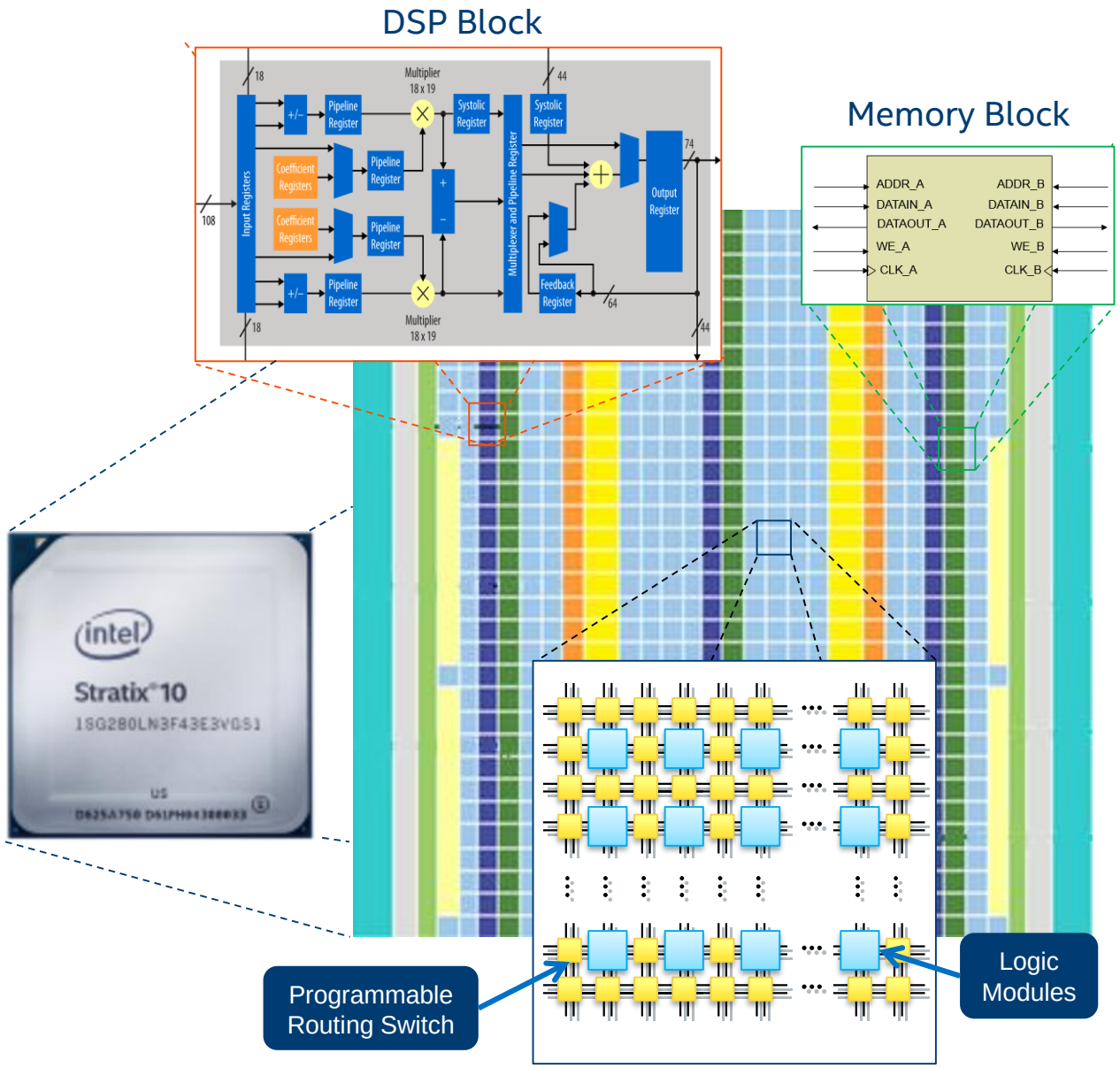
ACCELERATORS BASED ON INTEL® ARRIA® FPGA

FPGA OVERVIEW

- Field Programmable Gate Array (FPGA)
  - Millions of logic elements
  - Thousands of embedded memory blocks
  - Thousands of DSP blocks
  - Programmable routing
  - High speed transceivers
  - Various built-in hardened IP
- Used to create **Custom Hardware!**



INSIDE INTEL® DISTRIBUTION OF OPENVINO™ TOOLKIT

Deep Learning

Intel® Deep Learning Deployment Toolkit

Model Optimizer
Convert & Optimize

IR

Inference Engine
Optimized Inference
+ samples

IR = Intermediate Representation file

Open Model Zoo

- Intel & Public Pretrained Models
- Demos
- Model Downloader
- Accuracy Checker

- Deployment Manager
- Post Training Optimization Toolkit
- Benchmark App
- DL Workbench
- DL Streamer

Traditional Computer Vision

- OpenCV*
- Samples

For Intel® CPU & GPU/Intel® Processor Graphics

Tools & Libraries

Increase Media/Video/Graphics Performance

- Intel® Media SDK
  Open Source version
- OpenCL™ Drivers & Runtimes

For GPU/Intel® Processor Graphics

Optimize Intel® FPGA (Windows & Linux)

- FPGA RunTime Environment
  (from Intel® FPGA SDK for OpenCL™)
- Bitstreams

OS Support: CentOS* 7.4 (64 bit), Ubuntu* 16.04.3 LTS (64 bit), Microsoft Windows® 10 (64 bit), Yocto Project* version Poky Jethro v2.0.3 (64 bit), macOS* 10.13 & 10.14 (64 bit)

Intel® Architecture-Based Platforms Support

Intel® Vision Accelerator Design Products & AI in Production/ Developer Kits

An open source version is available at 01.org/openvinotoolkit (some deep learning functions support Intel CPU/GPU only).

INTEL® DISTRIBUTION OF OPENVINO™ TOOLKIT WITH DLA USER FLOWS

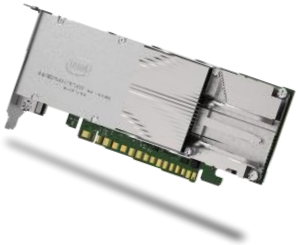
Application Developer

Design

Intel® Distribution of OpenVINO™ toolkit

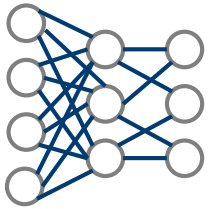
Run

Program



Turnkey Software Deployment Flow

Neural Net



Bitstream Library

Choose from many **precompiled** FPGA Images

Or

Custom create FPGA bitstream

FPGA Architecture Development Flow

IP Architect

Design

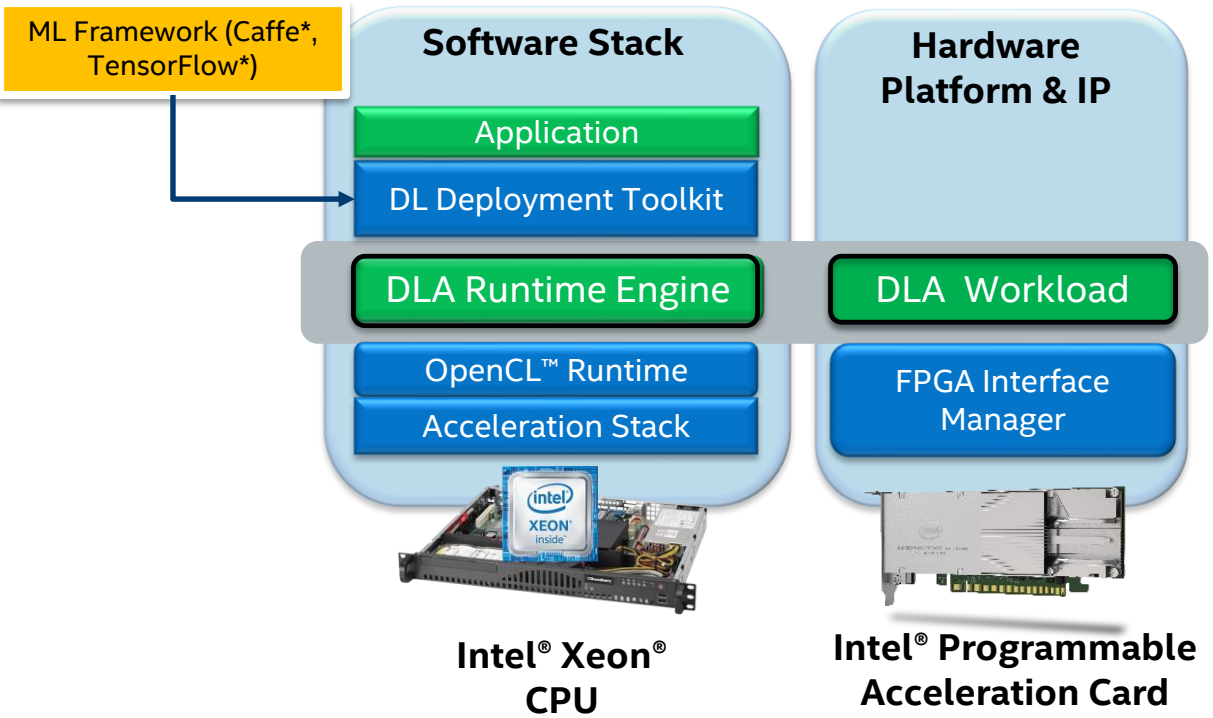
User Customization of DLA Suite Source Code

Compile

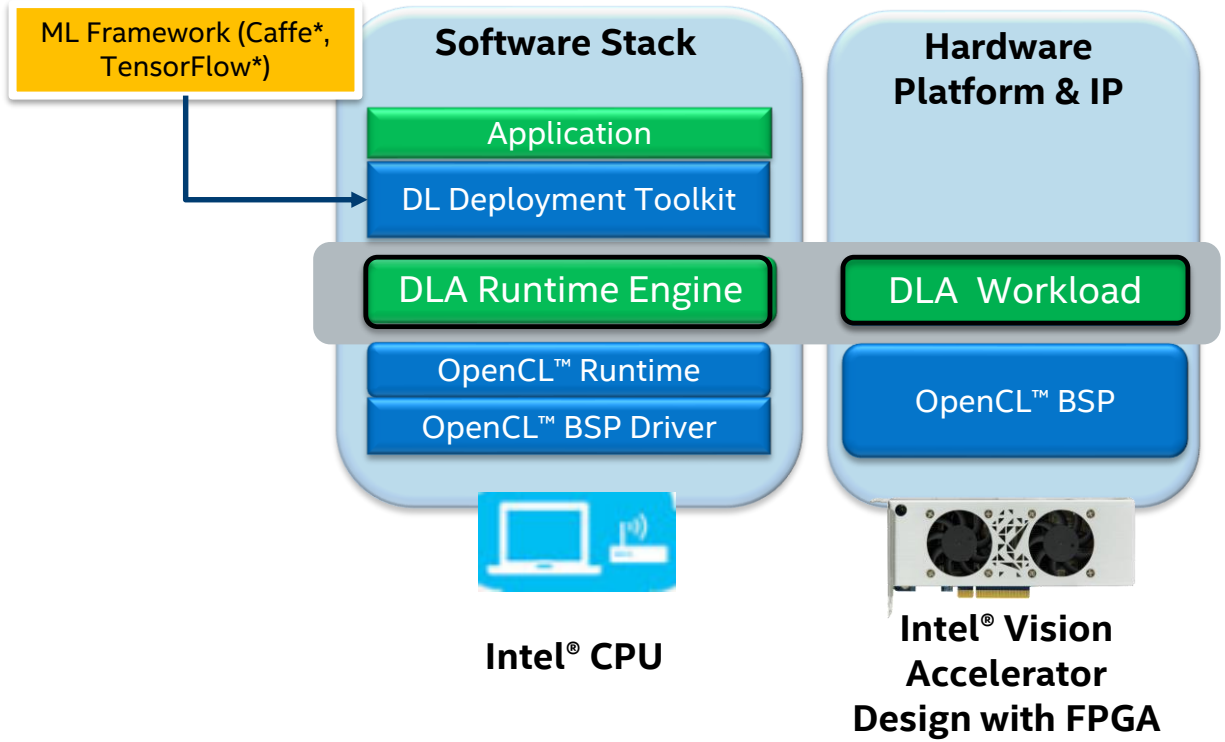
Intel® FPGA SDK for OpenCL™

MACHINE LEARNING ON INTEL® FPGA PLATFORM

Acceleration Stack Platform Solution



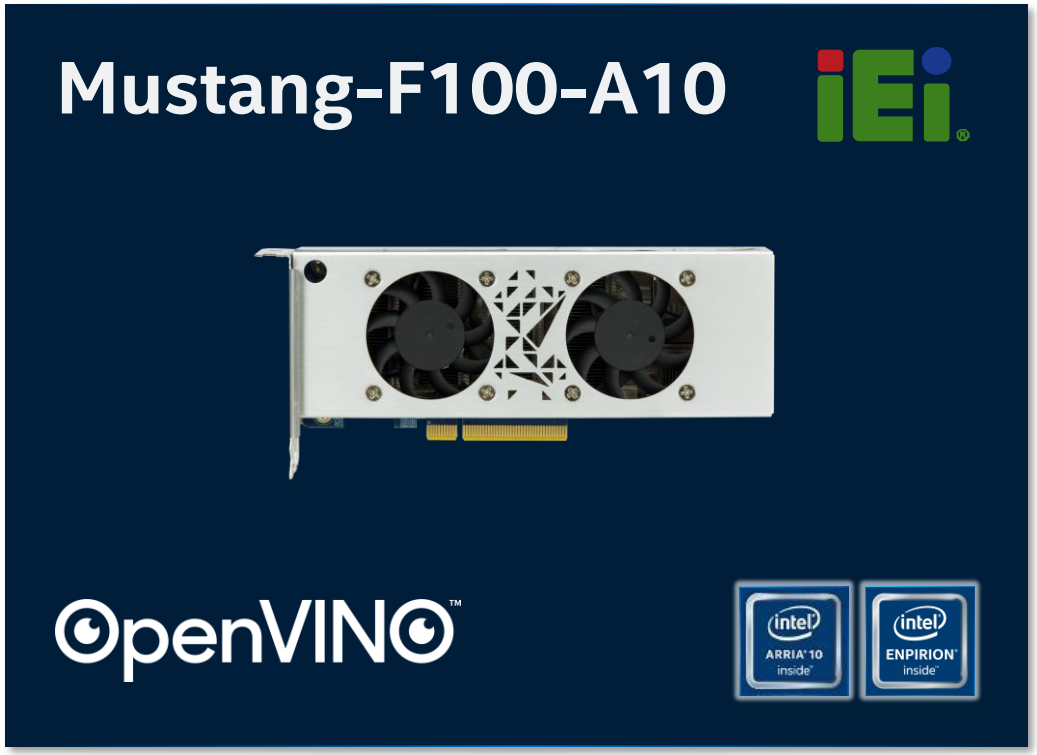
Edge Computing Solution



[Intel® FPGA Acceleration Hub]

[Intel® Vision Accelerator Design Products]

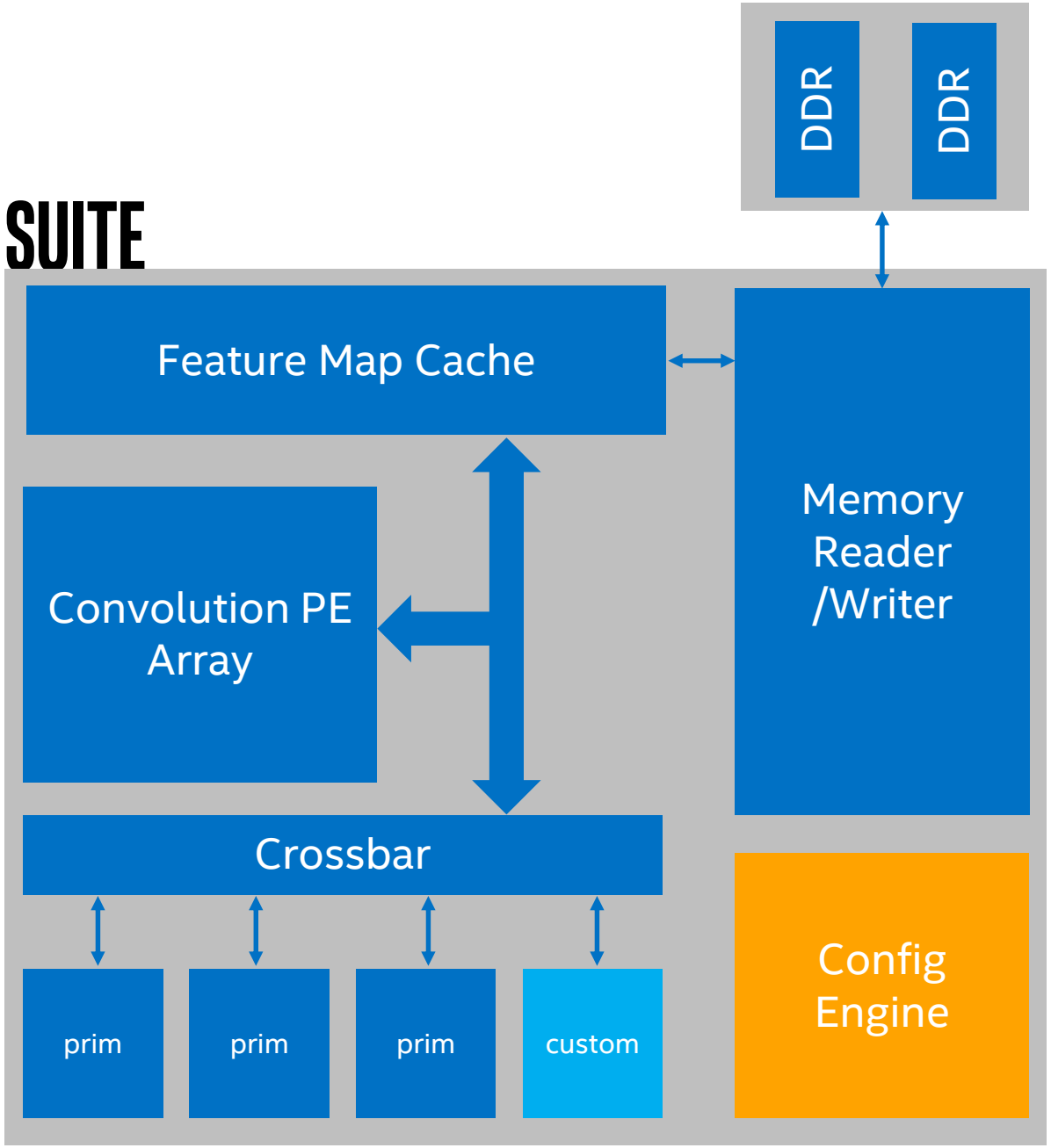
INTEL® VISION ACCELERATION DESIGN WITH INTEL® ARRIA® 10 FPGA
KEY DIFFERENTIATORS



- High performance, low latency
- Flexibility to adapt to new, evolving, and custom networks
- Supports large image sizes (e.g., 4K)
- Large networks (up to 4 billion parameters)
- Wide ambient temperature range (0° C to 65° C)
- 24/7/365 operation
- Long lifespan (8–10 years)

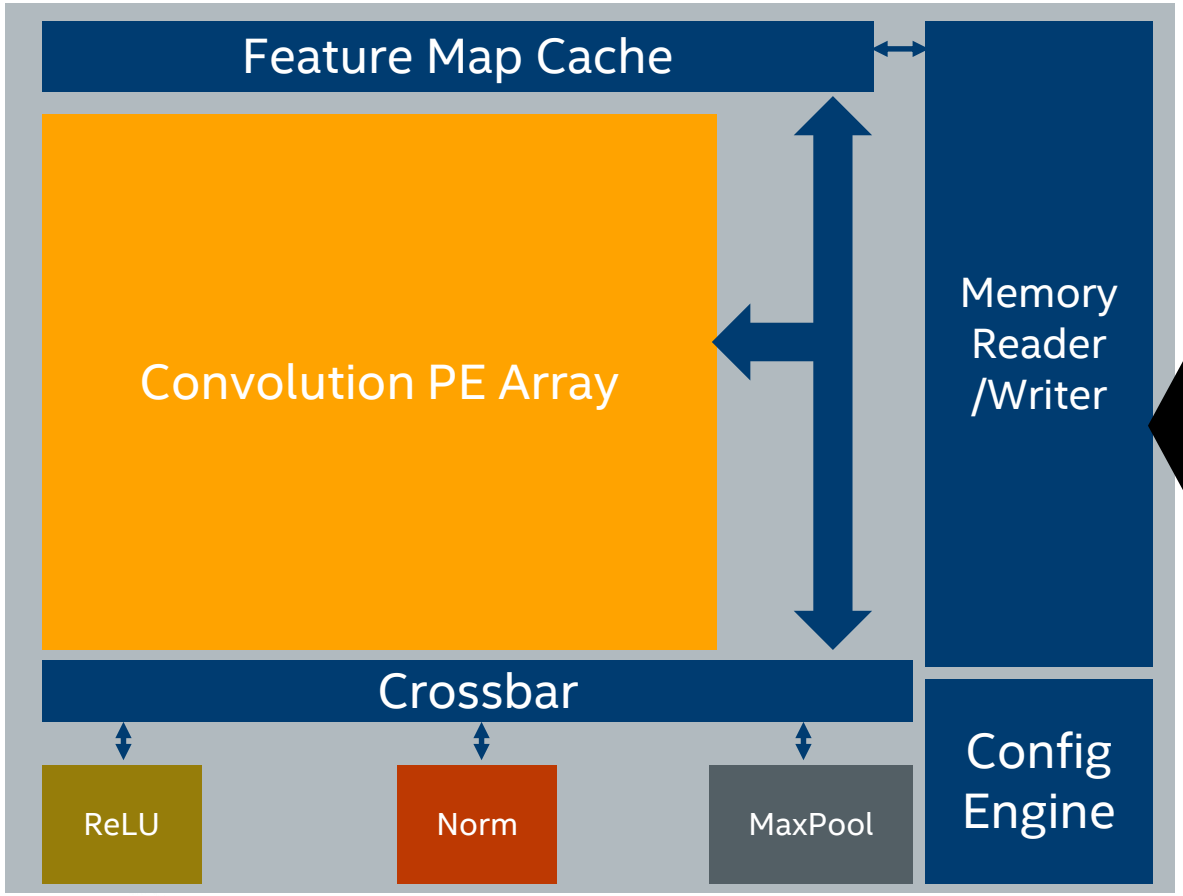
INTEL® FPGA DEEP LEARNING ACCELERATION SUITE

- CNN inference acceleration engine for topologies executed in a graph loop architecture
  - AlexNet, GoogleNet, SqueezeNet, VGG, ResNet*, MobileNet*, Yolo, SSD, …
- Software Deployment
  - No FPGA compile required
  - Run-time reconfigurable
- Customized Hardware Development
  - **Custom architecture creation w/ parameters**
  - Custom primitives using OpenCL™ flow

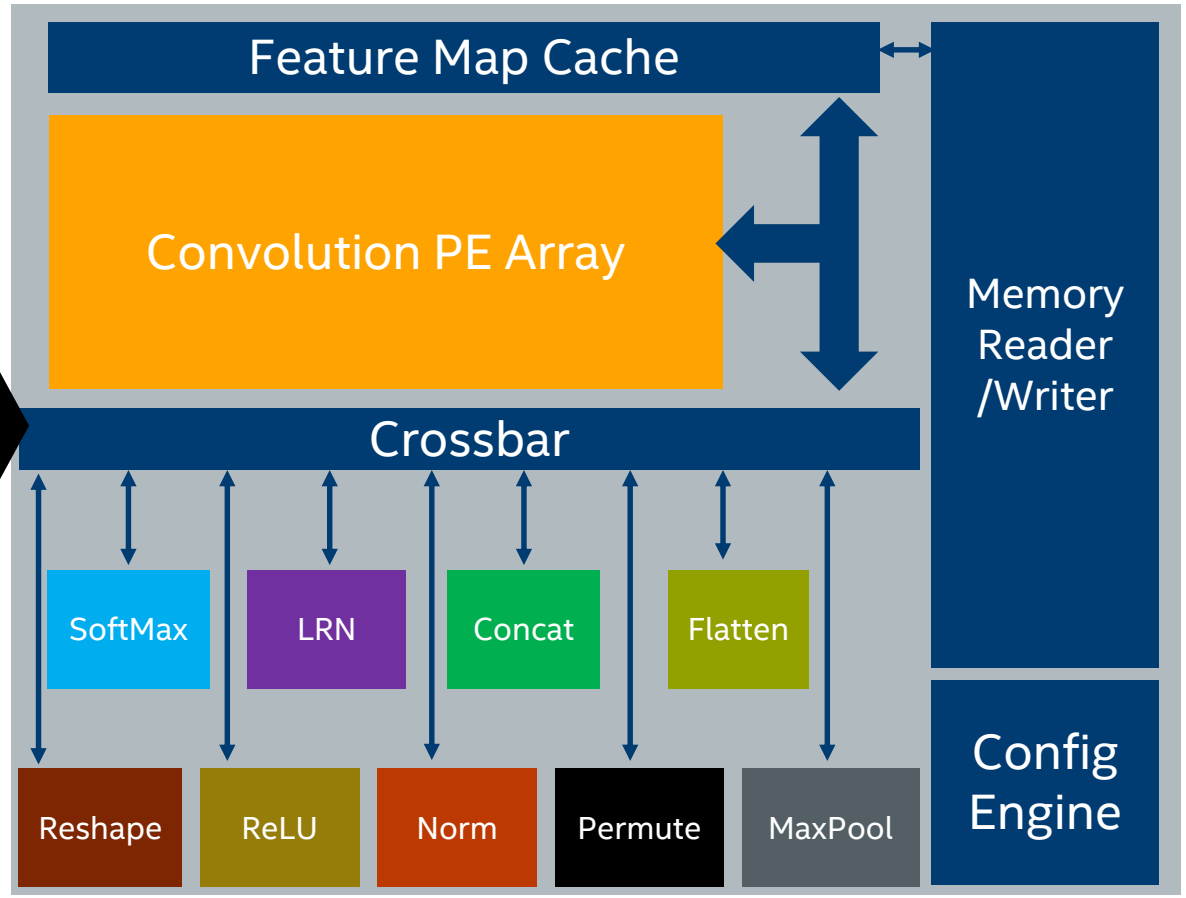


SUPPORT FOR DIFFERENT TOPOLOGIES

Adapts to support new or evolving networks

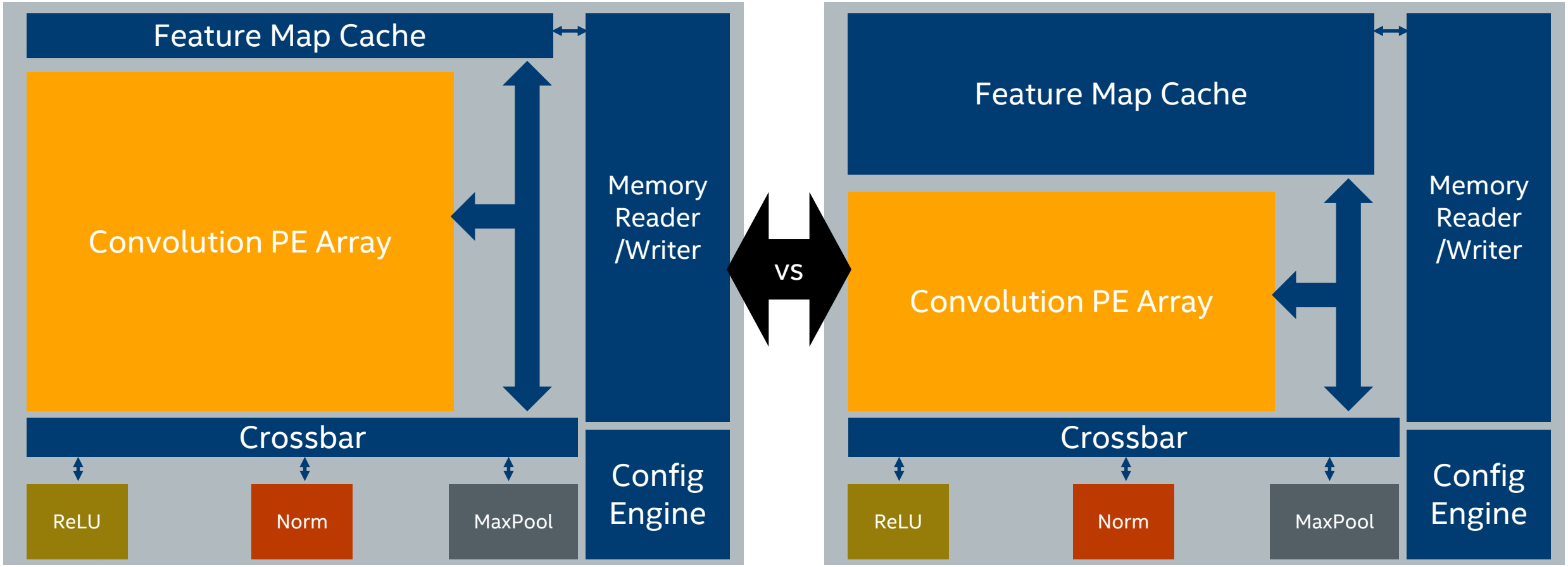


vs



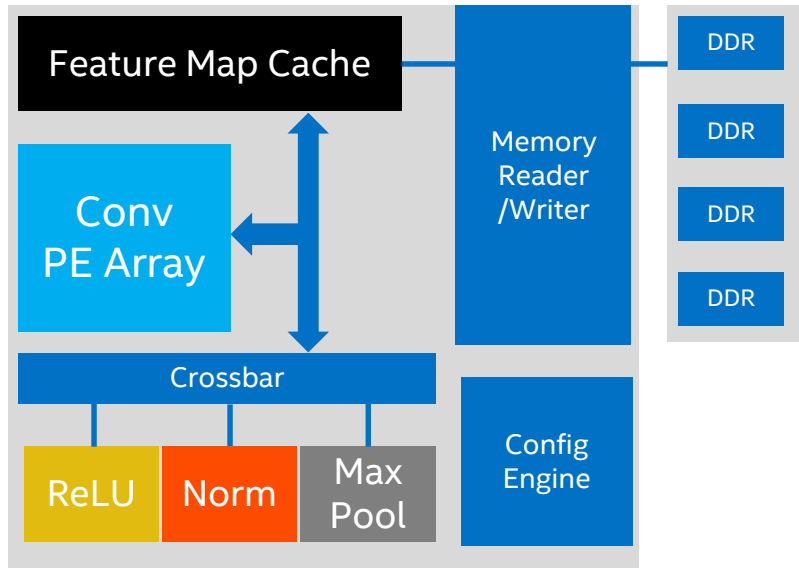
OPTIMIZE FOR BEST PERFORMANCE

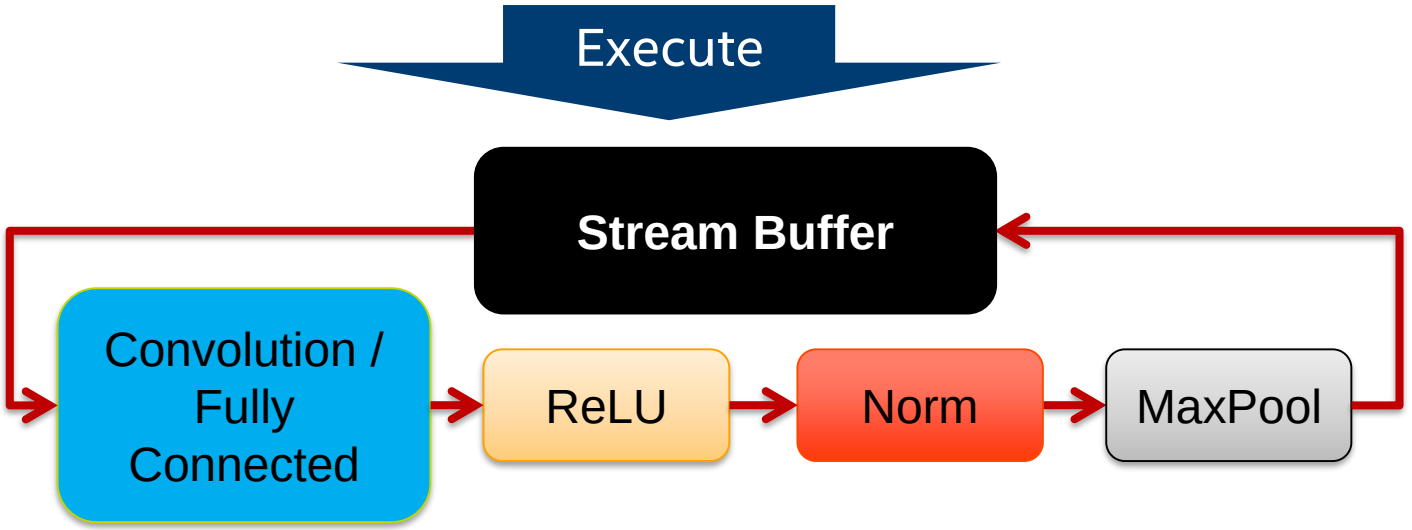
Tradeoff between size of Feature Map cache and convolutional PE array



DLA ARCHITECTURE: BUILT FOR PERFORMANCE

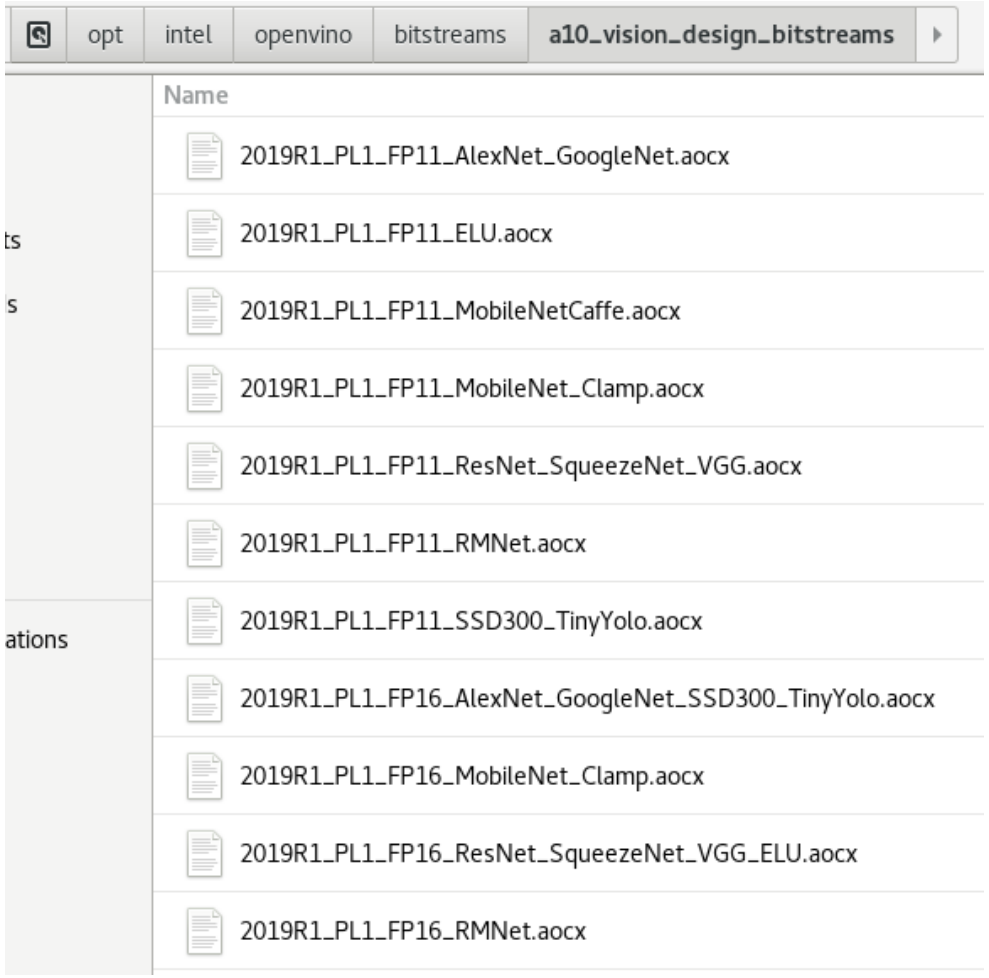
- Maximize Parallelism on the FPGA
  - Filter Parallelism (Processing Elements)
  - Input-Depth Parallelism
  - Winograd Transformation
  - Batching
  - Feature Stream Buffer
  - Filter Cache
- Choosing FPGA Bitstream
  - Data Type / Design Exploration
  - Primitive Support





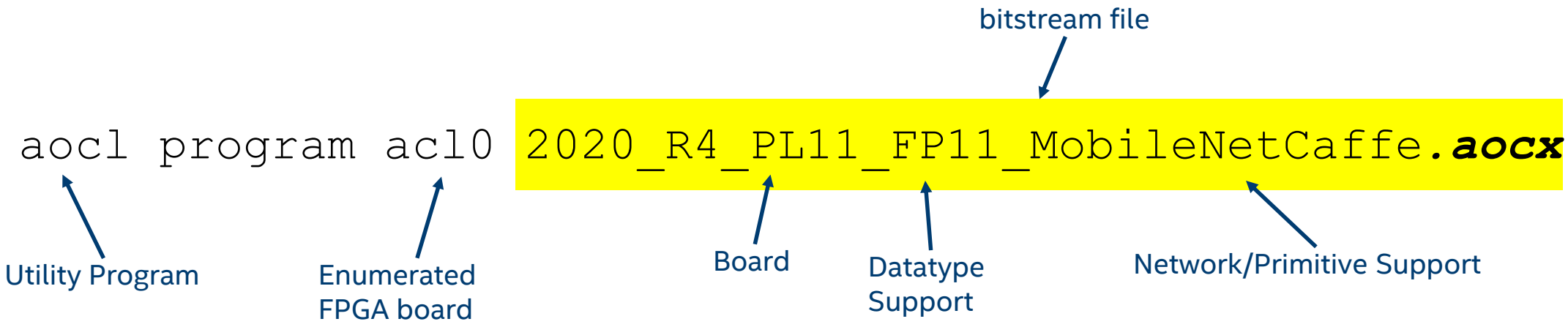
DLA ARCHITECTURE SELECTION

- Intel® Distribution of OpenVINO™ toolkit ships with many FPGA images for various boards/data types/topologies
  - <version>_<board>_<data type>_<Topologies/Feature>.aocx
- Find ideal FPGA image that meets your needs
- Check documentation for list of FPGA images and supported topologies
  - https://docs.openvinotoolkit.org/latest/_docs_IE_DG_supported_plugins_FPGA.html
- Example: ResNet* focused image does not have Norm (better performance)



LOAD SELECTED BITSTEAM PRIOR TO EXECUTION

- Program the FPGA with the selected FPGA bitstream



INTEL® DISTRIBUTION OF OPENVINO™ TOOLKIT FOR INTEL® VISION ACCELERATOR DESIGN WITH AN INTEL® ARRIA® 10 FPGA AND THE INTEL® PROGRAMMABLE ACCELERATION CARD WITH INTEL® ARRIA® 10 GX FPGA SUPPORT CHANGE

Intel will be transitioning to the next-generation programmable deep-learning solution based on FPGAs in order to increase the level of customization possible in FPGA deep-learning.

As part of this transition, future standard releases (i.e., non-LTS releases) of Intel® Distribution of OpenVINO™ toolkit will no longer include the Intel® Vision Accelerator Design with an Intel® Arria® 10 FPGA and the Intel® Programmable Acceleration Card with Intel® Arria® 10 GX FPGA.

Intel® Distribution of OpenVINO™ toolkit 2020.3.X LTS release will continue to support Intel® Vision Accelerator Design with an Intel® Arria® 10 FPGA and the Intel® Programmable Acceleration Card with Intel® Arria® 10 GX FPGA.