

AI GPU Link 架構及供應鏈 專家會議

重點摘要

1. AI GPU 除了製程節點影響算力外，GPU 互聯數量、連接能力等亦將影響整體算力。
2. 專家表示 CPO 版本之每單位傳輸速率單價較目前光收發模組為低；預期矽光子將打破現有光收發模組成本結構，其中光晶片封裝將是成本較高的部分。
3. 專家表示目前 CPO 架構在電晶片、雷射、矽光子及封裝上皆已發展至一定程度，而 fiber 至矽光子的封裝之生產效能提昇，仍有待觀察。

結論

產業專家表示目前 NVLink 已可達 6.4T 架構，而台灣晶圓製造領導廠商之 2026 年 CPO roadmap 亦為 6.4T，因此若 CPO 跟上則可以直接替換目前的銅線架構。若單一機櫃算力成長 20 倍以上，則銅線版本應會消失，而 1.6T 光收發模組在不論幾倍算力都是關鍵產品，係因銅線轉換成光纖 Data rate 將提升，GPU 互聯數越高，算力及成本效益都會變好。

AI GPU 世代

- 製程節點：GB200 為 N4；Rubin 為 N3；AMD MI325X 及 Intel Gaudi 3 皆為 N5
- GB200 2024 年末到 2025 年初出來 Rubin 預期 2026 年初
- 16 浮點(FP16)：NV H100/AMD MI325X/Intel Gaudi 3 算力分別為 2 /1.3 /0.5 PetaFLOPS，製程節點決定算力
- 4 浮點(FP4) NV H100 為 18 PetaFLOPS；AMD MI325X 則為 5.2 PetaFLOPS；GB200 更已達 40 PetaFLOPS；FP4 算力更加強調軟體能力

GPU 互聯

- Hopper GPU 為 8 顆；GB200 NVL72 可達 72 顆
- AMD MI325X 為 8 顆；Intel 為 48 顆
- 互聯 data rate NV 為 100Gb per channel (NVLink 架構)；AMD 及 Intel 分別為 32Gb、25Gb，且皆為 PCIe 架構
- NVL72 中一個 Switch 為 72 個 800G 通道連接為基本架構
- Hopper 轉移至 Blackwell 總體算力提升 45 倍，整體算力與互聯數非常相關
- 機櫃間為使用 800G 光互聯



請將您珍貴的一票投給凱基！

凱基投顧

姜兆剛
886.2.2181.8742
jackson.chiang@kgi.com

孫瑞隆
886.2.2181.8745
sean.sun@kgi.com

重要免責聲明，詳見最終頁

- NVL72 機櫃中是用銅線去連接，當中有 2 顆 switch IC，1 顆 switch IC 旁有 9 顆 connector，共 $2 \times 9 = 18$ 個 connector，為美系連接器大廠產品；產品上有 64 組銅線 $\times 100G = 6.4T$ (總速率)
- 台灣晶圓製造廠 CPO roadmap：2026 年 CPO 為 6.4T，NVLink 銅線版本一開始即為 6.4T 架構，未來若光架構跟得上則可以直接替換
- 100G 訊號沒有在 PCB 傳輸，是用銅線拉出，機櫃後面可以看到很多銅線
- 一個 Switch tray 上有 9 組 6.4T DAC 加上另一個 9 組 6.4T DAC 架構，整體共有 $18 \times 9 = 162$ 組銅線

CPO

- 主 IC 旁放置矽光子，矽光子為調變器，將主 IC 電訊號經由矽光子調變器，經光纖跳線傳出，光學版本為 CPO；而銅線版本(電傳輸)是 CPE (co packaged electronics)
- CPO 6.4T 出得來的話可以直接替換 CPE
- 台灣晶圓製造廠 2026 年要達成 6.4T CPO，主因要配合美系晶片廠商於 2026 年推出 Rubin 架構，矽光子 1.6T 在 2025 年就要出現
- 美系連接器大廠 connector 一條銅線可以支援 200G (224G)，所以 6.4T CPE 可以升級為 12.8T；光學連接則是 800G 升級 1.6T，所以 2025 年光模組廠商重點為 1.6T
- 機櫃跟機櫃走 1.6T 光架構，機櫃中銅線則為 200G，Rubin 硬體架構已完成，看光學版本能不能跟上
- 每個 GPU 分配 4.5 個 6.4T CPE head，CPE 總成本約 3,000 美元，則一個 CPE head 約為 667 美元，換算 1Gb 約為 0.1 美元(銅線)；光版本則約為 0.6-0.7 美元，因此銅線成本優勢高(銅線無主動元件)，功耗亦較光電元件低
- 800G 光收發模組市佔份額中國廠商為 60%；美國廠商 40%
- 美系晶片廠商 2023 財年 800G 光模組代工費用約 3 億美元
- GB200 NVL 年度需求約 844 萬的 800G AOC head 及 1900 萬個 6.4T copper cable head，銅線需求可由美系連接器大廠財報驗證

Rubin 系列可能發展

- 假設由 Blackwell 轉換至 Rubin，其整體算力與 Hopper 轉往 Blackwell 算力成長同樣為 45 倍，以此計算 R 系列需要互聯 648 個 GPU，機櫃應放不下；DAC head 則有 10,000 多個，機櫃亦放不下，下個世代可能要分櫃，則有利光收發模組、連接器需求，由其 1.6T 光收發模組需求量大增
- 銅版本轉換成光版本 Data rate 會拉升，GPU 互聯數越高，算力及成本效益都會變好
- 單一 GPU 算力成長五倍，單就 N4 往 N3 走，基本算力可成長 3 倍
- 互聯數由 72 個 GPU 往 144 個 GPU (單位面積由 2 顆增加至 4 顆)，CPE 升級至 200G，DAC head 升級至 12.8T，此架構僅需 324 組 CPE cable，面積及體積上可達成上做得到的，單一機櫃算力可望成長 10 倍

- CPO(光學版本)648 組 6.4T 同樣放得下，因一個銅線版本尺寸可以放進兩顆 CPO
- 若單一機櫃算力成長 20 倍以上，則銅線版本應會消失，而 1.6T 光收發模組在不論幾倍算力都是關鍵產品
- 美系晶片廠商硬體架構已完成，台灣晶圓製造廠晶片沒問題，剩下需觀察 fiber connector 是否能跟上，若能跟上對台廠有利
- GB200 每 Gb(光+銅)成本約為 0.14 美元，總計需求約 80 億美元
- 假設 Rubin 平台算力需求與 GB200 NVL 相等，則需要約 1,944 萬個 CPO，整體金額規模 180 億美元；推算 6.4T CPO 約 900 美元(對比 CPE head 約為 667 美元)

Q&A

- 整體 CPO 光+電(銅)每 Gb 單價 0.14 美元；銅則為 0.1 美元，若對比目前光收發模組單價？
 - CPO 光+電版本確實較光收發模組價格低
 - 光晶片成本較高處為封裝，毛利率非常好；矽光子會將原本光收發模組成本結構打破，美系 GPU 廠商預測 CPO 1 個 bit 是 10^{-10} 美元
 - 技術成本可行，但供應鏈的量能否趕上是問題
- 什麼零組件在往 R 系列值得關注？(content value 提升)
 - CPO 起來會讓光通訊產業架構重塑，CPO 自己就是光收發模組，產業鏈具重整機會
- CPO 採用率提升，DSP 是否有持續採用機會？
 - DSP 為模擬 Switch IC 的晶片，CPO 整合度越高，DSP 就不需要，因為 DSP 較為耗能
 - 長遠來看 CPO 並非最佳解
 - CPO 已可達 6.4T 12.8T，目前 CPO 具 ELS 架構
 - 未來若雷射放到矽光子後，電的功率會降低，光纖會被拿掉，未來關注是否 fiber connector 有新的架構
- 光晶片未來是由什麼廠商來設計？
 - 目前是主晶片廠商(如美系 GPU 廠商或特定應用系統商)主導矽光子晶片
 - 未來會有矽光子 fabless 出現，2018-19 年有一批矽光子公司被併購，2030 年後有一批矽光子公司會出現在較好的環境，因為過往供應鏈尚未跟上
- CPO 產業鏈哪部分仍須跟上？
 - 電晶片由美系 GPU 廠商主導沒問題；雷射晶片 6.4T 全世界至少三家合規公司；矽光子台灣晶圓製造廠商及後端 CoWoS 沒問題
 - 光纖組裝則較值得觀察
- 有什麼美系晶片廠商/台灣晶圓製造商以外的生態系？
 - 目前仍由美系晶片廠商綁死

- 另一美系 GPU 廠商可能仍需要五年，主因進入時間過晚
- 合聖在 CPO 扮演的角色？
 - 合聖主要做 CPO 中的 ELS，2021 年開始做，目前有 6.4T CPO ELS 產品
- 調變器技術發展？
 - Silicon modulator 的速度極限在 200-400G，可能導致矽光子的訊號調制技術，改為鋇酸鋰、III-V、polymer 等材料發展。
 - 矽光子可能朝向光載板發展，主動元件本來就是 chiplet，如何將雷射整合到矽光子上是重要問題

凱基證券集團據點

中國	上海	上海市靜安區南京西路 1601 號越洋國際廣場 1507 室 郵政編號：200040
台灣	台北	104 台北市明水路 700 號 電話 886.2.2181.8888 · 傳真 886.2.8501.1691
香港		香港灣仔港灣道十八號中環廣場四十一樓 電話 852.2878.6888 · 傳真 852.2878.6800
泰國	曼谷	8th - 11th floors, Asia Centre Building 173 South Sathorn Road, Bangkok 10120, Thailand 電話 66.2658.8888 · 傳真 66.2658.8014
新加坡		珊頓大道 4 號#13-01 新交所第二大廈 郵政編號：068807 電話 65.6202.1188 · 傳真 65.6534.4826
印尼		Sona Topas Tower Fl.11 Jl. Jend. Sudirman kav.26 Jakarta Selatan 12920 Indonesia 電話 62 21 250 6337

股價說明

等級	定義
增加持股 (OP)	對個股持正面看法，預期個股未來十二個月的表現超越凱基證券集團所追蹤的相關市場的總報酬。
持有 (N)	對個股持中性看法，預期個股未來十二個月的表現符合凱基證券集團所追蹤的相關市場的總報酬。
降低持股 (U)	對個股持負面看法，預期個股未來十二個月的表現低於凱基證券集團所追蹤的相關市場的總報酬。
未評等 (NR)	凱基證券未對該個股加以評等。
受法規限制	受凱基證券集團內部政策和/或相關法令限制使凱基證券集團無法進行某些形式的資訊交流，其中包括提供評等給投資人參考。
未評等 (R)	
*總報酬 = (十二個月目標價-現價)/現價	

免責聲明

凱基證券投資顧問股份有限公司（以下簡稱本公司）為凱基金控集團之成員。本報告之內容皆來自本公司認可之資料來源，但不保證其完整性及精確性。報告內容所提及之各項業務、財務等相關檔案資料及所有的意見及預估皆基於本公司於特定日期所做之判斷，故有其時效性限制，邇後若有變更時，本公司將不做預告或更新。本報告內容僅供參考，並不提供或遊說客戶為買賣股票之投資依據。投資人應審慎考量本身之投資風險，並就投資結果自行負責。本公司及所屬集團成員，暨其主管或員工皆有可能持有報告中所提及的證券。本公司及所屬集團成員並可能經常提供投資銀行或其他服務給報告中提及之公司或向其爭取相關業務。本報告之著作權為本公司所有，非經本公司同意，本報告全文或部份內容，不得以任何形式或方式引用、轉載或轉寄。