Yamamoto Sano Laboratory, Graduate School of Information Sciences, Tohoku University

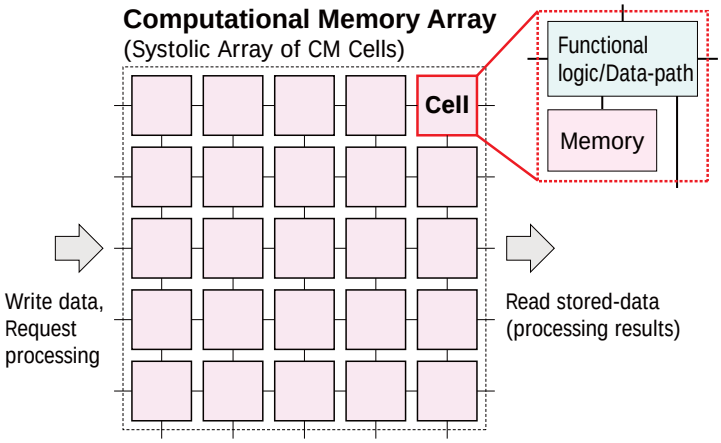
Custom Computing Machine for Computational Fluid Dynamics

Much more powerful computing resources are required in scientific computing fields including computational fluid dynamics. For such demand, we typically utilize supercomputers based on massively parallel processing with general-purpose processors. In General, parallel processing with too many processors tends to be very inefficient and saturate performance. This is because of the bottleneck in an interconnection network for data-communication/synchronization among processors. In order to efficiently use the power of given processors, the desired performance should be achieved with a small number of processors.

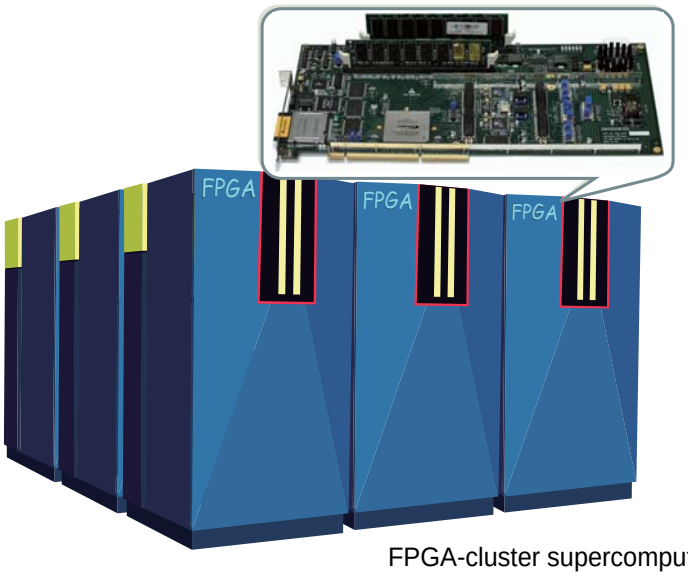
For the desired performance with a small number of processors, each processor has to be sufficiently powerful. However, recent general-purpose microprocessors suffer from increasing their performance with increased cost/resources. So far, performance improvement of microprocessors has been lead by 1) higher operation frequency, 2) higher integration of transistors and 3) advancement in architecture. 1) and 2) have been so far achieved by scaling in the semiconductor technology. However, it is getting difficult to further increase the operation frequency because of the wire-delay and the power density problems in the deep submicron semiconductor technologies. Moreover, insufficient parallelism in program codes prevents much more transistors integrated by the technology scaling from being efficiently utilized for higher performance. For 3), many architectures have been proposed to improve the memory-processor bottleneck. However, parallelism extraction from sequences of instructions, which is the key technology in increasing performance of a single-core microprocessor, is limited by their complexity.

Under these circumstances, the custom computing machines for specific computational applications are attractive because they map computing algorithms directly to circuits efficiently utilizing available on-chip hardware resources.

Especially, in computational dynamics where governing equations are numerically solved by the difference schemes, each computation is done with local data at each location and previous computing results of their neighborhoods. In our research project, we propose the systolic computational memory architecture, which fits such kind of computation. This architecture is composed of an systolic array of many cells for scalably and massively parallel computation. Each cell has a local memory tightly coupled with simple computing circuits to avoid the memory-processor bottleneck. The more cells are available, the higher parallel processing performance is obtained with sustained memory bandwidth per computing circuit. The array itself also behaves as a memory with computing function, which performs computations with written data independently of a main processor.



Systolic computational memory architecture



FPGA-cluster supercomputer

For real supercomputing on this architecture, we have to realize fine-grain parallelism in algorithms of target applications and map them to the systolic array. We focus on the difference schemes in CFD applications. We investigate the comptational depencences among grid points and the locality of computed data at each point, and then propose a parallel algorithm on an array. After deciding specifications for a cell; computing functions and data, we design hardware components of a cell for memory, computation and communication in an array.

Although theare are several ways to implement a custom computing machine as custom-VLSI and ASIC, we are going to implement a prototype system with an FPGA(field-programmable gate array). In the last decade many researchers have reported competitive or more performance of floating point computation on large-scale and leading-edge FPGAs. In addition, the distributed structure of logic blocks and embedded DSPs and SRAMs in recent FPGAs is very suitable for inplementing the systolic computational memory architecture. As future work, we are planning to build an FPGA-cluster supercomputer composed of parallel computers with FPGA-based accelerators for Petaflops supercomputing.