

# High Performance Computers And kits Used In Parallel Programming

Amir Shahshahani

ECSE 420

Prof. Zeljko Zilic



# Three Eras of Processor Performance



## Single-Core Era

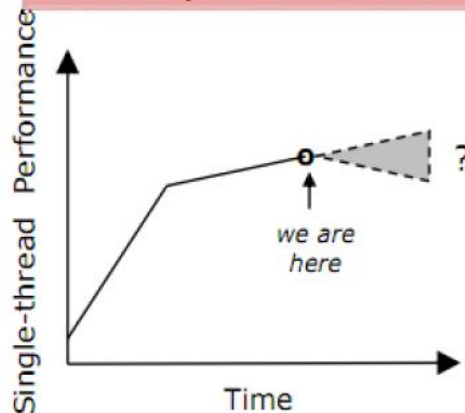
Enabled by

- Moore's Law
- Voltage Scaling
- Microarchitecture

Constrained by

- X Power
- X Complexity

Assembly > C > C++, Java



## Multi-Many Core Era

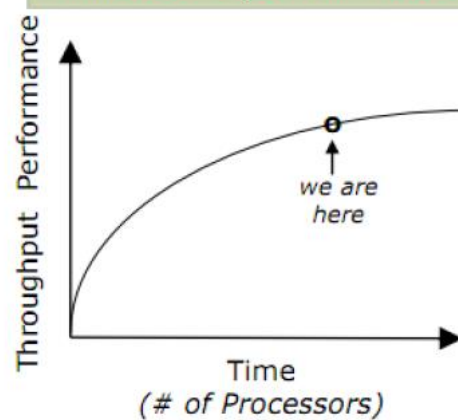
Enabled by

- Moore's Law
- Desire for throughput
- Desire for performance

Constrained by

- X Power
- X Parallel SW availability
- X Scalability

Pthreads > OpenMP/TBB



## Heterogeneous Systems Era

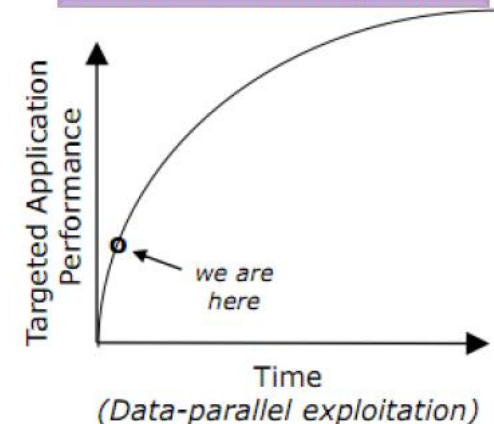
Enabled by

- Moore's Law
- Abundant data parallelism
- Power efficient GPUs

Constrained by

- X Power
- X Programming models
- X Communication overheads

Shader > Cuda > OpenCL



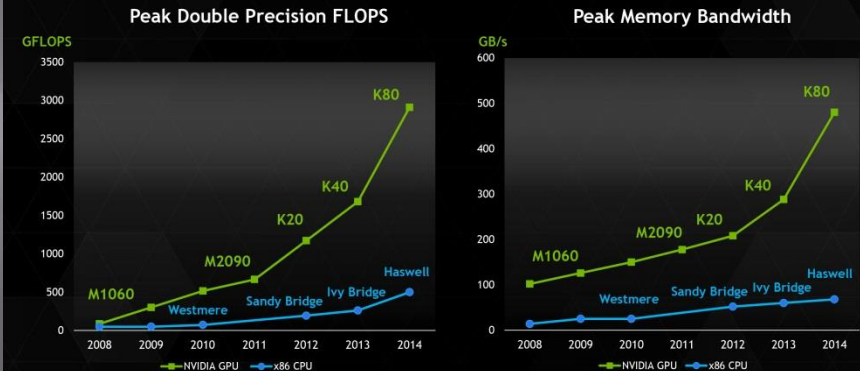
Source : The Salishan Conference on High Speed Computing, DATA PROCESSING IN EXASCALE-CLASS COMPUTER SYSTEMS  
Chuck Moore AMD Corporate Fellow & Technology Group CTO

# WHAT IS TESLA GPU ACCELERATOR?



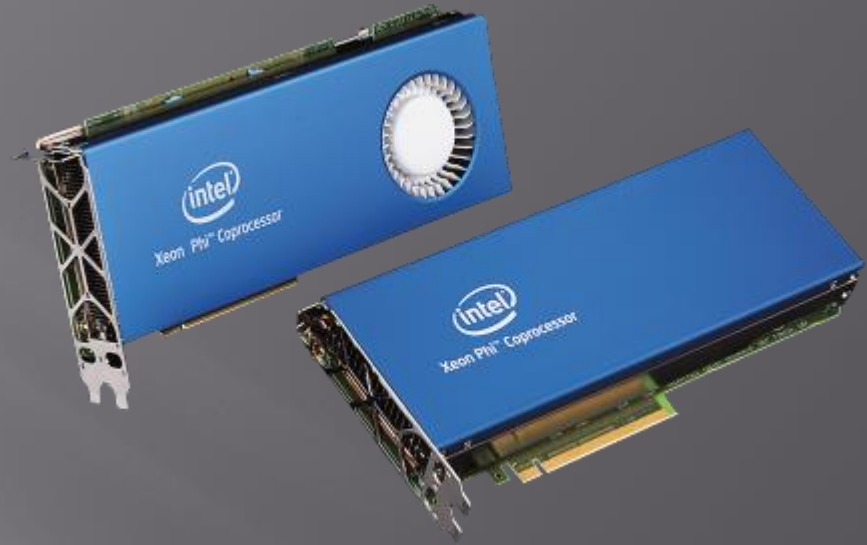
- World's fastest accelerators
- The ACCELERATED COMPUTING platform for datacenters
- Operate:
  - in simulations and in large scale calculations
  - for high-end image generation
  - with the use of [OpenCL](#) or [CUDA](#).

## Performance Lead Continues to Grow



Feature	Tesla K40	Tesla K20
Peak double precision floating point performance	1.43 Tflops	1.17 Tflops
Peak single precision floating point performance	4.29 Tflops	3.52 Tflops
Memory bandwidth (ECC off)	288 GB/sec	208 GB/sec
Memory size (GDDR5)	12 GB	5 GB
CUDA cores	2880	2496

# WHAT IS THE INTEL® XEON PHI™ COPROCESSOR?



- PCI Express add-in cards
- Enhance Parallel Application Performance
- Extends hardware capabilities and increases efficiency, all while optimizing power savings
- Shares parallel programming with general purpose processors

## 7120-A XEON PHI Coprocessor:

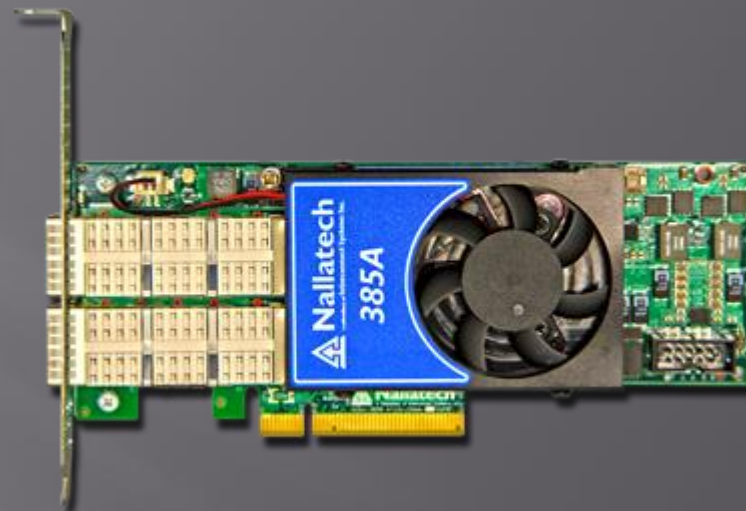
- ~ 4300 USD ( type 7120A)
- # of cores: 61
- Base frequency: 1.2 GHz
- 30MB L2 cache





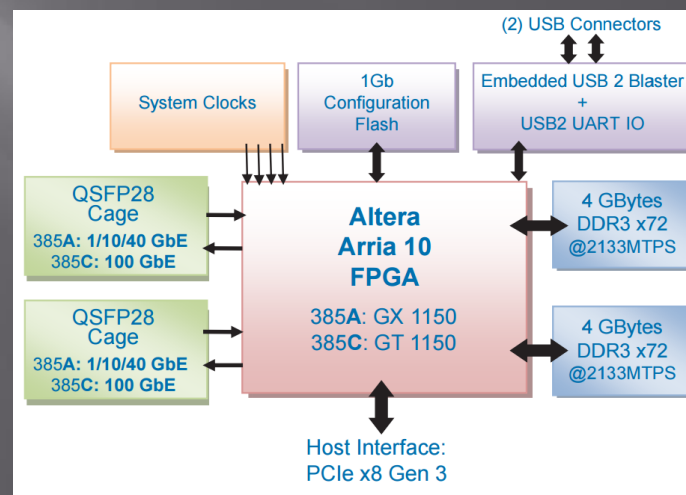
# WHAT IS NALLATECH FPGA BOARD?

- A powerful PCI-Express computing and I/O platform
- FPGA accelerated computing solutions
- Real-time embedded computing



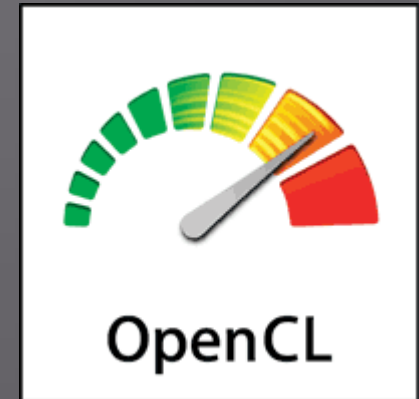
## NALLATECH 385A :

- 2 Banks 4G DDR3 SDRAM @ 2133 MTPS
- PCIe x8 Gen 3
- Arria 10 1150 GX FPGA with up to 1.5 Tflops
- OpenCL tool flow

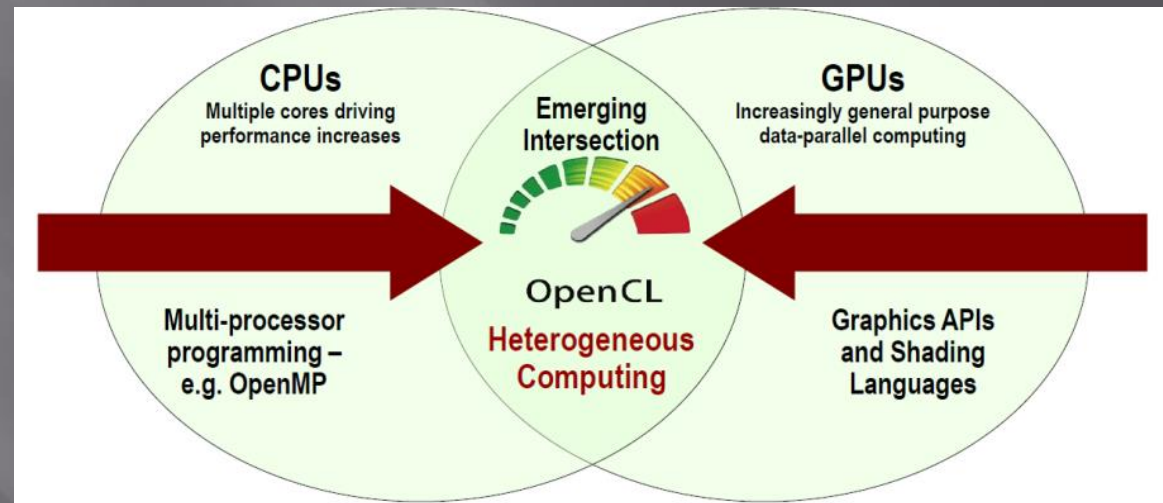


# OpenCL

## Open Computing Language



- Is a framework for writing programs that execute across: CPUs, GPUs, DSPs, FPGAs, etc.
- provides parallel computing using task-based and data-based parallelism
- Consists of a library that implements the API for C and C++,
- An OpenCL C compiler for the compute device(s) targeted.

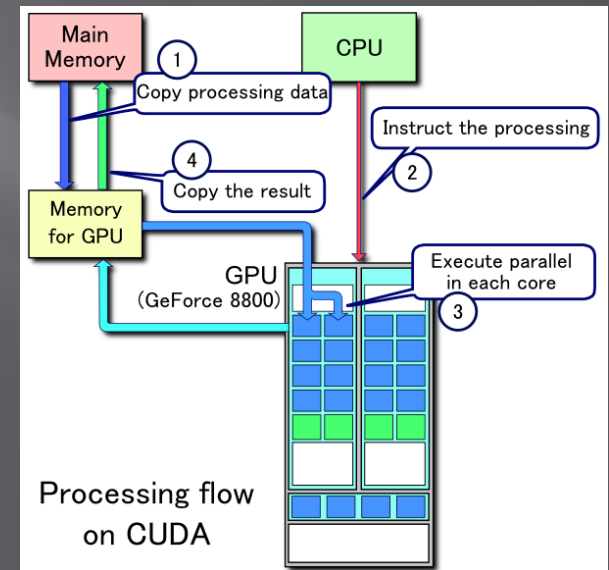


# CUDA

Compute Unified Device Architecture



- A parallel computing platform and API.
- Developers can use a CUDA-enabled (GPU) for general purpose processing.
- CUDA is designed to work with languages such as C, C++ and Fortran.
- Easier in parallel programming to utilize GPU resources, as opposed to previous API solutions like *Direct3D* and *OpenGL*.
- C/C++ programmers use 'CUDA C/C++', compiled with "NVCC".



## Example of CUDA processing flow

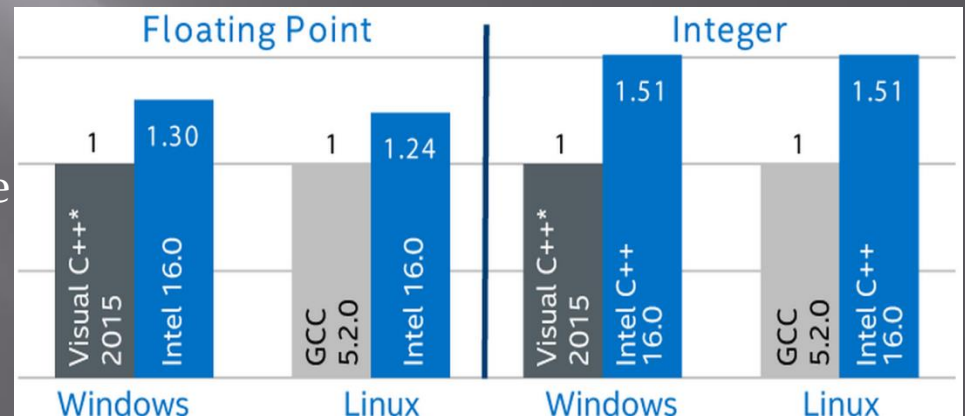
1. Copy data from main mem to GPU mem
2. CPU instructs the process to GPU
3. GPU execute parallel in each core
4. Copy the result from GPU mem to main mem



# Intel Parallel Studio



- Deliver top C++ and Fortran application performance with less effort
- Parallel Studio is composed of
  - Intel C++ Compiler with *Cilk Plus* and *OpenMP*
  - Intel Fortran Compiler with *OpenMP*
  - IDE plug-in integration with *Visual Studio* and *Eclipse*
  - Etc
- Boost your applications performance with Intel® C++ Compiler and Intel® Fortran Compiler



# Nallatech

## ALTERA PREFERRED BOARD for OPENCL



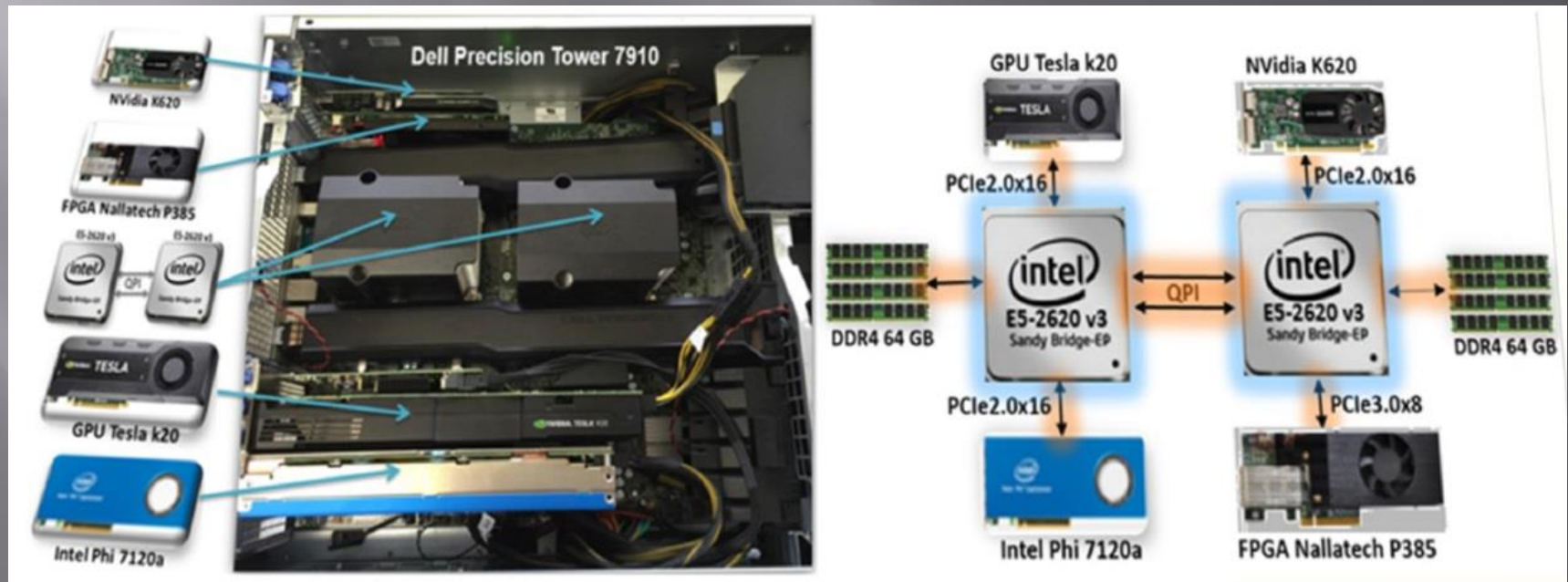
- Nallatech FPGA Accelerators support the Altera SDK for OpenCL
- Nallatech's partnership with Altera now enables programming with no knowledge on FPGA
- Altera Software Development Kit (SDK) for OpenCL provides a programming environment for programmers to develop their own massively parallel and compute intensive applications while reducing power consumption and total cost of ownership.



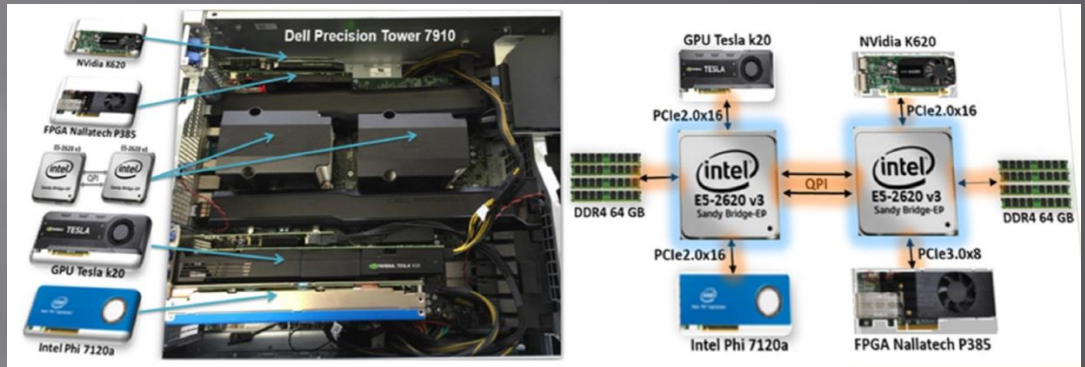
# 1-DELL TOWER 7910

## HPP-Heterogeneous Processing Platform

A Combination Of Multi-core,  
GPUs, FPGAs And Many-core  
Accelerators



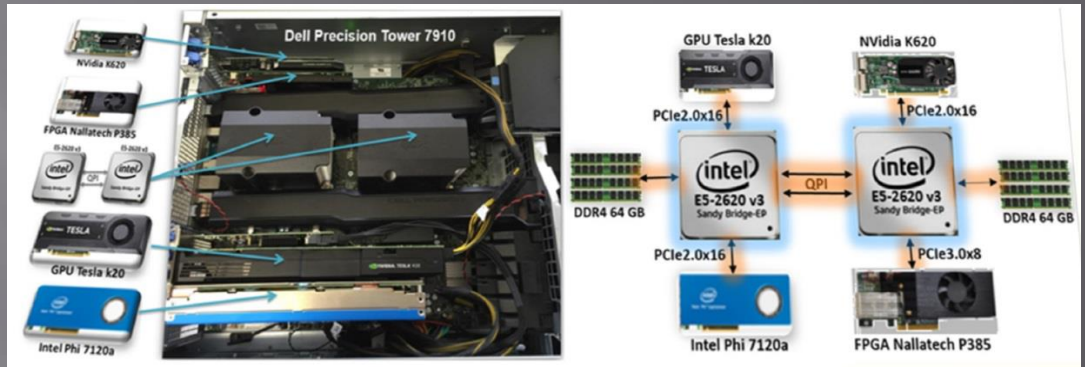
# Dell Tower 7910 Hardware Architecture



## The main components:

- Dual core Intel Xeon E5-2620 V3 ( 6 Cores each, 85 Watt)  
(CPU cache: 15MB / Processor)
- NVidia GPU - Tesla K20 (2496 CUDA cores , 225W)  
(5GB, 208 GB/s)
- Xeon Phi 7120A (61 Cores, 1.33 GHz, 300W)  
(Mem. 16 GB at 352 Gb/sec)
- FPGA board - Nallatech P385-A72 (Altera Stratix V)

# Dell Tower 7910 Hardware Architecture



## Other specifications:

- System Memory: 128GB ECC, 2133MHz DDR4 RDIMM
- Network controller: Intel i217 Gigabit Ethernet
- Video : Nvidia k620 2 GB
- Storage: 500GB SSD + 1TB SATA HDD



# Dell Tower 7910

## Software Architecture

### Pre-Installed Software components:

- RedHat Enterprise Linux 6.6 (Kernel version: 2.6.32-504.el6.x86\_64)
- Java Runtime Environment (JRE)
- gcc compiler and toolchain

Accelerator	Software and tools
GPU Tesla K20	NVIDIA CUDA 7 Toolkit
Nallatech FPGA P385	<ul style="list-style-type: none"><li>• Altera Quartus 15.0, Altera SDK for OpenCL 1.0</li><li>• Nallatech FPGA P385 Board Support Package</li></ul>
Xeon Phi 7120a	<ul style="list-style-type: none"><li>• Intel Manycore Platform Software Stack (MPSS)</li><li>• Intel Parallel Studio XE 2015, for C++<ul style="list-style-type: none"><li>o Intel C++/C/FORTRAN compilers, Intel Math Kernel Library</li><li>o Debuggers, Performance and Correctness Analysis Tools</li><li>o OpenMP, MPI, OFED messaging infrastructure (Linux only), OpenCL</li><li>o Programming Models: Offload, Native and mixed Offload+Native</li></ul></li></ul>

## 2-LENOVO THINKSTATION P900 MPAG-II , A Multiprocessor Array Platform



- Highest storage capacity in the industry
- Perfect for rendering and simulation
- High performance dual socket workstation



# Thinkstation P900

## Hardware Architecture

### The main components:

- Two E5 2603 v3 processors( 6 Cores , 85 Watt)  
(CPU cache: 15MB / Processor)
- NVidia GPU - Tesla K20 (2496 CUDA cores , 225W)  
(5GB, 208 GB/s)
- Xeon Phi 7120A (61 Cores, 1.33 GHz, 300W)  
(Mem. 16 GB at 352 Gb/sec)

# Thinkstation P900

## Hardware Architecture

### Other specifications:

- System Memory: 64 GB ECC 1600 MHz LP
- Network controller: Gigabit Ethernet dual port
- Video : Nvidia Quadro K420 Graphics Card
- Storage: 500GB SSD + 2 \* 1TB SATA HDD

# Thinkstation P900

## Software Architecture

### Pre-Installed Software components:

- RedHat Enterprise Linux 6.6 (Kernel version: 2.6.32-504.el6.x86\_64)
- Java Runtime Environment (JRE)
- gcc compiler and toolchain

Accelerator	Software and tools
GPU Tesla K20	NVIDIA CUDA 7 Toolkit
Xeon Phi 7120a	<ul style="list-style-type: none"><li>• Intel Manycore Platform Software Stack (MPSS)</li><li>• Intel Parallel Studio XE 2015, for C++<ul style="list-style-type: none"><li>o Intel C++/C/FORTRAN compilers, Intel Math Kernel Library</li><li>o Debuggers, Performance and Correctness Analysis Tools</li><li>o OpenMP, MPI, OFED messaging infrastructure (Linux only), OpenCL</li><li>o Programming Models: Offload, Native and mixed Offload+Native</li></ul></li></ul>

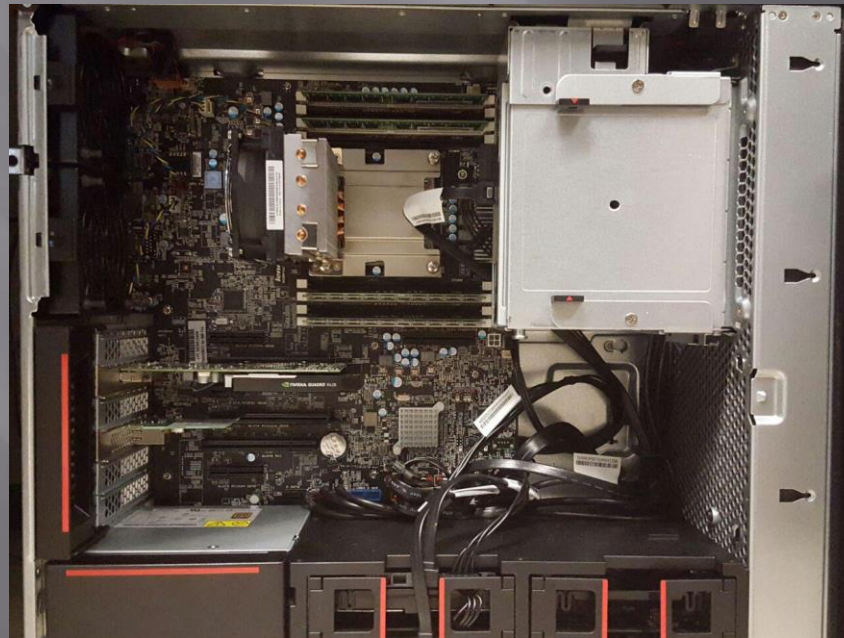


# 3-LENOVO THINKSTATION P500

## Mainstream Performance



- Incredible storage capacity
- Ample I/O slots
- Perfect for professional 3D design
- High performance single socket workstation



# Thinkstation P500

## Hardware Architecture

### The main components:

- E5 2603 v3 processors( 6 Cores , 85 Watt)  
(CPU cache: 15MB / Processor)
- NVidia GPU - Tesla K20 (2496 CUDA cores , 225W)  
(5GB, 208 GB/s)
- System Memory: 32 GB ECC 1866 MHz LP
- Network controller: Gigabit Ethernet dual port
- Video : Nvidia Quadro K420 Graphics Card
- Storage: 2 \* 1TB SATA HDD

# Thinkstation P500

## Software Architecture

### Pre-Installed Software components:

- Linux Ubuntu 10 ( 64 Bits) and Windows
- Java Runtime Environment (JRE)
- gcc compiler and toolchain

Accelerator	Software and tools
GPU Tesla K20	NVIDIA CUDA 7 Toolkit

## 4-LENOVO THINKSTATION S30



- Single-processor workstation
- Full advanced technology extended (ATX) motherboard



Processor	Intel Xeon E5-2620 ( 6 Cores, 12 threads)
Memory	16 GB DDR3
Graphic	NVIDIA Quadro K600
Storage	2 * 1-TB SATA HDD
Accelerator	GPU Tesla K20

Windows-7, Linux, All CUDA Toolkits, Microsoft Visual Studio



## 5-ANTEC CASE



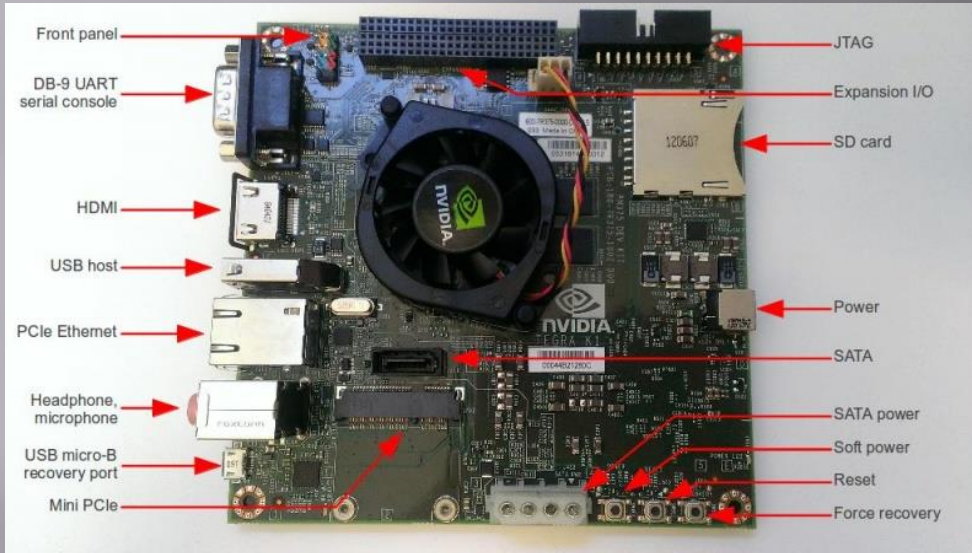
Processor	Intel Xeon E5-2630 V2( 6 Cores, 12 threads)
Memory	32 GB DDR3
Graphic	Gforce GT 630
Storage	2 * 1-TB SATA HDD + 500GB SSD
Accelerator	GPU <b>Tesla K40</b>

Windows-7, Linux, All CUDA Toolkits, Microsoft Visual Studio 2013



# 6-NVIDIA JETSON TK1

## An End-to-end Platform



- Only 192 USD
- Full-featured platform
- All-in-one package
- Ubuntu 14.04 host set up support
- Support CUDA 6.5 Toolkit
- Tegra Graphics Debugger 2.0
- Libraries for CUDA and OpenCV 2.4.10.1

### - Tegra K1 SOC

- Kepler GPU with 192 CUDA cores
- 4-Plus-1 quad-core ARM Cortex A15 CPU

-2 GB x16 memory with 64 bit width

-16 GB 4.51 eMMC memory

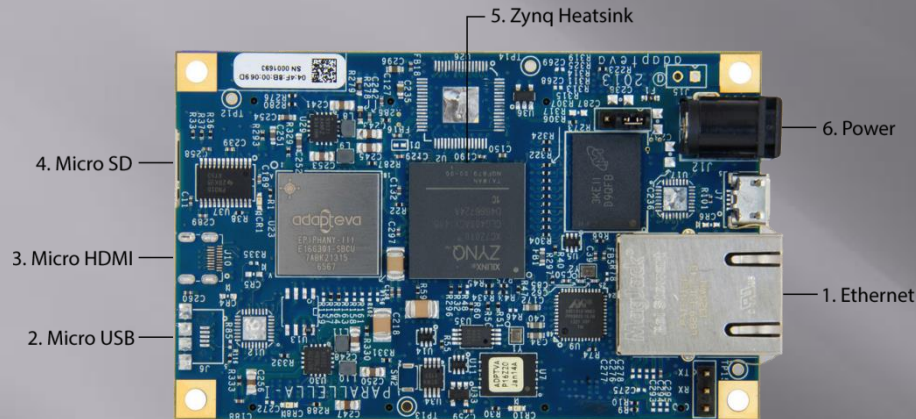
-1 Half mini-PCIE slot

-1 Full size SD/MMC connector

-1 Full-size HDMI port

# 7-PARALLELLA BOARD

## High Performance, Credit Card Sized Computer

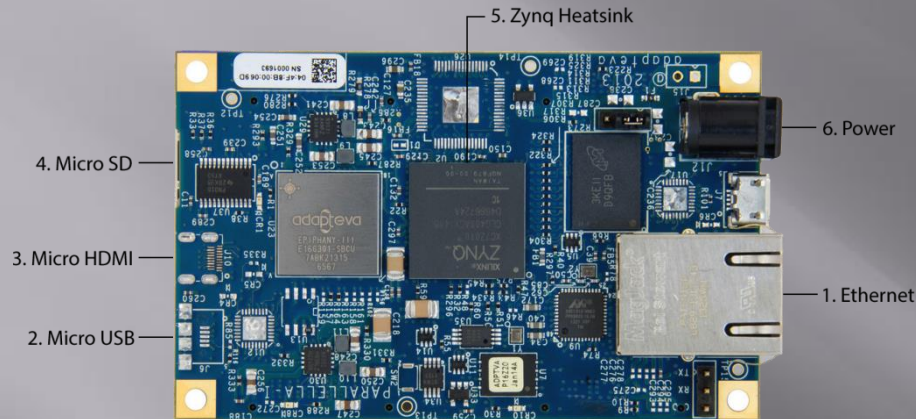


- Starting at 99 USD
- Useful in the field of parallel computing
- based on the Epiphany multi-core chips
- Can be used as a standalone computer, an embedded device
- Runs several of the popular Linux distributions, including Ubuntu

- Zynq-Z7010 or Z7020 Dual-core ARM A9 CPU
- 16-core Epiphany Coprocessor
- Up to 48 GPIO signal
- 1GB RAM
- MicroSD Card
- USB 2.0
- Linux Operating System

# 7-PARALLELLA BOARD

High Performance, Credit Card  
Sized Computer



## Applications:

- Smart-phones and tablet app acceleration
- High end audio
- Big Data Analytics
- Hyperspectral Imaging
- Medical applications
- Machine Vision
- High Speed Data Acquisition/Generation

[www.parallella.org](http://www.parallella.org)

# AMD FireStream

- Radeon-based product targeting stream processing and/or GPGPU in supercomputers

Released: 2008

Production statue: End-of-life !



## AMD FireStream™ 9270 stream processor: 2x Double Precision Performance\*

### AMD FireStream™ 9270



**\$1499**

- HPC-class performance deployable anywhere
- 240 double precision GFLOPS
- Optimal performance-per-watt

**Compute Power** 1.2 teraFLOPS

**Core Clock Speed** 750 MHz

**Single/Dual Slot** Dual Slot

**Memory Type** GDDR5 3.6 Gbps

**Memory** 2GB

**Typical Board Power** 160W

\* For more details please see benchmark specifications slide at the end of the deck



Students can take  
it home and  
install on their PC  
;)   
But return it !