Vision and AI DSPs for Ultra-High-End and Always-On Applications

Pulin Desai
Group Director, Tensilica Vision and AI Product Marketing
pulin@cadence.com

cadence®

Cadence Tensilica Processor and DSP IP Business

2021 embedded VISION summit

TENSILICA® CUSTOMERS

7B+ Processors SHIPPING Annually

DSP LICENSING REVENUE

#1 DSP IP LICENSING REVENUE

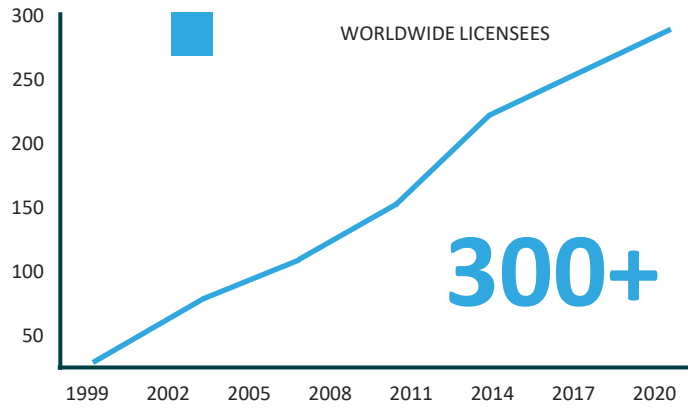
Processor LICENSING REVENUE

#2 Processor IP LICENSING REVENUE

SEMICONDUCTORS

19 19 of the Top 20 SEMICONDUCTOR VENDORS USE TENSILICA PROCESSORS

TENSILICA LICENSEES

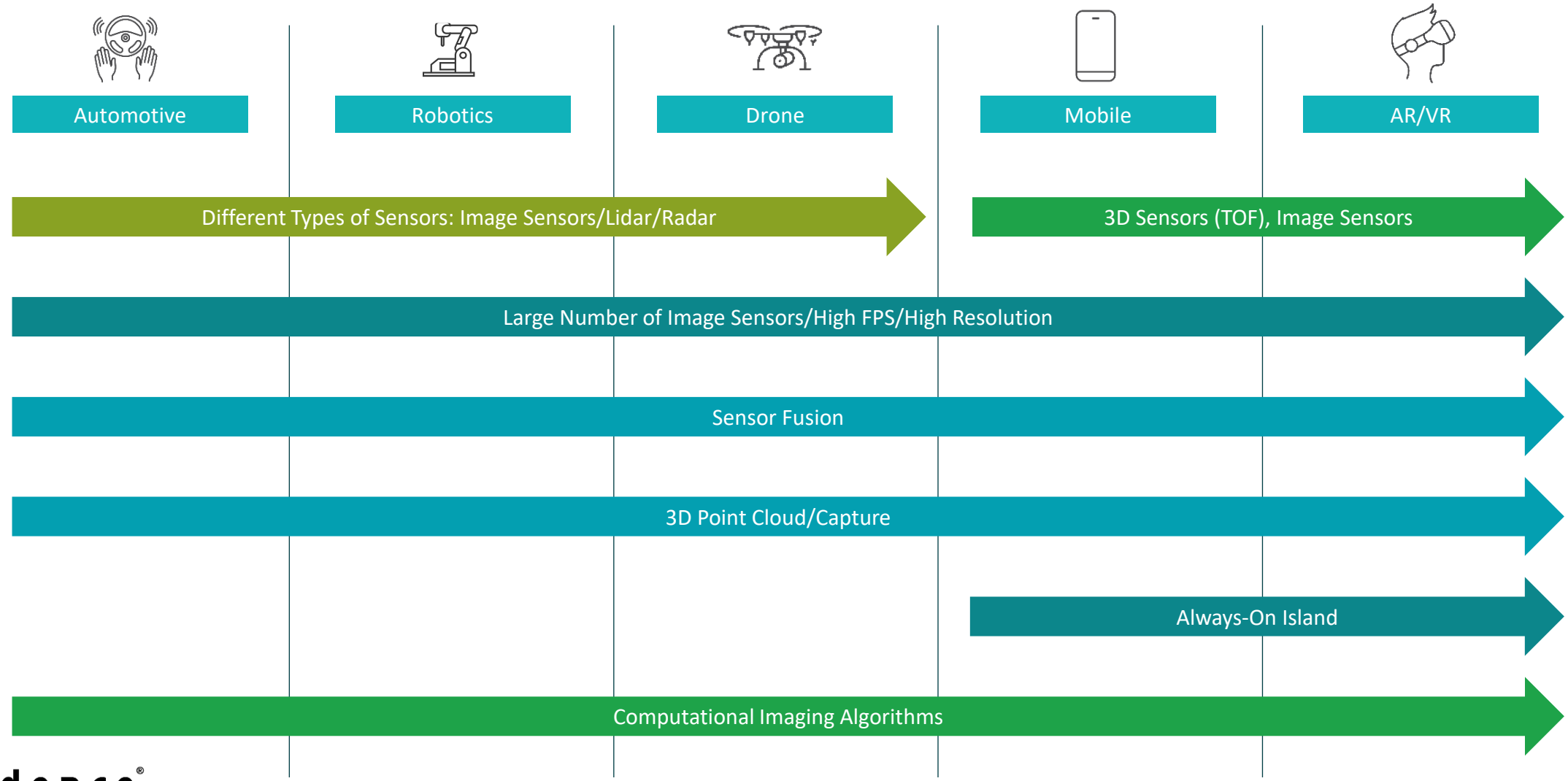


GLOBAL ECOSYSTEM

200+ ECOSYSTEM PARTNERS

cadence®

Major Trends in: Automotive/Robotics/Drone/Mobile/AR/VR



Need for Speeding Up Computer Vision Algorithms

Examples of image processing and computer vision (CV) applications

- Multi-frame HDR imaging
- Super-resolution imaging
- Bokeh effect

Examples of SLAM and CV applications

- 3D object detection and tracking
- Trajectory estimation

Constituent CV algorithms in these applications

- Feature detection, descriptor matching
- Perspective transformation
- Circle, bilateral filtering

Typical processing time:

- VGA resolution: typical 20 to 30ms/frame
- HD resolution: typical >200ms/frame

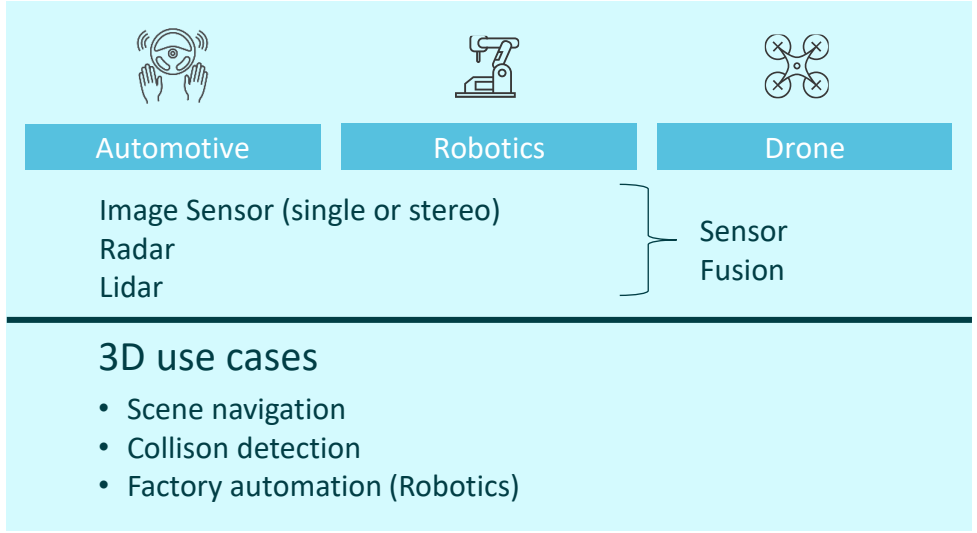
New architecture needed to speed up CV algorithms

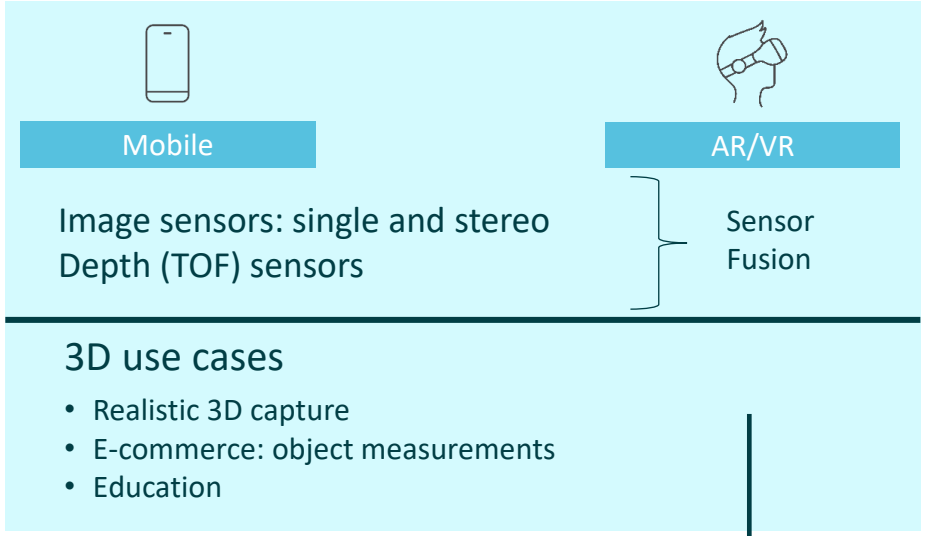
HDR Imaging

Low Light

Source: Visidon

Multiple Sensors and 3D Capture





Multiple sensors: requires sensor fusion

- Requires heavy floating-point and linear algebra calculations
- Object registration and key point detection

3D capture use cases

- Requires heavy floating-point and linear algebra calculation

Functional Measurements

Mobile and AR

- No need to wake up main CPU and compute complex, display, modem until user is authenticated
- User is authenticated using: voice command, face detection, fingerprint recognition (under the glass sensors) – requires AI
- User authentication block is always on to detect activity
- Low power mode in uW followed by mW mode to run authentication before turning on the rest of the device

Smart sensors

- Low power, always on, battery powered (Vision + AI workload)
- Examples: AI-IoT
  - Smart doorbell
  - Camera for object detection in kitchen appliance
  - Smart printers for authentication

Introduction to Tensilica Vision Q8 and Vision P1 DSPs

cadence®

7th-Generation Flagship Tensilica Vision Q8 and Vision P1 DSPs

Extend the product portfolio with 1024-bit SIMD Tensilica® Vision Q8 and 128-bit Vision P1 DSPs

- 2 new DSPs offer a complete portfolio of Vision DSPs from the high end (3.8TOPS) to low end (400GOPs)

7th-generation flagship Tensilica Vision Q8 DSP: 1024-bit SIMD

- Targeted at high-end mobile and high-resolution/high-end automotive markets
- >2X the computer vision/AI/FP performance compared to previous-generation Vision Q7 DSP
- Single core offers 3.8TOPS performance, 192GFLOP floating-point performance (FP32)

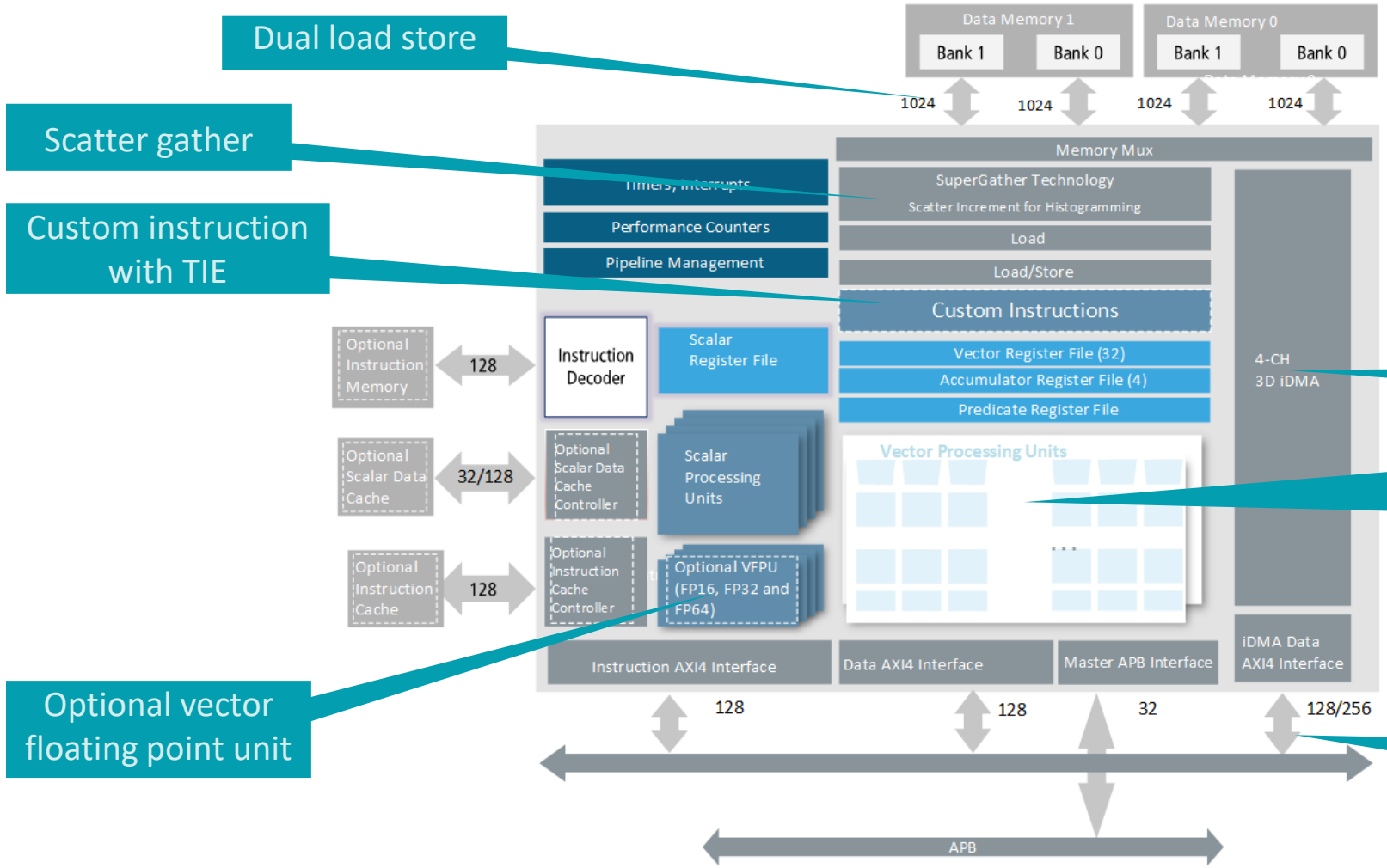
Tensilica Vision P1 DSP: 128-bit SIMD

- Targeted for always-on applications and smart sensors
- Offers one-third the area and power, 20% higher frequency compared to Tensilica Vision P6 DSP
- >0.256 TOPS AI performance

Both DSPs based on same SIMD and VLIW architecture, and instruction set used by highly successful Vision P6/Q7 DSPs

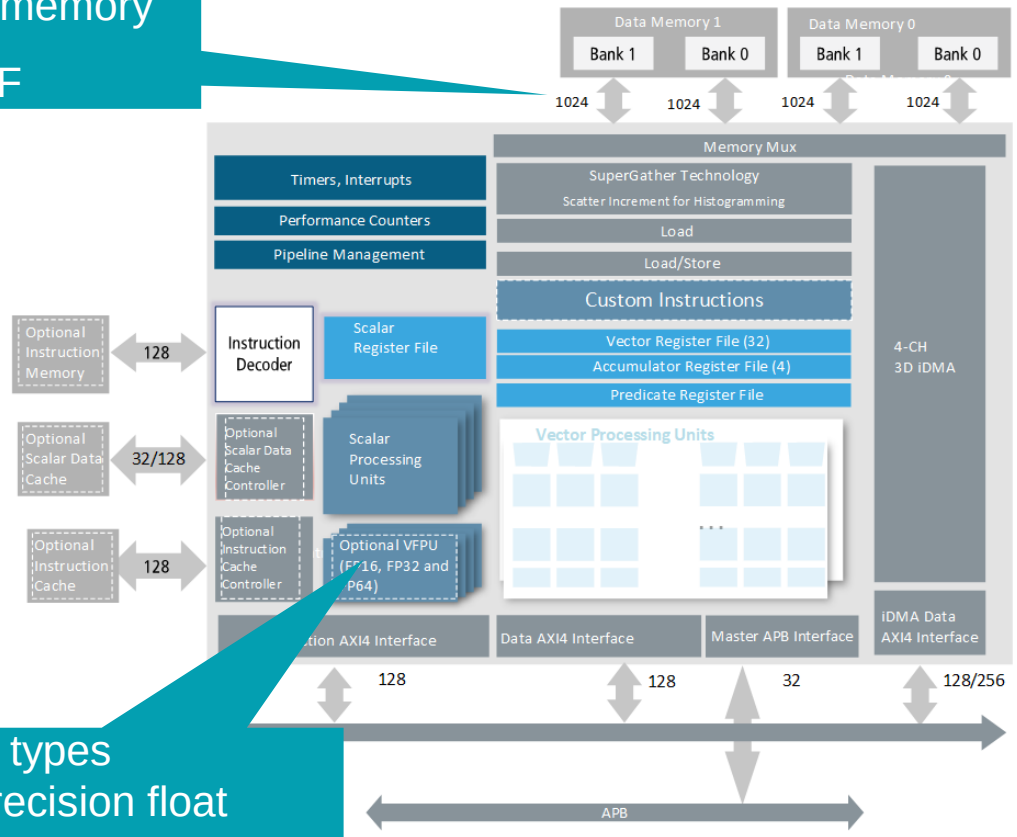
- Same software (Vision and neural network compiler) tools and library from the low end to the high end of the Vision DSP portfolio
- Access to larger software partners

Vision DSP Architecture: Common Across All Tensilica Vision DSPs



Tensilica Vision Q8 DSP: Base Architecture Improvements

2048-bit memory I/F

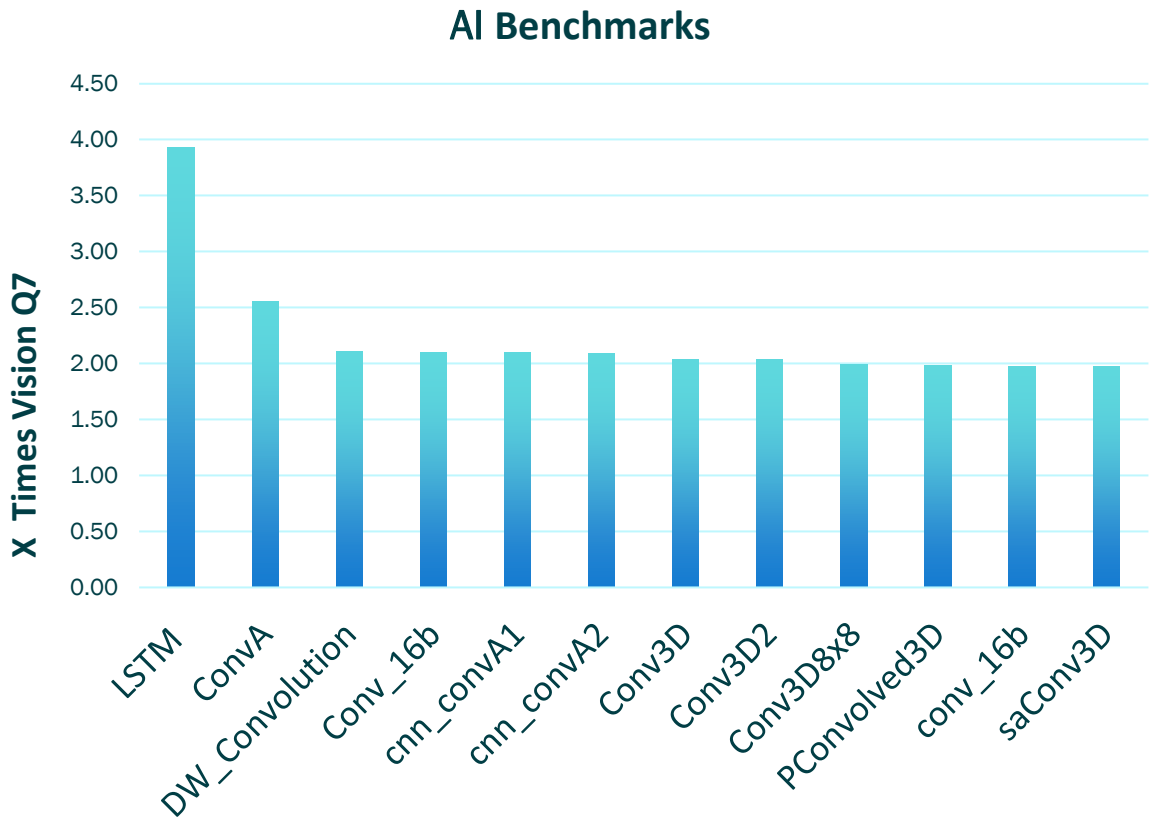


- New data types
- Double precision float (FP64)
- Complex float (FP64/FP32/FP16)

- **1024-bit** SIMD
- **2048-bit** data memory I/F
- 2X SIMD compared to Tensilica® Vision Q7/P6 DSPs translates into 2X performance at same MHz
- Allows SOC designers to use lower frequency and still achieve same performance as Tensilica® Vision Q7 and P6 DSPs
  - Leverage lower-level voltage rails/libraries
- New data types:
  - Double-precision float (FP64)
  - Complex float (FP64/FP32/FP16)
- Increased accumulator size for better accuracy
- Power measurement features for DVFS

Tensilica Vision Q8 DSP: AI Enhancements

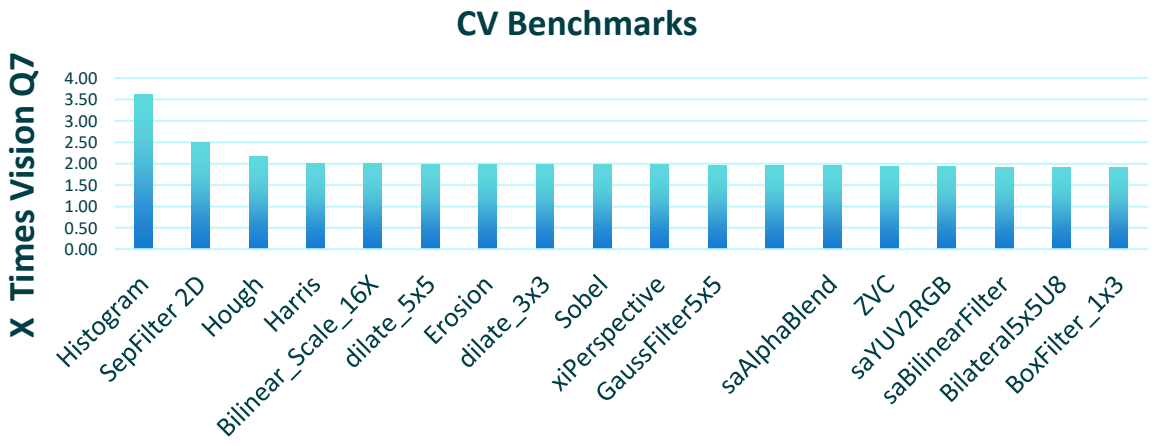
- 1K 8-bit MAC
- 256 16-bit MAC
- ISA optimized for efficient use of 1024-bit SIMD for multiple of 16 size depth convolution
- Enhancements for non-convolutional neural network layers
  - Example for leaky / parametric ReLU
- Multiply-accumulate operation improvements for asymmetric quantization

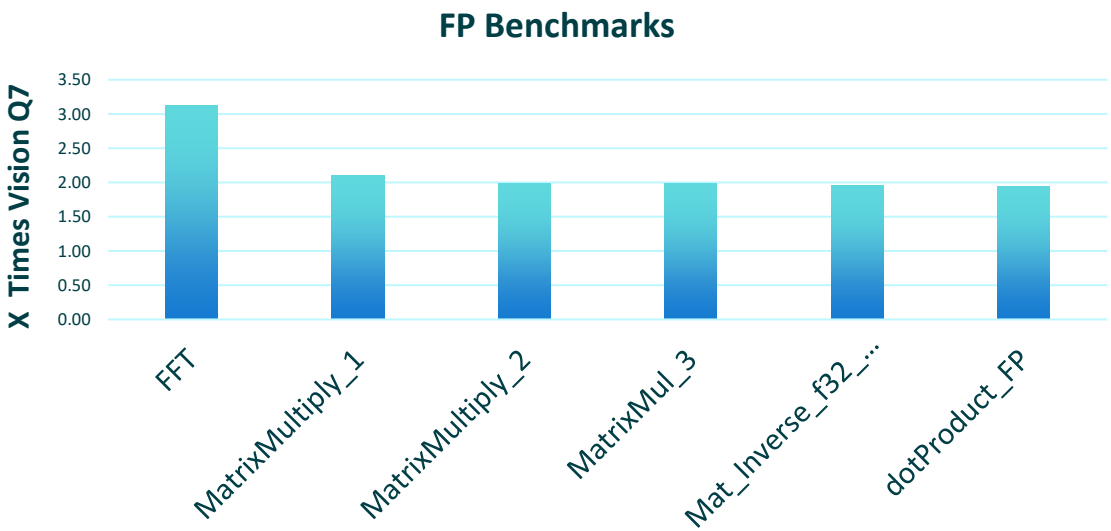


Tensilica® Vision Q8 DSP shows 2X performance improvements over Vision Q7 DSP in AI benchmarks

Tensilica Vision Q8 DSP: CV and FP Enhancements

- OpenCL and Halide performance improvements
  - Accumulator optimized for compiler requirements
- Multiply variants to improve filter performance (>2X performance)
- >2X FP64, FP32, and FP16 performance compared to Vision Q7 DSP (SLAM and linear algebra)
- Complex floating-point support for FP64, FP32, FP16
- FFT enhancements with ADDSUB (FP32, FP16)



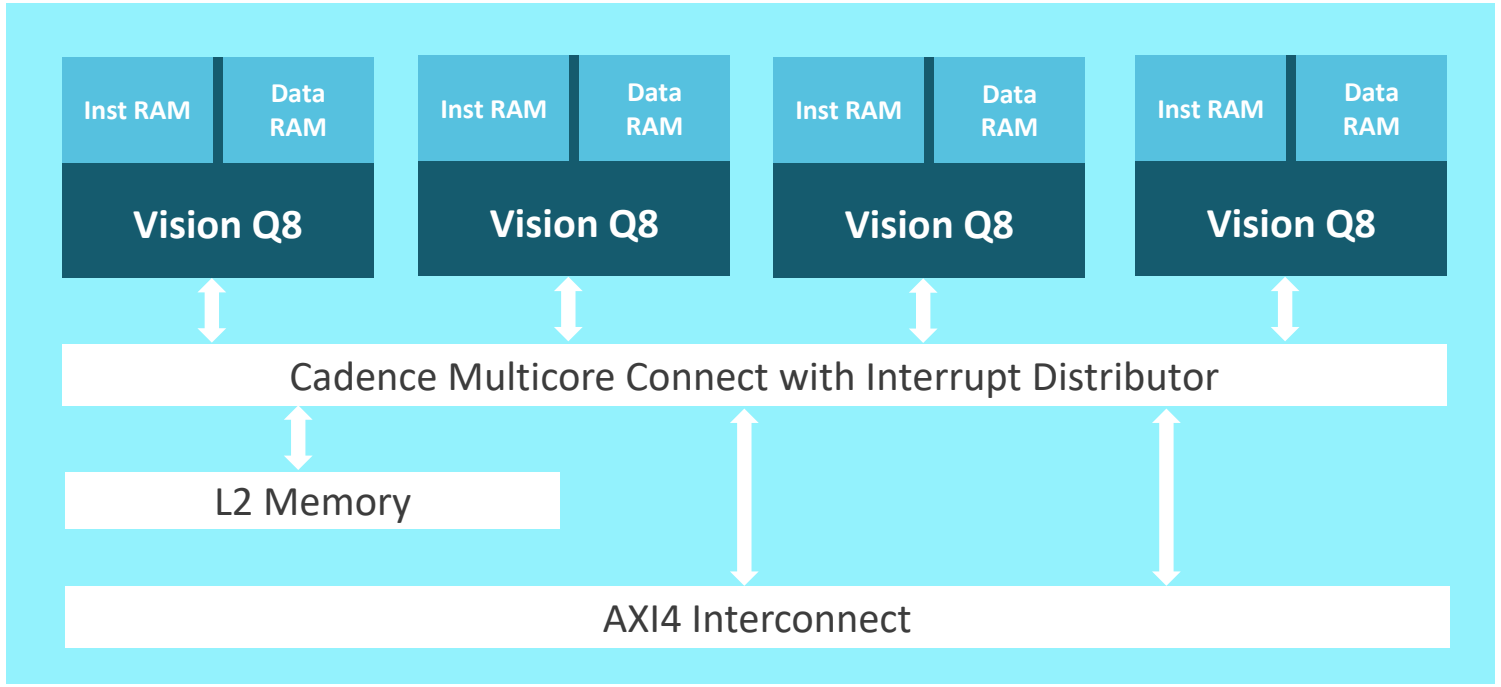


Maximum Performance of Tensilica Vision Q8 and Q7 DSPs

	Vision Q7	Vision Q8
SIMD width	512	1024
FP64 operations	16	32
FP32 operations	32	64
FP16 operations	64	128
Complex float for FP64, FP32, and FP16	NA	Yes
8-bit MAC	512	1024
16-bit MAC	128	256
SLAM acceleration	Yes	Yes

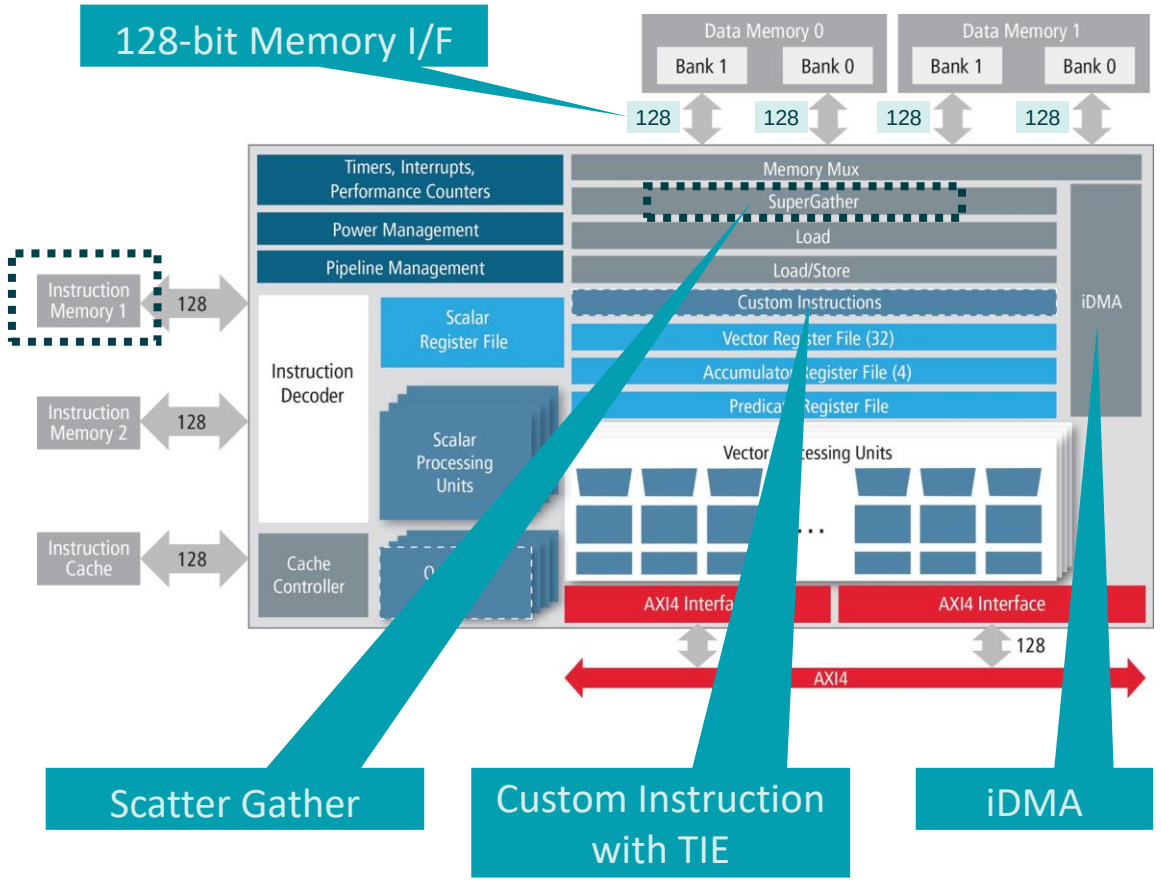
- Maximum configurations for both Vision DSPs
- Both Tensilica® Vision DSPs can be configured with lower FP and MAC count providing full flexibility

Multi-Core Solution with Tensilica Vision Q8 DSP



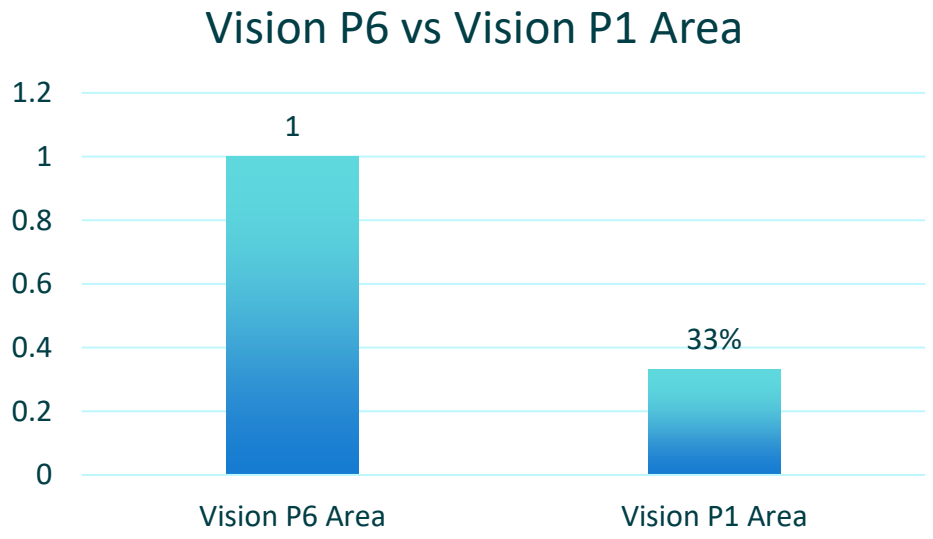
- Two- or four-core Tensilica® Vision Q8 DSP multicore from Cadence
- Cadence provides complete subsystem design
- Four-core Vision Q8 DSP offers 4K 8-bit MAC for AI
- ~800 GFLOP of FP32 performance

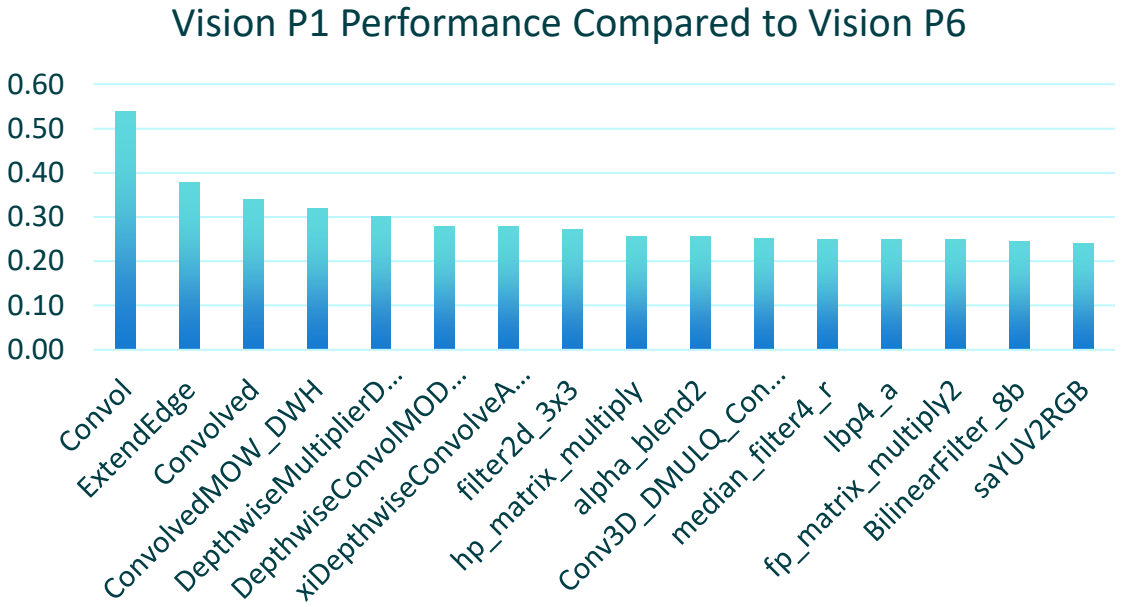
Tensilica Vision P1 DSP: Low-Power, Highly Optimized Vision and AI Core for Always-On and Smart Sensors



- Target market: always-on mobile, smart sensors, under screen mobile
- Offers up to 400GOPS
- 128-bit SIMD, 256-bit memory interface
- 128 8-bit MAC: low-end AI (lower MAC available)
  - ¼ SIMD compared to Vision P6 DSP but ½ MAC
- 1/3 area and power plus 20% higher frequency compared to Tensilica® Vision P6 DSP
- Instruction set compatible with Vision P6 DSP
- Same memory AXI interface, advance iDMA as Vision P6 DSP
- Same software libraries as other Tensilica Vision DSPs
- TensorFlow Lite Micro support
- Architecture optimized for small memory footprint and operation in low power mode

Tensilica Vision P1 DSP Performance and Area Compared to Tensilica Vision P6 DSP





- 1/3 area compared to 512-bit SIMD Tensilica® Vision P6 DSP
- Performance up to one-half compared to Vision P6 DSP with one-quarter SIMD width of Vision P1 DSP

Maximum Performance of All Tensilica Vision P6 and P1 DSPs

	Vision P1	Vision P6
SIMD Width	128	512
FP32 Operations	4	16
FP16 Operations	8	32
8-bit MAC	128	256
16-bit MAC	32	64
SLAM Acceleration	No	No

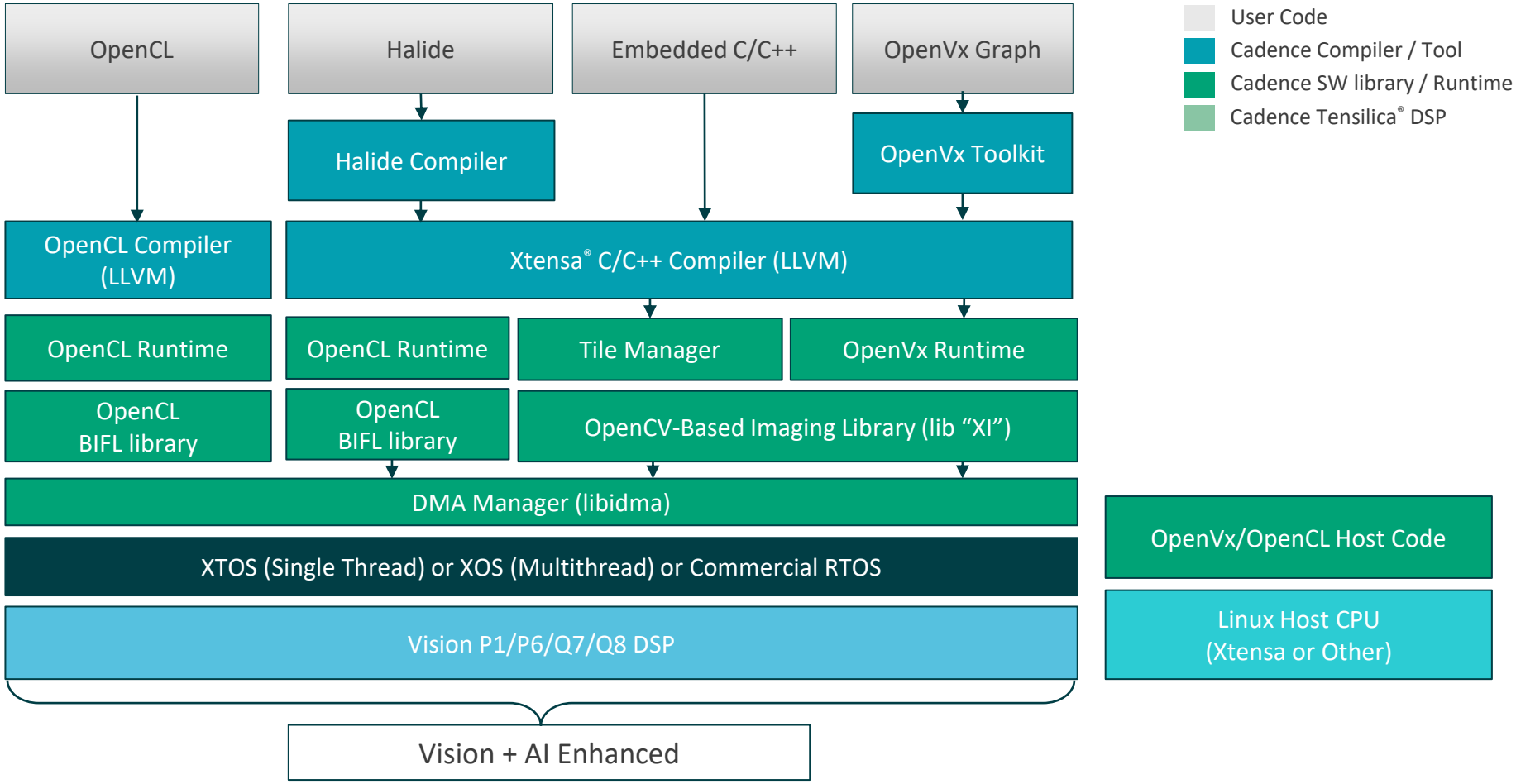
- Maximum configurations for both Tensilica® Vision DSPs
- Both Vision DSPs can be configured with lower FP and MAC count providing full flexibility

Software Migration and ISO 26262 Readiness

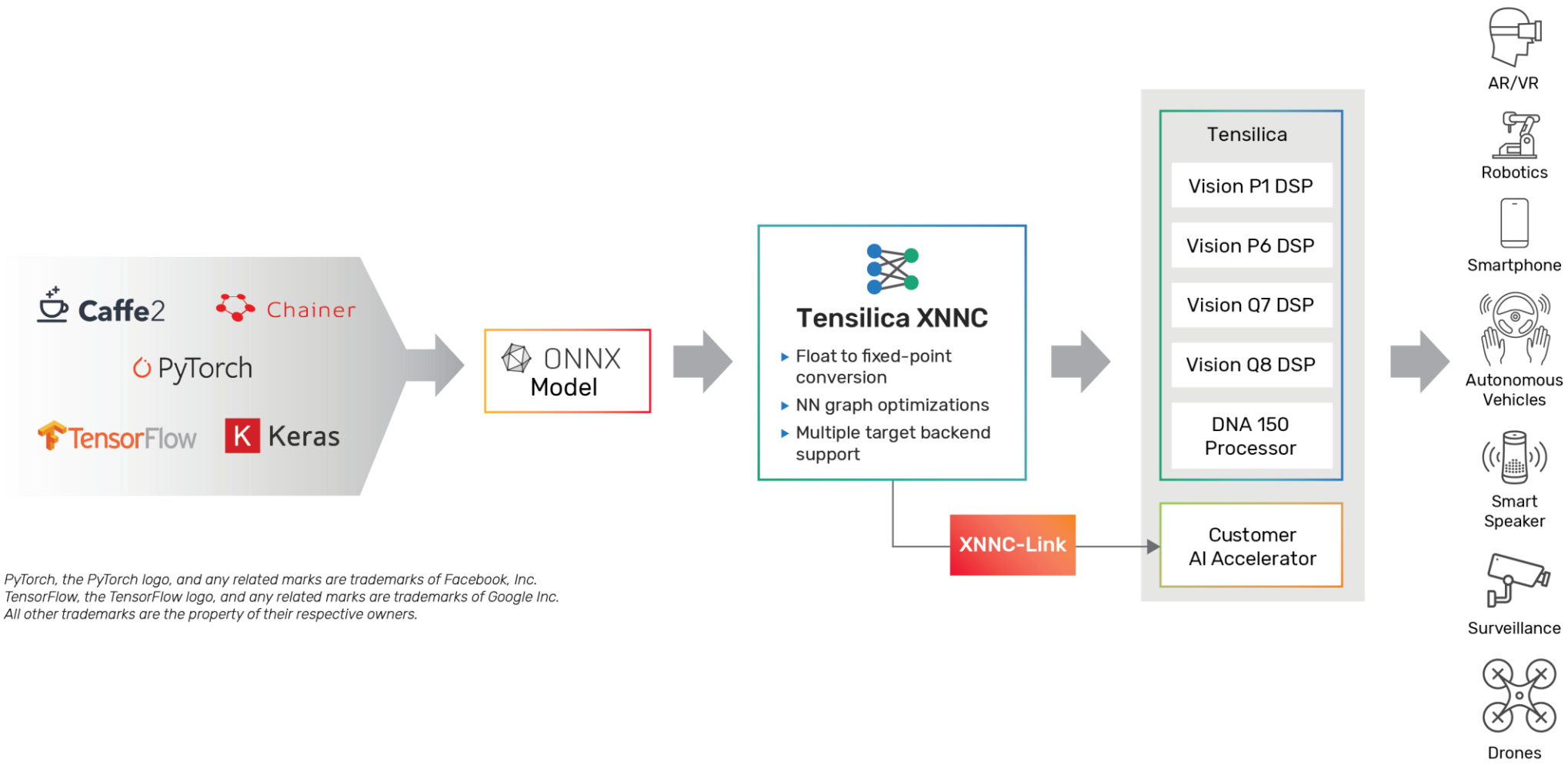
Software Migration from Tensilica Vision P6 and Vision Q7 DSPs	• N-way programming model • Preserves software investment with easy migration • Custom instructions using Tensilica® Instruction Extension (TIE) language
ISO 26262 Readiness	• IP (ASIL-B for Systematic/ASIL-D for Random fault) and tools designed for (ASIL-D) ISO 26262 certification • Customers can generate ISO 26262-compliant optimized DSP and design SoC • Customers can add custom TIE instructions while maintaining ISO 26262 certification

Tensilica DSPs: Comprehensive Vision Software Solutions

Full ecosystem of software frameworks and compilers for all vision programming styles



Cadence AI Software Ecosystem for Tensilica Vision DSPs and DNA Processor



PyTorch, the PyTorch logo, and any related marks are trademarks of Facebook, Inc.
TensorFlow, the TensorFlow logo, and any related marks are trademarks of Google Inc.
All other trademarks are the property of their respective owners.

Tensilica Vision and AI DSP Partner Ecosystem

Summary

Vision DSP Market

- Market needs high-performance vision DSP that supports various data types (fixed, float, complex float) and entry-level AI
- Driven by large number of sensors, higher fps, higher resolution
- Market also needs low-power vision DSP for always-on, smart sensor applications

Tensilica Vision DSPs

- 2 new Cadence® Tensilica® Vision DSPs offer a comprehensive Vision DSP portfolio from high end (3.8TOPS) to low end (400GOPS)
- 7th-generation flagship Tensilica Vision Q8 DSP: 1024-bit SIMD
- Tensilica Vision P1 DSP: 128-bit SIMD, offers 1/3 area and power plus 20% higher frequency compared to Tensilica Vision P6 DSP for always-on applications and smart sensors
- Both Vision DSPs based on same SIMD and VLIW architecture and instruction set used by highly successful Vision P6 and Vision Q7 DSPs
  - Enables fast time to market

Resource Slide

- **Cadence Resources**
  - https://www.cadence.com/en_US/home.html
  - https://www.cadence.com/en_US/home/tools/ip/tensilica-processor-ip.html
  - https://www.cadence.com/en_US/home/tools/ip/tensilica-ip/vision-dsps.html
  - https://ip.cadence.com/ai

cadence®