Tech Insights

Linley Fall Processor Conference 2022

November 1-2, 2022

Premier Sponsor

flexlogix Technologies, Inc.

Platinum Sponsor

ceremorphic®

Gold Sponsors

CEVA
cadence
ARTERIS IP
MOVELLUS
TetraMem
MARVELL
GlobalFoundries
SYNOPSYS® Silicon to Software
Achronix Data Acceleration
proteanTecs
ANDES TECHNOLOGY
expedera
kinara
UNTETHER AI
quadric
innatera
brainchip
Imagination
SIEMENS
BLUE CHEETAH
Qualcomm
NXP
SiFive
intel

TechInsights

Session 8: CPU IP and GPU IP

- **Architecture and Key Features of SiFive's Newest Out-of-Order Vector Processor**
  - Shubu Mukherjee, Vice President Architecture, **SiFive**
- **Andes Technology’s Next-Generation Scalable RISC-V Application Processor Family**
  - Charlie Su, President and CTO, **Andes Technology**
- **Power-Efficient Scalable Ray Tracing GPUs**
  - Kristof Beets, VP of Technology Insights, **Imagination**
- **Addressing Scalable Processor Performance in High-End Embedded Applications**
  - Kulbhushan Kalra, Engineering Manager, ARC, **Synopsys**

Enhancing the SiFive Performance Portfolio

Extended family of area & power efficient processors

Oct, 2022

Legacy ‘efficiency processors’ are not delivering on industry needs

- Latest market requirements are not being met by current suppliers
- Innovation not matching industry expectations and needs for over 5 years
- Latest innovations from SiFive brings significant upgrade opportunities
- SiFive vector compute brings performance boost and power efficiency
- SiFive Performance portfolio enables greater design flexibility

SiFive Performance™ Family

Market **leading RISC-V** Application **Processors**

- Performance density leadership
- First with latest RISC-V features, standards, and technology
- High performance with optimized power efficiency
- SiFive momentum with NASA, Google, and Intel Horse Creek

Market requirements for wearables

Smartwatch, sport watch, fitness tracker

Performance efficiency is critical

- Feature-rich OS demand aggressive design innovation
- Advanced features put stress on power envelope
- Physical dimensions require optimized area

SiFive solutions

- Best compute density enables greater flexibility
- Future-proof for next generation premium wearables
- Vector computing for AI/ML, media and sensor processing
- Path to Android Wear OS with RVA22/Platform-A

Market requirements for smart home appliances

Home assistant, smart TV, STB, smart speaker, thermostat, door bell, security camera

High processing power and edge AI required

- Audio processing & voice activation/recognition
- Edge AI vision for object detection & filtering
- 4K+ video encoding/decoding
- Network connectivity

SiFive solutions

- Broad portfolio to address full range of devices
- Vector compute for AI/ML, sound & media processing
- Auto-vectorizing compiler simplifies product development
- Vector cryptography for TLS/SSL acceleration
- Strong Linux community with standard software & RVA22

Market requirements for mobile

Feature phones, smartphones

Need for more performance & efficiency

- Requirement for dynamic scalable processor architecture
- Phone apps and UI drive constant innovation
- Battery life remains a key requirement

SiFive solutions

- Mixing high-performance and high-efficiency cores
- Vector computing for AI/ML and video workloads
- RISC-V standardized RVA22 to enable Android OS
- System-level virtualization support
- Advanced power management features

Market requirements for network appliances

Router, switch, WiFi AP, 5G base station

Need for performance & data throughput

- High throughput & massive parallel processing needs
- Hardware isolation for better software security

SiFive solutions

- Linux enabled high performance efficient processors
- Coherent multi-core and multi-clusters capability
- Flexible multi-level cache with cache stashing
- Vector crypto for TLS/SSL acceleration
- System-level virtualization support & WorldGuard

SiFive Performance Family

2023 Product Lineup

Hypervisor extension
IOMMU & AIA
WorldGuard
RVA22

Common features

P670

>12.6 SpecINT2k6/GHz
4-wide OoO core
2x 128b VLEN RVV
Vector crypto

P650

>11.5 SpecINT2k6/GHz
4-wide OoO core

P470

>8 SpecINT2k6/GHz
3-wide OoO core
1x 128b VLEN RVV
Vector crypto

P450

>6.6 SpecINT2k6/GHz
3-wide OoO core

Home Appliance | Wearable | Feature phone | Smartphone / Premium wearable | Network appliances

SiFive Performance™ P470

Boosted
Performance

Significant upgrade to legacy efficiency cores

Small
Area

Optimized area for power constrained applications

Efficient
Power

Highly-tuned for aggressively low power consumption

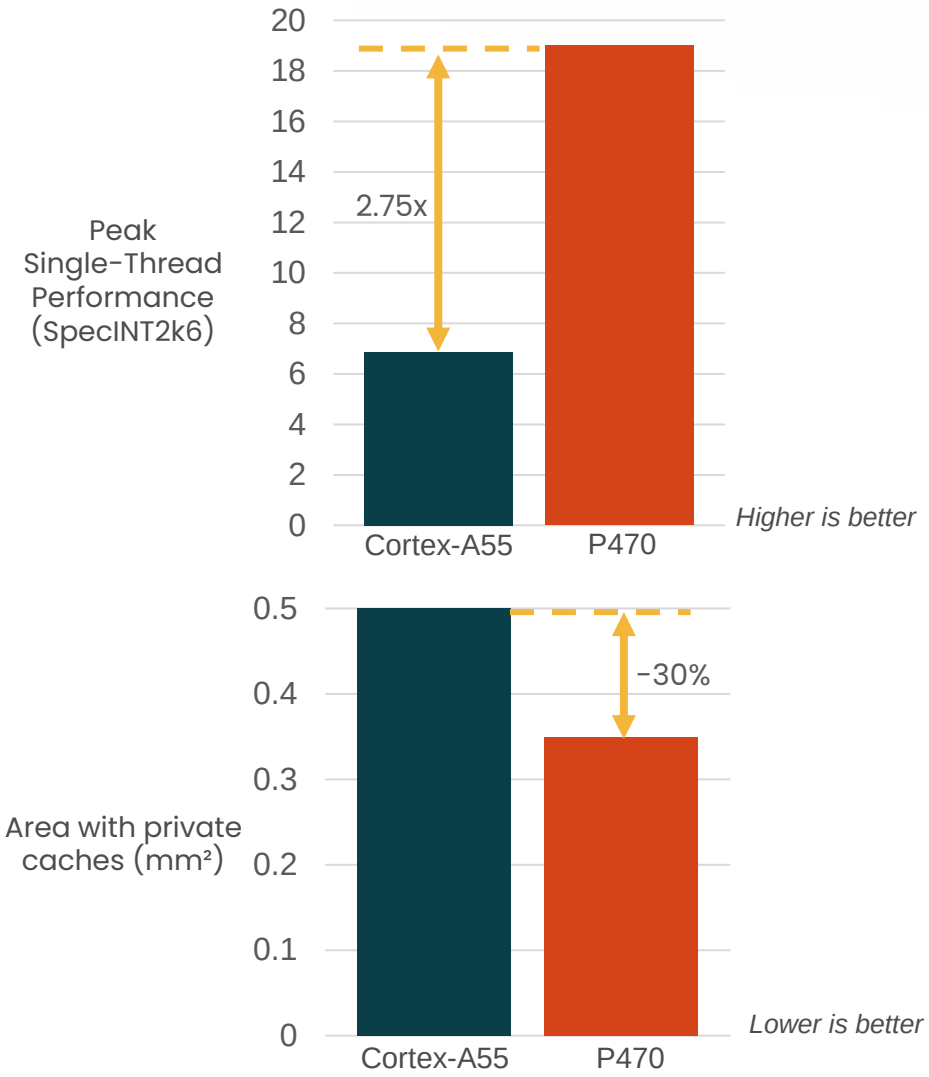
Optimized
Pipeline

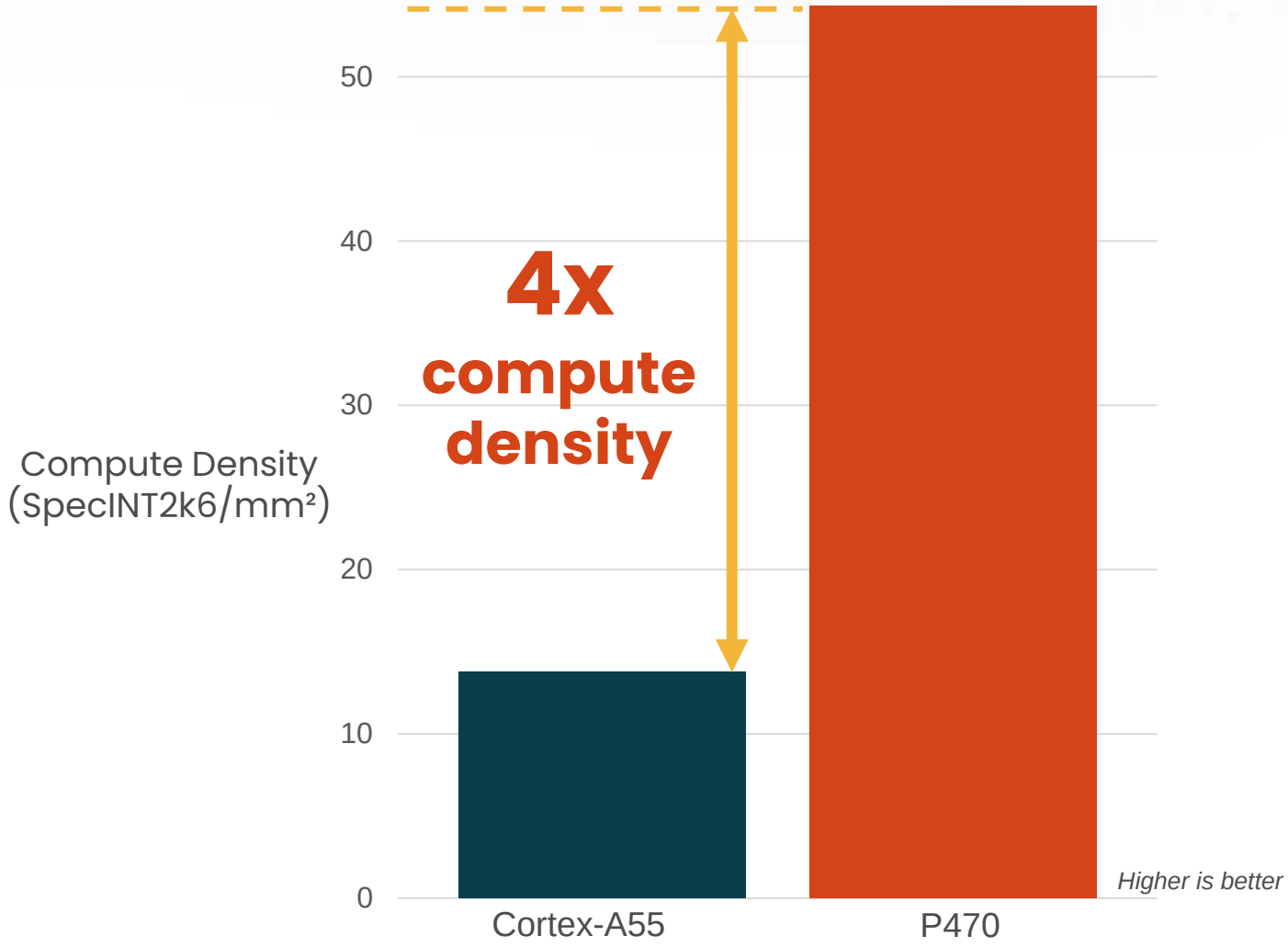
Out-of-Order pipeline enables optimal performance efficiency

RISC-V
Compliant

Compliant with RVA22 profile, with support for Vector and Vector Crypto extensions

P470 Peak Single-Thread Performance





*Cortex-A55: **Performance** 1.88GHz measured on Acer Spin 513 Chromebook with Qualcomm Snapdragon 7c. **Area**: 7nm TechInsights - Die photos show Cortex-A78 shortfall",*
*SiFive P470: **Performance** 2.97GHz, 0.95V 32KB L1 I$ and D$, 2MB L2$. **Area**: measured in 7nm*

Compute density matters

Flexibility to best meet application needs, power budget and cost envelope

SoC Cost Reduction

Optimal performance in the smallest area

Performance Increase

Higher performance in an equivalent area

Maximize Cores

Integrate more cores for optimal system design

SiFive P400-Series

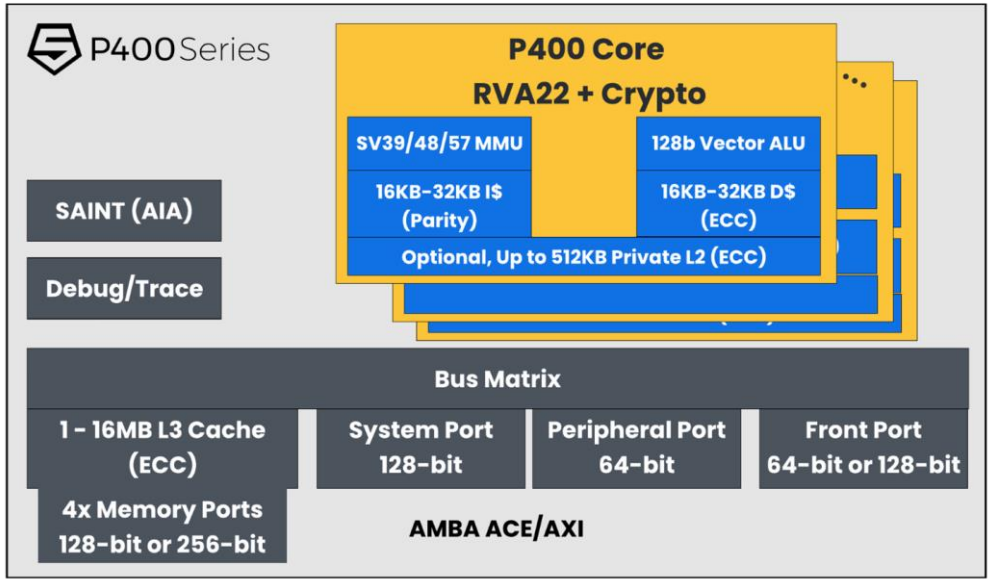
Out-of-Order Performance, Extreme Area and Power Efficiency

Based on an Out-of-Order pipeline finely tuned to bring double the performance of competing Efficiency cores while maintaining similar Area and Power footprints. P470 is compatible with P670 enabling heterogeneous implementations.

>8 SPECInt2k6/GHz in the highest performance configuration

RVA22 Compliant with RISC-V Vectors and RISC-V Vector Crypto

Extreme Compute per mm2 per mW

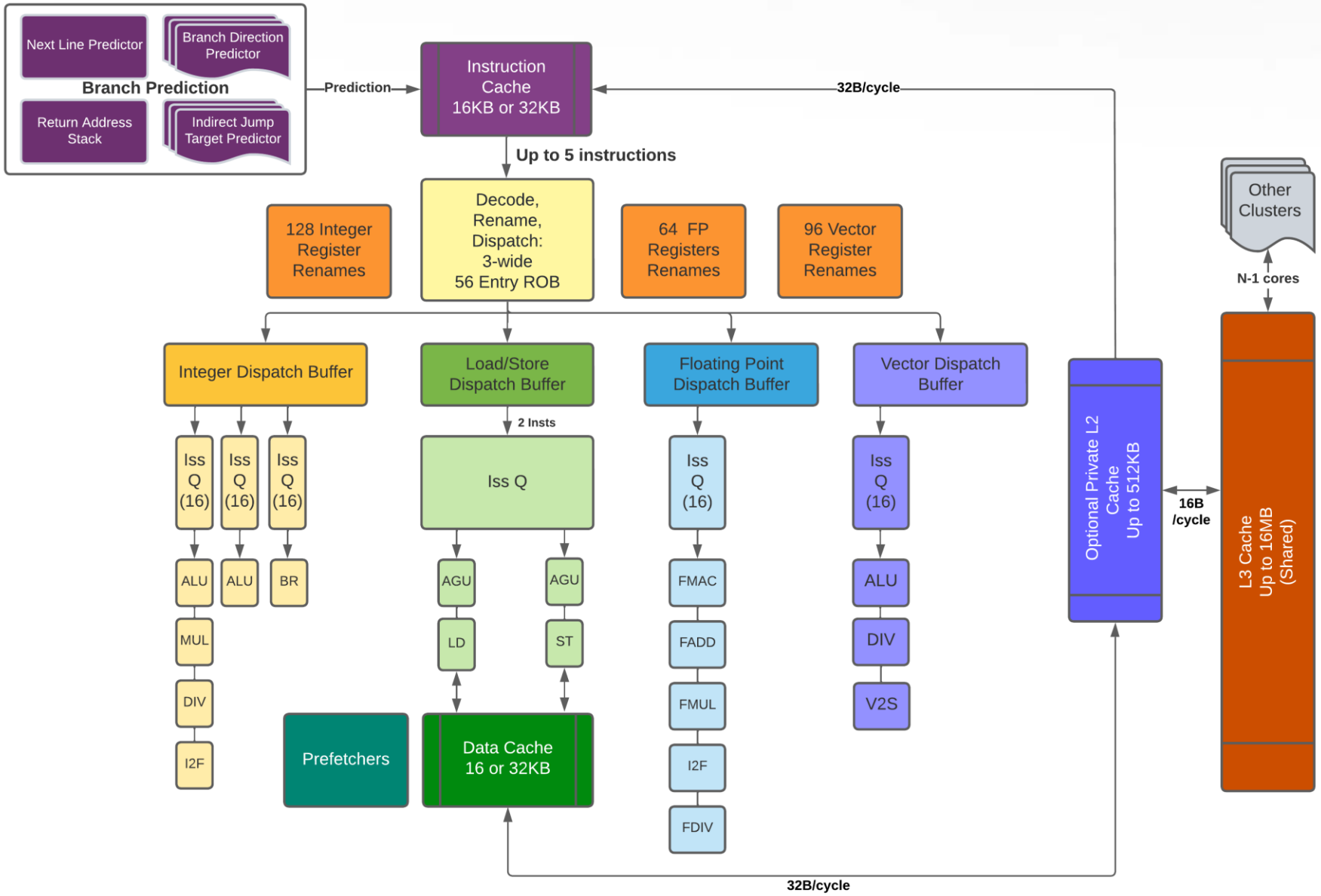


2.75x higher performance than Cortex-A55 in a similar area and power budget

WorldGuard advanced security capabilities

Supporting System IP Security, Virtualization, Trace and Debug

P470 detailed pipeline overview



SiFive Performance™ P670

Highest
Performance

Best-in-class performance

Balanced
PPA

Optimized performance within constrained area and power envelope

Vector
Extensions

Acceleration for media, crypto and data processing

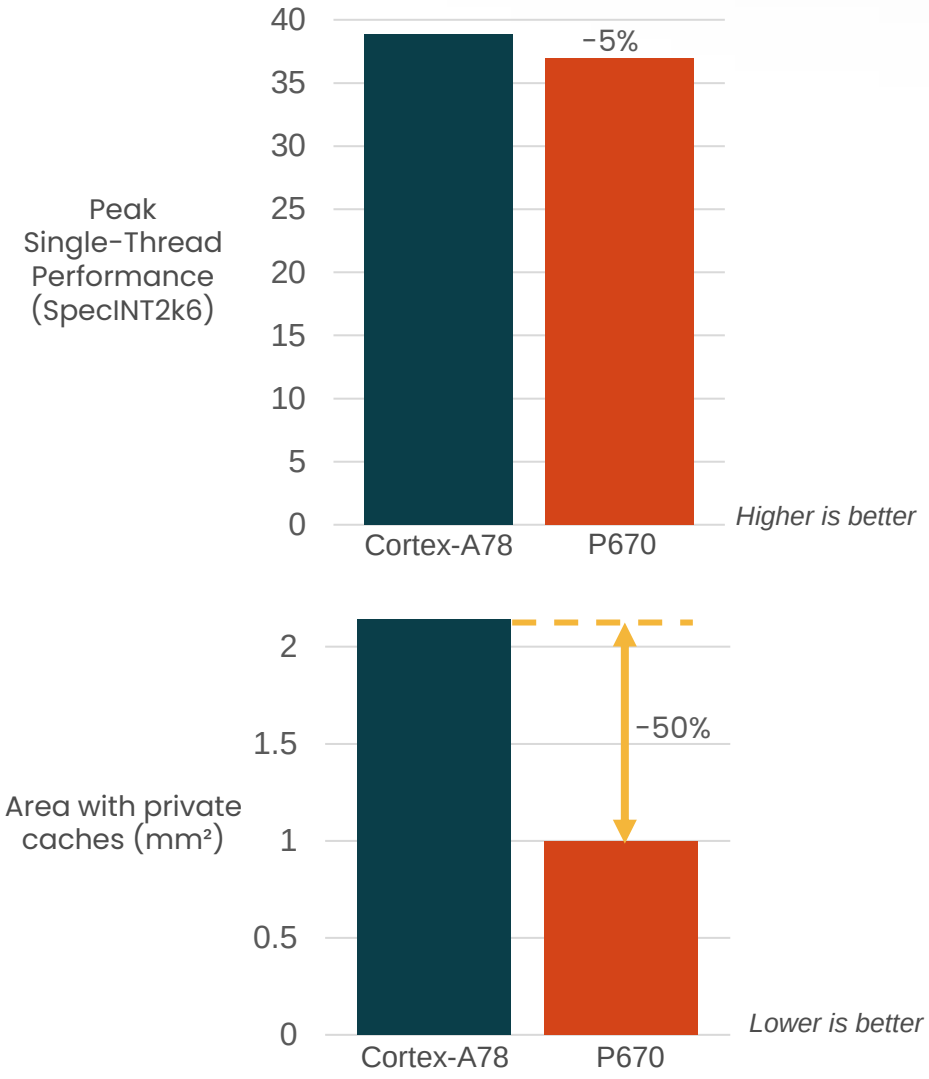
Feature
Rich

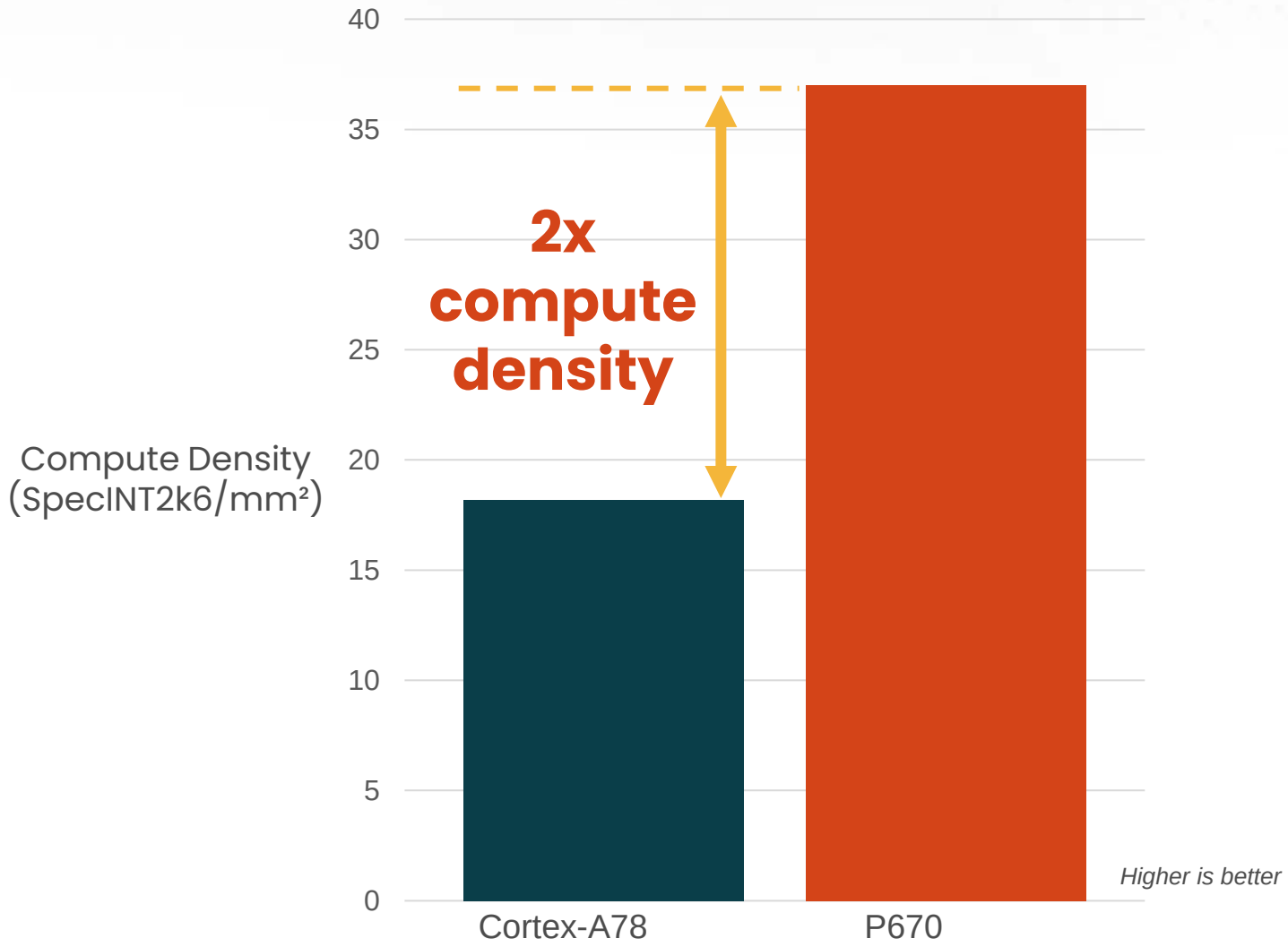
Virtualization, IOMMU, AIA, Debug & Trace, Security

RISC-V
Compliant

Compliant with RVA22 profile, with support for Vector and Vector Crypto extensions

P670 Performance & Efficiency





Cortex-A78: ***Performance*** *3GHz in 5nm with 32KB L1 I$ and D$, 512KB Private L2, 4MB L3 Anandtech.* ***Area****; 7nm TechInsights – "Die photos show Cortex-A78 shortfall", article*
SiFive P670: ***Performance*** *3.4GHz, 0.95V in 5nm with 32KB L1 I$ and D$, 256KB Private L2$, 4MB L3$.* ***Area****: measured in 7nm.*

SiFive P600-Series Application Processor

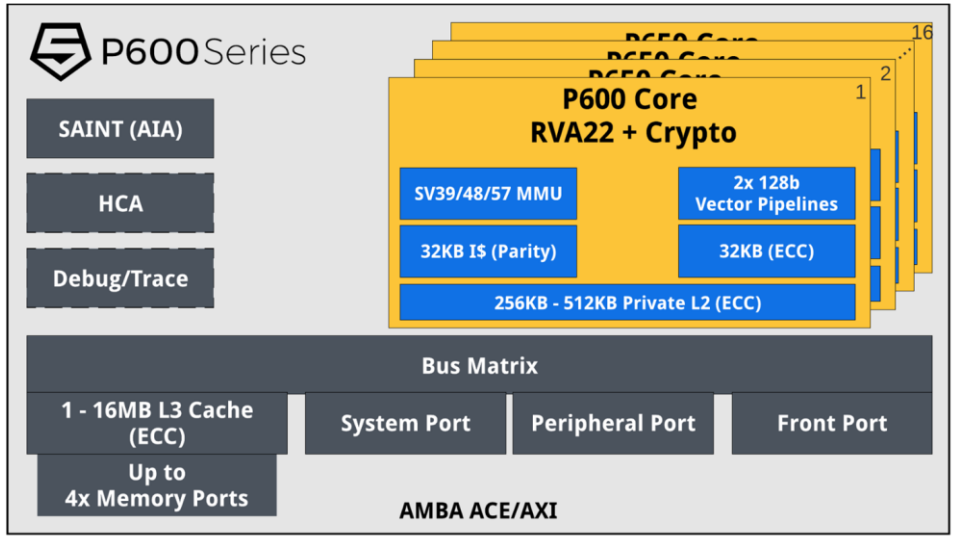
High-Performance Out-of-Order RISC-V Application Processor

The P600-Series is a quad-Issue, Out-of-Order processor, building on the highly successful P550 micro-architecture to reach even higher levels of performance. The P600-Series has class leading RISC-V capabilities such as Vector Processing, Virtualization, System Security, and higher core counts.

64-bit RISC-V ISA with Hypervisor Extension and MSI Interrupt Controller

Support for the RISC-V Vector Extension with 2x 128b Vector ALUs (P670 only)

Coherent multicore with up to 16 Cores in Core Complex and coherent interfaces

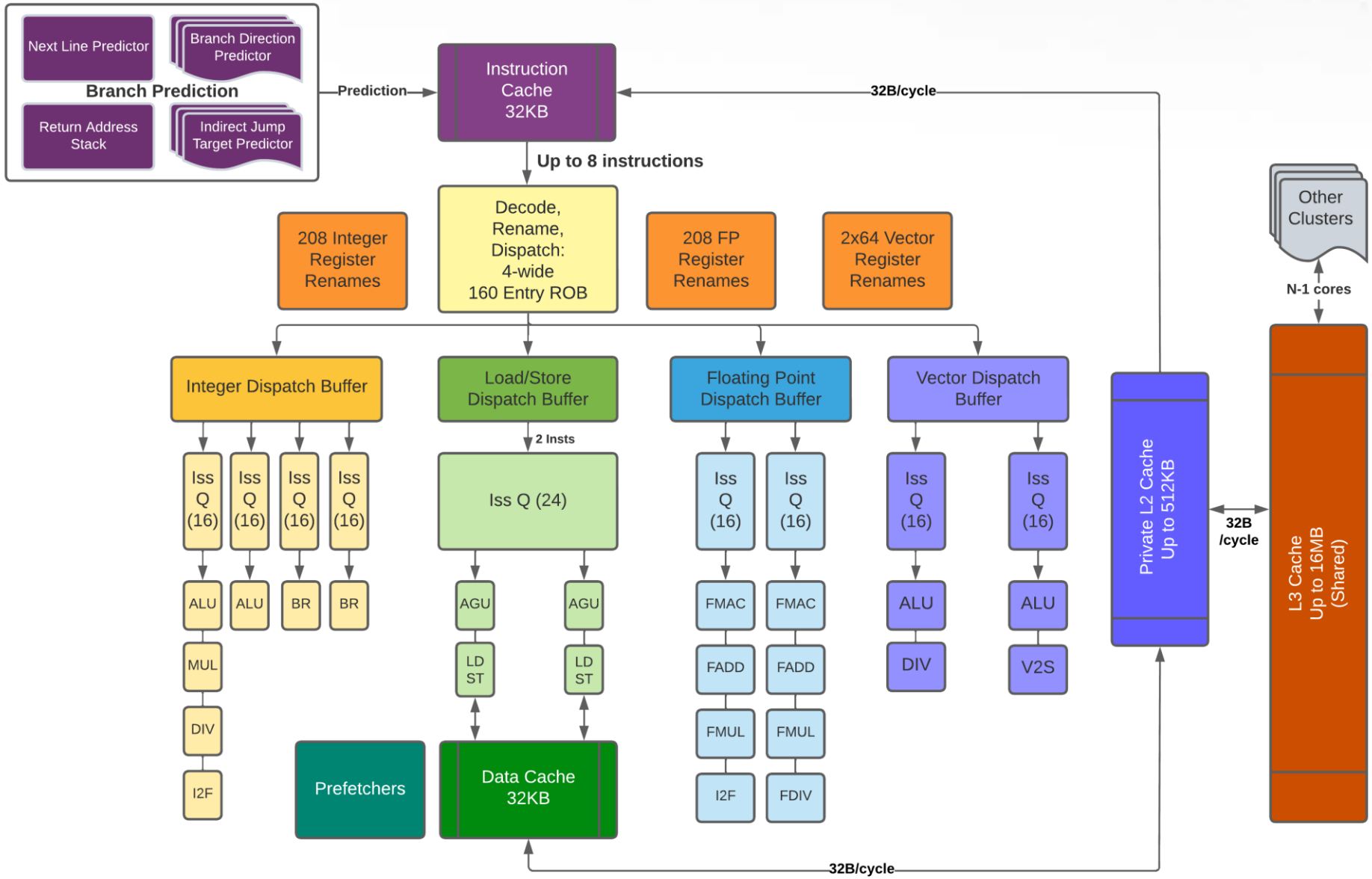


>12 SPECInt2k6/GHz with 50% less area than Cortex-A78

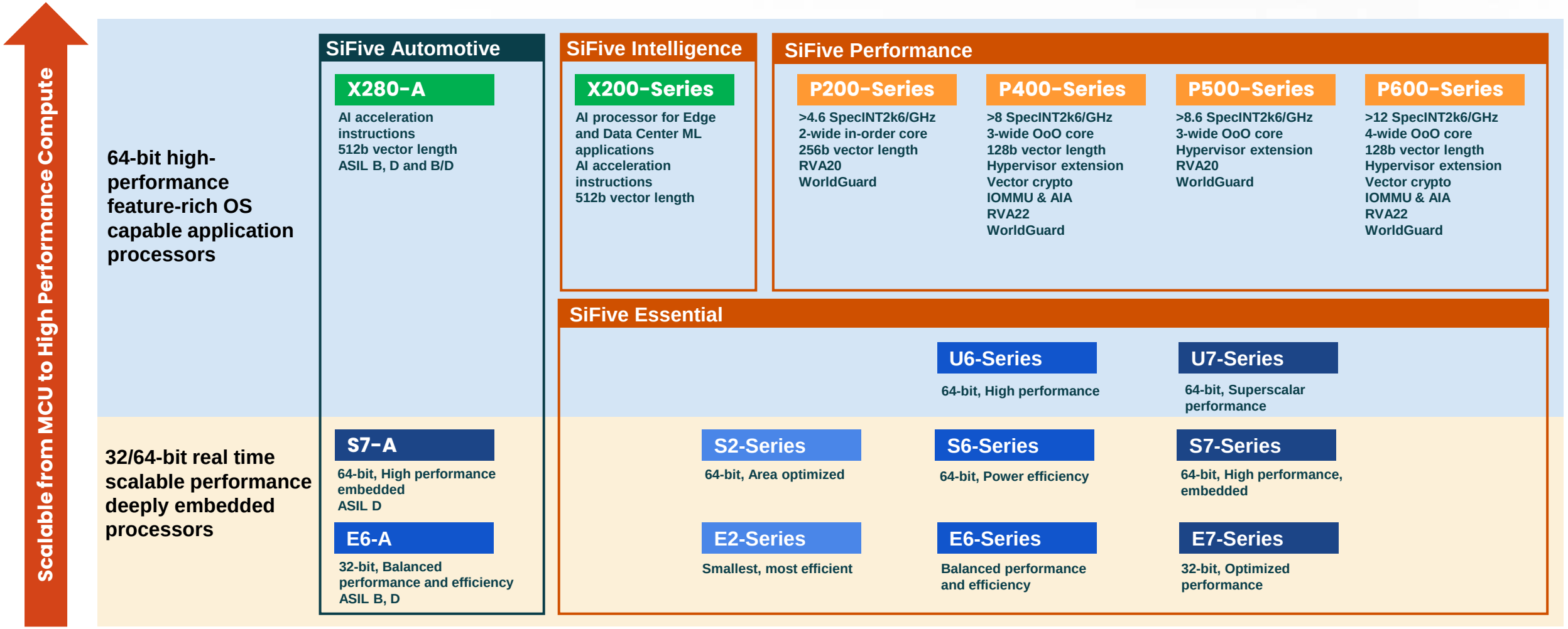
Sophisticated Trace and Debug capabilities

WorldGuard Security and RISC-V Vector Crypto

P670 detailed pipeline overview



SiFive broad IP portfolio



SiFive High Performance Solutions

Broadest RISC-V Application Processor Portfolio

High Performance Processors with Balanced & High Efficiency profiles

New Market Opportunities

High performance, high efficiency, better feature fit

First-to-market with RISC-V standards

RVA22 compliance for Android, System-level virtualization, Vector crypto extensions

SiFive continuous innovation

16-core support, cache optimization, ACE coherence, power management, WorldGuard

Upgrade to the SiFive Performance Family

Performance density leadership

High performance with optimized power efficiency

First with latest RISC-V features, standards, and technology

The P400-Series and P600-Series are available to Lead Partners in Q4 2022

SiFive

Thank you

SIFIVE.COM

Andes' Next-Generation Scalable RISC-V Application Processor Family

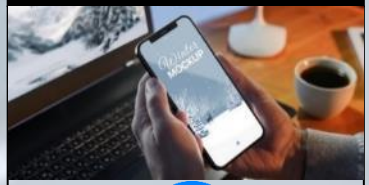
Charlie Su, Ph.D.
CTO and President, Andes
charlie@andestech.com

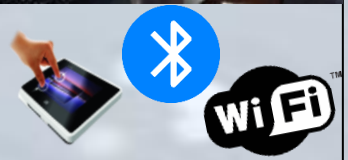
Linley Fall Processor Conference – Nov. 1-2, 2022

Andes RISC-V Adoptions From Edge to Cloud

■ Into the Space

Mobile





Performance, code size

MPU/MCU/AIoT

RENESAS
HPMicro
Kneron
Telink
Internet Company

Storage

PHISON

Performance, bandwidth, real-time

Above Cloud

Secure, control, compute, communicate, position

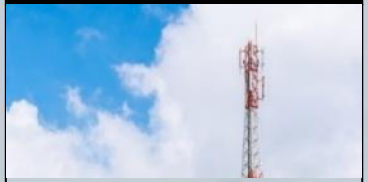
Datacenter AI

STREAM COMPUTING
SK telecom

Accelerate, accelerate, accelerate

Internet Company

5G Networks



EDGEQ
5G WITH AN EDGE

picocom
Empowering Wireless

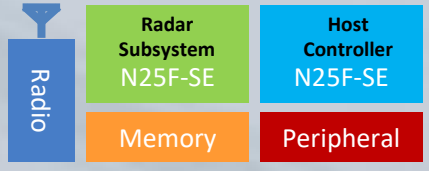
MEDIATEK
NOVATEK

On The Road Too

N25F-SE:

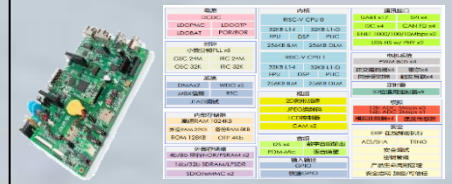
- Andes 1st Safety Enhanced core with ISO 26262 Full Compliance

In-Cabin Radar

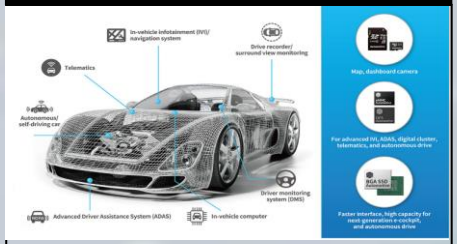


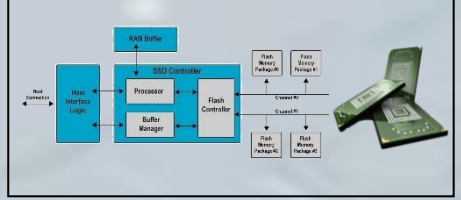
Auto TDDI

Auto MCU



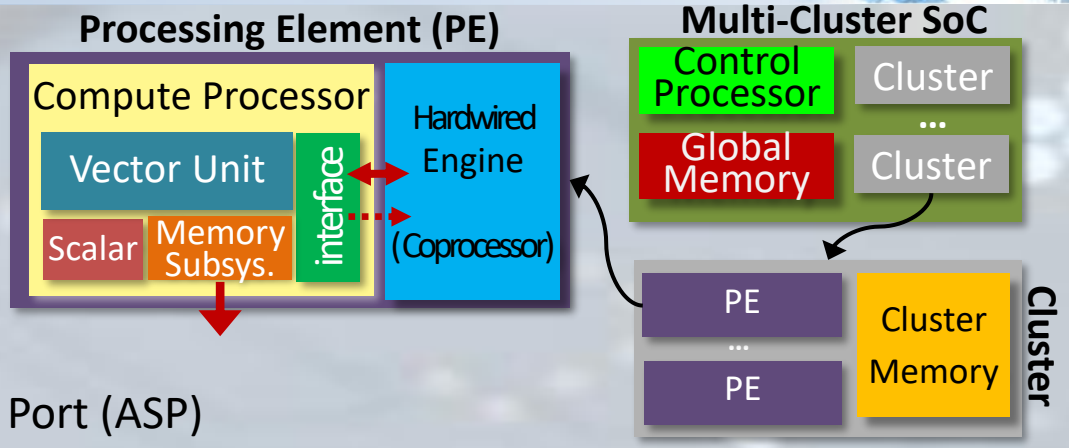
Auto Storage





Architecture for High-Speed Computing

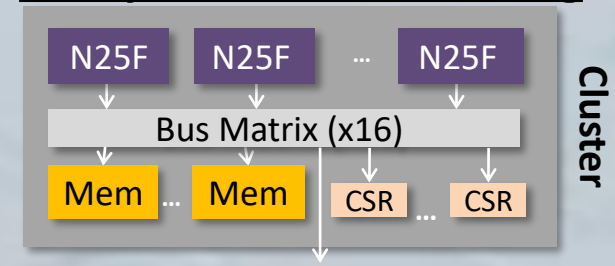
- **A scalable architecture**
- **Control processor: AX45MP 1c-8c**
  - OS, applications and cluster control
- **Compute processor: NX27V**
  - Powerful VPU
  - High-bandwidth memory subsystem
  - Extensibility: ACE & Andes Streaming Port (ASP)



VLEN,SIMD[1]: (bits)	512	256	128
Speedup[2] geomean[3]	78x	42x	22x
Speedup ratio	**3.59**	**1.92**	**1.00**
Area ratio (@7nm)	**2.40**	**1.43**	**1.00**

1: All data run on NX27V FPGA with 32KB I$, 512KB D$.
2: Compared with C scalar code compiled with high optimization
3: Geomean of F32 math functions/matmul, S8 CNN, and F16 MobileNet V1

Many-channel Processing



➔ **Separate decision for control and compute**

AndesCore® Lineup

- **N25F-SE:** The world 1st RISC-V core with **ISO 26262 Full Compliance**, not just *Ready*
- **NX27V:** The world 1st RISC-V **Vector** core
- **D25F:** The world 1st RISC-V **DSP-capable** core

45 Series 8-stage superscalar		N45, NX45	D45	A45(MP), AX45(MP)	A53/55, R52/82, M7
27 Series 5-stage MemBoost			**NX27V**	A27(L2) AX27(L2)	A5/7/35
25 Series 5-stage fast & compact	**N25F-SE**	N25F, NX25F	**D25F**	A25(MP) AX25(MP)	A5/7/35, R4/5, M4/33
Entry Series		N22			M0/0+/3/33/4
Categories	***FUSA***	*Embedded Control*	*DSP/Vector*	*Linux AP*	*References*

AndesCore® On the Horizon

AX60 Series 13-stage OOO MP			**AX65**		A72~A76; X1/V1
Categories	***FUSA***	*Power-efficient*	*Mid-range*	*Extended*	
45 Series 8-stage superscalar		N45, NX45	D45 NX45V[1]	A45(MP), AX45(MP) AX45MPV	A53/55, R52/82, M7
27 Series 5-stage MemBoost			NX27V	A27(L2) AX27(L2)	A5/7/35
25 Series 5-stage fast & compact	N25F-SE	N25F, NX25F	D25F	A25(MP) AX25(MP)	A5/7/35, R4/5, M4/33
Entry Series		N22	D23		M0/0+/3/33/4
Categories	***FUSA***	*Embedded Control*	*DSP/Vector*	*Linux AP*	*References*

Note 1: AX45MPV configured as one core

- AX60 Series: scale up and scale out

Note: roadmap subject to change without notice

Technologies Built-up

- **NX27V:**
  - In-order scalar unit
  - out-of-order vector unit, up to 4 VLEN results/cycle
- **AX27 and AX45:**
  - Support powerful **MemBoost** memory subsystem
    - Non-blocking caches with up to 16 outstanding requests
    - Sequential instruction prefetch and multiple stride-based data prefetch
    - Write-around to reduce latencies and avoid cache pollution

AX60-Series OOO Application Processors

Architecture:

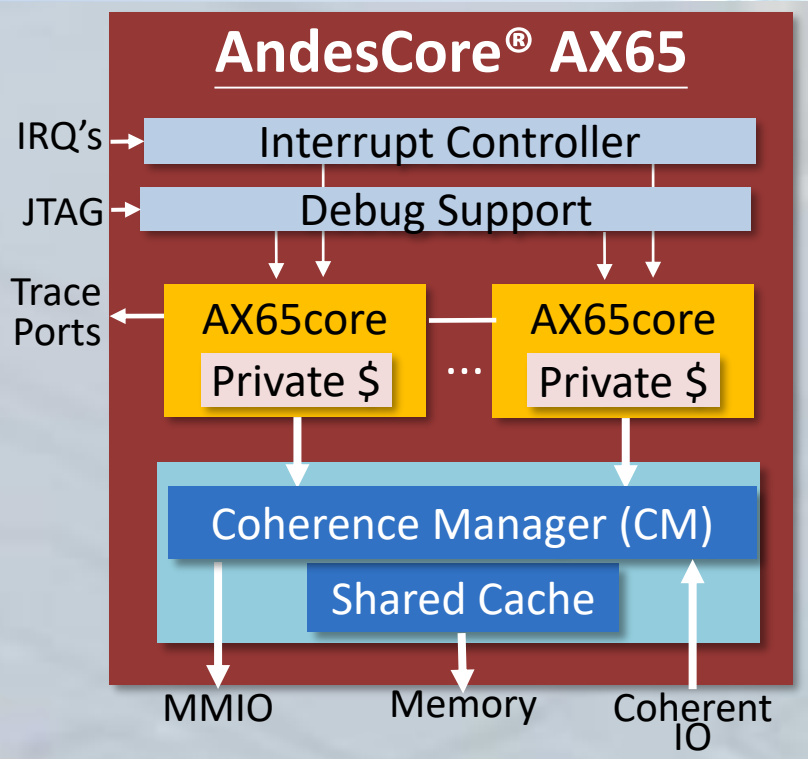
- RV64 GCBK
- Vector Extension
- SV39/48 (and other VM extensions)
- Cache Management Operations
- ePMP/PMA
- Andes Custom Extension™ (ACE)
- Hypervisor
- Interrupt controller: PLIC and AIA
- Debug and Trace

Microarchitecture

- Out-of-order superscalar
- Advanced memory subsystem
- Multicore cluster
- CHI-based scale-out:
  - Single-core as the building block
  - Cluster as the building block
- Error protection
- Power management (retention and PowerBrake)

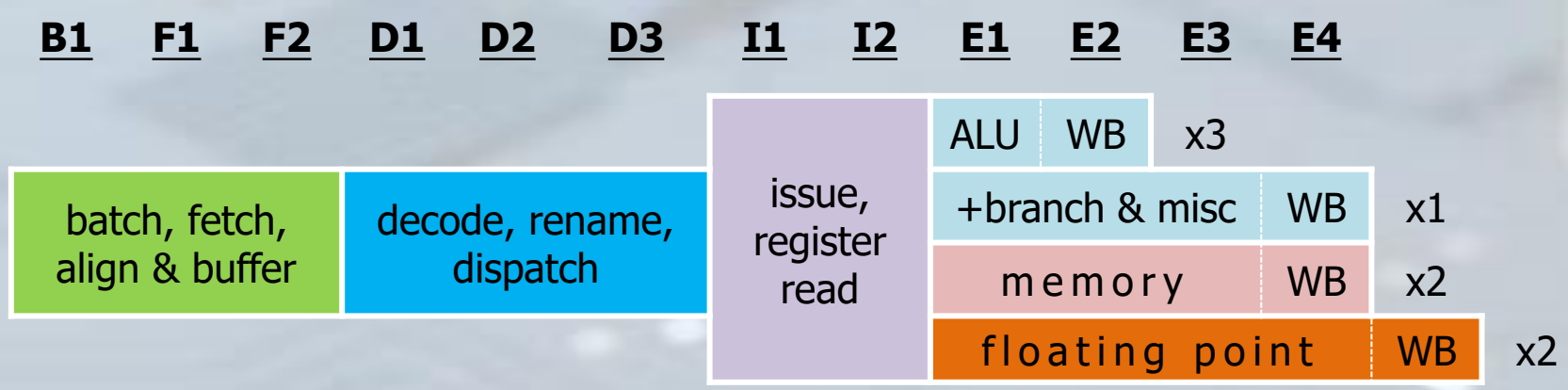
AX65: 1st Member of AX60-Series

- 4-way 13-stage superscalar
- Multicore cluster: 1~8 cores
- Private caches:
  - 64KB, multi-banked
  - Alias handling in HW
- Shared cache: up to 8MB
- CPU and CM: async clocks



- Bus Ports (Memory, MMIO, Coherent IO): 256 bits

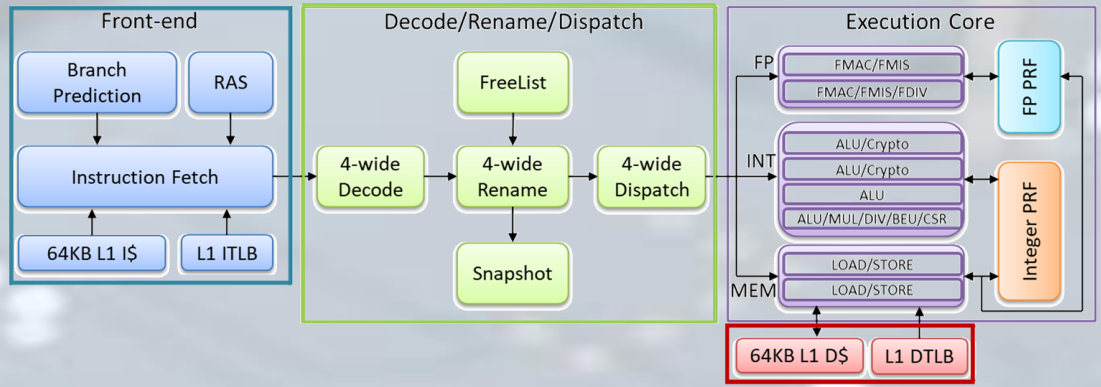
AX65 Microarchitecture: Overview



- **2 double-word fetches per cycle in 1 or 2 cache lines**
- **4 decode/rename/dispatch/graduate**
- **Execution pipes:**
  - 4 integer ALUs: 2 with scalar crypto, 1 with branch
  - 2 full load/store units
  - 2 FPUs: one full, one without divide/sqrt

AX65 Microarchitecture: Overview

- **Branch/return prediction:**
  - TAGE with loop prediction
  - 2-level BTB
  - 9-cycle misprediction penalty
  - RAS (Return Address Stack)
- **ROB/Freelist:** 128/128
- **Physical integer/FP registers (XPR/FPR):** 160/160
- **Split 2-level TLB:** L1 up to 32 entries, L2 up to 1024 entries
  - Combining 2 consecutive L1 entries
  - Aggressive concurrent table walkers
- **Load/store units:** up to 64 outstanding instructions
- **Total requests/core:** >20 outstanding requests

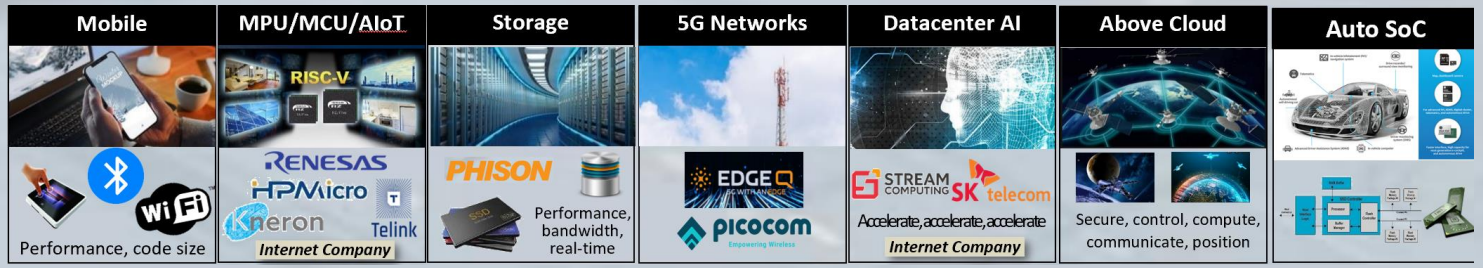


Preliminary Performance Results

AndesCore	AX27L2	AX45MP (over AX27L2)	AX65 (over AX45MP)
Micro-architecture	scalar in-order	dual-issue in-order	quad-issue OOO
Freq. (7nm)	~2 GHz	>2 GHz	>2.5 GHz
Coremark/MHz	3.55	5.63 (+59%)	>9.0 (+60%)
EEMBC FPMark/MHz	27.0	35.2 (+30%)	62.2 (+77%)
Mem Bandwidth/MHz	1.0x	1.40x (+40%)	2.76x (+97%)
Specint2k6/GHz	2.82	3.46 (+23%)	> 7 (>2x, target)

Concluding Remarks

- **AX60 series: most balanced PPA at various performance points**
  - AX65 is the mid-range of AX60 series
- **AX65 raises the control-plane performance for many applications**



  - Data plane: more NX27V/N25F, or upgrade to NX45V/N45

Thank You !

andestech.com

ANDES
TECHNOLOGY

DESIGNING POWER EFFICIENT SCALABLE RAY TRACING GPUS

KRISTOF BEETS – VP OF TECHNOLOGY INSIGHTS
KRISTOF.BEETS@IMGTEC.COM

LINLEY FALL PROCESSOR CONFERENCE – 2ND OF NOVEMBER 2022

Overview

Why include Hardware Ray Tracing ?

- Ray Tracing Benefits and Value
- Market Overview

Ray Tracing Architectures

- The Coherency Problem
- Hardware Solutions, Efficiency and Scalability

Ray Tracing Examples

- Visual Impact of Global Illumination with Ray Tracing
- Using Fragment Shading Rate to reduce the cost
- Performance Comparison

Why Include Hardware Ray Tracing ?

Ray Tracing Benefits and Value

Visual Quality…

Reflections, Shadows, Lighting – its how reality works

Simple…

Cast Ray versus 100s if not 1000s lines of shader/kernel code

No more complex approximations of ray tracing effects

Always been in use…

Baked Lightmaps since Quake (‘96), but now dynamic in real-time

Smaller App Sizes…

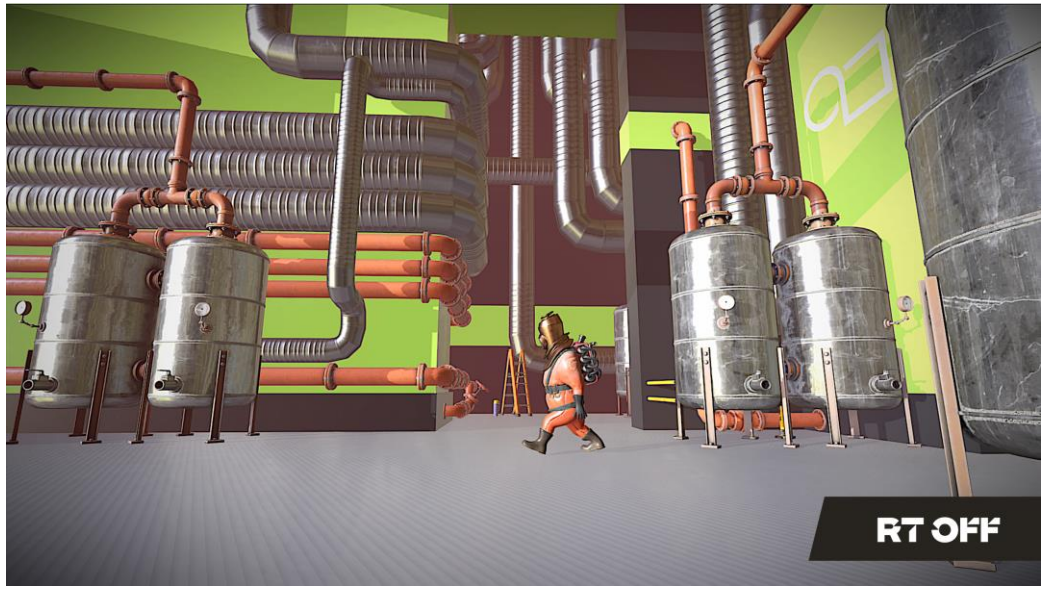
No GBytes of prebaked textures instead it’s all real-time and dynamic

More Artistic Freedom and Speed…

No more algorithm precision issues and artefact/bottleneck avoidance

Power, Processing and Bandwidth Efficiency…

Offload processing to specialised hardware versus general compute



Why Include Hardware Ray Tracing ?

Ecosystem Support

Enabled by Standardisation

- Enabled in DirectX
- Enabled in Vulkan
- Already in Apple Metal API also

Embraced by Console and PC Market

- Sony Playstation
- Microsoft XBOX
- All PC Market Vendors: Nvidia, AMD and Intel

Ramping up in Mobile Market already

- Not just by Imagination – OEMs driving innovation:
  - https://www.oppo.com/en/newsroom/press/oppo-unveils-ray-tracing-3d-wallpaper-at-gdc-2022/
  - https://consumer.huawei.com/uk/community/details/Huawei-Pheonix-will-bring-ray-tracing-tech-to-smartphone-gaming/topicId_41723/
  - https://news.samsung.com/global/samsung-introduces-game-changing-exynos-2200-processor-with-xclipse-gpu-powered-by-amd-rdna-2-architecture

"We are very pleased to see Imagination, the industry leader in ray tracing technology, release hardware ray tracing IP. We will work closely to explore the application of this technology in games."

Tencent Games

"At Carbonated, veterans from Zynga, Electronic Arts and Blizzard, we're excited about Imagination Technologies leading the future of mobile GPU and ray-tracing technologies."

Carbonated Inc.

"Working with companies such as Imagination helps drive the delivery of next-generation graphics capabilities to developers to shape the future of mobile gaming. Mobile GPU solutions often compromise performance in favour of maintaining power efficiency. However, our collaboration with Imagination has enabled us to navigate these constraints efficiently while also unlocking innovative hardware-accelerated ray tracing on the open-source Open 3D Engine. We look forward to continuing our great work together."

Open 3D Foundation

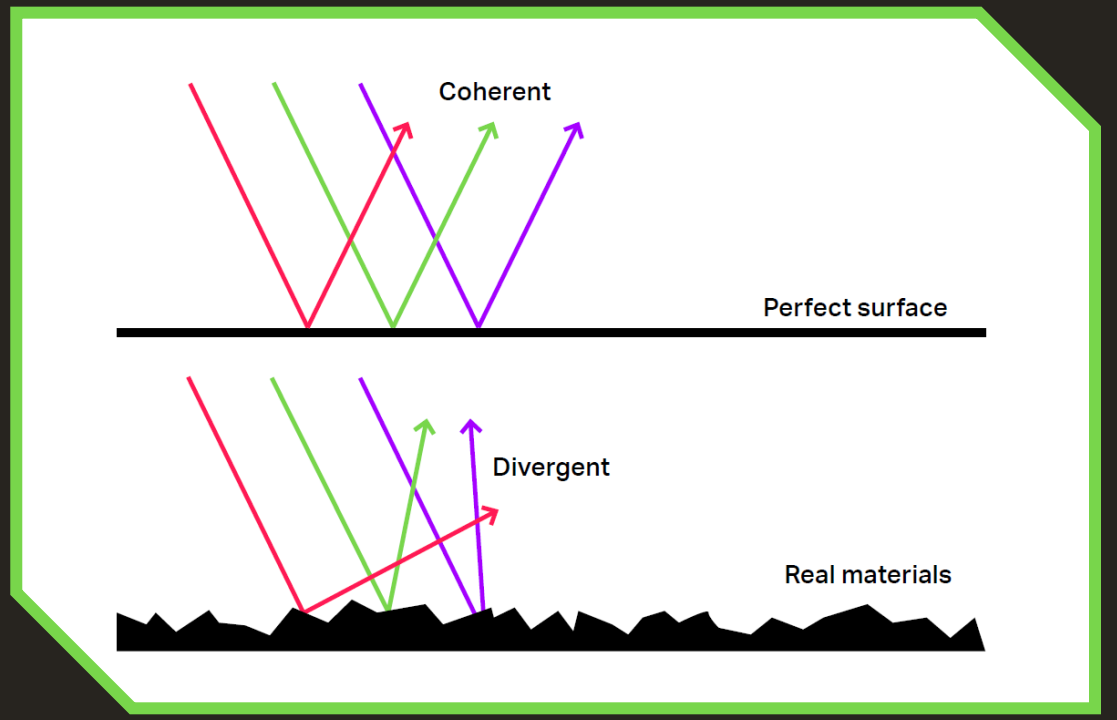
"Having developed chart-topping gaming titles for mobile across the years, we've established a keen understanding of what creates unique experiences for our players. A key mission when designing our games is to offer players an immersive and fun visual experience. We have been working with Imagination to give our developers more tools to create unique graphical gaming experiences for users across the world.

King Studios

https://www.imaginationtech.com/news/

THE RAY COHERENCY PROBLEM

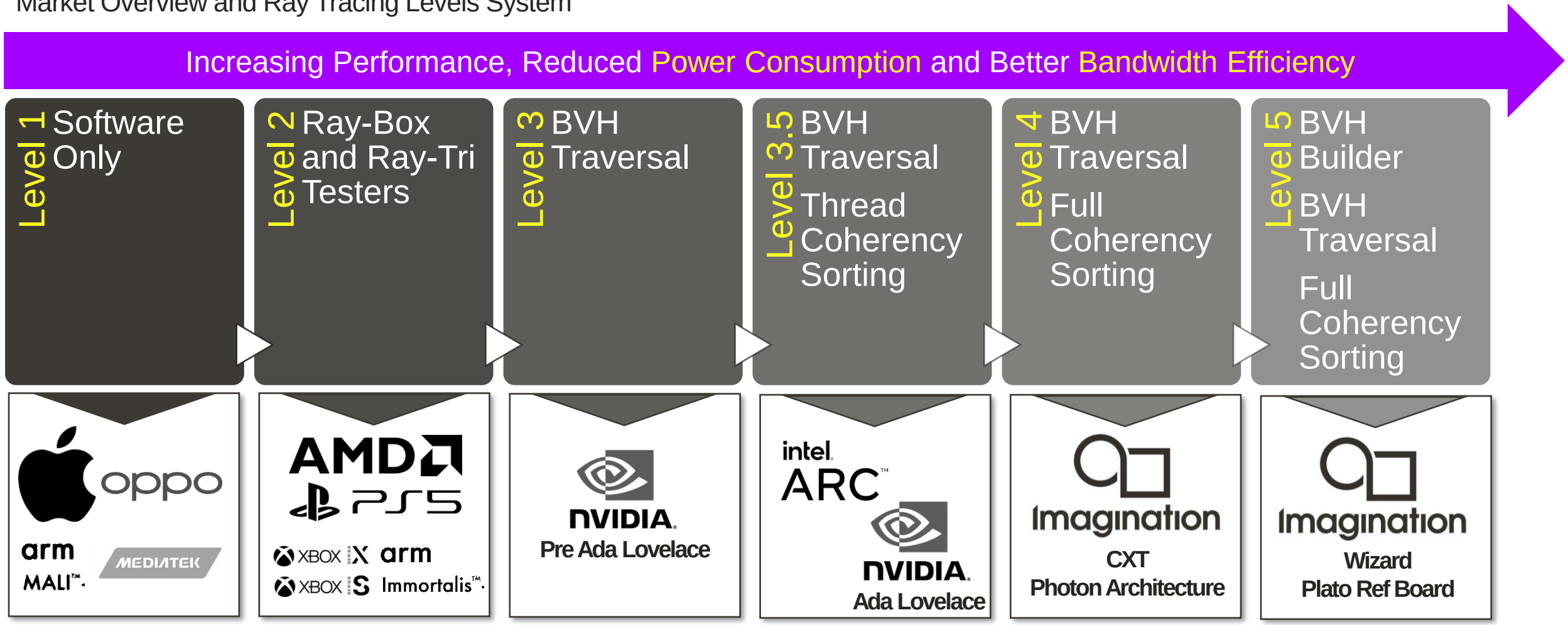
- Rays mimic behaviour of light in real world
- Rays bounce and traverse space in highly divergent way
- Mismatches with how traditional GPU is designed and optimised e.g. not SIMD – means diverging data access (bandwidth) and data processing (parallelism)
- Unique Imagination Solution is a Coherency Sorting Engine



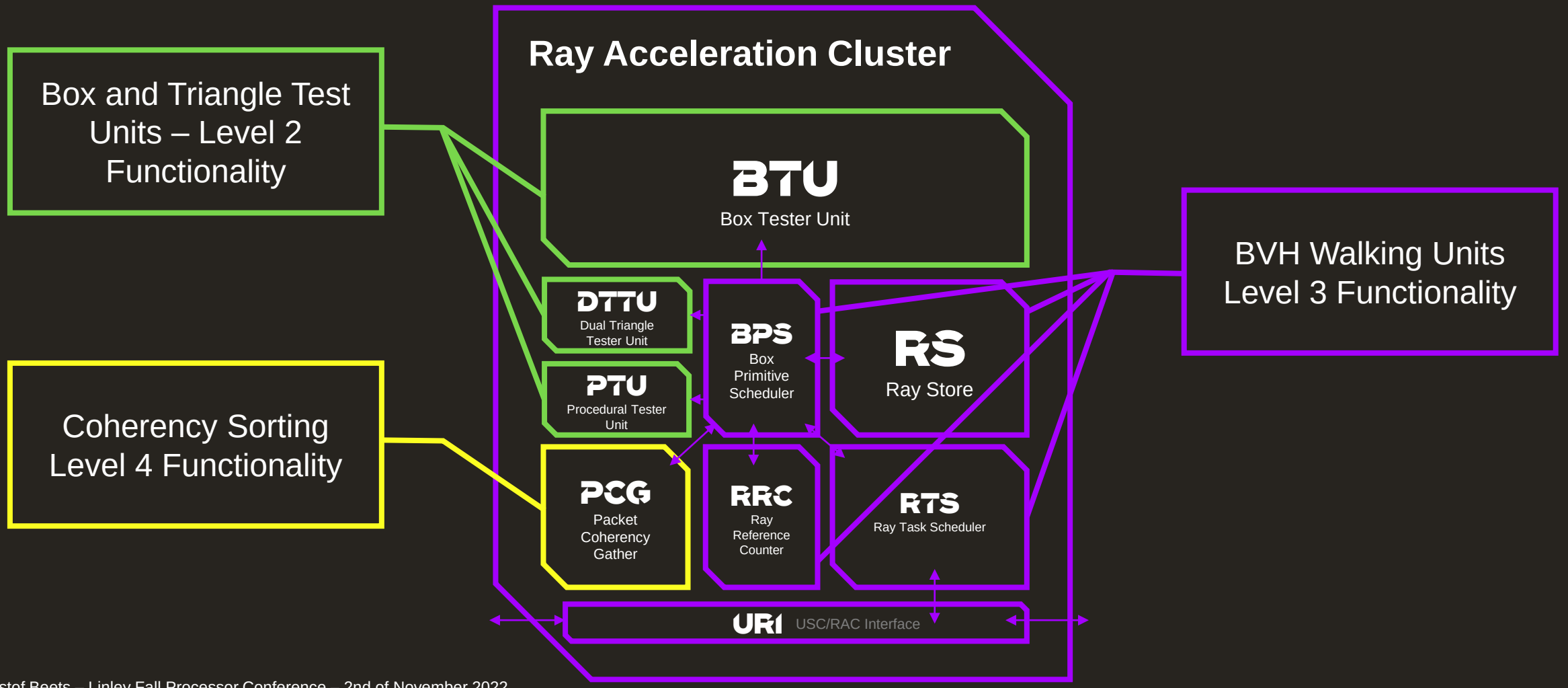
More details:
https://www.imaginationtech.com/resources/shining-a-light-on-ray-tracing/

Why Include Hardware Ray Tracing ?

Market Overview and Ray Tracing Levels System



PHOTON RAC

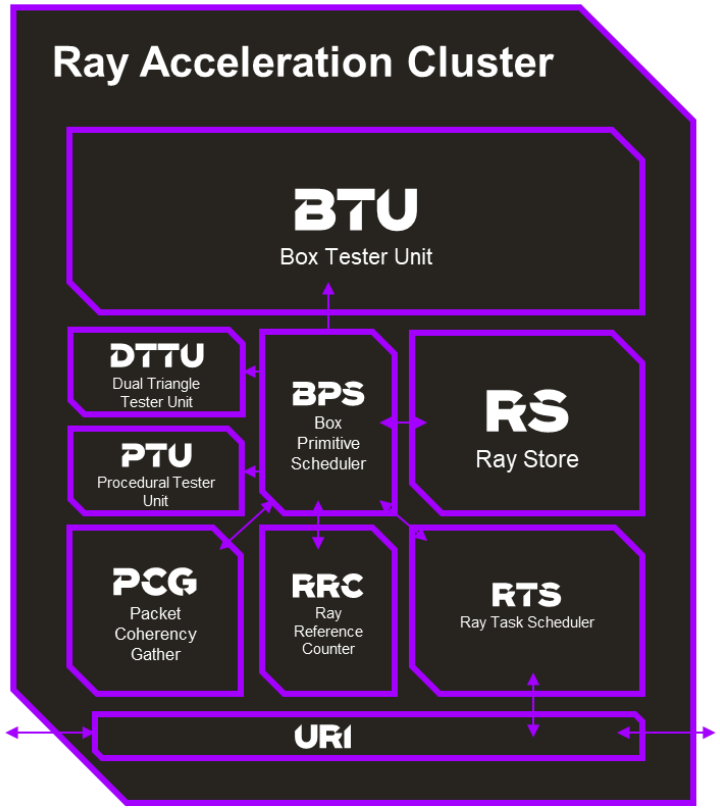

Box and Triangle Test Units – Level 2 Functionality
Coherency Sorting Level 4 Functionality
Ray Acceleration Cluster
BTU
Box Tester Unit
DTTU
Dual Triangle Tester Unit
PTU
Procedural Tester Unit
BPS
Box Primitive Scheduler
RS
Ray Store
PCG
Packet Coherency Gather
RRC
Ray Reference Counter
RTS
Ray Task Scheduler
URI
USC/RAC Interface
BVH Walking Units Level 3 Functionality

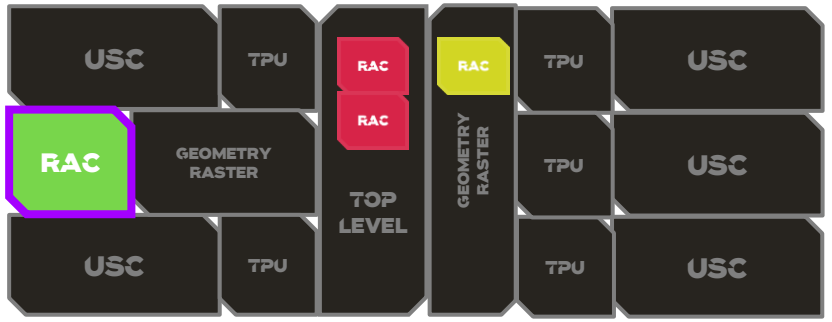
Imagination's "Photon" Ray Tracing Architecture Benefits

Stand-Alone "RAC" Solution

Ray Acceleration Cluster ("RAC")

- Self Contained Ray Tracing Unit – fully VK Ray Tracing compliant
- Includes all RTLS 4 Functionality
- Standard RAC enables:
  - 16 Box-Ray Tests/Clock
  - 2 Tri-Ray Test/Clock
  - Up to 4000 Rays in flight
- RAC Variants possible
  - Full, but also ½ and ¼ rate RACs for reduced area cost
  - Multiple RACs possible for higher performance e.g. desktop/cloud
- Flexible Integration
  - RAC shared between 2 or more USC/ALU engines
  - RAC can be located in different places inside the GPU and shared by more USC/ALU engines e.g. the Shared Logic Level or even the Top level – as shown ►



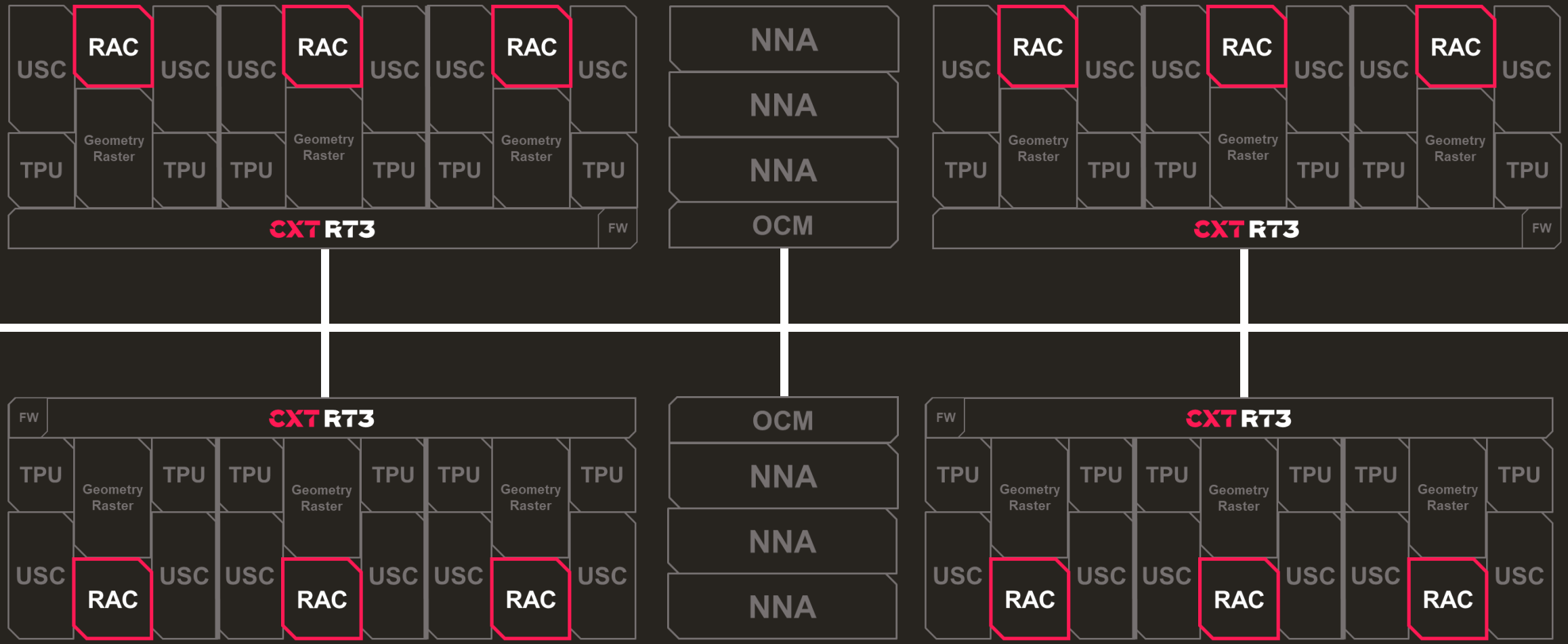


MULTI-CORE SCALABLE

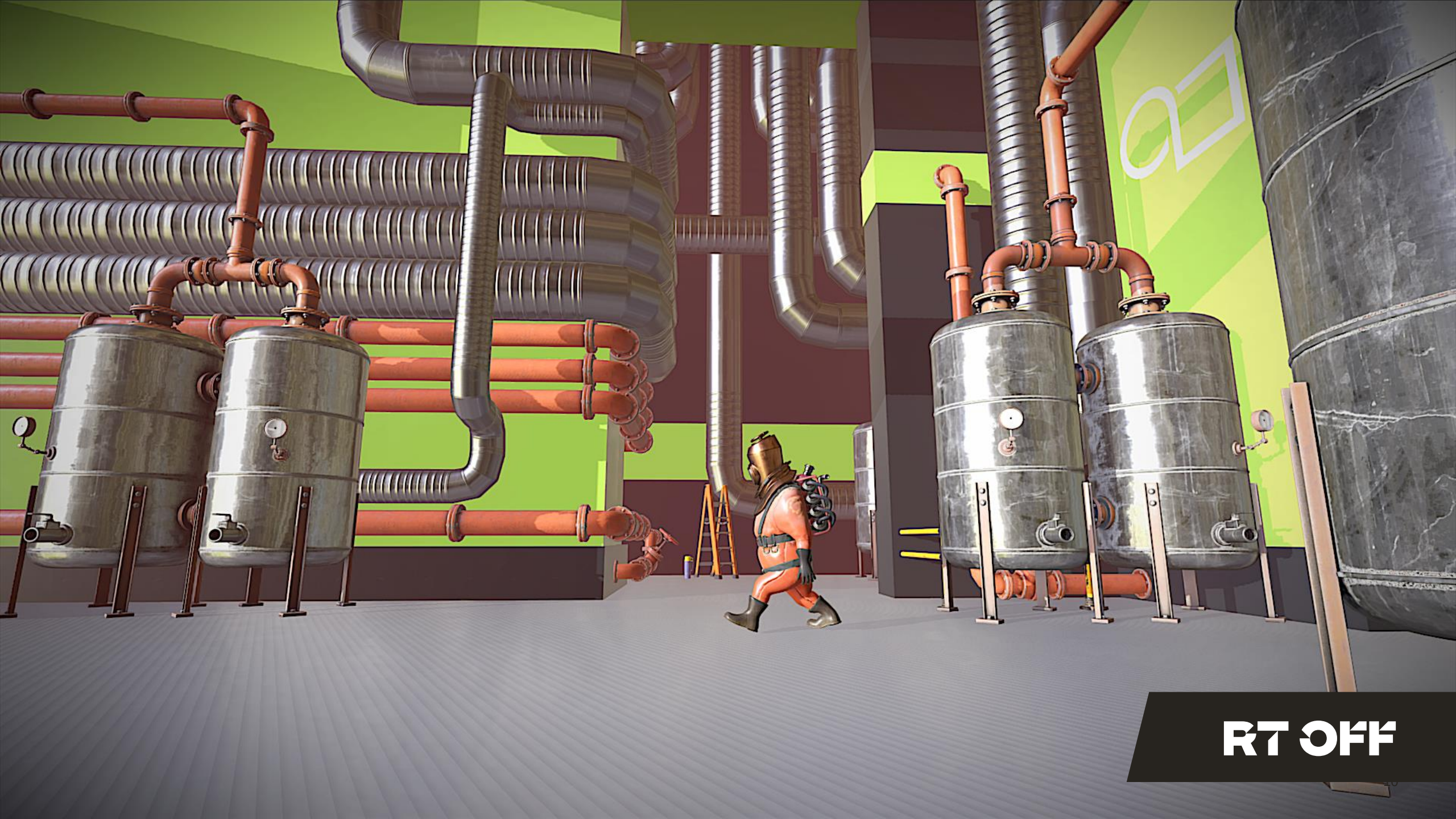
For Beyond Mobile Performance

Imagination

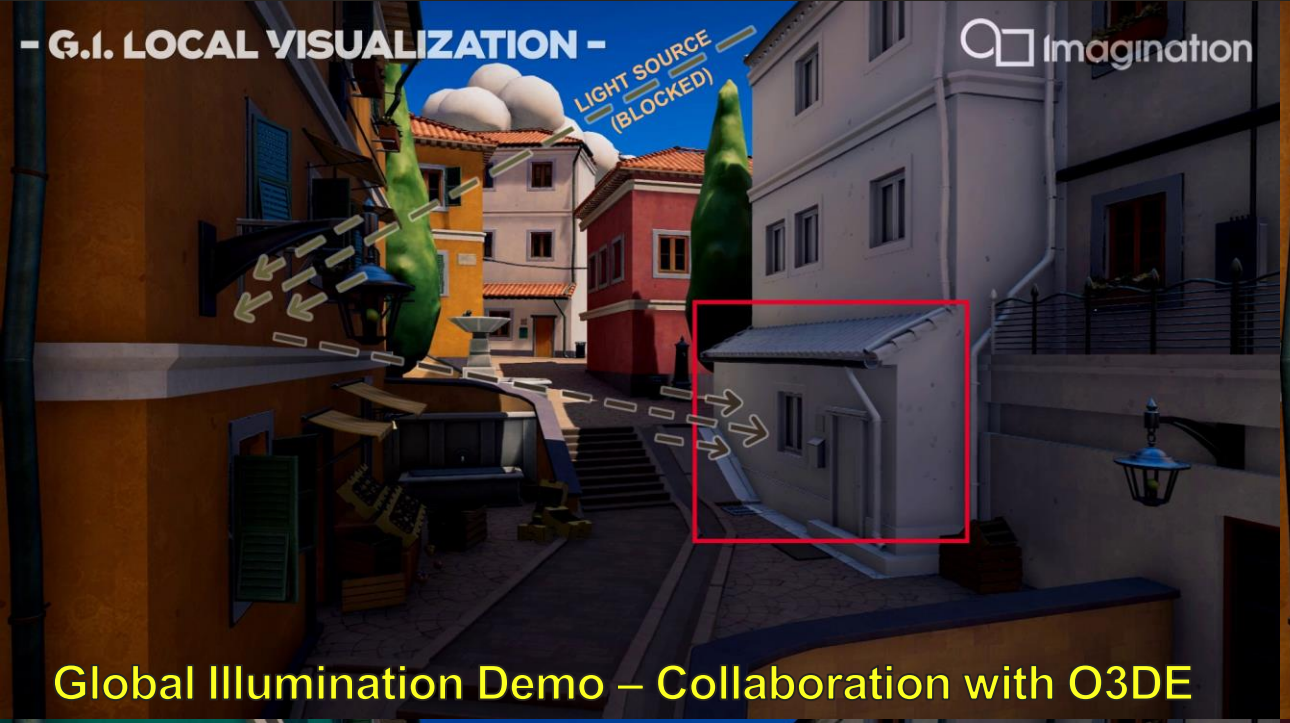
Additional NN IP for Super Res / Denoise

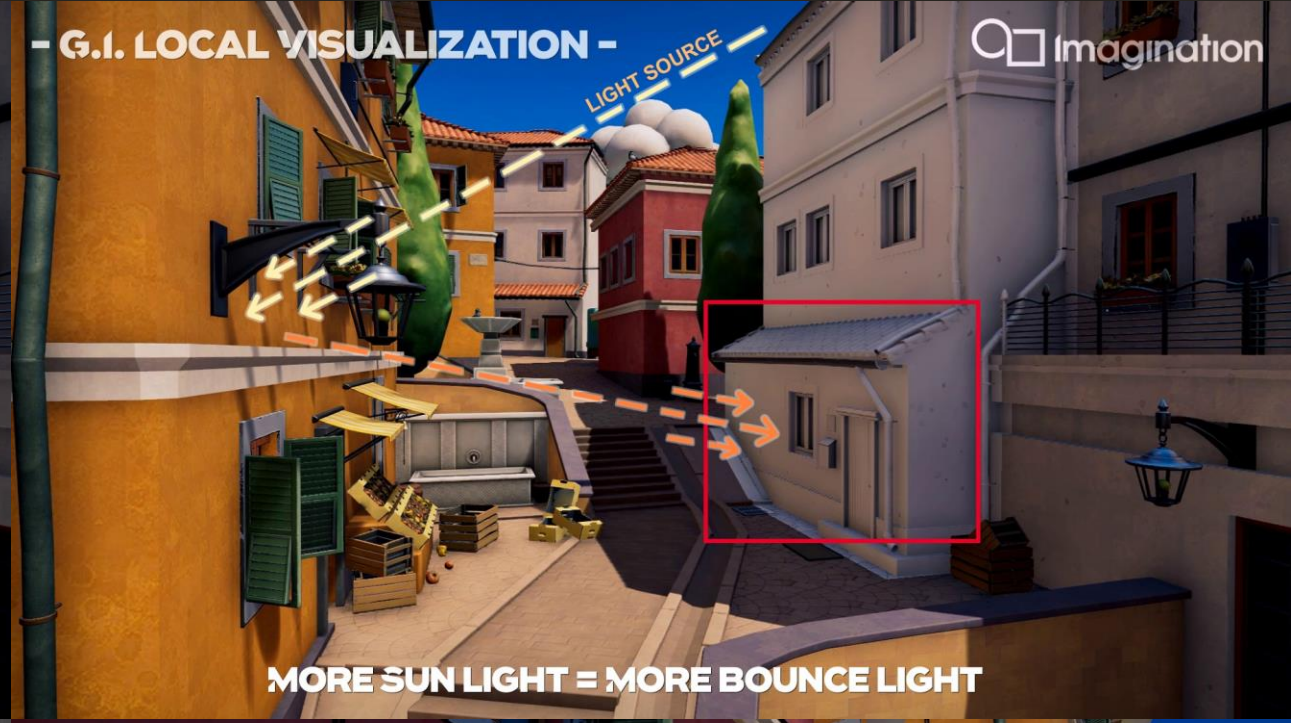


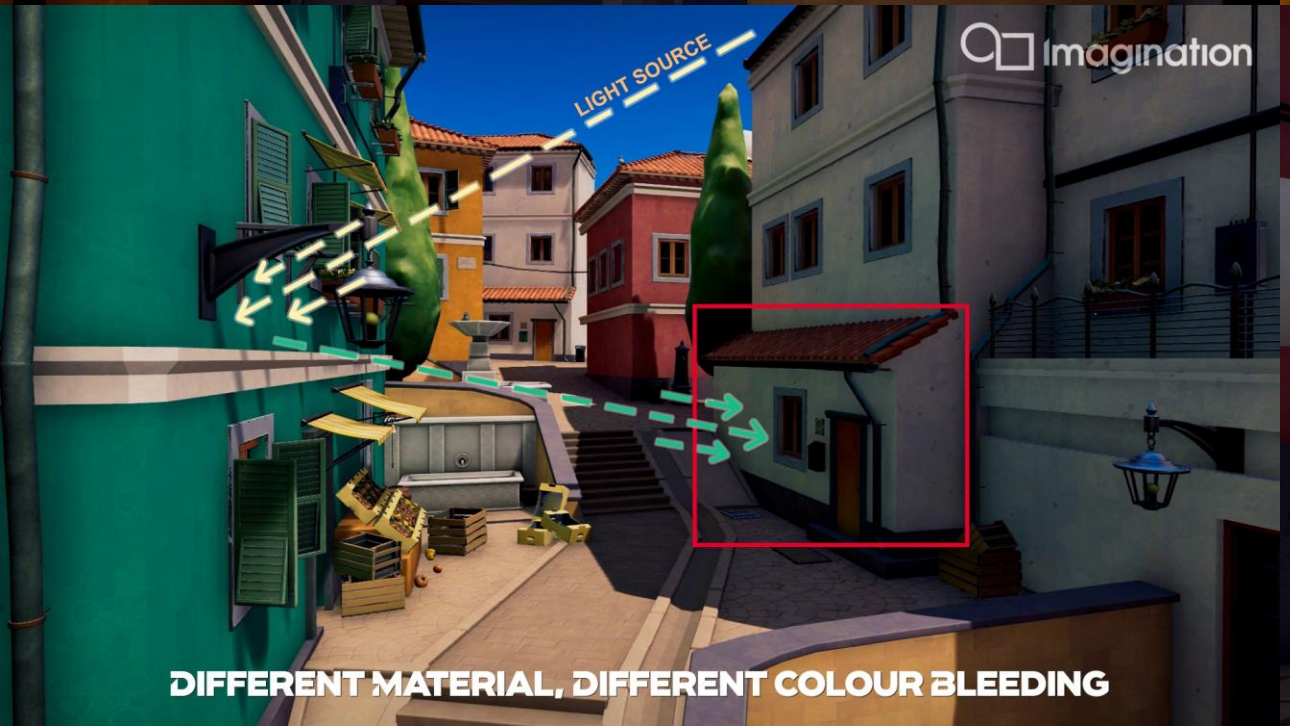
Up to 9TFLOPS FP32 Up to 2.5x ray tracing power efficiency of today's solutions Up to 7.8GRay/s


RT OFF


RT+GI ON


- G.I. LOCAL VISUALIZATION -
Imagination
LIGHT SOURCE
(BLOCKED)
Global Illumination Demo – Collaboration with O3DE


- G.I. LOCAL VISUALIZATION -
Imagination
LIGHT SOURCE
MORE SUN LIGHT = MORE BOUNCE LIGHT


Imagination
LIGHT SOURCE
DIFFERENT MATERIAL, DIFFERENT COLOUR BLEEDING


- G.I. PROBES VISUALIZATION -
Imagination
THE G.I. VOLUME IS FILLED WITH A GRID OF PROBES WHICH STORE LIGHTING INFORMATION
AT EACH POINT OF INTERSECTION BETWEEN RT RAYS AND THE SURROUNDING GEOMETRY.

Fragment Shading Rate – Balance Quality versus Performance, Bandwidth, Power Cost

Fragment Shading Rate

Enables shader execution based on “zones”

Normal/Full detail applies to 1x1 “zone”

Cost can be reduced by applying execution to zones

- 1x2, 2x1, 2x2, 4x1, 1x4, 4x2, 2x4, 4x4 pixel zones

Improves performance

Reduces Bandwidth and Power Consumption

But can reduce visual quality

Shader executed at a lower rate than per pixel

Compatible with ray tracing, send rays per zone

Fragment Shading Rate is controlled by developer

Can be set per draw call, per primitive

Or using an image map

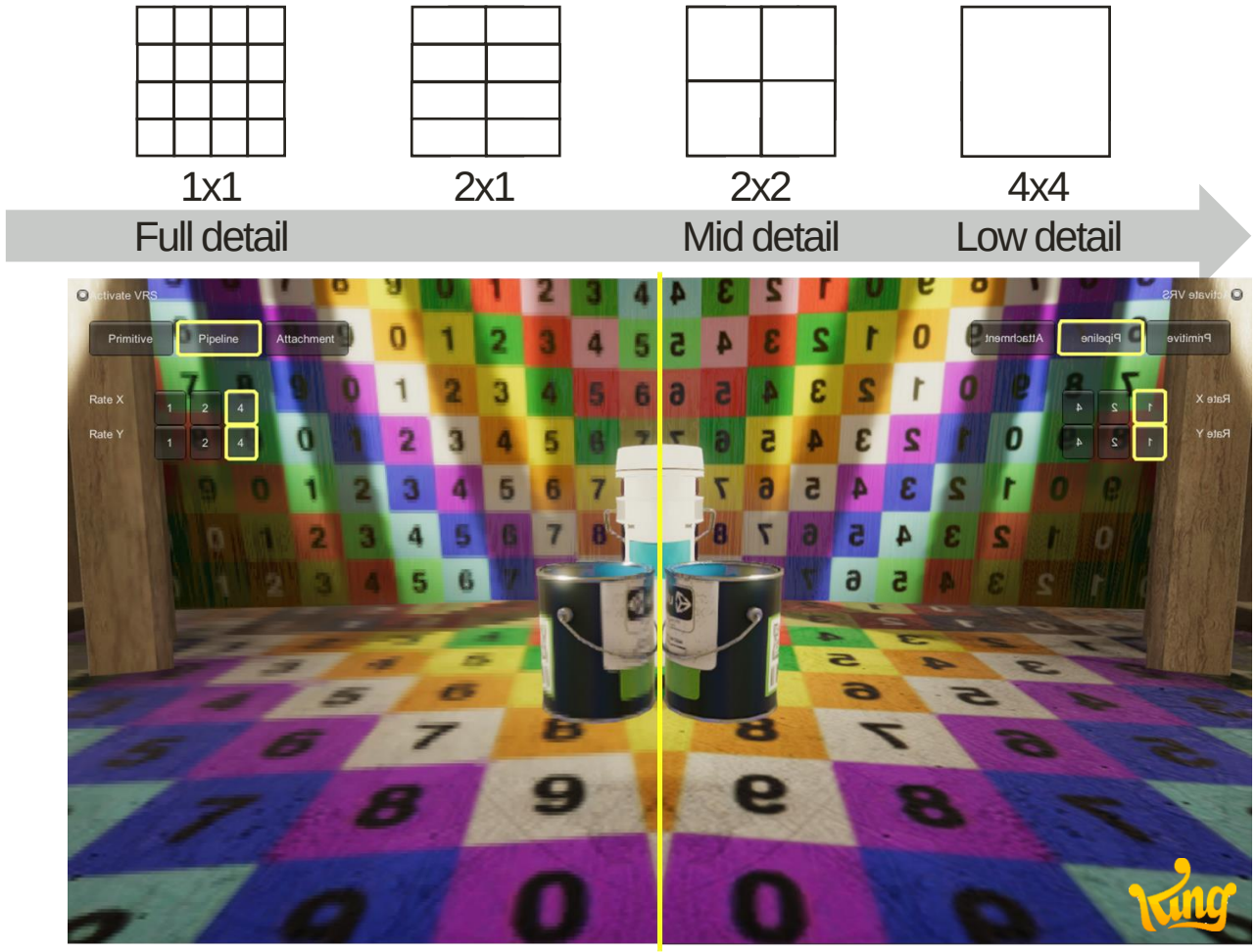
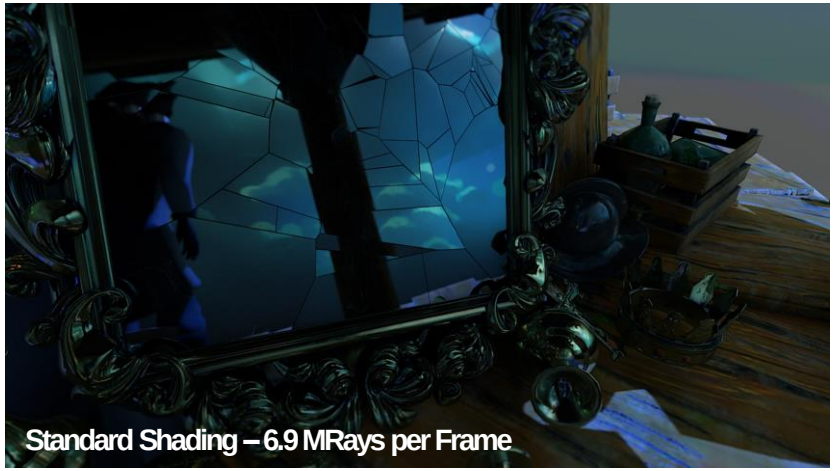
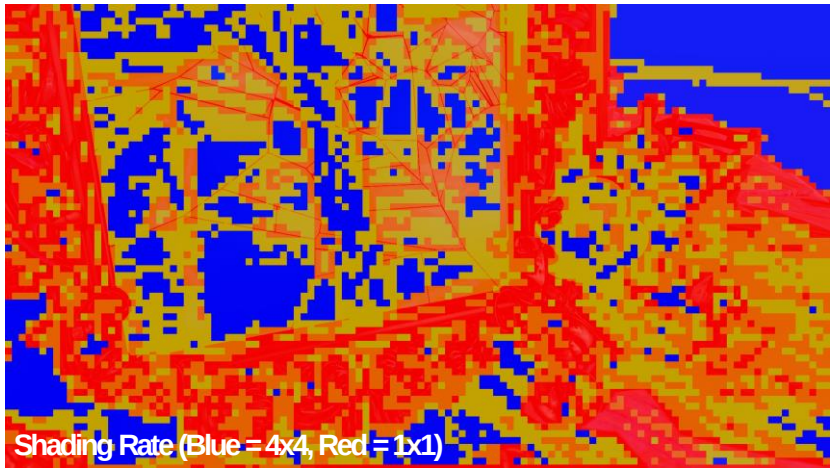


Image used with permission from https://www.king.com/

Ray Tracing combined with Fragment Shading Rate

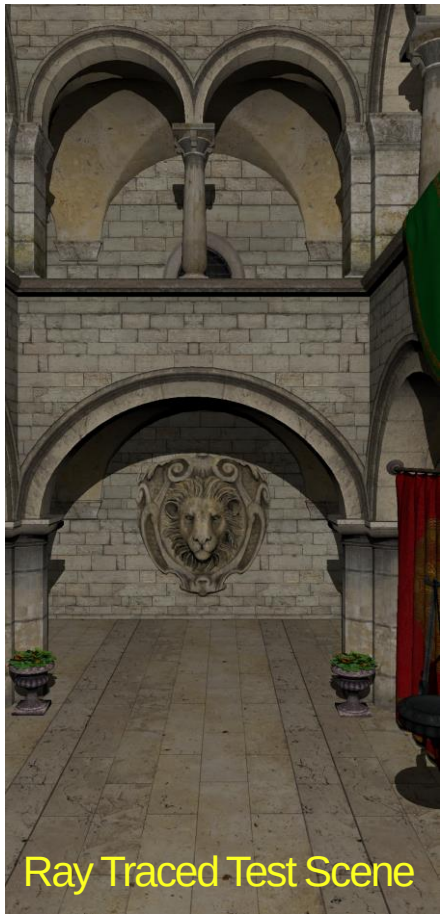
Improved Power Performance Efficiency with minimal impact on Quality





The Ray Tracing Difference – Level 4 versus Level 2 and 3 Ray Tracing Performance

Sponza Palace with Ray Traced Hard Shadow



Sponza Rendering Test Scene

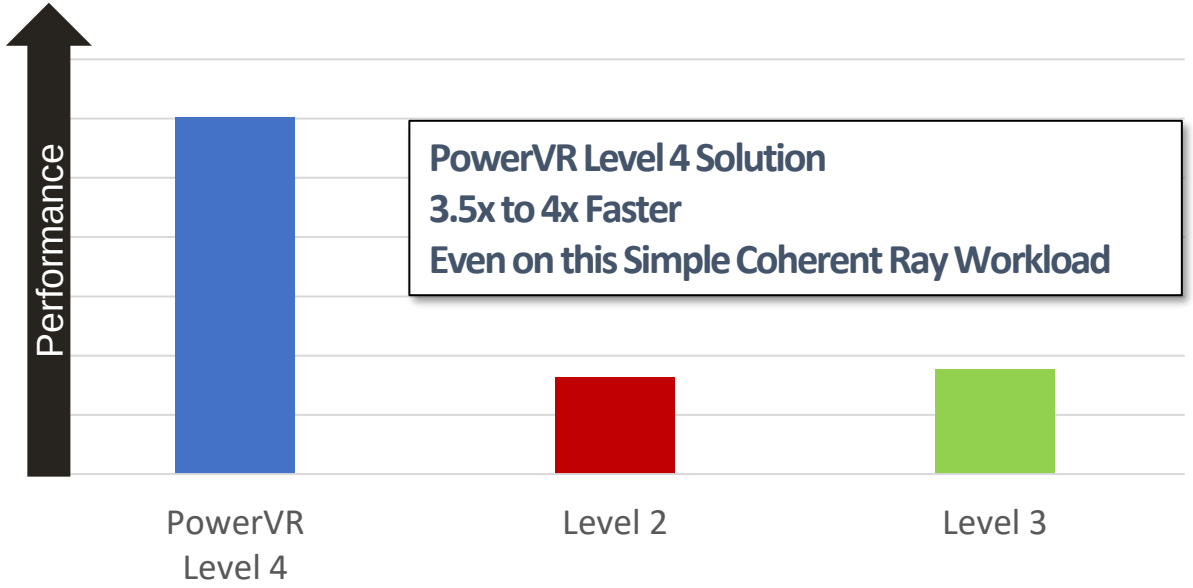
Commonly used for Rendering and Ray Tracing Tests

Here Ray Traced with Ray Query, single light, hard shadow (1 ray)

Simple scene with minimal ALU loading, coherent rays

Measurements on real platforms/implementations

Performance/Efficiency Comparison



Note: IMG Labs Measurements, platforms normalised to FP32 operations/clock, using latest drivers and application using Vulkan API with Ray Queries

Summary

Adoption of Ray Tracing is growing

PC and Console Today, Mobile emerging

Growing opportunities in Data Centre including Cloud Gaming

Ray Tracing offers high value

Improved visual quality

Improved efficiency vs complex approximations in shader code

Not all Hardware Solutions are Equal

Offload from programmable to fixed function units

Non-coherent processing and memory access behaviour

Brute force solutions will struggle especially in Mobile

Imagination

THANK YOU

SYNOPSYS®
Silicon to Software™

Addressing Scalable Processor Performance Requirements in High-End Embedded Applications

Linley Fall Processor Conference 2022

(Powered by TechInsights)

Kulbhushan Kalra, Director of R&D, Synopsys

November 1, 2022

Performance Tuning Critical for High Performance Embedded

Networking

- Home, edge, cloud usage growing
- Intelligence distributed across the network
- Processing and storage at each compute node
- Higher bandwidth, more data processing

SSD Storage

- Rapidly increasing drive capacity
- Higher bandwidth, lower latency, more IOPS/Watt
- Encryption, decryption
- In-storage compute & AI
- Soft RT processing

Artificial Intelligence

- Inference moving into the edge
- Dedicated chips for training in the data center
- Very high MAC throughput
- Very high memory bandwidth

Automotive

- Integrated application processors in complex SoCs (ADAS)
- Real-time safety processing
- V2X communications
- Highest reliability, security and safety

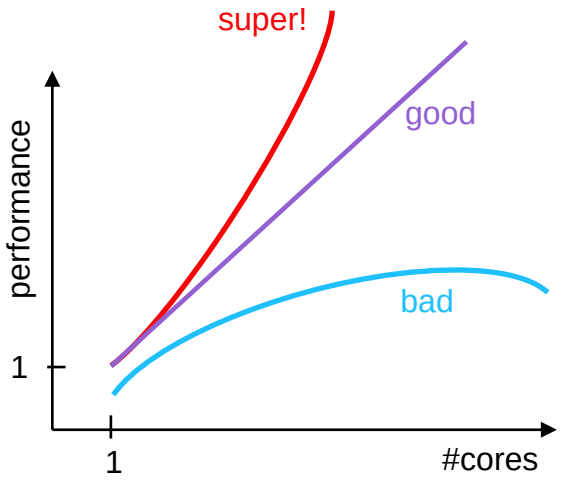
Multicore Performance Scaling

How to Effectively Use 12 Cores ?

- Application must be parallelizable
  - Ok for significant parts of SSD, networking, AI, wireless workloads
  - But: never forget Amdahl's law …

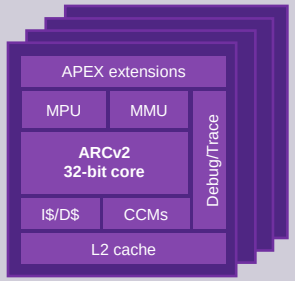
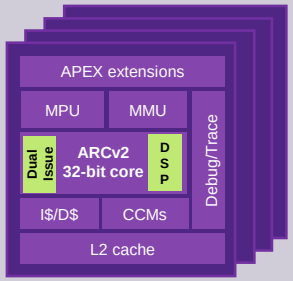
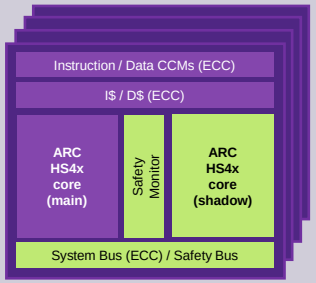
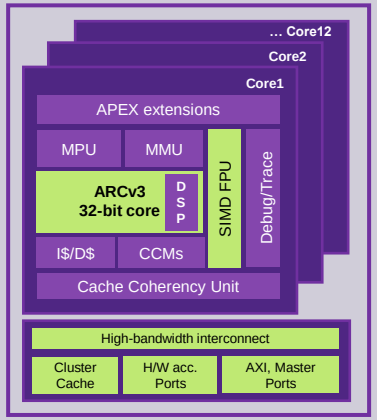
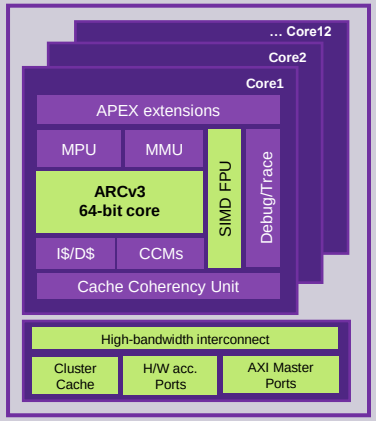
Speedup is limited by serial part of program

- Synchronization overhead must be low
  - Requires low-latency processor cluster architecture
  - Semaphores, mailboxes, …
- Communication bandwidth must be sufficient
  - Memory and interconnect architecture are key



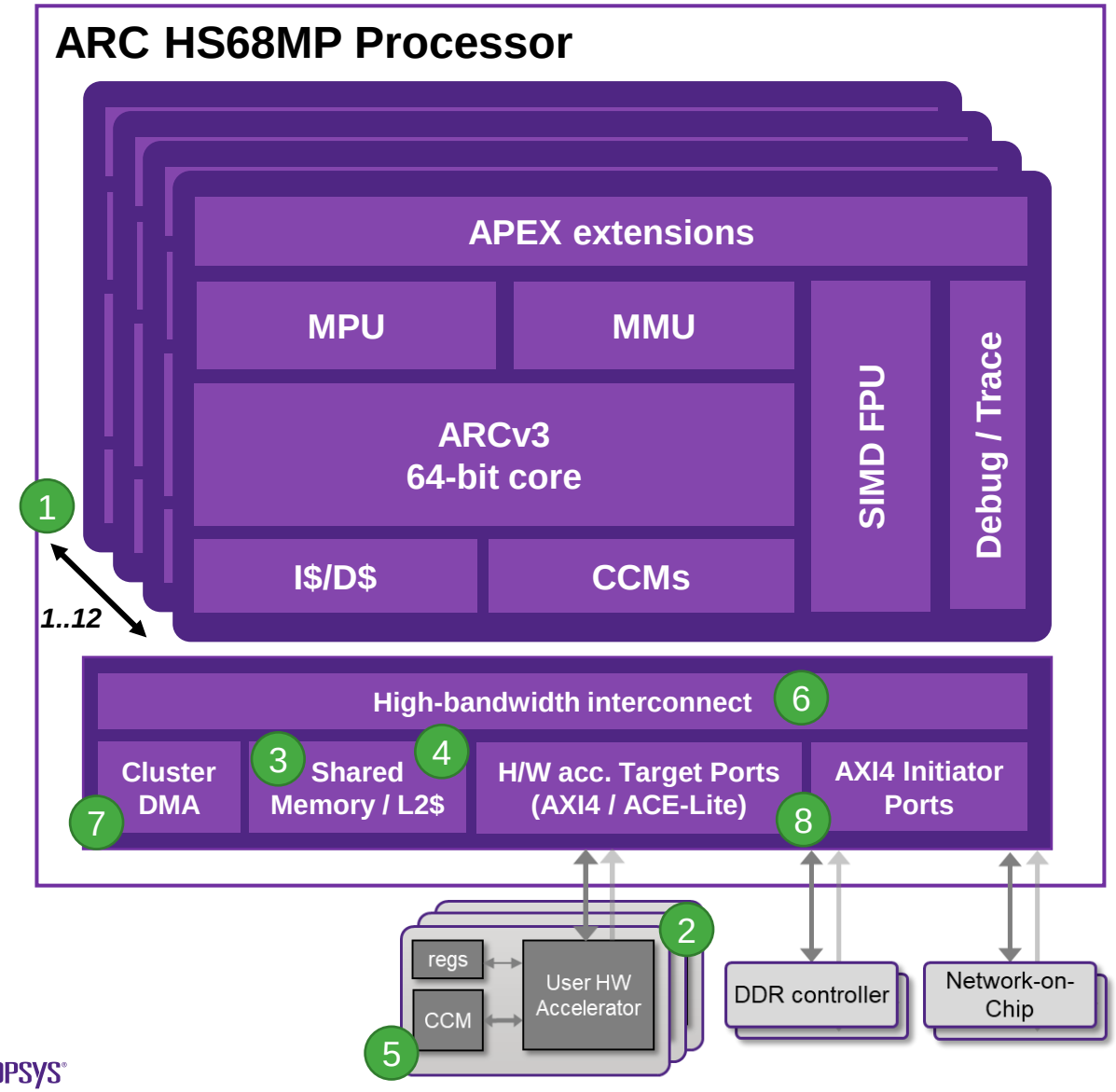
- In some cases, super-linear speedups can be achieved
  - Usually because the software working set suddenly fits in the aggregate L1 caches
  - …but this is rare

ARC HS Processors - Optimized for High-End Embedded

ARC HS3x	ARC HS4x / HS4xD	ARC HS4xFS Safety	ARC HS5x	ARC HS6x
				
ARCv2 based 32-bit processor • Single issue, 10-stage pipeline • Dual/quad-core implementations • APEX custom extension support	**ARCv2 based 32-bit processor** • Dual-issue, 10-stage pipeline • 150+ DSP instructions (HS47D) • 40% higher performance (HS3x)	**ARCv2 based safety processor** • Integrated H/W safety features • Single, dual and quad-core implementations (lock-step) • Optional FuSa S/W & certified MetaWare for Safety tools	**ARCv3 based 32-bit processor** • Up to 12 core coherent cluster with up to 16 H/W accelerators • Coherent, high bandwidth interconnect (800GB/s) • 150+ DSP instructions (HS57D)	**ARCv3 based 64-bit processor** • Up to 12 core coherent cluster with up to 16 H/W accelerators • Coherent, high bandwidth interconnect (800GB/s) • 3x cluster performance over HS4x (quad-core comparison)

* single core

Extreme Power / Performance Trade-offs with ARC Processors

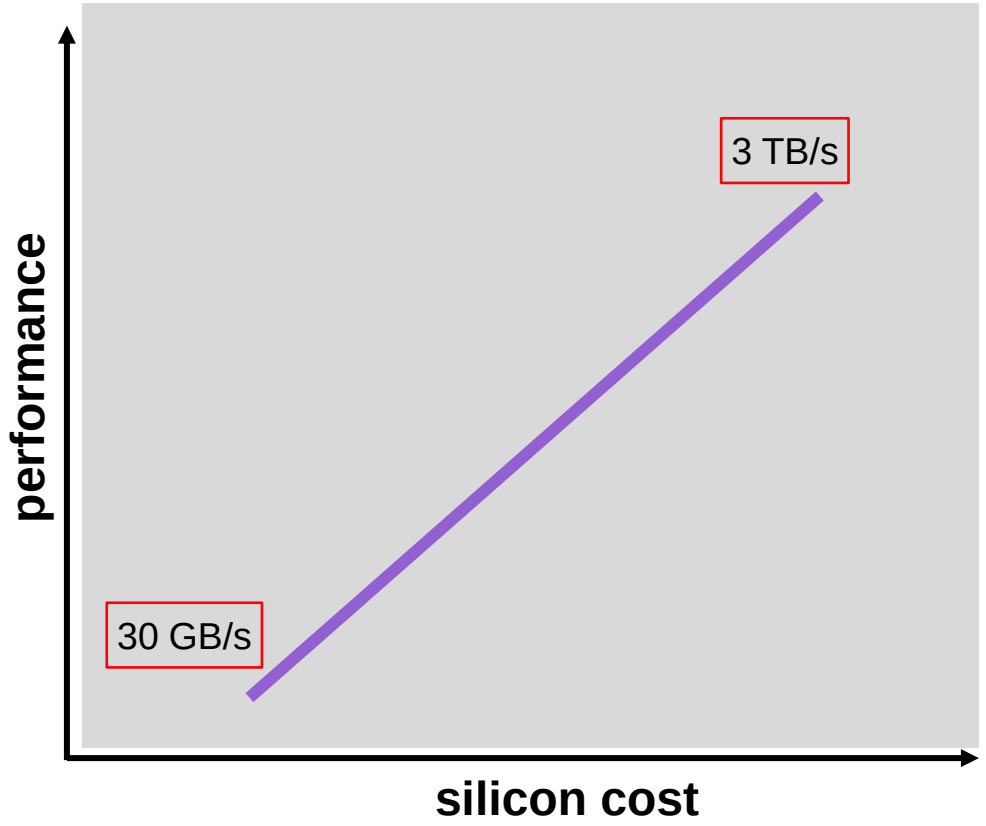


1. Unique interconnect supports up to 12 configurable cores (with I/O coherency)
2. Extensible support for up to 16 app-specific H/W accelerators
3. Flexible width and number of shared memory data banks
4. Memory sub-banking for power optimization
5. Large accelerator CCMs for predictable memory access times
6. Quality of Service (QoS) for user-controlled CPU bandwidth allocation
7. Cluster DMA for efficient data transfers
8. Interfaces with configurable data/address widths to fine-tine data throughput

Broadening the Performance Range

Flexibility to enable bandwidth ranges from 30 GB/S to over 3000 GB/s

Performance Impact Parameter	ARC HS6x Options
Bus widths of interfaces	32 to 512-bit
Width of shared cluster memory (SCM) data banks	128 to 512-bit
# of SCM data banks	2 to 32
# of NoC ports	1 to 4
# of outstanding transactions per port	Variable
Clock frequency	Variable



ARC HS6x provides a performance level parameter setting

- **A**rea-optimized, **M**edium, **B**andwidth, e**X**treme
- Controls degree of internal buffering
- Degree of concurrency in transport network

Adapting to a Wide Range of SoC Architectures

Ultra-flexible Interconnect Options

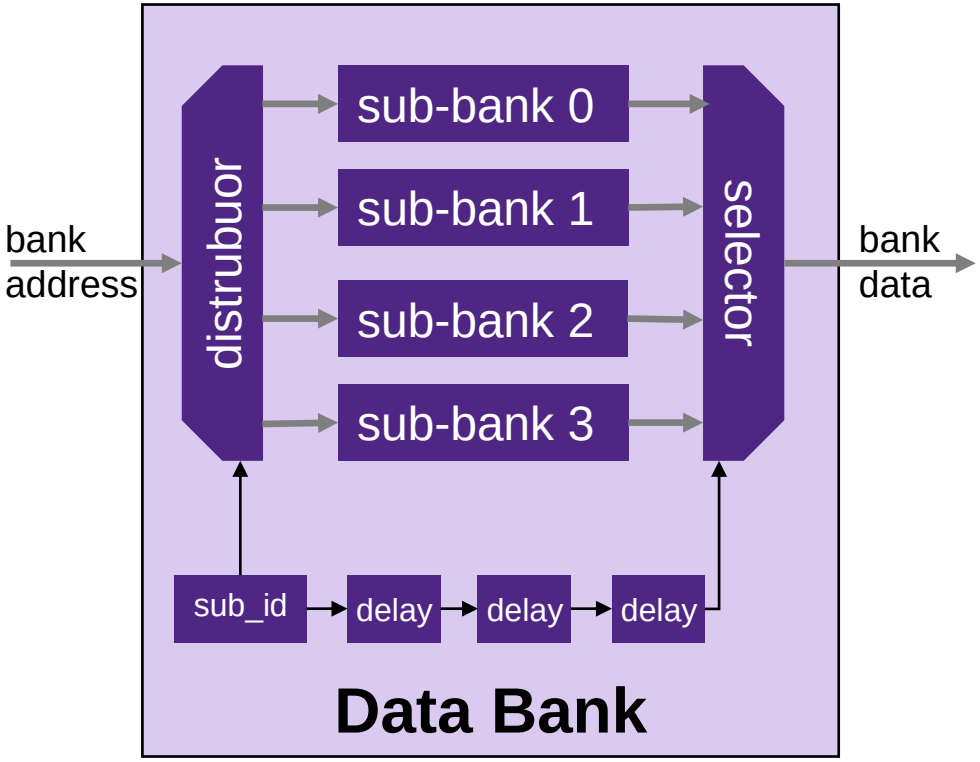
ARC HS68 Interface	# of Ports	Type	Description	Configurable data channel width (bits)	Speed*	Throughput (each interface)
Accelerator, DMI	0 .. 16	AXI4, ACE-Lite	Coherent high bandwidth, low-latency connections to the shared cache and memory, and to all CCMs	32, 64, 128, 256, 512	0 .. 2 GHz	up to 64B/cycle
NoC, DDR	1 .. 4	AXI4	High bandwidth memory interfaces to connect to DDR or NoC	128, 512	0 .. 2 GHz	up to 64B/cycle
Peripheral	0 .. 2	AXI4, AHB	Connections to the peripheral network	128	0 .. 2 GHz	up to 16B/cycle
CCM	0 .. 16	AXI4	Connections to small memories, such as Closely Coupled Memories (in accelerators), or a boot ROM.	32, 64, 128, 256, 512	0 .. 2 GHz	up to 64B/cycle

* *Each individual interface can be configured as synchronous, asynchronous, or source synchronous*

Maximizing Data Bank Bandwidth at Minimum Leakage Power

- Low-leakage SRAM may be too slow for single-cycle access

- ARC HS6x processors deploy sub-banking
  - *Each data bank still delivers 1 beat / cycle*
  - *2 or 4 sub-banks operating in interleaved mode*
  - *Sub-bank scheduler rotates access to next sub-bank every cycle*
  - *Capturing each sub-bank output after 1+WAIT cycles*
  - *WAIT = 1, 2 or 3 cycles*

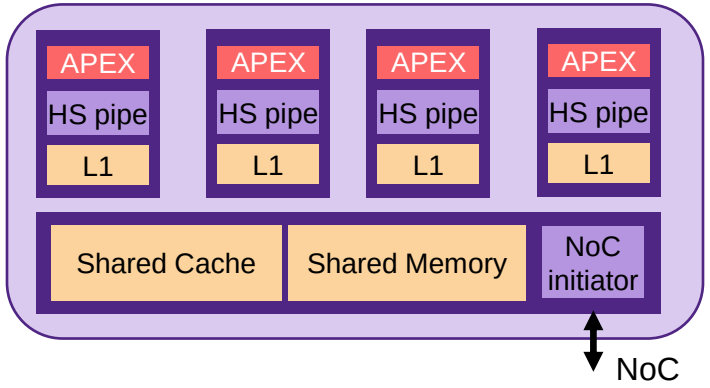


Example Configuration

Sub-banking optimizes power by accommodating high-Vt (slow) SRAMS

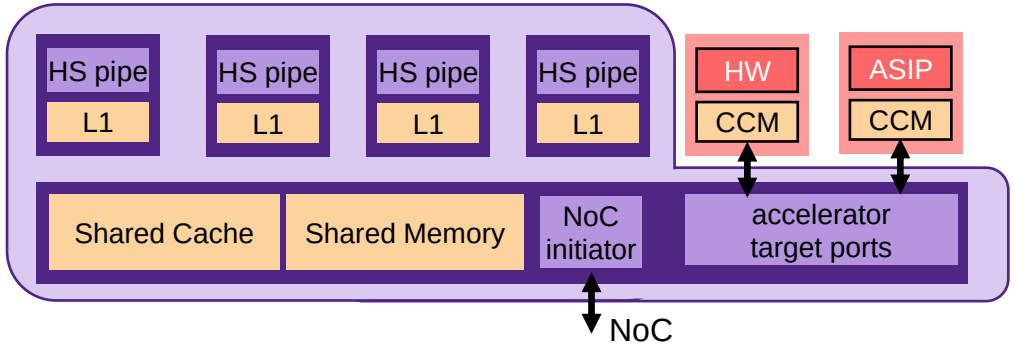
ARC HS6x Cores Offer Multiple Accelerator Options

Enabling customers to create unique implementations



APEX (ARC Processor EXtensions)

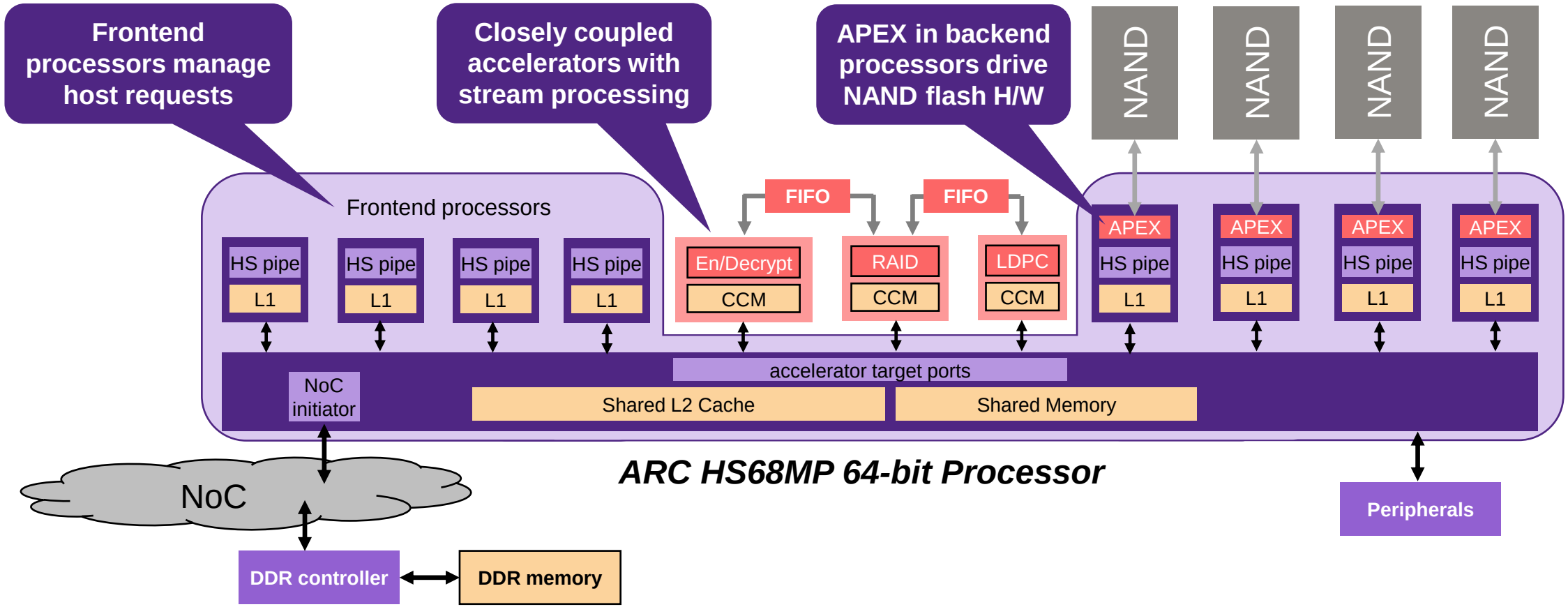
- Custom Verilog, integrated with ARC tools
- Integrated in the processor pipeline
- Operation controlled by APEX instructions
- Direct access to processor registers
  - *Source/destination operands in register file*
  - *Test/set flags in condition register*
- Use LD/ST instructions for memory access



Closely Coupled Accelerators

- Customer-designed hardware or ASIP
- Closely coupled to the processor cluster
- HW operation controlled via AUX registers
- Direct access to shared memory
  - *High bandwidth, low latency*
  - *Cached or uncached*
- Optional CCM, cluster accessible

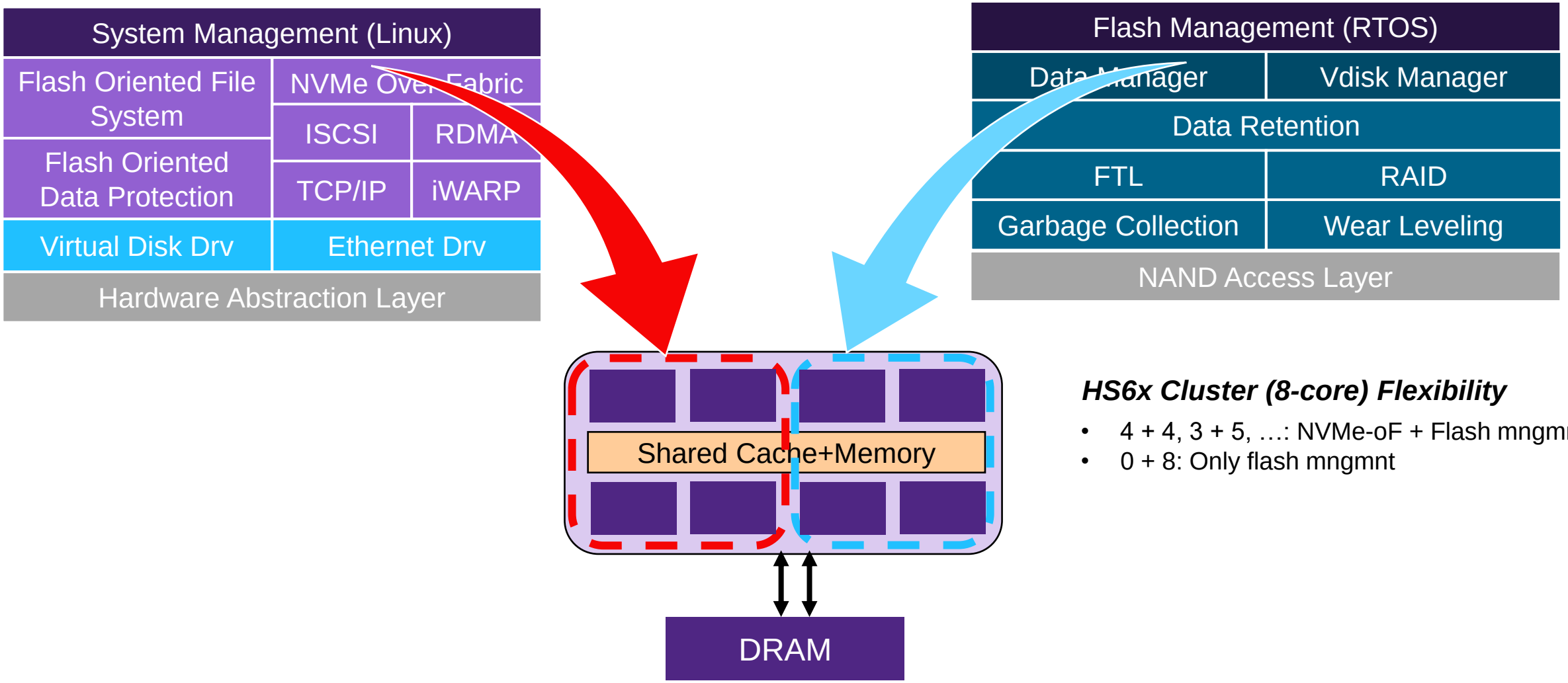
ARC HS68MP – A Complete Solution for Computational Storage



LPDC = Low Density Parity Check
RAID = Redundant Array of Inexpensive Disks

ARC HS6x Processors Deliver Software Flexibility

NVMe-oF and Flash Management S/W on Same Cluster



HS6x Cluster (8-core) Flexibility

- 4 + 4, 3 + 5, …: NVMe-oF + Flash mngmnt
- 0 + 8: Only flash mngmnt

Summary

- Applications requiring ever more performance
  - *SSD, Networking, AI, Automotive, Wireless, …*
- Performance can be scaled along several angles
  - *Multicore, Processor Extensions, HW acceleration, …*
  - *It is all about improving the architecture*
- ARC HS68 processors can scale to exceptional performance levels
  - *Integrating custom hardware accelerators*
  - *Including up to 12 HS6x cores, with the same or different configurations*
  - *Ultra-flexible interconnect options for performance tuning*
- Heterogeneous cluster provides best of both worlds
  - *Optimal power, performance and area balance*

SYNOPSYS®
Silicon to Software™

Thank You

Premier Sponsor

flexlogix Technologies, Inc.

Platinum Sponsor

ceremorphic®

Gold Sponsors

CEVA
cadence
ARTERIS IP
MOVELLUS
TetraMem
MARVELL
GlobalFoundries
SYNOPSYS® Silicon to Software
Achronix Data Acceleration
proteanTecs
ANDES TECHNOLOGY
expedera
kinara
UNTETHER AI
quadric
innatera
brainchip
Imagination
SIEMENS
BLUE CHEETAH
Qualcomm
NXP
SiFive
intel

TechInsights

Tech Insights

techinsights.com

Thank You!