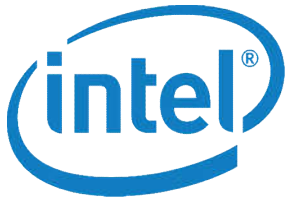
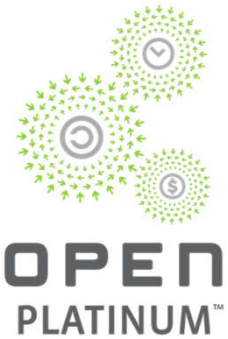
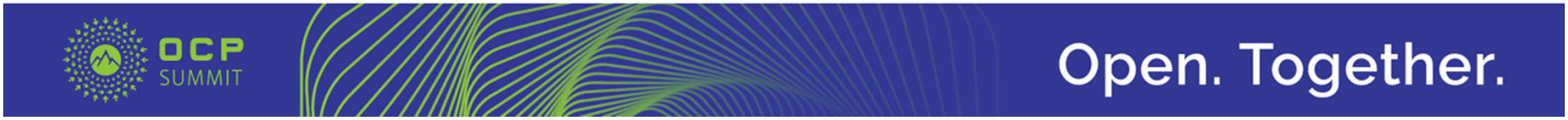

Open. Together.
OCP
SUMMIT



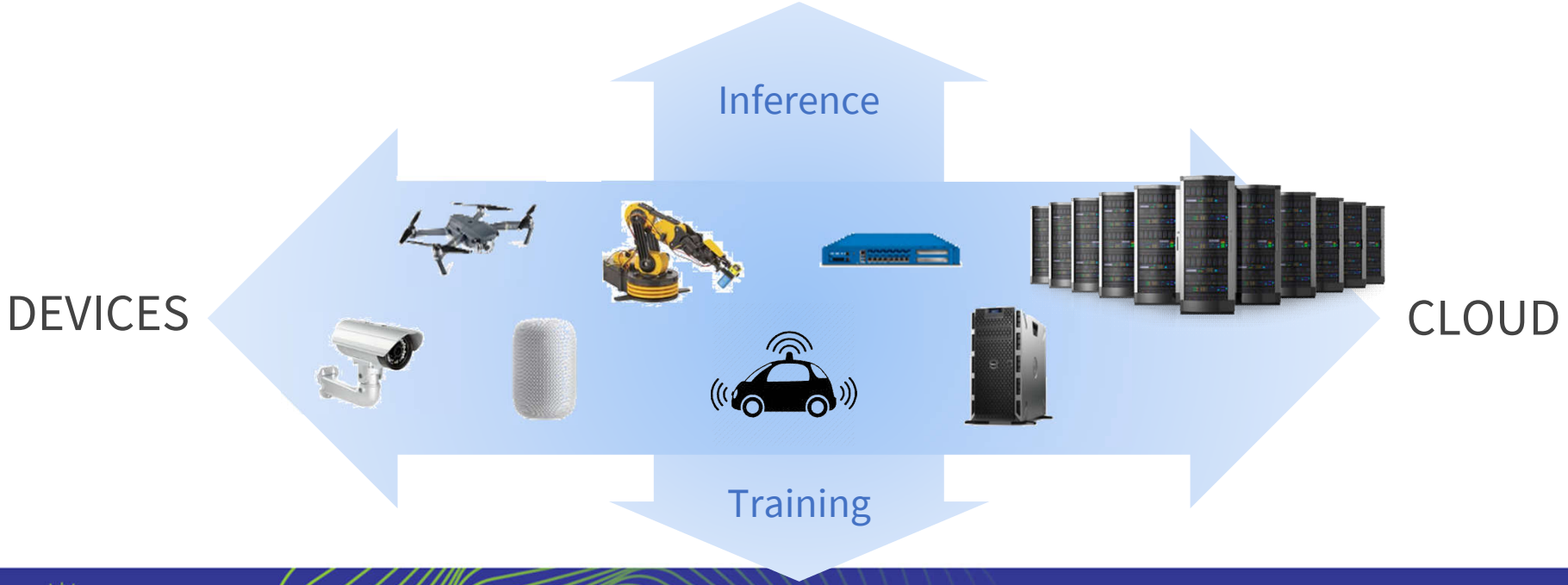
Breaking Barriers in AI: New Hardware, Standard Platform

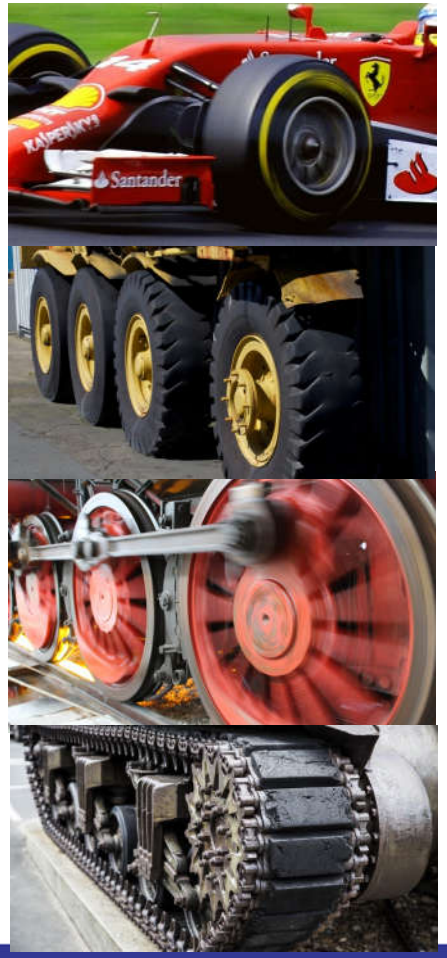
Tingwei Huang, Product Management Director, Intel



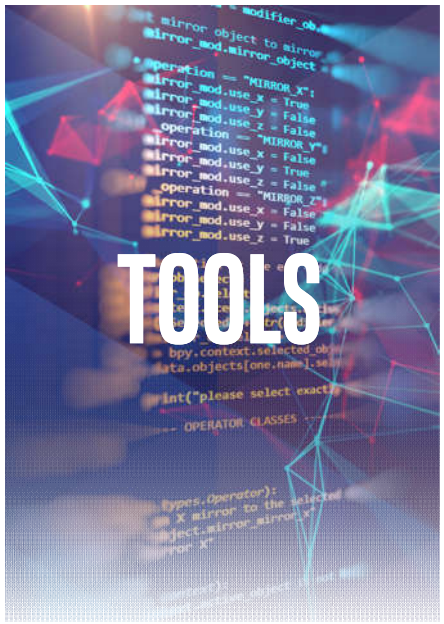


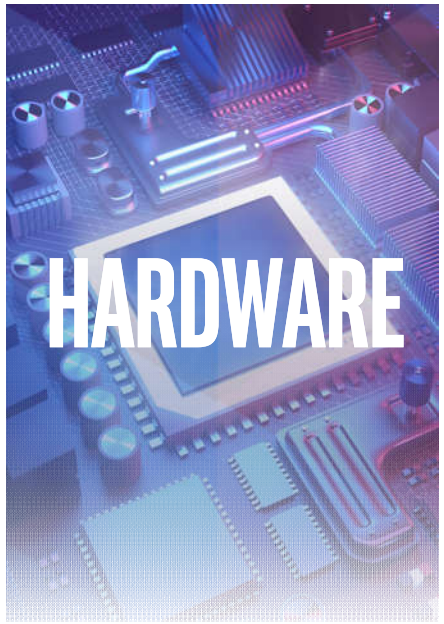
AI is Not a Workload
AI Enhances Workloads



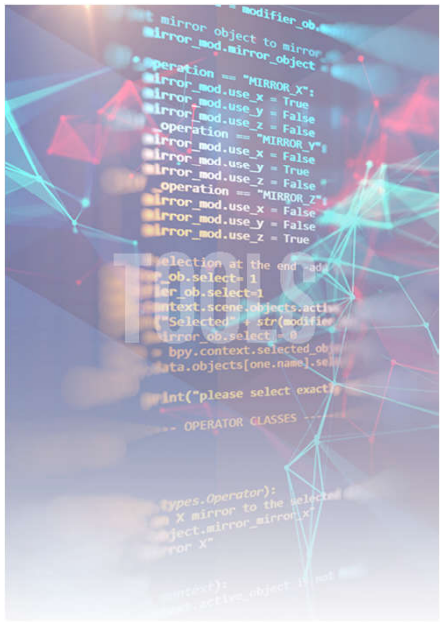

Santander

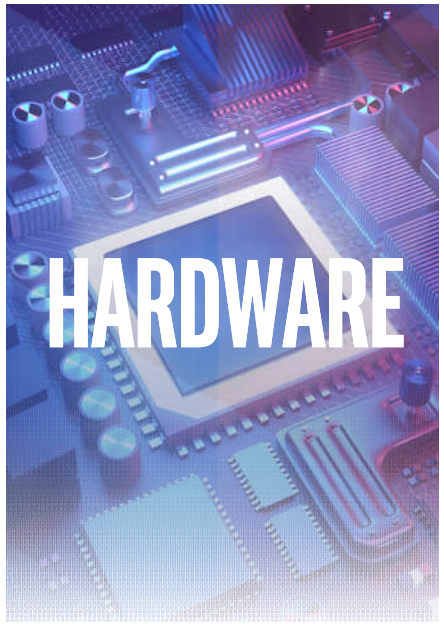
Enable Developers to Achieve Their AI Vision





Enable Developers to Achieve Their AI Vision





Intel® Xeon® is the Foundation of AI

Maximize Utilization
running data center and AI workloads side-by-side

Break Memory Barriers
in order to apply AI to large data sets and models

Train Complex Models
through efficient scaling to many nodes

Access Optimized Tools
including continuous performance gains for TensorFlow, MXNet, more

Run in the Cloud
including AWS, Microsoft, Alibaba, Tencent, Google, Baidu, more

Intel® AI Compute Hardware 2019

DEVICE

EDGE

MULTI CLOUD

+ MOBILEYE An Intel Company — **Automated Driving**

+ intel STRATIX 10 inside — **Flexible Acceleration**

+ intel MOVIDIUS inside — **Dedicated Media/Vision**

+ intel NERVANA inside — **NNP-I Intensive DL Inference**

+ intel NERVANA inside — **NNP-L Intensive DL Training**

Intel® Nervana™ Neural Network Processor
NNP L-1000 for Deep Learning Training

OCP Accelerator Mezzanine Module

Standard PCIE Module

Blazingly Fast Data Access

Leverage both on-die SRAM and on-package HBM

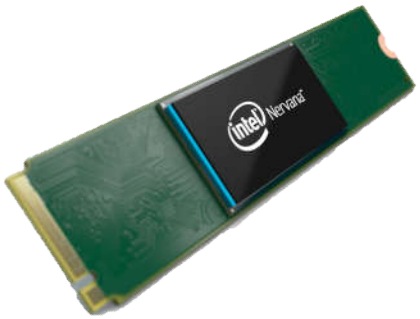
High Degree of Parallelism

Bfloat16 for enhanced performance and convergence

New Level of Scalability

Massive bi-directional data transfer through on and off chip interconnect

Intel® Nervana™ Neural Network Processor
NNP I-1000 for Deep Learning Inference

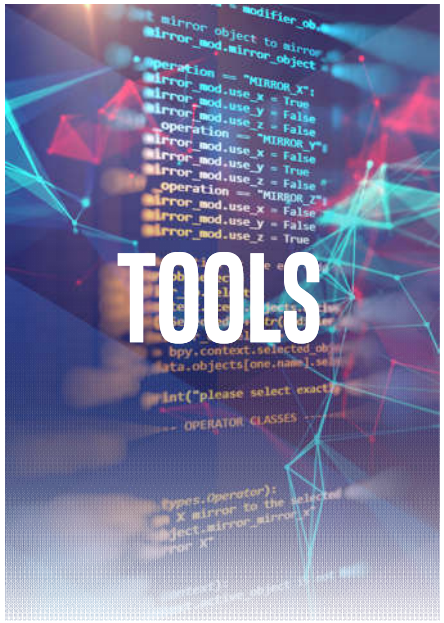


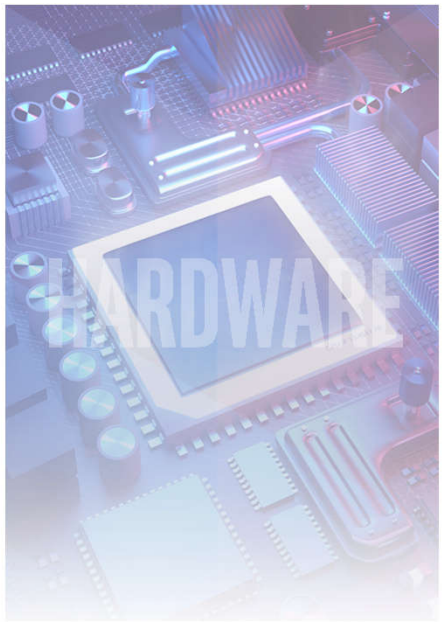
Deep Learning by Design

Intel® 10nm Ice Lake Core

In Production in 2019

Enable Developers to Achieve Their AI Vision

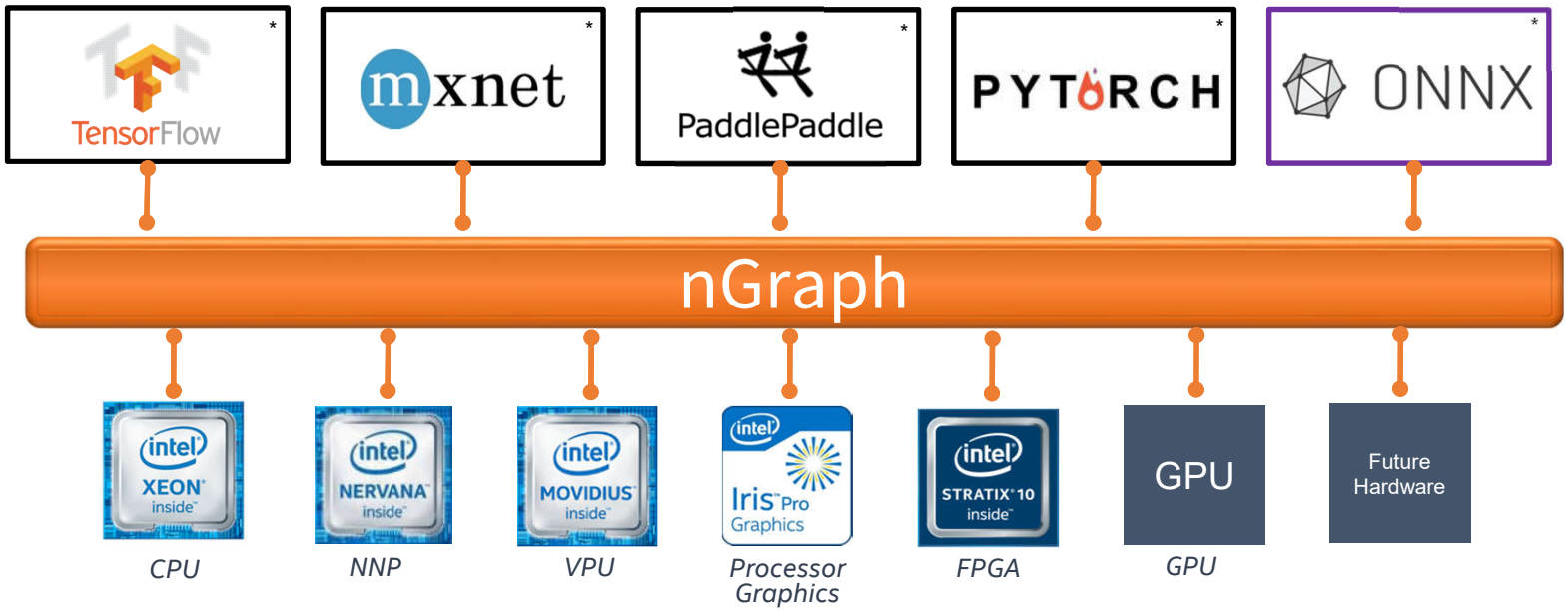




nGraph

BETA!

github.com/NervanaSystems/ngraph

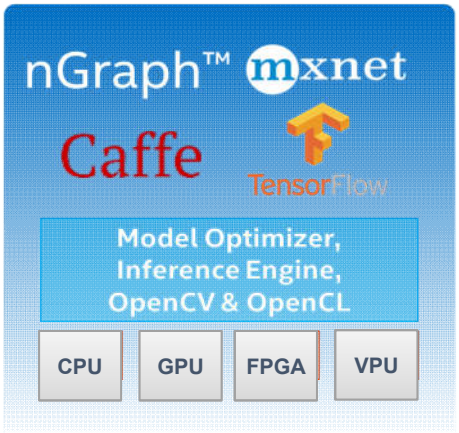


Open Source

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

OpenVINO™ Toolkit

Visual Inference and Neural Network Optimization



DEPLOY COMPUTER VISION AND DEEP LEARNING CAPABILITIES TO THE EDGE

Write Once, Scale to Diverse Accelerators

Broad Framework Support

High Performance, High Efficiency for Edge Inference

Free Download ▸ software.intel.com/openvino-toolkit

Open Source version ▸ 01.org/openvinotoolkit

VPU = Vision Processing Unit (Movidius)

NAUTA

BETA!

A Distributed Deep Learning Platform for Kubernetes*

github.com/IntelAI/nauta

BUILD

Multi-user collaboration
Interactive sessions
Template functionality

TRAIN

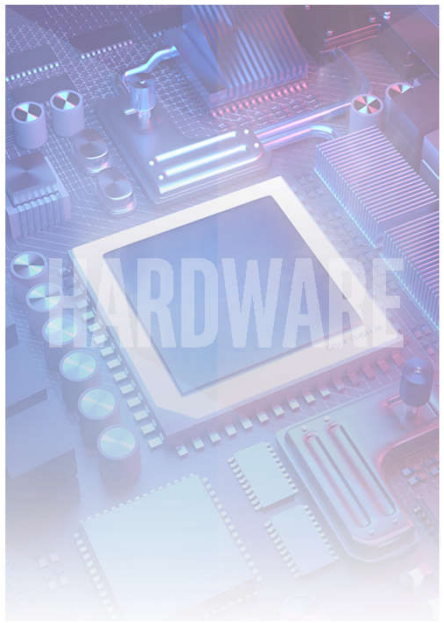
Fast training
Batch training
Experiment tracking
Multi-node distribution
Analytics & visualization using TensorBoard*

EVALUATE

Batch inference
Inference deployment
Export to edge devices

Open Source

Enable Developers to Achieve Their AI Vision



AI Builder Program

200+ Members: Software, System, Service and Solution Providers




Tech Enablement + **Co-marketing** + **Match-making** + **Investment**

https://builders.intel.com/ai

Notice and Disclaimer

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer. No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

Statements in this document that refer to Intel's plans and expectations for the quarter, the year, and the future, are forward-looking statements that involve a number of risks and uncertainties. A detailed discussion of the factors that could affect Intel's results and plans is included in Intel's SEC filings, including the annual report on Form 10-K.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

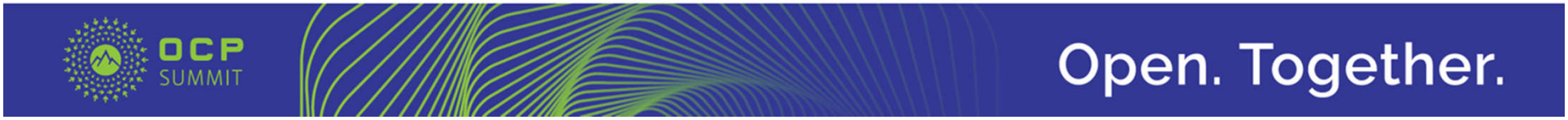
Performance estimates were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown." Implementation of these updates may make these results inapplicable to your device or system.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

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.

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, the Intel logo, Pentium, Celeron, Atom, Core, Xeon, Movidius, Saffron and others are trademarks of Intel Corporation in the U.S. and/or other countries.
*Other names and brands may be claimed as the property of others.




Open. Together.
OCP Global Summit | March 14–15, 2019
OCP
SUMMIT