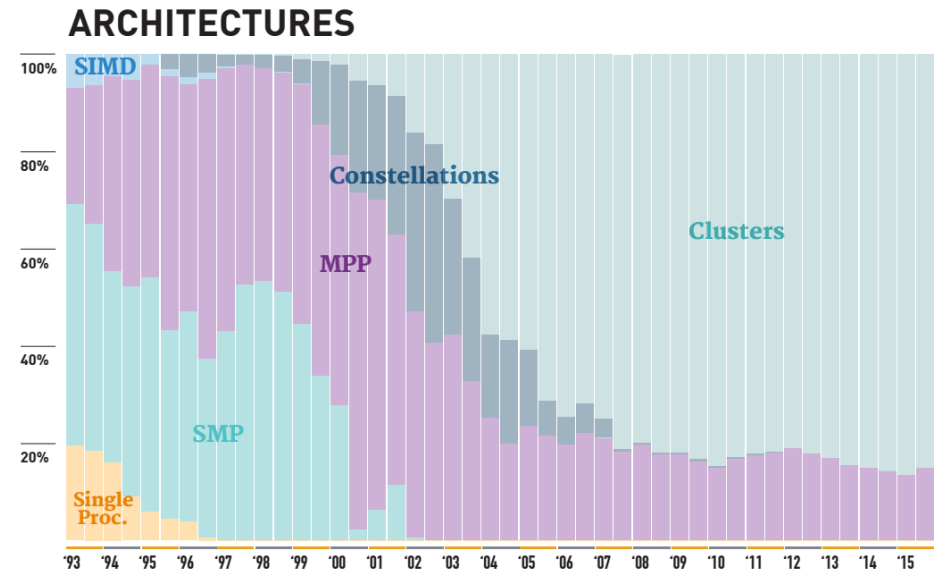
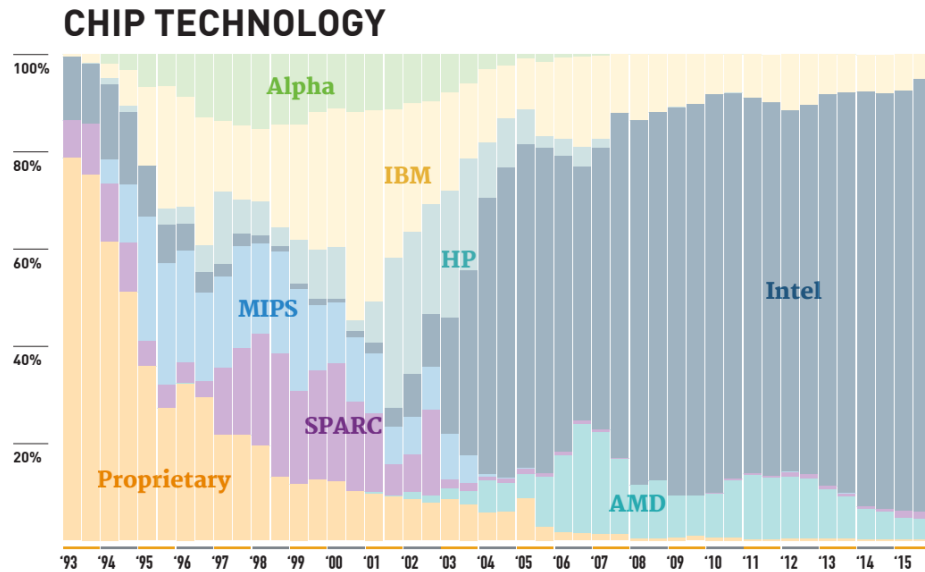


ACCELERATED COMPUTING: THE PATH FORWARD

Jen-Hsun Huang, Co-Founder and CEO, NVIDIA
SC15 | Nov. 16, 2015



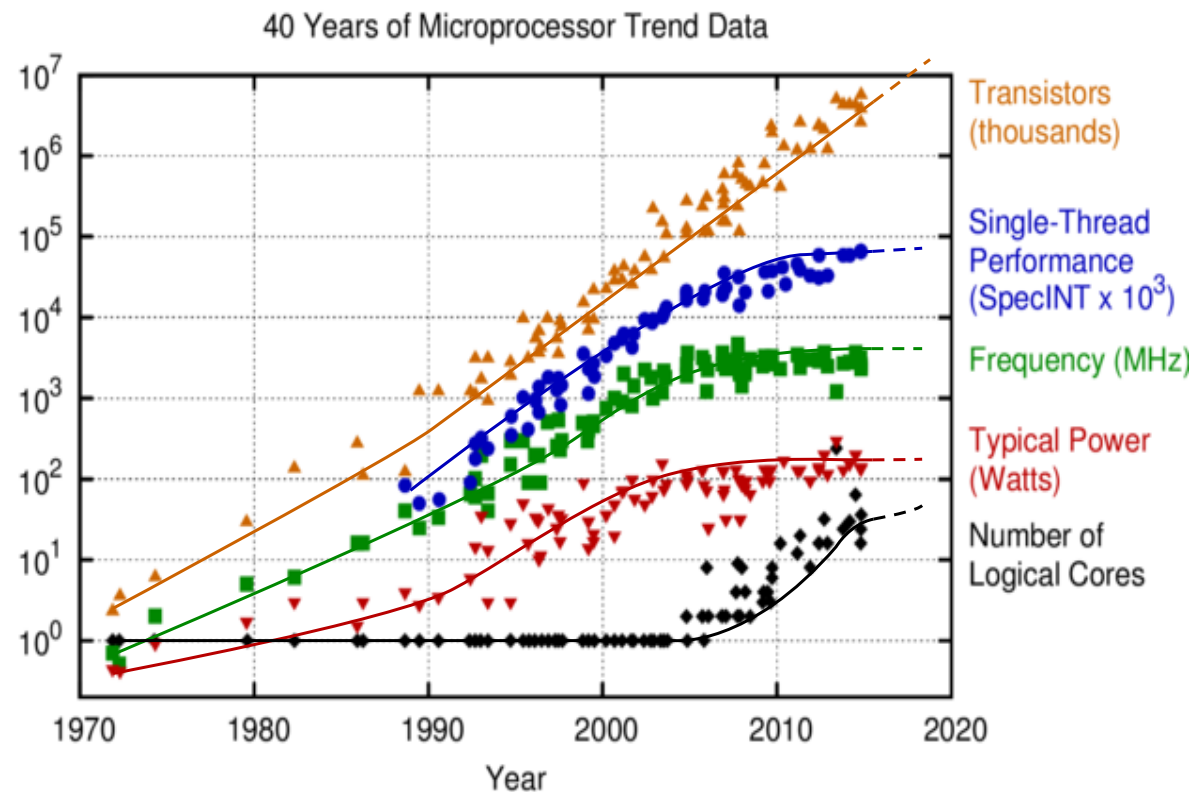
COMMODITY DISRUPTS CUSTOM



ACCELERATED COMPUTING: THE PATH FORWARD

“It’s time to start planning for the end of Moore’s Law, and it’s worth pondering how it will end, not just when.”

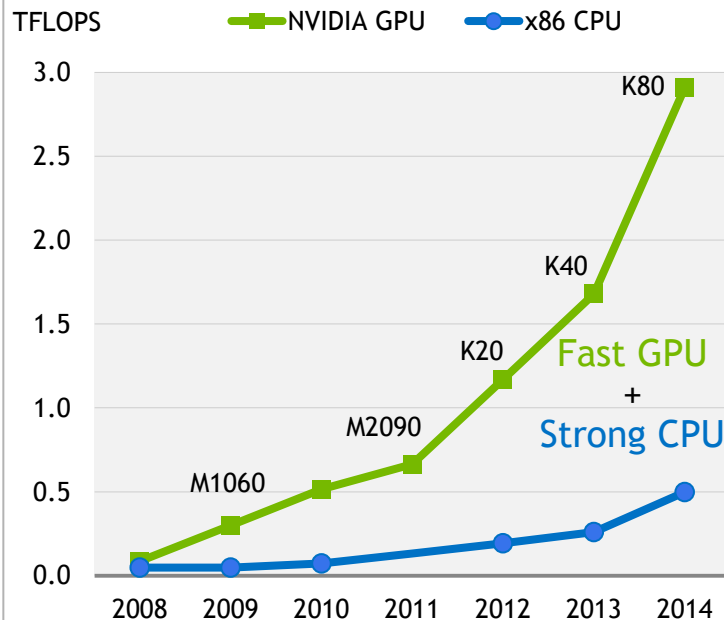
Robert Colwell
Director, Microsystems Technology Office, DARPA



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2015 by K. Rupp

NVIDIA ACCELERATES COMPUTING

Fast GPU Engineered for High Throughput



Productive Programming Model & Tools



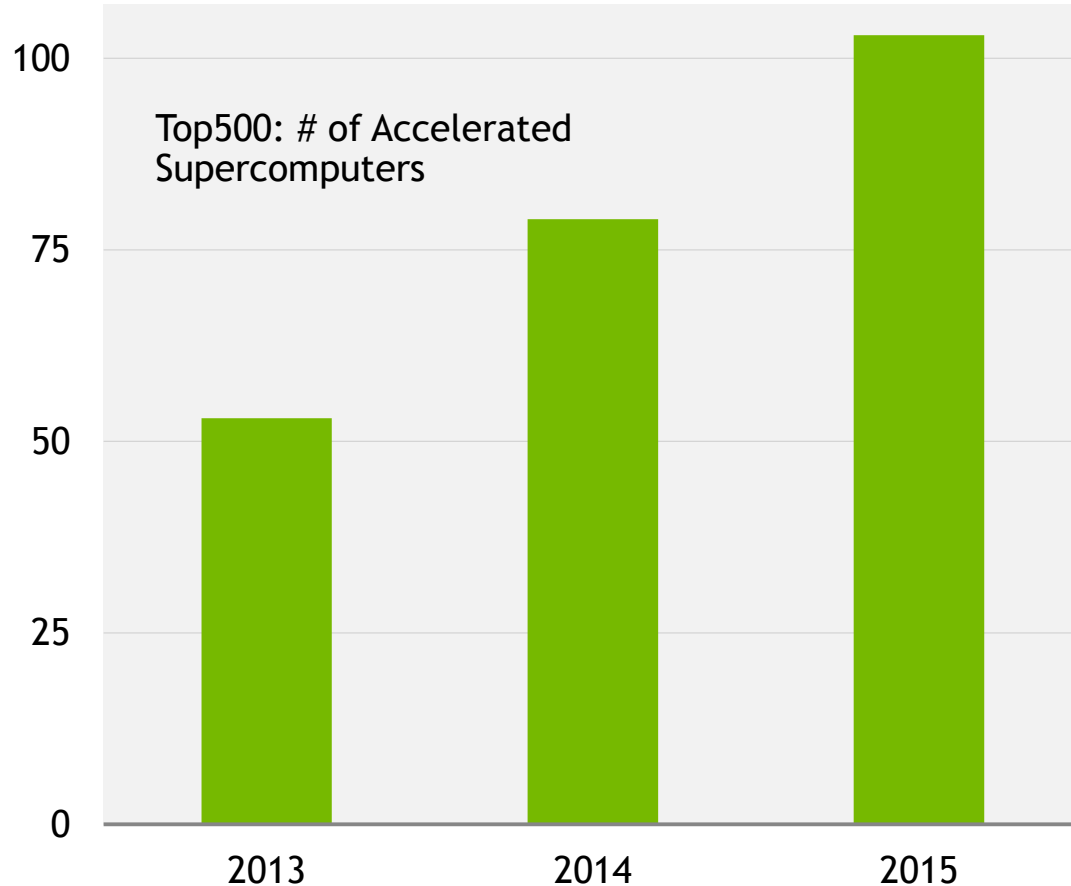
Expert Co-Design



Accessibility



ACCELERATORS SURGE IN WORLD'S TOP SUPERCOMPUTERS



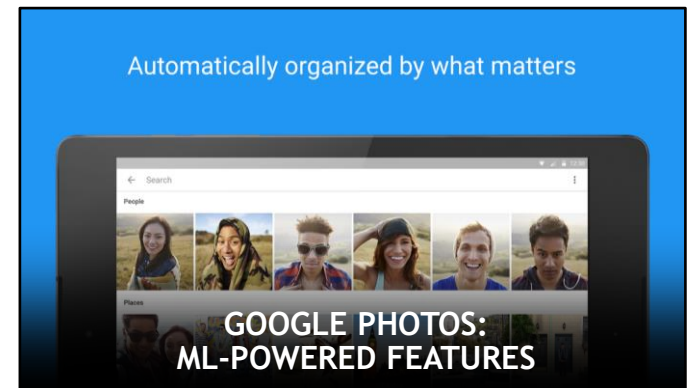
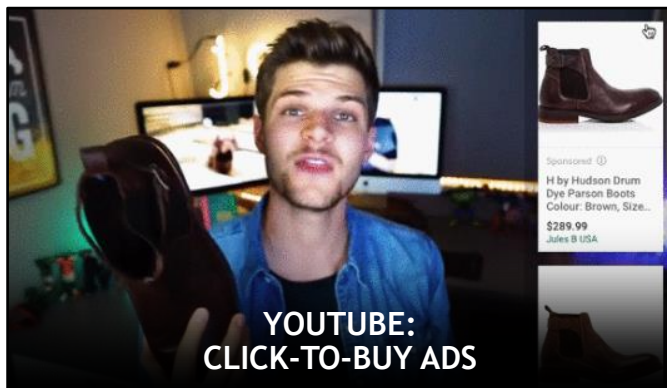
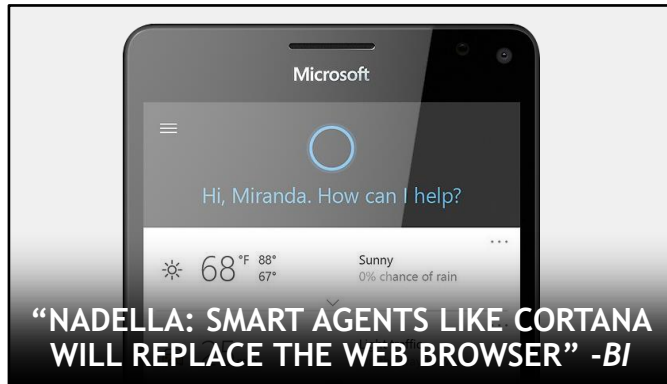
100+ accelerated systems now on Top500 list

1/3 of total FLOPS powered by accelerators

NVIDIA Tesla GPUs sweep 23 of 24 new accelerated supercomputers

Tesla supercomputers growing at 50% CAGR over past five years

MACHINE LEARNING HPC'S 1ST CONSUMER KILLER-APP



TESLA FOR MACHINE LEARNING

HYPERSCALE SUITE



GPU REST Engine



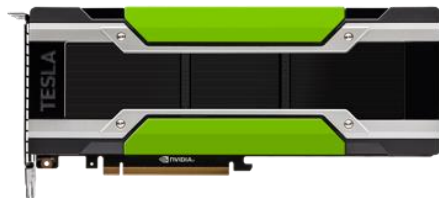
GPU Accelerated
FFmpeg



Image Compute
Engine

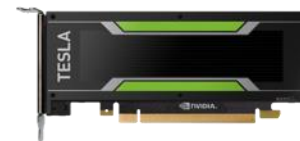
TESLA M40

POWERFUL: Fastest Deep Learning Performance



TESLA M4

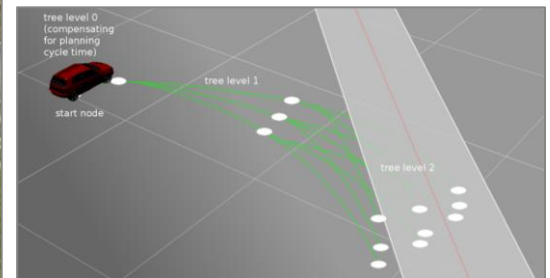
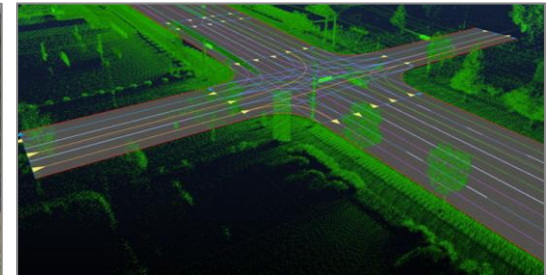
LOW POWER: Highest Hyperscale Throughput



MACHINE LEARNING REVOLUTIONIZING TRANSPORTATION

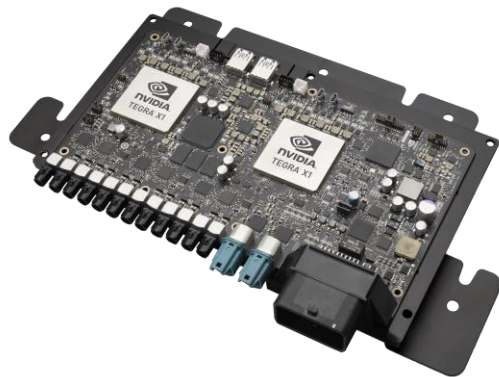
“Toyota Invests \$1 Billion in Artificial Intelligence in U.S.”

— U.S. News & World Report

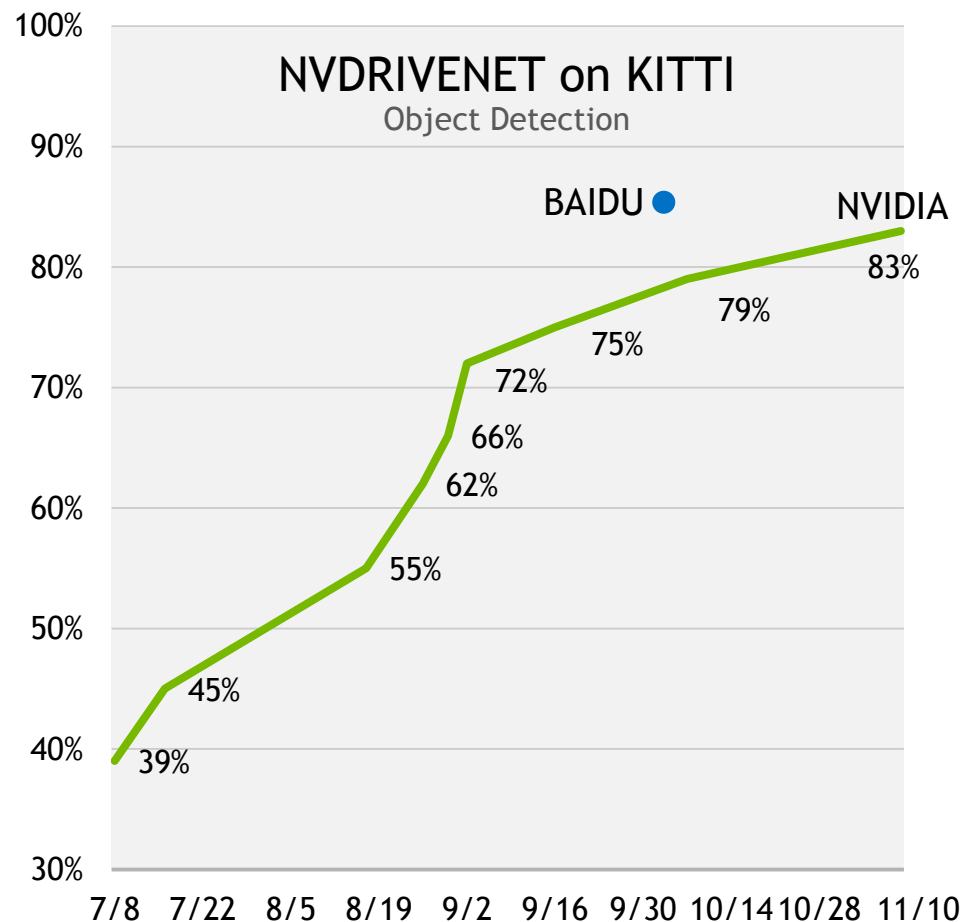


END-TO-END MACHINE LEARNING PLATFORM FOR AUTONOMOUS CARS

DRIVE PX



DIGITS DevBox



MACHINE LEARNING REVOLUTIONIZING AUTONOMOUS MACHINES



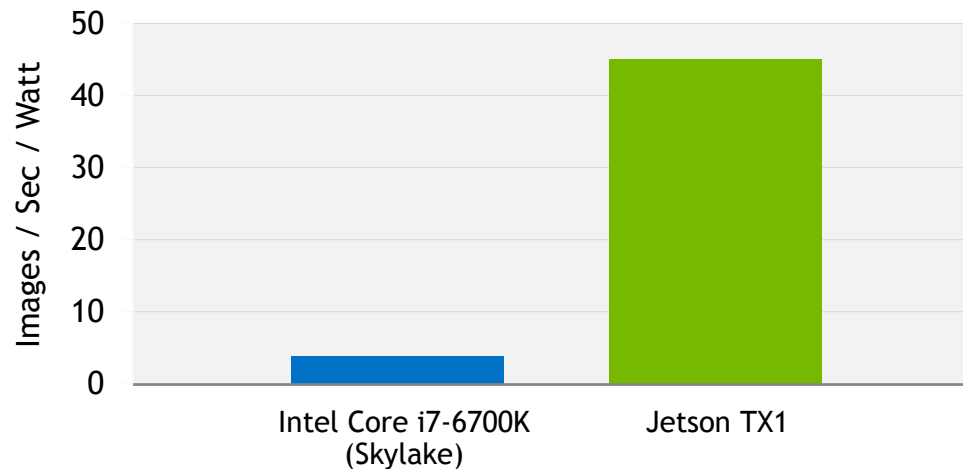
JETSON TX1

Supercomputer
on a Module



10x Energy Efficiency

Alexnet



GPU

1 TFLOPS
256-core Maxwell

CPU

64-bit ARM A57s

Memory

4GB LPDDR4
26 GB/s

Power

Under 10W

Under 10W for typical use cases

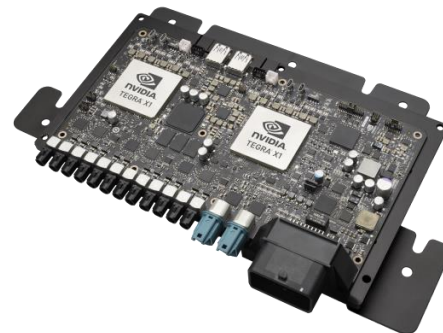
SUPERCOMPUTING EVERYWHERE



Titan X
for PC



Tesla
in the Cloud



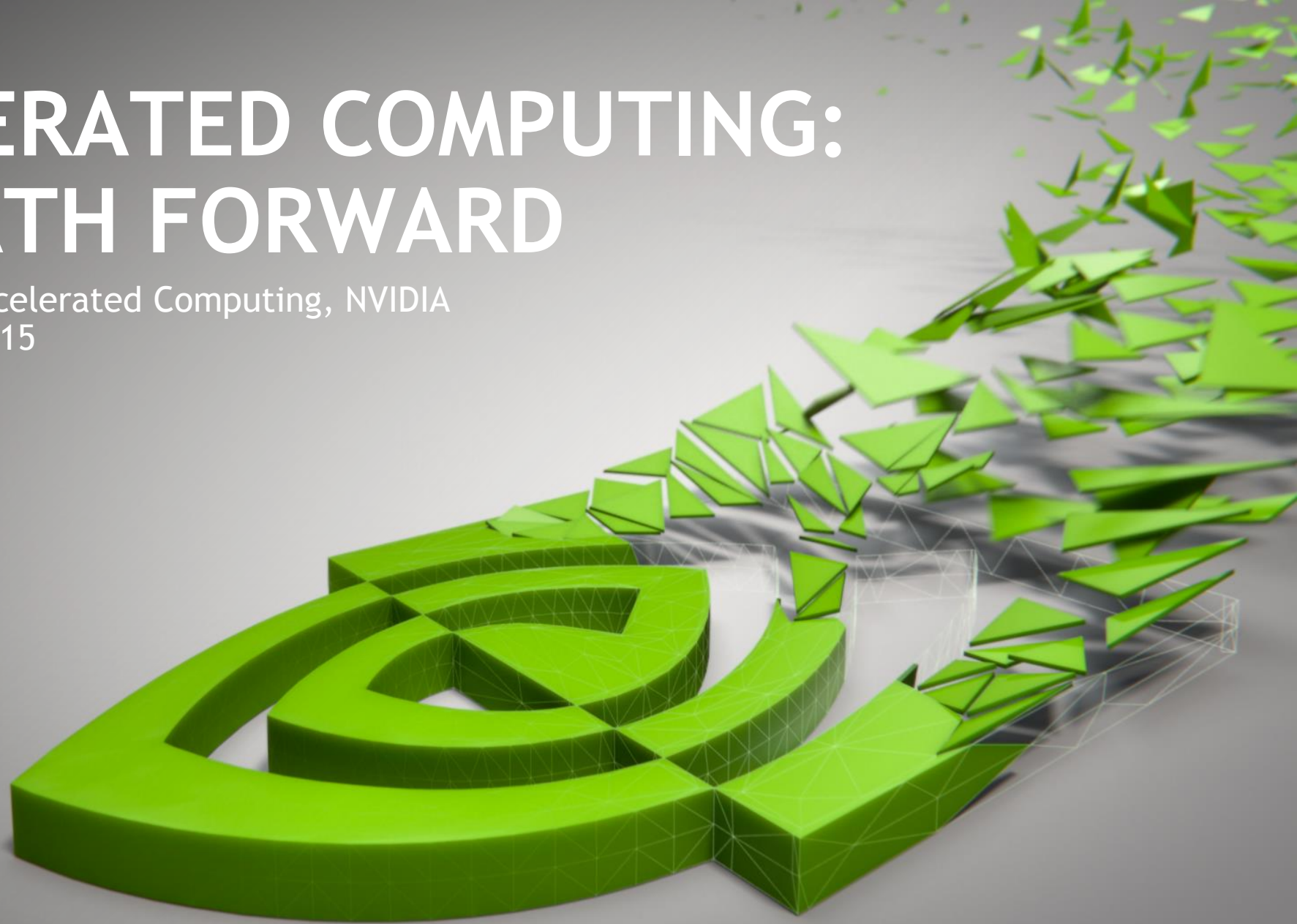
DRIVE PX
for Auto



Jetson TX1
for Robots

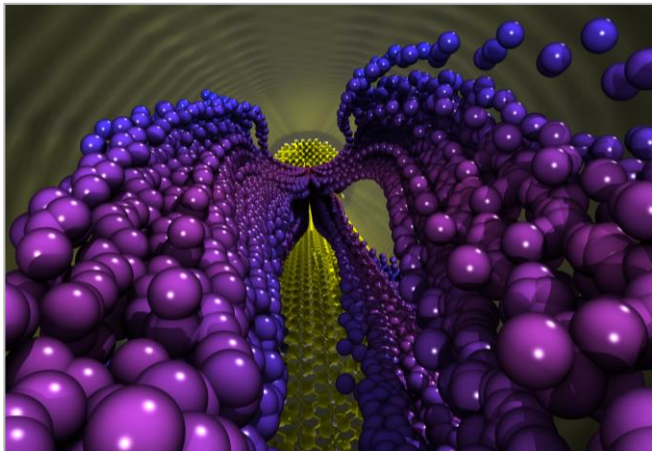
ACCELERATED COMPUTING: THE PATH FORWARD

Ian Buck, VP of Accelerated Computing, NVIDIA
SC15 | Nov. 16, 2015

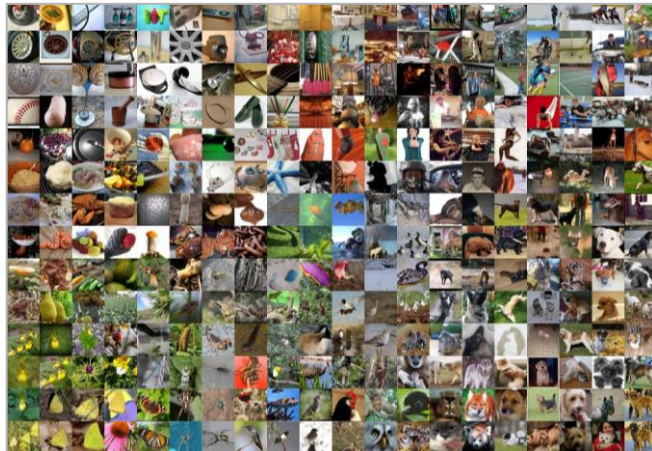


TESLA ACCELERATES DISCOVERY AND INSIGHT

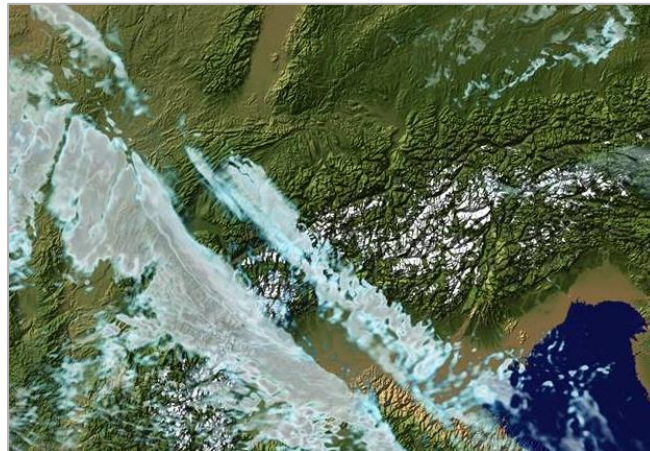
SIMULATION



MACHINE LEARNING



VISUALIZATION



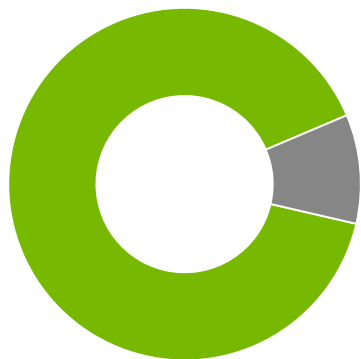
TESLA ACCELERATED COMPUTING

“Approximately a third of HPC systems operating today are equipped with accelerators and nearly half of all newly deployed systems have them.”

ACCELERATED COMPUTING: A TIPPING POINT FOR HPC, *Intersect360* Nov 2015

70% OF TOP HPC APPS NOW ACCELERATED

INTERSECT360 SURVEY OF TOP APPS



Top 10 HPC Apps
90%
Accelerated



Top 50 HPC Apps
70%
Accelerated

VASP NOW ACCELERATED



Typically Consumes
10-25%
of HPC System



1 Dual K80 Server
1.3x



4 Dual CPU Servers
1.0x

370 GPU-Accelerated Applications



TESLA FOR SIMLUATION

LIBRARIES



DIRECTIVES



LANGUAGES

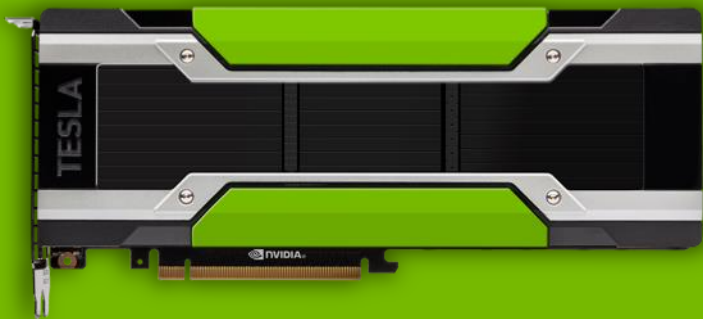


ACCELERATED COMPUTING TOOLKIT

TESLA ACCELERATED COMPUTING

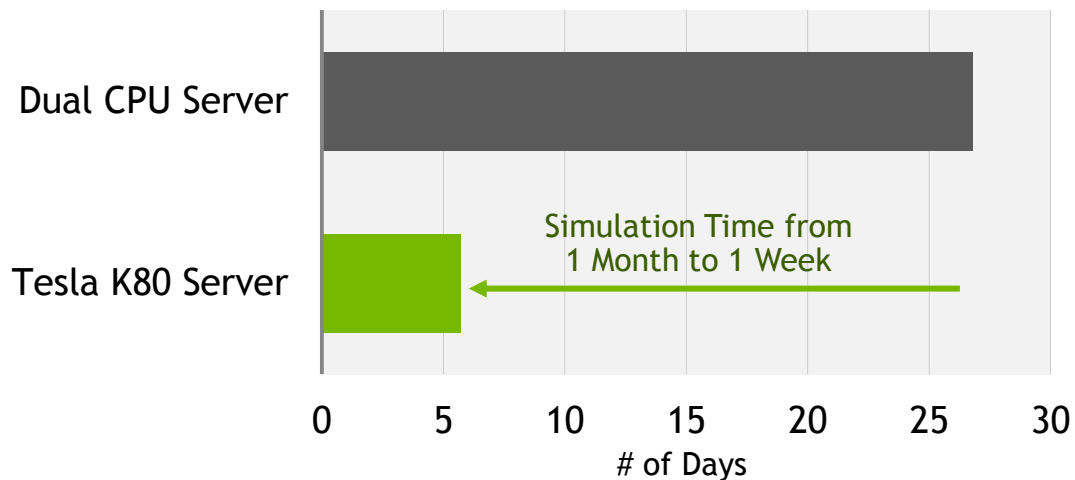
TESLA K80

World's Fastest Accelerator for HPC



5x Faster

AMBER Performance

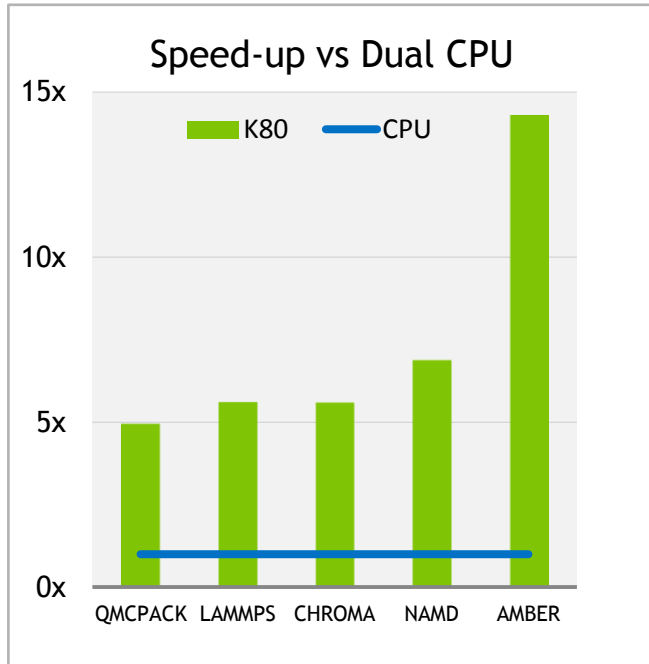


CUDA Cores	2496
Peak DP	1.9 TFLOPS
Peak DP w/ Boost	2.9 TFLOPS
GDDR5 Memory	24 GB
Bandwidth	480 GB/s
Power	300 W

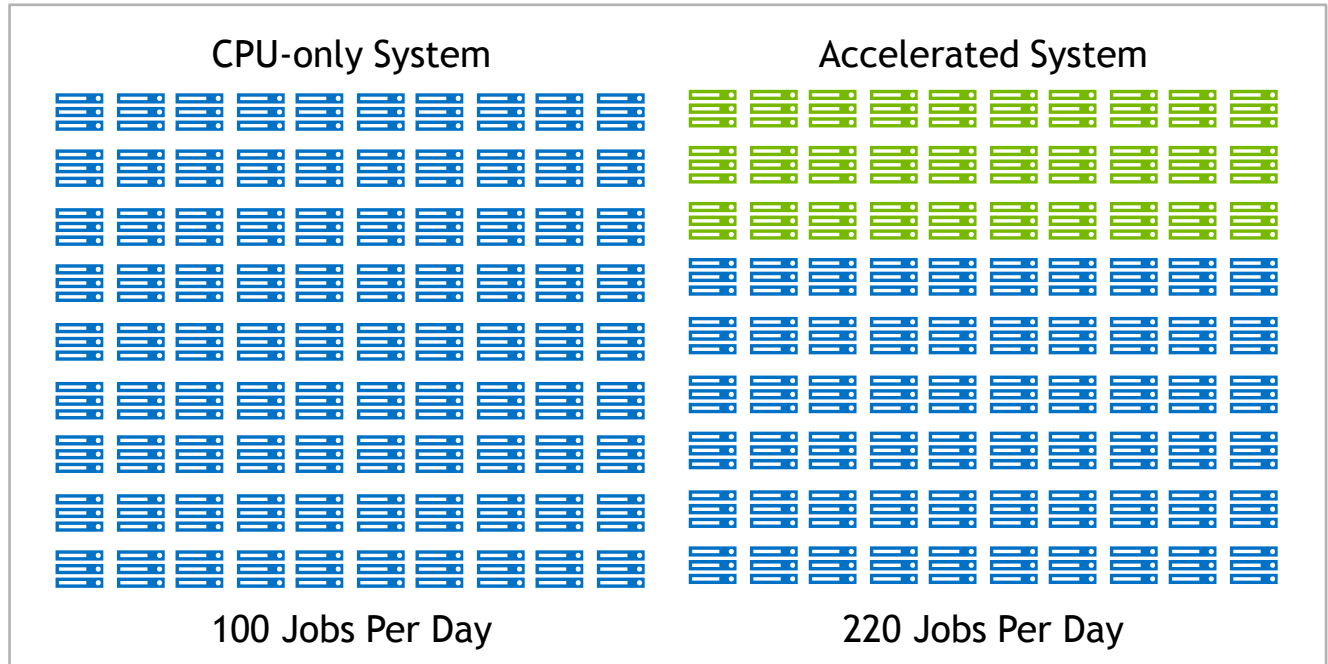
*AMBER Benchmark: PME-JAC-NVE Simulation for 1 microsecond
CPU: E5-2698v3 @ 2.3GHz. 64GB System Memory, CentOS 6.2*

APPLICATION PERFORMANCE BOOSTS DATA CENTER THROUGHPUT

TESLA K80: 5X FASTER



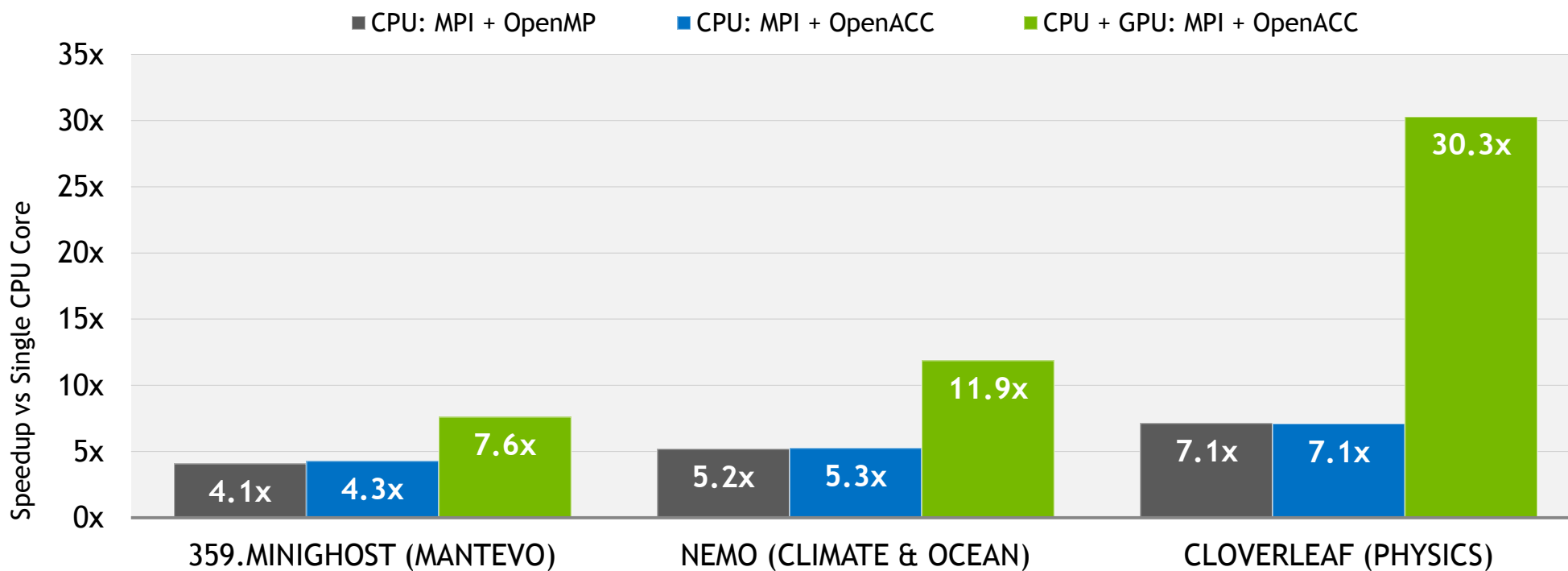
1/3 OF NODES ACCELERATED, 2X SYSTEM THROUGHPUT



OPENACC DELIVERS TRUE PERF PORTABILITY

Paving the Path Forward: Single Code for All HPC Processors

Application Performance Benchmark



359.miniGhost: CPU: Intel Xeon E5-2698 v3, 2 sockets, 32-cores total, GPU: Tesla K80- single GPU

NEMO: Each socket CPU: Intel Xeon E5-2698 v3, 16 cores; GPU: NVIDIA K80 both GPUs

CLOVERLEAF: CPU: Dual socket Intel Xeon CPU E5-2690 v2, 20 cores total, GPU: Tesla K80 both GPUs

TESLA HYPERSCALE FOR MACHINE LEARNING

HYPERSCALE SUITE



GPU REST Engine



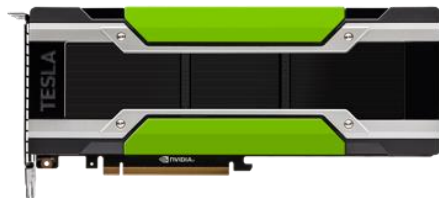
GPU Accelerated
FFmpeg



Image Compute
Engine

TESLA M40

POWERFUL: Fastest Deep Learning Performance



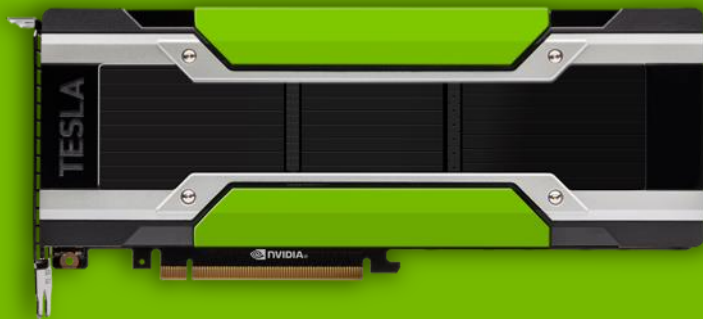
TESLA M4

LOW POWER: Highest Hyperscale Throughput



TESLA M40

World's Fastest Accelerator for Deep Learning



8x Faster Caffe Performance

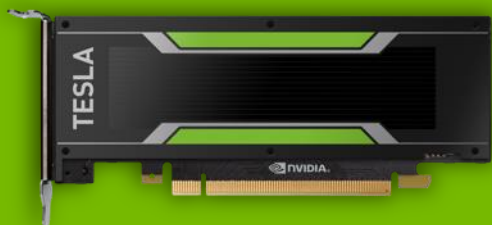


CUDA Cores	3072
Peak SP	7 TFLOPS
GDDR5 Memory	12 GB
Bandwidth	288 GB/s
Power	250W

*Caffe Benchmark: AlexNet training throughput based on 20 iterations,
CPU: E5-2697v2 @ 2.70GHz. 64GB System Memory, CentOS 6.2*

TESLA M4

Highest Throughput Hyperscale Workload Acceleration



Video
Processing

4x

Stabilization and
Enhancements



Image
Processing

5x

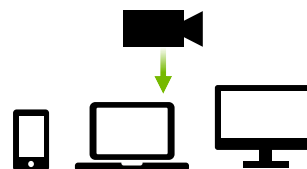
Resize, Filter, Search,
Auto-Enhance



Video
Transcode

2x

H.264 & H.265, SD & HD



Machine
Learning
Inference

2x



CUDA Cores

1024

Peak SP

2.2 TFLOPS

GDDR5 Memory

4 GB

Bandwidth

88 GB/s

Form Factor

PCIe Low Profile

Power

50 - 75 W

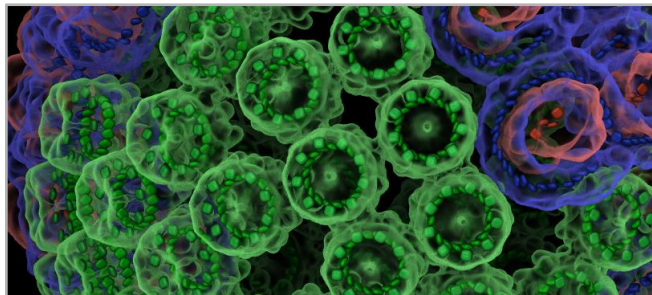
Preliminary specifications. Subject to change.

TESLA FOR VISUALIZATION

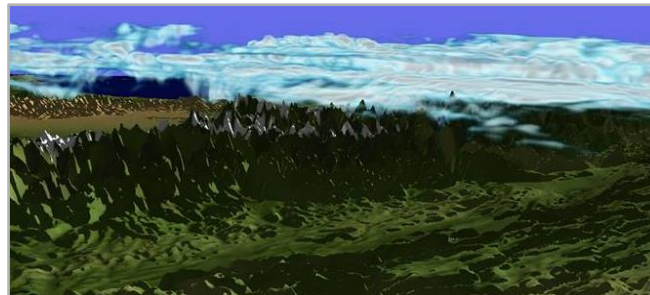
IRAY



OPTIX



INDEX



VISUALIZATION TOOLS FOR HPC

TESLA ACCELERATED COMPUTING

GROWING ADOPTION IN CLIMATE & WEATHER



MeteoSwiss Deploys World's First Accelerated Weather Supercomputer

2x higher resolution for daily forecasts

14x more simulation with ensemble approach for medium range forecasts



NOAA Chooses Tesla To Improve Weather Forecast Research

Develop global model with 3km resolution, five-fold increase from today's resolution

Improved resolution requires 40x higher in computational complexity

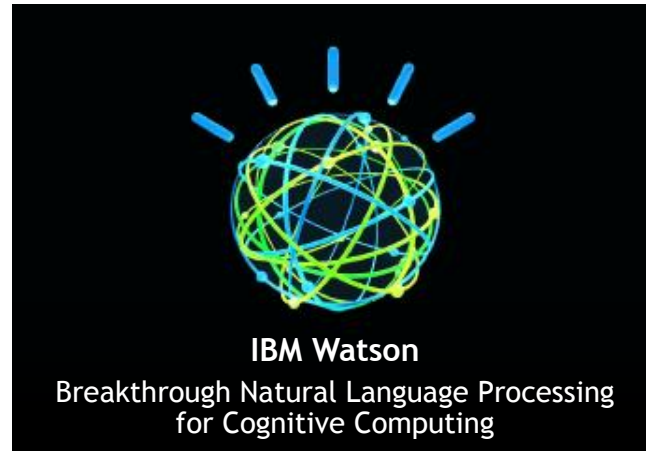


NEXT-GEN SUPERCOMPUTERS ARE GPU-ACCELERATED

SIMULATION



MACHINE LEARNING



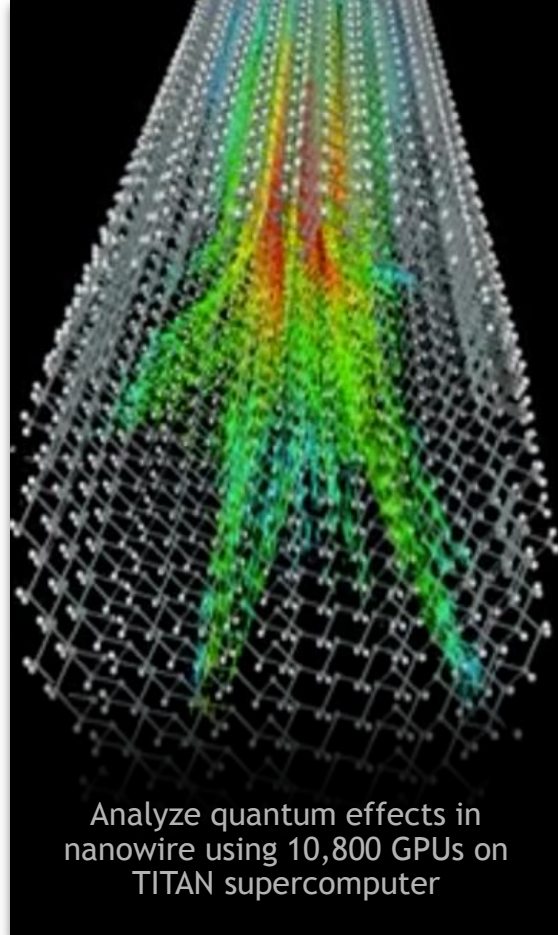
VISUALIZATION



TESLA ACCELERATED COMPUTING



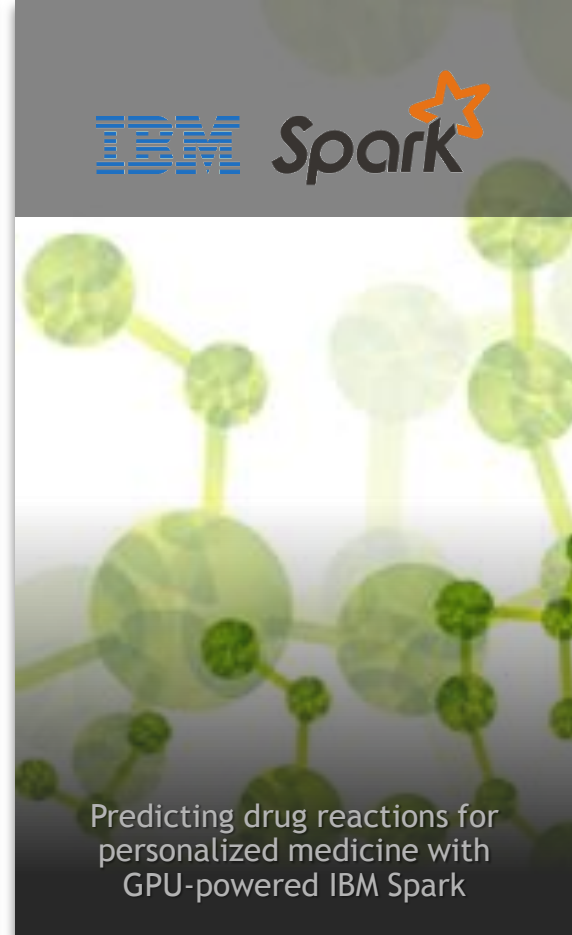
World's first in-situ weather simulation, running on Meteoswiss supercomputer



Analyze quantum effects in nanowire using 10,800 GPUs on TITAN supercomputer



Simulation of deadly tornado that hit El Reno, Oklahoma on May 24, 2011



IBM Spark

Predicting drug reactions for personalized medicine with GPU-powered IBM Spark

ACCELERATED SCIENCE AND DATA ANALYTICS ON DISPLAY AT SC'15

