How FPGAs Have Evolved to Address Compute Acceleration and Interconnects

Ken Chang

Xilinx at OFC Booth #1811
Agenda

- Why FPGA, Adaptable Compute?
- FPGA to Adaptable Compute Evolution
- FPGA Transceiver Evolution
- Summary
Compute and Bandwidth Needs Across All Industries

Cloud

Network

Edge

AI ADOPTION ACROSS MARKETS

EXPONENTIAL DATA BANDWIDTH GROWTH
The Slowdown of Moore’s Law and Technology Scaling

Processing Architectures are Not Scaling

40 YEARS OF PROCESSOR PERFORMANCE

Performance vs. VAX11-780

Source: John Hennessy and David Patterson, Computer Architecture: A Quantitative Approach, 6/e 2018
FPGA Evolution: From FPGA to SoC to ACAP

- **First FPGA Introduced**: 1985
- **First Virtex FPGA**: 1998
- **Virtex-2 Pro**: 2001
- **First 3D FPGA & HW/SW Programmable SoC**: 2011
- **First MPSoC & RFSoC**: 2016
- **ACAP**: 2019

**Moore’s Law Era**

**SoC Era**
Adaptive Compute Acceleration Platform

- Based on 7nm FinFET
- Platform Solution for Compute / Storage / Network Acceleration
- IP Subsystems and a Network-on-Chip
- Programmable AI Engine
AI Engine

- **Target Applications**
  - ML Inference for Cloud DC
  - 5G wireless: Radio, Baseband
  - ADAS/AD Embedded Vision

- **Each Tile:**
  - ISA-based Vector Processor
    (SW programmable, C++, e.g.)
  - Local Memory
  - Adaptable, Non-blocking Interconnect

---

Source: Juanjo Noguera, “HW/SW Programmable Engine: Increased Compute Density Architecture for Project Everest”, Hot Chips 2018
Industry’s First ACAP: Versal Architecture Overview

7nm FinFET

Scalar Engines
- Arm Dual-Core Cortex-A72 Application Processor
- Arm Dual-Core Cortex-R5 Real-Time Processor
- Processing System

Adaptable Engines
- Versal Adaptable Hardware
- Block RAM
- UltraRAM
- Accelerator RAM

Intelligent Engines
- AI Engines
- DSP Engines

Network On Chip
- PCIe & CCIX (w/DMA)
- DDR
- HBM
- 112G
- 58G
- 28G
- Multirate Ethernet
- 600G Cores
- MIPI
- LVDS
- GPIO
- Direct RF

© Copyright 2019 Xilinx
ACAP Meets Compute Demands for Cloud, Network, and Edge

**CLOUD**
- GoogleNet V1 Img/Sec (<2ms)
  - 16nm FPGA
  - 7nm ACAP
  - 10X

**NETWORK**
- Int 16x16 DSP Compute (TeraMAC/sec)
  - 16nm RFSoC
  - 7nm ACAP
  - 5X
- Single-Chip Encrypted Traffic (Gb/s)
  - 16nm FPGA
  - 7nm ACAP
  - 4X

**EDGE**
- ResNet50 Img/sec (batch=1)
  - 16nm MPSoC
  - 7nm ACAP
  - 15X

© Copyright 2019 Xilinx
Backplane Transceiver Trend from Industry

Backplane channel loss: > 25dB

Data source: ISSCC, VLSI Circuit Symposium

© Copyright 2019 Xilinx
Backplane Transceiver Trend: Date Rate vs Year

First 28G LR paper (IBM)

- 56G PAM4 RX
- 100G FEC
- 56G PAM4 RX

ADC based
Backplane Roadmap (From 2012)

- **Electrical-Optical Crossover Zone (1m Backplanes): 100 Gb/s-m**
- **180nm 3.125Gbps**
- **130, 90nm 12.5Gbps**
- **65, 40nm 25Gbps**
- **28, 20nm 50Gbps**
- **14, 10nm 100Gbps**

**Mainstream Backplane Transceiver Sampling Year**

- **100Gbps-m Per Link Optical Backplanes by 2020?**
- **Smaller BPs, Active Connectors, Non-NRZ Signaling?**

> Courtesy of E. Wu, “FPGA Optical Connectivity,” OFC Workshop 2012
Backplane Transceiver Trend: Date Rate vs Year
Electrical vs Optical

Ashok Krishnamoorthy, “Optical interconnects in computing and switching systems: the anatomy of a 20Tbps switch card”, ISSCC 2018
On-Board Optics Solution

> Capable of error free (BER < 10^{-12}) for compute application
> Electrical portion (16nm FinFET) based on CEI-56G-NRZ ➔ Published in ISSCC 2018
> Optical portion based on EAM silicon photonics with Driver/TIA in 16nm FinFET ➔ Submitted to VLSI 2019

<table>
<thead>
<tr>
<th>Total Power per Channel</th>
<th>mW</th>
<th>pJ/bit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>126</td>
<td>2.25</td>
</tr>
</tbody>
</table>

XSR Example: 50mm, 8dB
Optical …

Xilinx OFC Booth #1811

Error free at 50Gb/s!

Measured Bathtub 50Gbps
Tx7 Rx4 No EDFA Ext Loop

BER

4 6 8 10 12 14 16 18 20 22 24 26

PI Code

1.00E-12
1.00E-11
1.00E-10
1.00E-09
1.00E-08
1.00E-07
1.00E-06
1.00E-05
1.00E-04
1.00E-03
Transceiver Evolution: the “On Package” Stage

- Front Panel Pluggable
- Mid-Board Optics
- On Package, Electrically Pluggable
- On Package, Optically Pluggable

Slide courtesy of Samtec
Summary

Evolution of Compute Bandwidth

*Surpassed Moore’s Law with New Architecture*

Evolution of Serial I/O Bandwidth

*Surpass Electrical Limits with New Architecture?*

---

**Evolution of Compute Bandwidth**

- **Limits of Moore’s Law**
  - Log scale
  - Compute (TOPS)
  - Time
  - FPGA → ACAP re-architecture
  - To meet expectations of growth

**Evolution of Serial I/O Bandwidth**

- **Limits of Electrical Signaling**
  - Log scale
  - I/O BW/Package
  - Time
  - Optics?
Acknowledgments

> The speaker would like to thank the following people for their valuable contributions and discussions for the presentation

>> Anthony Torza, Ehab Mohsen, Kees Vissers, Vamsi Boppana, Gaurav Singh, Yohan Frans, Geoff Zhang, Ephrem Wu, Liam Madden
Adaptable. Intelligent.