

Naming	Mali-G71	Mali-G72	Mali-G31	Mali-G51	Mali-G52	Mali-G76	Mali-G57	Mali-G77	Mali-G78	Mali-G710	Mali-G510	Mali-G310	Mali-G715	Immortalis-G715	Mali-G720	Immortalis-G720
Architecture	Bifrost						Valhall								5 <sup>th</sup> Gen	
API Support	Mali-G71	Mali-G72	Mali-G31	Mali-G51	Mali-G52	Mali-G76	Mali-G57	Mali-G77	Mali-G78	Mali-G710	Mali-G510	Mali-G310	Mali-G715	Immortalis-G715	Mali-G720	Immortalis-G720
OpenGL ES 1.1 – 3.1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
OpenGL ES 3.2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Vulkan 1.0 – 1.3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
OpenCL 1.0, 1.1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
OpenCL 1.2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
OpenCL 2.0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
OpenCL 2.1					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
OpenCL 3.0									✓	✓	✓	✓	✓	✓	✓	✓
Core Features	Mali-G71	Mali-G72	Mali-G31	Mali-G51	Mali-G52	Mali-G76	Mali-G57	Mali-G77	Mali-G78	Mali-G710	Mali-G510	Mali-G310	Mali-G715	Immortalis-G715	Mali-G720	Immortalis-G720
ASTC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
AFBC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
AFBC – RGBA16										✓	✓	✓	✓	✓	✓	✓
AFRC											✓	✓	✓	✓	✓	✓
Shader framebuffer access	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Multiple Render Target <sup>[1]</sup>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2xMSAA	Automatically promoted to 4xMSAA														✓	✓
4xMSAA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
8xMSAA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
16xMSAA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
8-bit integer dot product	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
FP16 / R11G11B10 accelerated blending <sup>[5]</sup>							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Descriptor indexing							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Conservative rasterization													✓	✓	✓	✓
Variable Rate Shading													✓	✓	✓	✓
Ray tracing													○	✓	○	✓
Microarchitecture Features	Mali-G71	Mali-G72	Mali-G31	Mali-G51	Mali-G52	Mali-G76	Mali-G57	Mali-G77	Mali-G78	Mali-G710	Mali-G510	Mali-G310	Mali-G715	Immortalis-G715	Mali-G720	Immortalis-G720
Transaction elimination	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hidden surface removal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IDVS geometry pipeline	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DVS geometry pipeline															✓	✓

Core Config	<a href="#">Mali-G71</a>	<a href="#">Mali-G72</a>	<a href="#">Mali-G31</a>	<a href="#">Mali-G51</a>	<a href="#">Mali-G52</a>	<a href="#">Mali-G76</a>	<a href="#">Mali-G57</a>	<a href="#">Mali-G77</a>	<a href="#">Mali-G78</a>	<a href="#">Mali-G710</a>	<a href="#">Mali-G510</a>	<a href="#">Mali-G310</a>	<a href="#">Mali-G715</a>	<a href="#">Immortalis-G715</a>	<a href="#">Mali-G720</a>	<a href="#">Immortalis-G720</a>
Arithmetic units	3	3	1/2	2/3	2/3	3	2	2	2	4	3/4	1-4	4	4	4	4
Warp width	4	4	4	4	8	8	16	16	16	16	16	16	16	16	16	16
Thread count (max)	384	384	256/512	512/768	512/768	768	1024	1024	1024	2048	1536-2048	512-2048	2048	2048	2048	2048
FP16 operations/cycle	48	48	16/32	32/48	64/96	96	128	128	128	256	192-256	64-256	512	512	512	512
FP32 operations/cycle	24	24	8/16	16/24	32/48	48	64	64	64	128	96-128	32-128	256	256	256	256
Fragments/cycle	1	1	1/2	2	2	2	2	2	2	4	4	2-4	4	4	4	4
Pixels/cycle	1	1	1/2	1/2	2	2	2	2	2	4	4	2-4	4	4	4	4
Texels/cycle	1	1	1/2	1/2	2	2	4	4	4	8	4-8	2-8	8	8	8	8
Load/store cache size (bytes)	16K	16K	4K	16K	16K	16K	16K	16K	16K	32K	16K	8K-16K	32K	32K	32K	32K
Texture cache size (bytes)	8K	8K	16K	16K	16K	32K	32K	32K	32K	32K	32K	16K-32K	32K	32K	32K	32K
Tile bits/pixel <sup>[2]</sup>	128	256	256	256	256	256	256	256	256	256	256	256	256	256	256	256

Texturing	<a href="#">Mali-G71</a>	<a href="#">Mali-G72</a>	<a href="#">Mali-G31</a>	<a href="#">Mali-G51</a>	<a href="#">Mali-G52</a>	<a href="#">Mali-G76</a>	<a href="#">Mali-G57</a>	<a href="#">Mali-G77</a>	<a href="#">Mali-G78</a>	<a href="#">Mali-G710</a>	<a href="#">Mali-G510</a>	<a href="#">Mali-G310</a>	<a href="#">Mali-G715</a>	<a href="#">Immortalis-G715</a>	<a href="#">Mali-G720</a>	<a href="#">Immortalis-G720</a>
Bilinear samples/cycle	1	1	1/2	1/2	2	2	4	4	4	8	4-8	2-8	8	8	8	8
Trilinear filtering	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2
Nx anisotropic filtering <sup>[3]</sup>	N/A	xN <sup>(4)</sup>	xN	xN <sup>(4)</sup>	xN	xN	xN	xN	xN	xN	xN	xN	xN	xN	xN	xN
Depth format without reference	x1	x1	x1	x1	x1	x1	x1	x1	x1	x1	x1	x1	x1	x1	x1	x1
Depth format with reference	x1	x1	x1	x1	x1	x1	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2
Data size over 32 bits/texel	x1	x1	x1	x1	x1	x1	x2	x2	x2	x2	x2	x2	x2	x2	x1 <sup>(6)</sup>	x1 <sup>(6)</sup>
ASTC without EXT_decode_mode	x1	x1	x1	x1	x1	x1	x2	x2	x2	x2	x2	x2	x2	x2	x1 <sup>(6)</sup>	x1 <sup>(6)</sup>
3D format with linear filtering	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2
N channel 32bit/channel format with linear filtering	x4N	x4N	x4N	x4N	x4N	x4N	xN	xN	xN	xN	xN	xN	xN	xN	xN	xN
N plane YUV format	xN	xN	x1	x1	x1	x1	x2	x2	x2	x2	x2	x2	x2	x2	x2	x2

Bifrost ISA Config	<a href="#">Mali-G71</a>	<a href="#">Mali-G72</a>	<a href="#">Mali-G31</a>	<a href="#">Mali-G51</a>	<a href="#">Mali-G52</a>	<a href="#">Mali-G76</a>
Thread count (max)	384	384	256/512	512/768	512/768	768
Max work registers (32b)	64	64	64	64	64	64
Thread count with 0-32 work registers	384	384	256/512	512/768	512/768	768
Thread count with 33-64 work registers	384	384	128/256	256/384	256/384	384

Valhall ISA Config	<a href="#">Mali-G57</a>	<a href="#">Mali-G77</a>	<a href="#">Mali-G78</a>	<a href="#">Mali-G710</a>	<a href="#">Mali-G510</a>	<a href="#">Mali-G310</a>	<a href="#">Mali-G715</a>	<a href="#">Immortalis-G715</a>
Thread count (max)	1024	1024	1024	2048	1536-2048	512-2048	2048	2048
Max work registers (32b)	64	64	64	64	64	64	64	64
Thread count with 0-32 work registers	1024	1024	1024	2048	1536-2048	512-2048	2048	2048
Thread count with 33-64 work registers	512	512	512	1024	768-1024	256-1024	1024	1024

5 <sup>th</sup> Gen ISA Config	<a href="#">Mali-G720</a>	<a href="#">Immortalis-G720</a>
Thread count (max)	2048	2048
Max work registers (32b)	64	64
Thread count with 0-32 work registers	2048	2048
Thread count with 33-64 work registers	1024	1024

1. OpenGL ES has 4 render targets and Vulkan 8
2. Tile storage per pixel may be able to exceed this, but with reduced tile size. Theoretical limit is higher from Mali-G710 onward, but 256 is recommendation
3. Worst-case anisotropic filtering performance with a MAX\_ANISOTROPY = N
4. Mali-G72 r0p3 / Mali-G51 r1p1 or higher required
5. All have float blending. Valhall adds hardware acceleration for standard blend operations
6. Only fp16 and UNORM10 formats fully achieve x1

## **This reference sheet covers from the Bifrost Mali-G71, to 5th Gen GPUs up to Immortalis-G720.**

The API Support, Core Features and Microarchitecture Features tables cover which GPUs support which technologies. For more on given technologies see links below.

The Core Config table details the specs of the chips, rather than just whether features are available. As such for each GPU it has threads in a warp, total threads, and operations/ texels etc per clock cycle, as well as cache sizes. Note that for tile write rate on Arm chips this is both fragments written into the tile and the pixels written back out of the tile. Thread count is the total shader core hardware capacity; note that for OpenGL ES only 128 threads are exposed. For Mali-G310 and Mali-G510 Core Config has ranges depending on implementation – please check with device manufacturer for exact specification.

For Texturing, to work out cycles/sample for more complicated filters than bilinear, apply the multiplications in the tables on top of the bilinear performance to combine to the required filter. Remember to invert the bilinear samples/cycle to get cycles/sample. For example, a simple trilinear will be 2 x 1 cycles/sample on a Mali-G72, and 2 x 0.25 cycles/sample on a Mali-G77. To add in 4x anisotropic filtering, multiply by a further 4x. Note that anisotropic filter scaling is the worst-case number caused by the maximum number of sample taps, it will usually be less than this. Texture performance will differ from Image performance. Depth performance with/without reference refers to e.g., a shadow sampler with reference comparison returning a weighted bool vs a normal sample returning the actual depth value.

Finally, the architecture-specific tables give thread counts and registers for the chips. For more on the generations of Arm architectures see links below.

### **For a general picture of Arm GPU architectures see:**

+ [Arm GPU Architectures](#)

### **Specific architecture pages:**

- + [Bifrost \(Mali-G71 – Mali-G76\)](#)
- + [Valhall \(Mali-G57 – Immortalis-G715\)](#)
- + [5th Gen \(Mali-G720 – Immortalis-G720\)](#)
- + [Performance Counters](#)

### **For further reference on the technologies mentioned in the sheet, please refer to these webpages:**

- + [ASTC \(Adaptive Scalable Texture Compression\)](#)
- + [AFBC \(Arm FrameBuffer Compression\)](#)
- + [MSAA \(Multi-Sample Anti-Aliasing\)](#)
- + [Transaction Elimination](#)
- + [Hidden Surface Removal](#)
- + [IDVS \(Index-Driven Vertex Shading\)](#)
- + [DVS \(Deferred Vertex Shading\)](#)
- + [Shader Framebuffer Access \(GL ES\)](#)
- + [Shader Framebuffer Access \(Vulkan\)](#)

### **For free GPU profiling tools, see:**

- + [Arm Performance Studio](#)

**NOTE:** Mali-G78AE has the same base configurations and support as Mali-G78, but includes extra safety features. Mali-G6 series has the same specifications as Mali-G7 series for the values in this sheet.