

High Bandwidth Memory

Ketan Reddy and Tyler Krupicka

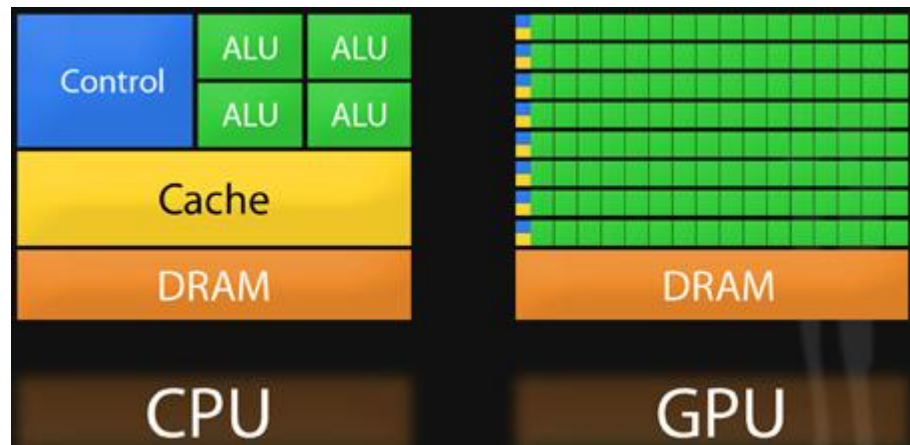
Contents

- Background on GDDR
- The Memory Bottleneck
- HBM Overview
- HBM Schematic Improvements
- Benchmarks
- HMC / 3DXpoint Comparison
- HBM2 Standard
- GPU Comparisons



Background Info - GPU and GDDR

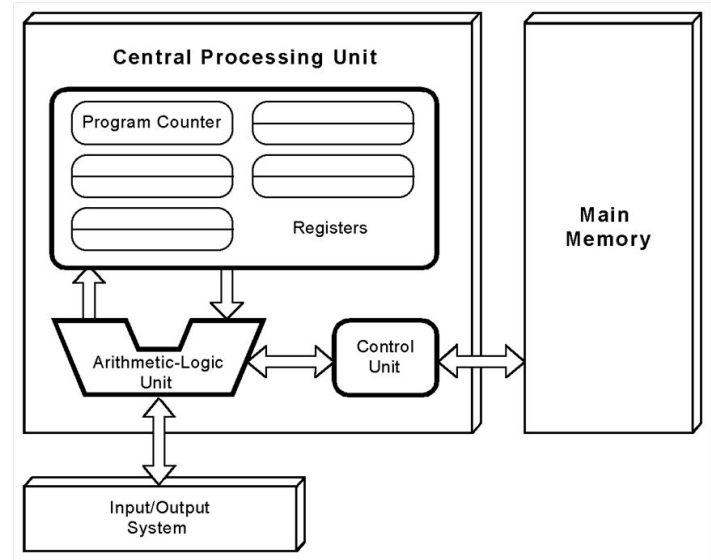
- Specialized hardware for rendering/encoding/decoding images and video
- Designed for highly parallel and computationally intensive operations
- Typically produced as standalone cards
- Does not use normal DDRx RAM
- GDDR has a higher bandwidth, wider bus, can request and receive data in the same cycle



The Problem

The Memory Problem

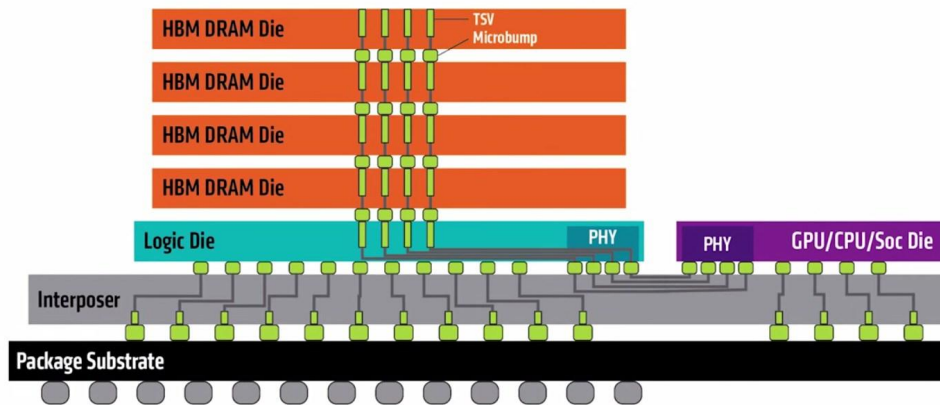
- The Von Neumann Memory Bottleneck
- Processor speeds have overtaken memory access speeds
- GDDR5 is rising in power consumption.
- Large footprint of GDDR5 chips expands form factor.



The Solution

HBM Overview

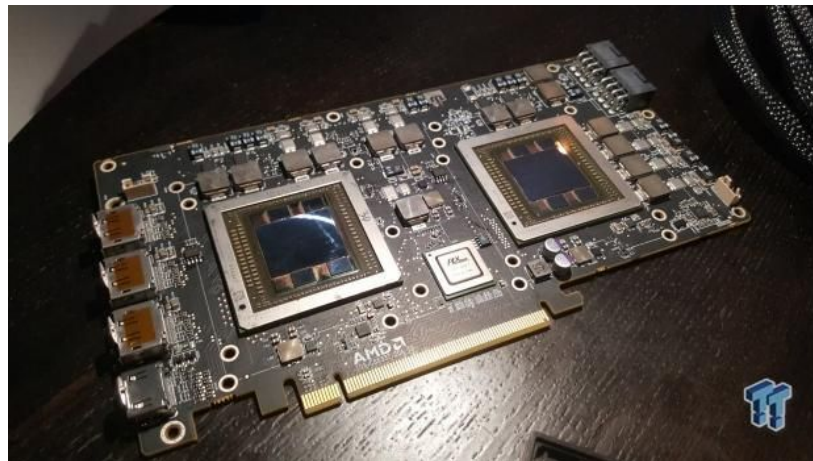
- Developed by AMD and SK Hynix
- JEDEC industry standard in October 2013
- Multiple DRAM dies stacked in a single package connected by TSV and Microbumps
- Connected to the processor unit directly via Interposer Layer
- Two fully independent channels between each stack and chip



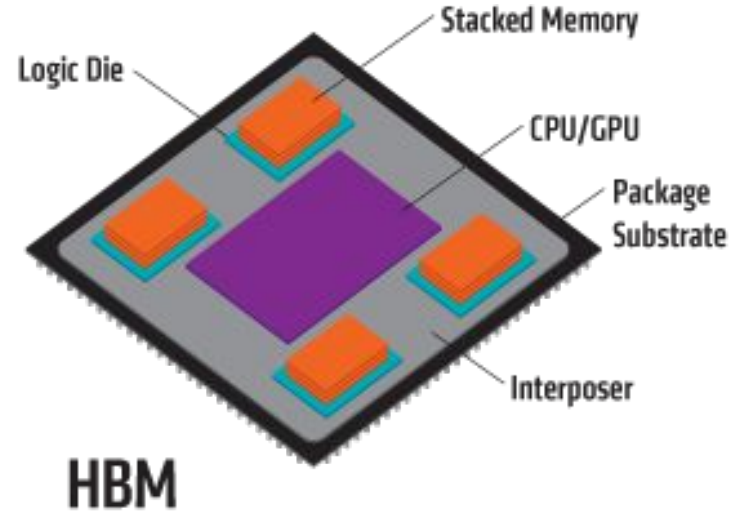
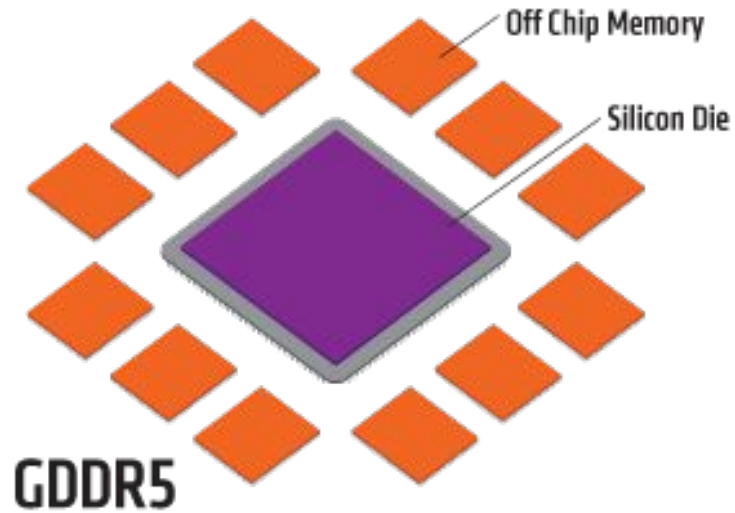
HBM Overview Continued

Improvements over Standard GDDR RAM

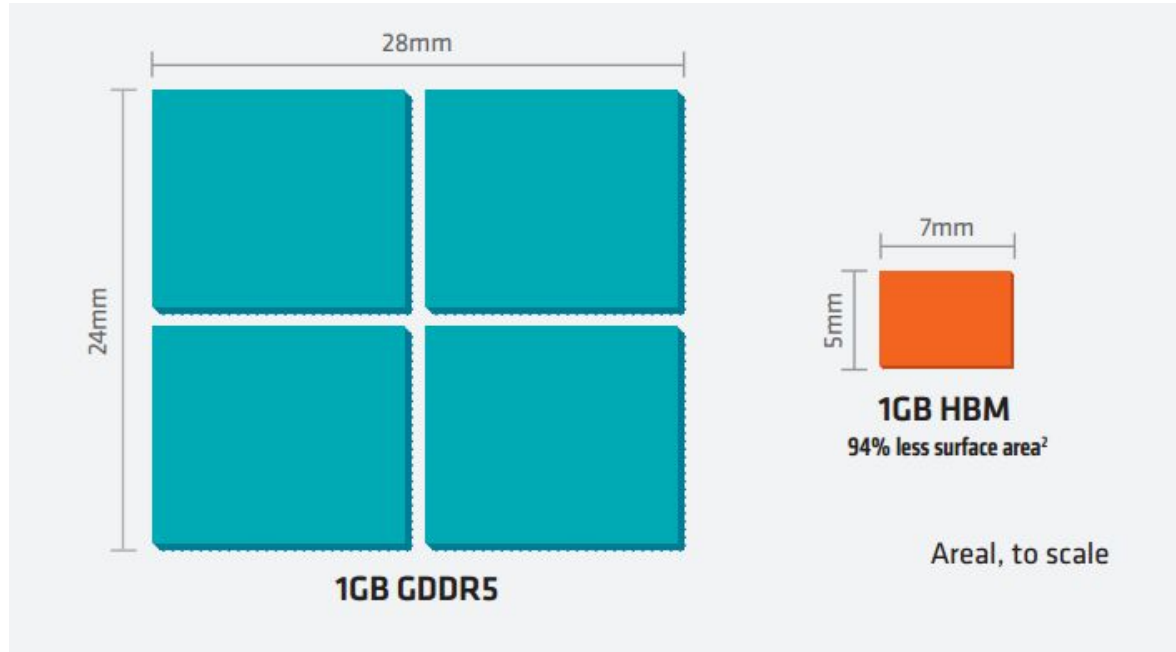
- Very High Bandwidth
- Lower Effective Clock Speed
- Smaller Package
- Lower Power Consumption
- Shorter Interconnect Wires
- Individual Banks Can Be Refreshed




HBM vs. GDDR5 Form Factor



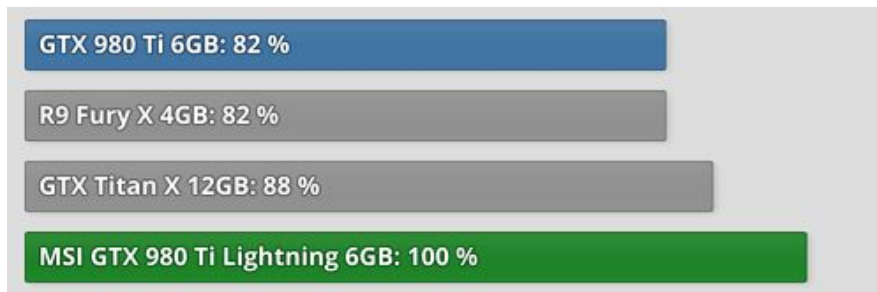
Space Savings of Cache Memory



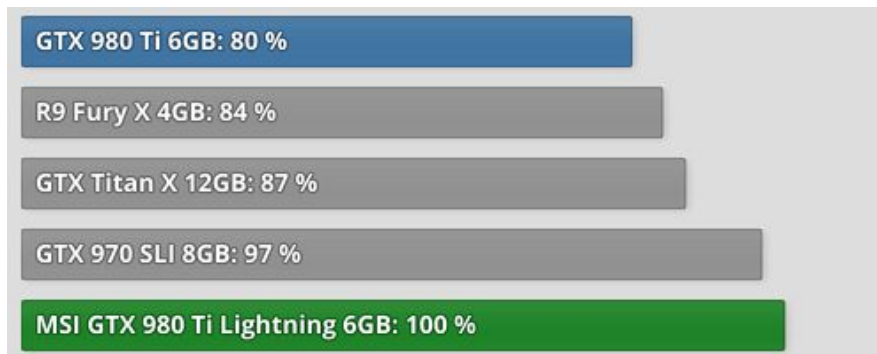
GDDR5 vs HBM1

- Bus Width: 32 bit
 - Clock Speed: 1750MHz
 - Transfer Rate per pin: 7GB/s
 - Bandwidth: 28GB/s per chip
 - Bandwidth per Watt: 10.5GB/W
 - Operating Voltage: 1.5V
 - Area: 24mm x 28mm
- Bus Width: 1024 bit
 - Clock Speed: 500MHz
 - Transfer Rate per pin: 1GB/s
 - Bandwidth: 128GB/s per chip
 - Bandwidth per Watt: 35GB/W
 - Operating Voltage: 1.3V
 - Area: 5mm x 7mm
- 

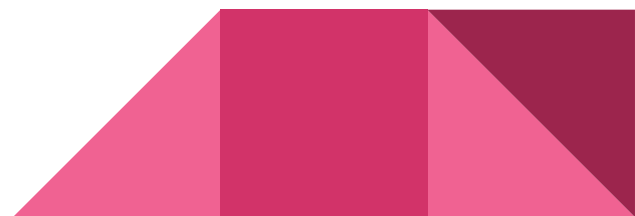
Benchmark of GDDR5 vs HBM (GTX 980ti vs R9 Fury X)



Stock 980ti and Fury X perform the same at 1080p/1440p



Stock Fury X outperforms the stock 980ti at 4k



Other 3D RAM Solutions: HBM vs HMC vs 3D XPoint

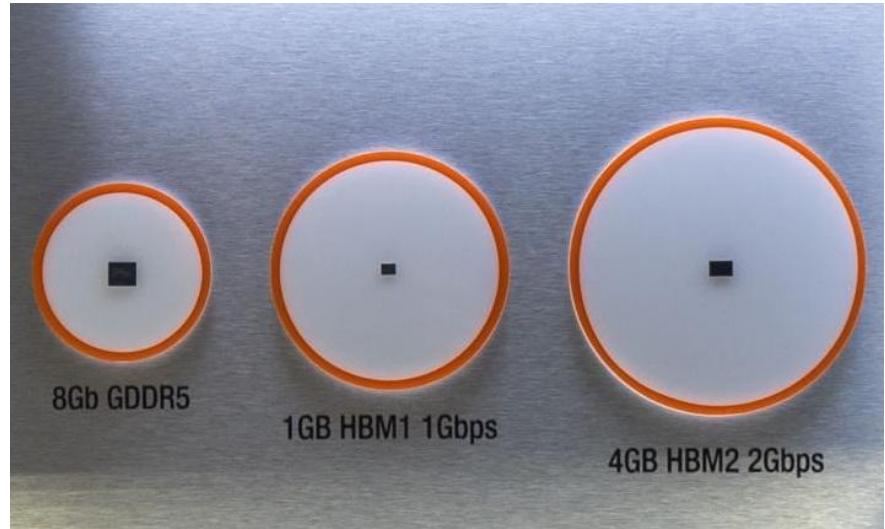
Type	HBM	HMC	3D XPoint
Developer	AMD, SK Hynix, Samsung	Arm, Micron, IBM, Samsung	Micron and Intel
Applications	VRAM	Multi-Core Servers, DRAM	Mass Storage, DRAM, Hybrid
Max Bandwidth	256 GB/s	480 GB/s	N/A (Application Based)
Other	JEDEC Standard, Per Bank Refresh	Also Uses TSV and Microbumps Stack 4-8 Memory Cells Not JEDEC Standard	Non-Volatile, High Read/Write Endurance



The Future

HBM Gen 2

- Finalized by JEDEC in January 2016
- Improvements over Gen 1
 - 8 Dies per stack
 - 2Gb/s per pin
 - 256 GB/s bandwidth
 - 8GB per package
- Already on market
 - NVIDIA Tesla P100
- Very important for high bandwidth applications such as VR and networking



GPU Memory Math

	AMD Radeon R9 290X	NVIDIA GeForce GTX 980 Ti	AMD Radeon R9 Fury X	Samsung's 4-Stack HBM2 based on 8 Gb DRAMs	Theoretical GDDR5X 256-bit sub-system
Total Capacity	4 GB	6 GB	4 GB	16 GB	8 GB
Bandwidth Per Pin	5 Gb/s	7 Gb/s	1 Gb/s	2 Gb/s	10 Gb/s
Number of Chips/Stacks	16	12	4	4	8
Bandwidth Per Chip/Stack	20 GB/s	28 GB/s	128 GB/s	256 GB/s	40 GB/s
Effective Bus Width	512-bit	384-bit	4096-bit	4096-bit	256-bit
Total Bandwidth	320 GB/s	336 GB/s	512 GB/s	1 TB/s	320 GB/s
Estimated DRAM Power Consumption	30W	31.5W	14.6W	n/a	20W



Questions?

Sources

- <http://www.pcworld.com/article/2922599/amd-talks-up-high-bandwidth-memory-that-will-power-its-next-gpus-pokes-nvidia-too.html>
- <http://www.amd.com/en-us/innovations/software-technologies/hbm>
- <https://www.amd.com/Documents/High-Bandwidth-Memory-HBM.pdf>
- <http://motherboard.vice.com/read/what-high-bandwidth-memory-is-and-why-you-should-care>
- <http://www.anandtech.com/show/9969/jedec-publishes-hbm2-specification>
- <http://www.overclock.net/t/1578881/fury-x-is-now-just-as-fast-as-gtx-980ti-in-1080p-1440p-and-faster-in-4k>
- <https://www.cs.utah.edu/thememoryforum/mike.pdf>
- Image Source: <http://zakarum.tistory.com/entry/The-von-Neumann-Architecture>
- Image Source: <http://www.ocdrift.com/amd-unleashes-radeon-fury-x-fury-and-nano-featuring-revolutionary-hbm-technology/>
- Image Source: http://images.anandtech.com/doci/9390/HBM_7_Interposer.png
- <https://dzone.com/articles/high-bandwidth-memory-vs-hybrid-memory-cube-what-i>
- Image Source: <https://www.extremetech.com/extreme/226240-sk-hynix-highlights-the-huge-size-advantage-of-hbm-over-gddr5-memory>