OpenCAPI Enabled FPGAs
The perfect bridge to a Data Centric World

Allan Cantle, CTO & Founder
Nallatech, a Molex Company
Nallatech at a Glance

Server qualified accelerator cards featuring FPGAs, network I/O and an open architecture software/firmware framework. **Design Services/Application Optimization**

- **Nallatech** – a **Molex** company
- 25 years of FPGA heritage
- **Intel** PSG (Altera) OpenCL partner
- **Xilinx** Alliance partner
- Server partners: Cray, DELL, HPE, IBM, Lenovo
- Application porting & optimization services:
  - Storage Acceleration
  - Energy-efficient HPC/Machine Learning
  - Real-time, low latency network and I/O processing
We are a global provider of innovative electromechanical solutions.

- Molex is part of one of the largest private companies in the United States.
- Annual Koch Revenue: $120B+
- Annual Molex Revenue: $6B+
- 21 countries
- 73 manufacturing locations
- 40,000+ employees

Vision:
To be a leading global provider of innovative electronic solutions that create value for our customers and society...
Organization

Molex Confidential Information
© Molex, LLC – All Rights Reserved. Unauthorized Reproduction/Distribution is Prohibited.

Molex
One company. A world of innovation.

- 40,000+ Employees
- $5B+ 2017 Revenue
- 73 locations in 21 countries
- Part of Koch Industries, one of the largest private companies in the world (Forbes estimate $115B)

ISI GROUP
- Camarillo, CA
- Concord, NH
- Cumbernauld UK

Commercial & Consumer Solutions (CCS)
- Tablets, Smartphones
- Appliances

Datacom & Specialty Solutions (DSS)
- Data/Networking
- Telecommunications
- Fiber Optics
- Medical

Transportation & Industrial Solutions (TIS)
- Automotive
- Robotics
- Industrial Automation

FPGA Network Processing
Real-Time Data Acquisition & Signal Processing
FPGA Accelerated High Performance Computing
Miniaturized Packaging & Integrated Solutions

Molex, LLC – All Rights Reserved. Unauthorized Reproduction/Distribution is Prohibited.
Why OpenPOWER™

Market Shifts

- Moore’s Law
- Workload Demands Up
- Numerous IT consumption models
- Mature Open software ecosystem

Strategy

- Vibrant ecosystem through open development
- Accelerated innovation through collaboration of partners
- Amplified capabilities driving industry performance leadership
“Domain Specific Architectures have to be the future (of computing)……………… a domain that’s rapidly changing and that’s why I think FPGAs have a particularly critical role to play as we begin to change those architectures”

– John Hennessy 8/21/2018

“It’s Time for New Computer Architectures and Software Languages”

– David Patterson 9/17/2018

1. https://www.youtube.com/watch?v=xnNs74tTHZU&feature=youtu.be
Domain Specific Architectures – Platform Flexibility is essential

» Todays Processors now include:
  » CPUs (large & small), GPUs, FPGAs & TPUs
» Memory & Storage are steadily becoming one
» Disaggregation will put a major strain on the Network
  » Re-aggregation is not scalable or power efficient with classic ToR Networking
» FPGA is the only processor that is truly data centric
  » Fundamentally a streaming processor
  » It’s IO can talk any fabric protocol
  » It can connect to any type of Memory & Storage
  » It can circuit switch and/or it can packet switch as required by the application

» **FPGAs provide both acceleration & the Glue for Domain Specific Architectures**
Data Center Architectures, blending evolutionary with revolutionary

Zero Power, High Bandwidth, Low Latency, Non Coherent, Streaming Data Plane, e.g. GenZ

Emerging Data Centric Enhancements

Existing Data Center Infrastructure

Traditional ToR Packet switched Network & Control Plane
OpenCAPI Enabled HyperConverged & Disaggregatable Server

» Leverage Google & Rackspace’s OCP Zaius/Barreleye G2 platform
» Reconfigurable FPGA Fabric with Balanced Bandwidth to CPU, Storage & Data Plane Network
» OpenCAPI provides Low Latency & coherent Accelerator / Processor Interface
» GenZ Memory-Semantic Fabric provides Addressable shared memory up to 32 Zetabytes

[Diagram showing the components and connections of a hyperconverged server with OpenCAPI enabled.]
Molex’s OpenCAPI Enabled Flash Storage Accelerator, FSA

» Xilinx Zynq US+ 0.5OU High Storage Accelerator Blade
» 4 FSAs in 2OU Barreleye-G2 OCP Storage drawer deliver:
  » 152 GByte/s PFD* Bandwidth to 1TB of DDR4 Memory
  » 256 GByte/s PFD* Bandwidth to 64TB of Flash
  » 200 GByte/s PFD* Bandwidth through the OpenCAPI channels
  » 200 GByte/s PFD* Bandwidth through the GenZ Fabric IO
» Open Architecture software/firmware framework

*PFD = Peak Full Duplex
250-SoC – FPGA NVMe-oF Storage Controller

- Xilinx Zynq US+ HHHL PCIe Gen 3 x16 / Dual PCIe Gen 4 x8 Card
- 1x 4/8 GB single rank Discrete DDR4 x72 MPSoC Memory @2400
- 1x 4/8 GB single rank Discrete DDR4 x72 Fabric Memory @2400
- 4x PCIe Gen 3 / 4 x8 Cabled I/O channels via OCULink x8 connectors
- 2x 100GbE Network IO

Flexible OCULink IO includes support for up to 2 OpenCAPI ports
PCISIG OCULink – An Universal Interconnect Family

Low Profile Add-in Card Applications
Connector and Cable Assembly Solutions
- 12 mm Max Height of Mated Cables is Below the 14.47 mm Max Height for Components on PCIe Add-in Cards

Mid-plane Applications
Low Profile Connector and Ribbon Cable Assembly
- Reduces possible interference with system board components

Point to Point Cabling
Lower Loss Solutions can simplify PCB construction and reducing the number of layers thereby lowering costs
250-SoC Photos

- **250-SoC including optional GbE SoC Network port**
- **250-SoC Card with OCuLink Cables attached**
- **250-SoC HHHL Card without GbE SoC Network port**
Full Featured 250-SoC System Configuration Examples

Config 1 – FlashGT+ legacy mode

Config 2 – NVMe Gen 4

Config 3 – Max PCIe Host Bandwidth A

Config 4 – Max PCIe Host Bandwidth B

Config 5 – OpenCAPI to NVMe Gen 4

Config 6 – OpenCAPI – Oversubscribed NVMe

Config 7 – Dual OpenCAPI Attach

Config 8 – PCIe G3 x16 Host Interface
Near Storage Accelerator
Shipping Now!

250S+

High-performance PCIe-based Flash SSD with localized FPGA acceleration capability

- NIC Form Factor - Low Profile, Half Length PCIe form factor
- PCIe Gen 3 or Gen 4 8-lane
- (1) Xilinx XCKU15P-2FFVA1156I FPGA-2 speed grade
- (1) bank of 4GByte 2400MTPS x80 DDR4 memory
- (4) M.2 connectors
- (4) M.2 to OCULink678mm cables
- Support for CAPI2.0 & CAPI SNAP
- Available pre-configured with NVMe-oF accelerated functions:
  - PCIe Gen 4 Host Bus Adaptor (HBA)
Near Storage Accelerator
Shipping Now!

**250S+**

**Additional Features:**

- FPGA Configuration Functions (JTAG, Remote Flash Update, Reconfiguration, Fallback Mechanism, etc.)
- Board Health Monitoring (Temperatures, Power, Voltages, etc.)
- Unique Board Data in PROM
- On-Board Anti-Bricking for Deployment in Data Centre
- Cabling options for OCuLink and MiniSAS-HD backplanes
- OpenPOWER CAPI SNAP 2.0 for POWER9

- DDR4: Maximum sustained read or write bandwidth based on total bus width (width includes ECC bits)
- PCIe: Peak Full Duplex show TX+RW bandwidth
- NVMe over M.2: Bandwidth to storage limited by disk performances, I/Os are PCIe Gen4 capable 8+8GB/s each
- x4, x8, etc. shows the number of lanes per interface
250-HMS – OpenCAPI Hybrid Memory Subsystem

- 32GBytes or DDR4 DRAM over 3x x72 Memory Channels
- 4TB FPGA enabled, M.2 NAND Media with 25.6GBytes/s peak bandwidth
  - Also supports regular supports standard M.2 NVMe Flash NAND SSDs
- Potential to Leverage Rambus’s Hybrid Memory IP
M.2 Format, FPGA enabled, NAND Media

- Up to 1TByte Flash Memory depth
- 6.4GBytes/s Peak NAND Bandwidth
  - 8x NAND Channels at 800MTPS
- 8GB/s peak M.2 IO Bandwidth
  - 16Gb/s transceiver speeds
- Potential to Support Open-Channel SSD Innovations
250-HMS – OpenCAPI Hybrid Memory Subsystem – Space Model
Bittware’s VU37P OpenCAPI capable FPGA Accelerator

**XUPVWH**

Xilinx UltraScale+ 3/4-Length PCIe Board

VU37P with Integrated HBM2, Quad QSFP, and 256 GBytes DDR4 on BittWare Viper Platform

**Features**

- **Ultra high-speed network interface**
  4x 100GbE and timestamping support

- **Latest generation 16nm FPGA**
  UltraScale+ FPGA with integrated HBM2

- **Optimized for thermal performance**
  BittWare Viper platform with passive heatsink supports large FPGA loads

- **OpenCAPI Support**
  over 2x SlimSAS IO Connectors
Bittware’s VU37P FPGA Accelerator – Block Diagram
HBM-enabled FPGA with support for 3DXPoint Optane DIMMs

520MX

High performance compute node featuring 3D-stacked High Bandwidth Memory (HBM2) and choice of external memory DIMMs

- **GPU/Phi** form factor (3/4 length) - dual slot
- Active, Passive or Liquid cooling options
- 16-lane PCIe Gen 3.0
- (1) Intel Stratix 10 FPGA
  - MX1650 or 2100
  - Choice of 2x 8GB HBM or 2x 16G HBM
- Up to (4) QSFP28 cages supporting 1G → 100G line rates
- Up to (2) DIMMs capable of supporting: DDR4 SDRAM, QDR-II+ SRAM or Intel Optane (3D-XPoint)
- Up to (2) OpenCAPI Enabled OCuLink connectors for board-to-board I/O
- Board Support Packages (BSP) for Intel OpenCL SDK
Xilinx Virtex US+ VU37P FPGA with 8GB of HBM2 DRAM

- 2x 50GByte/s PFD OpenCAPI Channels to Power 9 CPU through Mezzanine connector
- Up to 4x 50GByte/s PFD Channels to neighboring OpenCAPI Mezzanine Modules
- 4x Nearstack x4 Cabled IO + 4x OCULink x8 Cabled IO
- Supports IBM S922LC Witherspoon, Redbud & Sequoia, HPC Server Platforms

OpenCAPI to P9 OR FPGA to FPGA Channel through Mezz connector

OpenCAPI Channels to P9 CPU through Mezz connector

- Nearstack x4 Connector (No Sidebands) Up to 25Gbps
- OCULink x8 Connector (inc. Sidebands) Up to 25Gbps – supports PCIe x8
HBM FPGA OpenCAPI Mezzanine Module

Top View

- 4x Nearstack x4 Connectors for external Network IO
- 4x OCULink x8 Connectors for Internal connection to other modules or/and NVMe SSDs

Bottom View

- 2x FCI 400 Pin Megarray Connectors
IBM AC922 – Air cooled HPC Server with 4/6x OpenCAPI Mezz Modules

- 4x OpenCAPI Mezzanine Module Sites
- 2x additional OpenCAPI Mezzanine Module Sites on Watercooled variant
IBM AC922 with 4x FPGA OC Modules - Example HPC Configuration

POWER9 CPU

Virtex US+ VU37P FPGA

Virtex US+ VU37P FPGA

Virtex US+ VU37P FPGA

Virtex US+ VU37P FPGA

400GBytes/s peak Aggregate POWER9 to FPGA Bandwidth over OpenCAPI

100GB/s

150GB/s

100GB/s

400GBytes/s peak Aggregate NVMeoF Storage IO Bandwidth

Nallatech a molex company
Summary

» The Gods of Computer Architecture have spoken – It’s time to innovate!
» FPGAs provide a fertile ground for Computing Architecture Innovation
» Much of this innovation is at the System Level as much as the silicon level
» Molex has wide and deep support for OpenPOWER & OpenCAPI
  » Enabling the convergence of Storage and Memory
  » Enabling Modularity through cabled interconnect for flexible fabric configurations
  » Providing a range of FPGA Accelerated solutions to fit varying customer applications