Joint Webinar:

Deep Dive into the i.MX 95 Apps Processor's NPU and Vision Pipeline

With You Today…

Manish Bajaj
System Architecture Engineer
NXP

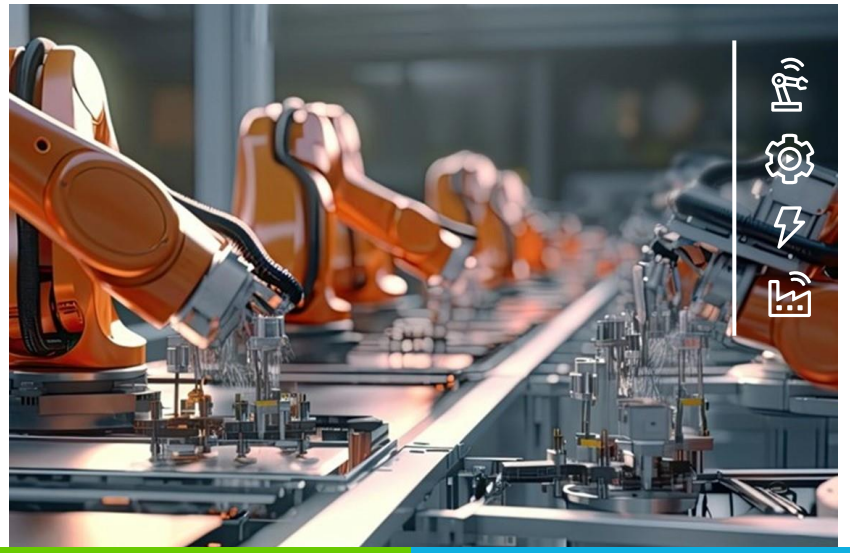
Mohamed Raad
Product Manager - Hardware
Toradex

Deep Dive into the i.MX 95
NPU and Vision Pipeline

Manish Bajaj

September 2024

NXP i.MX 95 Family for Automotive Edge, Industrial, & IoT



Safety

Ditch the hypervisor and simplify building safety capable platforms with the first-generation on-die i.MX functional safety framework. Featuring NXP Safety Manager, Safety Documentation, & NXP Professional support to enable ISO26262 (ASIL-B) / IEC61508 (SIL-2) computing platforms, including 2D display pipeline.

Intuitive Decisions

Deliver increased accessibility and augment complex interfaces with Generative AI-enhanced voice command & control with the first i.MX applications processor to integrate the new, efficient NXP eIQ® Neutron neural processing unit.

Connect & Secure

Build secure, private applications with peace of mind based on the combined capabilities of integrated security and authentication acceleration, including post-quantum cryptographic capabilities, and lifecycle management.

Visualize & Act

Responsive HMI for IoT, Industrial, and Automotive applications are easily created with NXPs partner ecosystem, unlocked by a powerful modern 3D graphics processor combined with strong, efficient hexacore application processor performance.

Learn More:
NXP.com/iMX95

Connectivity Leadership:
UWB, Wi-Fi, NFC, RFID, & BT

Co-Developed Platforms:
PMIC, Wi-Fi, Sensors, & More

Deep Application Insights:
26,000 Customers & Growing

i.MX 95 Block Diagram

Application Domain: **6x Arm Cortex-A55** @ Up to 2.0 GHz
- Arm v8.2 64-bit capable, 768kB total L2+L3 cache (ECC)

Real-Time Domains: **1x Cortex-M7** @ 800 MHz
- Arm v8-M supporting TrustZone-M / 32kB + 32kB cache (ECC) / 512kB TCM/on-chip SRAM (ECC)

Package:
- Large: 19 x 19mm FCBGA, 0.7 mm pitch, de-pop. array[1]
- Small: 15 x 15mm FCBGA, 0.5 mm pitch, de-pop. array[1], adjusted feature set[2]

OS targets: Linux OS, Android OS, GHS, QNX, FreeRTOS

Qualification: Commercial (0° C T_a to +95° C T_j); Auto / Ext. Industrial (-40° C T_a to +125° C T_j)

Safety: **1st Gen i.MX Safety Concept**
- NXP Safety Manager [Arm Cortex-M33] enabled by NXP Professional Services
- H/W Safety checkers & 2D display pipeline targeting ISO-26262 (ASIL-B) / IEC 61508 (SIL-2)

AI/ML: **NXP eIQ® Neutron NPU**

Security: **Edgelock Secure Enclave + Cryptographic Accelerator**

Internal Memory: 1376KB SRAM

External Memory:
- **Up to 6.4 GT/s x32 LPDDR5/4X (Inline ECC & Inline Memory Encryption)**
- 3x SD 3.0 / SDIO3.0 / eMMC5.1
- 1x Octal SPI, including support for SPI NOR and SPI NAND memories

3D Graphics: **Arm Mali G310 GPU (OpenGL® ES 3.2, Vulkan® 1.2, OpenCL 3.0)**

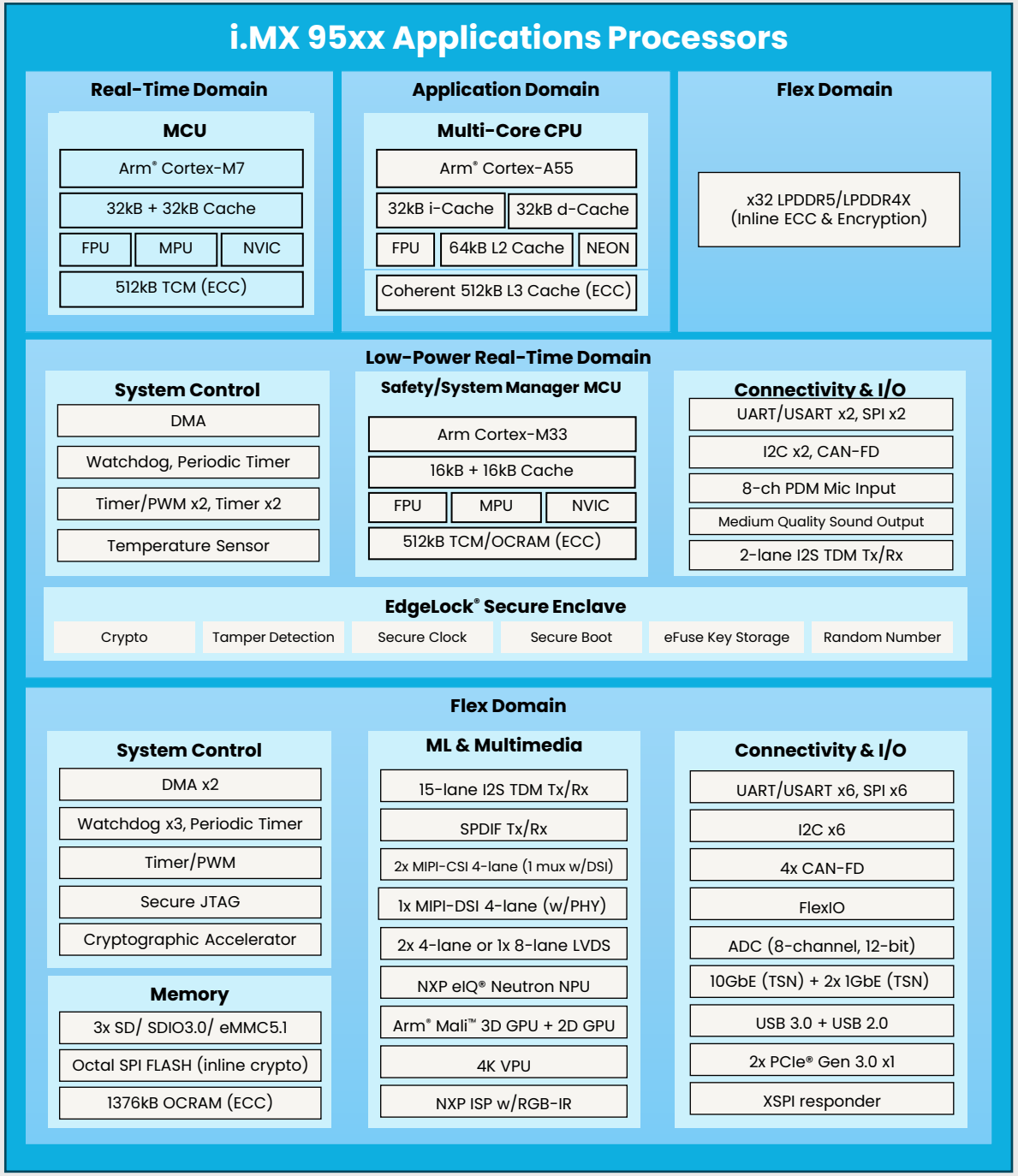
Video Processing:
- 4K60P H.265/H.264 decode or encode
  - 4K30P H.265/H.264 simultaneous encode & decode

Display Controllers (up to 3 simultaneous displays):
- 1x 333Mpixel/s MIPI-DSI (4-lane, 2.5 Gbps/lane) supporting 4K30P or 3840x1440P60
- Up to 1080p60 LVDS Tx (2x 4-lane or 1x 8-lane)

Camera & Image Signal Processing:
- **NXP ISP with RGB-IR support**
- MIPI-CSI (2x 4-lane, 2.5 Gbps/lane) with PHY (1 mux with DSI) with up to 8 virtual channels
  - 475 MPixel/s to scale from 1x 4K30P to 4x 1080P30 cameras
  - 500 MPixel/s Overdrive mode for 8 x1080P30 or 2x 4K30P cameras

Audio:
- 17-lane I²S TDM (32-bit @ 768 KHz), SPDIF Tx/Rx
- 8 channel PDM microphone input + MQS: Medium Quality Sound output

Connectivity & I/O:
- **2x PCIe Gen 3.0 (1-lane)**
- 1x USB 3.0 5Gbps Type C with PHY, 1x USB 2.0 with PHY
- **1x10GbE + 2x Gb Ethernet (w/TSN)**; AVB & IEEE 1588 for sync; and EEE
- 5x CAN-FD, 8x UART/USART/Profibus, 8x I2C, 8x SPI, 2x I3C, XSPI responder
- 1x 8-ch, 12-bit ADC
- 2x 32-pin FlexIO interfaces (bus or serial I/O)



Preliminary information subject to change without notice.
1 – Depopulated array or hybrid sphere placement allow PCB routing technologies normally reserved for larger pitch packages, e.g., standard plated through-hole (PTH) PCB that may not require micro-via, laser-drill, HDI blind or buried vias, or via in pad.
2 – i.MX 95 family in 15 x 15mm package includes: single PCIe x1 lane; LPDDR5 @ 4.2 GT/s; LPDDR4X @ 4 GT/s; does not include 10GbE interface.

VISION IP Highlights

i.MX 95 Family of Applications Processors

Powerful Immersive Graphics for HMI
Arm® Mali™ G310 3D GPU

Superb Code Portability & Ease of Use
- Strong App ecosystem
- Android Play Store
- Industry Support by HMI Engines

- The “Valhall” architecture is the basis of Mali-G310 GPU. It offers a new superscalar engine, simplified scalar ISA, and improved data structure alignment with modern APIs, such as Vulkan
  - Area- & Energy-efficiency Focus
  - QoS Support with finer granularity

Enhanced, Separate 2D GPU
- Enables Safety-context Stream
- Supports (de)warp function
- Higher pixel throughput
  - 1.3 gigapixel/sec
  - Faster than i.MX 8M Plus 2D GPU

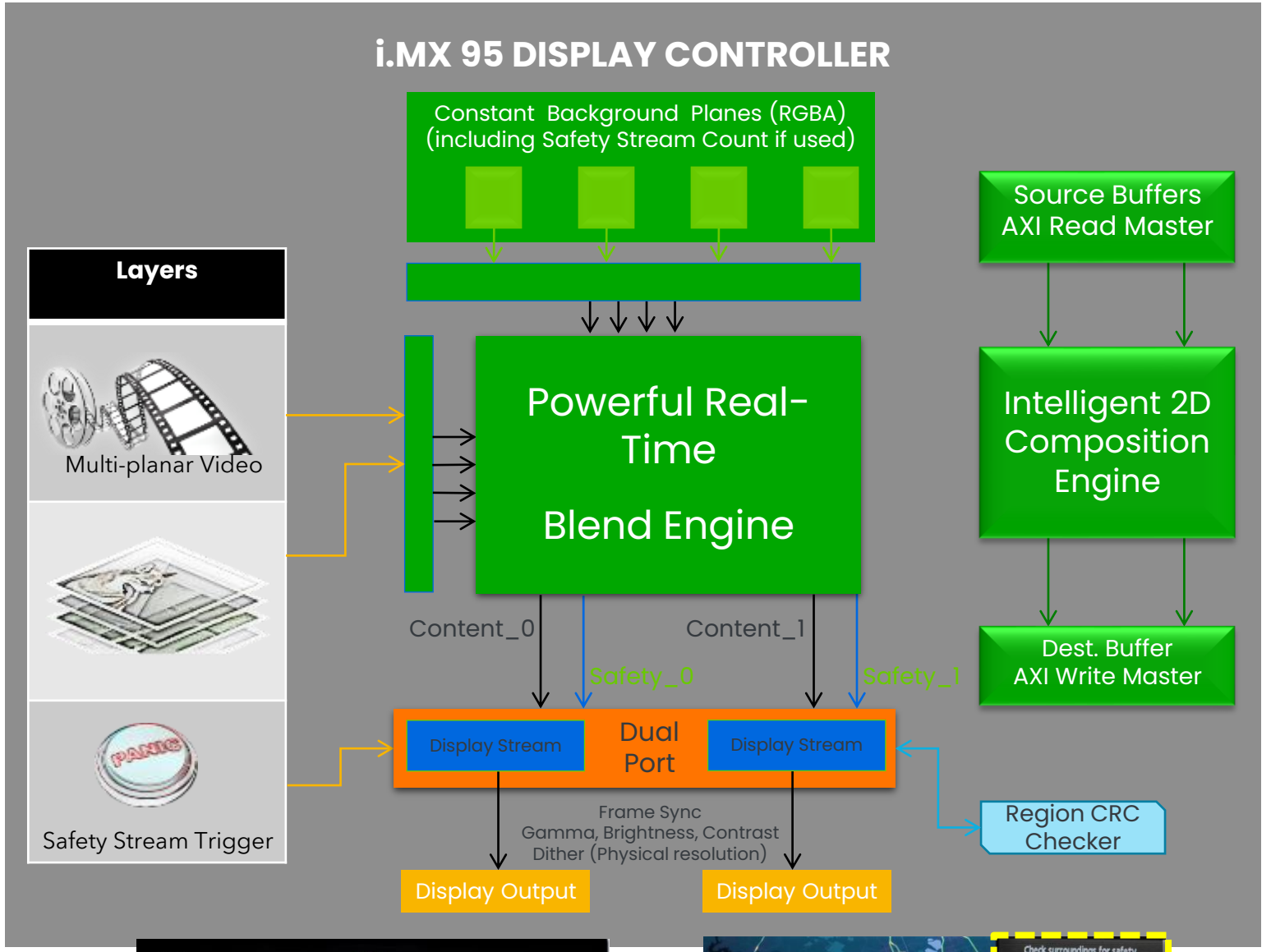
i.MX 95 3D GPU Benchmarks vs. i.MX 8/8x
First i.MX Applications Processor with Arm® Mali™ GPU

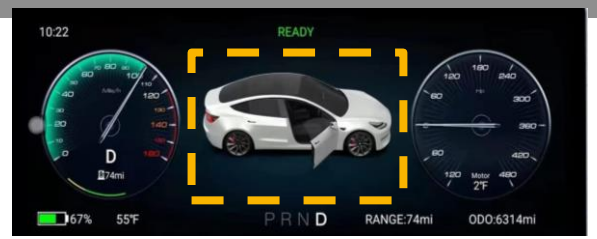
Specification/Benchmark	i.MX8 QXP/DXP GC7000 Lite	i.MX8DM GC7000XSVX	i.MX8QM 2x GC7000XSVX	i.MX 95 Mali-G310 v2[1]
GPU Clock Freq [core/shader; MHz]	700 / 850	800 / 1000	800 / 1000	**1000**
GFLOPS (FP32/FP16)	51.2 / 25.6	64 / 128	128 / 256	**64/128**
Triangle Fill-rate (Mega-triangles/sec)	234	267	267 + 267 (dual) 267 (bridged)	**400**
Pixel Fill rate (Giga-pixels/Sec)	1.4	1.6	3.2	**4**
GFXBench30 Manhattan offscreen1920x1080 (FPS)	5.4	10.7	16.3	**20.3**
GFXBench31 Manhattan31 offscreen (FPS)	3.2	6.9	12	**13**
GLBenchmark27 TREX Onscreen (FPS)	18	26	39	**35**
Antutu v8.4.x (Terracotta-Vulkan) score	TBD	522	534	**984**
Antutu v8.4.x (Refinery-OpenGLES31 + AEP) score	N/A	6849	7603	**18261**

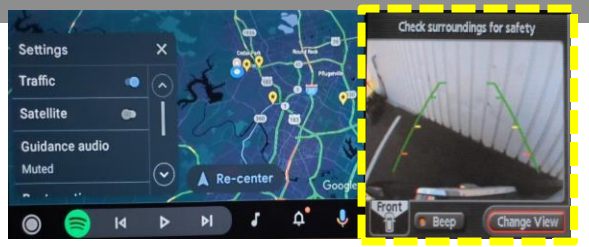
1- Preliminary results obtained by NXP internal team; subject to change. Platform configuration and BSP/software version may offer different results

i.MX 95 Graphics Capabilities

- **2 Display Output Streams**
  - Independent Content, used in combination or singly
  - Triple Independent Display Capability
    - MIPI-DSI 4-lane (4K30P/3840x1440P60)
    - LVDS 2x 4-lane (1080P60)
- **On-the-fly Blending up to 6 Layers**
  - **Video Plane (2x)**
    - Gamma, Brightness, Contrast, Saturation, Upscale, Deinterlace
    - RGB(A), Y(UV) 444, YUV 422, Index
  - **Gfx Plane**
    - YUV 444, RGBA, Index
  - **Background planes (Constant Color, Safety)**
- **Safety Streams**
  - Part of Background planes
  - Region CRC checker to detect fail-condition
- **Integrated 2D Composition engine**
  - High performance 2D operations
  - Independent of Arm® Mali™ GPU/ Arm® Cortex™-A
  - Can be used for warp/de-warp operations
- **Formats**
  - GPU-Tile, Super-Tile, VPU-tile, RGB, YUV, RGBA



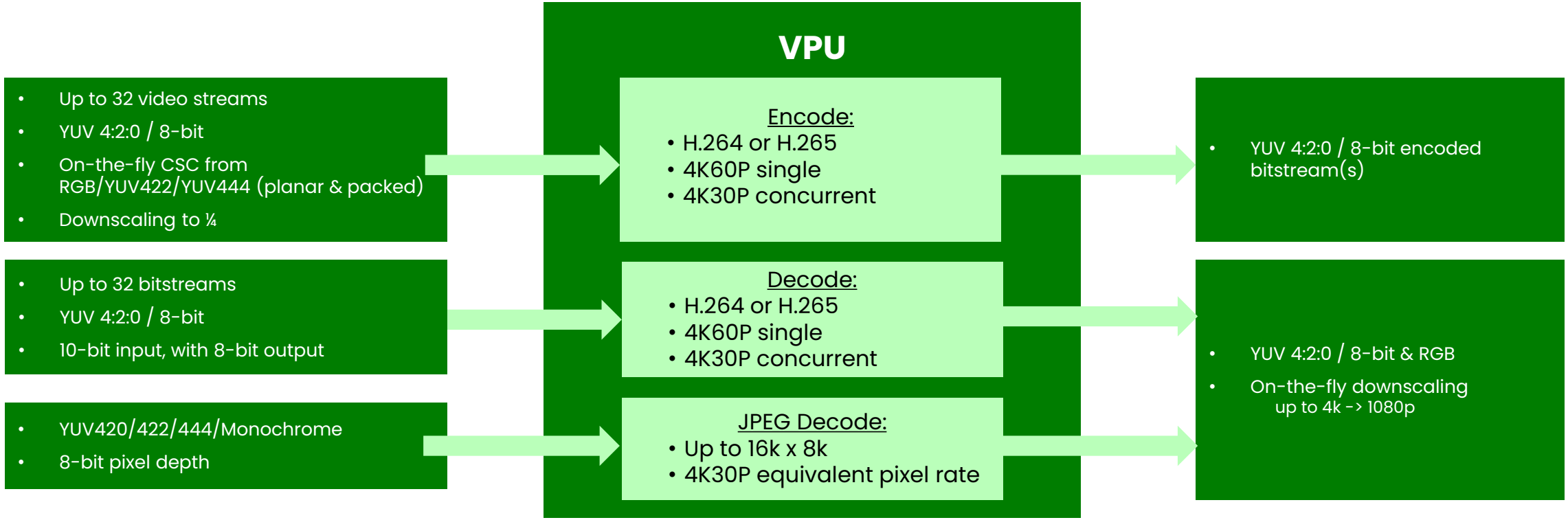




Preliminary specifications for pre-production products under development. Schedule, features, and enablement subject to change without notice.

Video Processing Unit (VPU)
Up to 32 streams from Ethernet, PCIe, USB, or Wi-Fi connected Smart Cameras

- Minimizes memory bandwidth usage:
  - Frame-buffer compression for internal VPU processing
  - On-the-fly color-space conversion & downscaling of input video streams
  - On-the-fly downscaling of output video streams
  - Video output in YUV and/or RGB raster format



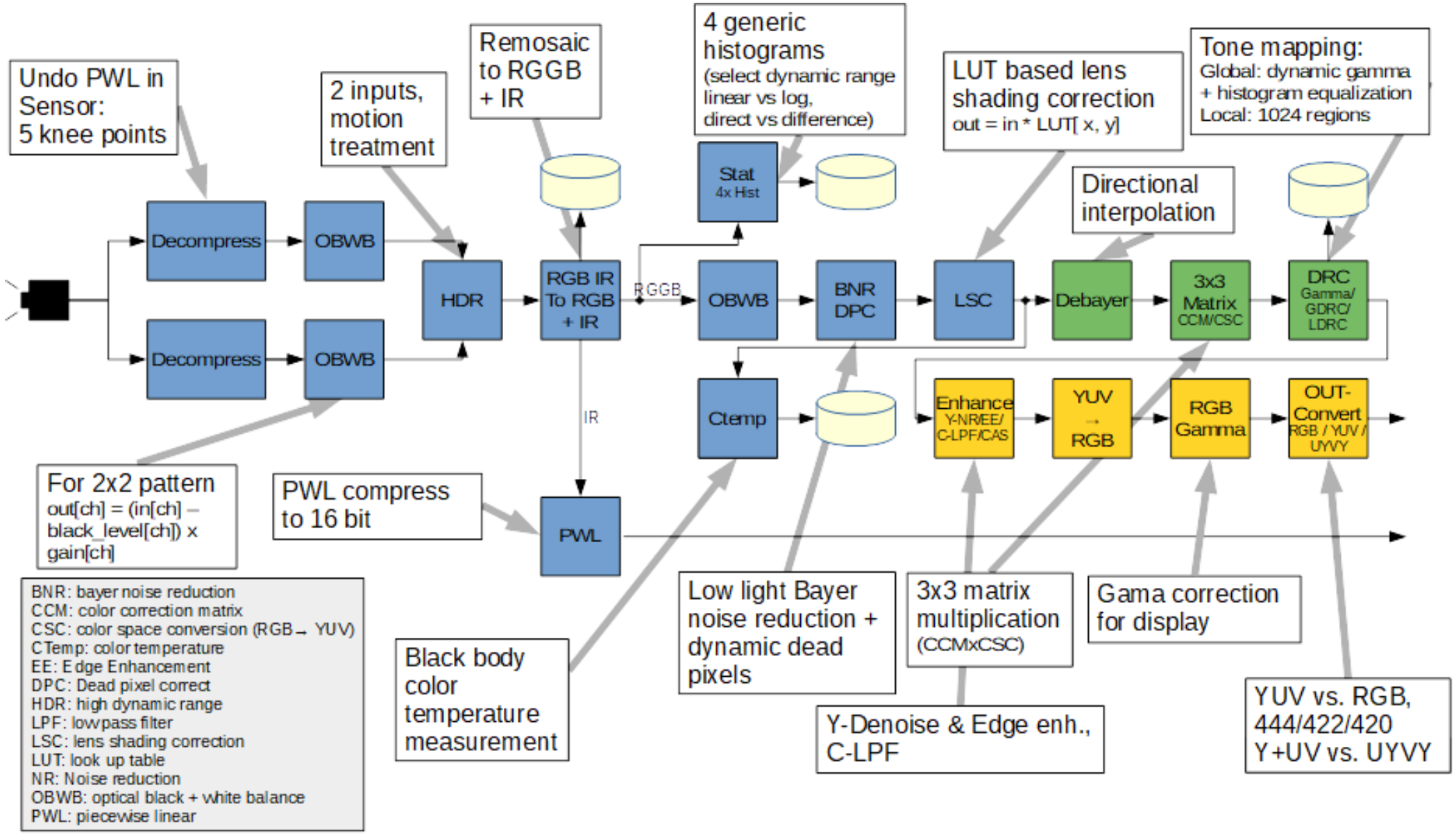
i.MX 95 ISP

i.MX 95 vs. i.MX 8M Plus - ISP Comparison
First i.MX Applications Processor with NXP designed ISP IP

Specification/Feature	i.MX8M Plus	i.MX 95
Pixel Throughput	375 Megapixels/Sec	**500 Megapixels/Sec**
Image Resolution	12MP @ 30fps 8MP @ 45fps	**12MP @ 45fps** **8MP @ 60fps**
Streaming Mode Support	Yes	**Yes**
Memory-to-Memory Support	No	**Yes**
RGB-IR Support	No	**Yes (4x4 Array Pattern)**
High Dynamic Range (HDR) Support	12-bit	**20-bit**
Chromatic Aberration Support	Yes	**No**
Statistics	Block	**Advanced Auto White Balance (AWB)**
Output Formats	YUV 420 YUV 422	**YUV 420, YUV 422** **YUV 444, RGB 888**
S/W Enablement	3rd Party	**NXP Provided Toolchain**
OS Support	Linux oriented	**OS Agnostic S/W Stack**
S/W Stack	V4L Layer provided on top of a native S/W Stack	**Direct Integration into V4L** **LibCamera support (Default)**

NXP Internal Enablement Sensors: OS08A20, OX05B1S, OX03C10
Preliminary specifications for pre-production products under development. Schedule, features, and enablement subject to change without notice.

ISP Pipeline in i.MX 95 Processor



Next-Generation NXP Image Signal Processor (ISP)
Optimized for Machine Vision

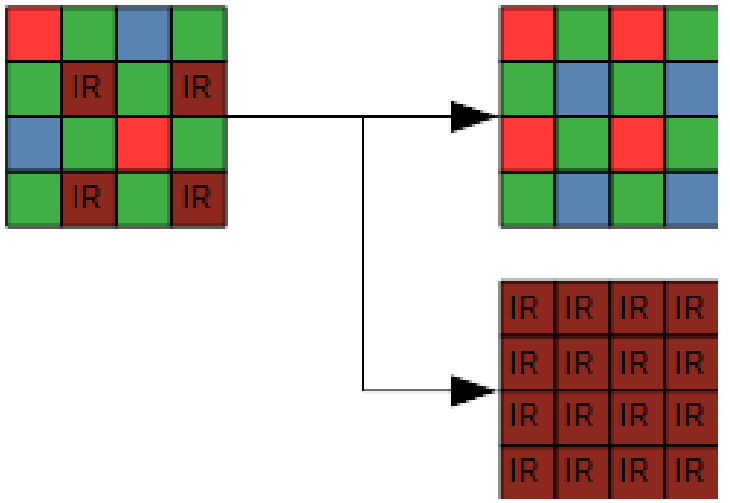
- Up to 12MP high-resolution camera 4096x3072p45 / 3820x2160p60
  - 4096 pixels per line
- 20 bit-per-pixel pipeline supports up to 8 sensors w/aggregate 500MPixel/s throughput
  - 1x 8MP@60FPS / 4x 2MP@60FPS / 8x 2MP@30FPS
  - DCG/HDR input up to 20-bit linear (+ compression)
  - VS input up to 16-bit linear (+ compression)
- HDR combining of 2 exposures
- Support Color, Monochrome & RGB-IR sensors
  - RGGB/RCCB, RCCC, Monochrome, RGB-IR (4x4) Image Sensors
  - Capture RGB + IR images simultaneously for better low-light image capture, small camera modules, and enhanced facial recognition
- Advance de-noising and edge enhancement for low-light conditions
- Single Camera memory-to-memory or streaming processing

Function	Description	Note
Decompress	Undo camera compression: 5 kneepoints	Per DCG / VS input
Black Level, Channel Gain (aka OBWB)	Multiply 2x2 input channels for white balancing Supports post-RGB-IR white balancing	Per DCG / VS input
High Dynamic Range (HDR)	Combine 2 exposures into HDR + motion treatment	HDR in camera for high resolutions HDR in ISP for <=2MP
Statistics	6x Generic Histograms: 2x at input of RGB-IR, 4x in RGGB path	
Region of Interest (ROI)	2x ROI supported	
Bayer Noise Reduction (BNR)	Bayer denoise for low-light conditions	
Lens Shading Correction (LSC)	LUT-based Lens Shading Correction	
Demosaic	Directional interpolation filters, local dynamics correction	
CCM-CSC	3x3 matrix multiplication for Color Calibration Matrix, Color Space Conversion	Combined matrix
Dynamic Range Compression	Compression from 20-bit to 12/10/8-bit GDRC: LUT-based LDRC: 1024 region gamma bases	Conversion from statistics to LUT in SW
Enhancement	Y: Denoise + edge enhancement	
Output converter	Semi-planar and interleaved, YUV420, YUV444, YUYV, RGB, selectable channel order, 16/8-bit buffer output	

RGB-IR 4x4 re-mosaic to RGGB + IR

Re-mosaic:

- Correct RGGB:
  - $R_o = R_i - IR * R\_Crosstalk$
  - $G_o = G_i - IR * G\_Crosstalk$
  - $B_o = B_i - IR * B\_Crosstalk$
- Compress IR from 20bit to 16 bit
- Process RGGB in pipeline
- Output compressed IR



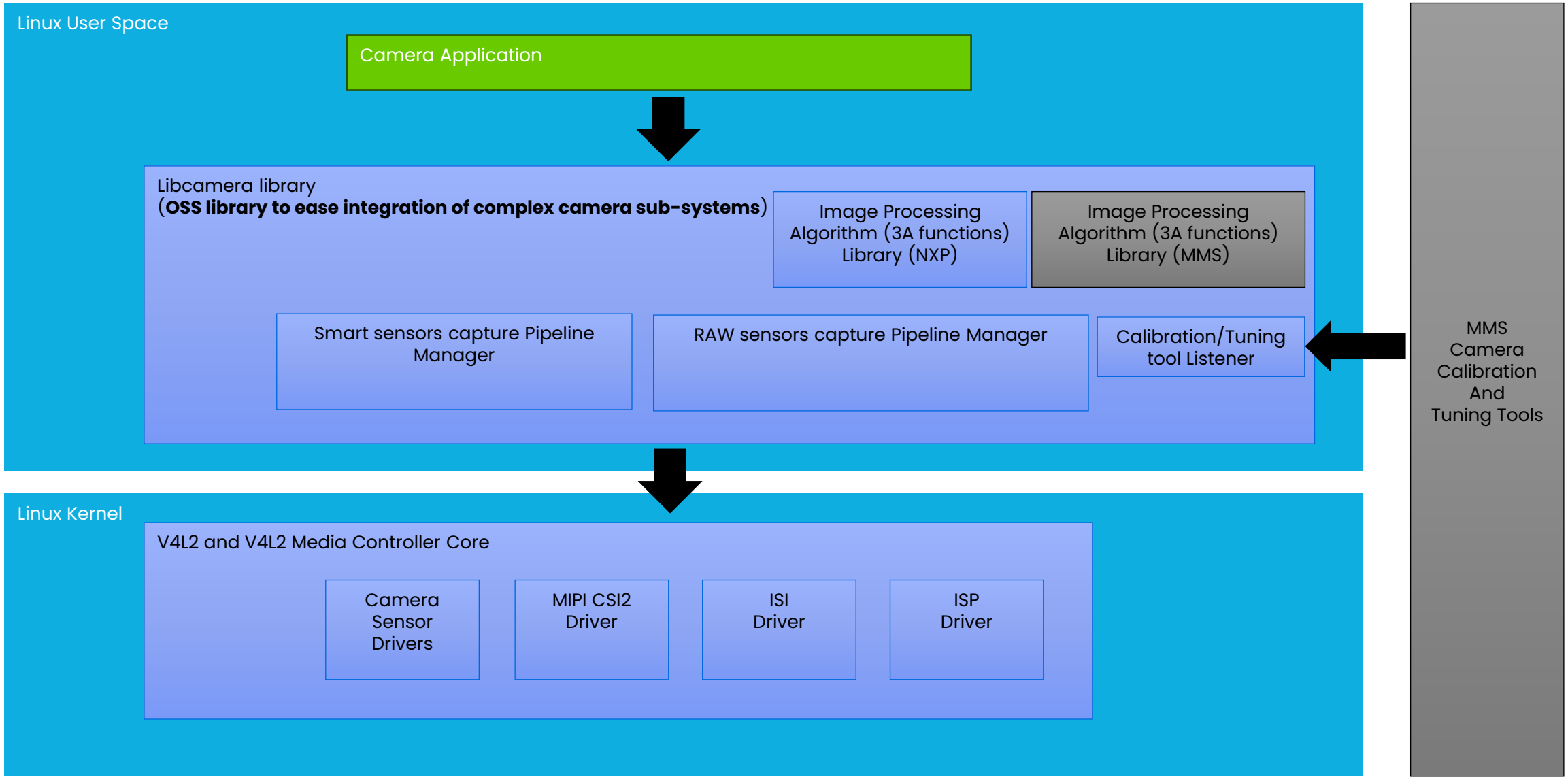
- IR can be processed in extra using pipeline

i.MX 95 ISP and Vision Pipeline

- i.MX 95 SoC is positioned for Vision applications, including NXP-owned ISP and Machine Learning HW accelerators.
- The enablement of the ISP will focus on:
  - Strong NXP ability to support its customers using the ISP
  - Ease of use:
    - SW architecture aligned with de-facto industry standards
    - Use of open-source SW where possible
    - Maximize upstream
  - Providing both high quality, as we well as free of charge options for critical ISP functions:
    - Partnering with MMS for high quality tuning and 3A
    - Open-source tuning tools and 3A
- End-to-end integration of camera, pre-processing, ML inferencing, post-processing, with HW acceleration.
- At this stage, NXP is looking for early partners, ready to engage in early Q4 for ISP usage.

NXP owned, fully equipped Camera tuning lab in Europe

i.MX 95 Linux®-based Camera Subsystem Software Architecture



ISP SUPPORT STRATEGY

R&D and Support delivering
ISP and camera enablement
New Sensor
porting collateral

EVK + BSP/SDK/DOC with
Out of the Box select
camera modules support.

ISP Tuning Enablement

Camera Module Partners

Portfolio of Pre-tuned
Camera Modules

Optional Custom
Tuning

Camera Software Pack

General/Mass Market NXP Customers

Tuning tool and 3A algos provided to NXP and partners by MMS.

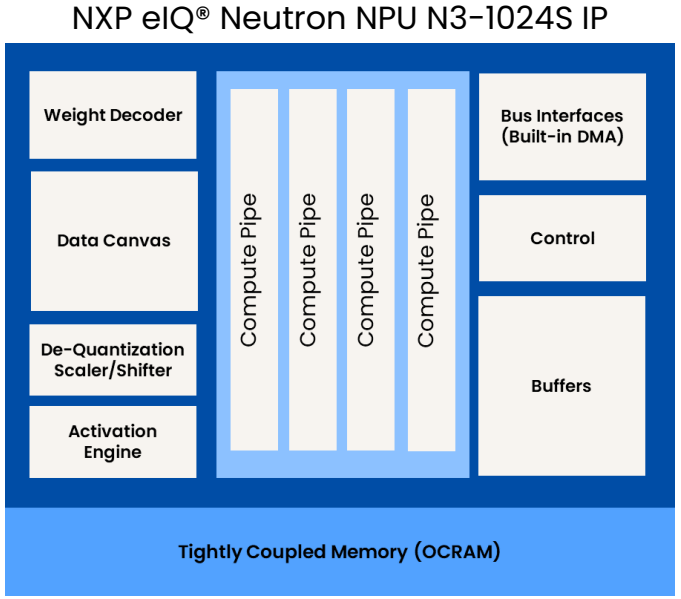
i.MX Processors: NXP eIQ® Neutron Neural Processing
On-die Acceleration for Edge AI/ML Workloads

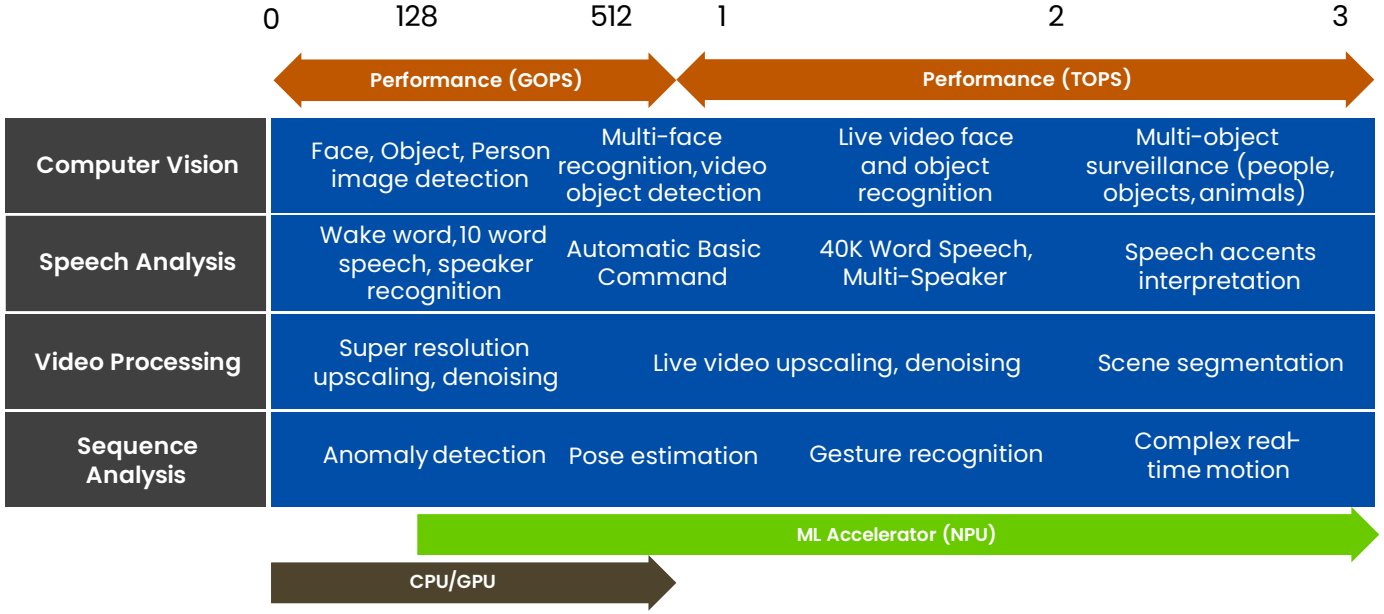
Enabling Edge AI Workloads:

- Predicative Maintenance; Driver/Occupant Monitoring; Speech-to-Intent; Object Classification; Presence Detection; Stereo Vision; Production Line Monitoring; & more
- Single Architecture With Great Scalability
  - Optimized for Edge AI inference workload performance & power efficiency
  - Supports major NN structures (CNN, MLP, RNN, LSTM, TCN, and more)
  - LLM support exploration underway: LLAMA v2 & Blenderbot

Why Build Our Own?

- Internal development provides flexibility to tune solution to better meet our customer needs and the ability to provide ongoing support and generational improvements for changing applications and operator support needs
  - Hardware scales from performance efficient 32 Ops/cycle to 2k Ops/cycle (and beyond) for portfolio coverage with a single architecture, and potential to provide future expansion
  - Software support is unified over multiple generations and device portfolio, creating consistent enablement and support solutions for our customers





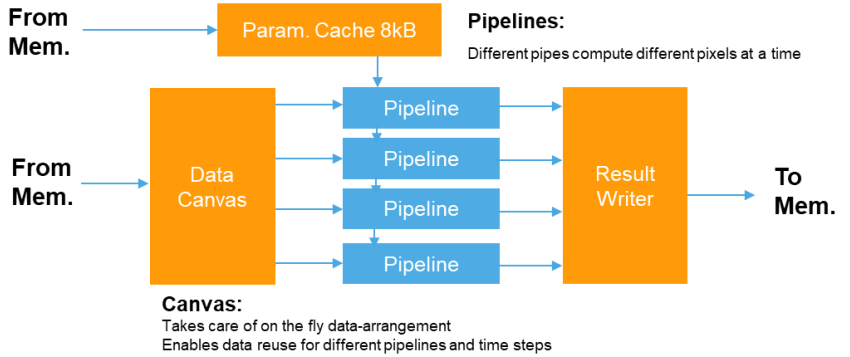
- Introducing the NXP eIQ® Neutron Neural Processing Unit (NPU) | NXP Semiconductors
- Preliminary specifications for pre-production products under development. Schedule, features, and enablement subject to change without notice.

Neutron – Data-Flow / Stationarity

Supports both output and weight stationary execution, incl. hybrid schemes

→ maximizes power-efficiency and performance

Block (simplified):

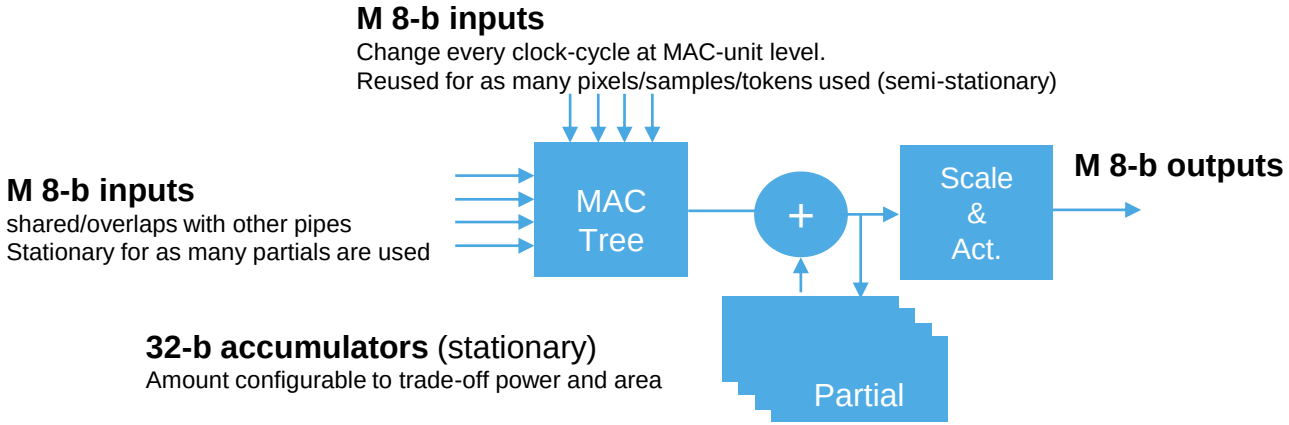


Output Stationary (typically, most efficient):

- outputs sticky to the pipeline until completely accumulated
- multiple output channels accumulated in parallel to reuse data (semi-stationary inputs)
- Pipelines work on different pixels/token/samples to share weights + weights cached for samples > pipelines (semi-stationary weights)

→ **reuse in all three dimensions** (inputs, weights, and outputs) **improves performance, power, and efficiency**

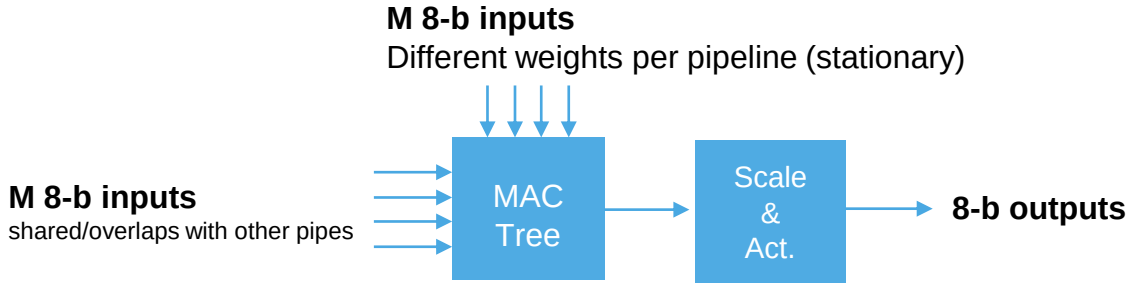
→ 32 outputs are stationary → only 8-bit inputs and weights flow through engine → minimizes area and power



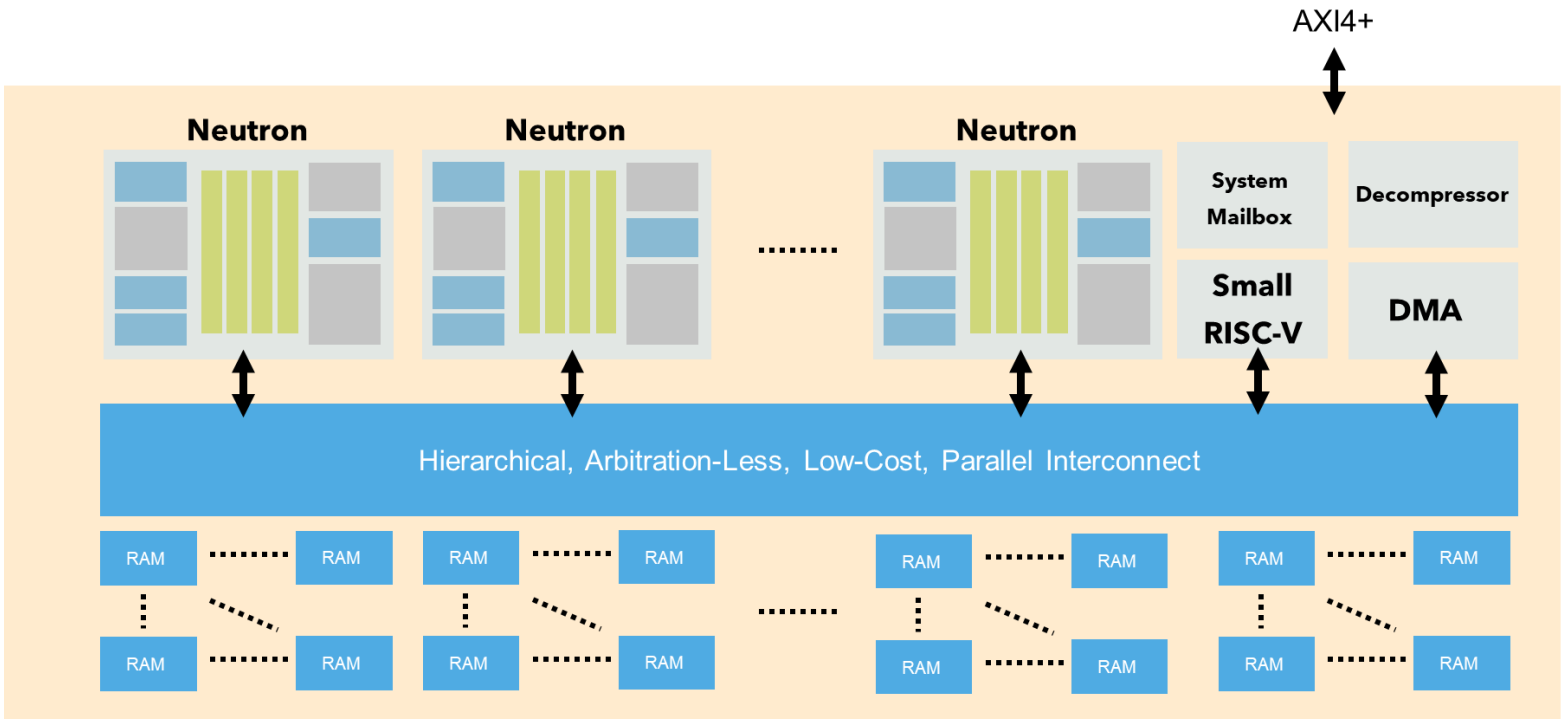
Weight Stationary

- weights sticky to the pipeline
- for one shot accumulation (i.e., number of weights less than M), more efficient than output stationary

(no need to move 32-bit values in and out of the pipelines, despite weight stationarity)



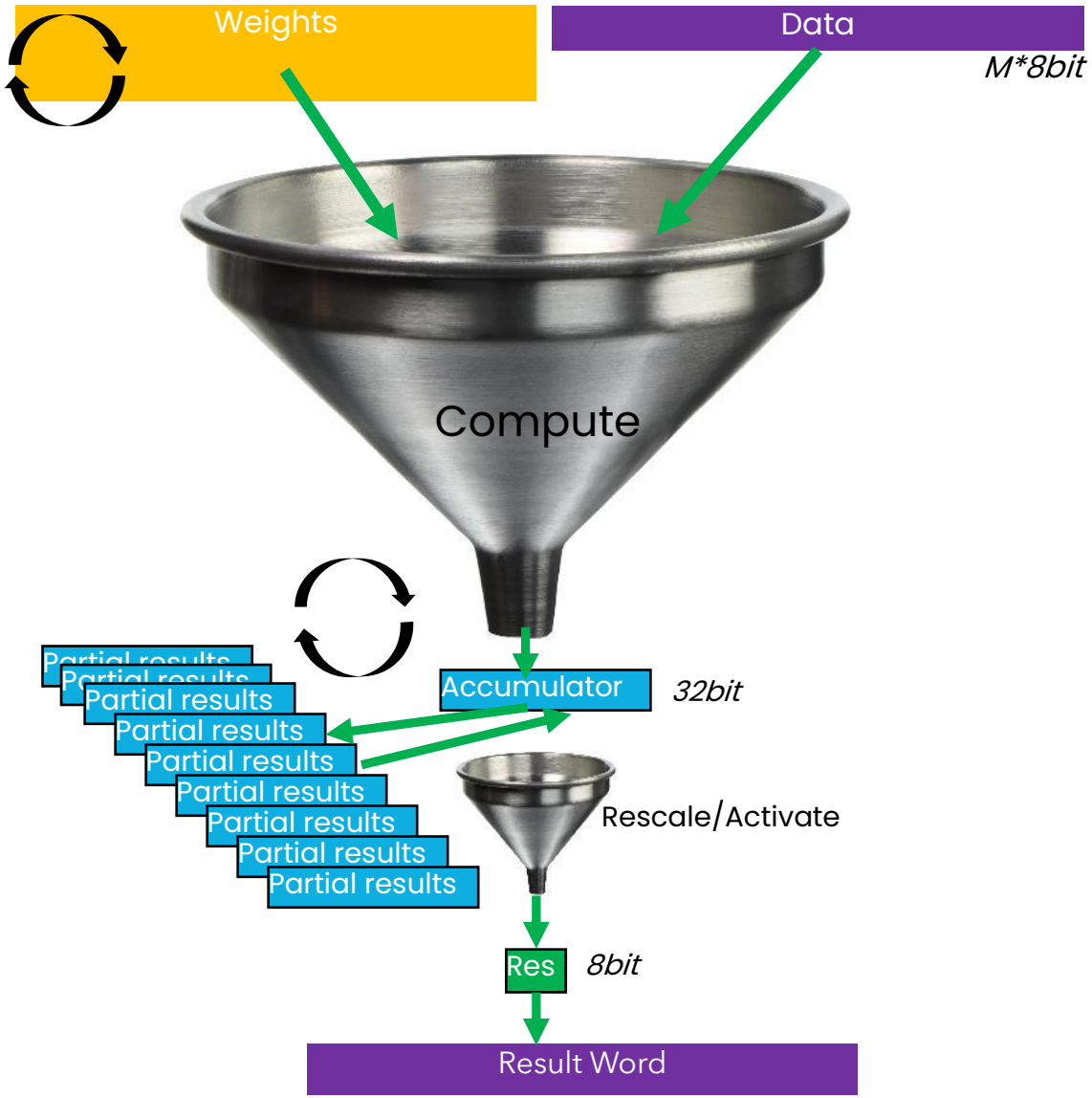
Neutron-S For MPU Class Devices



- More autonomous and scaling-up even further by having:
  - Dedicated high-throughput memories → Enables to integrate multiple of the base cores
  - Neural decompressor
  - Own controller core (RISC-V)
  - Own DMA to manage external memory (e.g., Flash or DDR)
- Works based on a job push queue (in system memory)

→ Same base Neutron core as MCU class enables software reuse!

Neutron – Core Pipeline



- **Matrix-by-vector** operation
- Build up results in partial buffers –reduce data by **dot-products**
- We **move** around **only the small operands**
- **Data re-use** within and across pipelines
- Bandwidth out is reduced by **re-scaling/Activating in pipeline**

Performance (Inferences/Sec)	i.MX 8M Plus VeriSilicon VIP8000 "2.3" TOPS[1]	NXP i.MX 93 ARM® ETHOS™ U65 "0.5" TOPS[2]	NXP i.MX 95 eIQ® Neutron N3-1024S "2" TOPS [3]
MobileNet-v1	368	236	**1102**[4]
MobileNet-v2	332	282	**721**
Inception v3	30	30	**101**
ResNet50-v1	60	20	**125**
SSD-MobileNet_v2	137	76	**350**
Performance relative to Neutron N3-1024S	0.48	0.35	**1**
Performance rel. to Neutron N3-1024S per physical TOPS	0.41	0.71	**1**

1. Measured on NXP i.MX8M Plus EVK
2. From benchmark document by Arm® shared with NXP.
3. Projected based on cycle-accurate simulations.
4. Measured on i.MX 95 A1 pre-validation silicon.

Preliminary results & specifications for pre-production products under development. Schedule, features, performance, and enablement subject to change without notice

eIQ® Neutron Performance:

More Than Simple TOPs

New NXP eIQ® Neutron NPU IP

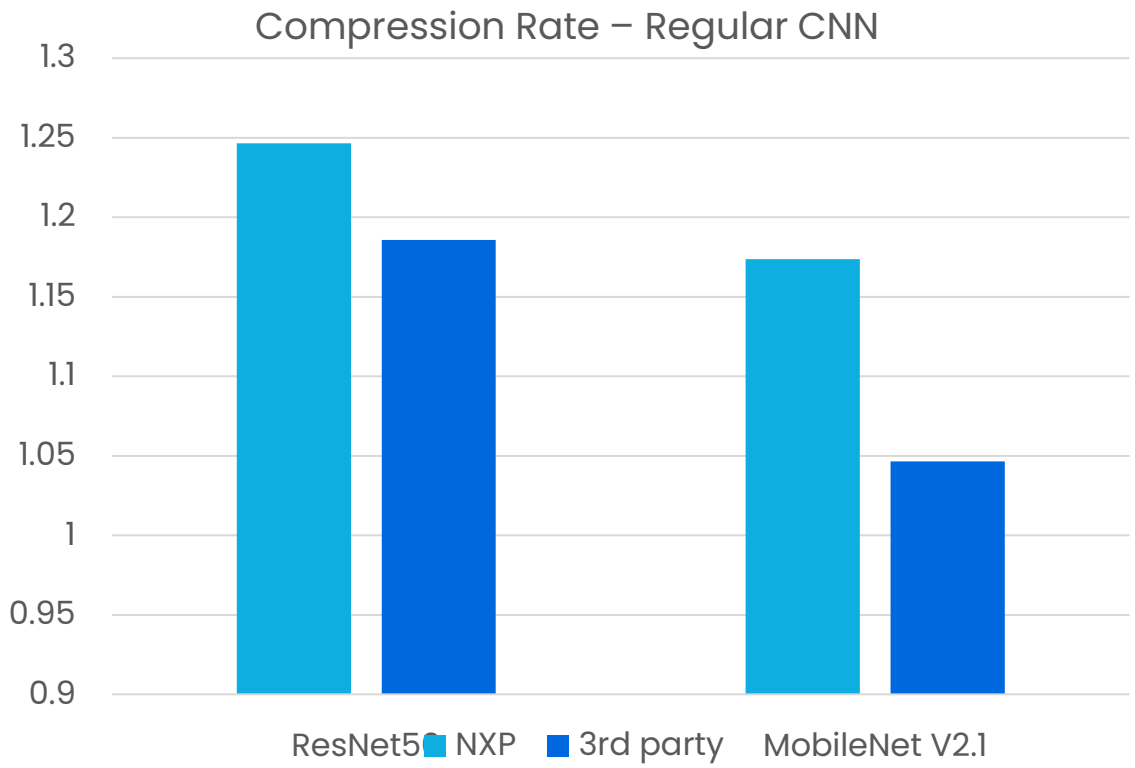
- Focused on Efficient Performance
  - Edge AI workload performance varies depending on model sparsity and data size – examine Inferences Per Second (Inf/s), not mechanical TOPs
- Support for Modern AI/ML Workloads
  - CNN, MLP, RNN, LSTM, TCN, and more;
- Award Winning eIQ Development Environment
  - Enables TensorFlow, Pytorch, Caffe, ONNX, etc.
- Accelerate machine vision workloads such as:
  - Object Detection & Classification
  - Predictive Maintenance
  - Production Line Inspection
  - Pose Detection
  - Facial Recognition &/Or Access Control
  - Voice-to-Intent
  - Driver/Occupant Monitoring
  - Traffic Monitoring & Identification
  - Pedestrian Detection & Identification
  - & More

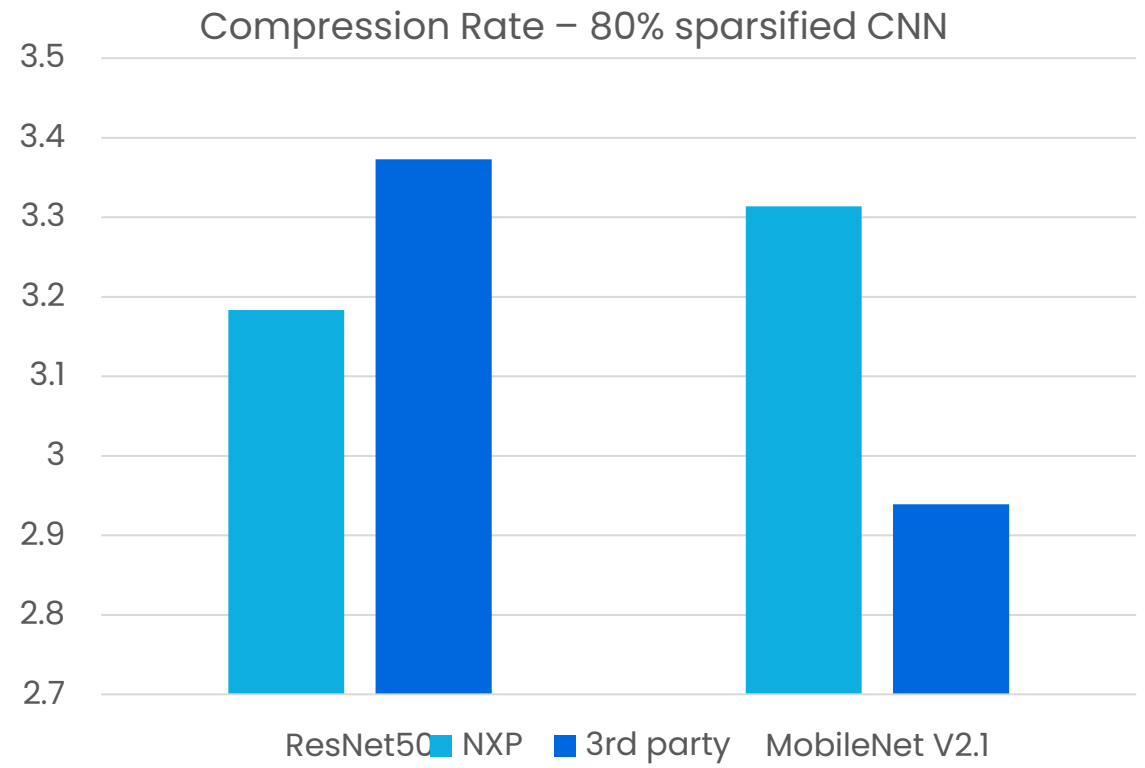
NXP ML Acceleration IP: eIQ® Neutron NPU weight Decompression compared to leading 3rd party IP

Variable-length encoding of weights at compile time. Decompression on-chip, on the fly, when weights are loaded into TCM

- Lossless (i.e., no impact on accuracy)
- Exploits normal distribution of weights caused by weight regularization (e.g. L2) during training
- Exploits uneven distribution from pruning weights

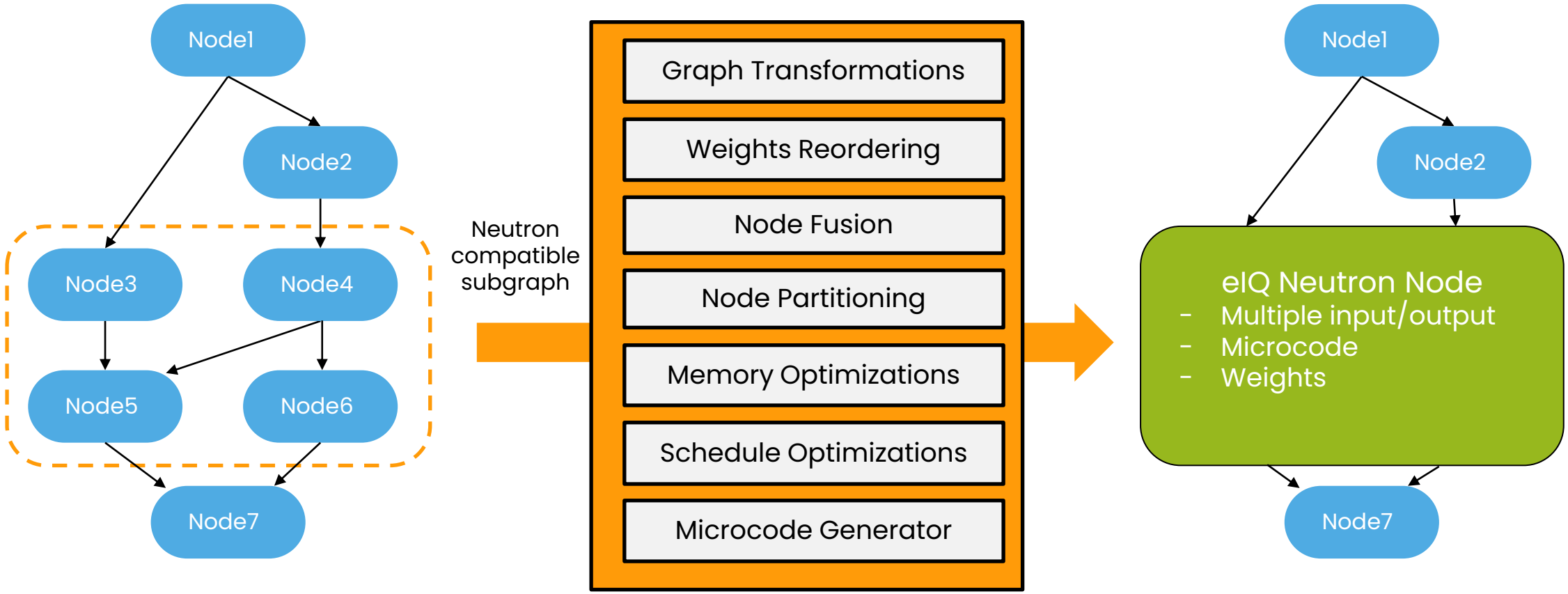
→ Memory savings for standard as well as pruned (i.e., sparsified) neural networks.





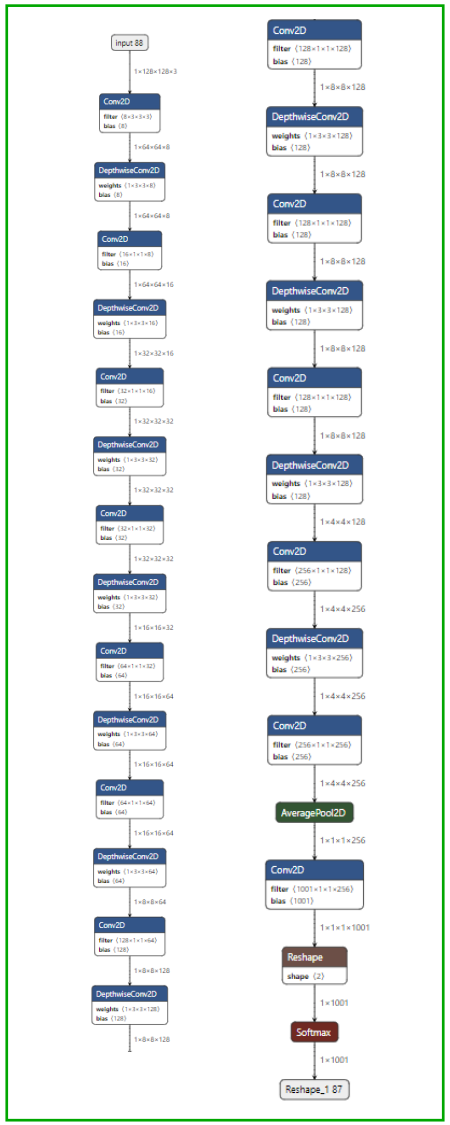
NPU SOFTWARE ARCHITECTURE

Software Architecture - Converter



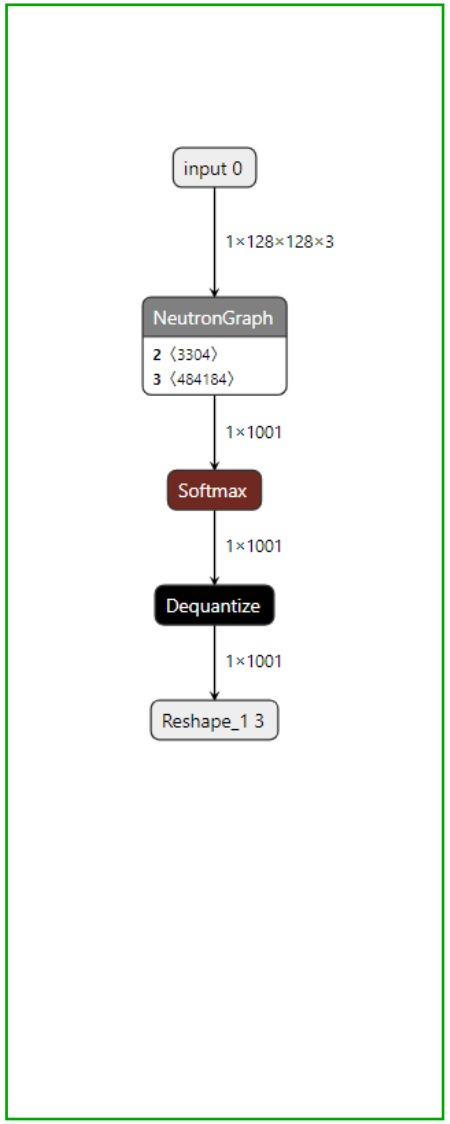
eIQ® Neutron converter Example #1

Before

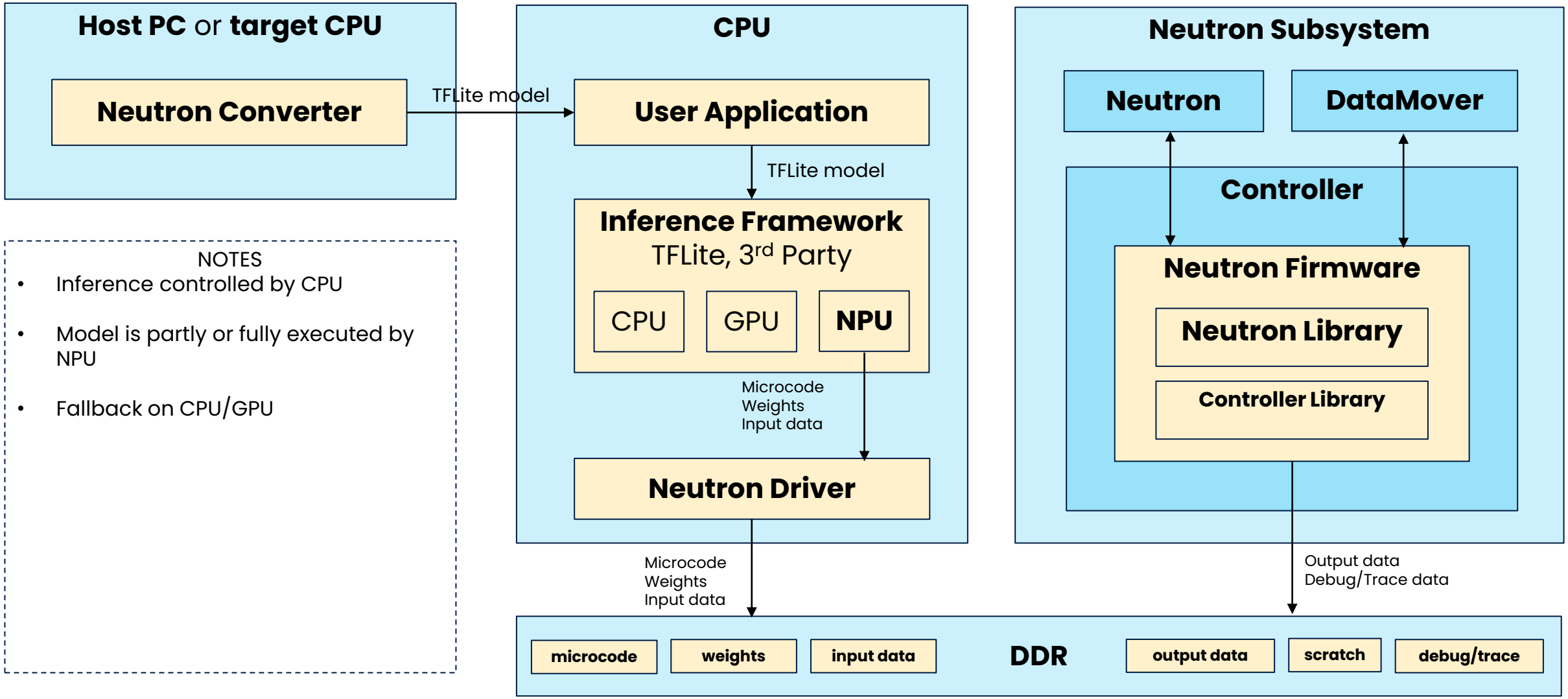


eIQ Neutron Converter

After



Neutron-S SW Flow

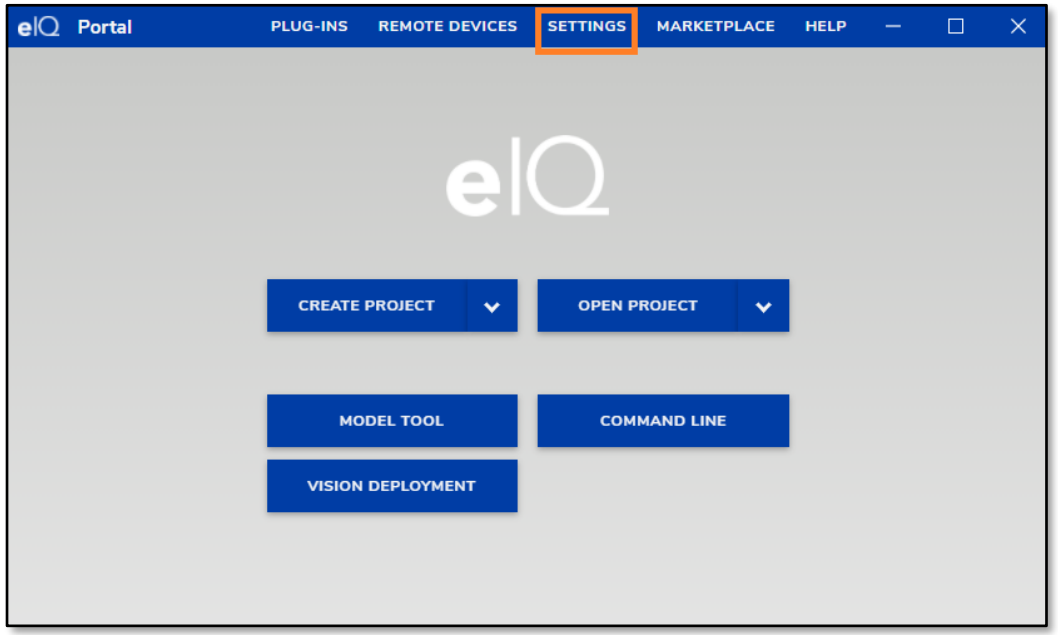


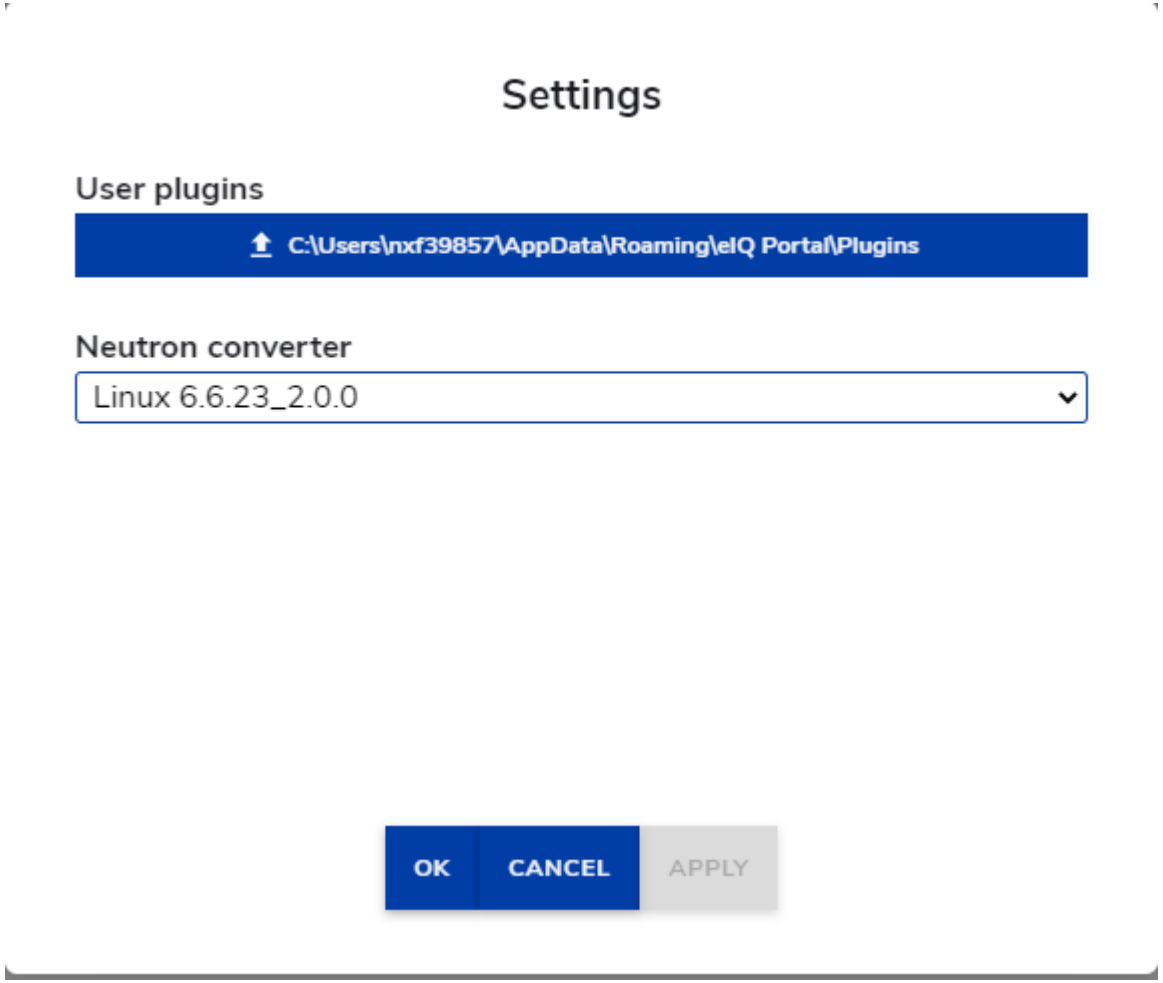
NOTES

- Inference controlled by CPU
- Model is partly or fully executed by NPU
- Fallback on CPU/GPU

Model Conversion: GUI

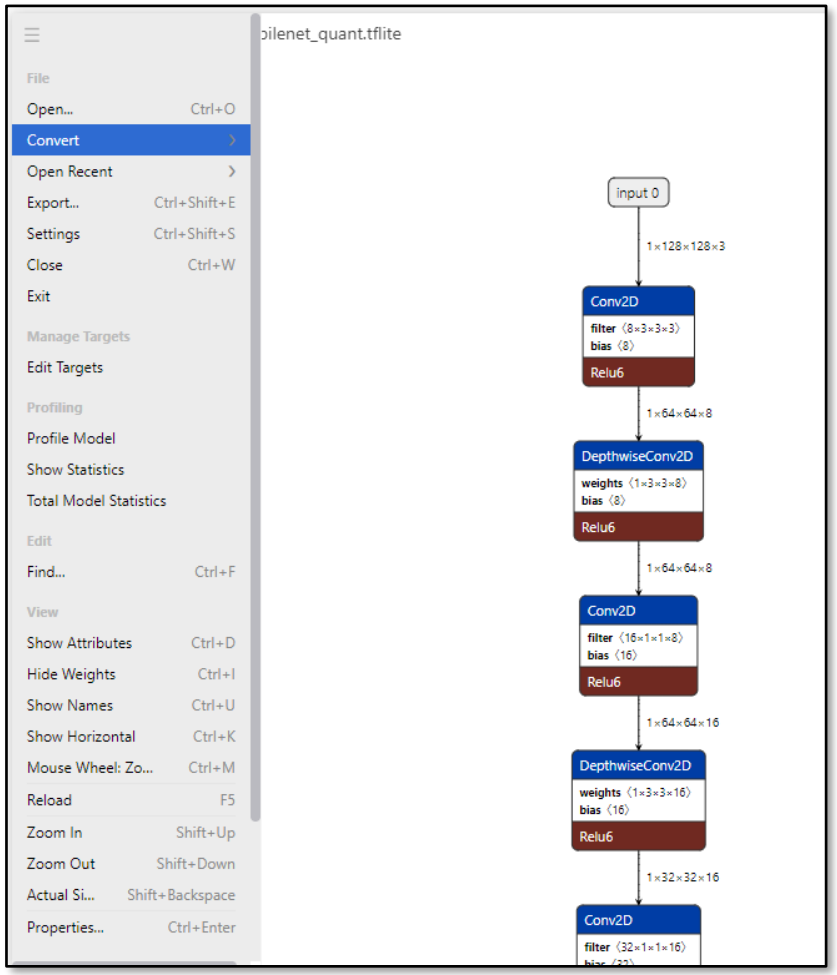
- Open eIQ Portal (part of eIQ Toolkit)
- Select which version of the Neutron Converter to use by clicking on **SETTINGS**

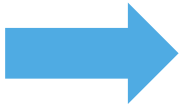


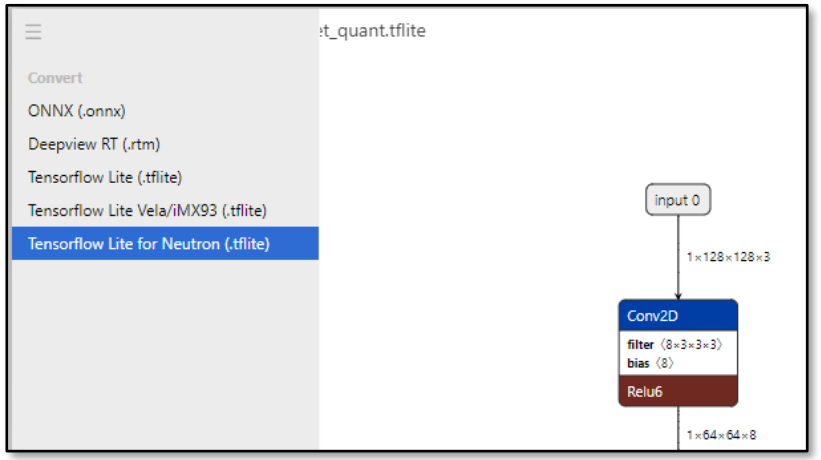


Model Conversion: GUI

Open model with Model Tool and use menu selection to **Convert**



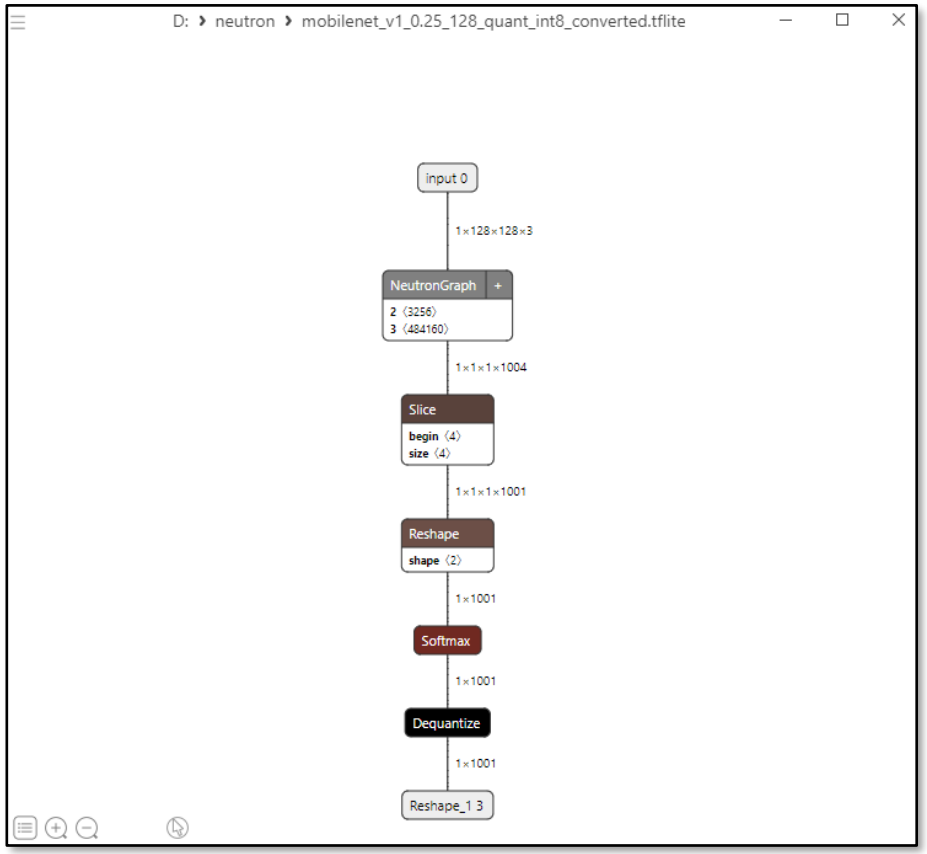




Model Conversion: GUI

In conversion options can select the target device and any custom options

- **dump-header-file** will be most common used option since will use that generated file to deploy model in SDK project



eIQ® Neutron-converter

- **neutron-converter** is the main tool used to convert a standard TensorFlowLite model into a custom TensorFlowLite model ready for Neutron execution by the Neutron backends.

```
neutron-converter --input <input_model_path> --output <output_model_path> --target <neutron_target>
```

- By default, the **neutron-converter** will try to specialize all the Neutron supported operators.
  - If you want to only specialize some operators or exclude some operators from being specialized, then there are command line arguments that can be used detailed in the User Guide

TAO/eIQ Workflow

1. Prepare the dataset.
2. Specify the experiment.
3. Train the model.
4. Export the model.

Action Recognition Example (TAO)

5. (Optional) Optimize the model.

Pruning Example (TAO)

6. Deployment.

Deployment with eIQ Toolkit (eIQ)

ONNX2TFLite

- NXP's own command-line tool to convert models directly from **ONNX to TF Lite** available in **eIQ Toolkit**
  - No need to go through TensorFlow (direct mapping from ONNX to TF Lite Flatbuffers)
  - Customers often use opensource tooling. Best converter experience based on NXP internal teams.
- Includes module for quantization called **ONNX2Quant**

Latest version:

https://bitbucket.sw.nxp.com/projects/AITEC/repos/onnx2tflite

Public releases (everyone):
https://www.nxp.com/design/design-center/software/eiq-ml-development-environment/eiq-toolkit-for-end-to-end-model-development-and-deployment:EIQ-TOOLKIT

Internal releases (+ alpha customers/qualified releases/other means of distribution):
https://nl-nxrm.sw.nxp.com:8443/#browse/browse:EIQTOOLKIT-raw_installer_release

Quantizing the ONNX Model

1. Convert calibration dataset - ONNX2Quant takes *.npy files (tensors) as input
2. Open **COMMAND LINE** from eIQ Portal

```
python –m onnx2quant <onnxModelPath> -o <onnxQuantModelPath> -c input.1::<calibrationDatasetPath>
```

Note:

ONNX2Quant uses ONNX RT Quantizer as an engine.

Converting to TF Lite

- eIQ Core supports acceleration only using TF Lite

```
python -m onnx2tflite <onnxQuantModelPath> -o <tfliteQuantModelPath>
```

Note:

ONNX2TFLite currently still outperforms tools such as Google's AI Edge Torch or the opensource ONNX2TF (model support). On top it is developed by NXP's and thus implements optimizations for NXP devices.

- ONNX2TFLite is guaranteed long-term support and development by NXP.

eIQ Toolkit Roadmap

2024

	Q1	Q2	Q3	Q4	Q1.25
	1.10 (22 Jan)	**1.11** (1 Apr)	**1.12** (1 Jul)	**1.13** (7 Oct)	
eIQ Toolkit	• Image Segmentation GUI • Modelrunner auto-installer • PyTorch enablement (ONNX2TFLite converter) • TAO Toolkit Extension (preview) • Splash Screen & Component Loading	• RT700 support • Cycle Estimator for Profiling • Custom Activation GUI • DeepView to eIQ Rebranding • TAO Toolkit Extension (production)	• i.MX 95 Support (EAR) • TensorFlow, ONNX/ONNX RT, Netron, node.js, ONNX2TFLite • MCX-N profiling • Neutron Converter (latest)	TensorFlow, ONNX/ONNX RT, Netron, node.js, ONNX2TFLite Neutron Converter (latest)	

i.MX 95 ISP and Vision pipeline

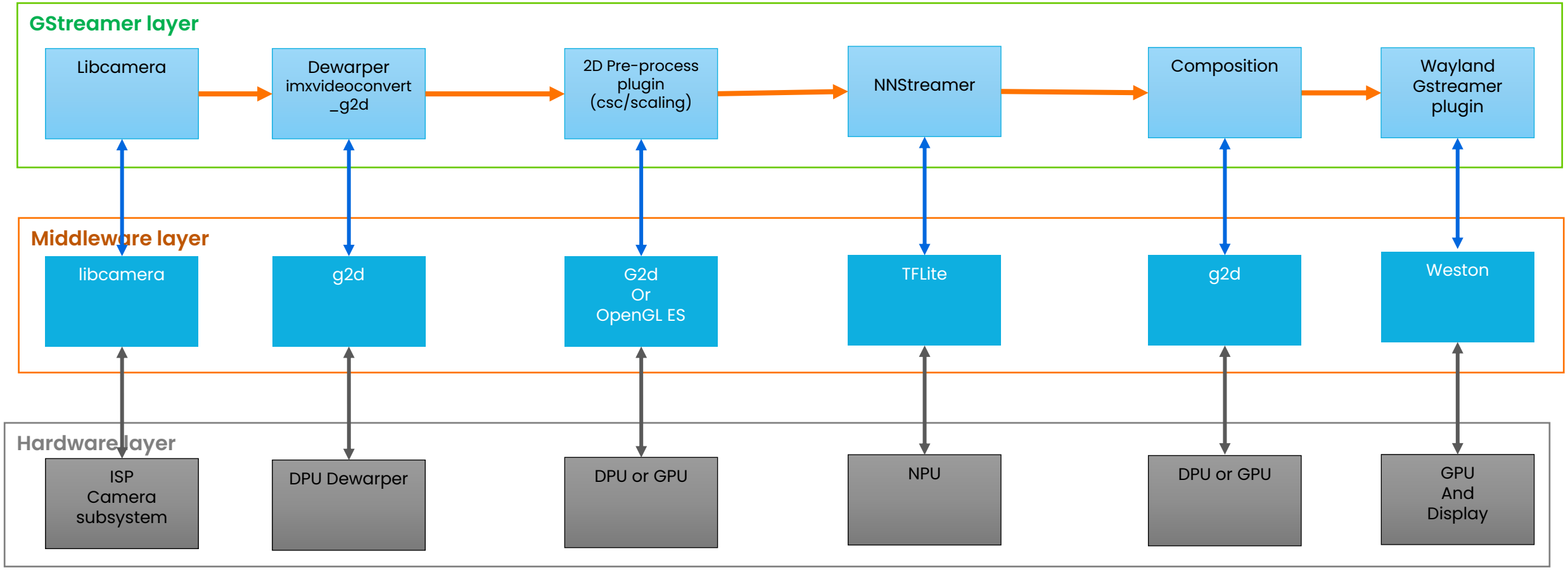
- i.MX 95 is positioned for Vision applications, including NXP-owned ISP and Machine Learning HW accelerators.
- The enablement of the ISP will focus on:
  - Strong NXP ability to support its customers using the ISP
  - Ease of use:
    - SW architecture aligned with de-facto industry standards
    - Use of open-source SW where possible
    - Maximize upstream
  - Providing both high quality, as we well as free of charge options for critical ISP functions:
    - Partnering with MMS for high quality tuning and 3A
    - Open-source tuning tools and 3A
- End-to-end integration of camera, pre-processing, ML inferencing, post-processing, with HW acceleration.

NXP owned, fully equipped Camera tuning lab in Europe

MPU Vision Pipeline

Features	Q1'2024	Q2'2024	Q3'2024	Q4'2024
NNStreamer	N/A	• i.MX95 Support to Neutron, DPU, GPU (alpha version) • NNStreamer v2.4.0	• i.MX95 DPU optimization • i.MX95 Libcamera support • i.MX 8M Plus/i.MX93 C++ examples • Merge nxp-nnstreamer-examples with GoPoint	• i.MX95 Performance optimization • i.MX95 Performance tools support
OpenCV Hardware Acceleration	N/A	N/A	• i.MX95 DPU acceleration for rotate, scaling and flip	• i.MX95 DPU acceleration for color space conversion • i.MX8 MP/i.MX93 2D GPU acceleration for color space conversion

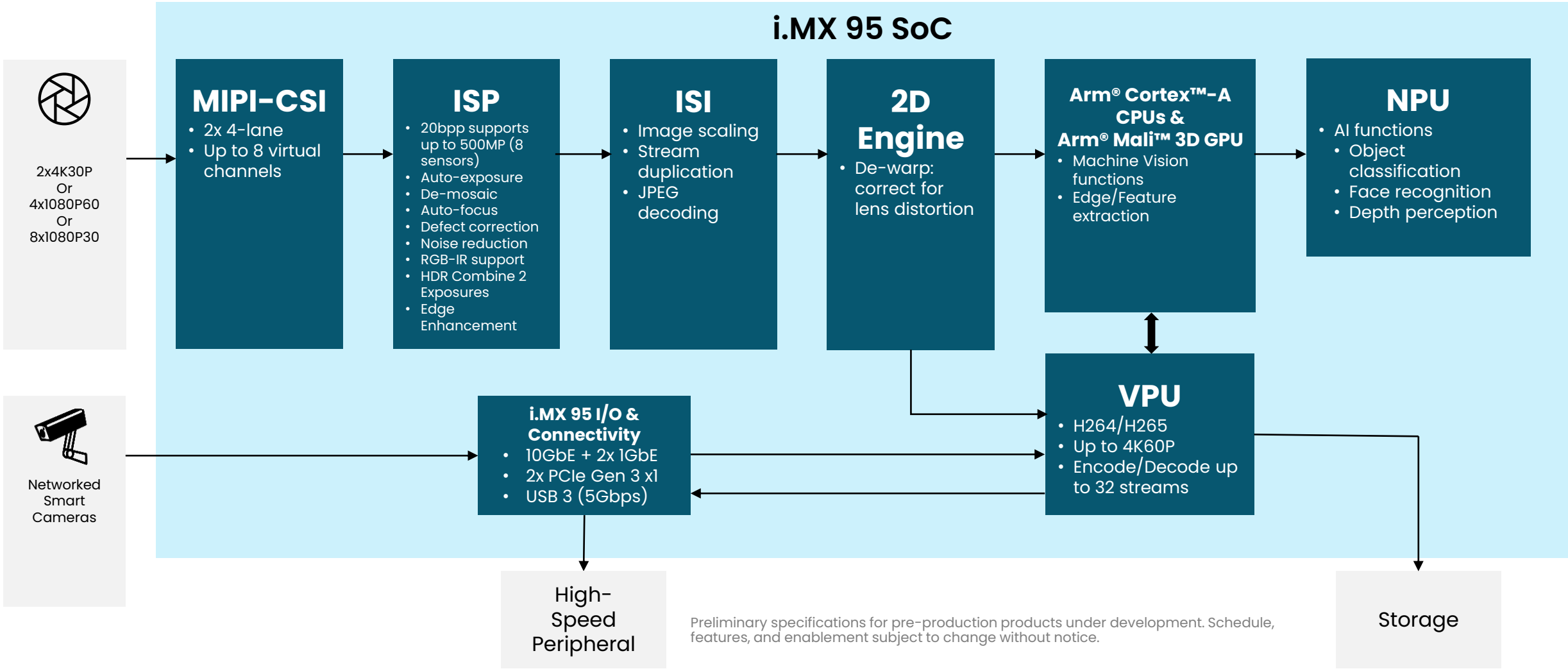
High-Level Architecture: end-to-end vision pipeline on i.MX95




```
$ gst-launch-1.0 libcamerasrc ! tee name=t t. ! queue \ # camera stream capture
imxvideoconvert_g2d ! video/x-raw,width=224,height=224,format=RGBA ! videoconvert ! video/x-raw,format=RGB ! \ # camera frames pre-processing
tensor_converter ! \ # camera frames conversion to NN tensors
tensor_filter framework=tensorflow-lite model=mobilenet_v1_uint8_converted.tflite custom=Delegate:External,ExtDelegateLib:libneutron_delegate.so ! \ #
NN inference using Neutron
tensor_decoder mode=image_labeling option1=labels_mobilenet_quant_v1_224.txt ! \ # NN inference post-processing
overlay.text_sink \ # classification results display on screen
t. ! queue textoverlay name=overlay font-desc=\"Sans, 24\" ! autovideosink sync=false # Composition: capture capture and classification labels
```

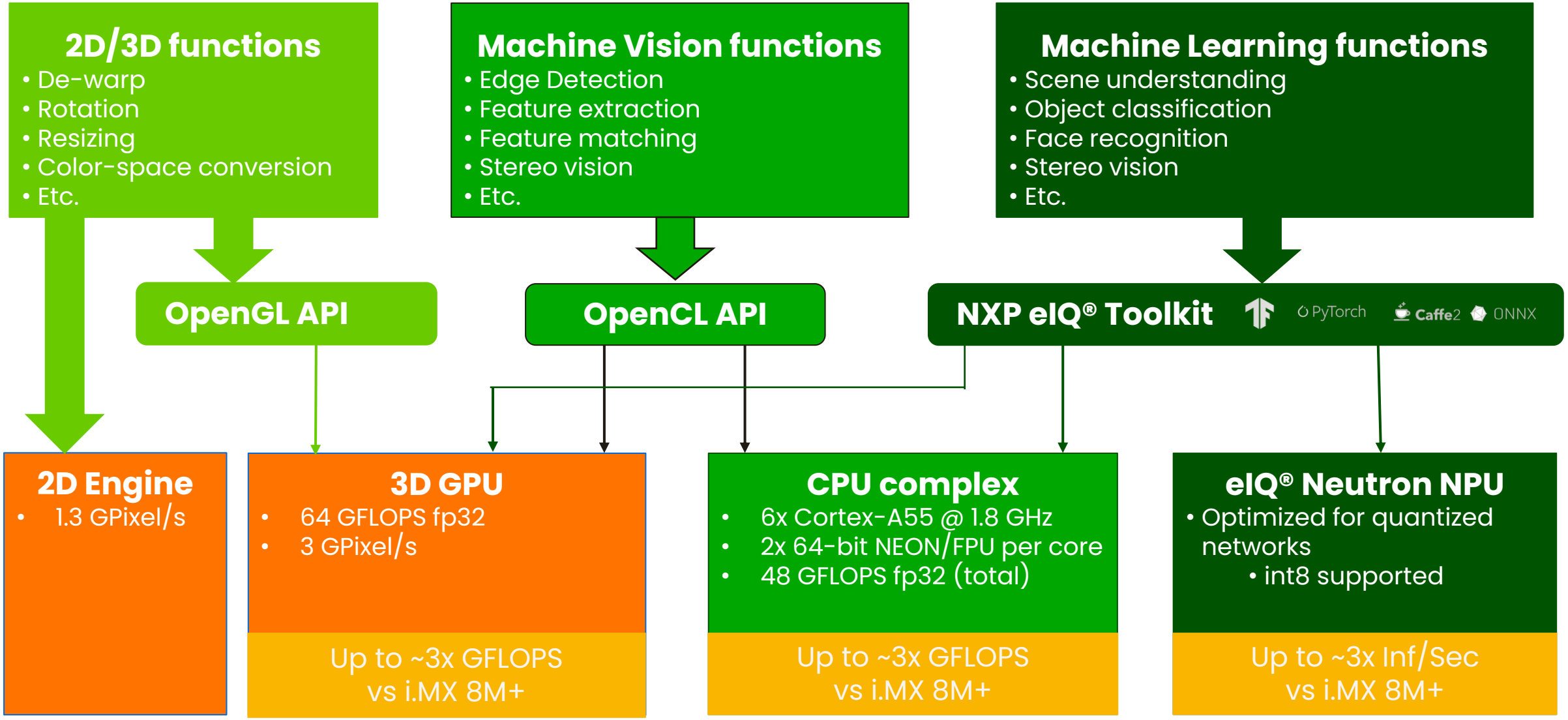
i.MX 95 Vision Processing Pipeline
Optimized for Machine Vision Applications

- Up to Single 12MP high resolution camera 4096x3072p30 / 3820x2160p60
  - 4K60P MIPI Camera with Overdrive Mode



Preliminary specifications for pre-production products under development. Schedule, features, and enablement subject to change without notice.

Embedded Vision Applications
Functional Mapping of Workloads to Compute Hardware



Learn More

NXP – Resources

- AI and Machine Learning at NXP Semiconductors (www.nxp.com/ai)
- eIQ® ML Software Development Environment (www.nxp.com/eiq)
- eIQ ML/AI Training Series (www.nxp.com/mltraining)
- eIQ Neutron Neural Processing Unit (NPU) | NXP Semiconductors (www.nxp.com/neutron)
- eIQ® ML Watermarking Model Protection Tool Demo | NXP Semiconductors

Introducing i.MX 95 Verdin EVK

Enhanced Reliability and Functional Safety

FuSa-dedicated **EEPROM** | **OctalSPI** Flash for fast boot

Expansibility and Connectivity

mPCIe and M.2 Key E slots, for Cellular Modems, Wi-Fi 6 / 6E, and other peripherals. **Micro-SD Card** slot | 8x **ADC** Inputs | 4x **I2C** via headers

High-Performance Compute

Featuring a 16GB **LPDDR5** RAM

Energy Flex Architecture

Power Consumption monitoring of SoM and Carrier Board **Control and optimization** of SoC voltage rails with NXP PF09 and PF53 **PMICs**

Display and Multimedia

Quad-Lane MIPI **DSI** | Octal-Lane **LVDS**
Audio Codec on-board | **PDM** mics

Machine Vision Capabilities

Quad-Lane MIPI **CSI** (up to 2x w/o DSI)

Rich Set of Interfaces

10-GbE port | 1x **GbE** port with **TSN** capability | 1x Automotive **1000BASE-T1** 2-Wire Ethernet **USB 3.1 Gen 1** Type C | 2x USB 2.0 Type A | 2x **CAN-FD**

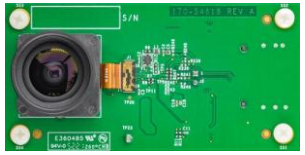
i.MX 95 Verdin EVK: HW Ecosystem

Riverdi

Toradex Capacitive
Touch Display 10.1” DSI

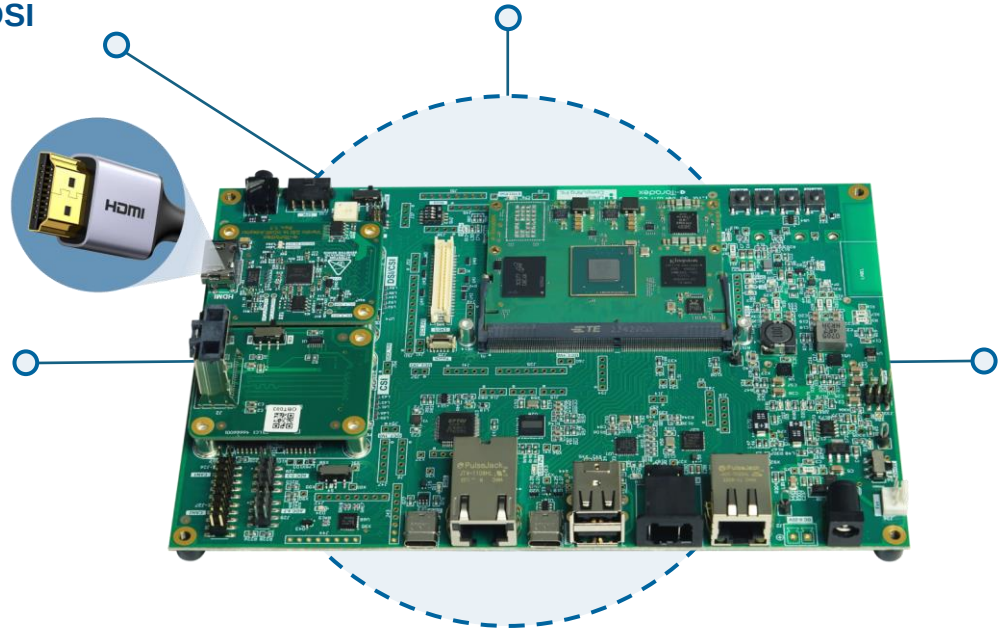
Toradex Capacitive
Touch Display 10.1” LVDS

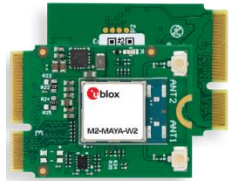
MX8-DSI-OLED1



4K MIPI CMOS (OS08A20)

Mini-SAS connector enabling NXP’s
Eco-system Peripherals





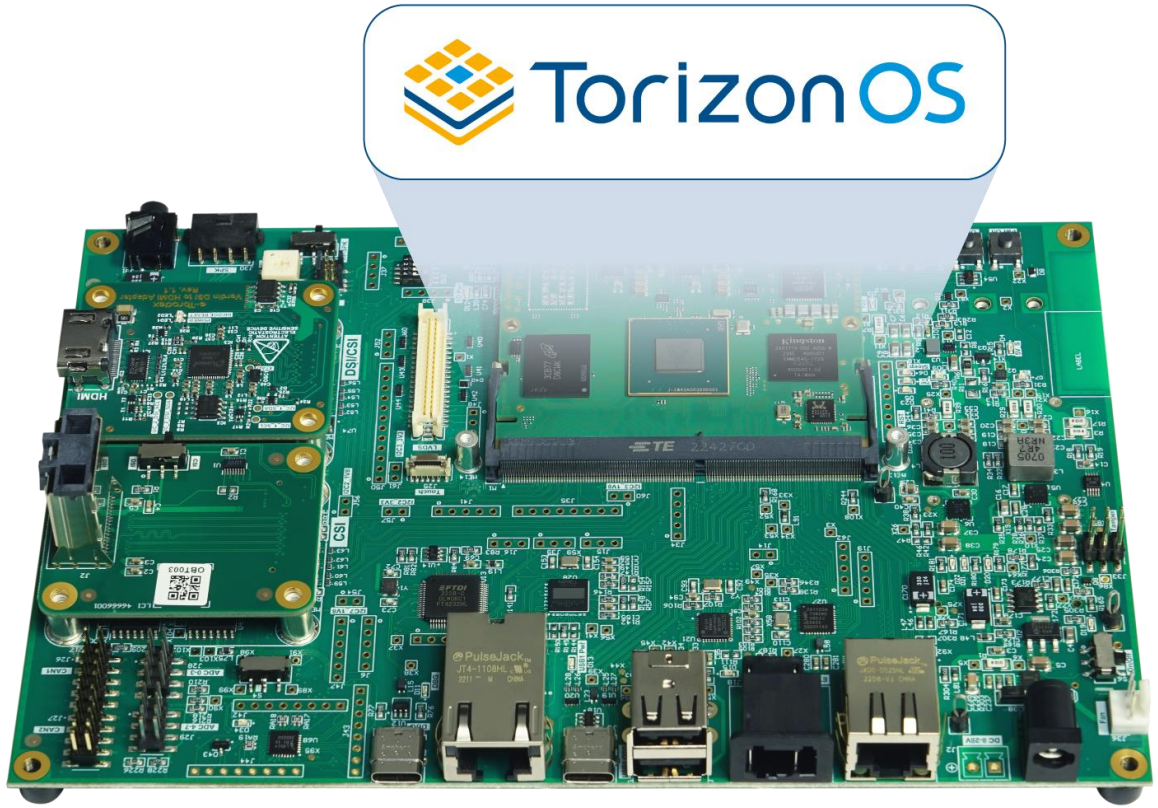
ublox

Wi Fi 6

Wi-Fi 6 (IW612) and future
Wi-Fi 6E modules
EVK will ship with u-blox
Maya-W2-W271 included

i.MX 95 Verdin EVK: SW Ecosystem

Out-of-the-box Support of a Wide Range of SW Frameworks



TorizonOS

yocto PROJECT PARTICIPANT

docker

Qt Group

CODESYS

Au-Zone Technologies Inc.

CRANK AMETEK

Toradex
Swiss. Embedded. Computing.

i.MX 95 Verdin EVK:
Enabling a Variety of Customer Use Cases

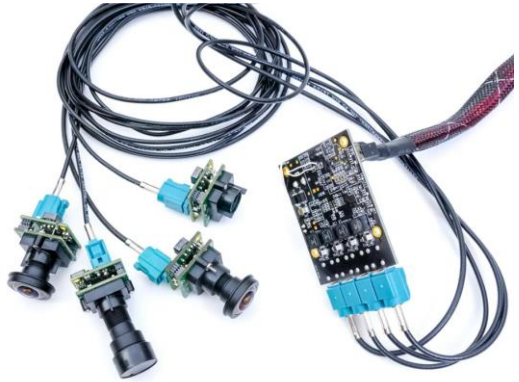
Example Automotive Use-Case: 2 Displays, 4 Cameras

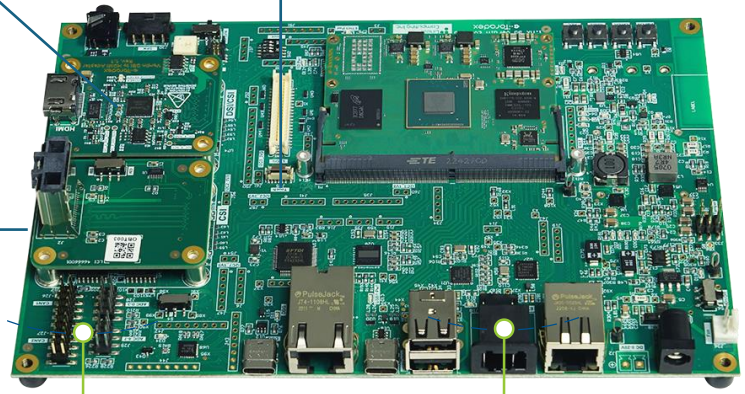
Riverdi

Toradex Capacitive
Touch Display 10.1” DSI

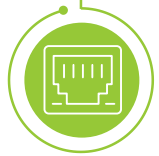
Toradex DSI
Display Adapter

Toradex Capacitive
Touch Display 10.1” LVDS





2x CAN-FD

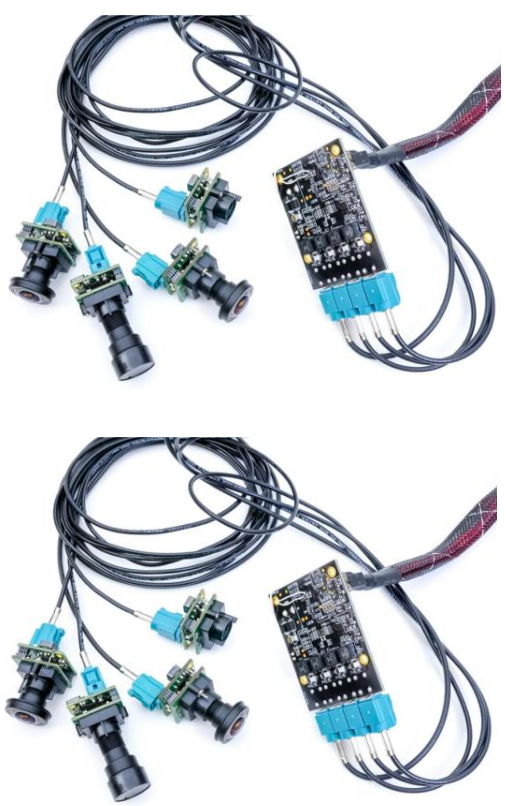


Auto-ETH port

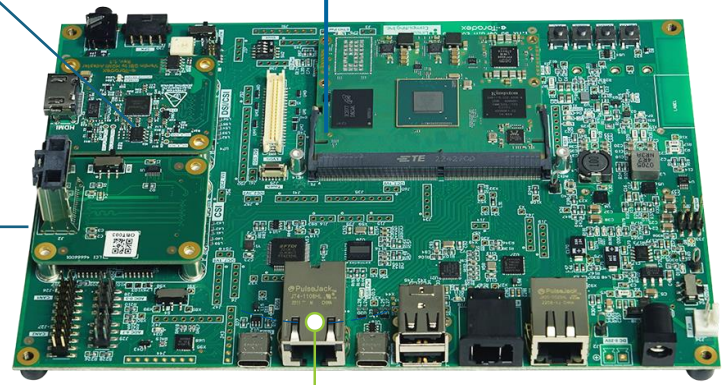
Toradex
Swiss. Embedded. Computing.

i.MX 95 Verdin EVK:
Enabling a Variety of Customer Use Cases

Example Industrial Use Case: 8 Cameras and High-Speed Networking



Toradex Capacitive
Touch Display 10.1” LVDS



10-GbE port

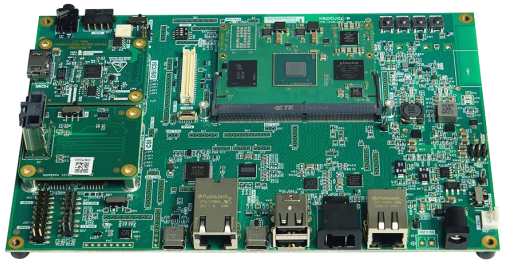
Toradex
Swiss. Embedded. Computing.

Apply for Early Access

Visit the i.MX 95 Verdin EVK landing page

Main Advantages for customers compared to official NXP EVK

- **Availability & Volumes**
  - Available and shipping now
- **Pricing at 750$ / 699€**
  - 50% the price of the official NXP EVK
- **Ready to go for prototyping**
  - Customers are free to build their own carrier board and use the SoM for prototyping
  - A migration guide towards the volume Verdin iMX95 will also be shared
- **Compact Size**
  - 200x136mm
  - Perfect fit inside a laptop bag



SCAN ME

Toradex
Swiss. Embedded. Computing.

Q&A

NXP | Toradex
Swiss. Embedded. Computing.

NXP | Toradex

Swiss. Embedded. Computing.

THANK YOU
FOR YOUR INTEREST

www.**toradex**.com | www.**torizon**.io
developer.toradex.com | **community**.toradex.com
labs.toradex.com

Torizon™