Powering the next level of graphics performance

Micron graphics double data rate 7 (GDDR7) memory elevates user experience by unleashing the potential of workloads in AI, gaming, and high-performance computing (HPC).

Built on Micron's industry-leading 1β (1-beta) DRAM process node, Micron GDDR7 is a high-performance memory solution that delivers an introductory data rate of up to **32 Gb/s**, achieving a total system bandwidth greater than **1.5 TB/s**. GDDR7 elevates user experience by delivering exceptional performance to achieve reduced processing times for complex calculations and simulations in HPC, lifelike graphics rendering in gaming, and lower latencies for AI inference workloads.

Figure 1: Micron GDDR7 memory

1. Assuming 12 placements with a 384-bit bus system running at 32 Gb/s. GDDR7 achieves 60% higher bandwidth than GDDR6, with >1.5 TB/s system bandwidth (GDDR7 at 32 Gb/s and GDDR6 at 20 Gb/s). Calculation based on component specifications.
2. Improvement comparing measured Micron 16Gb GDDR6 and measured Micron 16Gb GDDR7.
3. Inference projection for GDDR7 32Gb/s calculated using measured GDDR6 20Gb/s data.

Key benefits

The highest performance graphics memory with improved power efficiency and device reliability.

60%

60% higher bandwidth than GDDR6, with greater than 1.5 TB/s system bandwidth (GDDR7 at 32 Gb/s and GDDR6 at 20 Gb/s).[1]

>50%

Micron GDDR7 has greater than 50% improvement in power efficiency over GDDR6 along with new sleep modes to reduce standby power up to 70%.[2]

20%

Reduced response times up to 20% for inference workloads such as generative AI text to image generation.[3]

Micron 1β enabling next-gen graphics performance

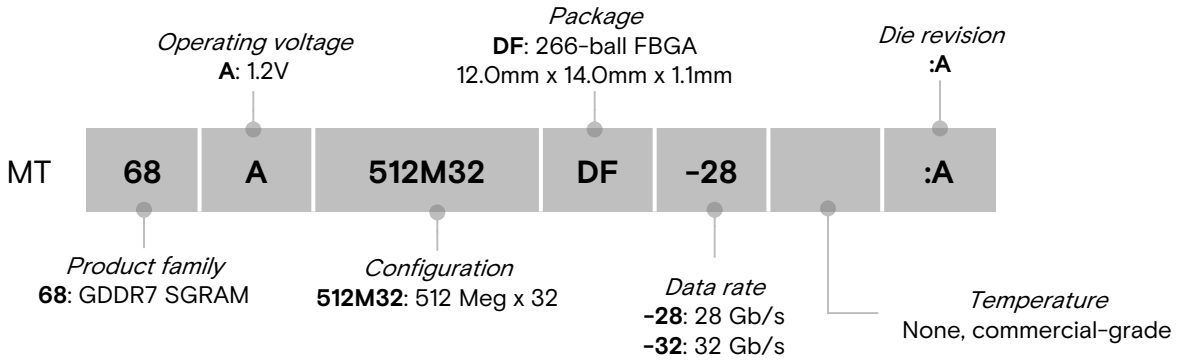


Figure 2: GDDR7 Micron part number (MPN)

PAM3 encoding for a data rate up to 32 Gb/s: Available with a speed per pin up to 32 Gb/s, enabling total system bandwidth greater than 1.5 TB/s.

Power: Lower operating voltage at 1.2V. Power-saving features include split voltage planes, partial device operation, and sleep modes.

Package: A thinner package height (1.1mm vs. 1.2mm) and high thermal EMC for packaging, providing 65% better thermal resistance.

RAS: Expanded RAS features such as OD-ECC, hPPR, CA parity, and 9-bit CRC for improved reliability.

Feature	GDDR6	GDDR7
DRAM die density	16Gb	16Gb
Max data rate (per pin)	18 Gb/s	32 Gb/s
BW/component	72 GB/s	128 GB/s
Signaling	NRZ	PAM3
Device width (I/O)	32 bits	32 bits
#Channels/package	2	4
Access/channel	32-byte	32-byte
Access/device	512b	1024b
Voltage (max)	1.35V	1.2V
Power efficiency (pJ/bit)	6.5 pJ/bit	4.5 pJ/bit
Package dimensions	12 x 14 x 1.2 mm	12 x 14 x 1.1 mm
Ball pitch	0.75mm	0.75mm/0.73mm
Clam shell	Yes	Yes
Ball count	180	266
RAS features		
On-die ECC	No	Yes
ECC transparency	No	Yes
CA parity	No	Yes
Error check and scrub	No	Yes
Host post package repair (hPPR)	No	Yes
Poison data handling	No	Yes

Table 1: Micron's GDDR6 and GDDR7 feature comparison

Multi-level signaling

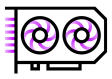
Memory	GDDR6	GDDR7
Encoding	Non-return-to-zero (NRZ)	PAM3
Bits per cycle	1	1.5
Levels	2 (0,1)	3 (-1, 0, 1)
System bandwidth	960 GB/s	1.5 TB/s
Speed per pin	20 Gb/s	32 Gb/s

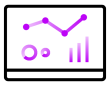
Table 2: Evolution of GDDR signaling techniques

Previous GDDR standards, which used non-return-to-zero (NRZ), relied on two signal levels to transmit data encoded as 1s or 0s. This methodology was sufficient for many generations until clock speeds and the corresponding system complexity became a bottle neck. To solve this challenge, Micron introduced multi-level signaling techniques with its innovative GDDR6X technology. Micron was the sole provider of GDDR6X, which utilized PAM4 signaling and provided an industry leading bandwidth of >1.1 TB/s. Micron's success and experience with GDDR6X laid the foundation for the next generation of GDDR using a similar signaling method. GDDR7 continues to scale bandwidth capability through advanced design architectures, including the industry's first implementation of PAM3 (pulse amplitude modulation 3) signaling. PAM3 provides higher data transmission rates than NRZ on GDDR6 by encoding more information (1.5 times more) per signal cycle using three distinct voltage levels (-1, 0, +1). While the number of bits transferred per cycle is lower in PAM3, as compared to PAM4 on GDDR6X, PAM3 allows for 50% higher voltage margin and has lower encoder complexity, which reduces the need to have higher frequencies for the memory bus and mitigates the resulting signal loss.

A wider range of applications for GDDR7

As gaming continues to become more lifelike and generative AI moves more workloads to edge devices, the need for high performance memory for purpose-built devices will only grow. GDDR7 offers the highest component pin speed at 32 Gb/s that, when combined with a 384-bit memory bus width, can deliver a whopping 1.5 TB/s of memory bandwidth for any application that requires peak system bandwidth, including gaming and content creation.





GAMING	GENERATIVE AI	HPC
PC, notebook, and console	AI inference, machine learning, large language models (LLMs)	3D modeling, HPC simulations, CAD design, animation
GDDR7 is expected to achieve greater than **30%** improvement in frames per second (FPS) for ray tracing and rasterization workloads.[4]	The high system bandwidth (>1.5 TB/s) offered by GDDR7 is expected to reduce response times by up to **20%** for generative AI text-to-image generation.[5]	Reduced processing times and seamless multitasking for complex workloads such as animation, 3D design, scientific simulations, and financial modeling.

Figure 3: GDDR7 applications

4. As compared to GDDR6/6X trend across 1080p, 1440p, and 4K resolutions.
5. Inference projection for GDDR7 32 Gb/s calculated using measured GDDR6 20 Gb/s data.

micron.com/GDDR7