

The logo for GPU Technology Conference, featuring the text "GPU TECHNOLOGY CONFERENCE" in white on a green trapezoidal background.

**GPU** TECHNOLOGY  
CONFERENCE

# LEAPS IN VISUAL COMPUTING

JEN-HSUN HUANG, CO-FOUNDER & CEO | GTC 2015

# FOUR ANNOUNCEMENTS

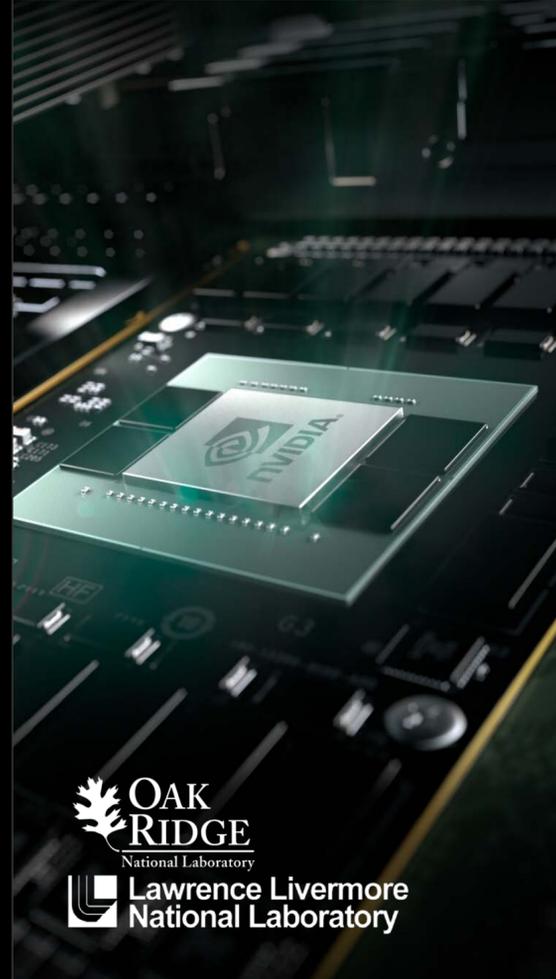
A New GPU  
and  
Deep Learning

A Very Fast Box  
and  
Deep Learning

Roadmap Reveal  
and  
Deep Learning

Self-Driving Cars  
and  
Deep Learning

# AMAZING YEAR IN VISUAL COMPUTING



# 10X GROWTH IN GPU COMPUTING

2008

**150,000**

CUDA Downloads



**27**

CUDA Apps



**60**

Universities Teaching



**4,000**

Academic Papers



**6,000**

Tesla GPUs



**77**

Supercomputing Teraflops



# 10X GROWTH IN GPU COMPUTING

2008

**150,000**  
CUDA Downloads



**27**  
CUDA Apps



**60**  
Universities Teaching



**4,000**  
Academic Papers



**6,000**  
Tesla GPUs



**77**  
Supercomputing Teraflops



2015



**3 Million**  
CUDA Downloads

# 10X GROWTH IN GPU COMPUTING

2008

**150,000**  
CUDA Downloads



**27**  
CUDA Apps



**60**  
Universities Teaching



**4,000**  
Academic Papers



**6,000**  
Tesla GPUs



**77**  
Supercomputing Teraflops



2015



**3 Million**  
CUDA Downloads



**319**  
CUDA Apps

# 10X GROWTH IN GPU COMPUTING

2008

150,000  
CUDA Downloads



27  
CUDA Apps



60  
Universities Teaching



4,000  
Academic Papers



6,000  
Tesla GPUs



77  
Supercomputing Teraflops



2015



3 Million  
CUDA Downloads



319  
CUDA Apps



800  
Universities Teaching

# 10X GROWTH IN GPU COMPUTING

2008

150,000  
CUDA Downloads



27  
CUDA Apps



60  
Universities Teaching



4,000  
Academic Papers



6,000  
Tesla GPUs



77  
Supercomputing Teraflops



2015



3 Million  
CUDA Downloads



319  
CUDA Apps



800  
Universities Teaching



60,000  
Academic Papers

# 10X GROWTH IN GPU COMPUTING

2008

2015

150,000  
CUDA Downloads



3 Million  
CUDA Downloads

27  
CUDA Apps



319  
CUDA Apps

60  
Universities Teaching



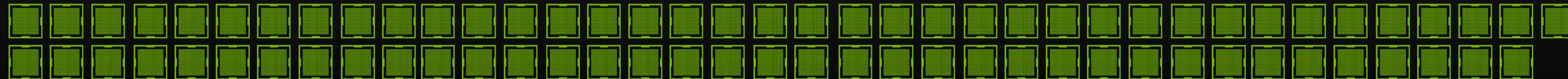
800  
Universities Teaching

4,000  
Academic Papers



60,000  
Academic Papers

6,000  
Tesla GPUs



450,000  
Tesla GPUs

77  
Supercomputing Teraflops



# 10X GROWTH IN GPU COMPUTING

2008

2015

150,000  
CUDA Downloads



3 Million  
CUDA Downloads

27  
CUDA Apps



319  
CUDA Apps

60  
Universities Teaching



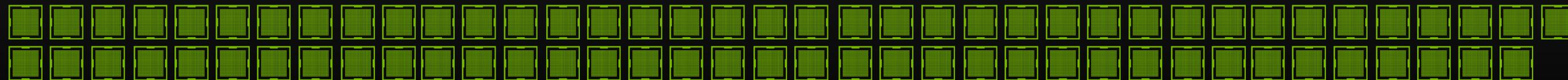
800  
Universities Teaching

4,000  
Academic Papers



60,000  
Academic Papers

6,000  
Tesla GPUs

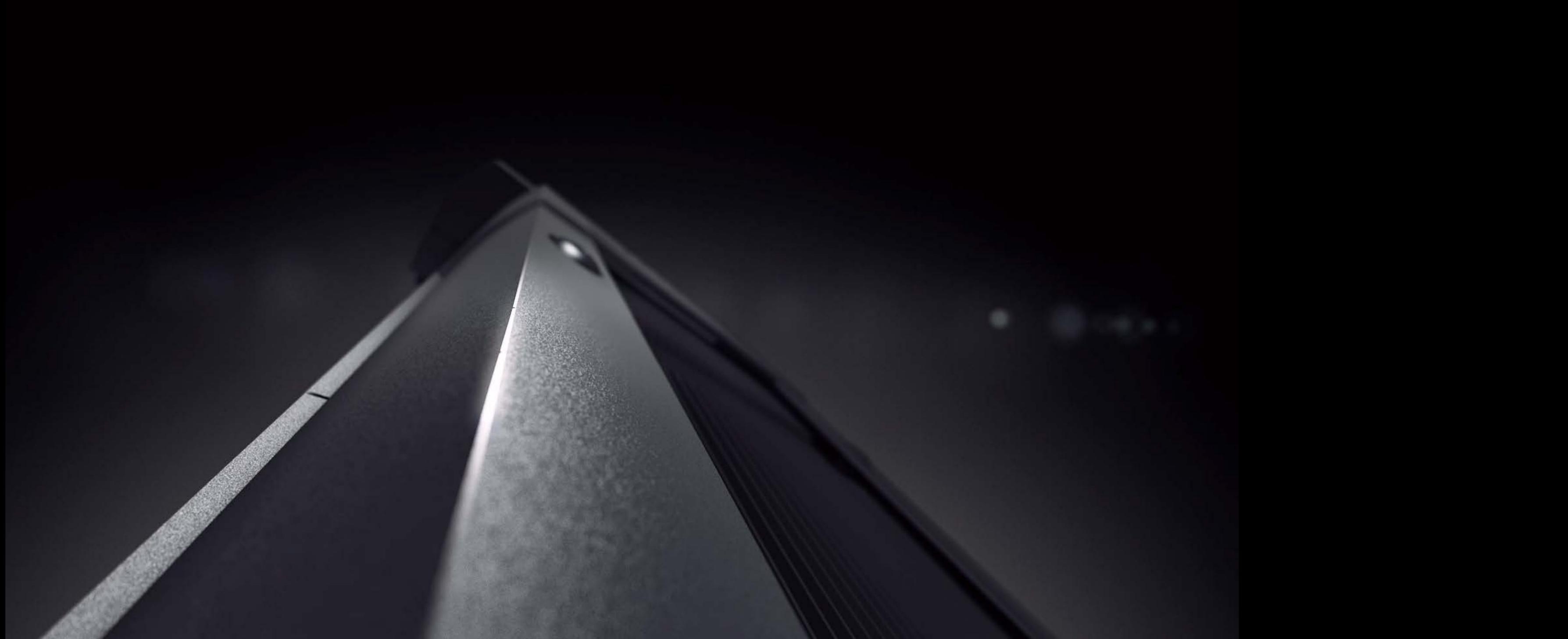


450,000  
Tesla GPUs

77  
Supercomputing Teraflops



54,000  
Supercomputing Teraflops



# TITAN X

THE WORLD'S FASTEST GPU

8 Billion Transistors

3,072 CUDA Cores

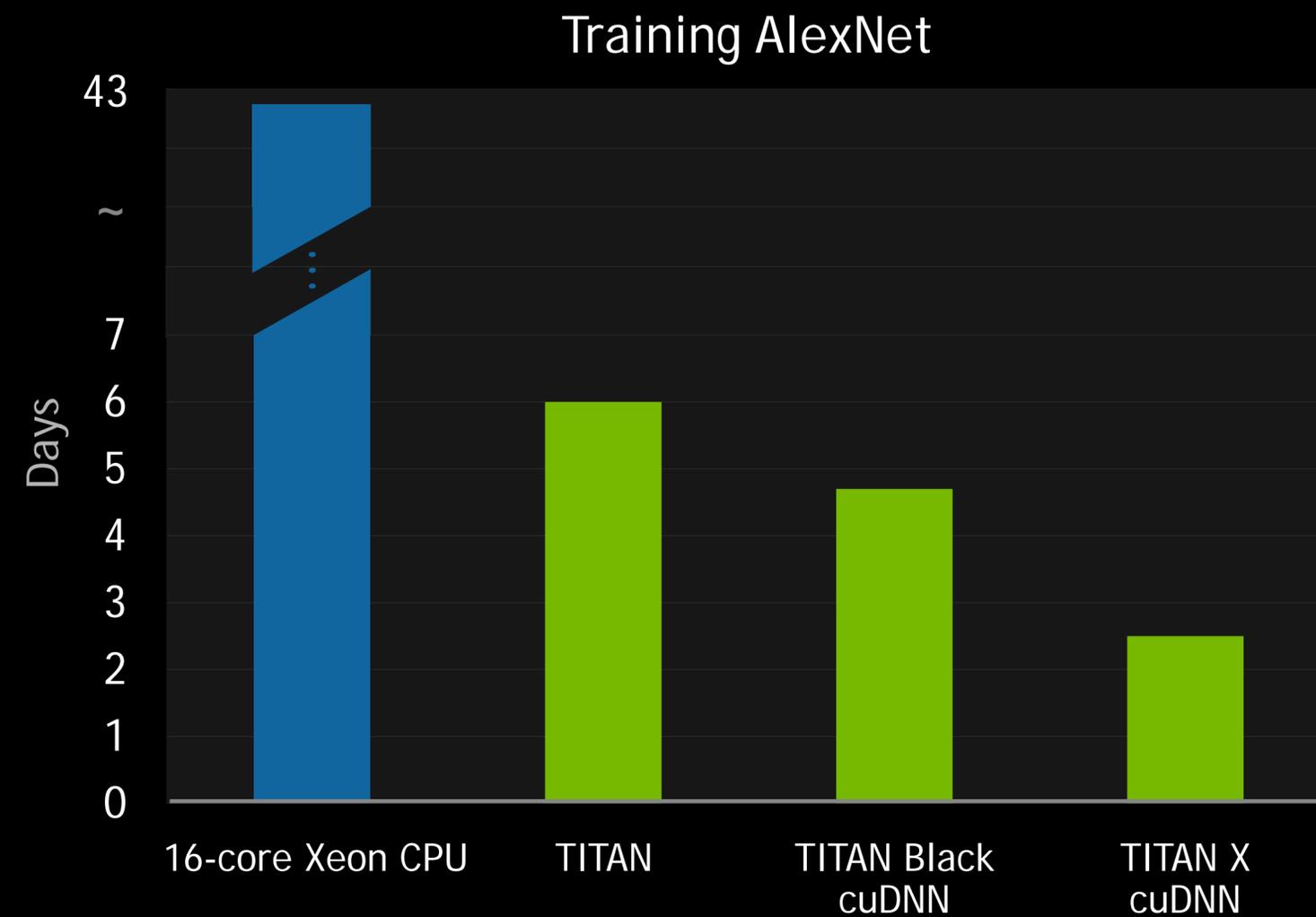
7 TFLOPS SP / 0.2 TFLOPS DP

12GB Memory





# TITAN X FOR DEEP LEARNING



# TITAN X

THE WORLD'S FASTEST GPU

8 Billion Transistors

3,072 CUDA Cores

7 TFLOPS SP / 0.2 TFLOPS DP

12GB Memory

\$999



# FOUR ANNOUNCEMENTS

A New GPU  
and  
Deep Learning

A Very Fast Box  
and  
Deep Learning

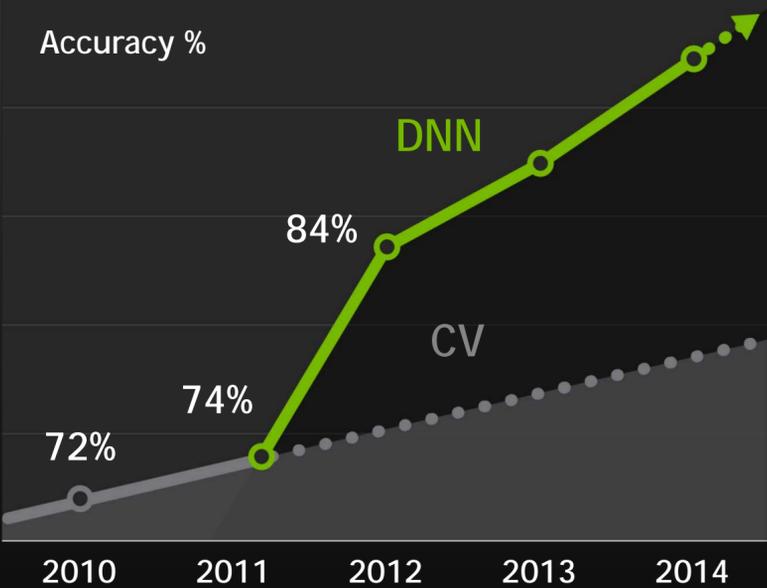
Roadmap Reveal  
and  
Deep Learning

Self-Driving Cars  
and  
Deep Learning

# A SHORT HISTORY OF DEEP LEARNING



**Convolutional Neural Networks for Handwritten Digit Recognition**  
LECUN, BOTTOU, BENGIO, HAFFNER, 1998



**ImageNet Classification with NVIDIA GPUs**  
KRIZHEVSKY, HINTON, ET AL., 2012

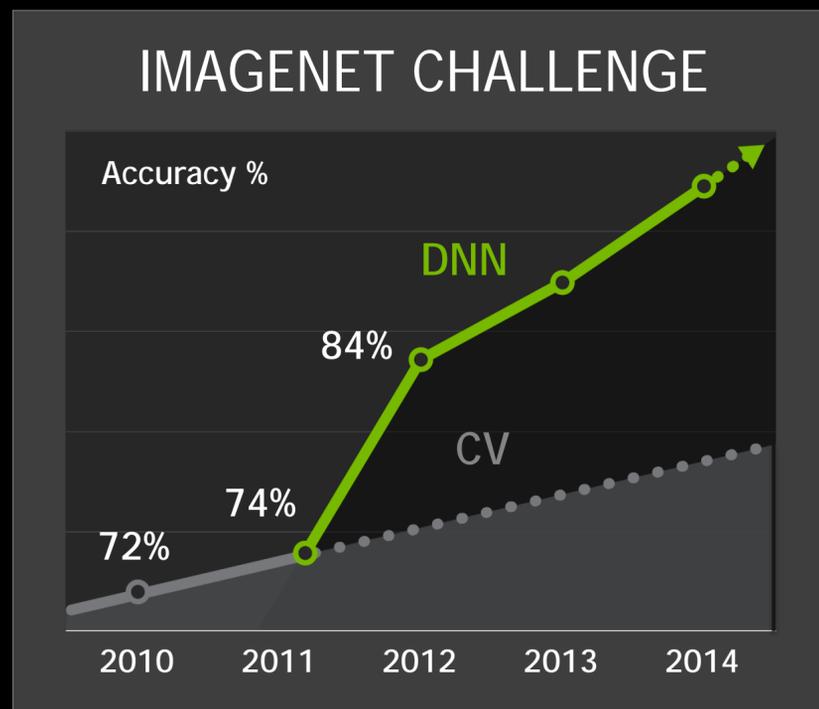
1995

2000

2005

2010

2015



*“Deep Image: Scaling up Image Recognition”*

— Baidu: 5.98%, Jan. 13, 2015

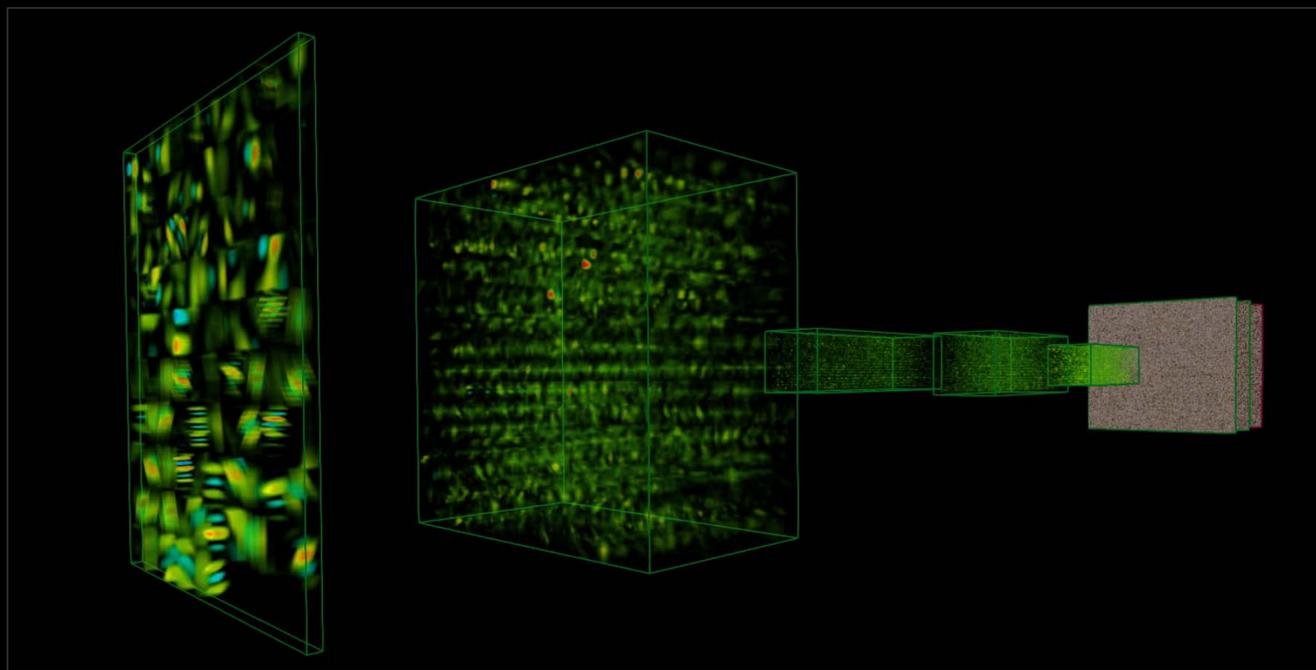
*“Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”*

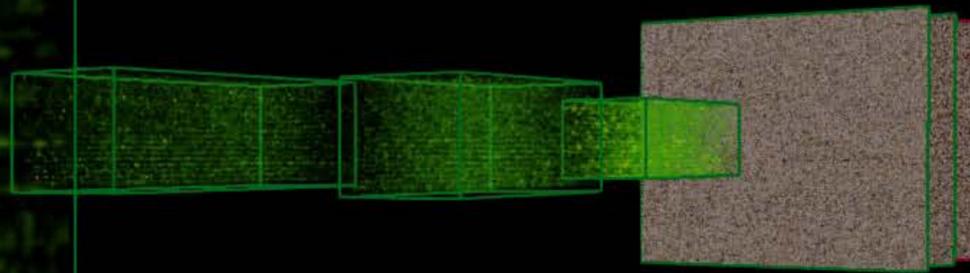
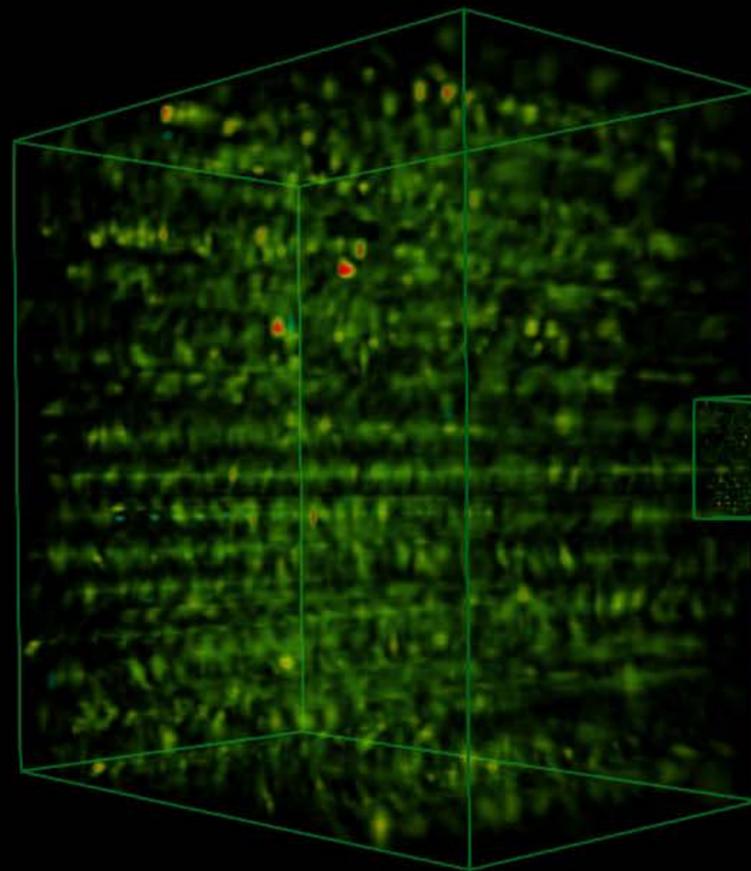
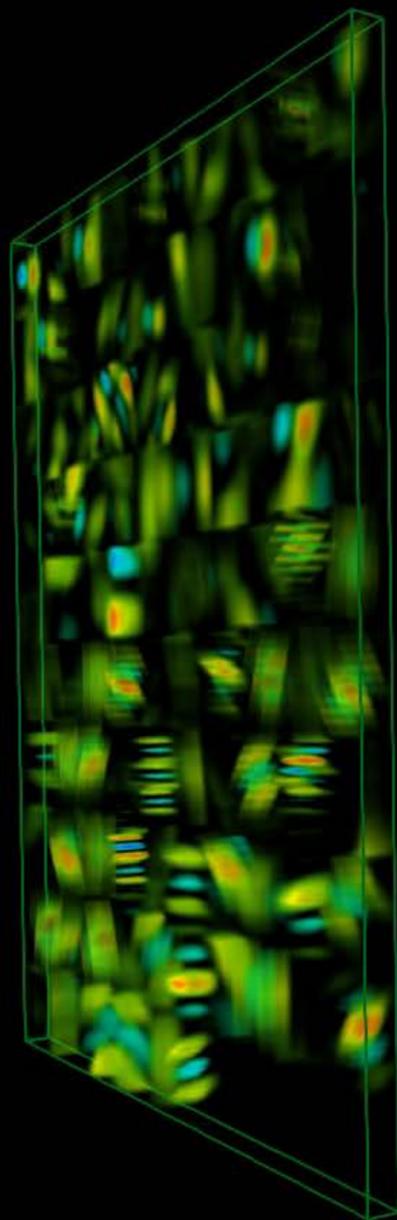
— Microsoft: 4.94%, Feb. 6, 2015

*“Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariant Shift”*

— Google: 4.82%, Feb. 11, 2015

# THE BIG BANG





DEEP LEARNING  
VISUALIZED

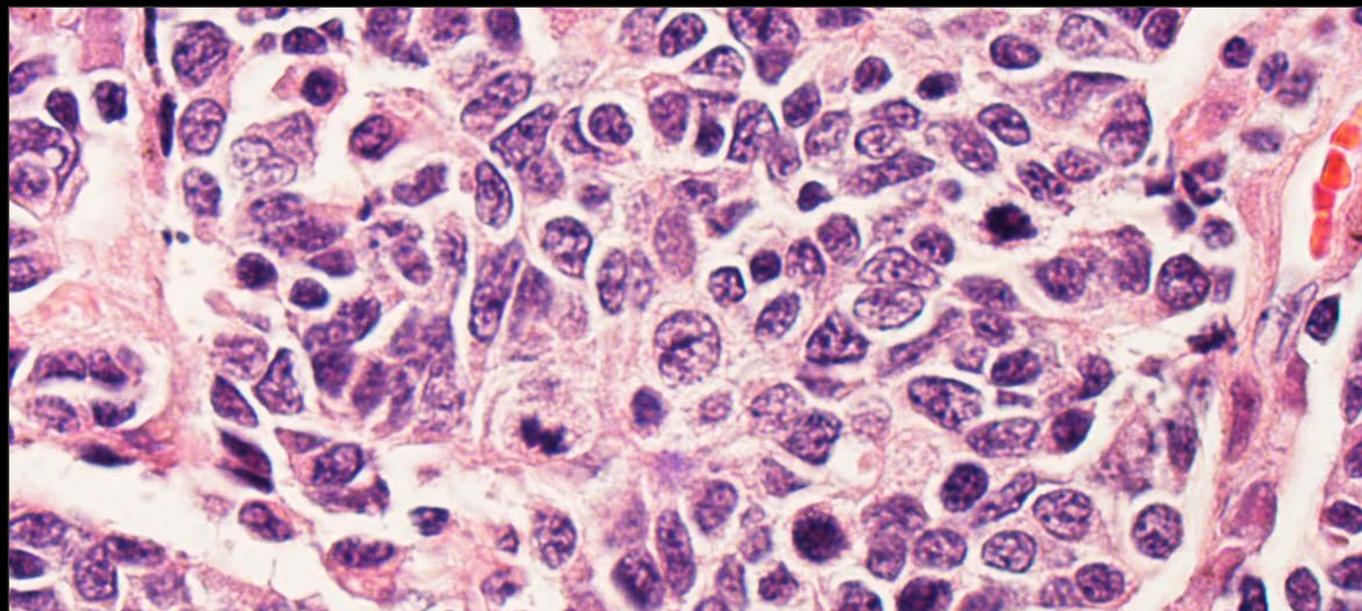
# GPU-ACCELERATED DEEP LEARNING

## START-UPS

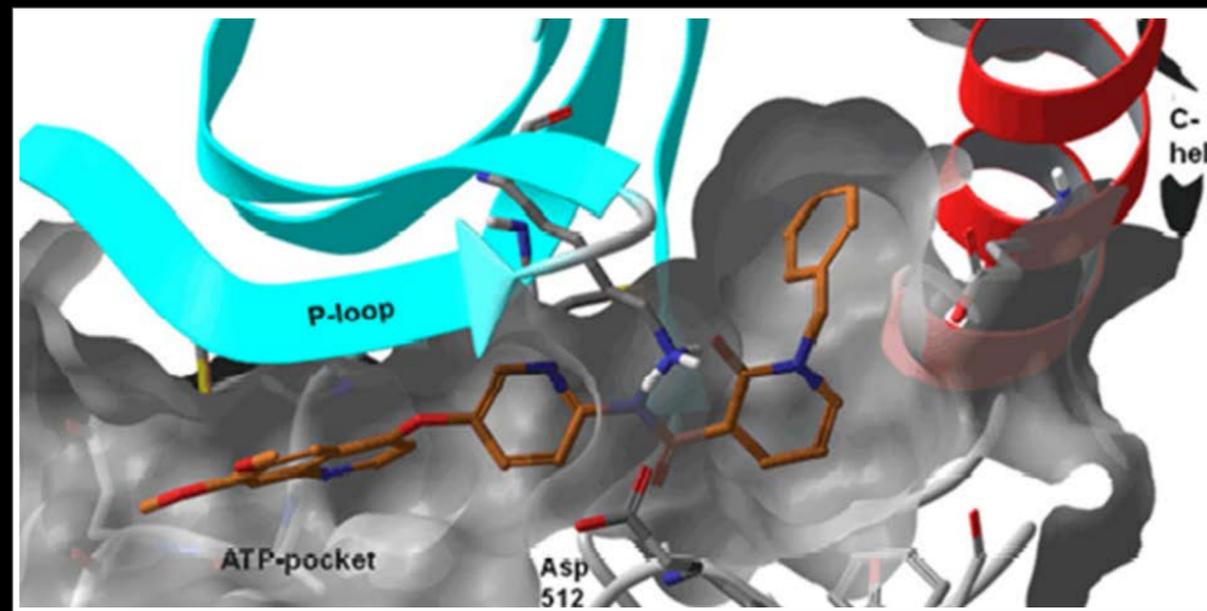


# DEEP LEARNING REVOLUTIONIZING MEDICAL RESEARCH

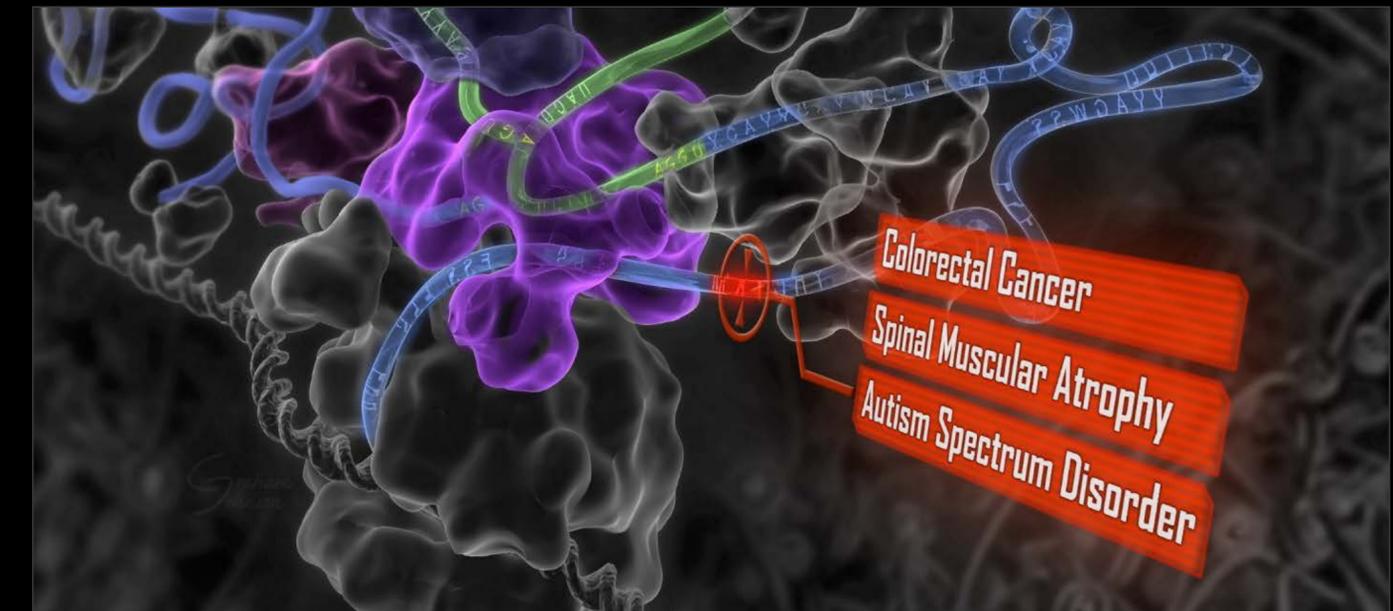
Detecting Mitosis in Breast Cancer Cells  
— *IDSIA*



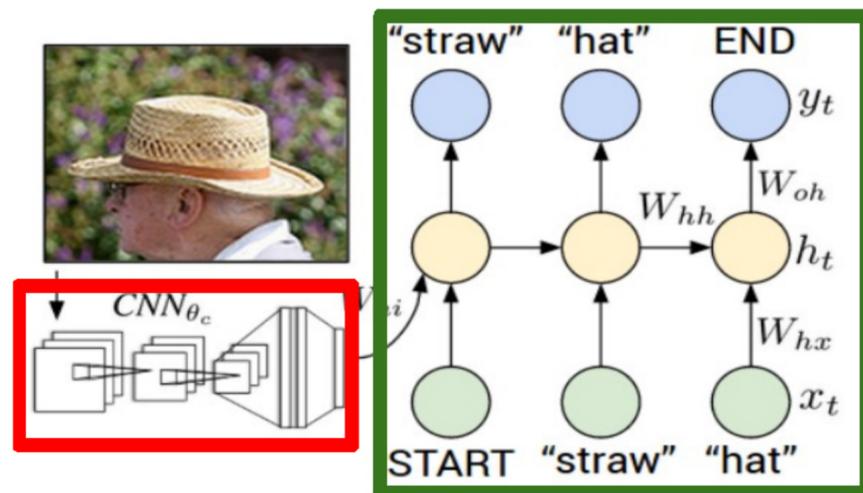
Predicting the Toxicity of New Drugs  
— *Johannes Kepler University*



Understanding Gene Mutation to Prevent Disease  
— *University of Toronto*



## Recurrent Neural Network



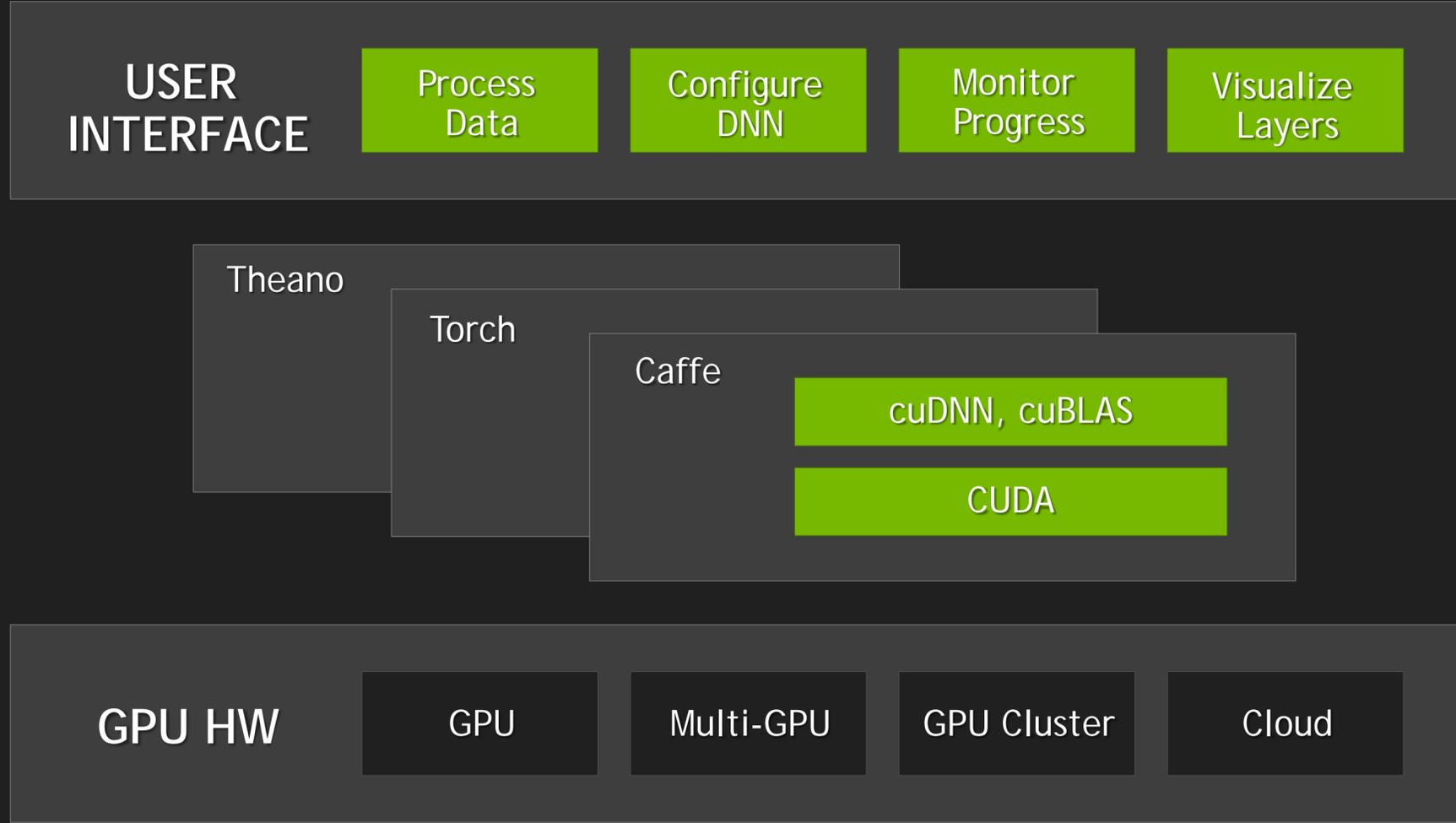
## Convolutional Neural Network

a bird perched on a branch of a tree



“Automated Image Captioning with ConvNets and Recurrent Nets”

—Andrej Karpathy, Fei-Fei Li



# DIGITS

## DEEP GPU TRAINING SYSTEM FOR DATA SCIENTISTS

- Design DNNs
- Visualize activations
- Manage multiple trainings

# DIGITS

## Process Data

**Image Classification Dataset**  
voc\_cropped@256x256  
Image Classification Dataset

**Job Information**  
Job Directory: /home/michaelo/.digits/jobs/20150311-171431-e0d8  
Image Type: Color  
Image Dimensions: 256x256  
Resize Mode: half\_crop

**Parse Folder (train/val)**  
Folder: http://sqr1/data/images/voc\_cropped/  
Number of categories: 20  
Training images: 26759  
Validation images: 8917 (25.0%)

**Job Status Done**  
• Initialized at Wed Mar 11, 05:16 PM (1 second)  
• Running at Wed Mar 11, 05:16 PM (2 minutes, 25 seconds)  
• Done at Wed Mar 11, 05:16 PM (Total - 2 minutes, 26 seconds)

**Create DB (train)**  
Input file: train.txt  
DB Entries: 26759

## Configure DNN

**Select Dataset**  
PASCAL VOC ILSVRC 2012 MNIST Dataset

**Data Transform**  
Crop Size: none  
 Subtract Mean

**Solver Options**  
Training epochs: 30  
Validation interval (in epochs): 1  
Batch size: 100  
Base Learning Rate: 0.01

**Custom Network**

```
layer {
  name: "conv1"
  type: "Convolution"
  bottom: "data"
  top: "conv1"
  param {
    lr_mult: 1
    decay_mult: 1
  }
}
```

**Model Name**  
ImageNet  
Create

## Monitor Progress

**Solver**  
solver.prototxt  
Network (train/val): train\_val.prototxt  
Network (deploy): deploy.prototxt

**Dataset**  
voc\_cropped@256x256  
Done Wed Mar 11, 05:16:57 PM  
Image Size: 256x256  
Image Type: COLOR  
Create DB (train): 26759 images  
Create DB (val): 8917 images

**Job Status Running**  
• Initialized at 08:28:15 AM (1 second)  
• Running at 08:28:16 AM

**Train Caffe Model Running**  
37%  
Estimated time remaining: 30 minutes, 31 seconds

• Initialized at 08:28:15 AM (1 second)  
• Running at 08:28:16 AM

**Graph:** Loss (train) (red), Loss (val) (orange), Accuracy (val) (blue) vs. Epochs.

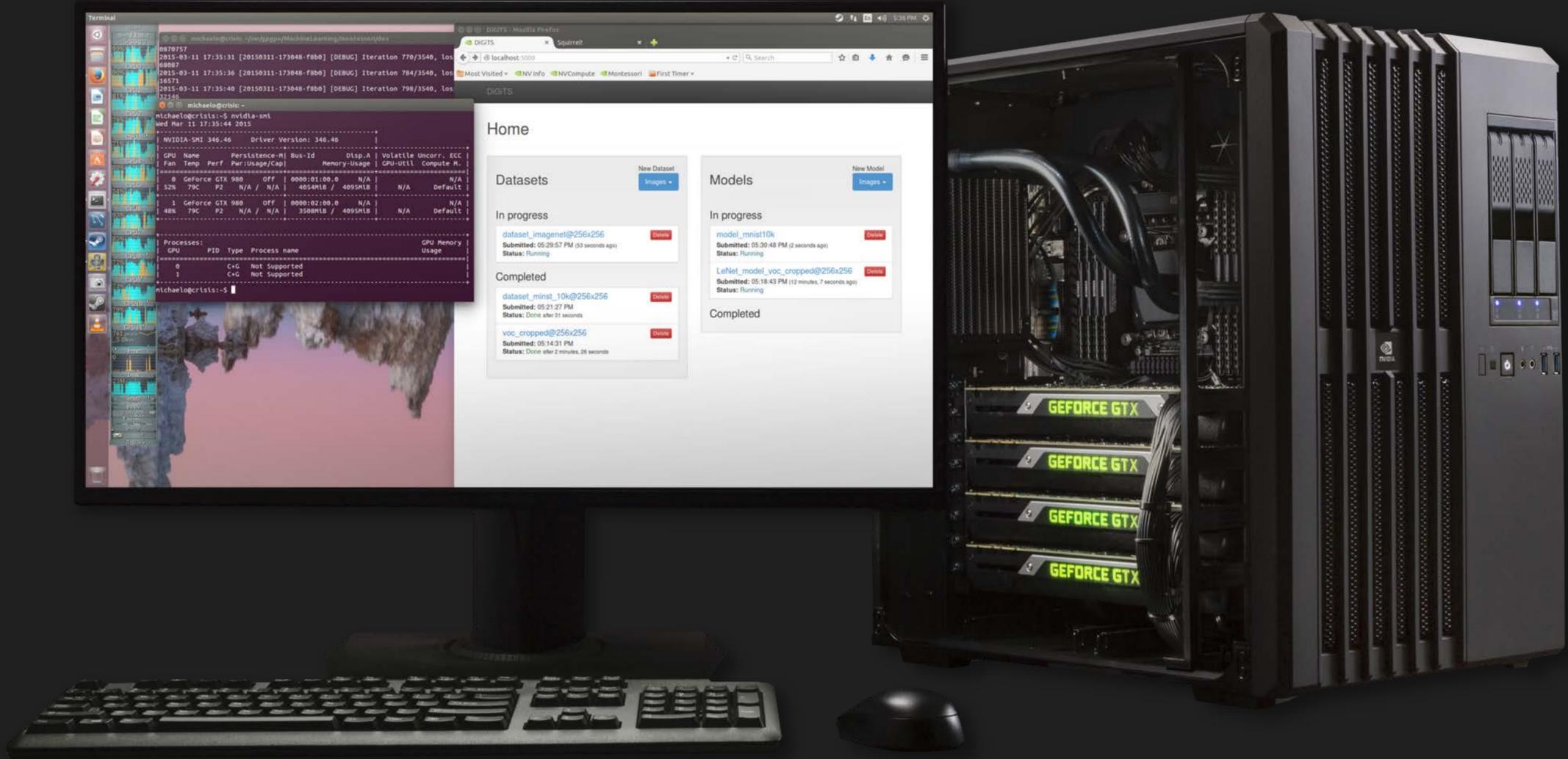
## Visualize Layers

**Test Image**  
8

**Predictions**  
8: 100.0%  
3: 0.0%  
0: 0.0%  
6: 0.0%  
4: 0.0%

**Layer Activations**  
conv1: [Heatmap of digit 8]  
pool1: [Three heatmaps of digit 8]

**Weights**  
[Weight matrix visualization]



# DIGITS DEVBOX

World's fastest GPU

Max GPU out of a plug

Multi-GPU training & inference

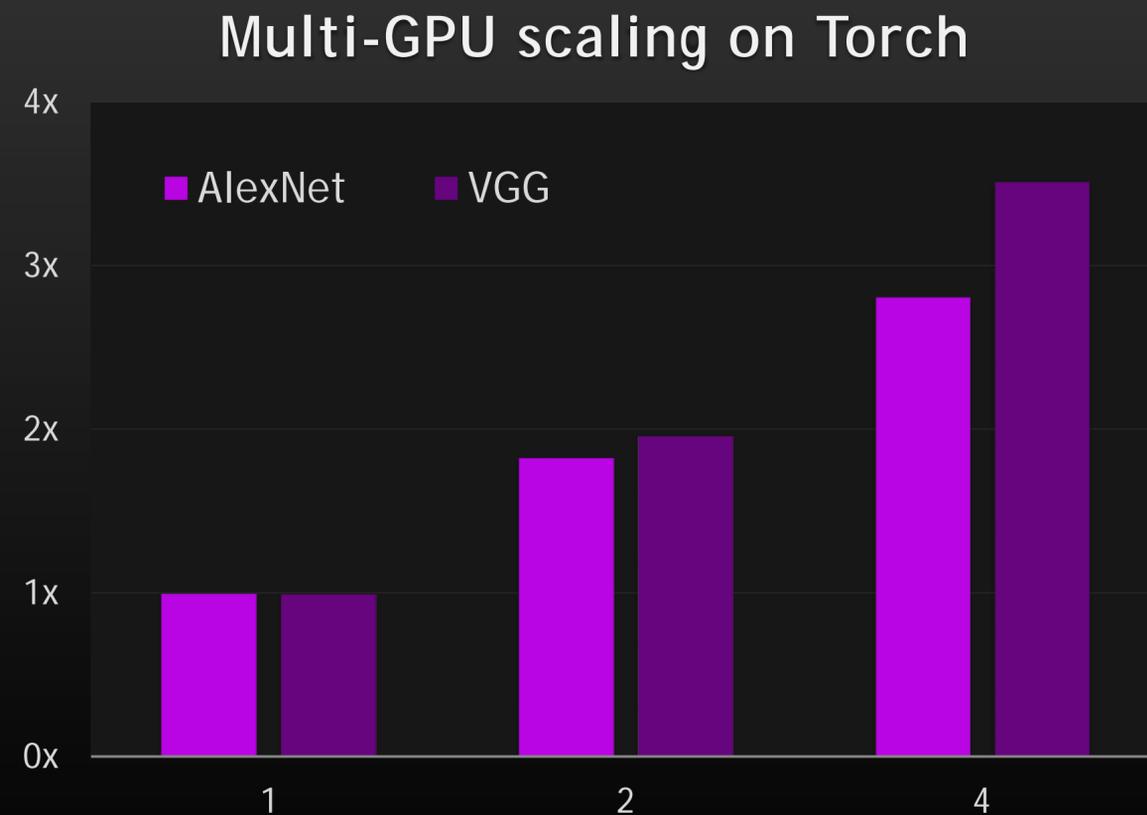
# DIGITS DEVBOX — EARLY RESULTS

*“DIGITS makes it way easier to design the best network for the job”*

— *Simon Osindero*

*A.I. Architech*

**flickr**



*“I’ve never seen AlexNet run this fast...TitanX is a monster, Crazy Fast”*

— *Soumith Chintala*

*Research Engineer*

**facebook**



Search NVIDIA

USA - United States

DRIVERS | PRODUCTS | COMMUNITIES | SUPPORT | SHOP | ABOUT NVIDIA

## DEEP LEARNING

Home > CUDA ZONE > Tools & Ecosystem > Key Technologies > NVIDIA® DIGITS™ DevBox

Subscribe

### NVIDIA® DIGITS™ DevBox The World's Fastest Deskside Deep Learning System



Deep learning is one of the fastest-growing segments of the machine learning/artificial intelligence field and a key area of innovation in computing. With researchers creating new deep learning algorithms and industries producing and collecting unprecedented amounts of data, computational capability is the key to unlocking insights from data.

DEVBOX ACCESS PROGRAM

BUILD YOUR OWN DEVBOX



# DIGITS DEVBOX

Available May 2015  
\$15,000

# FOUR ANNOUNCEMENTS

A New GPU  
and  
Deep Learning

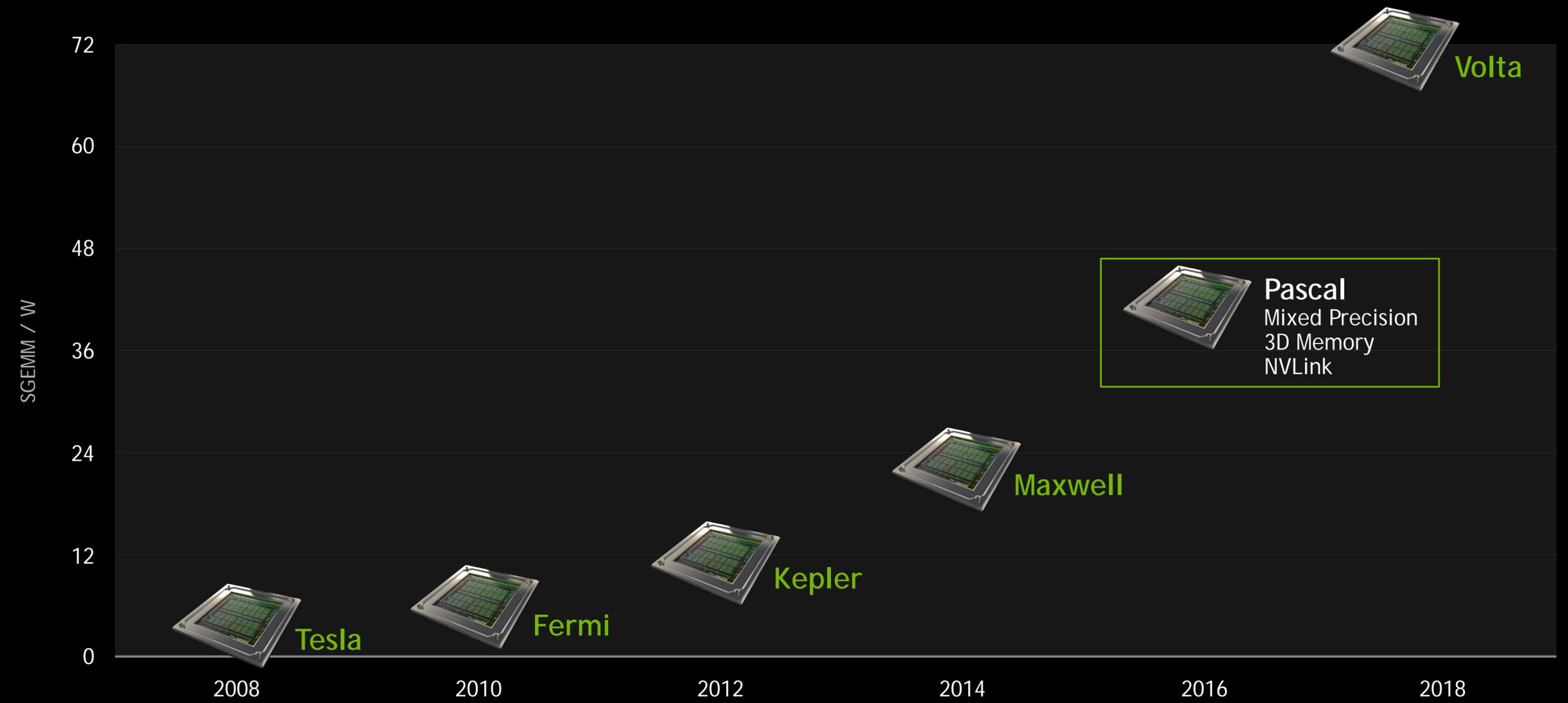
A Very Fast Box  
and  
Deep Learning

Roadmap Reveal  
and  
Deep Learning

Self-Driving Cars  
and  
Deep Learning

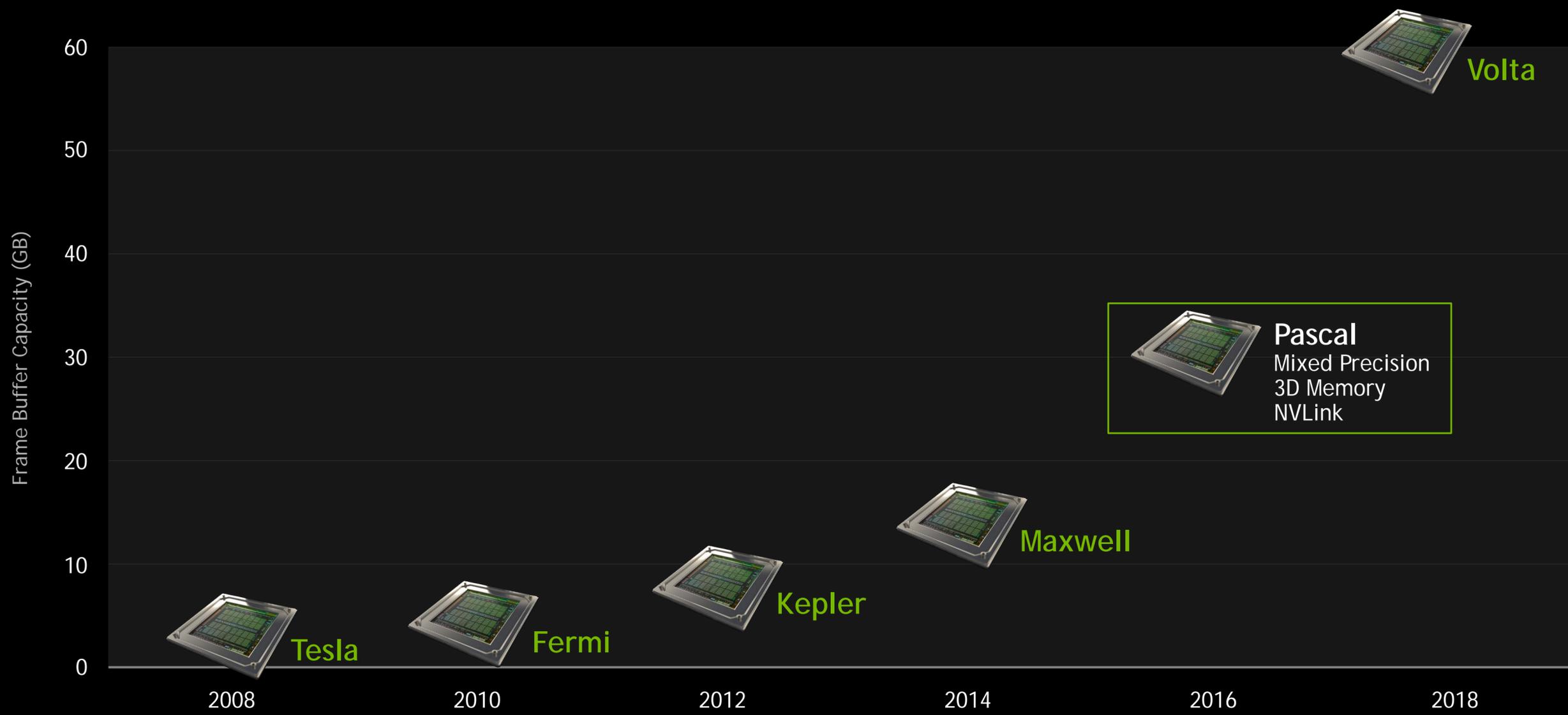
# GPU ROADMAP

Pascal 2x SGEMM/W



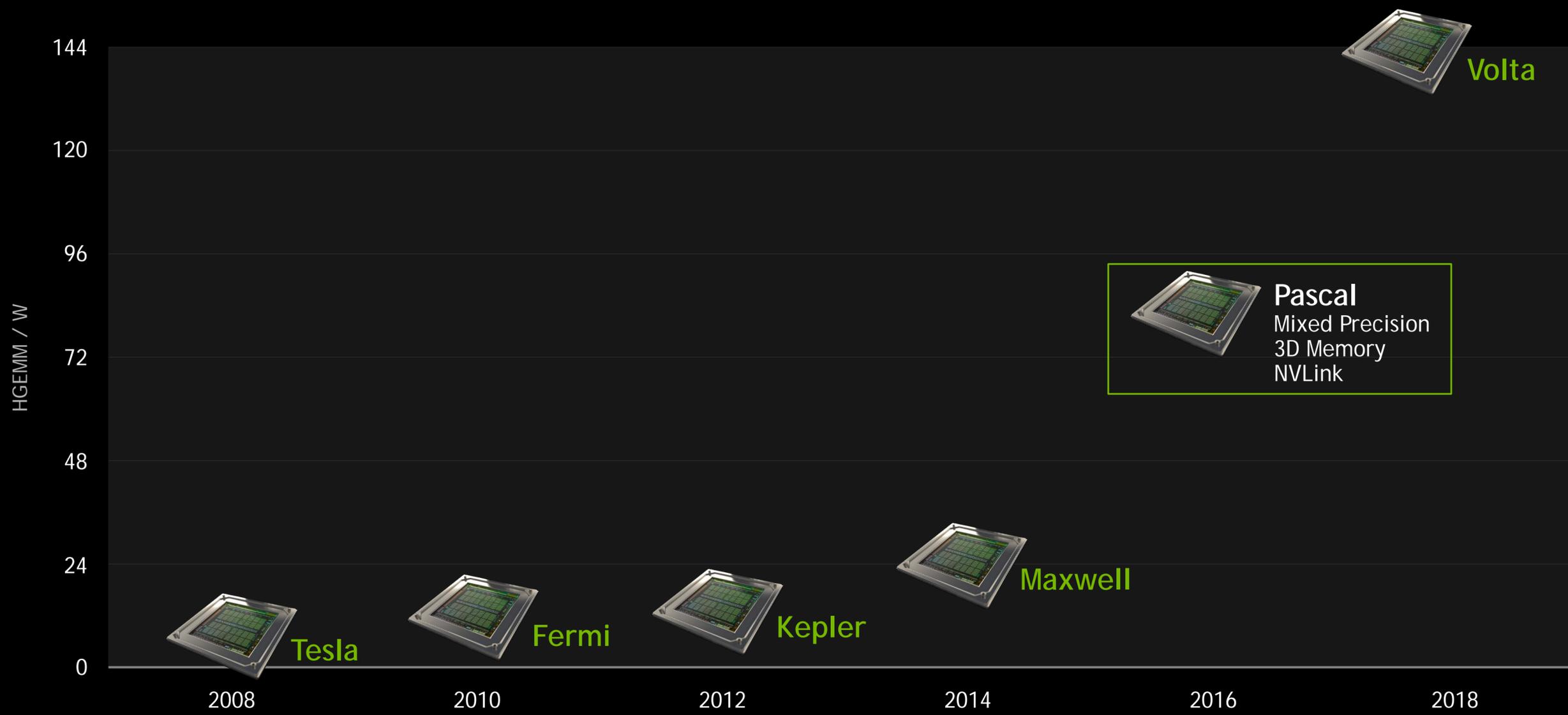
# GPU ROADMAP

Pascal 2.7x Memory Capacity



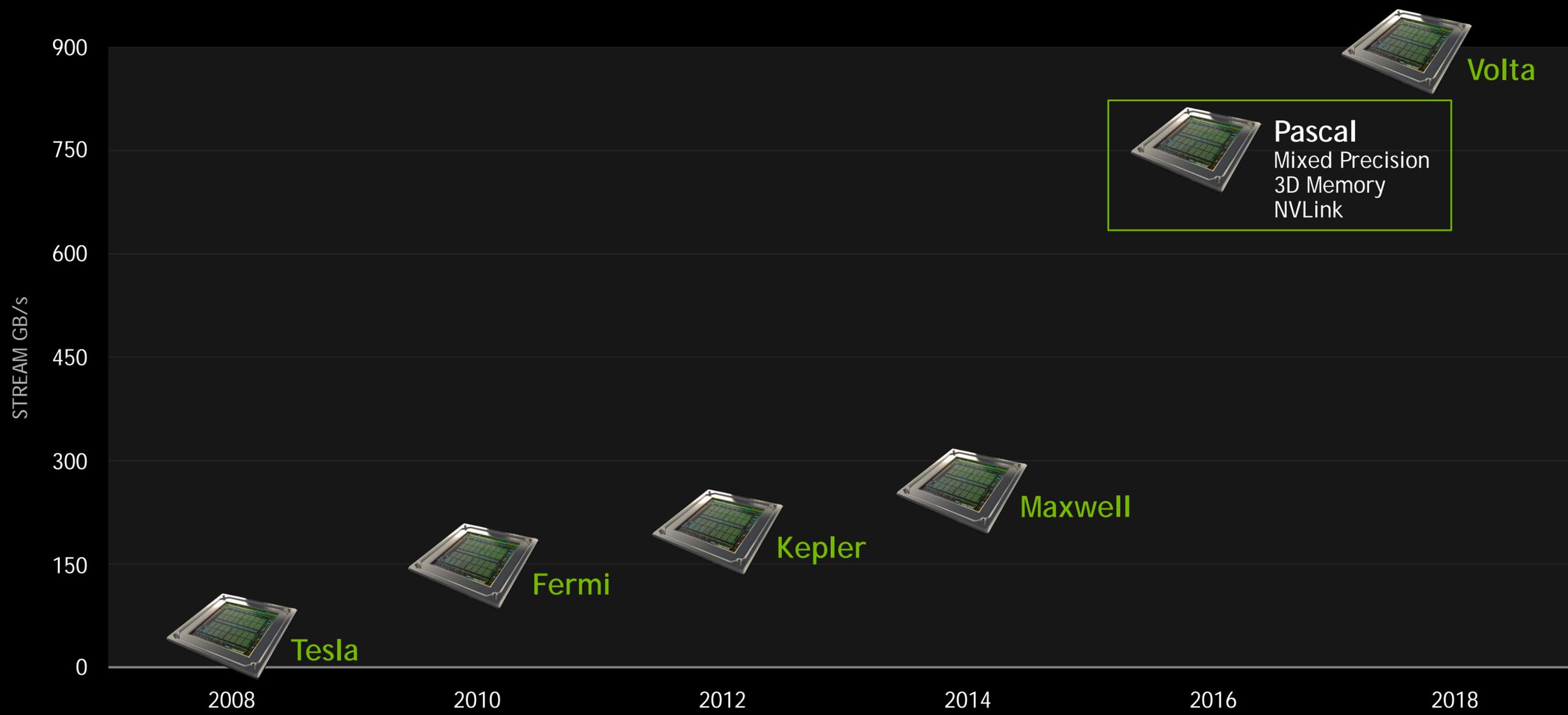
# GPU ROADMAP

Pascal 4x Mixed Precision



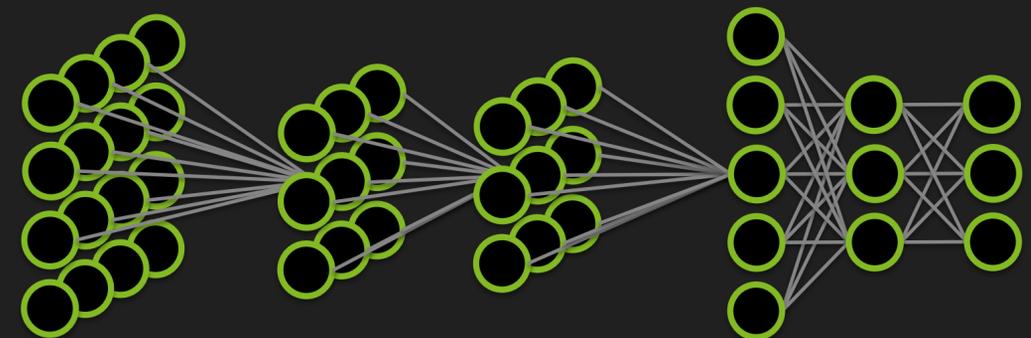
# GPU ROADMAP

Pascal 3x Bandwidth



# PASCAL 10X MAXWELL

forward →



CONVOLUTION  
(compute)

4x (FP16)

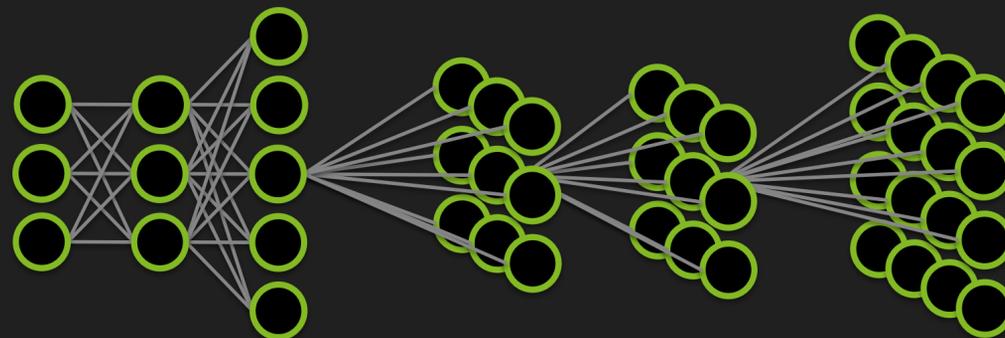
Mixed Precision

FULLY CONNECTED  
(bandwidth)

6x

3D Memory

backward →



FULLY CONNECTED  
(bandwidth)

6x

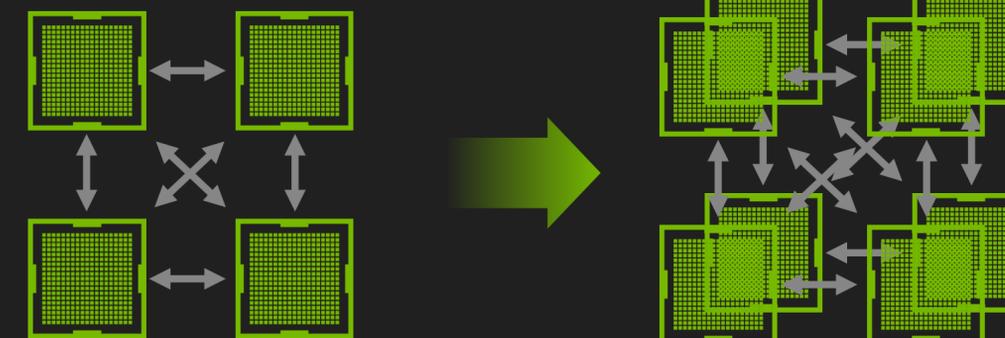
3D Memory

CONVOLUTION  
(compute)

4x

Mixed Precision

5x



WEIGHT UPDATE  
(interconnect)

10x

NVLINK

2x

\* Very rough estimates

# FOUR ANNOUNCEMENTS

A New GPU  
and  
Deep Learning

A Very Fast Box  
and  
Deep Learning

Roadmap Reveal  
and  
Deep Learning

Self-Driving Cars  
and  
Deep Learning

# TODAY'S ADAS

SENSE

FPGA  
CV ASIC

PLAN

CPU

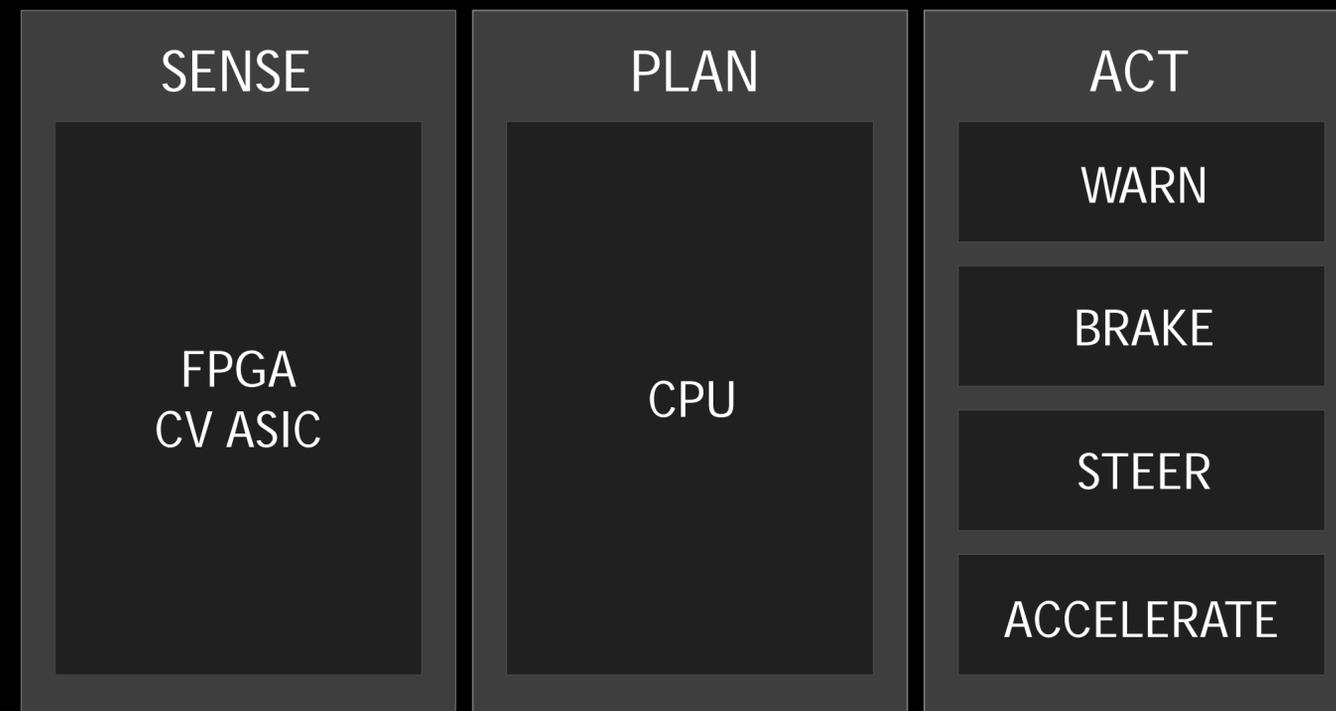
ACT

WARN

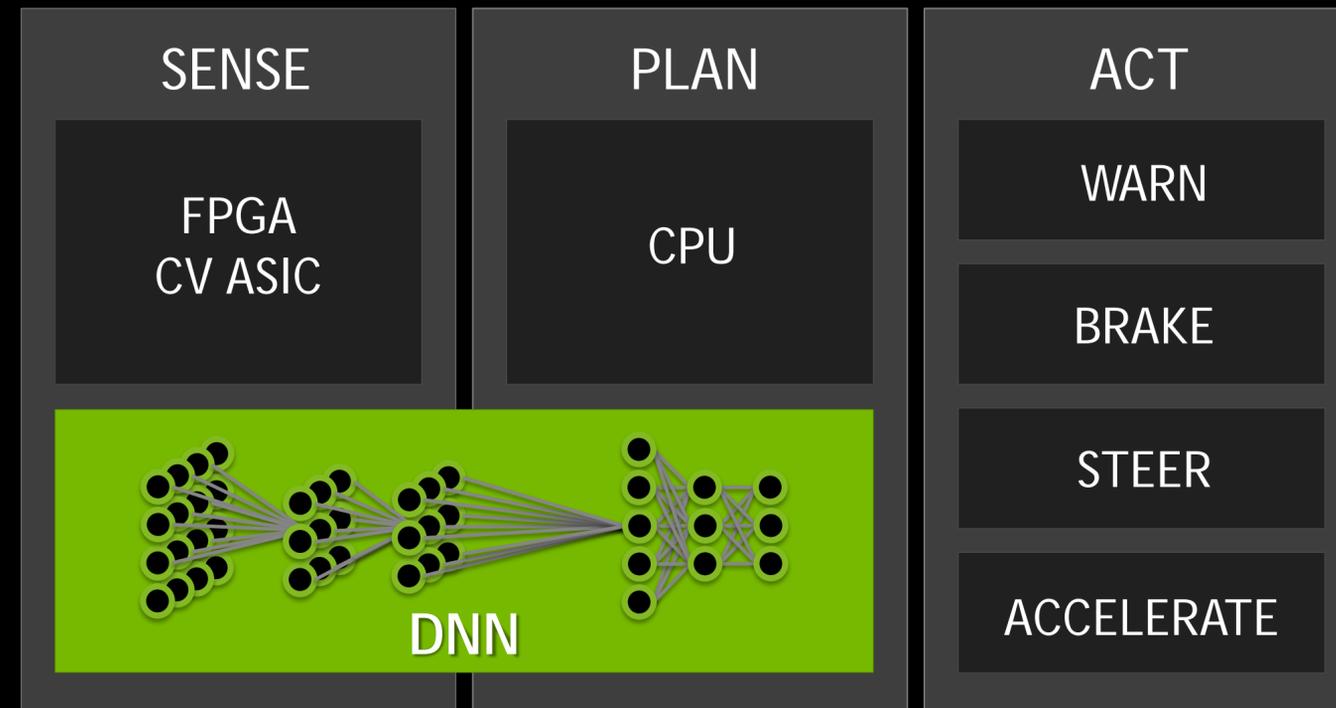
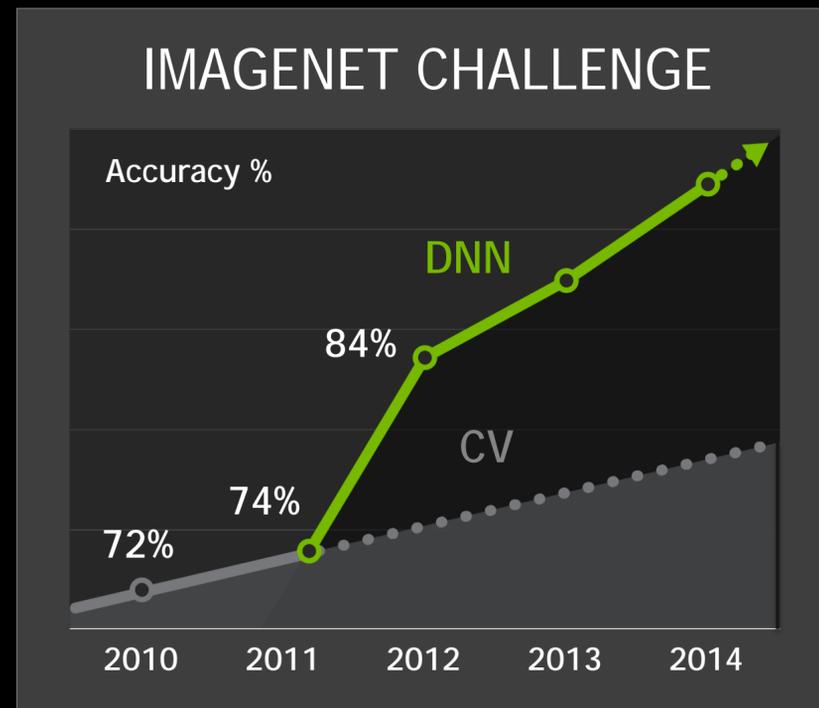
BRAKE



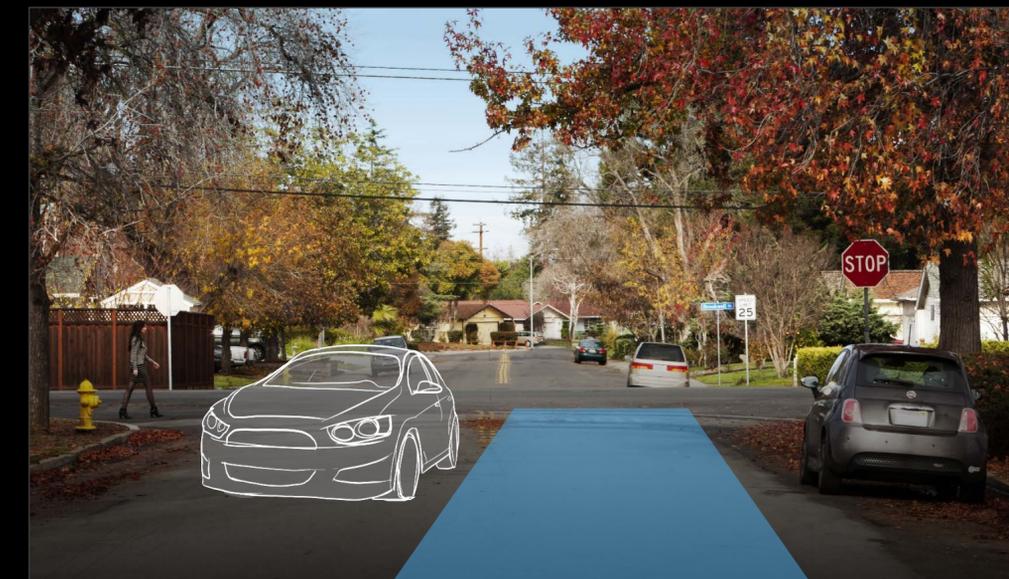
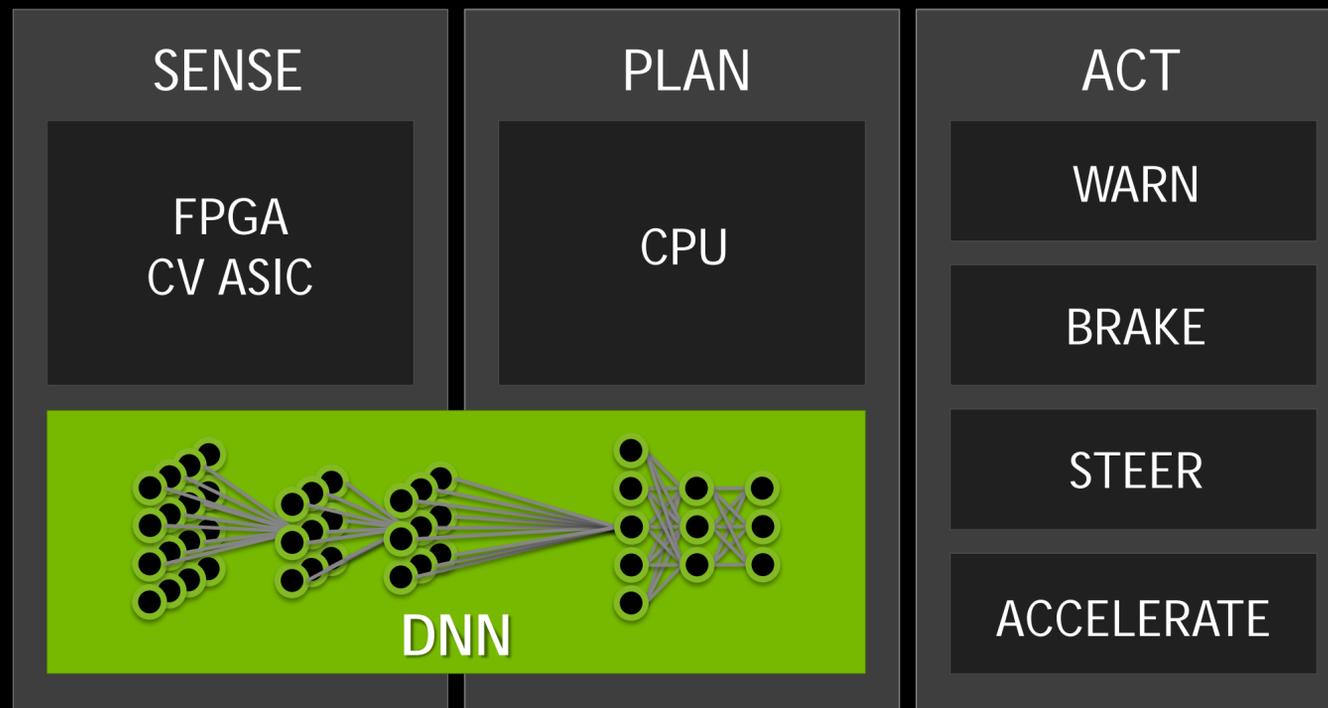
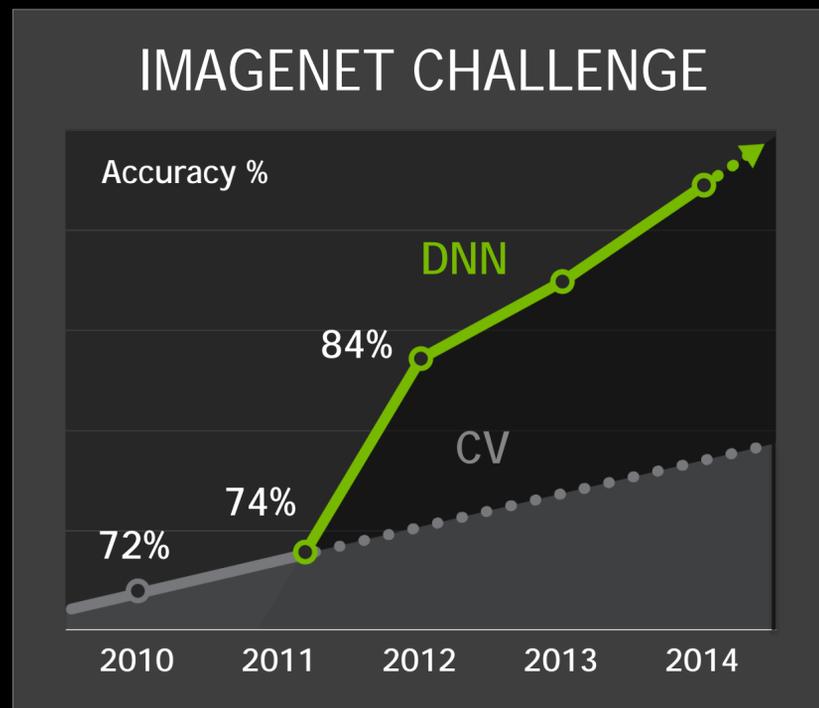
# NEXT-GENERATION ADAS



# NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER

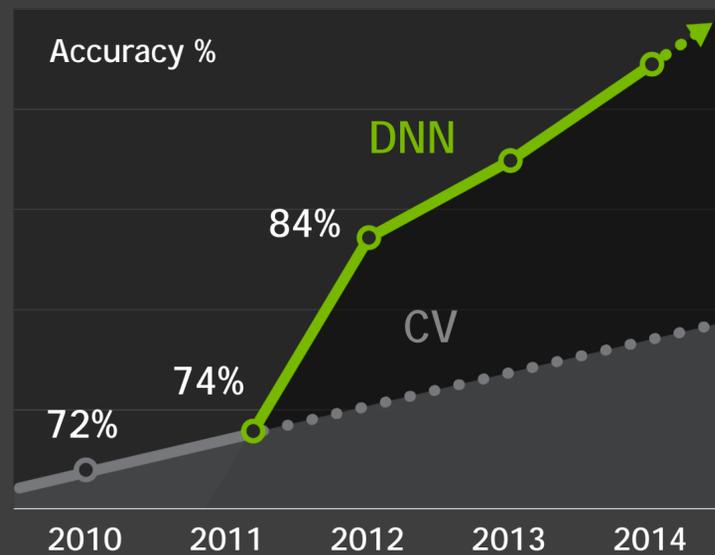


# NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER



# NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER

## IMAGENET CHALLENGE



## SENSE

FPGA  
CV ASIC

## PLAN

CPU

## ACT

WARN

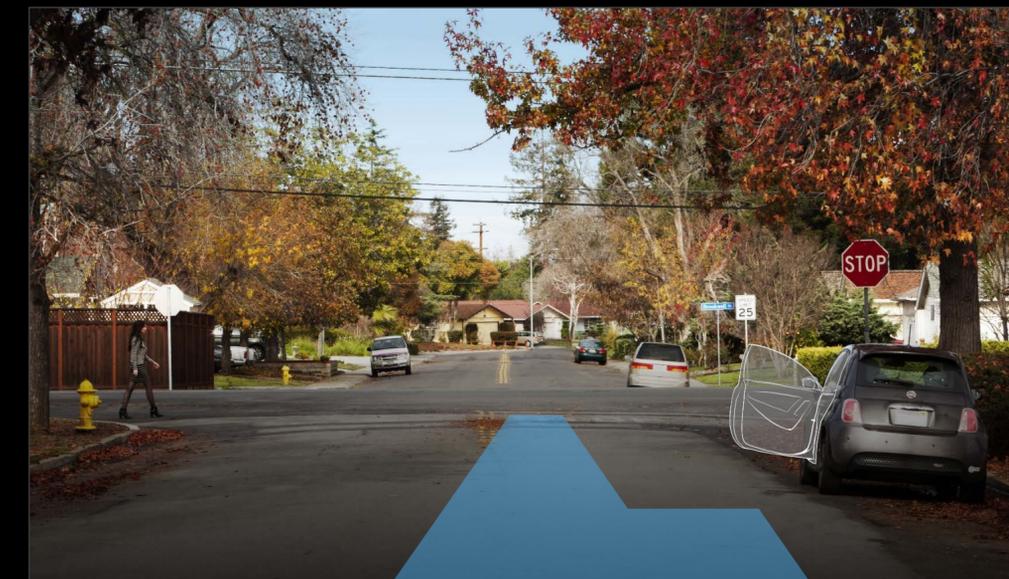
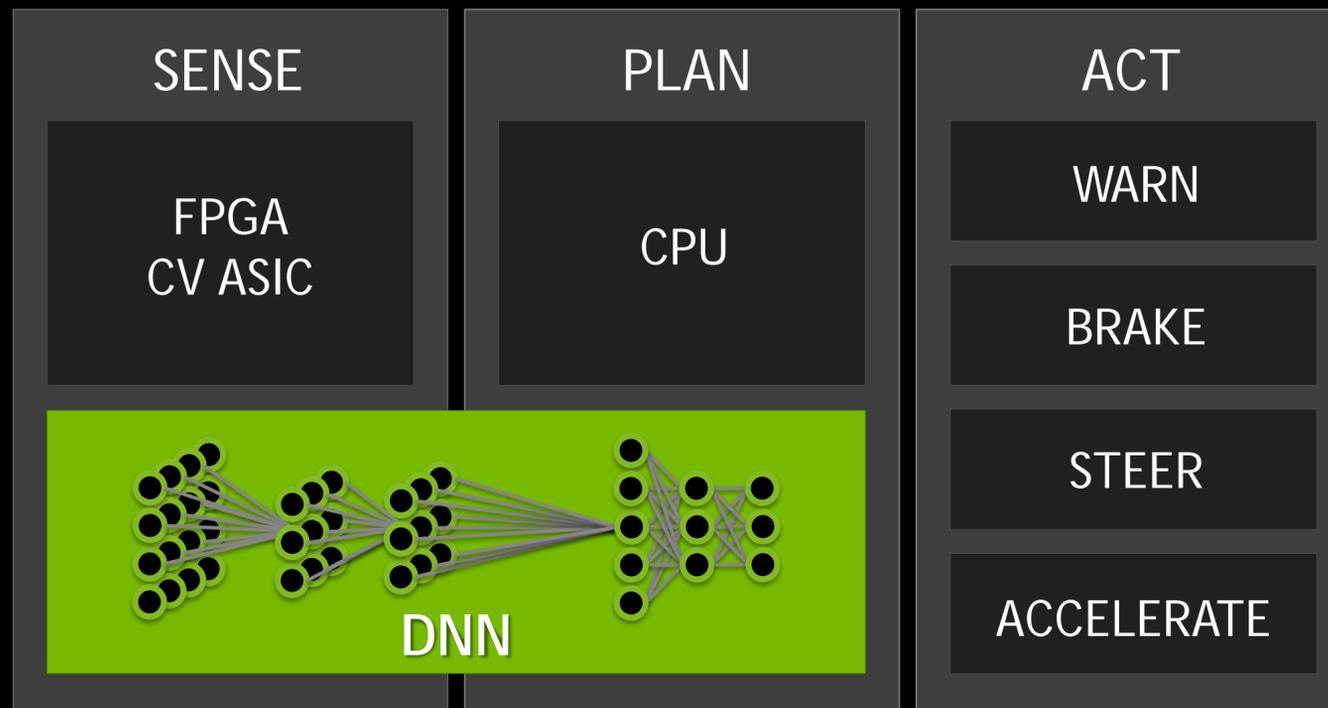
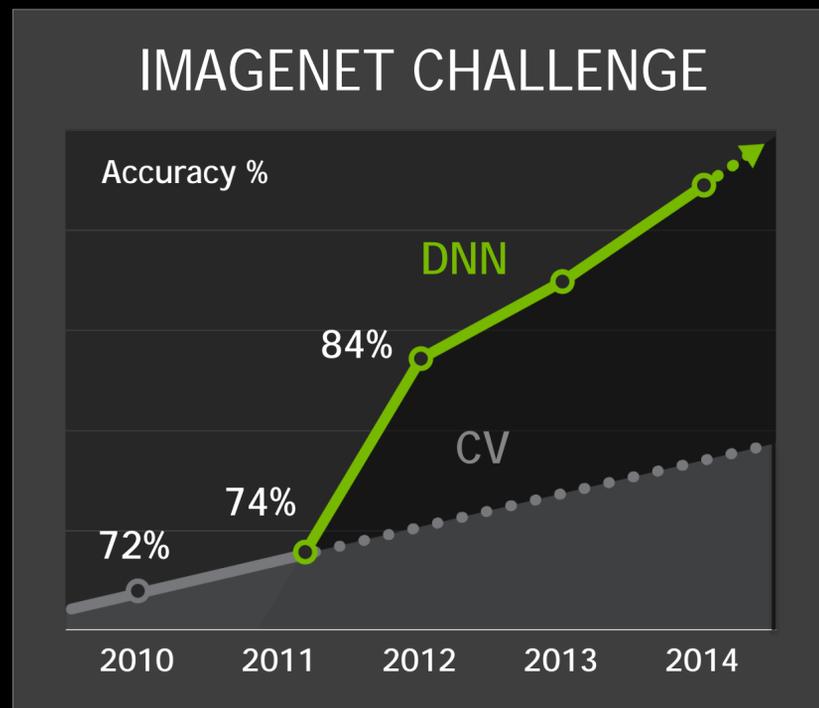
BRAKE

STEER

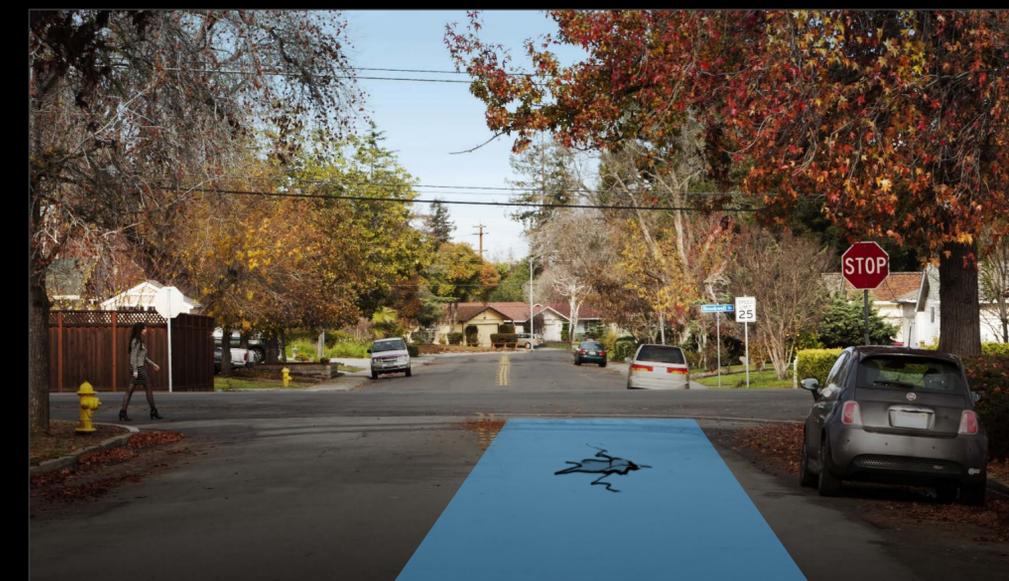
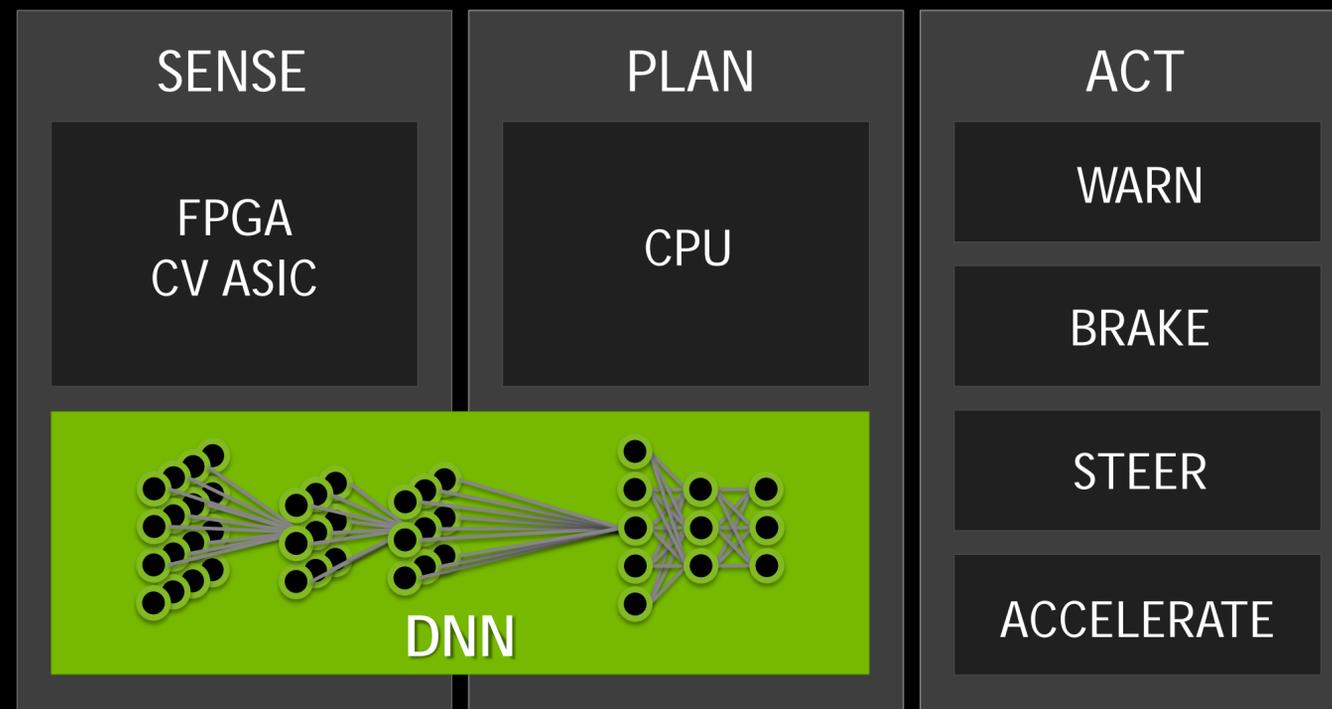
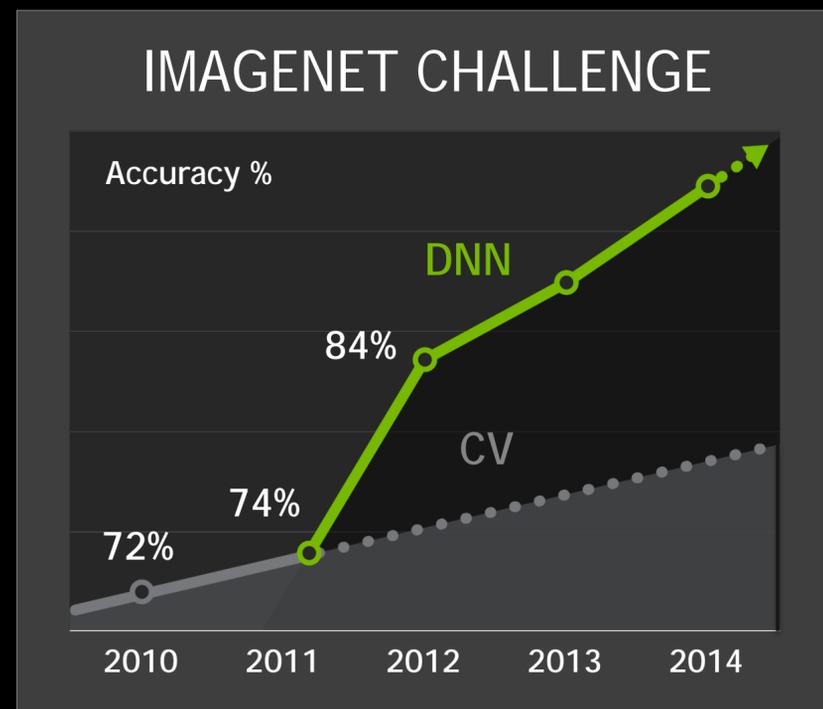
ACCELERATE



# NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER

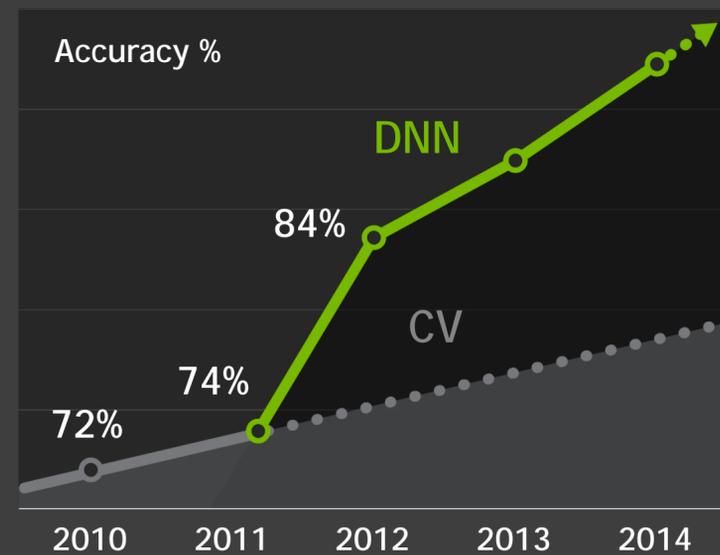


# NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER



# NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER

## IMAGENET CHALLENGE



## SENSE

FPGA  
CV ASIC

## PLAN

CPU

## ACT

WARN

BRAKE

STEER

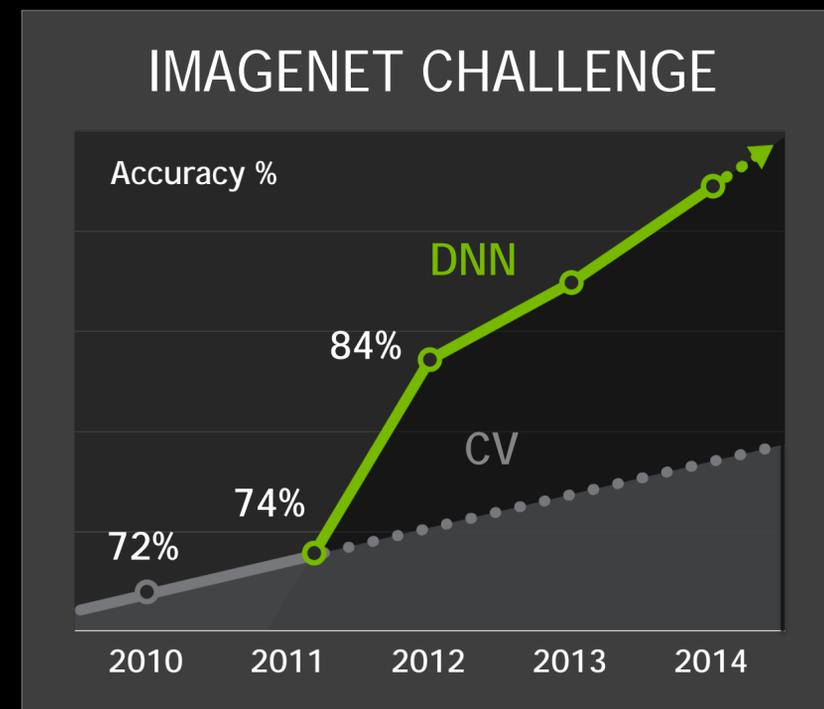
ACCELERATE



DNN



# PROJECT DAVE — DARPA AUTONOMOUS VEHICLE



DNN-based self-driving robot

Training data by human driver

No hand-coded CV algorithms

## PROJECT LEADS

Urs Muller: Chief Architect,  
Autonomous Driving, NVIDIA

Yann LeCun: Director,  
AI Research, Facebook

DAVE IN ACTION



# TRAINING DATA

225K Images



# TEST DRIVE

No Training



# TEST DRIVE

Partially Trained (52K images)



# TEST DRIVE

Fully Trained (225K images)





Number of Connections  
Frames / Second  
Connections / Second

### DAVE

3.1 Million

12

38 Million

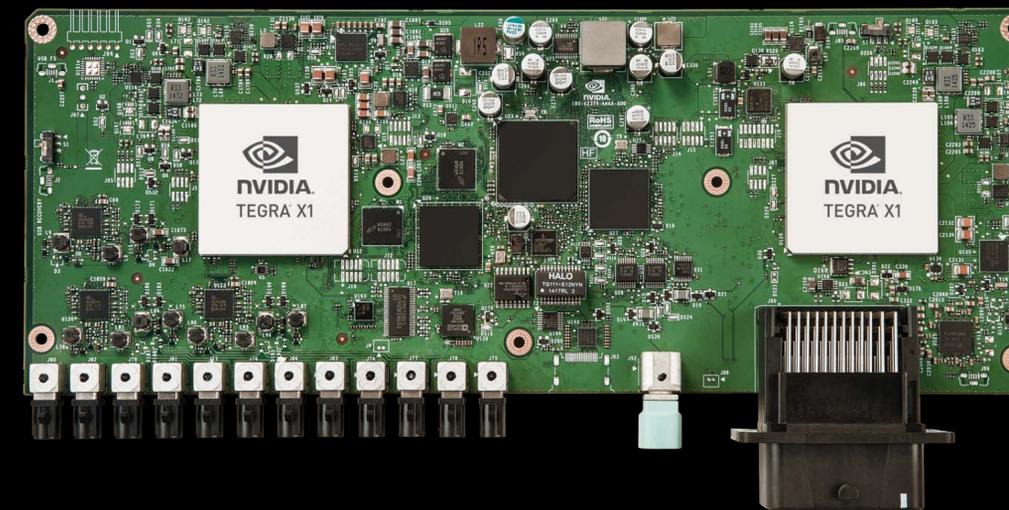
### AlexNet on DRIVE PX

630 Million

184

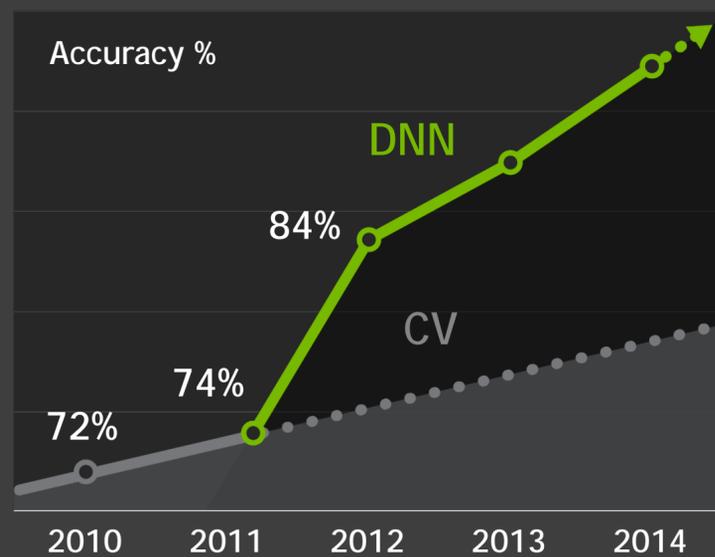
116 Billion

3,000x  
Faster



# NVIDIA DRIVE PX SELF-DRIVING CAR COMPUTER

## IMAGENET CHALLENGE



## SENSE

FPGA  
CV ASIC

## PLAN

CPU

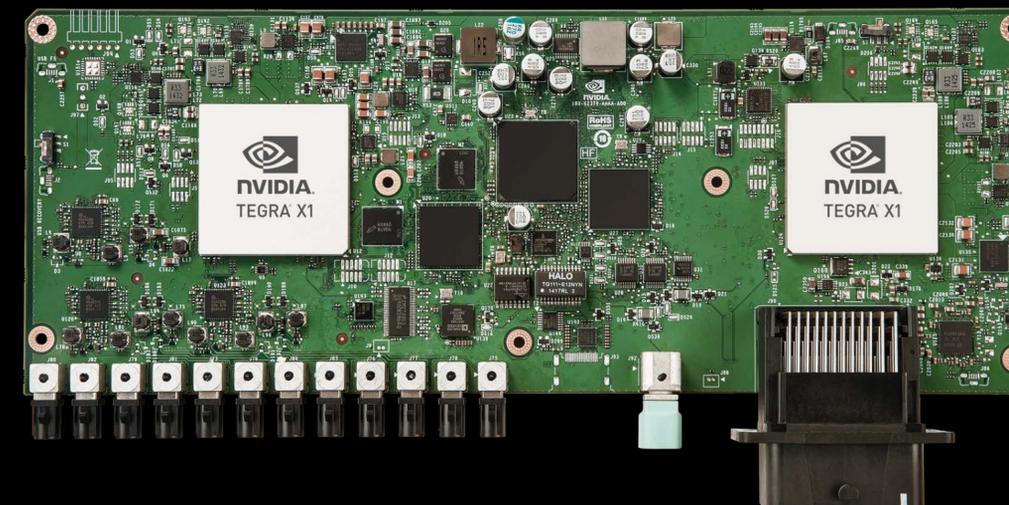
## ACT

WARN

BRAKE

STEER

ACCELERATE





Search NVIDIA

USA

DRIVERS ▶ PRODUCTS ▶ COMMUNITIES ▶ SUPPORT SHOP ABOUT NVIDIA ▶

AUTOMOTIVE

DRIVE PX

DRIVE CX

VISUAL COMPUTING  
MODULE

AUTOMOTIVE  
PARTNERS

NVIDIA Home > Products > Automotive > NVIDIA Drive PX

Subscribe

## NVIDIA DRIVE™ PX

Self-Driving Car Computer

### WELCOME TO THE FUTURE OF AUTOMOTIVE INNOVATION

#### SELF-DRIVING COMPUTING DEVELOPMENT PLATFORM

NVIDIA DRIVE™ PX is a powerful auto-pilot car computer designed to run the deep neural networks that will enable a car to see, think, and learn.

DRIVE PX features dual **NVIDIA Tegra™ X1** processors and delivers 2.3 teraflops of performance. Twelve camera inputs enable a wide range of ADAS features to run simultaneously, including surround view, collision avoidance, pedestrian detection, mirror-less operation, cross traffic monitoring, and driver state monitoring.

DRIVE PX is available to automakers, tier 1 suppliers, and research institutions working on developing systems that enable cars to drive themselves.

ORDER NOW



# NVIDIA DRIVE™ PX

## SELF-DRIVING CAR COMPUTER

Available May 2015  
\$10,000

TESLA

ELON MUSK

# LEAPS IN VISUAL COMPUTING

## TITAN X

The World's Fastest GPU



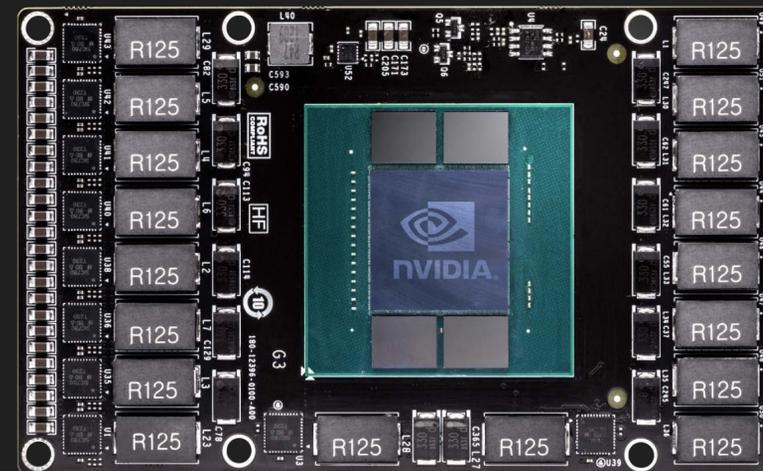
## DIGITS DevBox

GPU Deep Learning Platform



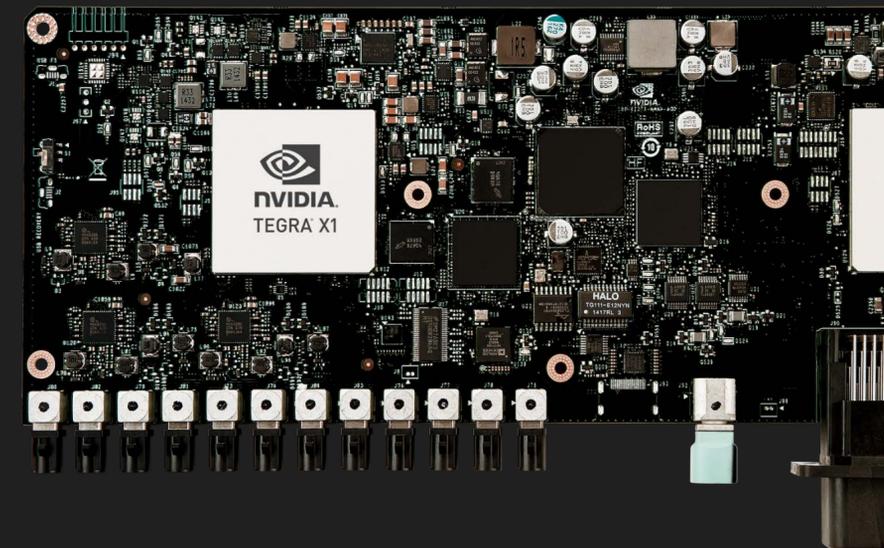
## Pascal – 10x Maxwell

For Deep Learning



## NVIDIA DRIVE PX

Deep Learning Platform for Self-Driving Cars



The logo consists of a green parallelogram with a white border. Inside the parallelogram, the text "GPU" is written in a large, bold, white sans-serif font. To the right of "GPU", the words "TECHNOLOGY" and "CONFERENCE" are stacked vertically in a smaller, white sans-serif font. A small teal triangle is located at the bottom-left corner of the parallelogram.

**GPU** TECHNOLOGY  
CONFERENCE

