Design Technology for Reconfigurable Computer Systems with Immersion Cooling

Ilya Levin, Alexey Dordopulo(✉), Alexander Fedorov, Yuriy Doronchenko

Scientific Research Centre of Supercomputers and Neurocomputers (LLC), Taganrog, Russia
{levin,dordopulo,doronchenko}@superevm.ru, ss24@mail.ru

Abstract. In this paper, we consider the implementation of reconfigurable computer systems based on advanced Xilinx UltraScale and UltraScale+ FPGAs and a design method of immersion cooling systems for computers containing 96–128 chips. We propose the selection criteria of key technical solutions for creation of high-performance computer systems with liquid cooling. The construction of the computational block prototype and the results of its experimental thermal testing are presented. The results demonstrate high energy efficiency of the proposed open cooling system and existence of power reserve for the next-generation FPGAs. Effective cooling of 96-128 FPGAs with the total thermal power of 9.6-12.8 kW in a 3U computational module is the key feature of the considered system. Insensitivity to leakages and their consequences, and compatibility with traditional water cooling systems based on industrial chillers are the advantages of the developed technical solution. These features allow installation of liquid-cooled computer systems with no fundamental change of the computer hall infrastructure.

Keywords: Immersion Cooling System · Liquid Cooling · Reconfigurable Computer Systems · FPGA · High-Performance Computer Systems · Energy Efficiency

1 Introduction

A reconfigurable computer system (RCS), which contains an FPGA computational field of large logic capacity[1-2], is used for implementation of computationally laborious tasks from various domains of science and technique[1-3], because it has a considerable advantage in its real performance and energetic efficiency in comparison with cluster-like multiprocessor computer systems. The RCS provides adaptation of its architecture to the structure of any solving task. In this case a special-purpose computer device is created. It hardwarily implements all computational operations of the information graph of the task with the minimum delays. Here, we have a contradiction between the implementation of the special-purpose device and its general-purpose usage for solving tasks from various problem areas. It is possible to eliminate these contradictions, combining creation of a special-purpose computer device with a wide range

adfa, p. 1, 2011.

of solving tasks, within a concept of reconfigurable computer systems based on FPGAs that are used as a principal computational resource [1].

Continuous increasing of the circuit complexity and the clock rate of each new FPGA family leads to considerable growth of power consumption and to growth of the maximal operating temperature on the chip. Practical experience of maintenance of large RCS-based computer complexes proves that air cooling systems have reached their heat limit.

According to the obtained experimental data, further development of FPGA production technologies and conversion to the next FPGA family Virtex UltraScale (power consumption up to 100 W for each chip) will lead to additional growth of FPGA overheat on 20…25°C. This will shift the range of their operating temperature limit (80…85°C), which means negative influence on their reliability when chips are filled up to 85-95% of available hardware resource. This circumstance requires a quite different cooling method which provides keeping of performance growth rates of advanced RCS.

2 Reconfigurable Computer Systems with Liquid Cooling

Today liquid cooling systems are the most promising design area for cooling modern high-loaded electronic components of computer systems[4-6], because of heat capacity of liquids which is better than air capacity (from 1500 to 4000 times), and higher heat-transfer coefficient (increasing up to 100 times).

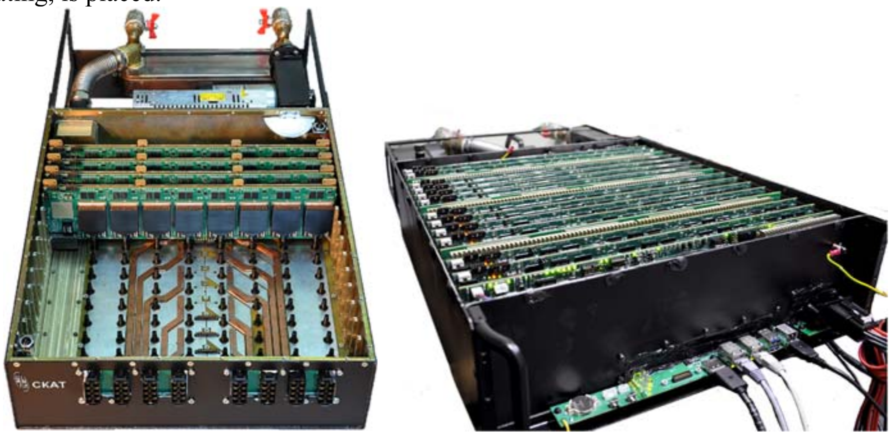
At present the technology of liquid cooling of servers and separate computational modules is developed by many vendors and some of them have achieved success in this direction [7-10]. However, these technologies are intended for cooling computational modules which contain one or two microprocessors. All attempts of its adaptation to cooling computational modules which contain a large number of heat generating components (an FPGA field of 8 chips), have proved a number of shortcomings of liquid cooling of RCS computational modules [5-6].

The special feature of the RCS produced in Scientific Research Centre of Supercomputers and Neurocomputers is the number of FPGAs, not less than 6-8 chips on one printed circuit board, and high packing density. Usage of an open liquid cooling system is efficient owing to the heat-transfer agent characteristics, and the design and specification of the used FPGA heatsinks, pump equipment, heat-exchangers. The heat-transfer agent must have the best possible electric strength, high heat transfer capacity, the maximum possible heat capacity and low viscosity. The heatsink must provide the maximum possible surface of heat dissipation, must allow circulation of the heat-transfer agent through itself, a turbulent heat-transfer agent flow in itself, manufacturability.

Since 2013 the scientific team of SRC SC and NC has actively developed the domain of creation of next-generation RCS on the base of their original liquid cooling system for computational circuit boards with high packing density and the large number of heat generating electronic components. Design criteria of the computational module (CM) of next-generation RCS with an open loop liquid cooling system are based on the principles as follows:

- the RCS configuration is based on a computational module with the 3U height and the 19” width and with self-contained circulation of the heat-transfer agent;
- one computational module can contain 12-16 computational circuit boards (CCB) with FPGA chips;
- each CCB must contain up to 8 FPGAs with dissipating heat flow of about 100 W from each FPGA;
- a standard water cooling system, based on industrial chillers, must be used for cooling the heat-transfer agent.

The principal element of modular implementation of an open loop immersion liquid cooling system for electronic components of computer systems is a reconfigurable computational module with liquid cooling (see the design in Fig. 1). The CM casing consists of a computational section and a heat exchange section. In the casing, which is the base of the computational section, a hermetic container with heat-transfer agent (dielectric cooling liquid) and electronic components with elements that generate heat during operating, is placed.



a)
b)

Fig. 1. Design of reconfigurable computer system with liquid cooling
(a – assembly of CM, b – assembled CM without top cover)

The computational section contains: boards of the computational module (not less than 12-16 (see Fig. 2)), control boards (see Fig. 3), RAM, power supply blocks (see Fig. 4), storage devices, daughter boards, etc. The computational section is closed with a cover. The CCB of the advanced computational module contains 8 Kintex UltraScale XCKU095T FPGAs. Each FPGA has a specially designed thermal interface and a small height heatsink for heat dissipation. To cool FPGAs we use heatsinks of an original design (see Fig. 2-b).

a)

b)

Fig. 2. Boards of computational module of reconfigurable computer system with liquid cooling (a – without heatsinks, b – with designed heatsinks)

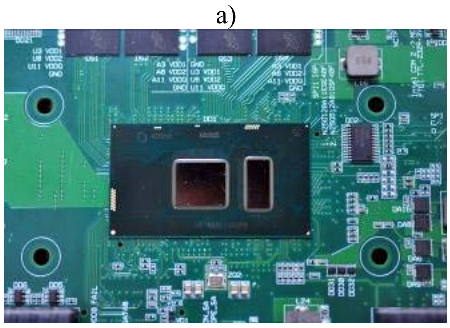
The specialists of SRC SC & NC have performed heat engineering research and suggested a fundamentally new design of a heatsink with original solder pins, which create a local turbulent flow of the heat-transfer agent. The used thermal interface cannot be deteriorated or washed out by the heat-transfer agent. Its coefficient of heat conductivity must remain permanently high. The specialists of SRC SC & NC have created an effective thermal interface which fulfills all specified requirements. Besides, the technology of its coating and removal was also perfected.

The computational section of the CM contains 12 CCBs with the power up to 800 W each, 3 power supply units. Besides, all boards are completely immersed into an electrically neutral liquid heat-transfer agent.

For creation of an effective immersion cooling system a dielectric heat-transfer agent was developed. This heat-transfer agent has the best possible electric strength, high heat transfer capacity, the maximum possible heat capacity and low viscosity.

The scientific team of SRC SC & NC has designed and produced an original motherboard based on an Intel Skylake® (Core I5-6300U) processor for the computational block with liquid cooling. The printed circuit board consists of 18 layers and has the minimum size of 490x109.7 mm for placement into the computational module.

а)



б)

Fig. 3. Board of loading and control block
(a – without heatsinks, b – with placed heatsinks)

For the motherboard we have designed an original basic input-output system (**BIOS**), which allows usage of all capabilities of the Intel Skylake® (Core I5-6300U) system-on-chip and external peripheral equipment. Besides, we have designed an immersion power supply block (see Fig. 4), which provides DC/DC 380/12 V transformation with power up to 4 kW for 4 CCBs. Power supply blocks for 4 CCBs are placed into the computational section.

The computational section adjoins to the heat exchange section, which contains a pump and a heat exchanger. The pump provides closed loop circulation of the heat transfer agent in the CM: from the computational module the heated heat-transfer agent passes into the heat exchanger and is cooled there. From the heat exchanger the cooled heat-transfer agent again passes into the computational module and there cools the heated electronic components. As a result of heat dissipation the agent becomes heated and again passes into the heat exchanger, and so on. The heat exchanger is connected to the external heat exchange loop via fittings and is intended for cooling the heat-transfer agent with the help of the secondary cooling liquid. As a heat exchanger it is possible to use a plate heat exchanger in which the first and the second loops are separated. So, as the secondary cooling liquid it is possible to use water, cooled by an industrial chiller. The chiller can be placed outside the server room and can be connected with the reconfigurable computational modules by means of a stationary system of engineering services.

a)

b)

Fig. 4. Power supply blocks for computational module circuit boards

The computational and the heat exchange sections are mechanically interconnected into a single reconfigurable computational module. Maintenance of the reconfigurable computational module requires its connection to the source of the secondary cooling liquid (by means of valves), to the power supply or to the hub (by means of electrical connectors).

In the casing of the computer rack the CMs are placed one over another. Their number is limited by the dimensions of the rack, by technical capabilities of the computer room and by the engineering services.

Each CM of the computer rack is connected to the source of the secondary cooling liquid with the help of supply return collectors through fittings (or balanced valves) and flexible pipes; connection to the power supply and the hub is performed via electric connectors.

Supply of cold secondary cooling liquid and extraction of the heated one into the stationary system of engineering services connected to the rack, is performed via fittings (or balanced valves).

The performance of one next-generation CM is increased in 8.7 times in comparison with the CM “Taygeta”. Such qualitative increasing of the system specific performance is provided by more than triple increasing of the system packing density owing to original design solutions, and increasing of the clock frequency and the FPGA logic capacity. Experimental results prove that the complex of the developed solutions concerning the immersion liquid cooling system provide the temperature of the heat-transfer agent not more than 33°C, the power of 91 W for each FPGA (8736 W for the CM) in the operating mode of the CM. At the same time, the maximum FPGA temperature during heat experiments does not exceed 57°C. This proves that the designed immersion liquid cooling system has a reserve and can provide effective cooling for the designed RCS based on the advanced Xilinx UltraScale+ FPGA family.

On basis of the designed solutions of the advanced computational module we have created a prototype of a computer system Nekkar (see Fig. 5). The RCS Nekkar contains 12 3U computational modules with liquid cooling.

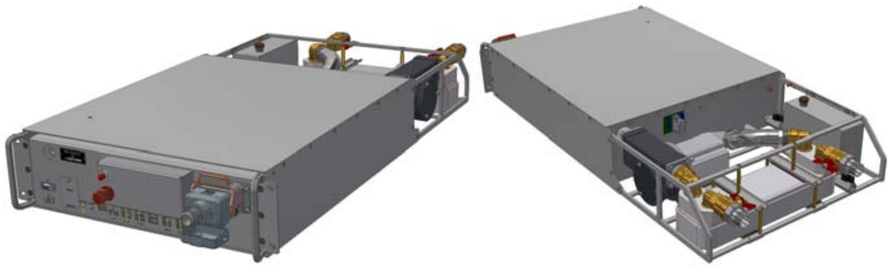
a)

Fig. 5. RCS Nekkar based on advanced computational modules
(a – without heatsinks, b – with placed heatsinks)

Each computational module contains 12 boards with the power consumption of 800 W each. Each board contains 8 Xilinx Kintex UltraScale XCKU095T FPGAs (100 million equivalent gates in each chip). The total performance of the RCS Nekkar is 1 PFlops, and its power consumption is 124 kW.

The improved design of the computational modules for serial production is shown in Fig. 6.

a)



b)

Fig. 6. Serial computational module Nekkar
(a – front panel, b – top and side view of front and back panels)

3 Advanced Reconfigurable Computer System ”Arctur” Based on Xilinx UltraScale+ FPGAs

Usage of the UltraScale+ FPGAs, which have been implemented on the base of the 16-nm technology 16FinFET Plus and produced by Xilinx since 2017, will provide up to triple ramp of the computational performance owing to increasing of clock frequency and FPGA circuit complexity; the size of the computer system remains unchanged. However, in spite of reduction of relative energetic consumption owing to new technological standards of FPGAs manufacturing, and owing to a certain power reserve of the designed liquid cooling system, it is possible to expect a new approach of FPGA operating temperatures to their critical values.

Besides, the new FPGAs of the UltraScale+ family have larger geometric sizes. The size of the FPGAs of the RCS “Nekkar” is 42.5x42.5 mm. The size of the FPGAs, which are going to be placed into the RCS “Arctur”, is 45x45 mm. Due to this circumstance it is impossible to use the existing design of the CCB, because the width of the printed circuit board will become larger and therefore will not fit for the standard 19’’ rack.

In this connection it is necessary to modify the designed open liquid cooling system and the CCB design that will lead to modification of the whole CM. During modification of the CCB design we have created a prototype of an advanced board shown in Fig. 7. The CCB contains 8 UltraScale+ FPGAs of high circuit complexity. To provide placement of a new CCB into a 19” rack possible, it is necessary to exclude its CCB controller from its structure. The CCB controller was always implemented as a separate FPGA and provided access to FPGA computational resources of the CCB, FPGA programming, condition monitoring of the CCB resources. One of FPGAs of the computation field will perform all functions of the controller.

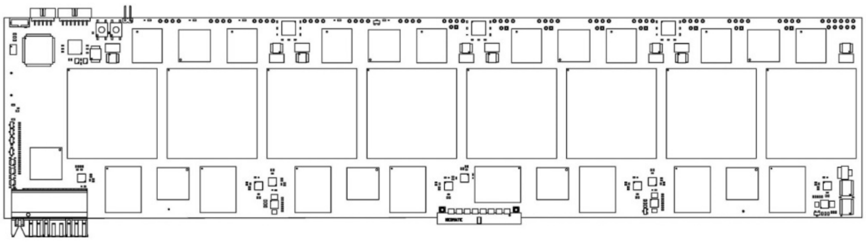


Fig. 7. Prototype of upgraded CCB assembly for UltraScale+ FPGAs

Within preliminary design projects, which deal with creation of an advanced RCS based on Xilinx UltraScale+ FPGAs and with improvement of the cooling system, we will solve the problems as follows:

1. Increase of effective surface of heat-exchange between FPGAs and the heat-transfer agent.
2. Increase of the performance of the heat-transfer agent supply pump.
3. Increase of reliability of the liquid cooling system with the help of immersed pumps.
4. Experimental improvement of the heatsink optimal design.
5. Experimental improvement of the technology of thermal interface coating.

We have designed a prototype of an advanced computational module with a modified immersed cooling system (Fig. 3). The distinctive feature of the new design is immersed pumps and the considerable reliability growth of the CM owing to reduction of the number of components and simplification of the cooling system. According to our plans, the heat exchange section will contain only the heat exchanger. We are working on experimental research of various pump equipment which can operate in the heat-exchange agent.

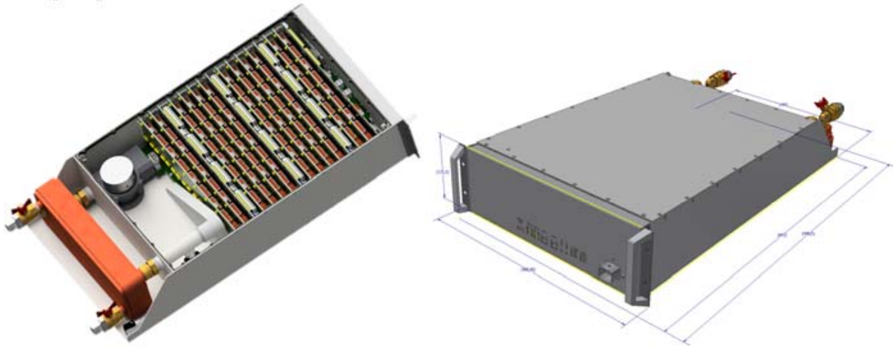


Fig. 8. A prototype of a computational module with a modified immersed cooling system

So, owing to breakthrough technical solutions which we have got during design of the RCS "Nekkar" with the immersed liquid cooling system, we can develop this direction of high-performance RCS design, and after some design improvements we can create a computer system which provides a new level of computational performance.

4 Conclusion

Usage of air cooling systems for the designed supercomputers has practically reached its limit because of reduction of cooling effectiveness with growing of consumed and dissipated power, caused by growth of circuit complexity of microprocessors and other chips. That is why usage of liquid cooling in modern computer systems is a priority direction of cooling systems perfection with wide perspectives of further development. Liquid cooling of RCS computational modules which contain not less than 8 FPGAs of high circuit complexity is specific in comparison with cooling of microprocessors and requires development of a specialized immersion cooling system. The designed original liquid cooling system for a new generation RCS computational module provides high maintenance characteristics such as the maximum FPGA temperature not more than 57°C and the temperature of the heat-transfer agent not more than 33°C in the operating mode. Owing to the obtained breakthrough solutions of the immersion liquid cooling system it is possible to place not less than 12 CMs of the new generation with the total performance over 1 PFlops within one 47U computer rack. Power reserve of the liquid cooling system of the new generation CMs provides effective cooling of not only existing but of the developed promising FPGA families Xilinx UltraScale+ and UltraScale 2.

Since FPGAs, as principal components of reconfigurable supercomputers, provide stable ramping of RCS performance, it is possible to get specific performance of RCS, based on Xilinx Virtex UltraScale FPGAs, similar to the one of the world best cluster supercomputers, and to find new perspectives of design of super-high performance supercomputers.

References

1. Tripiccione, Raffaele: Reconfigurable Computing for Statistical Physics. The Weird Case of JANUS// IEEE 23RD international conference on application-specific systems, architectures and processors (ASAP) (2012).
2. BAITY-JESI, M. et al.: The Janus Project: Boosting Spin-Glass Simulations using FPGAs. IFAC Proceedings Volumes, Programmable Devices and Embedded Systems, vol.12, №1 (2013).
3. Shaw, D.E. et al. Anton, a special-purpose machine for molecular dynamics simulation. Commun. ACM 51, 7, pp. 91–97 (2008).
4. Kalyaev, Igor A., Levin, Ilya I., Semernikov, Evgeny A., Shmoilov, Vladimir I.: Reconfigurable multipipeline computing structures. Nova Science Publishers, New York (2012). ISBN: 978-1-62081-462-8
5. Dordopulo, Alexey, Kalyaev, Igor, Levin, Ilya, Slasten, Liubov: High-performance reconfigurable computer systems. In: Lecture Notes in Computer Science, vol. 6873, **Chapter** Parallel Computing Technologies. pp. 272–283.
6. Dordopulo A.I., Levin I.I., Fedorov A.M., Kalyaev I.A. Reconfigurable computer systems: from the first FPGAs towards liquid cooling systems // Supercomputing Frontiers and Innovations. 2016. – C. 22-40. http://superfri.org/superfri/article/view/97 DOI: 10.14529/jsfi160102.
7. Technology https://www.coolitsystems.com/technology/, last accessed 2018/05/25.

8. Immers 6 R6 http://immers.ru/sys/immers6r6/, last accessed 2018/05/25.
9. Eurotech liquid cooling is hot! https://www.eurotech.com/en/hpc/hpc+solutions/liquid+cooling, last accessed 2018/05/25.
10. RSC http://www.rscgroup.ru, last accessed 2018/05/25.