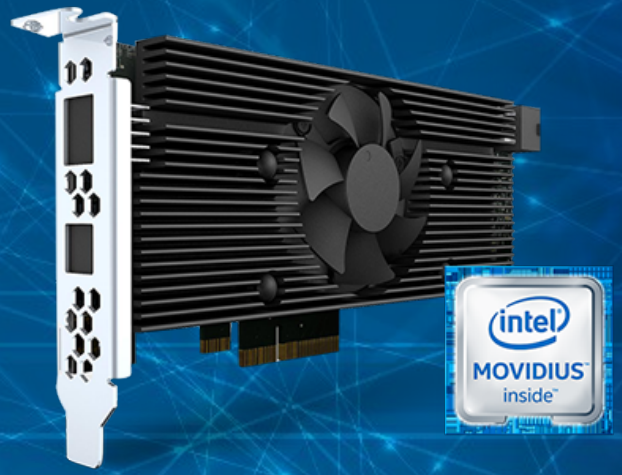
INTEL® HDDL VISION ACCELERATION

High Performance for AI Inference



LEGAL NOTICES AND DISCLAIMERS

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 www.intel.com.

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.

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.

Any forecasts of goods and services needed for Intel's operations are provided for discussion purposes only. Intel will have no liability to make any purchase in connection with forecasts published in this document.

Altera, Arria, the Arria logo, Intel, the Intel logo, Intel Atom, Intel Core, Intel Optane, Iris, Movidius, OpenVINO, Stratix, the Stratix logo and Xeon 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.

WHAT IS AI?

AI

MACHINE LEARNING

Regression
Decision trees

Classification
Extrapolation

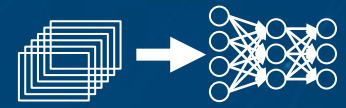
Clustering

DEEP LEARNING

Image Processing
Speech Recognition

Natural Language Processing

Computer Vision



SUPERVISED LEARNING

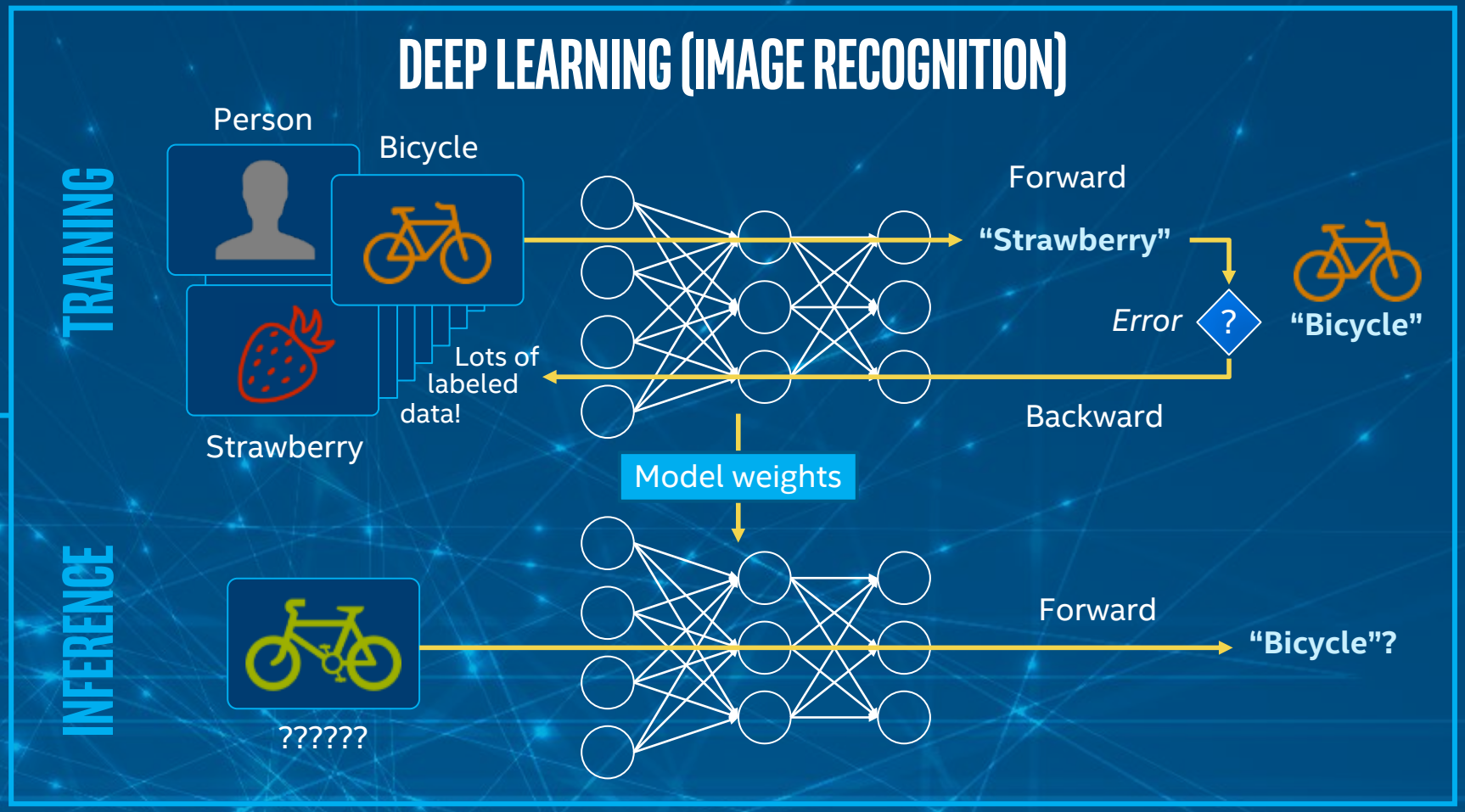
UNSUPERVISED LEARNING

REINFORCEMENT LEARNING

WHAT DOES AI DO?

DEEP LEARNING

Image Processing



FROM EDGE TO CLOUD, AI HAS CREATED A HUGE OPPORTUNITY FOR DEVELOPERS

KEY DRIVERS TO THE EDGE: BANDWIDTH, STORAGE, LATENCY, SECURITY

AI
THINGS

AI
EDGE COMPUTE

AI
NETWORK INFRASTRUCTURE

AI
DATA CENTER/CLOUD

45% of data will be stored, analyzed, and acted on at the edge by end of 2019[1]

43% share of AI tasks taking place on edge devices (vs. cloud) in 2023[2]

15X growth in devices with edge AI capabilities by 2023[2]

COMPUTER VISION AND ARTIFICIAL INTELLIGENCE ARE TRANSFORMING IOT DEVICES AT THE NETWORK EDGE

Public Safety & Surveillance
Traffic, Parking and LPR
Emergency Response

People Counting
Customer (i.e., Gender, Wait Time)
ATM Facial Recognition

Machine Vision
Asset Inspection (i.e., Pipeline)
Augmented Reality

Public Safety & Surveillance
Facial Recognition

Autonomous Vehicles
Public Safety (i.e., Bus/Rail)
Traffic & People Counting

Security & Surveillance
Responsive Retail Advertising
Digital Home Assitant

Manufacturing Automation
Industrial (i.e., Pipeline Welding)

Emergency Response
Asset Inspection (i.e., Windmill)

INTEL® AI TECHNOLOGY IN ACTION

Machine Learning and Mammography
Detecting invasive ductal carcinoma with convolutional neural networks showing how existing deep learning technologies can be utilized to train artificial intelligence (AI) to be able to detect invasive ductal carcinoma (IDC)[1] (breast cancer) in unlabeled histology images.

AI Assists with Skin Cancer Screening
Doctor Hazel, a skin cancer screening service powered by AI that operates in real time, relies on an extensive library of images to distinguish between skin cancer and benign lesions, making it easier for people to seek professional medical advice.

AI Helps Detect Bacteria in Water
Offline analysis is accomplished with a digital microscope connected to a laptop running Ubuntu* and the Intel® Movidius™ Neural Compute Stick. After analysis, contamination sites are marked on a map in real time

Shark Detection to Save Lives
Intel® Movidius™ Neural Compute Stick was used with Australia's Little Ripper Lifesaver UAV to monitor the New South Wales coastline for sharks.

Nudity Detection to Fight Child Pornography
The NaaS (Nudity-detection as a service) the result of a project by Haschek Solutions. It uses a cluster of three Raspberry Pis and two Intel Movidius Neural Compute Sticks to analyze photographs to identify pornographic images.

CORaiL: Coral Reef Restoration and Research
Prototype a fully functional modular AI-powered underwater camera unit that continuously counts the number of visible fauna, and, as possible, assign a taxonomy to help protect endangered coral reefs.

AI has the power to make a difference and change lives. What will you make?
Get Started Today ▸ intel.com/ncs

INTEL® VISION PRODUCT PORTFOLIO

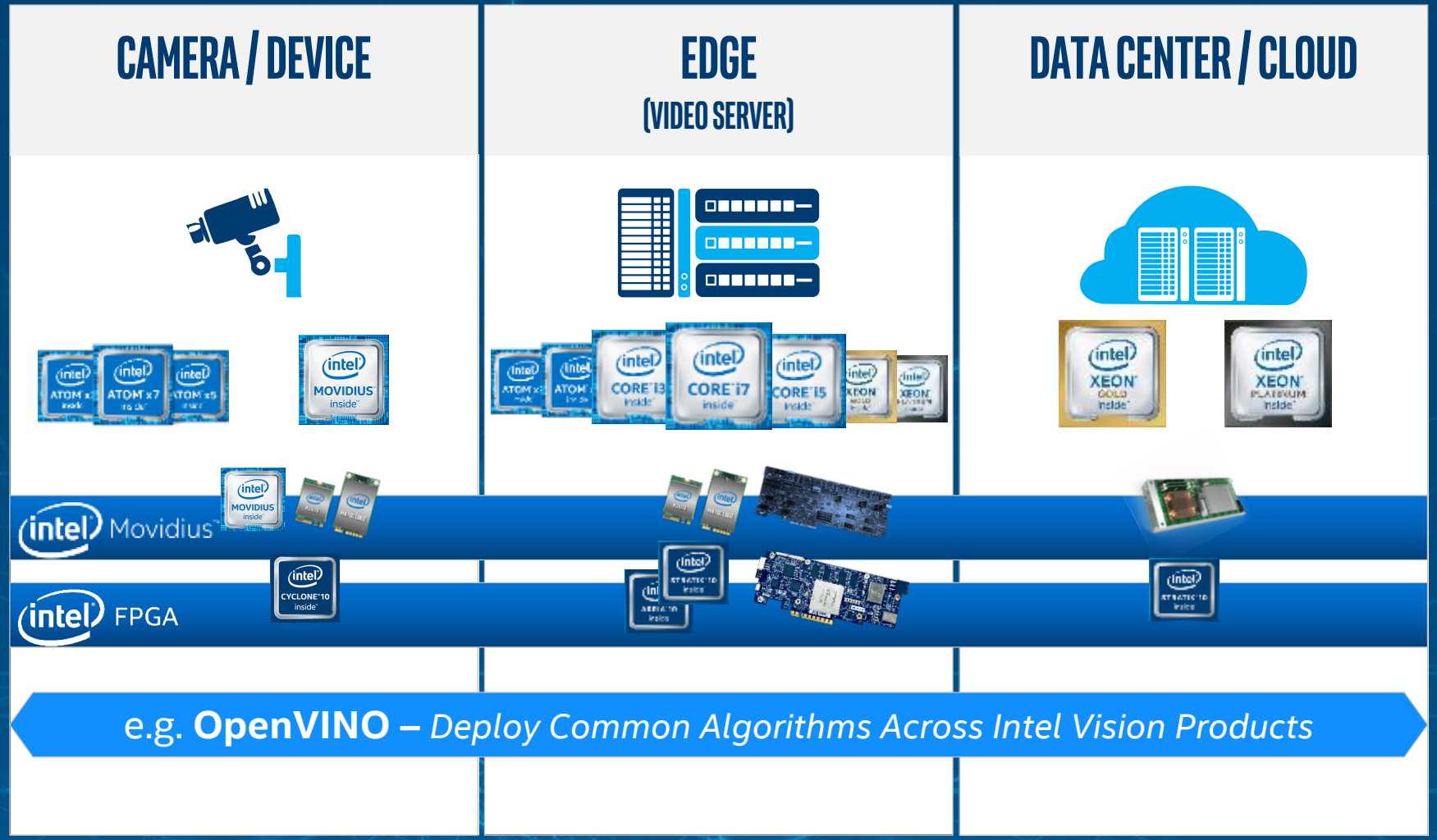
Host Processors

+

Discrete Accelerators

on

Unified SW Toolkits



INTEL® DEEP LEARNING DEPLOYMENT TOOLKIT

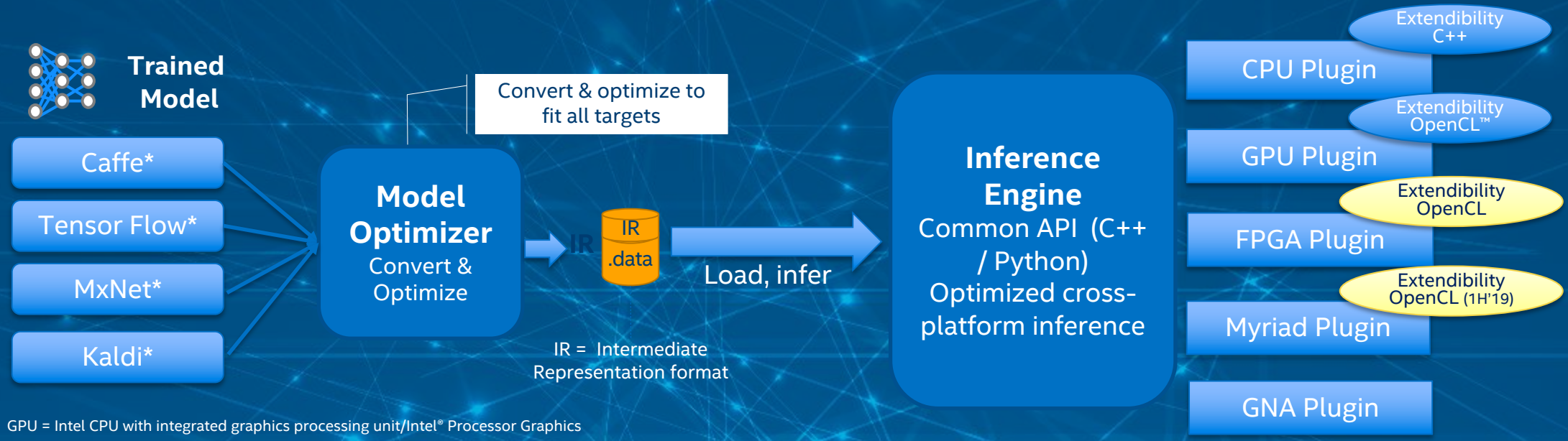
TAKE FULL ADVANTAGE OF THE POWER OF INTEL® ARCHITECTURE

Model Optimizer

- **What it is**: Preparation step -> imports trained models
- **Why important**: Optimizes for performance/space with conservative topology transformations; biggest boost is from conversion to data types matching hardware.

Inference Engine

- **What it is**: High-level inference API
- **Why important**: Interface is implemented as dynamically loaded plugins for each hardware type. Delivers best performance for each type without requiring users to implement and maintain multiple code pathways.



GPU = Intel CPU with integrated graphics processing unit/Intel® Processor Graphics

QUICKLY DEPLOY WITH PRE-BUILT PROJECTS
OPEN-SOURCED REFERENCE IMPLEMENTATIONS

Retail

Store Aisle Monitor
Capture video, generate a heat map, record the number of people present, and then integrate the results. The program can also create an output video and save snapshots.

Shopper Mood Monitor
Detect the mood of shoppers when looking at a retail or kiosk display.

Parking Lot Tracker
Receive or post information on available parking spaces by tracking how many vehicles enter and exit a parking lot.

Shopper Gaze Monitor
Build a solution to analyze customer expressions and reactions to product advertising collateral that is positioned on retail shelves.

Store Traffic Monitor
Monitor three different streams of video that count people inside and outside of a facility. This application also counts product inventory.

People Counter System
Create a smart video application using the Intel Distribution of OpenVINO toolkit. The toolkit uses models and inference to run single-class object detection.

Industrial

Object Size Detection
Use a video source such as a camera to capture images and validate an object's size.

Object Flaw Detector
Detect various anomalies of an object that is moving on a conveyor belt within a manufacturing facility, and then run analysis on what is detected.

Machine Operator Monitor
Send notifications when an employee appears to be distracted when operating machinery.

Motor Defect Detector
Predict performance issues with manufacturing equipment motors. Perform local or cloud analytics of the issues found, and then display the data on a user interface to determine when failures might arise.

Safety Gear Detector
For people working in hazardous conditions, wearing appropriate safety gear is critical. This solution observes workers as they pass in front of a camera, identifies them using facial recognition, and determines if they have adequate safety protection.

Restricted Zone Notifier
Secure work areas and send alerts if someone enters the restricted space.

Smart Cities

Industrial Anomaly Detection

Intruder Detector
Build an application that alerts you when someone enters a restricted area.

Facial Recognition Access Control
Develop a facial detection and recognition solution to grant access to designated areas.

INTEL® MOVIDIUS™ PRODUCT PORTFOLIO

	X1 Movidius	X2 Movidius	X8 Movidius
VPU	1 MA2485	2 MA2485	8 MA2485
Dimensions	M.2 2230 22mm x 30mm	M-PCIe* 30mm x 50mm	½ height PCIe* 69mm x 168mm
Minimum System Spec	Intel Atom® x7 E3950, 8GB DDR4, 64GB, USB 3.0, M.2 2230	Intel Atom® x7 E3950, 8GB DDR4, 64GB, USB 3.0, M-PCIe	Intel® Core™ i5 6500TE 8GB RAM, 500GB, USB3.0, 2 PCIe x4 x8 x16
Models	- AAEON* UP* AI Core X - Advantech* VEGA-320	- AAEON* UP* AI Core XM 2280 - ADLINK* EDL-mPCIe-MA2485 - Advantech* VEGA-330 - JWIPC* AI Accelerator Card - Uzel* UI-AL2	- IEI* Mustang-V100-MX8-R10 - NEXCOM* AIBooster-X8 - Uzel* UI-AR8

VISION ACCELERATION KIT + HDDL-R
POWERED BY IEI*

- Specialized Processors designed for high-perfromance machine vision at low power
- 8 VPUs x 16 SHAVE cores (Streaming Hybrid Architecture Vector Engine)
- Supports multiple network topologies (GoogLeNet, ResNet, etc.)
- Two Memory Banks (DDR4, 8GB in total)
- Preinstalled Intel® Distribution of OpenVINO™ toolkit
- Get started quickly with samples

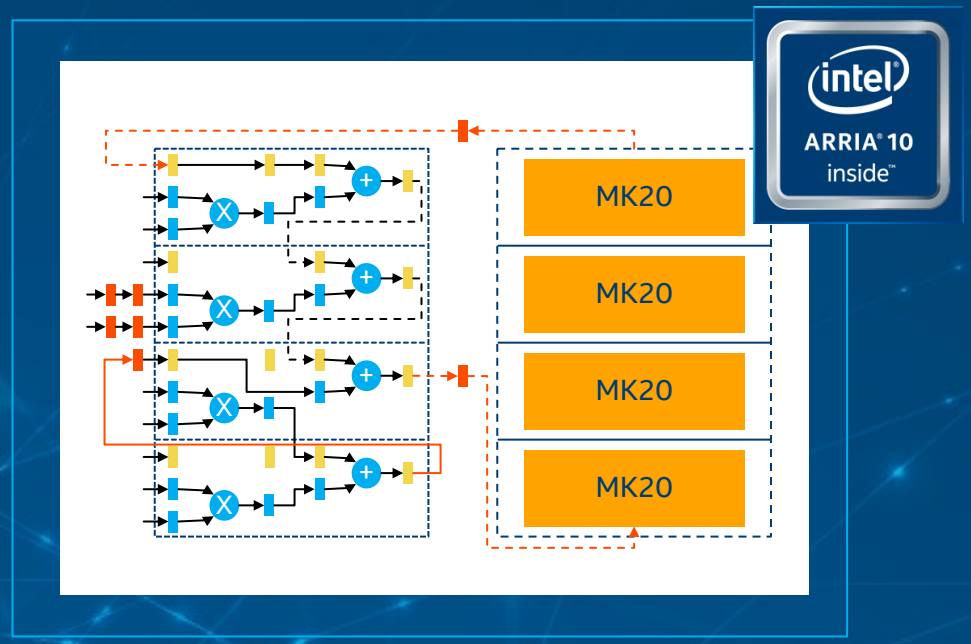
TANK-870-Q170

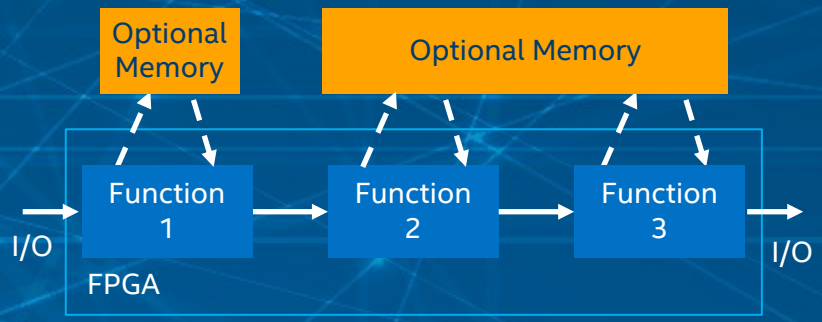
* Contact IEI Integration Corp. for pricing

INTEL® VISION ACCELERATOR HDDL-F

IEI * PCIE FPGA CARD

- Arria10 1150GX PCIe add-in card with OpenVINO support and bitstream update tool.
- Provides various precision options to support different performance target using the same hardware.
- Bitstream updates on a quarterly cadence for performance enhancements over time.





VISION ACCELERATION KIT + HDDL-F
POWERED BY IEI*

- Flexible, customizable processors designed to adapt to various display, video, and image processing workloads.
- Arria 10 GX 1150KLE FPGA
- Designed for high performance, low latency applications
- Supports multiple network topologies (GoogLeNet, ResNet, etc.)
- OpenCL BSP for accelerator customization
- Two Memory Banks (DDR4, 8GB in total)
- Passive or Active cooling
- ½ L, ½ H PCIe (Gen 3x8)

TANK-870-Q170

* Contact IEI Integration Corp. for pricing

DEEP LEARNING ENGINE DECISION TREE

Standard CPU?

Develop on CPU alone

Best for...

- *Convolutions*
- *Very small networks w/less weights*
- *Large input sizes (et. >720p)*
- *Operations with lots of sorting & matching*

Is this within parameters for performance, accuracy, allotted CPU utilization?

- Yes
- No

Available CPU cycles on pre-deep learning workload:

- <50% CPU with OpenVINO
- >50% of CPU

Higher-performance CPU or GPU?

Develop on higher performance CPU

Would a faster CPU provide the performance, accuracy, and allotted CPU utilization?

- Yes
- No

or

Is there an integrated GPU that would meet performance?

- No
- Yes

Develop on integrated GPU

HDDL-F, R, S, L Accelerator?

Accelerator needed

Determine whether to develop on HDDL-R/L/S series, or F-series based on characteristics like...

- *Network or "my network is most like..."*
- *TOPS required per stream/workload*
- *Memory size*
- *Image/input size*
- *Batch size*
- *Network precision*

HDDL-R, L and S

- *Batch Size: >1*
- ***Network Memory Size < 250MParams***
- ***Smaller image size***
- *Precision: FP16*
- *Accelerator Power Budget: 2-25W*
- *# of streams: 1-16*
- ***Networks Like:***
  - ***GoogLeNet v1/v2***
  - ***TinyYolo v1/v2***
  - ***ResNet-18, etc...***

Determine whether to develop on HDDL-R, L, or S based on...

- *Streams*
- *Power budget*
- *Memory footprint*

Develop on HDDL-R

Develop on HDDL-S

Develop on HDDL-L

Develop on HDDL-F

- *Batch Size: >1*
- *Network Memory Size: >250MParams*
- *Image size: up to FHD*
- *Precision: FP16/11/9*
- *Accelerator Power Budget: <50W*
- *# of streams: 3-15*
- *Networks Like:*
  - *VGG*
  - *Resnet-50*
  - *SqueezeNet*

All Green Boxes Use OpenVINO to optimize