DOI: 10.14529/jsfi230204

Multipurpose Reconfigurable Supercomputer with Immersion Cooling

Ilya I. Levin[1]***, Aleksandr M. Fedorov***[1]***, Yuriy I. Doronchenko***[1]***, Maksim K. Raskladkin***[1]

In the paper we consider a promising universal reconfigurable supercomputer, the computing nodes of which are reconfigurable computing device Arcturus. It was developed at the Supercomputers and Neurocomputers Research Center (Taganrog) and based on modern Xilinx FPGAs of UltraScale+ family of HBM-series. The purpose is to achieve the highest computational layout density, to ensure balanced power supply and cooling, as well as the implementation of powerful data exchange configuration. The supercomputer can have up to 1.5 thousand FPGAs of a single computing field. It has extensive information exchange capabilities between FPGAs within the device and between devices to solve tightly coupled problems. Differential lines with multi-gigabit transceivers connected to them are used as the main connections between FPGAs. It provides exchange at the velocity up to 25 Gbit/s. Information interaction between the RCD is performed through optical channels with the capacity up to 4.5 Tbit/s. Immersion technology is used for cooling components of computing system. It provides the removal of the total heat output up to 20 kW. The developed power supply configuration is based on the input constant voltage of 380 V and provides stable power supply to the components. Owning to the implementation of time-consuming algorithms for various scientific and technical problems with the high real performance, it is possible to widespread use of the Arcturus supercomputer. Scaling of computing nodes will allow designing an entire computing circuit of a supercomputer with the performance up to several tens of Petaflops.

Keywords: supercomputers, reconfigurable computing systems, computing performance, immersion cooling systems, computing energy efficiency, computational density, highly connected problems.

Introduction

The solution of modern time-consuming scientific and technical problems of various subject areas requires the design of computing systems not only with overall high chip performance, but also with a variety of high-capacity channels for direct information exchange between computing nodes. Solving such tightly coupled problems using calculators based on universal processors is extremely inefficient. This is because traditional supercomputer architectures do not correspond to the structure of implemented algorithms. Therefore, they provide the high computing performance for very small class of problems. In addition, the processor performance growth has now stopped due to the approach to the technological limit of production.

In this regard, it seems most appropriate to use reconfigurable computing systems (RCS) to adapt their architecture for solving problem structure [1, 2]. The RCS concept is the maximum structural implementation of the task information graph on the supercomputer hardware resource. Such implementation where the algorithm operation is assigned to each operating vertex of the graph, arcs between vertices determine the data transfer between operations, and information vertices of the graph are implemented by corresponding memory channels. The structural implementation of calculations ensures the organization of pipelined, fastest data processing [3].

Field-programmable gate arrays (FPGAs) as the element base of reconfigurable systems, in contrast to the element base of multiprocessor computing systems, continues to support the

[1]Supercomputers and Neurocomputers Research Center, Taganrog, Russia

RCS development by maintaining the growth rate of hardware resource and operating clock frequency. The reconfiguration possibility allows the user to create virtual specialized computers within the same architecture, adequate to the structure of each solved application problem. However, the main problem is the organization of the commutation system, which should provide a wide capacity, but not require large hardware costs and time delays. Existing reconfigurable supercomputers are characterized by powerful information exchanges only within the basic functionally complete nodes – calculating blades [4]. The bandwidth between the boards of computing modules is significantly lower. This limits the possibilities of efficient parallelization of calculations in application programs.

Currently, the employees of the Supercomputers and Neurocomputers Research Center are designing the experimental sample of a prospective universal reconfigurable computing device (RCD) Arcturus based on FPGA, which is a development of the commercially produced RCB Neckar [4]. A number of breakthrough technical solutions have been implemented in the new computer. Owning to them, it is possible to implement large graphs of complex problems in an entire computing circuit and perform calculations without interruptions due to the powerful information exchange system. At the same time, it has provided the necessary level of power supply and cooling.

The current level of energy consumption of chips does not allow obtaining effective solutions using the air cooling [5]. As a rule, the combined types of cooling, mainly liquid, are used at the development of new computing complexes. The most effective approach is the immersion technology with direct immersion of electronic components in a coolant [6]. It began to actively develop and received its implementation in computer technology samples. This technology ensures the absence of complex structural elements. This determines the highest layout density, as well as the use of low-cost coolants and guarantees the absence of critical leaks. However, constructs designed by various manufacturers [7, 8] for liquid cooling have a very low layout density of computing elements. As a rule, they are intended only for devices based on CPU or GPU. An original immersion cooling system has been developed.

The article is organized as follows. Section 1 presents the architecture of supercomputer, allowing increasing the performance efficiency at solving widespread tasks. In Section 2, the RCD Arcturus structure and its design features are described. Section 3 describes the organization of information interaction between RCB, as well as the features of the use of multi-gigabit transceiver technology implemented in them. Section 4 is devoted to the RCD Arcturus, which provides to problem solutions with special memory requirements. Section 5 describes the computing resources of the developed RCB. Section 6 is devoted to the description of the original motherboard, designed and manufactured for loading FPGAs and controlling the computing process in the RCB. Section 7 describes the design components of the RCB cooling system and the features of implemented immersion liquid cooling system. In Section 8, the real performance evaluation of the RCD Arcturus was performed using two floating-point tasks. In conclusion, the achieved results of the development of the promising RCD Arcturus are listed.

1. Supercomputer Architecture

For solving modern problems, it is necessary not only to have a large number of computing chips and memory, but also rapid data exchange between components of the computing system. It is known that the real performance is rapidly falling in tightly coupled problems due to the significant amount of transfers. For example, this class of problems includes the currently

widespread tasks of machine learning, simulation of complex physical processes and technological devices, problems of the Earth remote sensing and digital signal processing. To increase the efficiency of solving such problems, a supercomputer architecture is proposed to organize a commutation system with up to 1.5 thousand chips combined in a single computing circuit (Fig. 1). An architecture feature is the direct information communication between the chips in the horizontal and vertical directions, as well as ring connections in each horizontal layer.

The implementation of this architecture on a single printed circuit board (PCB) is not possible. The overall dimensions are too large. The components' installation is almost impossible. The cost of one board is prohibitively high; the maintainability is low. It is necessary to decompose into separate modules for implementation each row of chips on a separate PCB. The set of blades obtained in this way are placed vertically and ensure the interaction using the cross-board (Fig. 2). The horizontal placement of boards is impractical, as this will increase the depth, and lead to the impossibility of chips' cooling.

Besides the computing blades, the power blades and control module blade must be provided for power supply, monitoring and control. The power module can provide the power to several calculating blades (for example, four). Thus, it is possible to form an element of a regular structure – a processing unit (Fig. 2). Processing units are scaled in a single construct – a computing device. The number of such units in the computing device can be equaled to four. There can be one control module blade.

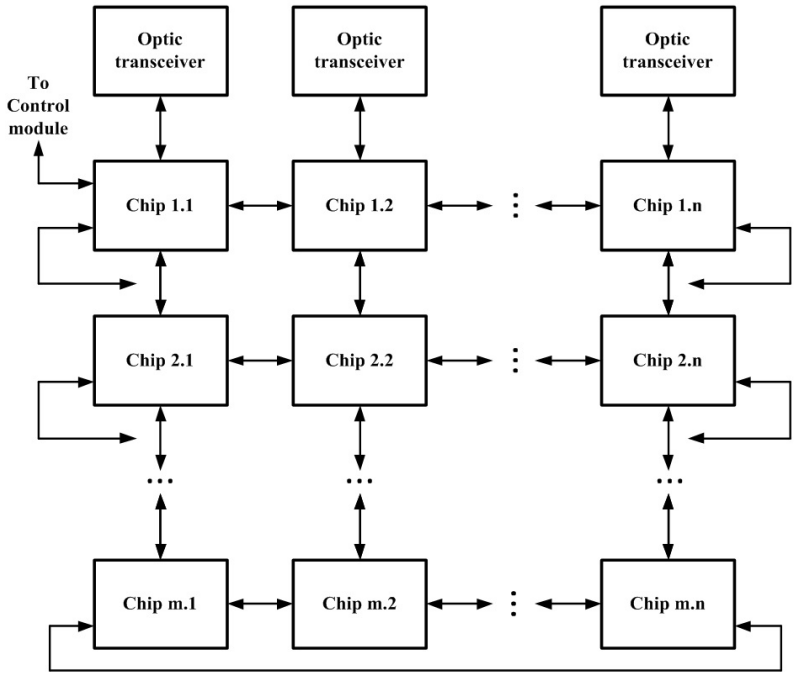

Figure 1. Supercomputer architecture

The decomposition of the proposed architecture leads to the rupture of vertical connections. The connections can be restored using the high-speed optical channels. Using the optical transceivers, a set of computing devices is combined with each other into a computing rack (Fig. 3). Therefore, the supercomputer can consist of several computing racks, in which computing devices jointly solve a problem.

The architecture shown in Fig. 1 corresponds to the "tor" architecture that does not contain vertical circuit information connections. Their implementation in general is impossible. Since the

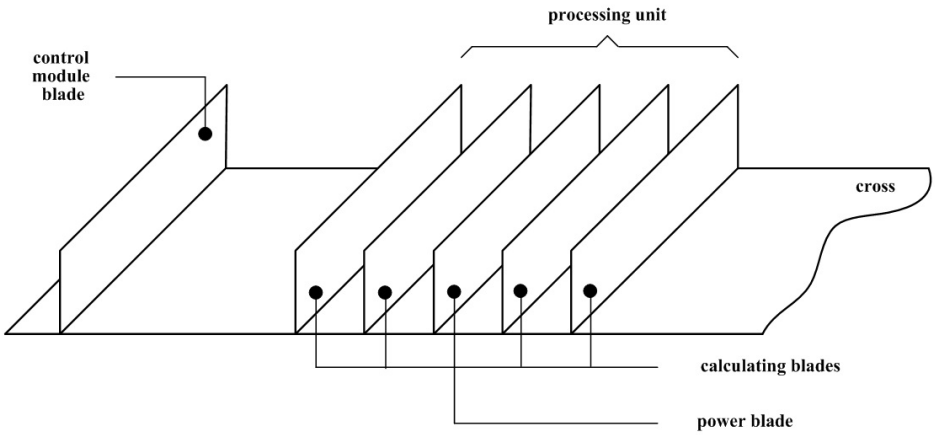

Figure 2. Blades placement

supercomputer is based on the modular expandability principle [9], and the number of processing units and computing devices is not known in each case.

At implementation of 6 chips on calculating blades, 4 processing units in the computing device will provide 96 chips. In the case of 16 computing devices in one computing rack, there will be more than 1.5 thousand chips.

This architecture can be used with ASIC, eASIC chips. The use of FPGA is most effective. Reconfigurable FPGAs allow solving many problems, which ensures the supercomputer versatility. In this case, the processing units combined in one construct called the RCD.

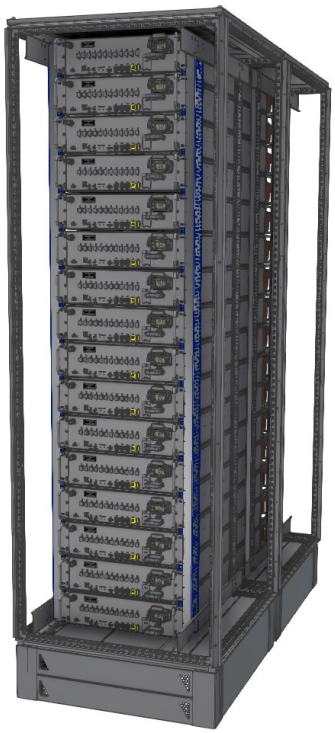

Figure 3. Computing rack

To solve time-consuming problems on the Arcturus supercomputer, it is necessary to use the paradigm of structural organization of calculations with parallelization by iterations and layers [10]. This approach provides the linear increasing of system performance at the linear increasing in hardware resource. This is not available to any modern computing system, based

on processors and graphics accelerators. This is because the number of information exchanges is often comparable to the number of performed operations in time-consuming problems. Accordingly, the data exchange time between computing nodes has a greater impact on the problem solution velocity with such connectivity in traditional systems. In some cases, when the number of information exchanges in problems is close to the number of performed operations, the data parallelization schemes used in traditional systems are critical to the RAM, the access speed and the data transfer rate in interfaces between computing nodes. There is a "bottleneck problem", and the data exchange time turns out to be longer than the calculation time. Therefore, the system performance is a significant decrease at increasing the number of computing nodes.

In contrast to traditional approaches, the communication is provided by a spatial commutation matrix at the structural organization of calculations with parallelization by iterations and layers between the functional system nodes. The problem of memory access speed is solved by using distributed memory technology and the independent multi-channel access.

The used approach has been repeatedly tested and has proven itself well at solving many time-consuming, tightly coupled problems, such as problems in the field of bioinformatics, artificial intelligence, molecular modeling, simulation of complex physical processes, bit processing, geophysics, the Earth remote sensing and many others.

2. RCD Arcturus Structure

The supercomputer is based on RCD. The main structure features of the developed RCD Arcturus were considered. The RCD Arcturus design has the standard 3U 19" dimension and contains 16 calculating blades (CBs) vertically arranged on a cross-board with 6 Xilinx UltraScale+ XCVU37P FPGAs in each. The control module blade based on the Intel Skylake processor is used for general management, configuration and control of computing process.

The power blade DC/DC has been developed for power supply of the processing unit (four calculating blades). It converted performing conversions from the input power supply of 380 V DC to 12 V. The maximum power consumption of the RCD Arcturus is 20 kW.

These electronic blades are placed on a single cross-board (Fig. 4), which provides inter-module information and control interaction.

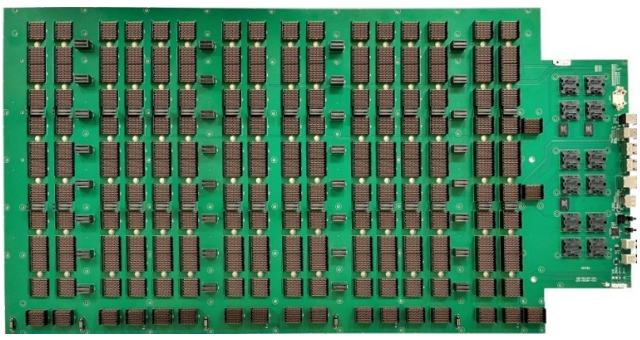

Figure 4. Cross-board

Cooling of the loaded electronic RCD components is provided by the immersion cooling system based on a dielectric coolant [11, 12]. The coolant has high electrical strength and thermal conductivity, as well as the highest possible heat capacity at low viscosity.

The construct consists of the main computing section with blades, filled by the coolant, and an additional dry section in front, which located user interfaces, connectors and optical modules. The heat exchanger for heat energy transferring from the first (coolant) to the second (water) cooling circuit is located behind the main section. The unique cross-board provides the high-frequency data transfer between blades. The components soldering is performed on an automatic line. The PCB dimensions are large and equal 845×436 mm. Therefore, the assembly of components and pressing of connectors is difficult at soldering paste application technology and pressing. The cross-board provides the sealed transition between the environments (coolant – air). The external dry-area of the cross-board implements the display output, control buttons, as well as interfaces for connecting a keyboard, mouse, display, Ethernet and optical information channels used for interdevice information exchange.

The 3D model of the RCD Arcturus is shown in Fig. 5. The RCD Arcturus with the open lid is shown in Fig. 6.

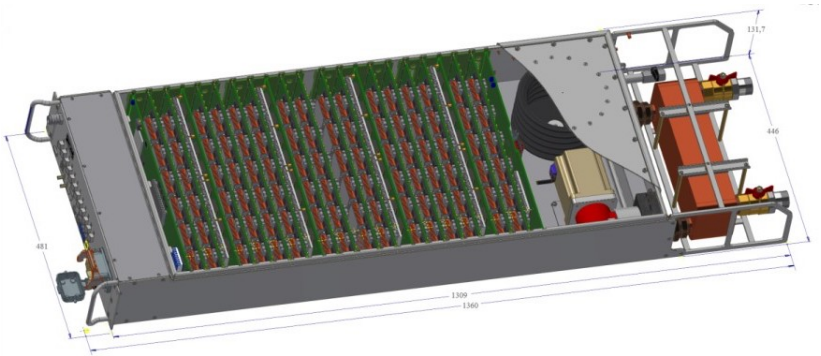

Figure 5. 3D model of the RCD Arcturus

Figure 6. RCD Arcturus

3. Data Exchange Configuration

The most important feature of the promising reconfigurable supercomputer is the support of unique computing power by the extensive capabilities of data exchange configuration. It represents multiple communication channels between FPGAs within the CB, as well as between FPGAs of neighboring CBs [13].The structure of information links has been clarified at operating with the CB topology and designing technical solutions. Communication between FPGAs is

provided via differential lines using multi-gigabit transceivers (MGT), integrated into the FPGA. Auxiliary connections are differential LVDS lines connected to the FPGA HP-banks (Fig. 7). 24 differential lines with data transfer rates up to 24 Gbit/s within the CB and 24 differential lines with data transfer rates up to 16 Gbit/s between the CBs are implemented between the pair of FPGAs. The total capacity of CB communication channels is 15.6 Tbit/s, including between CBs is 9 Tbit/s. Therefore, the RCD implements a universal orthogonal high-speed information communication system between computing FPGAs (Fig. 8).

At creation the computing complexes, the information interaction between RCD is performed through optical channels. Amphenol multichannel optical transceivers are installed on the cross-board and connected to the first calculating blade via the split connection, providing the capacity up to 4.5 Tbit/s (Fig. 8, 9).

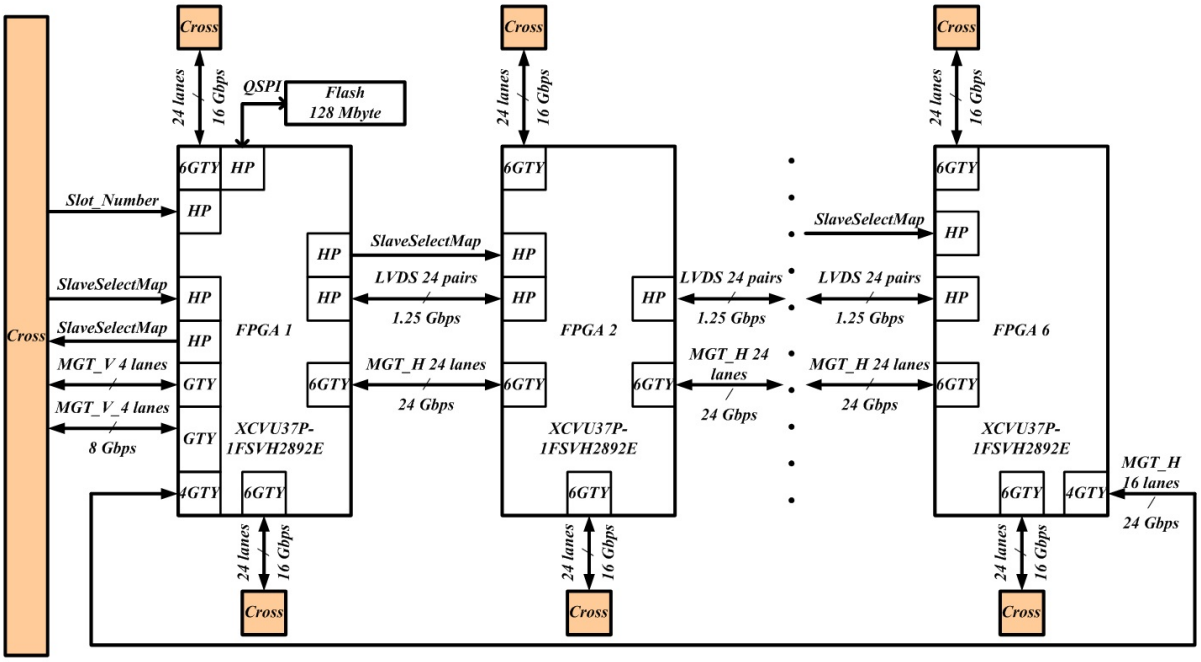

Figure 7. CB communication structure

Separately note the application features of the multi-gigabit transceiver technology. The part of the XCVU37P FPGA architecture is 96 high-speed (up to 25.785 Gbit/s) transceivers. For high data transfer rate, it is necessary to use special topological solutions at PCB designing: the limit on the line length, on the amount and various versions of transition holes, rounding corners, the use of a special PCB material that provides a low dielectric loss coefficient, etc.

However, the use of only topological solutions is not enough for high data transfer rates close to the maximum possible. The Xilinx FPGA manufacturer provides a special procedure for transceiver configuration to improve the transmitted data integrity. This setup of communication lines is a very difficult operation. It includes the searching procedure of transmitter optimal parameters and the receiver operating modes, as well as the margin estimation using an eye diagram and comparing the obtained value with a test mask. The setup procedure must be repeated for each high-speed line. The complete search for one communication line of all possible combinations of receiver and transmitter parameter values is about 100 thousand options. Obviously, manual configuration requires an unacceptably long time.

An original algorithm of progressive setup by the sequential approximation method have been developed by specialists of the Supercomputers and Neurocomputers Research Center.

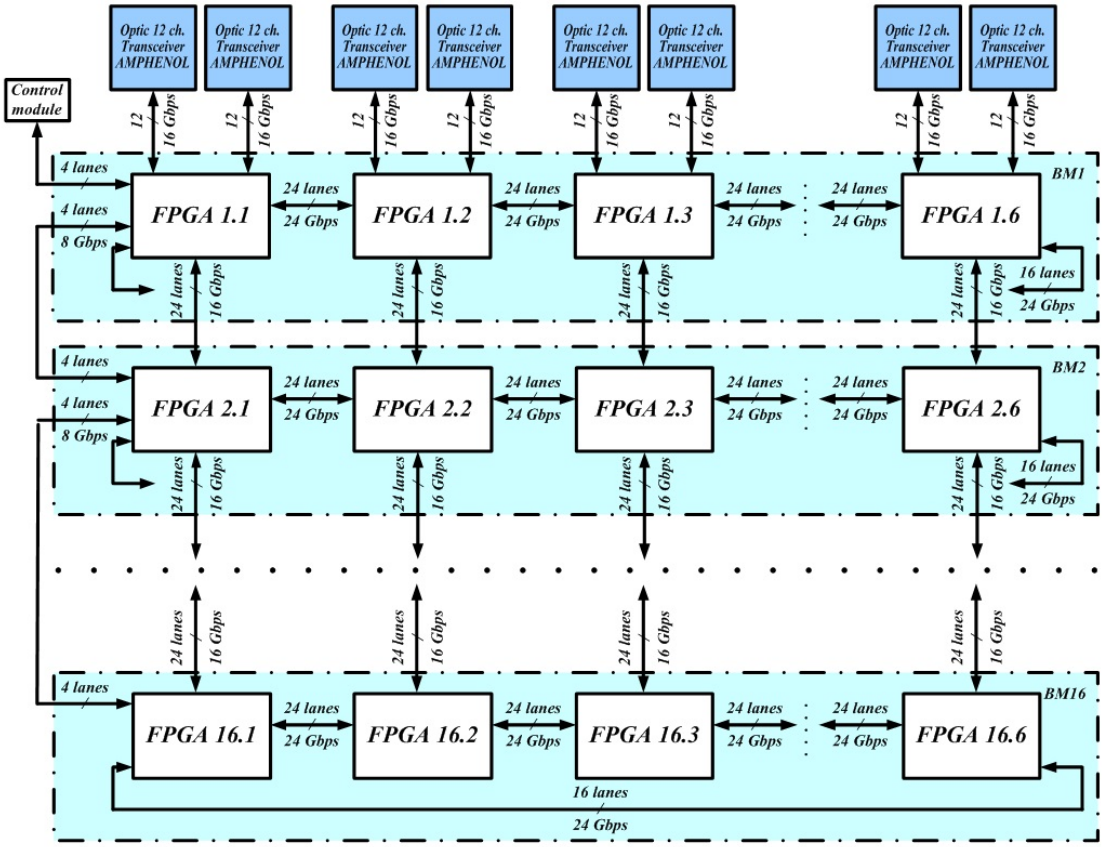

Figure 8. RCD communication structure

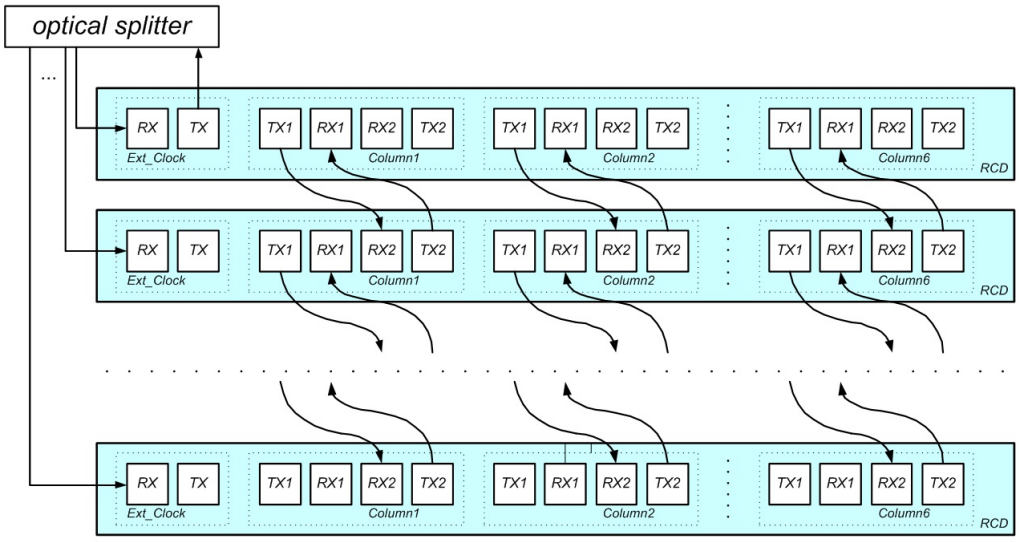

Figure 9. RCD interaction

Depending on the number of combinations, it is possible to reduce in 16 times for testing by iteratively searching the best set of settings, with the gradual decreasing in the range of considered parameters.

The algorithm consists of three iterations. At the first iteration, verification combinations are formed by sampling the transceivers settings every five possible values. Reference values are selected after receiving and measuring the eye diagrams of each combination – a set of settings gives the best result. At the next iteration, the parameters are sampled in steps of 3;

the search range around the selected reference values is narrowing. The following combinations are formed, the verification and adjustment of which gives a new set of reference values. The last iteration is an exact setup with the selection of each parameter from the required neighborhood of reference values. As a result, the number of checking combinations is significantly reduced using the progressive tuning method.

The progressive setup algorithm, as well as the developed functions for quality evaluation the of the eye diagram, logging the setup process and creating reports on the obtained results are implemented as a set of Tcl scripts, executed in the Xilinx Vivado.

The use of developed scenarios makes it possible to implement a fully automatic process of transceiver individual configuration. It does not require the participation of a highly qualified specialist and ensures the highest possible quality and transmission speed for each communication channel between FPGAs.

4. Calculating Blade

The RCD Arcturus should provide the problem solutions with special memory requirements. A typical technical solution used earlier was the placement of static or dynamic memory chips connected with computing FPGAs on the calculating blade [4]. Xilinx, as part of the UltraScale+ family, has released a special HBM line (High Bandwidth Memory). It includes the FPGA with a new hardware resource within the same package – the "HBM2" as integrated DDR. The HBM2 memory has a capacity up to 16 GB and can provide multi-channel access. This technology will significantly expand the possibilities for rational memory use in applications for solving the learning problems and mathematical physics.

The XCVU37P chip, used in the RCD calculating blade, has 8 GB HBM with the peak capacity up to 460 GB/s. For connection to the HBM controller, 256-bit API interfaces are used, the total number of which is 32. Each interface can either operate with its own dedicated 256 MB address space, or access any HBM address via the special switch (32×32 AXI crossbar switch). Experimental researches have shown the exchange possibility at the 12 GB/s speed on each channel.

The dimensions of the UltraScale+ FPGA case are 5 mm larger than the dimensions of the UltraScale FPGA case in previous RCD, for example, the RCD Skat [14], the RCD Neckar [4]. The dimensions of the PCB calculating blade are limited. Therefore, it is impossible to place, as usual, 8 FPGAs. Here 6 FPGAs are implemented. To preserve the computational density of the product, the number of processing units has been increased from 3 to 4.

Currently, the configuration work of the pilot calculating blade is being completed. The photo of CB is shown in Fig. 10.

(a) top view (b) bottom view

Figure 10. Arcturus calculating blade

All 136 differential MGT communication lines between FPGAs have the capacity of 6.6 Tbit/s. 288 differential MGT lines for the cross-board communication have the capacity of 9 Tbit/s. They are configured and tested using the progressive setup algorithm.

Experimental researches have confirmed the possibility of achieving the required error-free data transfer rates between FPGAs in different directions (Tab. 1).

Table 1. Error-free data transfer rates

Direction	Data rate, Gbit/sec	Number of MGT lines	Test duration	Transmitted bits	Bit error rate
FPGA – FPGA	25.6	192	96	$1.7 \cdot 10^{18}$	$0.6 \cdot 10^{-18}$
FPGA – Impel – FPGA	16	288	24	$0.8 \cdot 10^{18}$	$1.2 \cdot 10^{-18}$
FPGA – Impel – Amphenol – Impel – FPGA	16	144	24	$0.2 \cdot 10^{18}$	$5.0 \cdot 10^{-18}$

5. Power Supply, Power Blade

The new UltraScale+ chips computing resource is approximately twice the FPGA resource of the UltraScale previous generation. According to research, the maximum consumption of one FPGA is 160–180 W taking into account the increase in energy efficiency. It has required a radical redesign of the secondary power blade configuration of the calculating blade.

The contradictory requirements for the power supply system include: the high power; the maximum efficiency to minimize parasitic heat generation; the minimal overall dimensions to accommodate the required number of FPGAs on the PCB of the calculating blade. There are no existing solutions that provide the required current and voltage and are placed in one chip of the power blade. An original multiphase power blade system has been developed. Several chips are combined to obtain the required power. The difficulties of synchronization during parallel activation, interaction of radio-electronic components, PCB topological tracing, correct allocation of power polygons, etc. have been overcome. It was necessary to define a certain balance in the PCB between the number of FPGA power layers and the number of signal circuit layers. In addition, the problems of uniform current distribution from each source near the FPGA and ensuring no more than 80A current flows on each individual source FPGA section were solved.

The RCD power configuration ensures the unit's operability in the input voltage range from 350–400 V DC. The power blade is equipped with the soft start system, which ensures a smooth increase in voltage at the converter input and eliminates commutation interference.

The power configuration is started stepwise. The control module blade (CMB) is first started at the device beginning. In this, the primary converter generates the intermediate voltage of 12 V. Further, under the control of a microcontroller (MC), the following cascades are powered: FPGA on CMB, INTEL, cooling system, soft start system, pump motor and others. Information interaction between microcontrollers is implemented via the CAN bus.

To provide the power to the CBs FPGA, the primary conversion is performed on separate boards power blades (PB). The CB power system consists of two identical nodes, each of which receives 11–12.5 V from its converter located on the PB. Each node provides the power to three FPGAs. The power blade is shown in Fig. 11.

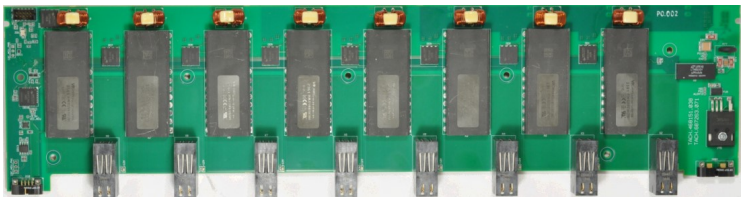

Figure 11. Power blade

The monitoring of voltages, currents and temperatures is the important node of the power blade configuration. A separate microcontroller is used to control and monitor the FPGA power. Data from all FPGA power microcontrollers are fed into the leading CB power microcontroller. Such distributed monitoring scheme makes it possible to increase the power configuration response rate to possible crashes and failures, as well as automatically disable only the failed FPGA, and not the entire calculating blade or RCD. At the same time, the communication system allows to redistribute the monitoring data flow bypassing the disabled FPGA. Note that in such cases it is possible to redistribute the user data flow bypassing the disabled FPGA, if this is available in the FPGA configuration created by the user.

6. Control Module Blade

An original motherboard (control module blade, CMB) was developed and manufactured to FPGA load and control the RCD computing process (Fig. 12). It is based on the Intel SkylakeCore I5-6300U processor. The CMB ensures the RCB functional completeness. It is realized as a separate board. An original basic input-output system (BIOS) has been developed for CMB. It allows using all capabilities of the Intel Skylake processor and external peripheral equipment. A radiator by the original design is used for heat removal from the processor.

Figure 12. Control module blade

The main CMB nodes are the Intel Core i5-6300U processor and the Kintex Ultrascale FPGA. The FPGA CMB performs the controller function and provides interaction between the processor and the first FPGA in the first calculating blade. Communication between the CMB and CB is performed via the cross-board. The CMB also provides hardware monitoring of sensors, pump parameters and other data with using a microcontroller.

7. Cooling System

The RCD case is a sealed container with a coolant, in which electronic modules are immersed. The cooling system is functioning as follows. The pump circulates the coolant in the RCD along the following closed circuit: the heated coolant enters the lamellar heat exchanger and cools there. Then, the coolant re-enters the RCD case under the cross-board under the nec-

essary pressure, with the help of special flow direction setters flows through heatsinks of cooled electronic components, which heats up itself and re-enters the heat exchanger (Fig. 13).

Figure 13. Heat exchanger

The heat exchanger is connected to the secondary cooling circuit through fittings. It was designed for coolant usually with water (Fig. 14).

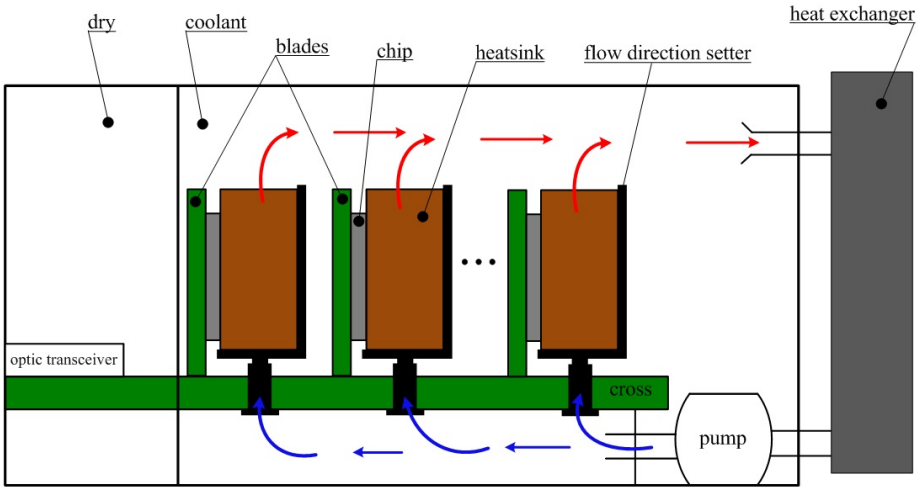

Figure 14. Cooling

The effectiveness of the immersion liquid cooling system is determined by the effectiveness of technical solutions for implementation each component: the used coolant, the construction and parameters of the used FPGA radiators, pumping equipment and heat exchangers.

The coolant must have the best electrical strength, high thermal conductivity, the highest possible heat capacity at low viscosity. The radiator must provide the maximum heat removal surface, the refrigerant circulation through the radiator, the refrigerant flow turbulence in the radiator, manufacturability. The used thermal interface must not degrade and be washed out by the refrigerant, have a consistently high thermal conductivity coefficient. The heat exchanger must provide the high heat transfer coefficient between the main and secondary cooling circuit. The pump with necessary performance is used for refrigerant circulation in the RCD volume.

This complex problem of RCD Arcturus cooling subsystem implementation provides the removal of thermal power up to 20 kW. It has been successfully solved and tested on the number of layouts. First, a new radiator construction was designed. Due to it, the effective heat exchange surface area was increased 2.5 times, compared to the previous design, at slight increasing of radiator dimensions. This became possible by reducing the thickness of ribs (thereby increasing the number of channels by 1.8 times) and increasing the number of grooves in the channels from 252 to 792; the grooves increase the heat removal area and form a special flow pseudo-turbulence (Fig. 15). Therefore, the heat removal capabilities have increased from 80 to 160 W.

The following important problems have been solved during the cooling system implementation:

- a submersible pump has been developed (Fig. 16) with the increased capacity up to 60 l/min, which makes it possible to abandon the cold air forced circulation in computing racks. It significantly simplifies the maintenance, and increases the reliability by reducing the number of open hydraulic connections;
- the cooling efficiency was improved using new heat exchanger (Fig. 12);
- a compensator of coolant volumetric expansion has been developed in the form of a corrugated rubber sleeve. The contact of the coolant with atmospheric air is completely excluded, and potential leaks are minimized;
- significant actions have been performed to change the RCD case and lid design. Due to it, the construction rigidity was increased, the product manufacturability was ensured, the assembly and product repair were simplified, the possibility of leaks was minimized, the construction weight was reduced.

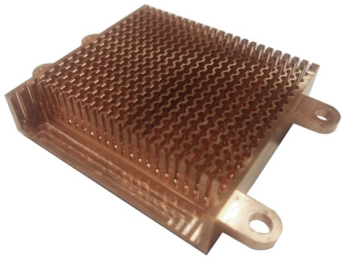

Figure 15. Heatsink

Figure 16. Submersible pump

Note that the distinctive feature of the used FPGA is its lidless-design (without the lid). The radiator installation technology directly on the FPGA crystal using the modern PSG graphite thermal interface by Panasonic with the thickness of 0.1 mm was proposed. Researches have confirmed the advantage of this technology in comparison with the radiator installation on the FPGA lid.

Currently, preparations are underway to test the developed cooling system on the manufactured experimental RCD sample. According to preliminary estimates, the thermal resistance of the "FPGA – coolant" transition is 0.27 C/W, then the temperature drop of the "FPGA – coolant" will be 43C at the average load of 160 W. When the water temperature in the second cooling circuit is 16C, the temperature on the FPGA will be 59C.

Therefore, the developed cooling system makes it possible to divert up to 20 kW of thermal power in the 3U construction. The RCD curb weight is about 150 kg. For comparison, one of the most effective modern solutions with the internal closed loop of liquid cooling is the specialized for sparse computing Cerebras CS-2 system [15] with the similar power consumption of up to 23 kW with the weight of 300 kg and placed in the 15U construction.

8. Results of Problems Implementation

The real performance of the RCD Arcturus was evaluated using two floating-point problems.

The first problem is to estimate the achievable real performance, close to the RCD peak performance. Its algorithm consists mainly of computational operations. Its implementation almost completely involves the FPGA computing resource. For this, a conveyor computing structure

was used that implements the filter function with a finite impulse response (FIR) with symmetric coefficients. The computational pipeline is a series of connected filtration cascades. The performance of 150 Tflops (single precision) can be achieved at implementation of this test problem of digital data processing. Therefore, the RCD construction of 16 RCD Arcturus will provide performance of 2.4 Pflops (single precision) at solving this problem. This assessment confirms the RCD application prospects to solve time-consuming problems.

The second problem is the algorithmic basis of the LINPACK computer performance test. The SLAE solution was performed using the pipeline computing structure for implementation of the LU-decomposition function with the matrix traversal by columns. The performance of 30 Tflops (double precision) can be achieved at implementation of this test problem. The performance will be 480 Tflops (double precision) in RCD construction based on 16 RCD Arcturus at solving the LU-decomposition problem. Obviously, this assessment cannot be a characteristic for comparison with other computing architectures. However, it confirms the RCD feasibility of test problems for traditional architectures and provides the performance level.

The advantage of the Arcturus supercomputer at solving deep machine learning problems is the performance linear increase. It is fundamentally impossible for graphics accelerators (such as Tesla A100). Analysis of the Nvidia official website data [16] shows that the increasing of Tesla A100 accelerators leads to decreasing of their specific performance at the problem of the ResNet-50 v1.5 neural network training. The Arcturus is 2-6 times faster than Tesla A100 at scaling. Besides, the analysis showed that the RCD Arcturus in terms of performance/power consumption is twice as efficient as the most modern Nvidia DGX system for problems of training neural networks of image classifiers and neural networks of natural language processing at the same performance.

Compared to the Dell R7910 RACK [17], the Arcturus RCD provides acceleration more than 610 times, and efficiency in terms of performance/power consumption is 31 times higher at solving complex problems of mathematical physics.

Compared to the SuperServer 2028U-TR4T+ server [18] manufactured by Supermicro Computer based on Intel Xeon CPU E5-2699A v4, the Arcturus RCD provides acceleration more than 114 times, and efficiency in terms of performance/power consumption is 16 times higher at solving the Earth remote sensing problems.

Conclusion

The promising RCD Arcturus, developed at the Supercomputers and Neurocomputers Research Center, has 96 FPGAs of the high integration degree in its construction (3U 19”). It includes a unique layout density of the computing resource – more than 270 million logic cells in total.

At the same time, appropriate power supply and immersion cooling of the product elements are provided at solving time-consuming problems. Ensuring the required RCD characteristics in the given design required a significant complication of the PCB topology and manufacturing technology of its components.

The RCD Arcturus production is scheduled to begin in 2024. Currently, the implementation of model applied problems in the field of high-speed machine learning, the Earth remote sensing, and mathematical physics problems are being prepared at the RCD.

Note that reconfigurable supercomputers require the appropriate engineering infrastructure. Their operation requires connection by means of supply and return manifolds through fittings

(or balancing valves) and flexible pipelines to the secondary cooling circuit, as well as to the power supply source and to the network hub.

The Arcturus RCD features are the wide possibilities of information exchange within the device and between devices for solving highly related problems. The achieved characteristics ensure the versatility of the RCD architecture for a wide class of problems and various fields of application. Combining many devices into a single computing circuit will allow creating collective use centers for solving time-consuming problems, world-class computing complexes with the performance of up to several dozen Petaflops.

Acknowledgements

The research is partially funded by the Ministry of Science and Higher Education of the Russian Federation as part of state assignments "Development of a Multi-Agent Resource Manager for a Heterogeneous Supercomputer Platform Using Machine Learning and Artificial Intelligence" (topic FSEG-2022-0001).

References

1. Kalyaev, A.V., Levin, I.I.: Modular-stackable multiprocessor systems with structural and procedural organization of calculations. Janus-K Publishing house, Moscow (2003). 380 p.

2. Kalyaev, I.A., Levin, I.I.: Reconfigurable computing systems based on FPGAs. SSC RAS Publishing House, Rostov-on-Don (2022). 475 p.

3. Kalyaev, I.A., Levin, I.I.: Reconfigurable multiconveyor computing systems for solving streaming problems. Information technologies and computing systems 2, 12–22 (2011)

4. Levin, I.I, Dordopulo, A.I., Fedorov, A.M., Doronchenko, Yu.I.: Development of technology for constructing reconfigurable computing systems with liquid cooling. Supercomputer Technologies. 5th All-Russian Scientific and Technical Conference. Materials, vol. 1, pp. 184–187. SFedU Publishing House, Taganrog (2018)

5. Kalyaev, I.A., Levin, I.I.: Reconfigurable computing systems with high real performance. In: International Scientific Conference, Parallel Computational Technologies, PaCT-2009. Proceedings, 186 p. SUSU Publishing House, Chelyabinsk (2009)

6. BiXBiT immersion cooling (2018). `https://bixbit.io/assets/docs/ru/investment_project.pdf`, accessed: 2023-04-19

7. Abramov, S., Amelkin, S., Klyuev, L., Chichkovsky, A.: Liquid cooling of computing complexes. Radioelectronic technologies 5, 79–82 (2017)

8. Klyuev, L., Hrebtovsky, I.: Experience in building HPC clusters of IMMERS using immersion liquid cooling (2015). `http://www2.sscc.ru/Seminars/paper/2015/2015-05-14_IMMERS.pdf`, accessed: 2023-04-19

9. Besedin, I.V., Dmitrenko, N.N., Kalyaev, I.A., *et al.*: A family of basic modules for building reconfigurable computing systems with a structural and procedural organization of computing. In: Scientific service on the Internet. Proceedings of the All-Russian Conference. pp. 47–49. MSU, RSU, IVT RAS (2006).

10. Levin, I.I., Pelipets, A.V.: Effective implementation of the parallelization on reconfigurable computer systems. Vestnik komp'iuternykh i informatsionnykh tekhnologii (Herald of computer and information technologies), 11–16 (2018). https://doi.org/10.14489/vkit.2018.08.pp.011-016

11. Levin, I.I., Dordopulo, A.I., Fedorov, A.M., Gulenok, A.A.: Reconfigurable computer based on Virtex UltraScale+ FPGAs with immersion cooling system. In: International Scientific Conference, Parallel Computational Technologies, PaCT-2017. Proceedings. pp. 27–41. SUSU Publishing Center, Chelyabinsk (2017)

12. Kalyaev, I.A., Dordopulo, A.I., Levin, I.I., Fedorov, A.M.: Evolution domestic of multichip reconfigurable computer systems: from air to liquid cooling: from air to liquid cooling. Trudy SPIIRAN 1(50), 5–31 (2017). https://doi.org/10.15622/sp.50.1

13. Levin, I.I., Doronchenko, Yu.I., Dordopulo, A.I., Levina, M.G.: High-performance reconfigurable computing system based on FPGA XILINX ULTRASCALE+. In: International Conference, XXII Kharitonov thematic scientific readings. Collection of scientific papers. pp. 284–296. FSUE "RFNC-VNIIEF", Sarov (2022).

14. Levin, I., Dordopulo, A., Fedorov, A., Doronchenko, Yu.: High-Performance Reconfigurable Computer Systems with Immersion Cooling. In: Sokolinsky, L., Zymbler, M. (eds.) Parallel Computational Technologies. PCT 2018. Communications in Computer and Information Science, vol. 910, pp. 62–76. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99673-8_5

15. Cerebras. https://www.cerebras.net/product-system/, accessed: 2023-04-20

16. NVIDIA Data Center Deep Learning Product Performance (2023). https://developer.nvidia.com/deep-learning-performance-training-inference, accessed: 2023-04-20

17. Dell Precision Rack 7910 Owner's Manual. https://www.karma-group.ru/upload/iblock/d16/precision-r7910-manual.62BF24D6F1B748C9BB6DA43755132D49.pdf, accessed: 2023-04-21

18. SPEC CPU2017 Floating Point Rate Results (2017). https://spec.org/cpu2017/results/res2017q3/cpu2017-20170828-00069.pdf, accessed: 2023-04-21