Dimensity 9000 – A Flagship Smartphone SoC

Presenter: Mediatek Ericbill Wang
Co-author: Arm Stefan Rosinger, Saurabh Pradhan
Brand: Dimensity  
Model: 9000  
Memory: 12GB+256GB  
OS: Android 12

1007396  
SOC: MT6983
MediaTek Dimensity 9000

**CPU**
- 1x Arm Cortex-X2 3.05GHz
- 3x Arm Cortex-A710 2.85GHz
- 4x Arm Cortex-A510 1.8GHz
- 8MB L3 + 6MB system-level cache

**APU**
- 4 Performance cores
- 2 Flexible cores

**Display**
- WQHD+ 144Hz / FHD+ 180Hz
- HDR10+ Adaptive

**Connectivity**
- 2x2 Wi-Fi 6E
- BT5.3 / Bluetooth LE audio
- GPS (L1 / L5), Galileo (E1 / E5a), Glonass, BeiDou (B1i / B1c / B2a), NavIC, QZSS

**GPU**
- Arm Mali-G710 MC10

**Memory**
- 4-channel LPDDR5x 7500Mbps

**Camera**
- 320MP capture
- 3-core 3-exp HDR-ISP
- 32MP + 32MP + 32MP triple cam
- 4K 3-exp video HDR
- AI-Video architecture

**Video**
- Decoder: 8K30, AV1 / VP9 / H.265 / H.264
- Encoder: 8K24, H.265 / H.264

**Modem**
- 5G Release 16
- DL 3CC 300MHz up to 7Gbps
- UL 2CC with R16 UL enhancement
- MediaTek 5G UltraSave 2.0

**TSMC 4nm**
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
Learnings on Tri-gear CPU Architecture - 2

- Energy Aware Scheduler
- Event Based Operating Frequency Decision

![Diagram of Tri-gear CPU Architecture]

- Task Wait Queue
- Energy Aware Scheduler
- Max-gear Run Queue
- Mid-gear Run Queue
- Min-gear Run Queue
- Per die power model
- Operating Frequency
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%)
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
Arm Mali-G710

- Bigger Core: 2X Texel & FMA rate
- Redesigned execution engine
- 1.2X performance density
Gaming Improvement vs. Dimensity 1200

Genshin Impact 60FPS

+20fps

D1200
D9000
AI Applications on Smartphone

Detection
Classification
Segmentation
Depth Estimation
Noise Reduction
Super Resolution
Speech Recognition
Speech Translation

Visual Perception
Object Construction
Image Quality
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 Dimensity 9000
Acknowledgements

- Dimensity 9000 teams for the Outstanding Work

- Co-authors
  - Arm: Stefan Rosinger, Saurabh Pradhan
  - Mediatek: Hugh Mair, Arthur Lin, Ted Lin