The All-Big-Core CPU Architecture of MediaTek Dimensity 9300

White Paper
March 2024

Table of Contents

List of Tables

List of Figures

1. Introduction

MediaTek premiered the first tri-gear CPU design with big, medium, and little CPU cores to the smartphone industry in 2017. With Arm's introduction of the Cortex-X CPU series, this has allowed flagship Android smartphones to adopt a tri-gear design with big (Cortex-X), medium (Cortex-A7), and little (Cortex-A5), reflecting the difference in architecture and performance capabilities of each. The little CPU cores have always existed in some capacity since the advent of Android smartphone SoC's, with the little (in-order execution) cores being used for low intensity tasks.

MediaTek is yet again revolutionized the design of smartphone SoCs with its-Latest Android smartphone SoC, the Dimensity 9300, which eschews the little CPU cores for the first time by adopting a pioneering 'All-Big-Core' design, in which all CPU cores are equipped with out-of-order execution pipelines. This industry-leading design harmoniously integrates the latest advancements in Arm CPU architecture with MediaTek's state-of-the-art power efficiency technologies. It positions MediaTek's SoC to exceed the expectations of the flagship Android smartphone market, delivering an unprecedented upgrade in performance and user experience. In this whitepaper, we explore the rationale behind MediaTek's All-Big-Core architecture, evaluate the results, and discuss the optimized software design necessary to fully capitalize on the unique advantages offered by this groundbreaking SoC.

2. Modern Mobile Application Behavior

The performance of flagship smartphones is mainly driven by two factors: mobile gaming and form factor. They are both pushing for greater performance with increasing thread level parallelism (TLP). For example, the game "Genshin Impact 60 - Sumeru", released in 2022, requires a TLP of over 5 for more than half of its gameplay duration. The fast-growing foldable phone market makes use of multi-window experiences, which require many cores to fully support a variety of multiple content combinations.

2.1. Modern mobile gaming experiences

The mobile gaming market is projected to achieve a remarkable CAGR of 15.5% from 2023 to 2032[1], and is expected to be the primary driver of increasing computing demand. Smartphones now boast display resolutions and frame rates that rival traditional gaming consoles; the adoption of 90+Hz display refresh rates has become a standard feature of premium devices, with Android phones reaching nearly 50% user preference share in Q2 2022.[2] This trend towards higher refresh rate, resolution, and more complex in-game visuals necessitates more CPU performance (shown in Figure 2 1). Games like "Genshin Impact" exemplify this trend by utilizing up to six parallel main threads for CPU-intensive tasks, alongside numerous auxiliary threads.

[1] Precedence Research. Mobile Gaming Market (By Platform: iOS, Android; By Age Group: Below 24 Years, 24-44 Years, Above 44 Years; By Business Model: Freemium, Paid, Free, Paymium) - Global Industry Analysis, Size, Share, Growth, Trends, Regional Outlook, and Forecast 2023-2032. July 2023

[2] ANTUTU benchmark. Global Users Preferences for Android Phones, Q2 2022. July 2022

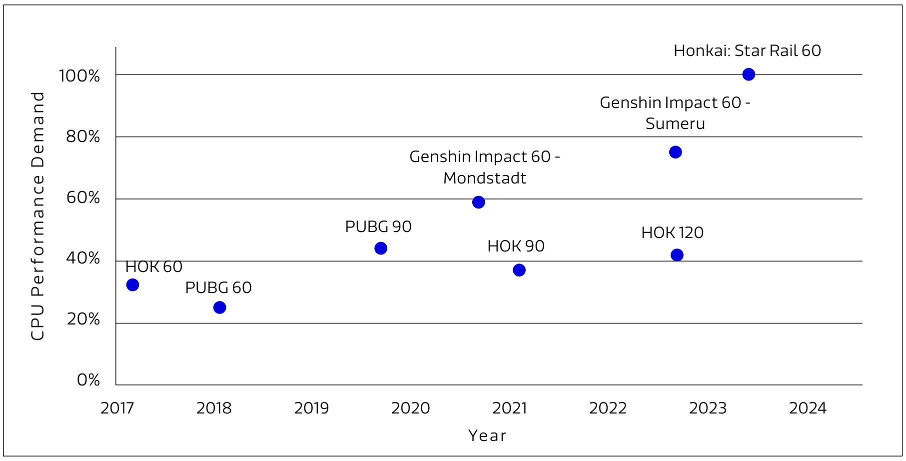

Figure 2.1. CPU performance demand trend for mobile gaming

"PUBG 90" has TLP remaining below 4 through almost 50% of its gameplay, however just three years later in 2022, "Genshin Impact 60 - Sumeru" required a TLP of over 5 during more than half of its gameplay duration (shown in Figure 2.2).

MediaTek's advanced All-Big-Core CPU architecture is designed to cater to these requirements, meeting the demands of high performance, while also being energy efficient.

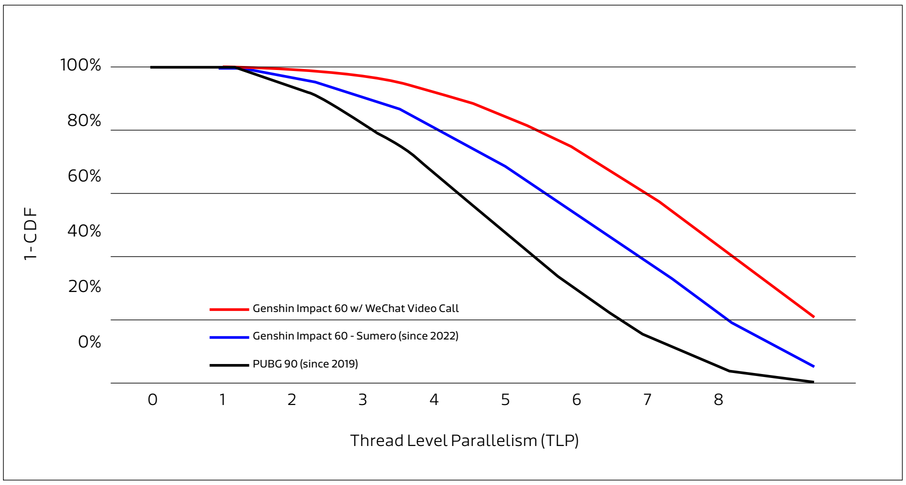

Figure 2.2. CPU TLP demand for mobile game trend and concurrent case based on Dimensity 9300

The TLP demand spikes even more in scenarios where users engage in concurrent tasks, such as gaming while on a video call. For instance, when playing "Genshin Impact 60" and simultaneously conducting a WeChat video call (demonstrated in Figure 2.3), a common multitasking scenario, the TLP demand exceeds 6 for 60% of the time (shown as the red line in Figure 2.2). This change signifies a substantial increase in the need for parallel processing capabilities. As users increasingly expect their devices to seamlessly run multiple tasks, the imperative for a CPU capable of meeting these rigorous TLP demands is clear.

Figure 2.3. Concurrent Case for Gaming (Genshin Impact) with Video Call (WeChat)

2.2. Supporting the foldable smartphone trend

Foldable devices are redefining the smartphone form factor, with a market increase of 62% within 2022[3]. The expanded screen real estate enhances the user experience by allowing for more detailed and immersive content. Multi-window capabilities enable users to multitask effectively, a feature that is particularly beneficial for productivity and entertainment. For instance, users can simultaneously watch a streaming video feed while posting comments on social media, as illustrated in Figure 2.4. With the rise of foldable devices, developers are focusing on optimizing apps to make use of the additional screen space. This shift underscores the need for CPUs capable of handling multiple high-performance tasks at once, which is ideally suited to MediaTek's unique CPU architecture, which excels in both performance and energy efficiency.

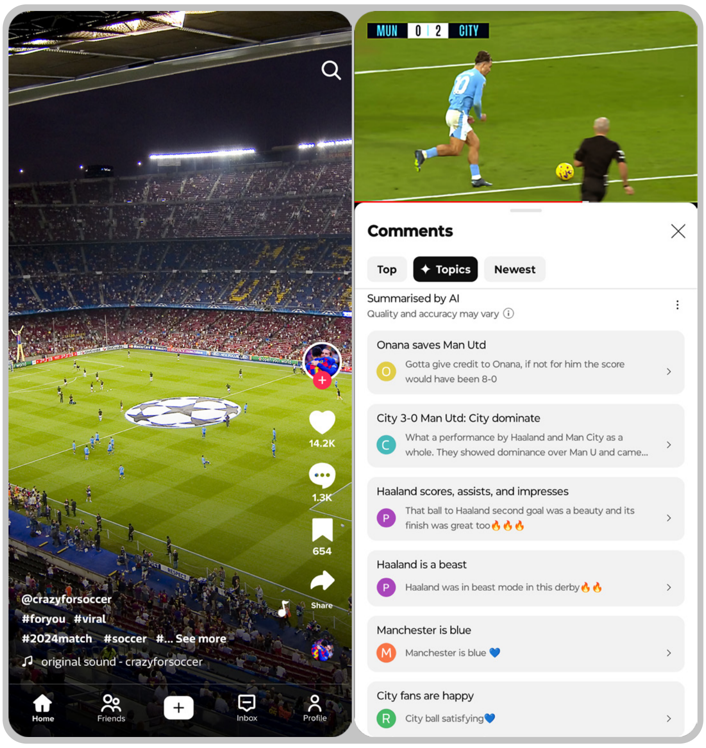

Figure 2.4. Multi-window use case in foldable device

[3] THE TIMES OF INDIA. Foldable phone market declines in the last three months of 2022: Report. Feb 2023

3. Evolution in hardware design

The trend in smartphone user experience drives the evolution of MediaTek SoC design, which has been committed to providing the best-in-class flagship 5G smartphone SoC since the introduction of the Dimensity series in 2019.

3.1. Tri-gear CPU architecture for best user experience

Table 3.1 lists MediaTek's flagship Dimensity CPU architecture designs before the Dimensity 9300. To avoid confusion, this whitepaper uses "gear" to refer to the classes of CPU cores in an SoC, such as min-gear, mid-gear, and max-gear.

Starting from the Dimensity 1200, a tri-gear CPU architecture is adopted, where the max-gear is an A-core with peak-clock speed targeting short-burst performance, the mid-gear is for sustainable gaming, and the min-gear for maximal number of DoU (day of use).

The next generation Dimensity 9000 introduced two major changes: firstly, the CPU ISA migrated from Armv8.2 to new Armv9.0 which improves security (e.g., PAC, BTI, MTE) and ML compute capability (e.g., BF16, MATMUL); secondly, it used the Cortex-X series CPU core for the max-gear to keep up with competitors in single core performance. The Dimensity 9200 followed the same CPU architecture design, but with Arm's latest annual CPU IP upgrade.

Table 3.1. MediaTek flagship smartphone CPU architectures before Dimensity 9300

	Dimensity 1000	Dimensity 1200	Dimensity 9000	Dimensity 9200
CPU ISA	Armv8.2	Armv8.2	Armv9.0	Armv9.0
Max-gear		Cortex-A78 x1	Cortex-X2 x1	Cortex-X3 x1
Mid-gear	Cortex-A77 x4	Cortex-A78 x3	Cortex-A710 x3	Cortex-A715 x3
Min-gear	Cortex-A55 x4	Cortex-A55 x4	Cortex-A510 x4	Cortex-A510 x4
L3$ Size	2MB	2MB	8MB	8MB

3.2. The All-Big-Core CPU architecture

Although the tri-gear CPU architecture helps achieve an optimal user experience, the specific core design of the gears is critical to the design goal of achieving better performance with less power.

During this time we observed the effective range of the min-gear, with a basic in-order CPU core architecture become smaller and smaller through each flagship SoC generation. In the Dimensity 1200, as shown in Figure 3.1, the min-gear only outperforms the mid-gear power efficiency in the sub one third of the graph. Moreover, in the Dimensity 9000, as shown in Figure 3.2, the mid-gear is always more efficient than the min-gear in this implementation. There are two major reasons for this trend: Firstly, Arm v9 in-order CPU (CortexA510) boosts IPC by 35%, but at the cost of power efficiency; Secondly, MediaTek keeps pushing higher frequencies at the lowest possible operating voltage via implementation methodology and circuit technology, which improves mid-gear power efficiency significantly.

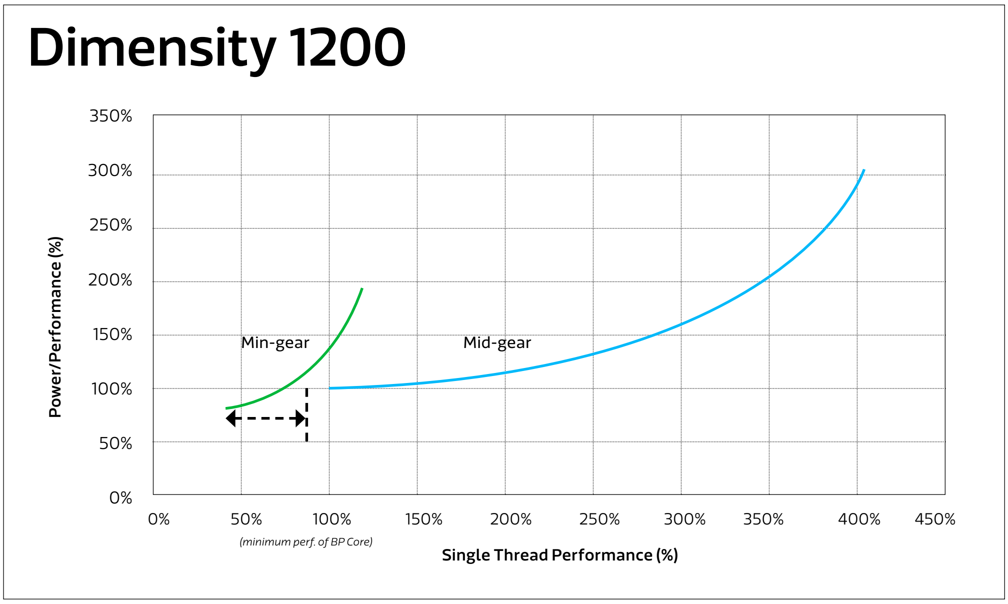

Figure 3.1. Dimensity 1200 CPU power efficiency

This observation triggered the design of an All-Big-Core architecture, where the mid-gear out-of-order cores replaces the in-order cores for min-gear CPU, with comparable power efficiency in the low performance range. The All-Big-Core design extends the usable range of the min-gear CPU, and it also simplifies both the cluster (DSU) design and the energy aware scheduler. Given the min-gear now provides a wider usable performance range, we further adopted power optimized Cortex-X CPUs for the mid-gear to achieve a better high frame rate gaming experience.

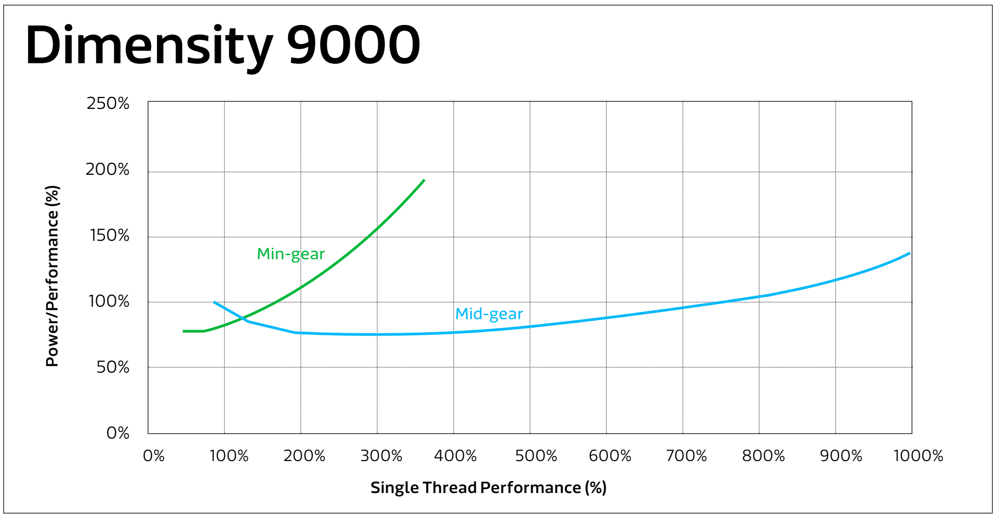

Figure 3.2. Dimensity 9000 CPU power efficiency

4. Evaluation results

In this section, we give both our evaluation results of the All-Big-Core architecture, and the community's third-party comparative analysis.

4.1. Power efficiency curves of the Dimensity 9300

Figure 4.1 shows the min-gear CPU power efficiency curve comparison between the Dimensity 9200 and 9300. Using SpecInt2k6, at the same performance point the Dimensity 9300 min-gear CPU observes a 60% lower power, while also capable of scaling to provide 3X peak performance vs. Dimensity 9200.

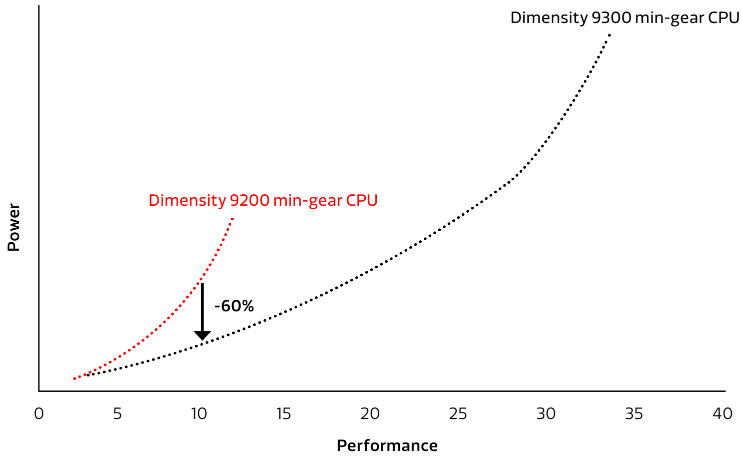

Figure 4.1. Min-gear CPU power efficiency curve comparison between Dimensity 9300 and 9200

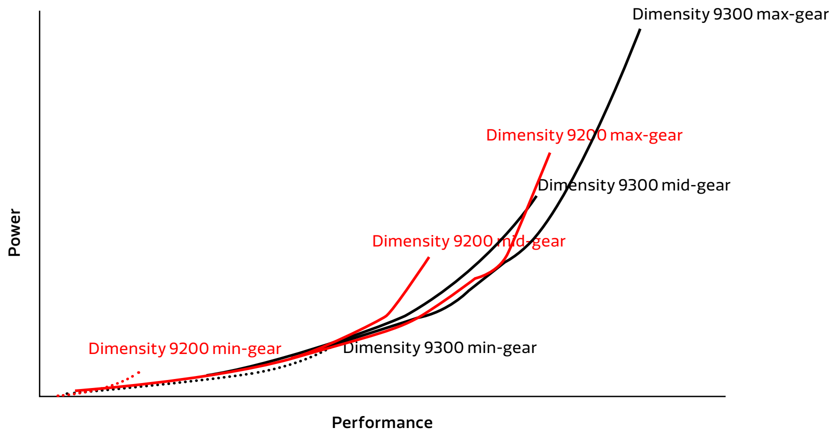

Figure 4.2. Multi-gear CPU power efficiency and performance curve comparison between Dimensity 9300 and 9200

Figure 4.2 shows the multi-gear CPU power efficiency curve comparison between Dimensity 9300 and 9200. The Dimensity 9300 has seven cores that can serve a performance range between 10 to 35 in SpecInt2k6, while the 9200 has only four. The additional cores delivering the 10 to 35 performance range can benefit more concurrent workloads.

Table 4.1 shows the power usage comparison between Dimensity 9300 and 9200 in key user scenarios, where the Dimensity 9300 observes about 10% lower power. In conclusion, the All-Big-Core architecture can provide higher multi-thread performance without sacrificing the value of DoU (Day of Use).

Table 4.1. key scenarios power (mA) comparison between Dimensity 9300 and 9200

Category	Scenario	Dimensity 9300 1st All-big-core (Measurement)	Dimensity 9200 (Measurement)
Basic	Flight Mode Suspend	5.0	5.7
	Home Screen Idle (CMD)	11.0	12.7
	MP3 (ADSP offload)	22.0	25
	Video - H.264 (1080P, 30FPS) HW mode	80	90
Camera	Capture Preview (1080P, 30FPS)	251	275
	Video Record 1080P EIS 30fps	308	331
	Video Record 4K EIS 30fps	386	428
	Video Record 4K EIS 60fps	625	707
Game	*Genshin 60 (fps/mA) *latest heavy scene	59.9/1231	56/1259

4.2. Third-party comparative analysis: Dimensity 9300 versus direct flagship competitor platform

Media reviewer, 'Geekerwan', published a review[4] comparing the Dimensity 9300 to the Qualcomm Snapdragon 8 Gen 3, and the Apple A17 Pro SoCs.

In this independent evaluation, the Dimensity 9300 achieved a greater Geekbenchv5 multi-core score than both competitor SoCs. The scores (using Apple A17 Pro as the baseline):

- MediaTek Dimensity 9300: 7368 (116%),
- Qualcomm Snapdragon 8 Gen 3: 6782 (107%),
- Apple A17 Pro: 6342 (100%).

The report also detailed the power efficiency curves of both the Dimensity 9300 and Snapdragon 8 Gen 3, where X-axis is total power in Watt and Y-axis is Geebnechv5 multi-core score as shown in Figure 4.3. Here we can see the Dimensity 9300 delivered greater multi-core performance at the same power consumption.

[4] Geekerwan (online media). MediaTek Dimensity 9300 Review. 2023.11.7

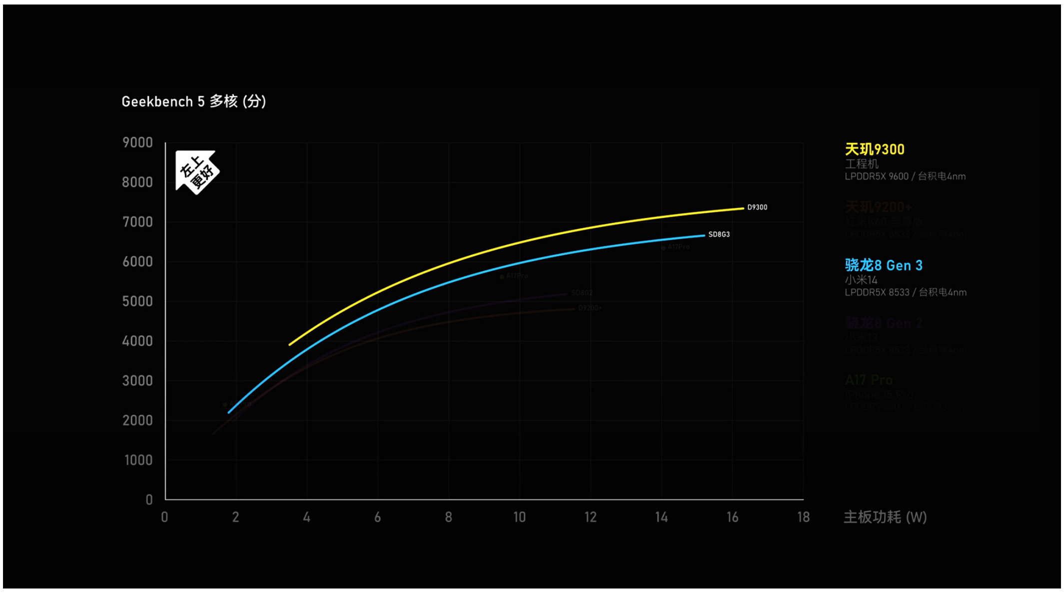

Figure 4.3. Geekbenchv5 multi-core: Dimensity 9300 achieves better power efficiency than Qualcomm Snapdragon 8 Gen 3

In the evaluation of heavy-loading games, such as Genshin Impact, the Dimensity 9300 continues to outperform Snapdragon 8 Gen 3, achieving a slightly higher frame rate while also at lower power consumption.

- MediaTek Dimensity 9300: 59.7fps @ 5.5W,
- Qualcomm Snapdragon 8 Gen 3: 59.2fps @ 5.7W

The runtime frame rate over time is given in Figure 4.4. As you can see the MediaTek SoC (orange line) provides a more consistent framerate with significantly fewer large dips (micro-stutter), which equates to a better user experience

Figure 4.4. Game Genshin Impact: Dimensity 9300 - has better FPS and power than Snapdragon 8 Gen 3

The results cross-verify the design goal of All-Big-Core: achieving better, more consistent performance with less power.

5. Requirements to software design

The All-Big-Core CPU architecture provides a robust hardware foundation for an optimized user experience. Certain software scenarios directly reap the benefits of this architecture without necessitating any modifications within Android. Prime examples include the execution of the dex2oat compilation thread and the zram compression thread.

The dex2oat thread is responsible for converting application code from the dex format into executable binaries, thereby enhancing the performance of application code execution. Meanwhile, the zram thread compacts memory pages that are not in active use, freeing up valuable memory space. Both tasks are typically considered non-critical by the system scheduler and are often allocated to min-gear cores to economize power consumption. However, their execution can indirectly affect the responsiveness of foreground applications. With the All-Big-Core architecture, this impact is minimized as these "non-urgent" tasks are executed more swiftly.

Nevertheless, to truly tap into the All-Big-Core hardware's full potential, systematic and thoughtful software design is essential to find the perfect balance between power consumption and performance. It is crucial that the CPU cores transition into sleep mode promptly and deeply whenever feasible.

As a smartphone processes numerous tasks simultaneously, originating from applications, the system framework, and platform drivers, the software design principles are as follows:

- For **single task**, a core should activate swiftly upon a task scheduled onto it and enter sleep mode rapidly once no active tasks remain.
- For **multiple tasks**, grouping them should be implemented to prevent repetitive activation/sleep cycles.
- For **interrupts**, the system ought to minimize the number of IRQs to the application processor and consider offloading them to an internal dedicated microprocessor.

In addition to these principles, conventional best practices in tri-gear CPU scheduling remain relevant. For instance, tasks for application's immediate responsiveness are typically assigned to max-gear cores for prompt responsiveness.

5.1. Closed-loop scheduling in system software design

To align with software design principles, a closed-loop approach in Android system software is favored. This closed loop can be primarily broken down into three functional components, as shown in Figure 5.1:

- **Budget/Demand Identification**: This component is responsible for discerning the application behavior, setting performance targets, and evaluating resource requirements.
- **Task/Resource Management**: This component involves mapping tasks across the platform's resources and dynamically modulating resource attributes, such as operational frequencies.
- **Thermal/Battery Regulation**: The focus here is on dynamically overseeing system operations and resource attributes within the constraints imposed by thermal conditions and battery life.

To execute closed-loop scheduling, the Android system must undergo modifications to enable the semantic information flow between applications and the hardware platform.

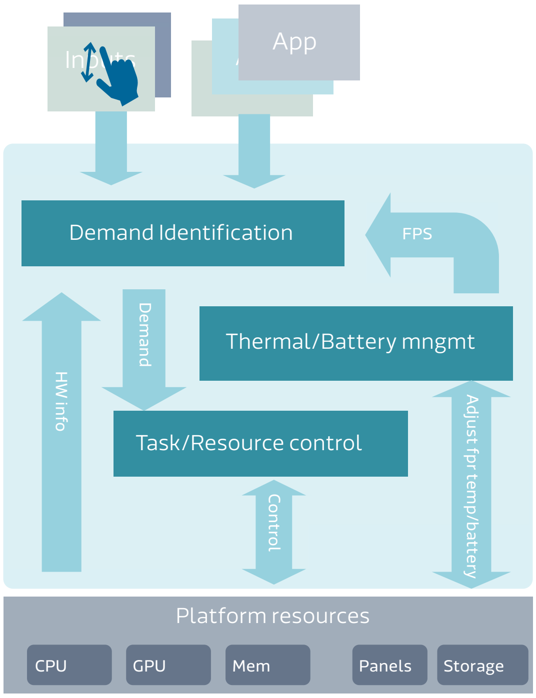

Figure 5.1. Closed-loop scheduling in the system

5.2. Software supports to closed-loop scheduling

There are several key requirements for the Android system to facilitate closed-loop scheduling:

- **Performance Targeting**: Applications may explicitly state their performance objectives, such as frame rate or execution time budget. This information is vital for the scheduler to allocate tasks effectively.
- **Critical-path identification**: The system needs to identify critical path tasks, e.g., those pivotal to frame generation. This allows the scheduler to prioritize these tasks appropriately and prevent them from being blocked by less urgent tasks.
- **Task Alignment**: The scheduler must support mechanisms to align tasks with an event that indicates when a new frame starts, like the vsync signal.
- **Task Grouping**: To minimize CPU wakeup times, the system should support task grouping across multiple frames, provided the upcoming tasks for subsequent frames are known in advance.

MediaTek has committed considerable efforts towards system enhancements. Given the breadth of support needed for software, ranging from application development to system-level optimizations, collaboration within the Android community is critical to evolve these capabilities.

6. Summary

The groundbreaking All-Big-Core CPU architecture deployed in MediaTek's Dimensity 9300 represents a confluence of the latest developments in Arm CPU architecture and MediaTek's power efficiency innovations. MediaTek has created an SoC that goes beyond all market expectations, providing a user experience that not just accelerates mobile applications, but creates a milestone of performance for future apps and experiences to leverage.

To fully exploit MediaTek's All-Big-Core CPU architecture, refinements in software system design are imperative to achieve an ideal equilibrium between power efficiency and performance. The crux lies in enabling the processor cores that can enter deep sleep states swiftly and efficiently whenever the opportunity arises. Fostering collaboration within the Android community is crucial for augmenting these competencies.