Tensilica Vision DSP Product Line Introduction

- Cadence Imaging/Vision Group
- July 2017 Update

cadence®

Application Targeted Processor Portfolio

Broad Range of Application Specific DSPs

HiFi DSPs

- Audio Pre and Post-Processing
- Voice trigger
- Noise Reduction
- Audio Encode & Decode

Fusion DSPs

- Auto Radar
- Always-alert Sensor processing
- Low-end Imaging
- Audio, Video and Speech

ConnX DSPs

- Narrow to wide band Wireless
- LTE/LTE-A/5G
- WiFi, Smart Grid
- Infrastructure & Terminals

Vision DSPs

- Image Pre and Post-Processing
- Convolutional Neural Networks
- AR/VR
- ADAS

Custom

Custom ISAs

- High Performance DSPs, NPUs, CPUs
- Application specific data types
- Custom ISA
- Special Functions

Automated User-Defined Customization

Interfaces, Instructions, State & Registers, Unique and Secure features

Xtensa® Processor Generator

Configurable and Extensible Common Foundation Technology of all Tensilica Processors

cadence®

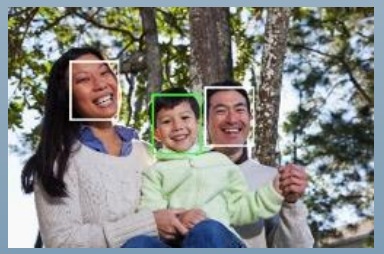
Emerging use cases: Cameras everywhere

Mobile

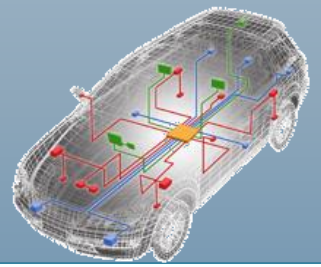
HDR

Video Stabilizer

Face Detection

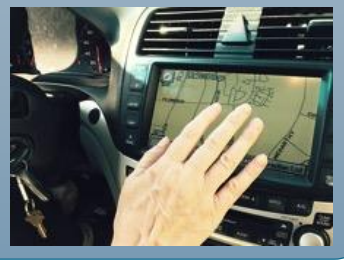


Automotive



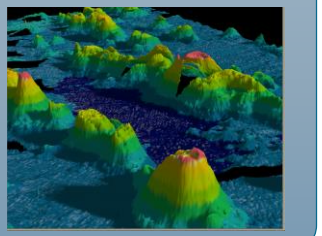
Traffic Sign Recognition

Gesture Control



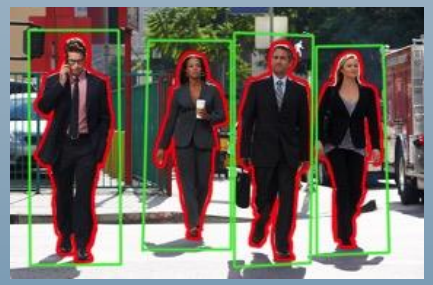
Drone

3D Vision



Security

People Detection



Wearables and IoT

cadence®

Imaging and computer vision for embedded systems

Imaging Is Still Important!	Required before **any neural network processing**
	Higher resolution: 32MP in Mobile, 4K video, 720p to 1080p in Automotive
	Higher frame rates: 240, 120fps
	Killer use cases: Stereo camera on mobile devices, high FOV (120, fisheye lens), 360^0 camera
Computer Vision	Traditional computer vision is based on feature detection
	Neural networks are better at many tasks and are being applied more today
	Compute, power consumption for embedded devices is a challenge

The Basics of Real-Time Neural Networks:

Tensilica® Vision C5 DSP aimed at *inference* applications in embedded systems

Training: Runs once per database, *server-based*, **very** compute intensive

Server Farm

Labeled dataset

Selection of layered network

Iterative derivation of coefficients by stochastic descent error minimization

10^4-10^{10} TMAC/s

Set of coefficients (1M-1B weights)

Embedded

0.5 to 10TMAC/s

Single-pass evaluation of input image

Most probable label

Deployment ("Inference"): Runs on every image, device based, compute intensive

Most important for embedded systems: Power

cadence

Vision DSPs: Family of Imaging and NN DSPs

Vision P5 DSP

Imaging Optimized

Optimized for Computational Photography Algorithms, Very **General Purpose Imaging**

64 MACs

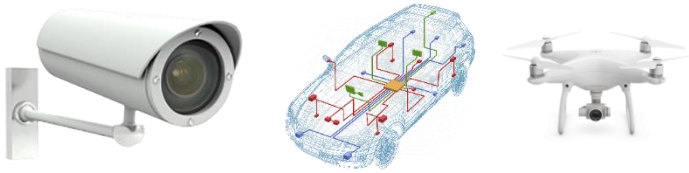
Vision P6 DSP

Imaging Plus NN Vision

General Purpose Imaging plus Occasional Use **Neural Network** recognition

256 MACs

Vision C5 DSP



Heavy-duty, Always-On NN

Full-time NN

1024 MACs

cadence

Tensilica® Vision C5 and Vision P6 DSPs:
Cadence Addressing All Market Segments for Neural Networks

Processing Power

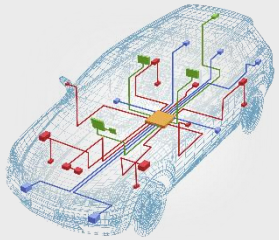
Up to 10TMAC/sec

1TMAC/sec

<200 GMAC/sec

Automotive (towards autonomous)

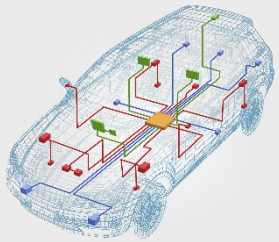
Multiple Vision C5 DSPs



Runs multiple NNs all the time

Surveillance / Automotive (semi-autonomous)

Vision C5 DSP



Runs a couple of NNs all the time

Mobile

Vision P6 DSP

Runs a NN once in a while

cadence®

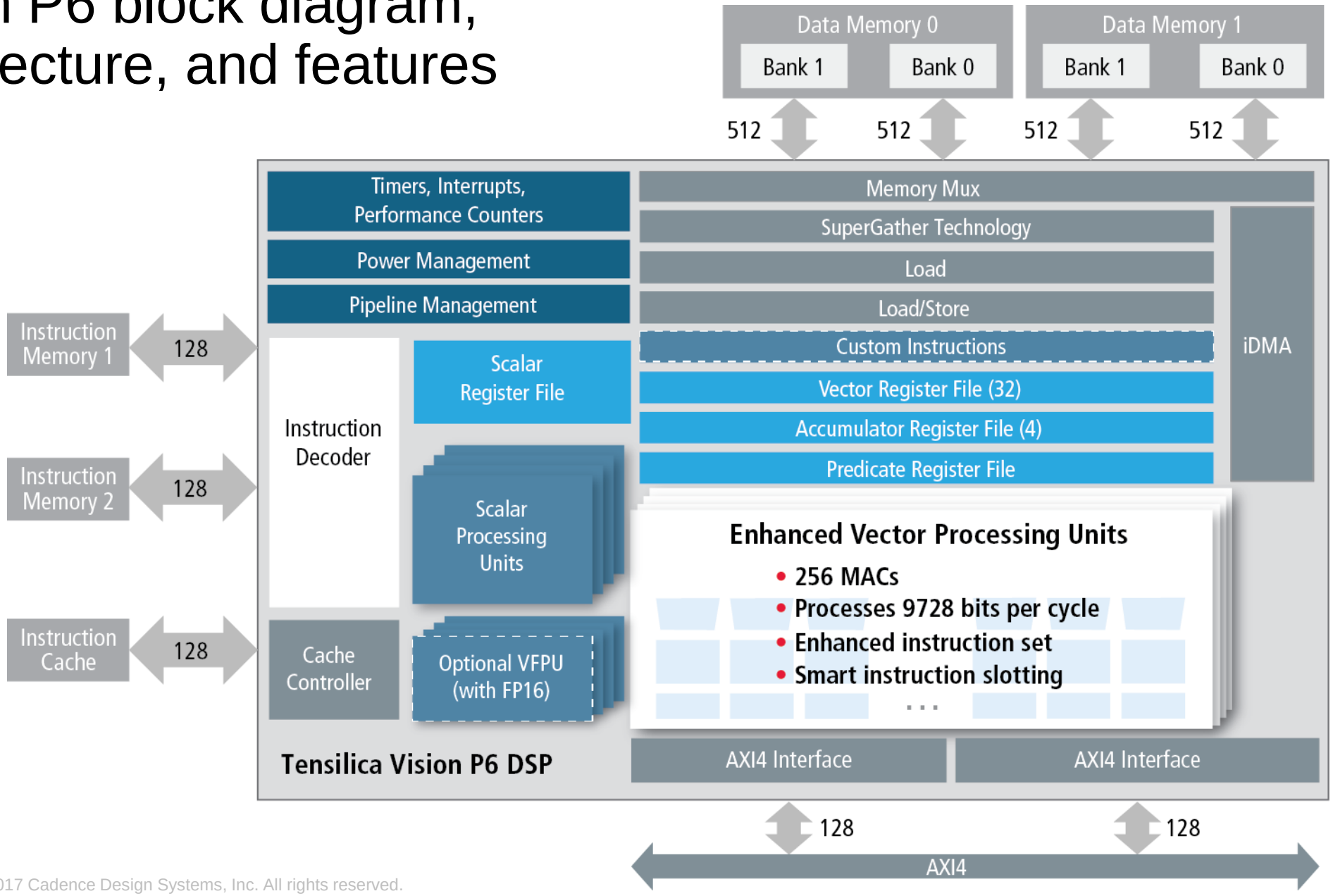
Vision P6 DSP: A complete imaging/vision and CNN processor

Neural Network Performance	• Up to 4X neural network performance
Imaging and Vision Benchmarks	• Up to 4X performance in well-known benchmarks
Multiply Accumulate (MAC)	• 4X MAC count
Vector Floating-Point Support	• 32-way SIMD vector FPU with 16-bit (FP16) • Easy GPU code porting
Software compatible	• Software compatible to Vision P5 • Up to 25% improvement with Vision P5 software

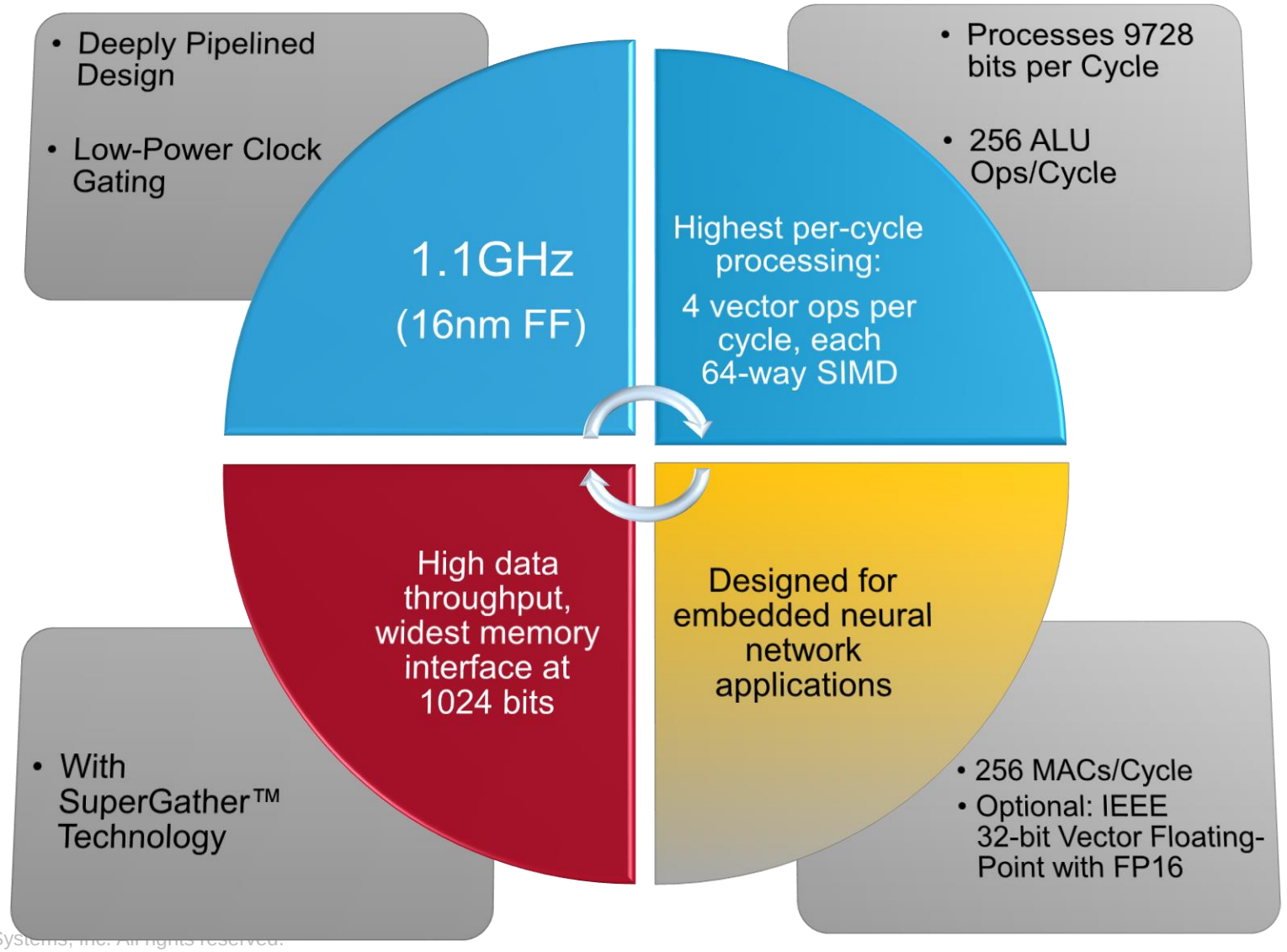
Relative numbers presented are in comparison with Tensilica Vision P5 DSP

cadence®

Vision P6 block diagram, architecture, and features



Vision P6 DSP: Flagship performance



Tensilica® Vision C5 DSP for Neural Networks

Complete, standalone DSP that runs all layers of NN (convolution, fully connected, normalization, pooling…)

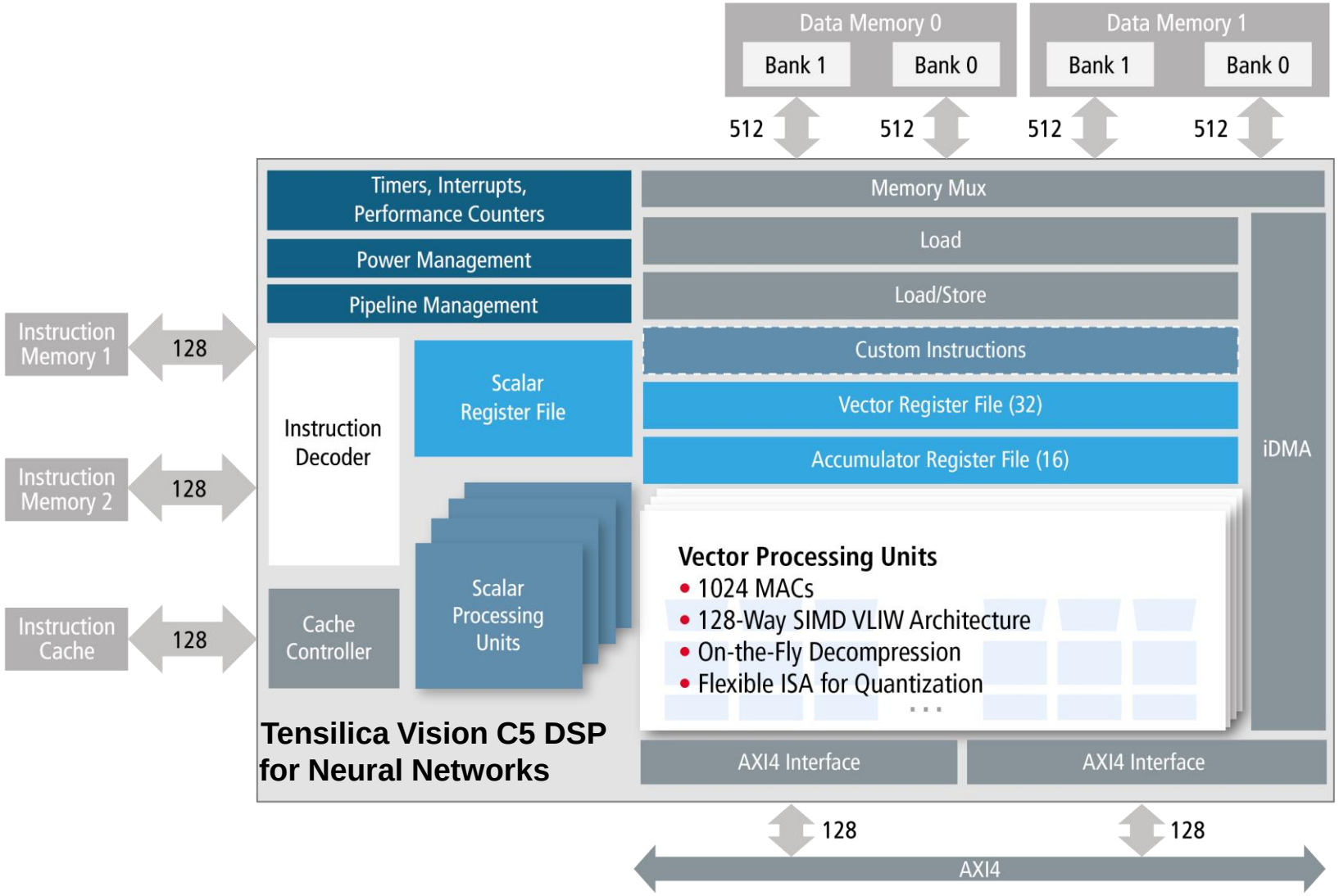
Building a DSP for changing NN field – general purpose and programmable

Not a “hardware accelerator” paired with a vision DSP, rather a dedicated, NN-optimized DSP

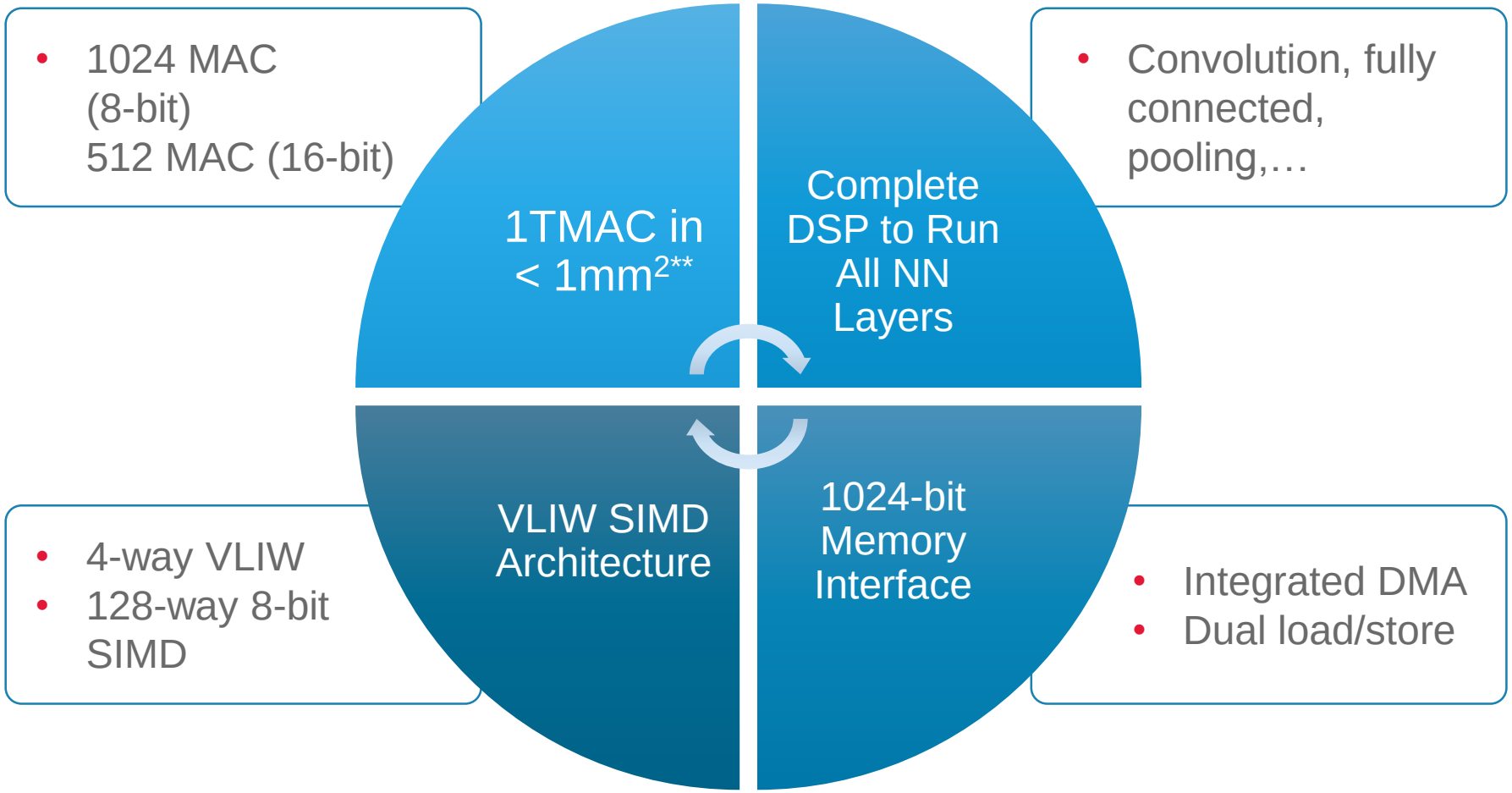
Architected for multi-processor design – scales to multi-TMAC/sec solution

Same proven software tool set as Vision P5/P6 DSP

Tensilica® Vision C5 DSP for Neural Networks



Tensilica® Vision C5 DSP for Neural Networks



**16nm

Features and Benefits of Vision C5

Feature	Benefit
First stand-alone, dedicated CNN DSP	Faster and lower power than conventional CPU/GPU/DSP
Programmable and extensible	Flexible and future proof, can do anything vs hardwired accelerators Support for new layers as they evolve
All layers run on Vision C5 compared to running the convolutions on the accelerator Normalization, max pooling	No need to move data between Vision C5 and main vision/imaging DSP Simple programming model Simplifies hardware architecture
1 TMAC in <1 mm2 in 16nm Vision C5 is architected for multi-processor clusters	Computational capacity and scalability for all CNN applications in surveillance and automotive

Vision Product Comparison Chart

		Vision P5	Vision P6	Vision C5
Use Case		**Imaging**	**Imaging and low-end CNN**	**Mid and high end CNN (Always-on CNN)**
MACs (Higher MAC = Higher Compute)	8x8	64	256	1024
	8x16	64	128	
	16x16	32	64	512
Vector Floating Point Unit	16b Half Precision	No	32 way SIMD (optional)	No
	32b Single Precision	16 way SIMD (optional)	16 way SIMD (optional)	No
MAX SIMD Width		64 way 8bit	64 way 8bit	128 way 8bit
Scatter Gather - Optimal LD/ST for dispersed pixels		Yes	Yes	No
Coefficient Decompression - Saves Memory Bandwidth		No	Yes	Yes
Data rearrangement – Efficient switch between vectorization schemes		Limited	Limited	Extensive

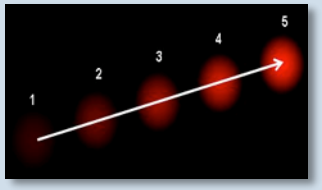
Broad software ecosystem for imaging/vision applications

Libraries

- Familiar programming model
- 1000 optimized library functions
- CNN, OpenVX, and OpenCV-based library functions

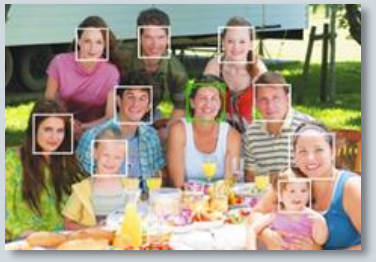
Kernels

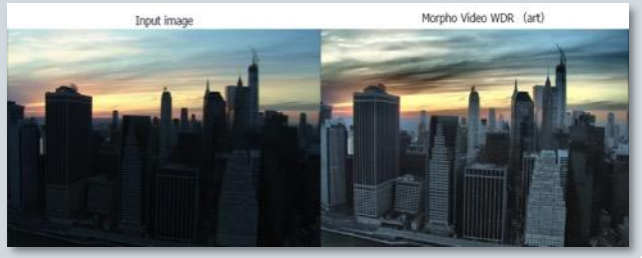
- High-performance Sobel, Median, Gaussian filters
- CNN Lib: Convolution, RELU
- SIFT, SURF, Harris Corner: Detection algorithm
- HOG, HAAR: Object detection and classification
- LK: Optical flow



Applications

- HDR, video WDR
- Image stabilization
- Face/people detection
- Face recognition
- CNN: Vehicle detection





XI-Library: Accelerates Commonly used OpenCV functionality

XI-Library

- Tile based processing for chaining of functions in local memory to efficiently use memory bandwidth
- Specific kernel sizes, data types and modes as part of API to avoid excessive checking inside the function for every tile
- Function works on image planes to avoid frequent deinterleaving of interleaved image data

XI Library Deliverables

- XI Library source code workspace: Delivered with XPG
- XI Library performance
- XI Library user's guide

XI-Library Example Functions

- Image Proc Modules: Convolution functions, Geometric Functions,…
- Features Modules: Feature Detection, Feature Descriptor
- Motion Analysis & background analysis
- Core modules: Binary elements, bitwise operation, vector operations…

Cadence OpenVX for Vision P DSP

Fully compliant to OpenVX 1.1

Highly optimized kernels leveraging the Cadence XI Library
- Performance benchmarks provided

Automatic tiling, DMA, local memory management, scatter-gather kernels
- Via provided Graph Mapper and Runtime
- Automatic overlapping of data transfer (DMA) and compute

Supported on Tensilica Vision P5 and Vision P6 processor cores
- Compliance tested for 2-core system with Xtensa host running Linux 4.3 connected to Vision DSP

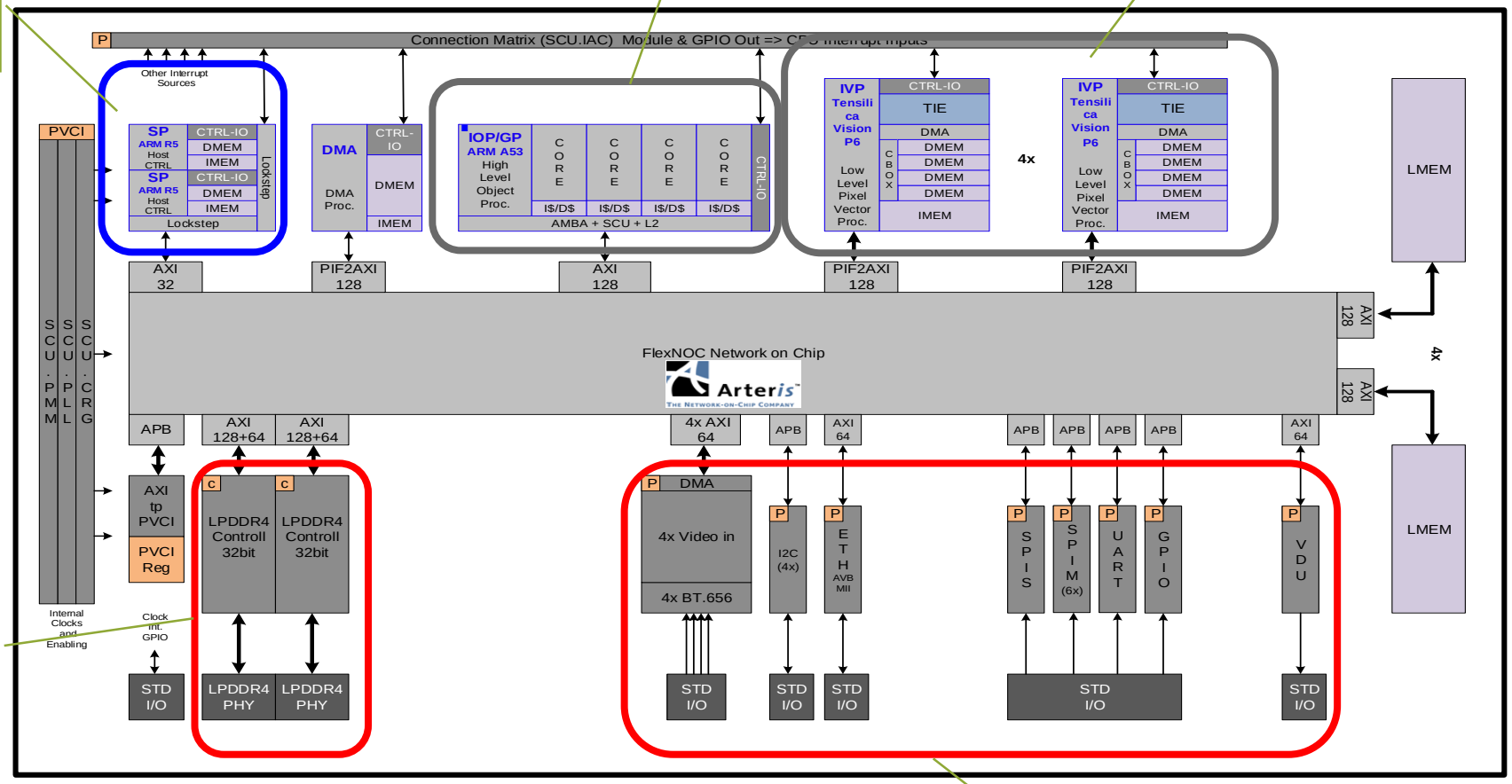
Vision DSP IP license includes source code for all APK modules

Cadence & Partner Developed Full-Speed Development Platform

Lock Step ARM R5

Quad ARM A53

Quad Vision P6



Memory I/F

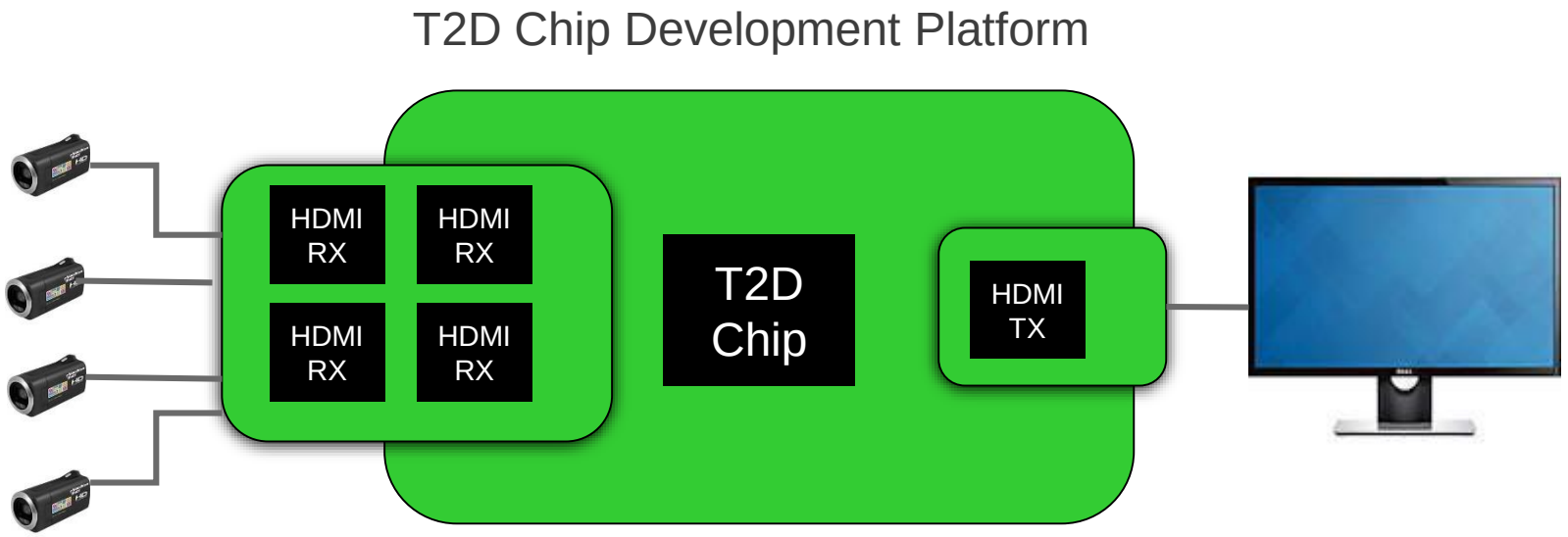
IO I/F

Imaging Performance

- 1024 MAC per/cycle: 512GMAC (@500MHz)
- 1024 ALU per cycle (8bit)
- 1280 8 bit fixed point operation per cycle
- 128/256 GFLOP Single precision/half precision

cadence

Cadence & Partner Developed Full-Speed Development Platform



T2D Development Platform
-T2D Chip @500MHz
-Four HDMI I/F for Camera Connection
-HDMI Monitor output

Board available in September 2017

Vision DSP Partner Ecosystem (Public)

Morpho

- WDR (wide dynamic range)
- Super video image stabilization

sensory

- Face and voice authentication
- Face detection

Almalence

- Super-resolution zoom, HDR
- Camera processing

KHRONOS GROUP
CONNECTING SOFTWARE TO SILICON

ArcSoft®

- Mobile imaging experts
- Low-light enhancement
- Dual-camera solutions

Dream CHIP

Vvurmi SYSTEMS
creating waves for tomorrow

- ADAS suite
- Fog removal, object detection
- System integrator

MULTICOREWARE

- CNN neural networking
- Imaging algorithm expertise

IRIDA LABS

- Imaging and vision experts
- Low light enhancement
- Advanced noise reduction
- Face detection

Summary

Imaging DSP and Neural Network DSP

- Market needs both Imaging DSP and Neural Network DSP
- Imaging processing requirement continues to increase: higher resolution and dual sensor
- Neural Network are evolving – computational capacity continue to increase

Vision C5 DSP for Neural Network

- Complete, standalone DSP that runs all layers of NN : Not an accelerator
- General purpose and programmable
- 1 TMAC/Sec computational capacity: 4X MAC capacity compare to Vision P6

Vision P6 DSP

- Up to 4X neural network performance compare to Vision P5
- 4X MAC capacity
- Easy migration of GPU floating-point code, with optional 16-bit floating-point unit
- Up to 4X performance improvement on well-known imaging/vision benchmarks

Relative numbers presented are in comparison with Tensilica Vision P5 DSP