DOI: 10.14529/jsfi160102

Reconfigurable computer systems: from the first FPGAs towards liquid cooling systems

Ilya I. Levin[1], *Alexey I. Dordopulo*[1], *Alexander M. Fedorov*[1], *Igor A. Kalyaev*[2]

The paper covers the history of development of design technologies of reconfigurable computer systems based on FPGAs of various families. Five generations of reconfigurable computer systems with high placement density, designed on the base of various FPGA families, from Xilinx Virtex-E to modern Virtex UltraScale, are described. The last achievements in the domain of design of energetic effective reconfigurable computer systems with high real performance are presented. One of such achievements is the developed liquid cooling system for Virtex UltraScale FPGAs. It provides independent circulation of the cooling liquid in the 3U computational module with the 19 height for cooling of 96-128 FPGA chips that in total generate 9.6-12.8 kWatt of heat. The distinctive features of the designed immersion liquid cooling system are high cooling efficiency with power reserve for the designed perspective FPGA families, resistance to leaks and their consequences, and compatibility with traditional water cooling systems based on industrial chillers.

Keywords: FPGA, reconfigurable computer systems, immersion liquid cooling system.

Introduction

One of perspective approaches to achieve high real performance of a computer system is the adaptation of its architecture to the structure of a solving task, and creation of a special-purpose computer device which hardwarily implements all computational operation of the information graph of the task with the minimum delays. A natural requirement for a modern computer system is hardware support of modification of both the algorithm of the solving task and the task itself, that is why FPGAs are used as a principal computational resource of reconfigurable computer systems [1].

The main advantages of programmable logic devices (PLD) are:

- possibility of implementation of complicated parallel algorithms;
- availability of CAD-tools for complete system simulation;
- possibility of programming or modification of in-system configuration;
- compatibility of various design projects when they are converted in a VHDL, AHDL, Verilog descriptions or in any other hardware description language.

1. The history of development of FPGA and FPGA-based reconfigurable computer systems

The history of the PLD architectures started at the end of the 70s, when the first PLDs with programmable-AND and programmable-OR arrays appeared. Such architectures were called FPLAs (Field Programmable Logic Array) and FPLSs (Field Programmable Logic Sequencers)[2]. Their main disadvantage is weak use of programmable-OR array.

At the end of the 80s PLD developers suggested new architectures that were simpler and cheaper: PALs (Programmable Array Logic) and GALs (Gate Array Logic). Such architectures

[1]Scientific Research Centre of Supercomputers and Neurocomputers, Taganrog, Russia
[2]A.V. Kalyaev Scientific Research Institute of Multiprocessor Computer Systems at Southern Federal University, Taganrog, Russia

were used in the PLDs of Intel, Altera, AMD, Lattice, etc. The PLDs had rather low integration density, a programmable-AND array and a fixed-OR array [3]. Another approach to reduction of programmable-OR array redundancy was a programmable macro-logic. Macro-logic-based chips contained either a programmable-NAND array or a programmable-NOR array, but it was possible to form complicated logic functions owing to numerous feedbacks.

At the beginning of the 80s there were three leading PLD vendors on the programmable logic device world market. Altera Corporation [4] was founded in June 1983, Xilinx Inc. [5] in February 1984, Actel Corporation [6] in 1985. At present these three companies hold about 80% of PLD world market and determine the ideology of PLD applications. In the past the PLD chips were only one product among various products of such enterprises as Intel, AMD, etc. But in the middle of the 80s leading positions on the PLD market were taken by enterprises specializing solely in design and production of the PLDs.

New vendors offered new PLD architectures, such as CPLDs (Complex Programmable Logic Devices) [7]. The CPLD has rather high integration density and contains several functional blocks connected by a switch matrix. Each functional block contains a programmable-AND array and a fixed-OR array. This PLD class includes Altera MAX7000, Xilinx XC9500 and various PLDs of other vendors, such as Atmel, Vantis, Lucent, etc.

Introduction of FPGAs (Field Programmable Gate Array) [8] ignited revolution of devices with programmable logic. The FPGA class includes Xilinx XC2000, XC3000, XC4000 and Spartan, Actel ACT1, ACT2, Altera FLEX8000 family and some Atmel and Vantis PLDs.

The FPGA configurable logic blocks (CLBs) are connected by a programmable switch matrix. The logic blocks consist of one or several rather simple logic cells based on a 4-input look-up table (LUT), a program-controlled multiplexer, a D-flip-flop. Input/output blocks (IOB) that provide bidirectional input/output, tri-state, etc., are typical for the FPGA-architectures. The FPGA chips have the following advanced features:

– a JTAG port that supports all mandatory boundary-scan instructions specified according to the IEEE 1149.1 standard;

– a master configuration mode (that required a build-in oscillator).

The FPGAs with a dedicated block RAM were the result of further development of the FPGA architecture, owing to which the FPGAs can be used with no external memory devices. The FPGAs have a high logic capacity, an easy-to-use architecture, a quite high reliability and an optimal ratio price/logic capacity, therefore they match various requirements, claimed by circuit engineers.

During the last years, the FPGAs along with custom VLSIs have become the basis of systems-on-chip. IP-cores of such systems have been designed separately and can be used in various projects. The final structure of an FPGA-based SoC is implemented by means of CAD tools.

The ideology of SoC design spurred the leading FPGA vendors on, and at the end of 1998 at the beginning of 1999 they introduced products with an equivalent logic capacity about a million equivalent gates and more. The Altera ApEX20K family is an example of new FPGA families for SoC design.

The Xilinx Virtex [9] family has a similar architecture with a great variety of high speed routing resources, a dedicated block RAM, an optimal high-speed carry logic. The Virtex family provides a high speed of data exchange between the chips up to 200 MHz (HSTL IV standard). The chips of the Virtex family have a relatively low price (not more than 40% from the equiv-

alent price of XC4000XL series), owing to the advanced production technique and perfected verification methods.

In 1998-1999 the augmentation of FPGA equivalent logic capacity changed the attitude to CAD-tools of both software developers and users. Till the end of the 90s the main tool of project description and schematic entry was a graphic editor and libraries of standard primitives and macros, such as logic elements, elementary combinational and sequential functional units, analogues of standard integrated circuits of small-scale and medium-scale integration. At present, circuit engineers use widely the hardware description languages for FPGA-based implementation of algorithms. Besides, up-to-date CAD-tools support both standard hardware description languages (such as VHDL, Verilog) and specialized hardware description languages developed by FPGA vendors specially for their own needs, CAD-tools and FPGA families with special architecture features. Such example is AHDL (Altera Hardware Description Language), which is supported by the Altera CAD-tools MAX PLUS II and Quartus.

Xilinx corporation designs IP-cores of frequently used functional blocks, including blocks of digital processing, bus interfaces, processors, etc. Owing to design of the IP-cores, based on the use of the Xilinx LogiCORE^{TM} development package and on the use of some software analogues developed by second-party vendors, it is possible to reduce the time of project design, minimize risks, and achieve the highest performance. In addition, with the help of the CORE Generator^{TM} the circuit engineer can design his own IP-cores with predictable and repeatable timing characteristics. The CORE Generator^{TM} has a quite simple user interface for generation of parameterized IP-cores specially perfected for use in the Xilinx FPGAs.

2. Research and design of FPGA-based computational systems

At the same time appearance of high-performance FPGA chips on the market gives wide perspectives for their use as principal components of high-performance computer systems (supercomputers).

At present, there are two kinds of high-performance computer systems which use FPGAs as principal components. The first kind is so called hybrid computer systems, i.e. classic cluster computers which contain FPGAs in their microprocessor nodes and use them as accelerators of calculations. The examples of such hybrid supercomputers are XT4 by Cray and RASC by SiliconGraphics. In these systems blocks of programmable co-processors are implemented in FPGAs, interconnected with themselves and the principal processors by high-speed busses.

The second kind of computer systems, which use FPGAs as principal components, is reconfigurable computer systems (RCS). In RCSs, FPGAs are principal computational components, whereas general-purpose processors are minor components, which control the operation of the reconfigurable part of the system.

The acknowledged leader in the domain of design of reconfigurable computer systems based on FPGA computational fields is Taganrog scientific school, founded by academician A.V. Kalyaev. Today it is represented by various RCSs of the supercomputer class designed in Scientific Research Institute of multiprocessor computer systems at Taganrog State University of Radio-Engineering (now at Southern Federal University) and Scientific Research Centre of Supercomputers and Neurocomputers. The principal computational resource of such systems is not microprocessors, but a set of FPGA chips united into computational fields by high-speed data transfer channels. The spectrum of produced and designed products is rather wide: from completely stand-alone small-size reconfigurable accelerators (computational blocks), computa-

tional modules of desktop and rack design (based on Xilinx Virtex-6, Virtex-7, and UltraScale FPGAs) to computer systems which consist of several computer racks placed in a specially equipped computer room.

Since 2001 four generations of FPGA-based reconfigurable computer systems have changed one another owing to the production of new FPGA families and growth of computational complexity of problems that require continuous increasing of RCS performance.

RCSs with macroprocessor architecture (RCS MPA) were the first generation of RCSs. They consisted of a number of basic modules implemented on FPGAs and a personal computer. Each basic module (BM) is a reconfigurable computational device designed according to the same architectural principles that as the whole system. Such approach provided natural implementation of structural procedural parallel programs for different granularity of parallelism and piping of calculations. First single-board RCS MPA basic module was designed and created in 2001. Its principal components were Xilinx Virtex- FPGAs. The board of the BM was produced according to 12-layer technology with double-side mounting of components. There are six signal layers and six layers of potentials: two ground layers, two layers for the power of 1.8 V, and two layers for the power 3.3 V. Fig. 1 shows the front and the back sides of the board of the RCS MPA basic module.

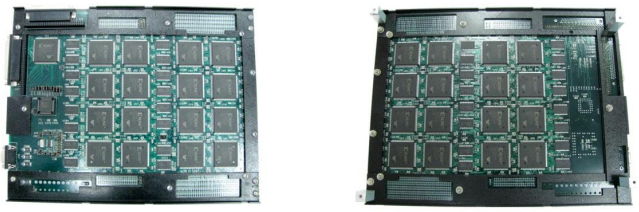

Figure 1. Front and back sides of the board of the RCS MPA basic module

The BM printed circuit board contained 32 FPGAs and 32 RAM chips, mounted on its both sides. Placement of the components on the BM printed circuit board provided the minimum length of connections between them. The performance of the RCS MPA basic module was $2.5 \text{x} 10^{10}$ op/sec, 64 processing elements, the power consumption was 30 Watt. On the basis of the BM printed circuit board of the RCS MPA a number of modular-scalable multiprocessor computer systems were designed and created [1].

RCS MPAs gave way to reconfigurable computer systems of the second generation, such as RCS with macroobject architecture. Using the RCS with macroobject architecture, the developer has two levels of its architecture programming [10].

On the first level (the lower one) the developer creates functional objects, which are necessary for implementation of a certain problem and which are physically placed in the FPGAs of the computational field. As a rule, the place and the function of the object remain constant not only for implementation of a certain task, but also for implementation of tasks of a certain class. In other words, on the first level the RCS is adjusted to some special-purpose architecture.

On the second level, with the help of the communication system the functional objects are united into computing structures, similar to the information graph of the solving task. Owing to such programming of RCS architecture, it is possible to increase the efficiency of the computational process in ten times in comparison with RCS MPA.

The first representative of RCS of the second generation with macroobject architecture was a modular-scalable RCS Bear (see its basic module in Fig. 2).

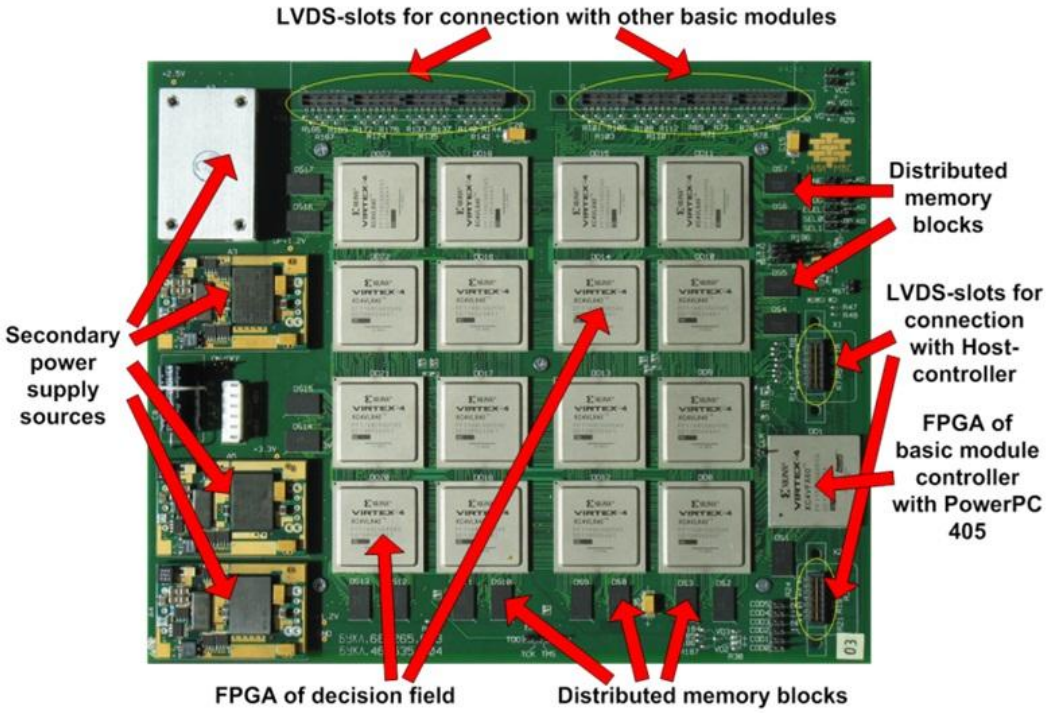

Figure 2. A general view of the basic module of the RCS Bear

The base of the basic module is a 20-layer printed circuit board with a double-side mounting of components, a computational field of 16 Xilinx XC4VLX40-10FF1148 FPGAs, a basic module controller, 17 chips of dynamic memory SDRAM, a programmable clock generator, four small-size DC-DC voltage transducers, LVDS-connectors for connection of basic modules via data channels, and other elements of surface mounting. The performance of the basic module of the RCS Bear was 50 GFlops, the frequency of the basic module was 160 MHz, and power consumption did not exceed 150 Watt.

The principles of macroobject architecture were used during the design of a small-size reconfigurable accelerator of the personal computer, intended for implementation of computationally laborious fragments of tasks of various problem domains. Fig. 3 shows the small-size reconfigurable accelerator of the personal computer.

The base of the accelerator was a basic module 4V4-25, implemented on a 18-layer printed circuit board of 150190 mm with double-side mounting of components. The performance of the basic module 4V4-25 was 25 GFlops, the frequency of the basic module was 160 MHz, and its power consumption did not exceed 145 Watt.

Owing to design experience of implementation of the RCS MPA and the RCS Bear, it became possible to implement RCSs of the third generation. The first representatives are RCSs

Figure 3. The small-size reconfigurable accelerator of the personal computer

of a family Ursa Major, which were designed on the basis of three types of basic modules, such as 16V5-75, 16V5-50, and 16S3-25. The basic module 16V5-75 (the most high-performance one) was used in such computer systems of the family as RCS-5, RCS-1R, and RCS-0.2-WS. The basic modules 16V5-50 and 16S3-25 were the components of such accelerators of the personal computer as RASC-50 and RASC-25.

The computational field of the basic module 16V5-75 consists of 16 Xilinx Virtex-5 XC5VLX110-2FF1153 FPGAs, which contain 11 million equivalent gates. The FPGAs are placed in the nodes of a 44 2D-lattice and are interconnected by a lattice-like communication system. Such communication system simplifies considerably the printed circuit board and improves its frequency characteristics because connections between the adjacent chips do not exceed 4 centimeters. Between remote FPGAs the data are transferred via transit channels through transit chips according to the lattice-like connections. The FPGAs placed on the borders of the 44 2D-lattice of the computation field are connected with 20 SDRAM DDR2 chips, which form distributed memory with the total volume of 1.25 GByte. Fig. 4 shows the printed circuit board of the basic module 16V5-75 with mounted electronic elements and the assembled basic module with its cooling system.

For heat dissipation and supporting all required temperature modes of the chips of the basic module, a combined cooling system is designed, which contains radiators, placed on FPGAs of the computation field, and air-fans for blowing them round. In general, the basic module 16V5-75 is a high-performance computational node with the performance above 75 (140) GFlops.

The basic module 16V5-75 is the base of the workstation RCS-0.2-WS and the computational block RCS-0.2-CB with the performance 300 GFlops.

Figure 4. The printed circuit board of the basic module 16V5-75

Fig. 5 shows the workstation RCS-0.2-WS without its top cover and the computational block RCS-0.2-CB.

Figure 5. The workstation RCS-0.2-WS and the computational block RCS-0.2-CB

The computational blocks RCS-0.2-CB are the base of the reconfigurable computer system RCS-1R with the peak performance of 1200 GFlops. This computer system is intended for scientific centres, which deal with research in such domains as physics, chemistry, biology, space, design of information-control systems for control of potentially dangerous industry, air-space industry, automobile industry, energetics, etc.

The high-end representative of the RCSs of the third generation was the system RCS-5 (see Fig. 6) with the peak performance of 6000 GFlops, which consisted of five racks RCS-1R, interconnected by Ethernet-switchers with general control.

The computer system RCS-5, placed in Research Computing Centre at Lomonosov Moscow State University, contains 20 computational blocks RCS-0.2-CB, 80 basic modules 16V5-75, 16384 processing elements (IEEE-754), which perform processing of 64-digit IEEE-754 data at the frequency of 330 MHz with the performance of more than 6000 GFlops. Between the blocks

a) b)

Figure 6. The RCS-5) the front view; b) the back view

the data exchange is performed via external LVDS and Gigabit Ethernet interfaces at frequency of 640 MHz.

Due to growing demands to RCS performance the density of placement of RCS components increased in 4 times. In 2012-2014, on the basis of Xilinx Virtex-7 FPGAs and computational modules (CM) 24V7-750 (Pleiad) and Taygeta [11], the fourth generation of RCS was designed.

The computational module 24V7-750 contains 4 printed circuit boards 6V7-180 (see Fig. 7); a control unit CU-7; a power supply subsystem; a cooling subsystem and other subsystems. The performance of the CM 24V7-750 is 2.58 TFlops for processing of 32-digit floating point data.

The CM 24V7-750 was used for the creation of a reconfigurable computer system RCS-7 (the state contract 14.527.12.0004 from 03.10.2011). The system RCS-7 contains 24 CMs 24V7-750 with a computation field of 576 Xilinx Virtex-7 XC7V585T-FFG1761 FPGAs (58 million equivalent gates each), interconnected and placed in one 47U computer rack with the peak performance of 1015 fixed point operations per second. The performance of the RCS-7, which contains from 24 to 36 24V7-750 CMs is from 62 to 93 TFlops for processing of 32-digit floating point data, and is from 19.4 to 29.4 TFlops for processing of 64-digit floating point data. The application area of the RCS-7 and computer complexes, created on its basis, is digital signal processing and multichannel digital filtration.

The 2U computational module Taygeta (see Fig. 8a) was also designed on the basis of Virtex-7 FPGAs and can be placed in a standard 19" computer rack. The CM Taygeta contains 4 boards of the CM 8V7-200 (see Fig. 8b), connected by high-speed LVDS-channels, an embedded host-computer, a power supply system, a control system, a cooling system and other subsystems.

The computational module Taygeta is based on the board of the CM 8V7-200, which is a 20-layer printed circuit board with double-side mounting of components. The board contains 8 Xilinx XC7VX485T-1FFG1761 FPGAs (48.5 million equivalent gates), 16 chips of distributed memory SDRAM DDR2 with its total volume of 2 GByte, LVDS and Ethernet interfaces, and other components. The performance of the CM Taygeta is 2.66 TFlops for processing of 32-digit floating point data. The performance of an RCS, which is based on the CM Taygeta and which contains 18 CMs Taygeta, is 48 TFlops for processing of 32-digit floating point data and 23 TFlops for processing of 64-digit floating point data.

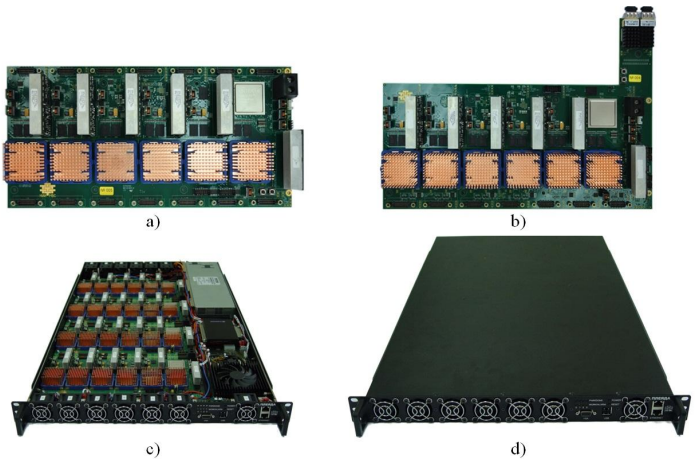

Figure 7. The computational module 24V7-750 (a – the boards 0-2, b – the board 3 with optical connector for connection with other CMs 24V7-750, c – the CM 24V7-750 without its top cover, d – with its top cover)

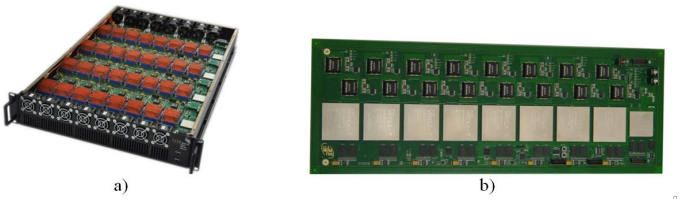

Figure 8. The CM Taygeta (a – without its top cover, b – the board 8V7-200)

Due to the number of problems with the cooling system, the fourth generation of RCS was designed. According to the obtained experimental data, conversion from the FPGA family Virtex-6 to the next family Virtex-7 leads to growth of the FPGA maximum temperature at 1115^o C. Therefore, the further development of FPGA production technologies and conversion to the next FPGA family Virtex Ultra Scale will lead to the growth of FPGA overheat at additional 1015^o C. This will shift the range of their operating temperature to 8085^o C, which means that their operating temperature exceeds the permissible range of the FPGA operating temperature ($65...70^o$ C), and hence, this will have negative influence on their reliability.

Practical experience of maintenance of large computer complexes based on CM RCS proves that air cooling systems have reached their heat limit. That is why use of air cooling for the next generation of FPGAs - Virtex UltraScale, which contain about 100 million equivalent gates and have power consumption not less than 100 Watt per FPGA chip, will not provide stable and reliable operating of the RCS when its chips are filled up to 85-95% of available hardware

resource. This circumstance requires a quite different cooling method which provides keeping of growth rates of the RCS performance for promising designed Xilinx FPGA families: Virtex UltraScale, Virtex UltraScale+, Virtex UltraScale 2, etc.

3. Liquid cooling for reconfigurable computer systems

Development of computer technologies leads to the design of computer technique which provides higher performance, and hence, more heat. Dissipation of released heat is provided by a system of electronic element cooling, that transfers heat from the more heated object (the cooled object) to the less heated one (the cooling system). If the cooled object is constantly heated, then the temperature of the cooling system grows and some time will be equal to the temperature of the cooled object. So, heat transfer stops and the cooled object will be overheated. The cooling system is protected from overheat with the help of cooling medium (heat-transfer agent). Cooling efficiency of the heat-transfer agent is characterized by heat capacity and heat dissipation. As a rule, the heat transfer is based on the principles of heat conduction, that require a physical contact of the heat-transfer agent with the cooled object, or on principles of convective heat exchange with the heat-transfer agent, that consists of physical transfer of the freely circulating heat-transfer agent.

To organize the heat transfer to the heat-transfer agent, it is necessary to provide the heat contact between the cooling system and the heat-transfer agent. For this various radiators facilities for heat dissipation in the heat-transfer agent are used. Radiators are set on the most heated components of computer systems. To increase the efficiency of heat transfer from an electronic component to a radiator, a heat interface is set between them. The heat interface is a layer of heat-conducing medium (usually multicomponent) between the cooled surface and the heat dissipating facility, used for reduction of heat resistance between two contacting surfaces. Modern processors and FPGAs need cooling facilities with as low as possible heat resistance, because at present even the most advanced radiators and heat interfaces cannot provide necessary cooling, if an air cooling system is used.

Till 2012 the air cooling systems were used quite successfully for cooling supercomputers. But due to the growth of performance and circuit complexity of microprocessors and FGAs used as components of supercomputer systems the air cooling systems have practically reached their limits for designed perspective supercomputers, including hybrid computer systems. Therefore, the majority of vendors of computer technique consider liquid cooling systems as an alternative decision of the cooling problem. Today liquid cooling systems are the most promising design area for cooling modern high-loaded electronic components of computer systems.

A considerable advantage of all liquid cooling systems is the heat capacity of liquids, which is better than air capacity (from 1500 to 4000 times) and higher than the heat-transfer coefficient (increasing up to 100 times). To cool one modern FPGA chip, 1 m3 of air or 0.00025 m3 (250 ml) of water per minute is required. Transfer of 250 ml of water requires much less of electric energy, than transfer of 1 m3 of air. Heat flow, transferred by similar surfaces with traditional velocity of the heat-transfer agent, is in 70 times more intensive in the case of liquid cooling than in the case of air cooling. Additional advantage is the use of traditional, rather reliable and cheap components such as pumps, heat exchangers, valves, control devices, etc. In fact, for corporations and companies, which deal with equipment with high packing density of components operating at high temperatures, liquid cooling is the only possible solution of the problem of cooling of modern computer systems. Additional possibilities to increase liquid

cooling efficiency are improvement of the initial parameters of the heat transfer agent: increasing of velocity, decreasing of temperature, providing of turbulent flow, increasing of heat capacity and reducing of viscosity.

Heat transfer agent in liquid cooling systems of computer technique is liquid such as water or any dielectric liquid. Heated electronic components transfer the heat to the permanently circulating heat transfer agent liquid, which, after its cooling in the external heat exchanger, is used again for cooling of heated electronic components. There are several types of liquid cooling systems. Closed loop liquid cooling systems have no direct contact between liquid and electronic components of printed circuit boards. In open loop cooling systems (liquid immersion cooling systems) the electronic components are immersed directly into the cooling liquid. Each type of the liquid cooling systems has its own advantages and disadvantages.

In the closed loop liquid cooling systems all heat-generating elements of the printed circuit board are closed by one or several flat plates with a channel for liquid pumping. So, for example, cooling of a supercomputer SKIF-Avrora is based on the principle of one cooling plate for one printed circuit board. The plate, of course, had a complex surface relief to provide tight heat contact with each chip. Cooling of a supercomputer IBM Aquasar is based on the principle of one cooling plate for one (heated) chip. In each case the channels of the plates are united by collectors into a single loop connected to a common radiator (or another heat exchanger), usually placed outside the computer case and/or rack or even the computer room. With the help of the pump the heat transfer agent is pumped through the plates and dissipates the heat, generated by the computational elements, by means of the heat exchanger. In such system it is necessary to provide the access of the heat transfer agent to each heat-generating element of the calculator, which means a rather complex piping system and a large number of pressure-tight connections. Besides, if it is necessary to provide the maintenance of the printed circuit boards without any serious demounting, then the cooling system must be equipped with special liquid connectors which provide pressure-tight connections and simple mounting/demounting of the system.

In closed loop liquid cooling systems it is possible to use water or glycol solutions as a heat transfer agent. However, the leak of heat transfer agent can lead to possible ingress of electrically conducting liquid to unprotected contacts of printed circuit boards of the cooled computer, and this, in its turn, can be fatal for both separate electronic components and the whole computer system. To eliminate failures the whole complex must be stopped, and the power supply system must be tested and dried up. Control and monitoring systems of such computers always contain multiple internal humidity and leak sensors. To solve the leak problem a method, based on negative pressure of liquid in cooling system, is frequently used. According to this method, water is not pumped in under pressure, it is pumped out, and this practically excludes the leak of liquid. If air-tightness of the cooling systems is damaged, then the air ingresses the system but no leak of liquid occurs. Special sensors are used for the detection of leaks, and modular design allows the maintenance without stopping of the whole system. However, all these capabilities considerably complicate the design of hydraulic system.

Another problem of closed loop liquid cooling systems is a dew point problem. In the section of data processing the air is in contact with the cooling plates. It means that if any sections of these plates are too cold and the air in the section of data processing is warmer and not very dry, then moisture can condense out of the air on the plates. Consequences of this process are similar to leaks. This problem can be solved ether by hot water cooling, which is not effective,

or by control and keeping on the necessary level of the temperature and humidity parameters of the air in the section of data processing, which is complicated and expensive.

The design becomes even more complex, when it is necessary to cool several components with a water flow proportionally to their heat generation. Besides the branched pipes, it is necessary to use complex control devices (simple T-branches and four-ways are not enough). An alternative approach is the use of an industrial device with flow control, but in this case the user cannot considerably change configuration of cooled computational modules.

Advantages of closed loop liquid cooling systems are:

- use water or water solutions as the heat transfer agent, which are available, have perfect thermotechnical properties (heat transfer capacity, heat capacity, viscosity), simple and comparatively safe maintenance;

– a large number of unified mechanisms, nodes and details for water supply systems, which can be used;

– great experience of maintenance of water cooling systems in industry. However, closed loop liquid cooling systems have a number of significant disadvantages, which restrict their widespread use:

– difficulties with detection of the point of water leakage;

– catastrophic consequences that are the result of leakages not detected in time;

– technological problems of leakage elimination (a required power-off of the whole computer rack, that is not always possible and suitable);

– required support of microclimate in the computer room (a dew point problem);

– a problem of cooling of all the rest components of the printed circuit board of the RCS computational module. Even slight modification of the RCS configuration requires a new heat exchanger;

– a problem of galvanic corrosion of aluminum heat exchangers or a problem of mass and dimensions restrictions for more resistant copper heat exchangers (aluminum is three times as lighter than copper);

– air removal from the cooling system that is required before starting-up and adjustment and during maintenance;

– complex placement of the computational modules in the rack with a large number of fittings required for plug-in of every computational module;

– necessity of the use of specialized computer rack with significant mass and dimension characteristics.

In open loop liquid cooling systems the heat transfer agent is the principal component, a dielectric liquid based, as a rule, on a white mineral oil that provides much higher heat storage capacity of the heat transfer agent, than the one of the air in the same volume. According to their design, such system is a bath filed with the heat transfer liquid (also placed into a computer rack) and which contains printed circuit boards and servers of computational equipment. The heat, generated by electronic components, is dissipated by the heat transfer agent that circulates within the whole bath. Advantages of immersion liquid cooling systems are simple design and capability of adaptation to changing geometry of printed circuit boards, simplicity of collectors and liquid connectors, no problems with control of liquid flows, no dew point problem, high reliability and low cost of the product.

The main problem of open loop liquid cooling systems is chemical composition of the used heat transfer liquid which must fulfil strict requirements of heat transfer capacity, electrical

conduction, viscosity, toxicity, fire safety, stability of the main parameters and reasonable cost of the liquid.

Open loop liquid cooling systems have the following advantages:

– insensibility to the leakages and their consequences, capability of operating even with local leakages of the heat transfer agent;

– insensibility to climate characteristics of the computer room;

– solution of the problem of cooling of all RCS components, because the printed circuit board of the computational module is immersed into the heat transfer agent;

– capability of modification of the configuration of the printed circuit board of the computational module without modification of the cooling system;

– simplicity of hydraulic adjustment of the system owing to the lack of complex system of collectors;

– possibility of the use of unified mechanisms, nodes and details produced for hydraulic systems of machine industry, and know-how of maintenance of electrical equipment that uses dielectric oils;

– increasing of the total reliability of the liquid cooling system.

The disadvantages of open loop liquid cooling systems are the following:

– necessity of additional pump and heat exchange equipment for improvement of thermotechnical properties (heat transfer capacity, heat capacity, viscosity) of the heat transfer agent. Here special dielectric organic liquids are used as the heat transfer agent;

– necessity of training of the maintenance of staff and keeping increased safety precautions for work with the heat transfer agent;

– necessity of more frequent cleaning of the computer room because of high permeability of the heat transfer agent, especially in the case of leakage;

– necessity of special equipment for scheduled and emergency maintenance operations (mounting/demounting of the computational module, loading/unloading of the heat transfer liquid, etc.);

– increasing of the maintenance cost because of the necessity of regular changeout of the heat transfer liquid when its service life is over and necessity of heat transfer agent management (transporting, receipt, accounting, storing, distribution, recovery of the heat transfer agent, etc.) in the corporation.

Estimating the given advantages and disadvantages of the two liquid cooling systems we can note more weighty advantages of open loop cooling systems for electronic components of computer systems. That is why for RCS computational modules designed on the basis of promising FPGA families, it is reasonable to use liquid cooling, particularly immersion of printed circuit boards of computational modules into liquid heat transfer agent based on a mineral oil.

At present the technology of liquid cooling of servers and separate computational modules are developed by many vendors and some of them have achieved success in this direction. However, most of 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 [12]. 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 density of placement. This increases considerably the number of heat generating compo-

nents in comparison with microprocessor modules, complicates application of the technology of direct liquid cooling IMMERS along with the other end solutions of immersion systems, and requires additional technical and design solutions for effective cooling of RCS computational modules.

4. Reconfigurable computer system based on Xilinx UltraScale FPGAs

Since 2013 the scientific team of SRC SC and NC has actively developed the domain of creation of next-generation RCS on the basis of their original liquid cooling system for printed circuit boards with high density of placement and the large number of heat generating electronic components. The basis of design criteria of the computational module (CM) of next-generation RCS with open loop liquid cooling system are the following principles:

– the principal configuration of the computer rack is the computational module with the 3U height and the 19 width and with self-contained circulation of the cooling liquid;

– one standard 47U computer rack can contain not less than 12 computational modules with liquid cooling;

– one computational module can contain 12-16 printed circuit boards with FPGA chips;

– each printed circuit board must contain up to 8 FPGAs with dissipating heat flow of about 100 Watt from each FPGA;

– a standard water cooling system, based on industrial chillers, must be used for cooling the liquid.

The principal element of modular implementation of open loop immersion liquid cooling system for electronic components of computer systems is a reconfigurable computational module of a new generation (see the design in Fig. 9). The CM of a new generation consists of a computational section, a heat exchange section, a casing, a pump, a heat exchanger and a fitting. In the casing, which is the base of the computational section, a hermetic container with dielectric cooling liquid and electronic components with the elements that generate heat during the operation, is placed. The electronic components can be as follows: computational modules (not less than 12-16), control boards, RAM, power supply blocks, storage devices, daughter boards, etc. The computational section is closed with a cover. The computational section adjoins to the heat exchange section, which contains a pump and a heat exchanger. The pump provides circulation of the heat transfer agent in the CM through the closed loop: 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 the stationary system of engineering services. The design of the computer rack with placed CMs is shown in Fig. 9b.

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).

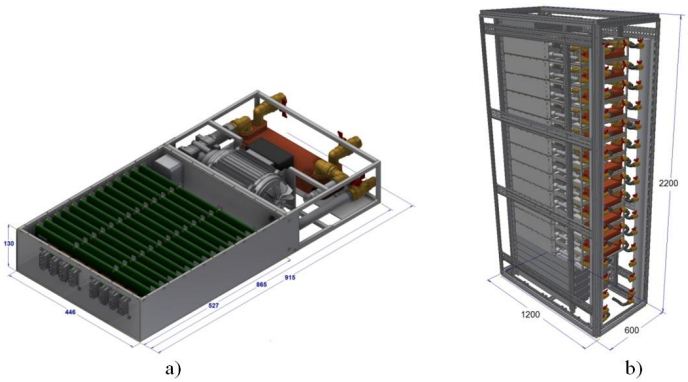

Figure 9. The design of the computer system based on liquid cooling a – the design of the new generation CM, b – the design of the computer rack

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. The supply of cold secondary cooling liquid and the extraction of the heated one into the stationary system of engineering services connected to the rack is performed via fittings (or balanced valves). A set of computer racks placed in one or several computer rooms forms a computer complex. To maintain the computer complex, it is connected to the source of the secondary cooling liquid, to the power supply, and to the host computer that controls this computer complex.

Besides the advantages, which are typical for open loop liquid cooling systems, the considered modular implementation of the open loop liquid cooling system for electronic components of computer systems has a number of additional advantages:

– printed circuit boards of computational modules and reconfigurable computational modules are identical, relatively stand-alone and interchangeable. If one of the CMs fails or if technical diagnosis is required, then it is not needed to disconnect completely the computer rack and to stop the execution of a task;

– high placement complexity of FPGAs in the CMs;

– the proposed technical solution allows, if it is necessary, increasing of performance of reconfigurable computational modules without significant increasing of dimensions (a more high-power pump and a heat exchanger can be placed into the selected dimensions). Growth of the number of printed circuit boards of the computational modules will slightly increase the dimensions (depth) of the reconfigurable computational module, but the density of placement will remain unchangeable.

Owing to simplicity of design of the heat exchange section of the reconfigurable computational module, its reliability grows significantly.

The 19 computer rack of the supercomputer (see the design in Fig. 9b) has the following technical characteristics:

– a standard 47U computer rack;

– 12 3U computational modules with liquid cooling;

– each computational module contains 12 printed circuit boards with the power of 800 Watt each;

– each printed circuit board contains 8 Kintex UltraScale XCKU095-1FFVB2104C FPGAs, 95 million equivalent gates (134 400 logic blocks) each;

– the performance of the new generation computational module is 105 TFlops;

– the performance of the computer rack, which contains 12 CMs, is 1 PFlops;

– the power of the computer rack, which contains 12 CMs is 124 kWatt.

The performance of one computer rack with the liquid cooling system, which contains 12 CMs with 12 printed circuit boards each, in 6.55 times exceeds the performance of the similar rack with CMs Taygeta. Here the performance of one CM of a new generation is increased in 8.74 times in comparison with the CM Taygeta. Such qualitative increase of the specific performance of the system is provided by the density of placement, increased more than in three times owing to the original design solutions, and by increasing of the clock rate and the number of gates in one chip.

For testing technical, technological solutions and for determination of expected technical and economical characteristics and service performance of the designed high-performance reconfigurable computer system with liquid cooling, a number of models, experimental and technological prototypes. Fig. 10 shows the technological prototype of the new generation CM. For this CM new designs of printed circuit boards and computational modules with high density of placement were created.

Figure 10. The technological prototype of a new generation CM

The printed circuit board of the promising computational module contains 8 Virtex UltraScale FPGAs of logic capacity of not less than 100 million equivalent gates each. The CM computational section contains 12-16 printed circuit boards of computational modules with the power up to 800 Watt each. Besides, all boards are completely immersed into electrically neutral liquid heat-transfer agent; the heat exchange section contains pump components and the heat exchanger, which provide the flow and cooling of the heat-transfer agent. The design height of the new generation CM is 3U.

For creation of an effective immersion cooling system a dielectric heat-transfer agent was developed. This heat-transfer agent has the best electric strength, high heat transfer capacity, the maximum possible heat capacity and low viscosity. On the basis of the transformer oil, a new oil called Dielectric oil with reduced viscosity MD-4.5 for cooling of electronic components of computers with reduced viscosity was created according to the method of vacuum distillation. For the oil DM-4.5 are determined technical specification TU 38.401-58-421-2015 and recommendations of its use. The oil MD-4.5 was completely tested in the heat engineering laboratory of SRC SC and NC on the technological prototype of the computational module with immersion open loop liquid cooling system. The performed set of laboratory and service tests proved reasonability of use of the oil MD-4.5 for cooling of electronic computer components and of the use of low-power pumps for its circulation (due to its reduced viscosity).

During the design of the new generation CM we have obtained a number of breakthrough technical solutions, such as an immersion power supply block for the voltage of 380 V and a transducer DC/DC 380/12 V, the minimum height of the printed circuit board of the CM of 100 mm is provided, an original immersion control board is designed and produced. For the cooling subsystem of the new generation CM we have determined required components of the cooling system such as an original heat interface, a low-height FPGA radiator of an original design for convective heat exchange, a pump and a heat exchanger optimal for the used heat-transfer agent. Besides, we have determined the design of a volume compensator of the heat-transfer agent and control elements of the cooling subsystem, such as optical level sensors and a flow sensor. The developed implementations of the cooling system design and circulation of the heat-transfer agent provide effective solution of the heat dissipation problem from the most heated components of the CM.

The complex of the developed solutions concerning immersion liquid cooling system will provide the temperature of the heat-transfer agent not more than 33^o C, the power of 91 Watt for each FPGA (8736 Watt for the CM) in the operating mode of the CM. At the same time, the maximum FPGA temperature does not exceed 55^o C. This proves that the designed immersion liquid cooling system has a reserve and can provide effective cooling for promising families of Xilinx FPGAs (UltraScale+, UltraScale 2,etc.).

Conclusion

The use of air cooling systems for the designed supercomputers has practically reached its limit because of the reduction of cooling effectiveness with growing of consumed and dissipated power, caused by the growth of circuit complexity of microprocessors and other chips. That is why the use of liquid cooling in modern computer systems is a priority direction of perfection of cooling systems 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 the 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 55 ^{o}C and the temperature of the heat-transfer agent not more than 33 ^{o}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, such as principal components of reconfigurable supercomputers, provide stable, practically linear growth 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.

This paper was financially supported in part by the Ministry of Education and Science of the Russian Federation under Grant 14.578.21.0006 of 05.06.2014, ID RFMEFI57814X0006

Referenses

1. I.A. Kalyaev, I.I. Levin, E.A. Semernikov, V.I. Shmoilov. Reconfigurable multipipeline computing structures. Nova Science Publishers, New York, USA, 2012.

2. Vicyn N. Sovremennye tendencii razvitija sistem avtomatizirovannogo proektirovanija v oblasti jelektroniki // Chip News, 1, 1997. S. 1215. [N.Vitsyn. Modern trends of development of CAD-systems in electronics // Chip News, 1, 1997, pp. 12–15.]

3. In the Beginning. By Ron Wilson, Editor-in-Chief, Altera Corporation. https://www.altera.com/solutions/technology/system-design/articles/_2013/in-the-beginning.html/ (accessed : 04.04.2016)

4. https://www.altera.com/(accessed: 04.04.2016)

5. www.xilinx.com/(accessed: 04.04.2016)

6. www.actel.com/(accessed: 04.04.2016)

7. Grushvickij R.I., Mursaev A.H., Ugrjumov E.P. Proektirovanie sistem na mikroshemah programmiruemoj logiki. Peterburg: BHV-Peterburg, 2002. 636 s. [R.I. Grushvitskiy, A.Kh. Mursayev, E.P. Ugrumov. System design on chips of programmable logic. St.-Petersburg: BHV-Petersburg, 2002, 636 pp.]

8. Steshenko V., Shipulin S., Hrapov V. Tendencii i perspektivy razvitija PLIS i ih primenenie pri proektirovanii apparatury COS // Komponenty i tehnologii, 2000, 8. [V. Steshenko, S. Shipulin, V. Khrapov. Trends and perspectives of FPGA development and their application for design of DSP devices // Components and technologies, 2000, 8.]

9. Tarasov I. Jevoljucija PLIS serii Virtex // Komponenty i tehnologii, 2005, 1. [I. Tarasov. Evolution of Virtex FPGAs // Components and technologies, 2005, 1.]

10. V.A. Gudkov, A.A. Gulenok, V.B. Kovalenko, L.M. Slasten. Multi-level Programming of FPGA-based Computer Systems with Reconfigurable Macroobject Architecture // IFAC Proceedings Volumes (ISSN 14746670), Programmable Devices and Embedded Systems, Volume 12, part 1, 2013, pp. 204–209.

11. Kalyaev I.A., Levin I.I., Dordopulo A.I., Slasten L.M. Reconfigurable Computer Systems Based on Virtex-6 and Virtex-7 FPGAs // IFAC Proceedings Volumes, Programmable Devices and Embedded Systems. vol.12, 1, 2013, pp. 210–214.

12. I.I. Levin, A.I. Dordopulo, Y.I. Doronchenko, M.K. Raskladkin. Reconfigurable computer system on the base of Virtex UltraScale FPGAs with liquid cooling // Proceedings of international scientific conference Parallel computer technologies (PaCT2016), 2016 pp. 221–230.