



# RISC-V CON

ONLINE WEBINAR

## Andes Infuses into Artificial Intelligence

**High-Efficiency and High-Flexibility  
Processor IPs + NN SDK for AI**

**John Min**

**Director of Field Application Engineering**

**Andes Technology**

**July 9, 2020**



# Agenda

- The Diversity of AI Use-Cases
- Andes RISC-V Processors for AI
- Andes NN SDK for AI
- Summary

# Andes at A Glance

## Who We Are



Pure-play CPU  
IP Company



RISC-V Founding  
Premier Member



Taiwan Stock  
Exchange Listed



Major Open-Source  
Contributor/Maintainer



Running Task Groups  
Vice Chair of TSC  
Director of the Board  
RISC-V Ambassador



## Quick Facts

**15**  
years old  
company

**200+**  
Licensees  
Worldwide

**80%**  
R&D  
employees

**5B+**  
accumulated Andes-  
embedded SoC shipped

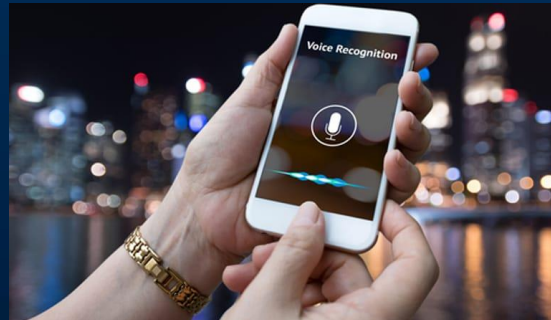
**17K+**  
AndeSight IDE  
installation





## Vision

- Image classification
- Object detection
- Image segmentation
- Spoof detection
- Face unlock
- Eye tracking
- Avatar
- SLAM
- ...



## Voice and Speech

- Audio front-end processing
- Keyword spotting
- Voice command
- Speech to text
- Natural language processing
- Text to speech
- ...

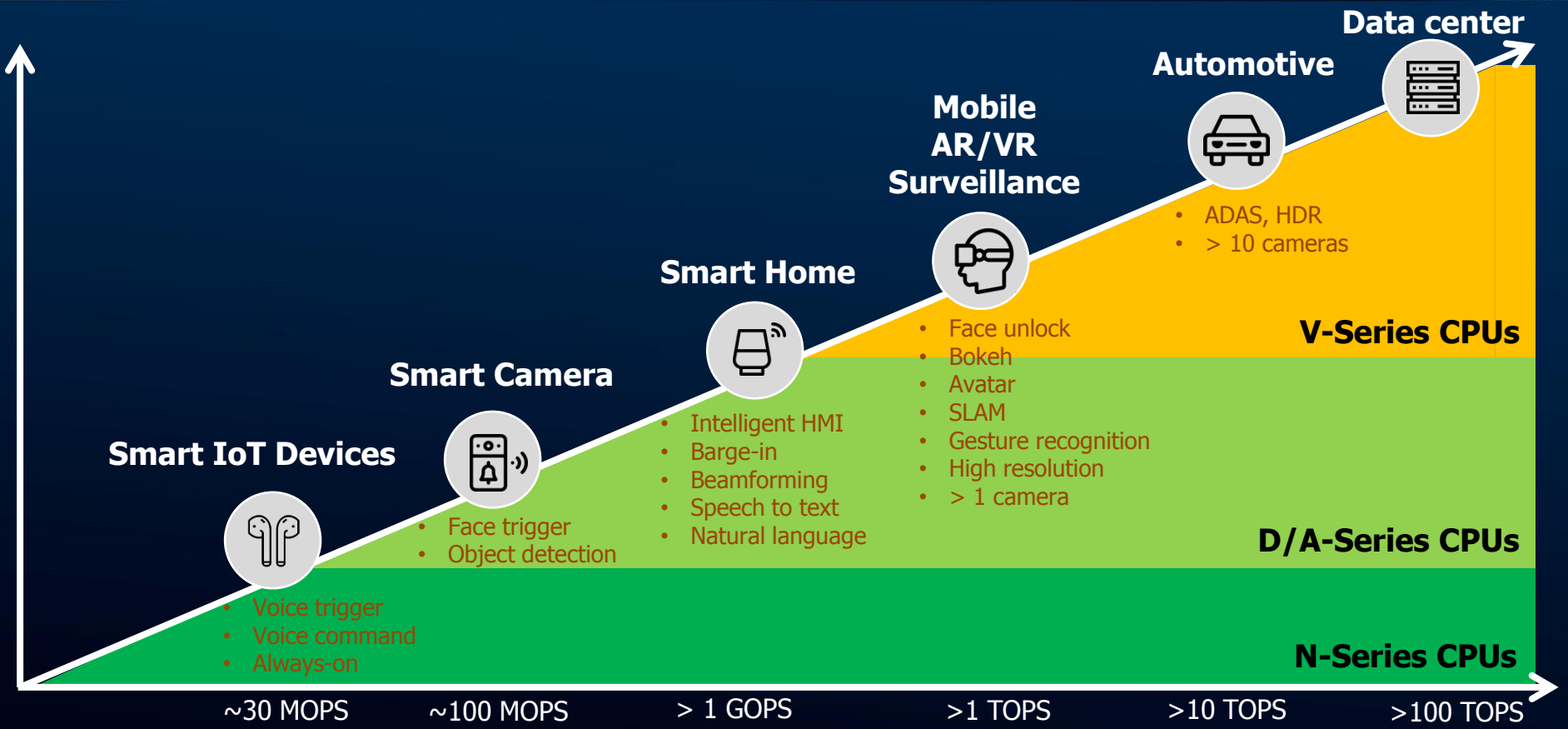


## Any signal

- Sensor fusion with force, pressure, accelerometer, gyro, ampere meter, vibration, radar/lidar, sonar, temperature, ...
- Pattern recognition
- Predictive maintenance
- Healthcare
- ...



# Andes Processors to Fit Your AI







# Andes RISC-V Processors Family

## N-Series Baseline

RISC-V baseline 32/64-bit SMP

+ Andes V5 instructions  
(RV-EIMACFD-XV5)

FPU, cache, local memory, ECC

2-stage to 8-stage pipeline

Frequency up to 1.2GHz  
@28nm worse case

## D/A-Series DSP/SIMD

RISC-V baseline 32/64-bit SMP

+ Andes V5 instructions  
+ DSP/SIMD instructions (RVP)

MMU (A-Series)

SIMD width: 32, 64

Data types: INT8, INT16, INT32

## V-Series Vector

RISC-V baseline 64-bit

+ Andes V5 instructions  
+ Vector instructions (RVV)

VLEN/SIMD width: 128, 256, 512

LMUL(Length Multiplier): 1, 2, 4, 8

Data types: INT4/8/16/32/64,  
BF16, FP16/32/64

- ✓ Leading PPA and high efficiency CPU
- ✓ Control logic and simple data computation

- ✓ Efficient SIMD for data computation
- ✓ Compact MCU AI and basic edge AI applications

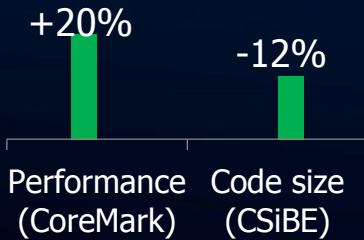
- ✓ High performance, efficiency and configurability
- ✓ Enable data intensive computing from edge to cloud

# RVP and RVV for Data Computation

## Andes RISC-V Baseline

- Clean state
- Compact
- Modular
- Andes V5 ISA extension

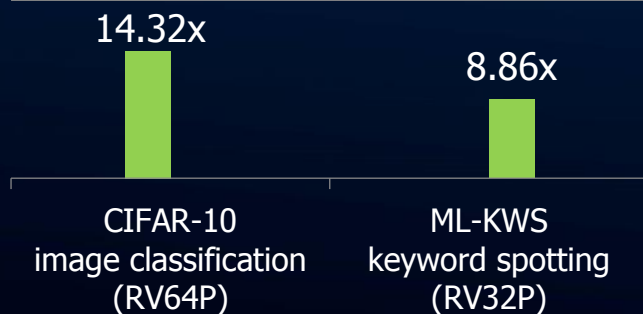
Relative benchmark  
with Andes V5 instructions



## RISC-V DSP/SIMD P-ext

- Andes contributed market-proven DSP/SIMD to RVP
- Use RV32 and RV64 XLEN-bit GPRs
- SIMD with 8b, 16b, 32b
- Complex DSP operating on 16/32/64-bit
- Saturation and rounding
- Min, max, shift, byte swap, bit reverse, pack, unpack, ...

Speedup with RVP



Taking RISC-V® Mainstream

## RISC-V Vector V-ext

- Follow RVV latest standard
- >300 vector instructions
- Scalable vector registers
- 2x/4x data expansion arithmetic
- Load/store, integer, fixed-point/floating-point operations

Speedup with RVV



Note, Based on N25F, Andes/mainline GCC v7.4



# Typical Andes CPU Usages for AI from Edge to Cloud

## Best-fitting control logic

- RISC-V Compact and modular design
- Remove the components which not needed (e.g. FPU, multiplier)

## MCU edge AI

- Single MCU with small data computation (e.g. voice/face trigger)
- Always-on, low power, and cost-sensitive devices (e.g. smart doorbell, ear pod)

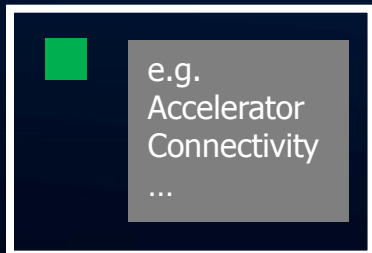
## Performance edge AI

- Application SoC for large data process of CV/ML (e.g. AR/VR, surveillance)

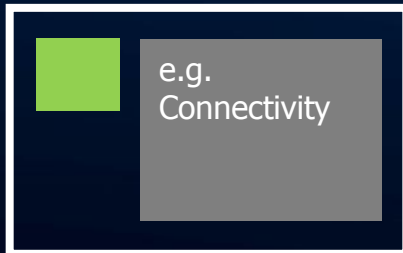
## Cloud AI

- Heterogeneous and cluster computing for AI data center

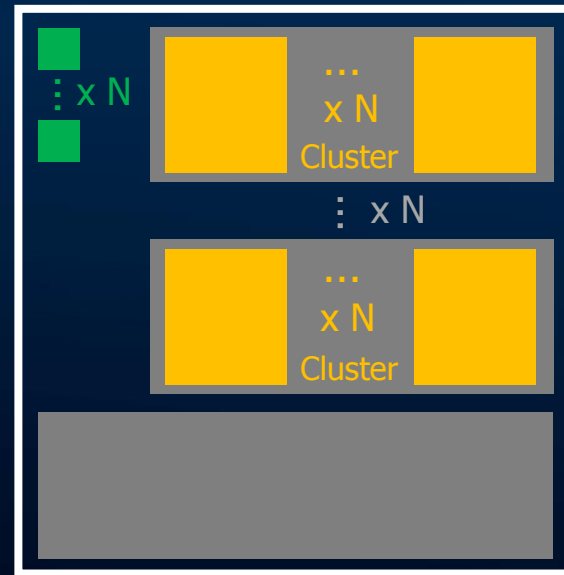
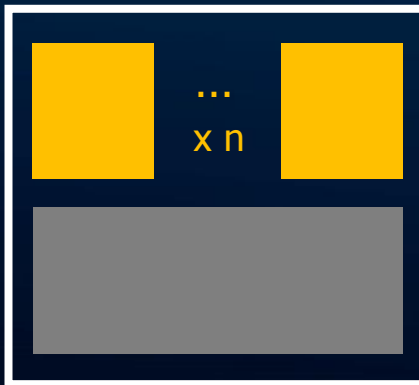
### Baseline



### Baseline + RVP



### Baseline + RVV



Control logic

Data computation





# Efficiency Boost with Andes Custom Extension™

## Compute kernel functions

- Extend instructions for kernel functions (e.g. CONV, GEMM)
- Typical case: implement few dedicated kernel functions which consumes heavy computing power
- Could fit in low power and cost-sensitive devices

## Control ports

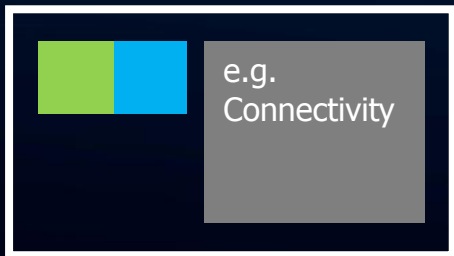
- Extend instructions to control ports (e.g. send command, ack, wait-for-result)
- Typical case: a very compact CPU as a powerful accelerator controller which can send 90-bit commands in one cycle

## Streaming ports

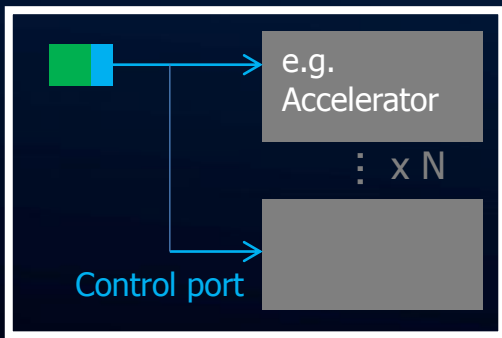
+ control ports + compute kernel functions

- Extend instructions for high volume information transferring between vector processors and external compute units
- Typical case: increase data bandwidth and shorten data latency when using vector to offload hard-wired AI compute unit (e.g. sigmoid)

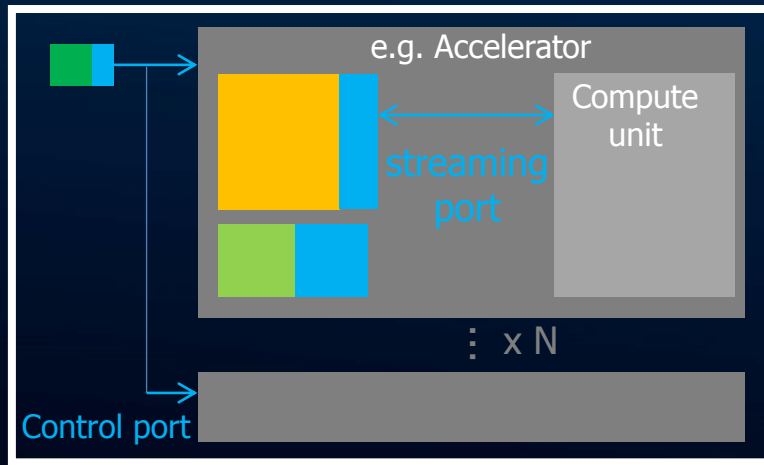
Baseline + RVP + ACE



Baseline + ACE



Baseline + RVV





# Voice-Based Human Machine Interface Use Case

## Pre-Processing

- Echo cancellation
- Noise reduction
- Beamforming
- Auto gain control
- ...

Simple data computation

## Feature extraction

- FFT
- Mel-Frequency Cepstral Coefficients
- Filter bank
- ...

Simple neural network model

## Voice trigger

- Keyword spotting (always-on)

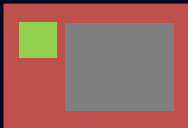
↓ wakeup

## Speech/Text transform

- Automatic Speech Recognition
- Speech synthesis

Intensive data computation  
Complex neural network model

Baseline + RVP



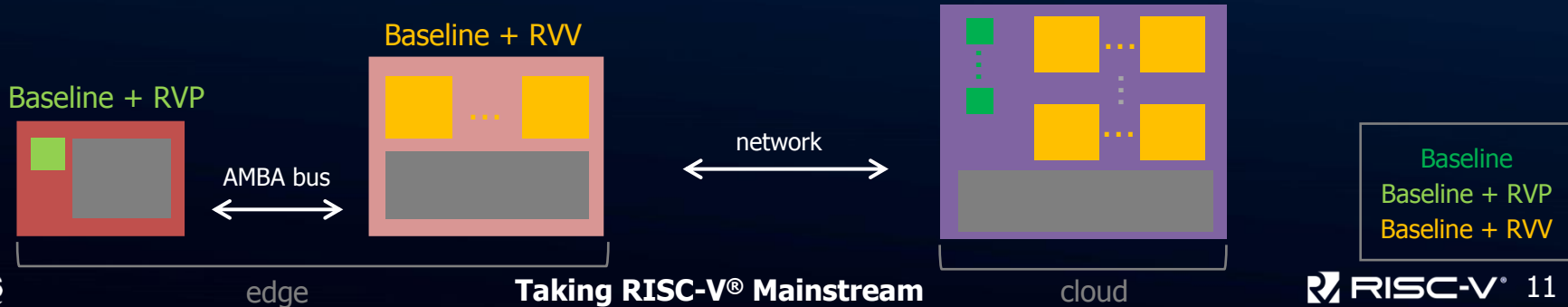
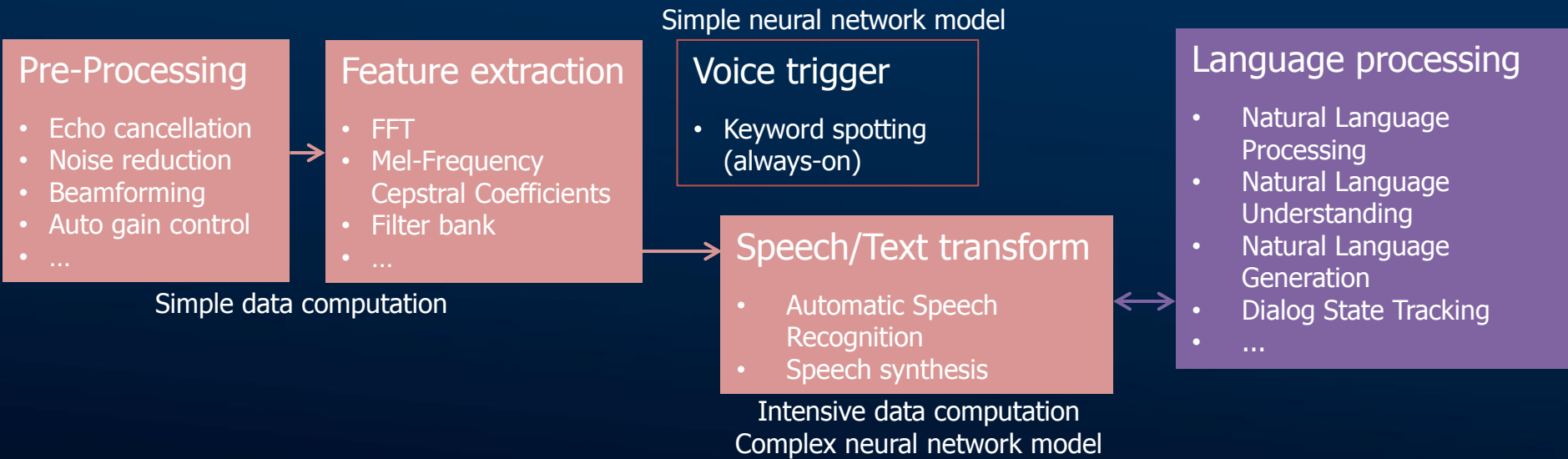
Baseline

Baseline + RVP

Baseline + RVV

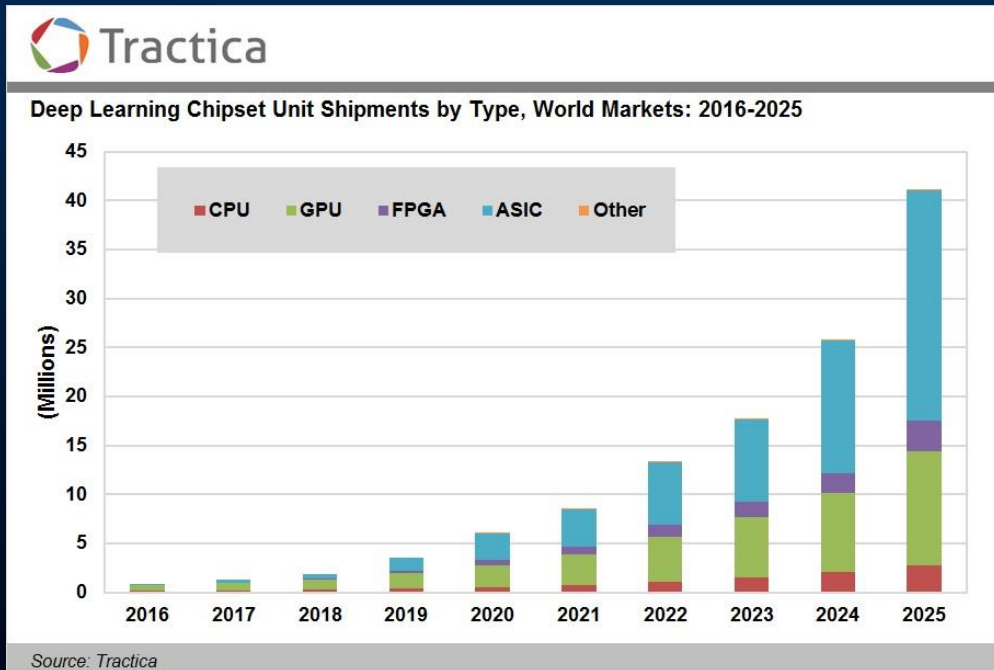


# Voice-Based Human Machine Interface Use Case





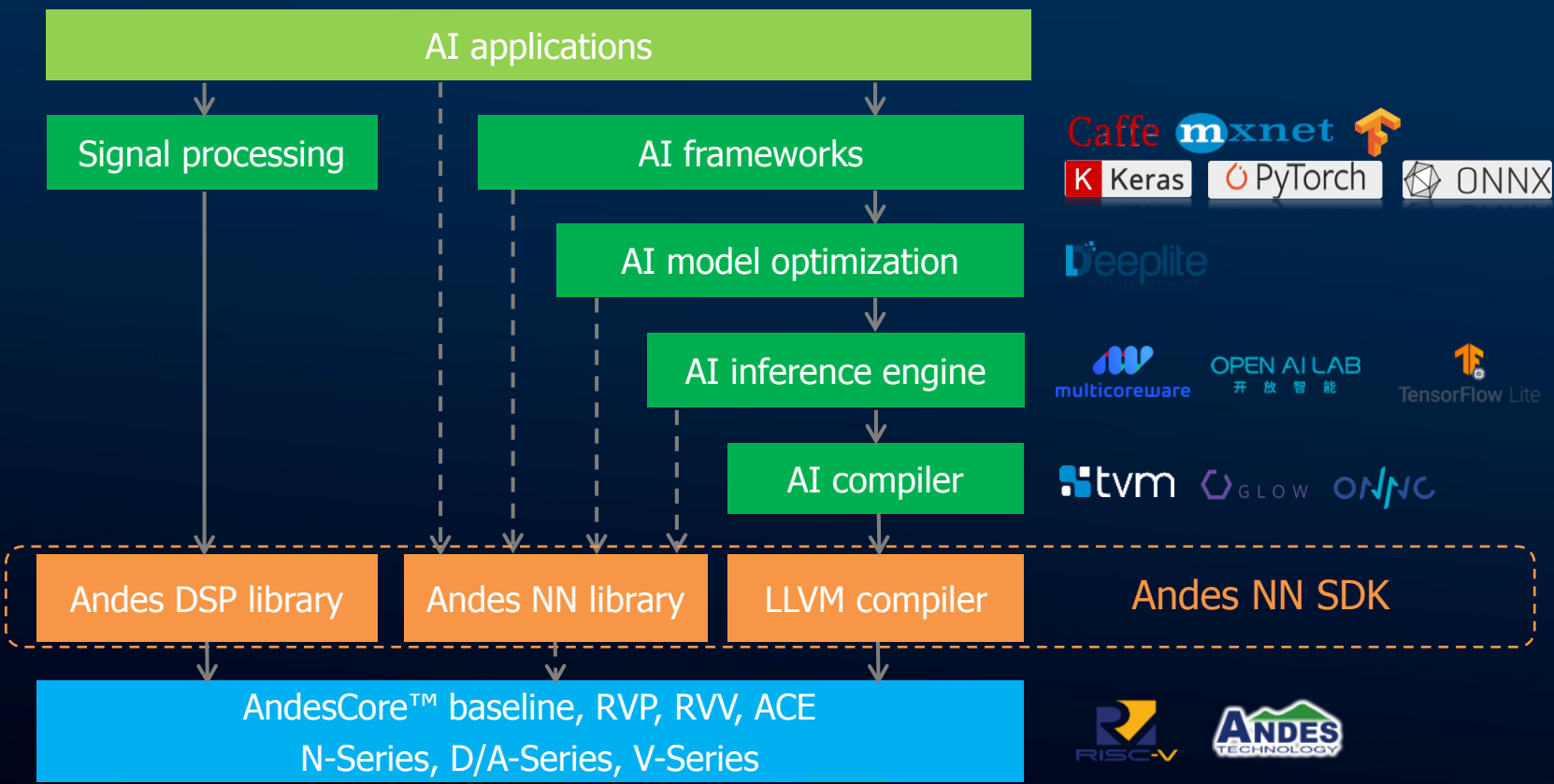
# Deep Learning Chipset Global Market



Tractica, March, 2017

<https://tractica.omdia.com/newsroom/press-releases/deep-learning-chipset-shipments-to-reach-41-2-million-units-annually-by-2025/>

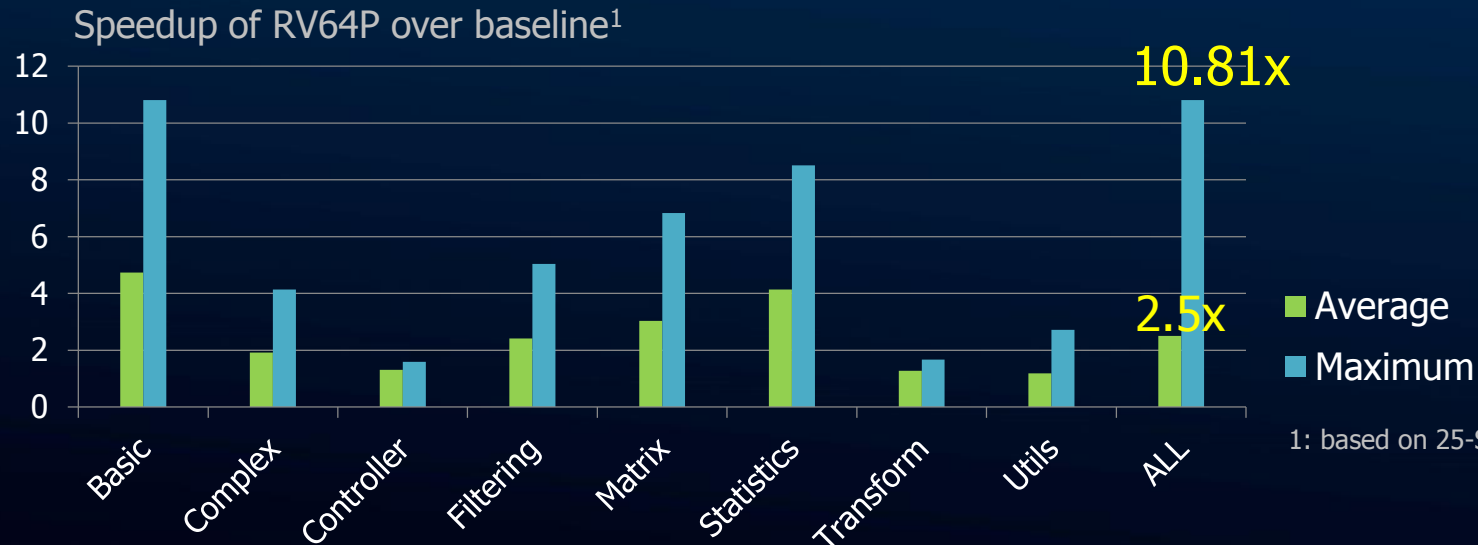
- Deep learning chipset market growing at 42.2% CAGR from 2016 to 2025
- Largest growth coming from ASIC including:
  - CPU
  - DSP
  - Vector processing unit
  - Hard-wired engine
  - ...





# Andes DSP Library

- Optimized low-level DSP functions for RISC-V baseline and RVP processors
- Boost signal processing performance
- >200 functions in 8 categories
- CMSIS-DSP like APIs



1: based on 25-Series, FPGA





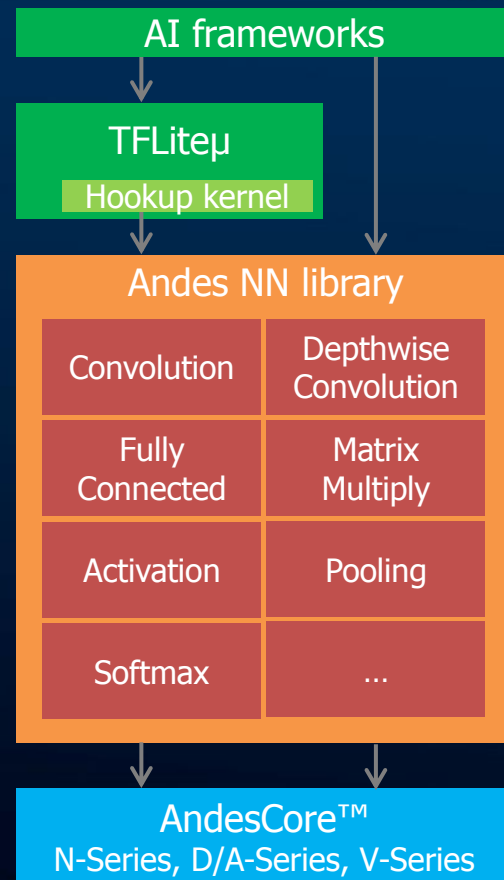
# Andes NN Library and TensorFlow Lite Micro

## Andes NN library

- An optimized low-level NN functions for RISC-V baseline, RVP and RVV processors
- Boost NN performance by using SIMD and vector instructions
- CMSIS-NN like API

## TensorFlow Lite for Microcontroller (TFLiteμ)

- Create bare-metal binary with offline flow
- Major kernel functions hooked up with Andes NN library

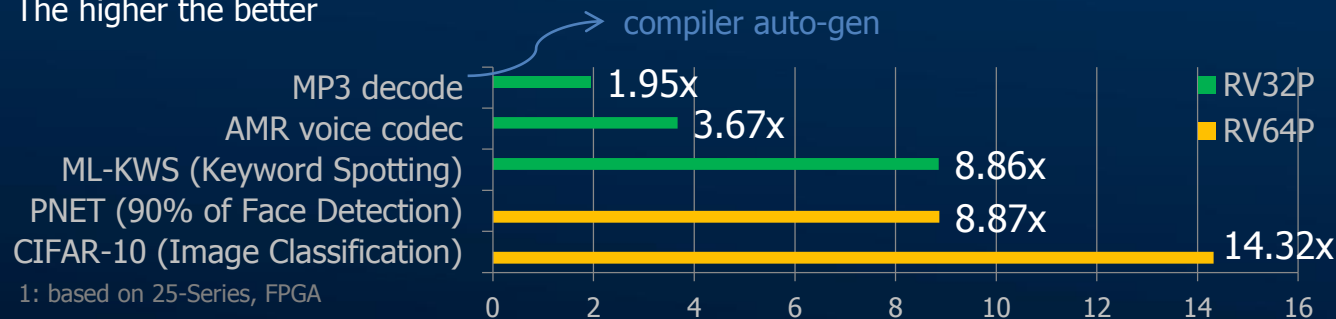




# RVP DSP/SIMD Processors Speedup

## Speedup of RVP over baseline<sup>1</sup>

The higher the better



- ✓ Performance boost with Andes NN/DSP libraries
- ✓ Increase power efficiency
- ✓ Higher response time

## CIFAR-10 image classification speedup<sup>2</sup>

The higher the better





# RVV Vector Processors Speedup over Baseline



## Note

- Compared to pure C scalar code compiled with high optimization
- Both vector and scalar code ran on the NX27V FPGA with 512-bit VLEN, 256-bit bus

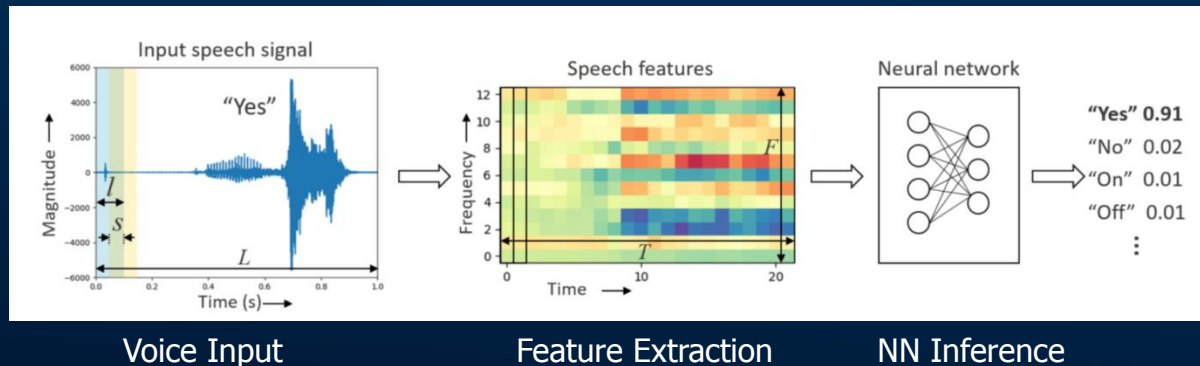


# Andes KWS Solution

<https://arxiv.org/pdf/1711.07128.pdf>

## ■ Voice trigger

- To wakeup the system
- Consume lower power than ASR for always-on usage
- Reduce false alarms



## ■ Voice command

- Hands-free solutions
- Simple and offline HMI

```
Please press any button to start calibrating Silence sound ...
cal=20461610, threshold=44003
```

Commands: yes, no, up, down, left, right, on, off, stop, go

Please press any button and speak command (within 1.25 seconds) now ...

Silence	Unknown	yes	no	up	down	left	right	on	off	stop	go
0	0	127	0	0	0	0	0	0	0	0	0

Detected yes (99%)

```
0000 000 .00000. .0000.0
`88. .8' d88' `88b d88( "8
`88..8' 888000888 `Y88b.
`888' 888 .o o. )88b
.8' `Y8bod8P' 8""888P'
.o..P'
`Y8P'
```

ADP-XC7K





# Andes KWS Solution

## ■ KWS software stack

- Andes NN/DSP library accelerated by Andes RISC-V DSP/SIMD P-ext

## ■ KWS application

- Feature extraction: MFCC
- AI model: DNN, DS-CNN, GRU

## ■ KWS tools

- KWS TensorFlow training script
- KWS quantization tool
- KWS model code-gen to .c/.h

Model	DS-CNN	DNN	GRU
Accuracy	94.4%	84.6%	93.5%
Flash size (code+rodata +data)	186 KB	243 KB	243 KB
SRAM size (data+bss)	35 KB	35 KB	36 KB
Cycles <sup>1</sup>	3,498,638	179,136	5,055,417

1: collected only from one inference sample of WAV file on D25F FPGA

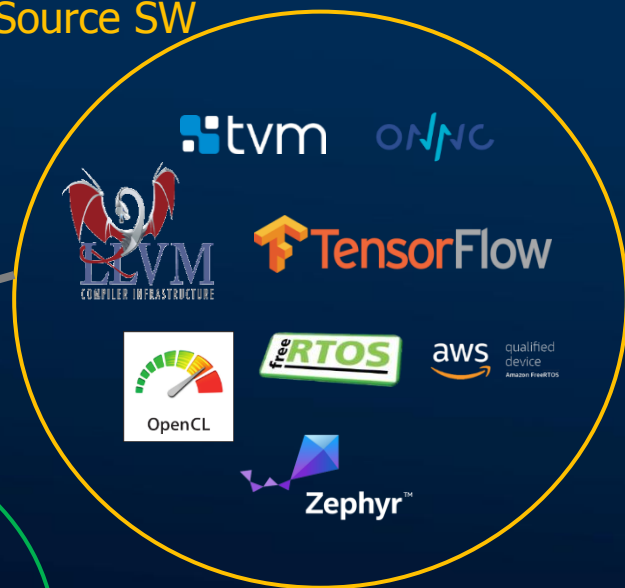


# Andes Partners for AI

AI tools and IP



Open Source SW



DSP and Vision



Development tools

Taking RISC-V® Mainstream





# Summary

- Andes RISC-V processors support the diversity of AI use-cases
  - **Baseline:** compact and modular control logic
  - **Baseline + RVP:** efficient DSP/SIMD for simple data computation
  - **Baseline + RVV:** high performance, efficiency and configurability to enable data intensive computing from edge to cloud
- Andes NN SDK targets to boost your SoC AI performance, achieve outstanding hardware utilization and most importantly, improve time-to-market
  - **Andes DSP library** for signal processing
  - **Andes NN library** for NN operators
- Ecosystem further advances your AI project developments



# RISC-V CON

ONLINE WEBINAR

**Thank you,  
See you next webinar!**