

MOBILE CHIP *report*

Insightful Analysis of Mobile System Design

MALI-G71 ENABLES COHERENT COMPUTING

ARM Builds New High-Performance GPUs on Bifrost

By Mike Demler (May 30, 2016)

ARM's Mali-G71 extends the Norse-mythology lineage of its predecessors—Midgard and Utgard—by using the new Bifrost architecture to launch its attack on the realm of high-performance GPUs. The G71 (previously code-named Mimir) implements an extensively redesigned microarchitecture that supports full hardware cache-coherent operation for tightly coupled CPU/GPU workloads. It also adds TrustZone support, ensuring content protection throughout the media-processing pipeline.

Mobile-processor vendors are increasingly adopting Mali GPUs to complement their Cortex-A CPUs. Mali's strength has been in the low-cost smartphone and tablet markets, where many vendors employ older Utgard GPUs such as the Mali-400. But in the past year, the newer Midgard Mali has won significant designs in both the high-performance and midrange GPU segments. For example, Samsung uses a massive 12-core Mali-T880 in the Exynos 8 chip that powers some versions of its flagship Galaxy S7 phone (see [MCR 12/7/15](#), "Exynos 8 Integrates Modem"). Huawei and Spreadtrum use quad-core versions of the GPU in the Kirin 950 and SC9860, respectively (see [MCR 11/23/15](#), "Kirin 950 Takes Performance Lead," and [MCR 3/14/16](#), "Spreadtrum Moves Up to Midrange").

With its new Bifrost design, ARM aims to gain more share in the high-end graphics segment and to address the requirements of performance-intensive tasks such as mobile gaming, augmented reality (AR), and virtual reality (VR). To enable these applications, the Mali-G71 GPU can scale from 4 to as many as 32 shader cores—twice the maximum number in the current Mali performance leader, the T860/880 (see [MCR 11/24/14](#), "Mali-T800 Boosts ARM GPUs"). It's also a more efficient design, requiring 20% less power and 40% less die area than the T880 under the same process and performance conditions.

Smartphone power budgets are likely to restrict G71-equipped processors to 16 shader cores, but tablets could use the maximum configuration. Mali is also popular in digital TVs, where ARM estimates it has approximately a 75% market share. The G71 can drive UltraHD displays, which require 120Hz screen-refresh rates at 4K resolution. It also supports the OpenCL 2.0 API for heterogeneous CPU/GPU computing, as well the Khronos Group's OpenGL ES 3.x and new Vulkan 3D-graphics API (see [MPR 4/6/15](#), "Graphics APIs Reach a New Low").

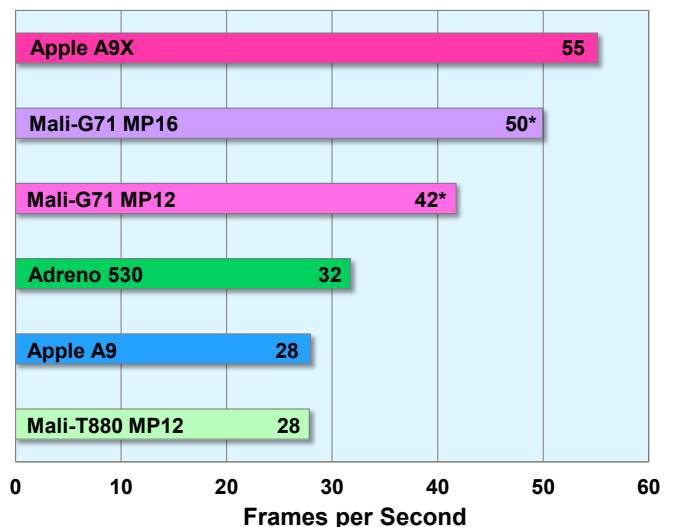


Figure 1. Mobile-GPU performance comparison. ARM's new Mali-G71 offers performance exceeding that of the current top smartphone GPU. It scales to as many as 32 shader cores, enabling it to also compete with the highest-performing tablet GPUs. (Source: Kishonti GFXBench 3.0 Manhattan test scores, except *ARM estimates)

Price and Availability

Production RTL for the Mali-G71 is currently available for licensing. ARM does not disclose its license fees. For more information, access www.arm.com/products/multi-media/mali-gpu/index.php.

The Mali-G71 is now available for licensing; in fact, we believe early-access customers have had the RTL since at least mid-2015. ARM expects the first application processors employing the core to reach production in 4Q16—likely in time for Samsung’s Galaxy S8, among others.

Bridging the Gap

In Norse mythology, Bifrost is the bridge connecting the human world (Midgard) to the world of the gods (Asgard). Although the G71’s predecessor, the Mali-T880, is already a strong competitor in the premium smartphone segment, a large gap resides between Midgard GPUs and the highest-performing tablet GPU—the Imagination-sourced PowerVR Series7XT in Apple’s A9X processor. A comparison of GFXBench scores on the Manhattan 3.1 Offscreen 1080p test shows that the Mali-G71 can bridge that gap.

Kishonti’s GFXBench web site reports that the Samsung Galaxy S7 scores 28fps on that test, matching Apple’s A9 processor in the iPhone SE, which uses a six-core PowerVR Series7XT GPU. In smartphones, those scores are exceeded only by the Adreno 530 GPU in Qualcomm’s Snapdragon 820 processor (serving in some versions of the Galaxy S7), which scores 32fps on the Manhattan test, as Figure 1 shows.

In 2017, we expect to see the first mobile processors manufactured in 10nm CMOS technology. ARM estimates that a 10nm Mali-G71 MP12 GPU, configured to run in the same power envelope as the current-generation 16nm Mali-T880, will deliver a 50% performance boost. The new process technology and GPU architecture will raise the smartphone bar to about 42fps. The architectural improvements include a new shader design and optimizations in

the geometry-data flow, which we will cover in detail in a forthcoming article on Bifrost.

In tablets, Apple’s A9X iPad Pro GPU takes advantage of the bigger power budget to run 12 PowerVR Series7XT graphics cores, achieving 55fps on GFXBench. According to ARM’s hardware simulations in a 16nm process, a 16-core Mali-G71 comes within 10% of Apple’s premium tablet processor, achieving approximately 50fps.

Building a Heterogeneous System

Applications such as augmented reality, computational photography, and image recognition are driving greater demand for general-purpose-GPU (GPGPU) computing. The OpenCL 2.0 API enables software written for such applications to take advantage of heterogeneous CPU/GPU computing with shared coherent virtual memory. The new Vulkan 3D-graphics API mandates coherent shared memory.

Before Bifrost, Mali GPUs provided only partial GPU/CPU coherence. The Midgard cores integrate Amba Ace-Lite interfaces that enable I/O coherence, which allows the GPU to snoop the CPU cache; the GPU’s cache can’t be snooped, however. The benefit of I/O coherence is it allows the GPU to read graphics commands and vertex data directly from the CPU cache, reducing external-memory transactions and eliminating unnecessary flushing of the CPU cache after every frame. I/O coherence cuts the power and processing time that result from the cache flushing and memory transactions.

As Figure 2 shows, Bifrost GPUs upgrade to the full Ace interface, so the GPU can act as a cache-coherent master peer to the CPUs. Thanks to full cache coherence, the CPU and GPU MMUs map virtual addresses to identical physical addresses, thus creating a single shared memory space. The coherent memories eliminate the need to manually copy data between cores, enabling programmers using OpenCL or Vulkan to create more tightly coupled applications that shift the CPU/GPU workload as needed. This approach simplifies programming and further reduces power and performance bottlenecks associated with accessing external DRAM.

ARM’s cache-coherent system employs the CoreLink cache-coherent interconnect (CCI) to link the CPU, GPU, and any I/O-coherent masters. The CCI-550 is a 128-bit bus that supports up to six Ace interfaces (see [MPR 11/9/15](#), “ARM CoreLink Snoops Six Aces”). The CoreLink network interconnect (NIC) links a Mali video processor (VPU) and display processor (DPU) into the memory system to access TrustZone-protected external DRAM.

The CCI-550 integrates the central snoop directory, which reduces power consumption by limiting snoop traffic to only those caches that contain copies of the data being updated. According to ARM’s measurements, the Mali-G71’s microarchitecture improvements and cache-coherent system hardware reduce run time for heterogeneous operations by up to 90% compared with a software-based system.

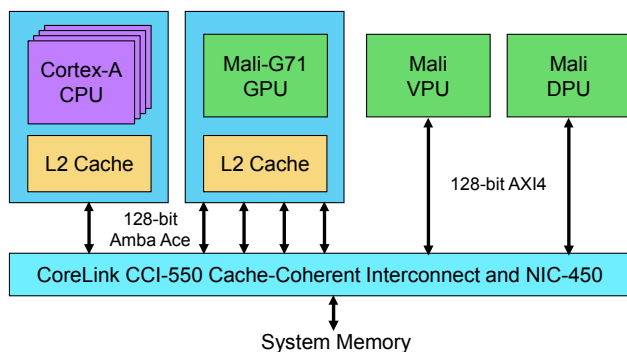


Figure 2. Block diagram of ARM CPU/GPU cache-coherent system. The Mali-G71 adds support for the Ace interface, allowing it to operate with full CPU/GPU cache coherence.

Challenging at the High End

ARM's success in developing low-power CPU cores for the mobile market has paved the way for attracting licensees to its other processor cores. Owing to its recent acquisition of Apical, which brings image-signal-processor (ISP) and computer-vision IP, the company now offers a complete one-stop shop for multimedia-processor designers.

Growth in Mali shipments has been particularly dramatic, as Figure 3 shows. Just five years ago, ARM graphics cores shipped in fewer than 50 million devices, but that number exploded in 2015 to 750 million—more than a 35% increase from the prior year. We estimate Mali shipments comprised approximately 50% of the licensable-GPU market last year, surpassing those from second-place Imagination by 9%.

Nevertheless, Mali still has plenty of room to grow. Until recently, the high-end mobile-GPU market remained a stronghold for Imagination's PowerVR licensable cores and Qualcomm's in-house Adreno GPU; the new Mali-G71 enables licensees to achieve performance levels high enough to match or even beat those processors.

ARM designed Bifrost to target emerging applications for AR and VR headsets, many of which suffer from the screen-door effect that exposes the pixel-density limitation of their displays. Because it supports 120Hz refresh rates and 4K resolution, the Mali-G71 is a good solution for that market. It also has a more efficient pixel-processing pipeline that reduces latency between the user input and display to less than four milliseconds, improving game response.

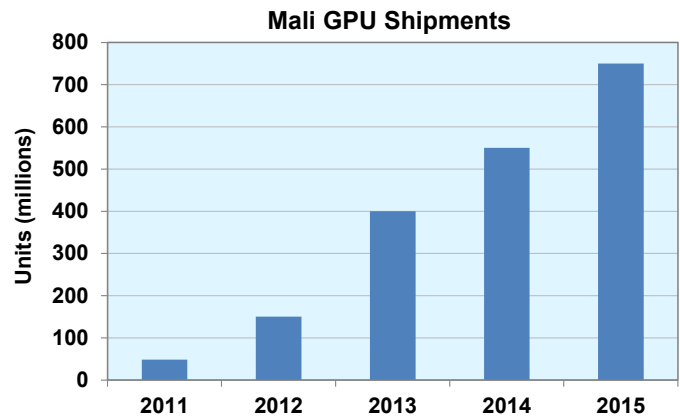


Figure 3. Shipments of ARM Mali GPUs. Since 2011, the company has increased Mali shipments from fewer than 50 million to 750 million. (Source: ARM)

By supporting heterogeneous CPU/GPU computing, the Bifrost architecture will enable developers to better balance their power budgets for high-performance graphics tasks. OpenCL 2.0 capability allows the company to combine Cortex-A CPUs with Mali GPUs to implement deep neural networks for embedded-vision applications, advancing its efforts to gain share in the advanced driver-assistance systems (ADAS) market. Bifrost is more configurable and scalable than previous Mali architectures, and we expect ARM will use it to develop a wide range of area-, power-, and performance-optimized versions that will simplify its catalog of graphics cores to a single family. ♦

To subscribe to *Mobile Chip Report*, access www.linleygroup.com or phone us at 408-270-3772.