

# GDDR6 給 FPGA 帶來的大頻寬存儲優勢以及性能測試

■作者：黃侖 / Achronix 高級應用工程師

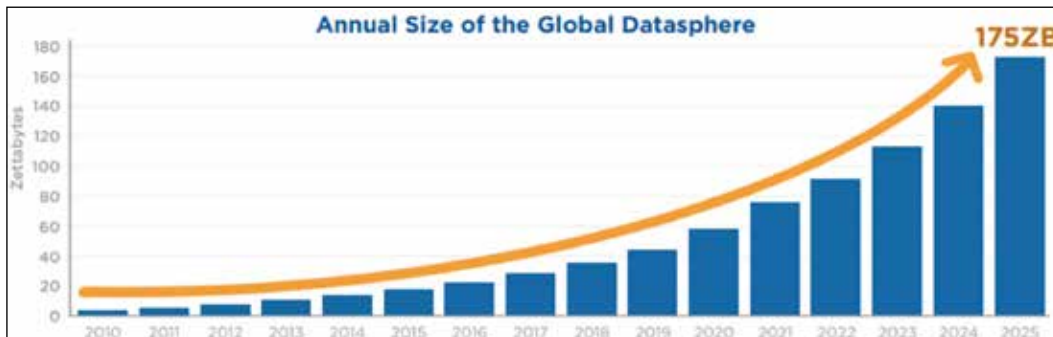
## 概述

隨著互聯網時代的到來，人類所產生的資料發生了前所未有的、爆炸性的增長。IDC 預測，全球資料總量將從 2019 年的 45ZB 增長到 2025 年的 175ZB<sup>[1]</sup>。同時，全球資料中近 30% 將需要即時處理，因而帶來了對 FPGA 等硬體資料處理加速器的需求。如圖 1 所示。

在這樣的資料高速增長的情況下，用於傳輸資料的網路頻寬和處理資料所需要的算力也必須急速增長。傳統的 CPU 已經越來越不堪重負，所以用硬體加速來減輕 CPU 的負擔是滿足未來性能需求的重要發展方向。未來的硬體發展需求對於用於加速的硬體平臺提出了越來越高的要求，可以概括為三個方面：算力、資料傳輸頻寬和記憶體頻寬。

Achronix 的新一代採用台積電 7nm 製程的 Speedster 7t FPGA 晶片根據未來硬體加速和網路加速的需求，在這三個方面都做了優化，消除了傳統 FPGA 的瓶頸。下面我們重點說一說為了提高記憶體頻寬，Achronix 通過採用硬核 GDDR6 控制器所帶來的優勢。

圖 1: 全球資料增長預測



## GDDR6 的發展

在 GDDR 的設計之初，其定位是針對圖形顯示卡所特別優化的一種 DDR 記憶體。因為 2000 年後電腦遊戲特別是 3D 遊戲的發展和火爆，使運行電腦遊戲的顯卡需要大量的快速圖像資料交互需求，GDDR 在這種情況下應運而生。第一個 GDDR 標準是基於 DDR 的 GDDR2，隨後發展到了基於 DDR3 的 GDDR5，在一段時間中非常流行。

2016 年，GDDR5X 正式發佈，它引入了具有 16n 預取的四倍數據速率模式，但代價是訪問細微性從 GDDR5 的 32Byte 提高到了 64Byte。2018 年，GDDR6 發佈，資料速率達到了 16Gbps，頻寬幾乎是 GDDR5X 的兩倍，同時採用了雙通道設計，訪問細微性和 GDDR5 一樣是 32Byte。

## GDDR6 和 DDR4/5 的比較

GDDR 一直以來是針對圖形顯示卡所優化的一種 DDR 記憶體。因為顯卡處理圖像資料，特別是 3D 圖像資料對顯存頻寬的要求更高，GPU 和 GDDR 之間的資料交換非常頻繁。而 DDR 記憶體

專注於與 CPU 進行資料交換的效率，因此對於整體存取性能、低延遲更為看重，所以在 CPU 和傳統的 FPGA 中基本都是用 DDR4。

隨著硬體加速需求對於記憶體頻寬提出了越來越高的要求，傳統的 DDR4 頻寬顯然已經無法滿足要求，Achronix 看重了 GDDR6 在資料存儲中的頻寬優勢，創新地將 GDDR6 引入到了 FPGA，徹底解決了傳統 FPGA 存儲頻寬不夠的瓶頸。

2020 年 7 月 15 日，JEDEC 存儲協會正式發佈了 DDR5 SDRAM 的標準 (JESD79-5)，記憶體的頻率相對 DDR4 的標準頻率有了大幅的提升，總傳輸頻寬也提升了 38%，但是還是和 GDDR6 的頻寬有一定的差距。如圖 2 所示<sup>[2]</sup>，GDDR6 和 DDR4/5 的頻寬對比。

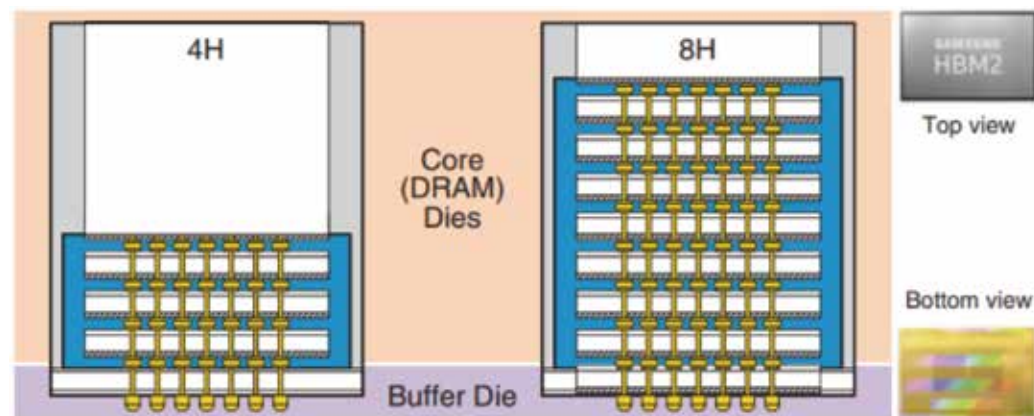
圖 2: GDDR 與 DDR 頻寬發展對比



圖 3: GDDR6 和 DDR4 性能對比



圖 4: HBM Die 的堆疊



如果實現同一個大頻寬存儲的應用，在提供相同的記憶體頻寬的情況下，無論在設計複雜度，PCB 佔用面積，還是在功耗方面，與 DDR4 相比，GDDR6 的性能都有很大的提高，如圖 3 所示<sup>[2]</sup>。

## GDDR6 和 HBM2 的比較

HBM 全稱 High Bandwidth Memory，最初的標準是由 JEDEC 在 2013 年發佈。2016 年 1 月，HBM 的第二代 HBM2 正式成為工業標準。HBM 的出現也是為了解決記憶體頻寬問題。與 GDDR6 不同的是，HBM 記憶體一般是由 4 個或者 8 個 HBM

的 Die 堆疊形成，我們稱之為一個 Stack。如圖 4 所示<sup>[4]</sup>。

我們以市面上帶有 HBM2 的高端 FPGA 為例，這個系列的 FPGA 集成了 1~2 個這樣的 HBM2 Stack。兩個 Stack 之間是相互獨立的，各自有自己的位址空間。每個 Die 都有獨立的兩個 128bit 的 Channel，所以 4 個 Die 8 個通道就是 1024bit 的位寬，HBM2 的頻率是 900MHz，按 DDR 的方式訪問，一個 Stack 總共頻寬是 900 (MHz) x 2 (DDR) x 1024 (位寬) / 8 = 230GB/s，兩個 Stack 最高可以到 460GB/s 的頻寬。

Achronix 的 Speedster 7t FPGA 集成了 8 個 GDDR6 的硬核，每個 GDDR6 的硬核支持雙通道。總的頻寬是 16Gbps x 16 (位寬) x 2 (通道) x 8 (控

制器) / 8 = 512 GB/s，略高於帶 HBM2 的 FPGA 記憶體頻寬。

從成本上來看，目前 GDDR6 與 HBM2 相比有著很大的優勢，HBM2 技術工藝要求高，目前晶片的良率和產量都會受到很大的影響。同時 GDDR6 使用起來更靈活，使用片外的 DRAM，可以根據應用要求，選擇不同速率，不同容量的 GDDR6 顆粒。HBM2 的優勢在於集成度高，不佔用 PCB 板的面積。圖 5 是 DDR4、GDDR6 和 HBM2 在成本上的一個綜合比較。

圖 5: DDR4 vs GDDR6 vs HBM2

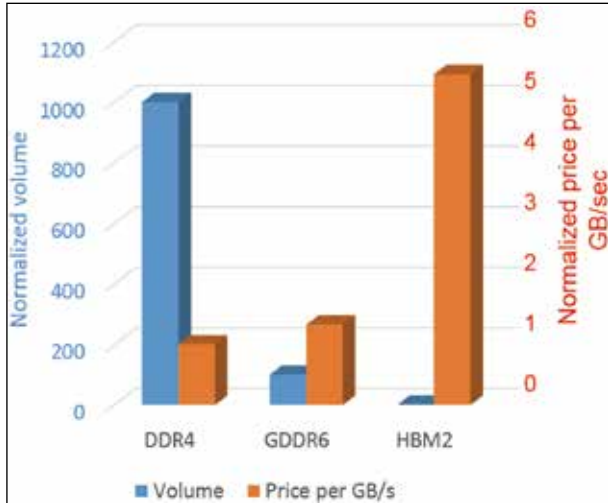


圖 6: GDDR6 顆粒結構

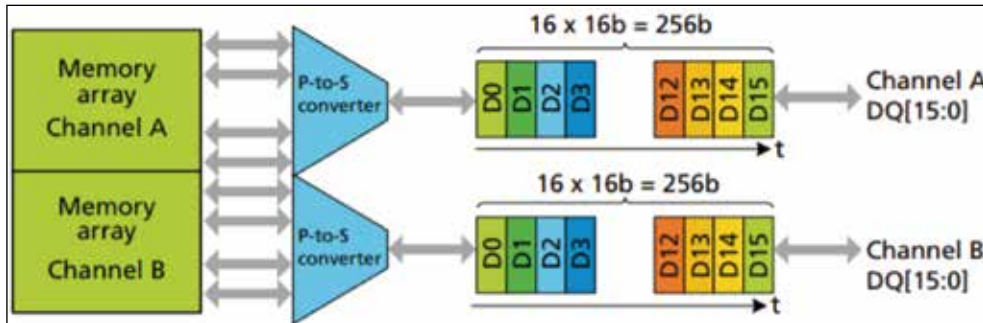
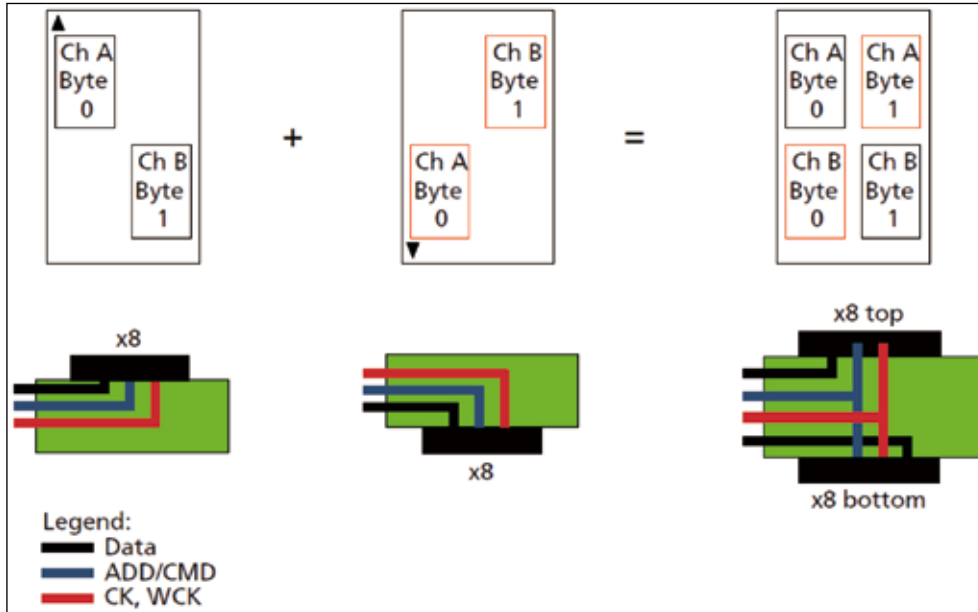


圖 7: GDDR6 的 Clamshell 模式



## GDDR6 技術細節以及 Clamshell 模式

GDDR6 結構如圖 6 所示<sup>[3]</sup>。它是採用 16n

Prefetch 結構，一次寫操作或者讀操作的資料是 16n。每個 GDDR6 顆粒有兩個獨立的通道，每個獨立的通道訪問獨立的記憶體空間。對於每個通道，讀或者寫的位寬是 256bit 或者 32Byte。P-to-S converter 是一個並變串的轉換器，把每個 256bit 位元寬的資料轉換成 16 位元匯流排，每位元匯流排上傳輸 16bit 的資料。這樣 GDDR6 每個通道最小的訪問細微性是 256bit 或者 32Byte。

根據 GDDR6 這樣 16n 預取結構，內部存儲陣列如果訪問週期是 1ns，則 I/O 上的資料率則是 16Gbps。

一個 GDDR6 控制器支持兩個獨立通道，一個 GDDR6 顆粒也是兩個獨立的通道，所以在通常模式下，一個 GDDR6 控制器對應一個 GDDR6 的顆粒，用 x16 模式，實現最高 512Gb/s 的頻寬。

因為目前市面上 GDDR6 顆粒的最大容量是 16Gb，在有些應用中如果對容量有一定的要求，可以使用一種叫 Clamshell 的連接方式，如圖 7<sup>[5]</sup> 所示，每個 GDDR6 控制器連接兩個 GDDR6 顆粒，每個 GDDR6 的顆粒用 x8 模式，這樣在這種 Clamshell 模式下，頻寬不變，但是支援的 GDDR6 的容量翻倍了。

### GDDR6 在 7t1500 上的讀寫效率

最後，我們測試一下 7t1500 上 GDDR6 控制器的讀寫效率，所有的測試結果基於模擬資料。測試環境如圖 8 所示。因為 7t1500 包含了片上網路 (NoC)，並且 NoC 已經實現了仲裁，時鐘域轉換的邏輯，我們用三個使用者邏輯通過 NoC 去訪問同一個 GDDR6 Channel，得到的綜合讀寫效率更能反映用戶實際運用中的場景。

在不同的突發長度和不同的位址訪問方式下的測試結果如圖 9 所示。

(如果需要進一步聯絡請發送郵件到：dawson.guo@achronix.com)

圖 8: GDDR6 讀寫效率測試架構

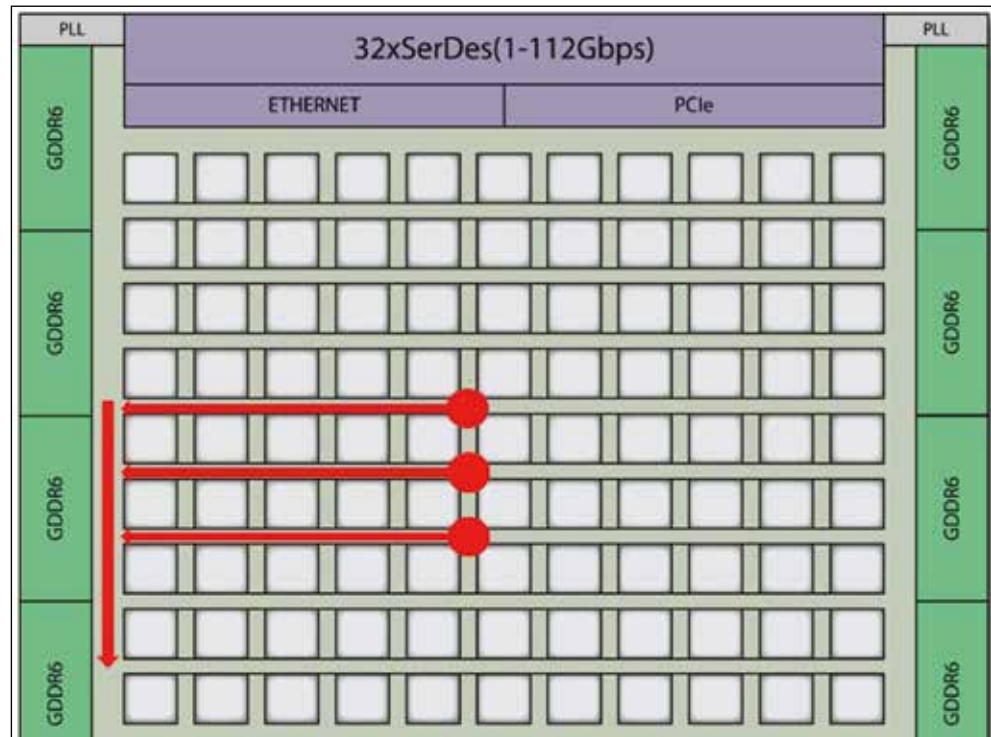


圖 9: GDDR6 讀寫效率

Linear address:				
Burst Length	64 Byte	128 Byte	256 Byte	512 Byte
GDDR6 100% write efficiency	98.33%	98.20%	98.11%	98.17%
GDDR6 100% read efficiency	98.16%	98.15%	98.19%	98.09%
Random address:				
Burst Length	64 Byte	128 Byte	256 Byte	512 Byte
GDDR6 100% write efficiency	45.98%	77.81%	94.03%	93.86%
GDDR6 100% read efficiency	43.82%	81.66%	97.17%	97.43%

### 參考文獻：

- 1.The Digitization of the World From Edge to Core 2018
- 2.Extending the Benefits of GDDR Beyond Graphics by Micron
- 3.TN-ED-03: GDDR6: The Next-Generation Graphics DRAM Memory Array Prefetch and Access Granularity
- 4.Samsung 網站：www.samsung.com
- 5.Micron 網站：www.micron.com
- 6.Achronix 網站：www.achronix.com 