GTC 2015 – Session S5429

Creating Dense Mixed GPU and FPGA Systems With Tegra K1s Using OpenCL & CUDA

Lance Brown, Director - HPC

ColoradoEngineering.com

Lance.brown@coloradoengineering.com

719-641-7287 Cell

We Can Solve Really Cool Problems Now

- Heterogeneous computing is more than CPU + GPU
- ARM processors changed the game
  - NVIDIA - GPU + ARM - CUDA
  - TI - DSP + ARM - OpenCL
  - Altera - FPGA + ARM – OpenCL
- Scalable from handheld to Enterprise & HPC

Why Listen to CEI?

- Been using FPGAs since 1985
- Been solving massively parallel problems for over 30 years
- We have/are designing multiple 24 & 32 layer boards featuring Altera FPGAs & NVIDIA GPUs
- Early adopter of new technologies and experts at marrying existing technologies in new ways

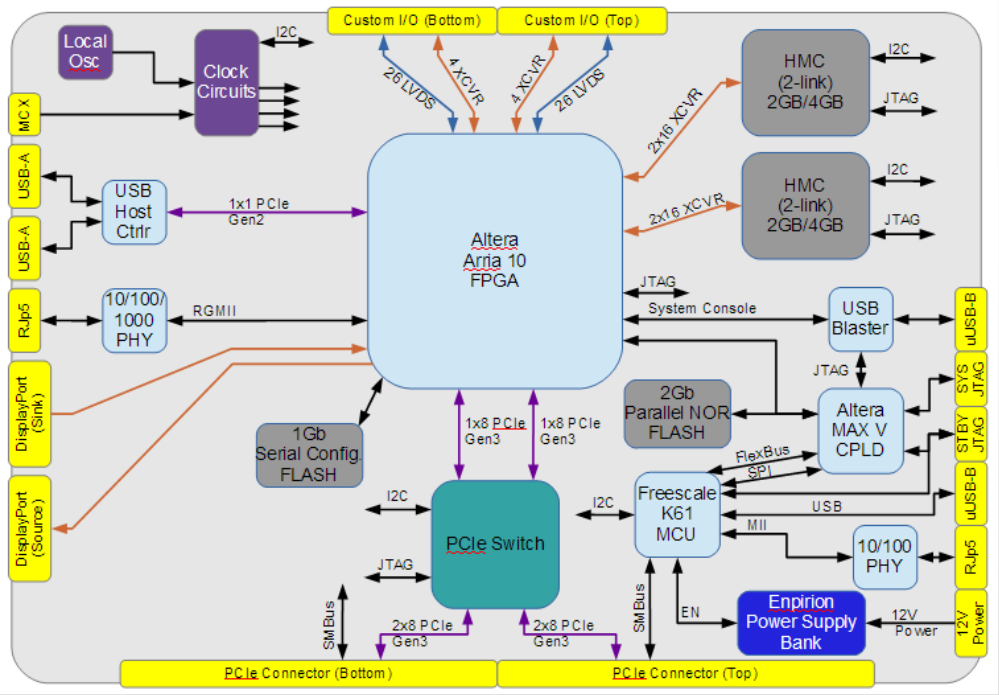
Game Changer #1
Altera's Hard Floating Point Unit IP & OpenCL

- FPGAs have traditionally supported soft floating point
- Altera introduced IEEE 754 Hard Floating Point with Arria 10
- Arria 10 FPGAs are rated from 140 GigaFLOPS (GFLOPS) to 1.5 TeraFLOPS (TFLOPS)
- Details at: https://www.altera.com/en_US/pdfs/literature/po/bg-floating-point-fpga.pdf
- OpenCV & Suricata Implementations Using OpenCL
- Partial Reconfiguration for Streamlined OpenCL Development
- On Intel's 14 nm FinFET Fab

Game Changer #2
NVIDIA Makes Tegra K1 Available

- GPU + ARM @ low power
- Very important – camera interfaces galore
- Can do significant processing at each edge node now
- Jetson Kit – awesome eval kit & affordable
- More importantly – chipset available through Arrow!
- Details at: https://developer.nvidia.com/hardware-design-and-development

CEI's Epiphany – Ultimate CV Platform
Altera Arria 10 & NVIDIA Tegra K1?

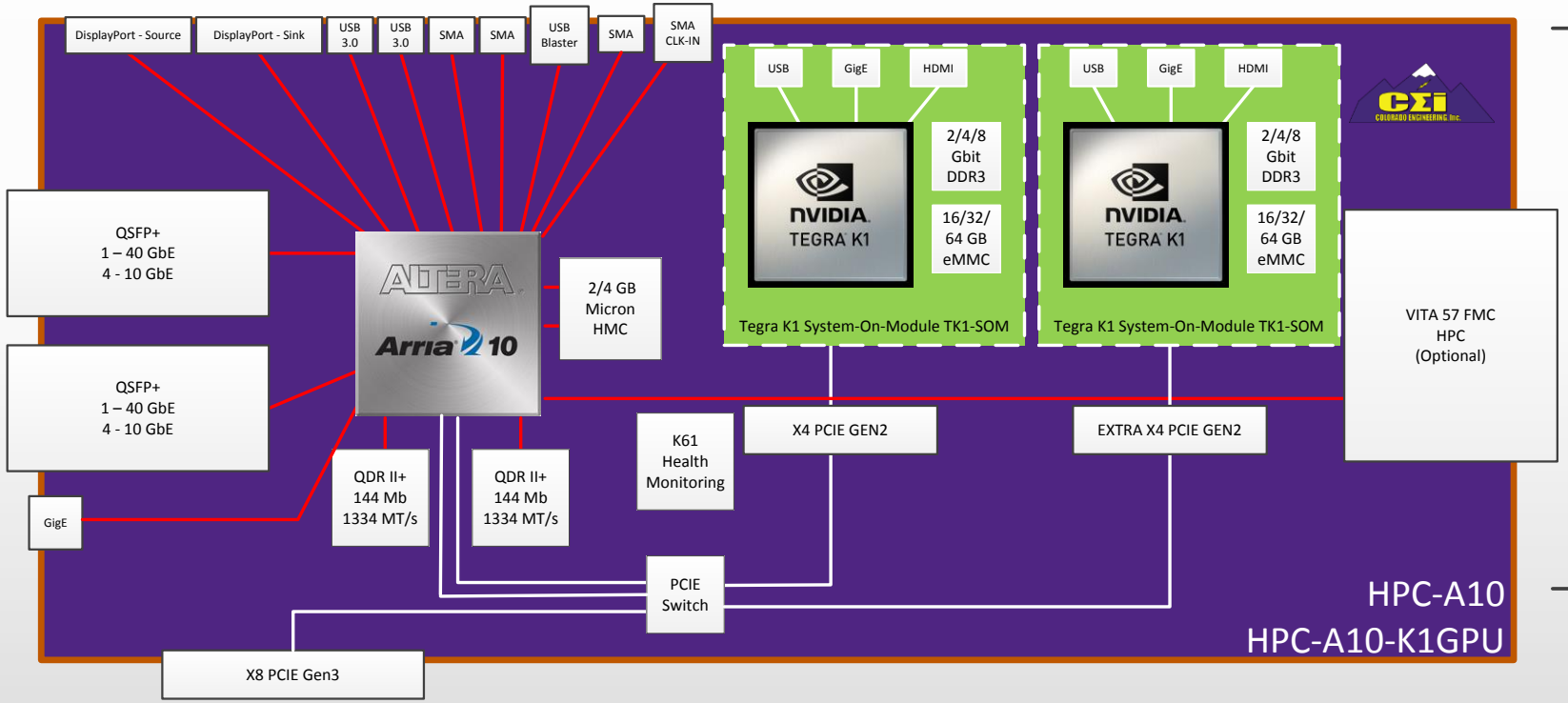


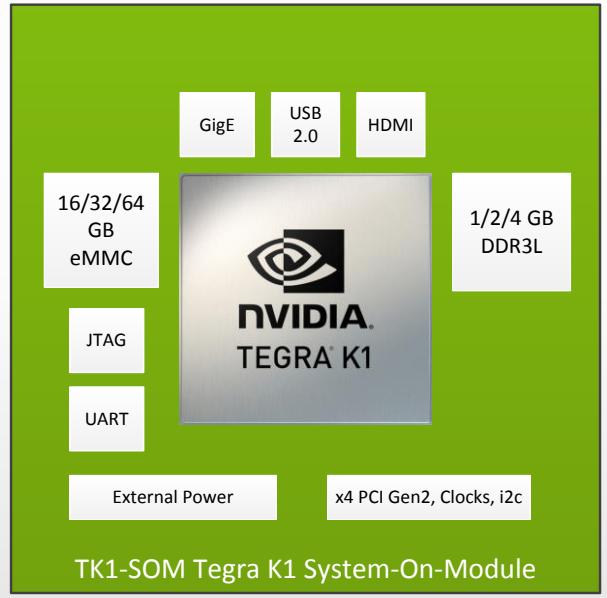
+

1500 GFLOPS

326 GFLOPS

First Union – Dual TK1s + Arria 10 HPC-A10-K1GPU





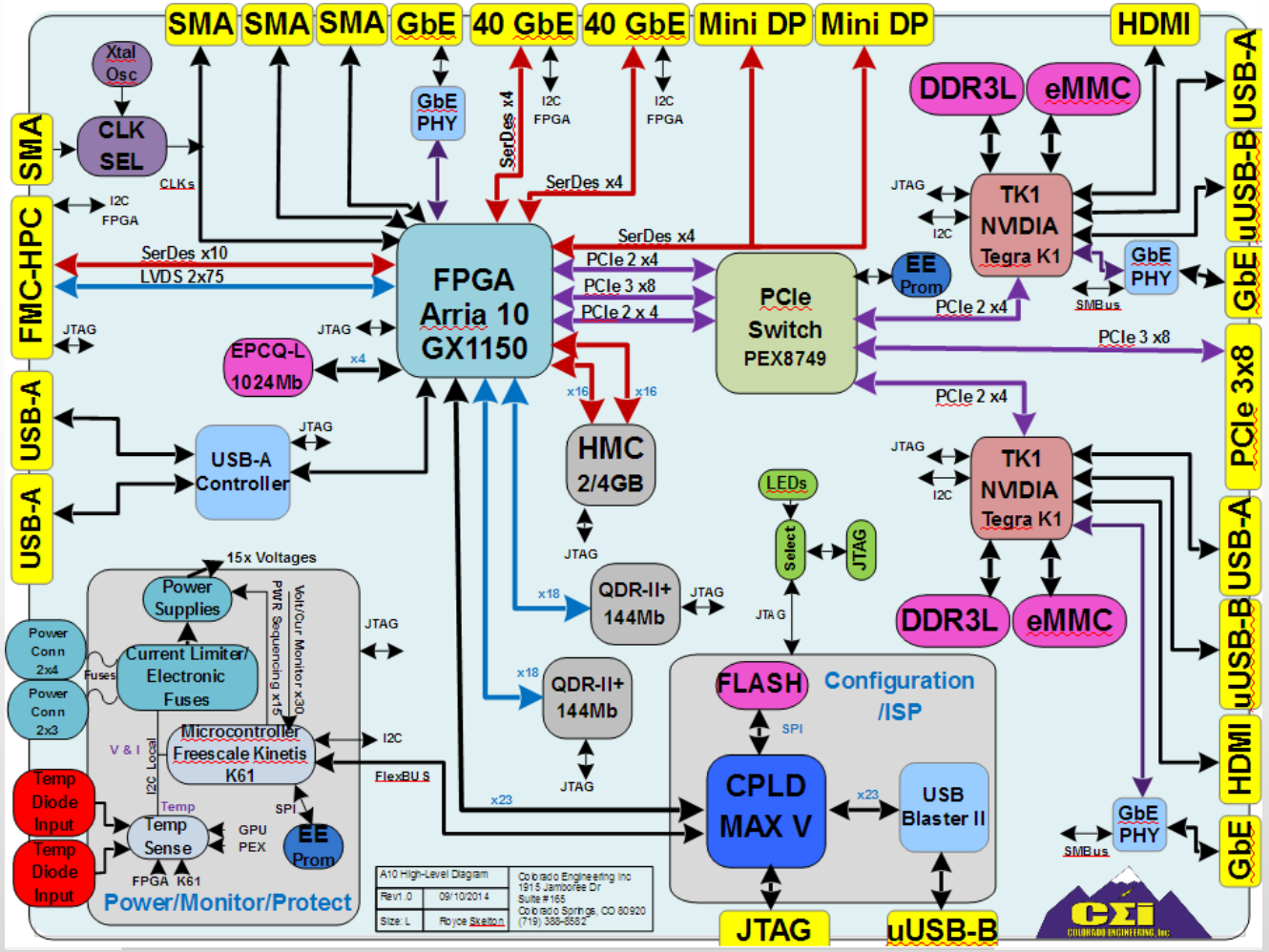
2 Inches

2 Inches

Available Stand-alone

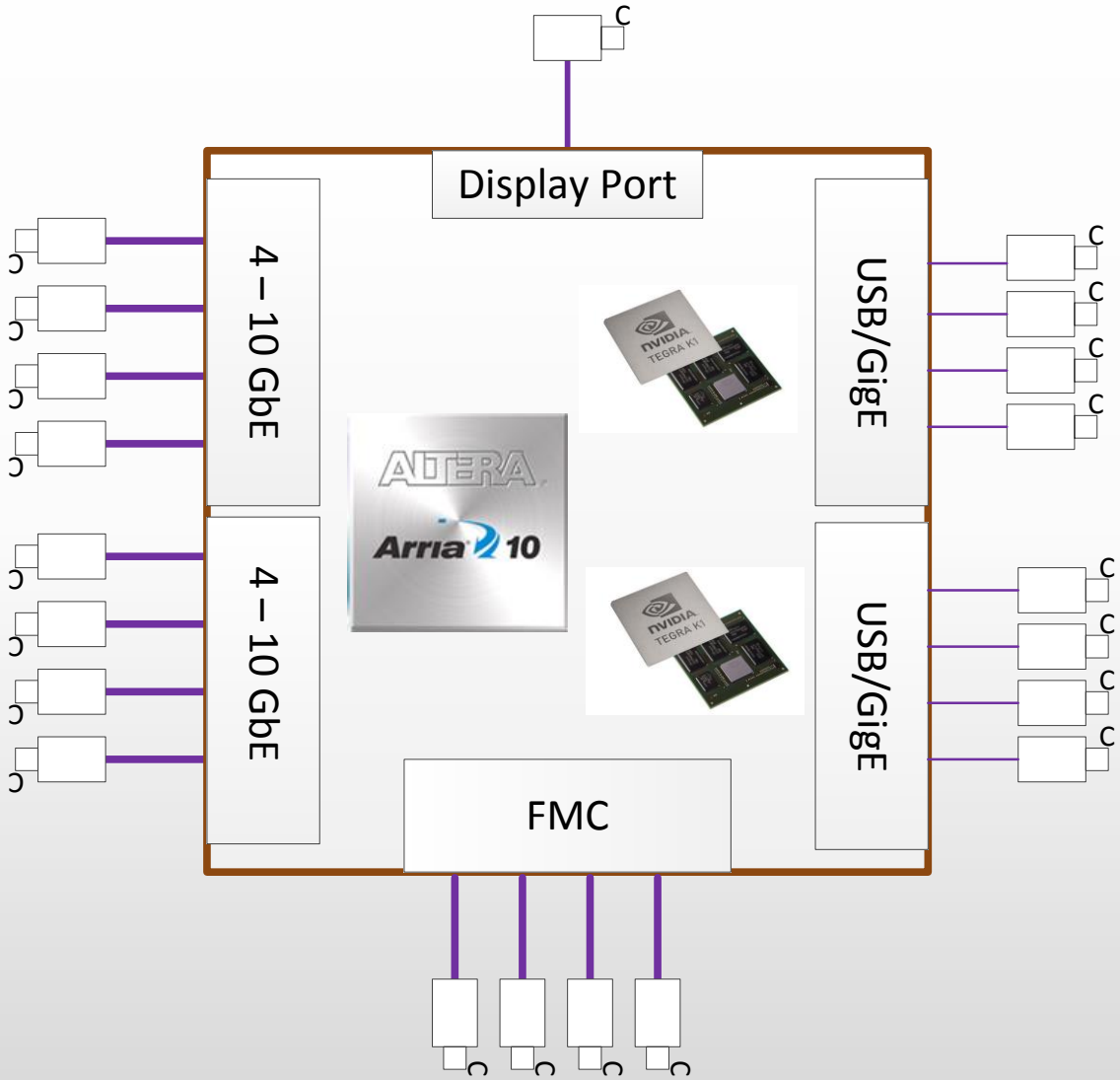
HPC-A10-K1GPU Design Details

- NVIDIA GPUDirect Support
- TK1's are root nodes
- TK1's can be field upgraded
- 8 - High Speed 10GbE Ports
- CUDA on TK1
- OpenCL on Arria 10
- 2 GB/s to each TK1
- HMC is 17X faster than DDR3
- 12 to 25 Camera/Sensor I/Os

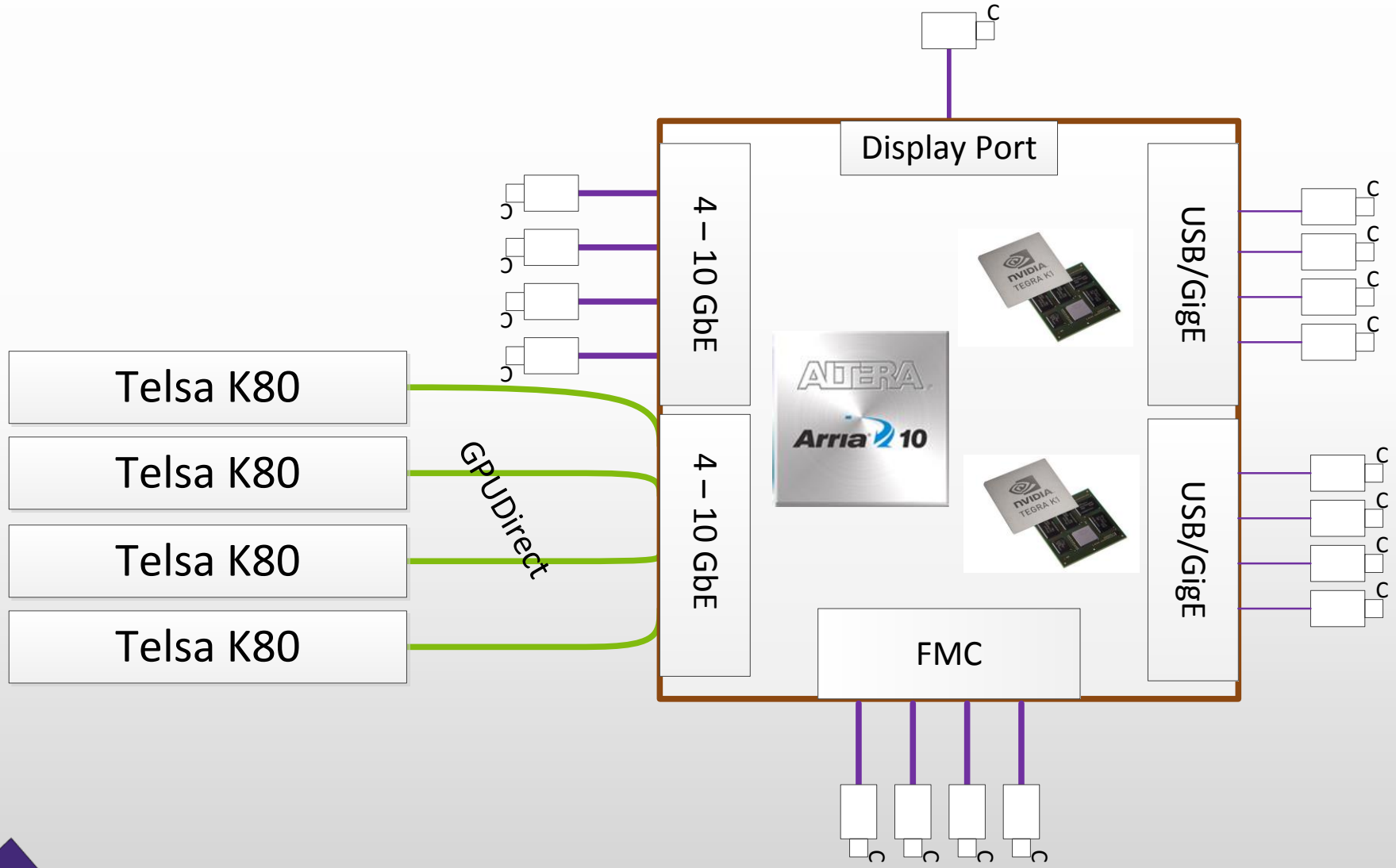


Single Node

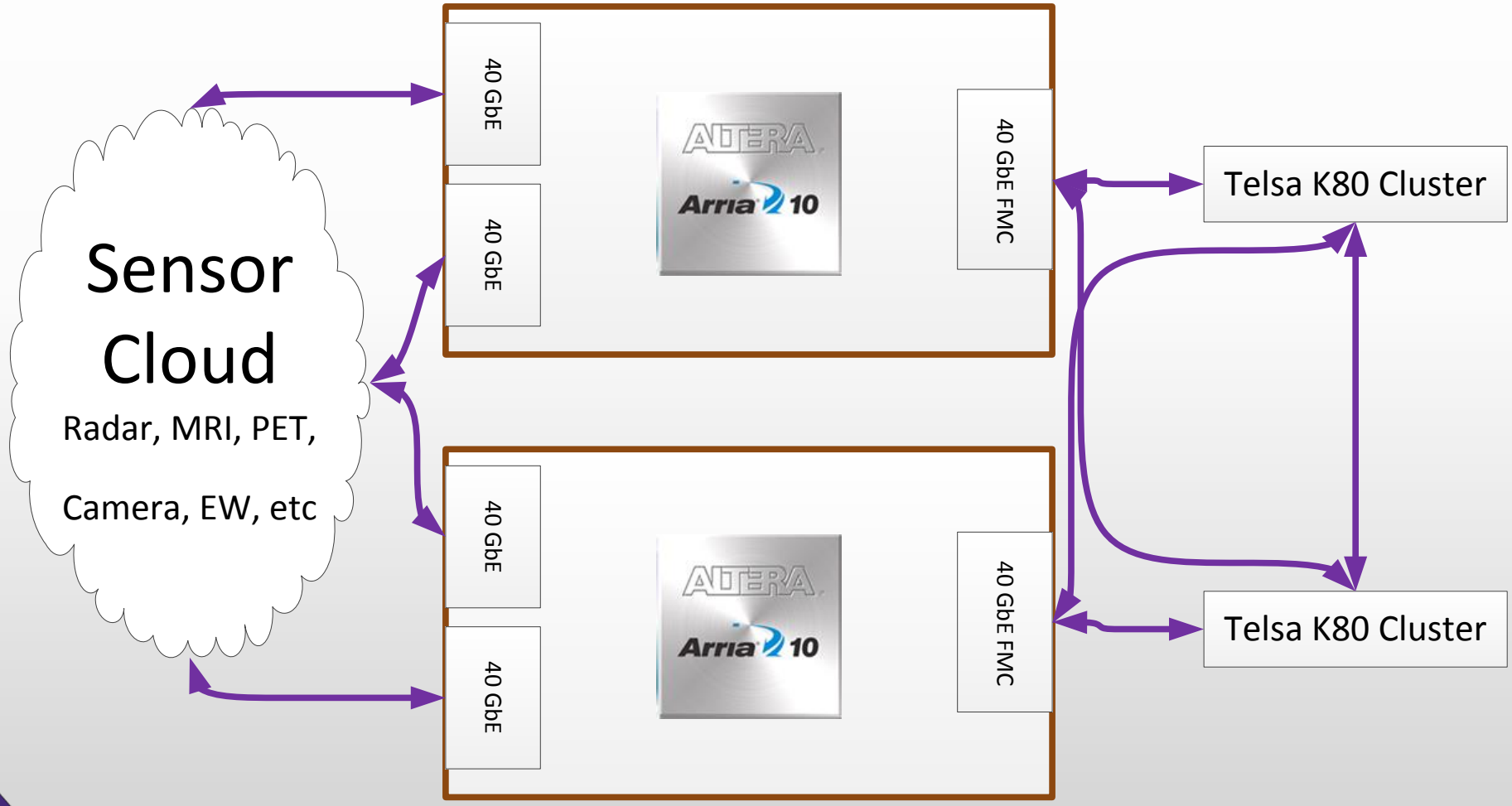
- 1 to 21 Cameras/Sensors
- Makes dumb cameras smart
- 10/40 GbE Sensors
- OpenCL on FPGA
- CUDA on Tegra



Tesla K80s + HPC-A10-K1GPU

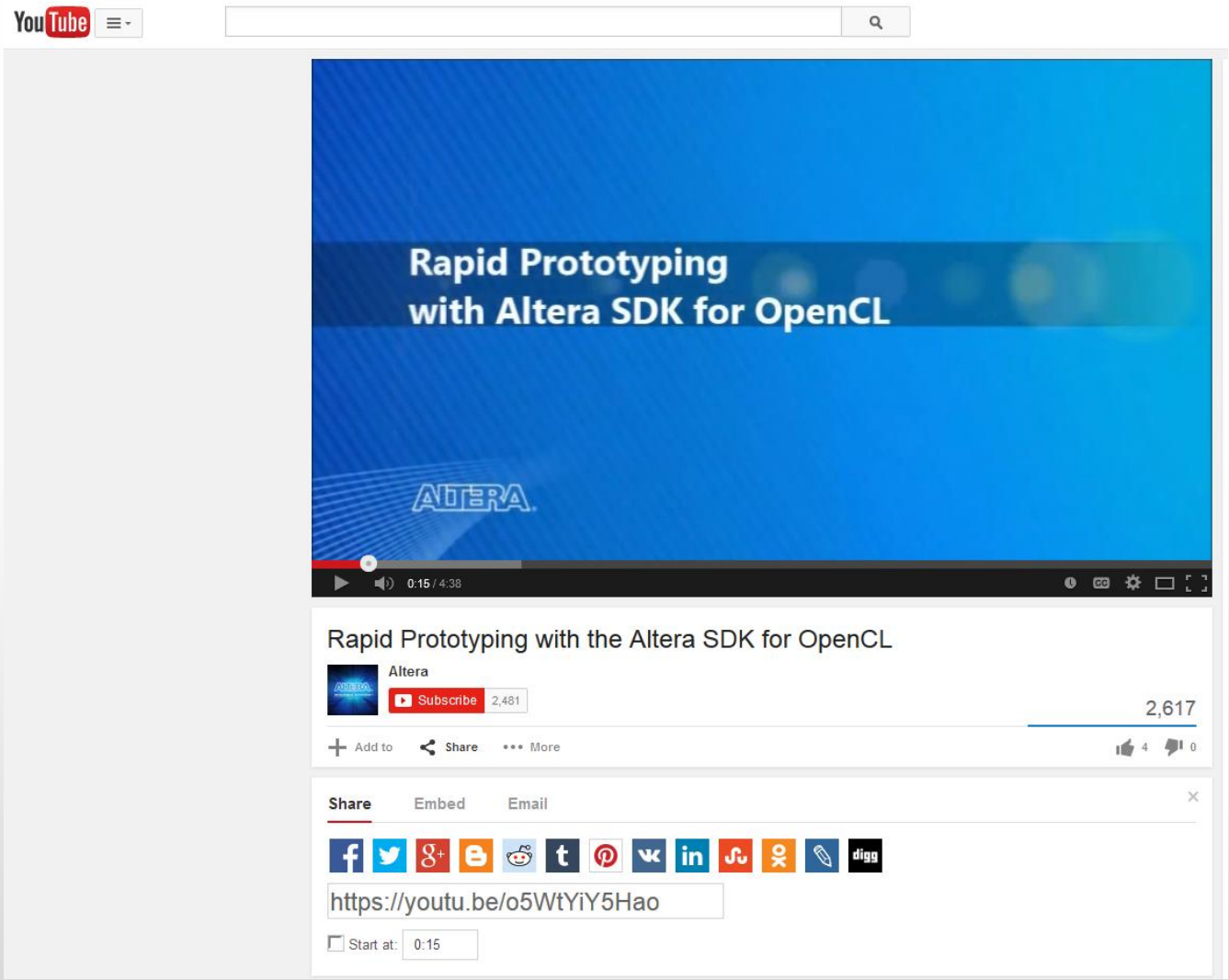


Sensor Gateway
Smart Host Bus Adapter (HBA)



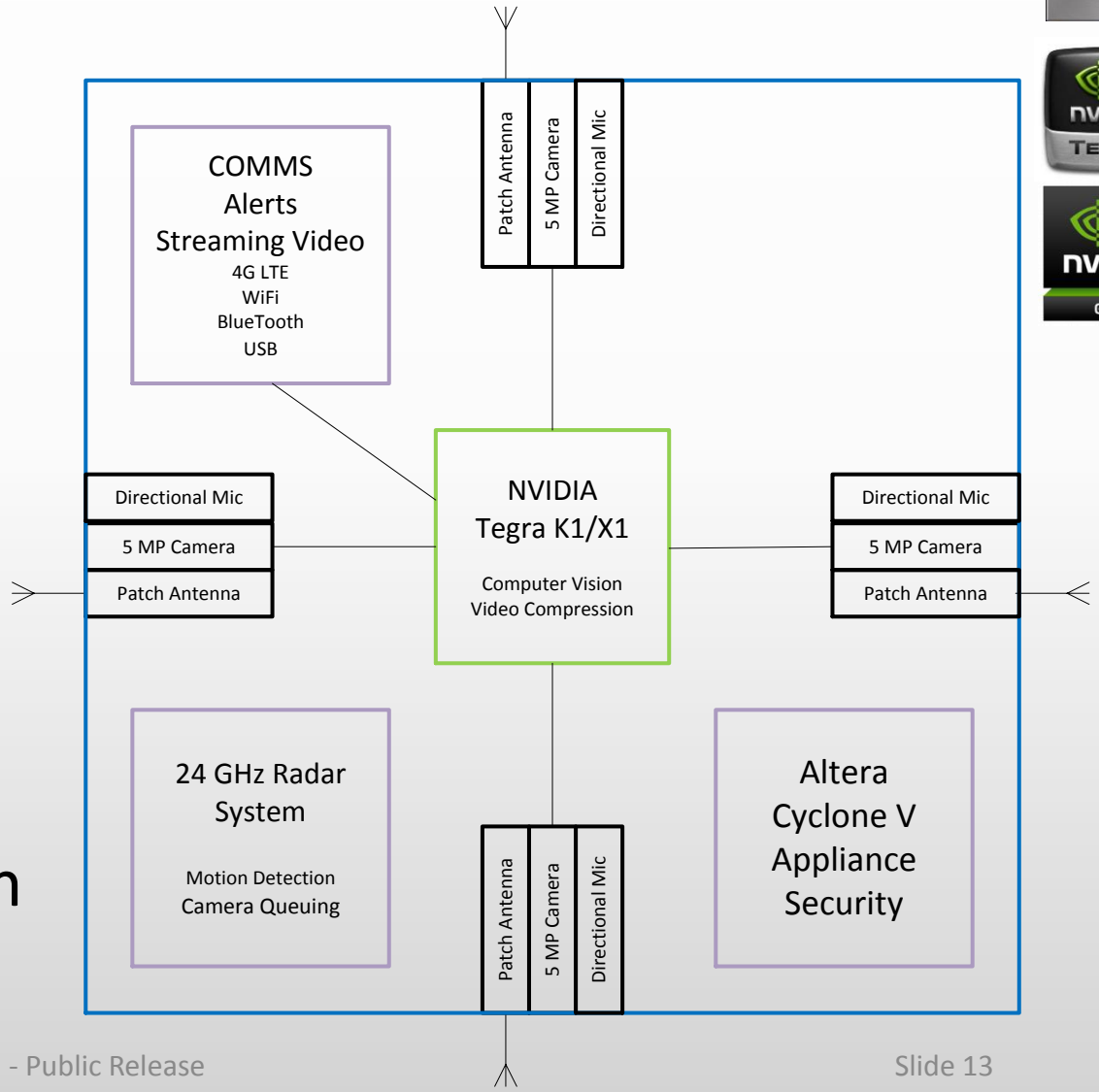
Programming FPGAs with OpenCL

- Easy to do now
- https://youtu.be/o5WtYiY5Hao
- Proficient in a day or two
- CAPI support too
- 95% to 99% Efficient as VHDL



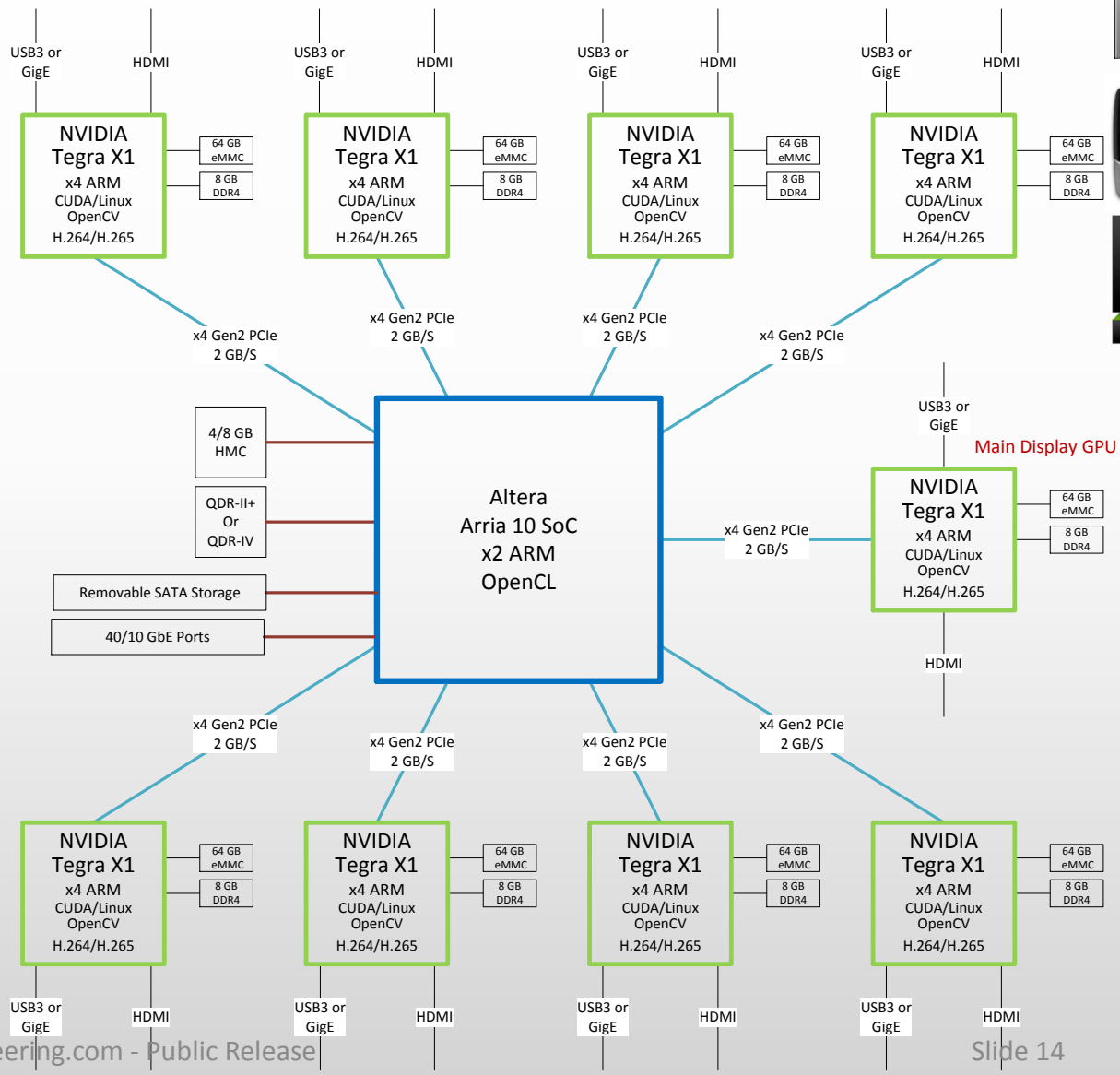
EDGE Node Processing

- Process on the EDGE using GRID
- Distributed deep learning node
- Low cost
- 4G enabled
- Fusion of Radar, EO, IO and Sound
- Download apps from Google Play
- Feedback to Tesla K80s via GRID
- SmartCity Ready
- Military Level Device Security Built-in



Distributed Aperture System
Distributed Sensors

- Large vehicle/Military ADAS
- SA360 systems
- Retrofit casino camera systems
- Make any sensor system smart
- Tegra K1/X1's Scalable
- Mixture of CUDA & OpenCL



Challenges
Hardware, Interconnects & Software

- FPGA + GPU
  - CUDA, OpenCL or CUDA + OpenCL
  - Working with MDA & AFRL on solutions
- Bandwidth
  - Tegra K1/X1 are x4 Gen2 PCIe – limits number and resolution of sensors attached to the Tegra.
  - More processing has to be done of Tegra, but that is okay since Tegra's keep increasing in power every year
  - Gen3 PCIe would be awesome
  - PCIe backplane – Using 40 GbE ports eliminates PCIe bottleneck
- Root Nodes
  - Tegra wants to root complex. Non-transparent switches need to be used
  - If Tegra could be an endpoint, a whole new world would open up

Future Architectures
Even Cooler Designs Possible

- Altera
  - Arria 10 SoC
    - Eliminates need for x86 CPU to run OpenCL
    - Truly stand-alone appliances
    - 100 GbE interfaces
  - Stratix 10 and Stratix 10 SoC
    - >10 TFLOPs for 100W
    - Details: https://www.altera.com/products/fpga/stratix-series/stratix-10/overview.html
- NVIDIA VOLTA
  - Looking for NVLink intermingling with FPGAs
- Virtual FPGAs + Virtual GPUs
  - Allow instant scaling and data protection

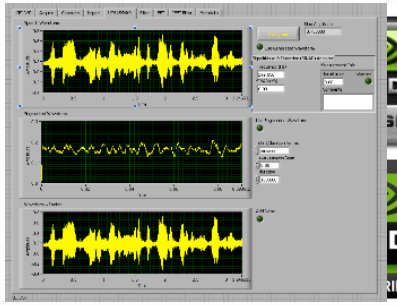
Summary

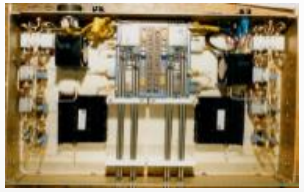
- GPU + FPGA can solve amazing and fun problems
- Tegra K1/X1 provide incredible capability at low cost which reduces the size of FPGA needed.
- OpenCL and Hard Floating Point IP make the Altera FPGAs a great partner with NVIDIA GPUs
- CEI is making scalable solutions to allow application developers to deploy from handheld to enterrpise/HPC

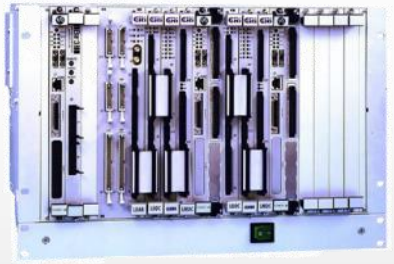
Hardware & Software Capabilities

- Enterprise & Embedded SW
  - Net Centric, SOA, web services, J2EE,SQL
  - C/C++
  - CUDA & OpenCL
  - Embedded real time code, RTOS, hardware drivers, Fault Detection / Fault Isolation, etc.
  - Simulations, APIs, and GUIs
  - Cognitive Software
  - Device Drivers
  - National Instruments Labview
  - DO-178C
- FPGA designs (VHDL/Verilog/Simulink)
- RF Design

- System / Subsystem Designs
- 30+ complex board designs
  - 32 layer PCBs with blind and buried vias
  - High speed (100s MHz → x GHz)
  - Analog (RF & I/Q Receivers)
  - Digital (FPGAs, DSPs, general purpose)
  - ADC and DAC
  - Standard and custom IO (busses, fabrics, SerDes, etc.)
  - Ruggedization and thermal management
  - CSWaP
  - Serial I/O (e.g. PCIe, Serdes)
  - DO-254







For More Information on Standard Products and Custom Engineering Services

Call Us – 719-388-8582 Office

Emails Us – lance.brown@coloradoengineering.com

Visit Us – Colorado Springs, CO (Sunny 300+ Days)

Browse Us – www.ColoradoEngineering.com