Stratix10 FPGA Cluster as Off-loaded Custom Computing Engine for Supercomputers

Kentaro Sano

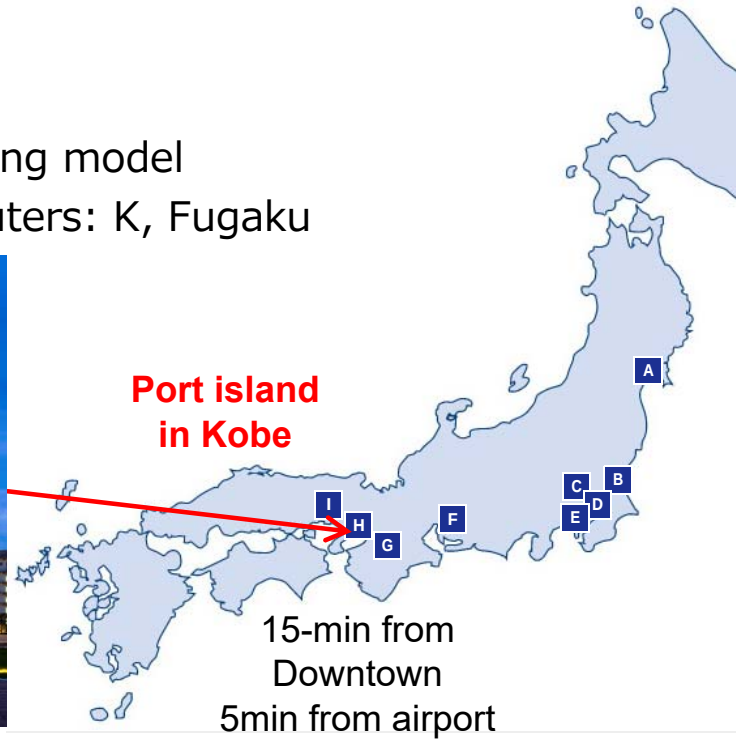
Processor Research Team, R-CCS Riken

June 20, 2019

Riken Center for Computational Science

Processor Research Team (2017.4~)

- ✓ Future HPC architectures
  - ➢ *Spatial custom computing*
  - ➢ *Data flow*
- ✓ Highly-productive programming model
- ✓ Advanced use of Supercomputers: K, Fugaku



Outline

- **Introduction**
  - ✓ Why spatial custom computing with FPGAs?
- **Architecture for spatial custom computing**
- **Feasibility study for extension of an existing machine**
- **Preliminary results**
- **Summary**

Stratix10 FPGA board (PAC)

Introduction

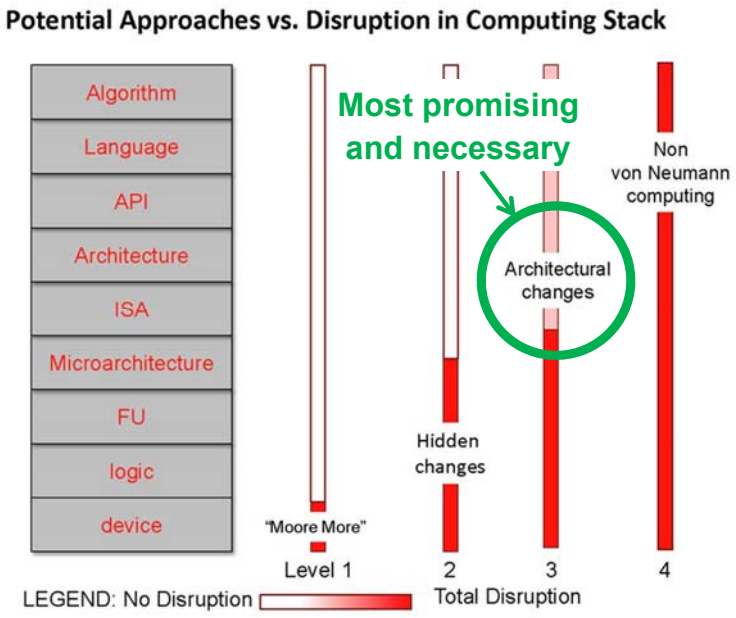
Need to change architectures for post-Moore era

- ✓ Re-consider "computing"
- ✓ We'll hit a barrier relying on inappropriate architectures.

What is fundamental but possible change?

- ✓ Dark silicon (power / Tr. not decrease)
- ✓ Limited integration at increasing cost / Tr.

Should we really change many-core architecture?



IEEE Rebooting Computing

[REF] http://materials.dagstuhl.de/files/17/17061/17061.ErikP.DeBenedictis1.Slides.pdf

Weakness of Many von-Neumann

Inefficient utilization of Tr.

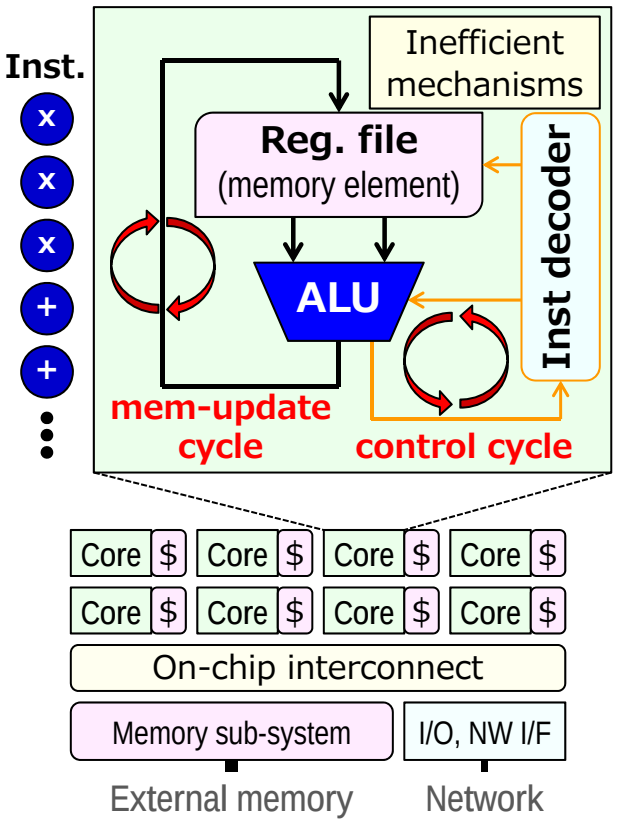
- ✓ Mechanisms to boost IPC
- ✓ But, no more transistors & no more power budget on a chip

Latency-sensitive characteristic

- ✓ von Neumann with + memory-update cycle + control cycle

Inefficient data-movement among cores

- ✓ Read and write in memory hierarchy (scratch pad / cache)



Many-core von-Neumann archi.

How Should We Change?

Inefficient utilization of Tr.

- ✓ Mechanisms to boost IPC
- ✓ But, no more transistors & no more power budget on a chip

More efficient use of transistor & switching for computation

Latency-sensitive characteristic

- ✓ von Neumann with + memory-update cycle + control cycle

Latency-tolerant architecture w/o cycles

Inefficient data-movement among cores

- ✓ Read and write in memory hierarchy (scratch pad / cache)

Data-movement w/o memory access

Answer: Spatial Custom Computing

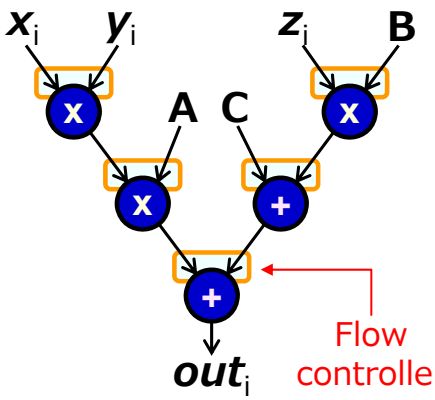
Customization

✓ Reconfigurable computing (with FPGAs)

More efficient use of transistor & switching for computation

Spatial compt. w/ Data-flow

✓ Flow instead of cycles

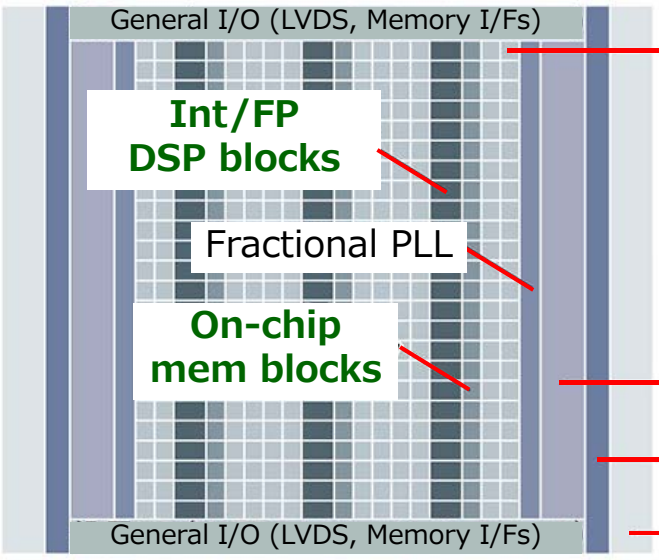


Latency-tolerant architecture w/o cycles

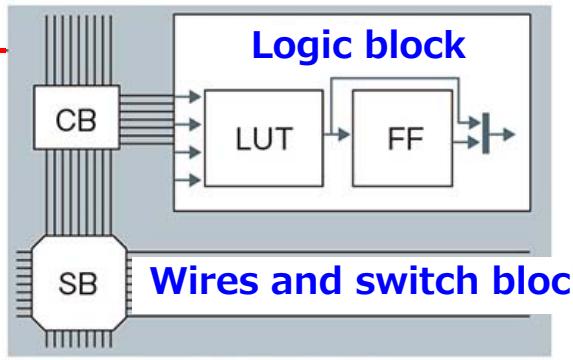
Data-movement w/o memory access

We consider data-movement first.

FPGA (Field-Programmable Gate Array) As a Platform for Custom Computing



FPGA



High-speed I/O (PCI-Express, 100G Ethernet Interlaken, serial transceiver)

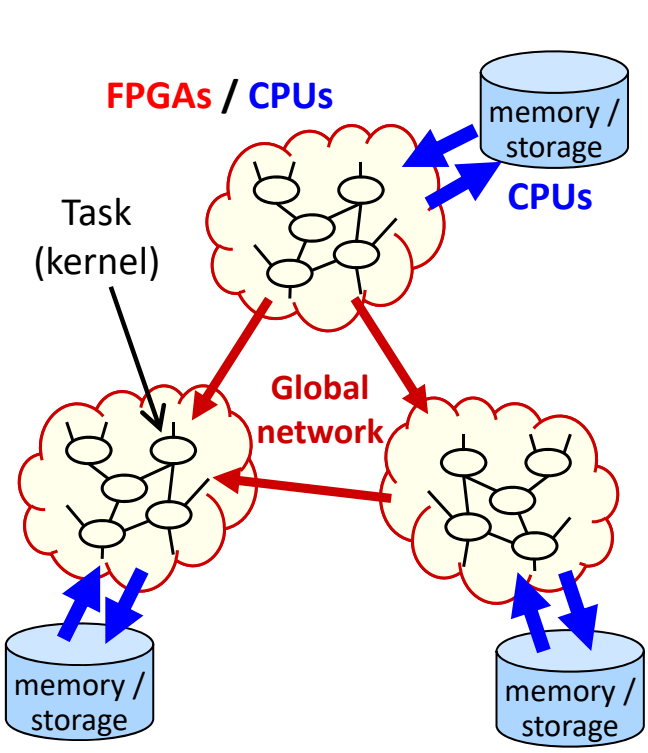
Memory I/F (DDR4, HBM, etc.)

Big array of floating-point DSPs, memories, and high-speed I/Os, rather than big array of logics

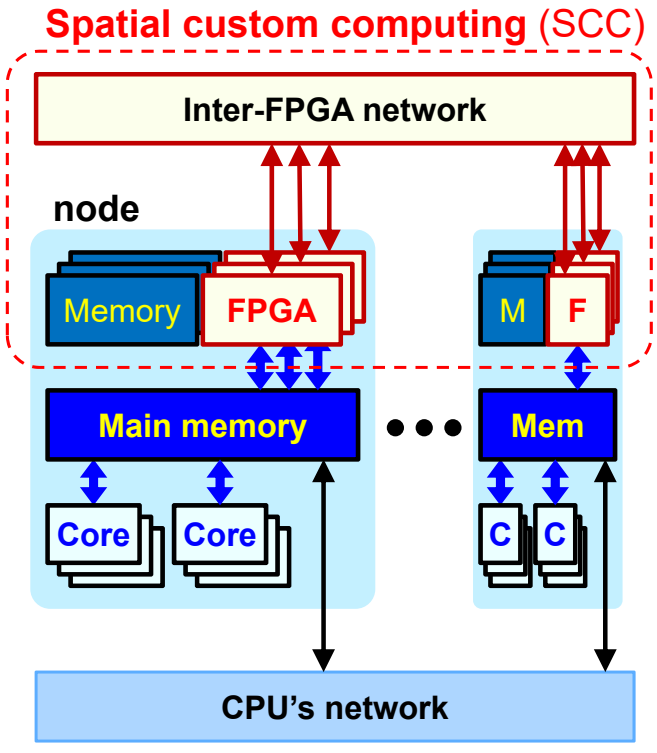
Outline

- Introduction
  - ✓ Why spatial custom computing (with FPGAs)?
- **Architecture for spatial custom computing**
- Feasibility study for extension of an existing machine
- Preliminary results
- Summary

Architecture for Spatial Custom Computing



System-wide spatial custom computing
(stream data through custom data-paths)



Architecture example
(CPUs + FPGAs)

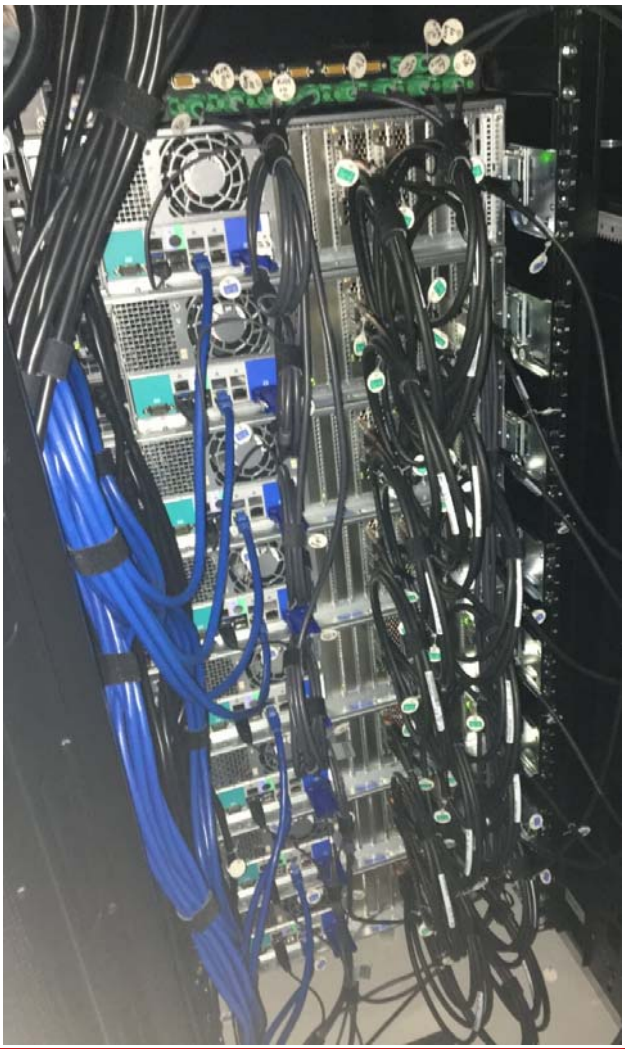
Outline

- Introduction
  - ✓ Why spatial custom computing (with FPGAs)?
- Architecture for spatial custom computing
- **Feasibility study for extension of an existing machine**
- Preliminary results
- Summary

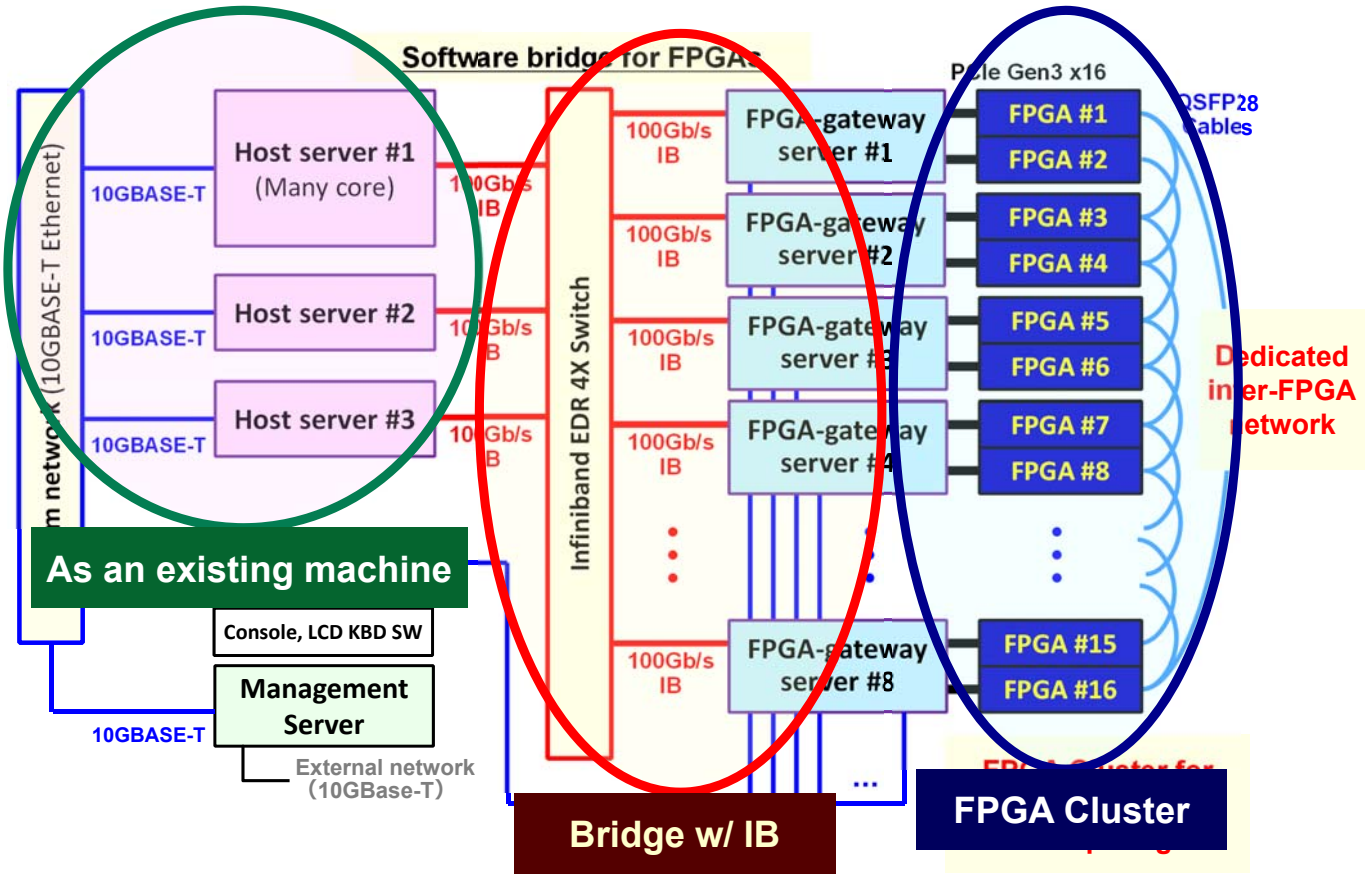
Proof of Concept

Feasibility study with
experimental prototype system

- ✓ How can we extend existing gen-purpose machines with FPGAs?
- ✓ How should we program custom data-paths on FPGAs?
- ✓ Eco system?



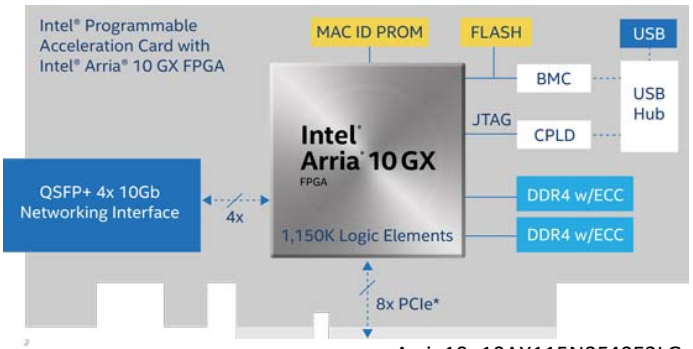
Experimental Prototype System



PAC : Programmable Acceleration Card

Rush Creek (preliminary eval.)

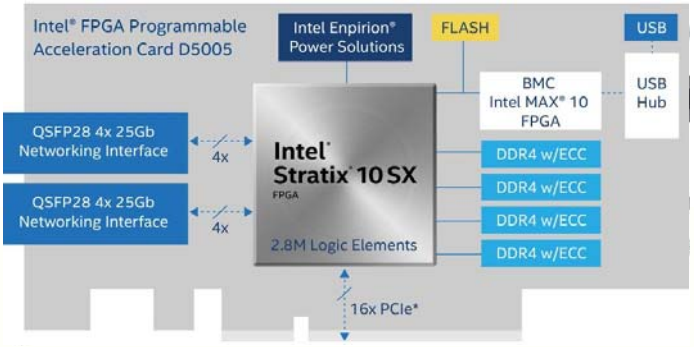
- ✓ **Arria10 FPGA** (20nm)
  - ➢ 1150K LEs, 53 Mb BRAMs
  - ➢ 1518 FP DSPs (1.5 TF in SP)
- ✓ 8GB DDR4 x 2ch
- ✓ PCIe Gen3 x8
- ✓ 1x QSFP+ (40Gb/s)



Arria10: 10AX115N2F40E2LG

Darby Creek (prototype system)

- ✓ **Stratix10 FPGA** (14nm)
  - ➢ 2753K LEs, 229 Mb BRAMs
  - ➢ 5760 FP DSPs (10 TF in SP)
- ✓ 8GB DDR4 x 4ch
- ✓ PCIe Gen3 x16
- ✓ 2x QSFP28 (100Gb/s)
- ✓ ARM Cortex-A53 1.5 GHz



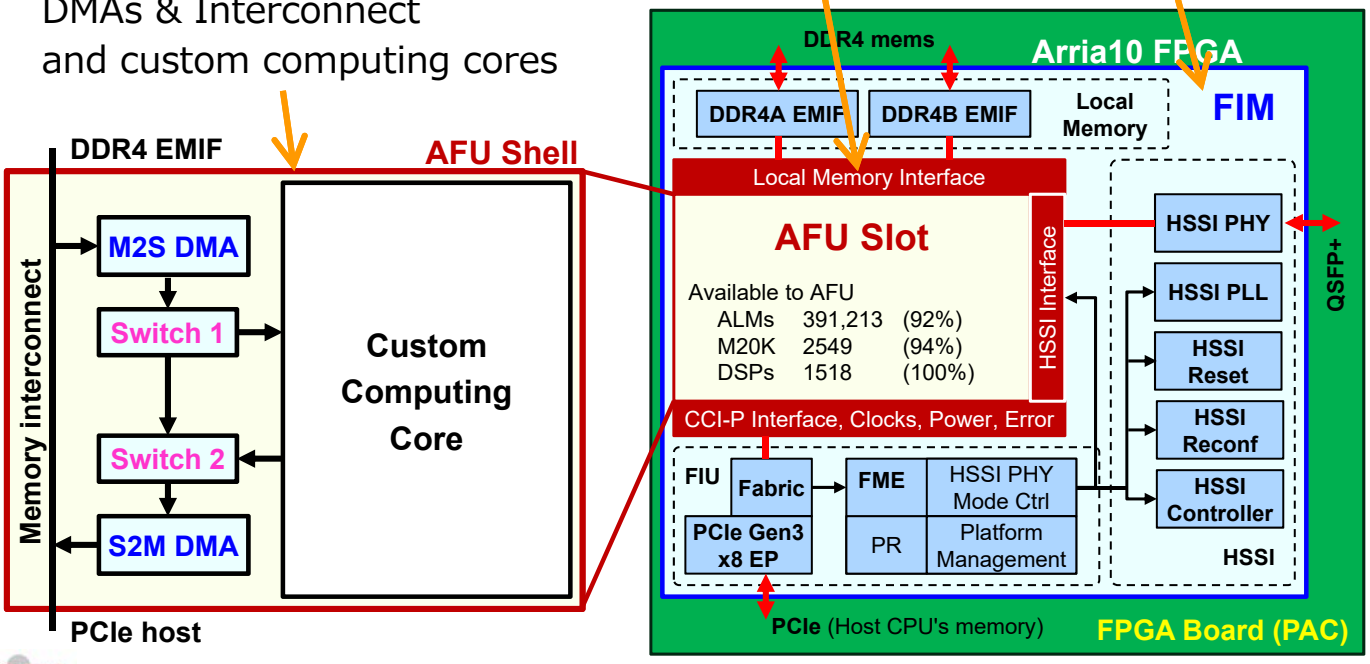
Stratix10: 1SX280HN2F43E2VG

FPGA Shell and User Logic on FPGA

FIM (FPGA Interface Manager) : **HW shell made by Intel**

AFU (Acceleration Function Unit) : **Reconfigurable logic region**

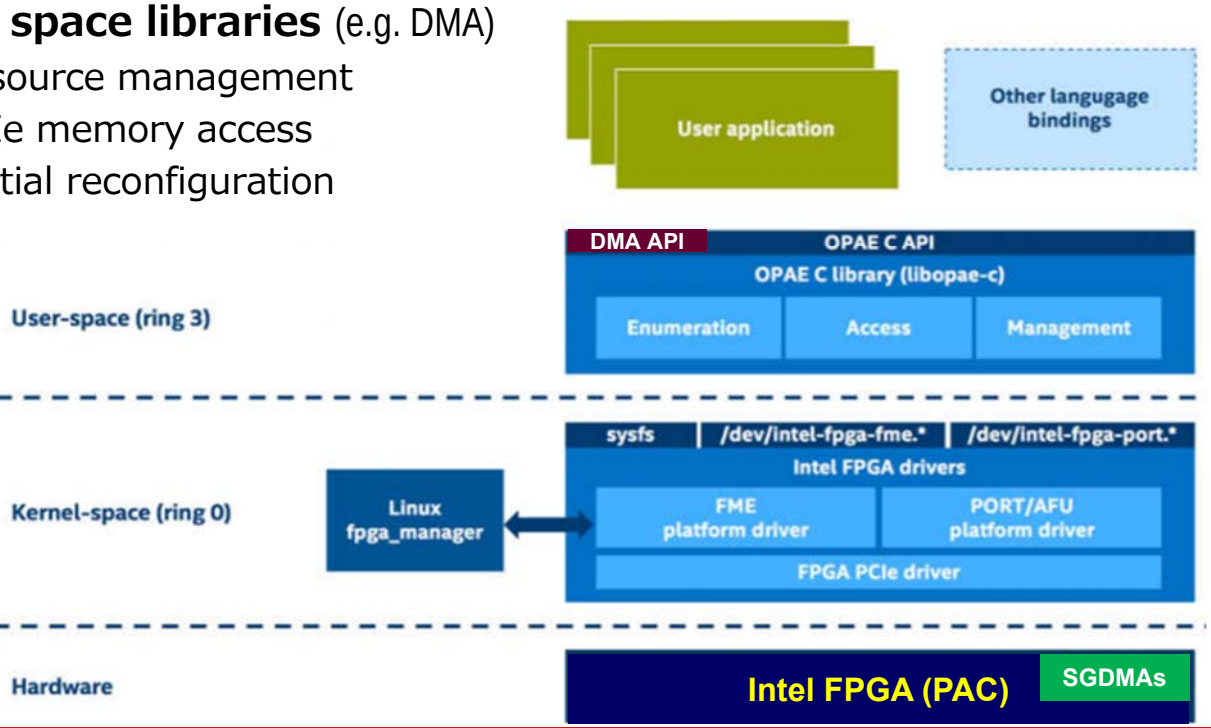
AFU Shell : **Our own HW shell**
DMAs & Interconnect
and custom computing cores



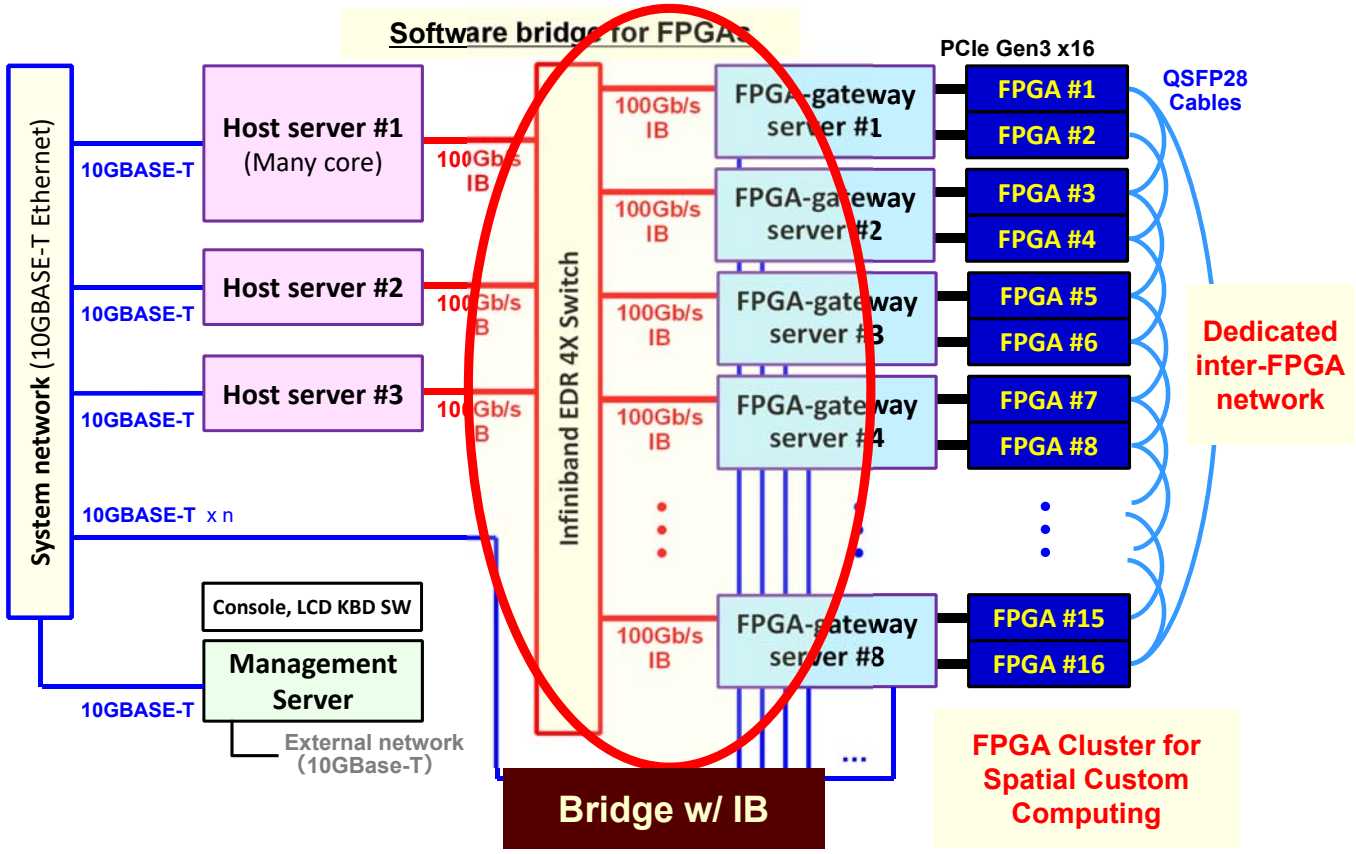
Intel OPAE (Open Programmable Accel. Engine)

Light-weight software interface for PAC

- ✓ **PCIe driver, APIs,**
  User space libraries (e.g. DMA)
  - ➢ Resource management
  - ➢ PCIe memory access
  - ➢ Partial reconfiguration



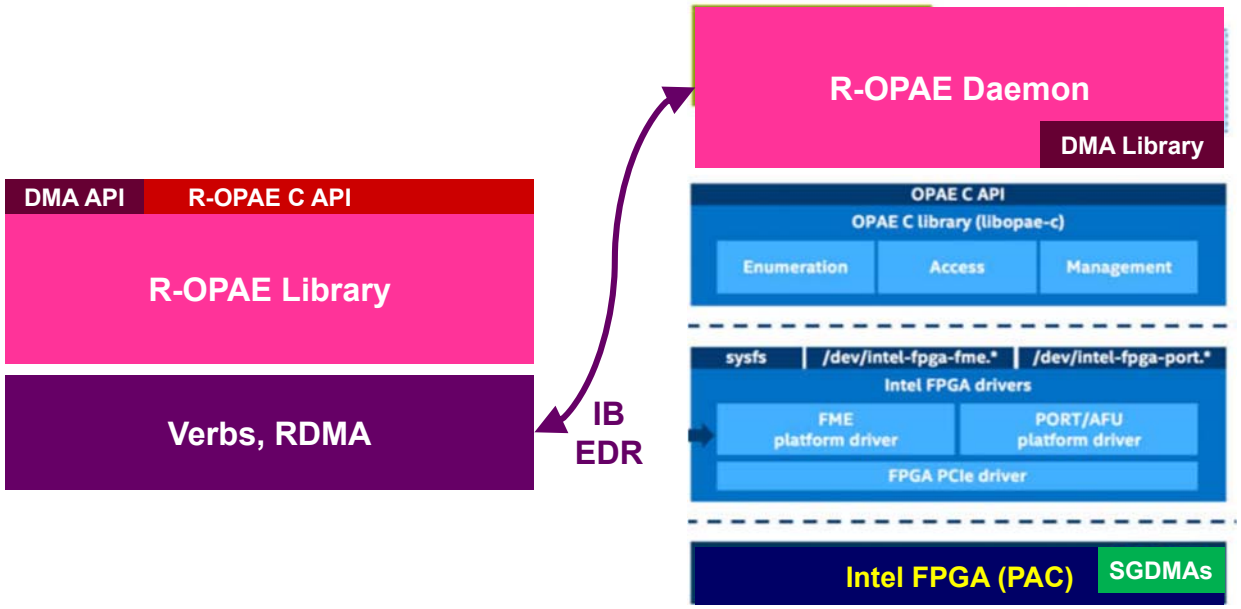
Experimental Prototype System



Remote-OPAE (developed by us)

Software bridge to connect hosts and FPGAs by IB (EDR)

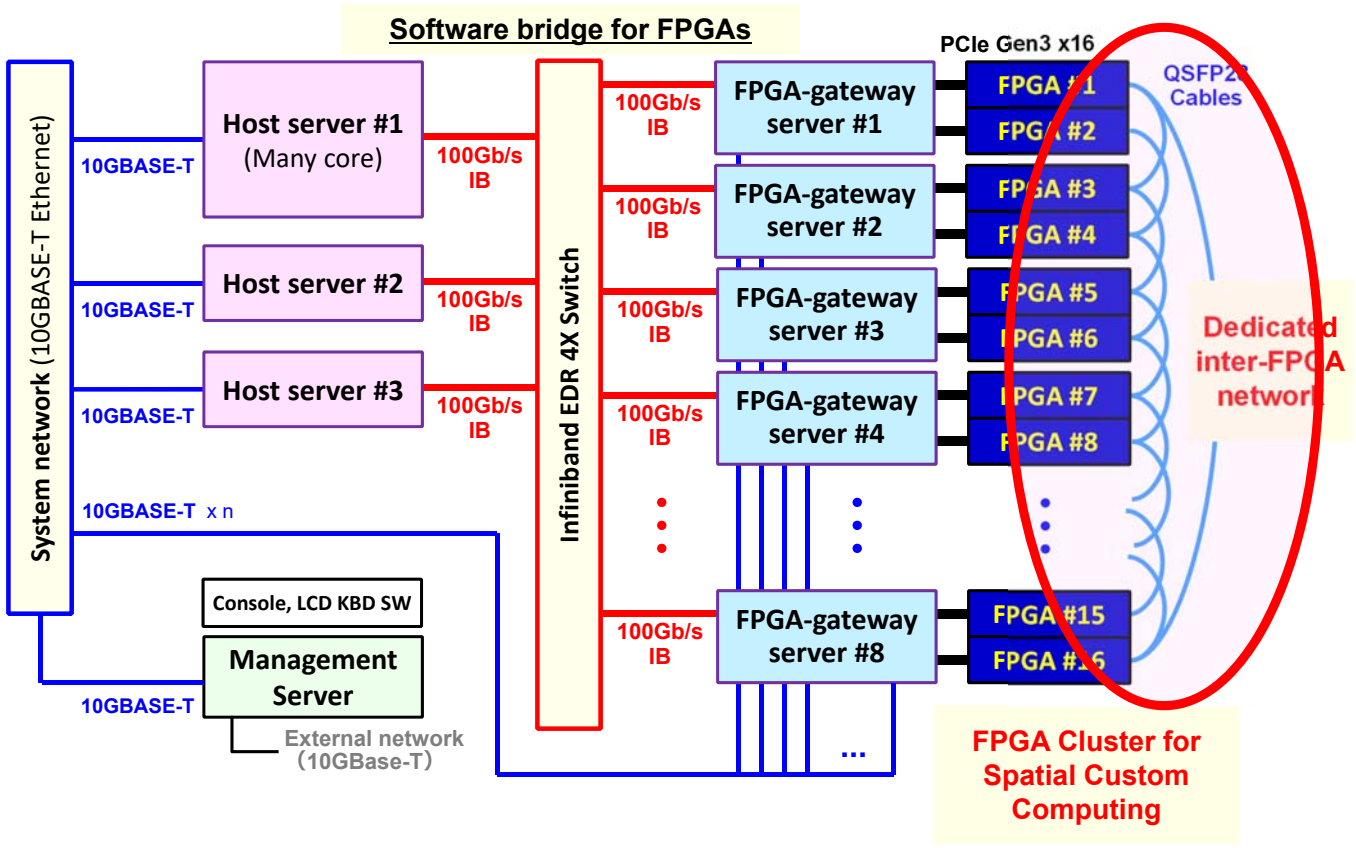
- ✓ Apps on hosts can use remote FPGAs transparently.



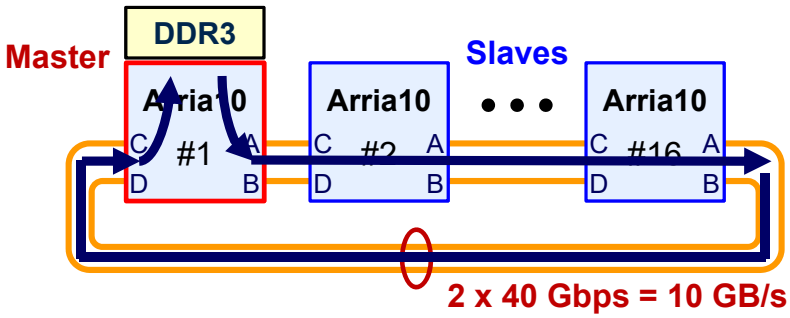
Host server

FPGA-gateway server (x86)

What Kind of NW should be Deployed?



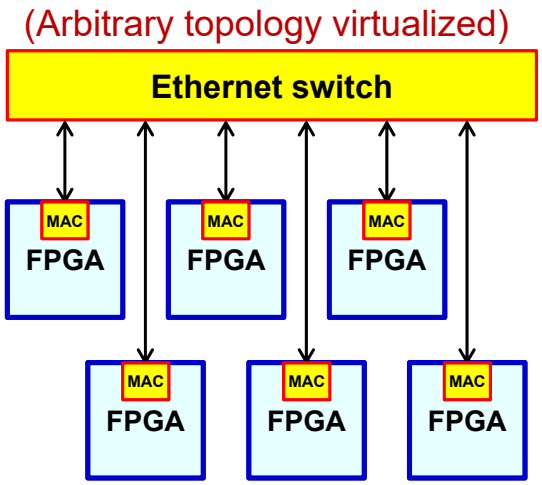
1D Ring for Stream Computing

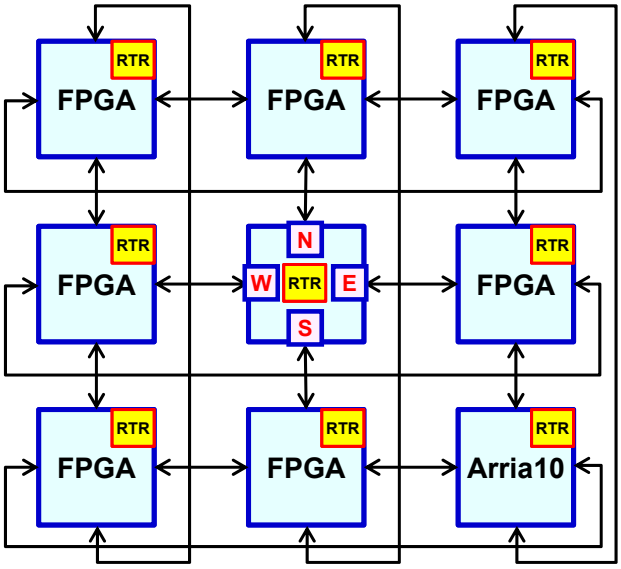


Pipelining with multiple FPGAs

- **Achieved with Arria10 FPGAs**
  - ✓ Computing pipelines on FPGAs
- **Lossless compressor**
  - ✓ Compress FP data-stream, Bandwidth enhanced to x2 ~ 3

Other Choices for Inter-FPGA Network





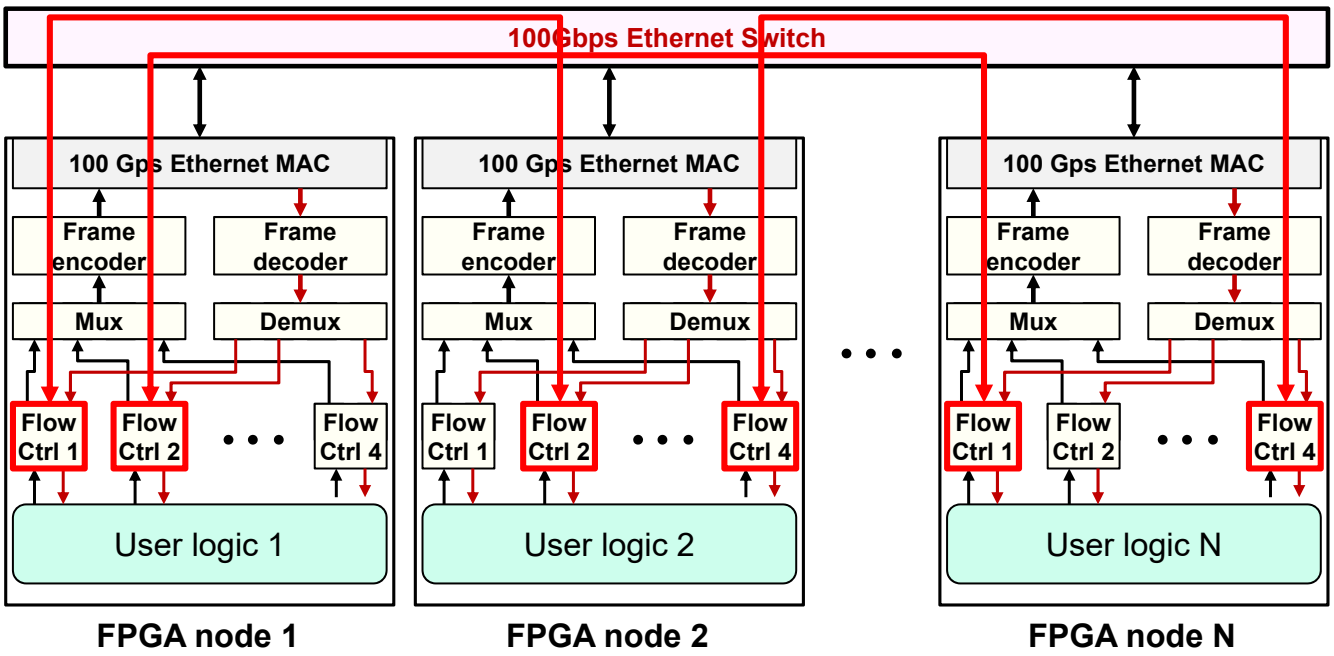
Indirect network : Ethernet

Pros) Flexibility, cutting-edge tech.
Cons) Overhead of ethernet frames, higher and variable latency, difficulty in flow-control

Direct network : 2D torus

Pros) Smaller overhead, lower and fixed latency
Cons) Inflexibility, more resources, difficulty to catch up

Virtual Topology on Ethernet Network



- **We can make virtual links on the top of Physical network.**
  - ✓ We can form arbitrary "logical" topology.

Outline

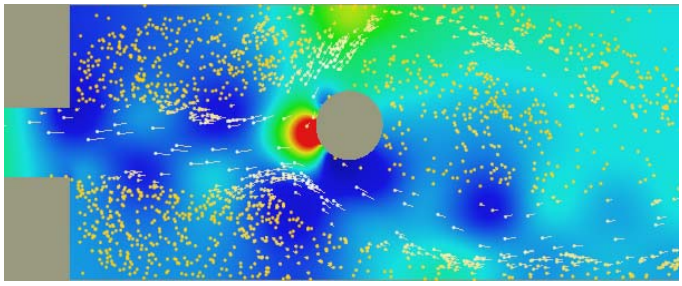
- Introduction
  - ✓ Why spatial custom computing (with FPGAs)?
- Architecture for spatial custom computing
- Feasibility study for extension of an existing machine
- **Preliminary results**
- Summary

Case Study:
Stream Computing for Fluid Dynamics Simulation

Preliminary Evaluation of Computing Core with Stratix10 FPGA

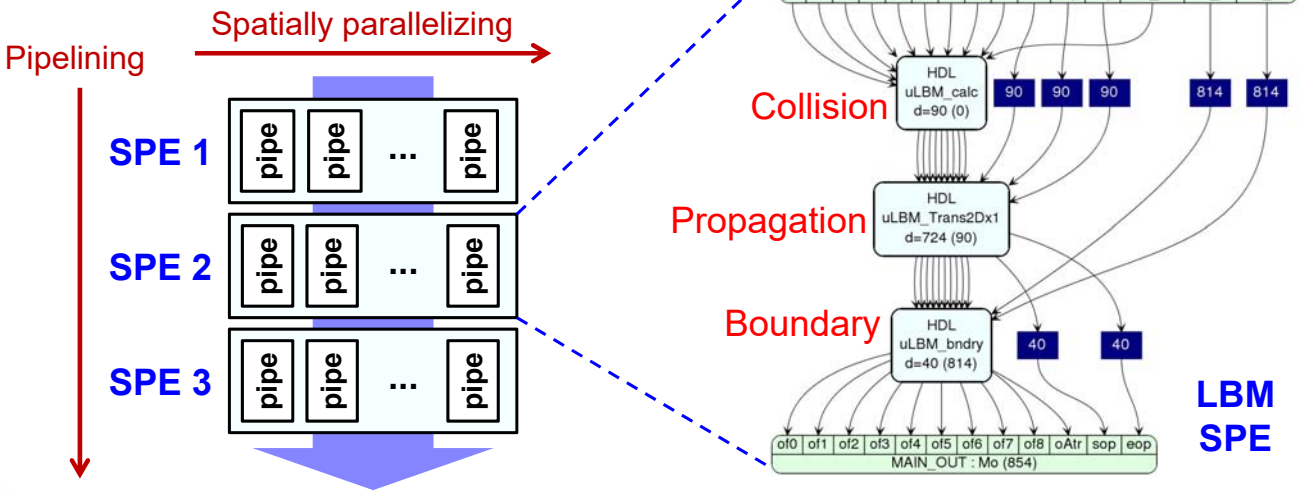
Fluid dynamics simulation

- ✓ **LBM** (lattice Boltzmann method)



Array architecture of SPEs

- ✓ **S**tream **P**rocessing **E**lements
- ✓ Generated by SPGen compiler



Kentaro Sano and Satoru Yamamoto, IEEE TPDS, Vol.28, Issue.10, pp.2823-2837, 2017.

Summary

- **Need to change architecture for post-Moore era**
  - ✓ Solution : Spatial custom computing ?
- **Feasibility study for extension of existing machines**
  - ✓ Experimental system with Stratix10 FPGAs
  - ✓ Remote-OPAE
  - ✓ Dedicated inter-FPGA networks
  - ✓ Preliminary evaluation for LBM computation w/ Stratix10
- **Future work**
  - ✓ Application/kernel implementation and evaluation
  - ✓ Tools/compilers to program FPGAs
  - ✓ System software (data-flow task scheduler)
  - ✓ Overlay architectures for FPGA, future processor architectures