Cadence Tensilica Product Overview

Tensilica Days Hannover
September 2019

cadence®

Tensilica® Products

Common Tools, Models, Debug, Trace

HiFi



- Audio pre- and post-processing
- Voice trigger
- Noise reduction, audio encode and decode

Vision




- Image and vision pre-/post-processing
- AI at the edge
- SLAM, SGM

Fusion


- Always-on sensor processing
- Multi-purpose
- IoT, consumer and industrial

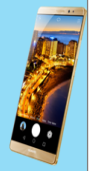
DNA
Standalone AI processor




- inference at the edge
- CNN, RNN …

ConnX






- Baseband processing
- Radar, lidar, communications
- 5G/4G, UE, V2X, IoT and Infrastructure

Controllers/ Custom ISAs

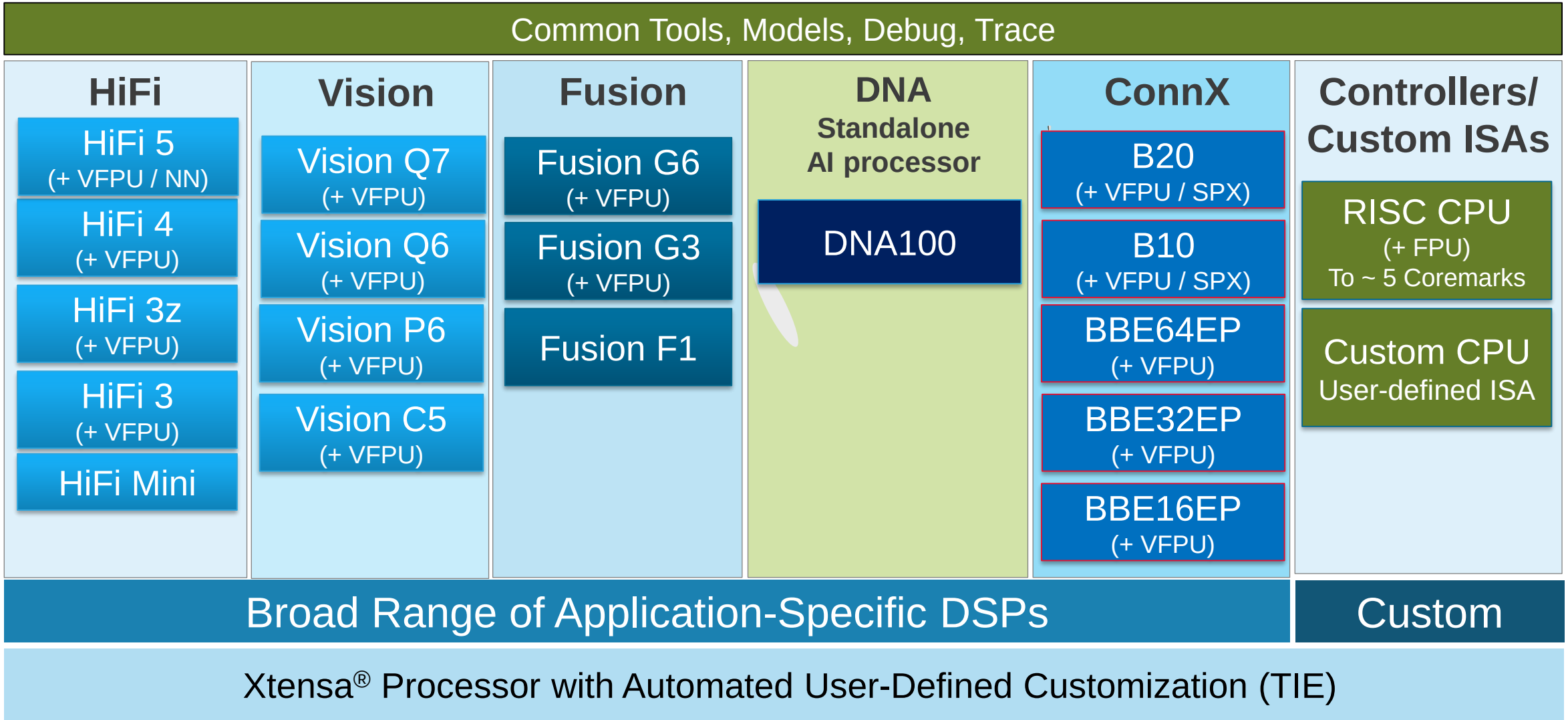

- High-performance DSPs, NPUs, CPUs
- Application-specific data types
- Custom ISA, special functions

Broad Range of Application-Specific DSPs

Custom

Xtensa® Processor with Automated User-Defined Customization (TIE)

Tensilica® Products



Automated Tool, ISS, Model, RTL, and EDA Script Generation

Base Processor
Dozens of templates for many common applications

Pre-Verified Options
Off-the-shelf DSPs, interfaces, peripherals, debug, etc.

Tensilica® IP

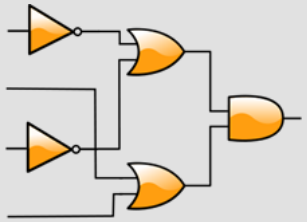
Optional Customization
Create your own instructions, data types, registers, interfaces

Customer IP

Iterate in minutes!

Processor Generator

Complete Hardware Design

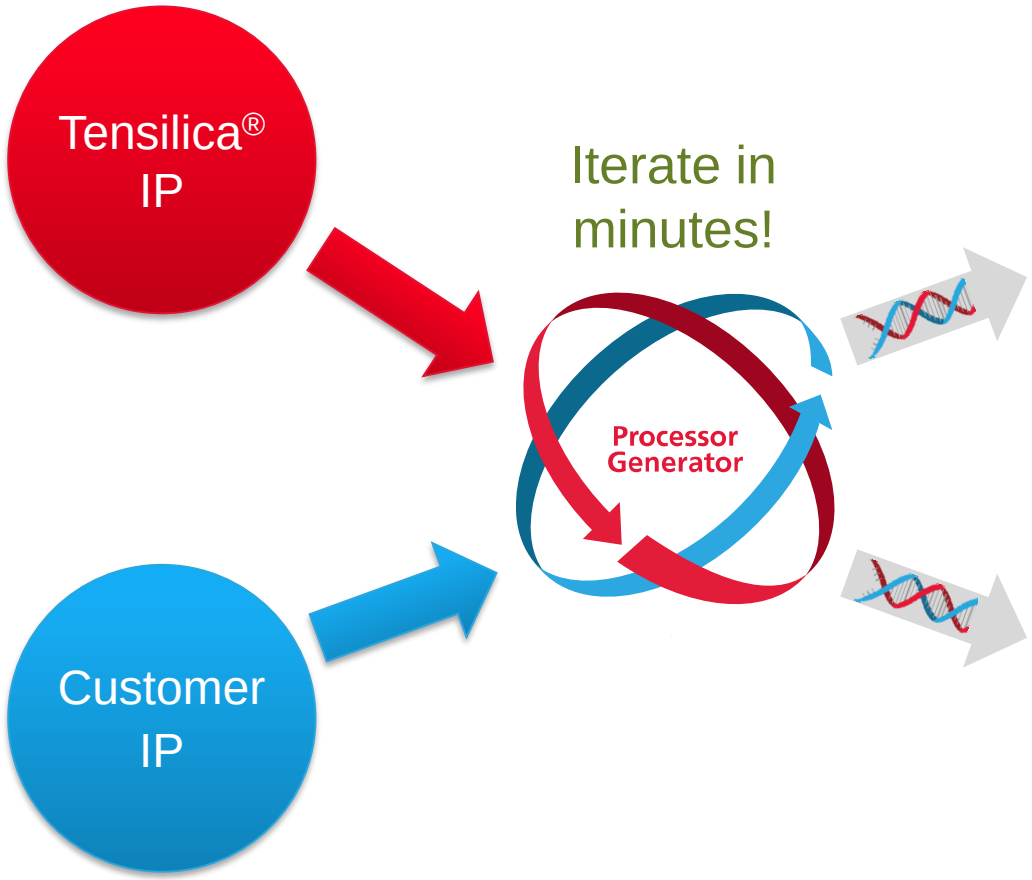


Pre-verified
Synthesizable RTL
EDA scripts
Test suite…

Advanced Software Tools

IDE
C/C++ compiler
Debugger
ISS simulator
SystemC® models
DSP code libraries

Automated Tool, ISS, Model, RTL, and EDA Script Generation



Xtensa Processor Generator Outputs

Hardware
- EDA scripts
- RTL

Synthesis

Block Place & Route

Verification

Chip Integration / Co-verification

→ To Fab / FPGA

System Modeling / Design
- Instruction Set Simulator (ISS)
- Fast Function Simulator (TurboXim)
- XTSC SystemC System Modeling
- XTMP C-based System Modeling
- Pin Level co-simulation

→ System Development

Software Tools
- Xplorer IDE Graphical User Interface to all tools
- GNU Software Toolkit (Assembler, Linker, Debugger, Profiler)
- Xtensa C/C++ (XCC) Compiler
- C Software Libraries
- Operating Systems

→ Software Development

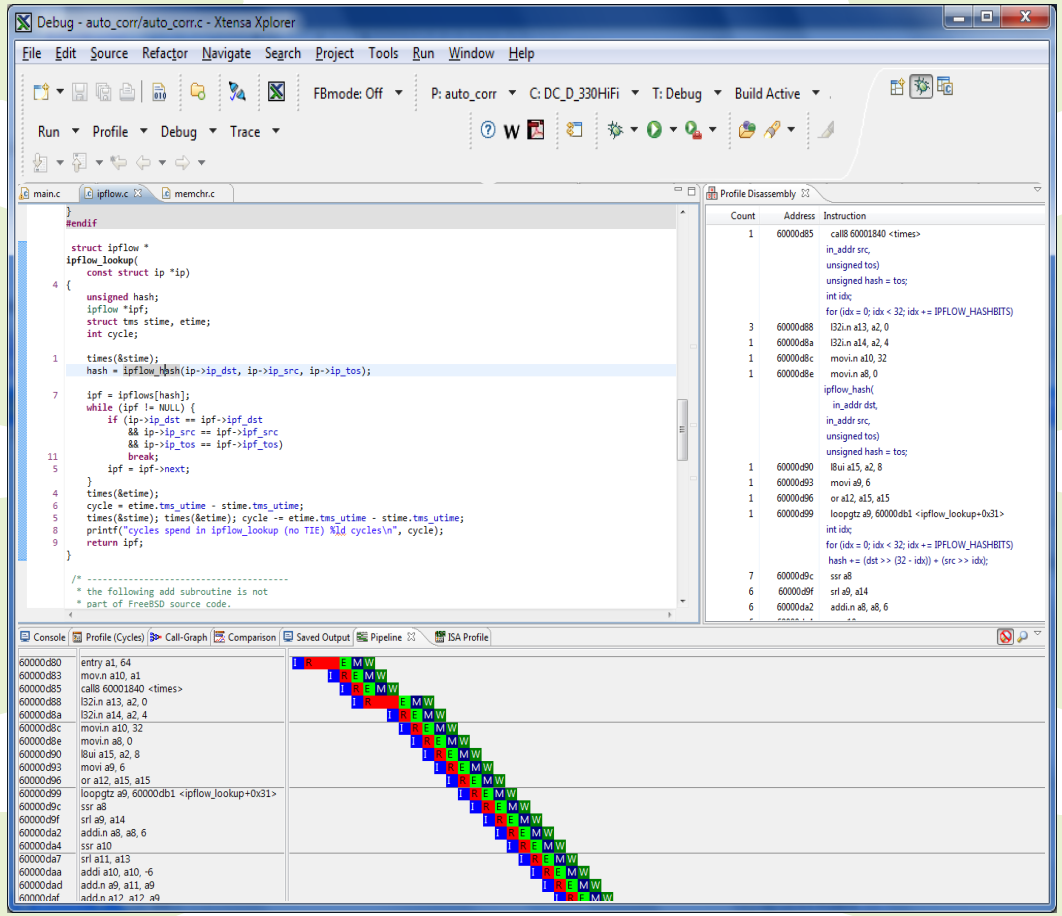
Easy DSP Software Development with Xtensa Xplorer

High-performance optimizing C/C++ Compiler

Cleanly map C/C++ to SIMD & VLIW with no assembly

Extensive software DSP library & examples

Familiar Eclipse-based GUI



Tools “know” your Processor configuration

Launch on ISS, SystemC, RTL, FPGA, or Silicon

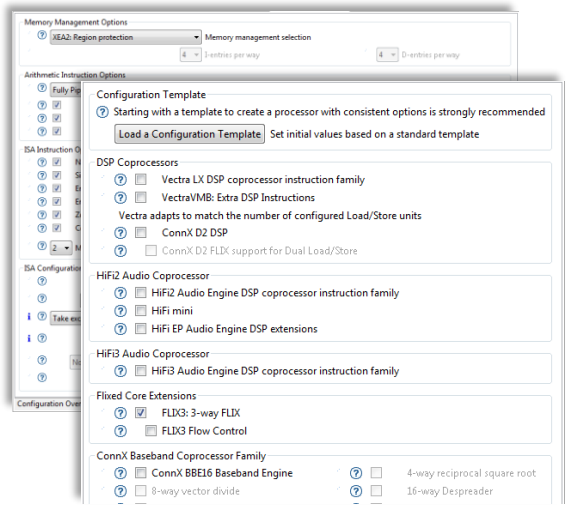
Code coverage, profiling, PC trace, multi-core support

3rd-party JTAG debug and real-time trace

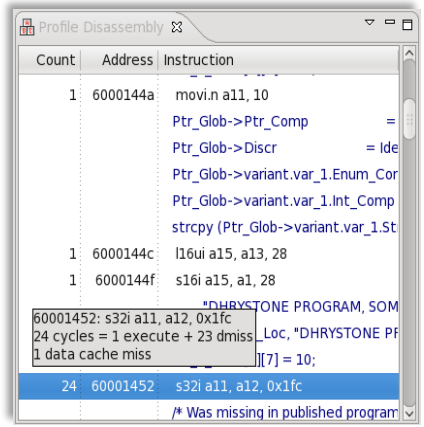
Benchmark and Analyze

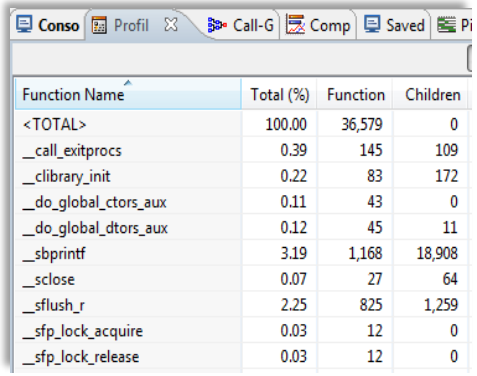
Use Xplorer to Interactively Check Performance As You Build

Check boxes to select pre-designed options

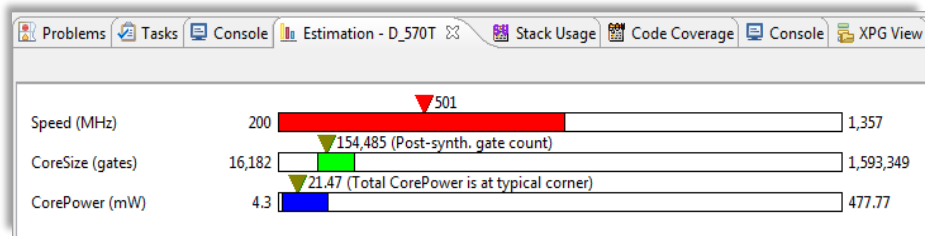


Profile your software to see the critical loops and what operations are used most

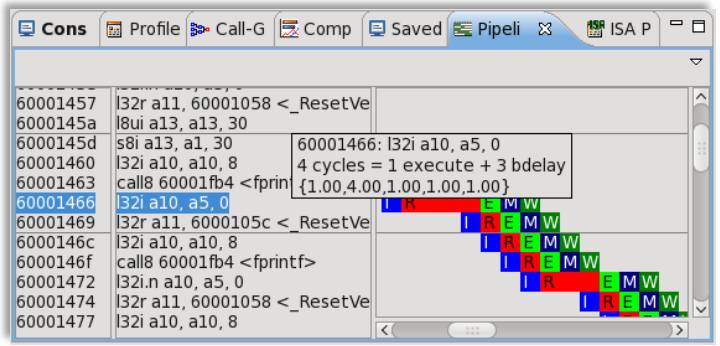




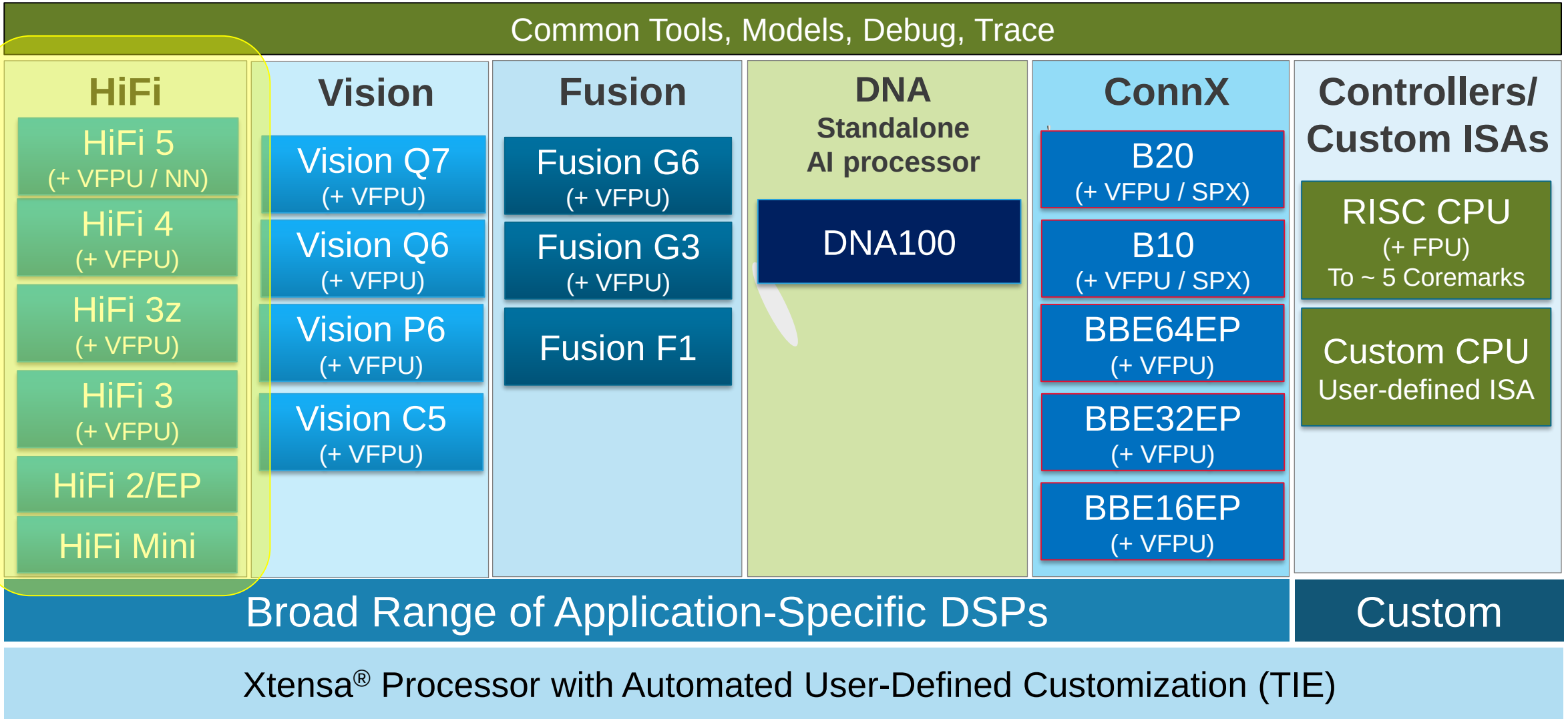
Instantly view PPA impact as you select options



Analyze performance bottlenecks with pipeline view



Tensilica® Products - HiFi



HiFi DSPs for Audio, Voice and Speech Applications

#1 AUDIO DSP IP

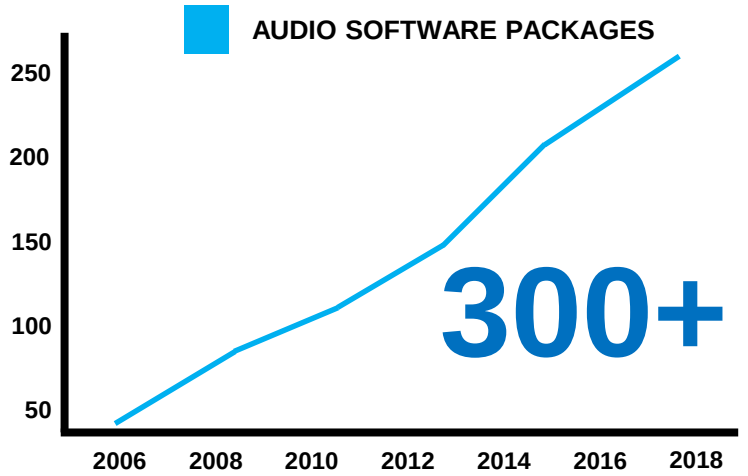
#1

AUDIO DSP
CHOICE

AUDIO ECOSYSTEM

130+ ECOSYSTEM **PARTNERS**

DIGITAL AUDIO SOFTWARE



HiFi LICENSEES WORLDWIDE

105+ HiFi DSP **FAMILY** Licensees

HiFi DSP SHIPMENTS

1B+ HiFi DSPs **SHIPPING** **WORDLWIDE** **ANNUALLY**

HiFi DSPs in Key Markets

Main Functions

- Front end processing
- Audio/Speech codecs
- Keyword/AI Speech
- Post Processing
- 3D Audio

HiFi Advantage

- High Performance
- Low Power/Energy
- Software Ecosystem
- Ease of Programming
- Flexibility

300+ Codecs and Audio/Voice Enhancement Packages

Stereo Audio

AAC
HE-AAC
HE-AAC Plus
MP3
FLAC
WMA 9
WMA 10 Pro
REAL Audio 8, 9, 10
Ogg Vorbis
AMR WB+
SBC Bluetooth
BSAC
DAB
DAB+
DRM (xHE-AAC)
APE

Voice

G.711
G.722
G.723.1
G.726
G.729AB
AMR-NB
AMR-WB
EVS
GSM-HR
GSM-FR
GSM-EFR
AAC-ELD
Opus
mSBC
CVSD
Skype SILK
AEC
LEC

Multi-Channel

AAC, HE-AAC (Plus)
WMA Pro, FLAC
MPEG-H

Dolby
MS10, MS11,MS12
Dolby AC-4
Dolby Atmos
Digital AC-3 and Plus
DDCE, True HD
Pro Logic II/IIx
Dolby Mobile 3+
DS1, DAa2, DAx3
Dolby Digital Live

DTS
HD Master Audio
Express
Transcoder, Neo:6
Broadcast, DMP, M6
Neural Surround
DTS Interactive
Headphone:X

Enhancement

Accusonus
Focus-MD, Focus-DNR
Alango
VCP8, VEP
Arkamys
ImmerseU
Sound Staging
AM3D
Diesel Power, Zirene
Audyssey
Dynamic Volume and EQ
Cadence
Sample Rate Converter
PDM=>PCM Converter
Conexant
AudioSmart
Cywee
Sensor Fusion Hub
Cyberon
Cspotter

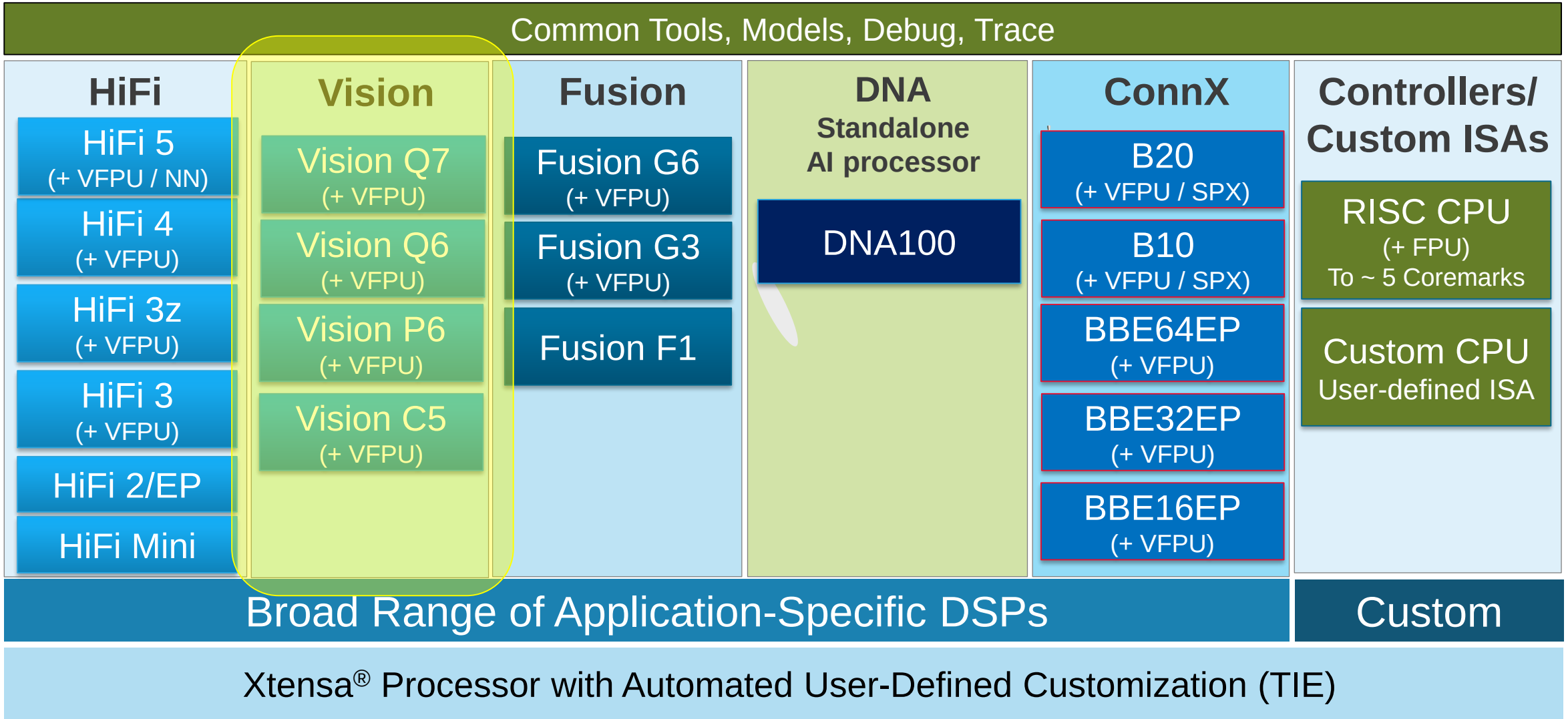
Enhancement

dbx-tv
Total Technology
DTS
TruSurround
TruVolume
TruDialog
TruTools
TruGaming
WOW XT
Fortemedia
iS620, iS700, iS800
Harman
Clari-Fi
Hillcrest Labs
Freespace
Kronoton
HDSX
Malaspina
VoiceBoost
Maxim
Dynamic Speaker Management
MightWorks
Call Solution+SRE

Enhancement

Müller-BBM
m|klang® ANC, ASD
NXP Software
LifeVibes
VoiceExperience
QNX
ANC
QSound
microQ
mQSynth, mQ3D
MQFX
Qvoice
Retune DSP
Beamform, AEC
Rubidium
Speech Processing
Sensory
TrulyHandsFree 3.0
Sonic Emotion
Absolute 3D
SPL
Vitalizer
Waves Audio
MAXXAudio,Voice, Nx

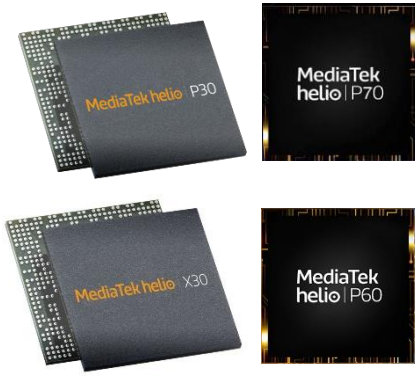
Tensilica® Products - Vision



Vision Recent Customer Success

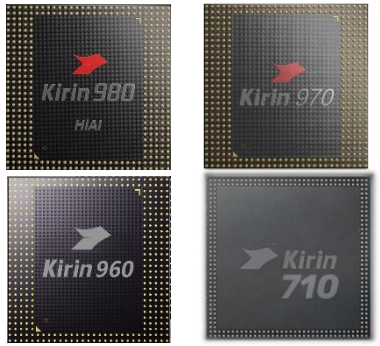
MEDIATEK

Panasonic Blackview
oppo realme uleFone
NOKIA vivo vernee
GIONEE MEIZU DOOGEE



- Vision P5 & P6 DSP
- Integrated in AI Processing Unit (P60)[1]
- 35+ Smartphone designs[2]

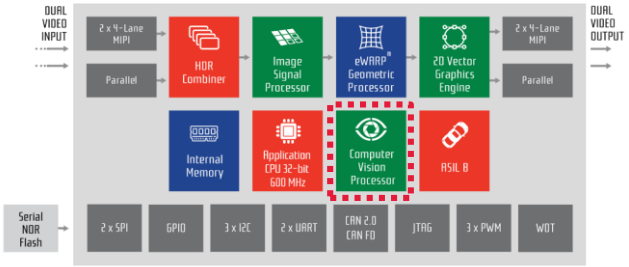
HUAWEI



HUAWEI
honor

- Vision P5 & P6 DSP
- For 3D sensing, HMI, AR/VR apps
- 100+ Smartphone designs[3]

GEO Semiconductor



- GW5400, the World's First Automotive Smart Viewing Camera Processor
- Vision P5
- Zero airflow environment, smart rear-view mirror, backup-camera

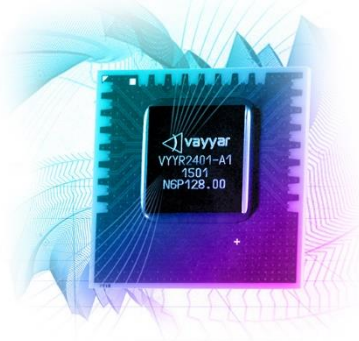
TOSHIBA

- Vision P6 used as Image Recognition Processor; Quad-core configuration
- Up to 1024 GOPs & 3.8x more power efficiency than CPUs,
- Vision P6 enables ISO26262 ASIL D certification

HISILICON

- Hi3559AV100, Hi3519AV100 new-generation intelligent video processor
- Hi3359A (4x P6), Hi3519A (1x P6)
- High-End Surveillance, professional camera, drone camera, extreme sports motion DV

vayyar



- Vision P5 DSP[5]
- Creates high-resolution 3D images in real time based on advanced RF technology
- Targeted for smart home, automotive, smart retail, robotics markets

Source: 1, 2, 3, 4, 5

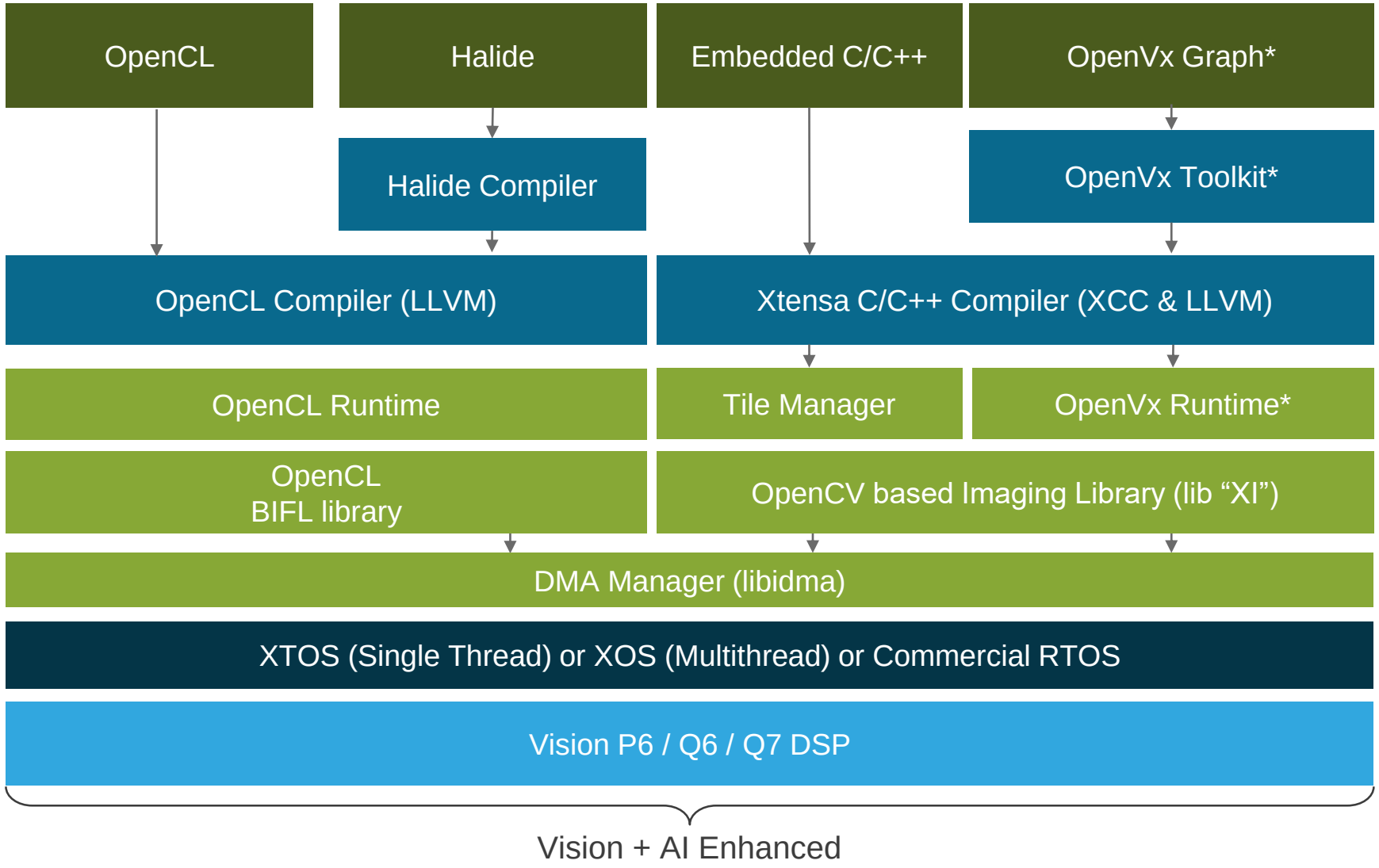
cadence

Vision Product Comparison Chart

		Vision P6	Vision Q6	Vision Q7
Use Case		Vision and Low-end AI (up to 256GMAC/sec)	Vision and Low-end AI (up to 384GMAC/sec) (maximizing MHz)	Vision and Low-end AI (up to 768GMAC/sec) (maximizing MHz) **ISA Improvements for SLAM**
MACs (Higher MAC = Higher Compute)	8x8	256		256/**512** (Optional)
	8x16	128		128
	16x16	64		64/**128** (Optional)
Vector Floating Point Unit	16b Half Precision	32 way SIMD (optional)		**2X 32 way SIMD (Optional)**
	32b Single Precision	16 way SIMD (optional)		**2x 16way SIMD (Optional)**
MAX SIMD Width		64 way 8bit		64 way 8 bit
SuperGather (Scatter Gather)		Yes		Yes

Tensilica: Comprehensive Vision Software Solutions

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

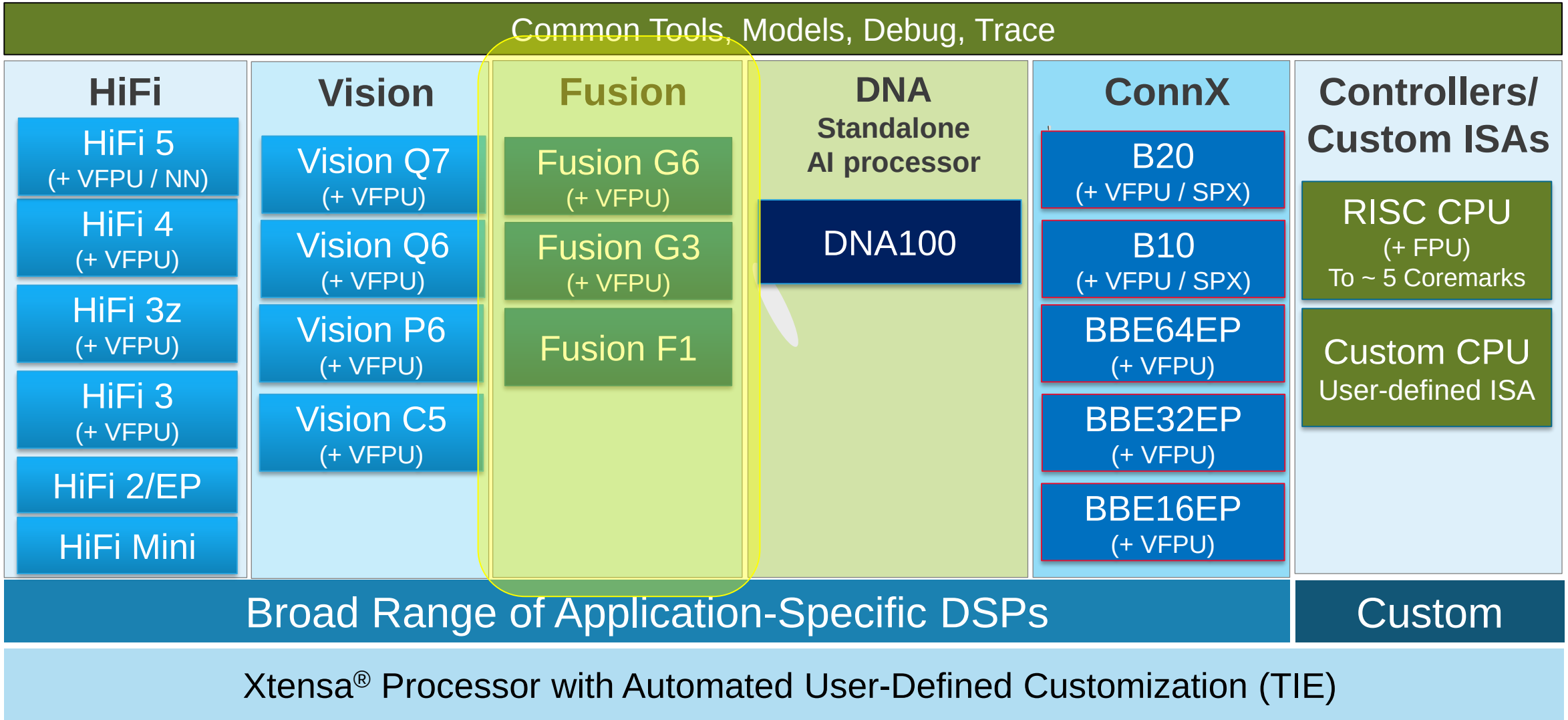


User Code
Cadence Compiler / Tool
Cadence SW library / Runtime
Cadence Tensilica DSP
*Not supported on Vision Q7

OpenVx/OpenCL Host Code

Linux Host CPU (Xtensa or Other)

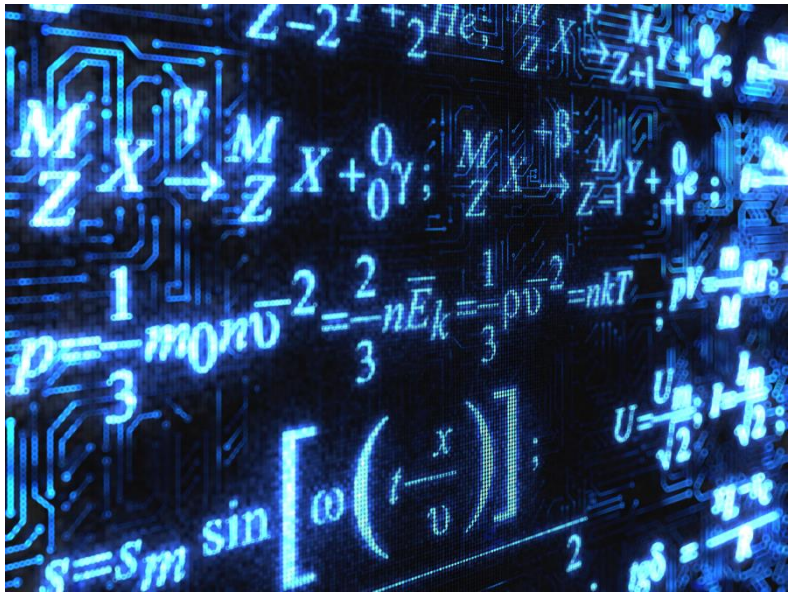
Tensilica® Products – Fusion G



Trends in Multiple Markets Driving the Need for Fusion G Family

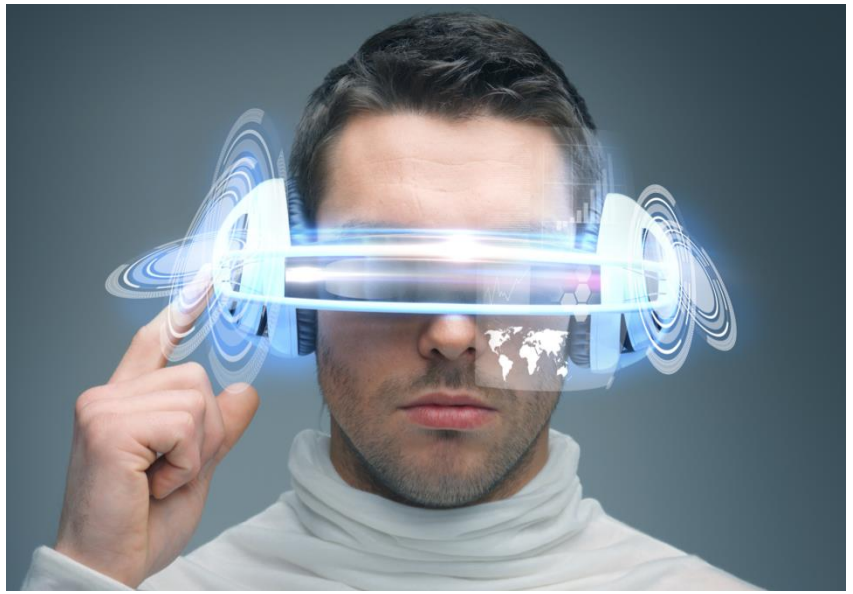
More floating-point use…

- Algorithms and codecs in floating point
- Time-to-market benefits
- Out-of-box performance
- Easy to program and optimize



Multi-purpose…

- Run multiple algorithms on one ISA
- Fixed- and floating-point DSP
- Support for multiple data types
- Efficient real-time control



Target Applications for Fusion G Family

Noise/Echo Cancellation

Surround Sound

Motor Control

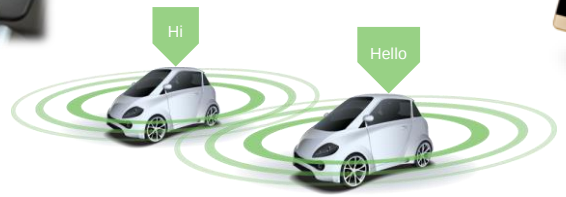
Radar Processing

Baseband Processing

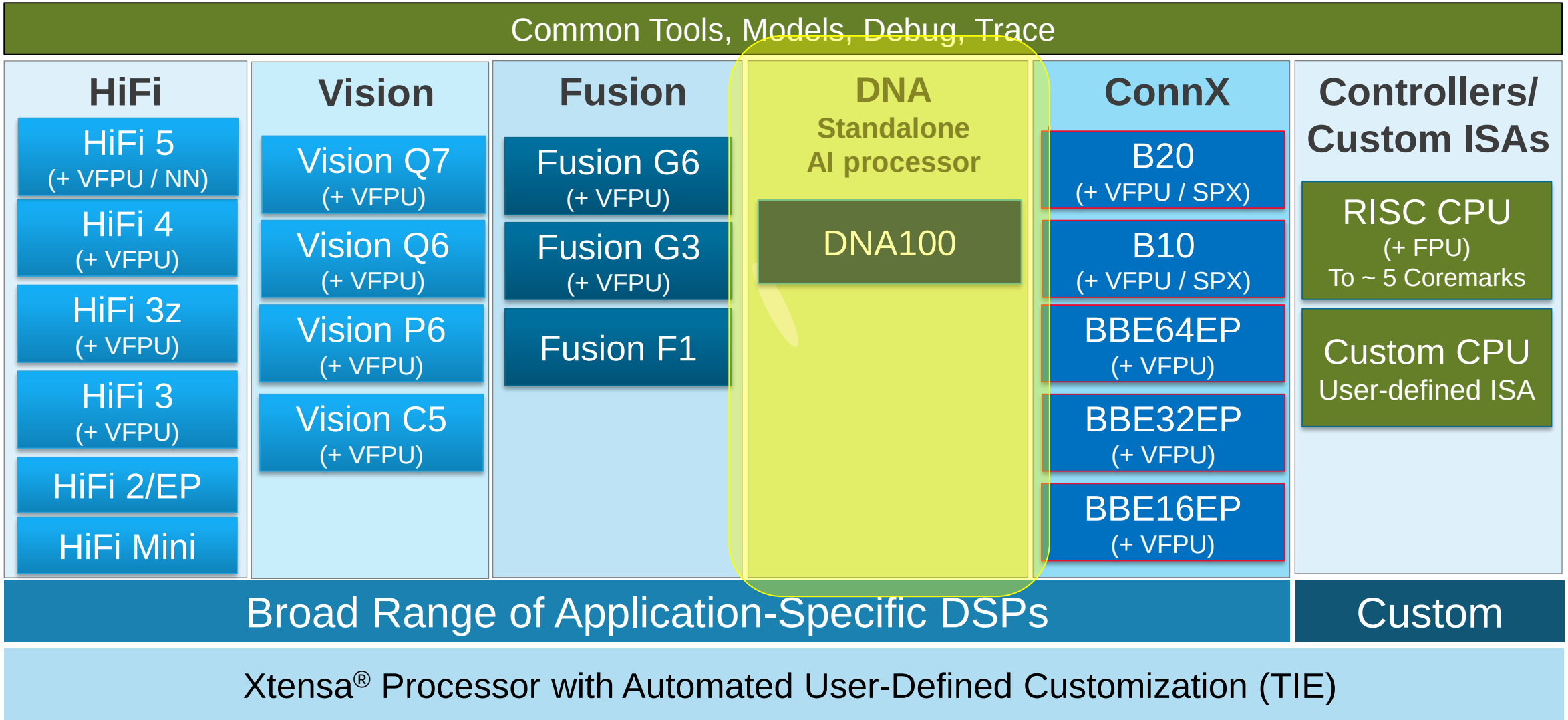
Low-End Image Processing

General Signal Processing

Sensor Fusion



Tensilica® Products – DNA (AI inference)



AI Inference at the Edge Processing Needs

Scalability	• IoT, Mobile, Surveillance, AR/VR, Automotive • <0.5TMAC to 100sTMAC
Increase in Compute Requirement	• Higher Resolution • Move from Traditional to AI
Latency Critical Applications	• AR/VR • Automotive
Higher Power Efficiency	• Battery Life • Bandwidth Constraints
Flexibility	• Continuously adapting/changing market needs

Tensilica® DNA 100 Key Features

Scalability

- Supports 256, 512, 1K, 2K Physical 8b MAC configurations
- Single click configuration select

Throughput

- Sparse compute hardware engine avoids multiply by 0 (activation & weights)
- Significant higher throughput and lower power compared to traditional dense compute engines

Bandwidth

- Compressor/Decompressor hardware block avoids data movement of 0 (activation & weights)
- Efficient memory management

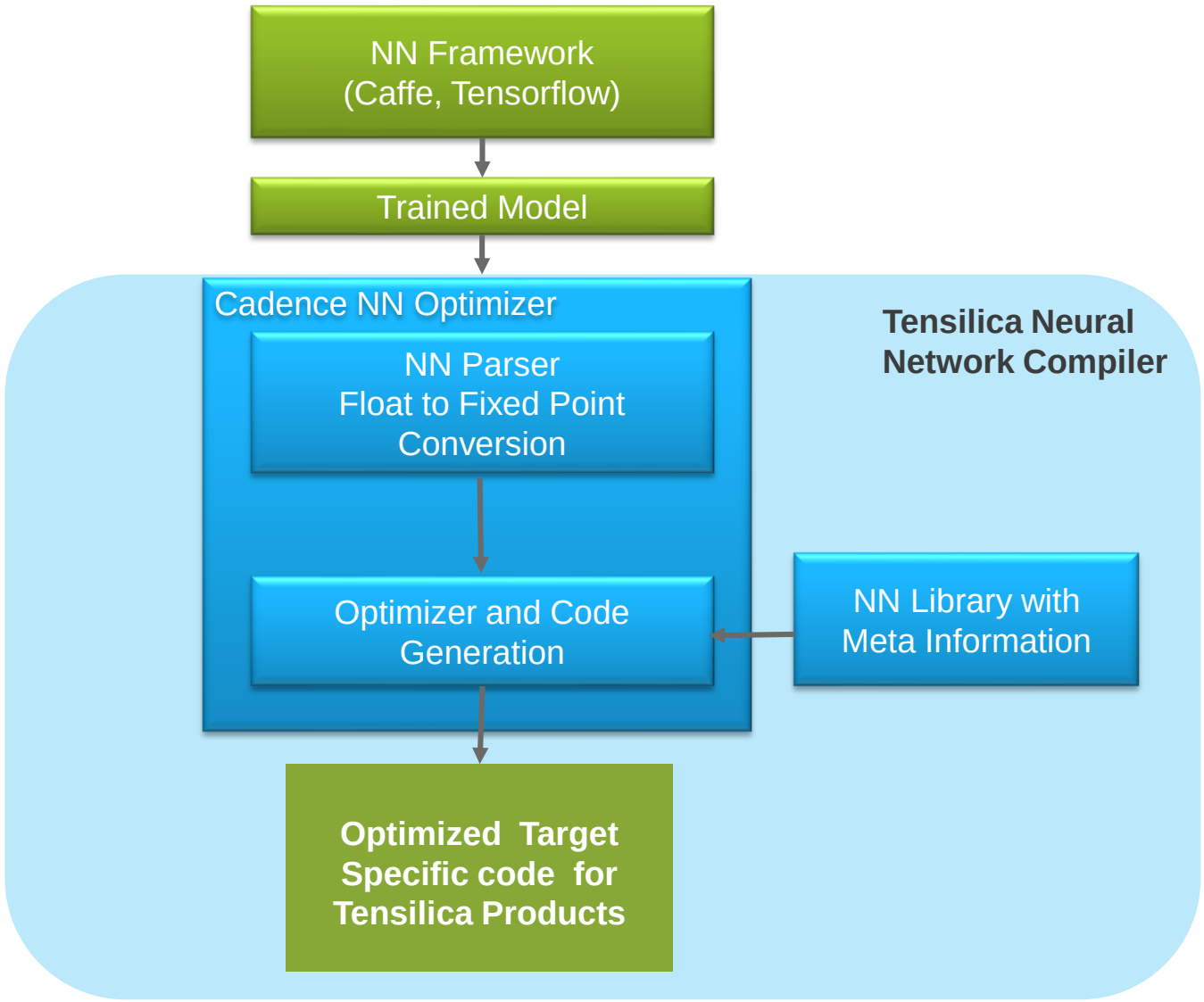
Programmable

- Integrated Vision P6 provides flexibility to support any hardware unsupported layer efficiently
- Integrated Vision P6 provides extensibility through TIE

Standalone

- Efficient convolution & fully connected layer support through high MAC occupancy rate
- Integrated VPU supports commonly used non-convolution layers

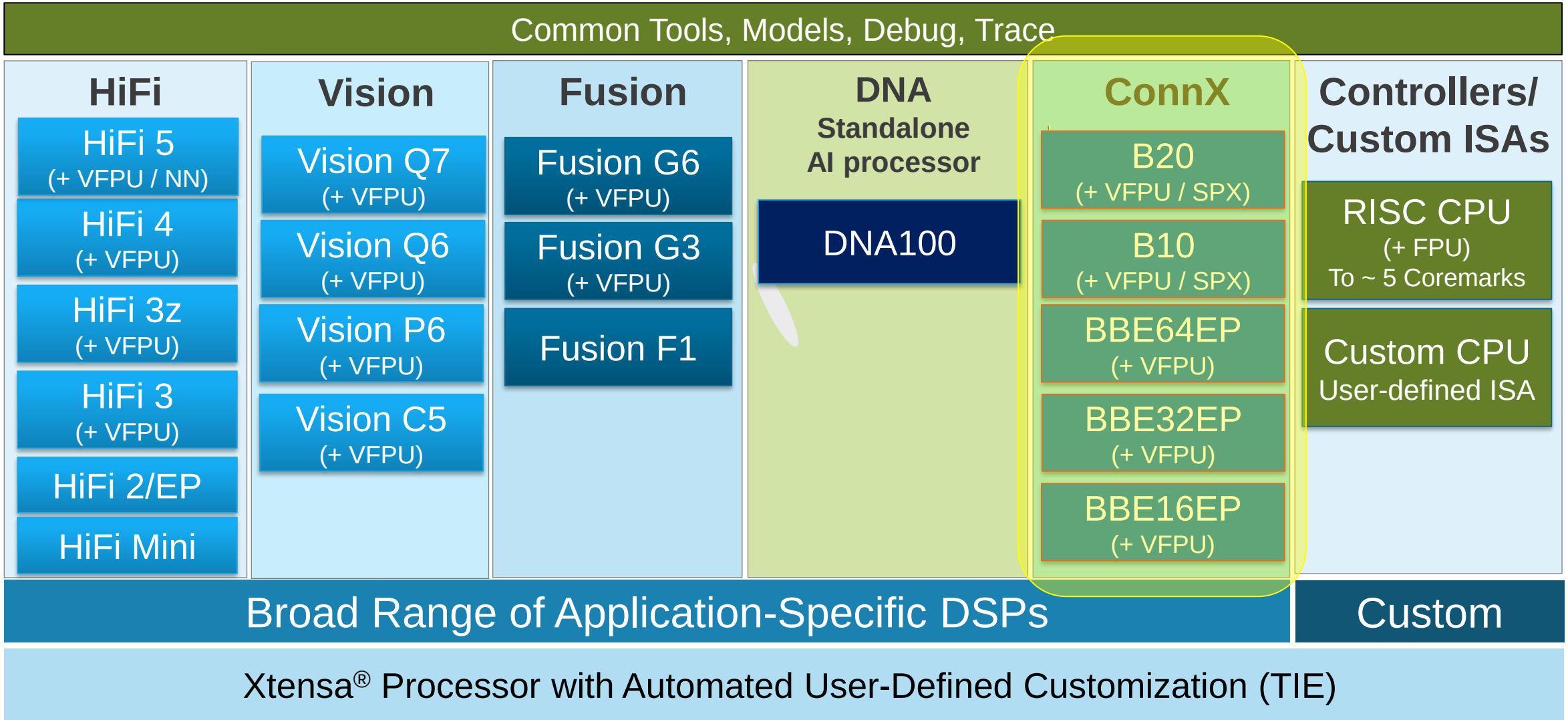
Tensilica Neural Network Compiler



Key Features

- Support for DNA 100, Vision Q6, Vision C5, Vision P6 DSPs
- Support for Windows and Red Hat Linux
- Custom Layer
- Analyzer enables:
  - prioritize accuracy vs performance
  - Min/Max range control to guide quantization
- Optimizer enables:
  - Efficient memory management (ex. avoid unnecessary roundtrips, local memory partitioning and buffering)
  - Tile & DMA management
  - Graph transformation (ex. fusion, elimination)
- NN library showcasing processor specific optimized functions

Tensilica® Products - ConnX



Cadence Tensilica DSPs - ConnX Family

ConnX BBE & ConnX B for complex vector computation

ConnX

- Baseband processing
- Radar, lidar, communications
- 5G/4G, UE, V2X, IoT and Infrastructure

Automated User-Defined Customization

Xtensa® Processor Generator

- Radar / Comms / OFDM domain-optimized DSP with rich feature set
  - Optimal for many classes of algorithm
    - FFT, filters, matrix operations, math functions
  - Native support for complex and real datatypes
    - In fixed point or optional vector floating point
- Scalable solution – 5 family members BBE16/32/64EP & B10/20
  - A family of DSPs with source code compatibility
    - Wide choice of PPA points for multiple applications with same source and tools
  - The same ISA is implemented in SIMD8, SIMD16 or SIMD32
    - 16-bit real and complex values (datapath is 128b, 256b, 512b)
    - 32-bit real and complex values on B10/20 (datapath is 256b, 512b)
- Datapath-width-tuned SIMD & VLIW architecture + local memory access
  - L1 access is 128b, 256b, 512b
  - Two L1 memory accesses per cycle allowed – up to 1024b / cycle
  - Optional integrated DMA engine has its own AXI4 master port
    - With access to L1 memories and can run “in the background”

Highlights of ConnX DSP enhanced performance

BBE16/32/64EP and B10/20

Optimal DSP engines

- Highly efficient MAC/ALU usage with broad DSP ISA

Broad range of DSPs buildable from a single platform

- Code compatible for ease of portability for performance scalability
- Scale to meet your needs: 16, 32, 64 and 128 MACs…
- Easily select useful options with click boxes
- Software tools and verification is automatic

Optimized for low power/energy

- Clock and data gating implemented, plus logic and memory bus gating are also options
- Loop buffer can reduce instruction memory accesses
- ISA acceleration reduces cycle counts

Limited area and power increase as frequency is increased

Libraries for ConnX DSPs

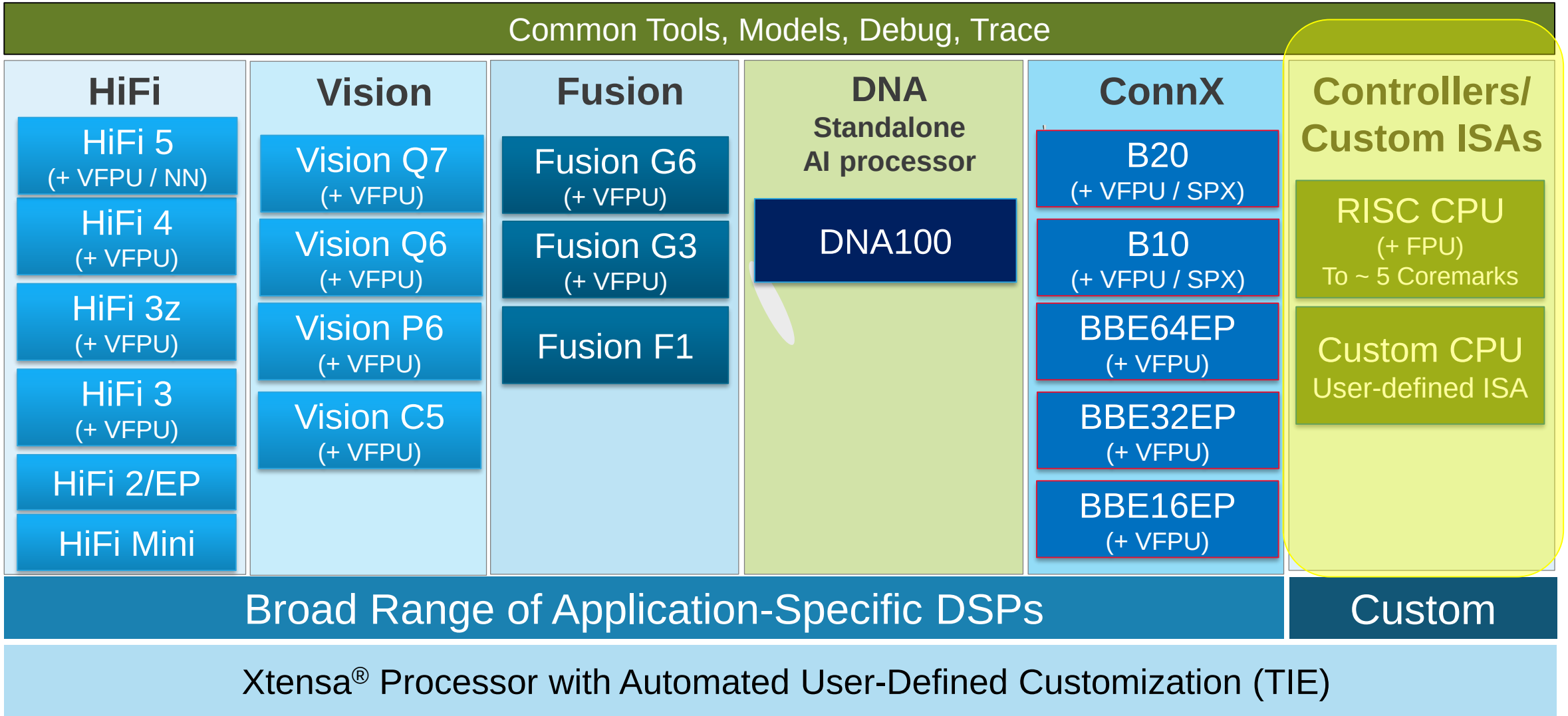
Rich and optimized DSP libraries for basic and advanced DSP functions

- Vector fixed/floating-point, real/complex support
- Library source code available

Note: To cover the complete library routine list, all optional configurations have to be selected

Category	Functions
Math	Scalar and vector for: arc cosine/sine/tangent, cosine/sine/tangent/cotangent, hyperbolic sine/cosine/tangent, sigmoid, logarithm, anti-logarithm, exponential, square root, reciprocal, reciprocal square root
Complex	Magnitude, phase, conjugate, exponent, combined cosine and sine, normalization, division, polar to cartesian and cartesian to polar conversion
FIR Filters	Convolution, auto/cross correlation, interpolation, decimation, polynomial fitting/interpolation
IIR Filters	Biquad, lattice block IIR
FFT	Complex, real, FFT/IFFT, DFT
Vector	Product, sum, magnitude, reciprocal, division, transcendental, peak, mean
Matrix	Multiply, transpose/Hermitian, Cholesky/QR decomposition and recursion, determinant, inverse, linear equation solution
Communication Systems	CRC, de-spreading, modulation, slicer, convolutional encoding, bit manipulation, PRBS Generation, space time coding

Tensilica® Products – Controllers and Customisation



Base RISC Capabilities

- Xtensa defines Powerful RISC ISA
- Available in LX or NX pipelines
  - Trade clock speed vs extreme flexibility
- Can be used for tiny control tasks
- Can be used for high performance general processing
- Good debug, trace features

Configurability – features and performance

- High degree of configurability
  - Base RISC ISA – tune to likely requirements (MUL, DIV, interrupt, debug, trace ..)
  - Supports optional single and double precision Floating point
  - Caches or TCM
- Can have user-defined extensions through TIE language
  - Could be new calculation operations
  - Register files, addressing modes
  - Interfaces to HW engines
- Up to 5 Coremarks
  - High performance multi-issue controllers

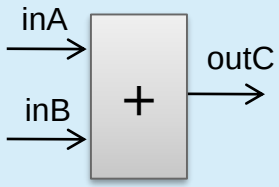
Differentiate with a custom ISA using “TIE”, automatic tools support

Tensilica Instructions Extensions

- Simple Verilog-like language, where you can define…
  - Input/output queues and ports
  - Custom register files
  - Fast lookup tables and local memories
  - Simple single-cycle or multi-cycle instructions
  - SIMD for vectorization
  - FLIX (VLIW) for grouping parallel operations into one instruction

Example: I/O Queue adder

Create an **addq** instruction with three 256 bit queues and an “add” operation:




```
queue inA 256 in
queue inB 256 in
queue outC 256 out
operation addq {} {in inA, in inB, out outC} {
    assign outC = inA + inB;
}
```

High throughput without using system bus:
64 bytes in and 32 bytes out **per operation**
Virtually unlimited bandwidth

Example: Popcount acceleration

Create a **pop_count** instruction that counts the one’s in a 32-bit register by adding the bits together.
This simple Verilog-like code is all it takes to create both the pre-verified adder RTL (175 gates) and the instruction:

```
operation pop_count {out AR co, in AR ci}{}{
   wire [3:0] a0 = ci[0]  + ci[1]  + ci[2]  + ci[3]  + ci[4]  + ci[5]  + ci[6]  + ci[7];
   wire [3:0] a1 = ci[8]  + ci[9]  + ci[10] + ci[11] + ci[12] + ci[13] + ci[14] + ci[15];
   wire [3:0] a2 = ci[16] + ci[17] + ci[18] + ci[19] + ci[20] + ci[21] + ci[22] + ci[23];
   wire [3:0] a3 = ci[24] + ci[25] + ci[26] + ci[27] + ci[28] + ci[29] + ci[30] + ci[31];
   wire [5:0] sum = a0 + a1 + a2 + a3;
   assign co = {26'b0, sum};
}
```

Best hand-coded ASM using standard instructions **takes >10 cycles.**
This simple instruction takes it down to just **one cycle** for **10x speedup**
Easily add custom acceleration into the processor ISA with full tools support

Vision Q7

Tensilica Vision Q7 DSP: 6th-Generation Vision and AI DSP

Follow on to the Vision Q6 DSP for Vision and AI:
Achieves **1.7GHz**/1.55** GHz** peak frequency

Up to 1.7X higher TOPS in the same area
Delivers up to 2.06** TOPS/ 1.82* TOPS

Up to 2X performance for vision/AI applications, including floating-point performance

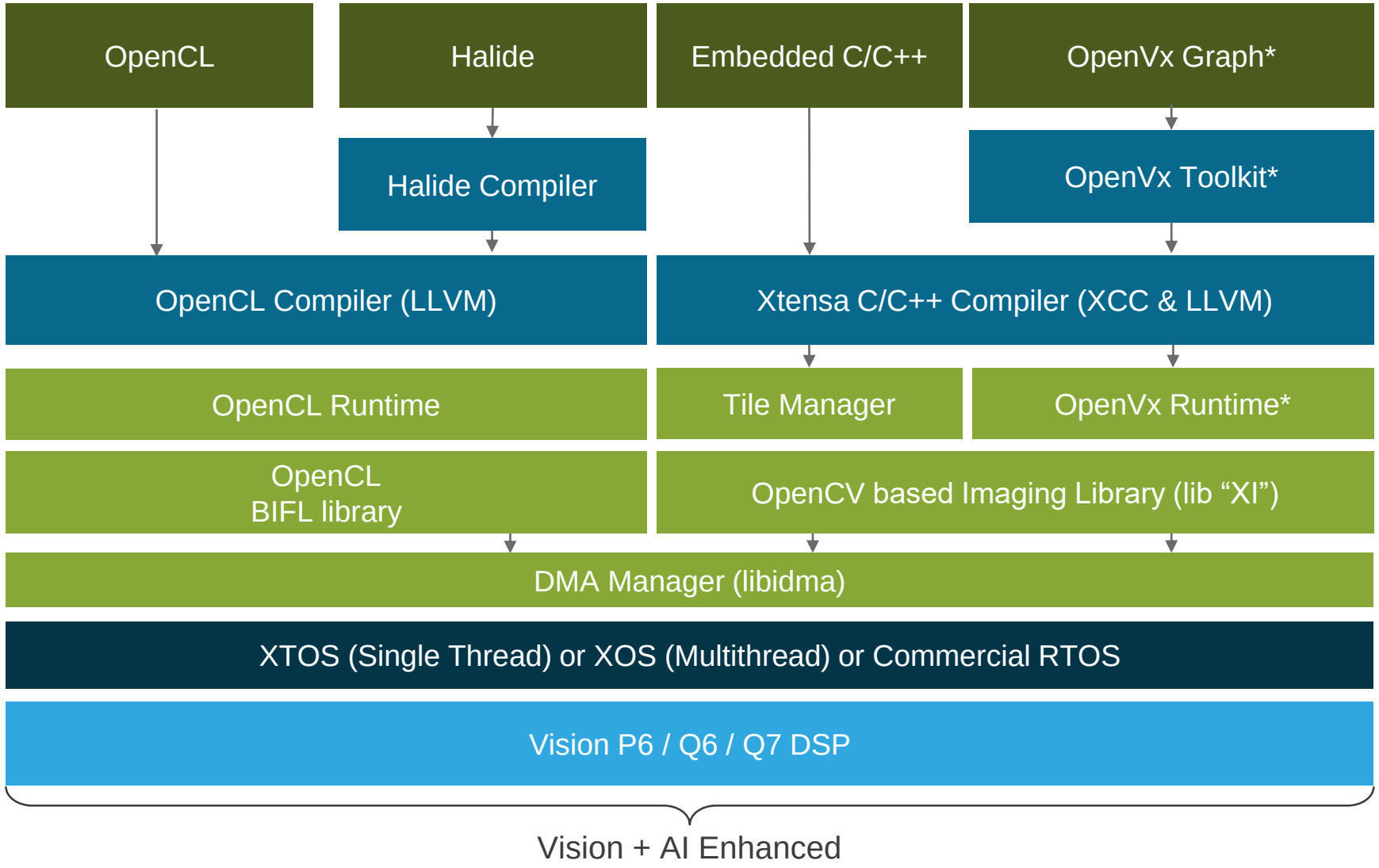
Up to 2X GMAC/mm^2 and GFLOPS/mm^2

- * 16nm process with OD
- ** 1.7GHz in 7nm with OD
- ® Vision Q7 DSP performance is relative to the Vision Q6 DSP

Software Ecosystem

Tensilica: Comprehensive Vision Software Solutions

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



User Code
Cadence Compiler / Tool
Cadence SW library / Runtime
Cadence Tensilica DSP
*Not supported on Vision Q7

OpenVx/OpenCL Host Code

Linux Host CPU (Xtensa or Other)

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
- Uses more efficient data structures for images and tiles instead of OpenCV structures that waste local memory
- In most cases, 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

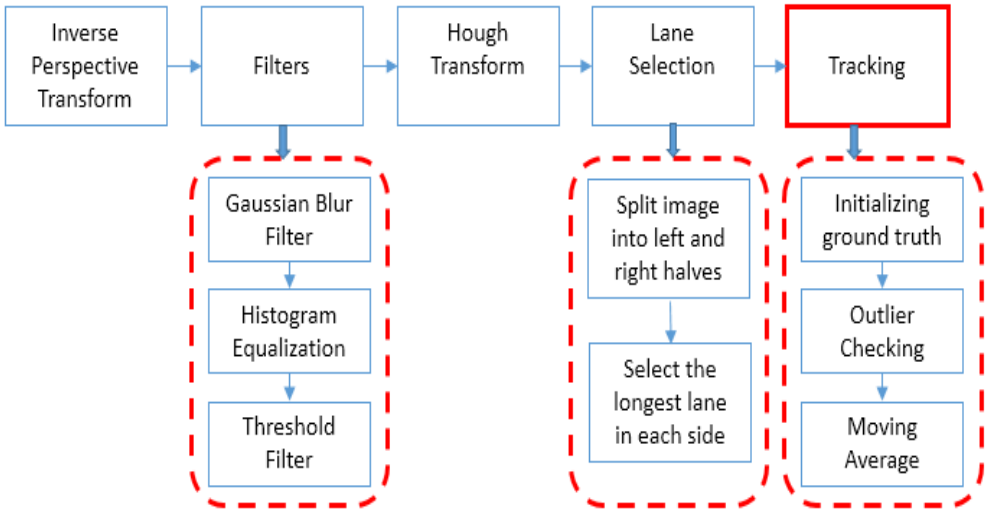
XI-Library Example Functions

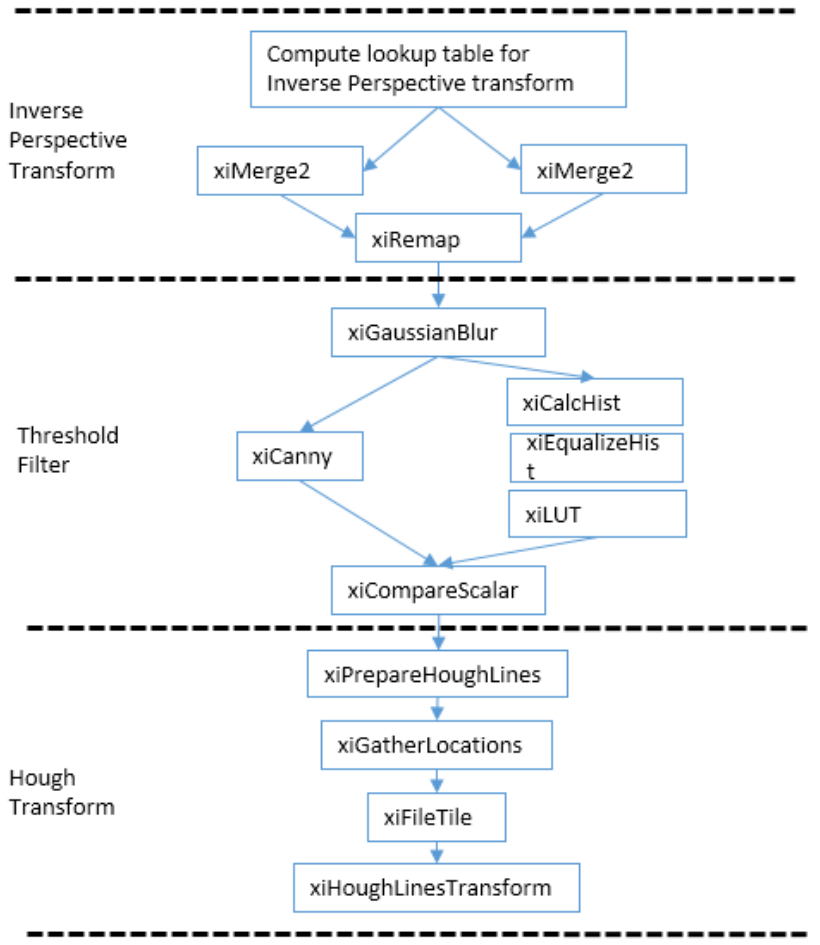
- Image Proc Modules: Convolution functions, Geometric Functions,…
- Features Modules: Feature Detection, Feature Descriptor
- Motion Analysis & background analysis
- Core modules: Binary elements, bitwise operation, vector operations…

Detail Description of XI Library Functions (few examples)

XI Library Module	Sub Module	Example functions (subset)	Description
Image Proc module	Convolution functions	xiBilaterfilter, xiBoxfilter, xiCannyedge	“Image processing” module contains functions related to geometrical image transformations, scaling, 2D convolution, edge detection
	Geometric transformation	xiResizeBilinear, xiTranspose, xiWarpAffine	
	Structural analysis	xiConnectedComponents	
	Miscellaneous transforms	xiCvtColor, xiIntegral, xiIntegralSqr	
Features module	Feature detection	xiCornerHarris, xiFAST	Feature detection and description functions are contained in the “features2D” module
	Feature descriptor	xiBRIEF	
Video/Motion Analysis module	Motion analysis	xiAccumulateWeighted, xiMeanShift, xiOpticalFlow_TrackPoint	“Video” module contains functions related to video processing.
	Background subtraction	xiBS_MOG	
Core module	Utility functions	xiErrStr , xiCopytile, xiFillTile	“Core” module includes low-level support and helper functions
	Unary element wise operations	xiAbs, xiBitwiseNot, xiClip, …	
	Binary element wise operations with scalar value	xiAbsdiffScalar, xiAddScalar, xiBitwiseAndScalar, …	
	Binary element wise operations	xiAbsdiff, xiAdd, xiBitwiseOr, …	
	Reduction operations	xiCountNonZero, xiGatherLocationsEQ, xiMaxLoc, …	
	Vector operations	xiMagnitude, xiPhase, xiPolarToCart, …	
	Miscellaneous operations	xiExtractChannel, xiLUT, xiMerge2/xiMerge3, ...	

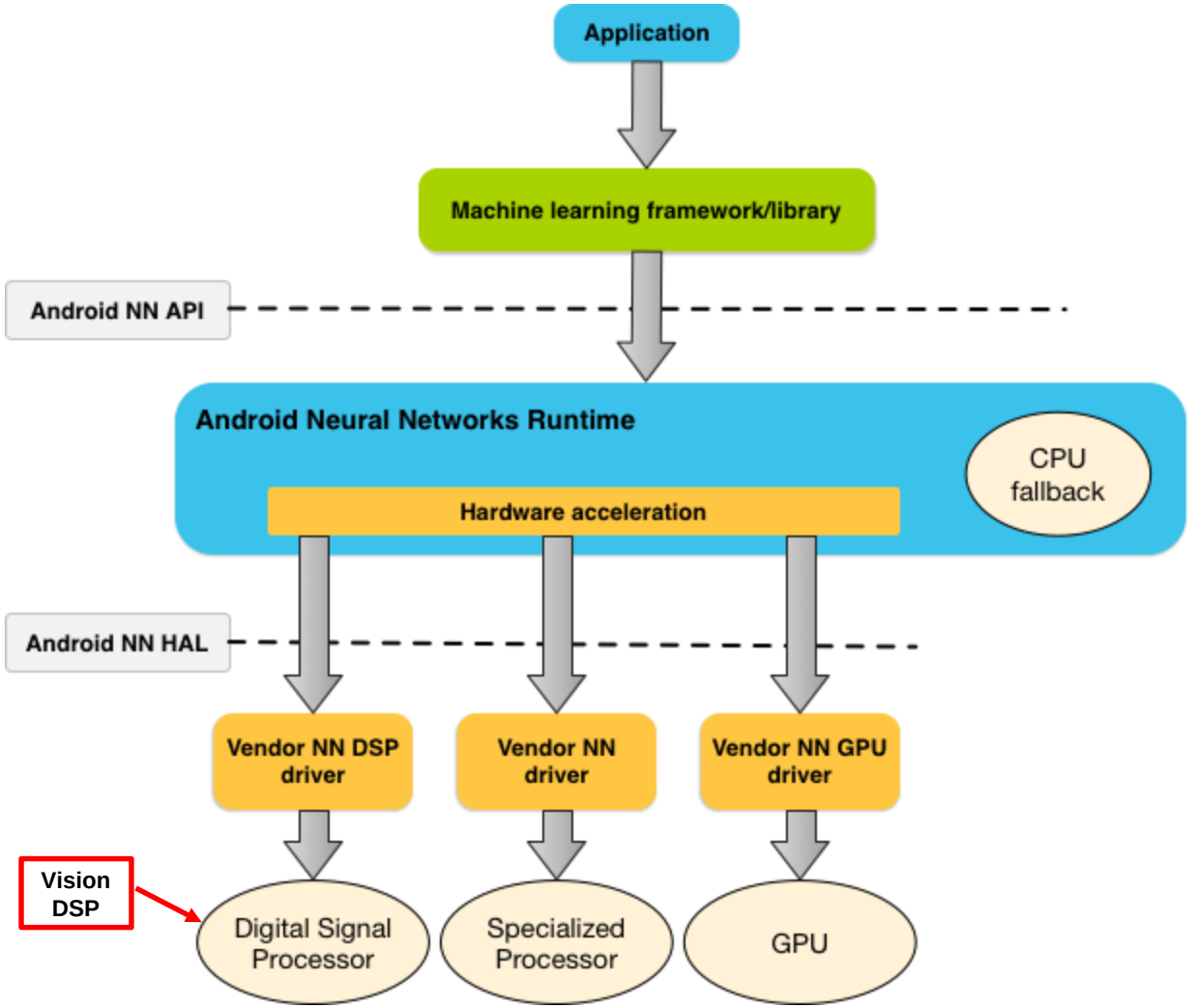
Example of Lane Departure Warning: Work done at Cadence Use of XI-Lib





Android Neural Network (ANN) API

Support for Android 8.1 Oreo ANN release



Source: https://developer.android.com/ndk/guides/neuralnetworks/index.html

Host

- Graph Executor decides to
  - execute graph/sub-graph/layer
  - Selects full layer kernel
  - Tile & DMA management*
  - Data/weight rearrangement
- XRP sends commands to DSP

DSP

- NN HAL Command Interpreter
  - Interprets commands from XRP to full layer kernel
- Full layer kernel
  - Allocates local memory
  - Loops over tiles, DMAs data, invokes library
- XI CNN Library
  - Optimized library using Google's Quantization scheme

* Not needed with Vision DSP

Software Status

Vision and AI Software Status: Vision Q7

	Current Release	Next Release/Comment
SW Pkg	Example SW package is available today with XPG Released with RI.1	
Various SW Kernels	All SW kernel listed in P6vs Q7 comparison are available today (17,18,19). They are all optimized for Vision Q7	
XI-Lib	First release : Released 8.0.1 released with RI.1 Available with XPG	Optimized for Vision Q7: Oct 2019 Also adding Matrix and FP functions in Nov 2019
XNNC	XNNC 1.5 for baseline 256 MAC available now	512MAC: Estimates for Resnet50 performance available **Working on other networks** **XNNC release Jan 2020 support for various networks with 512 MAC**
ANN	Baseline 256 MAC for Android Q: to be release by end of August 2019 Planning to align with Android Q release After we release Vision P6 for Android Q by end of August 2019	512MAC Plan: Plan in August 2019
OpenCL	Available now Released with 2019.1	
Halide	End of Aug 2019	


cadence®