**FlexFPGA: A Case for Multi-Tenant Disaggregated Cloud FPGA Architectures**

**STUDENTS:** MISAES LOPEZ, RAHAUL POTLURI

---

**The Problem**

- Traditional data centers restrict FPGA access to a single user, leading to resource underutilization and power wastage.
- *FlexFPGA* presents FPGAs as individual compute nodes integrated into data centers, enhancing efficiency, flexibility, and real-time resource allocation for diverse AI and cloud computing applications.

---

**FlexFPGA**

- *FlexFPGA* is a research cloud FPGA architecture leveraging Zynq 7000 SoC and partial reconfiguration to treat FPGAs as stand-alone compute nodes, integrating them into data centers via network attachment.
- Multi-tenancy is enabled by allowing users to upload partial BIT streams and push them onto reconfigurable modules, presenting a unified FPGA space to users while utilizing multiple reconfigurable regions in the background.

---

**State of Research and Industry**

- The problem of resource underutilization and power wastage in traditional data centers.
- Industry leaders like AWS, Alibaba, and Microsoft employ FPGA Co-Processor architectures, connecting FPGAs to hosts via PCIe, but still reserve them for single users, leading to potential resource and energy wastage.

---

**Future Work, References, and Acknowledgments**

- Develop data center software architecture for multi-tenant support.
- Gather additional power consumption data for architecture analysis, comparing against traditional setups.
- Evaluate processing speed and scalability of the architecture and comparing with other alternatives.

---

**Methodology**

- To evaluate the feasibility of our proposed architecture, we plan to measure processing speed and power consumption across disaggregated FPGAs and compare against traditional cloud computing architectures.

---

**Virtualization**

- In data center environments, virtualization is a key technology for resource allocation and management.

---

**Architecture Distribution**

- Co-Processor 20%
- Disaggregated 10%
- Hybrid 10%
- FPGA Reconfiguration 40%
- Container 10%
- Other 5%

---

**Disaggregated Reconfigurable Hardware**

- **Partial Reconfiguration preserves FPGA integrity while modifying specific areas with BIT files.**
- **FPGA shells can be created in Vivado and offer customizable infrastructure.**
- **Zynq 7000 devices integrate ARM Cortex-A9 processors with FPGAs, offering high performance and flexibility, suitable for seamless integration into data center and cloud environments with rich peripherals, facilitated by Xilinx’s Vivado software platform.**

---

**Leveraging Zynq SoC and Dynamic Function Exchange**

- Disaggregated architecture prototype
- Power consumption in disaggregated servers
- Power consumption (W), FPGA with SoC Only
- Power consumption (W), FPGA with Prototype
- Architecture for multi-tenant support
- System Architecture for Network-Reconfigurable FPGAs in the Cloud using Partial Reconfiguration
- System-Aware Reconfiguration for Network-Reconfigurable FPGAs in the Cloud
- System-Aware Reconfiguration for Network-Reconfigurable FPGAs in the Cloud

---

**Conclusion**

- FlexFPGA architecture combines the benefits of traditional FPGAs with the flexibility and scalability of cloud computing.
- It addresses the problem of resource underutilization and power wastage in traditional data centers.
- Future work includes developing a software architecture that fully leverages the capabilities of FlexFPGA, gathering additional power consumption data for comprehensive analysis, and evaluating the architecture’s performance and scalability against other alternatives.