

MULTIFUNCTIONAL INFORMATION AND COMPUTING COMPLEX OF JINR: STATUS AND PERSPECTIVES

**A.G. Dolbilov, I.A. Kashunin, V.V. Korenkov, N.A. Kutovskiy,
V.V. Mitsyn, D.V. Podgainy, O.I. Streltsova, T.A. Strizh, V.V. Trofimov,
A.S. Vorontsov**

Joint Institute for Nuclear Research, 6 Joliot-Curie, 141980, Dubna

E-mail: strizh@jinr.ru

The implementation of the MICC (Multifunctional Information and Computing Complex) project in 2017-2019 laid foundation for its further development and evolution taking into account new requirements to the computing infrastructure for JINR scientific research. The rapid development of information technologies and new user requirements stimulate the development of all MICC components and platforms. Multi-functionality, high reliability and availability in a 24x7 mode, scalability and high performance, a reliable data storage system, information security and a customized software environment for different user groups are the main requirements, which the MICC should meet as a modern scientific computing complex. The JINR MICC consisting of four key components - the grid infrastructure, the central computing complex, the computing cloud and the HybriLIT high-performance platform, which includes the “Govorun” supercomputer, ensures the implementation of a whole range of competitive studies conducted at the world level at JINR in experiments: MPD, BM@N, Alice, ATLAS, CMS, NOvA, BESIII, STAR, COMPASS and others. The MICC includes the Tier1 grid centre, which is the only one in the JINR Member States and one of the 7 world data storage and processing centres of the CMS experiment (CERN). The JINR Tier1 and Tier2 grid sites are elements of the global grid infrastructure used in the WLCG project for processing data from the LHC experiments and other grid applications.

Keywords: Grid, Cloud, HPC, LHC, Tier1, Tier2, EOS, monitoring

Andrey Dolbilov, Ivan Kashunin, Vladimir Korenkov, Nikolay Kutovskiy, Valery Mitsyn,
Dmitry Podgainy, Oxana Stretsova, Tatiana Strizh, Vladimir Trofimov, Alexey Vorontsov

Copyright © 2019 for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Introduction

The research program of the Joint Institute for Nuclear Research (JINR) for the next decades is aimed at conducting ambiguous and large-scale experiments on the Institute basic facilities and in frames of the worldwide cooperation. This program is connected with the implementation of the NICA (Nuclotron-based Ion Collider fAcility) megaproject [1], the construction of new experimental facilities, the JINR neutrino program, the modernization of the Large Hadron Collider (LHC) [2] experimental facilities (CMS, ATLAS, Alice), programs on condensed matter physics and nuclear physics. The implementation of the mentioned above projects requires adequate and commensurable investments in systems providing processing and storage of increasing data volumes. It should be mentioned that there are a big variety of computing architectures, platforms, operational systems, network protocols, and software products used for the tasks fulfilment in computing for high energy and nuclear physics. The experience of recent years shows that the progress in obtaining research results directly depends on the performance and efficiency of computing resources. In order to meet the needs of users in 2017 JINR launched a project of the Multifunctional Information and Computing Complex (MICC) [3].

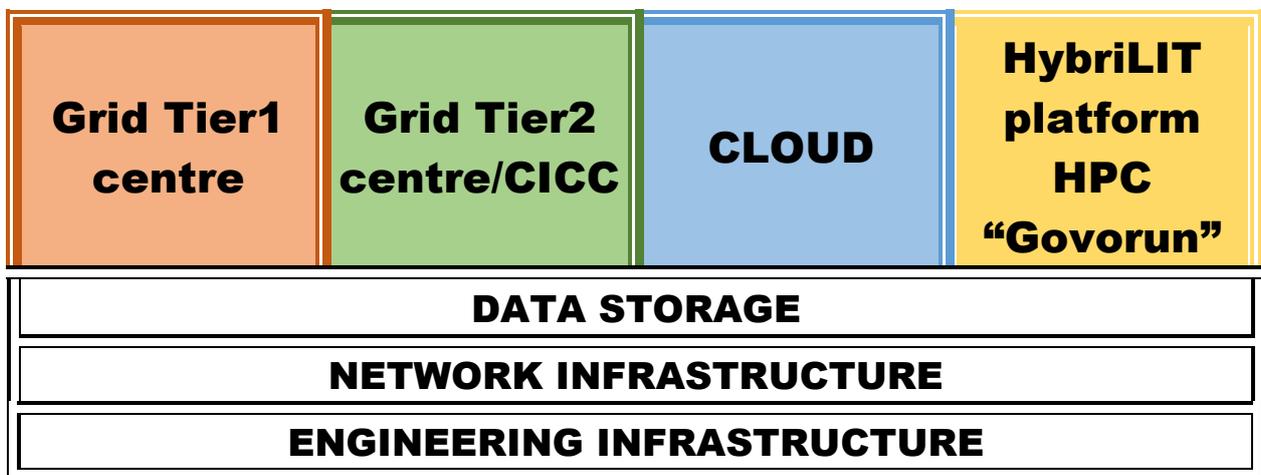


Figure 1. MICC main components

The MICC JINR was developed as a heterogeneous scalable data centre with common engineering and network infrastructures, data storage and monitoring systems. Thus, the main computing components of the MICC (Fig.1) are:

- Central Information and Computing Complex (CICC) of JINR with in-house build computing and mass storage elements and Tier2 for all experiments at the Large Hadron Collider (LHC) and other virtual organizations (VOs) in the grid environment [4],
- Tier1 for the CMS experiment [5],
- HybriLIT heterogeneous platform for High-Performance Computing (HPC) (including the “Govorun” supercomputer) [6],
- cloud infrastructure [7].

The JINR MICC resources are used for data storage, processing, analysis and modeling. The grid center resources of the JINR MICC are part of the global grid infrastructure WLCG (Worldwide LHC Computing Grid) [8]. The mission of the WLCG project is to provide global computing resources to store, distribute and analyze ~50-70 Petabytes of data expected every year of operations from the LHC. Nearly 170 sites in 42 countries are contributors in this project. More than 1,000,000 cores and 1 EB of storage were used to perform > 3 million jobs/day for computations on this infrastructure.

The next challenge for JINR is the NICA megaproject computing organization. Computing for the NICA megaproject should provide data acquisition from detectors and data transmission for processing and analysis. To perform the given tasks, computing has certain requirements including requirements to the network infrastructure, computing architectures, storage systems as well as

appropriate software of the system and for data processing and analysis. Developed computing models should take into account the trends in the development of network solutions, computing architectures and IT solutions, which allow combining supercomputer (heterogeneous), grid and cloud technologies and creating distributed, software-configured HPC platforms on their basis.

Another challenge is the requirements of modern neutrino experiments in data storage volumes and computing power, which have significantly increased. To effectively use resources of the information and computing environment of neutrino experiments, in which JINR employees participate, a decision on the creation of a unified neutrino information and computing platform (environment) based on the MICC resources was made. At present and over the next several years, Baikal-GVD, JUNO and NOvA neutrino scientific experiments with the JINR participation are the main consumers of storage resources and computing power of the JINR cloud.

2. MICC components

2.1. Network infrastructure

One of the most important components of the MICC as a multifunctional facility that provides access to resources and the opportunity to work with big data is the network infrastructure. All communications and control within the MICC are performed by means of fiber optic connections. In order to bring the MICC network infrastructure and telecommunication channels to the requirements of reliability and availability asked by international collaborations, which use the MICC resources for research, a mandatory backup of all connections and telecommunication channels is requested.

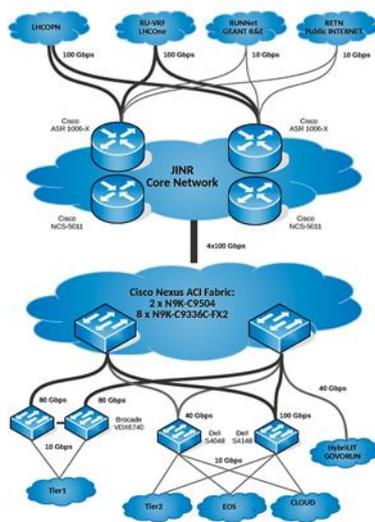


Figure 2. JINR network infrastructure

First years of the MICC project implementation laid foundation for the further development of the JINR network infrastructure, in particular, the projects related to increasing the bandwidth of the Moscow-JINR telecommunication channel to 3x100Gb/s, installing and configuring the equipment of the Institute backbone computing network to 2x100Gb/s and the distributed computing cluster network between JINR facilities to 400 Gb/s were carried out. The external JINR channel was built on the DWDM technology. The following telecommunication links are used for the connection with scientific networks and the Internet: LHCOPN/LHCONE/CERN (2x100Gb/s), Russian scientific networks and international scientific networks (4x10 Gb/s) (Fig.2). The reliability of external channels was improved due to the addition of two supplementary routers Cisco ASR-1006-X.

The inner MICC network has several dedicated network segments. The network segment of Tier1 was built of the Brocade equipment, in which the IS-IS (Intermediate System to Intermediate System) routing protocol is used for the calculation of the network segment on the second level of the OSI reference model (OSI - Open Systems Interconnection). This segment

allows a smooth interaction between 160 disk servers, 25 computational blade servers, 100 servers to support both the grid infrastructure and the tape robot.

Today, the Tier2/CICC network segments, the cloud environment, and the HybriLIT platform are no less important to meet the needs of the MICC computing complex and are built by analogy with the Tier1 network segment on the Dell and Cisco equipment. Ports up to 40 Gb/s are used to connect server components on access level switches. The core of the MICC network is built on Cisco Nexus 9504 and Nexus 9336C switches with a 100 Gb/s port bandwidth.

2.2. Engineering infrastructure

The MICC engineering infrastructure is designed to ensure the reliable, uninterrupted and fault-tolerant operation of the information and computing system and the network infrastructure. The use of the integrated approach to building the MICC engineering infrastructure allowed one to elaborate algorithms of the equipment operation and interaction of separate systems both in a normal operation mode and in emergencies, which ensured the uninterrupted performance regardless of

external factors. The guaranteed power supply system created during first years of the MICC project implementation ensures the guaranteed power supply of connected consumers, the automatic start of diesel-generator units (DGU), automatic load switching from the major external power supply network to DGUs and backward, sending an alarm to the dispatcher's post in case of an emergency with DGUs. The complex of the system consisting of uninterruptible power supplies (UPS) and two DGUs is a guaranteed power supply system, which ensures the complete energy independence of consumers from the external power supply network.

The MICC existing cooling control system is a complex of interconnected installations of various air and liquid cooling schemes, with the help of which the corresponding temperature regime ensuring the MICC functioning in a 24x7x365 mode is created. At present, the MICC climate control system has the following components: free cooling of the equipment with cooled air of the server hall; raised floor supply of cold air with a forced exhaust of hot air by ventilation panels; cooling of the cold corridor of the module by inter-row conditioners; liquid cooling of computing machine elements. According to the type of heat removal, the MICC climate control system refers to the mixed type of performance that combines systems with the evaporation of a refrigerant and systems with an intermediate coolant.

The features of the engineering infrastructure of the "Govorun" supercomputer, the CPU component of which is installed in the universal computing racks "RSC Tornado" [9] with a record energy density and a precision liquid cooling system, balanced for continuous operation with a high-temperature coolant with an optimal mode at a constant coolant temperature of +45 °C at the entrance to computing nodes (with a peak value of up to +57 °C), should be noted. The operation in a "hot water" mode for this solution allowed one to apply a year-round free cooling mode, using only dry cooling towers, cooling liquid with the help of ambient air on any day of the year, as well as completely get rid of the freon circuit and chillers. As a result, the average annual PUE of the system is less than 1.06.

2.3. Grid infrastructure

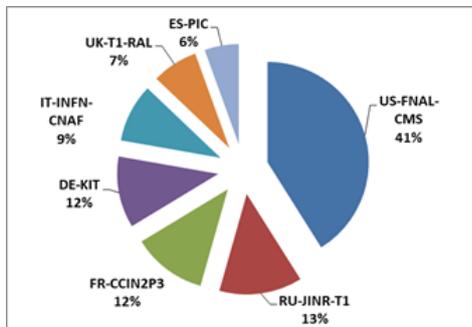


Figure 3. World Tier1 for CMS — Sum CPU Work (HS06 hours) by Tier1 centres and Years 2017-2019

The data processing system at JINR Tier2 consists of 4,128 cores for batch processing and provides a performance of 55.489 kHS06. The JINR Tier2 site is the best one in the Russian Data Intensive Grid (RDIG) federation and more than 11 million jobs were processed in 2017-2019, which accounts for 46.7% of the total CPU work of RDIG (Fig.4).

The MICC grid components are Tier1 for CMS at the LHC and the Tier2 center for Alice, ATLAS, CMS, LHCb, BES, BIOMED, COMPASS, MPD, NOvA, STAR, ILC, and others [10]. The data processing system at JINR CMS Tier1 consists of 10,688 cores and provides a performance of 151.97 kHS06. Tier1 (T1_JINR) took a second place among other Tier1 centers for CMS by completing over 19.9 million jobs and over 359 billion processed events, which is 19.3% of the total number of CMS events processed during 2017-2019, which accounts for 13% of the Sum CPU Work (HS06 hours) of the CMS experiment data processed by all Tier1 sites.

The data processing system at JINR Tier2 consists of 4,128 cores for batch processing and provides a performance of 55.489 kHS06. The JINR Tier2 site is the best one in the Russian Data Intensive Grid (RDIG) federation and more than 11 million jobs were processed in 2017-2019, which accounts for 46.7% of the total CPU work of RDIG (Fig.4).

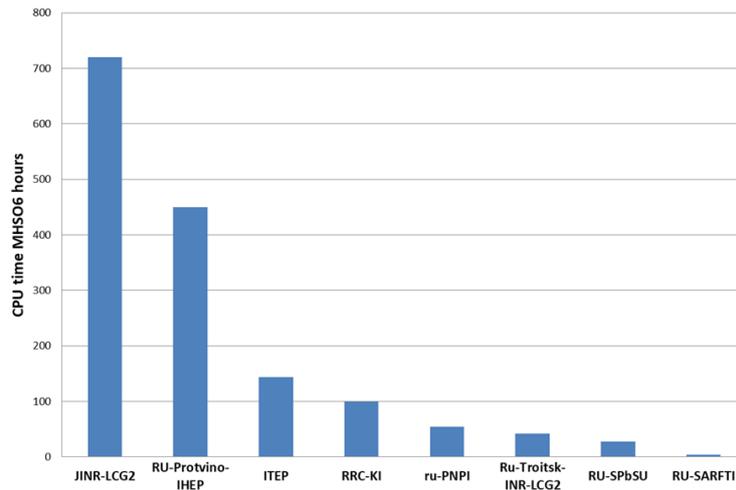


Figure 4. Sum CPU Work (MHS06 hours) by RDIG Tier2 centres and Years 2017-2019

2.4. Cloud infrastructure

The cloud infrastructure of JINR [7] operates on the basis of the OpenNebula 5.8 release which enables a relatively easy integration of supporting a container virtualization based on OpenVZ, as well as own extensions and additions, and is compatible with the Linux operating system (OS). It has the possibility of running virtual machines with this OS, the required functionality and quality of the software product, the license that permits modification and free use, the availability of clear and accessible documentation, and support from developers when modifying the software. In addition, OpenNebula appears to be an optimal choice in terms of the interrelationship of the hardware functioning in the infrastructure on its basis and the effort required for the development and maintenance of the cloud.

The JINR cloud resources were increased up to 1,564 CPU cores and 8.54 TB of RAM in total. Current hardware resources: 66 servers for VMs, 10 servers for ceph-based software-defined storage (SDS), 3 servers for front-end nodes in high availability setup.

The JINR cloud grows not only in the amount of resources, but also in the number of activities it is used for, namely, COMPASS production system services [11], a data management system of the UNECE ICP Vegetation [12], a service for scientific and engineering computations [13], a service for data visualization based on Grafana, jupyterhub head and execute nodes for it, gitlab and its runners as well as some others. Apart from that, there was a successful attempt to deploy a virtual machine in the JINR cloud with a GPU card transmitted from the server for developing and running machine and deep learning algorithms for the JUNO experiment. Moreover, the JINR distributed information and computing environment, combining resources from clouds of the JINR Member State organizations with the help of the DIRAC grid interware [14], began to be used to run BM@N and MPD experiment jobs.

2.5. HybriLIT platform and “Govorun” supercomputer

The commissioned in 2018 “Govorun” supercomputer named after Nikolai Nikolaevich Govorun, with whom the development of information technologies at JINR has been connected since 1966, became a natural elaboration of the HybriLIT heterogeneous cluster being the MICC component. The “Govorun” supercomputer is aimed at the significant speed-up of complex theoretical and experimental studies in the field of nuclear physics and condensed matter physics underway at JINR including the development of the NICA megaproject computing. The supercomputer commissioning led to the significant performance increase of CPU, as well as GPU, components of the HybriLIT heterogeneous cluster and, together with it, the supercomputer formed a heterogeneous platform. The platform consists of two elements, i.e. the education and testing polygon and the “Govorun” supercomputer, combined by the unified software and information environment.

The “Govorun” supercomputer [15] is a heterogeneous computing system containing the GPU component based on the NVIDIA graphics accelerator and the CPU component based on two Intel computing architectures.

The GPU component consists of 5 NVIDIA DGX-1 servers. Each server has 8 GPU NVIDIA Tesla V100 based on the latest architecture NVIDIA Volta. Moreover, one server NVIDIA DGX-1 has 40960 cores CUDA, which are equivalent to 800 high-performance central processors.

CPU computing nodes are based on Intel server products, namely, the most powerful 72-core server processors Intel® Xeon Phi™ 7290 (KNL), processors Intel® Xeon® Scalable (models Intel® Xeon® Gold 6154) (Skylake) and novel high-speed solid-state disks Intel® SSD DC P4511 with the NVMe interface and a capacity of 1 TB. For high-speed data transfer between computing nodes the supercomputer uses an advanced switching technology Intel® Omni-Path providing the speed of non-blocking switching up to 100 Gbit/s based on 48-port switches Intel® Omni-PathEdgeSwitch 100 Series with 100% liquid cooling.

The peak supercomputer performance is 1 PFlops for single-precision operations and 500 TFlops for double-precision operations. The average load on the computing components is the following: 80.58% on Skylake, 38.41% on KNL, and 73.58% on the component with GPU.

Currently 40 users from all Institute laboratories are performing calculations on the supercomputer. In total, over 135,000 tasks on all computing components were completed by all groups carrying out calculations on the supercomputer within the commissioning period.

2.6. MICC data storage

Systems of data storage and access, such as dCache, EOS and XROOTD, ensure work with data for JINR local users as well as WLCG users and collaborations. The JINR Tier1 storage system contains disk arrays and long-term data storage on tapes and is supported by the dCache-5.2 and Enstore 4.2.2 software. The total usable storage system capacity of disk servers is 10.4 PB; the IBM TS3500 tape robot is 11 PB. The storage system of JINR Tier2 contains disk arrays and is supported by dCache-5.2 and EOS. The total usable capacity of disk servers is 2,789 TB for ATLAS, CMS and Alice, and 140 TB for other VOs.

JINR joined a group of research centers that develop the WLCG data lake prototype for the HL-LHC [16]. The data lake prototype was built as a distributed EOS storage system and is used for storing and accessing big arrays of information. Global access to EOS is carried out by means of the WLCG software. The data lake prototype turned out to be small in terms of resources, but geographically distributed. The results were positive, and EOS was integrated into the MICC structure successfully. There is currently 3,740 TB of disk space available for EOS. EOS is visible as a local file system on the MICC working nodes and allows authorized users (by the kerberos5 protocol) to read and write data.

3. Monitoring system

A multi-level monitoring system was created for the JINR MICC [17]. It is based on different technologies such as Nagios, Icinga2, Grafana, and systems that were developed from scratch in JINR. The Icinga2 system is a suitable and reliable tool for hardware monitoring of all the MICC components. The Tier1 service monitoring system [18] and the HybriLIT monitoring system were developed in JINR. In order to provide the secure accessibility to all monitoring and control systems mentioned above, the MICC Operational Center was thoroughly designed and created [19]. It is a reliable place where operators can effectively control the MICC in a wide range of critical situations like a global power cut or a network failure. At the moment, the monitoring system controls all types of the MICC equipment. The number of nodes included in monitoring is more than 1,200. To ensure such extensive monitoring, a cluster monitoring system based on the Icinga2 software is used. Visualization is carried out with the help of add-ons such as Grafana and NagVis. With the help of these systems, data is also obtained from the JINR cloud infrastructure. Thus, despite the presence of their own monitoring systems in various components of the MICC, operators can acquire data on each element of the computer complex in a single access point.

4. Future plans

The further development and performance extension of the JINR MICC, as well as the provision of novel IT solutions to MICC users, and the increase in its operation efficiency are the uppermost tasks of the JINR Laboratory of Information Technologies.

The first main direction is related to the efficient extension of the MICC resources being a combination of various computing complexes and infrastructure units, namely, the LHC data processing centers of Tier1 and Tier2 levels, the cloud infrastructure, the “Govorun” supercomputer, the multicomponent data storage, the data transmission network, the MICC engineering infrastructure and the monitoring system. The implementation of this direction is impossible without creating a unified resource management system designed to provide the following: a unified user authorization system with support for user groups and virtual organizations, integrated with external authentication systems; a unified interface to computing resources and data storage systems; the ability to manage quotas and priorities of the resource usage.

Another important direction in the MICC development plan is the modernization of data storage systems; it is connected with the significant increase of information volumes expected in 2020-2021, which are to be stored and processed. In this regard, it is necessary to provide the following: a sufficient resource for storage and fast access to information during processing; a constantly expanding resource for long-term data storage; the ability to use a data management system that automates processes of interaction with storage systems.

References

- [1] NICA (Nuclotron-based Ion Collider fAcility), <http://nica.jinr.ru>
- [2] LHC (Large Hadron Collider) <https://home.cern/science/accelerators/large-hadron-collider>
- [3] V. Korenkov, A. Dolbilov et al, The JINR distributed computing environment, EPJ Web of Conferences 214, 03009, 2019, <https://doi.org/10.1051/epjconf/201921403009>
- [4] N.S. Astakhov, A.S. Baginyan, A.I. Balandin, et al., JINR grid TIER-1@TIER-2, CEUR-WS.org/Vol-2023/68-74-paper-10.pdf, 2017
- [5] A.S. Baginyan, A.I. Balandin, S.D. Belov, et al. The CMS TIER1 at JINR: five years of operations, CEUR-WS.org/Vol-2267/1-10-paper-1.pdf, 2018
- [6] Heterogeneous platform “HybriLIT”, <http://hlit.jinr.ru/en/>
- [7] A.V. Baranov, N.A. Balashov, N.A. Kutovskiy, et al., Present status and main directions of the JINR cloud development, *ibid.*
- [8] WLCG (The Worldwide LHC Computing Grid): <http://wlcg.web.cern.ch/LCG>
- [9] RSC Group. <http://www.rscgroup.ru/en/company>
- [10] A.S. Baginyan, A.I. Balandin, A.G. Dolbilov, et.al., Grid at JINR, *ibid.*
- [11] A.Sh. Petrosyan, COMPASS Production System Overview, EPJ Web of Conf., Vol. 214, 2019, <https://doi.org/10.1051/epjconf/201921403039>
- [12] A. Uzhinskiy, G. Ososkov, M. Frontsyeva. Management of the environmental monitoring data: UNECE ICP Vegetation case, *ibid.*
- [13] N. Balashov et al, Service for parallel applications based on JINR cloud and HybriLIT resources, EPJ Web of Conferences 214, 07012 (2019), <https://doi.org/10.1051/epjconf/201921407012>
- [14] A.V. Baranov, N.A. Balashov, A.N. Makhalkin, et al., New features of the JINR cloud, CEUR-WS.org/Vol-2267/257-261-paper-48.pdf, 2018
- [15] Supercomputer “Govorun”, http://hlit.jinr.ru/en/about_govorun_eng/
- [16] A.S. Baginyan, N.A. Balashov, A.V. Baranov et al. Multi-level monitoring system for multifunctional information and computing complex at JINR, CEUR-WS.org/Vol-2023/226-233-paper-36.pdf, 2017
- [17] I. Kadochnikov, I. Bird, G. McCance, et al., WLCG data lake prototype for HL-LHC, CEUR-WS.org/Vol-2267/509-512-paper-97.pdf, 2018
- [18] I. Kadochnikov, V. Korenkov, V. Mitsyn, et al., Service monitoring system for JINR Tier-1, EPJ Web of Conferences 214, 08016, 2019, <https://doi.org/10.1051/epjconf/201921408016>
- [19] A. Golunov, A. Dolbilov, I. Kadochnikov, et al., CEUR-WS.org/Vol-1787/235-240-paper-39.pdf