Accelerating HPC applications on FPGAs using OpenCL and FPGA Network Norihsa Fujita⁽¹, Ryohei Kobayashi^{(1,2}, Yoshiki Yamaguchi^{(2,1}, Makito Abe⁽⁴, Kohji Yoshikawa^{(1,3}, Masayuki Umemura^{(1,3})

1: Center for Computational Sciences, University of Tsukuba

2: Graduate School of Systems and Information Engineering, University of Tsukuba

3: Graduate School of Pure and Applied Sciences, University of Tsukuba

4: Astronomical Institute, Tohoku University

Accelerator in Switch (AiS)

- Accelerator in Switch (AiS) is a concept proposed by Prof. Amano, Keio University, Japan
 - It couples communication and computations tightly
 - FPGAs can act as both of computation accelerators and network switches
- FPGA programming cost using Hardware Description \bullet Language (HDL) is very expensive
- Due to improvement of High Level Synthesis (HLS), programming cost of FPGA is decreasing

OpenCL-ready High Speed FPGA Networking

- Intel FPGA supports OpenCL programming environment as an HLS
- Board Support Package (BSP) is a hardware component to support multiple different boards
 - Which FPGA chip is used on the board
 - What kind of peripherals are support by the board
- Basically, only minimum interfaces are supported
 - To perform inter FPGA communication, implementing network controller and
- Inter-node ping-pong communication latency through an Ethernet switch
 - Approximately 1µ sec of communication latency
 - Much faster than traditional method (CPU Copy + InfiniBand)



No HDL code is required

ullet

- Application programmers can program FPGAs
- We consider we can realize AiS system using FPGAs
- Pre-PACS-X (PPX) is a test-bed system in Center for Computational Sciences, University of Tsukuba
 - It is a prototype of the next generation system of their PACS series supercomputer
 - Each node has 2 CPUs, 2 GPUs and 2 FPGAs
 - Not only InfiniBand network for CPUs but also 40GbE network for FPGAs





integrating it into the BSP are required



- Communication is performed with I/O channel API
 - Vendor extension to OpenCL language
 - Enables control peripherals I/O from OpenCL

sender

// Set MAC Addresses
<pre>write_channel_intel(SET_SRC , src_addr);</pre>
<pre>write_channel_intel(SET_DST, dst_addr);</pre>

// Set send data for (i = 0; i < data_size; i++) write_channel_intel(SEND, send_data[i]);

// Get recy data	receiver
for (i = 0 ; i < data_size ; i++) recv_data[i] = read_channel_	_intel(RECV);

Himeno Benchmark (3D poisson equation solver) Halo data exchange in stencil computation and allreduce are implemented on our mechanism



Authentic Radiation Transfer (ART) on FPGA

- Accelerated Radiative transfer on grids Oct-Tree \bullet (ARGOT) has been developer in Center for Computational Sciences, University of Tsukuba
 - ART is one of algorithms used in ARGOT and dominant part (90% or more of computation time) of ARGOT program
- ART is ray tracing based algorithm
 - problem space is divided into meshes and reactions are computed on each mesh
 - Memory access pattern depends on ray direction
 - Not suitable for SIMD architecture



- Problem space is divided into small blocks
 - e.g. $(16, 16, 16) \rightarrow 8 \times (8, 8, 8)$ \bullet
 - PE is assigned to each of small blocks \bullet
 - To improve BRAM performance



- PEs are connected by channels each other
 - PE: Processing Element
 - **BE: Boundary Element**



- Performance evaluation on various problem sizes
 - Values are "M meshes/s" (throughput), higher is better
 - FPGA is 4.9 times faster than CPU on 64³ size
 - Although FPGA is almost equal performance on 64³ and 128³, FPGA is much faster than GPU on 16³ and 32³



Size	CPU(14C)	CPU(28C)	P100	FPGA
(16, 16, 16)	112.4	77.2	105.3	1282.8
$(32,\!32,\!32)$	158.9	183.4	490.4	1165.2
(64, 64, 64)	175.0	227.2	1041.4	1111.0
(128, 128, 128)	95.4	165.0	1116.1	1133.5

Mesh size # of PEs | ALMs | Regs. | M20K | DSP Freq.



- Our implementation uses channel based approach
- One of extensions to OpenCL for FPGAs by Intel
- It enables inter kernel communication much faster \bullet
 - No external memory (DDR) access is required
 - Lower resource utilization than DDR access





- Kernel of PEs and BEs are started automatically by autorun attribute
 - Lower control overhead and resource usage because of decreasing number of host controlled kernels

(16,16,16)	(2,2,2)	31%	31%	27%	21%	193.2MHz
(32,32,32)	(2,2,2)	40%	40%	29%	21%	173.8MHz
(64,64,64)	(2,2,2)	40%	40%	29%	21%	167.0MHz
(128,128,128)	(2,2,2)	40%	40%	29%	21%	170.4MHz



- Applying OpenCL-ready network to ART
 - Ray will be transferred through the network
- Using next-generation Intel Stratix10 FPGA \bullet
 - Faster frequency, 3.8 times more DSPs and 4.3 times more M20Ks
 - Up to 4x100 Gbps networking capability

We thank Intel University Program for providing us both of software and hardware. ACKNOWLEDGEMENT