FPGAs em HPC e ML

João Dullius
Field Application Engineer – BP&M
joaodullius@bpmrep.com.br

XXI Simpósio em Sistemas Computacionais de Alto Desempenho
Industry Trends

Heterogeneous Compute
Key challenge: Programming & integration of accelerators

Cloud to Edge
Key challenge: Need for retargetability

AI Proliferation
Key challenge: Efficient ML acceleration and integration in the whole application
Heterogeneous Compute

Cloud to Edge

AI Proliferation
Solution Providers Need Adaptive Compute Platform

Clear Need for HW and SW to merge to create Differentiated and high-performance Real-time Solutions
Xilinx Real-Time Compute Acceleration

Video Transcoding for VP9 Live Stream
- CPU: 1x
- GPU: 1x
- ASIC: 1x
- FPGA: 30x Frames per second

Genomic Data Analytics
- CPU: 1x
- FPGA: 90x Performance

Real-time AI Inference
- CPU: 1x
- GPU: 1x
- FPGA: 20x Performance/
- ASIC: 100x
Xilinx Real-Time Storage Acceleration

Compute Acceleration
Storage Acceleration
Network
PCle
CPU

Relative Performance

QUERY PERFORMANCE

1x 4x
13x
None 1 2 4

# FPGA Accelerators

Server: Dual Socket Xeon E5-2697 (24 Cores) - Data-Plane: ~20 Mpps - # Cores Used: 10
Xilinx Real-Time Network Acceleration

6X
More Packet processing throughput

Server + FPGA (VU9P) - Data Plane: ~100 Mpps - # Cores Used: 2
Vitis Unified Software Platform
Evolution of Xilinx’s Software Platform

- Vivado
- OS and Firmware SDK
- SDSoC, Embedded
- SDAccel, Data Center (FaaS, Alveo)
- AI inference Acceleration
- Vitis Unified Software Platform

2012

2019
Unified: All Developers can Build and Deploy to All Platforms

Build
- Embedded Developers
- Enterprise Application Developers
- Enterprise Infrastructure Developers
- Data & AI Scientists

Deploy
- Zynq
- Ultrascale
- Alveo
- Data Center Rack

XILINX
VITIS™
Versal: Adaptive Compute Acceleration Platform

Scalar Processing Engines

Adaptable Hardware Engines

AI Engine

Network on Chip
SW Programmable Infrastructure
Vitis: Unified Software Platform

- Domain-specific development environment
  - Vitis accelerated libraries
    - OpenCV Library
    - BLAS Library
    - Finance Library
  - Vitis core development kit
    - Compilers (ARM, HLS, AI Engines)
    - Analyzers
    - Debuggers
  - Xilinx runtime library (XRT)

- Vitis target platform

- Coming soon...
  - TensorFlow
  - Vitis AI
  - FFmpeg
  - Vitis Video
  - Partners (Genomics, Data Analytics, And more)

Partners: Genomics, Data Analytics, And more

- Finance Library
- Analyzers
- Debuggers

Vitis: Unified Software Platform
Industry Trend: Adaptive Computing

Silicon Design Cycle

Innovation Cycle

Top-1 Accuracy (%)

2012 2018

Pace of AI/ML Innovation
Industry Trend: Adaptive Computing

- Classification
- Object Detection
- Segmentation
- Speech Recognition
- Recommendation Engine
- Anomaly Detection

Diverse models over a broad range of applications
Industry Trend: Adaptive Computing
Build: Comprehensive Development Tool Suite

Host Application (C/C++/Python/OpenCL)

Libraries

C/C++/RTL

Target Platform

Models

Host CPU

ARM Compiler

Vitis HLS

AIE Compiler

DPU

Build

System Compile/Link

Xilinx runtime library (XRT)

System level Simulation

Device/Card

Device/Card

Device/Card
Platform-Based Development Approach

- Development is performed in the context of a platform
  - A pre-configured system containing I/O, status monitoring, and lifecycle management
- Standardized interfaces allow for automated composition of user functionality
Xilinx Runtime (XRT)

> **Platform- and OS-independent APIs for**
  >> Device management
  >> Memory management and data transfers
  >> Accelerator execution management

> **OpenCL wrappers, media frameworks, and domain-specific APIs built on top of base APIs**

> **Open source and available on Github**
Vitis HLS Compiler

- Higher level of abstraction
- Enables supporting new languages, e.g. C++ 14, OpenMP
- Planning to open source HLS front-end

Vivado HLS Code

```c
void vadd(
    const unsigned int *in1,
    const unsigned int *in2,
    unsigned int *out,
    int size)
{
    for (int i = 0; i < size; i++)
        out[i] = in1[i] + in2[i];
}
```

Vitis HLS Code

```c
void vadd(
    const unsigned int *in1,
    const unsigned int *in2,
    unsigned int *out,
    int size)
{
    for (int i = 0; i < size; i++)
        out[i] = in1[i] + in2[i];
}
```
400+ functions across multiple libraries for performance-optimized out-of-the-box acceleration
Deploy: Embedded, Single Server, Scale-out

Executable

Single Server

Scale Out

Xilinx Docker Registry

Kubernetes plugin and single node resource manager
Unified Software Platform

Cloud to edge, software and AI

Comprehensive tools, runtime, libraries and models

Standards, Open Source, Free

Embracing & participating in open source

Use of standard environments & APIs

How to get started

How to get access to Vitis and Vitis Libraries


Getting Started Tutorials

Thank You