FPGAs for HPC: Open Challenges for Infrastructure and System Operation

FPGA & GPU: compensate with each other toward perfect HPC hardware solution

Taisuke Boku
Director, Center for Computational Sciences
University of Tsukuba
taisuke@ccs.tsukuba.ac.jp
FPGA in HPC

- **Goodness of recent FPGA for HPC**
  - True codesigning with applications (essential)
  - Programmability improvement – HLS (High Level Synthesis): OpenCL, C, C++
  - High performance interconnect: ~100Gb (xn)
  - Precision control is possible (ex. AI)
  - Relatively low power

- **Problems**
  - Programmability: OpenCL is not enough, not efficient
  - Low standard FLOPS: still cannot catch up to GPU
    -> “never try what GPU works well on”
  - Memory bandwidth: 1-gen older than high end CPU/GPU
    -> be improved by HBM2 (Stratix10~)

BittWare 520N with Intel Stratix10 FPGA equipped with 4x 100Gbps optical interconnection interfaces
## GPU vs FPGA as HPC solutions

<table>
<thead>
<tr>
<th>device</th>
<th>GPU</th>
<th>FPGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>parallelization</td>
<td>SIMD (x multi-group)</td>
<td>pipeline (x multi-group)</td>
</tr>
<tr>
<td>standard FLOPS</td>
<td>😞 😞 (1000x cores)</td>
<td>😞 (~100x pipeline)</td>
</tr>
<tr>
<td>conditional branch</td>
<td>😞 (warp divergence)</td>
<td>😞 (both direction)</td>
</tr>
<tr>
<td>memory</td>
<td>😞 😞 (HBM2e)</td>
<td>😞 (DDR)→ 😞 (HBM2)</td>
</tr>
<tr>
<td>interconnect</td>
<td>😞 (via host facility)</td>
<td>😞 😞 (own optical links)</td>
</tr>
<tr>
<td>programming</td>
<td>😞 (CUDA, OpenACC, OpenMP)</td>
<td>😞 (HDL)→ 😞 (HLS)</td>
</tr>
<tr>
<td>controllability</td>
<td>😞 (slave device of host CPU)</td>
<td>😞 (autonomic)</td>
</tr>
<tr>
<td>HPC applications</td>
<td>😞 (various fields)</td>
<td>😞 (not much)</td>
</tr>
</tbody>
</table>
CHARM: Cooperative Heterogeneous Acceleration with Reconfigurable Multi-devices

multi-physics/multi-scale complicated problem

Basic cluster with GPUs (by InfiniBand)

invoke GPU/FPGA kernels

data transfer via PCIe (invoked from FPGA)

Application oriented FPGA-FPGA communication

100Gbps direct optical link

FPGA Network
Cygnus supercomputer at Center for Computational Sciences, Univ. of Tsukuba (2019~)  
81 nodes in total including **32 “Albireo” nodes with GPU+FPGA** (other “Deneb” nodes have GPU only)
Single node configuration (Albireo)

- All nodes in Cygnus are equipped with both IB EDR and FPGA-direct network
- Some nodes (Albireo) are equipped with both FPGAs and GPUs, and other nodes (Deneb) are with GPUs only
- GPU: NVIDIA V100 x4
- FPGA: Intel Stratix10 x2
Albireo node

IB HDR100 x4 ⇒ HDR200 x2

100Gbps x4 FPGA optical network

IB HDR200 switch (for full-bisection Fat-Tree)
Current research around Cygnus

- **FPGA-network**: CIRCUS (Communication Integrated Reconfigurable CompUting System)
  - direct interconnect facility among FPGA boards by multi-dimensional optical link (~100Gps) with router and OpenCL-ready API
  - pipelining all computation and communication seamlessly

- **GPU-FPGA DMA**: kicked by FPGA (without CPU)
  - PCIe-protocol base DMA engine to reduce multi-device high speed data transfer

- **Programming**: MHOAT (Multi-Hetero OpenACC Translator)
  - breaking single OpenACC code for multi-devices (GPU + FPGA) for each compiler
  - FPGA: OpenACC -> OpenCL -> aocx (by OpenARC research compiler by ORNL)
  - GPU: OpenACC -> NVPTX (PGI compiler)

- **Application**: ARGOT, application on astrophysics for early-universe object generation
  - two main parts are executed by GPU and FPGA
**CIRCUS performance**

**Throughput (1hop~7hops)**
- Max. throughput: 90.2 Gbps
- Min. latency: 500 ns

**Latency (1hop~7hops)**
- Latency+/hop: ~250 ns

---

Evaluated on up to 8 Bittware 520N FPGA boards in Cygnus supercomputer at CCS, University of Tsukuba [14]

Bandwidth of GPU-FPGA DMA

- Data size: 4 ~ 2G Bytes (PCIe gen3x8*: 7.8GB/s peak)
  * due to Intel FPGA BSP limitation

![Bandwidth Graph]

- Up to 7.2 GB/s (FPGA → GPU)
- Up to 3.7 GB/s (GPU → FPGA)
GPU-only vs GPU-FPGA coworking on astrophysics simulation ARGOT

Mesh size: 16$^3$

17.4x speed-up

Mesh size: 32$^3$

10.2x speed-up

Mesh size: 64$^3$

2.84x speed-up

Mesh size: 128$^3$

1.32x speed-up

Single node GPU (V100) + FPGA (Stratix10) evaluation on Cygnus
Topics for Discussion

- What are the most promising feature or advantages of present or future HPC using FPGAs?
  ⇒ multi-hetero solution, not FPGA stand-alone but coupling with other accelerators (GPUs) for perfect acceleration

- What technology is missing in infrastructure and operation of HPC systems using FPGAs?
  ⇒ programming framework to release users from tough low-level programming, keeping a certain efficiency: ex. HLS or DSL?

- How can we or should we tackle the above challenge in the community with academia and industries?
  ⇒ true codesigning with application users, open source facilities
CFP for HPC FPGA Workshop 2021 (Cluster2021)

https://sites.google.com/view/hpcfpga2021/

- HPC FPGA Workshop 2021 in IEEE Cluster 2021 (Sep. 7th, 2021, virtual)
  - topic: FPGA for HPC (exactly with same as this session!)
  - any hardware, software, middleware, application on FPGA utilization toward HPC
  - proceedings will be published as IEEE Cluster2021 workshop volume

- Important dates
  - paper submission due: June 25th ⇒ to be extended ???
  - acceptance notification: July 19th
  - camera ready due: July 30th
  - workshop: September 7th

- Organization
  - Taisuke Boku (chair), Martin Herbordt (PC chair), Franck Cappello, Kentaro Sano