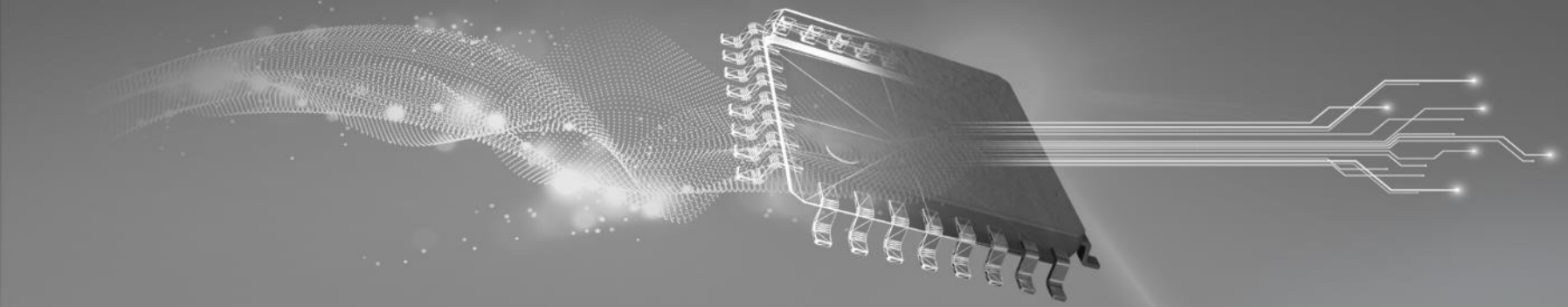
TI TECH DAYS



Deep Learning with Jacinto™ 7 SoCs: TDA4x

Tarkesh Pande

ADAS Business Unit

Agenda

- Introduction to Jacinto™ 7 processors
- Three step process
  - Network selection
  - Optimization
  - Deployment
- TI deep learning tools
- Open source frameworks

Introducing – Jacinto 7 processors

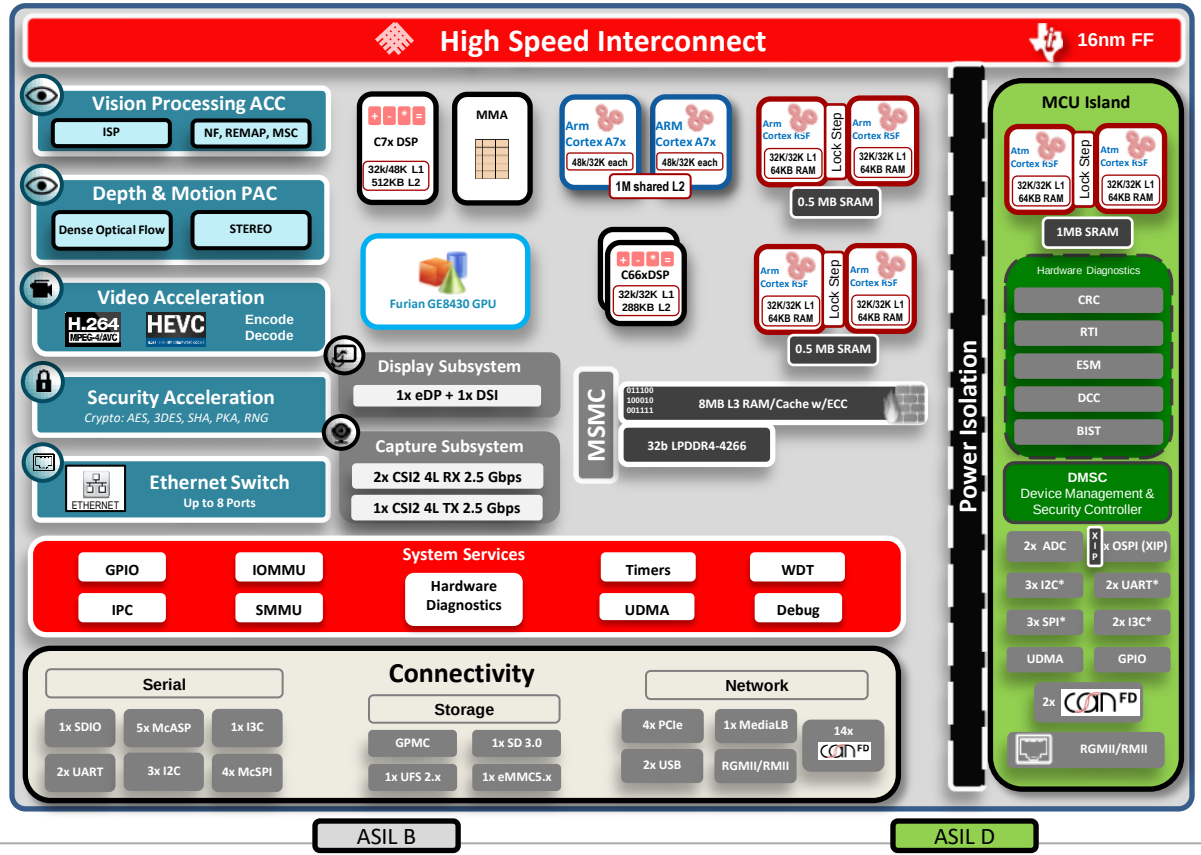
Accelerating key functions lowers power

- DSP for Computer Vision
- Vision Processing
- Video, Graphics
- Deep Learning

Industry's most efficient DL architecture

- Enables passively-cooling designs
- 90% utilization of deep learning accelerator due to smart memory system

Automotive Quality-ready process technology



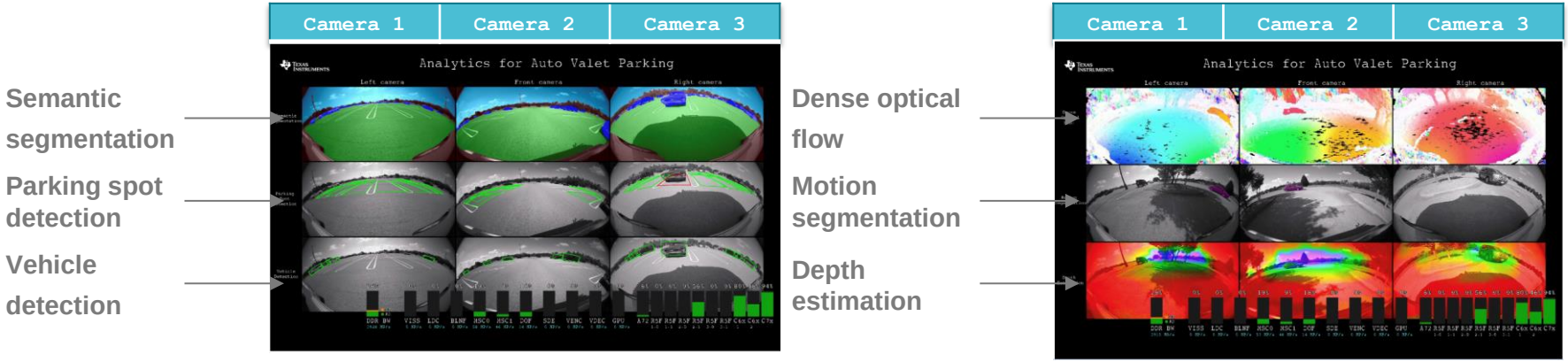
Deep learning system performance on TDA4VM

5 simultaneous deep learning algorithms on 3x 1MP camera each @ ~16 fps

- Parking spot detection
- Vehicle detection
- Semantic segmentation
- Motion segmentation
- Depth estimation

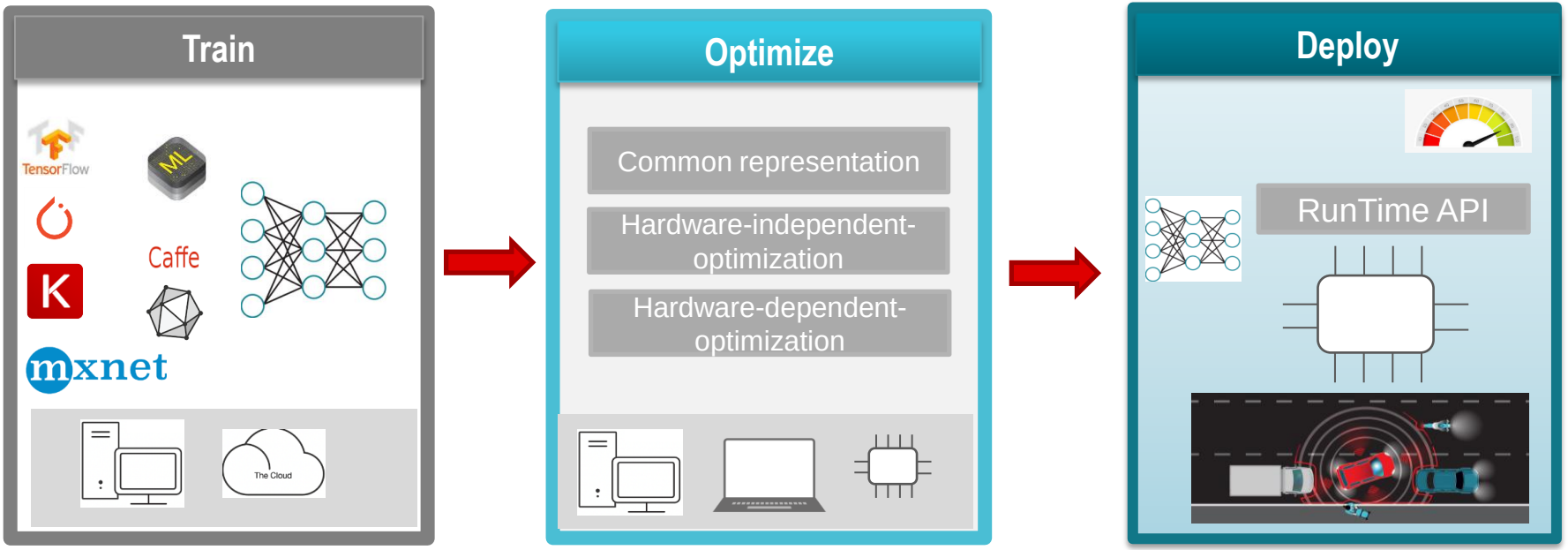
Resource loading

A72:	6%	**C7x+MMA:**	94%
		A:	
1xR5:	56%	**DOF:**	18%
5xR5s:	free	**DDR BW:**	26%



Deep learning inference at the edge

TI Deep Learning (TIDL) library provides custom tools for high-performance fixed-point inference on TDA4x hardware accelerators

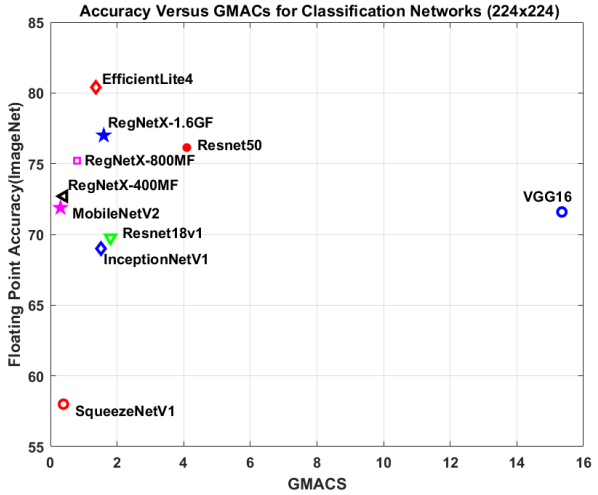


Training and network selection

- Choose and train network architecture that meet metrics
  - Accuracy and performance
  - Complexity is proportional to performance (time for execution)
- New architectures and ideas evolve quickly
  - Want to minimize time for test and deployment

Training and network selection

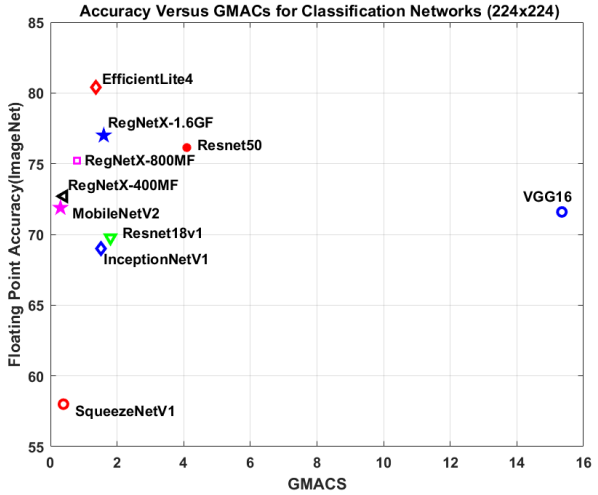
- Choose and train network architecture that meet metrics
  - Accuracy and performance
  - Complexity is proportional to performance (time for execution)
- New architectures and ideas evolve quickly
  - Want to minimize time for test and deployment



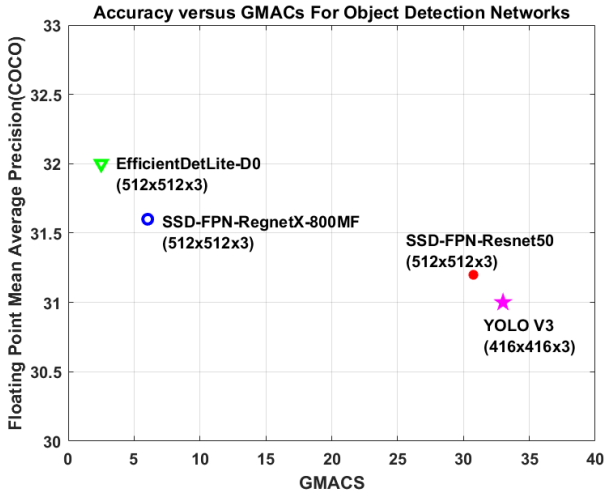
Classification

Training and network selection

- Choose and train network architecture that meet metrics
  - Accuracy and performance
  - Complexity is proportional to performance (time for execution)
- New architectures and ideas evolve quickly
  - Want to minimize time for test and deployment

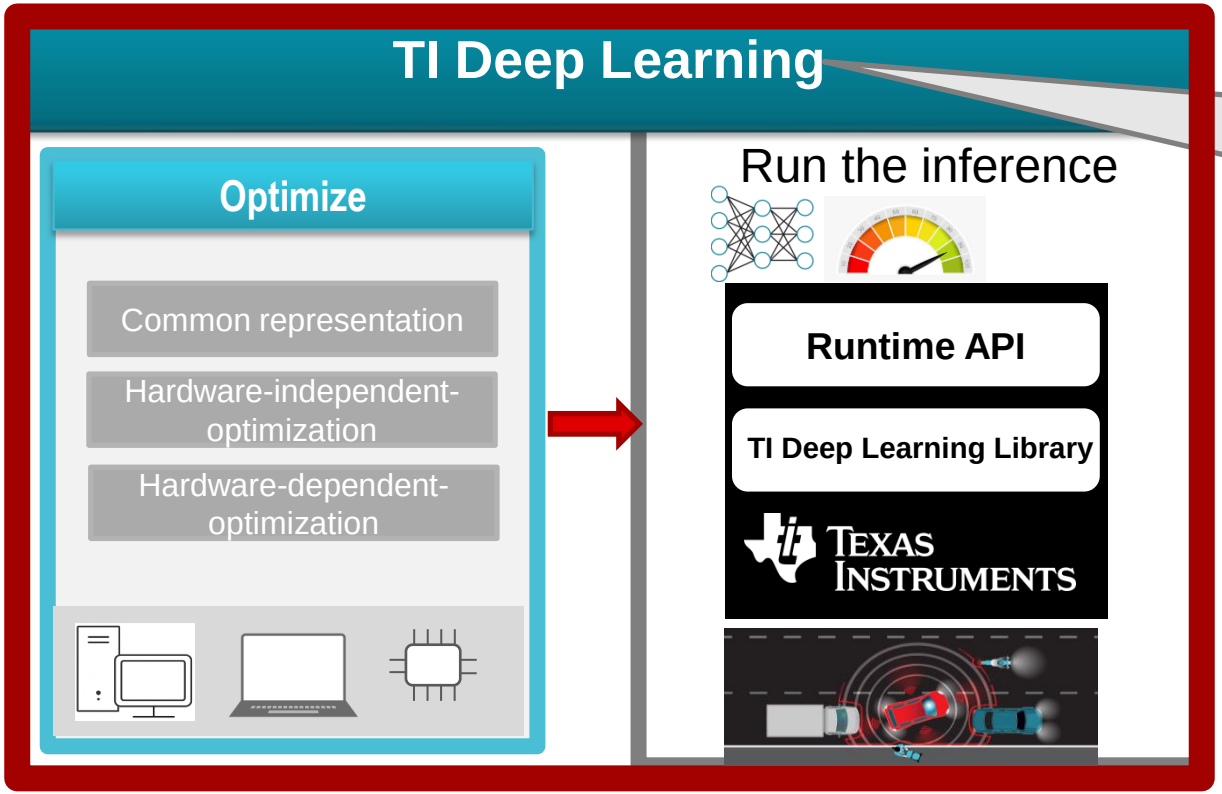


Classification



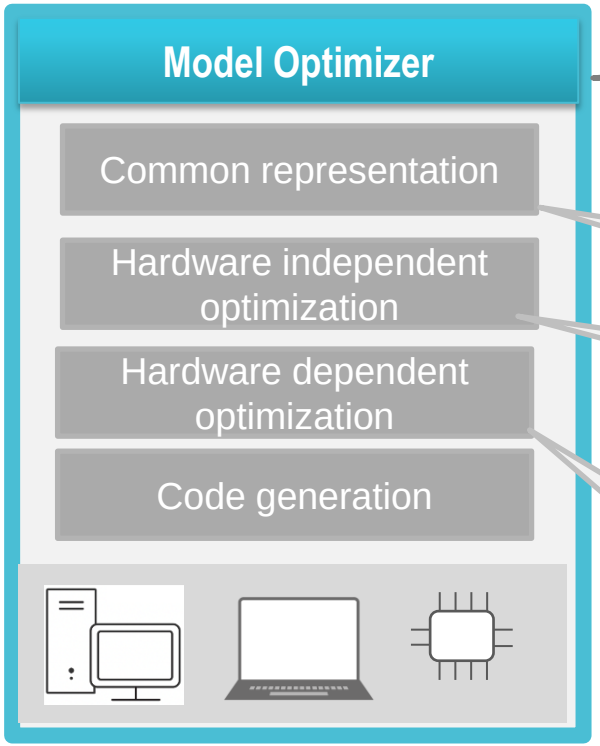
Object Detection

TI deep learning (TIDL)



- Enable high-performance inference at the edge on TI SoC

TI deep learning (TIDL) - model optimizer



PC based tools for :

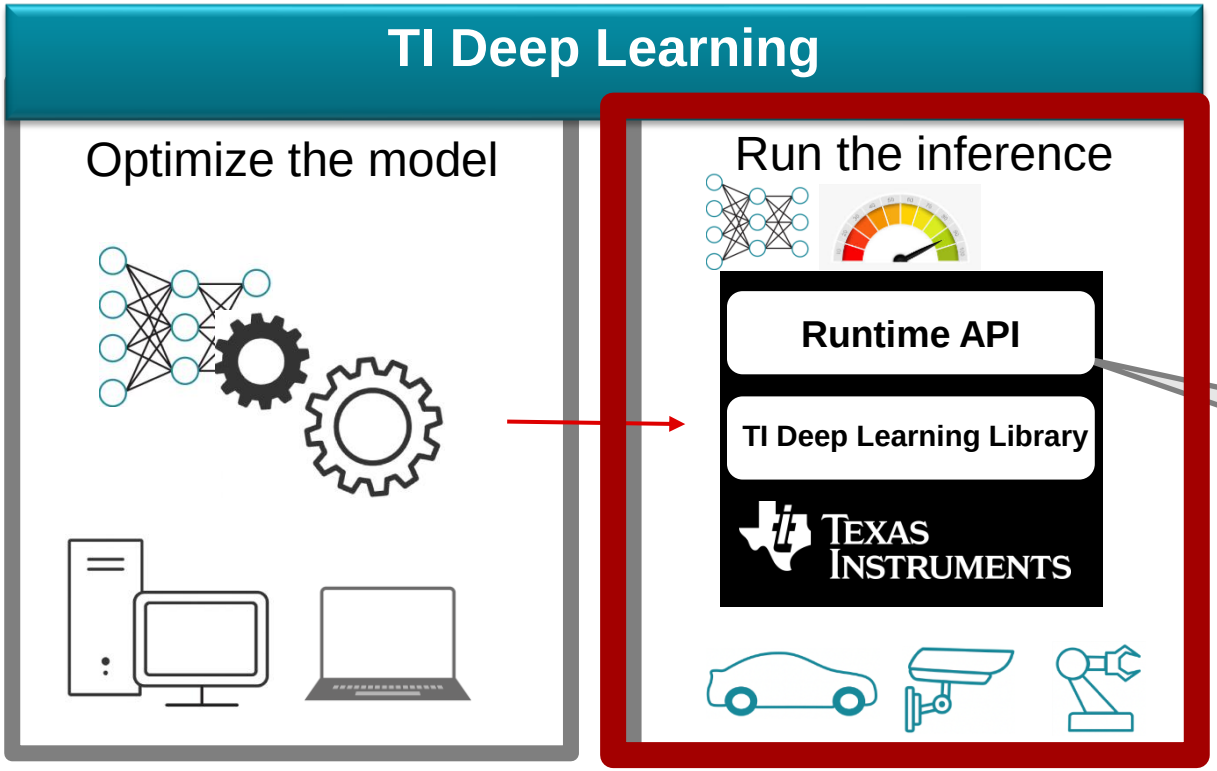
- Model optimizations
- Performance estimation
- Accuracy benchmarking

TensorFlow · K · mxnet · Caffe · TensorFlowLite

- Quantization
- Fusion
- Constant folding
- simplifications, dead code elimination
- Layout transformation

- Block based processing
- Vectorization
- Memory management

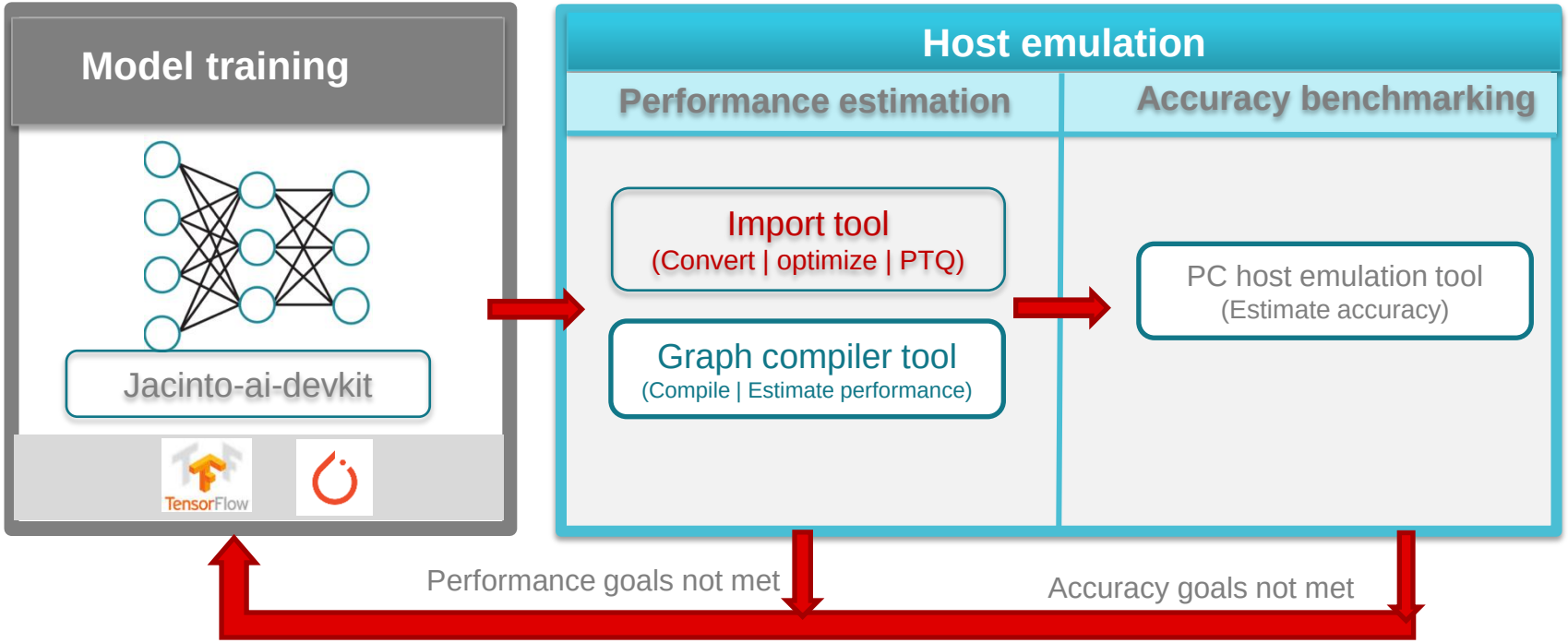
TI deep learning (TIDL)



Comprise of:

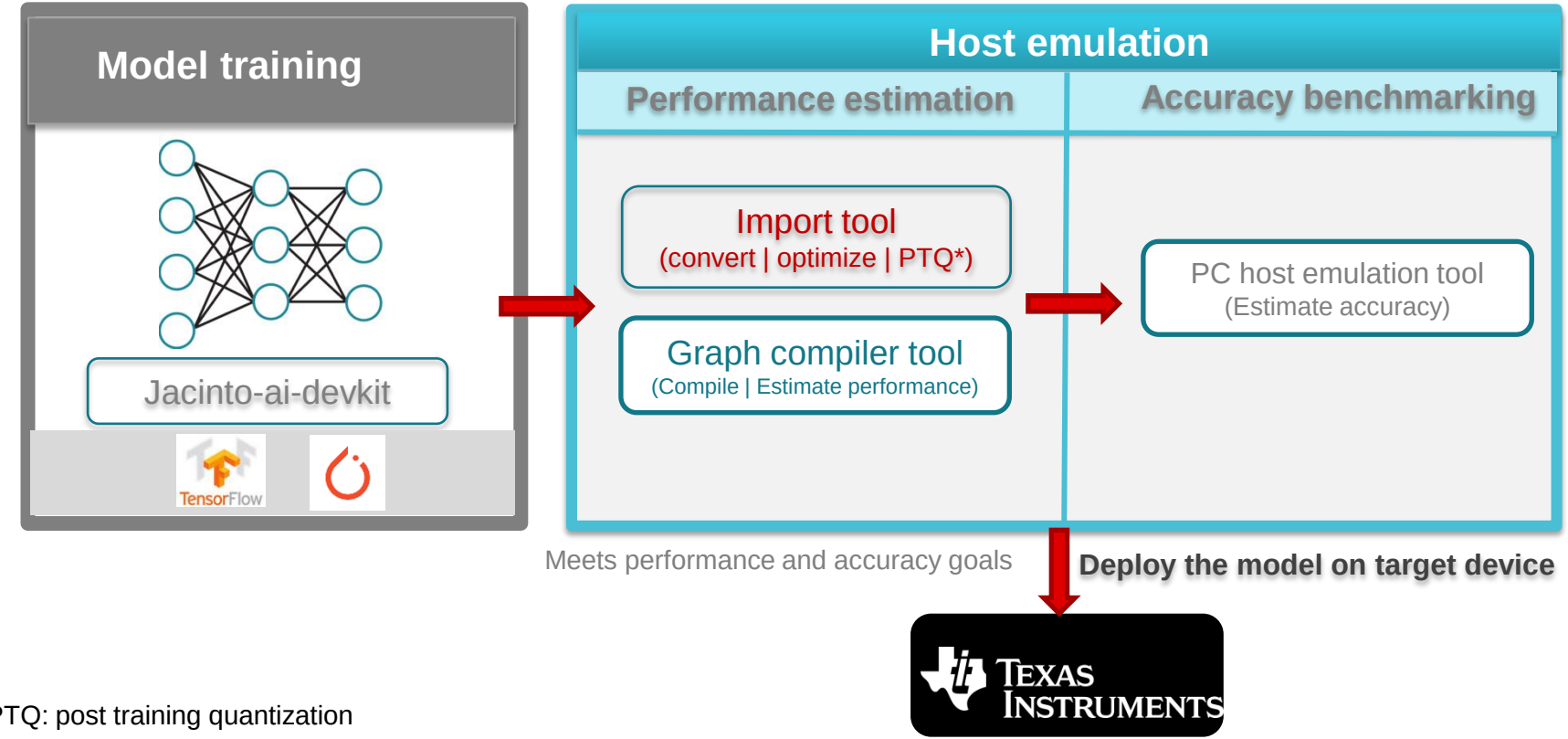
- PC based tools for model optimizations and performance estimation
- Runtime API and hand-optimized library for efficient inference execution on TI SoC

Network model development flow using TIDL tools



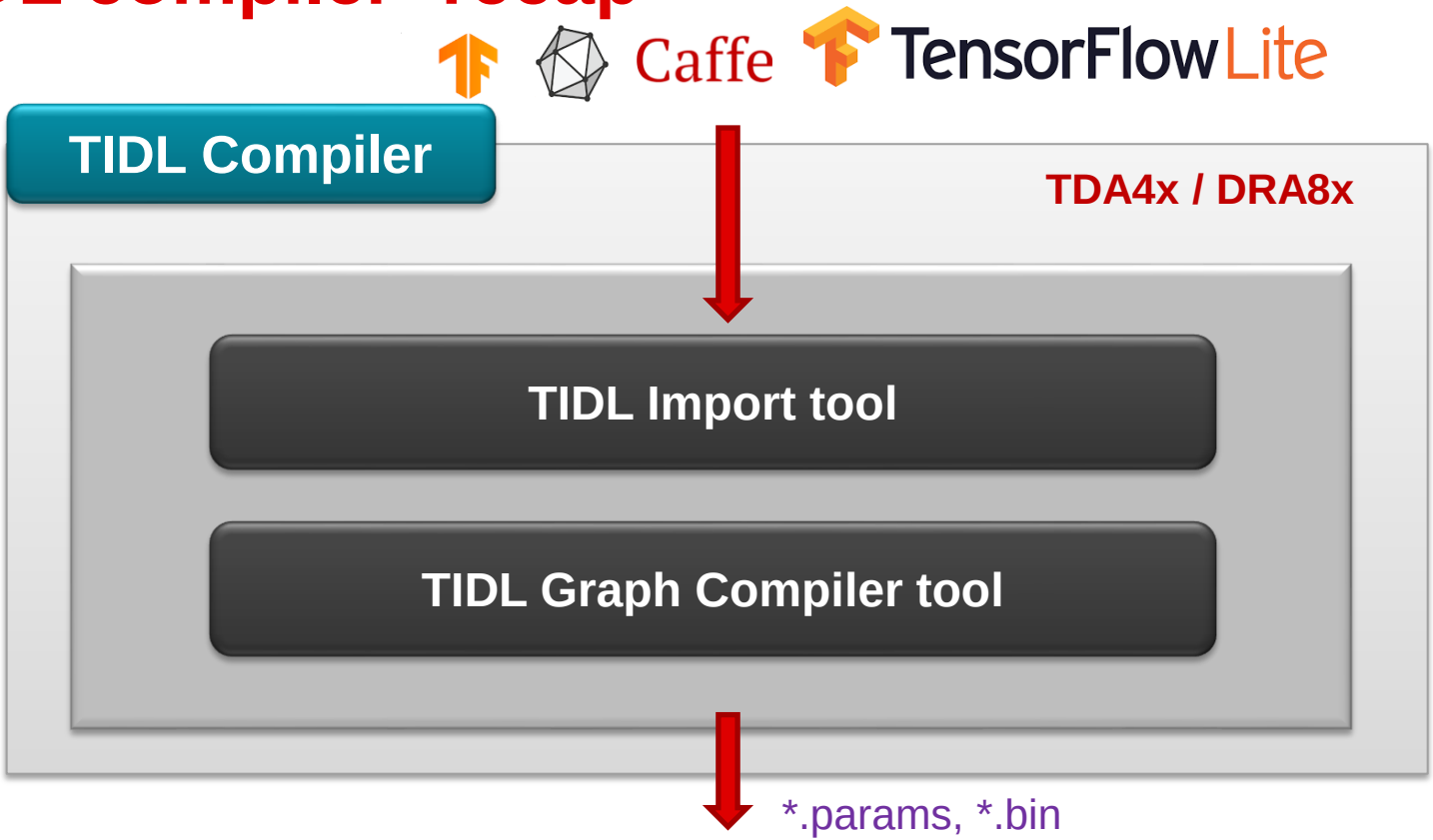
*PTQ: post training quantization

Network model development flow using TIDL tools

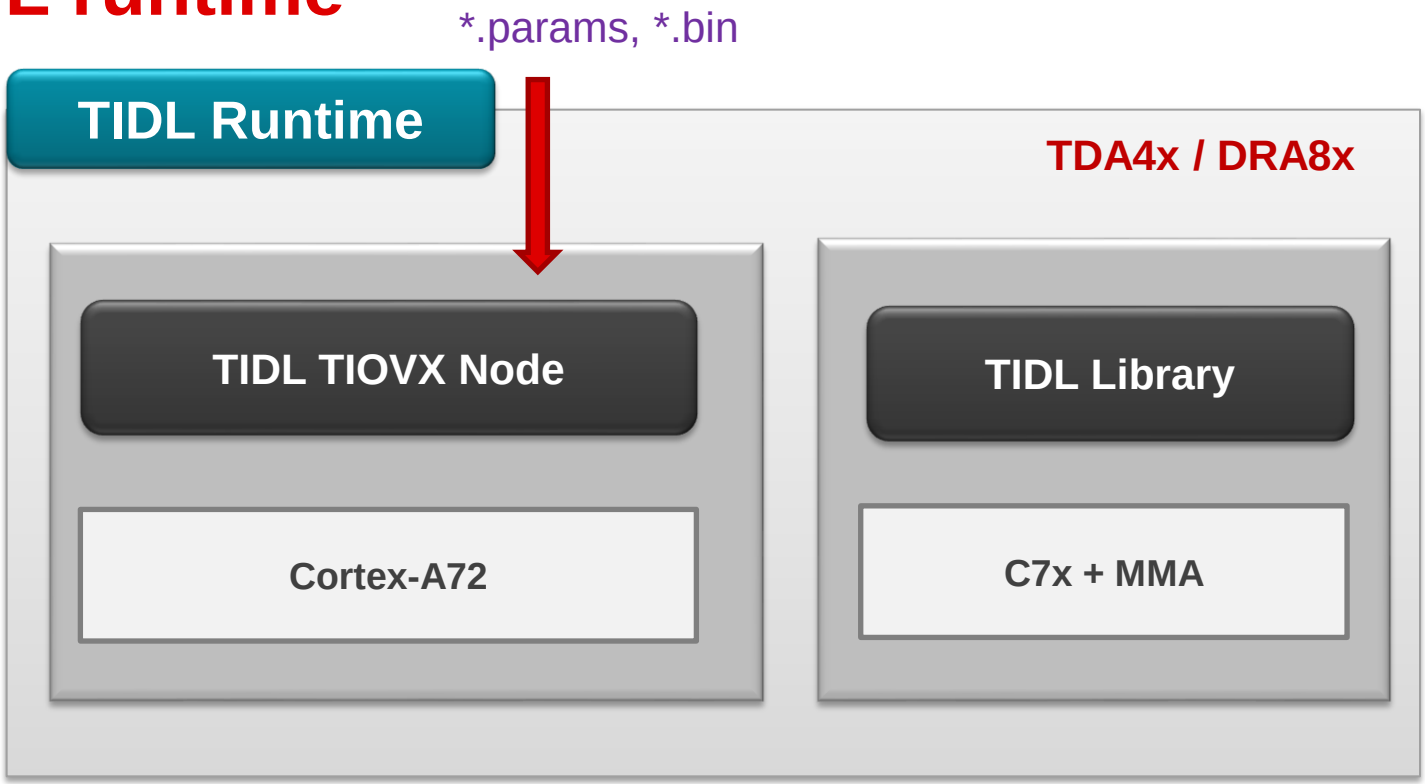


*PTQ: post training quantization

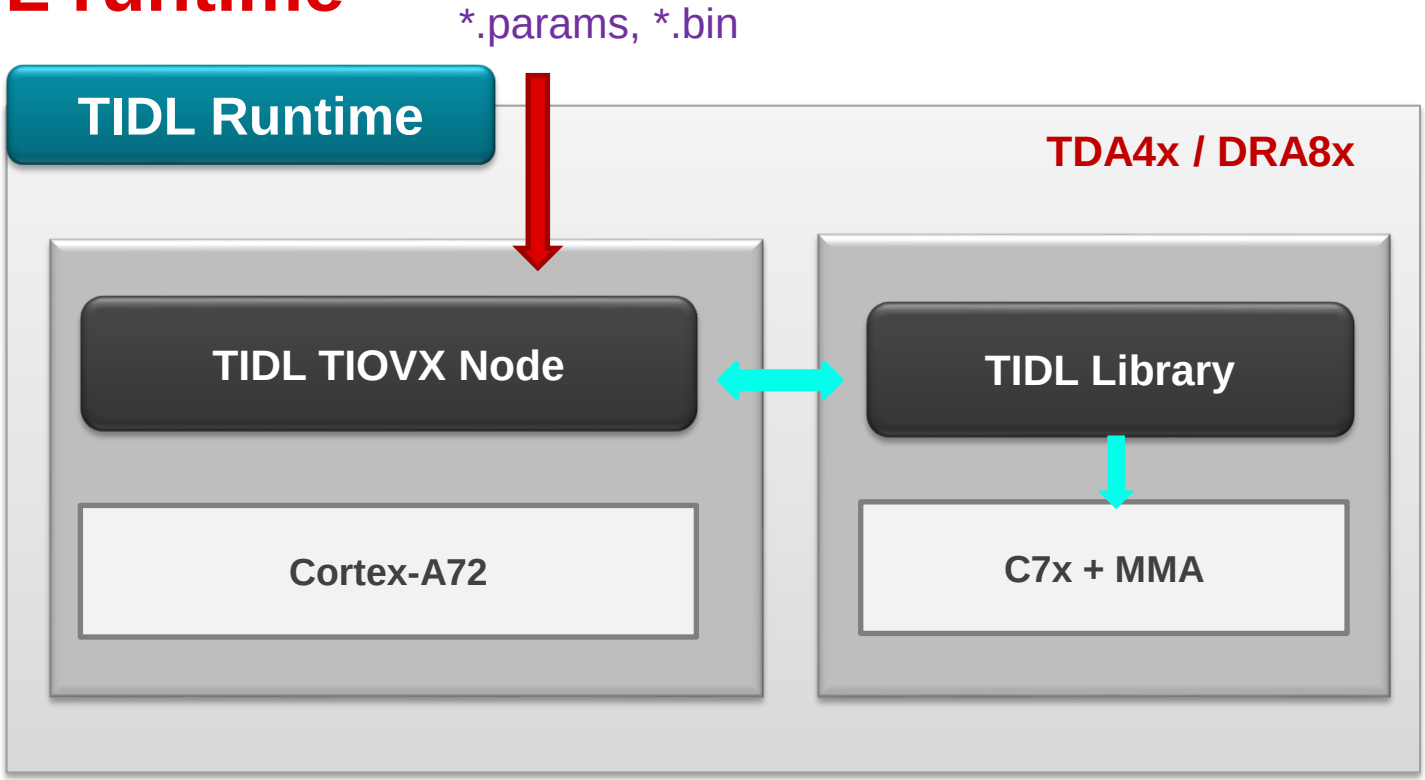
TIDL compiler- recap



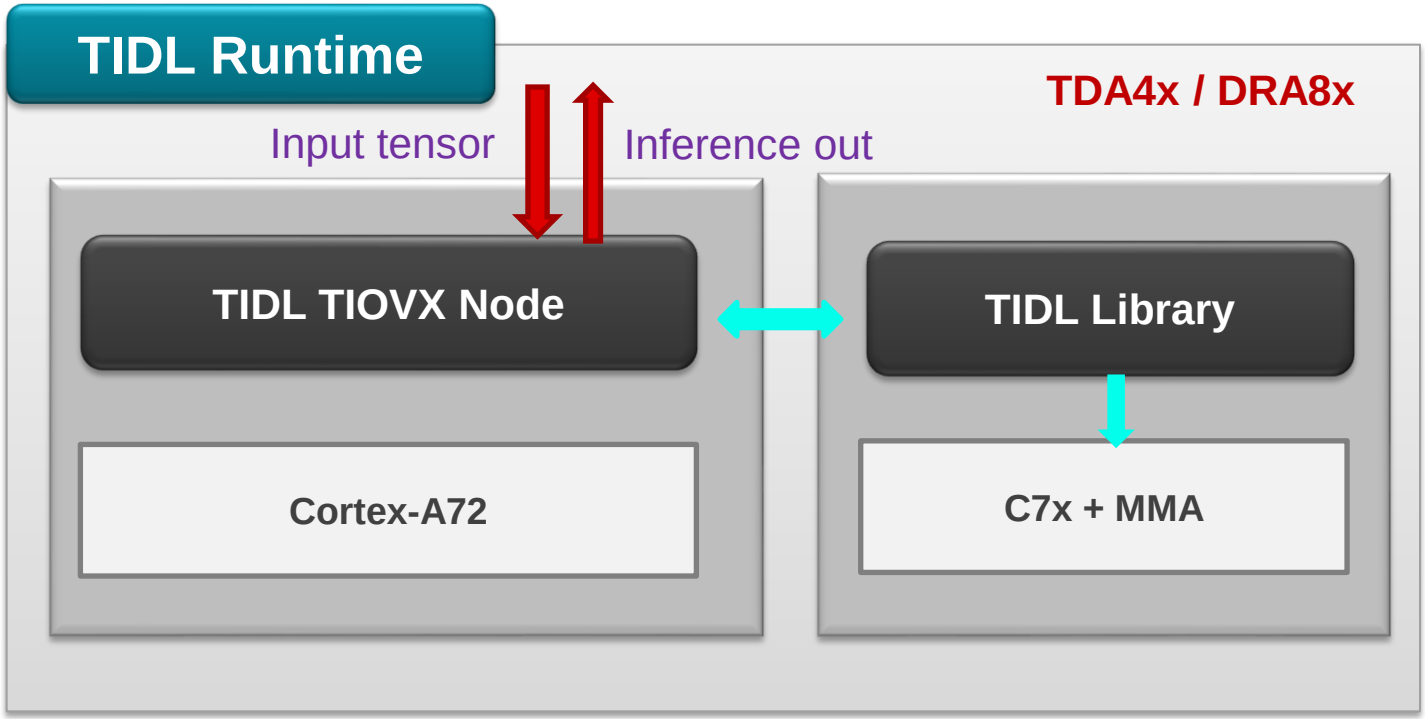
TIDL runtime



TIDL runtime



TIDL runtime



TIDL accelerate all operators you rely on

TIDL features:

- Accelerates all operators commonly used by CNN vision models on our deep learning accelerator cores
- Out-of-box support for 35+ pre-trained CNN models
- List continues to grow

Popular operators supported include:

- Convolution
- Pooling
- Element Wise
- Inner-Product
- Soft-Max
- Bias Add
- Concatenate
- Scale
- Batch Normalization
- Re-size
- Arg-max
- Slice
- Crop
- Flatten
- Shuffle Channel
- Detection output
- Deconvolution/Transpose convolution

Refer to latest Processor SDK user guide document for complete list of accelerated operators and tested models:

https://software-dl.ti.com/jacinto7/esd/processor-sdk-rtos-jacinto7/latest/exports/docs/tidl_j7_01_02_00_09/ti_dl/docs/user_guide_html/md_tidl_layers_info.html

Open source run-time support

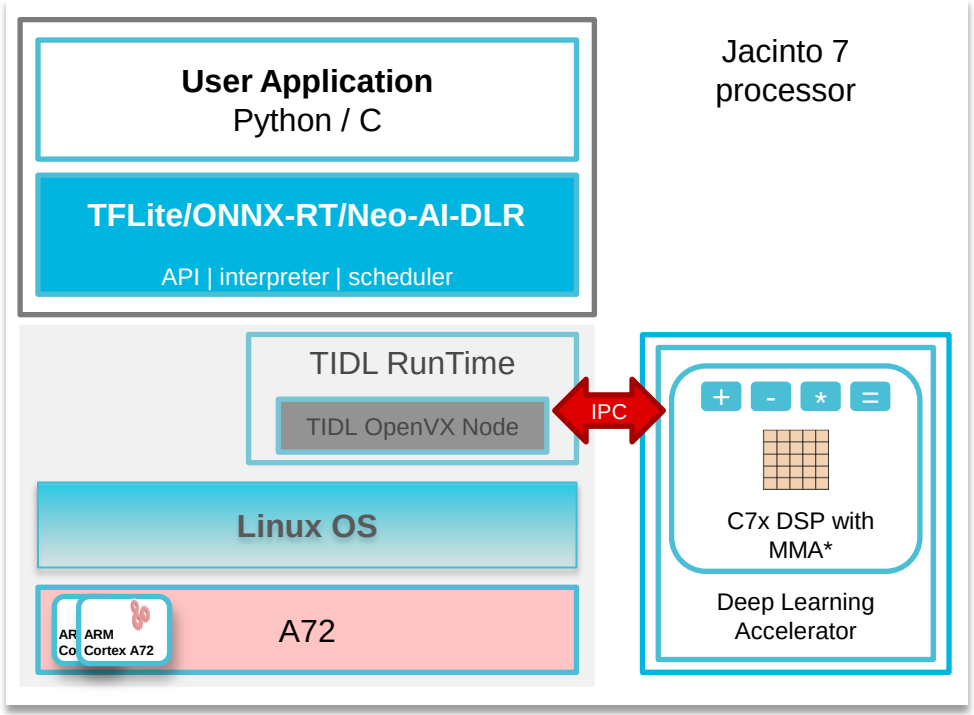
- Work in Progress

Training framework	Optimizer / Compiler / Graph format	RunTime Engine
TensorFlow	TensorFlowLite™ format	TFLite RunTime
	ONNX™ format	ONNX-RunTime
TensorFlowLite	TVM / SageMaker Neo™ compiler	Neo-AI-DLR

Now accelerate all your models with open source API

- Ease of use by providing Open Source programming interface
- TIDL unsupported operators run on Cortex-A core

TensorFLow Lite models	→	TFLite RunTime
ONNX models	→	ONNX-RunTime
TVM/SageMaker Neo compiler	→	Neo-AI-DLR



*MMA: Matrix Multiplication Accelerator (Tensor Processing Unit)

Summary

- TIDL minimizes time to deployment of Deep Learning Networks onto TI SOCs
- Aim for a seamless user experience
- New features and support constantly added

IMPORTANT NOTICE AND DISCLAIMER

TI PROVIDES TECHNICAL AND RELIABILITY DATA (INCLUDING DATASHEETS), DESIGN RESOURCES (INCLUDING REFERENCE DESIGNS), APPLICATION OR OTHER DESIGN ADVICE, WEB TOOLS, SAFETY INFORMATION, AND OTHER RESOURCES “AS IS” AND WITH ALL FAULTS, AND DISCLAIMS ALL WARRANTIES, EXPRESS AND IMPLIED, INCLUDING WITHOUT LIMITATION ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT OF THIRD PARTY INTELLECTUAL PROPERTY RIGHTS.

These resources are intended for skilled developers designing with TI products. You are solely responsible for (1) selecting the appropriate TI products for your application, (2) designing, validating and testing your application, and (3) ensuring your application meets applicable standards, and any other safety, security, or other requirements. These resources are subject to change without notice. TI grants you permission to use these resources only for development of an application that uses the TI products described in the resource. Other reproduction and display of these resources is prohibited. No license is granted to any other TI intellectual property right or to any third party intellectual property right. TI disclaims responsibility for, and you will fully indemnify TI and its representatives against, any claims, damages, costs, losses, and liabilities arising out of your use of these resources.

TI’s products are provided subject to TI’s Terms of Sale (www.ti.com/legal/termsofsale.html) or other applicable terms available either on ti.com or provided in conjunction with such TI products. TI’s provision of these resources does not expand or otherwise alter TI’s applicable warranties or warranty disclaimers for TI products.

Mailing Address: Texas Instruments, Post Office Box 655303, Dallas, Texas 75265