SymGrid: a Framework for Symbolic Computations on the Grid

Abdallah Al Zain
Heriot-Watt University, Scotland

OSGC, Oakland, California, May 2008

Symbolic Computation Infrastructure for Europe

Overview

1. Introduction
   - Background and Motivation
2. The SymGrid System
   - SymGrid Objectives
   - SymGrid-Services
   - SymGrid-Par
3. *Recent* Single-Cluster and Multicore Results
4. Linking SymGrid with Web Services
5. Conclusions and Further Work

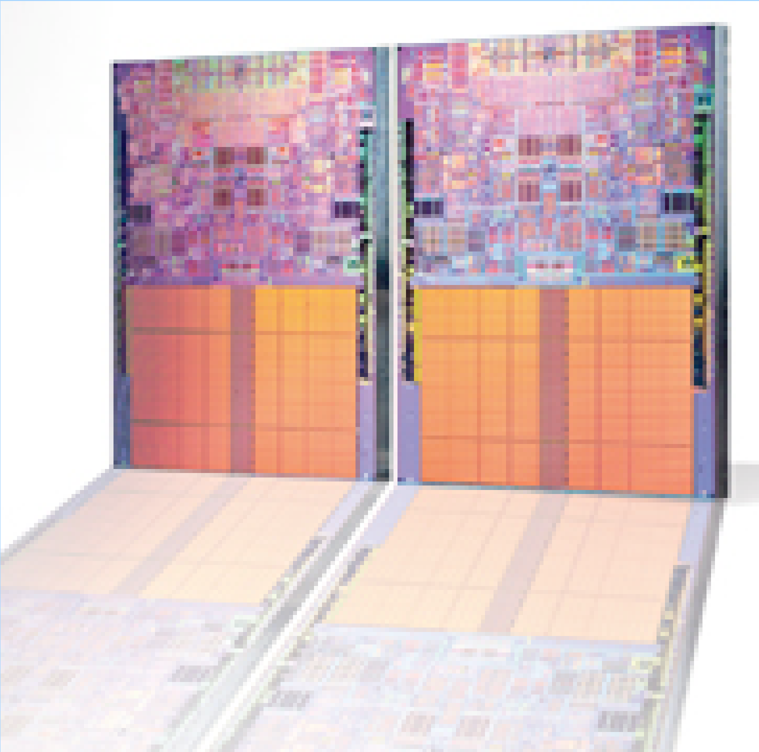
1. Introduction

The Past: Wacky Parallel Machines

Illiac IV, 1962 (or 1974?)

- **Special purpose parallel hardware**
- **Specialist programming notations**
- **Hard to program**
- **Single-processor systems much cheaper and almost as fast**

The Present: Parallel Programming becomes Mainstream



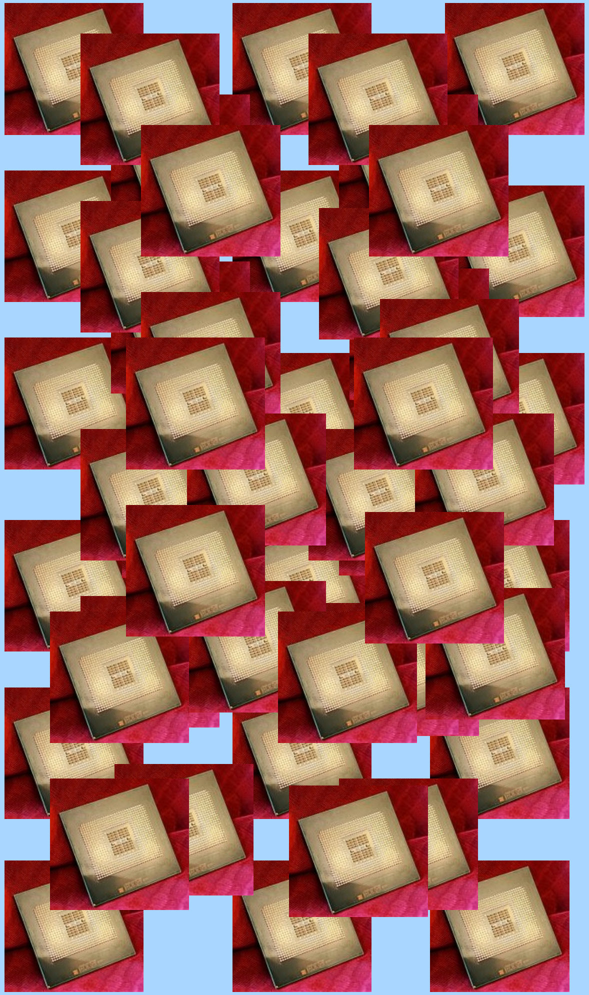
Intel Xeon 5340, Quad Core, 2007

- **All modern architectures are parallel**
- **Standard, but low-level programming notations (C+MPI)**
- ***Still hard to program/exploit***
- **Single processor systems now used only in low-end applications**

The Future: Massive Parallelism …

- 8 Cores
  - available as twin quad-core
- 16 Cores
  - available as four quad-core
- 80 Cores?
  - Intel research project
- 96 Cores?
  - Clearspeed accelerators
- 1000 Cores?

- 1000000 Cores?



SCIENCE
Symbolic Computation Infrastructure for Europe

… and the dream of Grid computing

- Seamless access to computing resources
  - Parallel performance available on demand
  - easy to use
  - no need to configure local systems
    - service-based framework
  - easy access to databases/other system resources
  - pluggable computing power

- Wide-area accessibility
  - integrates national/international infrastructures
  - world-wide availability

The Problems for Large-Scale Parallelism

- **Difficult to program parallel computations**
  - **task identification**
  - **placement**
  - **scheduling**
  - **migration**
  - **…**

- **Symbolic programs can be worse than numeric ones:**
  - **Dynamic data dependencies**
  - **Irregular parallelism (structure and workload)**
  - **Complex symbolic data structures**
  - **Difficult to estimate resource requirements**
  - **…**

Symbolic Computation Infrastructure for Europe

Finding the Right Tool

- Users *may not be aware* of existing tools that solve their problems
- Users may not be able to make the *best choice* of tool
- Not realistic to install *all* possible packages locally
  - Maintaining up-to-date working versions
  - Software licensing issues
- Not possible to understand *all* possible packages
  - need to discover operations performed by symbolic packages
- Need a standard taxonomy/semantic interface (à la Frege)
  - different packages have different names for the same operation
  - different packages wrap data differently

2. SymGrid

The SCIEnce Project:
Symbolic Computing Infrastructure in Europe

Five-year, €3.2M EU FP6 Project
RII3-2005-026133, 2006-2011

9 partners:

Heriot-Watt (UK), St Andrews (UK), IeAT (Romania), RISC (Austria), CNRS (France), Kassel (Germany), TU Berlin (Germany), TU/Eindhoven (the Netherlands), MapleSoft (Canada)

4 Systems Initially

Maple **MuPad** **Kant** **GAP**

SCIENCE Symbolic Computation Infrastructure for Europe

"It is reasonable to expect that in the year 2010, the predominant way of doing math will no longer be by pen and paper, but in an integrated web-based math-development system that supports the mathematician in all aspects of mathematics. "

Michael Kohlhase.

“And it will be parallel” Abyd Al Zain

SCIENCE
Symbolic
Computation
Infrastructure for
Europe

Overall SymGrid Objectives

1. produce a portable framework (SymGrid) that will
   - allow symbolic computations *to access* Grid services
   - allow symbolic components *to be exploited* as part of larger Grid service applications on a computational Grid
2. develop resource brokers that will support the irregular workload and computation structures that are frequently found in symbolic computations
3. implement a series of applications that will demonstrate the capabilities and limitations of Grid computing for symbolic computations

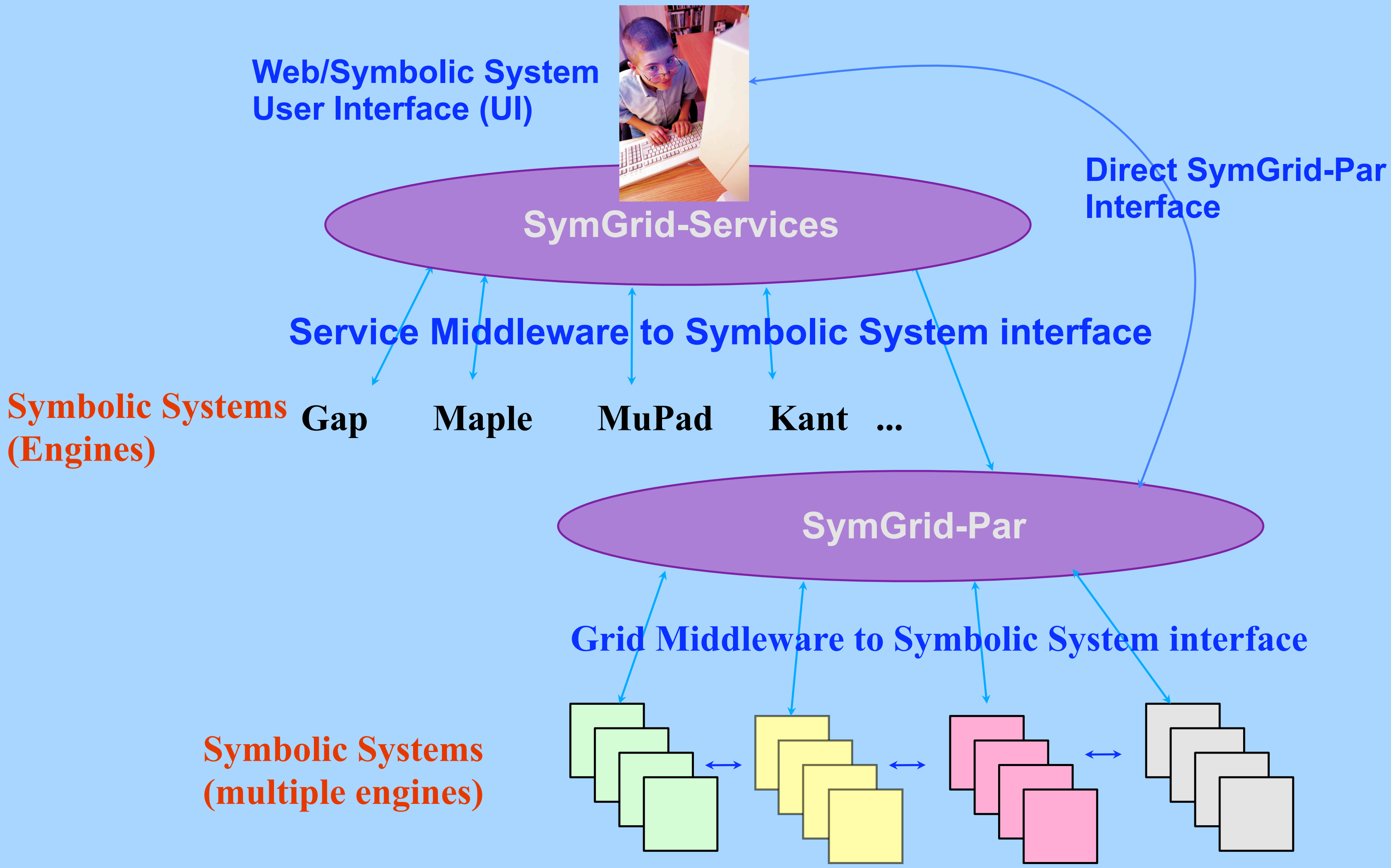
K. Hammond , A. Al Zain , G. Cooperman , D. Petcu , P.W. Trinder:
SymGrid: a Framework for Symbolic Computation on the Grid.
EuroPar 2007*, Rennes, France, Springer LNCS,* August 2007

SCIENCE Symbolic Computation Infrastructure for Europe

Novel Features

- Integrate multiple symbolic systems in a single framework
  - systems heterogeneity
  - platform heterogeneity
- Sophisticated integrated *computational steering* interface
- New domain-specific patterns of parallel computation
  - user-centric view of parallelism
- New user community
  - may have massive computational demands
  - yet exposure to parallelism/Grids is not common

SymGrid Design Overview



Access to Grid Services: SymGrid-Services

- Middleware to allow generic access to symbolic Grid services.
- Service discovery for symbolic Grid components.
- Higher-level system interfaces allow straightforward SymGrid access from within symbolic systems. Designed in association with the providers of the Maple, GAP, MuPAD and Kant systems in order to meet the generic requirements of symbolic system providers.
- Standard *OpenMath* protocols used to transmit mathematical data (SCSCP)
  - XML and native formats
- Services may be constructed from heterogeneous components
- Services may be *steered* interactively through an integrated user interface

- A. Carstea, M. Frincu, G. Macariu, D. Petcu, K. Hammond: Generic Access to Web and Grid-based Symbolic Computing Services – the SymGrid-Services Framework. *International Symp. on Parallel and Distributed Computing, Linz, Austria, IEEE Press,* July 2007

SCIENCE Symbolic Computation Infrastructure for Europe

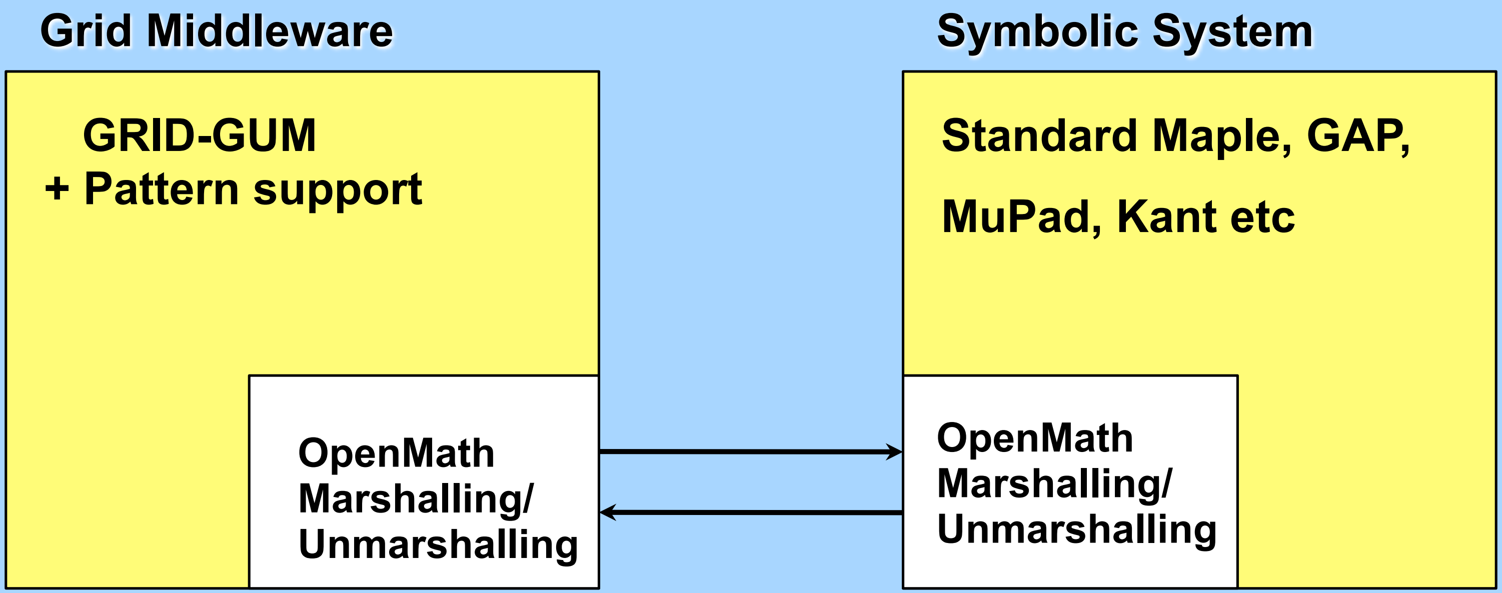
Symbolic Grid Components: SymGrid-Par

- **Orchestrates symbolic components into (possibly parallel) Grid apps.**
- **Built around GRID-GUM (J. Symb. Comp, 2006)**
  - **ultra-lightweight thread *(filament)* creation**
  - **distributed virtual shared memory**
  - **multi-level scheduling support**
  - **automatic thread placement**
  - **automatic datatype-specific marshalling/unmarshalling**
  - **extends parallel implementation of Haskell**
- **Links with SymGrid-Services to allow complete Grid-enabled applications to be constructed.**

A. Al Zain, P. Trinder, G. Michaelson, and H-W. Loidl: mGrid-Par: Evaluating a High-Level Parallel Language (GpH) for Computational GRIDs.
IEEE Transactions on Parallel and Distributed Systems, 19(2):219--233, 2008.

SCIENCE
Symbolic Computation Infrastructure for Europe

Grid Middleware to Symbolic System Interface



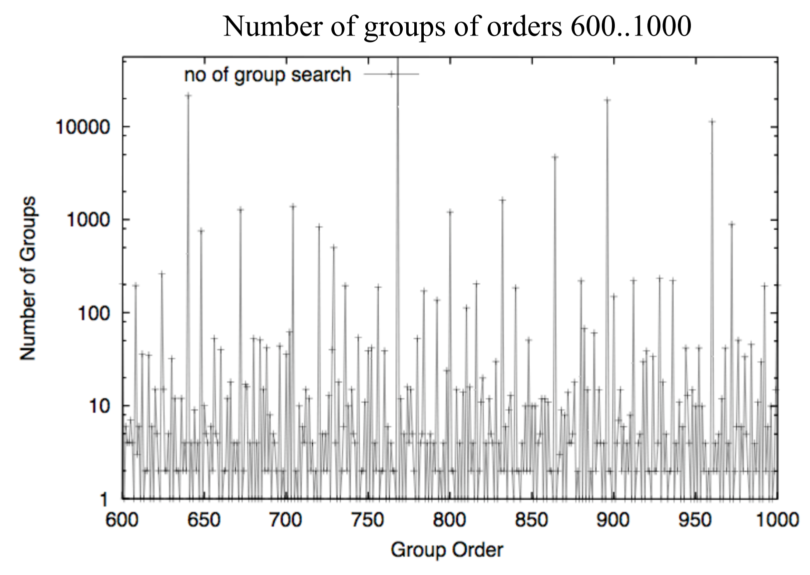
Symbolic Computation Infrastructure for Europe

3. Results

Example: smallGroup using the GAP system

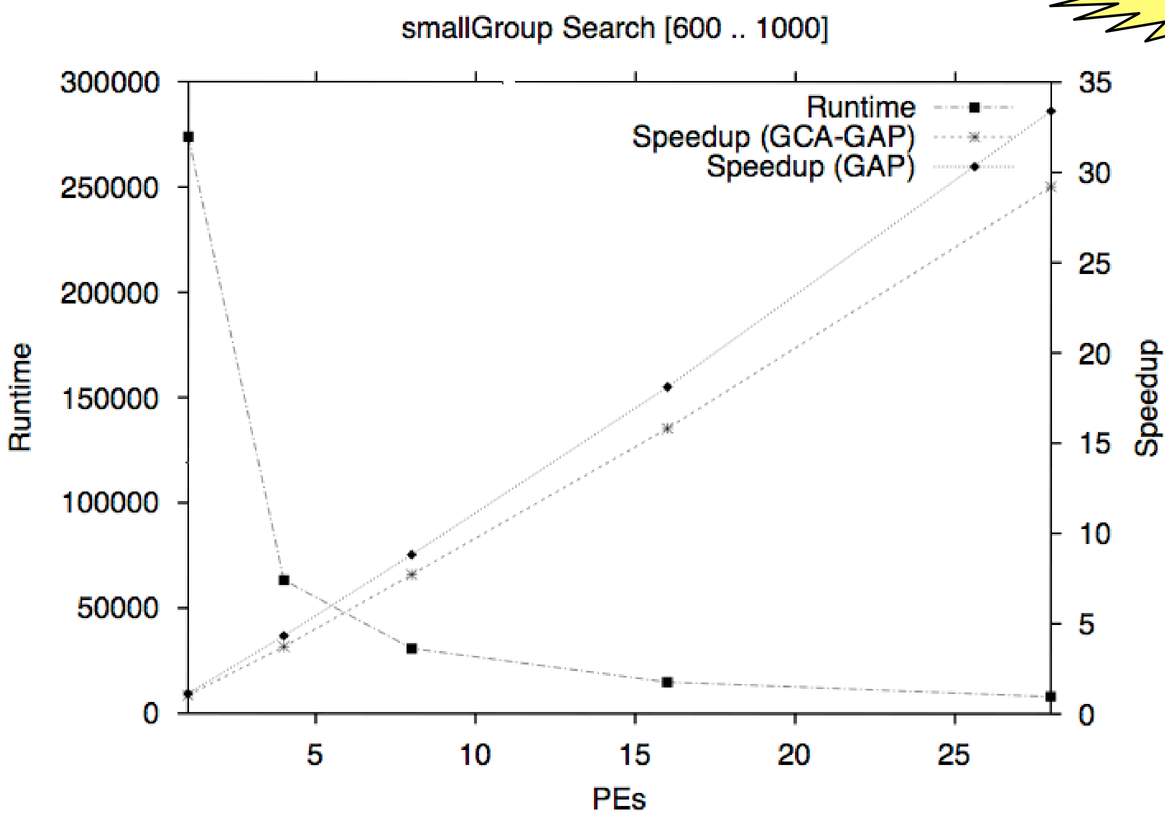
- **Search for *groups* with orders within a given range of values, s.t. the average order of the elements is an integer**
- **Two nested levels of parallelism: search across/within families of groups of the same order**
- **Two kinds of *irregularity:***
  - ***different complexity for different groups***
  - ***variation in number of groups of each order***

Irregularity in Number of Groups



Single-Cluster Results

1,163,006 candidate groups

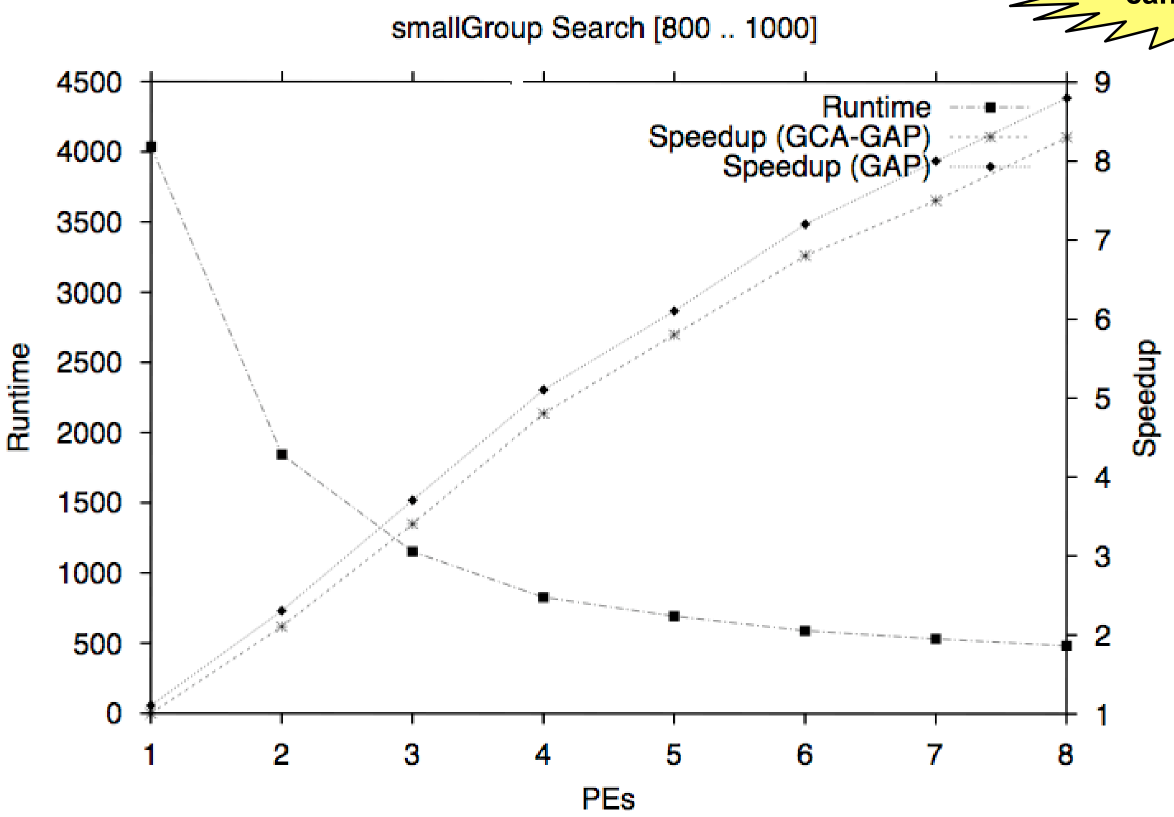


32 x Intel Pentium IV @ 3.0GHz. 100Mb/s Ethernet network. Fedora Linux, kernel version 2.6.10-1.

Symbolic Computation Infrastructure for Europe

Multicore Results

42,473 candidate groups



8-core Dell Poweredge 2950, 2 x Intel Xeon 5355 quad-core @ 2.66GHz
16GB fully-buffered 667MHz DIMMs. Centos Linux 4.5.

SCIENCE
Symbolic
Computation
Infrastructure for
Europe

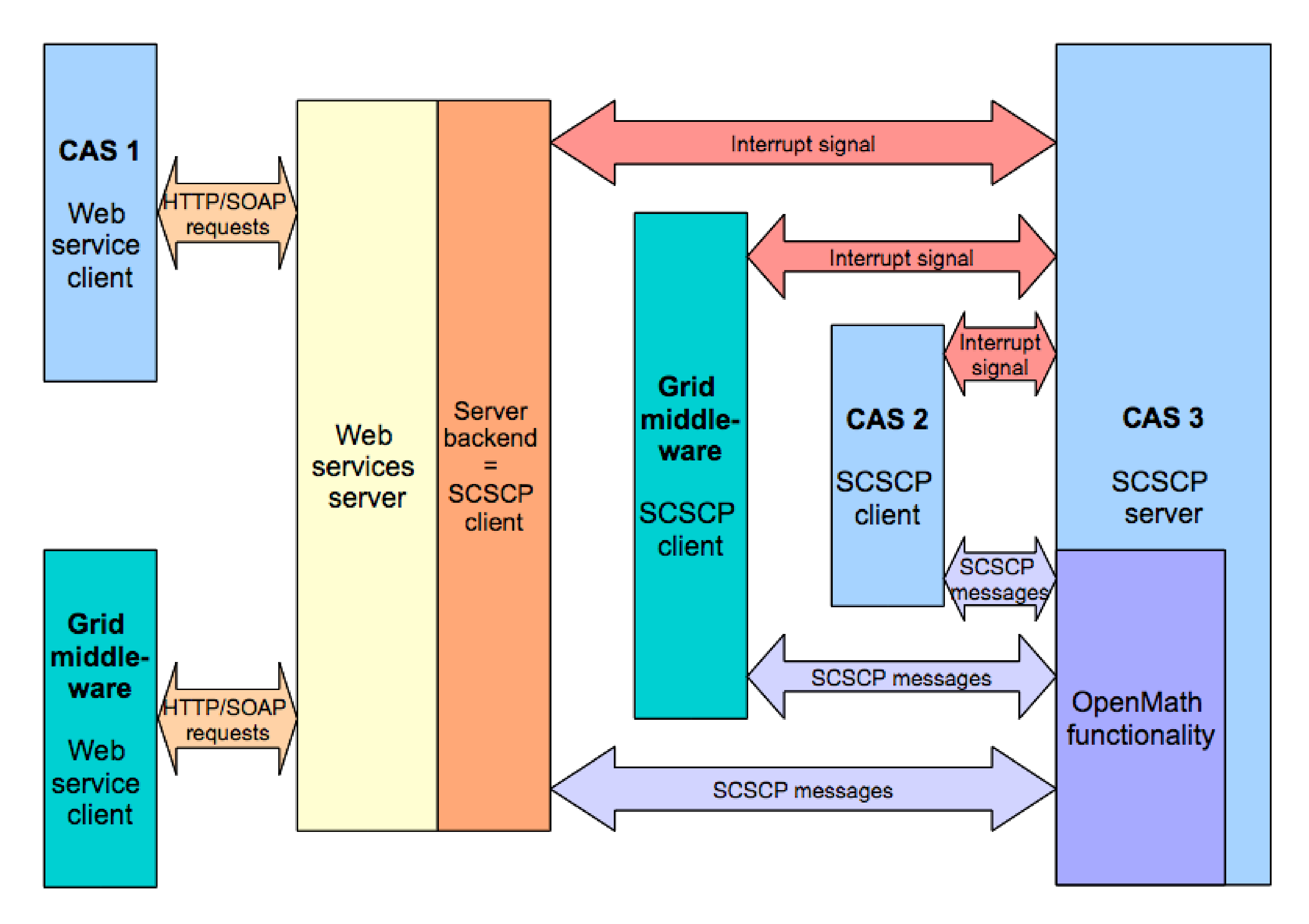
4. Composing Grid and Web Services

SCSCP:
an OpenMath-based communication protocol

- **Using the Symbolic Computation Software Composability Protocol (SCSCP), a computer algebra system (CAS) may offer services for the following clients:**

  i. **A Web server which passes on the services as Web services using SOAP/HTTP protocols to other clients**
  ii. **Grid middleware (e.g. SymGrid)**
  iii. **Another instance of the same CAS**
  iv. **Another CAS running on the same computer or remotely**

The BIG Picture:
Linking Grid + Web through SCSCP



SCIENCE
Symbolic
Computation
Infrastructure for
Europe

5. Conclusions and Further Work

SCIENCE Symbolic Computation Infrastructure for Europe

Conclusions

- New framework for symbolic computation
  - builds on sophisticated, flexible runtime system, GRID-GUM
- Highly generic
  - two symbolic systems now working: GAP and Maple
  - other symbolic systems possible (e.g. Mathematica, …)
  - *minimal/no change to underlying sequential system*
- Supports heterogeneity
  - systems built from multiple components
- Promising parallel results on single cluster/multi-core
- SCSCP allows linkage to web services

SCIENCE
Symbolic
Computation
Infrastructure for
Europe

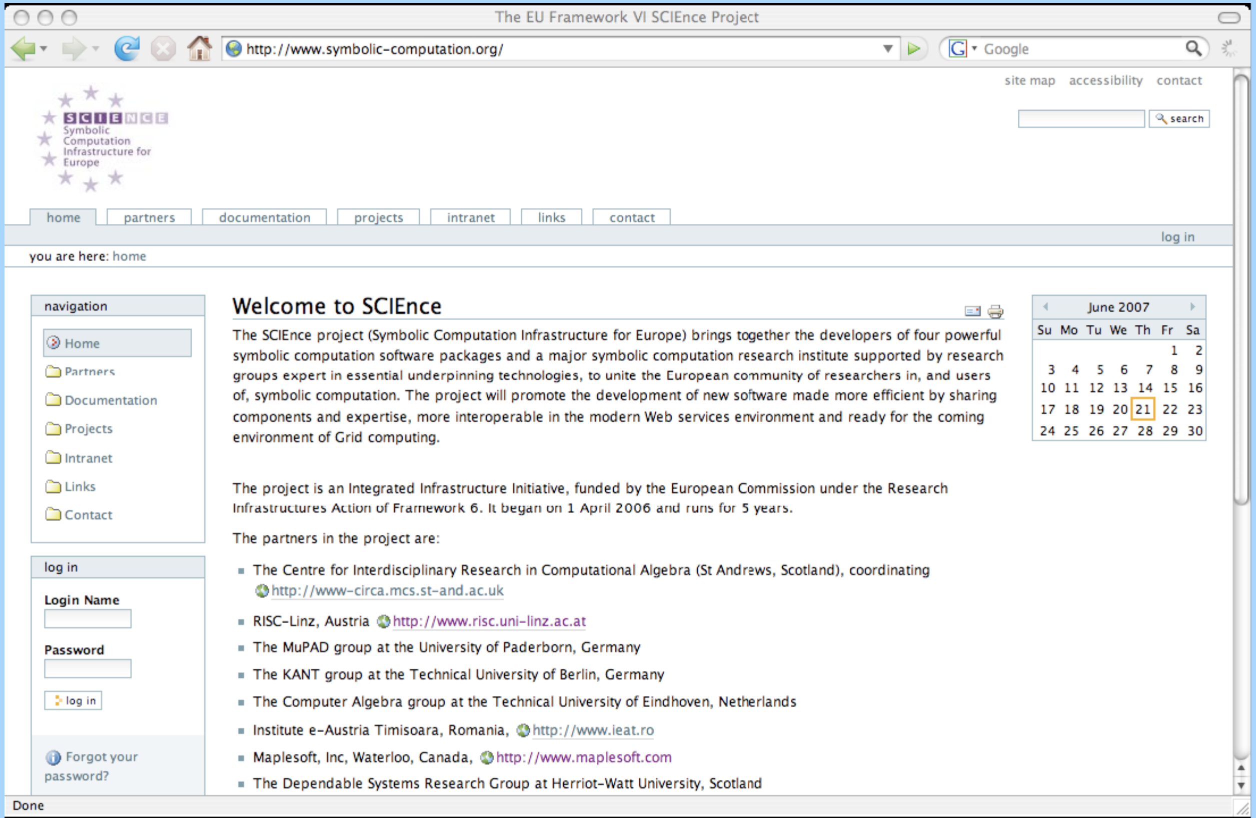
Future/Ongoing Work

- **Multi-cluster, geographically distributed**
  - **In hand: mainly a matter of practicalities/time**
  - **Grid clusters in Scotland, Germany, *Romania***
  - **Determine whether single-cluster results can be replicated**
- **Confirm parallelism results for Maple, …**
- **Extend to MuPad, Kant, (*your favourite system*)**
  - **New UIs for each system**
  - **New APIs for each system**
- **Develop heterogeneous symbolic applications**

SCIEnce Training schools at RISC, 2007-2010

- The 1st school was held on February 5-18, 2007
- The 2nd school was held on June 25 – July 6, 2007
- One school each year in 2008–2010

- http://www.symbolic-computation.org



Questions