





NAVEEN RAO

**CORPORATE VICE PRESIDENT AND GENERAL MANAGER
INTEL AI**

WELCOME TO
INTEL AI DEVCON

COMPUTE AND THE BRAIN

**OPERANT
CONDITIONING**
1938

**SPIKING
NEURON**
1952

**FIRST
NEUROSCIENCE
DEPARTMENT**
1964

**DEEP
LEARNING
PREVALENCE**
mid 2000s

**THE
TURING
MACHINE**
1936

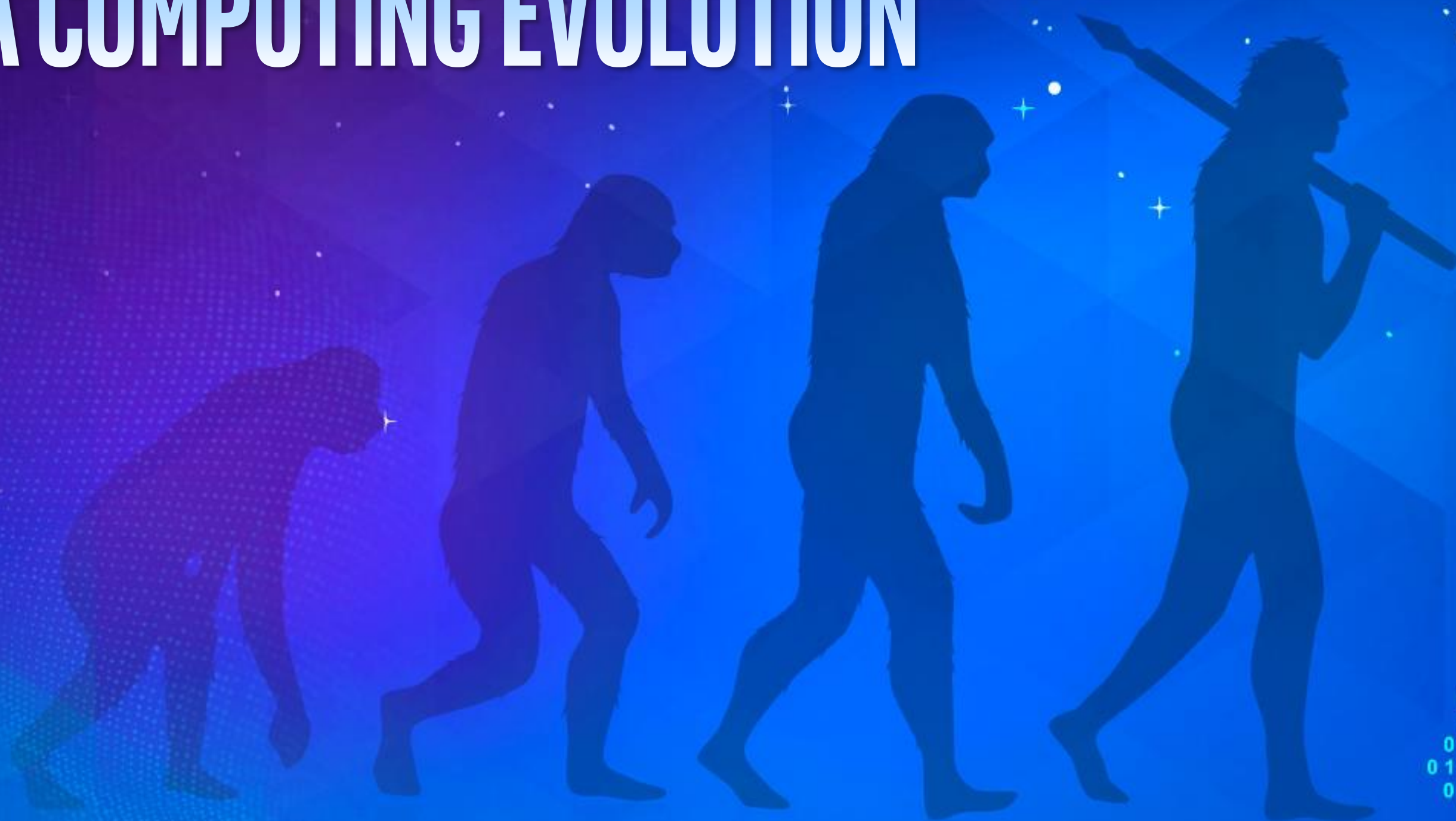
TRANSISTOR
1947

**FIRST
COMPUTER
SCIENCE
DEPARTMENT**
1962

**INTEL
FOUNDED**
1968

**FIRST 1
BILLION
TRANSISTOR
PROCESSOR**
mid 2000s

THE AI REVOLUTION IS REALLY A COMPUTING EVOLUTION



OUR VISION OF AI

A photograph of two women in a dimly lit office environment, focused on their work. The woman in the foreground, with dark curly hair and glasses, is wearing a pink sweater and typing on a laptop. The woman behind her, with reddish hair, is also looking at the laptop screen. In the background, a large monitor displays lines of code. The scene is overlaid with a blue geometric pattern on the right side.

BEST TOOLS TO BUILD TOWARDS
THE BEST POSSIBLE FUTURE

ENABLE DEVELOPERS TO ACHIEVE THEIR AI VISION



A vertical panel with a blue and red geometric background. It features a snippet of Python code for a Blender Mirror modifier. The code includes comments and property settings for the mirror object's use_x, use_y, and use_z properties, along with a 'please select exact' message and a list of operator classes.

TOOLS



A vertical panel with a blue and purple background. It features a 3D rendering of a circuit board with various components like capacitors, resistors, and integrated circuits. The board is shown from an isometric perspective.

HARDWARE



A vertical panel with a blue and white background. It features a photograph of several people in a workshop or office setting, working on laptops. The image is slightly blurred and has a blue tint.

COMMUNITY

ENABLE DEVELOPERS TO ACHIEVE THEIR AI VISION



TOOLS

```
... = modifier_ob...
... mirror object to mirror...
... mirror_mod.mirror_object...
... operation == "MIRROR_X":...
... mirror_mod.use_x = True...
... mirror_mod.use_y = False...
... mirror_mod.use_z = False...
... operation == "MIRROR_Y":...
... mirror_mod.use_x = False...
... mirror_mod.use_y = True...
... mirror_mod.use_z = False...
... operation == "MIRROR_Z":...
... mirror_mod.use_x = False...
... mirror_mod.use_y = False...
... mirror_mod.use_z = True...
... bpy.context.selected_ob...
... data.objects[one.name].sel...
... print("please select exact...
... --- OPERATOR CLASSES ---...
... types.Operator):...
... X mirror to the selected ob...
... ject.mirror_mirror_x"...
... mirror X"...
... context):...
... active_object is not...
```



HARDWARE



COMMUNITY



JASON KNIGHT
HEAD OF SOFTWARE PRODUCTS
INTEL AI

OPEN SOURCE

INTEL[®] MATH KERNEL LIBRARY FOR DEEP NEURAL NETWORKS (INTEL[®] MKL-DNN)

HELPS REALIZE THE INCREDIBLE BENEFITS OF DIRECT OPTIMIZATION

MATRIX MULTIPLICATION

BATCH NORM

POOLING

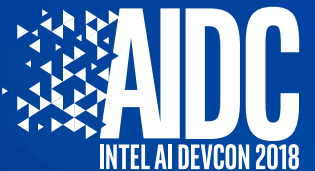
NORMALIZATION

ACTIVATION

CONVOLUTION



Intel/mkl-dnn



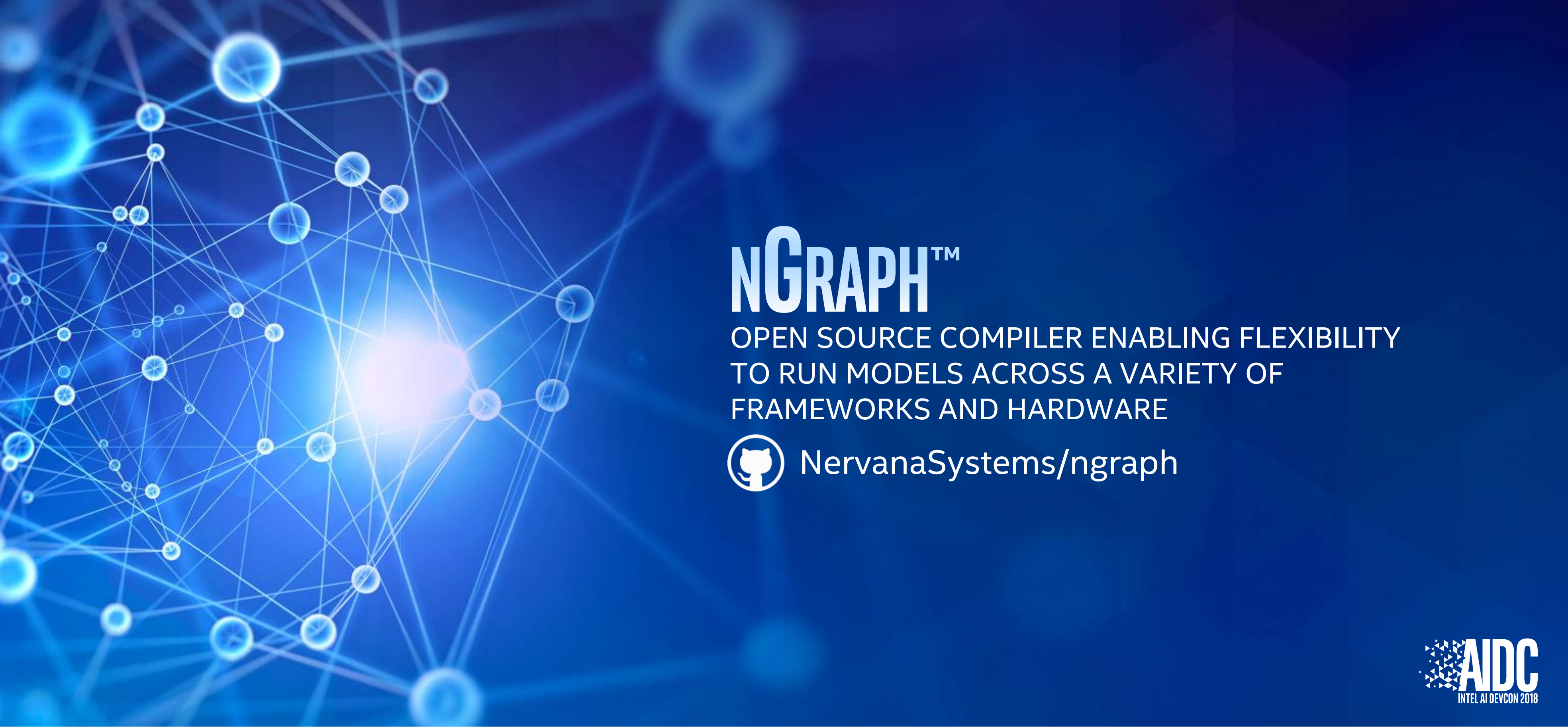


Google Open Source

OPTIMIZING TENSORFLOW TO SUPERCHARGE AI WORKLOADS

RAJAT MONGA
TENSORFLOW ENGINEERING DIRECTOR
GOOGLE



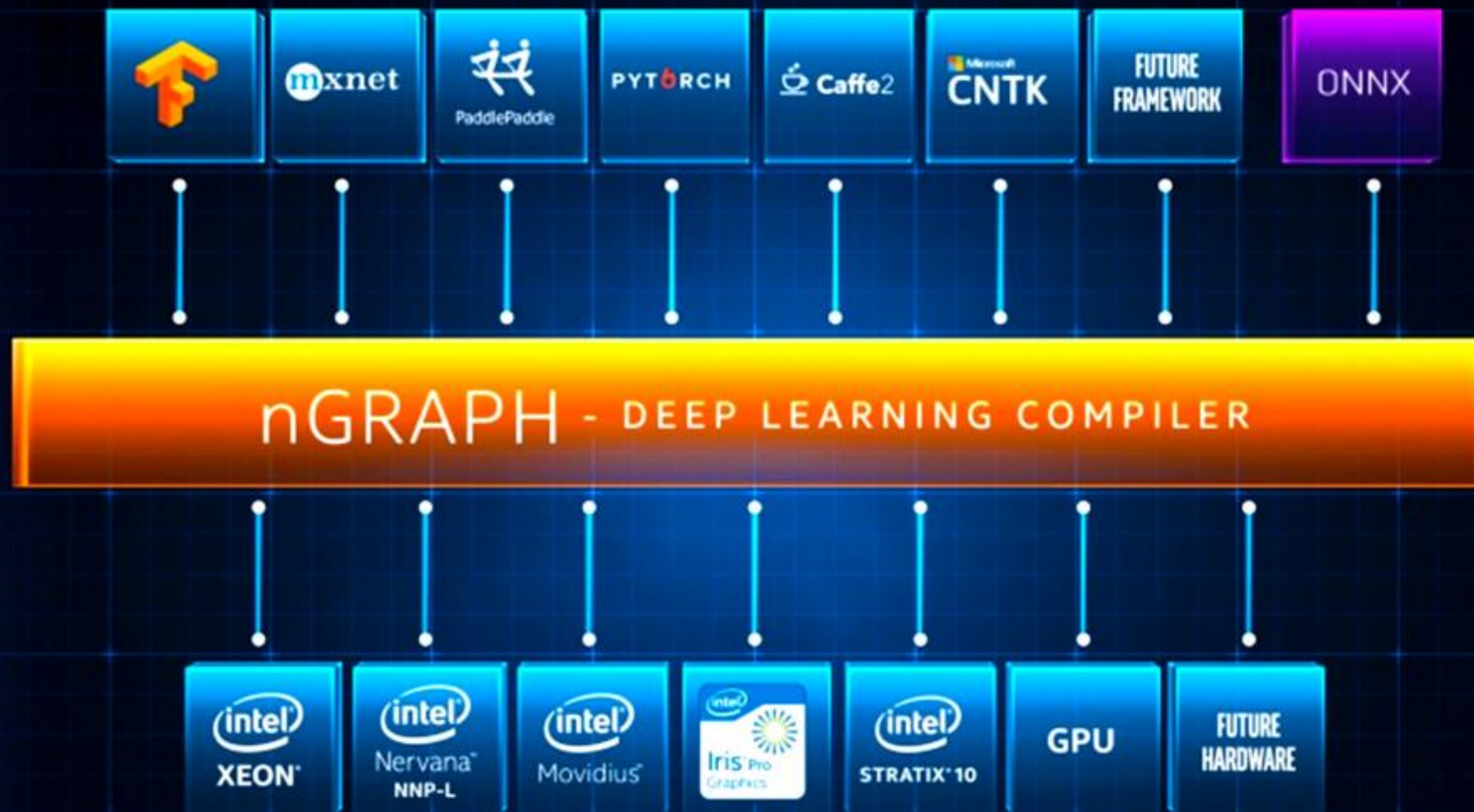
The background of the slide is a dark blue gradient with a complex, glowing network graph. The graph consists of numerous circular nodes of varying sizes, some of which are brightly lit with a white or yellow glow, and they are interconnected by a dense web of thin, light blue lines. The overall effect is one of a dynamic, interconnected digital space.

NGRAPH™

OPEN SOURCE COMPILER ENABLING FLEXIBILITY
TO RUN MODELS ACROSS A VARIETY OF
FRAMEWORKS AND HARDWARE



[NervanaSystems/ngraph](https://github.com/NervanaSystems/ngraph)

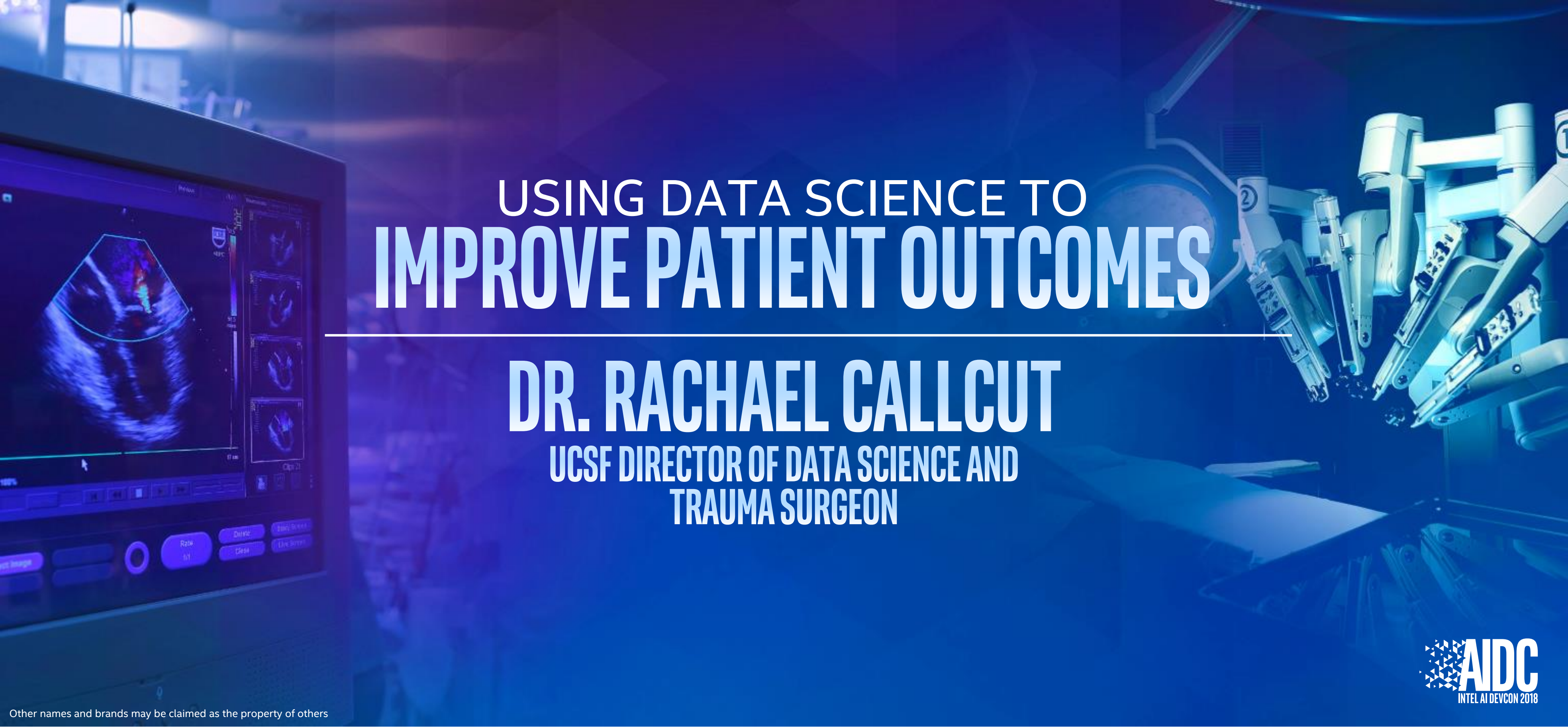


BIGDL: DISTRIBUTED DEEP LEARNING

ENABLING DEEP LEARNING TO TAKE ADVANTAGE
OF SCALABLE SPARK AND HADOOP CLUSTERS



intel-analytics/BigDL



USING DATA SCIENCE TO IMPROVE PATIENT OUTCOMES

DR. RACHAEL CALLCUT
UCSF DIRECTOR OF DATA SCIENCE AND
TRAUMA SURGEON



USING DATA SCIENCE TO IMPROVE PATIENT OUTCOMES

VIDEO IS THE ULTIMATE IOT SENSOR



EMERGENCY RESPONSE



FINANCIAL SERVICES



MACHINE VISION



CITIES/TRANSPORTATION



AUTONOMOUS VEHICLES



RESPONSIVE RETAIL



MANUFACTURING



PUBLIC SECTOR

ANNOUNCING: OPENVINO™ SOFTWARE TOOLKIT

VISUAL INFERENCE AND NEURAL NETWORK OPTIMIZATION

nGraph™ mxnet

Caffe TensorFlow

Model Optimizer,
Inference Engine,
OpenCV & OpenCL

CPU

GPU

FPGA

VPU ¥

DEPLOY COMPUTER
VISION AND DEEP
LEARNING CAPABILITIES
TO THE EDGE



WRITE ONCE + SCALE TO DIVERSE ACCELERATORS



BROAD FRAMEWORK SUPPORT



HIGH PERFORMANCE, HIGH EFFICIENCY FOR THE EDGE

ENABLE DEVELOPERS TO ACHIEVE THEIR AI VISION



TOOLS

```
... = modifier_ob...
... mirror object to mirror...
... mirror_mod.mirror_object...
... operation == "MIRROR_X":...
... mirror_mod.use_x = True...
... mirror_mod.use_y = False...
... mirror_mod.use_z = False...
... operation == "MIRROR_Y":...
... mirror_mod.use_x = False...
... mirror_mod.use_y = True...
... mirror_mod.use_z = False...
... operation == "MIRROR_Z":...
... mirror_mod.use_x = False...
... mirror_mod.use_y = False...
... mirror_mod.use_z = True...
... bpy.context.selected_ob...
... data.objects[one.name].sel...
... print("please select exact...
... --- OPERATOR CLASSES ---...
... types.Operator):...
... X mirror to the selected ob...
... ject.mirror_mirror_x"...
... mirror X"...
... context):...
... active_object is not...
```



HARDWARE



COMMUNITY

ENABLE DEVELOPERS TO ACHIEVE THEIR AI VISION



A snippet of Python code for Blender's mirror modifier, showing how to set up mirror operations for X, Y, and Z axes. The code includes comments and property assignments for the mirror_mod object.

```
... = modifier_ob.  
... mirror object to mirror  
... mirror_mod.mirror_object =  
... operation == "MIRROR_X":  
... mirror_mod.use_x = True  
... mirror_mod.use_y = False  
... mirror_mod.use_z = False  
... operation == "MIRROR_Y":  
... mirror_mod.use_x = False  
... mirror_mod.use_y = True  
... mirror_mod.use_z = False  
... operation == "MIRROR_Z":  
... mirror_mod.use_x = False  
... mirror_mod.use_y = False  
... mirror_mod.use_z = True
```

selection at the end
ob.select= 1
for ob.select=1
context.scene.objects.active
"selected" = str(modifier
selected_ob.select= 0
bpy.context.selected_obj
data.objects[one.name].sel
print("please select exact
--- OPERATOR CLASSES ---
...
... types.Operator):
... ask X mirror to the selected
... ject: mirror_mirror(X)
... mirror X"

TOOLS



A close-up photograph of a complex electronic circuit board, likely a GPU or AI accelerator, featuring various components like capacitors, resistors, and integrated circuits.

HARDWARE



A photograph showing a group of people, likely developers or researchers, working together at a desk with multiple laptops, illustrating a collaborative community environment.

COMMUNITY



FOUNDATIONAL FOR ARTIFICIAL INTELLIGENCE



Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective

Kim Hazelwood, Sarah Bird, David Brooks, Soumith Chintala, Utkar Dutt, Dmitry Dzhalozov,
Mohamed Fawzy, Bill Jia, Yanguang Jia, Aditya Kales, James Law, Kevin Lee, Jason Lu,
Pieter Noordhuis, Misha Smelyanskiy, Liang Xiong, Xiaodong Wang
Facebook, Inc.

Abstract—Machine learning sits at the core of many essential products and services at Facebook. This paper describes the hardware and software infrastructure that supports machine learning at global scale. Facebook's machine learning workloads are extremely diverse: services require many different types of models in practice. This diversity has implications at all layers in the system stack. In addition, a sizable fraction of all data stored at Facebook flows through machine learning pipelines, presenting significant challenges in delivering data to high-performance distributed training flows. Computational requirements are also intense, leveraging both GPU and CPU platforms for training and abundant CPU capacity for real-time inference. Addressing these and other emerging challenges continues to require diverse efforts that span machine learning algorithms, software, and hardware design.

1. INTRODUCTION

Facebook's mission is to "Give people the power to build community and bring the world closer together." In support of that mission, Facebook connects more than two billion people as of December 2017. Meanwhile, the past several years have seen a revolution in the application of machine learning to real problems at this scale, building upon the virtuous cycle of machine learning algorithmic innovations, enormous amounts of training data for models, and advances in high-performance computer architectures [1]. At Facebook, machine learning provides key capabilities in driving nearly all aspects of user experience including services like ranking posts for News Feed, speech and text translations, and photo and real-time video classification [2], [3].

Facebook leverages a wide variety of machine learning algorithms in these services including support vector machines, gradient boosted decision trees, and many styles of neural networks. This paper describes several important aspects of datacenter infrastructure that supports machine learning at Facebook. The infrastructure includes internal "ML-as-a-Service" flows, open-source machine learning frameworks, and distributed training algorithms. From a hardware point of view, Facebook leverages a large fleet of CPU and GPU platforms for training models in order to support the necessary training frequencies at the required service latency. For machine learning inference, Facebook primarily relies on CPUs for all major services with neural network ranking services

Facebook funnels a large fraction of all stored data through machine learning pipelines, and this fraction is increasing over time to improve model quality. The massive amount of data required by machine learning services presents challenges at the global scale of Facebook's datacenters. Several techniques are used to efficiently feed data to the models including decoupling of data feed and training, data/compute co-location, and networking optimizations. At the same time, Facebook's scale provides unique opportunities. Diurnal load cycles leave a significant number of CPUs available for distributed training algorithms during off-peak periods. With Facebook's compute fleet spread over ten datacenter locations, scale also provides disaster recovery capability. Disaster recovery planning is essential as timely delivery of new machine learning models is important to Facebook's operations.

Looking forward, Facebook expects rapid growth in machine learning across existing and new services [4]. This growth will lead to growing scalability challenges for teams deploying the infrastructure for these services. While significant opportunities exist to optimize infrastructure on existing platforms, we continue to actively evaluate and prototype new hardware solutions while remaining cognizant of game-changing algorithmic innovations.

The key contributions of this paper include the following major insights about machine learning at Facebook.

- Machine learning is applied pervasively across nearly all services, and **computer vision represents only a small fraction** of the resource requirements.
- Facebook relies upon an **incredibly diverse set of machine learning approaches** including, but not limited to, neural networks.
- **Tremendous amounts of data** are funneled through our machine learning pipelines, and this creates engineering and efficiency challenges far beyond the compute nodes.
- Facebook currently relies heavily on **CPUs for inference**, and **both CPUs and GPUs for training**, but constantly prototypes and evaluates new hardware solutions from a performance-per-watt perspective.
- The worldwide scale of people on Facebook and corresponding diurnal activity patterns result in a huge number of machines that can be harnessed for machine learning tasks such as **distributed training at scale**.

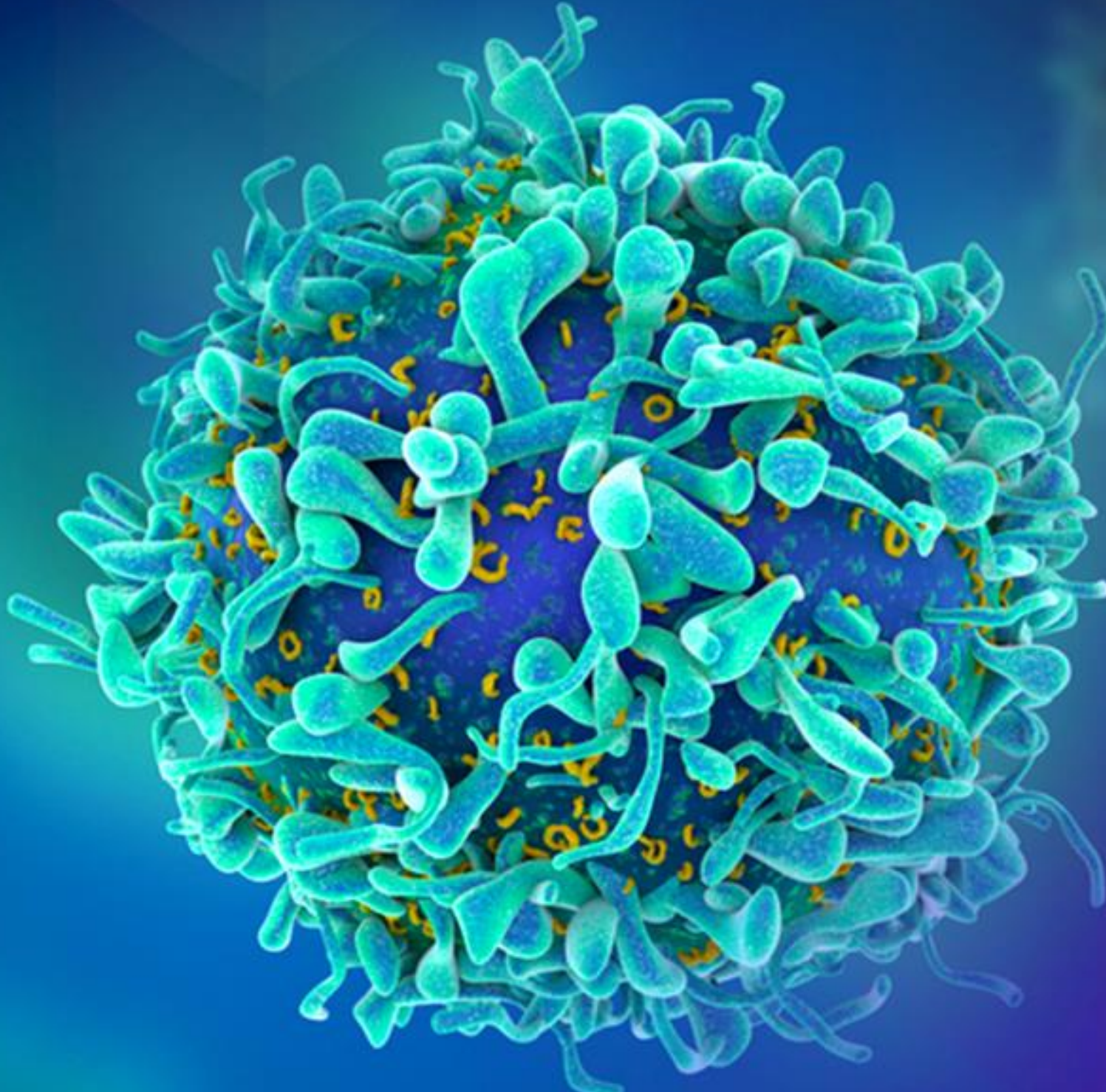
SO LET'S GET GROUNDED IN REALITY

facebook®

KIM HAZELWOOD
HEAD OF AI INFRASTRUCTURE FOUNDATION,
FACEBOOK

Services	Relative Capacity	Compute	Memory
News Feed	100X	Dual-Socket CPU	High
Facer	10X	Single-Socket CPU	Low
Lumos	10X	Single-Socket CPU	Low
Search	10X	Dual-Socket CPU	High
Language Translation	1X	Dual-Socket CPU	High
Sigma	1X	Dual-Socket CPU	High
Speech Recognition	1X	Dual-Socket CPU	High

TABLE III
RESOURCE REQUIREMENTS OF ONLINE INFERENCE WORKLOADS.

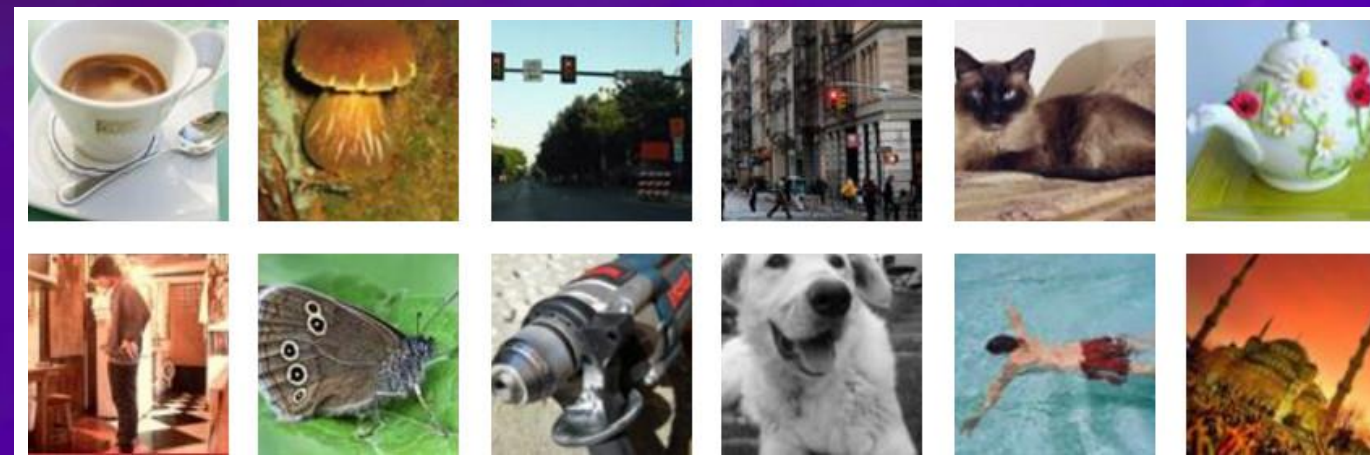


A SCIENTIFIC COLLABORATION BETWEEN INTEL AND NOVARTIS

KUSHAL DATTA, PHD

**RESEARCH SCIENTIST
INTEL AI**

IMAGENET



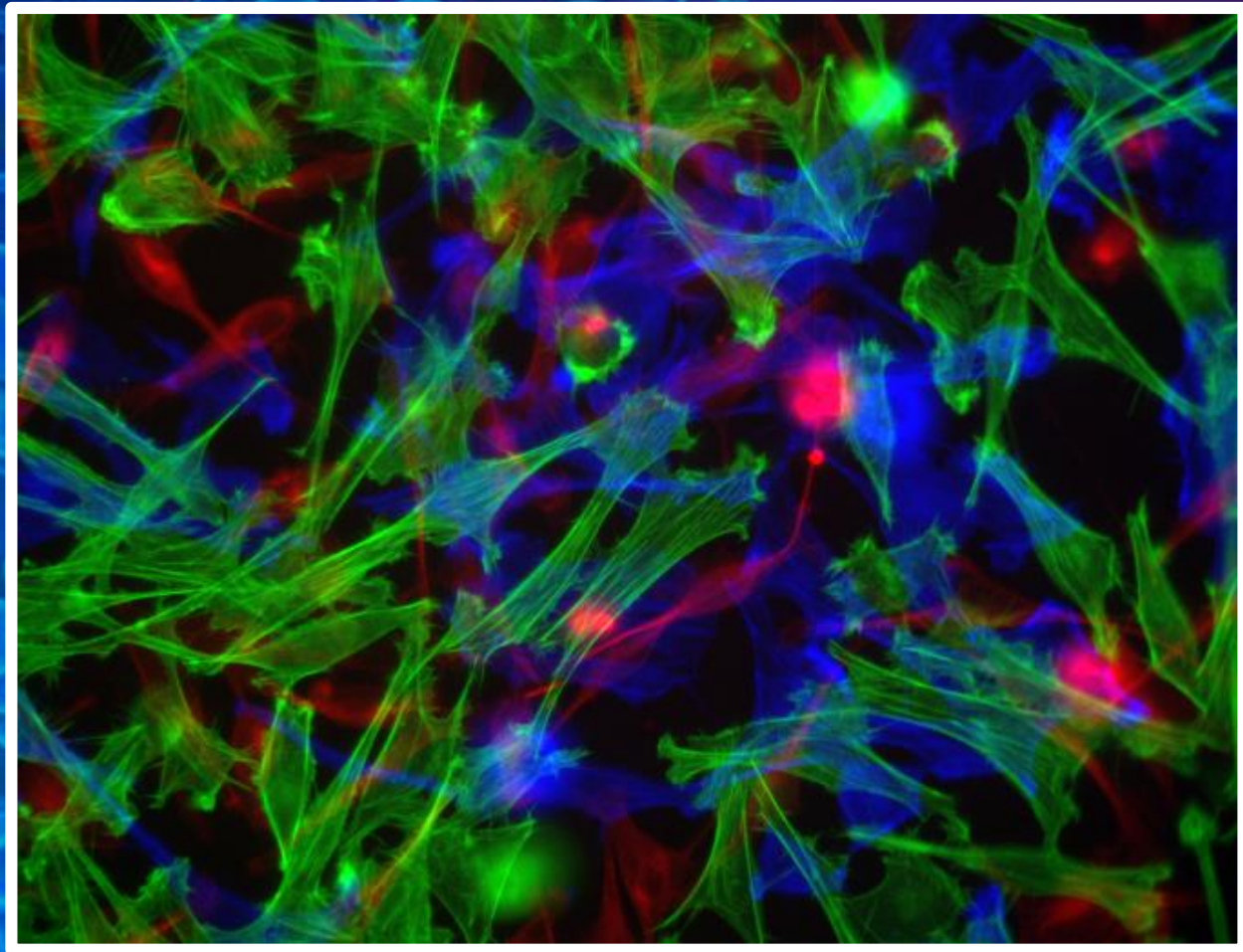
224 X 224 X 3

Other names and brands may be claimed as the property of others

Software and workloads used in performance tests may have been optimized for performance only on Intel® microprocessors.

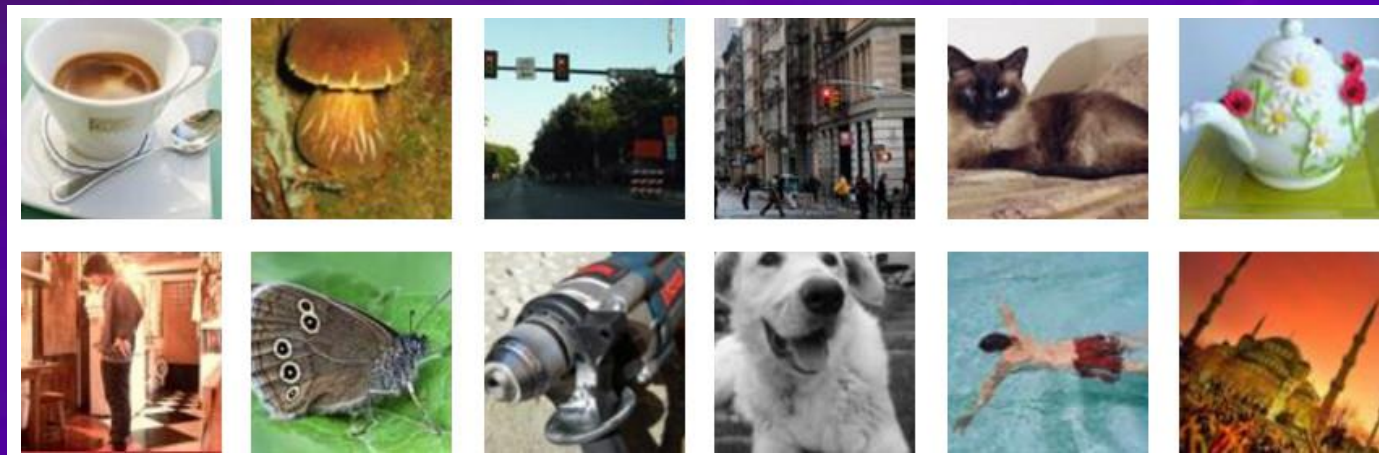
Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.

26X LARGER



1024 X 1280 X 3

IMAGENET



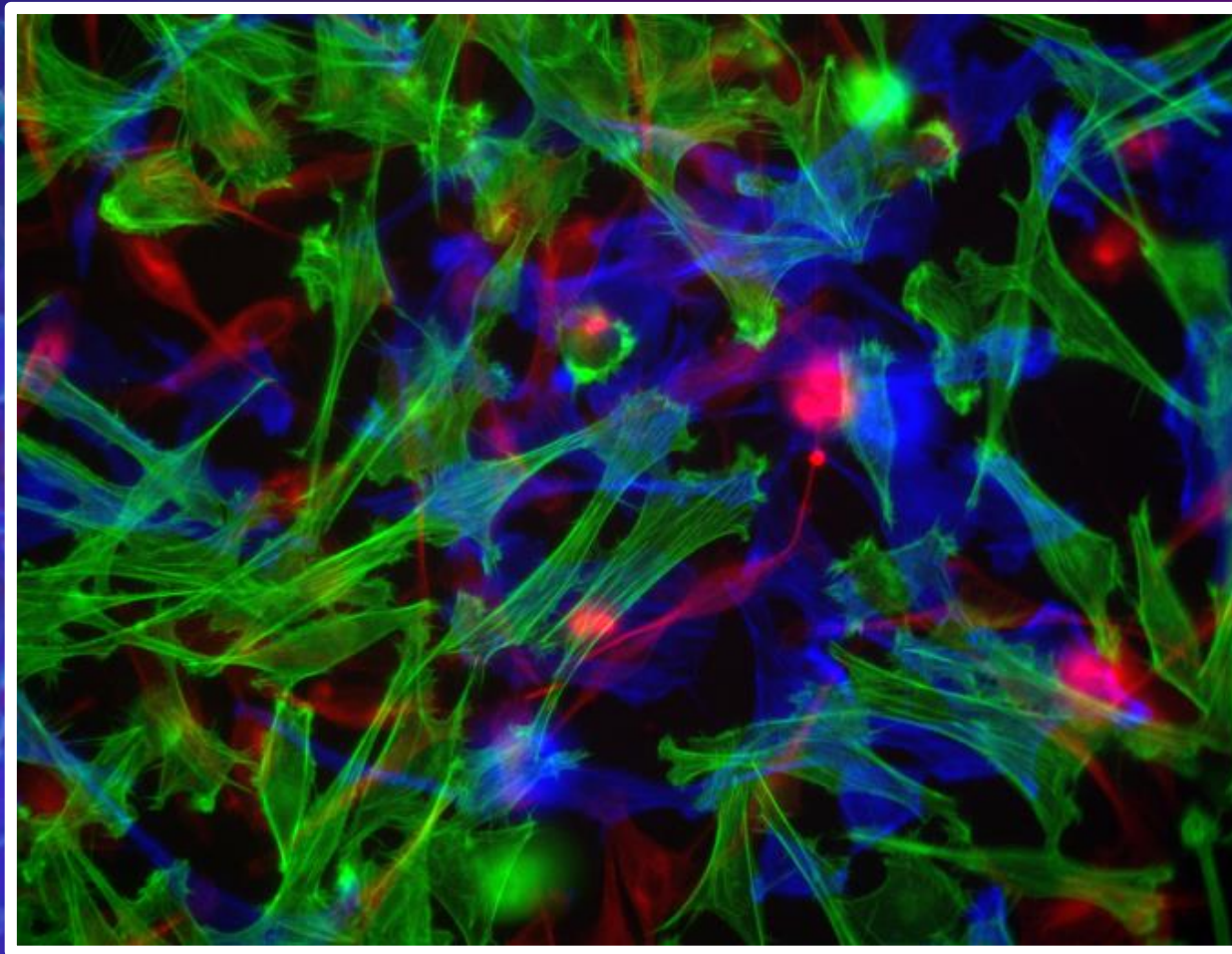
224 X 224 X 3

Other names and brands may be claimed as the property of others

Software and workloads used in performance tests may have been optimized for performance only on Intel® microprocessors.

Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.

26X LARGER MULTIPLE OBJECTS



1024 X 1280 X 3

Other names and brands may be claimed as the property of others

Software and workloads used in performance tests may have been optimized for performance only on Intel® microprocessors.

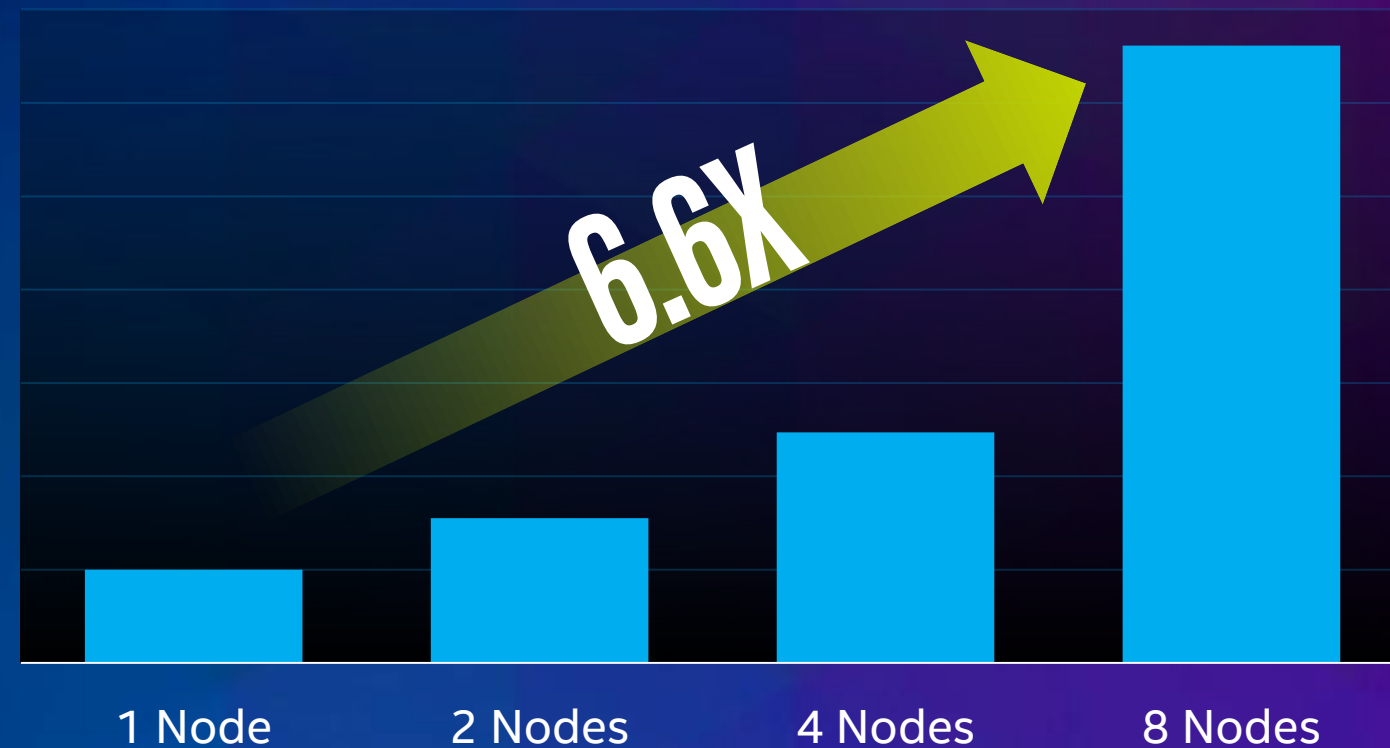
Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.

HIGH PERFORMANCE AT SCALE

SCALING OF TIME TO TRAIN

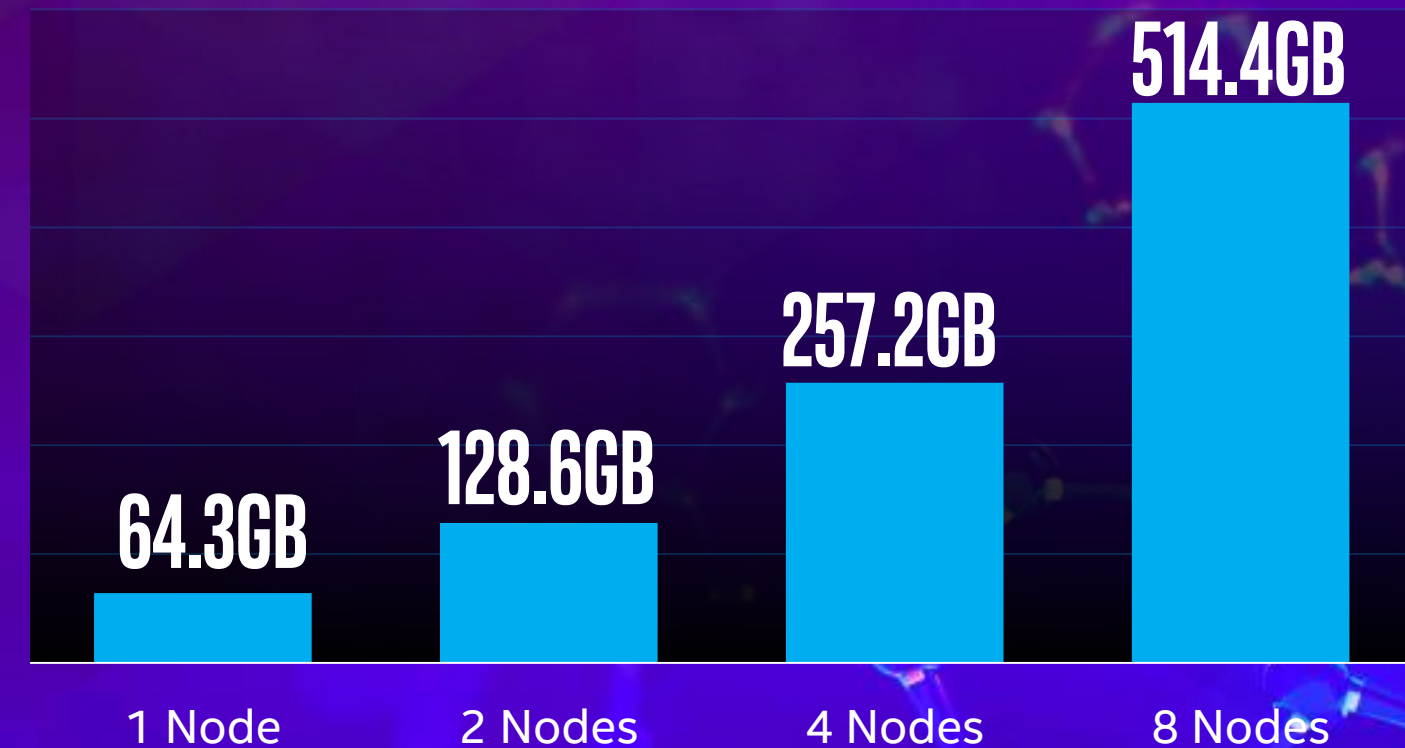
Intel® Omni-Path Architecture, Horovod and TensorFlow®

Speedup compared to baseline 1.0
measured in time to train in 1 nodes



TOTAL MEMORY USED

192GB DDR4 PER INTEL® 2S XEON® 6148 PROCESSOR



MULTISCALE CONVOLUTION NEURAL NETWORK



OPTIMIZED LIBRARIES

Intel® MKL/MKL-DNN,
cLDNN, DAAL

INTEL® OMNI-PATH ARCHITECTURE



Other names and brands may be claimed as the property of others
\$ Configuration: CPU: Xeon 6148 @ 2.4GHz, Hyper-threading: Enabled. NIC: Intel® Omni-Path Host Fabric Interface, TensorFlow: v1.7.0, Horovod: 0.12.1, OpenMPI: 3.0.0. OS: CentOS 7.3, OpenMPU 23.0.0, Python 2.7.5
Time to Train to converge to 99% accuracy in model
Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit <http://www.intel.com/performance>.

ARTIFICIAL INTELLIGENCE AND **INTEL® XEON® PROCESSORS**





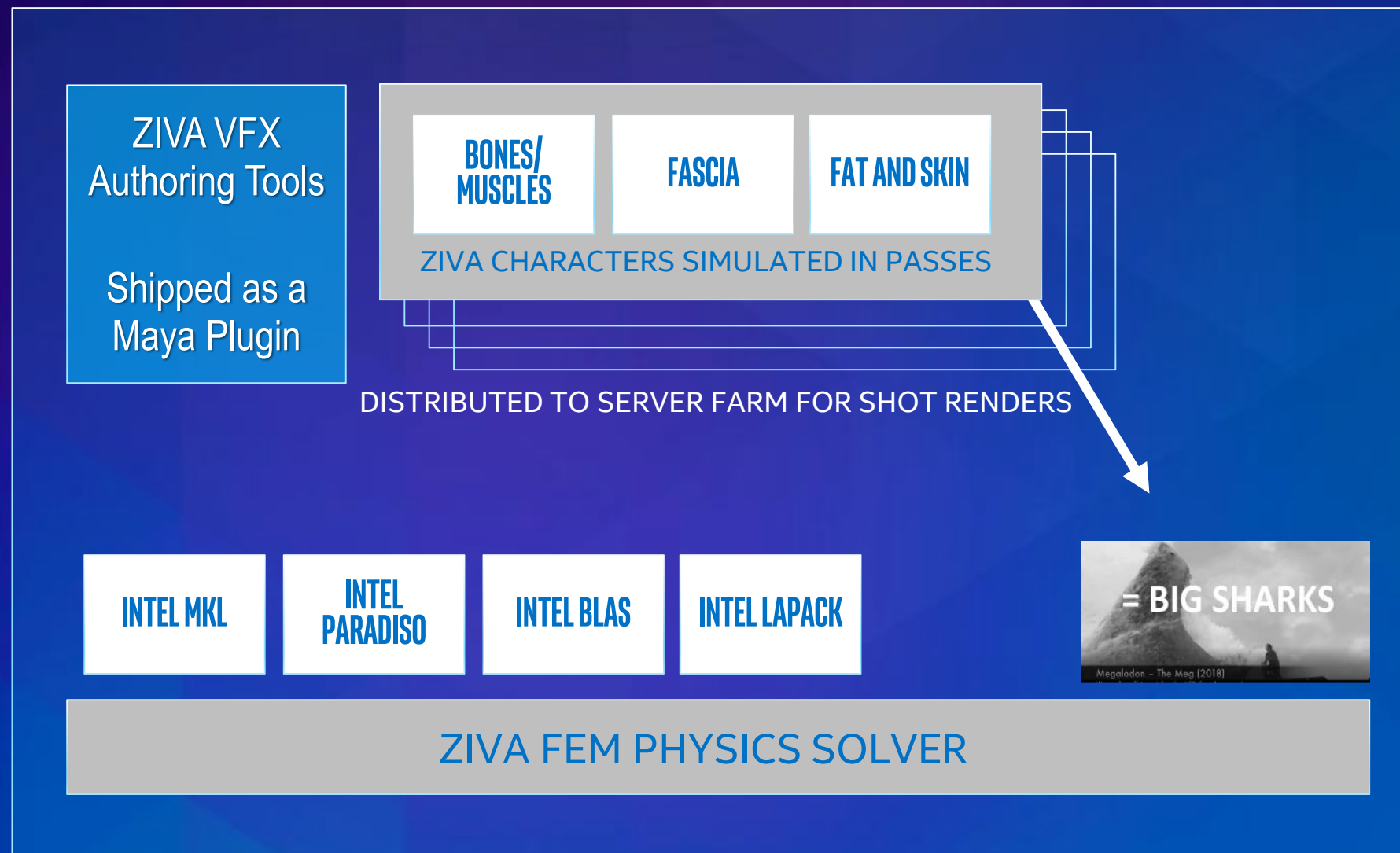
ZIVA

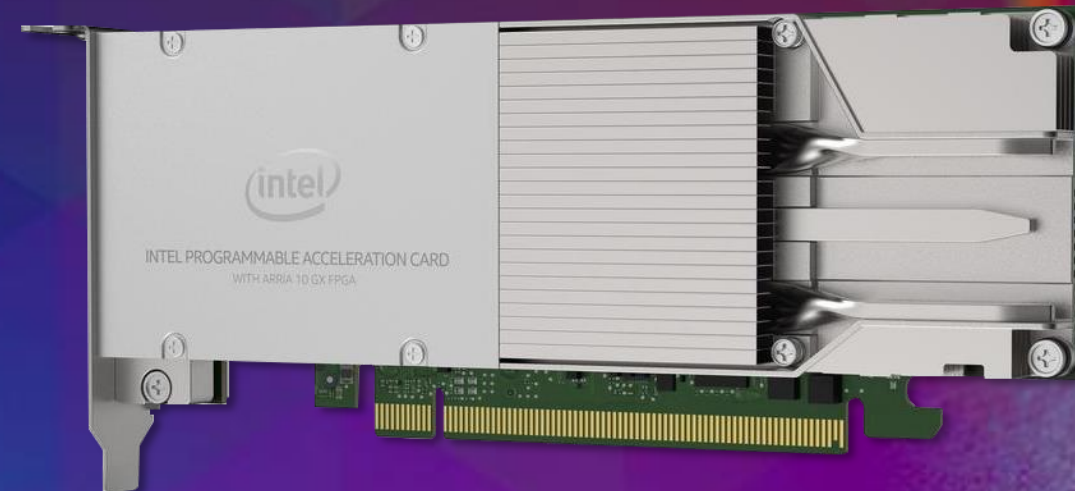
MACHINE LEARNING IMAGE RENDERING

JAMES JACOBS
CEO OF ZIVA



THE POWER OF ZIVA SIMULATION AND COMPUTE IN TODAY'S MEDIA & ENTERTAINMENT





FLEXIBLE REAL-TIME INFERRNCING FPGA PRODUCTS



// AI For Earth



GPU

STOP

1

1

10

45,546 images processed



FPGA

STOP

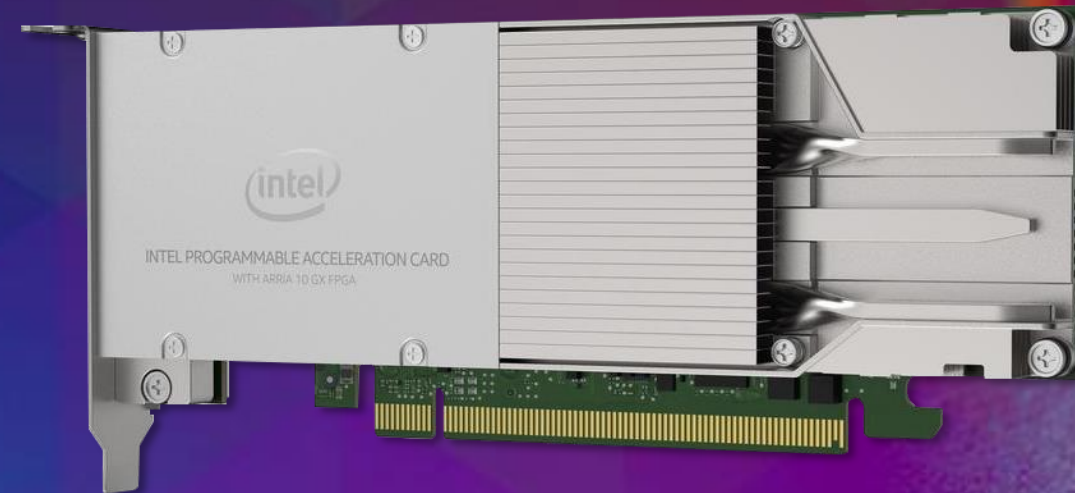
1

1

500K

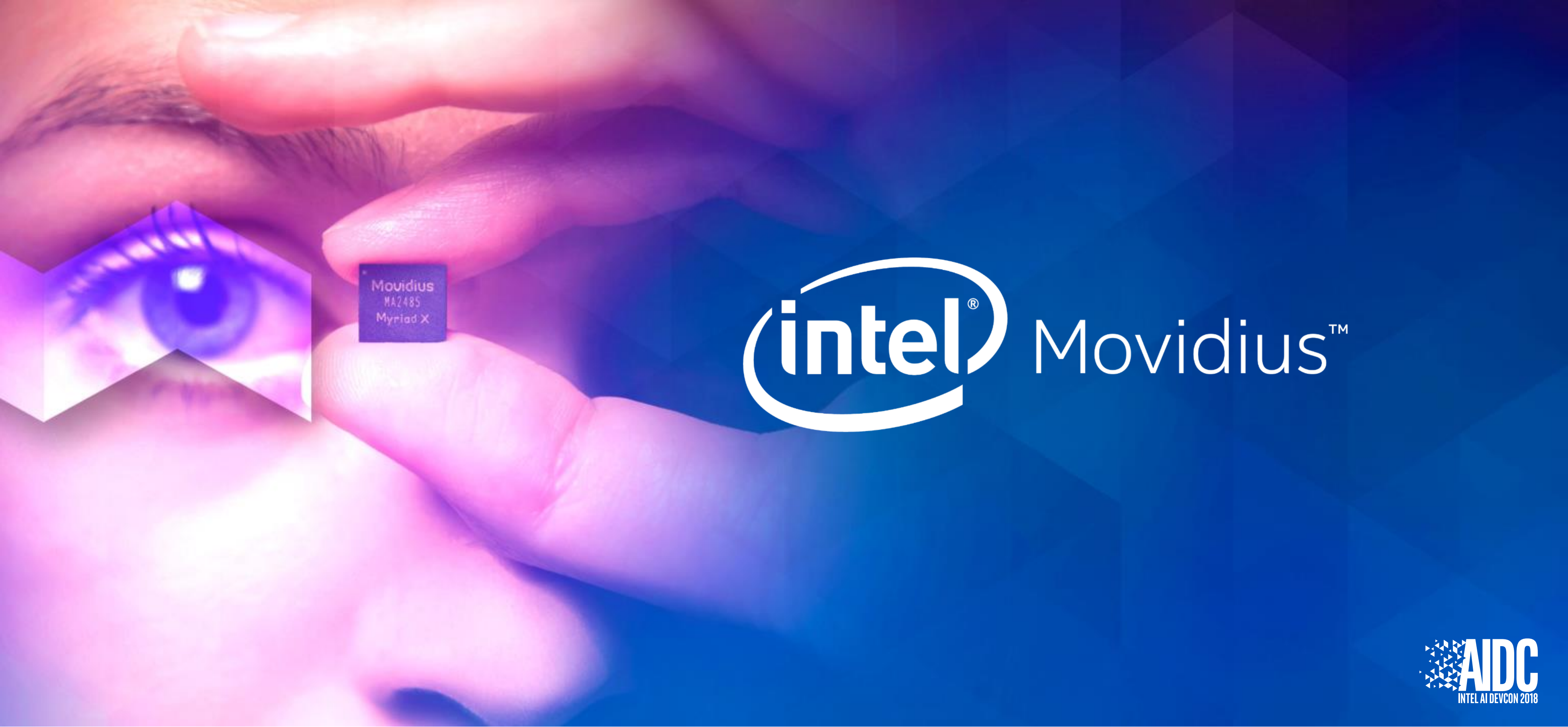
308,500 images processed





FLEXIBLE REAL-TIME INFERRNCING FPGA PRODUCTS





intel® Movidius™



- Deploy DNN and Computer Vision at the Edge
- Native FP16 and Fixed Point 8 bit support
- 4 TOPS with 1 TOPS of DNN Compute at 1W



AI GENERATED MUSIC



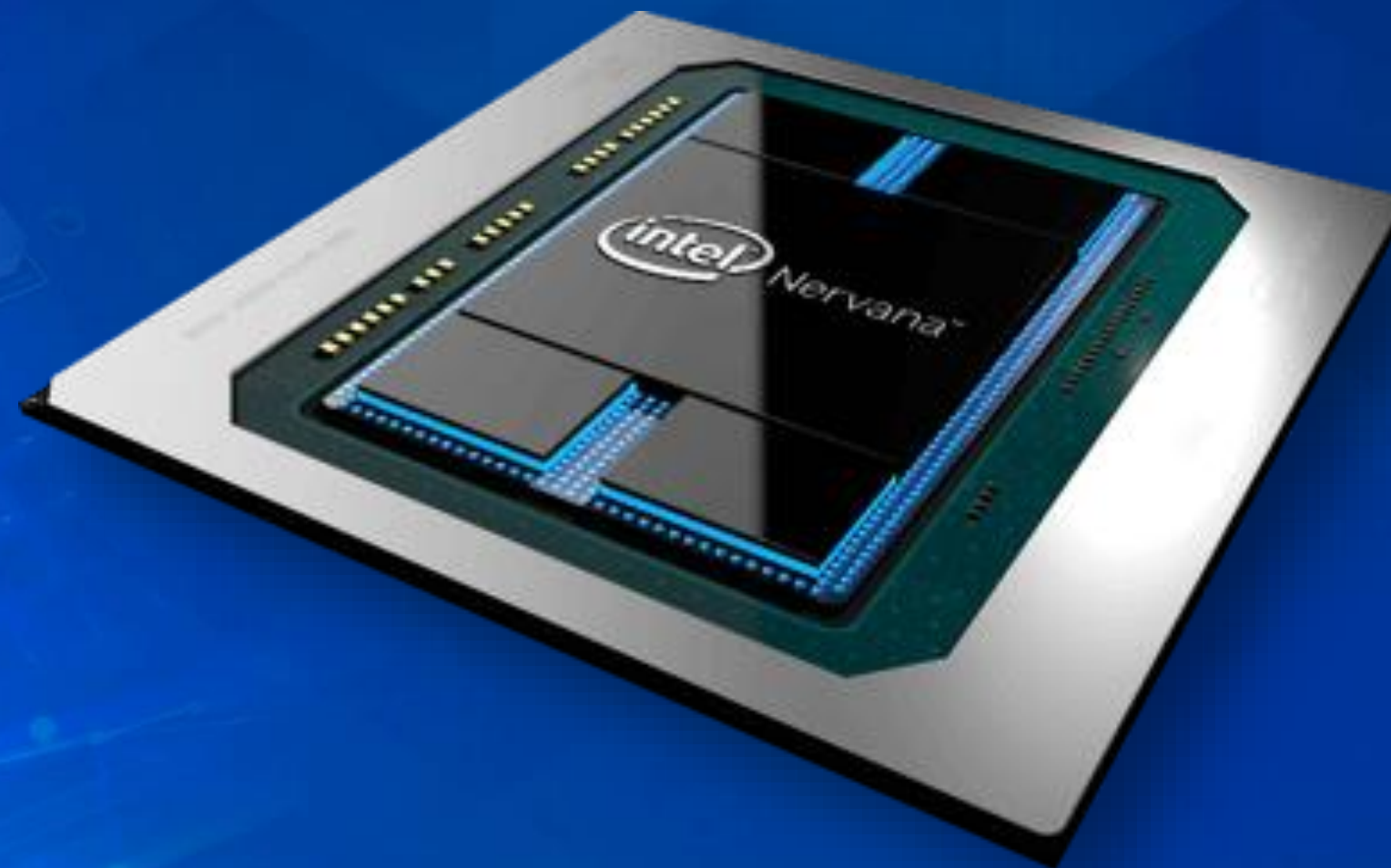
AI GENERATED MUSIC

KEVIN DOUCETTE

KATE KALLOT



NEURAL NETWORK PROCESSORS





GOALS – HIGH UTILIZATION & MODEL PARALLELISM

TOPs

■ Theoretical
■ Reality

0

Chip X

Lake Crest



GOALS – HIGH UTILIZATION & MODEL PARALLELISM

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit <http://www.intel.com/performance>

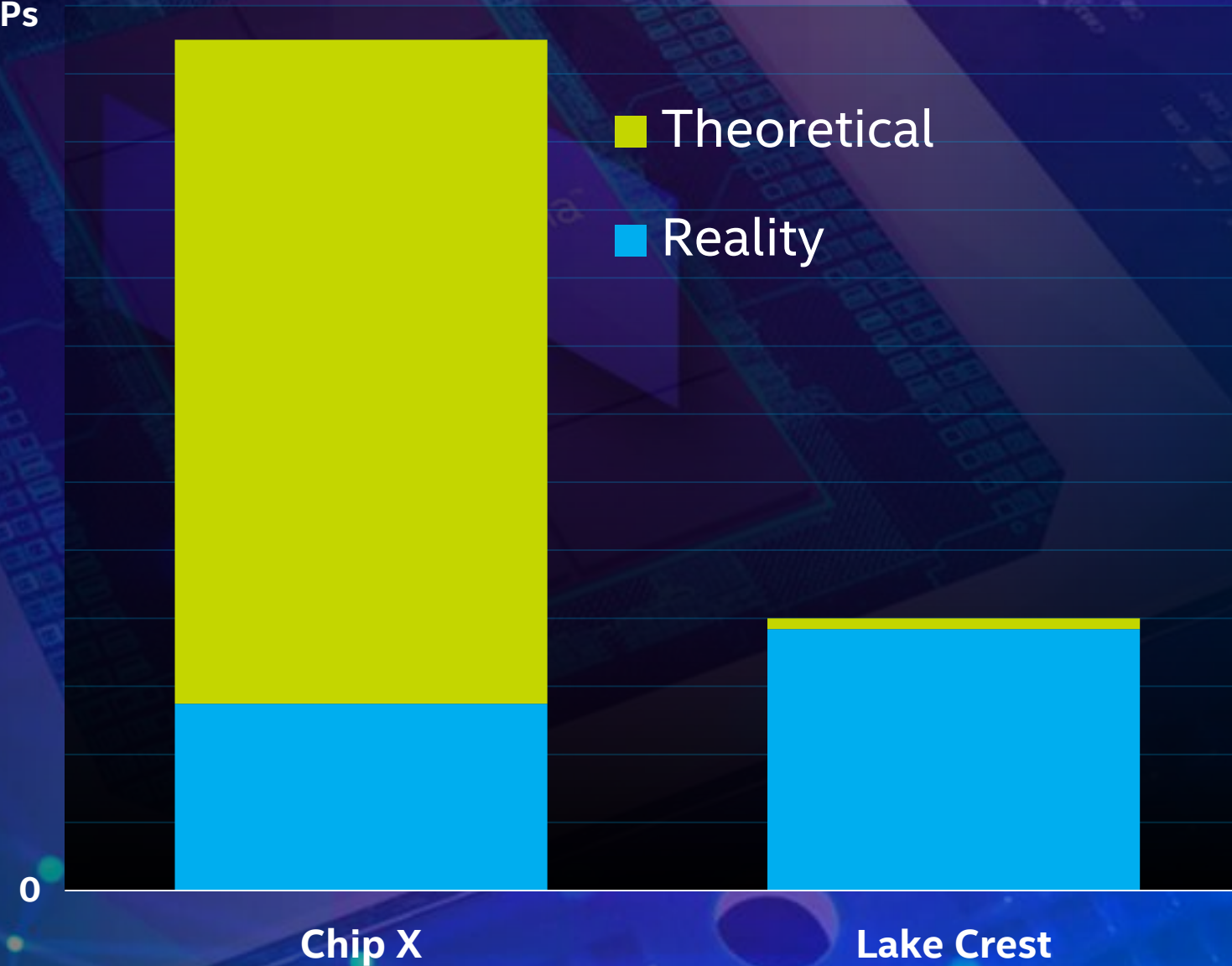
Chip X GEMM based on [DeepBench training](#) data for A(5124, 9124), B(9124,2560) matrix size GEMM operations performing DeepSpeech using FP16+ mixed precision at 27.43 TOPs.

Source: Lake Crest, Based on Intel measurements on limited distribution SDV, General Matrix-Matrix Multiplication; A(1536, 2048), B(2048, 1536).





TOPs



HIGH UTILIZATION & MODEL PARALLELISM

GEMM OPERATION
UTILIZATION¹

96.4%

A(1536, 2048)
B(2048, 1536)

MULTI-CHIP
SCALING²

96.2%

A(6144, 2048)
B(2048, 1536)

MULTI-CHIP
COMMUNICATION³

2.4 TB/s

OFF CHIP BANDWIDTH
<790ns LATENCY

POWER <210W

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit <http://www.intel.com/performance>

Chip X GEMM based on [DeepBench training](#) data for A(5124, 9124), B(9124,2560) matrix size GEMM operations performing DeepSpeech using FP16+ mixed precision at 27.43 TOPs.

Source: Lake Crest: Based on Intel measurements on limited distribution SDV

1 General Matrix-Matrix Multiplication; A(1536, 2048), B(2048, 1536)

2 Two chip vs. single chip GEMM performance; A(6144, 2048), B(2048, 1536)

3 Full Chip MRB-CHIP MRB data movement using send/rcv, Tensor size = (1, 32), average across 50K iterations



TOPs

■ Theoretical
■ Reality

0

Chip X

Lake Crest

Spring Crest (Estimate)



LATEST PROCESS NODE AND INCREASED COMPUTE DENSITY

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit <http://www.intel.com/performance>

Chip X GEMM based on [DeepBench training](#) data for A(5124, 9124), B(9124,2560) matrix size GEMM operations performing DeepSpeech using FP16+ mixed precision at 27.43 TOPs.

Source: Lake Crest - Based on Intel measurements on limited distribution SDV

Source: Spring Crest - Intel measurements on simulated product



FIRST COMMERCIAL NNP
INTEL[®] NERVANA[™]
NNP L-1000
in 2019

3-4x training performance
of first generation
Lake Crest product

SPRING CREST

PURPOSE BUILT DESIGN OPTIMIZED ACROSS
MEMORY BANDWIDTH, UTILIZATION, AND POWER

ENABLE DEVELOPERS TO ACHIEVE THEIR AI VISION



A snippet of Python code for Blender's mirror modifier, showing how to set up mirror operations for X, Y, and Z axes. The code includes comments and property assignments for the mirror_mod object.

```
... = modifier_ob.  
... mirror object to mirror  
... mirror_mod.mirror_object =  
... operation == "MIRROR_X":  
... mirror_mod.use_x = True  
... mirror_mod.use_y = False  
... mirror_mod.use_z = False  
... operation == "MIRROR_Y":  
... mirror_mod.use_x = False  
... mirror_mod.use_y = True  
... mirror_mod.use_z = False  
... operation == "MIRROR_Z":  
... mirror_mod.use_x = False  
... mirror_mod.use_y = False  
... mirror_mod.use_z = True  
...  
... at the end of the  
... ob.select= 1  
... ob.select= 1  
... context.scene.objects.active  
... "selected" = str(modifier)  
... bpy.context.selected_objects  
... data.objects[one.name].select  
...  
... print("please select exactly  
...  
... OPERATOR CLASSES ---  
...  
... types.Operator):  
... ask X mirror to the selected  
... object mirror mirror(X)  
... mirror X"
```

TOOLS



A close-up photograph of a complex electronic circuit board, likely a GPU or AI accelerator, featuring various components like capacitors, resistors, and integrated circuits.

HARDWARE



A photograph showing a group of people, likely developers or researchers, working together on laptops in a collaborative environment.

COMMUNITY

ENABLE DEVELOPERS TO ACHIEVE THEIR AI VISION



A vertical panel with a blue-tinted background showing a snippet of Python code for Blender's mirror modifier. The code includes comments and property assignments for 'MIRROR_X', 'MIRROR_Y', and 'MIRROR_Z' operations. The word 'TOOLS' is overlaid in large, white, semi-transparent letters.

TOOLS



A vertical panel with a blue-tinted background showing various AI hardware components, including circuit boards, chips, and cooling fans. The word 'HARDWARE' is overlaid in large, white, semi-transparent letters.

HARDWARE



A vertical panel with a blue-tinted background showing a group of people in a meeting or workshop setting, looking at a laptop screen. The word 'COMMUNITY' is overlaid in large, white, semi-transparent letters.

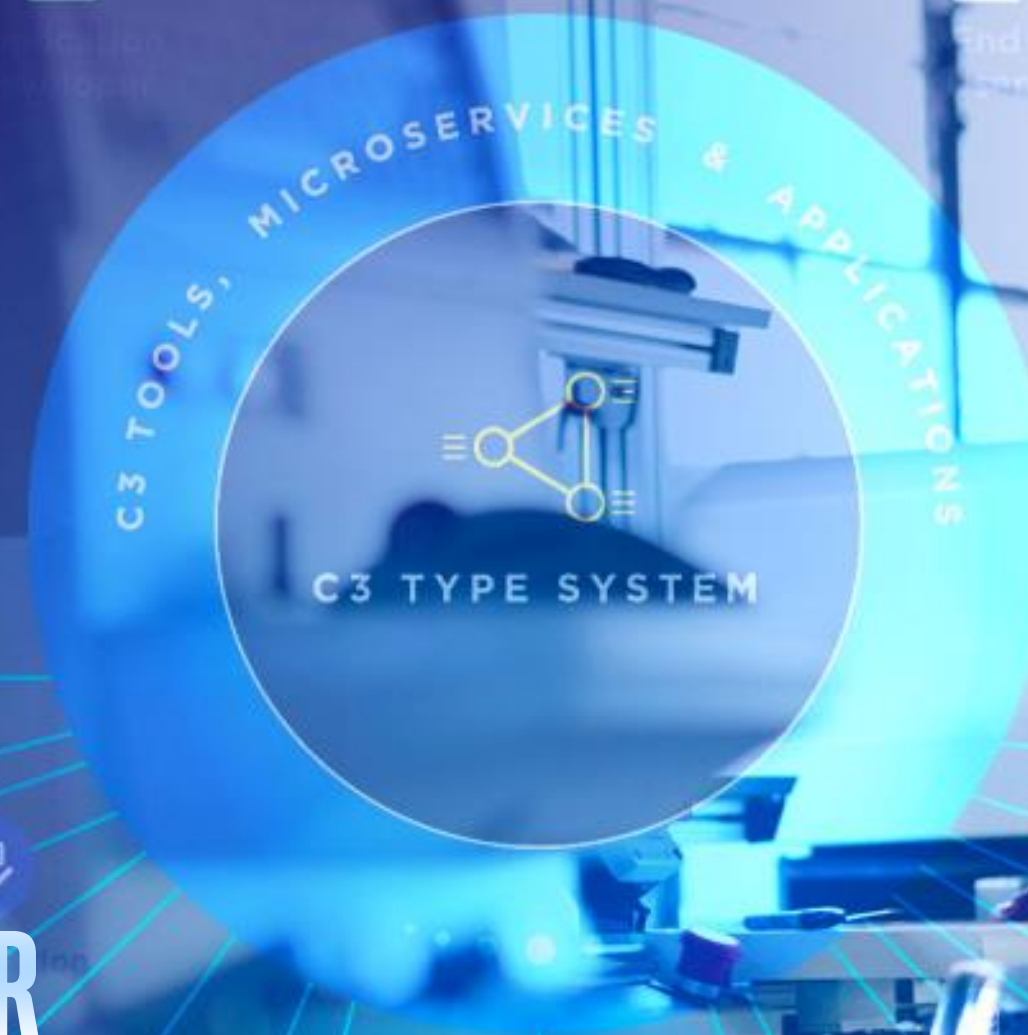
COMMUNITY

PARTNERSHIPS AND OUR DEVELOPER COMMUNITY



C3 IoT AI AND IOT

THOMAS SIEBEL
CHAIRMAN AND CHIEF EXECUTIVE OFFICER
C3 IOT





AMAZON AI AND MACHINE LEARNING

BRATIN SAHA

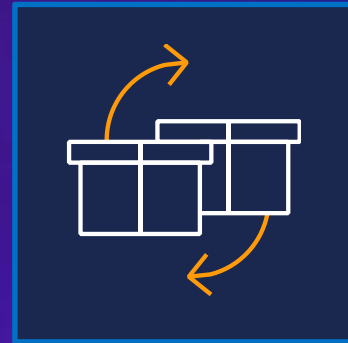
**VICE PRESIDENT AND GENERAL MANAGER,
MACHINE LEARNING PLATFORMS
AMAZON AI - AMAZON**



Machine Learning at Amazon: a long heritage



Personalized
recommendations



Fulfillment automation
/ inventory management



Cargo



Voice driven
interactions



Inventing entirely new
customer experiences

Machine Learning Platform

Feature
Staging

REAL TIME
DATA

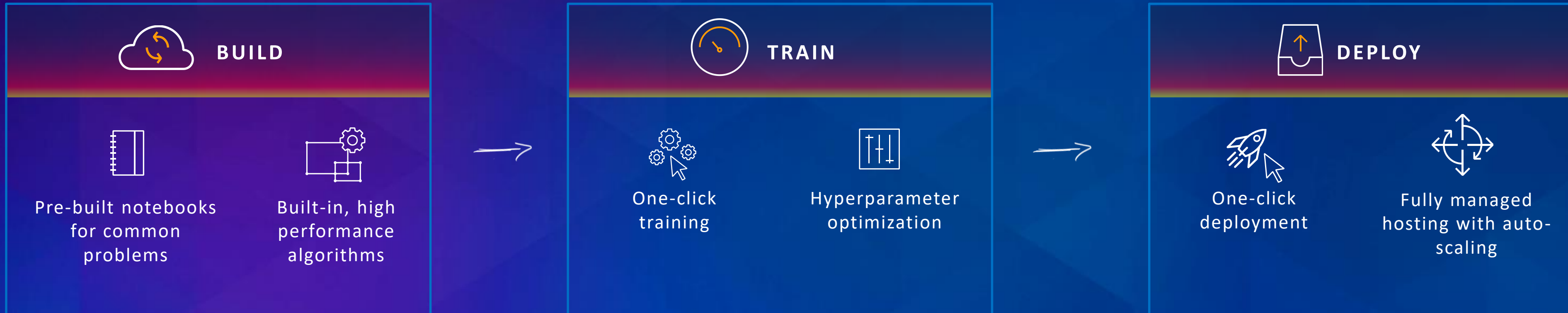
Amazon EC2

Developers

Task Management

Input/Output
AIDC
INTEL AI DEVCON 2018

Amazon SageMaker



AWS Machine Learning Stack

APPLICATION SERVICES



REKOGNITION



REKOGNITION
VIDEO



POLLY



TRANSCRIBE



TRANSLATE



COMPREHEND



LEX

PLATFORM SERVICES



AMAZON
SAGEMAKER

FRAMEWORKS AND INTERFACES



Frameworks



KERAS



Interfaces



AWS
DEEPLENS



INFRASTRUCTURE



EC2 GPUs



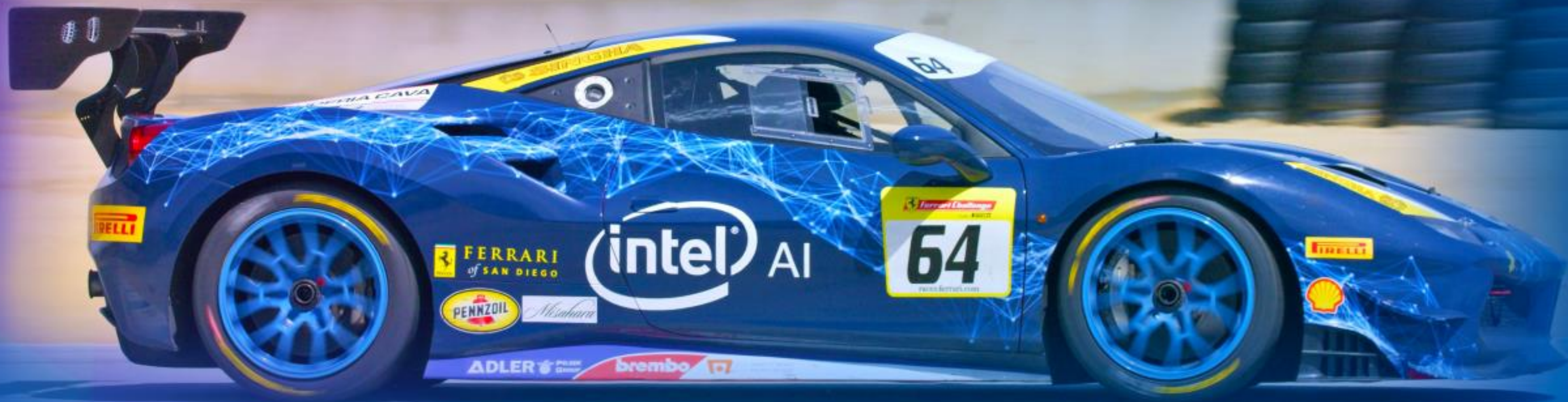
EC2 CPUs



IoT



Edge

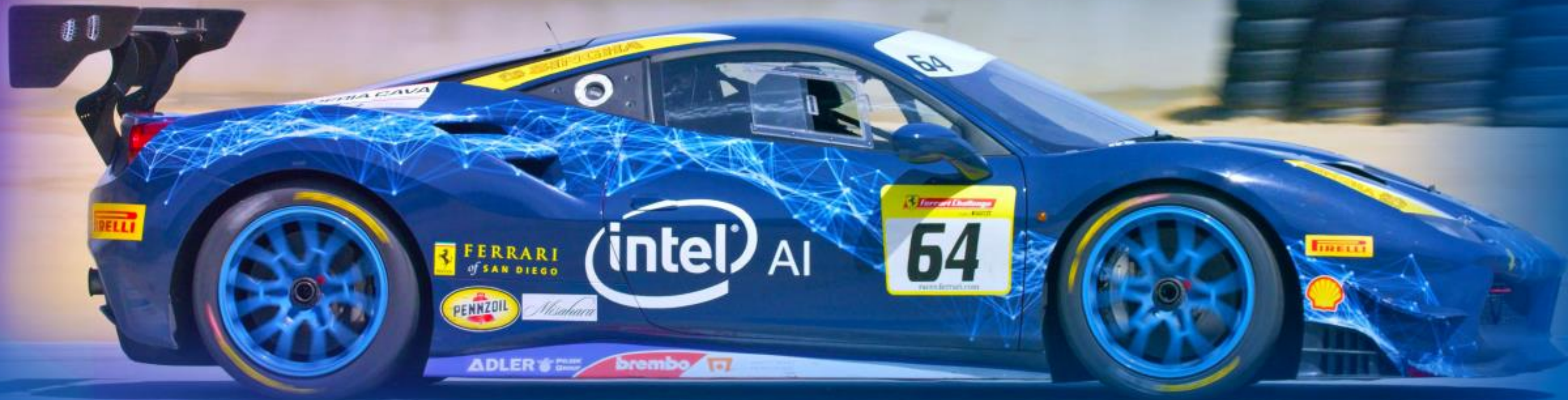




ANDY KELLER

DATA SCIENTIST

INTEL AI





YINYIN LIU
HEAD OF DATA SCIENCE
INTEL AI

OPEN SOURCE LIBRARIES

**NLP
ARCHITECT**



NervanaSystems/nlp-architect

**RL
COACH**



NervanaSystems/coach

**NEURAL
NETWORK
DISTILLER**



NervanaSystems/distiller



MLPerf

GOAL IS TO SELECT A SET OF ML PROBLEMS, EACH DEFINED BY A DATASET AND QUALITY TARGET, THEN MEASURE THE WALL CLOCK TIME TO TRAIN A MODEL FOR EACH PROBLEM.

For more information, see: <https://mlperf.org/>



OFFICIAL ARTIFICIAL INTELLIGENCE
PLATFORM PARTNER

SEE YOUR CONCEPTS* BECOME
A REALITY IN TOKYO 2020

[AICHALLENGE.INTEL.COM](https://aichallenge.intel.com)

*Idea implementation at the Olympic Games subject to approval





HELPING MAKE YOUR AI VISION A REALITY

TOOLS

HARDWARE

COMMUNITY

RISK FACTORS

Today's presentation contains forward-looking statements. All statements made that are not historical facts are subject to a number of risks and uncertainties, and actual results may differ materially. Please refer to our most recent earnings release, Form 10-Q and 10-K filing available on our website for more information on the risk factors that could cause actual results to differ.

DISCLAIMERS

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to www.intel.com/benchmarks.

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice Revision #20110804 Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at [\[intel.com\]](http://intel.com).

Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, the Intel logo, Xeon, Intel Nervana and Intel Movidius are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

© Intel Corporation.

