

## **VIRTUAL TESTBED AS A CASE FOR BIG DATA**

**A.B. Degtyarev<sup>1,a</sup>, A.V. Bogdanov<sup>1</sup>, V.V. Korkhov<sup>1</sup>, I.G. Gankevich<sup>1</sup>,  
Yu.V. Pylnev<sup>2</sup>, A.V. Eibozhenko<sup>2</sup>**

<sup>1</sup> *Department of Computer Modeling and Multiprocessing Systems, St.Petersburg State University, 7/9 Universitetskaya nab., St. Petersburg, 199034, Russia*

<sup>2</sup> *Engineering company «NEOTECH MARINE», 23 Zastavskaya str., St. Petersburg, 196084, Russia*

E-mail: <sup>a</sup> a.degtyarev@spbu.ru

Complex modeling of the behavior of marine objects under the influence of real external excitation is the most important problem. At present, the accuracy of direct simulation of phenomena with known physics is comparable to the accuracy of the results obtained during the model experiment in towing tanks. Particularly relevant is the creation of such marine virtual testbed for full-featured simulators and when testing the knowledge base of onboard intelligent systems. Such integrated environment is a complex information object that combines the features of both the enterprise system and the high-performance modeling tool. Integrated environment based on these basic principles is designed to solve in real time the following problems: 1. Collection and analysis of information on the current state of dynamic object (DO) and the environment, remote monitoring of the state of objects. 2. Evaluation and coordination of joint actions of DOs, proceeding from current conditions, with the aim of optimally common problem solving. 3. Centralized support for decision-making by operators of DO control in non-standard situations, organization of information support for the interaction of decision-makers in the conduct of ongoing operations. 4. Computer modeling of possible scenarios of situation development with the aim of selecting the optimal control strategy. 5. Centralized control of technical means. 6. Cataloging and accumulation of information in dynamic databases. Modern architecture of computer systems (especially GP GPU) allows direct full-featured simulation of a marine object in real time. Efficient mapping to a hybrid architecture allows even the ability to render ahead of time under various scenarios. The report discusses general concept of high-performance virtual testbed development and the experience of creating on their basis full-featured simulators.

**Keywords:** virtual testbed, Big Data, complex system, naval hydrodynamics

© 2018 Alexandr B. Degtyarev, Alexander V. Bogdanov, Vladimir V. Korkhov, Ivan G. Gankevich, Yury V. Pylnev, Anatoly V. Eibozhenko

## 1. Introduction

Recently more and more technical cases are designed and modeled on computers and this tendency seems to be dominating. Nevertheless, there are situations very important for humanity that cannot be modeled properly and reliably yet. The most obvious cases are ships modeling and design and atomic power plants security. We shall try to analyze this problem from the point of view recent advances in fundamental informatics.

The key item for the approach of the modeling is mathematical foundation. In most cases the different pieces of event under consideration are described by different mathematical approaches. And thus to make the mathematical model of our object is the separate problem. The matching of different mathematical pieces is extremely important in production of the code. Very often the scale of the events varies to such degree that asymptotical methods are of a great help [1, 2, 3]. In most cases we have to deal with what is called "complex system simulation" [4].

Usually for different parts of the objects under consideration software packages were proposed. But they were produced by different approaches for different platforms and usually they are optimized for different software. The process of unification of all the approaches is called "consolidation" and it is the separate problem of itself [5, 6, 7].

Even a consolidation is achieved the procedure of the modeling is a taught one due to different of scales and necessity of memorizing large amount of data from one iteration to another. On this stage we claim that this problem is most effectively solved by retrieving of some ideas of so called "Big Data approach" [6]. More than that it seems, that the complex systems simulation is very effective case for "Big Data" analysis.

It is very important that this approach makes it possible to organize the visualization of intermediate results in the process of computation. It is important not only for correction of parameters, but it might be a key element and in modification of the model and consolidation of the procedure.

We shall try to illustrate main ideas of our approach on the case of ship modeling that is realized in a project "virtual testbed" in St. Petersburg State University.

## 2. Mathematical models

The concept of a virtual testbed is a single consideration of a technical object from its design to disposal. This concept combines the use of all heterogeneous information about an object, methods of storage and modeling, and the extraction and acquisition of new knowledge. General structure of virtual testbed is shown on Figure 1. Here A is a block of models, B is a block of control and interpretation, C is a block of dataware, D is a block of applied information technology. Here we have different kinds of models: models for estimation of dynamic object, navigation and operation situation ( $M_1$ ); control models ( $M_2$ ) including on-board systems models ( $M\Phi$ ), models of dynamic object elements/systems interaction (MU); block of planning ( $M_3$ ), where MP are models of long-term and operative planning, MS are control models for components of virtual testbed. In B block we have DS – dialog system for virtual testbed control; LS are local control systems; SA is a block of scenarios; RB is a block of practical recommendations. Dataware block represents in general a distributed database of various components of a virtual testbed (DB), obtained through consolidation. Block D combines virtual reality components (VR), mathematical components of information technology (soft computing, multiagent systems, etc. – SC&MAC) and components of distributed computer environment (Big Data, cloud computing, hybrid computer systems, etc. – GRID).

Already now, IT allows to fully compare the real results of human activity with their virtual counterpart. This means that it is possible to organize such an integrated system in which all the accumulated data are linked to each other and used when necessary. For example, drawings of technical object created during design process are used in the future to implement the simulator of the object. The data obtained as a results of direct simulation are used to form knowledge base of on-board intelligent system, etc. So we see that that this is problem of Big Data.

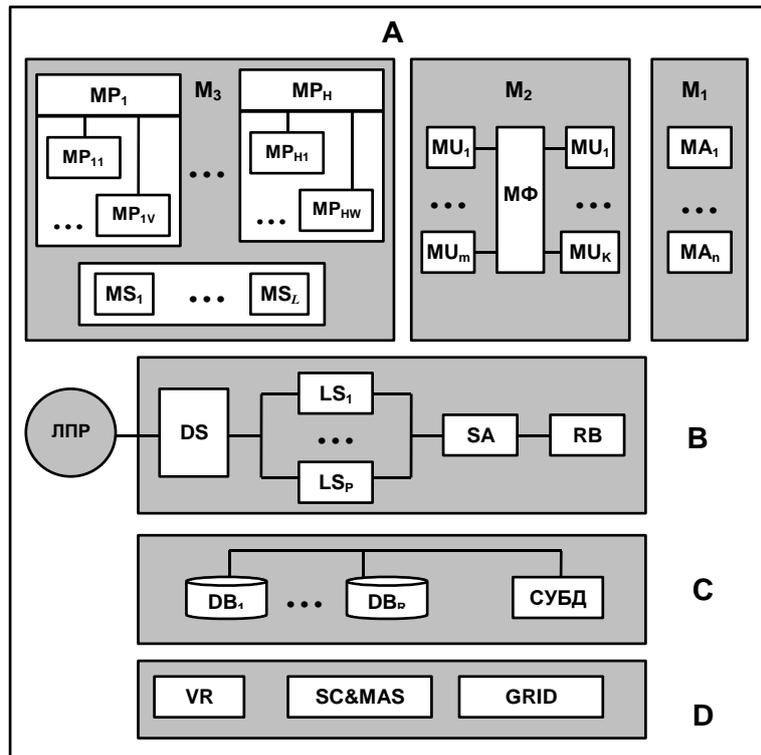


Figure 1. Structure of virtual testbed

It is obvious that such approach, when the simulation results can be fully matched with the behavior of a technical object in natural conditions, requires high accuracy and adequacy of the models used. For example, for naval hydrodynamics problems we can apply approach of direct computational experiment [8, 9].

Algorithmic implementation in this case is based on computational schemes of mixed Lagrangian and Eulerian approaches. This is expressed similar "large particles" and "final volume" methods in the double integration of first order equations of motions. Thus, the time cycle of computing experiment is divided into three conditional stages:

1 stage – Kinematic parameters are calculated for the centers of large fluid particles. For this purpose, the current source data at fixed nodes of Eulerian coordinates are used;

2 stage – Lagrangian or large deformable fluid particles are involved in free motion. They redistribute the internal properties of the original Euler cells to adjacent space;

3 stage – Laws of conservation of mass and energy are consistent. This is achieved by deformation of shifted fluid particles. The next step is reinterpolation of characteristics of current in initial nodes of the fixed Euler computational mesh.

The described algorithmic approach, mainly because of the proposed splitting of the solution on physical processes, gives the chance of explicit numerical schemes application at the first two stages. In this case, it is possible to substantially increase effectiveness of computing procedures through:

1. Natural parallelization of computation process;
2. Possibilities of adaptive correction of mesh area if needed;
3. Dynamic reconstruction of solution in accordance with fluid currents transformations in time.

Another approach [10] for sea wind waves simulation is based on fast mathematical models that are well suited for numerical calculations. Such kind models are not based on rigorous physical principles, as in the previous case. However, if preliminary studies provide an opportunity to prove their physical adequacy [11,12], they are also suitable for use as part of virtual testbed. However, such models are often based on an extensive statistical generalization of natural data. Therefore, they require careful preparation of the initial information, which is reflected in the parameters of the model used. In the considered problem of sea waves modeling, in order to adequately reproduce external influences, all conditions of wave formation are taken into account: geographical area, season, and the

alternation of storms characteristic of a given area. The advantage of such models, due to their simplicity, is the possibility of a relatively simple increase in their performance [10, 13, 14].

### **3. Consolidation technology**

Thus, a virtual testbed is designed to solve complex modeling problems and work with big amount of data. So the main aspects of virtual testbed are the following:

- Computing machinery – hardware;
- Uniform information environment – GRID, middleware;
- Program repository – libraries;
- System integration – principles of testbed operation;
- Concept of real time systems.

Grid management can be known as the distributed database management of the problem of data from a different point of view. We based our work on the object of oriented database management systems and its particular objectivity [15]. We compared the approach and find out the common for most efficient Data Grid that can manage the large of data stored in object-oriented databases. We used a reliable basis for such an effort the procedures and Data Grids are new in the Database research community after our test result and we know about the identifying the characteristics and requirements of Data Grids and how they can be met in the most efficient way.

Optimizing data is replication and access to the data over the wide area network (WAN), this is not addressed sufficiently in database research. Database Management System (DBMS) is normally one of the accessing data methods. For instance, a data server sends data to a client. For the Data Grid, such a single access method may be not optimal. By using an ODBMS, some restrictions are pointed to given some possible solution.

The Globus project provides tools for the Grid computing like a job scheduling and also working on Data Grid effort, that enables fast and efficient file transfer, a replica catalogue for managing files and some of the more replica management functionality is based on the more file. In the Grid community is a general tendency to deal with replication at the file level, such as a single file is the lowest granularity of replication [4]. It is the advantage of the structure file does not need to be known as the replication manager that is responsible for replicating files from one site to other sites over the Wide Area Network (WAN).

Middleware are all the services and applications necessary for efficient management of the data sets and files within the data grid, while the providing of users quick access to the data and files. Data access services are working hand in hand with the data transfer services to provide security service, access controls for management and any data transfers within the data grid. Security services provide mechanisms for authentication of users to ensure that properly identified. Using the password is one of the security services for authentication. Authorization services are the mechanisms that can control the user to be able to access after being identified after the authentication process.

Research can be combined from the both of communities by introducing a replication middleware layer that manage replication files by taking into the account of each site in a Data Grid, this can be managed data locally with a database management system. The middleware is responsible for site-to-site replication and synchronization in DBMS transactions on local data. In Grids, there are so many tools for monitoring applications and network parameters, which can be used for filling the gap. Hybrid solution is the replication middleware that has more restrictions for update synchronization and transparent access to data between database and Grid research. There will also be a performance of replica synchronization. However, the replication middleware is to provide several relaxations of the concept of transparent data access and data consistency. In this paper we assume that such a replication middleware is used for Data Grid.

Our project is carried out at the Saint Petersburg State University research center and it can be divided into 4 modules as followed:

- Module Network
- Random Dynamic process
- Data grids replacement
- Data Secure

The Clients server and Main servers are operated over a computer network on hardware separate in our project. A server machine is running one or more server programs with a high performance and sharing resources for clients. A client also shares any data from resources; clients need to initiate communication sessions with servers, which await incoming requests. The delivery packet is to the destination at a node in order to minimize the probability, that packets are eavesdropped over a specific link of a randomization process for packet deliveries. In this process, the previous next hop for the source node is identified in the first step of the process. That process can random pick up a neighboring node and the current packet transmission with the next hop. The exclusion for the next hop is a selection that avoids transmitting two packets of consecutive data.

We combined the data partition schemes with dynamic replication to achieve data security, and access performance in data grids processing. The partition data need to be properly allocated to achieve the actual performance benefits in replication process. Our project design can provide these following advantages:

- Data can be secured
- Enable the sharing of coordinated data from various resources and provides various services with distributed and data intensive computing
- Replication techniques are frequently used to improve the data and reducing of client response times and communication
- Single point can't accept in this system

#### **4. Some ideas of Big Data Approach**

As it was mentioned in previous sections of the article in considered case we are dealing with both data processing, and complex computations. In the first case, we should consider the data as a whole, in the second case, data is exchanged between different branches. We need to combine these components. At the same time, in the process of modeling, for example, it must be done at least twice – in preprocessing and postprocessing. As a result, we get a huge amount of heterogeneous data that change the original state of knowledge. With this data and knowledge it is necessary to work differently under different conditions.

So within the framework of virtual testbed we that Big Data are different as it was proposed in [16]. And for every kind of situations we must to have different tool. In [16] new definition of “Big Data” is characterized by the situation when the conditions for implementing the CAP theorem are relevant. The CAP theorem [17] is a heuristic statement that in any realization of distributed computations, it is impossible to provide the following three properties: Consistency, Availability and Partition Tolerance. We define different kinds of Big Data as an appropriate combinations of C, A and P on different stages of computation. Problems and features of implementations of virtual testbed fully confirm this approach.

#### **5. Numerical example**

One of the fundamental virtual testbed's functions in considered case is to simulate ship motion under the impact of ocean waves, and mathematical formulae and numerical methods used for that purpose are especially favorable to implement in a programme that runs on a GPU. These implementations use

- linear memory access patterns which are needed to coalesce memory loads and stores (i.e. help vectorise the code for GPU),
- large number of floating point calculations including transcendental mathematical functions which are slow to execute on a CPU, and
- geometrical transformations (rotation and translation) which are built-in functions in a GPU, but not in a CPU.

It is possible to implement every numerical solver in Virtual testbed to run on a GPU to eliminate costly data transfer of large multidimensional arrays and vector fields between CPU and GPU memory altogether. This is one of the goals of the project, but it is not achieved yet: only the

most demanding numerical solvers are rewritten for GPU to speed-up the programme. One of these solvers is presented in the following paragraphs.

The feature that distinguishes Virtual testbed from similar ship motion simulation programmes is the use autoregressive moving average (ARMA) model for wavy surface generation. Unlike linear models which represent wavy surface as weighted sum of cosines (which in turn allows to generate velocity potential directly), this model represents each point of the surface as weighted sum of previous in time and space points. In order to compute velocity potential, we derived a formula that works for any discretely given wavy surface and gives the same velocity potential field for linear models as the traditional approach. In order to compute wave pressure force acting on a ship hull we decompose the hull into triangular panels and add all individual forces acting on each panel.

ARMA model was found to be slow on GPU, but very fast on a CPU due to non-linear memory access pattern, however, the new velocity potential formula allowed us to use fast Fourier transforms to implement it efficiently on a GPU. Wave pressure computation involves geometrical transformations, large number of floating point calculations which made it easy to rewrite for GPU. These optimisations resulted in tenfold speed-up for velocity potential and fivefold speed-up for wave pressure on a computer with AMD FX-8370 CPU and GeForce GTX 1060 6GB GPU.

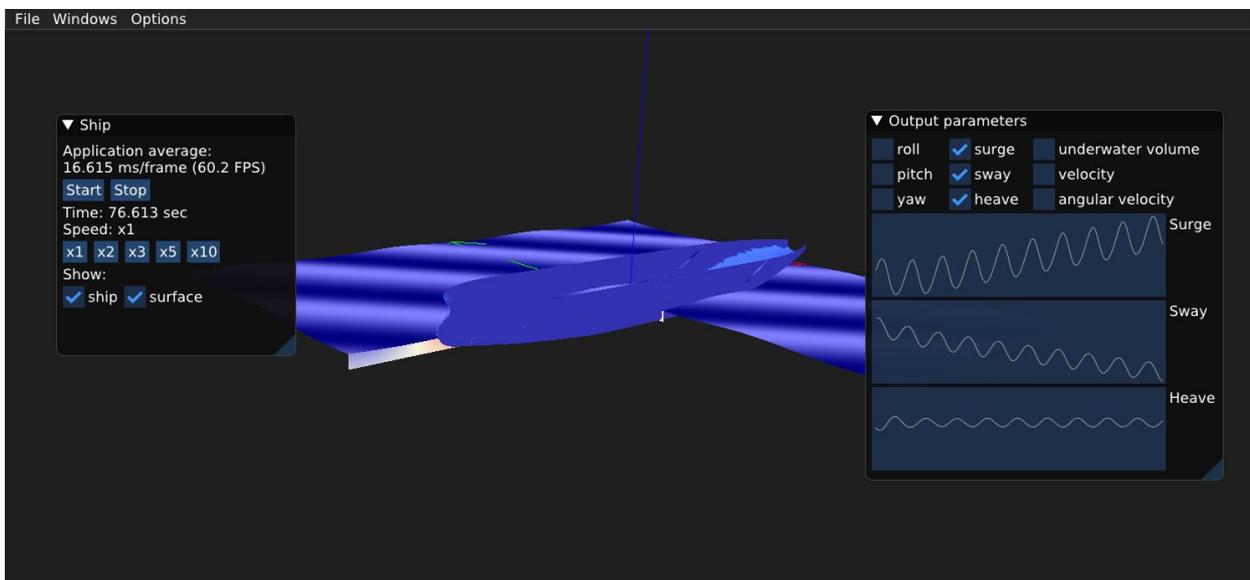


Figure 2. Example of naval virtual testbed realization

## 6. Acknowledgement

The work was supported by St.Petersburg State University (project id 26520170) and partly by Russian Foundation for Basic Research (RFBR), grants #17-29-04288, #16-07-00886.

## 7. Conclusion

It is clear that Big Data approach although is not the only one but is of a great help of large complex problems. It is an especially important to have a possibility of visualizing of behavior of technical objects on intermediate stages of calculations. That makes it possible to make a correction both to the model and parameters chosen without additional spending of computer time. We hope that very soon it will be possible to show the realistic technical objects in true natural environment.

## References

- [1] Van Dyke M., Perturbation methods in fluid mechanics. NY, Academic Press, 1964, 229p.
- [2] Nayfel Ali H., Perturbation Methods. WILEY-VCH, 2004, 425p.
- [3] Morse P.M., Feshbach H. Methods of Theoretical Physics. NY, McGRAW-HILL BOOK COMPANY, 1953.
- [4] Bogdanov, A., Degtyarev, A., Korkhov, V. New approach to the simulation of complex systems. // EPJ Web of Conferences, 2016, vol.108, 01002.
- [5] Bogdanov, A., Degtyarev, A., Korkhov, V. Desktop supercomputer: what can it do? // Physics of particles and nuclei letters, vol.14, is.7, 2017, pp.985-992.
- [6] Gankevich, I., Gaiduchok, V., Korkhov, V., Degtyarev, A., Bogdanov, A. Middleware for big data processing: test results. // Physics of particles and nuclei letters, vol.14, is.7, 2017, pp.1001-1007.
- [7] Bogdanov, A., Degtyarev, A., Korkhov, V., Gaiduchok, V., Gankevich, I. Virtual supercomputer as basis of scientific computing. In book: Horizons in Computer Science Research, vol.11, 2015, pp.159-198.
- [8] Bogdanov, A., Khramushin, V. Tensor Arithmetic, Geometric and Mathematic Principles of Fluid Mechanics in Implementation of Direct Computational Experiments. // EPJ Web of Conferences, 2016, vol.108, 02013
- [9] Degtyarev, A., Khramushin, V. Coordinate systems, numerical objects and algorithmic operations of computational experiment in fluid mechanics // EPJ Web of Conferences, 2016, vol.108, 02018.
- [10] Degtyarev, A., Gankevich, I. Simulation of Standing and Propagating Sea Waves with Three-Dimensional ARMA Model. In book: The Ocean in Motion: Circulation, Waves, Polar Oceanography. ed. by M.G.Velarde et al., Springer, 2018, pp.249-278
- [11] Degtyarev, A.B. New approach to wave weather scenarios modeling. In book: Fluid Mechanics and its Applications, vol.97, 2011, pp.599-617.
- [12] Boukhanovsky, A., Rozhkov, V., Degtyarev, A. Peculiarities of computer simulation and statistical representation of time-spatial metocean fields. // LNCS, vol.2073, 2001, pp.463-472.
- [13] Bogdanov, A.V., Degtyarev, A.B., Khramushin, V.N. High performance computations on hybrid systems: will "grand challenges" be solved? // Computer research and modeling, vol.7, is.3, 2015, pp.429-437 (in Russian).
- [14] Degtyarev, A. ; Gankevich, I. Hydrodynamic Pressure Computation under Real Sea Surface on Basis of Autoregressive Model of Irregular Waves. // Physics of particles and nuclei letters, vol.12, is.3, 2015, pp. 389-391.
- [15] Bogdanov A., Thurein Kyaw Lwin, Stankova E. Storage Database System in the Cloud Data Processing on the Base of Consolidation Technology. // LNCS, vol.9158, 2015, pp 311-320.
- [16] Bogdanov, A., et al. Big Data as the future of information technology. // Book of abstracts 8<sup>th</sup> Int. Conf. Distributed Computing and GRID-technologies in Science and Education, Dubna, 2018, p.25
- [17] Brewer, E.A. Towards robust distributed systems. // Proceedings of the XIX annual ACM symposium on Principles of distributed computing. — Portland, OR: ACM, 2000. — vol. 19, no. 7