Nimrod/E provides experimental design techniques for analysing parameter effects on an application's output. Used with Nimrod/G allows the ***experiment*** to be scaled up on grid and cloud resources.
- Nimrod/K provides all the Nimrod tools in a workflow engine called ***Kepler***. Nimrod/K adds all the parameter tools and grid/cloud services to Kepler while leveraging and enhancing all the existing grid tools already provided by adding dynamic parallelism in workflows.
- **SWIFT**

Swift, a highly ***scalable scripting language/engine to manage procedures*** composed of many loosely-coupled components that take the place of megajobs. It also ***throttles job submission*** as needed, and ***controls file transfers*** to ensure adequate performance.

2. IMAGER: *A Parallel Interface to Spectral Line Processing*

IMAGER is an ***interface to parallel implementation of imaging and deconvolution tasks*** of the Software Development Environment (SDE) of the NRAO. The interface is based on the ***MIRIAD interface*** of the ***BIMA*** (Berkeley-Illinois-Maryland Association) ***array*** and it allows for interactive and batch operation. ***The parallelization is carried out by distributing independent spectral line channels across multiple processors.***

Radio synthesis data reduction has been one of the most computer intensive operations in observational astronomy. ***In*** the common ***case of radio spectral line observations, large numbers of frequency channels lead to large amounts of data***. It is not uncommon with such instruments such as the VLA and the BIMA telescopes to have spectral line ***data sets in excess of a gigabyte***. ***Astronomers need access to fast processing to allow the analysis of such large data sets***. It is especially important to have analysis capabilities that allow astronomers to use different methods of ***non-linear deconvolutions*** in a timely basis to properly interpret their observations. ***The analysis of spectral line data, in which each channel is independent from every other channel, is an embarrassingly parallel problem.***

The IMAGER system has been used by astronomers at the University of Illinois to carry out analysis of data from the VLA and BIMA telescopes. The IMAGER package is currently supported on the SGI Power Challenge array at NCSA.

3. CoG Kit

Commodity Grid (CoG) Kits allow Grid users, Grid application developers, and Grid administrators to ***use, program, and administer Grids from a higher-level framework***. The ***Java and Python CoG Kits*** are good examples. These kits allow for easy and rapid Grid application development. It is fact that ***CoG Kits are also used within the Globus Toolkit***, and provide important functionality.

4. gLite

gLite (pronounced "gee-lite") is the next generation ***middleware for grid computing***. Born from the collaborative efforts of more than 80 people in 12 different academic and industrial research centers as part of the EGEE Project, *gLite* ***provides a framework for building grid applications tapping into the power of distributed computing and storage resources across the Internet***.

The gLite distribution is an integrated set of components designed to enable resource sharing. In other words, ***this is middleware for building a grid***.

The gLite middleware is produced by the EGEE project. The distribution model is to construct different services ('node-types') from these components and then ensure easy installation and configuration on the chosen platforms (currently ***Scientific Linux*** versions 4 and 5, and also ***Debian 4*** for the WNs).

gLite middleware is currently deployed on hundreds of sites as part of the EGEE project.

5. Globus Toolkit

The ***open source Globus Toolkit*** is a fundamental enabling technology for the "Grid," letting people share computing power, databases, and other tools securely online across corporate, institutional, and geographic boundaries without sacrificing local autonomy. The toolkit ***includes software services*** and l***ibraries*** for resource monitoring, discovery, and management, plus ***security and file management***. In addition to being a central part of science and engineering projects that manages total nearly a half-billion dollars internationally .

The Globus Toolkit has grown through an open-source strategy similar to the Linux operating system's, and distinct from proprietary attempts at resource-sharing software.

6. Parallel-Processing Astronomical Image Analysis Tools ***for HST and SIRTF***

NASA applied information system research develop and implement several ***parallel-processing astronomical image-analysis tools for stellar imaging data from the HST and the Space Infrared Telescope Facility***. The project combines the enabling image-processing technology of the ***new digital PSF-fitting MATPHOT algorithm for accurate and precise CCD stellar photometry*** with enabling technology of Beowulf clusters which offer excellent cost/performance ratios for computational power. ***Data mining tools to do quick-look stellar photometry and other scientific visualization tasks*** will also be written and used in order to investigate how such tools could be used at the data servers of NASA archival imaging data like the Space Telescope Science Institute.

7. Grid Observatory

The ***Grid Observatory is an open project***, keen to work with computing researchers. The EGEE grid offers an unprecedented opportunity to observe, and start understanding the new computing practices of e-science. With ***more than 40000 CPUs and 5PB of storage*** distributed worldwide, ***the management of 100.000 concurrent jobs***, and the perspective of a sustainable development, the EGEE grid is one of the more exciting artificial complex systems to observe.

The Grid Observatory collects, publishes, and analyzes, data on the behaviour of the EGEE grid. The ultimate goal of the Grid Observatory is to integrate data collection, data analysis, and the development of models and of an ontology for the domain knowledge. The Grid Observatory is part of the EGEE-III project. Because grid data and models are equally relevant for computer science, middleware development and system administration, the Grid Observatory is an open project.

Currently, the Grid Observatory provides only traces of the EGEE grid; it can be extended in the future to traces of other grids.

8. SEEGRID Infrastrucure – *South-Eastern European GRID*

Monitoring and Operational tools

SAM (Service Availability Monitoring)	https://c01.grid.etfbl.net/bbmsam/
GStat	http://goc.grid.sinica.edu.tw/gstat/seegrid/
GridIce	http://grid-se.ii.edu.mk/gridice/site/site.php
Accounting portal	**http://gserv4.ipp.acad.bg:8080/AccountingPortal**
Googlemap	http://www.grid.org.tr/eng/
Site map	http://see-grid.inima.al/see-grid-weather/
MonAlisa	http://monitor.seegrid.grid.pub.ro:8080
RTM (Real Time Monitor)	http://gridportal.hep.ph.ic.ac.uk/rtm/applet.html
Helpdesk	http://helpdesk.see-grid.eu/
HGSM (Hierachical Grid Site Management)	https://hgsm.grid.org.tr/
WatG Browser (What is at the Grid Browser)	http://watgbrowser.scl.rs:8080/
Nagios (Monitoring tool)	**https://portal.ipp.acad.bg:7443/seegridnagios/**

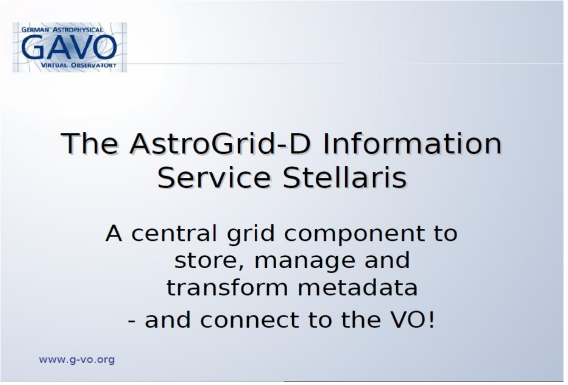
Core Services

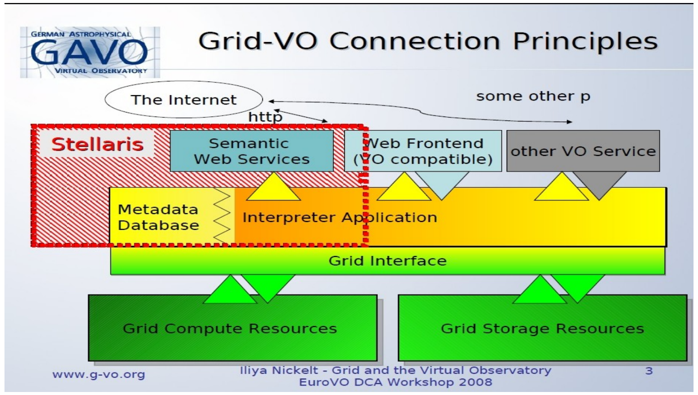
Service	Primary	Secondary
BDII	bdii.ipb.ac.rs	bdii.ulakbim.gov.tr
RB	rb.ipb.ac.rs	rb.ulakbim.gov.tr
WMS	wms.ipb.ac.rs	wms.ulakbim.gov.tr
VOMS	voms.irb.hr	voms.grid.auth.gr
LFC	**grid02.rcub.bg.ac.rs**	lfc.ipb.ac.rs
FTS	**grid16.rcub.bg.ac.rs**	
AMGA	**grid16.rcub.bg.ac.rs**	
MyProxy	myproxy.grid.auth.gr	myproxy.ipb.ac.rs
RGMA Registry and Schema	**gserv1.ipp.acad.bg**	

9. Stellaris - Enabling flexible metadata management for the Grid.

Stellaris is a metadata management service developed within the **AstroGrid-D project.** The aim is to ***provide a flexible way to store and query metadata*** relevant for e-science and grid-computing. This can range from resource description of grid resources (compute clusters, robotic telescopes, etc.) to application specific job metadata or dataset annotations. One can use common web-standards such as RDF (Resource Description Framework) to describe metadata and the accompanying query language SPARQL. ***Some features of the software*** include:

- A simple but powerful management interface for RDF-graphs
- Different backends for indexing through the use of RDFLib and Virtuoso
- Authentication using X.509-certificate verification
- Group-based authorization system
- SPARQL-protocol implementation with both XML/JSON result formats
- Graph lifetime management
- Stand-alone deployment or embeddable in Apache using WSGI





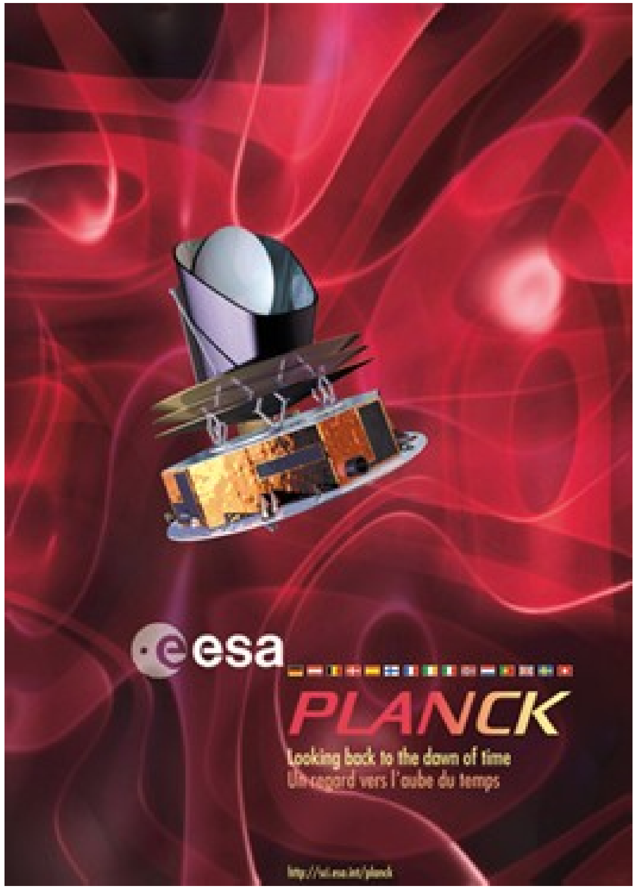
10. Virgo Consortium

The ***Virgo Consortium for Cosmological Supercomputer Simulations*** was founded in 1994 in response to the UK's ***High Performance Computing Initiative***. The Virgo Consortium has a core membership of about a dozen scientists in the UK, Germany, Netherlands, Canada the USA and Japan. The largest nodes are the Institute for Computational Cosmology in Durham, ***UK*** and the Max Planck Institute for Astrophysics in Garching, ***Germany***. Other nodes exist in Cambridge, Edinburgh, Nottingham and Sussex in the UK, Leiden in the ***Netherlands,*** McMaster and Queen's Universities in ***Canada***, Pittsburgh University in the ***USA*** and Nagoya University in ***Japan***. At any given time, an additional 20-25 scientists, mostly PhD students and postdocs, are directly involved in aspects of the Virgo programme.

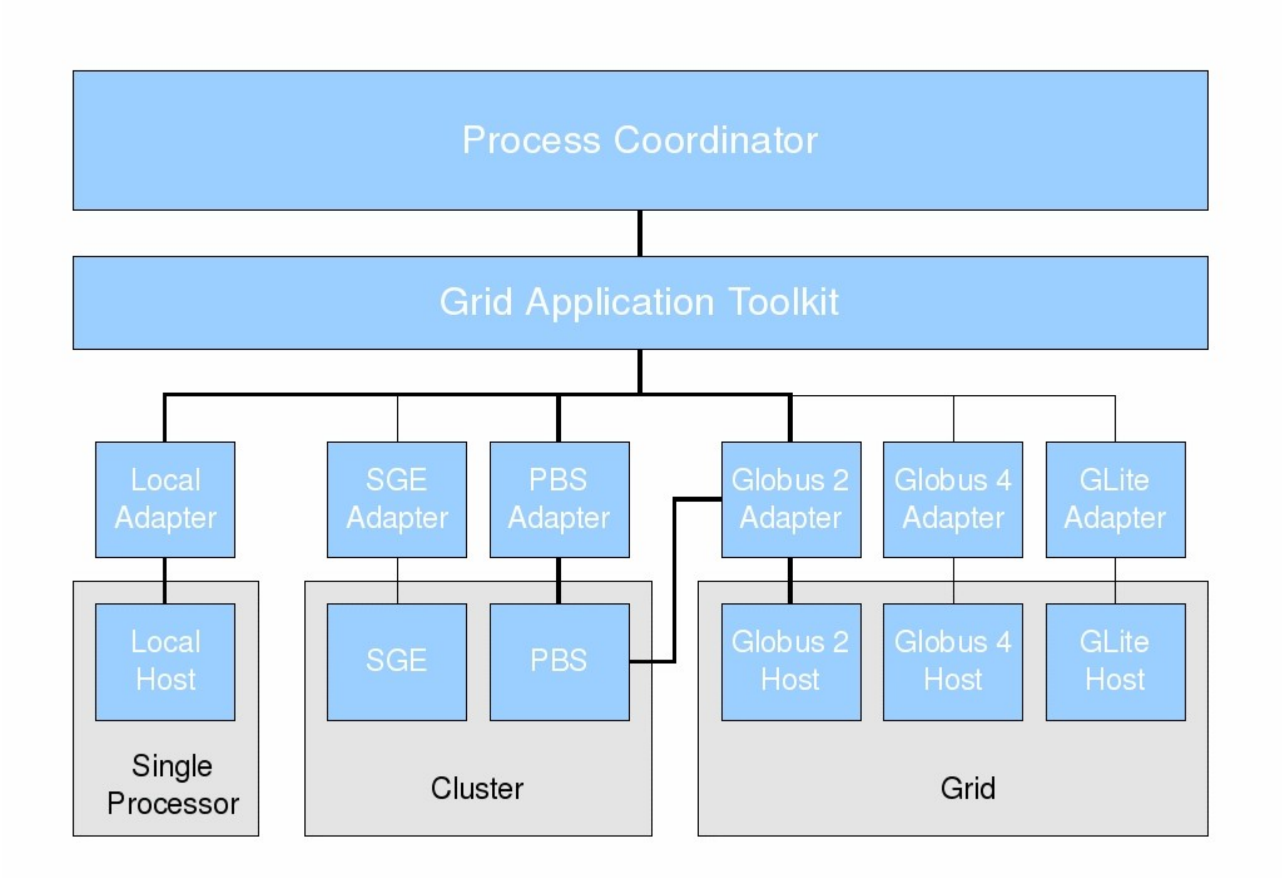
Science goals: The science goals of Virgo are ***to carry out state-of-the-art cosmological simulations***. The ***research areas*** include the l***arge-scale distribution of dark matter***, the ***formation of dark matter haloes***, the ***formation and evolution of galaxies and clusters***, the ***physics of the intergalactic medium*** and the ***properties of the intracluster gas***.

Virgo's current resources: Virgo has access to world class supercomputing resources in the UK and Germany, most notably the ***"Cosmology Machine"*** at Durham, which has a total of ***792 opteron cpus*** and more than ***500 ultra-sparcIII processors***, and the ***IBM Regatta system*** with ***816 power-4 processors*** at the Max-Planck Rechenzentrum in Garching. The ***main production codes*** are ***GADGET*** and ***MPI-HYDRA***.

11. *ProC* - *The Planck Process Coordinator Workflow Engine*



The vast amount of data produced by satellite missions is a challenge for any data reduction software in terms of complex job submission and data management. The demands on computational power and memory space qualify ***the satellite data reduction to be prototype of grid applications***. Therefore, the Planck Process Coordinator Workflow Engine ***ProC*** for the Planck Survey or satellite has been chosen as a ***Grid Use Case within the AstroGrid-D project***. The ***ProC is 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.



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.

****

4. SOFTWARE AND STANDARDS

1. AMEEPAR - *Parallel processing for hyperspectral imaging*

The wealth of spatial and spectral information provided by ***hyperspectral sensors (with hundreds or even thousands of spectral channels)*** has quickly introduced new processing challenges. In particular, ***many hyperspectral imaging applications require a response in (near) real time*** in areas such as environmental modeling and assessment, target detection for military and homeland defense/security purposes, and risk prevention and response.

At the time being only a few parallel processing algorithms exist in the open literature.

To address the need for integrated software/hardware solutions in hyperspectral imaging, a highly innovative processing algorithms on several types of parallel platforms, including commodity (Beowulf-type) clusters of computers, large-scale distributed systems made up of heterogeneous computing resources, and specialized hardware architectures is developed.

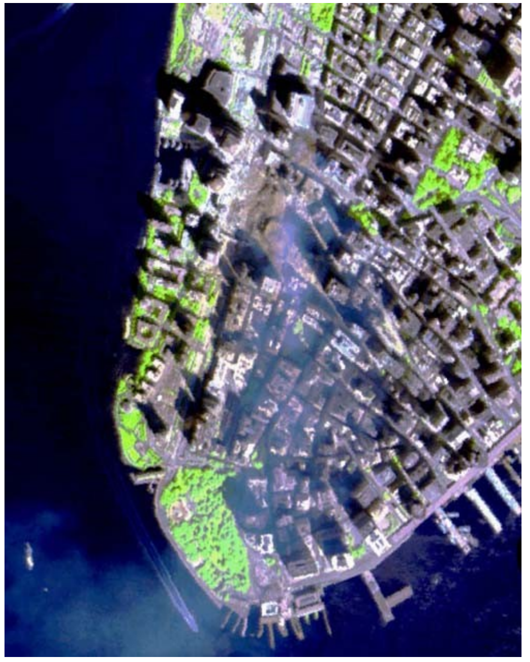
Several parallel algorithms to analyze the AVIRIS data were implemented.

- The battery of algorithms consisted of various target detection algorithms, such as the ***parallel automated target generation algorithm (P-ATGP)***
- Parallel classification algorithms based on the identification of pure spectral components , such as the ***fast***

pixel purity index (P-FPPI)

- ***The automated morphological extraction (AMEEPAR)***. This is one of the few available parallel algorithms that ***integrate spatial and spectral information***. Using ***256 processors***, AMEEPAR provided a 90% accurate debris/dust map of the full AVIRIS data in 10s, while the P-ATGP algorithm was able to detect the spatial location of thermal hot spots in the WTC area in only 3s.

On the Figure is ***hyperspectral image*** collected by the NASA Jet Propulsion Laboratory's AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) system ***over the World Trade Center*** (WTC) area in New York City on September 16, 2001. The data comprises ***614 samples, 3675 lines, and 224 spectral bands, for a total size of 964MB***. Figure shows a false-color composite of a portion of the scene, in which the spectral channels at 1682, 1107, and 655nm are displayed as red, green, and blue respectively. Here, vegetation appears green, burned areas appear dark gray, and smoke appears bright blue due to high spectral reflectance in the 655nm channel.



2. GADGET-2 – *a code for cosmological simulations of structure formation.*

Gadget **i**s a freely available ***code for cosmological N-body/SPH simulations*** on massively ***parallel computers*** with distributed memory. ***The code can be run on essentially all supercomputer systems*** presently in use, including ***clusters of workstations*** or ***individual PCs***.

GADGET ***computes gravitational forces with a hierarchical tree algorithm*** and represents fluids by means of smoothed particle hydrodynamics (SPH). **GADGET** follows the evolution of a self-gravitating collisionless N-body system, and allows gas dynamics to be optionally included. **GADGET** can therefore be used to address ***a wide array of astrophysically interesting problems***, ranging from ***colliding and merging galaxies***, to the ***formation of large-scale structure in the Universe and*** can also be used to ***study the dynamics of the gaseous intergalactic medium***, or to address ***star formation*** and its regulation by ***feedback processes***.

GADGET comes with a number of small examples that can be run to develop a feel for working with the simulation code. Here are the initial conditions for the following systems:

- ***A pair of colliding disk galaxies (collisionless)***
- ***A spherical collapse of a self-gravitating sphere of gas***
- ***Cosmological formation of a cluster of galaxies (collisionless, vacuum boundaries)***
- ***Cosmological structure formation in a periodic box with adiabatic gas physics***

3. CRBLASTER ***: a fast parallel-processing program for cosmic ray rejection***

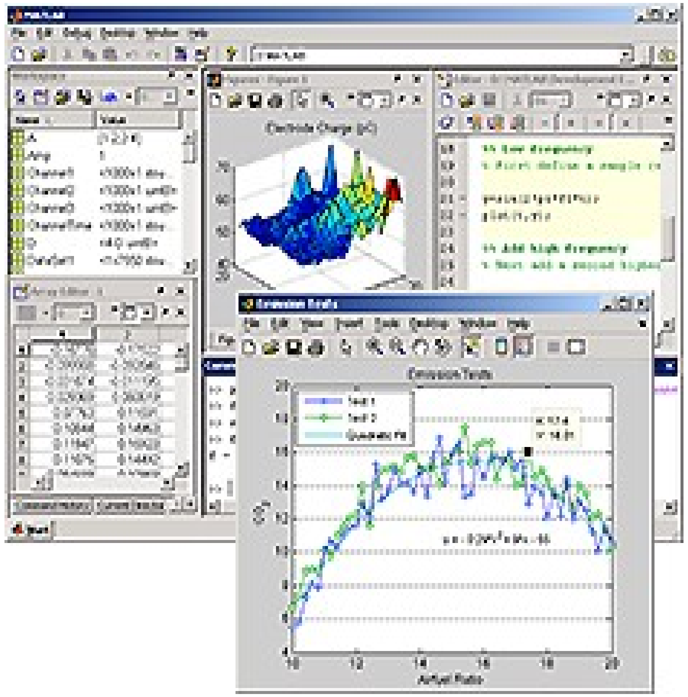
Many astronomical image-analysis programs are based on algorithms that can be described as being ***embarrassingly parallel***, where ***the analysis of one subimage*** generally ***does not affect the analysis of another subimage***. Yet ***few parallel-processing astrophysical image-analysis programs exis***t that can easily take full advantage of todays fast multi-core servers costing a few thousands of dollars. A major reason for the shortage of state-of-the-art parallel-processing astrophysical image-analysis codes is that the ***writing of parallel codes has been perceived to be difficult***.

CRBLASTER - a new fast ***parallel-processing image-analysis program does cosmic ray rejection*** uses van Dokknm's L.A.Cosmic algorithm**.** CRBLASTER is written in C using the industry standard Message Passing Interface (MPI) library. Processing a single 800×800 HST WFPC2 image takes 1.87 seconds using 4 processes on an Apple Xserve with two dual-core 3.0-GHz Intel Xeons; the efficiency of the program running with the 4 processors is 82%.

The code can be used as a software framework for easy development of parallel-processing image-anlaysis programs using embarrassing parallel algorithms: the biggest required modification is the replacement of the core image processing function with an alternative image-analysis function based on a single-processor algorithm.

4. MATLAB – more possibilities

MATLAB* can now run on Enabling Grids for E-sciencE (EGEE) computing power.** Widely regarded as ***a powerful piece of simulation software, for use in everything from optimizing rocket launch control settings to vector analysis, ***it is now fully compatible with any grid computing system using gLite middleware***.



MATLAB® is a high-level language and interactive environment that enables you to perform computationally intensive tasks faster than with traditional programming languages such as C, C++, and Fortran.

- Introduction and Key Features
- Developing Algorithms and Applications
- Analyzing and Accessing Data
- Visualizing Data
- Performing Numeric Computation
- Publishing Results and Deploying Applications

For now – only two examples:

Finance – University of Athens and **Making better lasers –** University of Bristol.

5. N_body-sh1p - *a parallel direct N_body code*

Educational N-body integrator with a shared but variable time step (the same for all particles but changing in time), using the Hermite integration scheme. The source code has been ***adapted for a parallel ring algorithm*** using the MPI library.

Typical command line (generates : n24body.out)

% nbody_sh1p < n24body.in > n24body.out

Small ***timing test*** (perfomed by A. Gualandris) for 128, 256 and 512 particles ***with up to 32 processors*** on the ***Blue (Boewulf) linux cluster*** at SARA.

6. VO-Software

Available VO-compatible applications for the immediate use to do science. The level of maturity of the applications depends on a high degree on the level of maturity of the corresponding IVOA protocols and standards. Care must be taken when using them for publications.

Application / Version (in alphabetical order)
Aladin v6.011a (January 2010)
Datascope v2.1 (March 2007)
Montage
Octet
Open SkyQuery
SkyView
Specview 2.14.4 (August 2009)
SPLAT 3.9.0 (May 2009)
TOPCAT/STILTS 3.5-1/2.1-1 (December 2009/December 2009)
VisIVO 1.5.7.1 (May 2009)
VOConvert 1.0 (June 2006)
VODesktop 1.3 (June 2009)
VOEventNet
VOPlot 1.5 (May 2009)
VOStat 1.1 (November 2008)
VOSA 1.0.2 (March 2009)
VOSED 1.3 (July 2009)
VOServices (Footprint, Spectrum, Filters, ...) 2.1.0.0
VOSpec V5.5 (September 2009)
WCSFixer

Functionality
Search for Images: Aladin, Datascope, SkyView, VODesktop
Search for Spectra: Aladin, Datascope, SPLAT, Specview, VOServices, VOSpec
Search for Catalogues: Aladin, Datascope, TOPCAT, VODesktop
Image visualisation: Aladin, SkyView
Spectra visualisation: SPLAT, Specview, VOServices, VOSpec
Catalogues visualisation: Aladin, TOPCAT, VOPlot
Cross-correlation: Aladin, Open SkyQuery, STILTS, TOPCAT
Scatter, 3D plots and histograms: TOPCAT, VOPlot
Statistics: VOStat
Footprint Service: Aladin, VOServices
Table format conversion: TOPCAT, VOConvert
Filter curves: VOServices
SED building: VOSA, VOSED, VOSpec
Fixing WCS: Aladin, WCSFixer

Other VO-compliant tools
DS9: Image visualisation
GOSSIP: SED fitting
Mirage: Table visualisation
VirGO: Search for Images and Spectra

Browse the Registries
EURO-VO Registry
AstroGrid Registry
NVO Registry

Manuals, Tutorials, How-tos
Aladin User manual
Datascope how to
Montage help
Open SkyQuery help
SkyView documentation
Specview examples
SPLAT documentation
STILTS documentation
TOPCAT documentation
VisIVO how to
VODesktop how to
VOSpec User manual

7. Experience WorldWide Telescope

WorldWide Telescope (WWT) enables your computer to function as a virtual telescope, bringing together imagery from the best ground and space-based telescopes in the world. Experience narrated guided tours from astronomers and educators featuring interesting places in the sky.

A web-based version of WorldWide Telescope is also available. This version enables seamless, guided explorations of the universe from within a web browser on PC and Intel Mac OS X by using the power of Microsoft Silverlight 3.0.

What is WorldWide Telescope?

WWT is an ***application that runs in Windows*** that ***utilizes images and data*** stored on remote servers ***enabling you to explore some of the highest resolution imagery of the universe available in multiple wavelengths***.

8. VirGO

VirGO is the next generation ***Visual Browser for the ESO Science Archive Facility*** developed by the VO Systems Department. ***It is a plug-in for the popular open source software*** Stellarium ***with added capabilities for browsing professional astronomical data.***

VirGO gives astronomers the possibility ***to*** easily ***discover and select data*** from millions of observations in a new visual and intuitive way. Its main feature is to perform ***real-time access*** and graphical display of a large number of observations by showing instrumental footprints and ***image previews***, and to allow their ***selection*** and ***filtering*** for subsequent retrieval. It reads FITS images and catalogues in VOTable format. It ***superimposes DSS background images*** and allows to view the sky in a *real life* mode as seen from the main ESO sites. Data interfaces are based on V***irtual Observatory standards*** enabling access to images and spectra hosted by other data centers and to exchange data with other VO applications through the PLASTIC messaging system.

VirGO-1.4.4 (Sept 09th 2009) is ***distributed as a binary compiled for linux-i386 and windows and MacOSX***. The package contains a binary version of Stellarium 0.10.3, the VirGO plug-in for ESO archive access and some extra star catalogs and landscapes.

Stellarium/VirGO is distributed under the GNU General Public License (GPL).

9. Message Passing Interface Standard ***(MPI)***

The Message Passing Interface Standard (MPI) is a message passing ***library standard*** based on the consensus of over 40 participating organizations, including vendors, researchers, software library developers, and users. ***The goal*** of the MPI i***s to establish a portable, efficient, and flexible standard for message passing*** that will be widely used for ***writing message passing programs***. The advantages of developing message passing software using MPI closely match the design goals of portability, efficiency, and flexibility. ***MPI*** is not an IEEE or ISO standard, but has ***in fact, become the "industry standard"*** for writing message passing programs on HPC platforms.

Reasons for Using MPI:

- **Standardization** - MPI is the only message passing library which can be considered a standard. It is supported on virtually all HPC platforms. Practically, it has replaced all previous message passing libraries.
- **Portability** - There is no need to modify your source code when you port your application to a different platform that supports (and is compliant with) the MPI standard.
- **Performance Opportunities** - Vendor implementations should be able to exploit native hardware features to optimize performance. For more information about MPI performance see the MPI Performance Topics tutorial.
- **Functionality** - Over 115 routines are defined in MPI-1 alone.
- **Availability** - A variety of implementations are available, both vendor and public domain.

10. Resource Description Framework ***(RDF)***

RDF is a standard model for data interchange on the Web. ***RDF*** has features that ***facilitate data merging even if the underlying schemas differ***, and it specifically ***supports the evolution of schemas over time*** without requiring all the data consumers to be changed.

RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple"). Using this simple model, ***it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.***

Here is ***an extraction from a list of almost all tools , that are marked as relevant to RDF***.

- 3Store (triple store).
- 4Suite (programming environment). Directly usable from Python
- 4store (triple store).
- ARC RDF Store (triple store). Directly usable from PHP
- ActiveRDF (programming environment). Directly usable from Ruby
- Allegro Graph RDF Store (triple store, programming environment, reasoner, development environment). Directly usable from Java, LISP, Python, Prolog

... and many others...

****

5. MODELLING AND SIMULATIONS

1. A star is born - thanks to supersonic turbulence

Using ***the largest simulation of supersonic turbulence*** to date, UC San Diego researchers have shown ***how fundamental laws of turbulent geophysical flows can also be extended to supersonic turbulence in the interstellar medium of galaxies***. The image shows the density field from one snapshot of the simulation, ***run on 4,096 processors for two weeks and resulting in 25 terabytes of data***. The brightest regions in the image represent ***gas at the highest density***. ***Dense filaments*** and ***cores***, created in such a way by supersonic turbulent flows, are subject to ***massive gravitational collapse*** – and that ***leads to the birth of stars***.

2. Astrophysical Thermonuclear Flashes

Advanced Simulation and Computing (ASC), Academic Strategic Alliances Program (ASAP) Center

The "FLASH Center" - is funded by the DOE ASC/Alliance Program to build ***a state-of-the-art simulator code for solving nuclear astrophysical problems related to exploding stars***. The website contains information about the astrophysics, the code, and related basic physics and computer science efforts.

FLASH Center scientists simulate a successful, fully-3D type Ia supernova explosion for the first time!

FLASH3.2 was released on July 2nd, 2009!

The ASC/Alliances Center for Astrophysical Thermonuclear Flashes at the University of Chicago ***runs simulations to solve the problem of thermonuclear explosions on the surfaces of compact stars***. Their simulations of ***Type Ia supernovae, exploding white dwarf stars***, have shown that ***an internal flame 'bubble' emerges at a point on the stellar surface, leading to surface waves that converge at the opposite point, and causing a shock and subsequent detonation of the entire star***. Previously, scientists thought that the original flame would directly transition to a detonation. Based only on well-known physical processes, these simulations exemplify the potential of numerical simulations for scientific discovery.

3. Cosmic simulation

Cosmic structure formation theory has passed test after test, ***predicting how many galaxies will form***, ***where they will form***, and ***what type of galaxy they will be***. But for almost 20 years, its predictions about the central mass of dwarf galaxies have been wrong. Worldwide, there are many teams working on their own versions; each attacks the problem from a different angle.

- E.g. ***Governato et al says:*** "Potentially, ***this is a very big problem for the model***. It might imply that ***the dark matter particle that we think is the correct one is not the correct one***, or ***maybe*** that ***gravity works differently*** than we think it does. So ***this is a very fundamental problem for physics.***"

A more realistic model of how stars form and die, incorporated into the existing cosmic structure formation theory were used. It turns out that ***when a star near the galactic center explodes, a lot of interstellar gas is blown away from the center of the galaxy***. As a result, ***less stars form at the center, because there is less gas***.

To create the simulation about ***a million computer hours were used***, which means that it would have taken close to a hundred years to run the same simulation on the average desktop. ***A simulation was running on computer resources at NASA*** Advanced Supercomputing Division, the Arctic Region Supercomputing Center, ***and TeraGrid.***

- ***Klypin's team*** is exploring ***the large-scale effects of energy released by young stars***.

Stars are forming, and ***young stars release large amounts of energy into the gas that surrounds them***. That energy finds its way to larger scales, ***affecting the motion of gas in the whole galaxy*** – even the way it is being accreted in the galaxy.

Over time, scientific understanding of processes such as ***star formation*** has evolved, ***yielding new equations***. The equations can ***in tur***n be ***used to refine the computational model***.

4. Flip-flopping of black hole accretion disks

The accretion disk of a black hole forms from gas attracted by the black hole's massive gravitational pull. For the last 20 years, astrophysicists have debated ***whether the whirlpool-like motion of the accretion disk will periodically reverse motion, a behavior called 'flip-flop'? According to*** a new ***simulation*** powered by ***TeraGrid***, the ***<u>whirlpools of gas flip-flop as they are sucked into black holes.</u>***

When ***flip-flopping first turned up in a 1988 numerical simulation***, some scientists argued that it explains recurrent x-ray flares observed by the European X-Ray Observatory in 1985. But in subsequent years, although ***some simulations showed flip-flop, others did not***, casting doubt on the existence of the phenomenon. The earlier work was criticized for a wide variety of reasons, but the chief among them was the lack of computer power and hence accuracy of the computation.

<u>Rem.:</u> The most basic form of the equation used in the simulation was originally formulated by Fred Hoyle and Ray Lyttleton in 1939.

The simulation found that ***the accretion disk reversed direction repeatedly***, confirming that ***at least in this model of black hole accretion disks, flip-flop does occur***.

5. Millennium Simulation Project

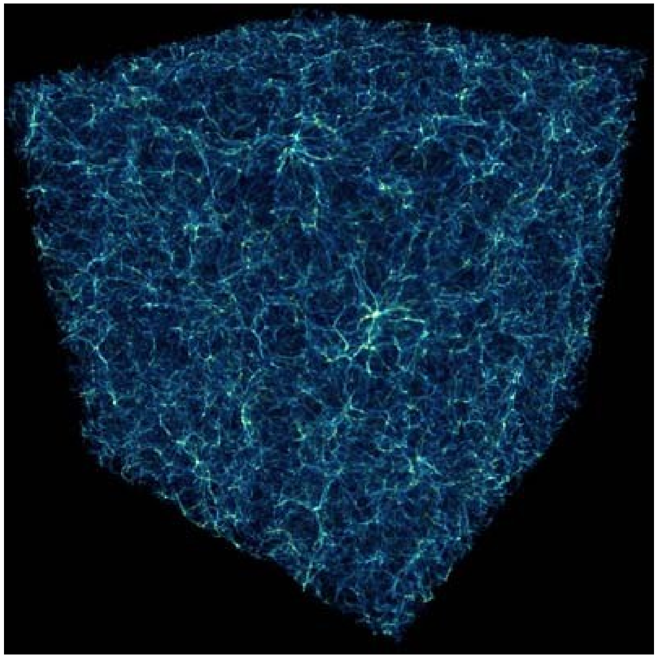
The Millenium Simulation Project is helping ***to clarify the physical processes underlying the buildup of real galaxies and black holes***. It has ***traced the evolution of the matter distribution***. The *Millennium Run* used more than ***10 billion particles*** to trace t***he evolution of the matter distribution in a cubic region of the Universe over 2 billion light-years*** on a side. ***It kept busy the principal supercomputer at the Max Planck Society's Supercomputing Centre in Garching, Germany for more than a month***. By applying sophisticated modelling techniques to the 25 Tbytes of stored output, ***Virgo scientists*** have been able to ***recreate evolutionary histories both for the 20 million or so galaxies*** which populate this enormous volume ***and*** for the ***supermassive black hole***s which occasionally power quasars at their hearts. By comparing such simulated data to large observational surveys, one can clarify ***the physical processes underlying the buildup of real galaxies and black holes***. Amongst them are:

- ***A journey through the simulated universe.***
- ***The dark matter distribution in the universe at the present time***
- ***The galaxy distribution in the simulation***
    - ***on very large scales***
    - ***for a rich cluster of galaxies***
- ***Slices of the dark matter distribution***
- ***Halo and semi-analytic galaxy catalogues***

How did the universe evolve into the structure we know? The very early universe consisted of homogeneous gas with tiny perturbations. As the gas cooled over time, it collapsed under gravity into clumps and then galaxies.

The researchers ran the largest detailed simulation of a cosmological structure to date. ***In the simulation***, ***the region of study collapses from about 2 billion light years across to form a region of galaxy clusters only 25,000 light years across***.

The ***distribution of galaxy clusters*** in the universe can actually help us to learn things about ***dark energy***, ***how much matter there is*** in the universe, and ***how fast the universe is expanding***...



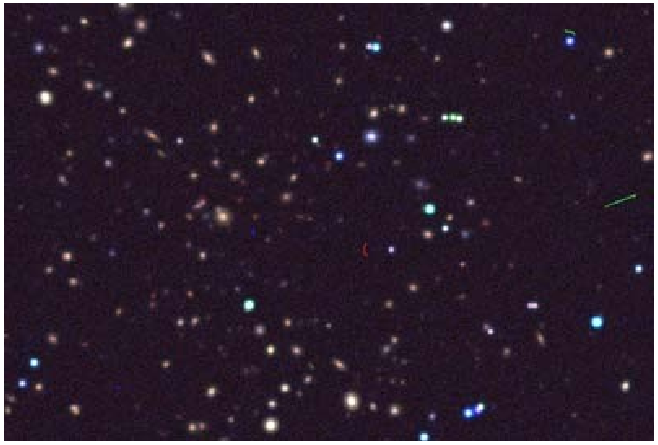
The filaments indicate "***warm-hot intergalactic medium***", or WHIM. WHIM constitutes about half of the universe's non-dark matter, yet we cannot see it very well. It emits and absorbs largely in the UV and soft X-ray portion of the electromagnetic spectrum, much of which is blocked by the earth's atmosphere.

The knot-like structures at the intersections indicate ***large groups and clusters of galaxies -*** important objects to study for understanding the fundamental properties of our universe such as the amount of matter, dark energy, and the expansion rate. The ***largest knot***, near the center, is a ***galaxy cluster***.

6. Dark Energy Survey - *Simulating starry images*

To better understand dark energy and its implications on our current knowledge of matter, energy, space, and time, scientists will conduct the ***large-scale Dark Energy Survey (DES), starting in 2012.***

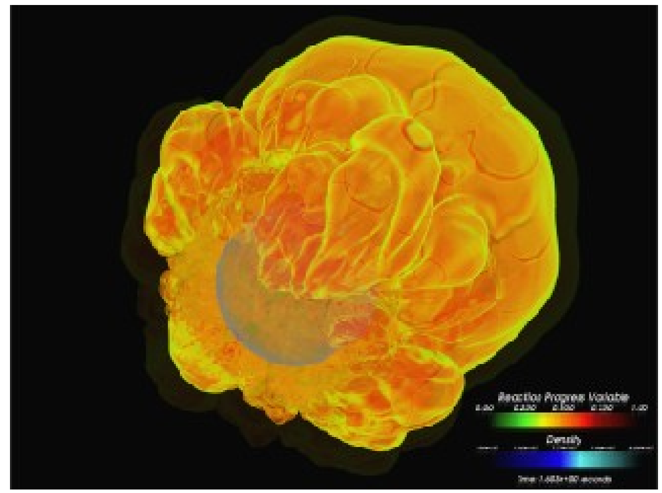
At the Cerro Tololo Inter-American Observatory in Chile, researchers will use the ***4-meter Blanco telescope,*** equipped with the ***Dark Energy Camera***, to capture brilliant ***images of more than 300 million galaxies.*** They expect to measure quantities related to pressure and energy density five times more precisely than currently possible.

Ever ***since the universe exploded*** into existence, ***it has been violently rushing outward***. Scientists ***expected*** the ***inward tug of gravity to slow this expansion over time, but the opposite is true***. The startling discovery that t***he universe's expansion is accelerating*** has led scientists to postulate ***the existence of an outward-pushing dark energy***.	

To test and debug the image processing programs, researchers use Open Science Grid to create complex simulations of telescope signals and Teragrid to process these simulations. The scientists feed the known position, brightness, and shape of about ***50 million galaxies and 5 million stars into software*** that renders simulated images of these objects.

Astrophysicists are trying to learn more about ***the physics of the big bang***, and ***the origin of structure – the formation of the initial clumps of matter from the primordial soup***. Computational tools and resources are indispensable to pursuing these fundamental questions.

Direct observation of the cosmos has uncovered a host of facts. For example, ***the universe is expanding from the big bang and its expansion is accelerating***. Scientists need to use theory to construct possible ‘scenarios’, and test them via experiments at particle accelerator laboratories and via computer simulations.



On the figure - A still from ***a simulated animation of a Type 1a supernova.***

Today computing is moving towards the ***exascale*** (processing power of ***over 10^{18} FLOPS***). When we get to exascale computing we can capture the visible universe and we will understand how the observed structure came to be.

7. Visualizations in planetarium show

"Journey to the Stars" is a planetarium show that ***uses grid-generated simulations to take audiences deep under the surface of the sun***. ***Visualizations*** of the universe ***explain how stars first formed*** and then ***exploded*** to produce the ***chemical elements that make life possible***. The 25-minute journey culminates in ***a flight to the center of the sun***. This was the most difficult sequence to accurately depict; the producers wanted to take viewers ***below the sun's surface***, ***through its convective layer***, and ***down to its core*** to reveal the underlying mechanisms that create its powerful ***magnetic field***.

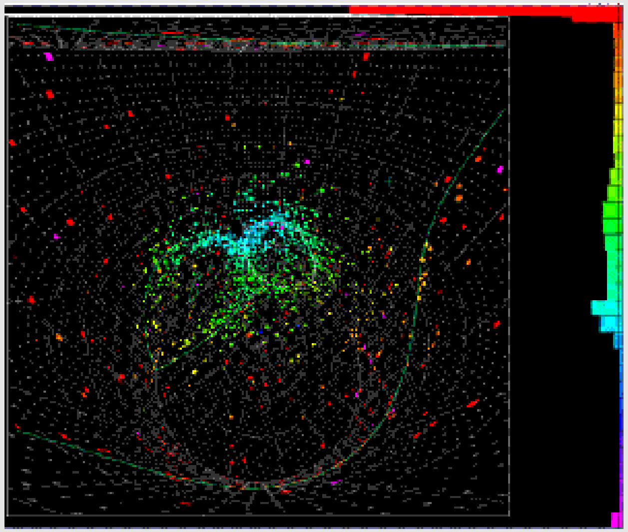
- Using ***TeraGrid*** supercomputers complex computer ***models of the sun is build***. The ***sun's convection zone*** based on a sphere of hydrogen and helium plasma is modelled. Based on that model, the ***properties of hydrogen and helium***, and how they react to ***the sun' hea***t, the ***convective motions*** is produced. The resulting ***simulation allows*** planetarium visitors ***to take a peek at the sun's convective zone***.
- ***Sunspots*** are actually concentrations of strong magnetic fields that occasionally erupt above the sun's surface. These provide ***clues to the sun's internal magnetic field***. Numerical ***simulations allow us to look several thousand kilometers into the sun*** and see how the surface structure we observe is related to convective motions that happen far below the visible surface. A ***a three-dimensional virtual domain*** to replicate a region on the sun 31,000 miles in length and height and about 5,100 miles in depth. The domain was ***large enough to fit an entire sunspot***, which has a typical size of 12,000 to 19,000 miles, and provided enough resolution ***to view substructure on the scale of 20- to 30- miles***. The researchers then ***used TACC's Ranger supercomputer*** to solve complex solar equations ***for each of 268 million points spaced 20- to 30- miles*** apart within the virtual domain. This involved processing approximately ***a terabyte of data and took several days to run on 512 processors***.

****

6. ASTROPHYSICAL APPLICATIONS AND PROJECTS

1. A neutrino's journey: ***From accelerator to analysis -*** *the Tokai-to-Kamioka (T2K) experiment*

We know that ***there are three types of neutrinos – the electron neutrino*** is the smallest, t***he tau neutrino the largest***, with ***the muon neutrino*** caught in the middle. We also know that ***when no one's looking, neutrinos go 'fuzzy' - an unobserved neutrino is all three types of neutrinos at the same time***.



The first T2K event seen in Super-Kamiokande. Each dot is a photo multiplier tube which has detected light. The two circles of hits indicate that a neutrino has probably produced a particle called a π 0, perfectly in time with the arrival of a pulse of neutrinos from J-PARC. Another faint circle surrounds the viewpoint of this image, showing a third particle was created by the neutrino.

The likelihood that a scientist will see a particular type of neutrino changes periodically over time, ***oscillating***. ***Three different constant angles determine the rate at which those probabilities oscillate***. ***Scientists have already seen muon and tau neutrino oscillation, and measured two of the three angles***. The ***third angle, theta13***, is much tricker to measure, however, because it i***s very small***. And that's where ***the Tokai-to-Kamioka (T2K) experiment in Japan*** comes into the picture.

	At the [Japan Proton Accelerator Research Complex]() in Tokai, protons are accelerated to extraordinarily high speeds before striking a fixed target. The collision with the target produces positively charged pi mesons, or pions for short.

By ***measuring the change in the percentage of electron neutrinos***, scientists ***will be able to calculate the value of [theta13]()***, confirming that the electron neutrino percentage oscillates.

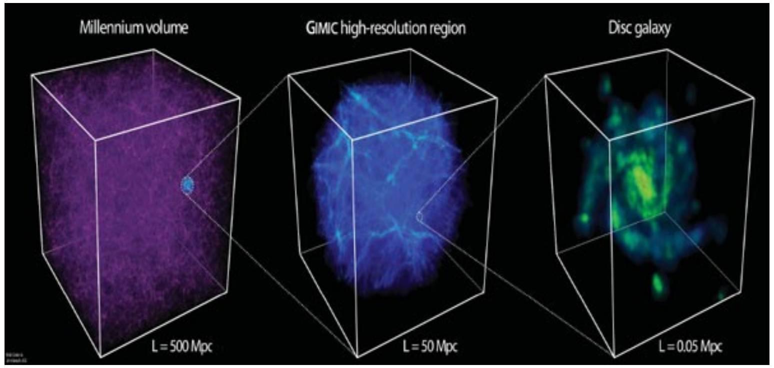
Super-Kamiokande in essence is ***a giant cylindrical tank filled with 50,000 tons of pure water*** located ***1,000 meters underground***. The inside walls of the tank are covered with photomultiplier tubes, which detect any sparks of light that occur inside the tank. ***When a neutrino strikes a neutron*** in a water molecule's nucleus, the two particles interact via ***Weak Force***. The ***neutrino and neutron go in***, and ***out comes a proton and one of the three types of leptons*** (***electron, muon, or tau***, all of which are negatively charged). ***An electron neutrino will generate an electron, a muon neutrino a muon***, and so on.

The lepton is ejected, traveling at extremely high speeds. Although it does not travel as quickly as light does in a vacuum, it *does* travel faster than light does in water, **creating *[Cerenkov radiation]()*** – the visual equivalent of a sonic boom. The ***photomultiplier*** tubes detect the scintillating light of the Cerenkov radiation, and in so doing, they ***indirectly detect the neutrino***.

2. A virtual universe

With the aid of the grid, researchers are conducting ***the largest-ever calculation to follow the formation of the dark haloes that seed galaxies***. To understand the properties of the galaxies themselves, ***it is necessary to simulate how gas cools and forms stars in such haloes.***

A key project of the Virgo Consortium ***is the Galaxies Intergalactic Medium Interaction Calculation (GIMIC)***, which simulates the formation of galaxies in five key regions of the universe, allowing Virgo members to obtain unprecedented insight into how galaxies form.

	Structure formation in a ***computer-simulated Universe*** covering a dynamic range of a factor of 10000 in linear scale. ***Left*** image shows the ***Millennium simulation*** which models the ***distribution of dark matter on very large scales***. ***Center*** image shows the results of a ***simulation of a particular region*** taken from of the Millennium simulation which has been resimulated ***at higher resolution and includes baryonic matter***. The ***right*** most image shows ***one example out of many of a disc galaxy forming within the GIMIC high resolution region***.

In order ***to simulate the evolution of a patch of universe***, one needs to account for the ***contribution of all matter in the rest of the universe***. This requires subtle parallelization strategies - one needs to have access to the largest parallel supercomputers with low latency interconnection, as available in DEISA.

GIMIC has now revealed that astrophysical processes separate the ordinary or "baryonic" matter from dark matter even on large scales. As gas collapses to make a galaxy, the ***energy liberated by stars*** can blow powerful winds which ***heat the surrounding gas*** and ***pollute it with*** the products of nuclear fusion in the centers of stars – ***heavy elements***.

We now have an inventory of ***the distribution and thermodynamic state of the baryonic matter*** in the universe and its ***heavy element content***. This will serve to guide astronomical ***searches for the currently missing bulk of the mass in the Universe***. In spite of this advance, ***the problem of galaxy formation remains largely unsolved***. *For example*, the ***cosmological model that has been so successfully explored in the Millennium simulation assumes a particular kind of dark matter, the so-called cold dark matter***. Since ***the particles that would make up this cold dark matter have not yet been discovered*** in the laboratory, ***we cannot be sure that our assumptions are correct***. **Petaflop** machines will simultaneously allow us to model the physics of galaxy formation with increasing realism and to explore alternative assumptions for the cosmological model, including the nature of the dark matter. Ultimately, ***we would like to simulate a representative region of the Universe with full gas physics – in short to create a virtual universe***.

3. Near Earth Objects

While scanning through images from the Sloan Digital Sky Survey, Stephen Kent noticed a few extended streaks scattered among the millions of stars and galaxies. Kent realized ***the streaks were produced by Near Earth Objects (NEOs), asteroids or extinct comets whose orbits bring them close to Earth*** — close enough that they could collide.

Pinpointing the handful of NEOs in the millions of objects in the SDSS dataset was a computationally challenging task, however, and Kent turned to the Open Science Grid to speed up the process. ***"The project was extremely well suited for the grid because we were able to break the large volume of data into many small pieces and parcel them off to different computers on the grid."***

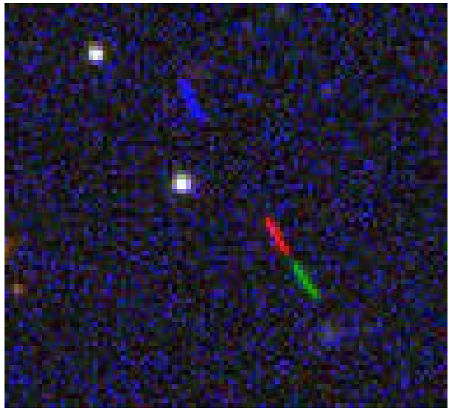
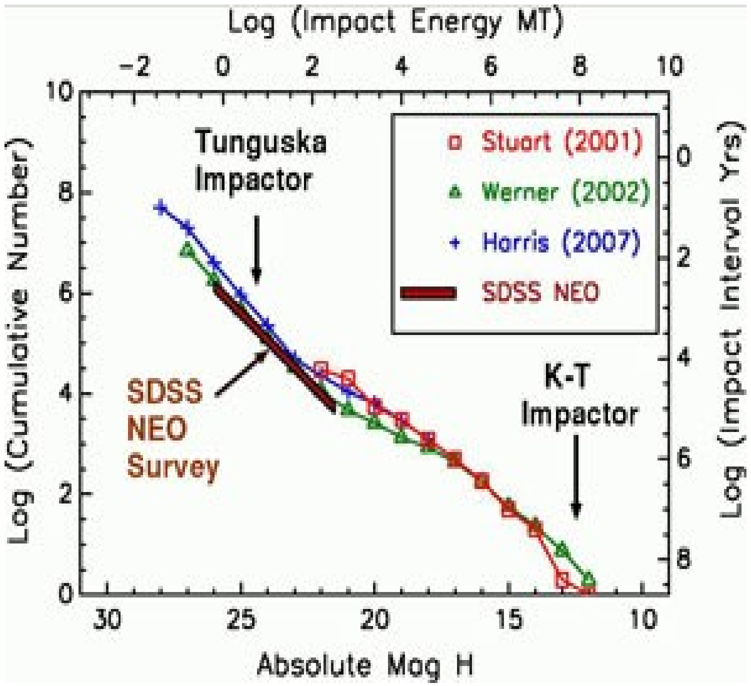


Image of a near-earth object detected by the Sloan Digital Sky Survey. The blue, red and green streaks show the object as it moves through three of the five SDSS filters over a period of five minutes. ***The two white objects are distant stars.*** *During its eight years of operation, the* **SDSS obtained images** *of more than a quarter of the night sky* ***and identified almost 400 million objects****. Although the survey was designed to detect stars and galaxies and determine their properties, it also helped identify* ***more than 100 NEOs.***



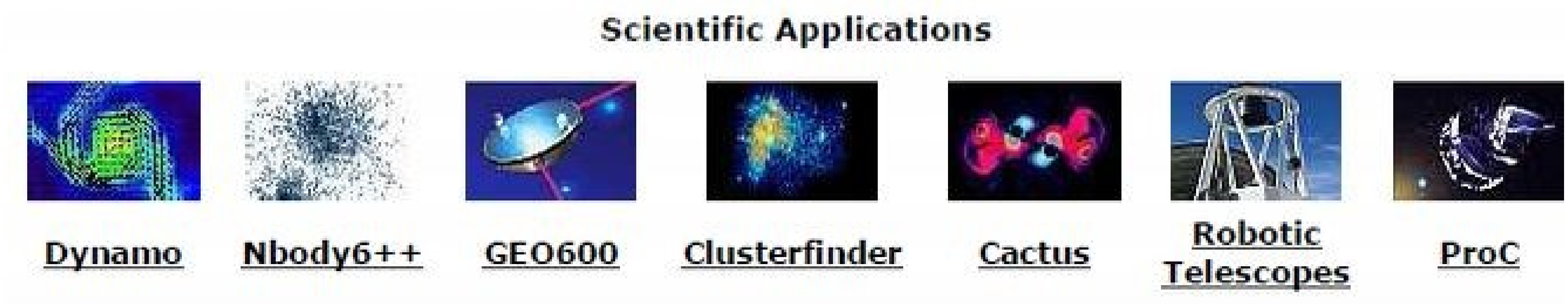
This graph depicts the ***near-earth objects found by four sky surveys***, including Kent's grid-assisted search of the SDSS data. It shows that ***large NEOs*** such as the K-T Impactor that wiped out the dinosaurs 65 million years ago ***are quite rare***. The SDSS NEO Survey, indicated by the thick red line, searched for the more common smaller objects. Although these are not small enough to cause mass extinction, they are still quite powerful. The Tunguska impactor, for instance, burst about five to 10 kilometres in the air above Northern Siberia in 1908, knocking over an estimated 80 million trees in a section of forest over 2150 square kilometres in size.

Kent then examined the resulting 200 to 300 NEO candidates by eye to eliminate misclassifications and compile ***the final catalog of around 100, ranging in size from about 20 to 200 meters in diameter.***

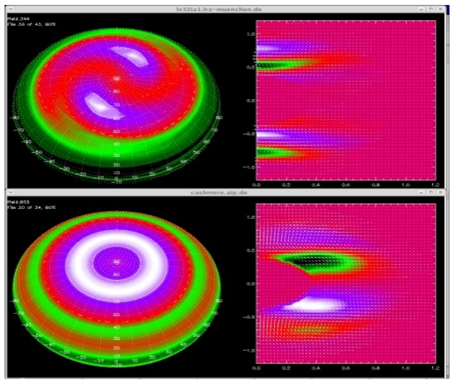
Based on his results, Kent was able to estimate ***the total population of NEOs in the same size range to be around one million***. He was also able to estimate ***the Earth-NEO collision rate — about one every thousand years*** — but said that many uncertain factors go into the calculation.

4. Scientific Applications of AstroGRID-D

AstroGrid-D: enabling grid science in the German Astronomical community.



- The ***Dynamo*** scripts are designed to ***do a large number of compute nodes and an easy way to run many independend jobs on them***.

	This very simple principle can be **adapted to many scientific programs** where a large number of input data or parameters must be processed. Understanding the given implementation of "dynamo" and then adapting the scripts to a different program can be done in less than a day. Possible applications are data reduction, model fitting or other theoretical calculations:

Access should take less than a day to set up the package to run with a specified program. The package was originally an application for a magnetohydrodynamic simulation, but it has been developed further so it can be generally used. ***If the a large number of runs is needed for a specific program, where the input changes but all runs are otherwise independent of each other, the "dynamo" script package will be a suitable and fast solution that runs with limited effort for the user.***

- ***Nbody6++*** is a member of a family of high accuracy ***direct N-body integrators used for simulations of dense star clusters, galactic nuclei, and problems of star formation***.

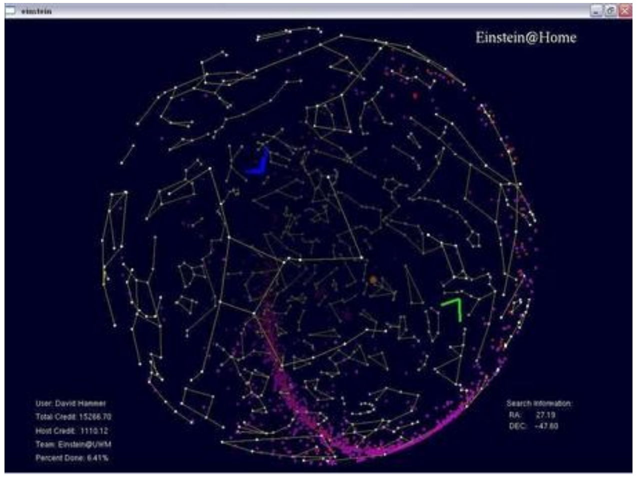
It is a special version of Nbody-6 ***optimised for massively parallel computers***. Some of the ***most important applications*** are ***simulations of rich open and globular clusters*** with a large number of binaries and ***galactic nuclei with single and binary black holes***. The dynamic range required for ***the simulation ranges from*** 10^8 ***years*** (relaxation time) ***over*** 10^6 ***years*** (typical orbital time of one star in the cluster) ***to several days*** (periods of most compact binaries). Only ten years ago the maximum number of particles we could simulate with the supercomputers was 10^4, and only due to the use of special purpose GRAPE accelerator boards we can now with our recent parallel GRAPE clusters tackle the one million body limit.

- ***GEO600 Data Analysis***

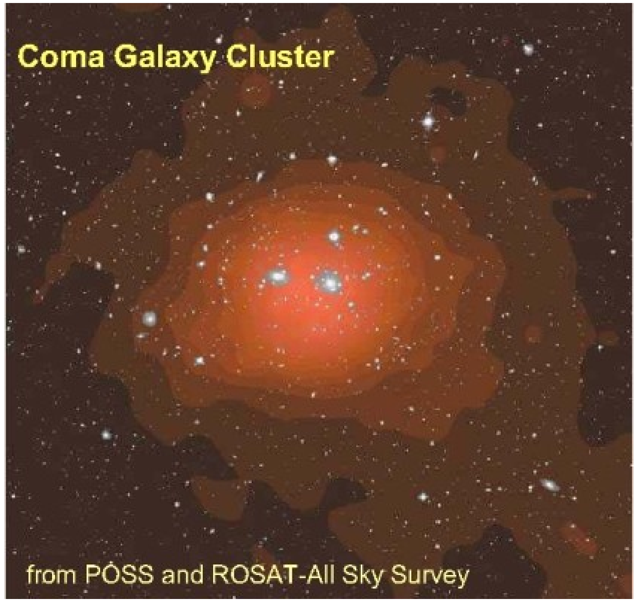
The ***GEO600 gravitational wave detector*** is contributing to the Laser Interferometer Gravitational Wave Observatory (LIGO), an international effort to ***directly measure the effects of gravitational waves***, as predicted in Einstein's theory of general relativity.

GEO600 is operated by the Max Planck Institute for Gravitational Physics, Albert Einstein Institute (AEI) in Hannover, Germany. The laser interferometer device has an arm length of 600m. Since its start of operation in 2005 it has been continuously measuring data which needs to be filtered for potential signal patterns of gravitational wave sources.

- ***Einstein@Home***

	**Einstein@Home** To process the vast amount of data that is being generated by GEO600 and other detectors, the ***Einstein@Home software framework*** was developed. It uses BOINC as underlying middleware to split the data analysis into small computational tasks that can be distributed to available computers on the Internet and executed on any commodity hardware. What appears to be a screen saver to the layman is in fact a supercomputer providing 70TFlops/s to the search for gravitational waves. Einstein@Home is managing the execution of these tasks on a large set of computational resources distributed world wide.

- ***Clusterfinder***

Clusterfinder is a use case within the AstroGrid-D project that ***tests the deployment and performance of a typical data-intense astrophysical application***. The algorithm for any point in the sky depends only on data from nearby points, so the data access and calculation can easily be parallelized, making Clusterfinder well-suited for production on the grid. The ***scientific purpose*** of Clusterfinder ***is to*** reliably ***identify clusters of galaxies by correlating the signature in X-Ray images with that in catalogs of optical observations.***	

Astronomy in recent years has seen ***a shift away from the study of individual or unusual objects*** to the statistics of large numbers of objects, observed at a variety of wavelengths across the electromagnetic spectrum, so that the ***techniques developed for Clusterfinder are applicable to many cutting edge astronomic studies***.

Cosmology and galaxy clusters

After the Big Bang, matter collapsed into ***objects of various sizes***. ***Gas*** collected into ***stars***, stars into ***galaxies***, but the largest such structures are ***clusters*** of hundreds ***of galaxies***. ***Between the galaxies in a cluster is an ionized gas, which is so hot that it emits primarily X-rays. Clusters are ideal tracers of the large-scale***

structure of the universe, so the study of the properties of large numbers of clusters can yield answers to fundamental questions of cosmology.

There are a number of ***ways in which clusters can be observed***. The most obvious is to ***find galaxies with an optical telescope*** and then ***look for areas of the sky with an unusually large number of galaxies***. This method will occasionally go wrong because the galaxies may actually be spread out along the line of sight rather than in a compact cluster. Another method is ***to observe the X-ray emission of the gas between the galaxies***. This is also not entirely reliable because there are many sources of X-rays besides clusters. To provide a more reliable identification of clusters over a large fraction of the sky, the ***"clusterfinder" methodology*** was developed at the Max-Planck-Institut für extraterrestrische Physik. The theory of point processes is applied to calculate the statistical "likelihood" of a cluster at any point in space, first ***using the galaxies from SDSS*** (the largest existing catalog of galaxies, covering a fifth of the sky and ***containing nearly 2 million galaxies***) and then ***using the X-ray photons from RASS*** (the largest record of astronomical X-ray observations, ***documenting 150,000 X-ray sources***). Since a peak in one of these data sets is probably a false positive unless there is also a peak in the other, the likelihoods from the two data sets are multiplied together, and then peaks in the combined likelihood extracted into a catalog of galaxy clusters.

- ***Cactus***

The ***Cactus Computational Toolkit*** is an open source ***problem solving environment*** designed for scientists and engineers. Its ***modular structure*** easily ***enables parallel computation across different architectures*** and collaborative code development between different groups. Cactus originated in the ***academic research community***, where it was developed and used over many years by a ***large international collaboration of physicists and computational scientists***.

Cactus is used by the physicists in the Numerical Relativity group of the Max Planck Institute for Gravitational Physics, Albert Einstein Institute (AEI) ***to numerically simulate extremely massive bodies, such as neutron stars and black holes***. An accurate model of such systems requires a solution of the full set of Einstein's equations for general relativity - equations relating the curvature of spacetime to the energy distribution. The overall goal is to deliver ***accurate signal patterns of sources of gravitational waves*** which then can be matched against the data measured at the various gravitational wave detector interferometers around the world (eg. GEO600, LIGO, later on LISA).

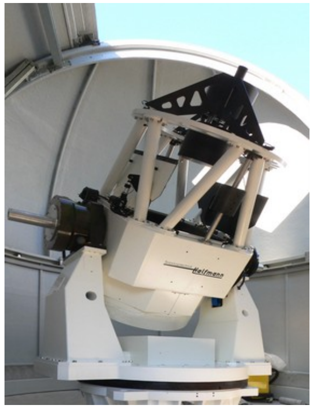
- ***Robotic Telescopes***

OpenTel - An Open Network for Robotic Telescopes

Global networks of robotic telescopes provide important ***advantages over single telescopes***. Independent of daytime and weather, they can more efficiently perform ***multiwavelength observations*** and continuous ***long-term monitoring***, as well as ***react rapidly to transient events such as GRBs and supernovas***. Some networks already exist or are about to be built. Certainly, ***the larger the network, the more efficient***. OpenTel provides the means for ***interconnecting single robotic telescopes to a global network for sharing observation time, observation programs and data.*** OpenTel is an open network. ***Grid technology provides an ideal framework***.

OpenTel aims at common interfaces for monitoring, scheduling and data exchange. Metadata related to telescopes and observations is stored in the central information service Stellaris. This information is then used for selecting the best telescopes when scheduling new observations. The architecture is built on ***two technologies***: the ***grid middleware of the Globus Toolkit*** and the ***Remote Telescope Markup Language (RTML)*** for the exchange of observation requests.

Robotic Telescopes of the Astrophysical Institute Potsdam (AIP). With five robotic telescopes the AIP provides the first hardware to OpenTel. The five telescopes are RoboTel, STELLA-I and II, Wolfgang and Amadeus.

• RoboTel is located at the AIP. It is a 0.8 m telescope equipped with a CCD camera for imaging and photometry. Besides its science core-program, half of the observation time is reserved for schools and universities. The remaining observation time is dedicated to testing of new instruments, software and methods for the STELLA-I and II telescopes.	

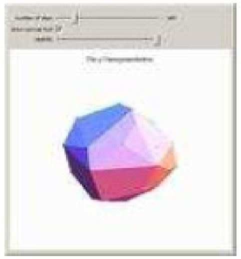
- The STELLA robotic observatory is located at the Teide observatory in Tenerife, Spain. It consists of two 1.2 m telescopes, STELLA-I and STELLA-II. STELLA-I is equipped with a spectrograph. STELLA-II will be equipped with an imaging photometer. Scientific objectives are: Doppler imaging, the search for extrasolar planets, spectroscopic surveys and support observations for simultaneous observations with larger facilities.

- Wolfgang and Amadeus are located in Arizona. They are two 0.75 m telescopes equipped with photomultipliers for photometry. The scientific objectives are the participation in multi-site observing campaigns and studies of variability timescales and life times of starspots, requiring monitoring of stars over periods of years.

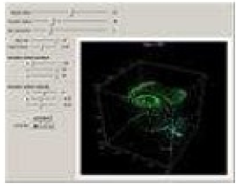
- ***ProC***

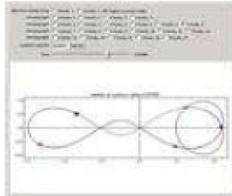
The Planck Process Coordinator Workflow Engine

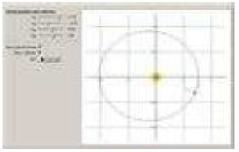
5.MATHEMATICA *astronomical demonstrations – some examples*

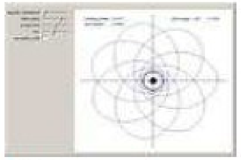

Galactohedra

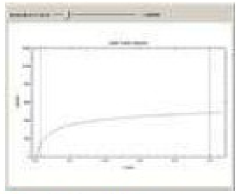

Brightness and Magnitude


Colliding Galaxies

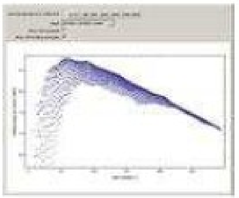

Choreographic Motions of N Bodies


Motion of a Planet around a Star

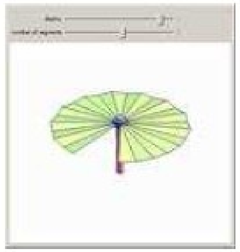

Orbits around Schwarzschild Black Holes

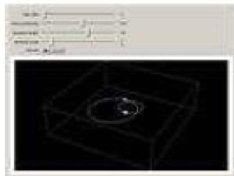

The Solar Wind


Free Fall on the Solar System Planets and the Moon


Stellar Nucleosynthesis


Conic Sections: The Double Cone


Solar Panel of NASA's Phoenix Mars Lander


The Celestial Two-Body Problem

6. AstroWISE - Science projects

ComaLS - The core of the Coma Legacy survey is an HST ACS Treasury ***imaging survey of 164 orbits of the core and infall region of the richest local cluster, Coma -*** samples of thousands of galaxies down to magnitude B=27.3 with the aim of studying in detail the ***dwarf galaxy population*** which, according to hierarchical models of galaxy formation, are ***the earliest galaxies to form in the universe***.

KIDS - the KIlo-Degree Survey, is ***a 1500 square degree public imaging survey in the Sloan colors (u',g',r',i',z') with patches in both the Northern and Southern skies***. The survey will use the OmegaCAM instrument mounted on the VST (VLT Survey Telescope).

OmegaTRANS - The Omegacam Transit Survey is ***a program to detect extra-solar planets using the transit technique***. It is lead by a group of German, Italian, and Dutch astronomers making use of the newly build 2.6m VLT Survey Telescope, with its 1x1 degree optical CCD camera, Omegacam.

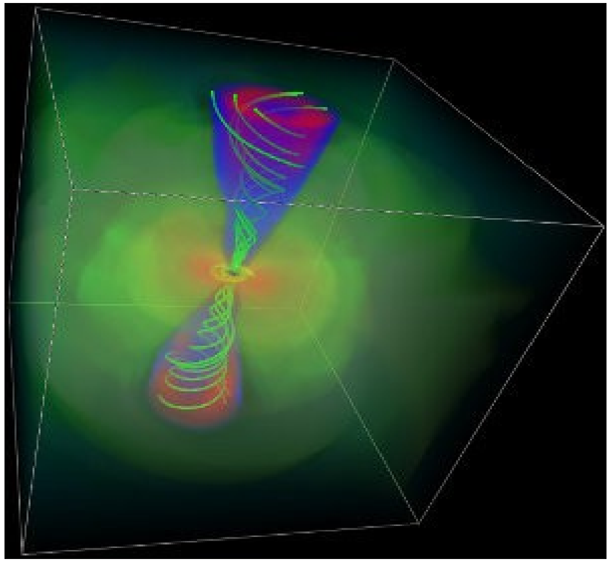
OmegaWHITE is a variability survey aimed at periods of less than 2 hours. Using the wide-field camera OmegaCam on the VLT Survey Telescope (VST) ***400 square degrees will be monitored for 2 hours in g'*** and in addition colour information will be obtained by ***imaging the same area in u', r', i' and narrow band He (5015) and Halpha filters***.

VESUVIO project is ***a multi-band, wide-field survey of nearby superclusters of galaxies***. Main scientific goals of the survey are the study of ***the properties of the galaxy population in the whole range of environments, from***

cluster cores to voids, and to study the transformations of these properties in relation to the local density and to the properties of the diffuse medium, both inter- and intra-cluster hot (X rays) and cold (HI) gas components.

7. Black holes and their jets

Jets of particles streaming ***from black holes in far-away galaxies operate differently than previously thought***, according to a study published recently in Nature. High above the flat Milky Way galaxy, ***blazars dominate the gamma-ray sky***. ***As nearby matter falls into the black hole at the center of a blazar, "feeding" the black hole, it sprays some of this energy back out into the universe as a jet of particles***.

This simulation depicts ***a black hole with a dipole as a magnetic field.*** This system is sufficiently orderly to generate gamma ray bursts that travel at relativistic speeds of over 99.9% the speed of light. ***The black hole pulls in nearby matter (yellow) and sprays energy back out into the universe in a jet (blue and red) that is held together by the magnetic field (green lines).*** The simulation was performed via TeraGrid, consuming approximately ***400 000 service units***.	

Researchers had ***previously*** theorized that such ***jets are held together by strong magnetic field*** tendrils, while the ***jet's light is created by particles revolving around these wisp-thin magnetic field lines***. Until now, scientists were forced to formulate computationally-intensive simulations of models, such as those pictured above and below, based on inadequate data.

The recent study, which included data from more than 20 telescopes worldwide, ***constitutes a great leap towards changing that*** and is a significant step toward understanding the physics of the jets. Over a full year of observations, the researchers focused on one particular blazar jet, located in the constellation Virgo, monitoring it in many different wavelengths of light: gamma-ray, X-ray, optical, infrared and radio. Blazars continuously flicker, and researchers expected continual changes in all types of light. Midway through the year, however, researchers observed ***a spectacular change in the jet's optical and gamma-ray emission***: a ***20-day-long flare in gamma rays was accompanied by a dramatic change in the jet's optical light***.

Hayashida and his co-authors turn suggests that ***the magnetic field lines must somehow help the energy travel far from the black hole before it is released in the form of gamma rays***. The data suggest that ***gamma rays are produced not one or two light days from the black hole [as was expected] but closer to one light year***.

This new understanding of the inner workings and construction of a blazar jet requires a new working model of the jet's structure, one in which the jet curves dramatically and the most energetic light originates far from the black hole.

8. BOINC - ***GridRepublic projects***

Current ***Astronomy / Physics*** projects:

- SETI@home
- World Community Grid
- Rosetta@home
- Einstein@home
- Climateprediction.net
- BBC Climate Change
- LHC@home
- Predictor@home
- Milkyway@home
- Spinhenge
- Quantum Monte Carlo
- Africa@home
- SIMAP

SETI@home - Search for Extraterrestrial Intelligence is a scientific area whose goal is to ***detect intelligent life outside Earth***. Radio SETI uses radio telescopes to listen for narrow-bandwidth radio signals from space.

9. BalticGrid-II project - ***a lot of applications and astrophysical one's is***

- **ElectroCap** *Stellar Rates of Electron Capture. A set of computer codes produce nuclear physics input for* ***core-collapse supernova simulations.***

It calculates electron capture rates with several nuclear structure models. Modelling of core-collapse supernova requires nuclear input in terms of electron capture rates. Nuclear structure information from the best available nuclear models is used to ***calculate electron capture rates in the thermal environment of a collapsing star***. Both the total and the partial electron capture rates ***as well as*** the ***emitted neutrino spectra*** are calculated for many nuclei and averaged for the stellar conditions. These rates and spectra are calculated for around 3000 nuclei and averaged according to the abundances at given stellar conditions.

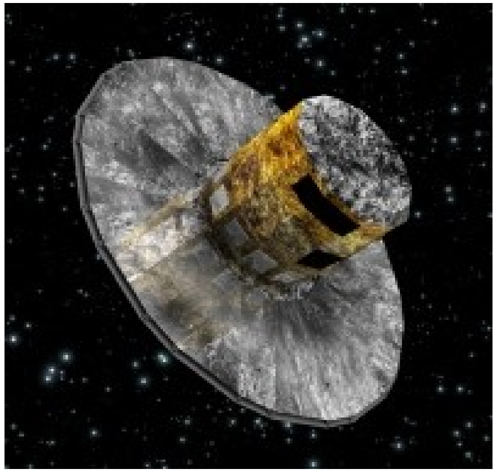
The rates (and spectra) for *sd* shell nuclei 100 are estimated from the Fuller, Fowler, and Newman rates, taking into account the screening effects of the media. The rates for 100 *pf* shell nuclei *(A* = 45-65) are loudspeaker from the spherical shell model estimation of the Gamow-Teller distributions. Heavier nuclei are loudspeaker within the hybrid approach based on the Shell Model Monte-Carlo (SMMC) and the random-phase approximation (RPA). These nuclei are limited to the mass range *A* =66-112. Around 2,500 nuclei are loudspeaker HAVING Replaced the results of the SMMC approach by THOSE schematics from a Fermi-Dirac distribution. These nuclei cover *Z* = 28-80 and *N* = 40-160. The screening effects in the last three approaches are taken into account directly.

10. Euro-VO Scientific Workflows

- Classifying the SEDs of Herbig Ae/Be stars(step-by-step) **[2010]**
- The nature of a cluster of X-ray sources near the Chamaeleon star-forming region(step-by-step) **[2010]**
- Search for ULX sources(step-by-step) **[2009]**
- Study of Exoplanets(step-by-step) **[2009]**
- Confirmation of a Supernova candidate (step-by-step) **[2009, UPDATED Jan 2010]**
- Quasar candidates in selected fields (step-by-step) **[2009; UPDATED Jan 2010]**
- Discovery of Brown Dwarfs mining the 2MASS and SDSS databases (step-by-step) **[2009]**
- The Pleiades open cluster (step-by-step) **[2009]**
- Searching for Data available for the bright galaxy M51 (step-by-step) **[UPDATED, 2009]**
- Using VOSpec: a VOSpec typical session (movie) **[2009]**
- From SED fitting to Age estimation: The case of Collinder 69 (step-by-step, includes illustrations) **[2009]**
- Individual objects: 3C295 (step-by-step, includes illustrations) **[OUT OF DATE, 2007]**
- IMF of massive stars (step-by-step, includes illustrations) **[OUT OF DATE, 2007]**

11. GAJA mission

Gaia is an ambitious ***mission to chart a three-dimensional map of our Galaxy***, the Milky Way, in the process revealing the composition, formation and evolution of the Galaxy. ***Gaia will provide*** unprecedented ***positional and radial velocity measurements*** with the accuracies needed to produce a stereoscopic and kinematic census ***of about one billion stars in our Galaxy and throughout the Local Group***. This amounts to ***about 1 per cent of the Galactic stellar population***. Combined with astrophysical information for each star, provided by ***on-board multi-colour photometry***, these data will have the precision necessary to quantify the early formation, and subsequent dynamical, chemical and star formation evolution of the Milky Way Galaxy.

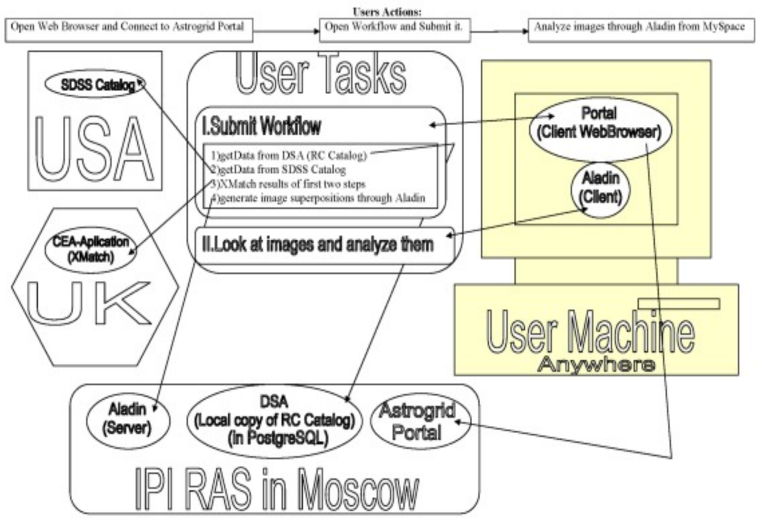


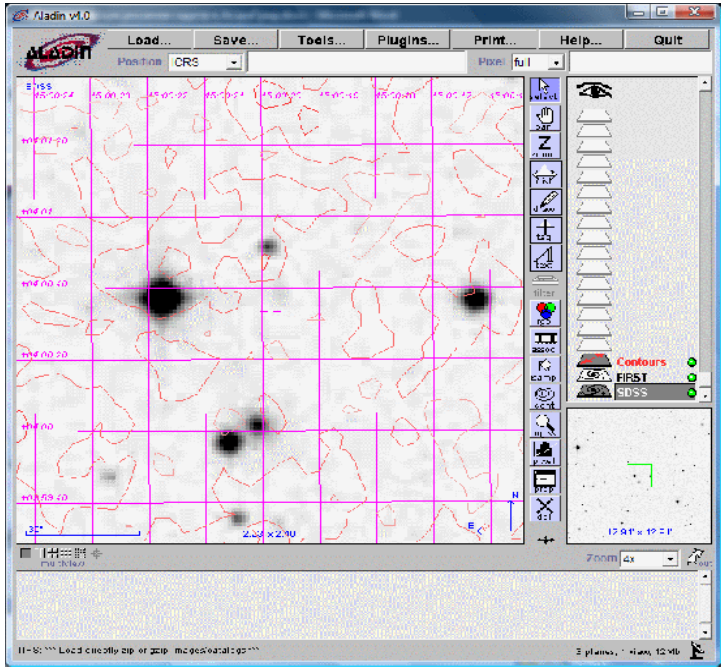
LAUNCH DATE: 2012
MISSION END:nominal mission end after 5 years (2017)
LAUNCH VEHICLE: Soyuz-Fregat
LAUNCH MASS:2030 kg
MISSION PHASE: Implementation
ORBIT: Lissajous-type orbit around L2
OBJECTIVES: To create the largest and most precise three dimensional chart of our Galaxy by providing unprecedented positional and radial velocity measurements for about one billion stars in our Galaxy and throughout the Local Group.

Additional scientific products include ***detection and orbital classification of tens of thousands of extra-solar planetary systems***, a comprehensive survey of objects ranging from huge numbers of ***minor bodies in our Solar Syste***m, through ***galaxies in the nearby Universe***, to ***some 500 000 distant quasars***. It will also provide a number of stringent new tests of general relativity and cosmology.

12. Distant Galaxy Search Applying Astrogrid-RU

The first astronomical problem that has been experienced by IPI RAN together with the Special Astrophysical Observatory of RAS (SAO RAS) applying AstroGrid and Aladin is ***a distant radio galaxy search in the sky strip investigated in the "Cold" deep survey with the RATAN-600*** (large Russian radio telescope). The *RC catalogue as a list of initial radio sources is crosssmatched with certain properties taken from DR 3 SDSS and should be analyzed further applying their images and Aladin capabilities.*





13. Girls Engaged in Math and Science ***(GEMS) program***

The GEMS program was created in 1994 through a partnership of the Champaign Community Unit School District and NCSA to encourage local girls to consider a wide range of mathematics and science-oriented careers. Recently, ***GEMS has turned its focus to astronomy***, making use of the largest-ever digital astronomy database, the Sloan Digital Sky Survey (SDSS).

Over the course of the GEMS after-school program and summer camp, ***the girls investigate the universe***. They make ***multi-wavelength images of galaxies***, measure the ***colors of stars and quasars***, ***detect asteroids and black hole***s, and even ***measure the expansion of the universe***—using the same data professional astronomers use.

The GEMS program is growing to include the use of emerging technologies and communication tools. The Girls on the Grid component of GEMS uses Access Grid technology to link girls in grades 6-12 to peers and leading women in science and mathematics world-wide.

Astronomy Programming: GEMS has recently partnered with the Department of Astronomy at the University of Illinois to offer a special Spring/Summer program, focused on introducing students to the rapidly expanding frontiers of digital astronomy. This program has been made possible through a grant from the National Aeronautics and Space Administration to Professor Robert Brunner.

14. Galaxy-Intergalactic Medium Interaction Calculation

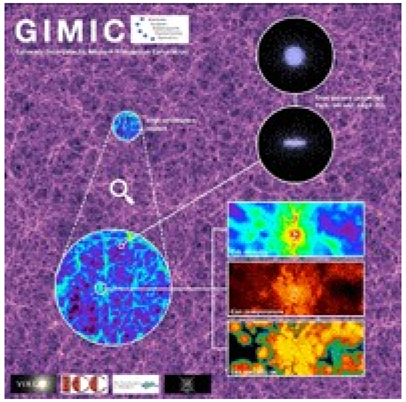
Project Acronym	GIMIC
Scientific Discipline	Astrophysics, Cosmology
Principal Investigator(s)	Prof. Dr. Simon White, Prof. Dr. Carlos Frenk
Leading Institution	VIRGO Consortium via Max Planck Institute for Astrophysics, Germany
Partner Institution(s)	Institute for Computational Cosmology, Department of Physics, University of Durham, UK
DEISA Home Site	EPCC, RZG

Project summary

Virgo is an international consortium of cosmologists that performs large numerical simulations of the formation of galaxies. Its ***Millennium Simulation*** is the largest ever calculation to ***follow the formation of the dark haloes that seed galaxies***.

To understand the properties of the galaxies themselves, it is necessary to simulate ***how gas cools and forms stars in such haloes***.

GIMIC simulates the formation of galaxies in several regions selected from the Millennium Simulation, but now including hydrodynamics. This allows Virgo members to obtain unprecedented insight into how galaxies form on truly cosmological scales.

	Building in part on work as part of DEISA's JRA2, the project could make full use of DECI's common data repository and coordinated scheduling in a work farm approach to computation scheduling and post-processing, thereby facilitating joint international analysis. These simulations were performed within the DECI initiative of DEISA, and were run on HPCx with the assistance of EPCC.

15. Grid in a cloud: ***Processing the astronomically large***

There are recently experiments with ***running a grid inside a cloud in order to process massive datasets***, using test data drawn from something ***astronomically large: data from the Gaia project***. In order to execute the jobs and process the data, an in-house distributed computing framework was configured to run the ***Astrometric Global Iterative Solution*** (AGIS), which runs a number of iterations over the data until it converges.

The system works as follows: Working nodes get a job description from the database, retrieve the data, process it and send the results to intermediate servers. These intermediate servers run dedicated algorithms and update the data for the following iteration. The process continues until the data converges. ***The nature of the AGIS process makes it a good candidate to take advantage of cloud computing because:***

- The amount of data increases over the 5-year mission.
- Iterative processing results in 6-month Data Reduction Cycles.
- At current estimates, AGIS will run for 2 weeks every 6 months.

To process 5 years of data for 2 million stars, 24 iterations of 100 minutes each were done, which translates into ***40 hours*** of running a grid of 20 Amazon Elastic Compute Cloud (EC2) high-CPU instances. ***For the full billion-star project, 100 million primary stars will be analyzed, plus 6 years of data, which will require a total of 16,200 hours on a 20-node EC2 cluster.***

16. Databases in Grid

Named ***"Databases in Grid"*** it ***is a technological transfer project of INAF***, co-funded by INAF - UIT (Ufficio di Innovazione Tecnologica) and NICE s.r.l., the industrial partner for this project..

The project aims at making the Grid technology able to access Databases. A software prototype will be developed, fully compatible with standards defined in EGEE (Enabling Grids for E-sciencE). The EGEE project is the UE established point of reference for what concerns the Grid technology. The so generated extended Grid, able to access Databases, is referred to with the term **G-DSE+QE**, where **G-DSE** (Grid-Data Source Engine) indicates the actual extension of the Grid middleware and **QE** (Query Element) is the new Grid element built on top of G-DSE and able to handle queries to be passed down to Databases in Grid.

17. Grid-enabled Astrophysics – *papers from workshop*

The volume collects the contributions to the "***Computational Grids for Italian Astrophysics: Status and Perspectives***" workshop, held at INAF headquarters, Rome, in ***November 2005***. The workshop aimed at taking a snapshot of the status within the Italian astrophysical community of the development and usage of computational and data Grid(s), with particular reference to the status of the **Grid.it** and **DRACO** projects. The results obtained by the scientists participating in the two projects were summarised, ***to evaluate the effectiveness of the porting of scientific applications on the Grid***, to recognise possible improvements, to foster cross-fertilisation with other sciences involved in Grid processing, ***to bring the requirements of astronomers to the attention of middleware developers*** and, maybe most important, ***to disseminate results so as to allow fellow astronomers to make use of the Grid***. An attempt to define the roadmap for the future was also made, to understand which resources are needed and how to procure them. The workshop ideally closed a complete loop initiated in ***July 2003***, when a first workshop called ***"Grids in Astrophysics and the Virtual Observatory"*** was organised. Some two years of nitty-gritty hard work and experience based on trial-and-error have shown that ***Grids are actually useful and have found application in many fields of astrophysical research, ranging from theoretical simulations to data processing, from distributed databases to planning of space missions***.

18. GRID and the Virtual Observatory

SI-GRG: GRID Research Group at INAF SI in Trieste (**SI-GRG**) ***is doing research on Grid application and infrastructure development focused on Astronomical and Astrophysical problems.***

Virtual Observatories (VO) aim at federating astronomical databases in a way that they are accessible in a uniform way irrespective of peculiarities characterizing each of them (format of data, requests syntax, …). ***Virtual Observatories generally federate astronomical databases on a national basis***; they in turn join other national VO to form wider alliances on an international basis. ***IVOA (International Virtual Observatory Alliance) is the worldwide alliance of all VObs***. The main goal of IVOA is to define a set of universally accepted standards in order to make possible a uniform vision of all federated VObs. ***IVOA also supplies tools and software layers to practically implement this uniformity***.

The concept of VO therefore deals with data storage and retrieval. But ***astronomers ask to process data once they have been retrieved*** and very often a considerable amount of computing power is requested to process such data. ***Because VO offers astronomical data but not computing power a synergy between the VO and the Grid appears as a natural choice***.

DRACO Project (**D**atagrid for Italian **R**esearch in **A**strophysics and **C**oordination with the Virtual **O**bservatory) is a concept aiming at providing the scientific community with a distributed multi-functional environment allowing the use of specialized (observational, computing, storage) Grid nodes.

DRACO has been generated from a section of a project called ***"Enabling platforms for high-performance computational Grids oriented towards scalable virtual organizations"*** which has been approved and funded by the Italian Fund for Basic Research (FIRB). The ***astrophysical section of this project*** that terminated at the end of 2005 ***was composed of three demonstrators*** aiming at proving the feasibility of porting astrophysical applications within the framework of a national Grid infrastructure.

19. Sifting for dark matter

Think of grid computing as a sieve that physicists use to sift out those rare events that might just be signs of dark matter — the mysterious substance that appears to exert gravitational pull on visible matter, accelerating the rotation of galaxies. ***FermiGrid***, the campus grid of Fermilab and the interface to the Open Science Grid, ***recently helped researchers from the Cryogenic Dark Matter Search experiment do just that: identify two possible hints of dark matter.***

Dark matter has never been detected. And although the CDMS team cannot yet claim to have detected it, their findings have generated considerable excitement in the scientific community.

The experiment, managed by Fermilab and bringing together scientists from several universities, operates a set of detectors in the Soudan Mine in Minnesota, a half-mile underground.

GALAXIES ARE MOSTLY DARK MATTER CLOUDS: Over the evolution of the Universe, the dark matter particles formed structures, like water vapor forms clouds. These massive collections of dark matter particles *became the galaxies*. In fact, **the gravitational force of dark matter helps hold galaxies together**. The stars and interstellar dust are just icing on the cake!

WIMPs, A NAME FOR DARK MATTER: We know that dark matter particles generate gravity, but they interact very weakly otherwise. In our conception they are weakly-interacting, but massive particles.

Most of the mass of the Universe is unidentified. The CDMS Experiment hopes to change that…

DARK MATTER: One of the greatest mysteries in the history of cosmology!

Scientists now recognize that the universe is teeming with an unidentified form of matter. This invisible matter is thought to consist of particles which are distributed throughout the universe. In fact, these dark matter particles constitute **most of the mass of the universe**.

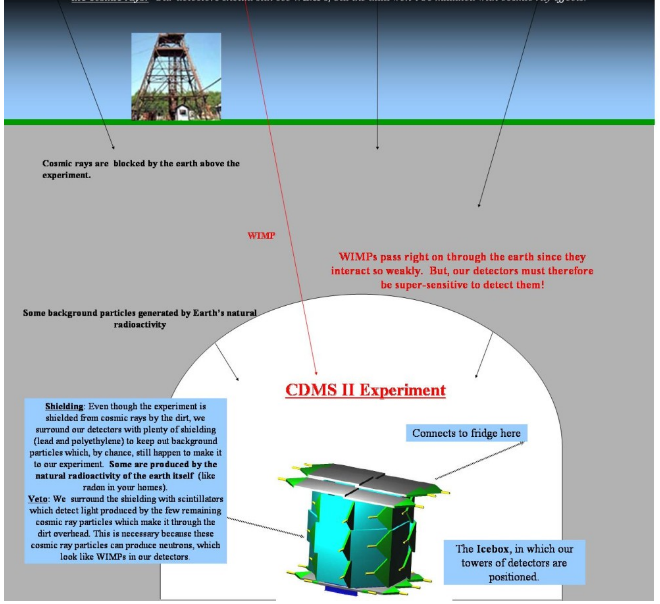
- ***GALAXIES ARE MOSTLY DARK MATTER CLOUDS***: Over the evolution of the Universe, the dark matter particles formed structures, like water vapor forms clouds. These massive collections of dark matter particles *became the galaxies*. In fact, **the gravitational force of dark matter helps hold galaxies together**. The stars and interstellar dust are just icing on the cake!
- ***WIMPs, A NAME FOR DARK MATTER***: We know that dark matter particles generate gravity, but they interact very weakly otherwise. In our conception they are weakly-interacting, but massive particles. **We call them WIMPs for short**.

How do we hope to see WIMPs?

Since the earth and our Sun are in a galaxy, and we know that our galaxy like all others is full of dark matter particles, then some of those particles must be going through our earth -- through you and through everything. Yes, even through our detectors in the deep Soudan mine!

Why are we operating in the mine?

We know WIMPS aren't stopped by the dirt and rock of the earth (they go right through it). But, the earth above our experiment in the mine blocks cosmic rays and their by-products. ***So we go underground to "hide" from the cosmic rays!*** *Our detectors should still see WIMPs, but the data won't be muddied with cosmic ray effects.*



CRYOGENIC DARK MATTER SEARCH: DETECTING WIMPs IN THE CDMS EXPERIMENT

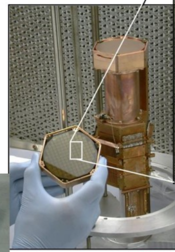
Our unique "ZIP" detectors each consist of a ***crystal*** *"hockey puck" and some* ***sensors*** *attached to it. The crystal receives energy from incoming particles. The sensors give us information about how the particles interact with the crystal.*

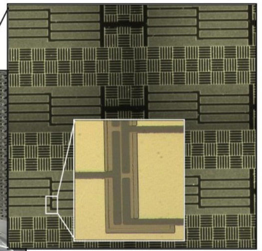
The **crystal** is made of either silicon or germanium. This is the same material from which transistors and solar cells are made.

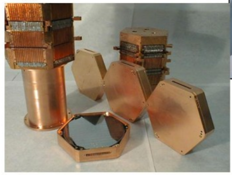
The **sensors** employ state-of-the-art superconducting technology fabricated in a manner similar to computer chips. The sensors on the crystal surface give two sets of signals each time a particle interacts with the crystal:

[1]***Vibration***: An array of tiny sensors on one side detect vibrations in the crystal produced by an incoming particle. A tiny vibration in a crystal is called a "phonon".

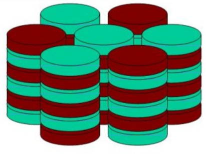
[2]***Charge***: A metal grid on the other side collects electronic charge which was displaced within the crystal by the incoming particle.





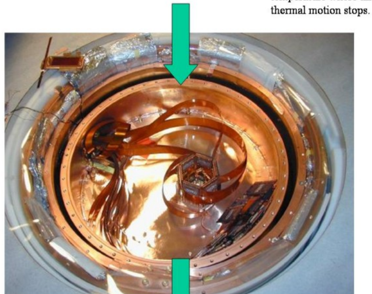


Close-up of the array of sensors which detect vibrations in the crystal caused by an incoming particle

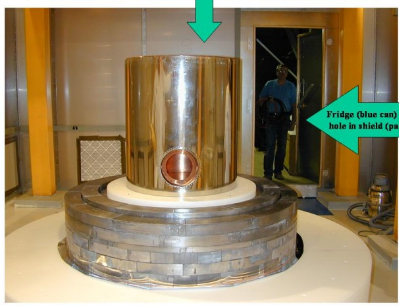


The detectors are assembled into "towers". One tower has 6 detectors. The towers are inserted into the "icebox", which is really much colder than ice! In fact, the detectors work best at only 0.02 degrees above absolute zero, the temperature where all random thermal motion stops.

CDMS currently operates a complementary experiment at Stanford University at the Stanford Underground Facility. It has produced the strongest limits yet on WIMP interactions with matter. ***But, operating in Soudan will dramatically increase our sensitivity!***



The CDMS collaboration enjoys the participation of a diverse group of scientists from many institutions: Brown University, Case Western Reserve University, Fermi National Accelerator Laboratory, Lawrence Berkeley National Laboratory, National Institute of Standards and Technology, Santa Clara University, Stanford University, University of California at Berkeley, University of California at Santa Barbara, University of Colorado at Denver, University of Minnesota



Fridge (blue can) connects to icebox through hole in shield (partially constructed, at left)

UNDERSTANDING THE DETECTOR SIGNALS: HOW TO IDENTIFY A WIMP!

The main point: A WIMP (dark matter particle) should produce a vibration signal when it bumps into an atomic nucleus in the detector crystal, but it should NOT produce a significant electronic charge signal since WIMPs themselves have no electronic charge interaction

What do these dots show?

The different colored dots (*see figure below*) represent different types of particles which interacted with one of our detectors. The purple dots are photons (which are just individual bits of light), the green dots represent stray electrons (called "beta rays") and the red dots were produced by neutrons. Each of these particles interacts with our detectors in a different way.

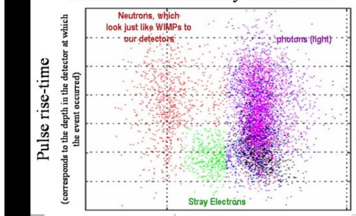
What does this mean?

It shows that **we can distinguish among different particles** by how much electronic charge they knock loose in the detector crystal.

The good news:

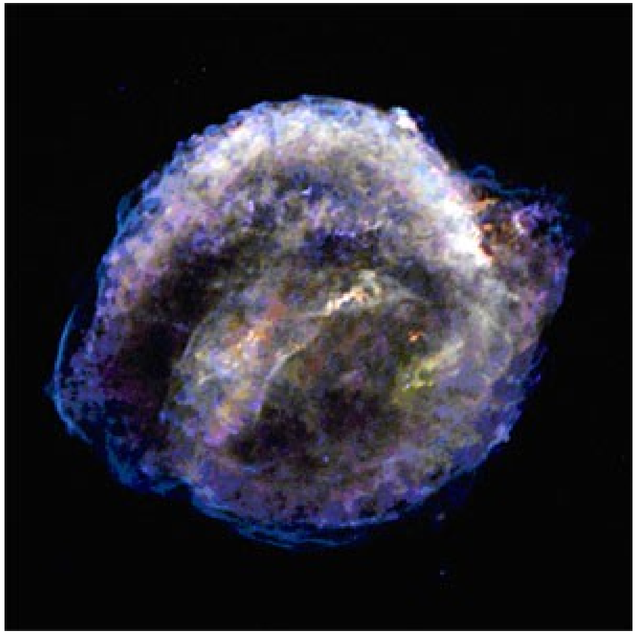
WIMPs should behave very much like neutrons (the red dots). *They displace very little charge in the crystal.* If we can identify neutrons, this means that we can also tell which signals are produced by WIMPs!

In Soudan there should be almost no neutrons around since we are deep underground. So, anything which looks like a neutron is likely to be a WIMP!



Amount of electronic charge displaced in the detector
(Events to the right displaced more charge while the events on the left displaced less)

20. Chandra X-ray observatory ***- Space viewed through X-ray glasses***



The remnants of the supernova created when Kepler (a star named after the famous astronomer Johannes Kepler) exploded. It is ***one of the youngest and brightest recorded supernovae in our Milky Way galaxy***.

Using Chandra data, astronomers have identified ***this as a Type Ia supernova formed when a white dwarf star, made up of carbon and oxygen, becomes unstable and ignites***.

... ***how the burning front of a white dwarf star propagates from the center - the structure of the front is very complicated and small compared to the start itself***.

Scientists at the Max-Planck institute use supercomputers to run ***3-D simulations of Type Ia supernovae***, the largest of which produce several Terabytes of data. ***The simulations predict observable properties, such as light curves and spectra***, which can then be compared to observation data to evaluate the modeling process. Comparing images from Chandra with such as this to the simulations offers the possibility of a more detailed evaluation, which will help refine the models and use them to better understand Type Ia supernovae.

21. STAR experiment

In an ***experiment called STAR***, collaborating scientists in Prague aim to ***recreate the quark-gluon plasma*** (a soup-like state of the matter) ***that permeated the universe less than a second after the Big Bang***. To do this, they analyze data from Brookhaven National Laboratory BNL's high-energy heavy nuclei collisions. Before installation of the Tier2 site at the Nuclear Physics Institute of the Academy of Sciences of the Czech Republic (NPI ASCR) in Prague, STAR collaborators had to connect to BNL remotely each time they needed to retrieve analysis data, and network latencies made this a tedious task.

Researchers at NPI ASCR retrieve the data from BNL via a physical fiber cable running between the two countries that provides Ethernet connectivity at 1 Gigabit line. The Tier2 data transfer framework allows the BNL datasets to be deposited into a "Disk Pool Manager," developed by the LHC Grid Computing project, where Prague collaborators can easily access them using tools developed by Open Science Grid.

22. The Networked Telescope: Progress Toward a Grid Architecture for Pipeline Processing

Pipeline processing systems for modern telescopes are widely considered critical for addressing the problem of ever increasing data rates. This might be particularly important with regard to radio interferometer data in which the ***post-calibration processing*** required to create ***an image for scientific analysis*** - yet not well defined.

BIMA Image Pipeline attempts to address this issue. The pipeline by default is automated and uses NCSA supercomputers to carry out the processing. ***This same system can also be used by the astronomer to create new processing projects*** using data from the archive.

Here are some ways we want to allow users to interact with the pipeline:

(a) **prior to observations:** the astronomer can override default processing parameters to better suit the scientific goals of the project;

(b) **during observations:** the astronomer can monitor the telescope and data via the web;

(c) **after observations:** the astronomer can browse the archive's holdings using customizable displays;

(d) **prior to processing:** the astronomer can create his/her own scripts for reprocessing archival data;

(e) **during processing:** optional viewers can be opened up to monitor, and possibly steer, the deconvolving process.

The processing is carried out using AIPS++. Its event-driven programming model (combined with the toolkit nature of AIPS++) makes it ideal for building automated processing in a distributed environment. ***An important role for NCSA***, as a member of the AIPS++ development consortium, ***is to enable support for parallel processing on a range of mildly to massively parallel machines***, with a particular emphasis ***on Linux clusters***.

****