

AI ASIC 產業概述

組員：

陳風年、楊子甫、鄭凱軒、邱歆媚、
吳秉翰、林御庭、謝亞臻、方翊倫

產業介紹

- 雲端 AI ASIC 市場持續高速成長，預估2030 年市場規模將達 500 億美元，CAGR 約 25%。
- ASIC 雖在絕對算力和晶片間互聯方面無顯著勝於，但 AI ASIC 的可訂製化及成本優勢依舊為其在推論端鞏固市場定位。
- AI ASIC 供應鏈採台美分工，美方掌握前端設計與 IP，台廠專注特定應用與後端整合設計，雙方互補推動產業發展。

成長動能

- 全球 CSP 持續擴張資本支出，自研 ASIC 以維持晶片自主與滿足低成本、高效運算需求。
- 隨著 AI 應用逐步進入推論階段，晶片設計將更注重成本效益與特定場景優化，呈現客製化、高性價比及區域化趨勢，進一步推動 ASIC 需求成長。

推薦個股

- 博通(AVGO)

產業介紹



產業介紹

成長動能

潛在風險

投資標的

結論

FIIRC ASIC 半導體供應鏈

上游 設計 & IP

IC 設計工具 & 服務

IP & EDA

ARM、SNPS、CDNS

IC 設計服務

世芯-KY、創意、智原

IC 設計

AVGO、MRVL、聯發科

中游 製造 & 代工 & 封測

晶圓製造

Pure-Play Foundry

台積電、聯電、GFS、中芯國際

IDM + Foundry

INTC、Samsung

封裝測試

日月光、力成科技、AMKR、台積電、京元電、INTC

下游 應用

數據中心

Google、Amazon、Meta、Microsoft、Enterprises

網通 & 5G

CSCO

消費性電子

車用電子

ASIC跟GPU的比較

- AI ASIC 以高效能與低功耗取代通用 GPU。
- 犧牲彈性與開發成本換取專用化優勢。
- AI ASIC成為雲端巨頭推進 AI 運算的核心架構。

AI ASIC 與 GPU 架構與應用定位比較

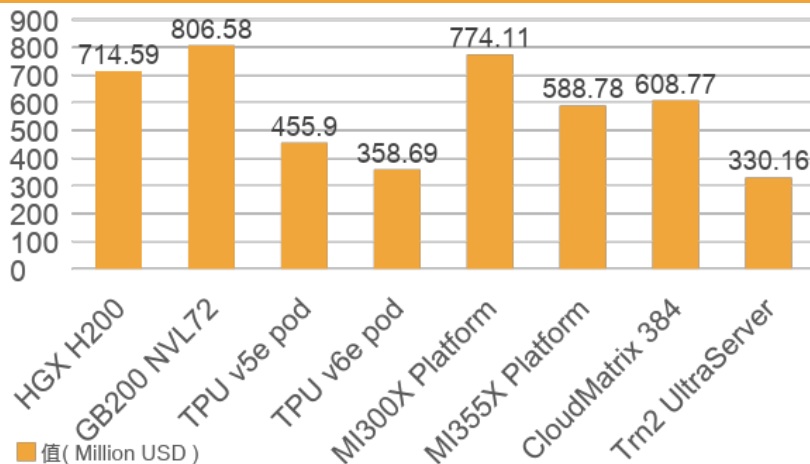
項目	GPU(通用架構)	AI ASIC(客製化架構)
運算架構	通用運算核心 (CUDA / ROCm) 支援多種 AI 模型	任務導向電路 (TPU / Trainium / MTIA) 針對特定模型優化
效能密度	中等	高 (>3 × GPU)
功耗效率	一般 (需大量 GPU 並行)	高 (能耗比提升 3–5 倍)
開發彈性	高，適合快速開發與多用途	低，設計固定但效能極佳
開發成本/週期	開發成本低、導入快	開發成本高、開發週期長
主要應用	AI 訓練	AI 推理



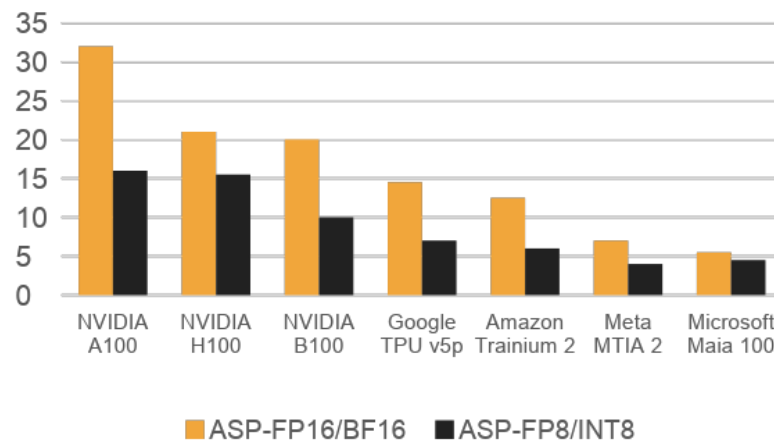
ASIC 具成本優勢

- ASIC 的硬體結構為特定任務量身定制，省去了許多針對通用加速計算的不必要硬體設計，其單位算力成本因此相較於 GPU 更低。
- 左表在相同目標吞吐量下，各平台的 TCO 存在明顯落差，ASIC 整體成本相對通用 GPU 平台可少到約 40–60% 的量級。
- 右表柱狀比較顯示 GPU 代際雖有降價趨勢，但圖表可以明顯看出雲端自研 ASIC 單位成本更低

ASIC、GPU 推理端(Llama 2 70B model)總體成本



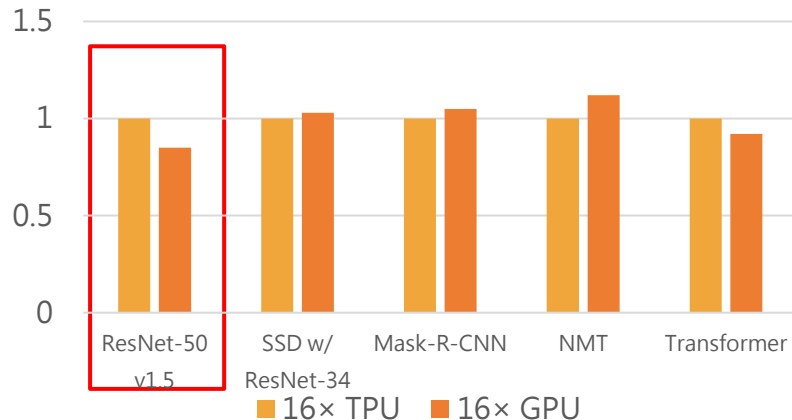
市面主流ASIC、GPU 算力成本對比



ASIC 對專門設計之效用有優勢，GPU則在總體算力佔優

- GPU 的通用性與相較於 ASIC 高，以及更加強大的並行運算能力使其在要求運算精度以及非線性運算的 AI 模型訓練端有更好的處理能力。而 ASIC 是一種針對特定應用訂製的晶片，在執行單一或特定類型的任務時具有更低的能耗表現。AI 模型推理端因模型權重已經固定，僅需執行大量矩陣運算，對運算精度的要求也較訓練端低，因此 ASIC 在此場景能夠有效發揮其優勢。

ASIC、GPU 訓練端對架構專門模型下效用



ASIC、GPU 之適用情景

GPU 更適用於訓練

- AI 訓練過程需要處理大量的資料與複雜的計算，對晶片的運算能力以及平行處理能力要求高。
- 而GPU 對於**並行運算能力較優**；高度靈活性也使其更適合這種頻繁的調試與反覆迭代。

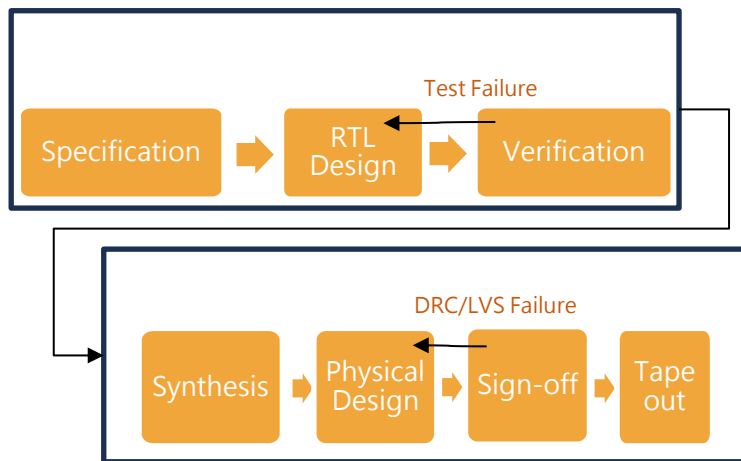
ASIC 更適用於推理

- 在推理階段，需要對輸入的資料進行快速的預測與分類。
- ASIC 高度客製化的特性針對推理任務進行最佳化，以較低的功耗實現高速的**推理運算**。因此ASIC 的成本優勢更加明顯，有效降低企業的營運成本。

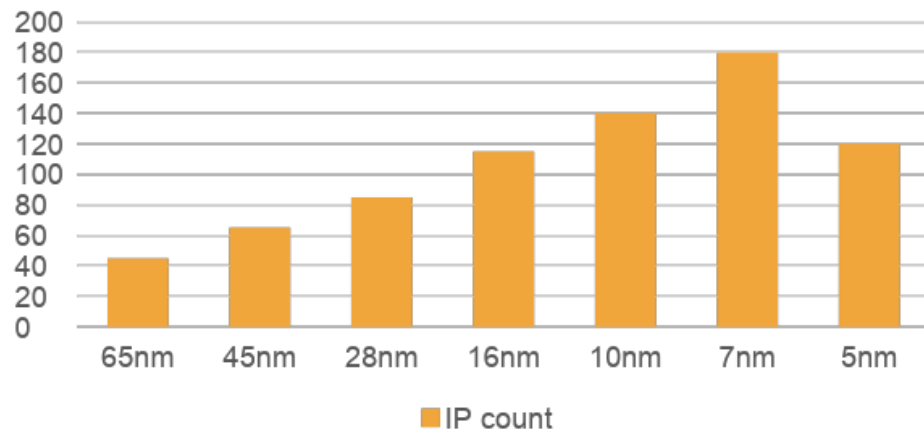
前後段設計流程及分工

- 晶片設計流程複雜，前段設計負責架構與RTL藍圖，包含關鍵的IP整合；後段設計則是將其轉為實體佈線並完成交付 (Tape-out)。美國 CSPs 競逐 AI 市場同時，必須掌握客製化 IP 技術含量來創造差異化。
- 先進晶片對 IP 的依賴度極高，因此通常由掌握核心 IP 技術的美廠負責，近年也不少新晶片專案由 CSP 親自操刀前段，將複雜耗時的後段工作外包，以專注雲端核心業務。

前後段設計流程

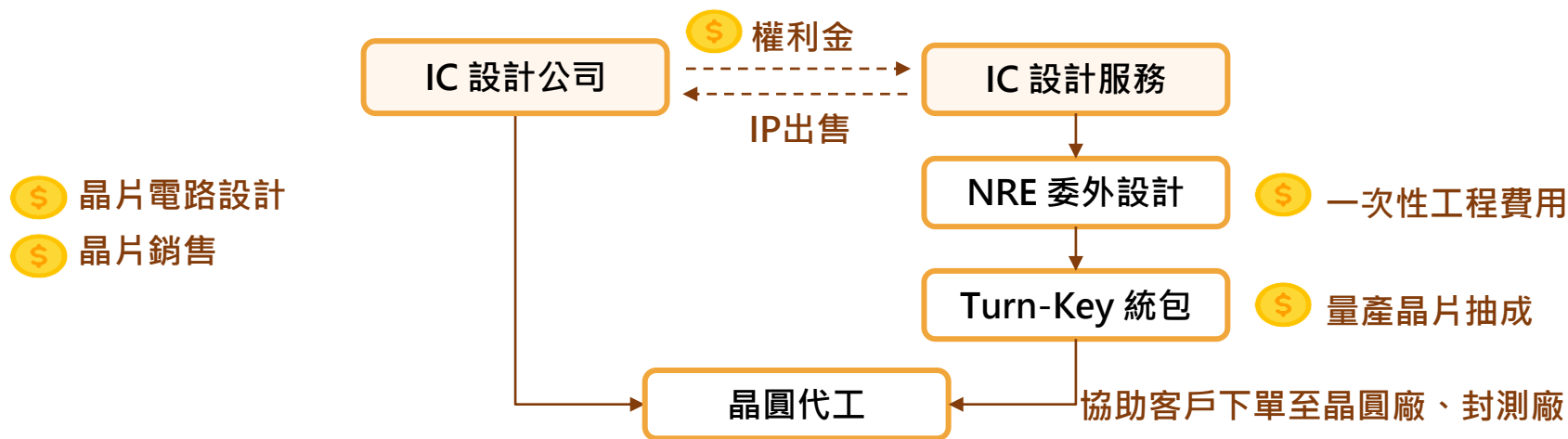


Marvell IP 資源 (SerDes)



FIIRC IC 設計產業分為晶片設計與設計服務

- **IC 設計公司**專注於晶片電路設計，負責將產品規格轉化為可製造的電路設計，使用 EDA 等工具並大量應用 IP 模組或外發 IC 設計服務公司來縮短開發周期、降低成本，並委由服務設計公司實施後端物理佈局與投片。
- 而 **IC 設計服務公司**如世芯-KY、創意則提供後端的物理佈局設計以及解決開發設計時的技術問題，並協助客戶完成設計以及至晶圓廠投片，主要以 NRE 收費，後續獲利來自授權與量產需求。

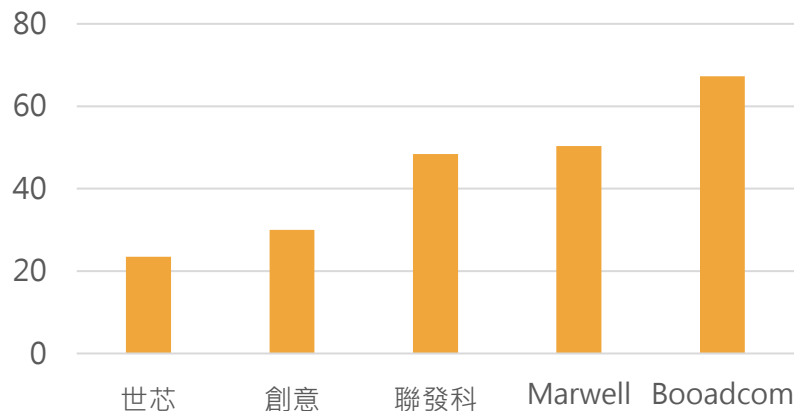




商業模式比較

- 在 AI ASIC 專案之中，台系廠商與美系廠商皆逐漸以 **Turnkey 模式** 作為主要營收來源，協助客戶完成從後端設計到量產交付的全流程服務。此模式強化了其在供應鏈中的主導角色。然而，從產品組合與獲利結構來看，美系廠商 **NRE 業務** 占比較高，毛利率因此普遍優於台系廠商。整體而言，Turnkey 模式雖提高營收穩定性，但相對壓低了單案獲利水準。

台美廠毛利率比較 (2025)



商業模式比較

項目	NRE	Turnkey	COT
定義	一次性費用	統包服務	一次性設計
客製化程度	高	低	高
收入模式	一次性	隨出貨量增加	一次性
量產責任	客戶	IC設計服務公司	客戶
毛利率	高	中	高

CSP投入ASIC研發

- AWS 與 Marvell 合作開發 Trainium v2，並攜手 世芯電子（Alchip）進行 Trainium v3，2025年 ASIC出貨量預估為美系CSP中成長最快。Meta 與博通（Broadcom）共同開發 MTIA v2，聚焦於能效最佳化與低延遲推論架構。Microsoft 初步研發 Maia v2，由 創意電子 負責設計並與 Marvell 合作開發進階版。Google 推出 TPU v6 Trillium，主打高能效比與大型模型最佳化，並新增與聯發科合作以強化新一代AI晶片研發。

各CSP自研ASIC

業者	晶片名稱	製程	2025採用比例
Google	TPU v5/v6/v7	5-3 nm	76%
AWS	Trainium、Inferentia	7-3 nm	23%
Microsoft	Maia	5-3 nm	4%
Meta	MTIA	5-3 nm	13%
Intel	Gaudi 2/3、Falcon	7-3 nm	2%

成長動能

產業介紹

成長動能

潛在風險

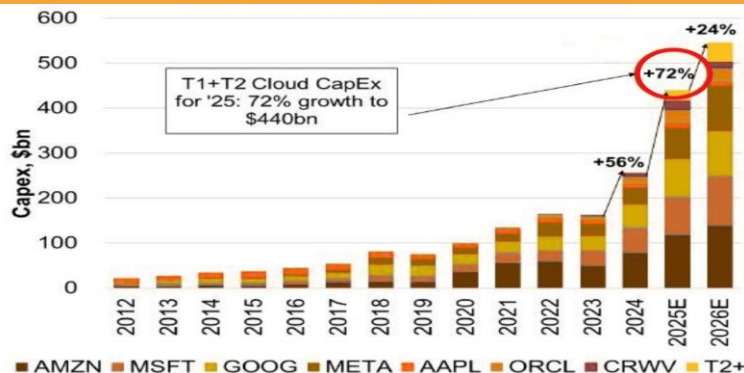
投資標的

結論

CSP資本支出持續擴張，AI 投資動能依然強勁

- 全球 CSP 正掀起新一波資本支出擴張潮，預估至 2026 年，全球 CSP 資本支出將突破 5,000 億美元，反映生成式 AI 持續推動產業投資。2025 年四大 CSP 的資本支出預估達 3,510 ~ 3,570 億美元，年增近 40%，AI 基礎設施為主要成長動能。四大 CSP 為 AI ASIC 的主要需求來源，且具備自研與大規模導入能力，目前在高階 AI 伺服器採購量已佔全球逾六成。
- 根據預測，2025 年 Tier 1 與 Tier 2 CSP 的資本支出年增率將達到 72%，顯示業者正積極投入 AI 伺服器與自研晶片。2026 年雖然增速放緩至 24%，但整體投資動能依然強勁，意味著 AI 已成為長期結構性趨勢。

全球CSP資本支出總額(十億美元)



公司	25年資本支出 (億美元)	YoY
Google	850億	61.9%
Amazon	1,000 (約80%在 AWS)	50.2%
Meta	660 ~ 720	68.2% ~ 83.5%
Microsoft	900 ~ 1,000	19.0% ~ 32.2%

CSP為避免輝達壟斷定價，有自研晶片之需求

NVIDIA 25 財年毛利率達 75%，即便在部分季度因庫存損失下降，也仍維持在 70% 左右的非 GAAP 水準。相較之下：Broadcom 約 66–67% Marvell 約 60–62%。這種**超過產業均值 10–20 個百分點的毛利率**，代表輝達不僅在設計能力上領先，也在供應鏈與價格談判上幾乎具有壟斷力量。

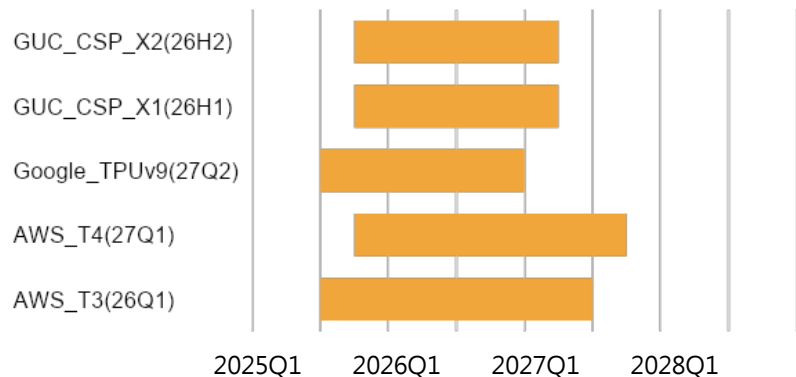
由於高階AI晶片的產能有限，晶片交期普遍出現遞延現象，導致 CSP 的資料中心建置時程被迫延後。這不僅影響伺服器部署與算力擴充進度，也使雲端服務上線與 AI 訓練計畫出現時程落差。為了降低對單一晶片供應商的依賴，CSP 需要具備議價能力，包括透過自研 ASIC、導入多元供應鏈或與晶片廠進行深度協同開發，以確保長期供貨穩定與成本可控。

長期趨勢：高毛利壟斷 → 尋求替代 → ASIC 生態崛起 2027 以後：CSP ASIC 市佔逐漸上升，壟斷結構轉為寡佔

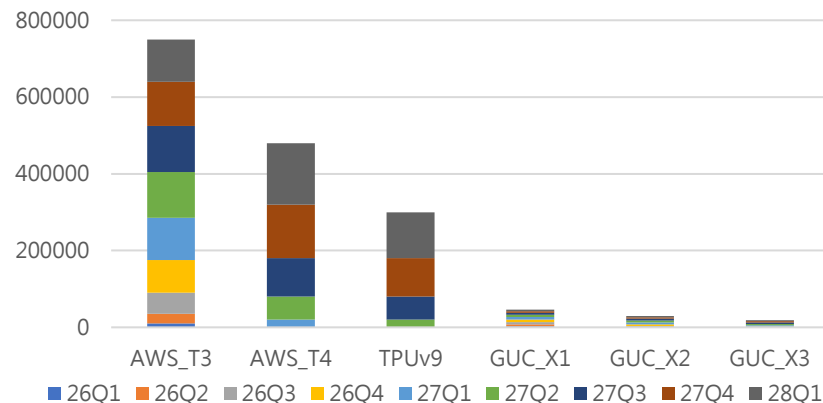
2026–2028：量產節奏與出貨曲線

- AI ASIC 供應鏈自上而下可分為上游的先進製程與先進封裝（TSMC / CoWoS / HBM）、中游的設計服務商（如 Alchip、GUC、MediaTek、Broadcom、Marvell），以及下游的雲端服務客戶（Google、AWS、Meta、Microsoft）。隨著 3nm 向 2nm 過渡與 HBM4 導入，設計服務與投片需求同步擴大，世芯與創意受惠於大型 CSP 專案推進；然而 CoWoS 產能與 HBM 供給仍構成短期瓶頸，而 CSP 為壓縮時程採 COT 模式，雖能加快落地但對中游毛利形成擠壓。後續觀察重點包括 2nm / HBM4 的量產時程、CSP 專案轉量進度，以及中游 ASP 與 NRE 的動能變化。

主要案別 Tape-out 與量產起始（季度）



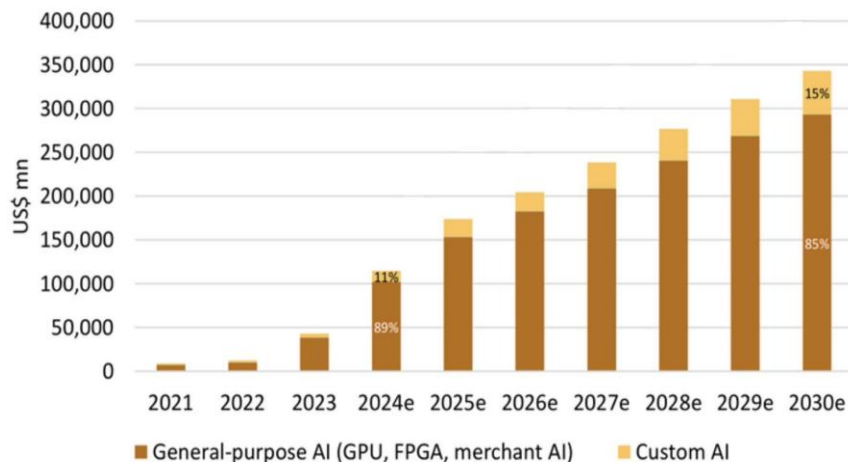
各專案季度出貨量貢獻



ASIC 市場規模將以 25% CAGR 穩定成長

- 根據預期，雲端 AI ASIC 市場將持續高速成長，市場規模將從 2024 年的 132 億美元提升至 2030 年的 500 億美元，年均複合成長率 (CAGR) 約 25%，屆時將佔整體雲端 AI 晶片市場約 15%。
- 在出貨量方面，2025 年至 2029 年將由 558 萬顆增加至 1,196 萬顆，CAGR 約 21%，顯示雲端業者對自研 AI 晶片的需求快速上升，成為 AI 基礎建設成長的重要推動力。

ASIC市場規模預估(百萬美元)



高階ASIC出貨量

K unit	2024	2025	2026	2027	2028	2029
出貨量	4082	5581	7620	9906	11391	11961
YoY	202%	37%	37%	30%	15%	5%

潛在風險

產業介紹

成長動能

潛在風險

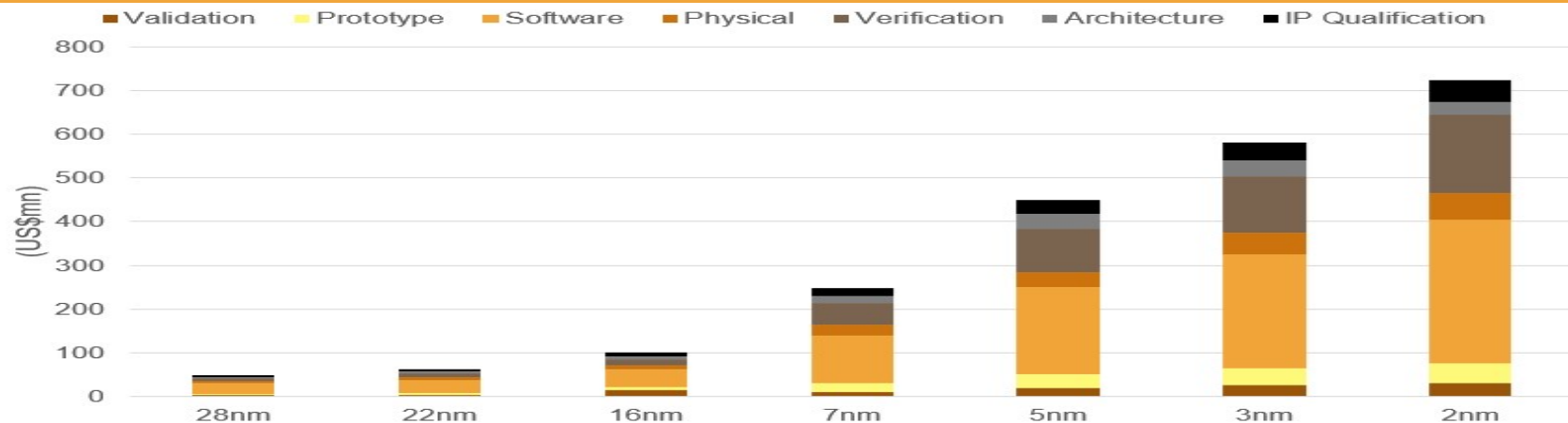
投資標的

結論

ASIC 高額建置成本以及開發失敗損失使其無法於中小企業站穩根基

- ASIC 設計成本隨製程節點呈指數級增加，而因為客戶在需花費大量時間進行晶片軟硬整合，使得軟體成本佔總體開發成本很大的比重。因此，ASIC 儘管在能耗以及大量部署成本上擁有優於 GPGPU 的表現，但其初始的佈置成本卻高於後者。而若開發出現延誤或者技術問題，前期投入之成本可能付之一炬。因此無法大規模部署實行成本效益的中小廠商可能更傾向購買輝達現成的解決方案，使得 ASIC 在與 GPGPU 的競爭中落入下風。

各節點晶片研發成本



投資標的

產業介紹

成長動能

潛在風險

投資標的

結論

Broadcom Inc (AVGO)

結論

公司主要業務	半導體解決方案、企業軟體 (VMware、CA、Symantec)
2025 Q3營收(十億美元)	16.0
公司市值 (兆)	1.77
收盤價	376.47
主要競爭對手	NVIDIA、AMD、Marvell、Intel、Cisco
產品應用	雲端運算、5G 基礎建設、企業級路由器與交換機 Wi-Fi 6、Wi-Fi 7、藍牙晶片

成長動能

AI 半導體業務持續增長

博通的 AI 半導體業務自 2025 財年第三季以來，營收持續增長。該季 AI 半導體營收達到 52 億美元，較前一年同期增長 63%，並預計第四季將增長 66%，達到約 62 億美元。其中 AI 半導體與 VMware 業務為主要成長動能

軟體業務成長亮眼

截至 2025 財年第 3 季，博通的基礎設施軟體部門（主要由 VMware 貢獻）營收達 68 億美元，較去年同期增長 17%，佔總營收的 43% 該部門的調整後 EBITDA 利潤率達 67%，顯示出強勁的獲利能力。

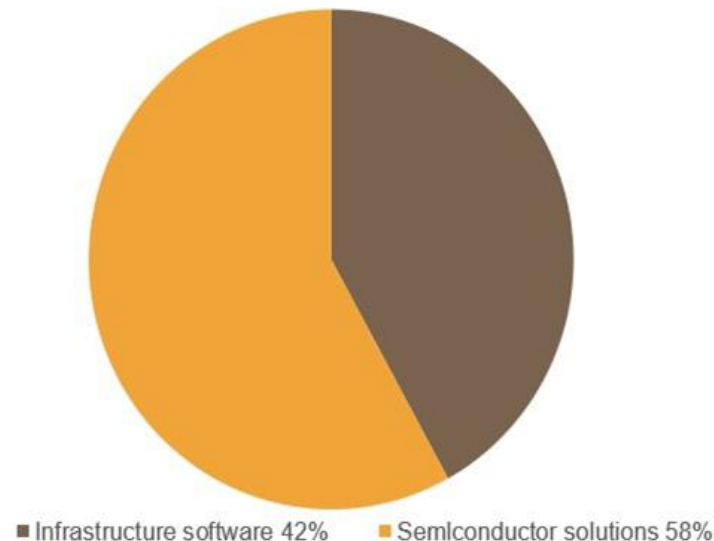


Broadcom Inc (AVGO.US)

2025Q3財報摘要

財務項目	數值(十億美元)	年增率	說明
淨利	5.5	37%	受 AI 晶片與 VMware 帶動
調整後 EBITDA	10.7	30%	盈利能力穩健
自由現金流	7.0	47%	現金流極大
毛利率	75%(非GAAP)	27%	居半導體產業高位
每股盈餘	1.69(美元)	36%	穩定且持續增長

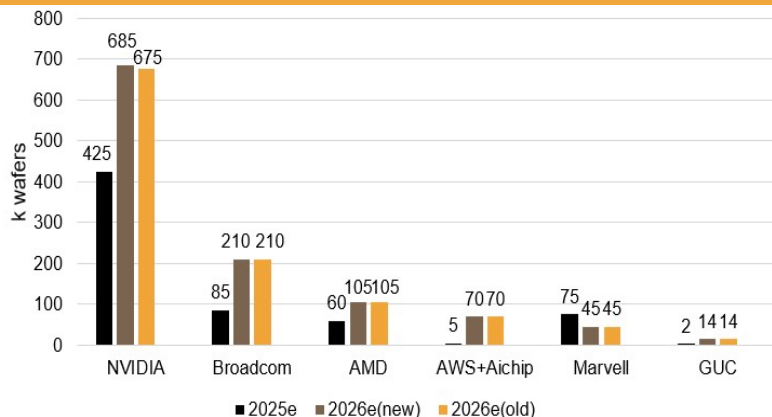
博通營收占比



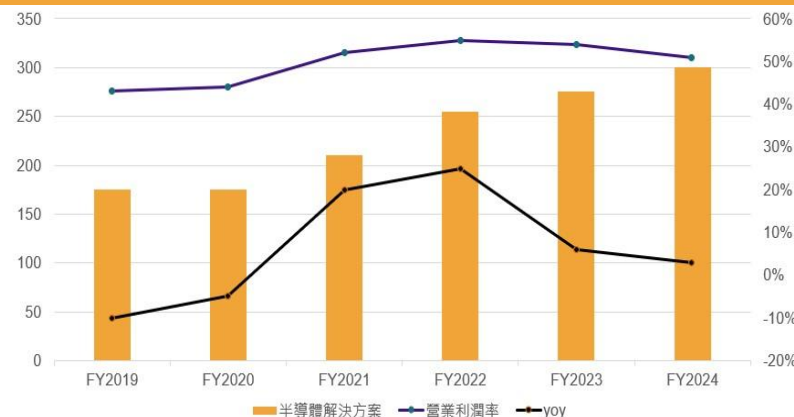
AI 晶片推動先進封裝與博通半導體營收加速

- 博通的成功故事，正是 AI 運算需求和先進封裝技術緊密結合的典範。由於來自 Google 3nm TPU v7 的強勁需求，博通 2026 年 CoWoS 晶圓預訂量已經從原先的 145k 上修至 205k 片。這充分證明了其訂單充足，並深受 CSP 青睞。
- 博通 AI 業務佔其半導體解決方案部門收入的比例，已經從 2019 財年的不到 5%，大幅提升至 2023 財年的約 15%。

全球 CoWoS 需求的年增長趨勢概況



博通半導體解決方案收入

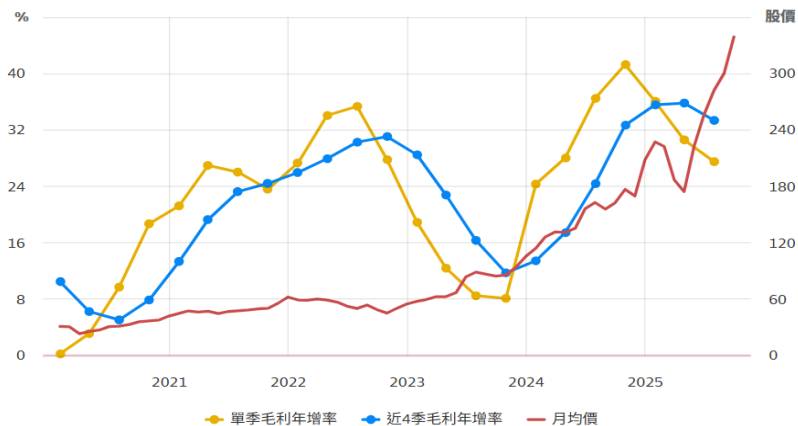




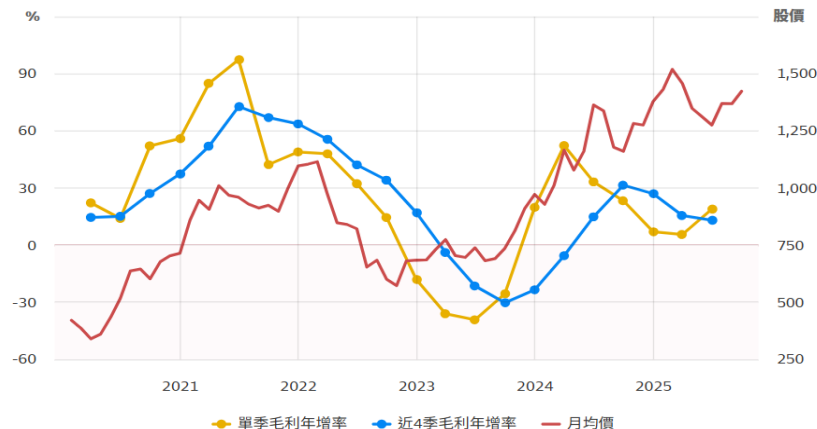
博通毛利率分析

- 博通的最新毛利率高達 67%。相較之下，作為 TPU 競爭者的聯發科的最新毛利率為 49%，儘管已創五季新高，兩者之間仍有約 18 % 的落差。
- 博通之所以能保持近七成的毛利，是因其在極高門檻領域建立的地位：聯發科雖然在 SerDes 技術掌握優勢，但博通憑藉著大量的 IP 儲備及 CoWoS 單量贏得 CSP 廠商信任。

Broadcom 歷年毛利率數據



聯發科歷年毛利率數據





Broadcom Inc (AVGO)

競爭優勢

技術	擁有 SerDes、CPO 等關鍵 IP，主導 AI ASIC 前端設計架構。
供應鏈	與台灣後段設計服務廠（如世芯、創意）協同合作，形成高效率 AI 晶片製造鏈
客戶綁定能力	深度合作於 AWS、Meta、Google 等大型 CSP，長期專案穩定。
商業模式	以 turnkey 模式（設計到量產統包）強化營收穩定性

潛在風險

市場競爭	NVIDIA 壟斷 AI 晶片訓練，短期內GPU替代難度高。
需求波動	若 CSP 資本支出增速放緩，短期成長可能受影響。
併購整合風險	VMware 整合期仍需時間，短期管理複雜度上升。
地緣政治風險	中美科技管制可能影響供應鏈與客戶布局。

結論

產業介紹

成長動能

潛在風險

技術演進

結論



產業介紹

- 雲端 AI ASIC 市場持續高速成長，預估2030 年市場規模將達 500 億美元，CAGR 約 25%。
- ASIC 雖在絕對算力和晶片間互聯方面無顯著勝於 GPU，但 AI ASIC 的可訂製化及成本優勢依舊為其在推論端鞏固市場定位。
- AI ASIC 供應鏈採台美分工，美方掌握前端設計與 IP，台廠專注特定應用與後端整合設計，雙方互補推動產業發展。

成長動能

- 全球 CSP 持續擴張資本支出，自研 ASIC 以維持晶片自主與滿足低成本、高效運算需求。
- 隨著 AI 應用逐步進入推論階段，晶片設計將更注重成本效益與特定場景優化，呈現客製化、高性價比及區域化趨勢，進一步推動 ASIC 需求成長。

推薦個股

- 博通 (AVGO)