John von Neumann Institute for Computing

Air Pollution Forecast on the HUNGRID Infrastructure

R. Lovas, J. Patvarczki, P. Kacsuk, I. Lagzi, T. Turányi,
L. Kullmann, L. Haszpra, R. Mészáros, A. Horányi,
A. Bencsura, Gy. Lendvay

published in

Parallel Computing:
Current & Future Issues of High-End Computing,
Proceedings of the International Conference ParCo 2005,
G.R. Joubert, W.E. Nagel, F.J. Peters, O. Plata, P. Tirado, E. Zapata (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **33**, ISBN 3-00-017352-8, pp. 121-128, 2006.

http://www.fz-juelich.de/nic-series/volume33

Air pollution forecast on the HUNGRID infrastructure

R. Lovas[a], J. Patvarczki[a], P. Kacsuk[a], I. Lagzi[b], T. Turányi[b], L. Kullmann[c], L. Haszpra[c], R. Mészáros[d], A. Horányi[c], Á. Bencsura[e], Gy. Lendvay[e]

[a]Computer and Automation Research Institute, Hungarian Academy of Sciences (MTA SZTAKI), Budapest, Hungary

[b]Department of Physical Chemistry, Institute of Chemistry, Eötvös University, Budapest, Hungary

[c]Hungarian Meteorological Service, Budapest, Hungary

[d]Department of Meteorology, Eötvös University, Budapest, Hungary

[e]Institute of Chemistry, Chemical Research Center of the Hungarian Academy of Sciences, Budapest, Hungary

Computational Grid systems are gaining more and more attention in the natural sciences but very often the end-users (biologists, chemists, physics) must tackle various problems when they want to deploy such systems. In this paper a unified software development family is presented, which is able to cover each stage of parallel software development as well as the seamless application migration from parallel systems to Grid platforms. Besides the recently established HUNGRID infrastructure, the development life-cycle is also presented through two air-pollution modelling applications, which enable the authorities to prevent the harmful effects of high-level ozone concentration and accidental releases. The developed computational models can play crucial role in the management of the photochemical smog episodes; they can be used to test the effects of the ozone fluxes and possible emission control strategies and accidents.

1. Introduction

Computational Grid systems [1] are gaining more and more attention in the natural sciences. In such systems, a large number of heterogeneous computer resources are interconnected to solve complex problems. The main aim of the national research project, 'Chemistry Grid and its application for air pollution forecast' [2], was the investigation of the feasible applications of Grid technology in computational chemistry from practical aspects; e.g. prevention of the harmful effects of high-level ozone concentration. The project relied on an academic product family of MTA SZTAKI; a Grid monitoring tool, called Mercury [3], and two integrated application development environments, called P-GRADE parallel programming environment [4] (see Figure 2), and P-GRADE Grid portal [5] (see Figure 3). These tools enable the parallelisation and 'gridification' of sequential applications in a more flexible and transparent way than other solutions [2,6,7] by means of their high level graphical approach, multi-grid support, performance analyzer and debugging tools. Recently, the P-GRADE portal has been developed further to provide support for the efficient execution of complex programs in various Grids. It includes the dynamic execution of applications across the Grid resources according to the actual state and availability conditions provided by the new information/broker system. In the project, the consortium applied these achievements for supporting a specific e-Science area; the established infrastructure can provide access for chemists to Hungarian computational Grid resources, called HUNGRID, which is not only a virtual organization within the EGEE project [8] but it contains new elements, such as P-GRADE portal [5] and MERCURY [3],

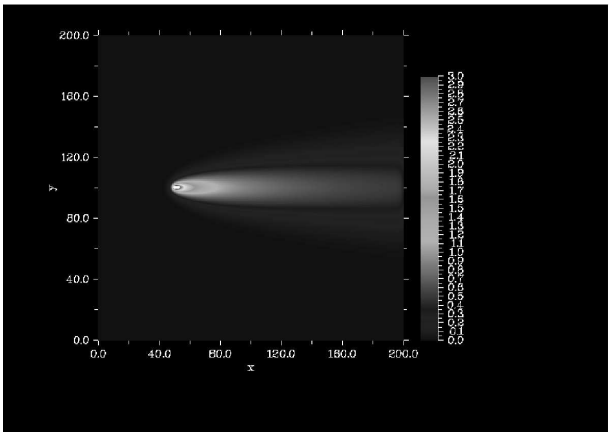

Figure 1. Result of simulation: distribution of chemical (radioactive) substances.

making easier the use of the infrastructure for solving complex problems, such as modelling of air pollution.

2. Application I: Accidental release of chemical substances

Modelling the accidental release of chemical (or radioactive) substances from a single source requires that the numerical simulations must be achieved obviously faster than in real case in order to use them in decision support. A feasible way is the parallelization of source code. Evolution of chemical species can be described by second-order partial differential equations:

$$\frac{\partial c_i}{\partial t} = K_{x,i}\frac{\partial^2 c_i}{\partial x^2} + K_{y,i}\frac{\partial^2 c_i}{\partial y^2} - u\frac{\partial c_i}{\partial x} - v\frac{\partial c_i}{\partial y} + R_i(c_1, c_2, \ldots, c_n), \quad i = 1, 2, \ldots, n, \tag{1}$$

where c_i is the concentration, $K_{x,i}$, $K_{y,i}$ are the turbulent diffusion coefficients, u, v are the components of the horizontal wind velocity and R_i is the chemical reaction term, respectively, of the ith chemical species. t is time, and x and y are the spatial variables. The chemical reaction term R_i may contain non-linear terms in c_i. For n chemical species, an n dimensional set of partial differential equations is formed describing the change of concentrations over time and space. These equations are coupled through the non-linear chemical reaction term.
The basis of the numerical method for the solution of the partial differential equations is the spatial discretisation of the partial differential equations on a two-dimensional rectangular grid. In these calculations, the grid spacing is uniform in both spatial directions. This approach, known as the 'method of lines', reduces the set of partial differential equations (PDEs) of three independent variables (x, y, t) to a system of ordinary differential equations (ODEs) of one independent variable, time. A second order Runge-Kutta method is used to solve the system of ODEs arising from the discretisation of the transport terms with chemistry.

2.1. Parallel implementation in P-GRADE

The graphical language of P-GRADE consists of three hierarchical design layers: (i) Application Layer is a graphical level, which is used to define the component processes, their communication ports as well as their connecting communication channels. Shortly, the Application Layer serves for describing the interconnection topology of the component processes or process groups (see Figure 2,

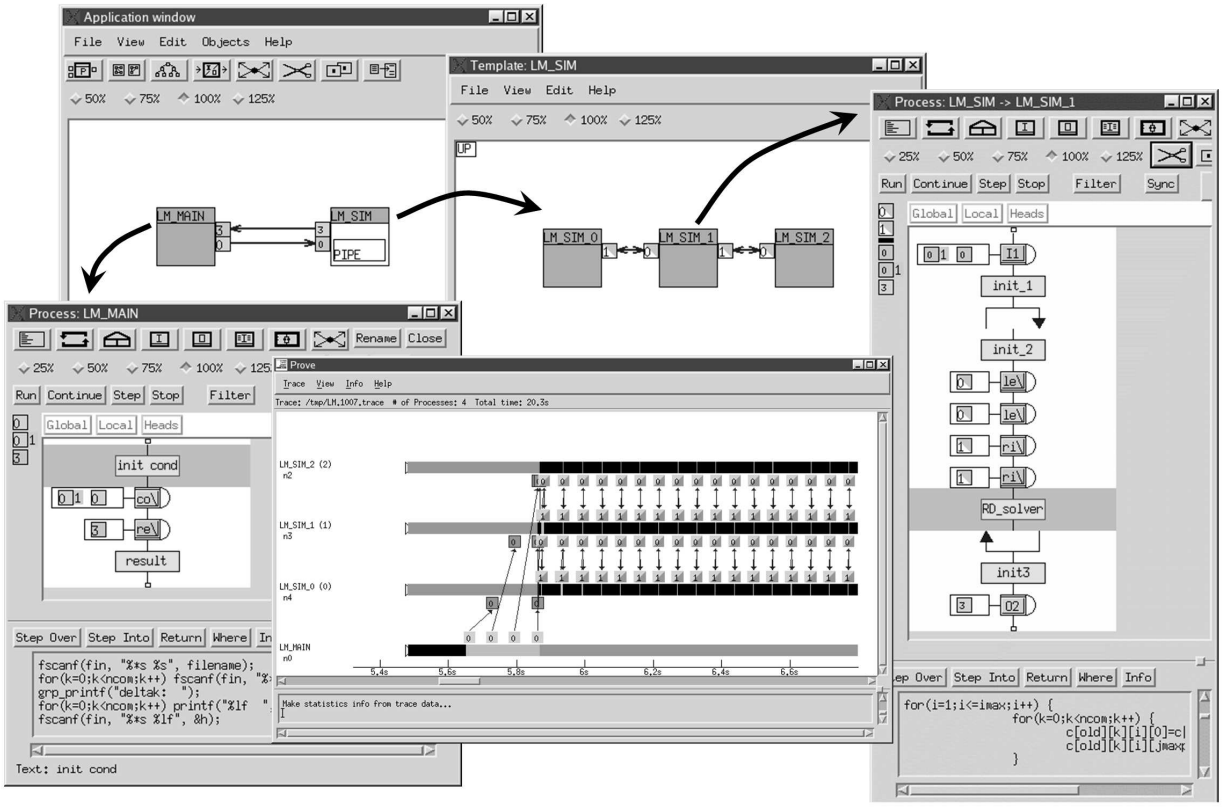

Figure 2. Accidental release simulation: parallel code in P-GRADE development environment.

Application window). (ii) Process Layer is also a graphical level where different types of graphical blocks are applied: loop construct (see Figure 2, blue arcs in the window labelled Process: *LM_SIM* → *LM_SIM_1*), conditional construct, sequential block, input/output activity block and macrograph block. The graphical blocks can be arranged in a flowchart-like graph to describe the internal structure (i.e. the behaviour) of individual processes (see Figure 2, Process windows). (iii) Text Layer is used to define those parts of the program that are inherently sequential and hence only pure textual languages like C/C++ or FORTRAN can be applied at the lowest design level. These textual codes are defined inside the sequential blocks of the Process layer (see Figure 1, at bottom of Process window labelled Process: *LM_SIM* → *LM_SIM_1*).

In order to parallelise the sequential code of the presented accidental release simulation the domain decomposition concept was followed; the two-dimensional grid is partitioned along the x space direction, so the domain is decomposed into horizontal columns. Therefore, the two-dimensional subdomains can be mapped onto e.g. a pipe of processes (see Figure 2, Template: *LM_SIM* window). An equal partition of sub-domains among the processes gives us a well balanced load during the solution of the reaction-diffusion-advection equations assuming a homogeneous cluster (see Figure 2, PROVE performance visualisation window) or a dedicated supercomputer as the execution platform. During the calculation of the diffusion of the chemical species communications are required to exchange information on the boundary concentrations between the nearest neighbour sub-domains, which are implemented via communication ports, channels (see Figure 2, Template: *LM_SIM* window, arcs between small rectangles), and communication actions (see Figure 2, Process: *LM_SIM* → *LM_SIM_1*, icons labelled as '*le*' and '*ri*' in the control flow like description). For the calculation the

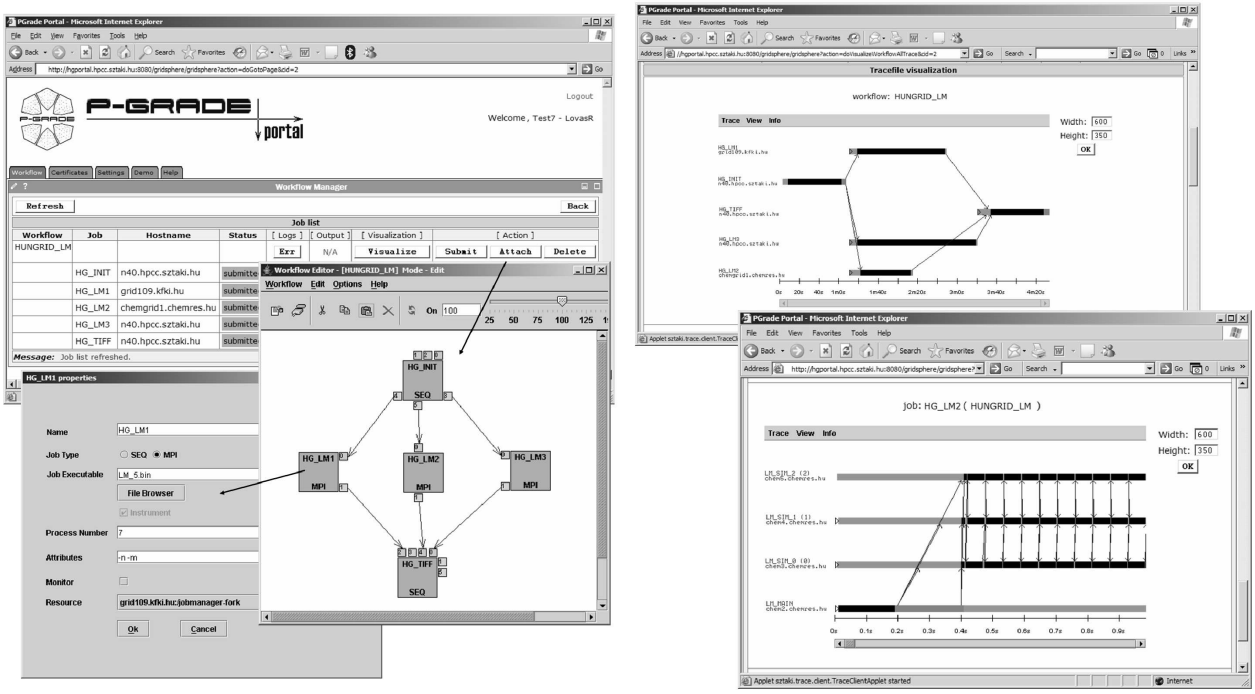

Figure 3. Accidental release simulation on HUNGRID: workflow and job descriptions in P-GRADE portal and execution visualisation at both levels.

process invokes sequential code segments (see Figure 2, bottom of Process: *LM_SIM* → *LM_SIM_1* windows). Based on the hierarchical graphical description and the given sequential code segments, we generated automatically the instrumented MPICH code with the P-GRADE environment. In the performance analysis phase, the underlying MERCURY monitor [3] collects the trace information for the on-line performance visualization. For illustration purposes, in Figure 2 the PROVE window depicts the space-time diagram of the execution with 4 processes, where the user can inspect the behaviour of the application; e.g. the initialization phase, the multicast of input parameters (blue arcs between the process bars), or the periodic exchanges of boundary conditions between the processes of pipe communication template. The black colour in process bars represents the periods when the process executes the sequential code segments of calculation, so it can be easily recognized that the simulation can perform well. The parallel version of reaction-diffusion-advection simulation has been tested and fine tuned similarly to the earlier developed chemical simulations [9] on SZTAKI cluster (a part of HUNGRID infrastructure) using it as a dedicated resource. This self-made Linux cluster contains 29 dual-processor nodes (Pentium III/500MHz) connected via Fast Ethernet.

2.2. Workflow description and execution on HUNGRID with P-GRADE portal

P-GRADE portal [5] is a workflow-oriented Grid portal with the main goal to enable users to manage the whole lifecycle of workflow-oriented complex grid applications. The P-GRADE portal supports the graphical development of workflows (see Figure 3, Workflow editor) created from various types of existing components (sequential, MPI or PVM jobs), executing these job-workflows in the Grid relying on user credentials, and finally analyzing the monitored trace-data by the built-in visualization facilities. The P-GRADE Portal provides the following functions: Creation and modification of workflow applications; Setting the Grid environment; Managing Grid certificates;

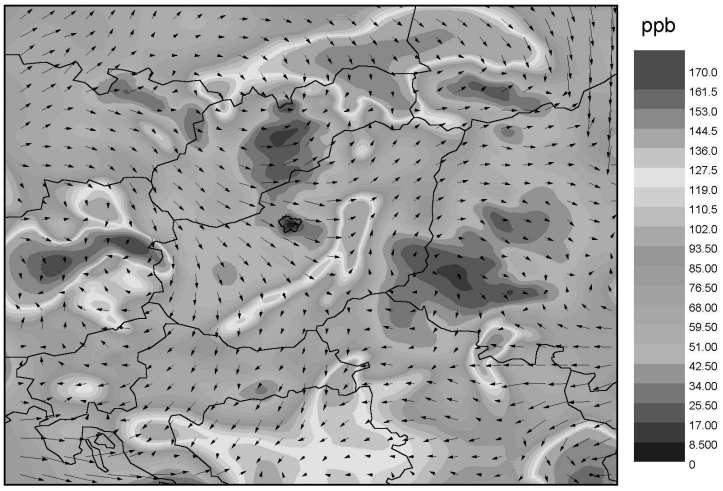

Figure 4. Calculated ozone concentrations on the 3rd of August, 1998 at 17.00 with wind field originated from the ALADIN weather prediction model.

Managing the execution of workflow applications on grid resources; Visualizing the progress of workflows and their component jobs (see Figure 3).
The simulation was executed on the HUNGRID infrastructure, which is the Hungarian Virtual Organisation (VO) of the EGEE Grid [8] based on LCG-2 with some extensions developed as part of the Hungarian Grid research efforts. HUNGRID was created in the framework of the Hungarian Grid project funded by IHM and has got about 500 machines from 5 sites and two other sites have already decided to join with another 150 machines.
The presented workflow version of accidental release simulation contains 5 jobs. The first sequential job, HG_INIT (see Figure 3) initialize the input parameters for the different simulation scenarios with varying emission factors, wind speed, altitude of the source, time limit, etc. Then, three parallel jobs, HG_LM1 HG_LM3 calculate the effects of the different accidental release scenarios. These parallel jobs were developed by the P-GRADE development environment (see Section 2.1), and the instrumented MPICH executables were generated by the P-GRADE run-time system. Finally, the last sequential job, HG_TIFF is responsible for the creation of emission diagrams based on the calculated results (see Figure 1).
The execution time was reduced radically, since each simulation job was executed on different sites of HUNGRID VO in parallel (see Figure 3, top window in right), the simulation jobs were parallel applications itself (see Figure 3, bottom window in right), and the communication and initialization overhead was quite acceptable as they can be seen in Figure 3. In the frame of SEEGRID projects, the accidental release simulation with 8 parallel jobs has been demonstrated successfully with 8 different sites collected from South-European countries at MIPRO conference [10].

3. Application II: Photochemical air pollution

The phytotoxic nature of ozone was recognized decades ago. Due to high emissions of ozone precursor substances, elevated ozone concentrations may cover large areas of Europe for shorter (episodic) or longer periods under certain meteorological conditions. These elevated concentrations

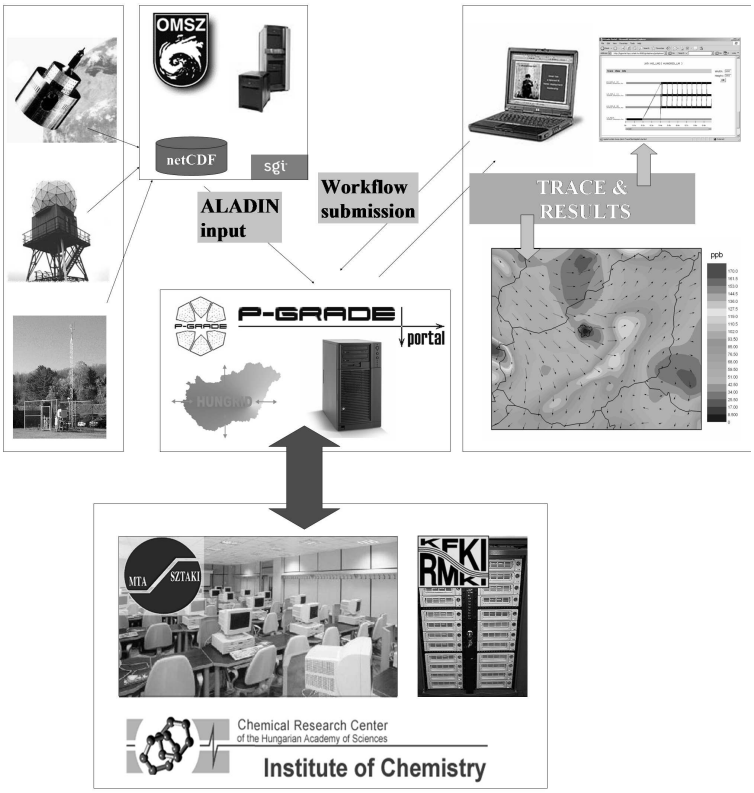

Figure 5. Air pollution forecast on the HUNGRID infrastructure.

can be potentially damaging to agricultural and natural vegetation. Occasional extreme concentrations may cause visible injury to vegetation, while long-term exposure, averaged over the growing season, can result in decreased productivity and crop yield. For the computational study of this phenomenon in Hungary, a coupled Eulerian photochemical reaction-transport model and a detailed ozone dry-deposition model were developed. The Eulerian air pollution model was developed through a co-operation between the Eötvös University, Budapest, The University of Leeds and the Hungarian Meteorological Service [11–13].

This model fully utilized the experience collected previously at the Leeds University on the use of adaptive gridding methods for modelling chemical transport from multi-scale sources. The model has been elaborated within a flexible framework where both area and point pollution sources can be taken into account, and the chemical transformations can be described by a mechanism of arbitrary complexity. The reaction-diffusion-advection equations relating to air pollution formation, transport and deposition are solved on an unstructured triangular grid. The model domain covers Central Europe including Hungary, which is located at the centre of the domain and is covered by a high-resolution nested grid. The sophisticated dry-deposition model estimates the dry-deposition velocity of ozone by calculating the aerodynamics, the quasi-laminar boundary layer and the canopy resistance. The meteorological data utilized in the model were generated by the ALADIN meso-scale limited-area numerical weather prediction model (see Figure 5), which is used by the Hungarian Meteorological Service [14]. For Budapest, the emission inventories for CO, NOx and VOCs were provided by the local authorities with a spatial resolution of 1 km 1 km and also include the most significant 63 emission point sources. For Hungary, the National Emission Inventory of spatial res-

olution 20 km 20 km was applied which included both area and point sources. Outside Hungary, the emission inventory of EMEP for CO, NOx and VOCs was used, having a spatial resolution of 50 km 50 km.
The work demonstrates that the spatial distribution of ozone concentrations is a less accurate measure of the effective ozone load than the spatial distribution of ozone fluxes. The fluxes obtained show characteristic spatial patterns, which depend on soil moisture, meteorological conditions, ozone concentrations and the underlying land use. The simulation of photochemical air pollution is based on the presented approach (see Section 2.1) as well as the experiences concerning the earlier developed P-GRADE version of ultra-short range weather prediction system [10]. The major difference is that their simulation jobs have not been parallelized due to the unavailable source code of some third-party modules in the model. Thus, we were not able to take advantages of multi-level parallelism, only the workflow level, which looks similar to the presented case (see Section 2.1). As it is depicted in Figure 5, the P-GRADE portal server is in the centre of air pollution simulation and dedicated to HUNGRID infrastructure. Currently, it provides access to three clusters located at different academic institutes; MTA SZTAKI, KFKI-RMKI, and CRC-HAS. The portal server has access to the meteorological data (ALADIN input files) as well, which are calculated numerically by the Hungarian Meteorological Service based on the available radar and satellite images, the observations (see Figure 5, left side), and results of other models. The portal server can be accessed remotely by submitting the simulations, i.e. the P-GRADE workflows, and by downloading the visualization of execution traces (see Figure 3, right side) and simulation results (see Figure 4) on the local machine.

4. Summary and related works

Several institutes apply clusters for environmental modelling (e.g. [15]), and some of them have already recognised the advantages of Grid technology (e.g [16]). There are several available workflow managers and Grid portals for e-Science [17,7,16] but the presented P-GRADE development environment together with the P-GRADE Grid portal provide one of the most flexible and unified ways for the parallel application development and deployment on various Grids. In this paper, we demonstrated that the HUNGRID platform with the presented development tools can provide efficient user support for complex applications, e.g. for air pollution forecasting. On the other hand, P-GRADE portal v2.1 is already working as service for HUNGRID (operated by SZTAKI), SEE-Grid (operated by SZTAKI) [10], and the UK National Grid Service (operated by University of Westminster) [18]. Recently, the developer alliance decided to start supporting the Hungarian ClusterGrid [19], the Croatian Grid and the Turkish Grid as well. Moreover, P-GRADE portal is already connected to the EU GridLab testbed [20], and the UK OGSA testbed [21] for demonstration purposes. Therefore, the P-GRADE portal (as one of its most outstanding features) is able to provide seamless multi-grid access to the end-users, and the application developers do not have to tackle the variety of emerging grid technologies and standards, such as Condor, Globus Toolkit, OGSA, GAT, LCG or gLite.

Acknowledgement

The research described in this paper has been supported by the following projects and grants: Hungarian IHM 4671/1/2003 project, Hungarian OTKA T042459, T043770, and D048673 grants, OTKA Instrument Grant M042110, Hungarian IKTA OMFB-00580/2003, GVOP-3.1.1-2004-05-0359/3.0, and EU INFSO-RI-508833 projects.

References

[1] I. Foster, C. Kesselman: Computational Grids, Chapter 2 of The Grid: Blueprint for a New Computing Infrastructure. Morgan-Kaufman. 1999.

[2] R. Lovas, I. Lagzi, L. Kullmann, Á. Bencsura: Chemistry GRID and its Applications for Air Pollution Forecasting. ERCIM News, 61, 18-19, 2005.

[3] N. Podhorszki, Z. Balaton, G. Gombás: Monitoring Message-Passing Parallel Applications in the Grid with GRM and Mercury Monitor, 2nd European Across Grids Conference, AxGrids 2004, Nicosia, Cyprus, 179-181, 2004.

[4] P. Kacsuk, et al.: PGRADE: a Grid Programming Environment. Journal of Grid Computing, 1, 171-197, 2003.

[5] Cs. Németh, G. Dózsa, R. Lovas, P. Kacsuk: The P-GRADE Grid portal. Computational Science and Its Applications - ICCSA 2004: International Conference, Assisi, Italy, Lecture Notes in Computer Science, 3044, 10-19, Springer-Verlag, 2004.

[6] O. Gervasi, A. Lagana: SIMBEX: a portal for the a priori simulation of crossed beam experiments, Future Generation Computer Systems. The International Journal of Grid Computing: Theory, Methods and Applications, 20, 703-715, 2004.

[7] I. Taylor, I. Wang, M. Shields, S. Majithia: Distributed computing with Triana on the Grid. Concurrency and Computation: Practice and Experience, 17, 1-18, 2005.

[8] http://www.eu-egee.org

[9] R. Lovas, P. Kacsuk, I. Lagzi, T. Turányi: Unified development solution for cluster and grid computing and its application in chemistry. Computational Science and Its Applications - ICCSA 2004: International Conference, Assisi, Italy, Lecture Notes in Computer Science, 3044, 226-235, Springer-Verlag, 2004.

[10] J. Patvarczki, et al.: Using the P-GRADE Grid Portal in the SEE-GRID project. 28th International Convention Mipro, Opatija, Adriatic Coast, Croatia 2005.

[11] A. Tomlin, et al.: On the use adaptive gridding methods for modelling chemical transport from multi-scale sources, Atmospheric Environment, 31, 2945-2959, 1997.

[12] I. Lagzi, et al.: The simulation of photochemical smog episodes in Hungary and Central Europe using adaptive gridding models, Lecture Notes in Computer Science, 2074, 67-77, 2001

[13] I. Lagzi, et al.: Modelling ozone fluxes over Hungary, Atmospheric Environment, 38, 6211-6222, 2004

[14] A. Horányi, I. Ihász, G. Radnóti: ARPEGE/ALADIN: A numerical weather prediction model for Central-Europe with the participation of the Hungarian Meteorological Service. Idöjárás, 100, 277-301, 1996.

[15] J. L. Palau, G. Pérez-Landa, M.M. Millán: Air pollution modelling in complex terrain: 'Els Ports-Maestrat' regional-scale study. ERCIM News, 61, 26-27, 2005.

[16] L. Fusco, J. van Bemmelen: Earth observation archives in digital library and grid infrastructures. Data Science Journal, 3, 222-226, 2004.

[17] M. Vanneschi: The programming model of ASSIST, an environment for parallel and distributed portable applications. Parallel Computing, 28, 1709-1732, 2002.

[18] A. Goyeneche, et al.: Experiences with Deploying Legacy Code Applications as Grid Services Using GEMLCA, Proc. of European Grid Conference 2005, Lecture Notes in Computer Science, 3470, Springer, Amsterdam, 851-860, 2005

[19] G. Sipos et al.: Connecting ClusterGrid and P-GRADE Portal. NETWORKSHOP 2005 Conference, Szeged, Hungary, 2005.

[20] http://www.gridlab.org

[21] T. Kiss, et al: Legacy Code Support for Production Grids, Grid 2005 - 6th IEEE/ACM International Workshop on Grid Computing, Seattle, Washington, USA, 2005. (to appear)