Accelerating RISC-V AI and IoT Development with Andes Software Solutions

Simon TC Wang
Technical Marketing Manager
Andes Technology
April 26, 2021

Agenda

- **AndeSight™ IDE**
- **AndeSoft™ BSP**
- **AI Software Stack**
  - DSP and Vector programming
  - NN library and real NN model use cases

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

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

Quick Facts

16 years old company

80% R&D employees

7B+ (By end of 2020) accumulated Andes-embedded SoC shipped

200+ Licensees Worldwide

20K+ AndeSight IDE installation

Taking RISC-V® Mainstream

AndeSight™ IDE

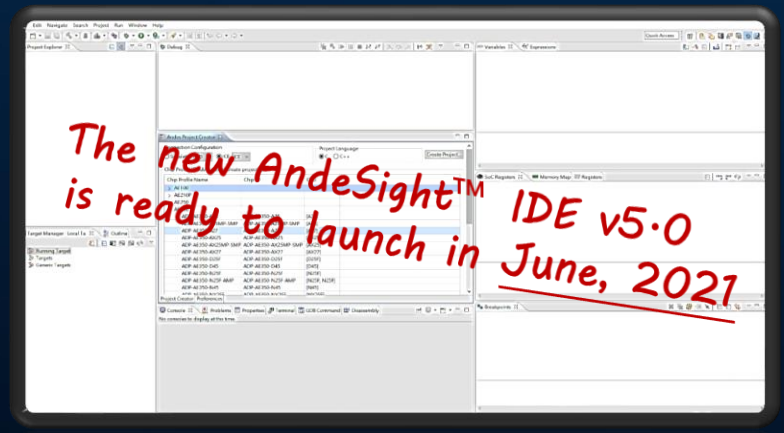
Comprehensive Development Environment

AndeSight™ IDE for RISC-V

Accelerate your RISC-V software developments
with the comprehensive development environment!

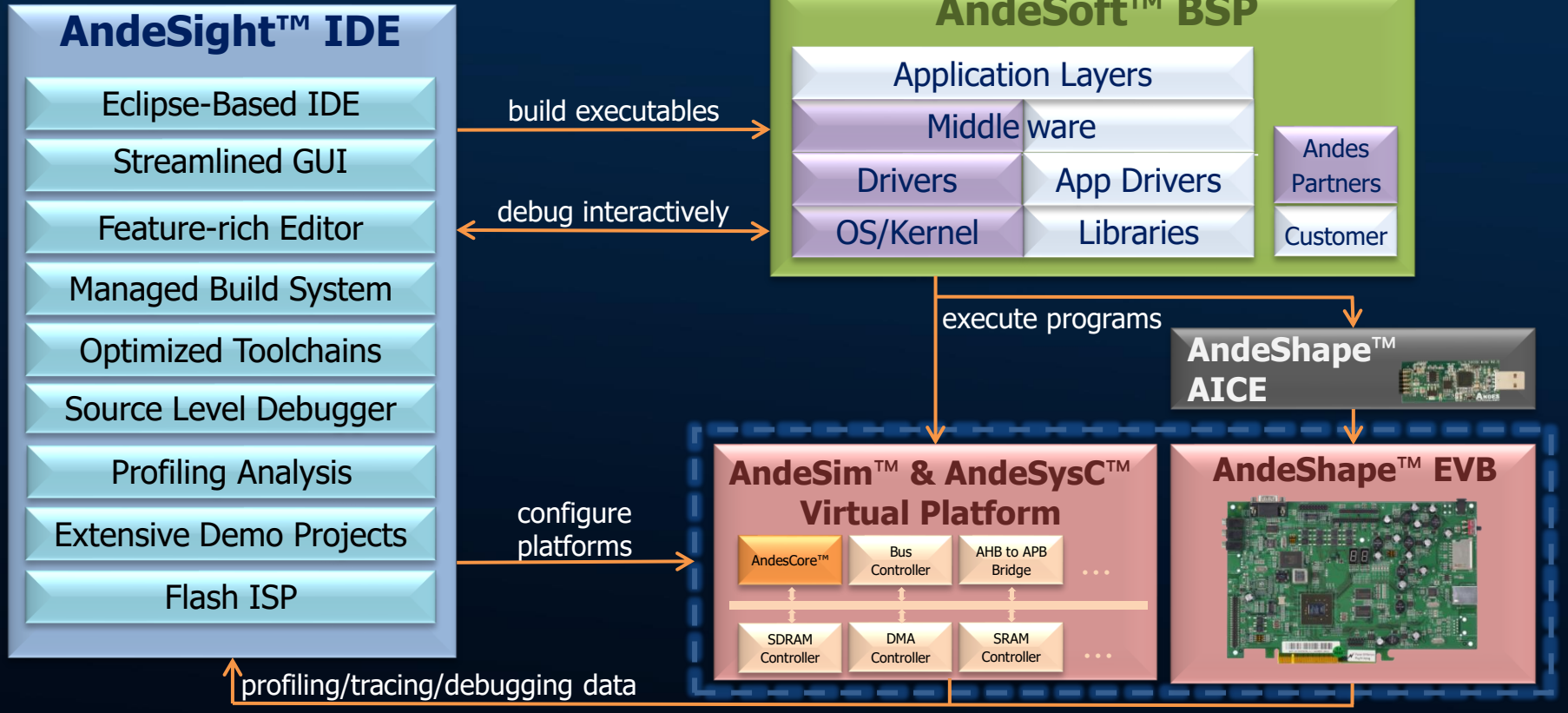
AndeSight™ IDE
with AndeSoft™ BSP inside

- User-friendly and easy-to-use IDE
- Highly-optimized GCC/LLVM toolchains for outstanding performance and compact memory footprint
- Abundant demos for boosting your development
- Rich RTOSes and Linux (LTP verified)
- Optimized compute library: DSP, Vector[1]
- Peripheral drivers for AndeShape™ platform
- Near cycle-accurate simulators: AndeSim™
- Arduino support for Andes Corvette EVB



1: official release in 2021/Q2
2: upcoming solutions

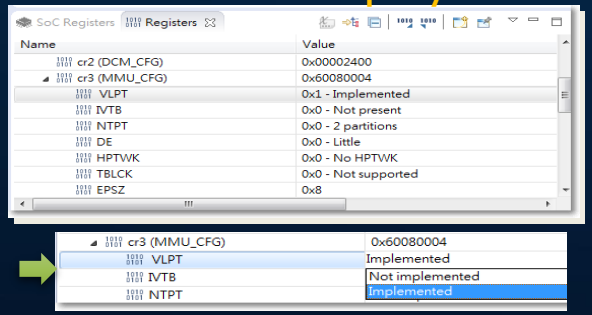
SoC SW Development Environment



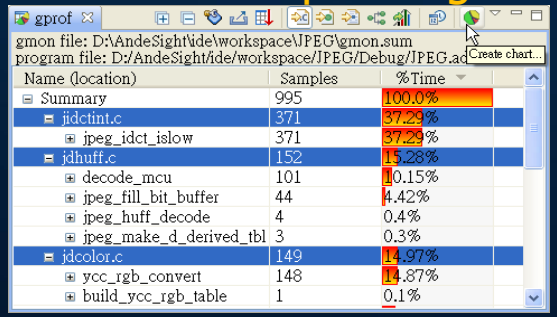
AndeSight™: Professional IDE

- **Eclipse-based IDE, enriched from 16-year continuous development**

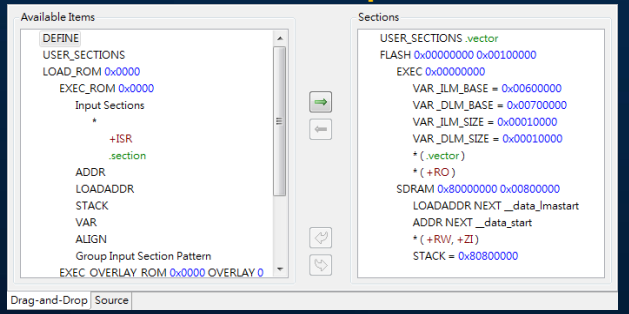
Bit-filed display



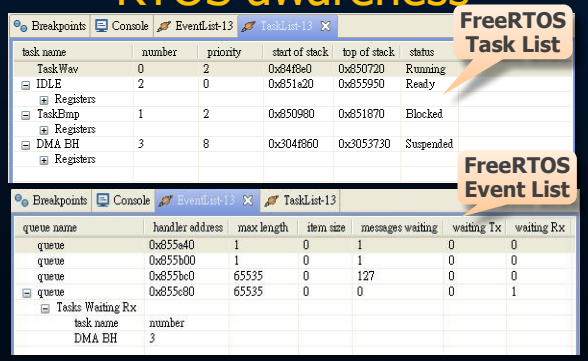
Function profiling



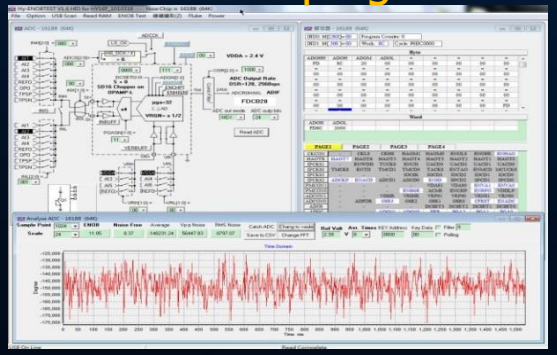
Meta linker script editor



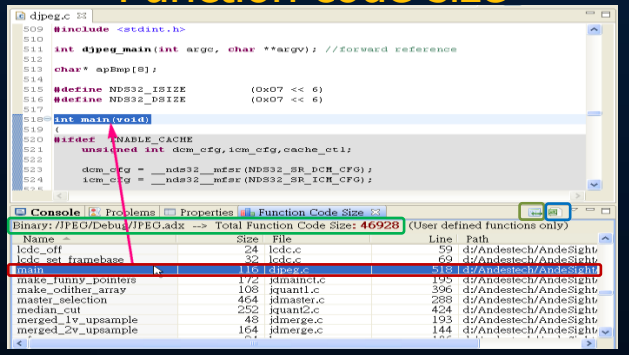
RTOS awareness



Custom plugin

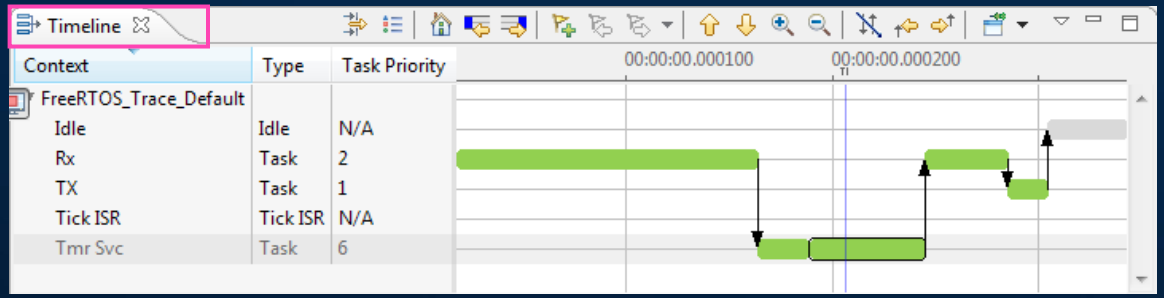


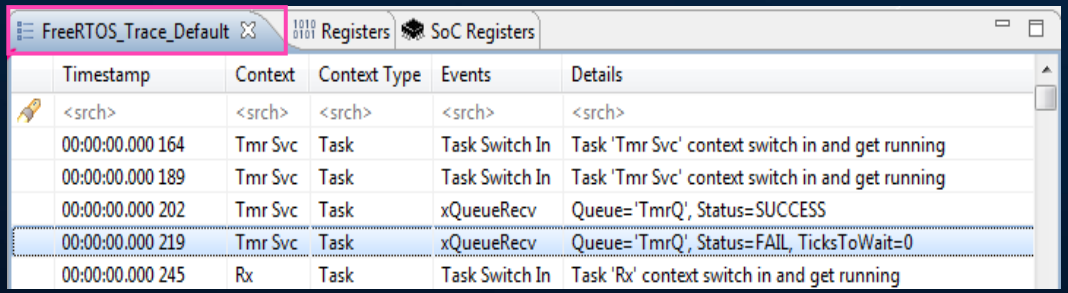
Function code size

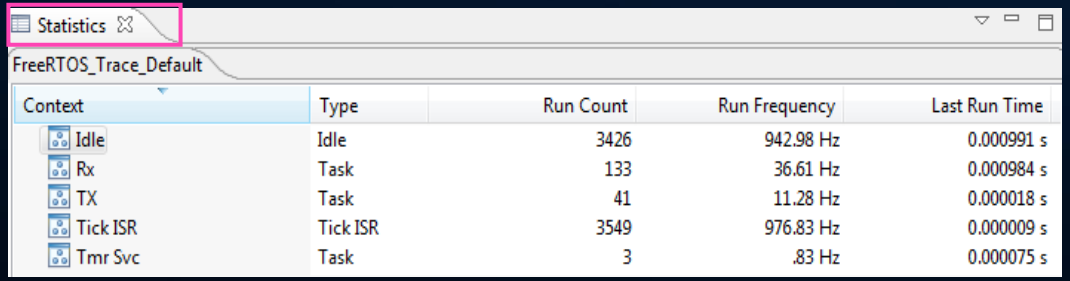


FreeRTOS Timeline Analyzer

- Visualize the runtime execution behaviors of task, interrupts and events within a period of time[1]



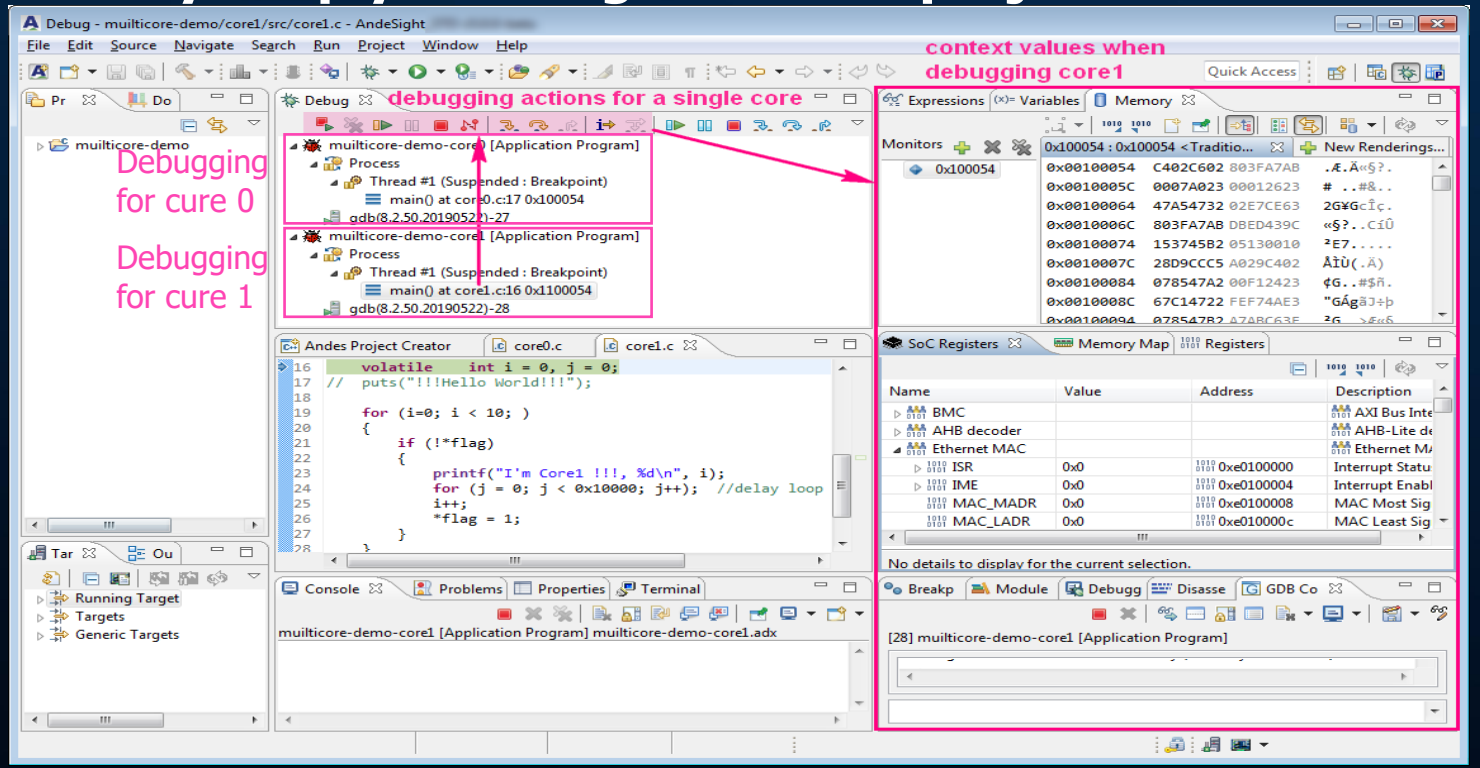




1: EVB target only

Multicore Development Support

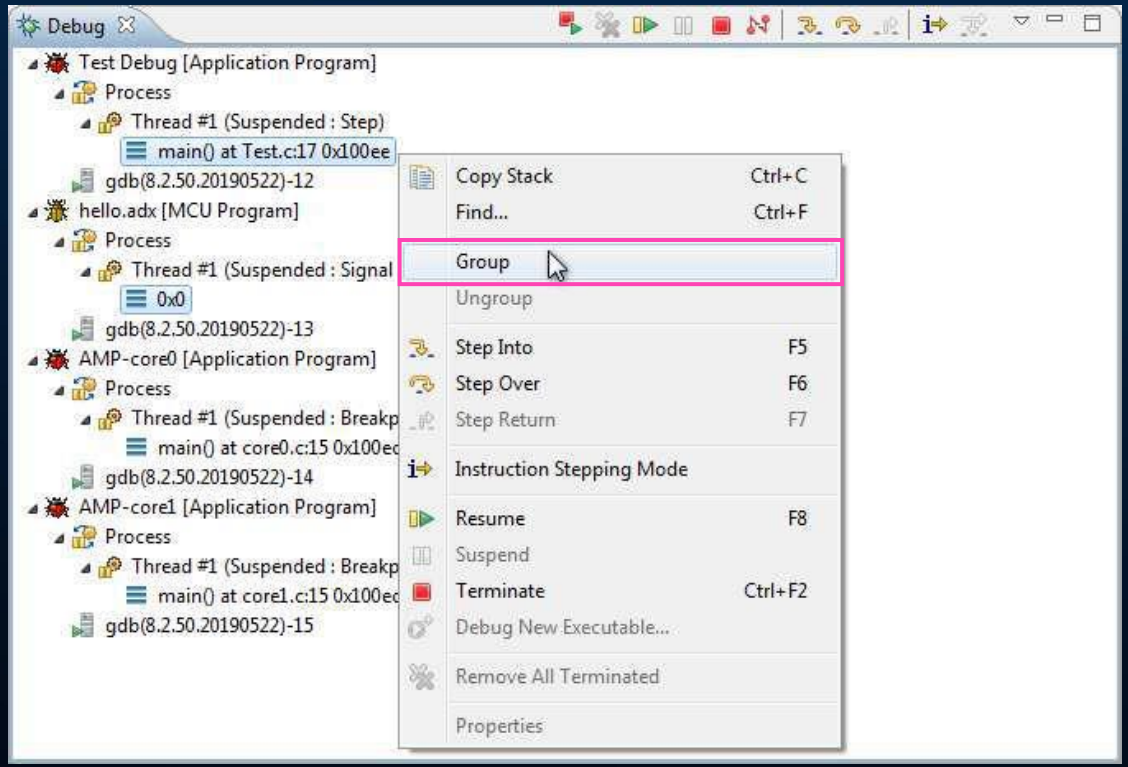
- **Build and debug the multicore software with separate configurations by simply creating multicore projects**



Core Grouping

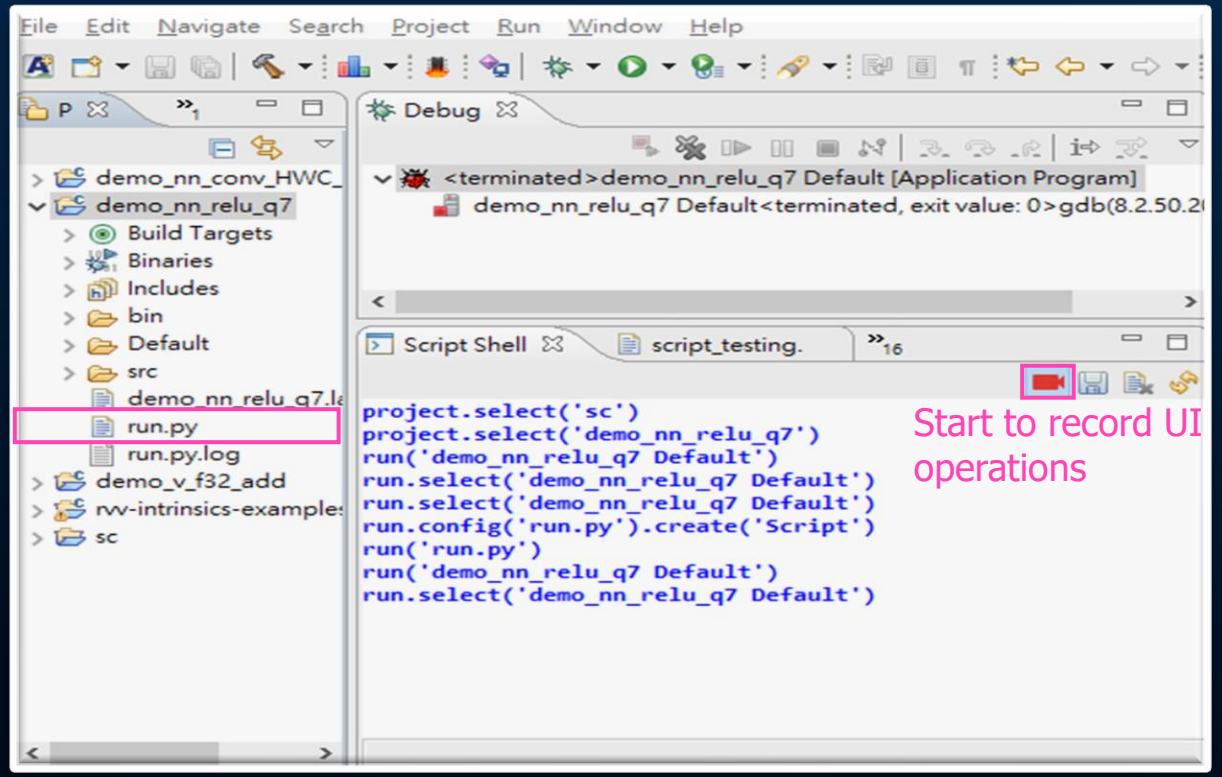
Grouping Cores for Efficient Debugging

- Debug commands can be sent to a specific set of cores at the same time.



AndeSight™ Scripting

- **Record and replay UI operations with Python script**
  - Test automation
  - Issue reproduction
  - …



Python script

Start to record UI operations

AndeSoft™ BSP

Application Building Blocks Inside of AndeSight™ IDE

AndeSoft™ BSP: Application Building Blocks

Fundamental

- Compiler/Toolchain are contributed to and supported officially by **GNU/LLVM** communities
- Optimized **MCUlib**, newlib and glibc
- Optimized low-level compute libraries for NN, DSP and vector processing: **libnn, libdsp, libvec**
- Fast and near cycle-accurate simulators: **AndeSim™, AndeSysC™, Qemu[1]**
- Debugging: **GDB and ICEman (speed-optimized OpenOCD)**
- Concise linker script and its tools, **Linker Scattering-and-Gathering (LdSaG)**
- **Bare-metal drivers and demo programs** to demo AndesCore™ features

Real-Time Operating Systems

aws qualified device Amazon FreeRTOS

- Open source port on Andes: **Zephyr**, **FreeRTOS**
- Commercial port on Andes: **Azure RTOS ThreadX, LiteOS, RT-Thread, SylixOS**
- **RISC-V ready:** VxWorks, µC/OS-[II/III], MyNewt, embOS, RTEMS, NuttX, seL4, uC3/Compact, AliOS Things, TencentOS Tiny, HarmonyOS, Nucleus

Linux, Middleware and SW Framework

- **Linux kernel** v4.17 and LTS v5.4, device drivers and advanced features: **strace**, **ftrace**, **Perf**, **SMU**, **power throttling**, **suspend to RAM, HIGHMEM** and **kernel module**
- **U-Boot, OpenSBI** and **BBL**

1: available upon request

AndeSoft™: Bare Metal

- **Rich startup demo projects for Andes-specific features**

Categories	Startup demo
Interrupt	PLIC, CLIC
Memory	MMU, PMP, PMA, cache, cache lock, ECC, bus matrix slave port
Power Management	PowerBrake, hibernate, WFI CPU standby/resume
Programming	DSP, printf UART redirect, C++ programming
Misc	StackSafe™, performance monitor, SMP

- **AMSI (Andes MCU Software Interface) driver APIs**
  - DMA, Flash, GPIO, I2C, PWM, RTC, SPI, UART and WDT

AndeSoft™: Linux

- **Linux kernel**
  - Mainline Linux compatible
  - V4.17 and LTS v5.4, RV[32|64]GC, SMP
  - Cache coherence and non-coherence support
  - Linux Test Projects (LTP) verified
  - Device drivers for AndeShape™ AE350 platform
- **Kernel features**
  - **strace/ftrace** to debug Linux applications easily
  - **Perf** to evaluate the system bottleneck
  - **Power Management**
    - Suspend-to-RAM: suspended by sysfs and wakeup by RTC and UART interrupt
    - PowerBrake: power throttling controlled by sysfs
  - **Kernel module** all RV32/RV64 relocation type
  - **HIGHMEM** support >1GB memory for RV32
  - **CPU hotplug**
- **U-Boot, U-Boot-spl, OpenSBI and BBL**

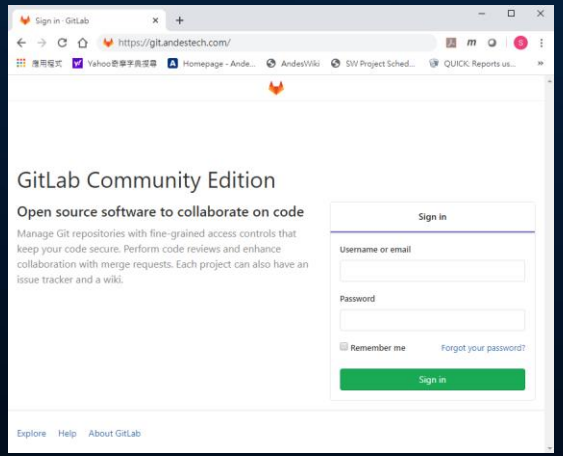
AndeSoft™: Linux

- **Linux Distribution and Build System***

fedora debian yocto PROJECT OpenWrt Wireless Freedom BuildRoot Making Embedded Linux Easy

*: available upon request

- **Andes GitLab Service for Linux Development Packages**
  - A public web service: git.andestech.com
  - Under Andes Maintenance Program

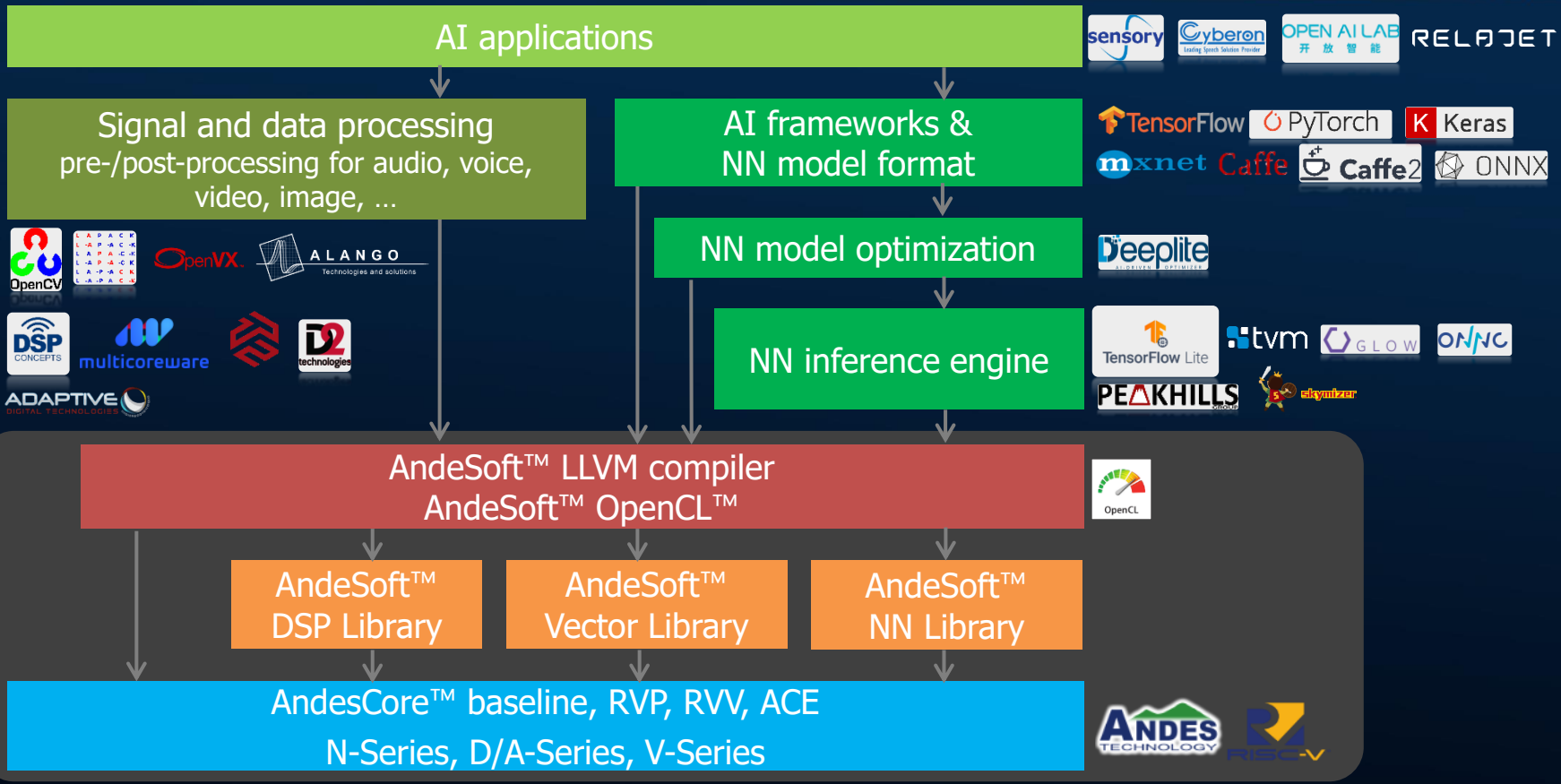


ANDES TECHNOLOGY GitLab

AI Software Stack

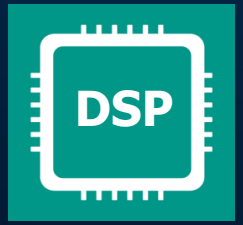
DSP, Vector and NN Programming

AI Software Stack and Ecosystem

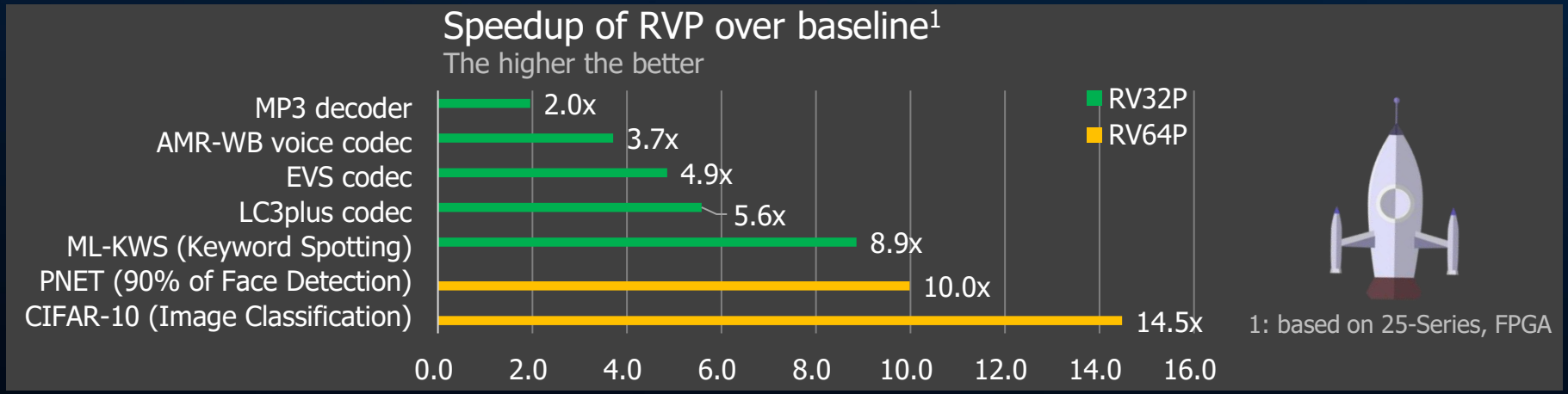


RISC-V DSP/SIMD Software Solutions

First implementation of RISC-V DSP/SIMD extension (RVP)

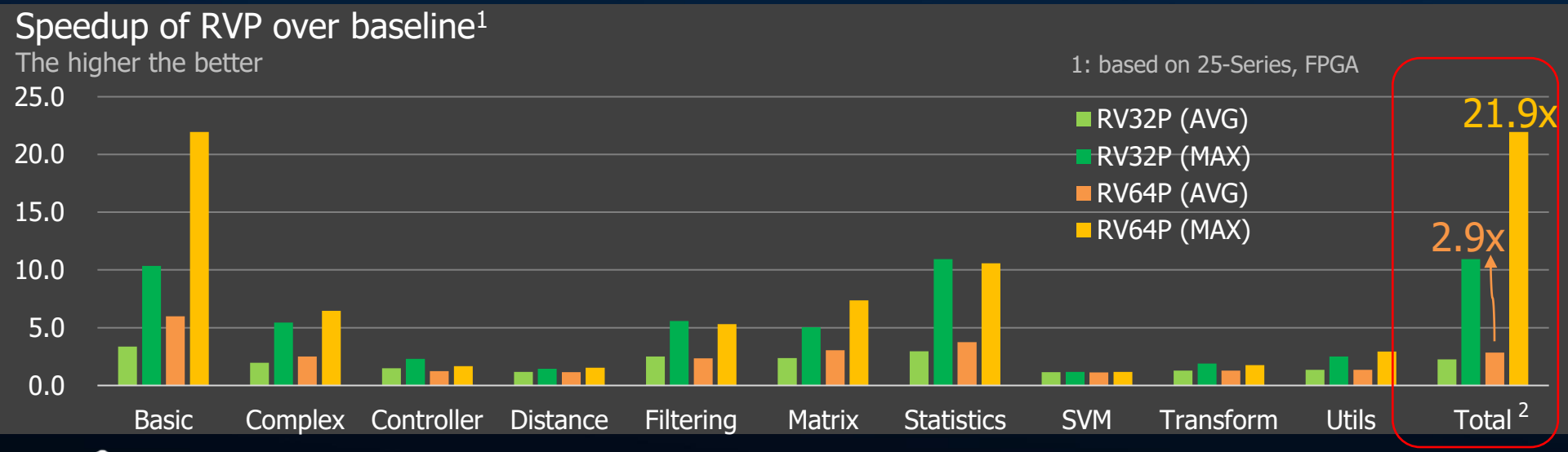


- DSP intrinsic functions: as C-like functions without writing error-prone assembly
- AndeSoft™ DSP library: ~300 functions in 10 categories
  - Basic, complex, controller, distance, filtering, matrix, statistic, SVM, transform, utils
- Increase power efficiency to your DSP applications: 2.0x-5.6x speedup for codecs, and 8.9x-14.5x speedup for NN models



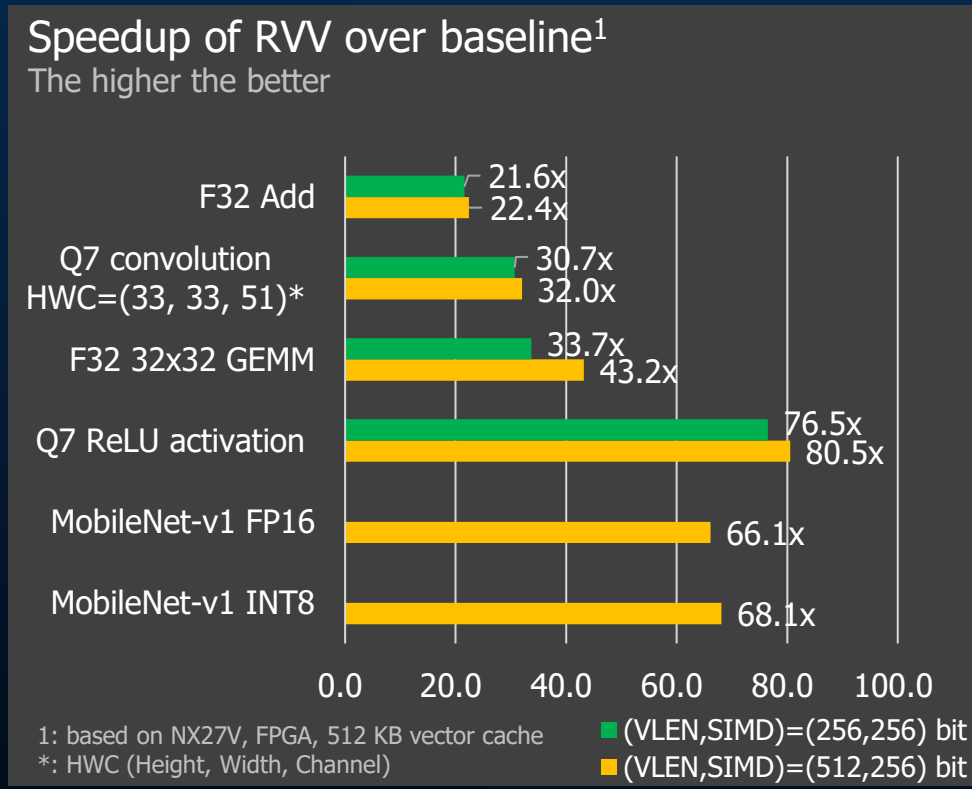
AndeSoft™ DSP Library

- Optimized low-level DSP functions for RISC-V Baseline and RVP processors
- Compatible with CMSIS-DSP API
- Compiler automatically generates partial RVP instructions to facilitate development
- Boost signal processing performance
- Speedup 2.9x in average and 21.9x in maximum

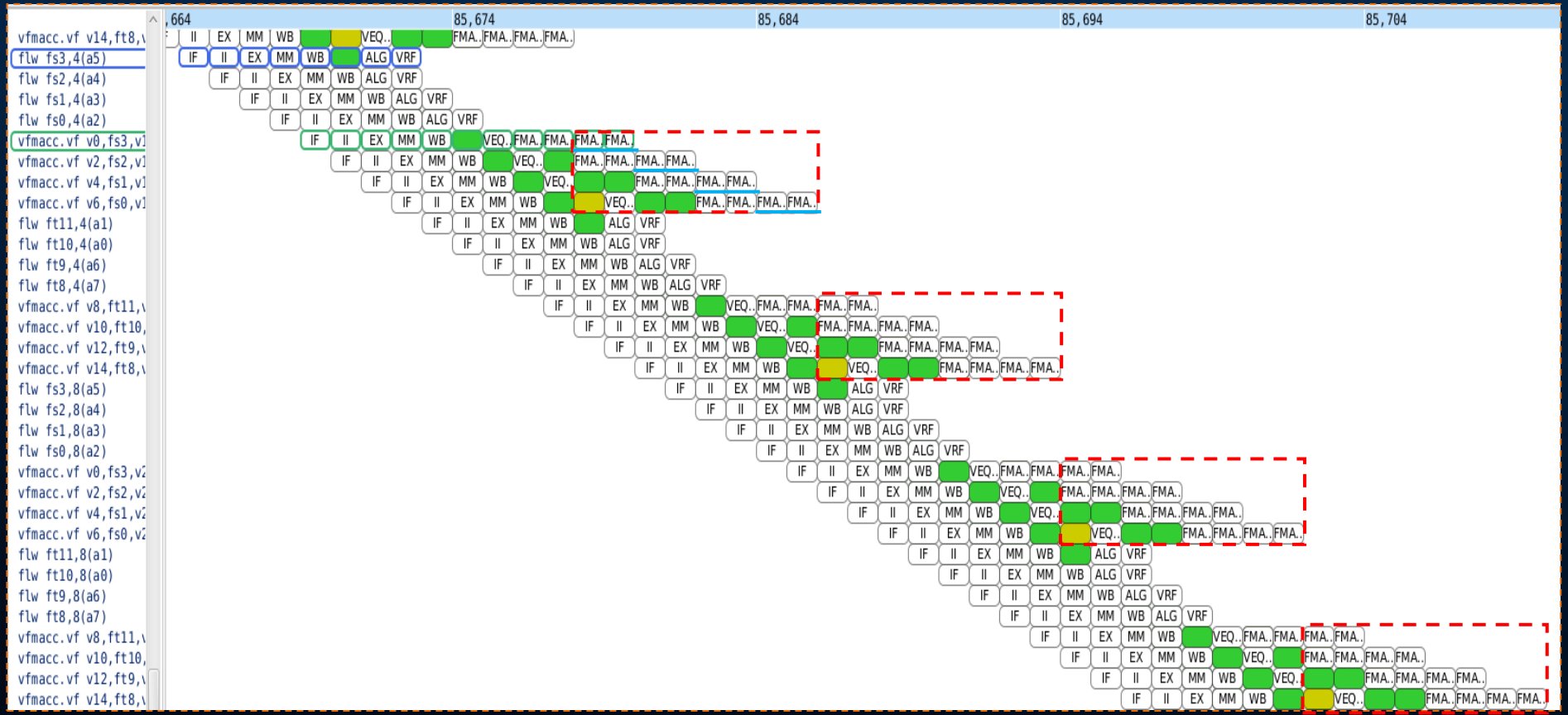


RISC-V Vector Software Solutions

- **Standard software solutions in AndeSight™ IDE**
  - AndeSim™: new cycle-accurate simulator
  - Toolchains: intrinsic functions, as C-like functions without writing error-prone assembly
  - AndeSoft™ Vector library:
    - Optimized for RISC-V baseline and RVV processors
    - > 100 functions in 5 categories: basic, filtering, image, matrix, and transform
    - Compatible with NE10 library APIs
- **Advanced software solutions**
  - AndeSysC™: SystemC support for AndeSim™
  - AndesClarity™: GUI-based processor pipeline visualizer and analyzer
  - AndeSoft™ NN library: optimized neural network functions

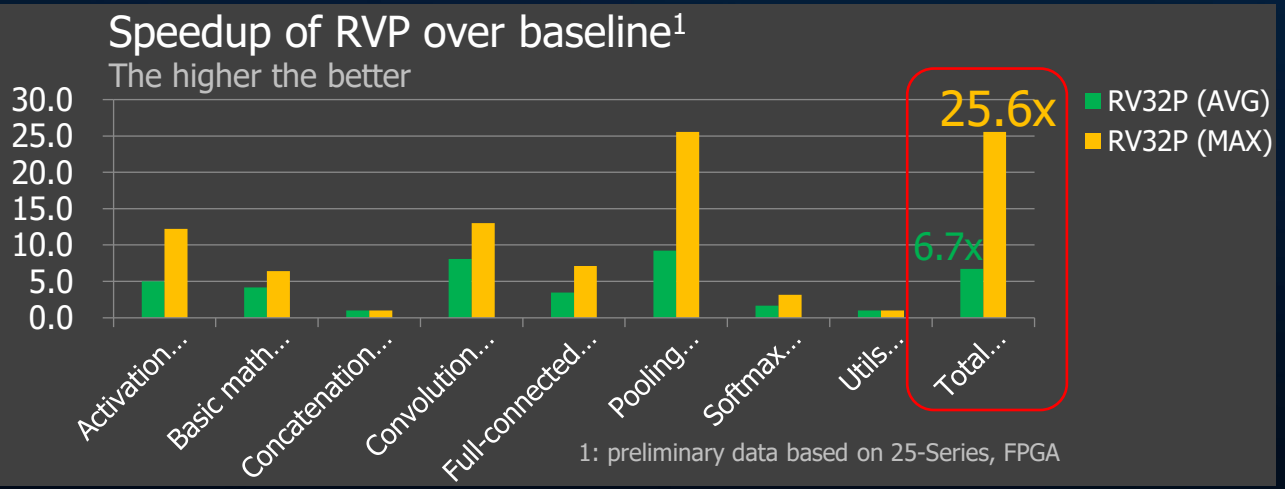


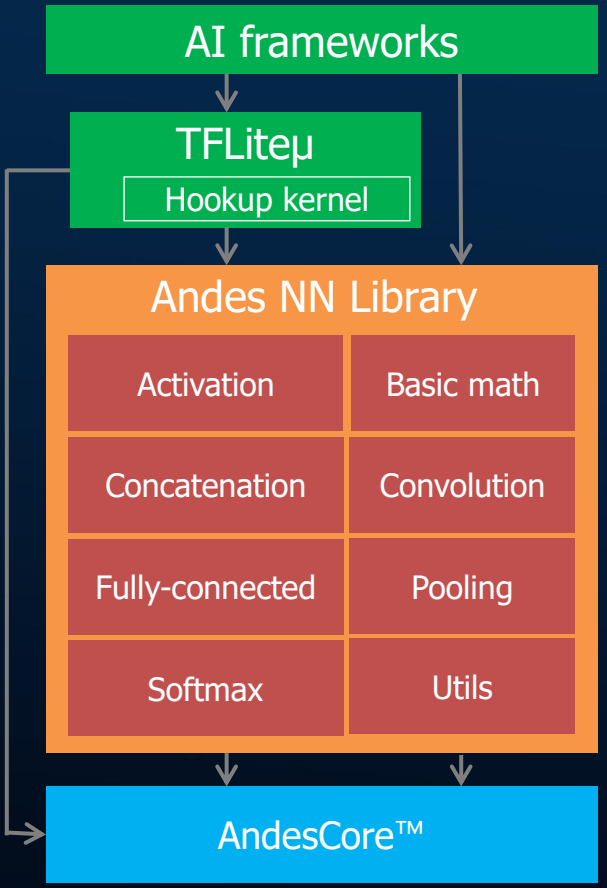
AndesClarity™ Processor Pipeline Analyzer



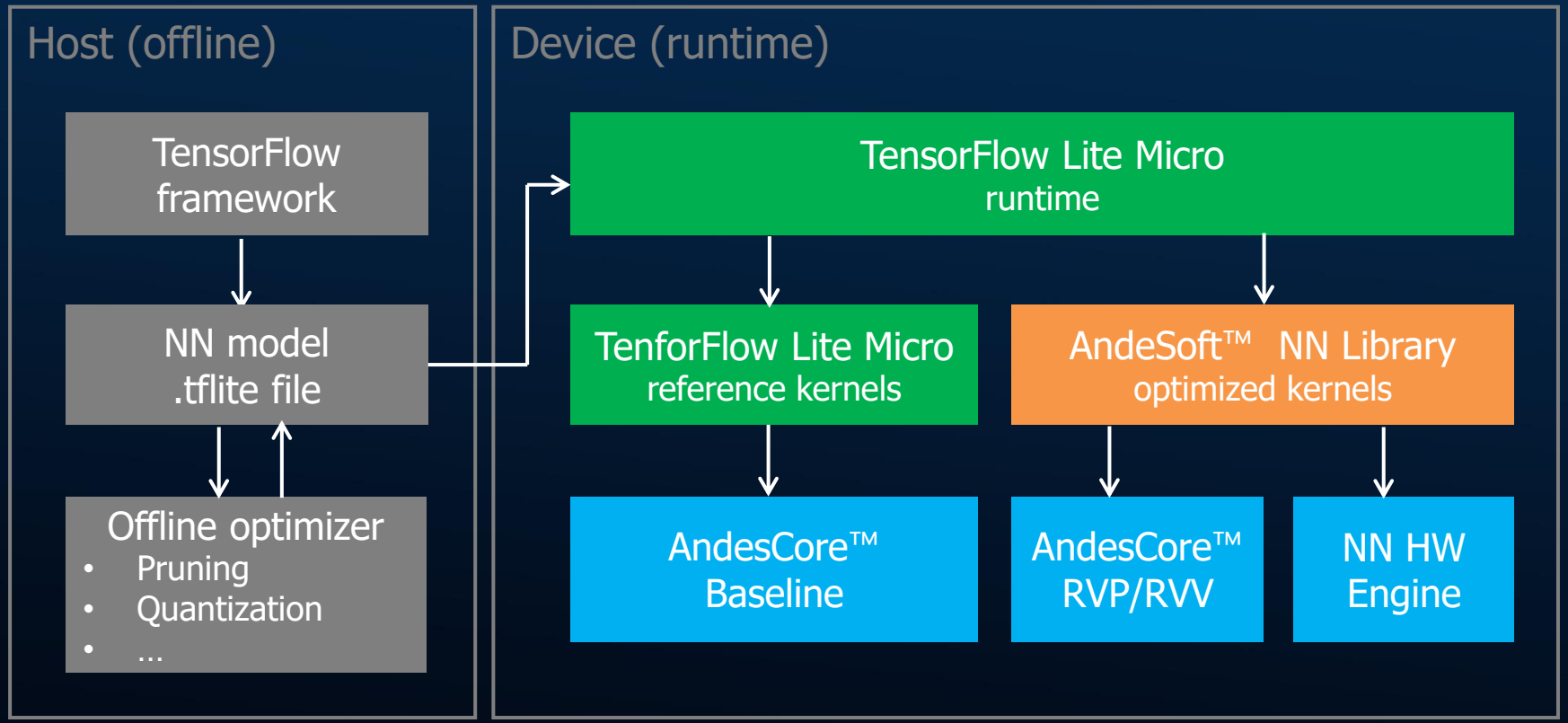
AndeSoft™ NN Library

- Optimized neural network functions for Pure-C, RVP, and RVV processors
- Boost NN performance by using SIMD and vector instructions
- >120 functions in 8 categories
- Compatible with CMSIS-NN APIs





Inference Flow with TensorFlow Lite Micro



AndeSoft™ NN Library API

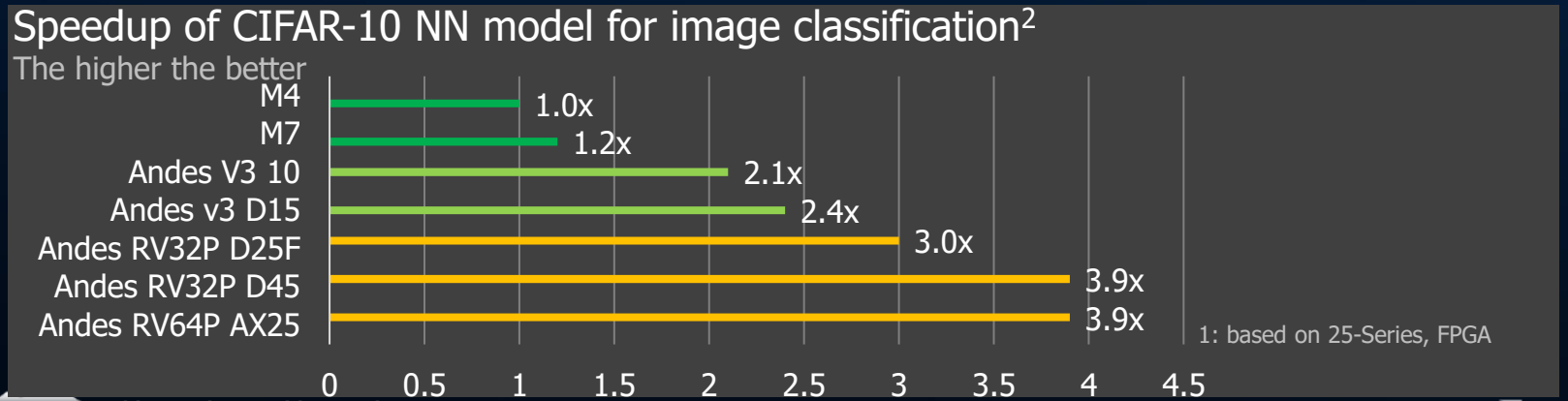
Categories	Operators	NN Library API[1]
Activation	Sigmoid Tanh	riscv_nn_activate_[q7|q15]
	ReLU Leaky ReLU	riscv_nn_relu_[q7|q15] riscv_nn_leaky_relu_q7
Basic	Element-wise	riscv_nn_ew_add_s8 riscv_nn_ew_mul_s8
Concatenation	Concatenation	riscv_nn_concate_s8_[w|x|y|z]
Convolution	Convolution Pointwise convolution Depthwise convolution	riscv_nn_conv_HWC_[q7|u8|s8|q15] riscv_nn_conv_1x1_HWC_[q7|u8|s8] riscv_nn_conv_dw_HWC_[q7|u8|s8]
Fully-connected	Fully-connected	riscv_nn_fc_[q7|u8|s8|q15]
Pooling	Maximum pooling Average pooling	riscv_nn_maxpool_[q7|s8]_HWC riscv_nn_avepool_[q7|s8]_HWC
Softmax	Softmax	riscv_nn_softmax_[q7|u8|s8|q15]
Utils	Reshape	riscv_nn_reshape_s8

1: including but not limited to the list

Use Case: CIFAR-10 NN Model

CIFAR-10	Operators	NN Library APIs
layer 1	Convolution Activation (ReLU)	riscv_nn_conv_HWC_q7_RGB_fast riscv_nn_relu_q7
	Pooling (maxpool)	riscv_nn_maxpool_q7_HWC
layer 2	Convolution Activation (ReLU)	riscv_nn_conv_HWC_q7_fast riscv_nn_relu_q7
	Pooling (maxpool)	riscv_nn_maxpool_q7_HWC
layer 3	Convolution Activation (ReLU)	riscv_nn_conv_HWC_q7_fast riscv_nn_relu_q7
	Pooling (maxpool)	riscv_nn_maxpool_q7_HWC
layer 4	Fully-connected	riscv_nn_fc_q7_fast
layer 5	Softmax	riscv_nn_softmax_q7

- ✓ Performance boost with AndeSoft™ NN library
- ✓ Increase power efficiency
- ✓ Lower response time



Use Case: MobileNet-v1 NN Model

Inference latency of MobileNet-v1 NN model for object detection

The lower the better

MobileNet-v1	Data Type	ISA	VLEN (bit)	SIMD (bit)	Real chip latency (ms)	Nomalized latency[3] (ms @1GHz, 1 core)
CA9 (Xilinx PYNQ)[1]	FP32	NEON		128	721.4 @650MHz, 2 cores	703.4
CA53 (Raspberry Pi-3B)[1]		NEON		128	121.8 @1.2GHz, 4 cores	438.5
CA72 (Firefly RK3399)[1]		NEON		128	77.9 @1.8Hz, 2 cores	210.3
CA73 (Kirin 970)[1]		NEON		128	41.3 @2.36GHz, 4 cores	292.4
Andes NX27V[2]	FP16	RVV	512	256		62.3
Andes A27L2[2]	INT8	RVP		32		564.5
Andes AX27L2[2]		RVP		64		358.5

Table 1. MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	3 × 3 × 3 × 32	224 × 224 × 3
Conv dw / s1	3 × 3 × 32 dw	112 × 112 × 32
Conv / s1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw / s2	3 × 3 × 64 dw	112 × 112 × 64
Conv / s1	1 × 1 × 64 × 128	56 × 56 × 64
Conv dw / s1	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw / s2	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 256	28 × 28 × 128
Conv dw / s1	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw / s2	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 512	14 × 14 × 256
5× Conv dw / s1	3 × 3 × 512 dw	14 × 14 × 512
5× Conv / s1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw / s2	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 1024	7 × 7 × 512
Conv dw / s2	3 × 3 × 1024 dw	7 × 7 × 1024
Conv / s1	1 × 1 × 1024 × 1024	7 × 7 × 1024
Avg Pool / s1	Pool 7 × 7	7 × 7 × 1024
FC / s1	1024 × 1000	1 × 1 × 1024
Softmax / s1	Classifier	1 × 1 × 1000

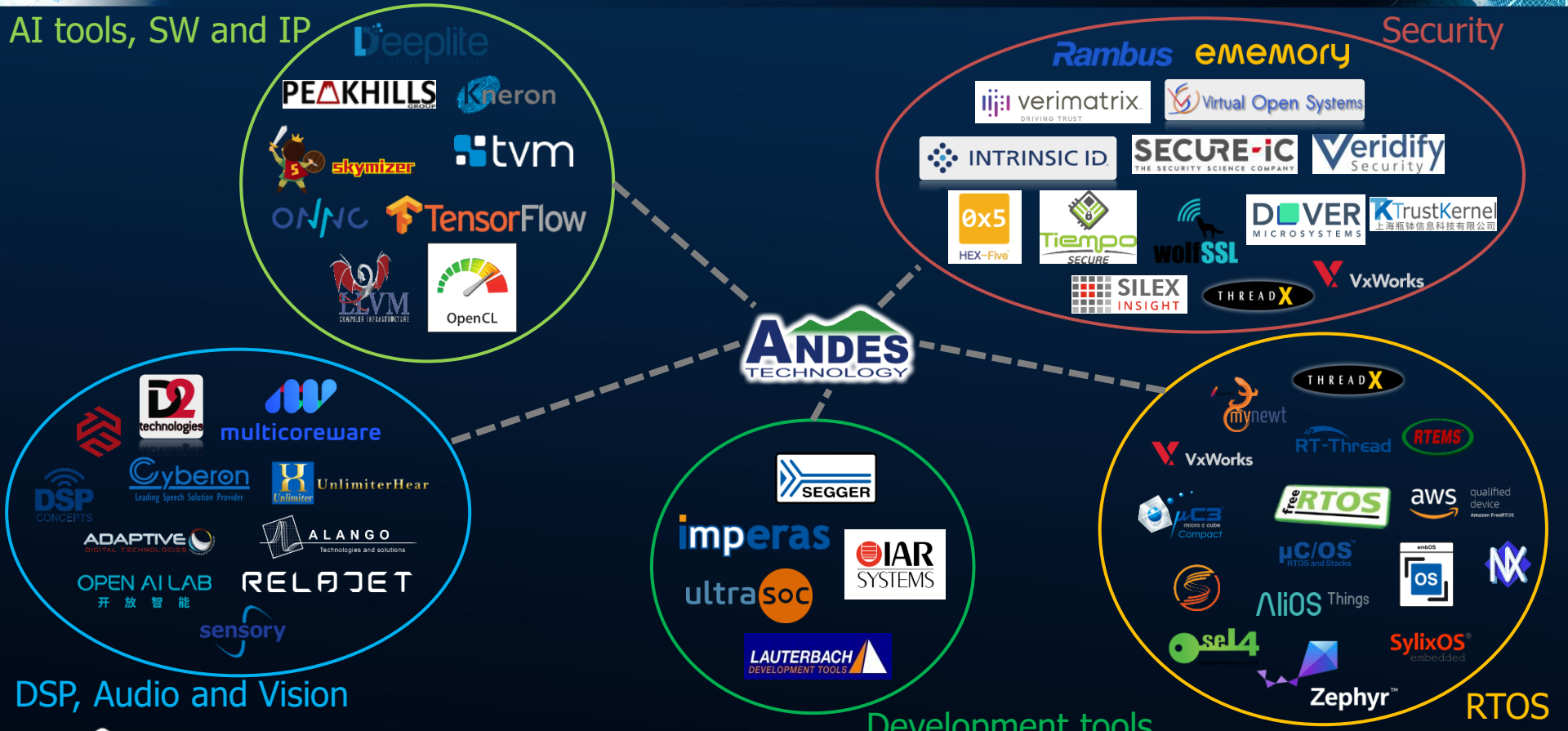
Table 2. Resource Per Layer Type

Type	Mult-Adds	Parameters
Conv 1 × 1	94.86%	74.59%
Conv DW 3 × 3	3.06%	1.06%
Conv 3 × 3	1.19%	0.02%
Fully Connected	0.18%	24.33%

1: TVM, https://github.com/apache/tvm/wiki/Benchmark#arm-cpu
2: AndeSoft™ NN Library, PyTorchCV imagenet-1k "MobileNet x1.0", https://pypi.org/project/pytorchcv/
preliminary data based on FPGA and scaling to 1.0 GHz. Real SoC performance will depend on memory subsystem
3: Scale to the same frequency; estimated performance 4 cores="3"*1 core, the magic number "3" is from experience

Andes Ecosystem

Andes Partners and Ecosystem

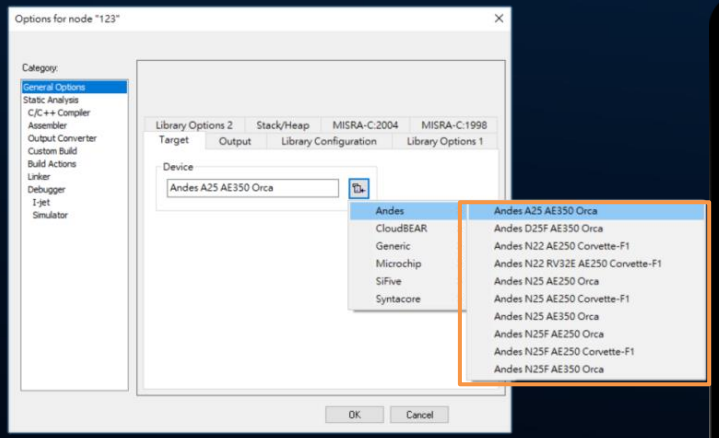


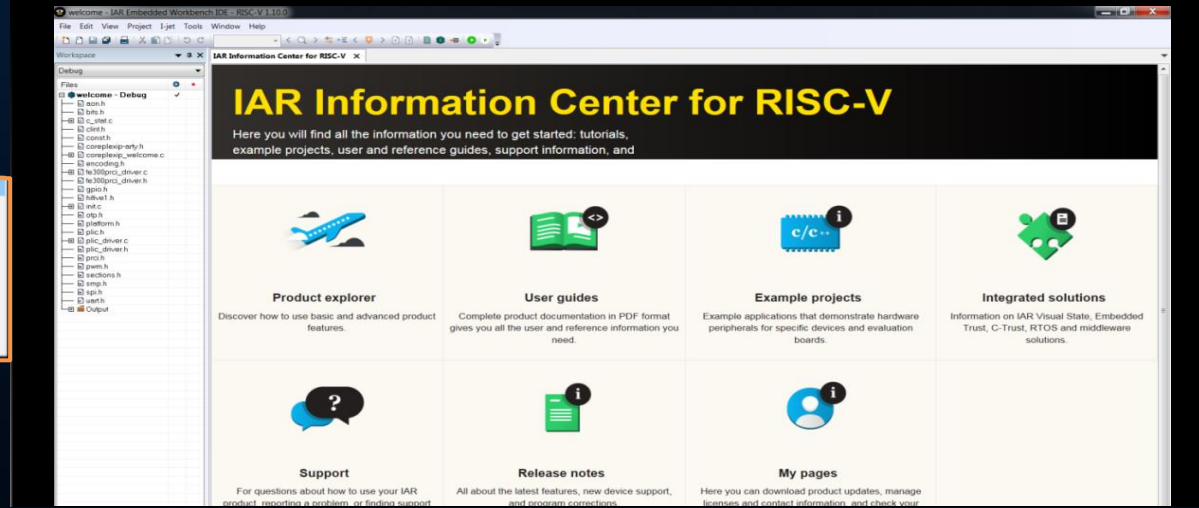
IAR Embedded Workbench for RISC-V

- **Complete build and debug toolchain for RISC-V**
- **Support all series of Andes RISC-V CPU[1]**
- **Functional Safety Certified**
  - IEC 61508
  - ISO 26262
  - EN 50128
  - IEC 62304

IDE	Debugger	Toolchain
Embedded Workbench	I-Jet ICE	Compiler
	C-SPY tool	C-STAT
		Functional Safety

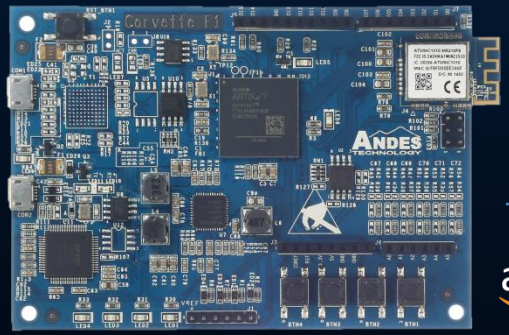
1: AndesCore™ 45-Series will be available in the mid-year of 2021
2: Part of Andes V5 RISC-V extensions are supported in the mid-year of 2021

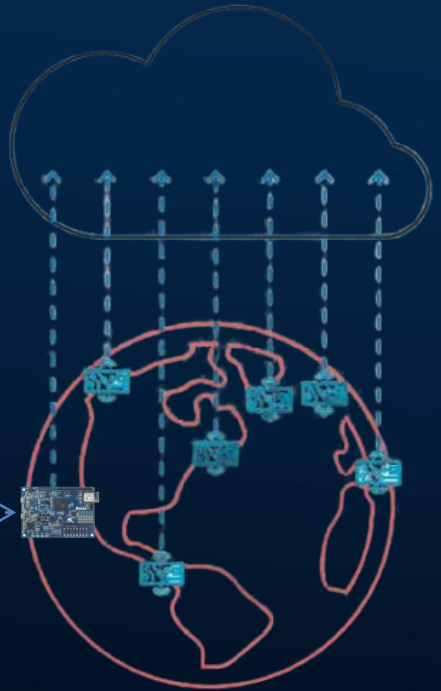




AWS Device Qualification Program for FreeRTOS

- **AndeShape™ Corvette-F1 N25 now is an AWS FreeRTOS-qualified evaluation platform**
- **The FreeRTOS port is verified and executes smoothly and consistently with AWS IoT**
  - FreeRTOS kernel, Wi-Fi management, TCP/IP stack, PKCS#11[1], TLS[2], MQTT[3]





aws qualified device Amazon FreeRTOS

freeRTOS

1: Public Key Cryptography Standards (PKCS) #11 refers to the cryptographic token interface
2: Transport Layer Security
3: Message Queue Telemetry Transport

Summary

- **AndeSight™ IDE**
  - User-friendly and easy-to-use IDE
  - Accelerate your RISC-V software developments with the comprehensive development environment
- **AndeSoft™ BSP**
  - Highly-optimized toolchains for better performance and smaller memory footprint
  - Well-integrated building blocks to reduce time-to-market
- **AI Software Stack**
  - Boost the performance for RISC-V AI and IoT applications with optimized AndeSoft™ DSP, Vector and NN libraries

AndeSight™ IDE Free Download

- Previous version for evaluation
- Three-month time limit

Thank you