

Specialized Hybrid Chinese Microprocessors for Solving Important Problems of Fundamental Medicine and National Defense

Andrey Molyakov*

Institute of information technologies and cybersecurity, Russian State University for the Humanities, Russia

***Corresponding author:** Andrey Molyakov, Institute of information technologies and cybersecurity, Russian State University for the Humanities, Russia



ARTICLE INFO

Received: 📅 March 08, 2021

Published: 📅 March 16, 2021

Citation: Andrey Molyakov. Specialized Hybrid Chinese Microprocessors for Solving Important Problems of Fundamental Medicine and National Defense. Biomed J Sci & Tech Res 34(4)-2021. BJSTR. MS.ID.005573.

ABSTRACT

In this paper author describes specialized hybrid Chinese microprocessors for solving important problems of fundamental medicine and national defense. CT-2 and CT-3 (codenamed "Thunderbolt") were developed by the University of Defense Technologies (NUDT) and built on hybrid multi-core massive multi-threaded microprocessors. Chinese hybrid microprocessors of this type are not only the basic building block of Petaflops supercomputer, but also a microprocessor for strategic tasks: national defense, fundamental medicine, molecular biology, genetic engineering and etc.

Keywords: NUDT; Codename "Thunderbolt"; Specialized Hybrid Microprocessors; ParallaX

Introduction

For the beginning we should talk about the directions in the organization of a higher level of the hierarchy of building a computing node, new strategic chinese microprocessors, as well as about future design solutions for 3D/4D assembly of nodes, the technologies of their cooling. It is obvious that there are many improvements in hardware characteristics that would not have been possible without using of complex system-level algorithms, as well as progressive application-level algorithms. This happened as a result of the joint work of system and application programmers with hardware developers – a worthy example of imitation in the organization of work on high-speed electronic components. This phenomenon will be typical for the next ten years due to the stagnation in the development of silicon technologies and, as a result, the need for optimizations at the expense of everything that is possible. The main problems are emerging now: the presence of a dangerous "memory wall", when large delays in the execution of memory accesses lead to downtime of functional devices due to the lack of data for them; the presence of a cumulative impact on the scalability of performance overhead for communication and nonlinear changes in processor performance due to changes in

the efficiency of using the cache memory during splitting due to the parallelization of the working data set of the task into subsets falling on the processors; the tendency of promising parallel computing models to work on shared memory, which sharply reduces communication losses. Next, an approach was chosen with using of processors with a multi-threaded architecture. Chinese hybrid microprocessors of this type are not only the basic building block of Petaflops supercomputer, but also a microprocessor for strategic tasks: fundamental medicine, molecular biology, genetic engineering and etc. They will also be used in the defense industry [1,2]. CT-2 and CT-3 (codenamed "Thunderbolt") were developed by the University of Defense Technologies (NUDT) and built on hybrid multi-core massive multi-threaded microprocessors.

Microprocessor CT-2

We should note that the ideas of the American project ParallaX have been chosen as the architectural basis for the next level of the hierarchy in relation to the strategic chinese microprocessors. This supercomputer contains 128K heterogeneous multithreaded streaming microprocessors (dodecatrons) and mass memory modules, which communicate via a WDM optical network such as

DataVortex. Each dodecatron contains a decisive field of 256 ALUs (32 GHz clock rate, 8 TFlops). Along the perimeter of the decisive field there are 12 multi-threaded cores with memory and network interfaces (clock frequency 16 GHz, thousands of threads, peak performance 4 Teraflops, operations on short vectors, memory available in each core - 512 MB [2,3]. This project has a rich and long history, currently eminent participants. The elements of the architecture proposed in the project meet the most modern requirements and have great prospects. According to the modern distribution of work under the UHPC program, this project is attributed to the work of the Sandia Laboratory, corresponds to the views on promising supercomputers of the specialists of this laboratory, and the main ideologist P.M. Kogge took and still takes an active part in this project [4,5]. A prototype of the CT-2 computing module has been created (Figure 1), which implements

such an architectural approach on a board. There are 32 sockets on the board for 16-core multi-threaded CT-2 microprocessors [6]. Thus, if we assume that there are 256 hardware threads in one core, then the parallelism level of this board is $32 \times 16 \times 256$, i.e., 131072 thread parallel processes. The mass-multithreaded hybrid microprocessor CT-2 is figuratively characterized by three hieroglyphs depicted on the microcircuit case – “tiger” (computing power), “hawk” (swiftness, the presence of many asynchronous threads), “dragon” (symbolizes the breakthrough nature of this microprocessor, dragon – symbol of progress). The comments indicate the nature of the refinement of the microprocessor (why the association with the tiger) – the introduction of powerful SIMD operations on short vectors, as well as the enhancement of computational capabilities, as is done in graphics processing units (GPUs).

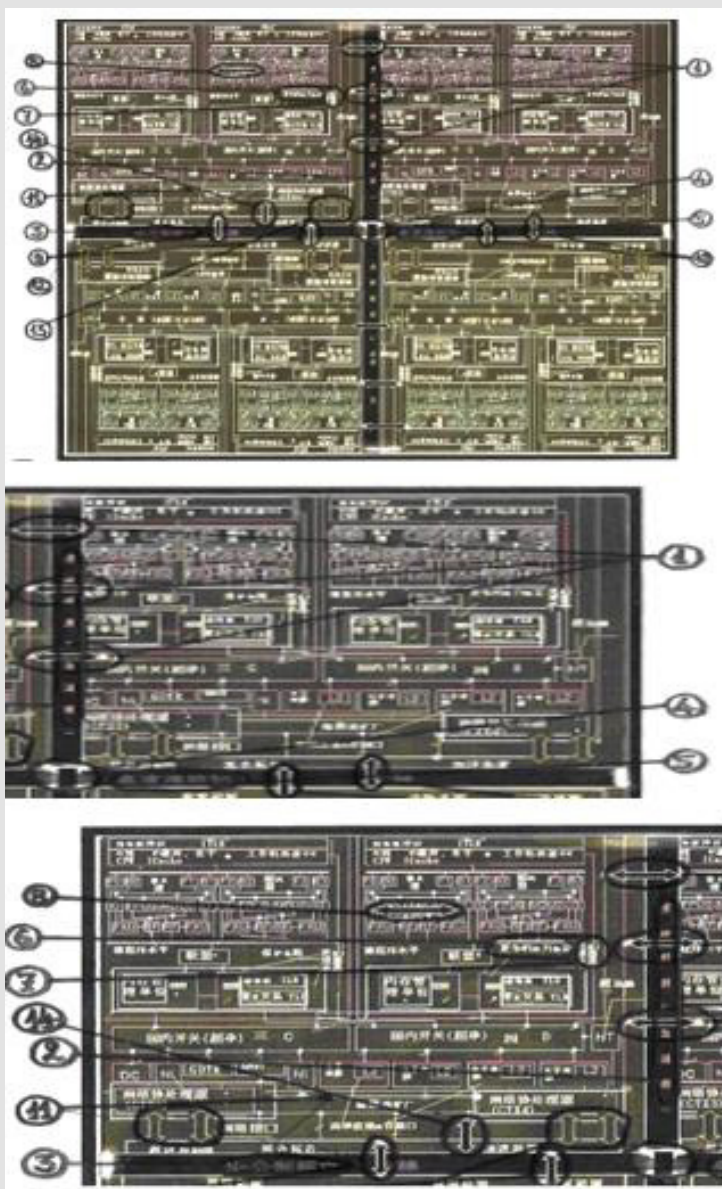


Figure 1: General view of the motherboard (CT-2 project).

Comments to Figure 1 (Fundamentally New Architectural Solutions):

- a) Means a direct connection of the microprocessor crystals at three levels, can be used to increase the nuclearity of the microprocessor, as well as to increase fault tolerance (duplication of resources involved in the execution of the program and control of the verification of the results they produce)
- b) High-speed substrate links made using nanotechnology, to provide direct connection of microprocessor crystals
- c) High-speed substrate links that are used to implement N-dimensional network of supercomputer modules and microprocessors inside modules, also used to implement fast connections of microprocessors single substrate (see comments 9 and 10), the N-dimensional network most likely means not an N-torus, but a multi-connected network of the N-cube type or a Clos network
- d) This is a transition to the next level of 3D assembly, a communication chip with pinless connections like technology, developed several years ago as part of the project "Hero" of the DARPA HPCS project
- e) Means that information is transmitted with a strong compaction and on high frequency
- f) Indicates that it is possible to set the data copying permission mask, for some data segments, you can specify that their data is not needed cache
- g) Colored lines represent intra-crystal bond lines
- h) Means that changes are possible in the composition of blocks and functional devices
- i) It is indicated that transfers between microprocessors of the same substrate can be accelerated in a special way and implemented considering providing fault tolerance (see comment a)
- j) Channel of direct connection of adapters / serializers / deserializers of packets, respectively, intermodular network (CTX3) and off-chip memory (CTX4)
- k) Links to the communication network of supercomputer modules, one of which is connected to the NI microprocessor unit (for example, for accessing fragments of non-cached data segments), additional broadcast in the MMU is not required, you can directly from the CTX3 network adapter send a request to the CTX4 off-chip adapter
- l) Links to output to off-chip DRAM memory, which are located on one another or more other 3D assembly underlays
- m) Changes in the communication subsystem are possible.

Chinese scientists are deploying experimental benches to study nanophotonics. Functional devices for performing operations on sparse matrices is a very urgent topic of research purposes and development, such devices are implemented in processor projects for working with graphs of a new generation. They are also important for a new generation of calculating tools for performing scientific and technical calculations, for example, on irregular dynamically changing grids and on-chip networks with different topologies. Further development of the CT-2 / CT-3 project codenamed Fenghuang. It represents PIM (Processor-In-Memory) with a 3D assembly technologies for manufacturing HBM memory modules with increased bandwidth, as well as PIM technologies of hybrid 3D assemblies with HBM memory chips and processors, technologies of high-speed and energy efficient SER/ DES transceivers for the implementation of inter-crystal link, based on alternative branches of the Data-Intensive Architecture (DIVA) systems. TSV connections will be nanophotonic. In the crystal plane of this assembly the compounds will be also nanophotonic. Explanation: Launched in 2019, a new "breakthrough" technologies project for PRC military intelligence is so named to denote a new area of research and development, along with the creation of massive multi-threaded and massive parallels systems with traditional links and pin contact pads. Fenghuang are mythological birds found in Sinospheric mythology that reign over all other birds. The males were originally called feng and the females huang but such a distinction of gender is often no longer made and they are blurred into a single feminine entity so that the bird can be paired with the Red Dragon, which is traditionally deemed male.

The next implementation of CT-2 is made using 22 nm technology, the number of cores will already be 48, i.e., the parallelism will increase to 393216, which is approximately 4×10^5 . The level of parallelism of an Exascale system should be about 109, then $10^4 - 10^5$ of such cards will be required. We should expect a sharp miniaturization, which is confirmed by research and plans for new constructs, discussed below. Initially, the massively multi-threaded CT-2 microprocessor with asynchronous threads for information systems has also become hybrid, it has enhanced the power of numerical processing – SIMD operations on short vectors have been introduced, as well as elements of modern graphics processors in the form of synchronous threads. A prototype of such a microprocessor modified in architecture was made at the Taiwanese TSMC factory using 22-45 nm technology, the number of cores (such as the microprocessor J7) on one chip is 12, each core should have 256 hardware threads. The hybridity in microprocessor CT-2 allows its use for a wider class of tasks. There is information that by changing the assembly of computing nodes (increasing the share of CT-2 microprocessors and removing the network data receiving and output units), it is possible to create a supercomputer for scientific and technical calculations with real performance on scientific and technical problems at the Exascale level [7].

Microprocessor CT-3

The new processor line is complemented by one more hieroglyph “Typhoon” (means a revolutionary approach in the field of design and assembly). The CT-3 project designs such motherboards on a ceramic substrate or crystal. On Figure 2 you can see motherboard and contact pads. Figure 3 is a side view of a cryogenic solution chamber in which 3D (in future 4D) composite board assemblies are immersed. Each section is a substrate with different VLSIs (or

simply with one VLSI, the size of the substrate is apparently about 40x40 mm), which has a certain functional orientation. In the comments to the figure, options for this orientation are indicated. Such substrates with VLSI are fixed on one and the other side of the composite board and also in the plane along the side contact pads. Contact pads have the following features (innovative assembly technology) there are 6 types of sections A, B, C, D, E, F (see Figure 3)

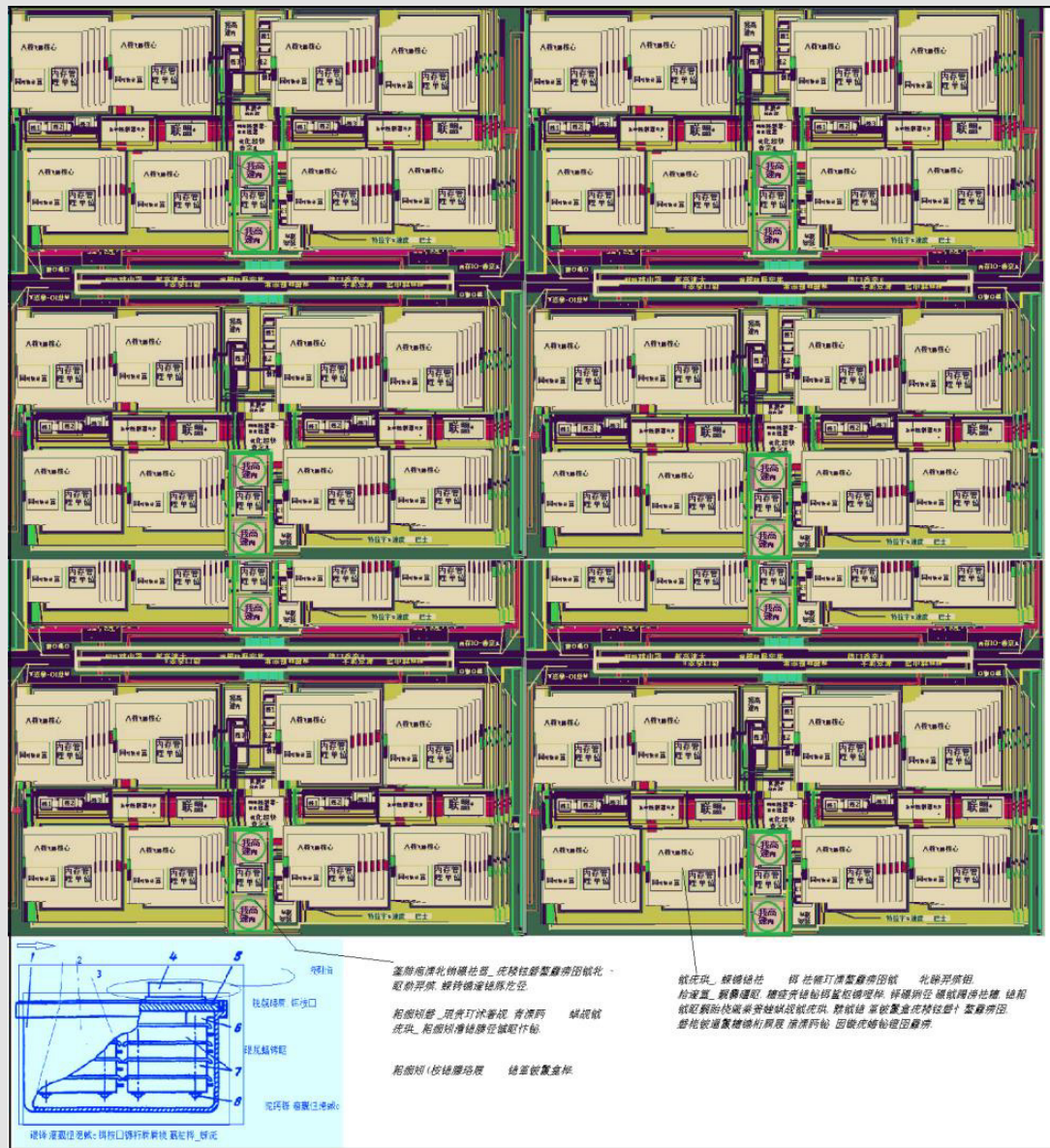


Figure 2: General view of the motherboard (CT-3 project).

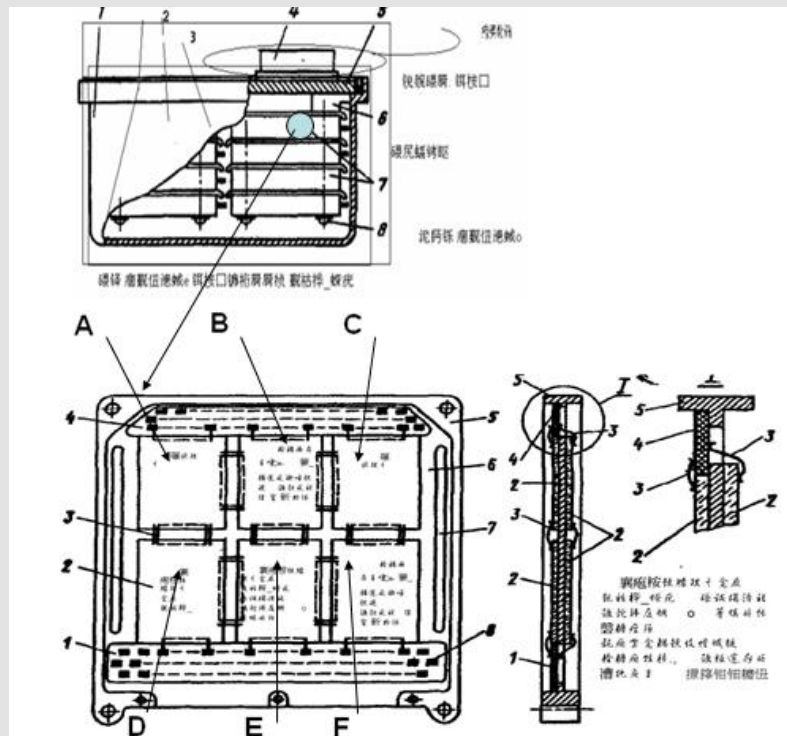


Figure 3: 3D assembly of composite boards in cryogenic liquid (side view).

Comments to Figure 3:

- a) 1 - framework;
- b) 2, 3 - sections filled with cryogenic solution;
- c) 4 - surge-holder;
- d) 5 - lid for cooling (general view);
- e) 6, 7 - 3D assembly composite boards;
- f) 8- clamp and fastening.

- a) A-General Purpose Data Processor (this includes MIPS, Power PC, Intel, AMD)
- b) B-General purpose processors: multi-threaded - for work with memory (FT-1000, FT-1500), streaming - for data processing (Tilera, Monarch)
- c) C-General purpose graphics processors (including GPU)
- d) Note: Zones D, E, F are used to install CT-2 class processors
- e) D-massively multi-threaded processors for working with memory
- f) E-Massively multithread processors for data processing
- g) F-Massively multi-threaded processors for networking
- h) 3-Pinless connections of processor modules Figures 1-3.

A set of substrates with VLSI are spaced between the two halves of the composite board, but in one row there are not three, but four substrates with VLSI. The results obtained in the development

of basic microprocessors allow China to reach a new level in the field of onboard systems for unmanned aerospace vehicles, as well as other autonomous systems and communications. There is even an opinion among chinese specialists that the appearance of these microprocessors makes it unnecessary to use FPGAs with their obvious modern advantage in the implementation of streaming computing schemes with data management, but also with obvious disadvantages in terms of programming complexity and cost, energy consumption. It is also known that the developed basic hybrid microprocessors allow software emulation of the PowerPC, X86 and Monarch microprocessor instruction systems (this microprocessor includes a decisive field of inhomogeneous elements and universal RISC-type processor cores), which ensures the portability of previously created complexes. Moreover, the specified project implements in the future the development of chinese mass-parallel processor on the base of multithread core with specialized accelerators. Currently, additional sections on the motherboard are implemented to install the following processors: GPU Volta, GPU Ampere, Colossus, PEZY-SC2 [9,12].

Conclusion

In-depth analysis of the reasons for this loss of leadership in the HPC field, carried out by American specialists, showed that the technology of cluster assembly from commercially available components is not entirely sufficient. This conclusion led in the United States to a number of actions at the state level, including the revival of deep architectural research and development, and the restoration of schools of computer architecture in American universities. Supercomputers CT-2/ CT-3 with several levels of hierarchy, high parallelism and globally addressable memory require powerful software with dynamic parallelization of programs, optimization tools that consider the heterogeneity of access to globally addressable memory. It is clear that the issues of parallelization and memory optimization should be hidden from the end user as much as possible. According to the available data, the middle level of the system software is being worked out now (the approaches GASNet (active messages), ARMCI (message aggregation), Charm ++ (object-oriented computing models with message transfer control), transactional memory (process localization) are being analyzed and modified. over memory and saving locks). At present, silicon (CMOS) technology of 22 nm is industrially developed, but only by four companies in the world - Intel, TSMC, Samsung and STMicroelectronics. Intel predicts the 5 nm level will be reached in 2020.

It is believed that further miniaturization is impossible, and Moore's law will cease to work no later than 2024 [13,14]. Effective using of new "revolution" technologies in the near future and until the end of Moore's law is associated with various optimization techniques: architecture (specialized processors and accelerators), microarchitecture; connections on-chip and between crystals; constructs in the form of 3D assembly of modules and 3D VLSI. Work of this type is underway and is considered as the main one in the creation of exascale supercomputers, but they are associated with overcoming too many problems. New solutions in the field of element-design base (technologies of the post-Moore Era) can facilitate the creation of Exaflops supercomputers and become the basis for creating supercomputers of the next performance

levels. The main tasks in the development of new options for the element base: increase the frequency of operation, reduce energy consumption, and achieve a high level of integration. Modern microprocessors operate at a frequency of 3-6 GHz. Devices based on Superconducting Technologies (RSFQ) can operate at a frequency of several hundred GHz; publications are known for samples operating at frequencies of 20 GHz and 80 GHz. Devices based on Quantum Cellular Automata (QCA) can operate at frequencies on the order of THz.

References

1. A Molyakov (2019) China Net: Military and Special Supercomputer Centers. Journal of Electrical and Electronic Engineering 7(4): 95-100.
2. Sterling T (2011) In Search of Exascale Roadmap.
3. T Sterling (2007) Architecture Paths to Exaflops Computing Is Multicore the next Moore's Law? What about the Memory?
4. A Molyakov (2019) Age of Great Chinese Dragon: Supercomputer Centers and High Performance Computing. Journal of Electrical and Electronic Engineering 7(4): 87-94.
5. A Molyakov (2019) Analysis of World Experience in Creating Parallel Computing Systems Designed to Effectively Solve DIS-tasks. Journal of Electrical and Electronic Engineering 7(4): 101-106.
6. XJ Yang, Y Dou, QF Hu (2006) Progress and Challenges in High Performance Computer Technology. Journal of Computer Science & Technology 21: N5.
7. PM Kogge (2007) Computer systems with lightweight multi-threaded architectures. United States Patent Application Publication US p. 22.
8. M Anderson, M Brodowicz, H Kaiser, T Sterling (2010) The ParalleX Model for Distributed Adaptive Mesh Refinement p. 21.
9. Durant L (2017) Inside Volta: The World's Most Advanced Data Center GPU.
10. Jia Z, Maggioni M, Staiger B, Scarpazza DP (2018) Dissecting the NVIDIA Volta GPU Architecture via Microbenchmarking. Technical report.
11. Jia Z, Tillman B, Maggioni M, Scarpazza DP (2019) Dissecting the Graphcore IPU Architecture via Microbenchmarking. Technical report.
12. PEZY-SC2-PEZY// <https://en.wikichip.org/wiki/pezy/pezy-scx/pezy-sc2>
13. Adamov A (2019) Main problem directions in the field of domestic element base of supercomputers. Cybersecurity Issues 4: 2-12.
14. Weiwu Hu (2010) A Multicore Processor Designed for PetaFLOPS Computation.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2021.34.005573

Andrey Molyakov. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>