



Dimensity 9000 – A Flagship Smartphone SoC

Presenter: **Mediatek** Ericbill Wang

Co-author: **Arm** Stefan Rosinger, Saurabh Pradhan



Antutu New Model Exposure

HOT

Brand: Dimensity

Model: 9000

Memory: 12GB+256GB

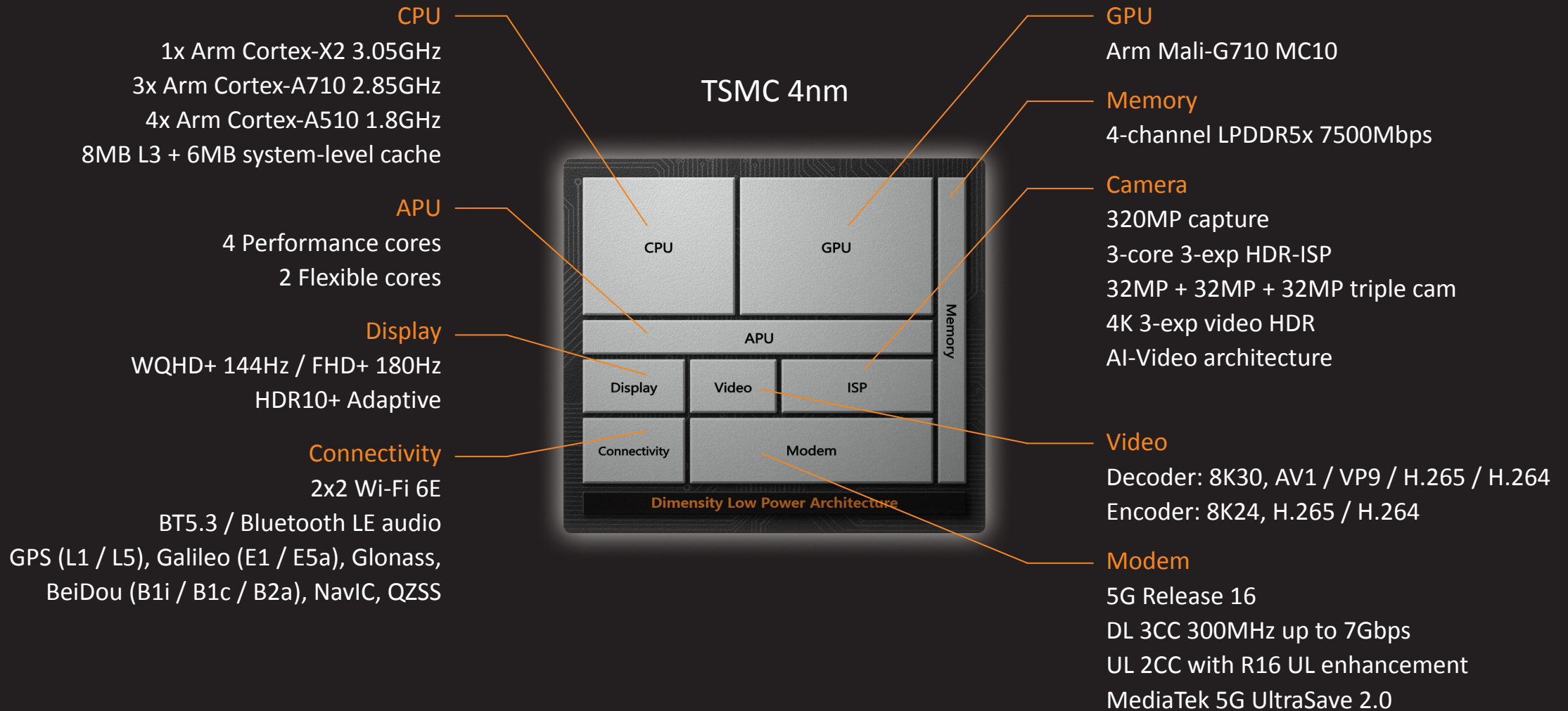
OS: Android 12



1007396

SOC: MT6983

MediaTek Dimensity 9000



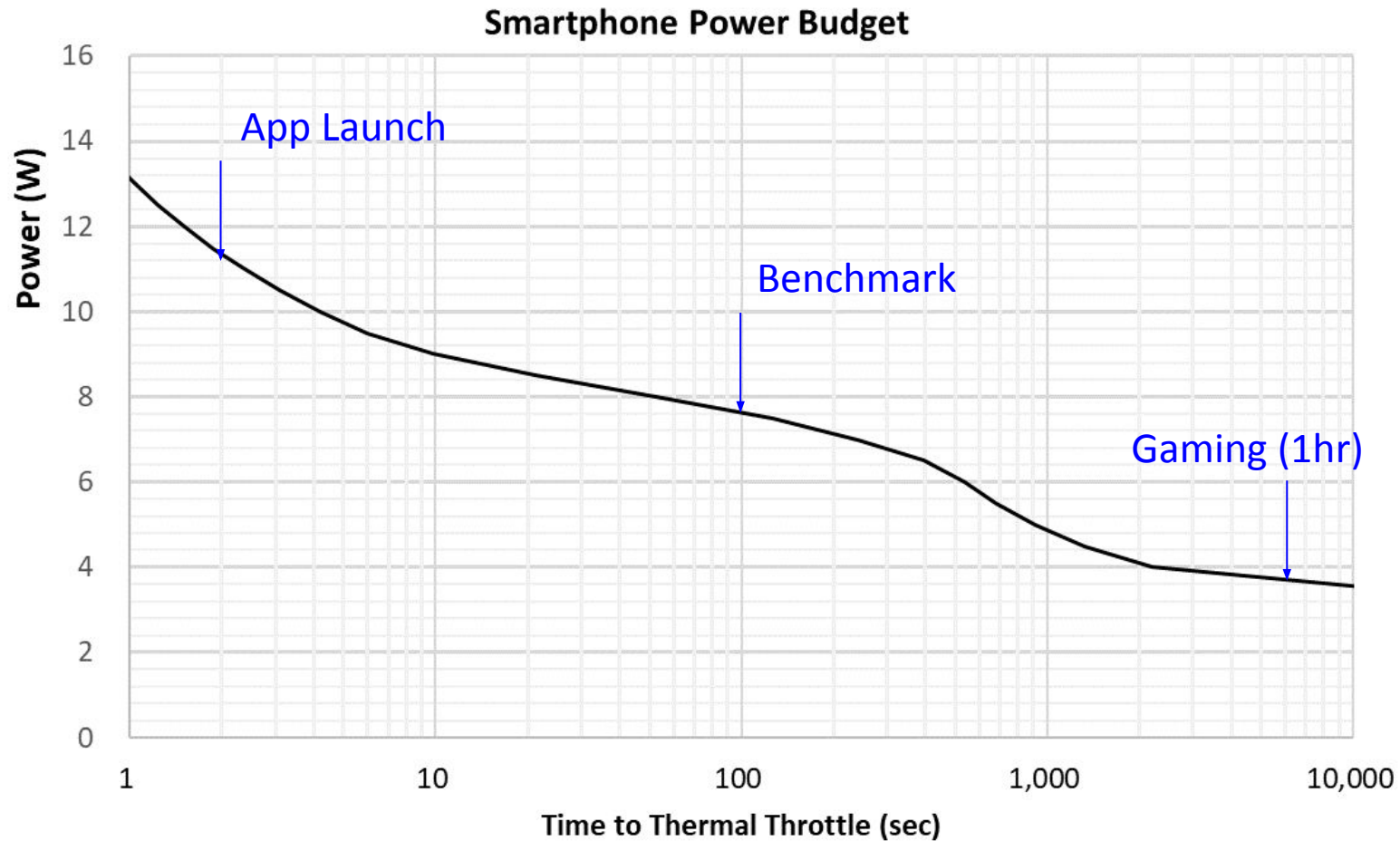
What Makes a Great Smartphone SoC

- Display, Camera & Gaming drive higher performance
- User demands responsive and sustainable operation experience
- Thin & Light device limit thermal power budget & battery size

Required High Performance **AND** Low Power Processors

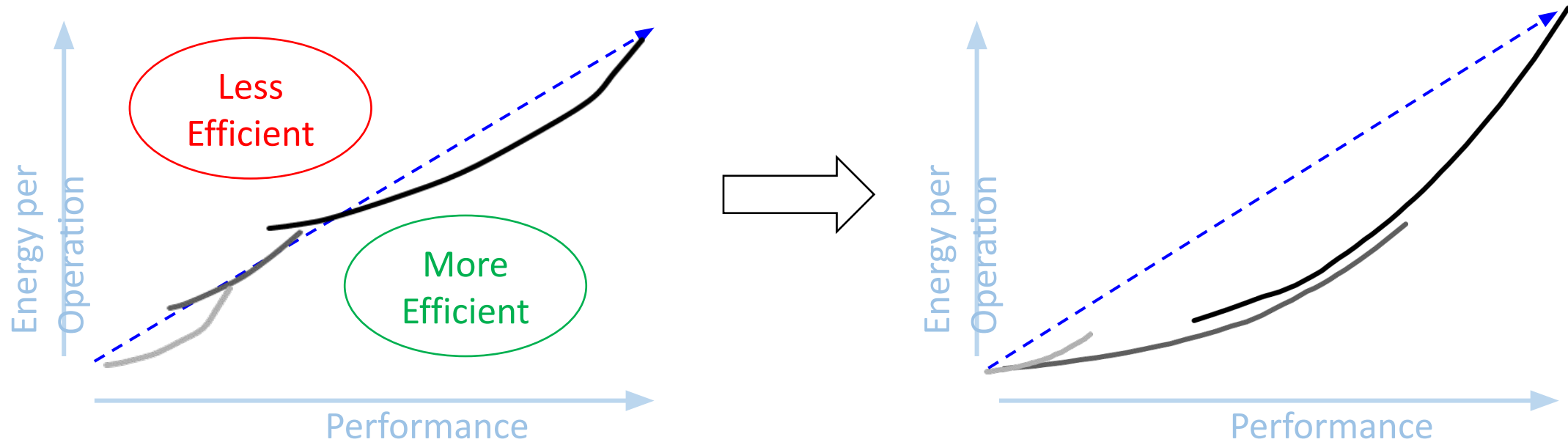
CPU Challenge: One Size Does Not Fit All

Workload: 1) Short-burst 2) Sustainable 3) Daily Use
Challenge: 1) Peak Perf 2) Perf & Power 3) Low Power



Learnings on Tri-gear CPU Architecture - 1

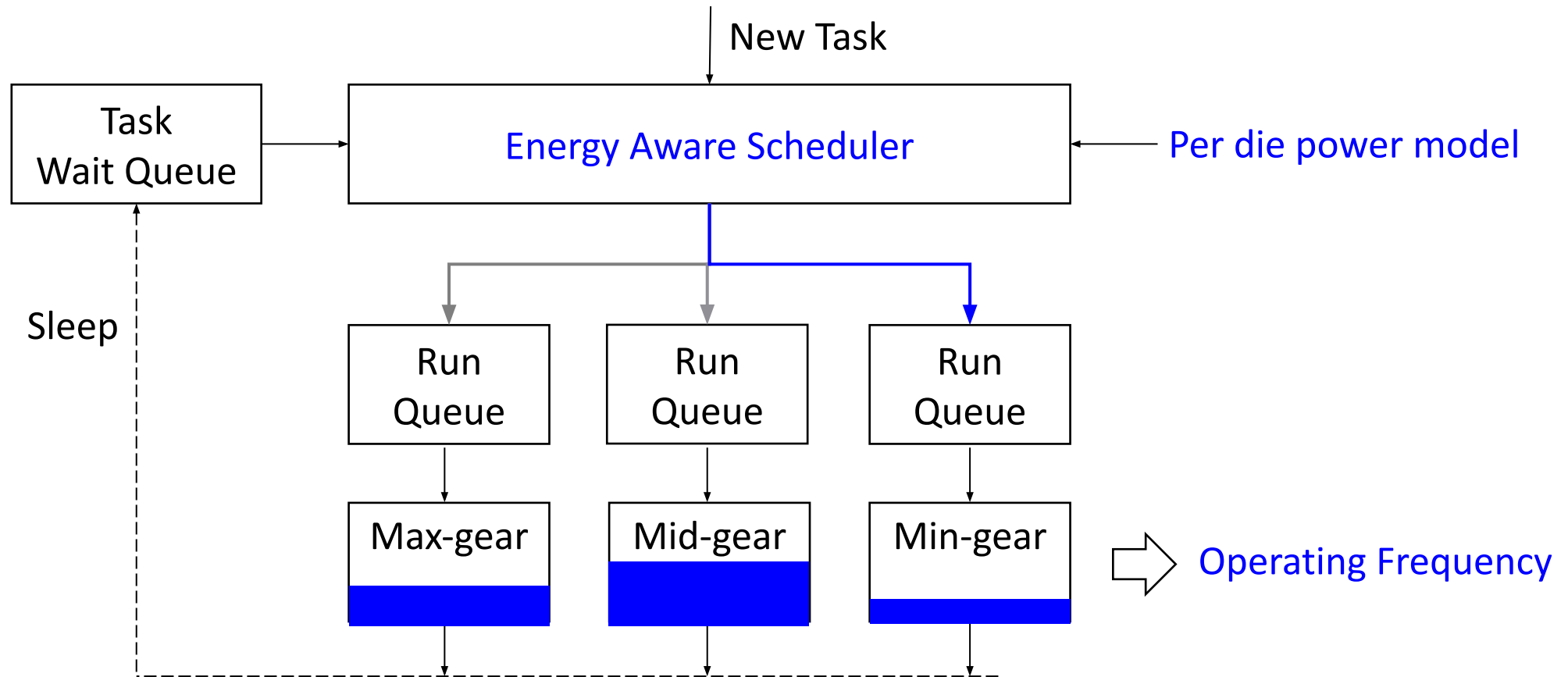
▪ Usable Tri-gear Power Efficiency Curve



- Min-gear CPU
- Mid-gear CPU
- Max-gear CPU

Learnings on Tri-gear CPU Architecture - 2

- Energy Aware Scheduler
- Event Based Operating Frequency Decision



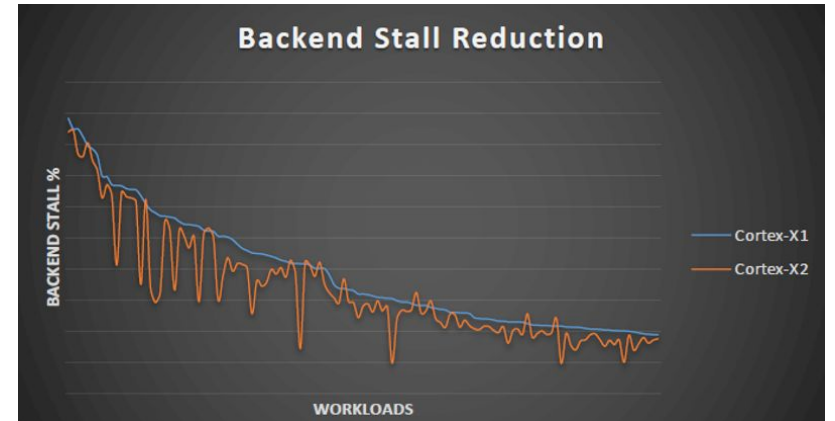
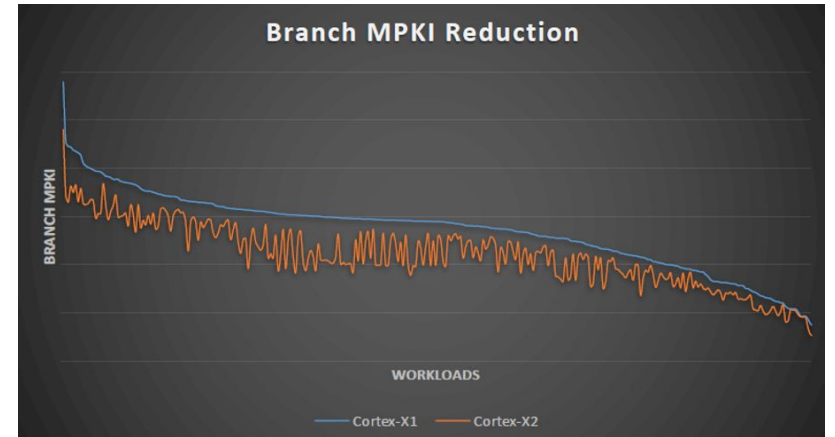
CPU Highlights vs. Dimensity 1200

- Arm Cortex-X2 +40% integer performance over Arm Cortex-A78
- Arm Cortex-A510 +35% integer performance over Arm Cortex-A55
- -50% CPU power @ iso-performance
- Geekbenchv5 single-thread 1278 (+36%) , multi-thread 4400 (+33%)

Cortex-X2 1MB L2	Cortex-A7 10 512KB L2	Cortex-A5 10	Cortex-A5 10	Cortex-A5 10	Cortex-A5 10
	Cortex-A7 10 512KB L2	Shared 512KB L2\$ Shared VPU		Shared 512KB L2\$ Shared VPU	
DynamIQ Shared Unit equipped 8MB L3\$					

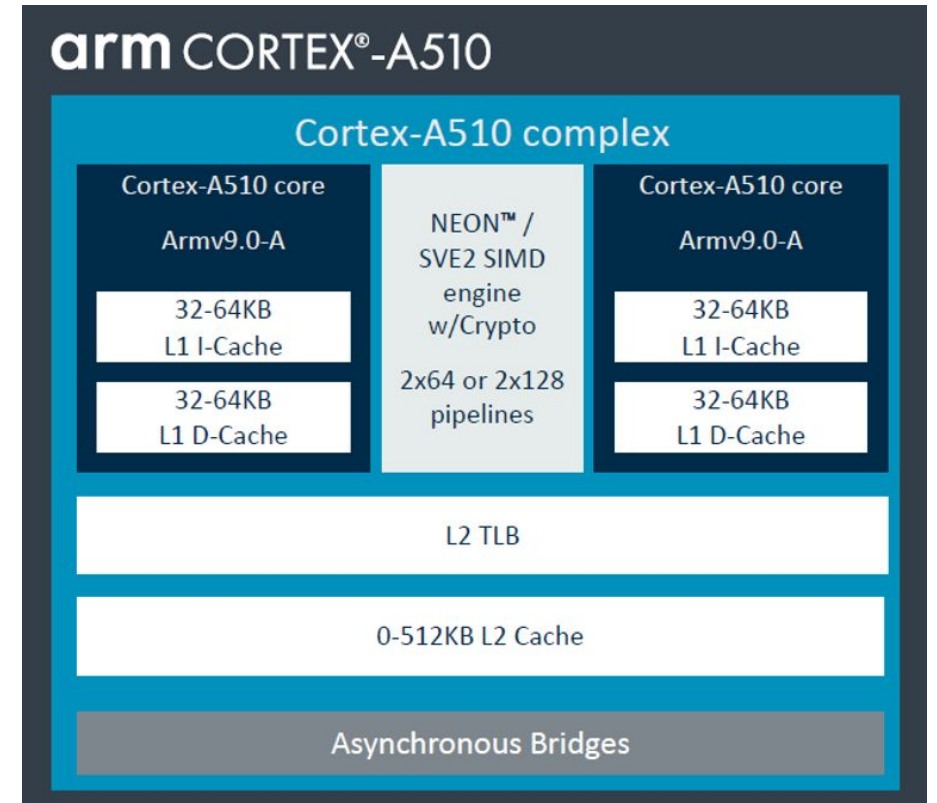
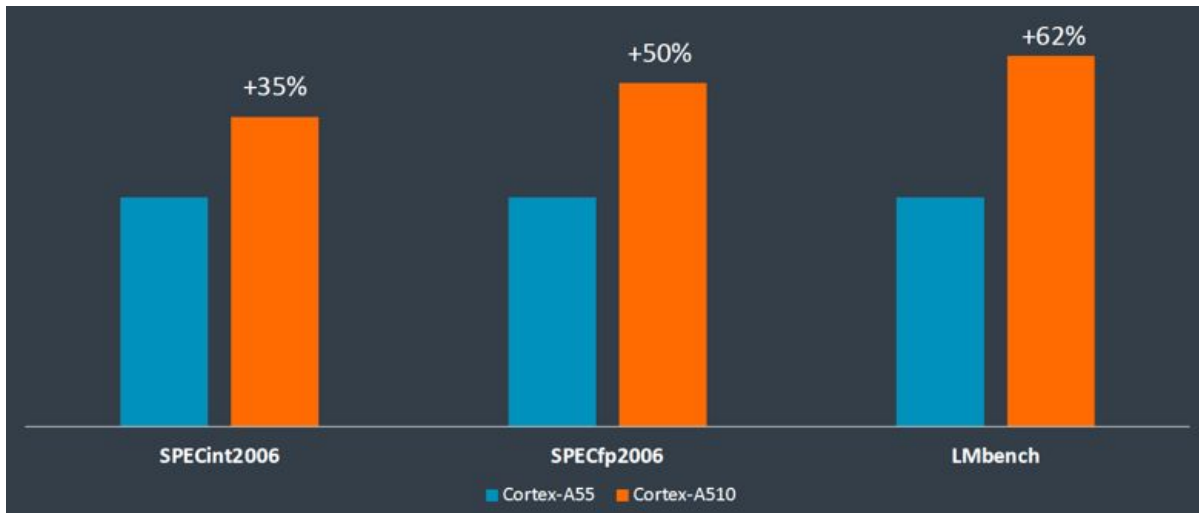
Arm Cortex-X2

- Performance: Arm Cortex-X2 +16% over Arm Cortex-X1
- Front-End
 - Decouple branch prediction from fetch
- Out-of-Order Core
 - Remove a pipeline stage at dispatch
 - +30% out of order window
- Back-End
 - +33% load/store structure size
 - Data prefetching enhancements



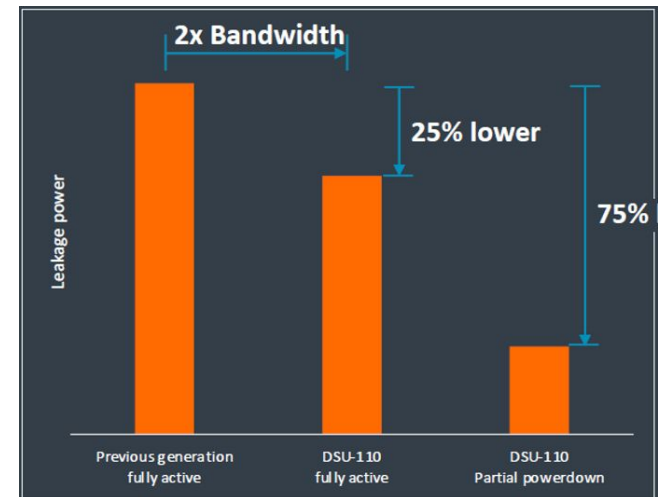
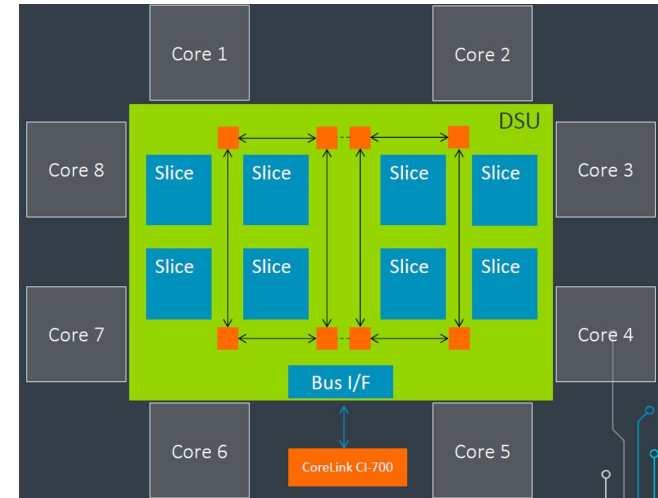
Arm Cortex-A510

- Performance: Arm Cortex-A510 +35% over Arm Cortex-A55
- Merged-Core Microarchitecture



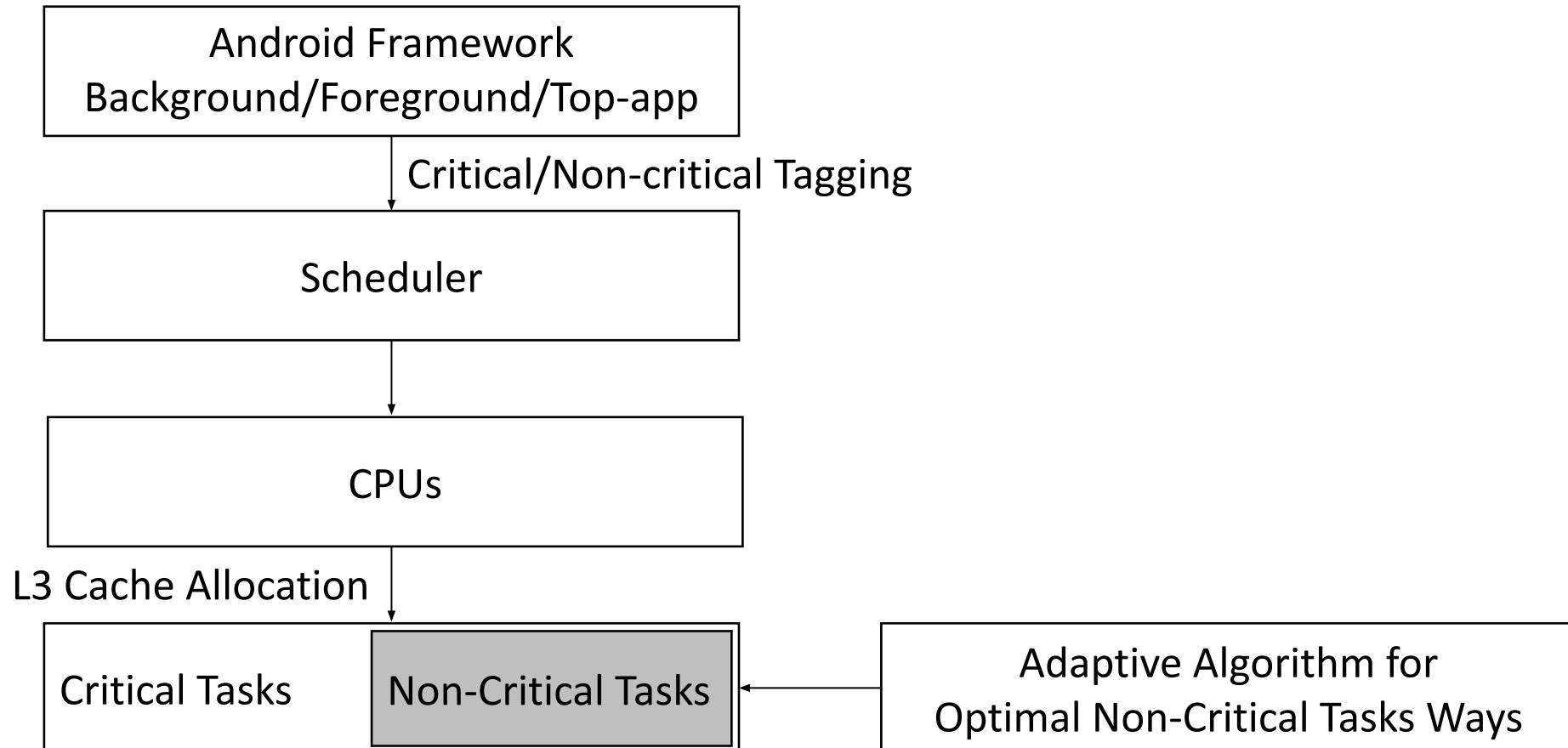
Arm DynamIQ Shared Unit-110

- DSU-110 for Cache Coherency & Shared L3\$
- Optimized Ring Transport Network
 - Bi-directional, Dual-ring
- 2X Bandwidth and 25% Lower Leakage
 - Support Partial SRAM & Logic Shutdown
- Cache Partition for QoS



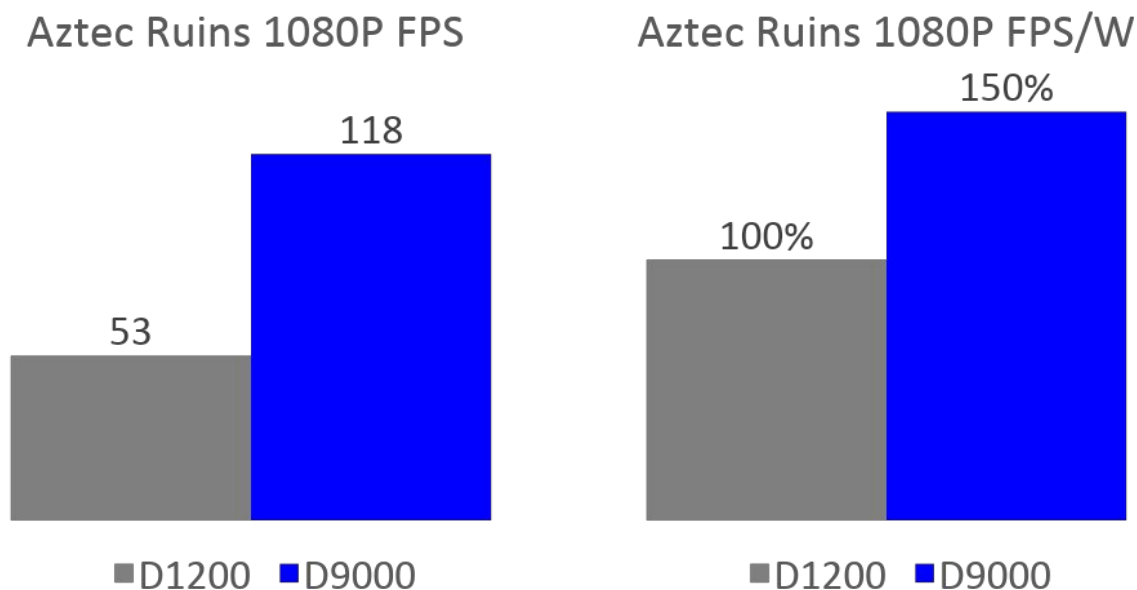
CPUQoS Technology

- Real system is noisy, mixed critical and non-critical tasks
- Speed-up 14% for application launch stress, -5% power on 120fps game



GPU Highlights vs. Dimensity 1200

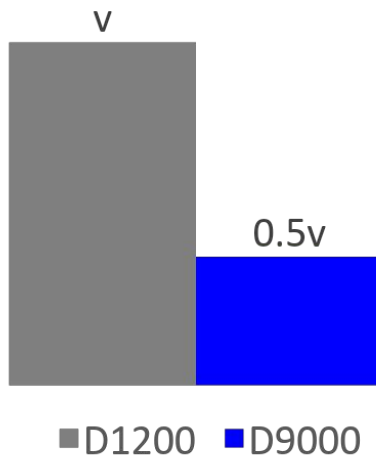
- 2.2X peak performance from Arm Mali-G77 9 cores to Mali-G710 10 cores
- 1.5X power efficiency from N4 process & new Mali-G710 IP
- Enable Genshin 60fps, PUBG HDR 90fps



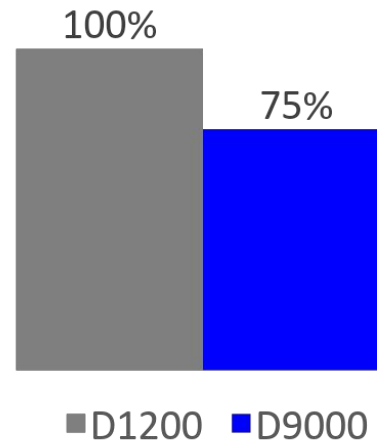
GPU Sustained Performance Optimization

- Lower minimum operating voltage to 0.5V
- -25% GPU bandwidth by larger GPU & system cache + compression
- -15% GPU driver loading on CPU by offloading to embedded processor

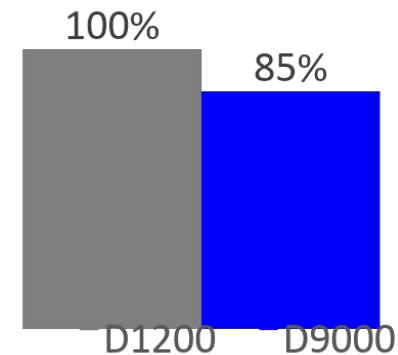
GPU Operating Voltage



GPU Bandwidth



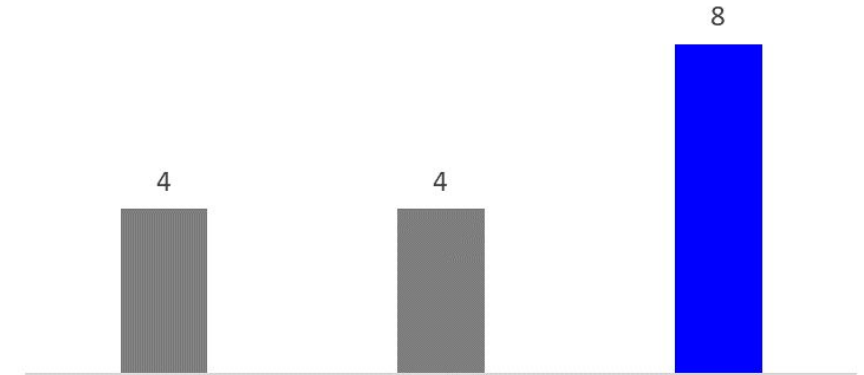
CPU Loading



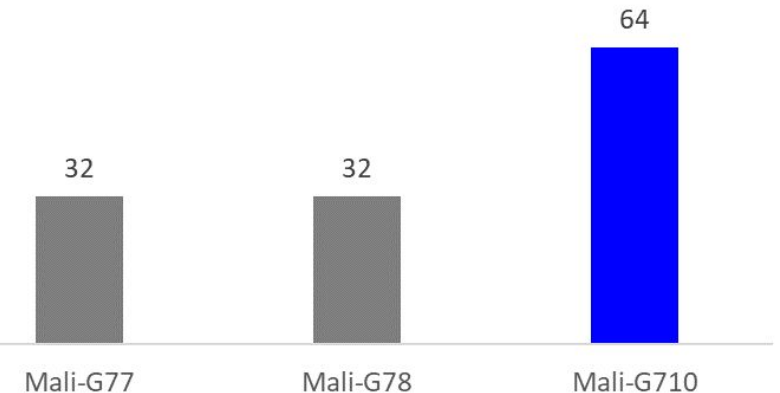
Arm Mali-G710

- Bigger Core: 2X Texel & FMA rate
- Redesigned execution engine
- 1.2X performance density

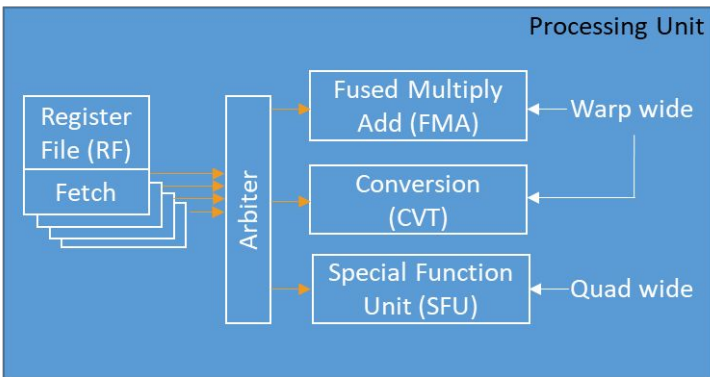
Bilinear Texels per clock, per core



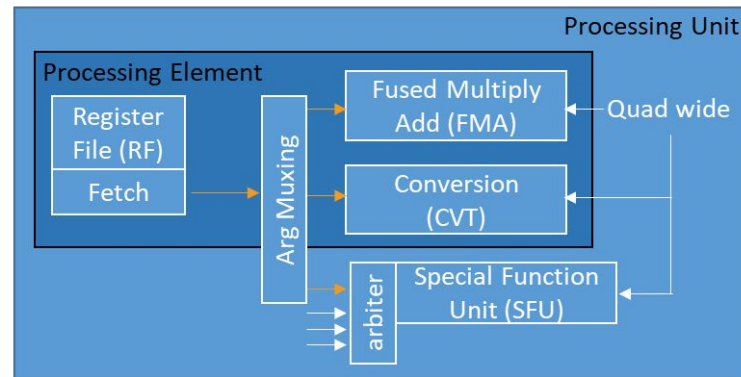
FMA per clock, per core



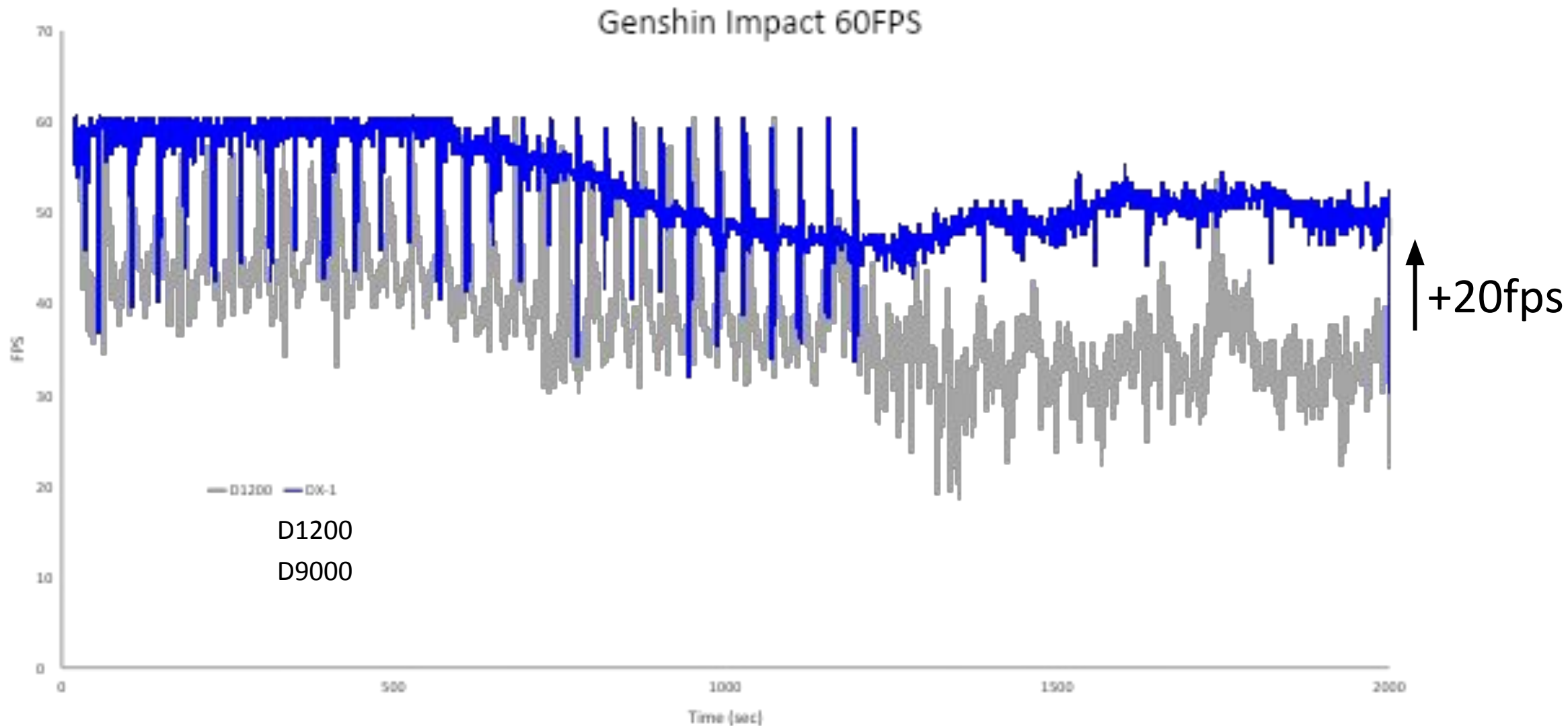
Mali-G78



Mali-G710



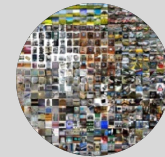
Gaming Improvement vs. Dimensity 1200



AI Applications on Smartphone



Detection



Classification

**Visual
Perception**

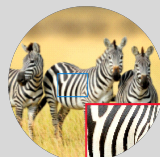


Segmentation



Depth Estimation

**Object
Construction**



Noise Reduction



Super Resolution

**Image
Quality**



Speech Recognition

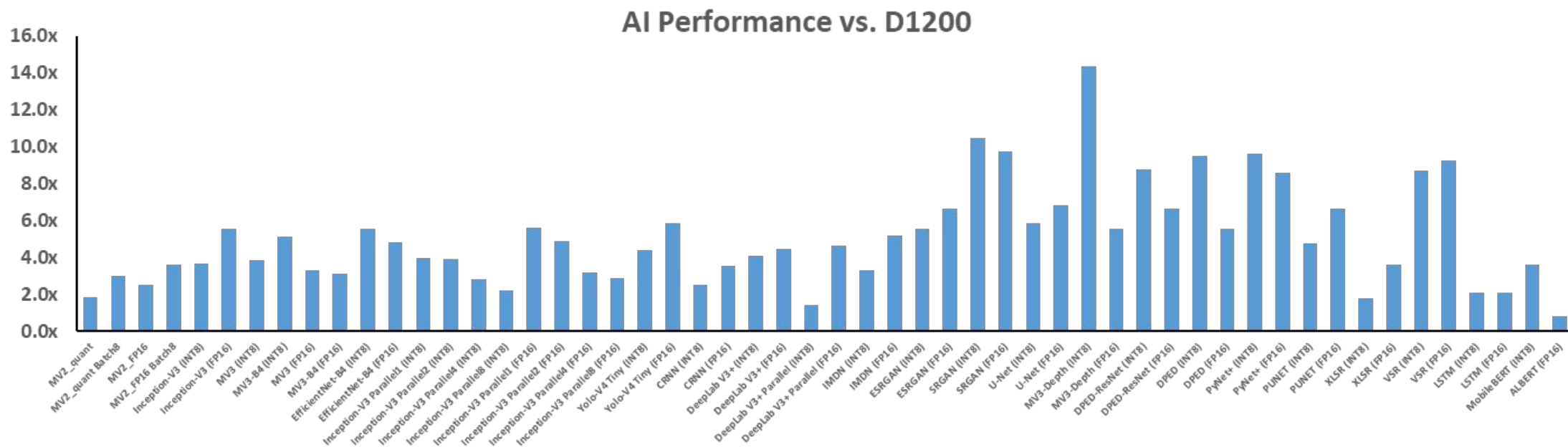


Speech Translation

**Speech
Applications**

APU Highlights vs. Dimensity 1200

- 4.3X performance
- 2.8X power efficiency
- Enable flagship AI features on camera capture and sustainable video



APU Feature Overview

Versatile data type & operators

4b/8b/16b integer, FP16, BF16

Power efficient MAC array

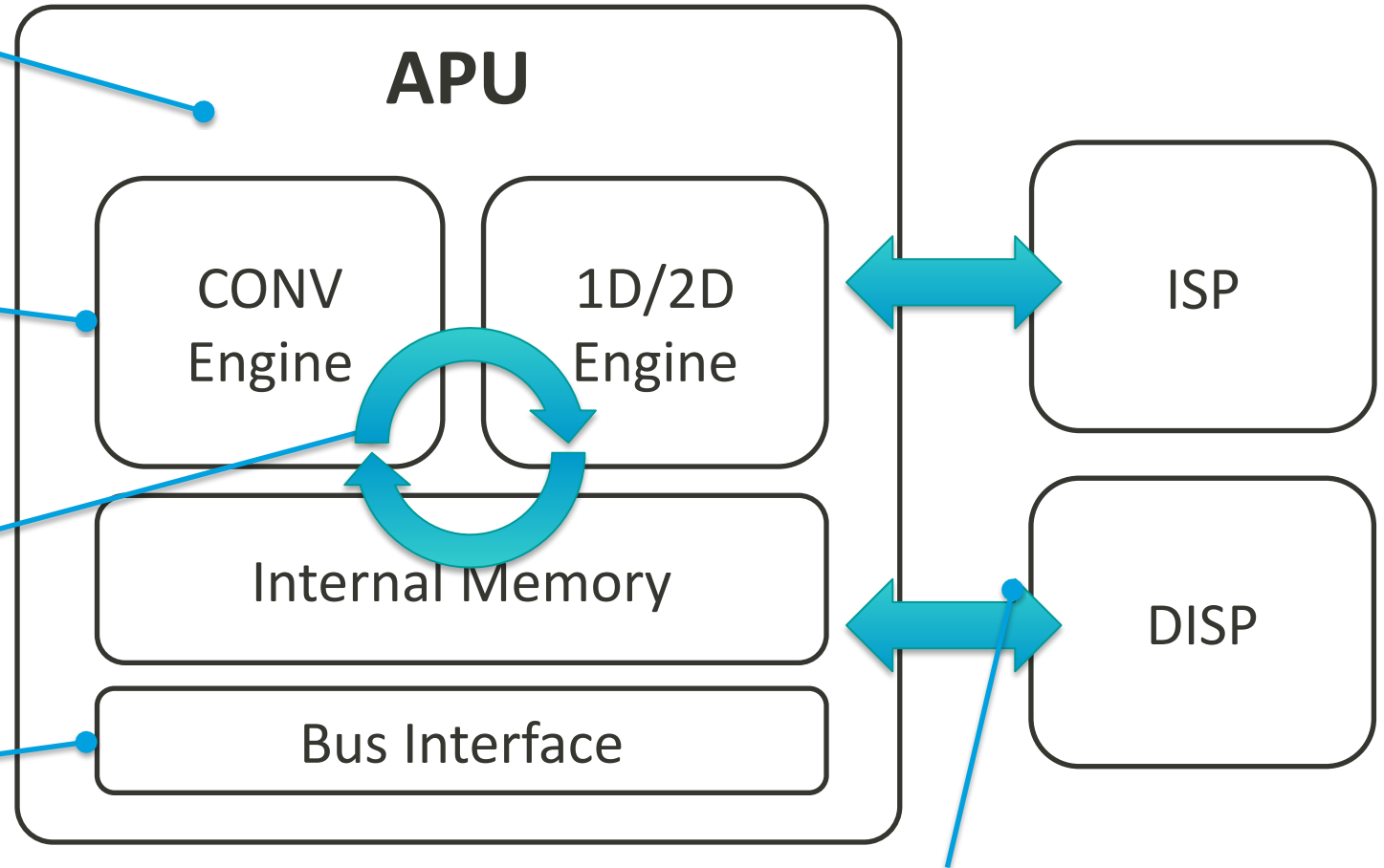
Efficient MAC arch., data reuse

In-APU data exchanging

Layer-fusion, flexible tile walking

Minimized DRAM Data

Data compression

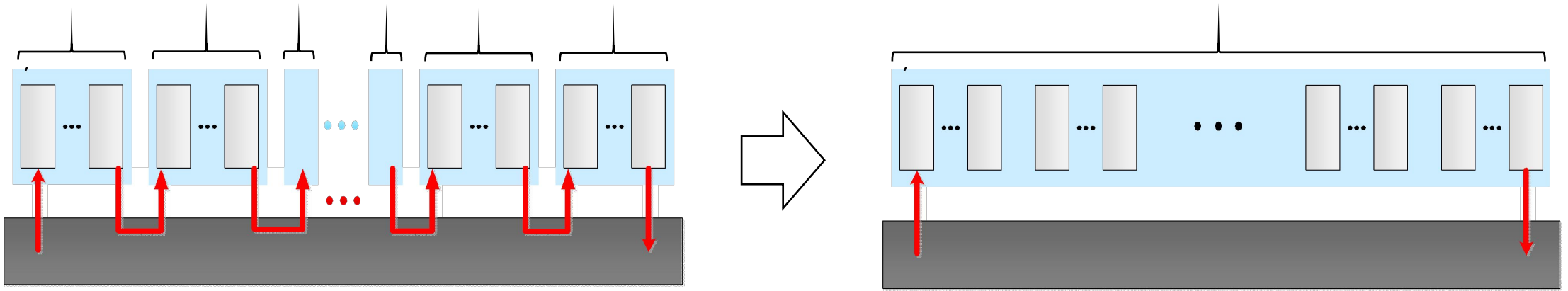


Inter-Subsys direct communication

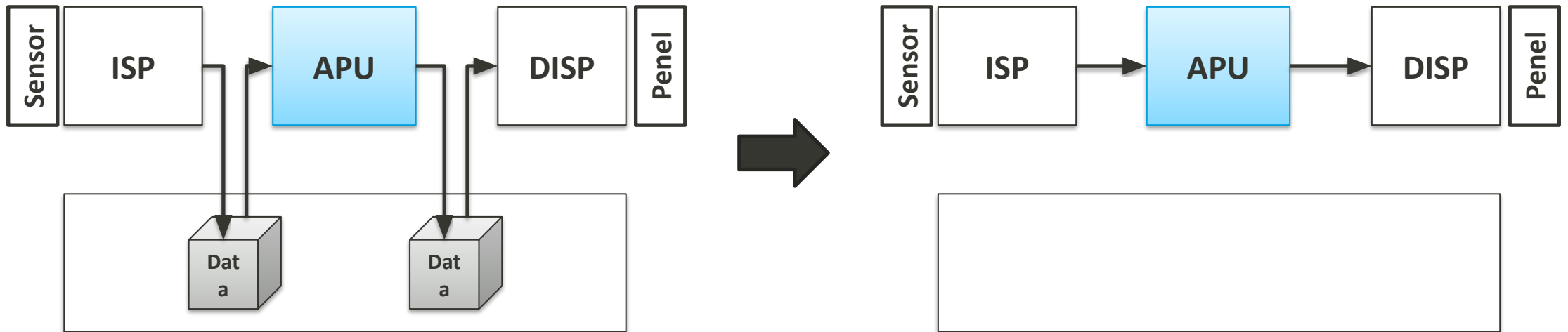
Direct control and data interface

APU Dram Bandwidth Reduction Techniques

- -65% bandwidth by Deeper Layer Fusion on 4K30 Video AINR

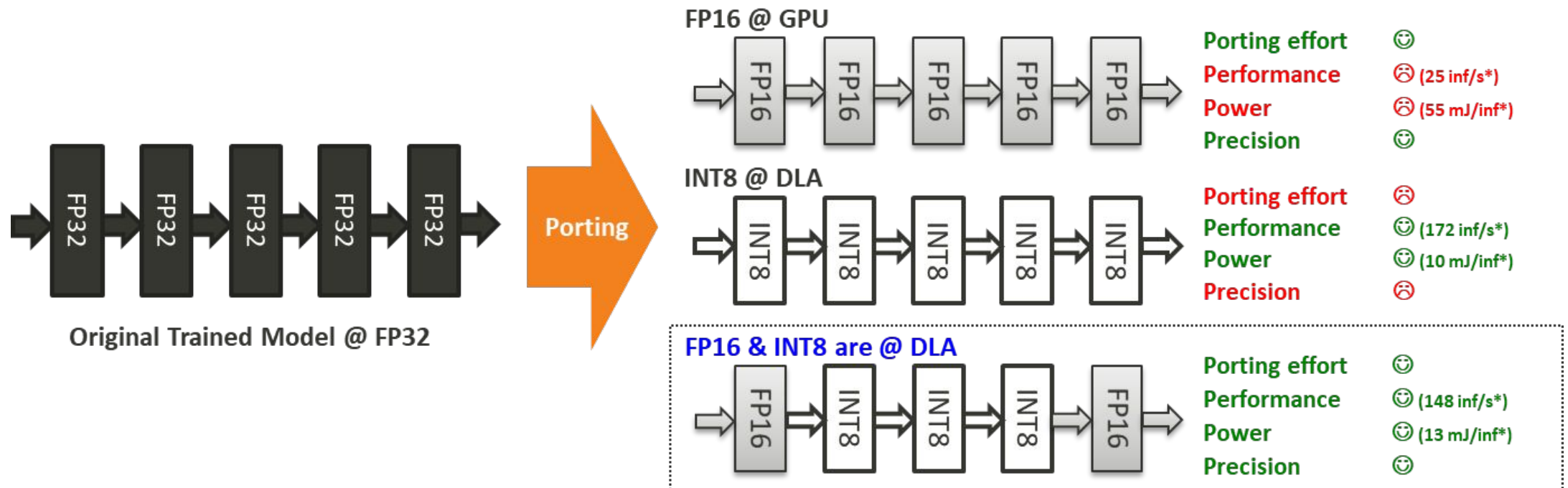


- -24% bandwidth by Tile-based Direct Link with ISP/DISP on FHD60 AISR



APU Data Type & Mixed Precision

- Optimal data type is application specific
 - e.g. 4/8-bit for classification, FP16 for speech
- Support mixed precision data type in one network architecture



**GET READY
FOR SOMETHING**

— **INCREDIBLE**
IN YOUR NEXT SMARTPHONE

MEDIATEK



Acknowledge

- Dimensionality 9000 teams for the Outstanding Work
- Co-authors
 - Arm: Stefan Rosinger, Saurabh Pradhan
 - Mediatek: Hugh Mair, Arthur Lin, Ted Lin