

## Research of Acceleration Calculations in Solving Scientific Problems on the Heterogeneous Cluster HybriLIT

E. I. Alexandrov\*, D. V. Belyakov\*, M. A. Matveyev\*,  
D. V. Podgainy\*, O. I. Streltsova\*, Sh. G. Torosyan\*,  
E. V. Zemlyanaya\*, P. V. Zrellov\*<sup>†</sup>, M. I. Zuev\*

\* *Laboratory of Information Technologies  
Joint Institute for Nuclear Research*

*6, Joliot-Curie str., Dubna, Moscow region, Russia, 141980*

<sup>†</sup> *Plekhanov Russian University of Economics  
36, Stremyanny per., Moscow, Russia, 117997*

The paper presents some test results of the heterogeneous computing cluster HybriLIT put into operation at the Laboratory information technologies of the Joint Institute for Nuclear Research. The structure of the cluster includes computational nodes with NVIDIA graphical accelerators and Intel Xeon Phi coprocessors. The necessity of integration of such a computational platform in the JINR Multifunctional Information and Computing Complex is determined by a global tendency to use hybrid computing architectures for carrying out massive-parallel computations in applied scientific problems solution. Test of the cluster aimed at: first of all, test of the efficiency of hardware and software settings that include operational system, resource manager, file system, compilers, and, secondly, test of the efficiency of using different architectures for the solution of particular applied problems in order to provide user guides on specialized libraries. For realization of the cluster test, an approach that includes test computations by means of standard program packages such as Linpack and program complexes established in LIT has been developed. The presented results show that the use of hybrid computing architectures allow accelerate the solution of applied scientific problems, and heterogeneous computing cluster HybriLIT is an effective means of accomplishing this aim.

**Key words and phrases:** high performance platform, Linpack benchmarks, technology of parallel programming, heterogeneous computing.

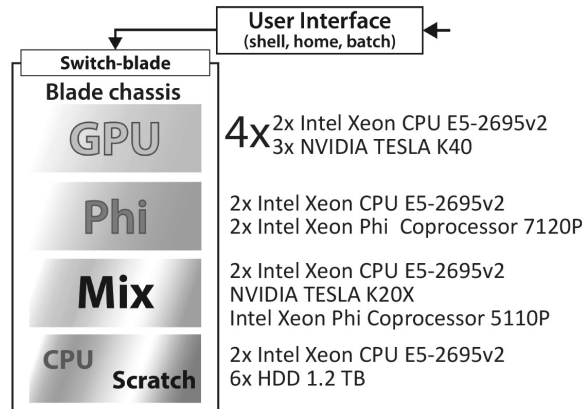
### 1. Introduction

Nowadays computations with the use of high performance computing systems that include different types of computation accelerators are becoming widespread in scientific and applied scientific researches. This is clear from the list of TOP500 supercomputer sites [1]; according to the 45th edition of the TOP500, 35% of all platforms performance is provided by coprocessors of various architectures; and half of them – by NVIDIA graphical processors [2]. Following this tendency, the producers of applied software are adjusting the developed packages and specialized mathematical libraries, as well as create new ones with regard to the development of hardware in the field of high performance computations. That provides an effective use of computation accelerators in scientific tasks. For example, nowadays NVIDIA graphical accelerators contain a wide range of applied software for carrying out scientific researches, computations, etc. [3]. Heterogeneous computing platform HybriLIT [4] has been developed in the Laboratory of Information Technologies to support scientific researches that require massive parallel computing. The platform contains NVIDIA graphical accelerators and Intel Xeon Phi coprocessors. This platform is used not only for carrying out computations, but also as a polygon for the development of hybrid applications that use new computing architectures. The current configuration of the cluster contains seven computational nodes with GPU and Intel Xeon Phi coprocessors that include:

- GPU nodes with three TESLA K40s graphical processors (4 nodes);
- Phi node with two Intel Xeon Phi 7120P coprocessors;
- MIX node with Tesla K20X graphical processor and Intel Xeon Phi 5110P coprocessor;

- Scratch node with two Intel Xeon E5-2695 v2 processors.
- Theoretical performance of the cluster is following:
- single precision – 77 TFlops;
  - double precision – 29 TFlops.

Fig. 1 shows the structure of the cluster.



**Figure 1. The structure of the cluster HybriLIT**

This paper presents the test results of the cluster HybriLIT. The research pursued two main aims: first of all, settings efficiency test of the hardware and basic software that includes OS Scientific Linux 6.6, SLURM [5], NFS4 file system [6], compilers, etc.; secondly, efficiency test of using different architectures for the solution of particular applied tasks in order to provide users some recommendations on the use of specialized libraries such as libraries for matrix-vector operations, fast Fourier transformation, random data generator, etc. that allow to accelerate computations.

The described test procedure allows to consider particularities of various architectures and includes test tasks of different types:

- computational tasks that may be a part of big programs or packages;
- computational tasks that combine intensive memory access, etc.;
- computational tasks that use mathematical libraries (BLAS, LAPACK, FFW, etc.).

The overall test of the cluster HybriLIT included:

- use of standard test programs (Linpack [7] and other);
- test tasks based on the most frequently used parallel programming technologies;
- program packages for the solution of applied tasks: GIMM\_FPEIVE [8] and Multi Configurational Time Dependnet Hartree (for) Bosons (MCTDHB) [9] packages.

The paper presents the test results with the use of Linpack package and GIMM\_FPEIVE program complex.

## 2. Test of the components of HybriLIT cluster with the Linpack benchmark

Linpack benchmark test detects the performance of the whole computational system as well as of the parts of its components [7]. The first test version was introduced in 1979 as an additional part to the Linpack library for solution of various systems of linear algebraic equations (SLAE). Ever since Linpack benchmark has undergone numerous upgrades according to the fast developing computational platforms. However, its main target based on the solution of SLAE  $Ax = f$  by means of LU-factorization with partial pivoting has remained. Nowadays, this test is used for the performance detection of supercomputers included in TOP500 [1].

There are several implementations of the test; in this paper, there were used two of them: Linpack benchmark is a test for systems with shared memory, and MP Linpack benchmark is a specialized expansion that allows testing systems with distributed memory. The first version of the test was used for testing multicore computational components of each node; the second version was used for testing coprocessors and graphical accelerators. It should be noted that the second version of the test can be used for performance detection of the cluster [7] as well.

Specifications of the HybriLIT components can be found at [4].

## 2.1. Description

### CPU test

Test of computational nodes with Intel Xeon processors was carried out by means of Linpack benchmark included into Intel Cluster Studio version 2013 sp1.2.144. The matrix size varied from 1000 to 45000. Maximal performance was reached at about  $N = 40000$ .

### GPU test

Tests were held by means of MP Linpack benchmark, hpl-2.0\_FERMI\_v15 version that allows to detect the overall performance of a computational node with GPU. In the research of the integral performance of CPU + GPU, the matrix size varied from 1000 to 120000. Maximal performance was reached at  $N = 120000$ .

### CPU + Intel Xeon Phi Co-Processor test

Test of computational nodes with Intel Xeon Phi coprocessors was carried out by means of Linpack benchmark included into Intel Cluster Studio version 2013 sp1.2.144. In the research of the integral performance of CPU + Xeon Phi, the matrix size varied from 1000 to 120000. Maximal performance was reached at  $N = 120000$ .

## 2.2. Results

### GPU test results

On the basis of the Linpack benchmark test for GPU, performances for different number of CPU cores per graphical accelerator were obtained (see Fig. 2).

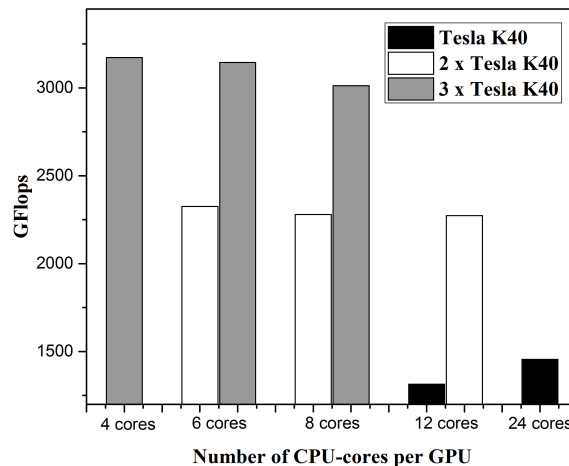


Figure 2. Dependence of the performance on the number of cores per GPU.

It should be noted that maximal performance for one graphical accelerator is reached at using cores of both processors, while the maximal performance for two or three GPU is reached by means of using half cores.

### Results for mixed configurations

It should be pointed out that as the integral performance of a computational node depends on the processor frequency, an additional research on the dependency of performance on fixed processor frequency has been carried out.

Figs. 3, 4 show the comparison of efficiency and performance of using different computational elements for two processor frequency – minimal and maximal; the efficiency is detected as the dependence of the obtained result of the test to the theoretical part that is given by the producer.

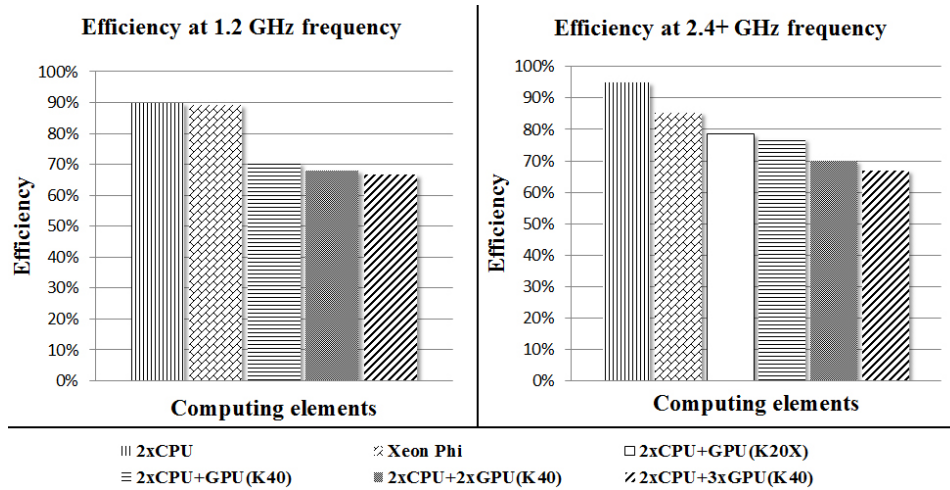


Figure 3. Comparison of the efficiency of using different computational elements at 1.2 GHz and 2.4+ GHz frequency.

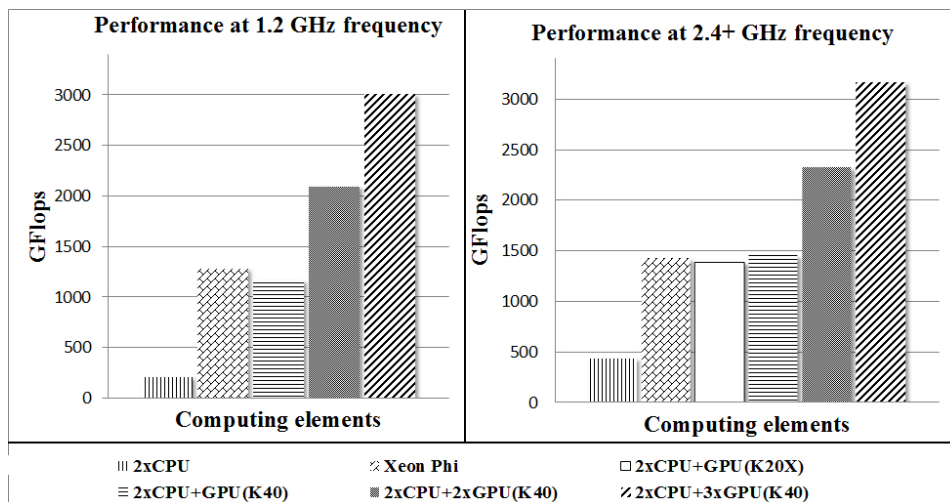


Figure 4. Comparison of performance of different computational elements at 1.2 GHz and 2.4+ GHz frequency.

As it can be seen in Fig. 3, the efficiency of using only CPU is higher than its use with accelerator. It is connected with the data transfer from one device to another. It should be noted that at minimal processor frequency 1.2 GHz, the efficiency of Intel Xeon Phi can be compared to CPU efficiency. It stems from the fact that the central processor core frequency and coprocessor frequency are almost equal.

From the performance comparison shown in Fig. 4 it can be seen that the performance of CPU is essentially lower than the performance of CPU + Xeon Phi and CPU + GPU hybrid architectures. As the number of accelerators increase, the performance of hybrid system is growing by times. The performance of a system consisting of two CPU and one coprocessor at 1.2 GHz frequency exceeds performance of 2CPU + GPU system, but at maximal processor frequency, the situation is opposite.

### 3. Test on the basis of GIMM\_FPEIVE package

Test of the cluster HybriLIT was also carried out on the basis of GIMM\_FPEIVE program complex that is meant for modeling of thermophysical processes that appear at material irradiation with ion beams [8]. A particular characteristic of the package is in its module structure that allows carrying out computations on different computational platforms.

Compilation of the package was carried out both by means of shareware compilers GCC and specialized ones such as: Intel Cluster Studio for the systems with Intel processors with various options that takes account of the particularities of the architecture and different levels of optimization. For Intel architecture there were used the following options: `-march=native` was used for GCC compiler and `-march=core-avx-i` was used for Intel Cluster Studio compilers. As a result of these options, computation time cutting for 38% was achieved.

Additional options `-env I_MPI_PIN_DOMAIN socket` `-env I_MPI_PIN_ORDER compact` were added in batch script for calculations with Intel Cluster Studio and option `-map-by core` for the calculations with GCC compiler. These options allowing to allocate the amount of processes between CPUs on the node.

Fig. 5 depicts the maximal and minimal computation time for the different size of the task at using GCC and Intel Cluster Studio compilers on the basis of MPI module of GIMM\_FPEIVE package on the cluster HybriLIT. Details of mathematical statement of problem and parallel algorithm are described in [10]. Numbers on each diagram show the reached acceleration for each calculation. For methodical purposes, the number of timesteps was taken relatively small. Maximal time on the diagrams corresponds a serial execution. Minimal time is chosen from parallel computation times for different number of MPI processes.

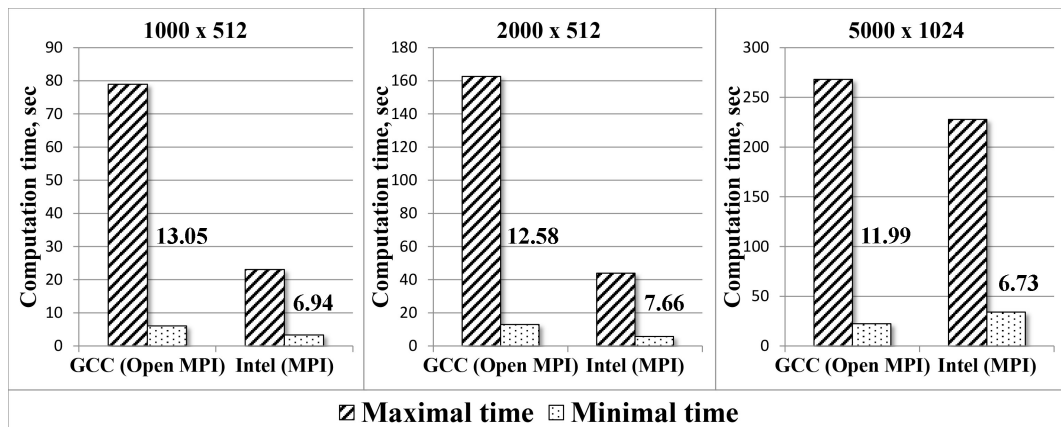
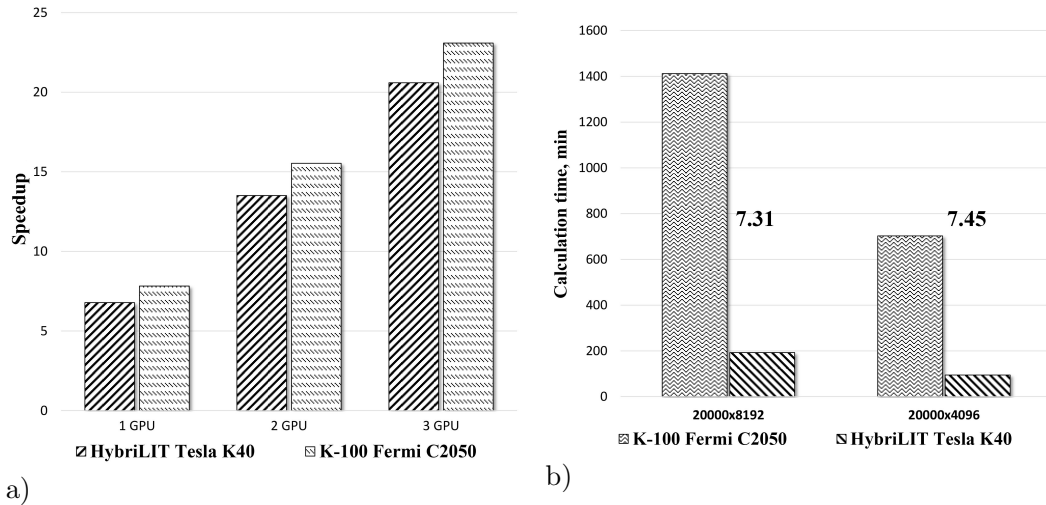


Figure 5. Value of maximal and minimal computation time for the different size of the task at using GCC (Open MPI) and Intel (MPI) compilers.

On the basis of CUDA module GIMM\_FPEIVE package, computations with the use of several NVIDIA graphical accelerators were carried out. Fig. 6 depicts computation speedup on different number of GPU according to computation in a serial mode. All computations were carried out at compute capability = 3.5.



**Figure 6. a) Dependence of computation speedup on the number of GPU according to serial computation b) Dependence of computation time on the task size**

One of the peculiarities of using GPU in computations is limited memory of the device. In order to estimate computations' efficiency on different types of GPU, computations for the task that exceeds the memory of the accelerator have been carried out. The comparison was made for Fermi C2050 accelerator with 2.5 Gb memory and Tesla K40s with 12 Gb memory. The task for Fermi C2050 was carried out in parts that were transferred to the device sequentially. Task with grid dimension  $2000 \times 4096$  requires 5 Gb GPU memory, and 10 Gb memory for  $2000 \times 8096$  dimension. Therefore, first case requires to perform an additional data transfer between CPU and GPU, and three transfers in the second case. Fig. 6b depicts computation time of tasks of the previously described dimensions on the cluster HybriLIT and K-100.

As it can be seen in Fig. 6b, data transfer between CPU and GPU takes a lot of time and computation time on Tesla K40s exceeds computation time on Fermi C2050 by 7 times.

## 4. Conclusion

The paper presents test results of the heterogeneous cluster HybriLIT for the solution of different applied scientific problems. Carried out test allowed to improve system setting of the cluster, resource manager SLURM, and also to provide user guides on effective use of the software such as compilers with various options, specialized libraries and applied program packages. As it can be seen from the obtained results, essential computation speedup for the chosen test problems have been reached. In particular, acceleration for MPI applications at using one node increased by 13 times, and acceleration by 50 times was reached for CUDA applications. Such results prove the efficiency of using hybrid computation platforms and specialized software for an essential acceleration of scientific computations to which heterogeneous computing cluster HybriLIT is related.

## References

1. Top500 supercomputer sites.  
URL <http://www.top500.org>

2. Nvidia company.  
URL <http://www.nvidia.com/page/home.html>
3. Gpu applications. high performance computing.  
URL <http://www.nvidia.com/object/gpu-applications.html>
4. Heterogeneous computing cluster hybrid of lit jinr.  
URL <http://hybrilit.jinr.ru>
5. The job management systems slurm: Simple linux utility for resource management.  
URL <http://slurm.schedmd.com/>
6. Network file system (nfs) version 4 protocol.  
URL <https://www.ietf.org/rfc/rfc3530.txt>
7. Intel optimized linpack benchmark and the intel optimized mp linpack benchmark.  
URL <https://software.intel.com/en-us/articles/intel-math-kernel-library-linpack-download>
8. E. I. Alexandrov, I. V. Amirkhanov, E. V. Zemlyanaya, et al., Principles of software construction for simulation of physical processes on hybrid computing systems (on the example of gimm\_fpeip, Bulletin of PFUR. Series "Mathematics. Informatics. Physics" (2) (2014) 197–205, in Russian.
9. Mctdhub-lab software.  
URL <http://QDlab.org>
10. I. V. Amirkhanov, E. V. Zemlyanaya, N. R. Sarker, et al., Mpi implementation of the 2d and 3d simulation of phase transitions in materials irradiated by heavy ion beams within the thermal spike mode, Bulletin of PFUR. Series "Mathematics. Informatics. Physics" (4) (2013) 80–94.

УДК 004.415.538

## Исследование ускорения вычислений при решении научных задач на гетерогенном кластере HybriLIT

**Е. И. Александров\***, **Д. В. Беляков\***, **М. А. Матвеев\***,  
**Д. В. Подгайный\***, **О. И. Стрельцова\***, **Ш. Г. Торосян\***,  
**Е. В. Земляная\***, **П. В. Зрелов\*†**, **М. И. Зуев\***

*\* Лаборатория информационных технологий  
Объединённый институт ядерных исследований  
ул. Жолио-Кюри, д. 6, Дубна, Московская область, Россия, 141980*

*† Российский экономический университет имени Г.В. Плеханова  
Стремянный пер., д. 36, Москва, Россия, 117997*

В работе представлены некоторые результаты тестирования введенного в эксплуатацию в Лаборатории информационных технологий Объединенного института ядерных исследований гетерогенного вычислительного кластера HybriLIT. В состав кластера входят вычислительные узлы с графическими ускорителями NVIDIA и сопроцессорами Intel Xeon Phi. Необходимость введения такой вычислительной платформы в состав Многофункционального информационно-вычислительного центра ОИЯИ обусловлена мировой тенденцией применения гибридных вычислительных архитектур для проведения массивно-параллельных вычислений при решении научно-прикладных задач. Проводимое тестирование кластера преследовало две основные цели, а именно тестирование эффективности настройки аппаратной части и базового программного обеспечения, включающего в себя операционную систему, планировщик задач, файловую систему, компиляторы, а также исследование эффективности использования различных архитектур для решения конкретных прикладных задач с целью выработки рекомендаций пользователям по использованию специализированных библиотек. Для реализации тестирования кластера была разработана методика, включающая в себя тестовые расчеты как с использованием стандартных пакетов программ, таких как Linpack, так и программных комплексов созданных в ЛИТ. Из представленных в работе результатов следует, что использование гибридных вычислительных архитектур позволяет существенно

ускорить решение научно-прикладных задач, а гетерогенный вычислительный кластер HybriLIT является эффективным средством для достижения этой цели.

**Ключевые слова:** высокопроизводительные вычислительные комплексы, технологии параллельного программирования, программные комплексы, Linpack benchmarks, гетерогенные вычисления.

## Литература

1. TOP500 Supercomputer Sites. — <http://www.top500.org>.
2. NVIDIA Company. — <http://www.nvidia.com/page/home.html>.
3. GPU Applications. High Performance Computing. — <http://www.nvidia.com/object/gpu-applications.html>.
4. Heterogeneous computing cluster HybriLIT of LIT JINR. — <http://hybrilit.jinr.ru>.
5. The job management systems SLURM: Simple Linux utility for resource management. — <http://slurm.schedmd.com/>.
6. Network File System (NFS) version 4 Protocol. — <https://www.ietf.org/rfc/rfc3530.txt>.
7. Intel Optimized LINPACK Benchmark and the Intel Optimized MP LINPACK Benchmark. — <https://software.intel.com/en-us/articles/intel-math-kernel-library-linpack-download>.
8. Принципы построения программного комплекса для моделирования физических процессов на гибридных вычислительных системах (на примере комплекса GIMM\_FREIP) / Александров Е. И., Амирханов И. В., Земляная Е. В. и др. // Вестник РУДН. Серия «Математика. Информатика. Физика». — 2014. — № 2. — С. 197–205.
9. MCTDHB-lab software. — <http://QDlab.org>.
10. MPI реализация алгоритмов для 2D и 3D моделирования фазовых переходов в материалах, облучаемых тяжёлыми ионами, в рамках модели термического пика / Амирханов И. В., Земляная Е. В., Саркар Н. Р. и др. // Вестник РУДН. Серия «Математика. Информатика. Физика». — 2013. — № 4. — С. 80–94.