Summaries of scientific publications

of Assoc. Prof. Dr. Emanuil Atanasov, IICT-BAS

candidate for "Mathematics" specialty "Computational Mathematics (High Performance Computing)", for the needs of the section "Grid Technologies and Applications", under the new name "High Performance Systems, Networks and Algorithms", for participation in a competition for academic position "Professor" announced in Official Gazette pc. 41/21.05.2019 г.

[1] Misev, A., Atanassov, E.,Performance analysis of GRID middleware using process mining, (2008) Lecture Notes in Computer Science, 5101 LNCS (PART 1), pp. 203-212, DOI: 10.1007/978-3-540-69384-0_26, ISSN: 0302-9743

Performance analysis of the GRID middleware used in a production setting can give valuable information to both GRID users and developers. This paper describes a new approach to this issue, using the process mining techniques. Analyzing logs of the middleware activities, performed on the SEE-GRID pilot production Grid infrastructure, objective qualitative and quantitative information on what actually happens has been obtained. Using the appropriate tools like ProM to apply the process mining algorithms, many interesting findings and conclusions have been drawn.

[2] Stoykov, S., Atanassov, E., Margenov, S.. Efficient sparse matrix-matrix multiplication for computing periodic responses by shooting method on Intel Xeon Phi. AIP Conference Proceedings, 1773, 110012, AIP Publishing, 2016, ISBN:978-073541431-0, ISSN:0094-243X, DOI:10.1063/1.4965016, 110012-110012

Many of the scientific applications involve sparse or dense matrix operations, such as solving linear systems, matrix-matrix products, eigensolvers, etc. In what concerns structural nonlinear dynamics, the computations of periodic responses and the determination of stability of the solution are of primary interest. Shooting method is widely used for obtaining periodic responses of nonlinear systems. The method involves simultaneously operations with sparse and dense matrices. One of the computationally expensive operations in the method is multiplication of sparse by dense matrices. In the current work, a new algorithm for sparse matrix by dense matrix products is presented. The algorithm takes into account the structure of the sparse matrix, which is obtained by space discretization of the nonlinear Mindlin's plate equation of motion by the finite element method. The algorithm is developed to use the vector engine of Intel Xeon Phi coprocessors. It is compared with the standard sparse matrix by dense matrix algorithm and the one developed by Intel for the library MKL and it is shown that by considering the properties of the sparse matrix better algorithms can be developed.

[3] Nedjalkov, M., Vasileska, D., Atanassov, E., Palankovski, V., Ultrafast Wigner transport in quantum wires, (2007) Journal of Computational Electronics, 6 (1-3), pp. 235-238, DOI: 10.1007/s10825-006-0101-y, ISSN: 1569-8025

In this work, two quantum-kinetic models, governing the transport of an initial highly non-equilibrium carrier distribution generated locally in a nanowire, are explored. It is known that dissipation processes due to phonons govern the carrier relaxation, which at early stages of the evolution is characterized by the lack of energy conservation in the collisions. The models are analyzed and approached numerically by a backward Monte Carlo method. The basic difference between them is in the way of treatment of the finite collision duration time. The latter introduces quantum effects of broadening and retardation, ultrafast spatial transfer and modification of the classical trajectories. These effects are demonstrated in the presented simulation results.

[4] Shterev, K., Atanassov, E., Stefanov, S., GPU calculations of unsteady viscous compressible and heat conductive gas flow at supersonic speed, (2014) Lecture Notes in Computer Science , 8353, pp. 549-556, DOI: 10.1007/978-3-662-43880-0_63, ISSN: 0302-9743

The recent trend of using Graphics Processing Units (GPUs) for high performance computations is driven by the high ratio of price performance for these units, complemented by their cost effectiveness. Such kinds of devices are increasingly being deployed not only as accelerators for supercomputer installations, but also in GPU-enabled nodes in Grid and Cloud installations. At first glance computational fluid dynamics (CFD) solvers match perfectly to GPU resources, because these solvers make intensive calculations and use relatively small memory. Nevertheless, there are scarce results about the practical use of this serious advantage of GPU over CPU, especially for calculations of viscous, compressible, heat conductive gas flows with double precision accuracy. In our work we present calculation of unsteady, viscous, compressible and heat conductive gas with double precision accuracy using GPU-enabled version of the algorithm SIMPLE-TS, written on standard OpenCL. As a test case the flow past a square in a microchannel at supersonic speed with Mach number M = 2.43 is modelled on an AMD Radeon HD 7950 GPU card, achieving 90 GFlops, which is 46 times faster than the CPU serial code run on Intel Xeon X5560.

[5] Ivanov, P., Atanassov, E., Jaime, C., Computational study on the conformations of CD38 and inclusion complexes of some lower-size large-ring cyclodextrins, (2014) Journal of molecular structure, 1056, pp. 238-245, DOI: 10.1016/j.molstruc.2013.10.048, ISSN: 0022-2860

The conformations of CD38 were examined by conformational search with molecular dynamics simulations using the Glycam04 force field. The results were compared with previous ones for CD26, the largest cyclodextrin for which crystal data are available. Principal component analysis (PCA) was applied for post-processing of the simulation

trajectories. Limited number of modes determine the overall deformations of the macroring of CD38. The longer perimeter of the macroring allowed the formation of a form not observed so far - a three-turn helix shaped as a short tube. In analogy with CD26, significant participation was monitored for conformations of CD38 with one-turn spirals at the opposite sides of the macroring linked together from the 'bottom' and from the 'top' with extended bridge spacers. Computationally were examined for the first time inclusion complexes of some lower-size LR-CDs, namely complexes of CDn (n = 13,14,26) with adamantane and of CD14 with 1-hydroxyadamantane. The macroring conformation of CD13 was not altered by the inclusion of the substrate molecule which acquired preferred positioning not in the middle of the cavity but rather close to the glucose residues at one of the sides. The same positioning of the small molecule in the cavity of the more flexible CD14 macroring enhanced the appearance of bent onto two conformation of this cyclodextrin. The most interesting behaviour presented the complex of CD26 with adamantane in which case the small molecule acts as a 'nucleation center' for the formation of a second helical turn about the substrate molecule.

[6] Balaž, A., Prnjat, O., Vudragović, D., Slavnić, V., Liabotis, I., Atanassov, E., Jakimovski, B., Savić, M., Development of Grid e-Infrastructure in South-Eastern Europe, (2011) Journal of Grid Computing, 9 (2), pp. 135-154, DOI: 10.1007/s10723-011-9185-0, ISSN: 1570-7873

Over the period of six years and three phases, the SEE-GRID programme established a strong regional human network in the area of distributed scientific computing and set up a powerful regional Grid infrastructure. It attracted a number of user communities and applications from diverse fields from countries throughout the South-Eastern Europe. From the infrastructure point view, the first project phase has established a pilot Grid infrastructure with more than 20 resource centers in 11 countries. During the subsequent two phases of the project, the infrastructure has grown to currently 55 resource centers with more than 6,600 CPUs and 750 TBs of disk storage, distributed in 16 participating countries. Inclusion of new resource centers to the existing infrastructure, as well as a support to new user communities, has demanded setup of regionally distributed core services, development of new monitoring and operational tools, and close collaboration of all partner institution in managing such a complex infrastructure. This paper presents an overview of the development and current status of SEE-GRID regional infrastructure and services and describes its transition to the NGI-based Grid model in EGI, with the strong SEE regional collaboration.

[7] Atanassov, E., Gurov, T., Karaivanova, A., Energy aware performance study for a class of computationally intensive Monte Carlo algorithms, (2015) Computers & mathematics with applications, 70(11), pp. 2719-2725, DOI: 10.1016/j.camwa.2015.07.014, ISSN: 0898-1221

The latest developments in the domain of HPC have lead to the deployment of complex extreme-scale systems, based on diverse computing devices (CPU, GPU, accelerators)

thus posing the question of scalability in the light not only of parallel efficiency, but also in terms of energy efficiency. In this paper a new metrics for energy aware performance estimation is proposed, based on authors' experiences and the analysis of the existing metrics. The performance of computationally intensive Monte Carlo applications deployed on heterogeneous HPC systems with focus on energy efficiency and equipment costs is studied and the the energy aware performance results of CPU and GPU variants of the tested algorithms are compared with respect to the introduced measures and metrics. The results of our study demonstrate the importance of taking into account not only scalability of the HPC applications but also energy efficiency and equipment cost. They also show how to optimize the selection of CPU computing or computing with GPGPUs. The results can be used by application developers/users and also by resource providers.

[8] Atanassov, E., Gurov, T., Karaivanova, A., Energy Performance Evaluation of Quasi-Monte Carlo Algorithms on Hybrid HPC, (2015) Large-scale scientific computing, 9374, pp.172-181, DOI: 10.1007/978-3-319-26520-9_18, ISSN: 0302-9743

The increasing demands of scientific applications and the increasing capacity of modern computing systems lead to the need of evaluating energy consumption and, consequently, to the development of energy efficient algorithms. In this paper we the energy-related performance of a class of quasi-Monte Carlo algorithms on hybrid HPC systems is studied. These algorithms are applied to solve quantum kinetic integral equations using Sobol and Halton sequences. The energy performance results are compared on a CPU-based computer platform and computer platforms with accelerators like GPU cards and Intel Xeon Phi coprocessors with respect to several metrics.

[9] Atanassov, E., Dimitrov, D., Gurov, T., Evaluation of Stochastic Algorithms for Financial Mathematics Problems from Point of View of Energy-efficiency, (2015) Application of mathematics in technical and natural sciences, 1684, DOI: 10.1063/1.4934300, ISSN: 0094-243X

The recent developments in the area of high-performance computing are driven not only by the desire for ever higher performance but also by the rising costs of electricity. The use of various types of accelerators like GPUs, Intel Xeon Phi has become mainstream and many algorithms and applications have been ported to make use of them where available. In Financial Mathematics the question of optimal use of computational resources should also take into account the limitations on space, because in many use cases the servers are deployed close to the exchanges. In this work various algorithms for option pricing are evaluated , in terms of their energy and space efficiency. Since it has been established that low-discrepancy sequences may be better than pseudorandom numbers for these types of algorithms, the Sobol and Halton sequences are also tested. The raw results are presented, the computed metrics and conclusions from the tests.

[10] Özturan, C., Kotroni, V., Atanassov, E., Development of virtual organizations, applications and services for earth science on grid e-Infrastructures, (2010) Earth Science Informatics, 3 (4), pp. 197-198, DOI: 10.1007/s12145-010-0074-z, ISSN: 1865-0473

Grids emerged as a platform to share computational resources in the early 2000s. The grid concept was generalized to include the sharing of not just computational resources but also other resources such as data, devices, and applications in the so called virtual organizations (VOs). Grids are essential in the move into the so-called e-Science era. E-Science refers to computational knowledge generation using various combinations of massive computational resources, data, and devices in highly distributed network environments. It is the next natural step beyond the web-based platforms of the last decade that offers distribution of scientific data and running of applications through web interfaces on client–server architectures. This paper considers the shortcomings of the traditional web-based platforms and describes the approaches used in the regional e-Infrastructure, with an accent on Earth Science applications.

[11] Atanassov, E., Karaivanova, A., Gurov, T., Ivanovska, S., Durchova, M., Dimitrov, D.S., Quasi-Monte Carlo integration on the grid for sensitivity studies, (2010) Earth Science Informatics, 3 (4), pp. 289-296, DOI: 10.1007/s12145-010-0069-9, ISSN: 1865-0473

This paper presents error estimations and performance analysis of quasi-Monte Carlo algorithms for solving multidimensional integrals (with up to 100 dimensions) on the grid using MPI. Taking into account the fact that the Grid is a potentially heterogeneous computing environment, where the user does not know the specifics of the target architecture, it is important that parallel algorithms should be able to adapt to this heterogeneity, providing automated load-balancing. Monte Carlo algorithms can be tailored to such environments, provided parallel pseudorandom number generators are available. In both cases the efficient implementation of the algorithms depends on the functionality of the corresponding packages for generating pseudorandom or quasirandom numbers. In the paper it is proposed an efficient parallel implementation of the Sobol sequence for a grid environment and it is used for numerical experiments on a heterogeneous grid. To achieve high parallel efficiency a newly developed special grid service called Job Track Service has been used, which provides efficient management of available computing resources through reservations.

[12] Georgiev, D., Atanassov, E., Extensible framework for execution of distributed genetic algorithms on grid clusters, (2014) 37th international convention on information and communication technology, electronics and microelectronics (MIPRO), pp. 301-306, DOI: 10.1109/MIPRO.2014.6859581, ISBN: 978-953-233-077-9

Genetic algorithms are effective metaheuristic optimization methods, based on the principles of natural selection and genetics. Although they are able to find adequate

solutions in acceptable time for small problems, their execution time increases rapidly for problems with large search space and complex fitness landscape. Computational grids provide an appropriate platform for the execution of parallel genetic algorithms. However, when using multiple grid clusters for a single computational task, one has to comply with a number of technical and administrative limitations. This paper presents a new implementation (framework) of Genetic Algorithms, written in C++11, which provides a mechanism for parallel execution and an extensible programming interface. This framework utilizes MPI, ZeroMQ and the newest generic programming facilities of C++, enabling the execution of a distributed GA on multiple grid clusters. Various tests were performed on clusters from the European Grid Infrastructure, which prove the efficiency of the new approach.

[13] Atanassov, E., Gurov, T., Karaivanova, A., Message Oriented Framework with Low Overhead for Efficient High-Performance Monte Carlo Simulations, (2013) 36th international convention on information and communication technology, electronics and microelectronics (MIPRO), pp. 169-171, ISSN:1847-3946

In the recent years Bulgaria acquired a substantial amount of HPC resources of various types. The biggest procurement has been the BlueGene/P supercomputer at SAITC with 8192 CPU cores, while the Bulgarian Academy of Sciences has now two HPC clusters with Intel CPUs and InfiniBand interconnection, which total more than 1000 logical cores. In addition some servers equipped with powerful GPUs are available for applications that can take advantage of them. The coordinated use of such resources by one application faces significant challenges due to the heterogeneity of the resources and the networking and security constraints.

In order to facilitate the coordinated use of all these resources where each resource is used for the parts of the application where it is most efficient, we have developed a framework that allows the researcher to interconnect resources of the above types with minimal overhead. In this paper we describe the architecture of the system and demonstrate its effectiveness for a semiconductor modeling application, showing numerical and timing results.

[14] Misev, A., Atanassov, E., User level Grid Quality of service, (2010) Lecture Notes in Computer Science, 5910 LNCS, pp. 507-514, DOI: 10.1007/978-3-642-12535-5_60, ISSN: 0302-9743

Improving the Quality of Service of the Grid infrastructure is one of the most important ongoing issues in the Grid community. It has many implications, from broader users' acceptance of the technology to the shift of the Grid usage from the scientific world toward the businesses. While best effort can be acceptable in the scientific community, the business applications must have clear and well defined service levels based on deterministic QoS metrics. In this paper a different type of Grid Quality of Service is

introduced, one under users' control. Using it, the users can have better control over the level of services they are using from the Grid infrastructure.

[15] Atanassov, E., Karaivanova, A., Ivanovska, S., Tuning the generation of Sobol sequence with Owen scrambling, (2010) Lecture Notes in Computer Science, 5910 LNCS, pp. 459-466, DOI: 10.1007/978-3-642-12535-5_54, ISSN: 0302-9743

The Sobol sequence is the most widely used low-discrepancy sequence for numerical solution of multiple integrals and other quasi-Monte Carlo computations. Owen first proposed scrambling of this sequence through permutations in a manner that maintained its low discrepancy. Scrambling is necessary not only for error analysis but also for parallel implementations. Good scrambling is especially important for GRID applications. However, scrambling is often difficult to implement and time consuming. In this paper an algorithm for fast generation of Sobol sequence with Owen scrambling is proposed, tuned to the specific computer hardware. Numerical and timing results, demonstrating the advantages of our approach are presented and discussed.

[16] Misev, A., Atanassov, E., ULMon - Grid monitoring from user point of view, (2009) Proceedings of the International Conference on Information Technology Interfaces, ITI, pp. 621-626, DOI: 10.1109/ITI.2009.5196158, ISSN: 1330-1012

Leveraging a complex infrastructure such as the computing Grids requires an appropriate level of monitoring. Since it is a technology in development, the current computing Grid middleware offers very little monitoring capabilities from the users' point Of view. Making the Grid closer to the users requires that the users have deeper insight of what is happening with the computational jobs they submit. In this paper we present a novel tool for monitoring the gLite based Grid infrastructure from the users' perspective

[17] Atanassov, E., Dimov, I.T., What Monte Carlo models can do and cannot do efficiently?, (2008) Applied Mathematical Modelling, 32 (8), pp. 1477-1500, DOI: 10.1016/j.apm.2007.04.010, ISSN: 0307-904X

The question "what Monte Carlo models can do and cannot do efficiently" is discussed for some functional spaces that define the regularity of the input data. Data classes important for practical computations are considered: classes of functions with bounded derivatives and Holder type conditions, as well as Korobov-like spaces.

Theoretical performance analysis of some algorithms with unimprovable rate of convergence is given. Estimates of computational complexity of two classes of algorithms - deterministic and randomized for both problems - numerical multidimensional integration and calculation of linear functionals of the solution of a class of integral equations are presented. (c) 2007 Elsevier Inc. All rights reserved.

[18] Atanassov, E., Georgiev, D., Manev, N., Number Theory Algorithms on GPU Clusters, (2014) High-performance computing infrastructure for south east europe's research communities:

results of the hp-see user forum 2012, 2, pp. 131-138, DOI: 10.1007/978-3-319-01520-0_16, ISSN: 2196-7334

Many algorithms from Number Theory and their implementation in software are of high practical importance, since they are the building primitives of many protocols for data encryption and authentication of Internet connections. Number theory algorithms are also the basic part of cryptanalytic procedures. Many of these algorithms can be parallelized in a natural way. This paper describes a software package that implements various Number Theory algorithms on GPU clusters and in particular the implementations of integer factorization using NVIDIA CUDA on clusters equipped with NVIDIA GPUs. Results of experiments regarding the performance of the implementation are also presented.

[19] Georgiev, D., Atanassov, E., Alexandrov, V., A Framework for Parallel Genetic Algorithms for Distributed Memory Architectures, (2014) 5th workshop on latest advances in scalable algorithms for large-scale systems (SCALA), pp. 47-53, DOI: 10.1109/ScalA.2014.13, ISBN: 978-1-4799-7562-4

Genetic algorithms are metaheuristic search methods, based on the principles of biological evolution and genetics. Through a heuristic search they are able to find good solutions in acceptable time. However, with the increase of the complexity of the fitness landscape and the size of the search space their runtime increases rapidly. Using parallel implementations of genetic algorithms in order to harness the power of modern computational platforms, is a powerful approach to mitigating this issue. In this paper several parallel implementations ranging from MPI to hybrid MPI/OpenMP and MPI/OmpSs are made. These implementations are optimized for execution on tightly coupled distributed memory systems. Issues that arise when running a distributed genetic algorithm are addressed and an adaptive migration scheme is presented. A comparison of their efficiency is also made.

[20] Atanassov, E., Durchova, M., Generation of the Scrambled Halton Sequence Using Accelerators, (2013) 36th international convention on information and communication technology, electronics and microelectronics (MIPRO), pp. 177-181, ISSN: 1847-3946

The Halton sequence is one of the most popular low-discrepancy sequences. In order to satisfy some practical requirements, the original sequence is usually modified in some way. The scrambling algorithm, proposed by Owen, has several theoretical advantages, but on the other hand is difficult to implement in practice due to the trade-off between high memory and high computational requirements. In our work we concentrate on the case when the number of coordinates is relatively high. The use of computational accelerators and especially GPUs is increasingly relevant for such practical applications, since more and more of the resources, available through the Grid and Cloud infrastructures, provide access to such accelerators, provided that the software can make use of them. In this paper we discuss our algorithm for generation of the Halton sequence

with Owen-type scrambling implemented using NVIDIA CUDA. We also show numerical results, achieved on our GPU-enabled nodes, which are equipped with NVIDIA M2090 cards.

[21] Atanassov, E., Ivanovska, S., Computation and Analysis of Sobol Coefficients for Air Pollution Concentrations over the Territory of Bulgaria, (2013) 36th international convention on information and communication technology, electronics and microelectronics (MIPRO), pp 234-237, ISSN: 1847-3946

One of the main tools for modeling air pollutions over the territory of Bulgaria is the US EPA Models-3 system. The main components of the system are MM5/WRF meteorological preprocessor, SMOKE emission preprocessor and CMAQ chemical transport model. TNO emission inventory is used as emission input. The Models-3 "Integrated Process Rate Analysis" option is applied to discriminate the role of different dynamic and chemical processes for the pollution for all SNAP categories. In this work we evaluate the influence of the different input parameters as concentration in different SNAP categories under constant meteorological conditions to the output concentrations over Bulgaria following methodology of Sobol-Saltelli. In order to obtain reliable estimates of the Sobol coefficients we perform a large of MPI jobs using the clusters in South-Eastern region. Using these coefficients we assess the relative importance of the various input parameters, and their interactions.