

產業焦點評論

# 解析 AMD Chiplet 晶片之發展重點

## 前言

AMD 在 2021 年 6 月的 COMPUTEX 中發表了首款利用 3D chiplet 技術製作的晶片 Ryzen 9 5900X CPU，並展示出其產品效能。這個由 AMD 與台積電共同開發完成的 3D chiplet 技術，顯示出對先進封裝的高度需求。本文將透過 AMD 在近期公開的專利申請案解析其在 chiplet 晶片的發展重點。

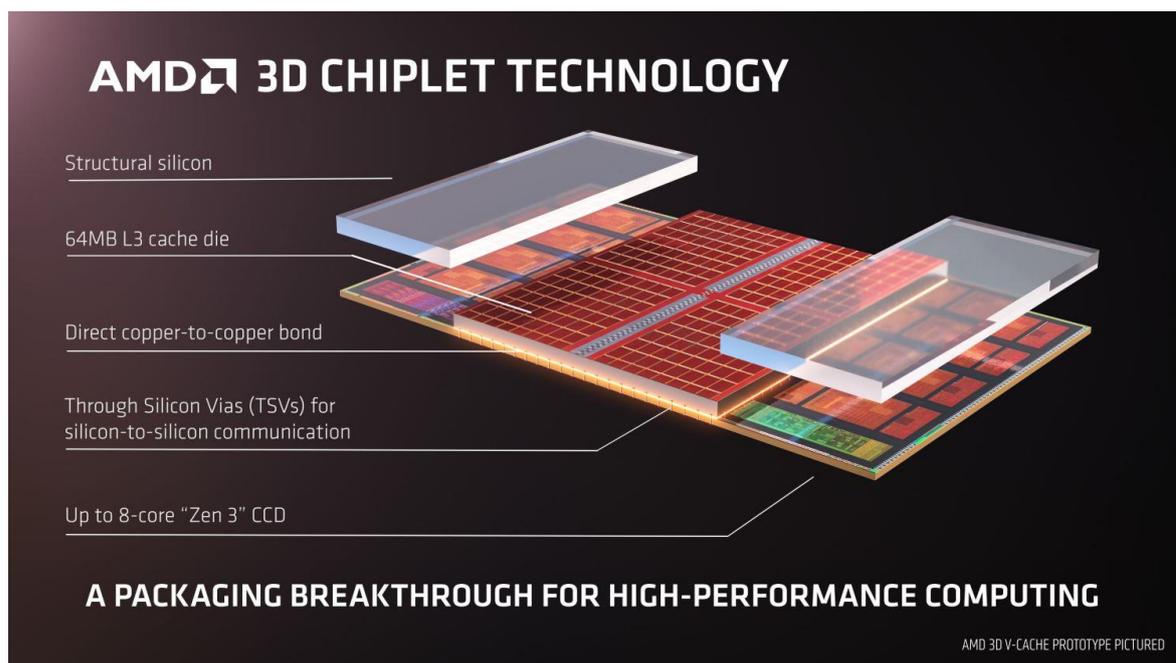
鄭凱安

# AMD 在 COMPUTEX 2021 發表 chiplet 晶片

AMD 總裁暨執行長蘇姿丰 2021 年 6 月 1 日在台北國際電腦展 ( COMPUTEX ) 的演講中發表了 AMD 首款採用 3D 小晶片 ( chiplet ) 技術的 AMD Ryzen 5000 系列處理器。這款晶片所採用的 3D chiplet 技術是由 AMD 與台積電所共同開發完成，將是 AMD 未來高階運算產品的關鍵技術。這個技術的實現，也使 AMD 可以依照既定規劃於 2021 年底前將 3D chiplet 技術導入高階運算晶片的實際生產。

圖一是此次 AMD 所發表採用 3D chiplet 技術的 Ryzen 5000 示意圖。蘇執行長表示這個晶片採用了領先業界的 hybrid bond 技術，將 chiplet 以 active-on-active 的 3D 晶片堆疊方式結合，並將快取記憶體 ( cache ) 晶片以 3D 結構實現，提供超過 2D chiplet 技術 200 倍的互連密度，並展現達到現有 3D 解決方案 1/3 的功耗。每一個核心小晶片 ( Core Complex Die, CCD ) 除原本配置的 32MB cache 外，將可透過 3D chiplet 堆疊配備 64MB 的 cache，合計有 96MB 的 cache 容量，使運算效能大幅提升。由 AMD 所提供採用 3D chiplet 技術與標準版 Ryzen 9 5900X 的測試比對資料來看，採用 3D chiplet 的 Ryzen 9 5900X 平均每秒可處理的影像幀數 ( Frame Per Second, FPS ) 多出了 12%，運算效能有明顯提升。

圖一、AMD 的 3D chiplet 技術說明



資料來源：AMD · 2021 年 07 月

## 由近期專利公開解析 AMD chiplet 技術

AMD 在 COMPUTEX 所發表的 active-on-active chiplet 技術可從其近期公開的專利申請案中一窺究竟。2021 年 4 月 1 日，美國專利與商標局 ( USPTO ) 將 AMD 的美國專利申請案「ACTIVE BRIDGE CHIPLET WITH INTEGRATED CACHE」公開，揭示了 AMD 以小晶片 ( chiplet ) 開發高階晶片的部分技術。這個專利案在 2019 年 9 月提出申請，聚焦於串接 GPU 小晶片的主動式橋接小晶片 ( Active Bridge chiplet，以下簡稱為 Active Bridge )，並在 Active Bridge 上建構了快取記憶體的结构。以下，將就此一專利技術的重點進行解析。

### GPU發展面臨的挑戰

AMD 在專利的說明中首先提到 GPU 架構設計上的挑戰。GPU 在運算時是分成許多小區塊 ( sections ) 同步進行平行運算，而這些區塊間必須保持溝通與記憶體同步共享，才能確保平行運算結果的一致性。然而對於應用程式而言，在建構圖像畫面時，如果必須同時溝通多個 GPU 小區塊，並兼顧這些區塊同步共享記憶體儲存的内容，在運算上是非常沒有效率的。因此，在實務上，應用程式的設計多半是以 CPU 透過

單一管道溝通 GPU 進行運算為原則，而 GPU 在接收 CPU 運算指令與資料後的分區、平行運算與區塊間的記憶體同步共享等則留在 GPU 內部處理。

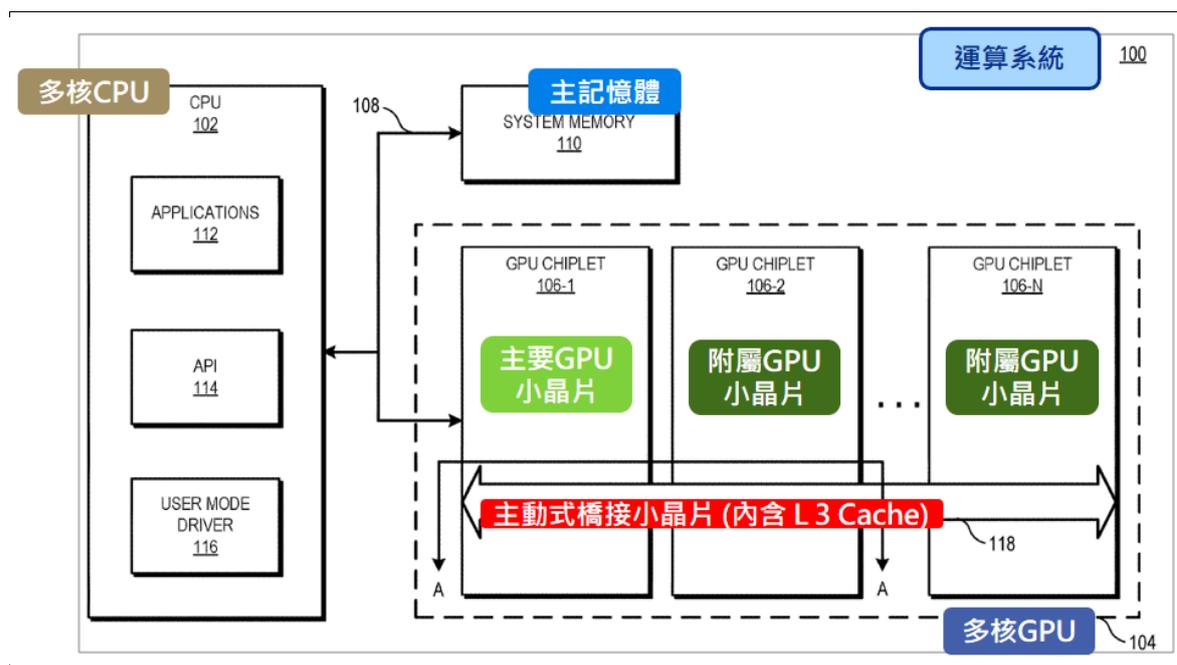
隨著高效能運算需求持續增加，GPU 需要負擔的運算工作更多，使得 GPU 內部的平行運算結構與電路更加複雜，使傳統以單一晶片製程製作，也就是所謂 monolithic 的 GPU 元件結構更加複雜，製程精度要求更加嚴峻。這也造成了 GPU 隨著運算效能提升，維持製程良率與控制開發成本成為一大挑戰。

因應這樣的挑戰，本次 AMD 的專利申請案提出以由多個小晶片組合代替小區塊形成 GPU，以 Active Bridge 串接 GPU 小晶片進行整合運算的設計，並運用 3D 封裝技術完成小晶片間的結構整合，以減少晶片製程的挑戰。

## 以Chiplet架構構建GPU

圖二是本次公開專利申請案的代表圖，圖中的運算系統由一個多核心 CPU、一個多核心 GPU 與主記憶體（通常是 DRAM）組合而成，其中多核 GPU 由一個主要 GPU 小晶片（Primary GPU chiplet）與其他附屬 GPU 小晶片組合而成，並以一主動式橋接小晶片串接所有 GPU 小晶片確保 GPU 小晶片之間的資料互通與記憶體同步。在操作時，CPU 只與主要 GPU 小晶片連結，後續運算由主要 GPU 小晶片透過 Active Bridge 與其餘附屬 GPU 小晶片溝通完成。在運算中，CPU 與 GPU 晶片內部各自配置有記憶體，稱為快取記憶體（cache），加上主記憶體，可以因應運算中資料儲存的需求；而在圖二的小晶片架構 GPU 中，連接 GPU 小晶片的 Active Bridge 上面也設置有快取記憶體，可做為 GPU 內部共享記憶體的一部分。

圖二、位元切換時，讓能階本身徐緩變化，以減少能量損失



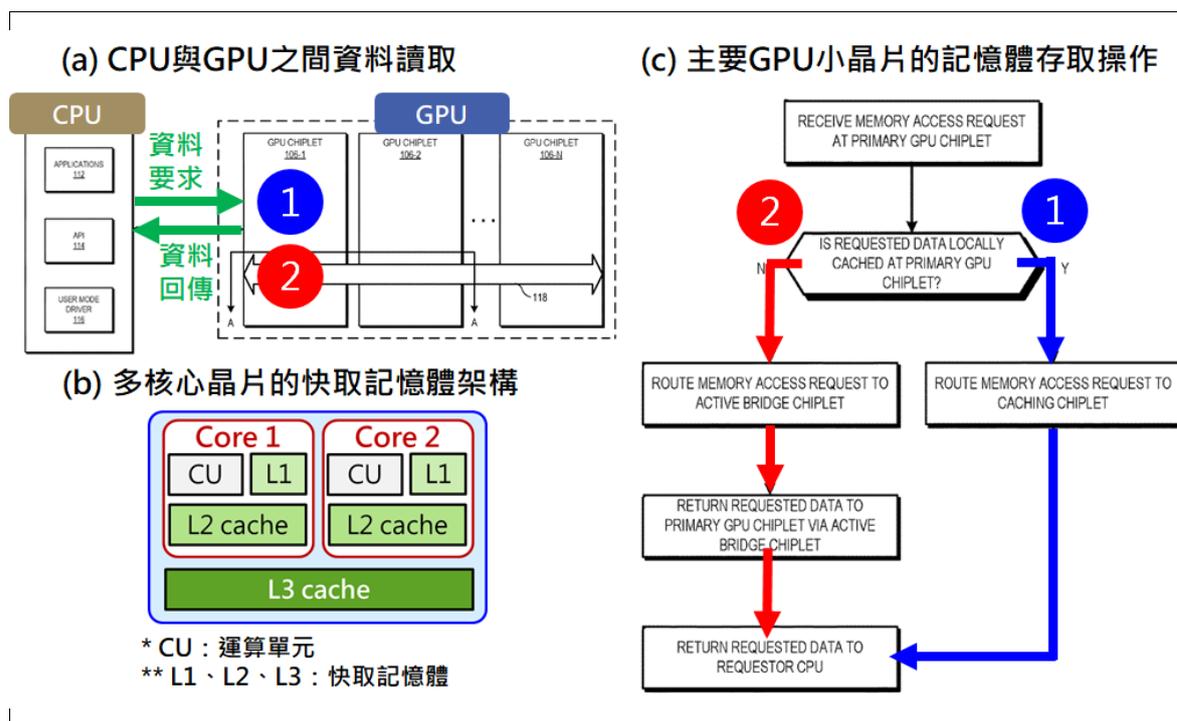
資料來源：USPTO · MIC 整理 · 2021 年 07 月

圖三進一步說明主要 GPU 小晶片與 Active Bridge 以及附屬 GPU 小晶片之間的記憶體共享模式，其中，左上 (a) 圖是 GPU 應 CPU 要求調取記憶體儲存內容後回傳的示意圖，左下 (b) 圖則是多核心晶片的快取記憶體架構，右側 (c) 圖則代表了 GPU 內部主要 GPU 小晶片處理 CPU 要求的流程。

首先以 (b) 圖說明多核心晶片內部的記憶體配置。一般將晶片內部快取記憶體分為 L1、L2、L3 三個等級，容量最小的 L1 cache 緊鄰運算單元 CU，其次為容量較大的 L2 cache，以及容量更大、所有核心共用 L3 cache。快取記憶體通常儲存運算單元正在使用或才使用過的資料，而運算單元 CU 有資料需求時，因為存取越靠近的快取記憶體運算速度越快，所以會依序檢查 L1、L2、L3 cache 是否存有所需資料，若有則直接取用進行運算；若沒有，則會從外部主記憶體調入該筆資料，暫存於 cache 中隨時取用。

由 (c) 圖來看，當主要 GPU 小晶片接收 CPU 所發出的記憶體資料調取要求後，會先判斷該資料所在位置，如果是在主要 GPU 小晶片的 L1 或 L2 cache 中，則直接取出回傳給 CPU；反之，則透過 Active Bridge，從附屬 GPU 小晶片的 L1 或 L2 cache 或者是 Active Bridge 上設置的 L3 cache 中取出所需資料回傳給 CPU。

圖三、位元切換時，讓能階本身徐緩變化，以減少能量損失



資料來源：USPTO · MIC 整理 · 2021 年 07 月

### Chiplet GPU的3D結構設計

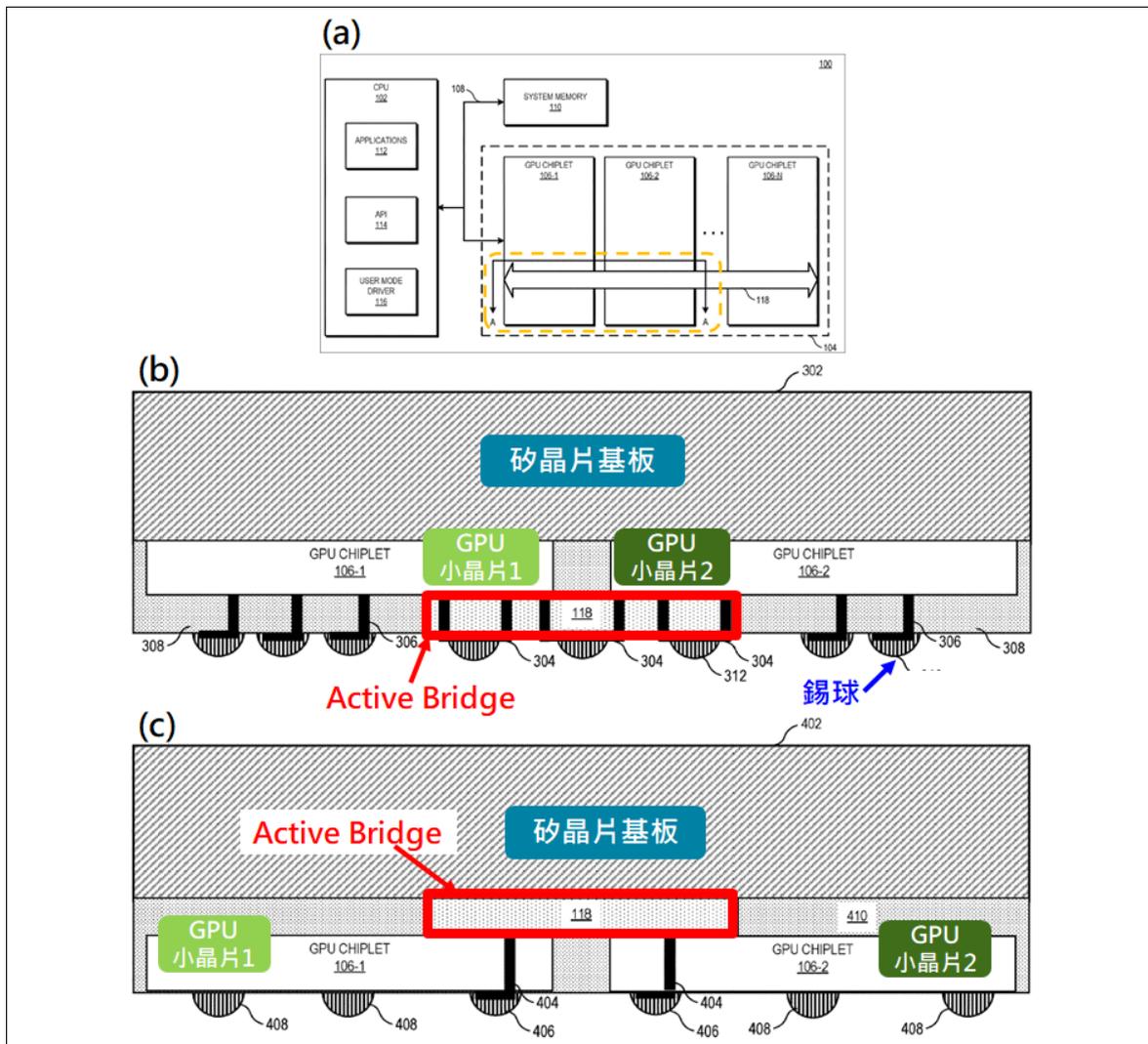
圖四說明 AMD Chiplet GPU 的 3D 結構設計，其中 ( b ) 圖、( c ) 圖呈現的二種結構設計各包含 2 個 GPU 小晶片以及 1 個 Active Bridge 小晶片，是 ( a ) 圖中黃色虛線位置的剖面結構。

在 ( b ) 圖的第一種設計中，2 個 GPU 小晶片貼附在矽晶片基板上，而 Active Bridge 則貼合在 2 個 GPU 小晶片的上方，並與 GPU 小晶片保持電路連結，GPU 小晶片上方及與 Active Bridge 間形成的空洞以介電 ( dielectric ) 材料填充，而整個矽晶片基板翻轉朝下，以便與下方電路板或載板連結封裝。在 Active Bridge 與介電材料上設有穿透式的導電通道( Through Silicon Via, TSV 及 Through Dielectric Via, TDV )，以便輸入輸出訊號連接 GPU，而與載板或電路板的連結則透過錫球實現。

而在 ( c ) 圖的第二種設計中，Active Bridge 先貼附在矽晶片基板上，而 2 個 GPU 小晶片則貼合在 Active Bridge 的上方，中間的空間以介電材料填充。由於 Active Bridge 在下方，GPU 小晶片必須設置穿透式的 TSV 以形成 Active Bridge 所需的對外訊號連接通道。在這個設計中，矽晶圓基板同樣必須翻轉向下，再透過錫球與載板或電路板形成封裝。

透過圖四的結構，可以清楚掌握本專利申請案中 Active Bridge 與 GPU 小晶片的 3D 堆疊封裝方式，有助於進一步了解 AMD 的 chiplet 晶片結構與封裝設計。

圖四、搭配 2 種基本電路，可實現任意邏輯運算



資料來源：USPTO · MIC 整理 · 2021 年 07 月

## MIC 觀點

### 將L3 cache安置在Active Bridge上除了增加記憶體容量，也可增加GPU運算單元的堆疊密度，有助於提升運算效能

從傳統 2D 的晶片中，快取記憶體在晶片中是占據一定面積的。隨著晶片運算效能需求的提升，晶片中運算元件的數量必須增加，無疑地將會與快取記憶體爭奪晶片可用空間，但快取記憶體又是高效能運算中資料暫存不可或缺的部分，一旦容量過小將會影響運算中資料的存取速度，導致運算效能降低。上述的兩難使得運算單元與快取記憶體的空間規劃成為高階晶片設計的重大挑戰。

AMD 利用 3D active-on-active 堆疊技術將 L3 cache 由平面的 GPU 小晶片中移出，轉而建置在 Active Bridge 上，這樣的設計除了使晶片設計者有空間規劃足夠的 L3 cache 容量，從記憶體層面提高運算效能外，事實上還有另一個重要的優點，就是在 GPU 小晶片中除去了 L3 cache 所需的空間，使 GPU 小晶片做到同樣運算單元數量下面積更小或是同樣面積下規劃更多的運算單元，從而增加 GPU 運算單元的堆疊密度。無論是記憶體容量還是運算單元密度的增加，都足以提升運算效能，是非常棒的設計。

### Chiplet成為高階晶片開發的重要技術選擇，台廠應提前布局

隨著資料中心與邊緣運算需求持續增加，不管是 CPU、GPU、ASIC 或 SoC，對於運算效能提升的需求將不斷增加，但在高密度的運算單元堆疊下，晶片製程的良率與運算中產生的大量熱能成為巨大挑戰。相較於 monolithic 單一晶片製程所製作的晶片，chiplet 較小的面積與相對簡單的電路將可提高製程良率，而不論是 2D 或 3D 封裝，小晶片也增加了利用封裝結構分散熱能的可能性。從 AMD 此次揭露的技術與產品來看，chiplet 的架構設計已成為高階運算晶片開發不可忽視的一個方向。

從 Active Bridge 結構來看，採用 chiplet 設計的高階運算晶片對於先進封裝有較高的需求，也將使 AMD 等晶片商對於台積電的依賴程度持續增加。在這樣的趨勢下，台積電以及其先進封裝的設備、材料夥伴等更需提前布局，強化先進封裝產能的良率與穩定供給，才能在未來與高階晶片供應商共榮發展。



發行所	財團法人資訊工業策進會 產業情報研究所(MIC)
地址	台北市 106 敦化南路二段 216 號 19 樓
電話	(02)2735-6070
傳真	(02)2732-1353
全球資訊網	<a href="https://mic.iii.org.tw">https://mic.iii.org.tw</a>
會員服務專線	(02)2378-2306
會員傳真專線	(02)2732-8943
E-mail	<a href="mailto:members@micmail.iii.org.tw">members@micmail.iii.org.tw</a>
AISP 會員網站	<a href="https://mic.iii.org.tw/aisp">https://mic.iii.org.tw/aisp</a>

以上研究報告經 MIC 整理分析所得，由於產業變動快速，並不保證上述報告於未來仍維持正確與完整，引用時請注意發布日期，及立論之假設或當時情境。  
著作權所有，非經 MIC 書面同意，不得翻印或轉載