NVIDIA'S DEEP LEARNING ACCELERATOR MEETS SIFIVE'S FREEDOM PLATFORM

Frans Sijstermans (NVIDIA) & Yunsup Lee (SiFive)
NVDLA — NVIDIA DEEP LEARNING ACCELERATOR

IP Core for deep learning - part of NVIDIA’s Xavier SOC
Optimized for Convolutional Neural Networks (CNNs), computer vision
Targeted towards edge devices, IoT
Industry standard formats and parameterized

Why open source NVDLA

   Encourage Deep Learning applications
   Invite contributions from the community
AREA, PERFORMANCE, AND POWER

**SMALL CONFIGURATION**
- INT8 data path
- 1 RAM interface
- No advanced features

**EXAMPLE**
- Config: 512 MACs, 256KB buffer
- Area: 1.4mm$^2$ in 16nm
- Perf: 93 fps ResNet50
- Power: 107mW

**LARGE CONFIGURATION**
- INT8, INT16, FP16 data path
- 2 RAM interfaces
- Weight compression

**EXAMPLE**
- Config: 1024 8b + 512 16b MACs, 256KB
- Area: 2.4mm$^2$ in 16nm (excl. TCM)
- Perf: 230 INT8/115 FP16 fps ResNet50
- Power: 348 / 475 mW for INT8 /FP16
SW ARCHITECTURE

Compile time

- Caffe model
- Compile time
- parser
- compiler
- loadable

Run time

- Application
- User Mode Driver
- Kernel Mode Driver
- DLA hardware

©2018 NVIDIA CORPORATION
Inserting video: Insert/Video/Video from File.

Insert video by browsing your directory and selecting OK.

File types that work best in PowerPoint are mp4 or wmv.
There are lots of custom chips you can build with NVDLA
Come to my Keynote talk “Opportunities and Challenges of Building Silicon in the Cloud” tomorrow morning at 9:20am!
SiFive Freedom Chip Platforms

- SiFive Tapes Out Multiple Base Platforms
  - Demonstrates silicon capability of each platform
  - Enables RISC-V software development
  - Reduces risk for customer
  - Proves our and matures Design flow for each platform

- Customization Capabilities
  - Add/remove DesignShare and SiFive IP
  - Customization of SiFive CPU IP
  - Customer can add own IP into Platform (accelerators, co-processors, other IP)

- From Prototype to Production
  - SiFive handles all logistics, incl. fab, package, test
  - SiFive scales to production
  - Final delivery is packaged, tested, qualified Silicon

Next talk in the same room will introduce brand new Freedom Revolution Chip Platform with HBM2 and 56-112Gb/s SerDes
Freedom Unleashed 64-bit Multi-Core RISC-V Linux Platform

• 1.5+ GHz U54-MC SiFive CPU
  • 1x E51: 16KB L1I$, 8KB DTIM with ECC support
  • 4x U54: 32KB L1I$, 32KB L1D$ with ECC support
  • Single- and Double-precision floating-point support
  • 2MB Banked L2$ with directory-based cache-coherence & ECC support

• ChipLink
  • Serialized Chip-to-Chip Coherent TileLink Interconnect

• DDR3/4, GbE, Peripherals

Freedom U540, FCBGA, manufactured in TSMC 28nm
HiFive Unleashed: World’s First Multi-Core RISC-V Linux Dev Board

- SiFive FU540-C000 (built in 28nm)
- 8 GB 64-bit DDR4 with ECC
- Gigabit Ethernet Port
- 32 MB Quad SPI Flash
- MicroSD card for removable storage
- MicroUSB for debug and serial communication
- Digital GPIO pins
- FMC connector for future expansion with add-in cards
HiFive Unleashed with Microsemi PolarFire Expansion Board
HiFive Unleashed with Xilinx VCU118 Evaluation Kit
Freedom Development Kit Comes with Linux BSP Based on Debian or Fedora
DEMO Setup: HiFive Unleashed + NVDLA

- **NVDLA small config**
  - 2048 MACs, 512 KB
- **NVDLA mapped onto Xilinx VU118 Evaluation Kit**
- **NVDLA running open-source YOLOv3 object recognition**
- **Linux OS running on HiFive Unleashed**
  - Easy to port over umd/kmd from ARM
- **Demo setup built with OpenCV thanks to Debian**
Check out HiFive Unleashed + NVDLA demo at SiFive Booth!
Customize your Freedom Chip with NVDLA Today!

- Open-source IP cores further lower the bar to implement RISC-V-based products
- Freedom chip platform offers a complete template SoC with software support
- Freedom Unleashed + NVDLA is a great starting point for smart IoT SoCs and devices

- Everything is open-sourced, so check it out and contribute yourself!
  - NVDLA
    - https://github.com/nvdla/hw
    - https://github.com/nvdla/sw
    - http://nvdla.org
  - Freedom Platform
    - https://github.com/sifive/freedom
    - https://github.com/sifive/nvidia-dla-blocks

- Once you’re ready, please come talk to us for your RISC-V AI chip needs!
End