

Strengthening Compute and Data Intensive Capacities of Armenia

Hrachya Astsatryan, Vladimir Sahakyan,
Yuri Shoukourian

Institute for Informatics and Automation Problems of the
National Academy of Sciences of the Republic of Armenia.
Yerevan, Armenia
{hrach,svlad,shouk}@sci.am

Jack Dongarra
University of Tennessee
Knoxville, USA
dongarra@eecs.utk.edu

Pierre-Henri Cros, Michel Dayde
Institut de Recherche en Informatique de Toulouse
Toulouse, France
{Pierre-Henri.Cros,michel.dayde}@irit.fr

Per Oster
IT Center for Science,
Espoo, Finland
Per.Oster@csc.fi

Abstract — Traditionally, Armenia has had a leading position within the computer science and Information Technology sectors in the South Caucasus region and beyond. Information Technology (IT) is also one of the fastest growing industries of the Armenian economy [1]. In 2000, the Government of Armenia recognized the IT sector as the primary constituent of the country's economic progress. Armenia is, more than ever, in need of cutting-edge and relevant e-infrastructures and e-services to tackle today's societal and scientific challenges. The Institute for Informatics and Automation Problems (IIAP) of the National Academy of Sciences of the Republic of Armenia (NAS RA) [2] is the only state supported structure for software, hardware, and brainware technologies in Armenia. The institute is responsible for Armenia's National research and education network (Academic Scientific Research Computer Network of Armenia, ASNET-AM) [3] and the National Grid Initiative (ArmNGI) [4], and provides computational and networking facilities and advanced services to users. The main objective of this article is to highlight key activities that will spur Armenia to strengthen its scientific computing capacity thanks to the analysis made of the current trends of e-Infrastructures in Europe and the USA.

Keywords — HPC; Big Data; ASNET-AM; ArmNGI; e-infrastructures; Armenia.

I. INTRODUCTION

Modelling and numerical simulation - considered to be the third pillar of the science after theory and experimentation - is at the heart of multiple domains, which are not only scientific, but also societal (e.g., energy, health, environment), economic, financial (e.g., industrial competitiveness), and life ethics (e.g., biology). They also appear increasingly as decision-making tools for critical cases like global warming, natural disasters, etc. Since modelling and simulation are essential for multiple scientific advances, the control of all the aspects of high performance computing (HPC) - as well as the capacity to

exploit the masses of data to tackle the solution of these complex models - is inescapable.

In most areas, the HPC challenges are related to the modelling of complex systems in order to better understand larger and more complex problems. The phenomena that are modelled are complex, coupled, multi-physics, and multi-scales in space (from molecular up to living system size, for example) or in time. The issues related to visualization, processing large amounts of data, and uncertainty quantification are all central. Coupling codes from different research areas allows for new perspectives and enables us to tackle complex multiphysics phenomena by taking advantage of the knowledge of each discipline involved.

Over the past few years resilience has become a major issue for HPC systems, especially for large petascale systems and future exascale systems. The roadmap to exascale and the evolution of supercomputers [5] requires major efforts in several key areas:

- Efficient exploitation of next generation architectures at the hardware, system, and runtime support levels because of their high degree of parallelism and their heterogeneity. Co-design is considered a key activity;
- Impact of recent technologies such as GPUs and Cloud computing in the HPC ecosystem;
- Scalability;
- Sustainable development including evolution of codes and numerical libraries within scientific communities;
- Programming paradigms and languages.

From 2015 to 2020, the HPC market should engender around 200 billion dollars of income, according to Terascale, resulting from both the growth of the HPC market as it moves towards Exascale, and from the deployment of academic and departmental machines - more and more often in connection with the multiplication of datacenters, with applications related to engineering, "Big Data" (e.g., genomic or financial

analyses), and with the deployment of resources in public and private cloud computing infrastructures.

"Big Data" is now considered to be the fourth pillar of science [6], and - just as HPC is - understanding and overcoming "Big Data" challenges is crucial for science and industrial advancements and competitiveness. The volume and the complexity of data does not stop growing. It is no longer possible to dissociate HPC from "Big Data" i.e., the exploitation of the masses of data stemming from digital simulations (climate, combustion, fusion, astrophysics), large instruments (LHC, ITER, LSST, LOFAR, genomic platforms), ground or space systems of observation (seismology and geodesy, RESIF, Euclid, WFIRST, GAIA, imaging and interferometry) or simply multiple devices of data acquisition (broadband sequencer, sensors' networks, social networks, etc.). In addition, HPC and "Big Data" cannot be separated from all the elements, which constitute the ecosystem of computing and data intensive infrastructures (computing infrastructures, storage, networks, computer rooms, etc.).

In multiple scientific and socioeconomic domains, the volume and the variety of the existing data, as well as the time constraints of calculation, revealed new challenges. To address these new challenges, new tools, scientific methods, and technologies are necessary in a wide range of areas, including: in fundamental sciences (e.g., physics of high energy, fusion, earth and universe sciences, bio-computing, neurosciences, etc.), digital economy (e.g., business intelligence, Web, e-commerce, social networks, e-government, health, telecommunications and media), ground and air transport, financial markets, environment (e.g., climate, natural risks, energy resources, smart cities, connected houses), security, consumer goods industry (e.g., smart industry, customized products, product design, and production chain).

In several areas, the volume and the complexity of data challenge our capacities to explore and to analyze them. Without adequate processing capacity for analysis and adequate storage for the data, the systems do not produce "science" or decision-making support. Some of the main challenges for "Big Data" lie in developing research in order to master all the technologies of the data value chain; offering improved and readily available training; reinforcing the industrial ecosystem (SME's and large groups); and developing bridges between research and industry. It is also crucial to reinforce links between researchers and industry leaders and to cultivate teams for multidisciplinary R&D; develop the exploitation of open data; encourage the implementation of training programs - for instructors and students - in universities and engineering schools.

Since it may be necessary to set up research initiatives for developing / deploying infrastructures, algorithms, mobile devices and applications serving billions of users, "Big Data" offers numerous economic opportunities. It is estimated that these applications could generate (according to the mission Afdel on big data) more than 40 Billion Euros of income in 2016 and a potential value creation of 200 Billion Euros in European public administration.

In recent years significant investments have been made by the European Commission and European member states to create pan-European e-Infrastructures in the following domains:

- HPC: The mission of PRACE (Partnership for Advanced Computing in Europe) [7] is to enable high impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society. The computer systems and their operations accessible through PRACE are provided by 4 PRACE members (BSC representing Spain, CINECA representing Italy, GCS representing Germany, and GENCI representing France) who committed a total funding of EUR400 million for the initial PRACE systems and operations;
- Grid: The mission of the EGI-InSPIRE project (Integrated Sustainable Pan-European Infrastructure for Researchers in Europe) [8] is to combine the new Distributed Computing Infrastructures (DCIs) - such as clouds, supercomputing networks, and desktop grids - for the benefit of user communities within the European Research Area, including 37 National Grid Initiatives (NGIs), 2 European International Research Organizations and eight partners from the Asia Pacific region;
- Cloud: Helix Nebula [9] is a new pioneering partnership between big science and big business in Europe that is charting the course towards the sustainable provision of cloud computing - the Science Cloud. The partnership brings together leading IT providers and three of Europe's leading research centers - CERN, EMBL, and ESA - in order to provide computing capacity and services that elastically meet big science's growing demand for computing power;
- Big Data: The EUDAT (Towards a pan-European Collaborative Data Infrastructure) [10] vision is to support a Collaborative Data Infrastructure, which will allow researchers to share data within and between communities and to provide a solution that will be affordable.

As the premier distributed cyberinfrastructure for open scientific research in the United States, XSEDE (Extreme Science and Engineering Discovery Environment) [11] supports supercomputers and high-end visualization and data analysis resources across the country, as well as data collections and computational tools that are critical to the success of those researchers. In the past three years, the United States, the European Union, and Japan have each moved aggressively to develop their own plans for achieving exascale computing in the next decade.

As a state scientific organization NAS RA unifies the scientific and research institutes, subsidiary services and governing body - the Presidium of the Academy includes more than 34 scientific institutions and other organizations. As the state supported structure of NAS RA for software, hardware and brainware technologies in Armenia, we introduce and discuss IIAP and its key activities that will spur Armenia to strengthen its scientific computing capacity.

The current State of e-infrastructures in Armenia is presented in Section 2 and the further activities to strengthen the scientific computing capacity in Armenia can be found in Section 3. Finally, the conclusion and directives for future research are drawn in the conclusion section.

II. CURRENT STATE OF E-INFRASTRUCTURES IN ARMENIA

The Armenian e-infrastructure is a complex national IT infrastructure consisting of both communication and distributed computing infrastructures. IIAP plays a key role at national level in the fields of networking and distributed and large-scale research infrastructures (Grid, HPC, Cloud) as a developer and operator of ASNET-AM and ArmNGI.

Since 1994, ASNET-AM has provided a high-quality infrastructure and services to the academic, research and educational community of Armenia by connecting more than 60 scientific research, academic, and cultural organizations. The ASNET-AM backbone consists of network communication nodes in 6 cities in Armenia, which are interconnected by fiber optics and wireless links (see fig. 1). Most of the sites located in Yerevan are connected via ASNET-AM's fibre at bitrates of 1 Gb/s. A few key sites, such as the Presidium of the National Academy of Sciences and the IIAP, are interconnected at 10 Gb/s. ASNET-AM offering to the Armenian research and educational community access to the pan-European GEANT network. ASNET-AM also hosts a wide variety of different services for connected organisations, such as the Eduroam that provides Internet access in the campuses to visiting students and staff members, and the mail filter that enables institutions to have all their e-mail messages checked or filtered for viruses or spam.

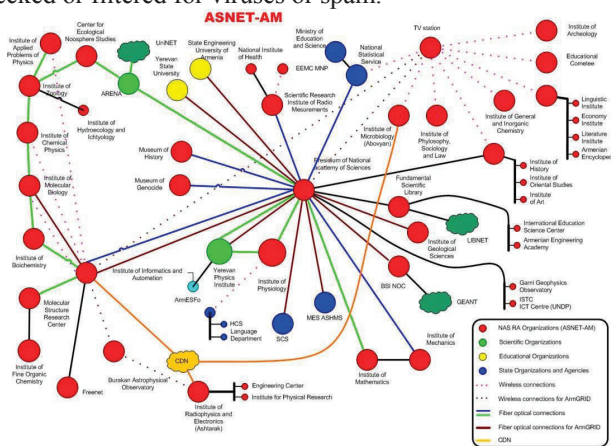


Fig. 1. Topology of ASNET-AM.

Being a long-time Armenian academic network operator, ASNET-AM participates in a series of international activities and is a part of a worldwide community of research and education networks. ASNET-AM is a member of several international organizations, such as the Trans-European Research and Education Networking Association, Reseaux IP Europeens, and the Central and Eastern European Networking Association. ASNET-AM is also an active participant in European networking projects within the Seventh Framework

Programmes, including, Multi-Gigabit European Research and Education Network and Associated Services and Black Sea Interconnection. ASNET-AM and the national research and education networks of all six Eastern Partnership (EaP) countries will establish and operate a high-capacity broadband internet network for research and education through the EaPConnect (Eastern Partnership Connect) Project [12] funded by the European Commission's Directorate-General for Neighbourhood and Enlargements Negotiations. With connections to the pan-European GÉANT network EaPConnect will enable scientists and academics in the region to engage in global collaborative projects and bring the EaP countries closer to EU.

In 2004, the first high Performance computing cluster (128 processors, Myrinet interconnection), called Armcluster [13], was developed in IIAP with the 523.4GFlops performance achieved by High Performance Linpack test. Being the most powerful computational resource in the field of science and education in Armenia and in the South Caucasus region, the Armcluster constitutes the core of the Armenian grid infrastructure.

Now IIAP coordinates the ArmNGI, which is a national Armenian effort to establish a nationwide grid environment for computational science and research. The goal of the ArmNGI is to pursue a variety of scientific users in utilizing the Grid for their applications. All these applications rely on a wide range of diverse computer science technologies composed from standard grid middleware and sophisticated high-level extensions. Now the computational resources (about 500 cores) of Armenian Grid infrastructure distributed among our leading research (National Academy of Sciences, Yerevan Physics Institute) and academic (Yerevan State University, State Engineering University of Armenia) organizations are located in the cities of Yerevan and Ashtarak (see fig. 2). ArmNGI is a member of the European Grid initiative, which establishes a sustainable European Grid Infrastructure.

The move to the cloud computing in Armenian academic sector has already started. Experimental cloud infrastructures based on OpenNebula and OpenStack platforms have been deployed in Armenia and beyond in the region. For example the integrated federated Cloud infrastructure in the Black Sea region mainly in Romania, Armenia, Georgia and Moldova [14].

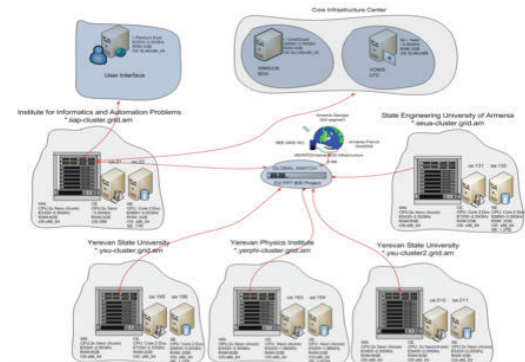


Fig. 2. Topology of ArmNGI infrastructure

Several services and tools have been developed to provide value-added abilities to the target research communities, such as:

- A single sign on an integrated multi-service environment for linear algebra calculus [15] that enables the solution of large-scale linear systems of equations using direct solvers, makes easier the use of parallel block iterative algorithm and provides an interface for parallel decision making algorithms;
- A portal for quantum optics technologies based on the numerical simulations and modeling of complex quantum systems [16] that makes possible to solve various problems for quantum open systems. Such systems are mainly constructed from photons, including laser fields, atom-photon interactions and nonlinear optical interactions in crystals and describe photonic devices in the presence of imperfection effects: quantum noise, decoherence, etc;
- An interoperable scientific gateway for parallel geoprocessing of satellite image indices [17], and NDVI time series analyzes [18] using spatial data infrastructures, etc. The gateway hides low-level access mechanisms to computational resources with high-level graphical interfaces, making even non-GIS expert users capable of defining and executing distributed applications using the HPC resources.

Recently the eduGAIN service has been developed under the umbrella of the Armenian Federation Identity for Research and Education (AFIRE), which is an identity federation that gathers teaching and research institutions in Armenia. Through AFIRE, a user keeps all its information at the home institution and can access services offered by the institutions participating in the federation. The eduGAIN service provides a secure sign-on system facilitates sharing the applications and services, such as providing cloud storage services based on pydio. Pydio is a mature open source software solution for file sharing and synchronization, which provides enterprise-grade features to gain back control and privacy: user directory connectors, legacy filesystems drivers, comprehensive admin interface, and much more.

The primary users of the distributed computing and storage resources consisted of the following domains:

- Life sciences [19-20], such as molecular dynamics simulation on the state of long chain alkyl sulfonate micelle which occurs in the presence of salt, or molecular dynamics simulations of amphi-phile/water system consisting of 128 sodium pentadecyl sulfonate and 2251 water molecules;
- Earth and climate sciences [21], such as the operational usage of a mesoscale Weather Research and Forecasting model for the territory of Armenia with fine spatial resolution, which gives an opportunity to improve weather forecasting and to promote research that advances understanding of weather processes, and increase the utility of forecast with an emphasis on high-impact weather;
- Astronomy and cosmology [22], such as the Armenian Virtual Observatory as a collection of interoperating data

archives (20 thousand photographic plates and films) and software tools which utilize the internet to form a scientific research environment in which astronomical research programs can be conducted;

- Particle physics (ATLAS and ALICE experiments at CERN);
- Computational engineering and other scientific disciplines (self-organized systems based on the results in the theory of cellular automata, classical spin glass theory, linear cryptanalyses of cryptosystems, etc.) [23].

Communication with European counterparts is a crucial element of integration into European Research Area and might create additional opportunities to increase visibility of the region.

Taking into account the status of the current e-infrastructure, and the relationships with the European and international societies and infrastructures, the priorities of Armenia in science and technology, it is advisable to strengthening the compute and data intensive capacities of Armenia to tackle the fundamental problems in science and engineering in Armenia and beyond, such as:

- Data intensive: astrophysics and cosmology (archives and sharing data received from the telescopes located at the Byurakan Astrophysical Observatory), digital libraries (archiving and sharing research data received from the Fundamental Scientific Library of the National Academy of Sciences of Armenia), Armenology (development and promotion a multilingual communication platform), high energy physics (support the scientific experiments at CERN), disaster management (responding to a risk-averse society), urban management (transport control and monitoring), telemedicine, digital tomography, etc.;
- Compute intensive: climate studies, weather forecasting, atmospheric chemistry and air pollution, modelling of Lake Sevak water resources: quantity and quality, computational simulations of advanced nanomaterials, molecular dynamics simulations, molecular simulations of biomolecules, Ab initio calculations, microbial genomes and community gene expression, satellite and biomedical image processing, etc.;

Despite the success to participate in several pan-European and regional EU projects (EGInSPIRE, GN4, HP-SEE, BSI, SEE-GRID-SCI), Armenia is still underprivileged compared to Western Europe in terms of available e-Infrastructure capacity and know-how. Thus to satisfy the growing needs of compute and data intensive applications, it is crucial to significantly extend the available resources in the following directions:

- Optical fibre networks: to deploy modern data communication networks based for ASNET-AM enabling the wide-scale, high-volume, low-latency data movements that are foreseen;
- 200 TB Data Storage: to satisfy the growing technical possibilities of experimental equipment and tools (telescopes, sensors, stations, etc.). The storage capacities will be built in a distributed and hierarchical way;
- 200TFlops HPC system: to support the large computational needs of national and regional research communities.

III. FURTHER ACTIVITIES TO STRENGTHEN THE SCIENTIFIC COMPUTING CAPACITY IN ARMENIA

One of the great opportunities in Armenia is that it is still possible to coordinate the development of the major IT infrastructures, which has not always been the case in other countries, leading to occasional inefficient investments. Another opportunity lies in the ability to design a real urbanisation of IT infrastructures, which is necessary in our application. The first step to strengthening the computational capacity of Armenia was the establishment of the Supercomputing Center, which took into account both the importance of expanding the usage of supercomputing in an effort to boost the economic competitiveness of Armenia, and also that the Government of Armenia declared ICT as one of the priority sectors of the Armenian economy - recognizing the progress made in Armenia in e-infrastructure development, especially in distributed computing (Grid, Cloud) and networking infrastructures.

A memorandum of understanding (MoU) regarding cooperation in the field of supercomputing has been signed between the Republic of Armenia's National Academy of Sciences, the Ministry of Education and Science, the Ministry of Economy, and the Ministry of Emergency Situations to set up a collaboration for the establishment of an Armenian National Supercomputing Center at the National Academy of Sciences. This supercomputing center would be, at least initially, administered and located at IIAP, in order to develop and apply high-performance information technologies for science and society in Armenia and beyond, and to deliver advanced scientific and industry research in the field of high performance computing. A MoU for exploring potential future collaboration in the fields of high-performance computing and computational science between the Université Toulouse III - Paul Sabatier (France), the Institute for Computer Science and Control of the Hungarian Academy of Sciences (Hungary), and IIAP has been signed and expresses the institutes' respective intentions for future collaborations. This exchange has already begun in several scientific domains, including:

- Medical Image Processing: The topic of this collaboration is the ill-posed inverse problems in medical imaging, which requires a complex optimization routine with appropriate regularization terms. The inverse problems are involved in several of our research axis, such as: deconvolution, super-resolution, compressive sampling, beamforming, and associations of them;
- Coding and cryptography: Cryptanalysis of lattice-based fully homomorphic encryption schemes, security proofs for alternative noise distributions in public key cryptosystems, and fast polynomial operations (FFT and noise generation) in the lattice-cryptography setting;
- Green Computing: Linux kernel governor for reducing HPC energy consumption, mathematical and analytical models of energy consumption, energy consumption reduction in hybrid (CPU/GPU) architectures, memory ballooning, etc.

IV. CONCLUSION

Recent years have brought a series of coordinated and complementary e-infrastructure initiatives that have been crucial for deploying state-of-the-art e-infrastructures in Armenia. In this paper, we have described the actual status of these e-Infrastructures in Armenia and beyond. Our goal is to focus on e-infrastructures and services that cut across a range of user disciplines and aim at empowering researchers with open online access to scientific resources, a range of facilities, networking, and collaboration tools. For the further the process of joining, sharing and exploiting the national e-Infrastructure resources and to satisfy the needs of user communities it is crucial to establish partnerships with European infrastructure projects (ESFRI: PRACE, ELIXIR, CLARIN, etc) and to become a member of other relevant international infrastructures and research organisations.

ACKNOWLEDGMENT

This work was supported by the INARMERA-ICT (Integrating Armenia into ERA: Information and Communication Technologies), GN4 (Multi-Gigabit European Research and Education Network and Associated Services), EaPConnect and VI-SEEM (VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean) Projects funded by the European Commission.

REFERENCES

- [1] "Information and Telecommunication Technologies Sector in Armenia," Enterprise Incubator Foundation, Yerevan, Armenia, 2013.
- [2] Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia, URL: <http://iiap.sci.am>.
- [3] Academic Scientific Research Computer Network of Armenia, URL: <http://www.asnet.am>
- [4] Yu. Shoukourian, V. Sahakyan, H. Astsatryan, "E-Infrastructures in Armenia: Virtual Research Environments," in IEEE Proc. CSIT, Yerevan, Armenia, 2013, pp. 1-7.
- [5] J. Dongarra, P. Beckman, et. al, "The International Exascale Software Roadmap," International Journal of High Performance Computing, vol. 25, no. 1, pp. 3-60, 2011.
- [6] T. Hey, S. Tansley, K. Tolle, "The Fourth Paradigm: Data-Intensive Scientific Discovery," vol. 1. Redmond, WA: Microsoft Research, 2009.
- [7] PRACE Research Infrastructure, URL: <http://www.prace-ri.eu/>.
- [8] The European Grid Infrastructure, URL: <http://www.egi.eu/>
- [9] Helix Nebula, URL: <http://www.helix-nebula.eu>
- [10] EUDAT, URL: <http://www.eudat.eu>
- [11] XSEDE, URL: <https://www.xsede.org>
- [12] EaPConnect, <http://www.eapconnect.eu>
- [13] H.V. Astsatryan, Yu. Shoukourian, V. Sahakyan, "The ArmCluster Project: Brief Introduction," In Proc. PDPTA, Las Vegas, Nevada, USA, 2004, pp. 1291-1295.
- [14] H. Astsatryan, A. Hayrapetyan, W. Narsisian, V. Sahakyan, Yu. Shoukourian, G. Neagu, A. Stanciu, "Environmental Science Federated Cloud Platform in the BSEC Region," International

- Journal of Scientific & Engineering Research, vol. 1, no.1, pp. 1130-1133, January 2014.
- [15] Hrachya Astsatryan, Vladimir Sahakyan, Yuri Shoukouryan, Michel Daydé, Aurelie Hurault, Ronan Guivarch, Arutyun Terzyan, Levon Hovhannisyan, "Large-Scale Linear Systems of Equations and Algorithms based on Integrated P-Grade Portlal," vol. 11, no. 2, pp. 239-248, 2013.
- [16] H. Astsatryan, T. Gevorgyan, A. Shahinyan, "Web Portal for Photonic Technologies Using Grid Infrastructures," Journal of Software Engineering and Applications, vol. 5, no. 11, pp. 864-869, 2012.
- [17] Hrachya Astsatryan, Wahi Narsisian, Andranik Hayrapetyan, Albert Saribekyan, Shushanik Asmaryan, Vahagn Muradyan, Yaniss Guigoz, Gregory Giuliani, Nicolas Ray, "An Interoperable Web Portal for Parallel Geoprocessing of Satellite Image Indices," Springer Earth Science Informatics, 2014, DOI: 10.1007/s12145-014-0165-3.
- [18] Hrachya Astsatryan, Andranik Hayrapetyan, Wahi Narsisian, Shushanik Asmaryan, Vahagn Muradyan, Yaniss Guigoz, Gregory Giuliani, Nicolas Ray, "An Interoperable Cloud-based Scientific Gateway for NDVI Time Series Analysis," Elsevier Computer Standards & Interfaces, 2015, doi: 10.1016/j.csi.2015.02.001.
- [19] A. Poghosyan, L. Arsenyan, A. Shahinyan, "Molecular Dynamics Study of Intermediate Phase of Long Chain Alkyl Sulfonate/Water Systems," Langmuir, vol. 29, no. 1, pp. 29-37, 2013.
- [20] L. A. Tavadyan, G. A. Martoyan, "Analysis of Kinetic Models of Chemical Reaction Systems," Value Approach, ISBN: 978-1-63321-081-3, Nova Science Publisher, 2014
- [21] T. Khotsanyan, et al., "Implementing and evaluating the weather research and forecast model for the territory of Armenia," in Proc. CSIT, Yerevan, Armenia, 2009, pp. 490-494.
- [22] H.V. Abrahamyan, A.M. Mickaelian, A.V. Knyazyan, "The IRAS PSC/FSC Combined Catalogue, Astronomy and Computing," Astronomy and Computing, 2014.
- [23] V. S. Poghosyan, S. Y. Grigorev, V. B. Priezhev, P Ruelle, "Logarithmic two-point correlators in the Abelian sandpile model," Journal of Statistical Mechanics: Theory and Experiment, vol. 2010, July 2010.