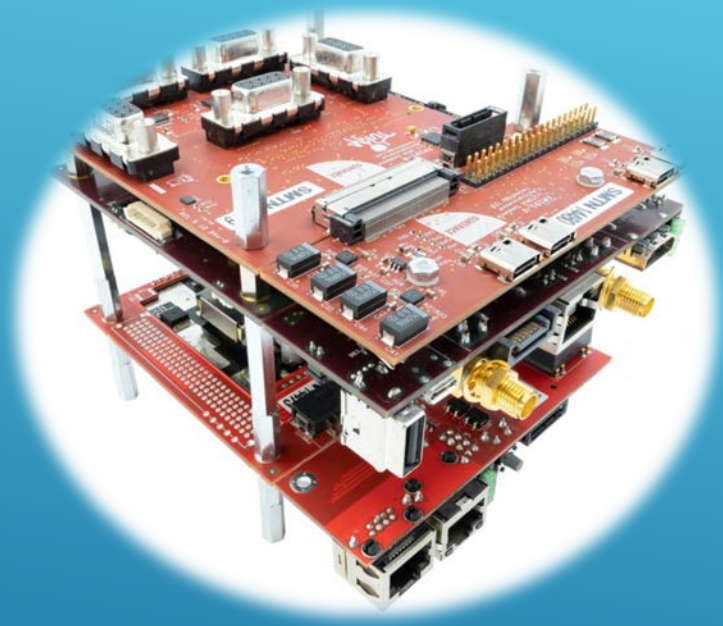


EM3V - Embedded Vision



Sundance Multiprocessor Technology, Ltd.

Pedro Machado pedro.m@sundance.com

Fatima Kishwar fatima.k@sundance.com

Flemming Christensen flemming.c@sundance.com

OVERVIEW

- Company profile
- Introduction
- Technologies
- VCS-1 (EMC2) System
- Discussion
- Future Work



SERVICE AND EXCELLENCE!

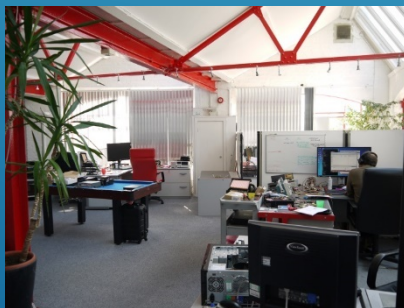
SUNDANCE

Established in 1989 by Flemming CHRISTENSEN

- Employee Owned and a 'Life-Style' company
- 10x people with 300+ years experience
 - 4x with accredited Xilinx FPGA training
- Always designed and built our own products
- BSI - ISO9001-2015 certified, since 2003

Techology Focus

- Acceleration, Vision, Sensor & Robotics



MODULAR, RECONFIGURABLE - LIKE LEGO®

SUNDANCE

- Reduced time-to-market
- Rapid Prototyping
- Reconfigurable
- Flexible
- Modular
- Scalable
- Reliable



Year 1989

TURN-KEY SYSTEM SOLUTIONS FROM A-Z

• Commercial-of-the-Shelves Systems (COTS)

- Flexible and Upgradeable
- Design for Excellence
- Maintenance with ease



• Custom Bespoke Systems

- Hazardous Environment
- Custom Integration Service
- Design of Enclosures



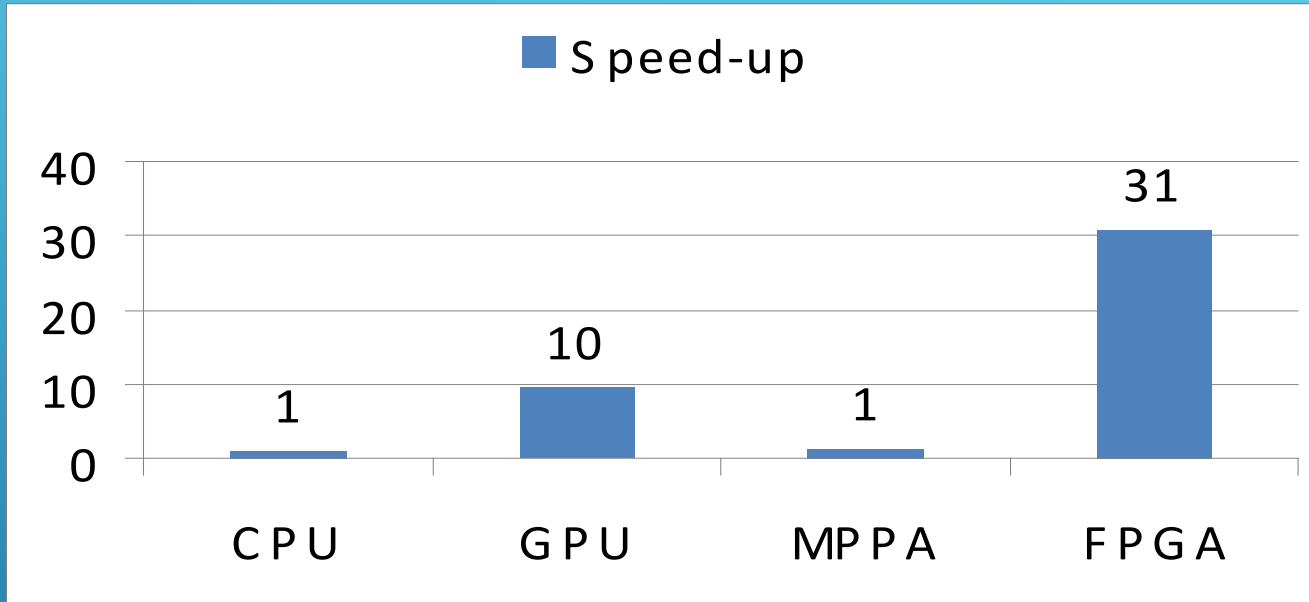
INTRODUCTION

- ▶ Increasing demand for High Performance Computing
 - ▶ Everyone wants more compute-power
 - ▶ Finer time-steps; larger data-sets; better models
- ▶ Decreasing single-threaded performance
 - ▶ Emphasis on multi-core CPUs and parallelism
 - ▶ *Do computational biologists need to learn PThreads?*
- ▶ Increasing focus on power and space
 - ▶ Boxes are cheap: 16 node clusters are very affordable
 - ▶ *Where do you put them? Who is paying for power?*
- ▶ How can we use hardware acceleration to help?

TYPES OF HARDWARE ACCELERATOR

- ▶ GPU : Graphics Processing Unit
 - ▶ Many-core - 30 SIMD processors per device
 - ▶ High bandwidth, low complexity memory – no caches
- ▶ MPPA : Massively Parallel Processor Array
 - ▶ Grid of simple processors – 300 tiny RISC CPUs
 - ▶ Point-to-point connections on 2-D grid
- ▶ FPGA : Field Programmable Gate Array
 - ▶ Fine-grained grid of logic and small RAMs
 - ▶ Build whatever you want

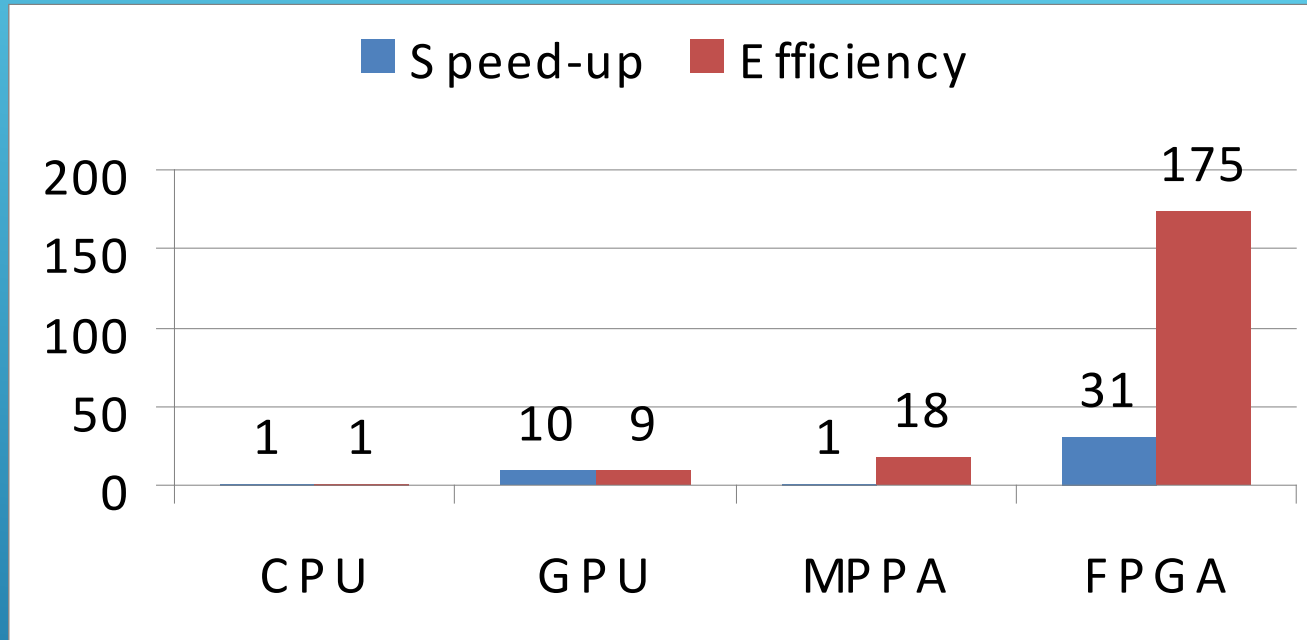
HARDWARE ADVANTAGES: PERFORMANCE



- More parallelism - more performance
- GPU: 30 cores, 16-way SIMD
- MPPA: 300 tiny RISC cores
- FPGA: hundreds of parallel functional units

A Comparison of CPUs, GPUs, FPGAs, and MPPAs for Random Number Generation,
D. Thomas, L. Howes, and W. Luk , In Proc. of FPGA, pgs. 22-24, 2009

HARDWARE ADVANTAGES: POWER



- GPU: 1.2GHz - same power as CPU
- MPPA: 300MHz - Same performance as CPU, but 18x less power
- FPGA: 300MHz - faster and less power

A Comparison of CPUs, GPUs, FPGAs, and MPPAs for Random Number Generation,
D. Thomas, L. Howes, and W. Luk , In Proc. of FPGA, pgs. 22-24, 2009

FPGA ACCELERATED APPLICATIONS

▶ Finance

- ▶ 2006: Option pricing: 30x CPU
- ▶ 2007: Multivariate Value-at-Risk: 33x Quad CPU
- ▶ 2008: Credit-risk analysis: 60x Quad CPU

▶ Bioinformatics

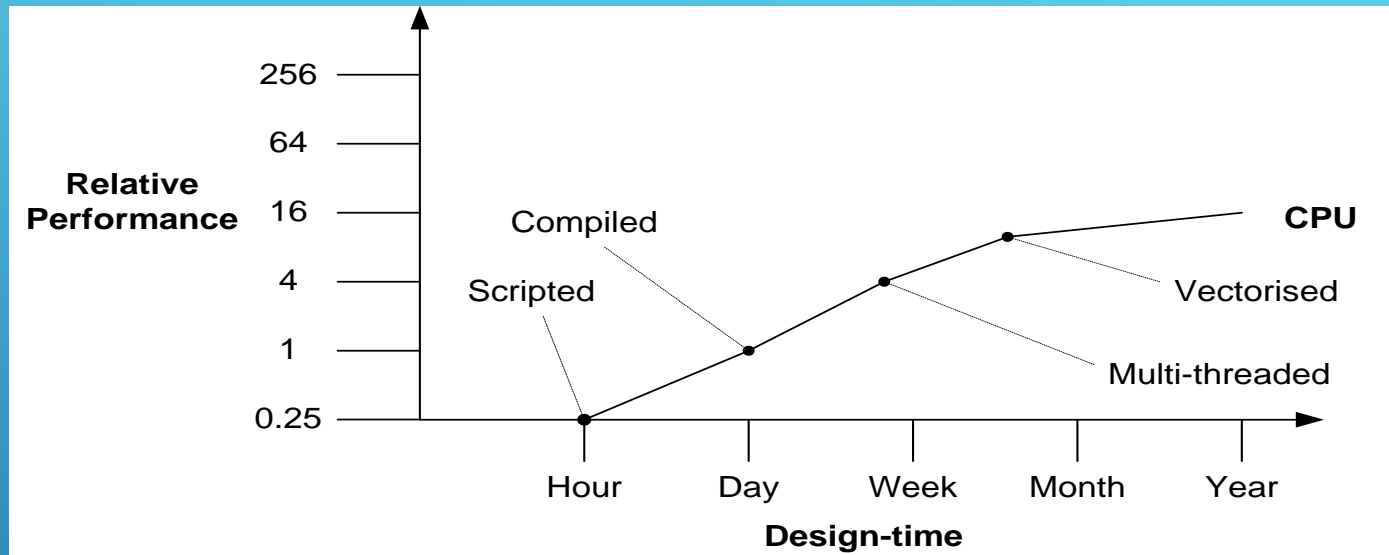
- ▶ 2007: Protein Graph Labelling: 100x Quad CPU

▶ Neural Networks

- ▶ 2008: Spiking Neural Networks: 4x Quad CPU
1.1x GPU

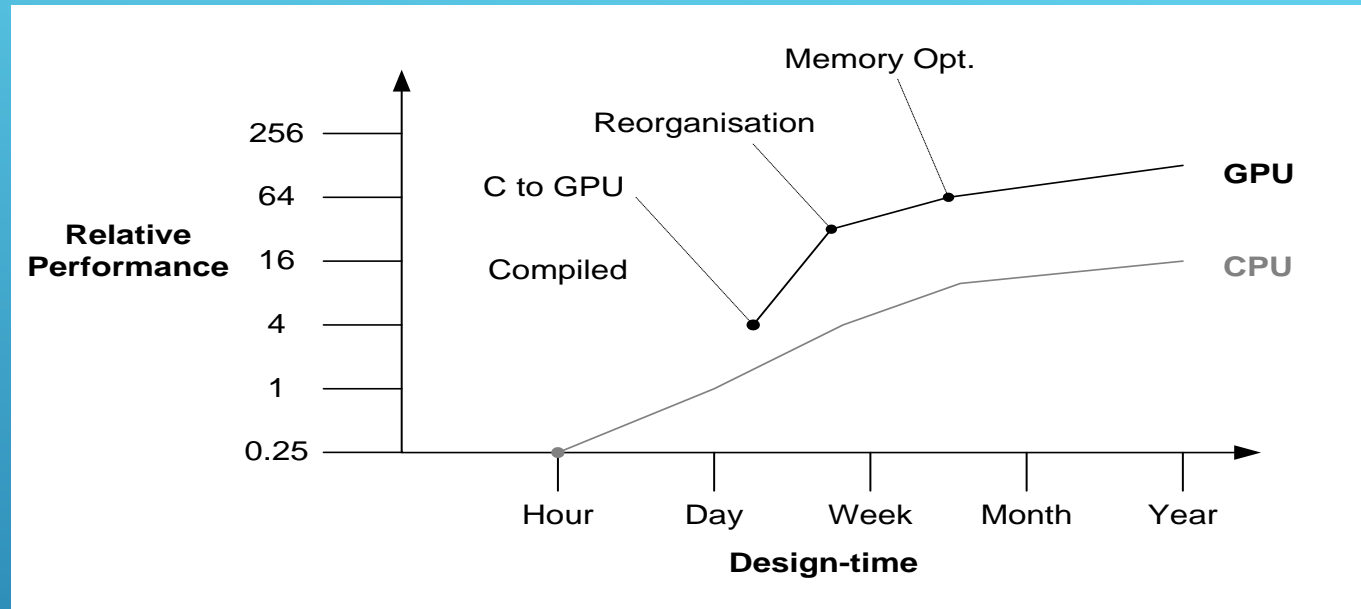
All with less than a fifth of the power

PROBLEM: DESIGN EFFORT



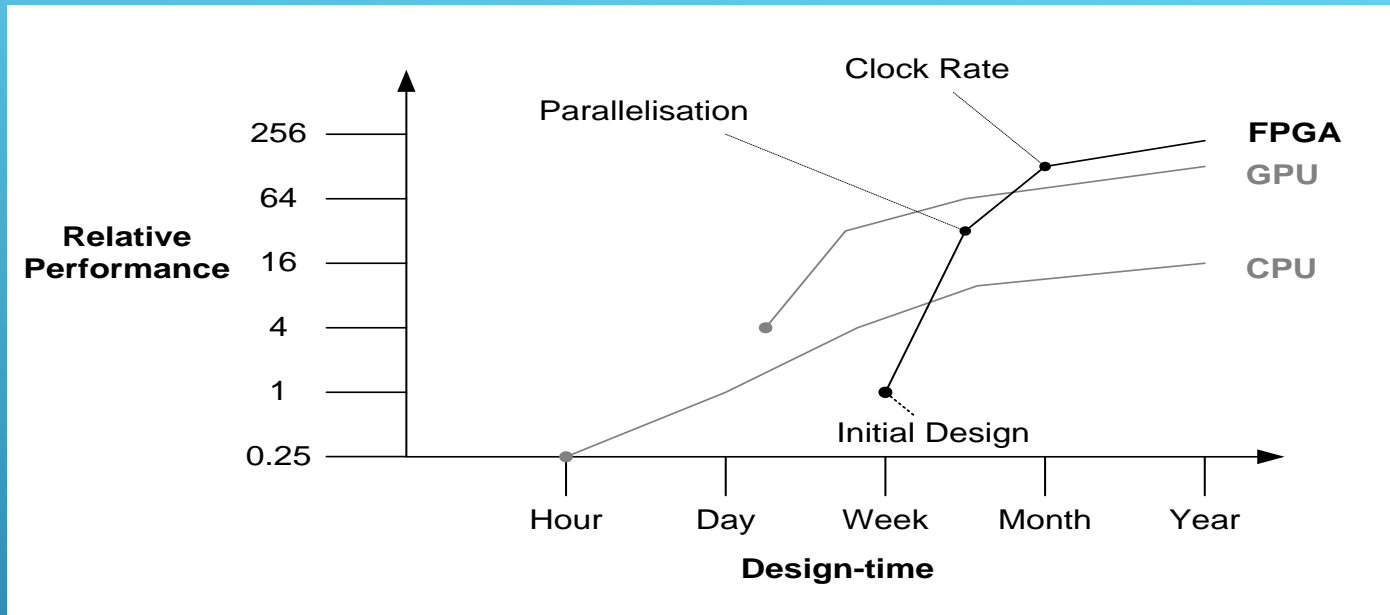
- ▶ Researchers love scripting languages: Matlab, Python, Perl
 - ▶ Simple to use and understand, lots of libraries
 - ▶ Easy to experiment and develop promising prototype
- ▶ Eventually prototype is ready: need to scale to large problems
 - ▶ Need to rewrite prototype to improve performance: e.g. Matlab to C
 - ▶ Simplicity of prototype is hidden by layers of optimisation

PROBLEMS: DESIGN EFFORT



- ▶ GPUs provide a somewhat gentle learning curve
 - ▶ CUDA and OpenCL *almost* allow compilation of ordinary C code
- ▶ User must understand GPU architecture to maximise speed-up
 - ▶ Code must be radically altered to maximise use of functional units
 - ▶ Memory structures and accesses must map onto physical RAM banks
- ▶ We are asking the user to learn about things they don't care about

PROBLEMS: DESIGN EFFORT



- ▶ FPGAs provide large speed-up and power savings – at a price!
 - ▶ Days or weeks to get an initial version working
 - ▶ Multiple optimisation and verification cycles to get high performance
- ▶ Too risky and too specialised for most users
 - ▶ Months of retraining for an uncertain speed-up
- ▶ Currently only used in large projects, with dedicated FPGA engineer

GOAL: FPGAS FOR THE MASSES

- ▶ Accelerate niche applications with limited user-base
 - ▶ Don't have to wait for traditional "heroic" optimisation
- ▶ Single-source description
 - ▶ The prototype code is the final code
- ▶ Encourage experimentation
 - ▶ Give users freedom to tweak and modify
- ▶ Target platforms at multiple scales
 - ▶ Individual user; Research group; Enterprise
- ▶ Use domain specific knowledge about applications
 - ▶ Identify bottlenecks: optimise them
 - ▶ Identify design patterns: automate them
 - ▶ *Don't try to do general purpose "C to hardware"*

WHICH TECHNOLOGY?

Huge Design Space

- Which Accelerator?
- Which Brand?
- Number of cores?
- Which Device?
- Device Cost?
- Design time?
- Which algorithm?
- Use case optimization?

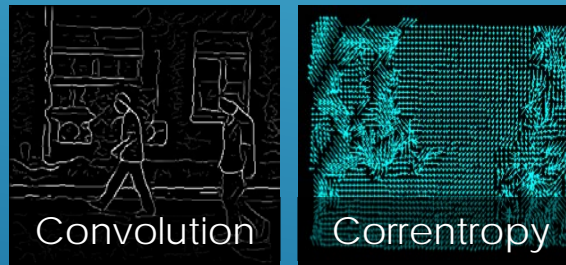
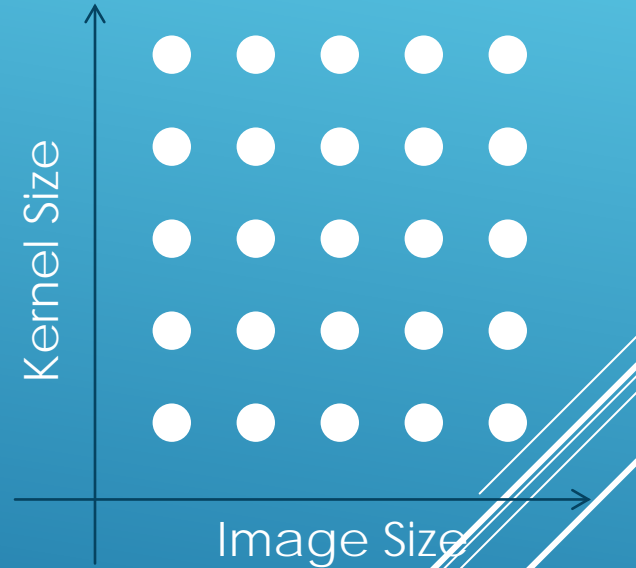
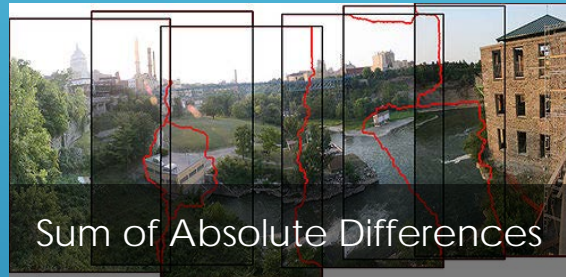
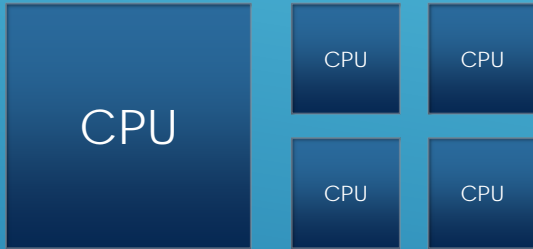
- ▶ Clear architectural trend of parallelism and heterogeneity
- ▶ Heterogeneous devices have many tradeoffs
- ▶ Usage cases also affect best device choice
- ▶ **Problem:** huge design space

TYPICAL CV APPLICATIONS: SLIDING WINDOW

Devices

Algorithms

Use Cases



- ▶ **Contribution:** thorough analysis of devices and use cases for sliding window applications
- ▶ Sliding window used in many domains, including image processing and embedded

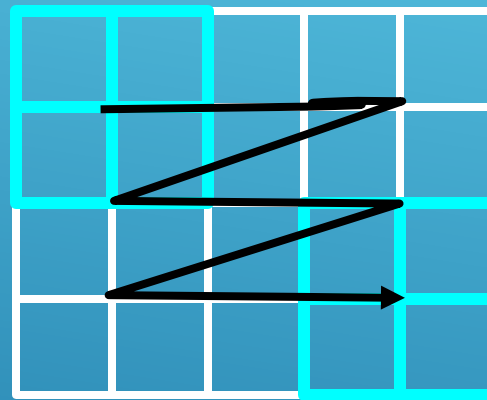
TYPICAL CV APPLICATIONS: SLIDING WINDOW APPLICATIONS

```

Input: image of size  $x \times y$ , kernel of size  $n \times m$ 
for (row=0; row < x-n; row++) {
  for (col=0; col < y-m; col++) {
    // get  $n \times m$  pixels (i.e., windows
    // starting from current row and col)
    window=image[row:row+n-1][col:col+m-1]
    output[row][col]=f(window,kernel)
  }
}

```

Window 1



Window

Kernel



Window Function



Window W-1

Output Pixel

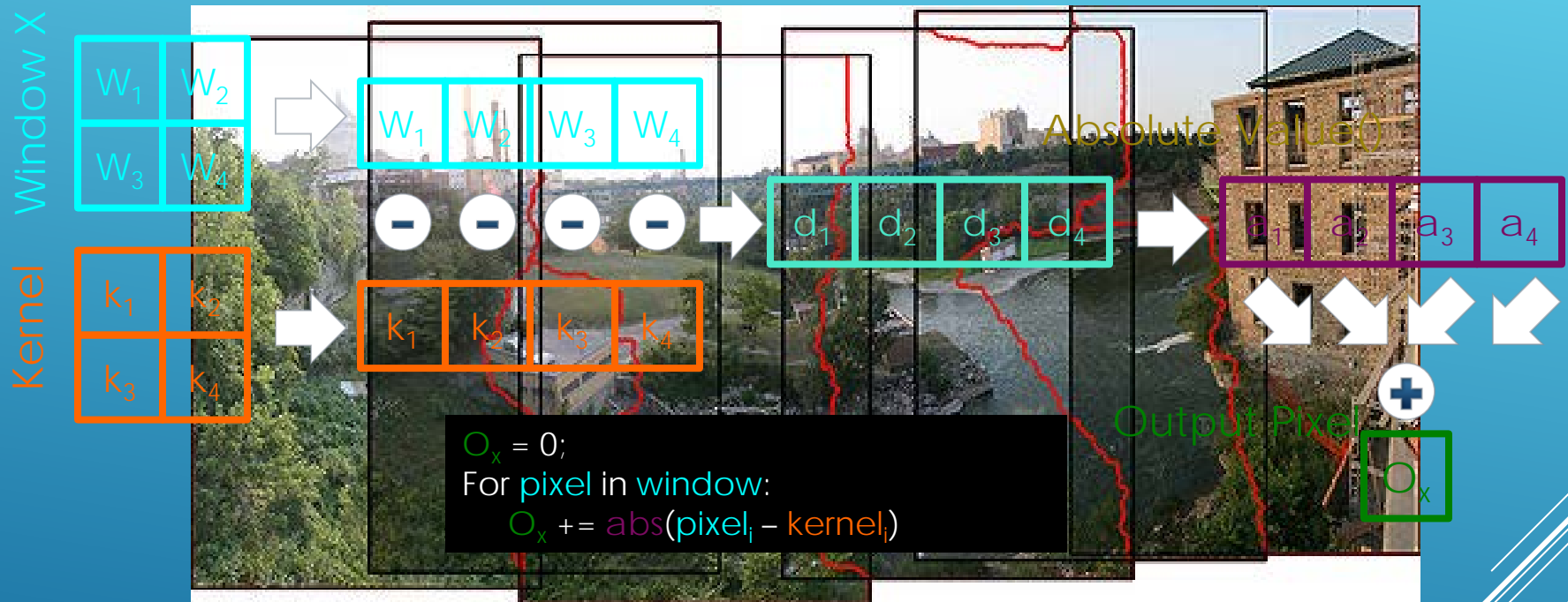
- ▶ Consider a 2D Sliding Window with 16-bit grayscale image inputs
- ▶ Applies window function against a window from image and the kernel
- ▶ “Slides” the window to get the next input
- ▶ Repeats for every possible window

▶ 15x15 kernel on 1080p 30-FPS video = 120 billion memory accesses/second

7/16/2018

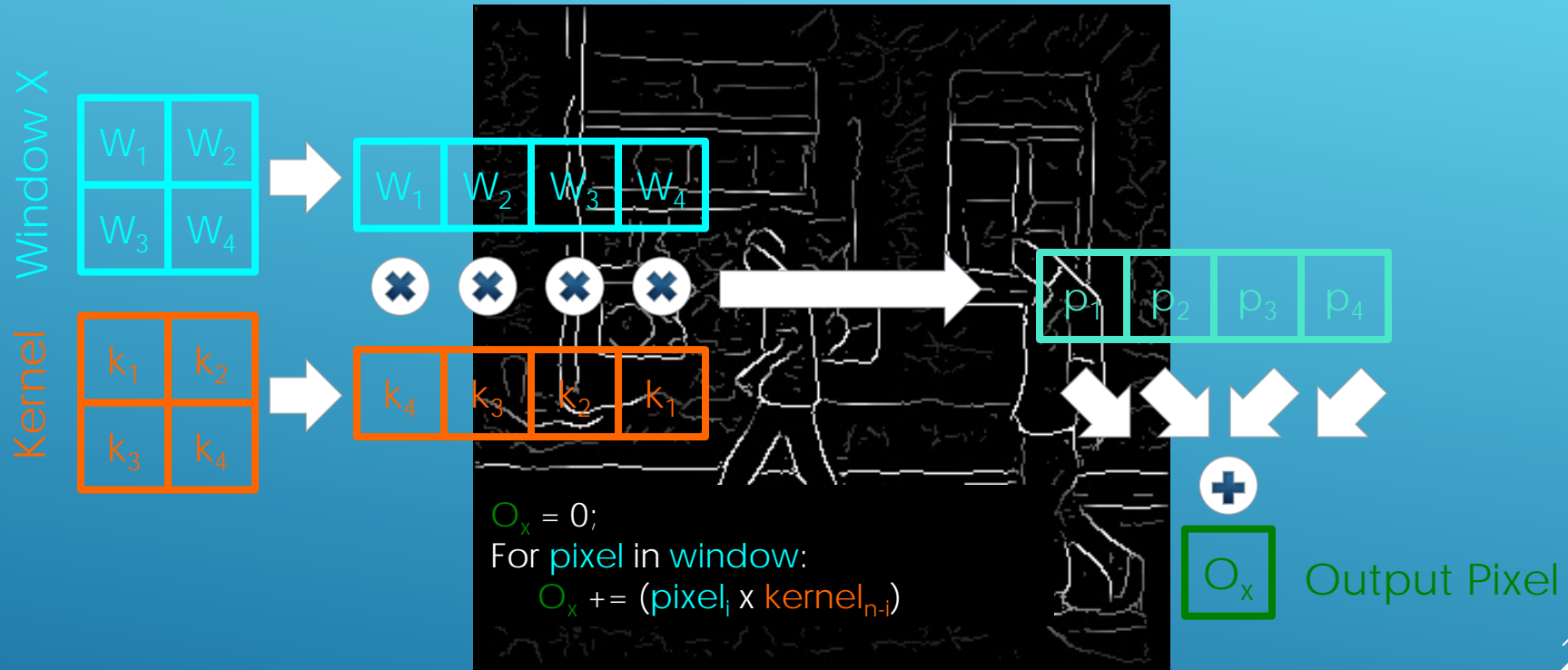
17

APP 1: SUM OF ABSOLUTE DIFFERENCES (SAD)



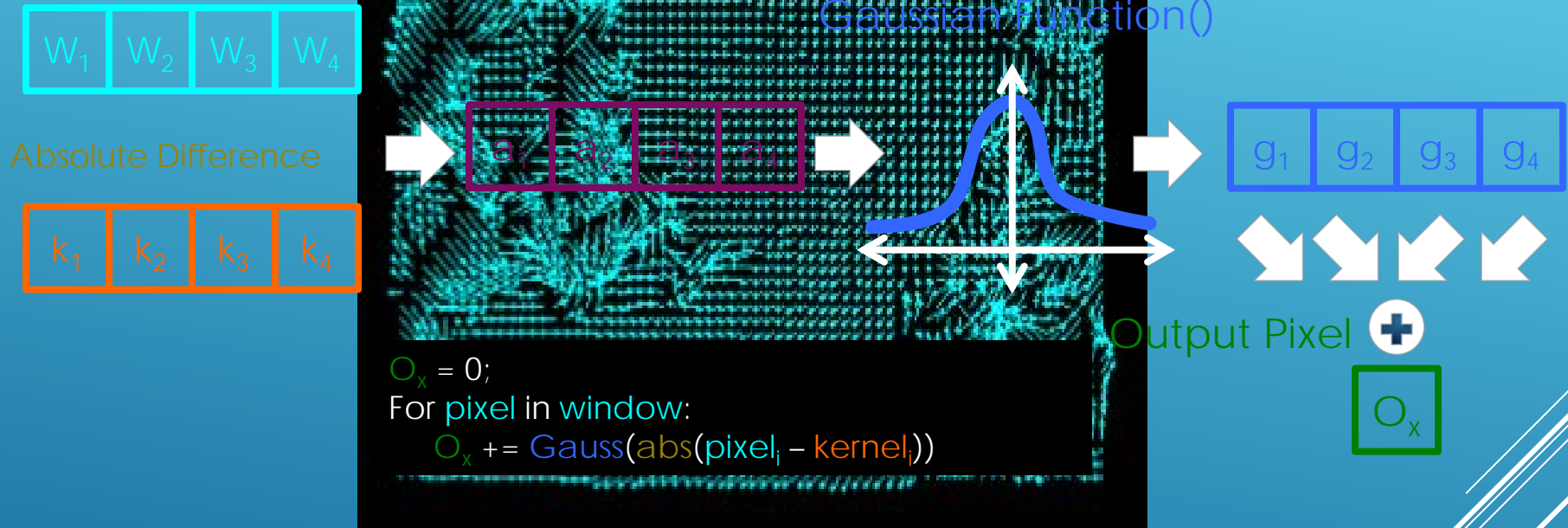
- ▶ **Used for:** H.264 encoding, object identification
- ▶ **Window function:** point-wise absolute difference, followed by summation

APP 2: 2D CONVOLUTION



- ▶ Used for: filtering, edge detection
- ▶ Window function: point-wise product followed by summation

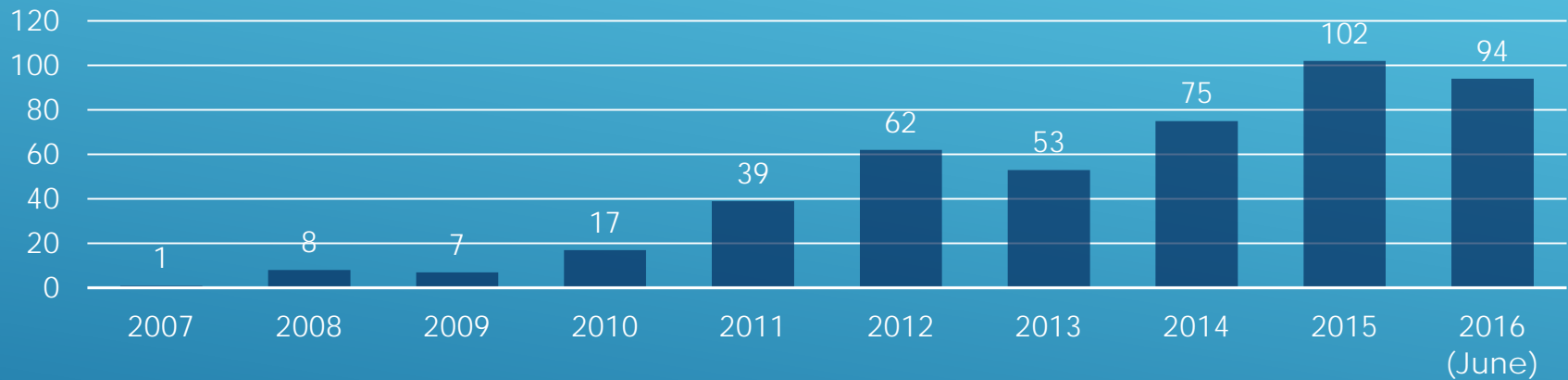
APP 3: CORRENTROPY



- ▶ **Used for:** optical flow, obstacle avoidance
- ▶ **Window function:** Gaussian of point-wise absolute difference, followed by summation

HETEROGENEOUS COMPUTING SYSTEM IN TOP500 LIST

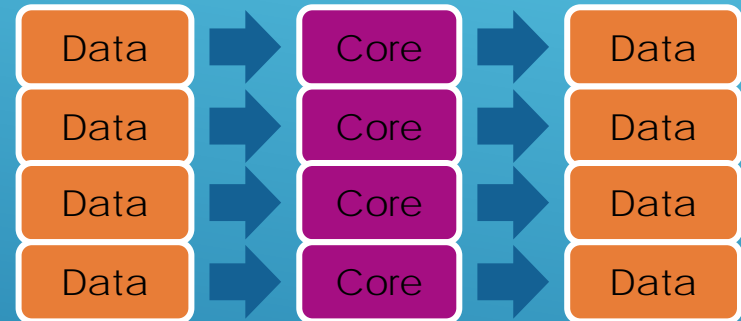
Heterogeneous computing system in Top500



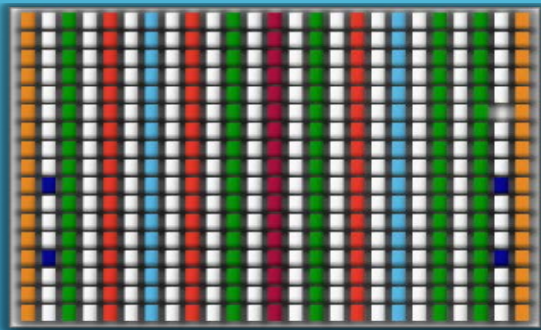
- ▶ Reason: Significant performance/energy-efficiency boost from GPU/CPU

GPU: SPECIALIZED ACCELERATOR FOR A SET OF APPLICATIONS

- ▶ Specialized accelerator for data-parallel applications
 - ▶ Optimized for processing massive data
- ▶ Give up unrelated goal and features
 - ▶ Give up optimizing latency for processing single data
 - ▶ Give up branch prediction, out-of-order execution
 - ▶ Give up large traditional cache hierarchy
- ▶ More resource for parallel are processing
 - ▶ More cores, more ALU



CREATING APPLICATION-SPECIFIC ACCELERATOR WITH FPGA



Virtex®-7 FPGA

	Logic Fabric LUT-6 CLB		Precise, Low Jitter Clocking MMCMs		On-Chip Memory 36Kbit/18Kbit Block RAM
	Enhanced Connectivity PCIe® Interface Blocks		DSP Engines DSP48E1 Slices		Hi-perf. Parallel I/O Connectivity SelectIO™ Technology
	Hi-performance Serial I/O Connectivity Transceiver Technology				

- ▶ Only provides primitive building blocks for computation
 - ▶ Register, addition/multiplication , memories, **programmable** Boolean operations and connections
- ▶ Build application-specific accelerator from primitives building blocks
 - ▶ Interconnection between primitive functional units
 - ▶ Timing of data movement between primitive functional units
- ▶ Opportunities for optimizations for a specific application!
 - ▶ Maximizing efficiency while throwing away redundancy

THE CHALLENGES OF PROMOTING FPGAS AMONG SOFTWARE ENGINEERS

- ▶ Require tremendous efforts
- ▶ Extensive knowledge of digital circuit design

AXI Master **Timing Closure**

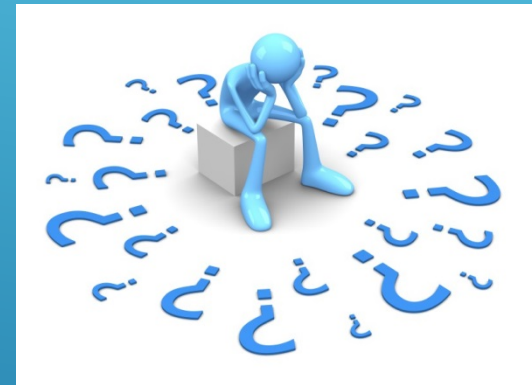
Burst inference

DSP48

Stable interface

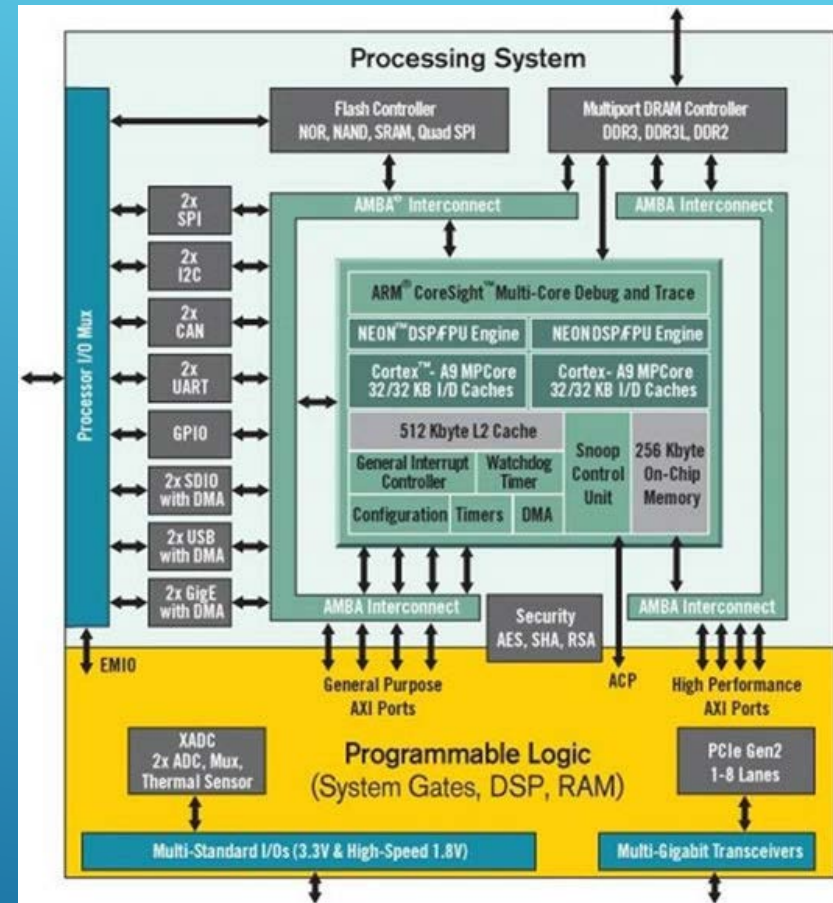
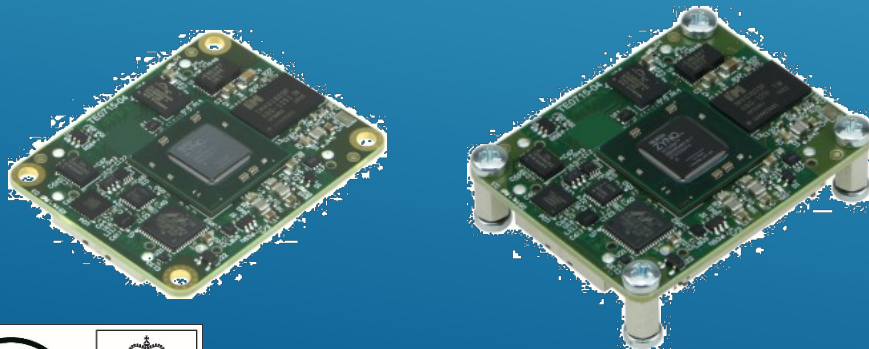
Loop rewind

- ▶ The potential of FPGAs is not easily accessible by common software engineers



HYBRID ARQUITECTURE

- Zynq-7000 devices are equipped with dual-core ARM Cortex-A9 processors integrated with 28nm Artix-7 or Kintex®-7 based programmable logic for excellent performance-per-watt and maximum design flexibility.
- Sundance's EMC2 carrier board is Compatible with all the Zynq-7000 Series.

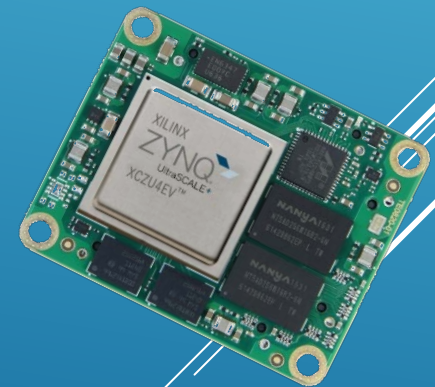
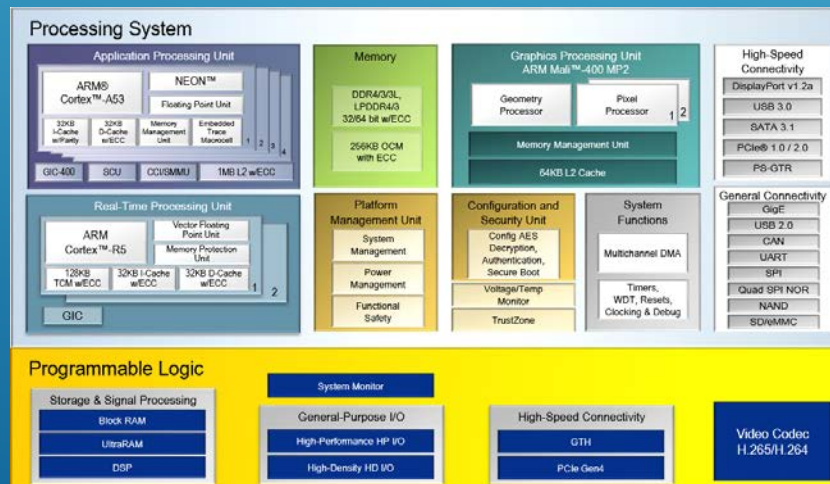


WHAT IS THE BEST SOLUTION?

SUNDANCE



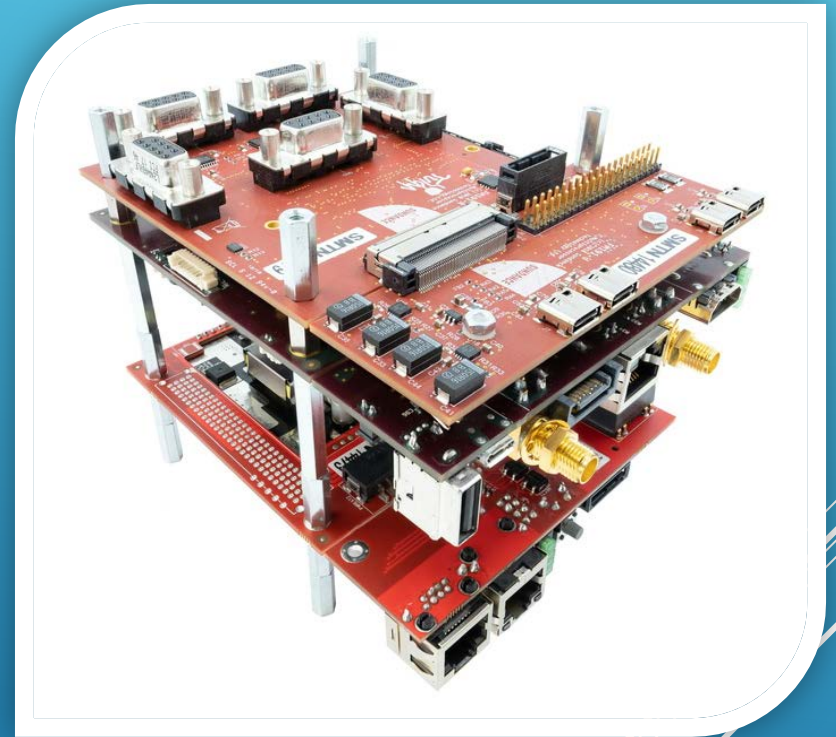
- Zynq® UltraScale+™ MPSoC devices provide 64-bit processor scalability while combining real-time control with soft and hard engines for graphics, video, waveform, and packet processing.
- Sundance's EMC2 carrier board is compatible with all the Zynq® UltraScale+™ MPSoC. Our focus is now on the XCZU4EV device (automotive grade).



VCS-1 (EMC2) HARDWARE FEATURES

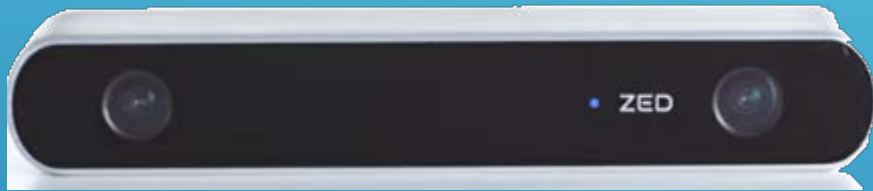
Connectivity:

- ▶ FM191-R; FMC-LPC to:
 - ▶ 15x Digital I/Os [DB9]
 - ▶ 12x Analogue Inputs [DB9]
 - ▶ 8x Analogue Outputs [DB9]
 - ▶ 1x Expansion [SEIC]
- ▶ FM191-U; SEIC to:
 - ▶ 4x USB3.0 [USB-c]
 - ▶ 28x GPIO [40-pin GPIO]
- ▶ FM191-A1; 40-pin GPIO
 - ▶ 28x GPIO [DB9]



VCS-1 (EMC2) SENSORS COMPATIBILITY

The ZU4EV MPSoC is compatible with a wide range of sensors.



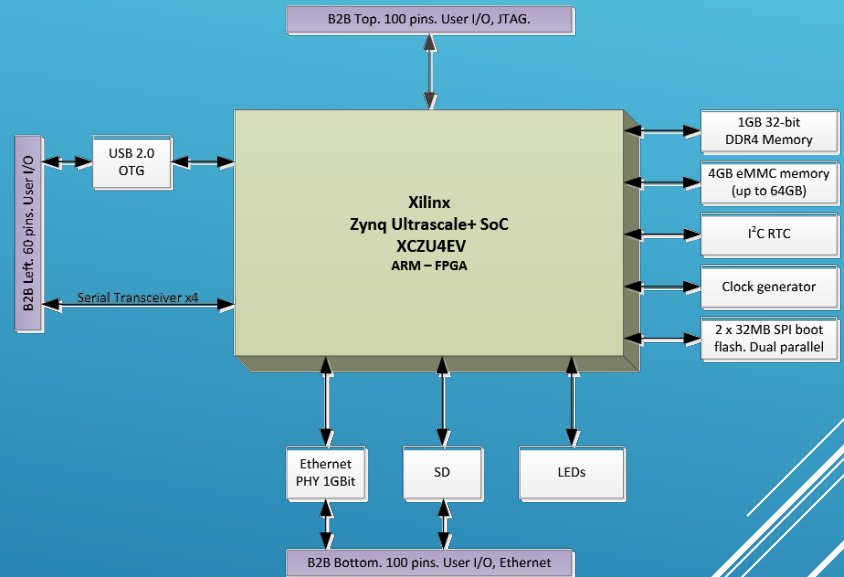
Depth sensor (up to 20m). Interface USB3.0



Thermal imaging
Interface: Eth



Movidius Neural Compute Stick
Interface: USB3.0

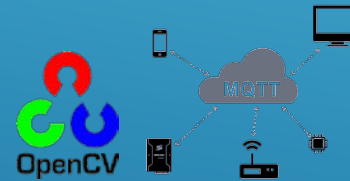


JAI AD-130GE
Interface: Eth

VCS-1 (EMC2) COMPATIBILITY

VCS-1 features:

- ▶ Raspberry PI and Arduino compatible;
- ▶ Compatible with most of the Arduino/RPI sensors and actuators;
- ▶ 4x USB3.0 ports for interfacing with a wide range of sensors;
- ▶ MQTT and OpenCV compatible
- ▶ ROS compatible
- ▶ ROS2 ready
- ▶ HIPPEROS ready

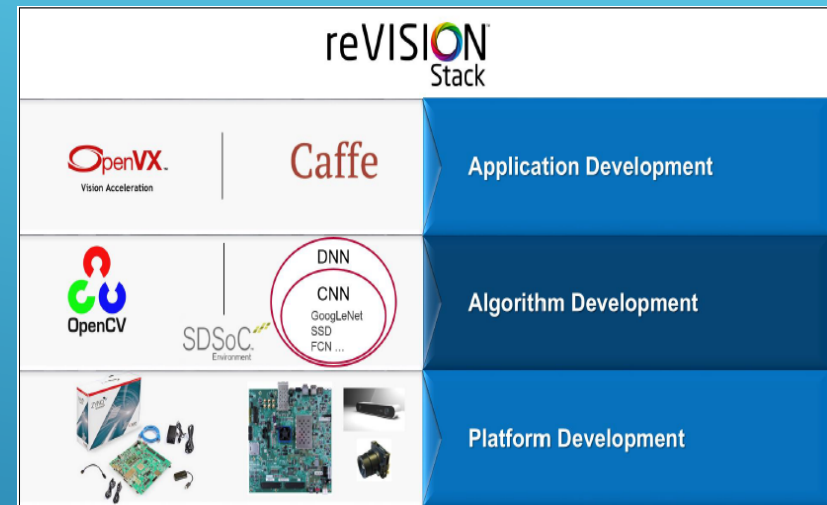


DEEP LEARNING ON THE VCS-1 (EMC2)

The VCS-1 will be fully compatible with the Xilinx reVision stack.

- ▶ Includes support for the most popular neural networks including AlexNet, GoogLeNet, VGG, SSD, and FCN.
- ▶ Optimized implementations for CNN network layers, required to build custom neural networks (DNN/CNN)

Xilinx reVISION stack



VCS-1 (EMC2) OPEN SOURCE SOFTWARE AND FIRMWARE

Open Source Hardware/software and online documentation:

- ▶ Open Hardware Repository

<https://www.ohwr.org/projects/emc2-dp>

- ▶ Microsoft Windows/Linux 64-bit SDK

https://github.com/SundanceMultiprocessorTechnology/VCS-1_SDK

- ▶ FM191 ARM SDK

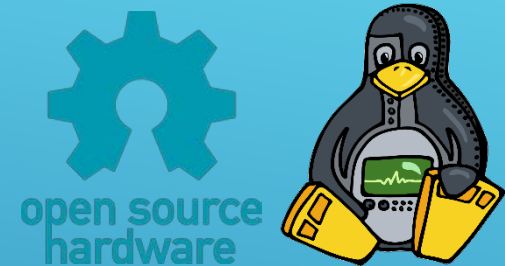
https://github.com/SundanceMultiprocessorTechnology/VCS-1_FM191_SDK

- ▶ FM191 Firmware

https://github.com/SundanceMultiprocessorTechnology/VCS-1_FM191_FW

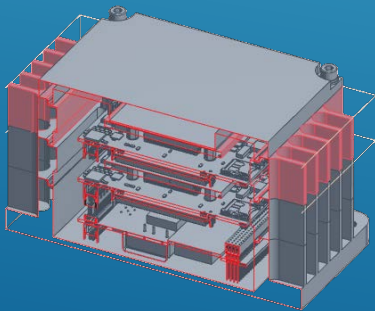
- ▶ EMC2 ROS

▶ https://github.com/SundanceMultiprocessorTechnology/VCS-1_emc2_ros



VCS-1(EMC2) ENCLOSURE

A custom enclosure was specially designed for accommodating the VCS-1 system.



DISCUSSION

The VCS-1 has the following characteristics:

1. High performance (24V@0.595A)
2. Low power consumption
3. Highly compatible with a wide range of commercially available sensors and actuators
4. Highly optimised for computer vision applications
5. Fully reconfigurable



WHAT NEXT?

- ▶ Be part of our customers family by ordering our products or hiring our services.
- ▶ Sundance University Program (SUP)
 - ▶ Access to hardware prototypes
 - ▶ Advisory Board Members
 - ▶ BSC/MSc/PhD Internships
- ▶ Funding capture
 - ▶ H2020
 - ▶ InnovateUK
 - ▶ EPSRC

Know more about SUP and on-going projects

<https://www.sundance.com/sundance-in-eu-projects-programs/>



UNIVERSITY CLIENTS

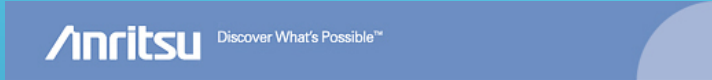
SUNDANCE



7/16/2018

INDUSTRIAL CLIENTS

SUNDANCE



7/16/2018

EM3V - Embedded Vision

QUESTIONS?



Sundance Multiprocessor Technology, Ltd.

Pedro Machado pedro.m@sundance.com

Fatima Kishwar fatima.k@sundance.com

Flemming Christensen flemming.c@sundance.com

7/16/2018