Google TPU 產業與供應鏈

晶片 × 板子 × 機櫃 × 網路 × 散熱電源 × 台股供應鏈 × 市場說法

stock_llm_wiki | 2026-10-14

報告大綱

01 **TPU 是什麼** 架構、歷代演進、和 GPU 差在哪

02 **需求與出貨** Capex、外部客戶、出貨量各家並列

03 **晶片與封裝** 世代路線、設計分工、封裝與 HBM

04 **板子與托盤** 托盤配置、Host CPU、PCB 與材料、板上電源

05 **機櫃與網路** 機櫃結構、拓樸、OCS 與光銅互連

06 **散熱與電源** 液冷價值量、鍍金冷板、PSU 與 HVDC

07 **供應鏈與市場** 封測分工、台股曝險、市場多空與追蹤訊號

十個結論：TPU 放量的錢，流向規格和 NVIDIA 最不一樣的地方

2027 年出貨約翻倍；Google 收回運算晶片設計、v8 起綁自家 Axion CPU、機櫃沒有交換器托盤

- **量**：2027 年 TPU 出貨各家估 735–910 萬顆，約為 2026 年 2 倍；Anthropic 部署 1→5→10 GW
- **設計**：v8t 起運算晶片由 Google 自己設計，聯發科做 I/O 與封裝；Broadcom 守 v8i / v9i 推論線
- **封裝**：v9 走 Intel EMIB-T，2028 年封裝與 ABF 載板是出貨上限
- **晶片越做越大**：v8i 2 顆運算晶粒 → v9 24 顆晶粒 → v10 46 顆、15 倍光罩
- **CPU**：v8 全系統改用 Axion，TPU:CPU 由 4:1 → 4:2，CPU 出貨 125 萬 → 400 萬顆
- **機櫃**：沒有 NVLink 背板與交換器托盤；櫃內銅纜、櫃外光纖＋OCS
- **網路**：TPU8 每顆網路硬體內含值是 TPU7 的 2–5 倍（光纖 8→18.2 條、光模組 1.8→3.0 顆）
- **散熱**：單晶片功耗低，但每層冷板 12 片＋鍍金，單層冷板價值各平台最高
- **板子與材料**：主板 24L→32–44L、M7→M8；每櫃 PCB 2 萬 → 3.8 萬美元；HVLP4、T-glass 缺貨
- **台股**：有份額數字的約十家；CPO / NPO 不是 v8 的驅動力，嘉基仍屬推論

一張表看懂：TPU 的錢流到哪裡

依層級列出主要內容、台股映射與證據強度；★ 為規格和 NVIDIA 差最多、彈性最大的環節

層級	主要內容	台股映射	證據
晶片設計	I/O 晶片、後段整合（v8t 起）；Axion CPU 設計服務	聯發科 2454、創意 3443	強
晶圓與封裝	N3 / N2 代工、CoWoS、oS 外包、EMIB-T 載板	台積電 2330、日月光 3711、欣興 3037、臻鼎 4958	強
封測	FT、燒機、探針卡、測試座、分選機、SLT	京元電 2449、旺矽 6223、鴻勁 7769、穎崴 6515、致茂 2360	強 / 中
板子與材料	主板 PCB、CCL、HVLP4 銅箔、Low-Dk 玻纖、BMC	金像電 2368、台光電 2383、金居 8358、南亞 1303、信驊 5274	強 / 中
機櫃與網路 ★	**光模組、光纖、OCS、DAC / ACC、交換器、組裝**	**光聖 6442、聯亞 3081、英業達 2356、鴻海 2317**	**中**
散熱 ★	**冷板、鍍金、快接頭、補強環、歧管**	**奇鋐 3017、匯鑽科 8431、富世達 6805、健策 3653、時碩 4566**	**強 / 中**
電源	PSU、power shelf、BBU、±400V 電源櫃	台達電 2308、光寶科 2301	中
潛在切入	NPO / CPC 插座、AEC、v10 設計服務	嘉基 6715、嘉澤 3533、貿聯 3665、世芯 3661	弱

資料來源：本簡報整理；證據強度依第 7 章個股表

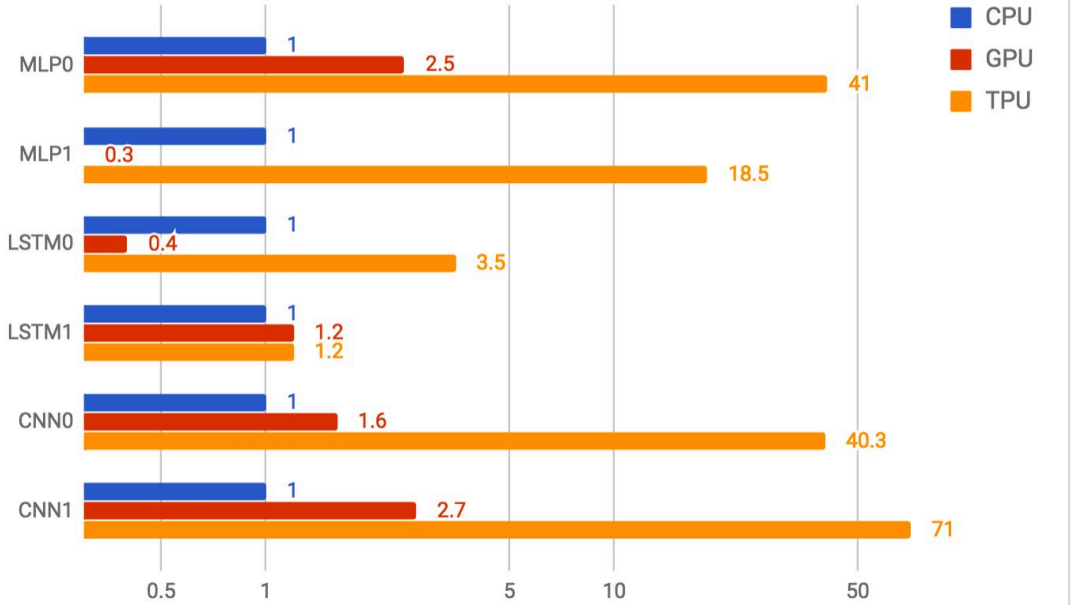
01

TPU 是什麼

只做矩陣乘法的專用晶片，靠「系統」而不是單顆晶片贏

為什麼 Google 要自己做晶片：從一張推論卡開始

2015 年起部署的 TPU v1 只做一件事——矩陣乘法，效能是同期 CPU / GPU 的 15–30 倍



TPU v1 相對效能（CPU＝1，對數軸）：六種推論模型上 TPU 為 1.2–71 倍，GPU 為 0.3–2.7 倍（Google，2017）

起點

- 2013 年 Google 估算：若全面上線語音辨識，資料中心要倍增 → 決定自研推論晶片
- v1：28nm、700MHz、約 40W；256×256＝65,536 個 8 位元乘加器，92 TOPS（INT8）
- 效能是同期 CPU / GPU 15–30 倍、每瓦 30–80 倍；PCIe 卡、無 HBM
- 之後每一代都往「訓練＋大規模叢集」走：v2（2017）起可訓練、v3（2018）起液冷、v4（2022）起 OCS
- Gemini 全系列在 TPU 上訓練與服務（Google 原文）

資料來源：Google Cloud blog〈An in-depth look at Google's first TPU〉（2017-05-12）；TPU v1 論文 arXiv 1704.04760；SemiAnalysis（2025-11-28）

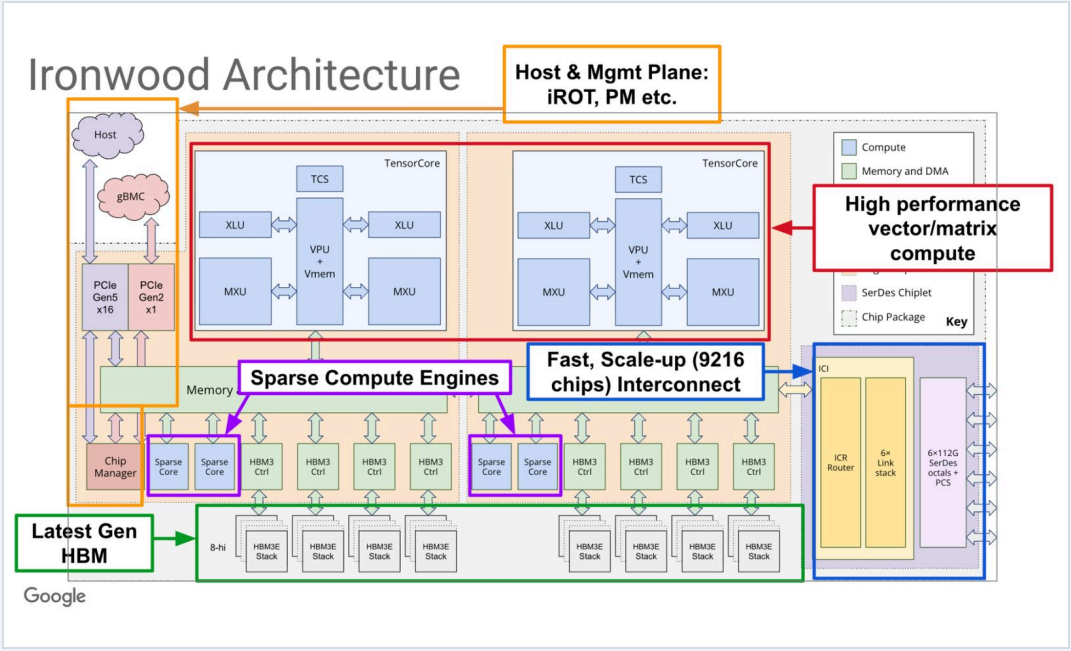
核心架構：脈動陣列讓資料在乘加器之間流動，不回記憶體

CPU 每次運算都存取記憶體，GPU 每次存取暫存器；TPU 的乘加器彼此直連，矩陣乘法中途不碰記憶體

三種晶片怎麼算矩陣（Google 官方說法）

	做法	代價
CPU	通用運算，每次運算都讀寫記憶體	馮紐曼瓶頸，平行度低
GPU	數千個 ALU 平行，每次仍讀寫暫存器 / 共享記憶體	彈性高、生態完整（CUDA）
TPU	MXU 脈動陣列：乘加器串成網格，資料像心跳一波波流過	只擅長矩陣乘法，通用性低

MXU：v6e 之前 128×128，v6e、Ironwood 為 256×256；每個 MXU 每週期 16K 次乘加；輸入 BF16、累加 FP32。



Ironwood 架構：雙 chiplet，各含 1 TensorCore（2 MXU＋VPU＋XLU）、2 SparseCore、4 疊 HBM3E；ICI 6×112G SerDes（Google Hot Chips 2025）

TensorCore

MXU＋向量單元（activation、softmax）＋純量單元（控制流）。Ironwood 每晶片 2 個 TC，JAX 視為 2 個 device。

SparseCore

處理 embedding 的不規則記憶體存取；約 5% 面積換來 embedding 模型 5–7 倍加速（v4 論文）。8i 改用集體通訊引擎 CAE。

記憶體階層

HBM（Ironwood 192GB、7.37 TB/s）→ 片上 SRAM（VMEM；8t 128MB、8i 384MB）→ 經 PCIe 的 host 記憶體。

歷代演進：十年算力放大約 25 倍、pod 從 256 顆到 9,600 顆

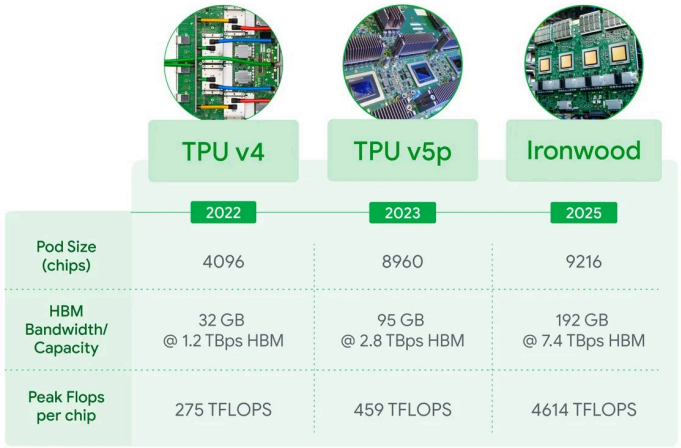
精度不可跨代直接比：v5p / v6e 為 BF16，Ironwood 標 FP8，TPU 8 標 FP4

世代（年）	用途	峰值算力 / 晶片	HBM / 頻寬	ICI / 晶片	Pod 規模	能效與備註
v1（2015）	推論	92 TOPS（INT8）	無（28 MiB 片上）	—（PCIe 卡）	單卡	比 CPU / GPU 每瓦 30–80 倍
v2（2017）	訓練＋推論	4 晶片裝置 180 TFLOPS	—	—	64 裝置 11.5 PFLOPS	首代可訓練
v3（2018）	訓練	123 TFLOPS（BF16）	32 GiB / 900 GB/s	70 GB/s	1,024 顆，2D torus	首代液冷
v4（2022）	訓練	275 TFLOPS	32 GiB / 1.2 TB/s	300 GB/s	4,096 顆，3D torus＋OCS	較 v3 每瓦 2.7 倍
v5e（2023）	性價比	197 TFLOPS	16 GB / 0.8 TB/s	400 GB/s	256 顆，2D torus	推論每美元為 v4 的 2.5 倍
v5p（2023）	旗艦訓練	459 TFLOPS	95 GiB / 2.77 TB/s	1.2 TB/s	8,960 顆，3D torus	LLM 訓練比 v4 快 2.8 倍
v6e Trillium（2024）	訓練＋推論	918 TFLOPS	32 GB / 1.64 TB/s	800 GB/s	256 顆，2D torus	MXU 擴到 256×256
v7 Ironwood（2025）	**推論優先**	**4,614 TFLOPS（FP8）**	**192 GB / 7.37 TB/s**	**1.2 TB/s**	**9,216 顆，42.5 EFLOPS**	**每瓦 2 倍於 Trillium；雙 chiplet**
TPU 8t（2026）	**預訓練**	**12.6 PFLOPS（FP4）**	**216 GB / 6.53 TB/s**	**2 倍**	**9,600 顆，121 EFLOPS**	**訓練每美元 2.7 倍**
TPU 8i（2026）	**推論 / RL**	**10.1 PFLOPS（FP4）**	**288 GB / 8.60 TB/s**	**Boardfly**	**1,152 顆（1,024 在線）**	**推論每美元 +80%**

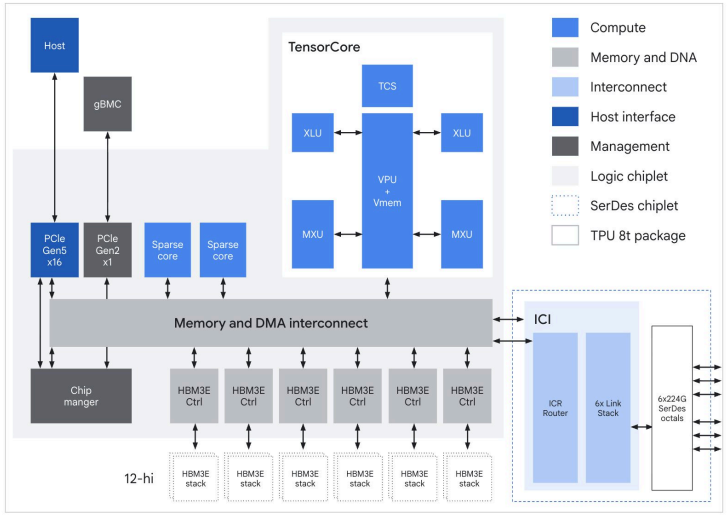
資料來源：Google Cloud 各世代 TPU 文件；ISCA 2023 論文表 4；Google Cloud blog（2024-05-15、2025-11-06、2026-04-22）；TPU v1 論文

最近三代對照：記憶體 6 倍、頻寬 6 倍、算力一路往低精度走

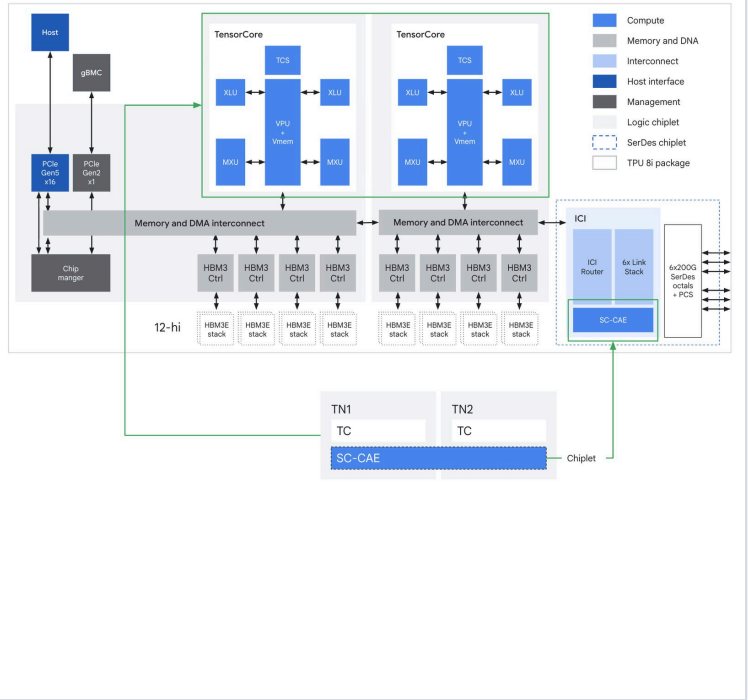
Ironwood 是第一顆多運算晶粒的 TPU；TPU 8 首度拆成訓練、推論兩顆專用晶片



Google 官方對照：v4（2022）4,096 顆、32GB；v5p（2023）8,960 顆、95GB；Ironwood（2025）9,216 顆、192GB、4,614 TFLOPS（FP8，其餘為 BF16）



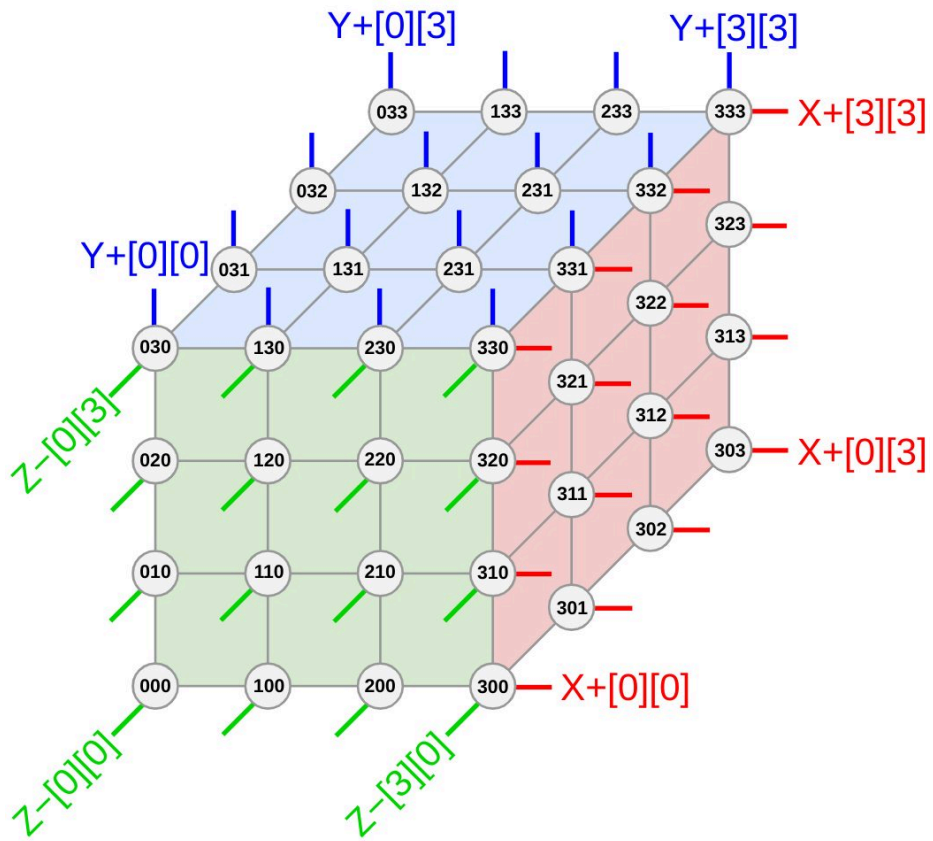
TPU 8t 方塊圖：1 TensorCore（2 MXU）、2 SparseCore、6 疊 12-hi HBM3E、PCIe Gen5 x16、ICI＋6 組 224G SerDes（Google 2026-04-22）



TPU 8i 方塊圖：2 TensorCore、8 疊 12-hi HBM3E、集體通訊引擎 CAE、6×200G SerDes（Google 2026-04-22）

TPU 靠系統贏：ICI 直連＋3D Torus＋OCS

OCS 只占系統成本不到 5%、功耗不到 3%，卻讓 pod 可以重組、繞過故障



4×4×4 cube：每顆晶片接 X / Y / Z 三向鄰居，邊緣環繞連線經光埠接到 OCS（Google 2025-11-06）

三個關鍵

- **ICI**：晶片間直連、不經交換器；Ironwood 每顆 9.6 Tb/s（1.2 TB/s 雙向）。SemiAnalysis 認為 ICI 是 NVLink 唯一真正的對手
- **3D Torus**：一個機櫃 64 顆＝一個 cube；滿 4×4×4 才有完整環繞；OCS 可在不重接線下改成 twisted torus
- **OCS**：v4 用 48 台 Palomar（136×136 埠 MEMS 鏡面）；可繞過壞掉的 cube、從 checkpoint 重組；機櫃可獨立上線、縮短部署時間
- **超出 pod**：Multislice 經資料中心網路（DCN）串多個 slice；Jupiter 網路本身也用 OCS（Apollo）

軟體與商業化：JAX / XLA 自家棧，價格比 GPU 便宜

框架（JAX、PyTorch）→ XLA 編譯器 → TPU；外部使用門檻高於 CUDA，但大客戶能自己寫 kernel

軟體棧

- JAX＋XLA：Google DeepMind 內部主力；XLA 把運算融合成 kernel 餵滿 MXU，Pallas 補自訂 kernel
- PyTorch：原生 PyTorch on TPU 進入 preview（與 Meta 合作）；TPU7x 不支援 TensorFlow
- Pathways：單控制器分散式執行層，JAX＋Pathways 可擴到 100 萬顆單一訓練叢集
- 推論：vLLM、SGLang 原生支援；MaxText（訓練）、Tunix（RL）

Cloud TPU 定價（美元 / 晶片小時，us-central1）

世代	隨需	3 年承諾
Ironwood	12.00	5.40
v6e Trillium	2.70	1.22
v5p	4.20	1.89
v5e	1.20	0.54

外部客戶（有原文者）

- Anthropic：最多 100 萬顆 TPU、2026 年逾 1 GW（官方 2025-10-23）
- Apple：AFM 用 8,192 顆 v4＋2,048 顆 v5p 訓練（論文）
- SSI、Citadel Securities、Salesforce、Hugging Face 等（Google 原文）
- Meta：租用已簽、直接購買洽談中（報導，非官方）

資料來源：Google Cloud TPU 定價頁（2026-10-14 抓取）；Google Cloud blog（2025-11-06、2026-04-22）；Anthropic（2025-10-23）；Apple arXiv 2407.21075；SemiAnalysis（2025-11-28）

02

需求與出貨

Google 自用＋對外銷售；2027 年各家估 735–910 萬顆

Capex 上修、backlog 翻倍，TPU 開始賣進客戶機房

2026 年 Capex 指引 1,950–2,050 億美元，2027 年「顯著增加」；TPU 系統銷售 2Q26 起交付客戶資料中心

450 億美元
2Q26 單季 Capex，年增 100%；60% 花在伺服器

1,950–2,050 億
2026 年 Capex 指引（美元），由 1,800–1,900 億上修

5,140 億美元
Cloud backlog；1Q26 4,620 億、4Q25 2,400 億

TPU 銷售模式在改變

- 2Q26 開始把 TPU 系統交付到客戶自己的資料中心，本季只認列少量；現有合約收入大多落在 2027 年
- 管理層：自研自製的價值在擴大 Cloud TAM，讓大型企業把 TPU 放進自家機房
- 算力優先序：前沿模型 / AGI → Search、YouTube → Cloud；Cloud 供給不足，3Q26 起向第三方租算力
- 國泰：2027 年 TPU 機櫃逾 10 萬櫃、首度外賣機櫃，組裝供應商由美商擴到 4–5 家

回本速度是 TPU 的賣點

- Alphabet 線上會議：AI 伺服器回收期不到 2 年，自研 TPU 可在 1 年內回本（永豐 09-16、群益 ECOC 10-06）
- Broadcom：XPU 效能不低於 GPU、成本不到 GPU 一半（富邦 Broadcom 法說 memo 09-03）
- SemiAnalysis：v7 全包 TCO 比 GB200 低約 44%；以 Anthropic 視角，每有效 PFLOP 比 GB300 低約 52%（假設 40% MFU）

資料來源：JPM、富邦（Alphabet 2Q26 法說，2026-07-23）；Nomura（2026-06-30）；國泰（2026-09-24）

外部客戶：Anthropic 每年新增 1 → 5 → 10 GW

Broadcom 法說揭露的部署節奏；TPU 從內部晶片變成有外部買家的產品線

1 GW

FY26：Anthropic 部署 Ironwood（v7）

+5 GW

FY27：再部署 TPU v8i，成 Broadcom 最大 XPU 客戶

+10 GW

FY28：再新增；管理層明言為 Anthropic 自身需求

時間軸

- 2025–10：Anthropic 宣布擴大使用 Google Cloud，最高 100 萬顆 TPU、2026 年上線逾 1 GW
- 2026–04：與 Google、Broadcom 簽新協議，取得數 GW 次世代 TPU（v8），2027 年起上線
- 2026–06：Broadcom 與 Apollo、Blackstone 推 AI XPV 平台，初始融資 350 億美元，首筆用於 Anthropic 逾 1 GW 部署
- Celestica：TPU 機櫃出貨同時服務 Google 內部與 Anthropic 等外部客戶

規模換算與其他客戶

- Broadcom：Google 另簽長約，未來數年每年 TPU 供應達數百億美元；FY27＋FY28 潛在需求約 30 GW（Google、Anthropic、OpenAI、Meta）
- Nomura：以每顆 1,000W 估，20 GW 約對應 700 萬顆以上 TPU / ASIC、36–46 萬片 CoWoS
- 永豐論壇：Whalefish / Sunfish Ultra 是賣給 Anthropic 的外賣機架，自用 TPU 給聯發科
- 庫內沒有 Meta 採購 TPU 的一手來源（市場傳聞見第 7 章）

資料來源：富邦 Broadcom Q3 法說 memo（2026-09-03）、Nomura（2026-06-30）、凱基（2026-09-30）、Celestica call memo（2026-09-24）

客戶占比：2027 年起外部客戶約占 TPU 一半

Aletheia：Google 自用 2026 年約 350 萬顆 → 2027 年 400–450 萬 → 2028 年 550 萬；總量 360 → 850 → 1,200 萬

Google 自用 vs 外部（Aletheia，萬顆）

年度	TPU 總量	Google 自用	外部（推算）	外部占比
2026	360	約 350	約 10	約 3%
2027	**850**	**400–450**	**400–450**	**約 47–53%**
2028	**1,200**	**550**	**約 650**	**約 54%**

外部＝總量減自用（本簡報推算）；Aletheia 的總量含 Anthropic 經 Broadcom 取得的量。Aletheia 2025 年總量 330 萬低於自用約 350 萬，口徑有落差。

外部主要是 Anthropic

來源	Anthropic 規模
Broadcom 法說（09–03）	FY26 1 GW → FY27 再 +5 GW → FY28 再 +10 GW；FY27 成 Broadcom 最大 XPU 客戶
Aletheia（10–04）	Anthropic 自有算力 2027 年 6 GW → 2028 年 13 GW，多數來自 Broadcom TPU（v8i、v8.5），另 5 GW 為 AMD
MS（07–23）	TPU 對外銷售約 3.5 GW 集中 2027 年、另約 4 GW 在 2028 年
SemiAnalysis	v7 階段 100 萬顆：40 萬顆 Broadcom 直賣整櫃、60 萬顆經 GCP 租用

其他外部客戶（無量化）

Meta：已簽數十億美元租用、2027 年直接購買洽談中（報導）｜黑石合資 TPU 雲：首批 500MW、2027 年上線｜SSI、Apple、Citadel 等為小量或舊案｜OpenAI 否認大規模採用。

讀法

2027 年起 TPU 一半以上給外部，且高度集中在 Anthropic——這是「量」的上檔，也是黃仁勳說「100% Anthropic」的客戶集中風險。Celestica 也說外部占比無法量化。

2027 年出貨：十家估 735–910 萬顆，方向一致是翻倍

差距來自「設計出貨 vs 封裝顆數 vs 測試顆數」口徑；4Q26–2027 是 v8 兩條線同時放量

來源（日期）	2026E	2027E	2028E	口徑
MS（08–24）	370 萬	735 萬	>650 萬	v7 20＋v8i 400＋v8t 300＋v9 15 萬
MS（09–27）	330 萬	750 萬	>900 萬	2028 年 v8t 200 萬＋v9 300 萬
JPM（07–01）	—	800 萬	—	由 500 萬上修
Nomura（06–30）	420 萬	830 萬	—	台積電產出換算
GFHK（08–24）	450 萬	884 萬	—	原圖單位標 mn，級距應為千顆
UBS（09–22）	410 萬	900 萬	—	Broadcom 500＋聯發科 400 萬
凱基（09–30）	**402 萬**	**910 萬**	**—**	**v8i 550＋v8t 360 萬**
國泰（09–24）	—	800–900 萬	—	機櫃 10 萬櫃以上
Aletheia（10–04）	360 萬	850 萬	1,200 萬	含 Anthropic；TPU8 生命週期 1,400–1,500 萬
定錨講座	—	—	1,000 萬	Google 內部目標

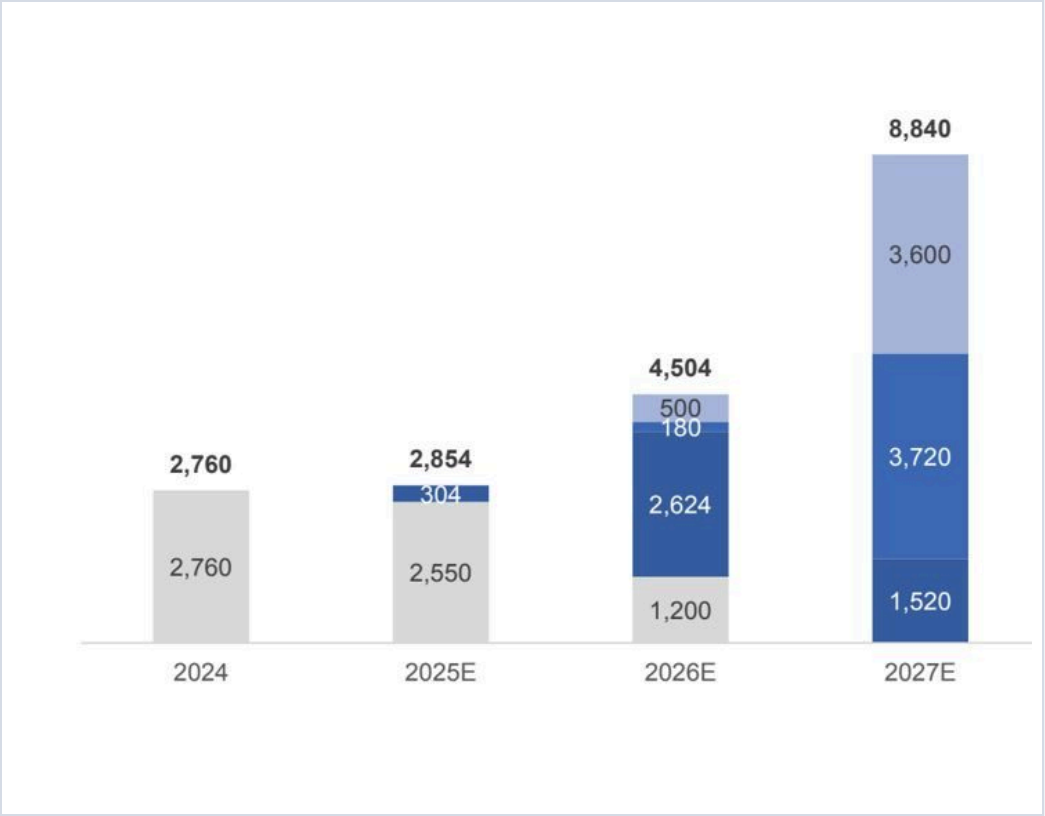
資料來源：MS（08–24、09–27）、GFHK（08–24）、UBS（09–22）、JPM（07–01）、Nomura（06–30）、凱基（09–30）、國泰（09–24）、Aletheia（10–04 / 10–06）、華南（09–18）

出貨拆世代：v8i（Broadcom）與 v8t（聯發科）2027 年同時放量

MS 世代別模型；v8i 量各家差最多（MS 396 萬 vs 凱基 550 萬）

MS 世代拆解（千顆）

世代	設計	2025e	2026e	2027e	2028e
v6 Trillium	Broadcom	1,200	—	—	—
v7 Ironwood	Broadcom	300	2,300	200	—
v8i Sunfish	Broadcom	—	900	4,000	2,500
v8t Zebrafish	聯發科	—	500	3,000	1,000→2,000
v9 Humufish	聯發科	—	—	150	3,000
合計		**1,750**	**3,700**	**7,350**	**>6,500→>9,000**

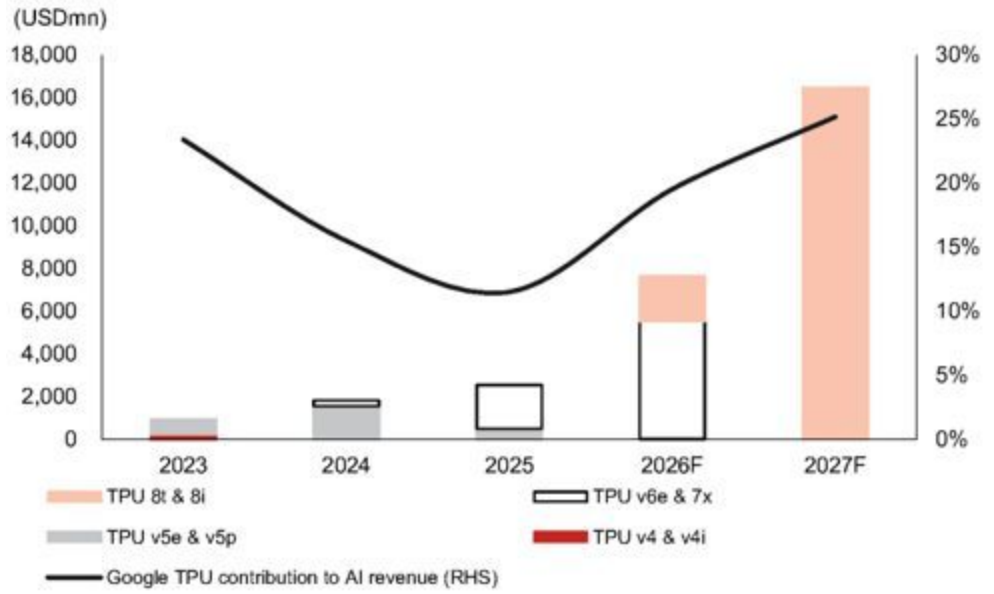


GFHK 出貨量估計（千顆）：2024 年 276 萬、2027E 884 萬；2027E 淺藍段為聯發科 v8t（GFHK，2026-08-24）

MS 09-27 把 v8t 2028 年由 100 萬上修到 200 萬顆，理由是台積電 CoWoS 續支援 3nm 世代；凱基 2027 年 v8i 550 萬、v8t 360 萬。v8 量產時程：凱基 4Q26，Aletheia 稱 v8i 機櫃 HVM 在 2027 年中。

從台積電看 TPU：2027 年占其 AI 營收約四分之一

Google 占 CoWoS 產出 2026F 約 20% → 2027F 約 27%；N3 供需 2027 年更緊



Google TPU 對台積電 AI 營收貢獻（百萬美元）與占比；2027F 約 165 億美元，以 8t / 8i 為主（Nomura，2026-06-30）

重點數字

- Nomura：Google TPU 對台積電營收 2025 年 25 億 → 2026F 76 億 → 2027F 165 億美元；2027F 8t 51 億、8i 113 億
- Nomura：Google 2027F CoWoS 48 萬片（+115%），Broadcom：聯發科 ≈ 66—68：32—34
- UBS：Google 占 CoWoS 需求 2025 年 11% → 2026E 16% → 2027E 20%（口徑不同於 Nomura）
- JPM（10-12）【新】：Rubin、TPU v8、Trainium 3 都還在 N3，N3 供需 2027 年惡化
- BofA（10-12）【新】：v8t / v8i 初期放量足以抵銷 Ironwood 降載

資料來源：Nomura（2026-06-30）、UBS（2026-09-22）、JPM 亞洲半導體（2026-10-12）、BofA 台積電（2026-10-12）

03

晶片與封裝

設計權回到 Google；晶粒越來越多、光罩越來越大，封裝和 HBM 是瓶頸

世代路線：每一代分成訓練（t）與推論（i）兩條線

Broadcom 從 v8t 起被擠出主線，只留推論用的 v8i / v9i；聯發科連拿 I/O 與後段整合

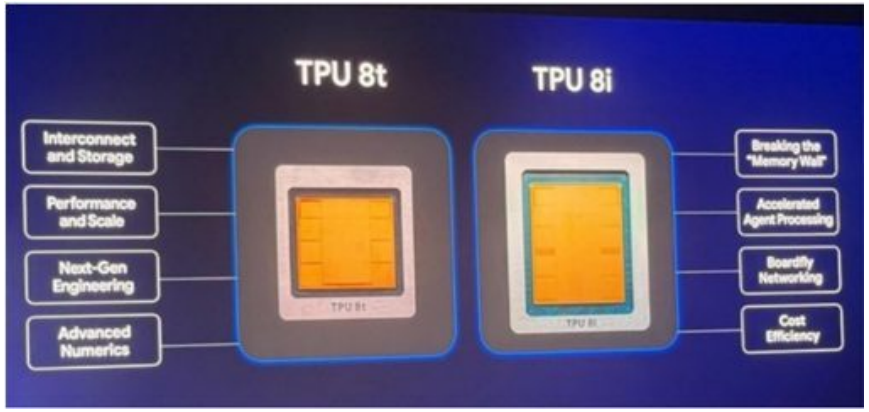
	v7 Ironwood	v8i Sunfish	v8t Zebrafish	v9t Humufish	v9t+ Triggerfish	v9i Whalefish	v10t Icefish
時程	4Q25	4Q26	4Q26	4Q27	4Q27	1Q28	4Q28
運算晶片設計	Broadcom	Broadcom	**Google**	**Google**	**Google**	Broadcom	**Google**
I/O＋後段	Broadcom	Broadcom	聯發科	聯發科	聯發科	Broadcom	聯發科 / Google
運算晶片製程	2×N3E	2×N3P	1×N3P	4×N2P	4×N2P	2×N3P	A14/N2P×8＋N2P×16
I/O 製程	1×N5	1×N4	1×N4P	4×N3P	4×N3P	2×N4	N2P×6
HBM	8×HBM3e 192GB	8×HBM3e 288GB	6×HBM3e 216GB	6×HBM4 288GB	6×HBM4e 384GB	6×HBM4e 384GB	16×HBM5
HBM 頻寬	7.4 TB/s	8.6 TB/s	6.5 TB/s	—	—	—	—
封裝	CoWoS-S	CoWoS-L	CoWoS-S	EMIB-T	EMIB-T	CoWoS-S / SoIC	EMIB-T 或 CoWoS
內部代號	Ghostfish / Hammer	Hell Cat	Mad Dog	A5922	—	Sunfish×2	—

命名各家不一：MS 只列「v9 Humufish」；A5922 在 CLSA / Nomura / 永豐指 Humufish、凱基指 Triggerfish；Merope 有 2nm 美系設計（MS）與 3nm AMD＋Broadcom（永豐）兩說。v8 製程另有 SemiAnalysis「N2P＋HBM4E」口徑，與多數券商的 N3 / HBM3E 不同。

官方規格：8t 拚訓練規模、8i 拚推論延遲

Google Cloud Next 2026 發表；8t 單一 superpod 9,600 顆、121 EFLOPS，8i 加大 SRAM 對付記憶體牆

	v7 Ironwood	TPU 8t	TPU 8i
定位	訓練＋推論	大規模預訓練、embedding	後訓練、即時推論、reasoning
HBM	192GB / 7.38 TB/s	216GB / 6.53 TB/s	288GB（12-Hi HBM3E） / 8.60 TB/s
片上 SRAM	—	128MB	384MB（前代 2.4 倍）
峰值算力	4.6 PFLOPS（FP8）	12.6 PFLOPS（FP4）	10.1 PFLOPS（FP4）
ICI 頻寬	9.6 Tb/s	19.2 Tb/s	19.2 Tb/s
拓樸	3D Torus	3D Torus＋Virgo	Boardfly＋OCS
單一 pod	9,216 顆	9,600 顆 / 121 EFLOPS	1,152 顆（1,024 在線）
Host CPU	Intel x86	Google Axion	Google Axion
效益	—	訓練 perf/$ 最高 2.7 倍	MoE 推論 perf/$ +80%；兩者 perf/W 約 2 倍

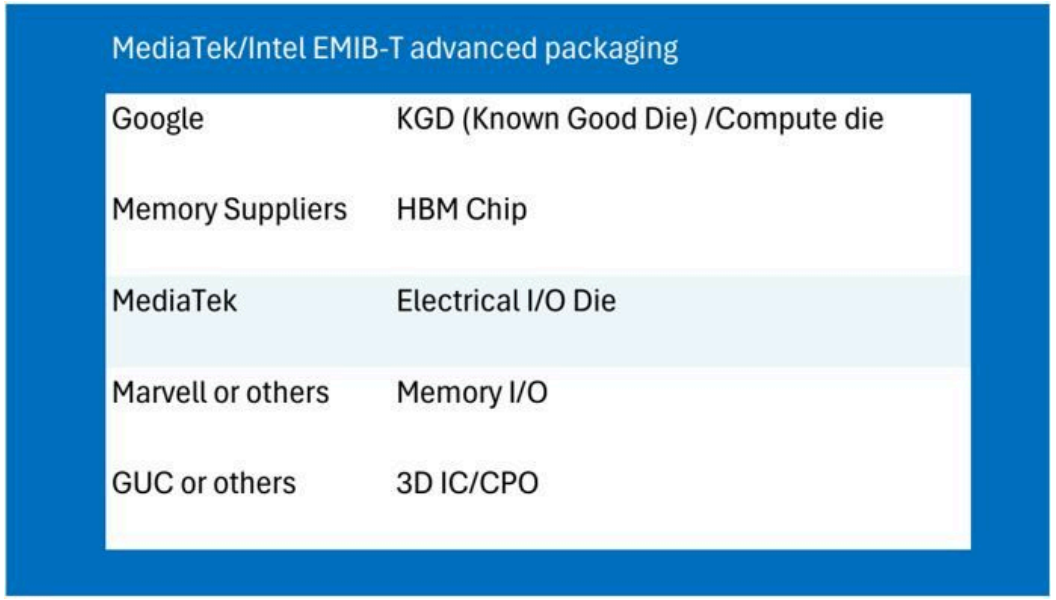


Google 投影片上的 TPU 8t（左）與 8i（右）封裝實物；中央運算晶粒、四周 HBM（引自 MS 2026-05-25）

資料來源：Google Cloud Next 投影片（引自凱基 2026-09-30）；Google Cloud blog〈TPU 8t and 8i technical deep dive〉（2026-04-22）

設計分工：Google 收回運算晶片，設計服務商改收 I/O＋封裝

v10 起 Google 直接交「已驗證良品晶粒」（KGD），周邊工作拆給更多家



v10 分工：Google 給 KGD、記憶體廠給 HBM、聯發科做電氣 I/O、Marvell 等做記憶體 I/O、創意等做 3D IC / CPO（MS，2026-08-24）

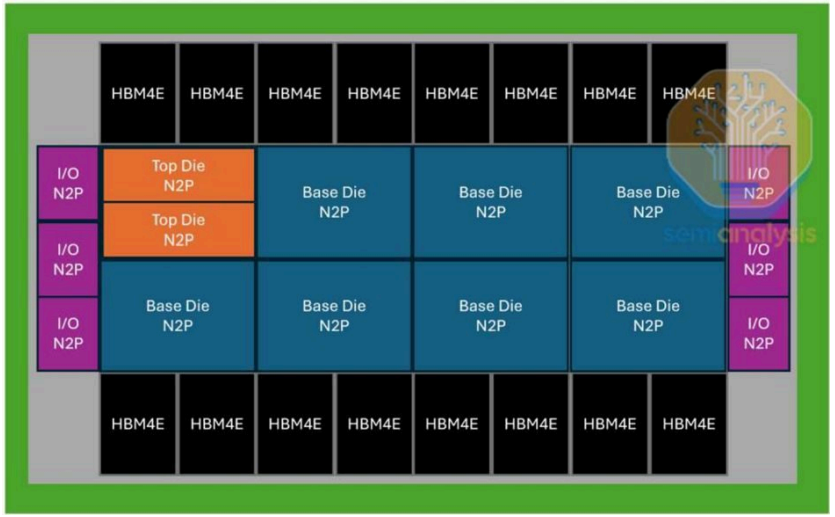
演變與原因

- v3—v7 Ironwood：全部 Broadcom
- v8：Sunfish（8i）Broadcom；Zebrafish（8t）運算晶片 Google 自做、I/O 與後段聯發科
- v9：Broadcom 原案 Pumafish 啟動約一個月後被取消，改做 Whalefish（≈ 兩顆 Sunfish）
- 聯發科獨拿 v9 的原因：Google 要 336G 特規 SerDes，聯發科做得出、Broadcom 押 448G 未 ready（永豐論壇）
- 價格：JPM 稱聯發科方案便宜 40—50%；Google 自己採購 HBM
- v10 候選：Marvell（記憶體 I/O）、創意（3D IC / CPO）、AMD、世芯；Citi 認為聯發科與台積電仍是主要來源

晶片越做越大：v8i 兩顆運算晶粒，v10 一個封裝 46 顆晶粒

光罩面積 2.5 倍 → 15 倍、功耗約 1,300W → 約 10,000W；v10 首度 3D 封裝

世代	晶粒組成	光罩	功耗	封裝	時程
v8i Sunfish	2 運算（N3E）＋1 I/O（N4）＋8 HBM3E	≈2.5x	≈1,300W	CoWoS	4Q26
v8.5 Whalefish	4 運算（N3）＋2 I/O（N4）＋12 HBM3E / 4	≈5x	≈2,600W	CoWoS	2027–28
v9 Humufish	4 運算（N2P）＋4 core logic＋4 I/O（N3P）＋12 HBM4E＝24 晶粒	9.5x	≈4,500W	EMIB-T / 面板級	2028 年中
v10 Icefish	16 運算＋8 core logic＋6 I/O（皆 N2P）＋16 HBM4E＝46 晶粒	15x	≈10,000W	EMIB-T＋首度 3D	2029



v10 Icefish 平面配置：中央 2 Top Die＋7 Base Die、左右各 3 I/O Die，上下各 8 顆 HBM（SemiAnalysis，2026-05-18）

Aletheia 同份報告內 v8i 功耗另寫 600W，與 1,300W 不一致；UBS 寫 8i 封裝約 900W、8t 約 1,000W、v9 約 2,400W；JPM 寫 v8 約 1,000W。MS 10-05【新】稱聯發科最初以 14x 光罩處理 v10 IceFish。

一顆 TPU 是一整個封裝模組，由上往下拆

從冷板到運算板 PCB，每一層都有台廠

冷板（鍍金底座）＋銅片導熱 ｜ 奇鋐、匯鑽科

均熱蓋 / 加強框（Lid / Stiffener） ｜ 健策、一詮

HBM ｜ 運算晶粒＋I/O 晶粒 ｜ HBM ｜ SK 海力士、三星、美光；台積電

微凸塊（μbump，間距數十 μm）

中介層 / 矽橋：CoWoS 矽中介層、RDL＋LSI、EMIB-T 矽橋 ｜ 台積電、Intel

C4 凸塊 ｜ EMIB-T bumping：力成

ABF 載板（v7 16L → v8 18–22L；EMIB-T 把矽橋埋在這層） ｜ 欣興、臻鼎、Ibiden、南電

BGA 錫球

運算板 PCB（v8 32–44L、M8；v9 轉 HDI） ｜ 金像電、欣興、ISU、WUS、TTM

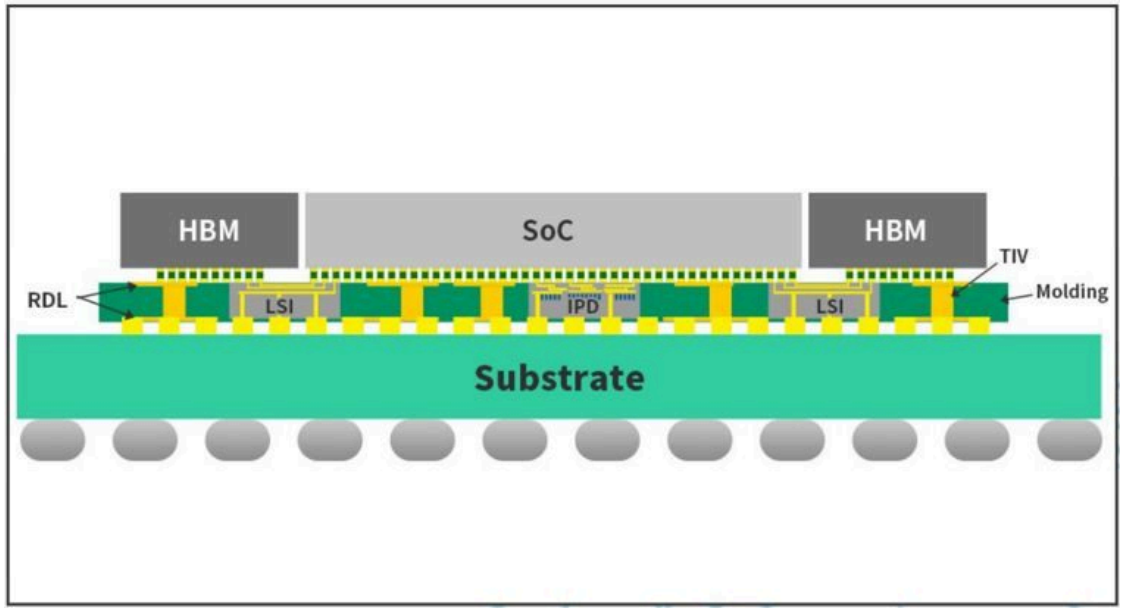
各元件做什麼

元件	功能	誰做
運算晶粒	矩陣乘法核心	v8t 起 Google 設計；台積電
I/O 晶粒	SerDes、ICI、對外介面	聯發科 / Broadcom
HBM	高頻寬記憶體，6–16 疊	SK 海力士、三星、美光
中介層 / 矽橋	晶粒間極細線路	台積電、Intel
ABF 載板	晶片訊號轉接 PCB	欣興、臻鼎、Ibiden
矽電容 / IPD	封裝內穩壓去耦	三星電機；愛普＋力積電

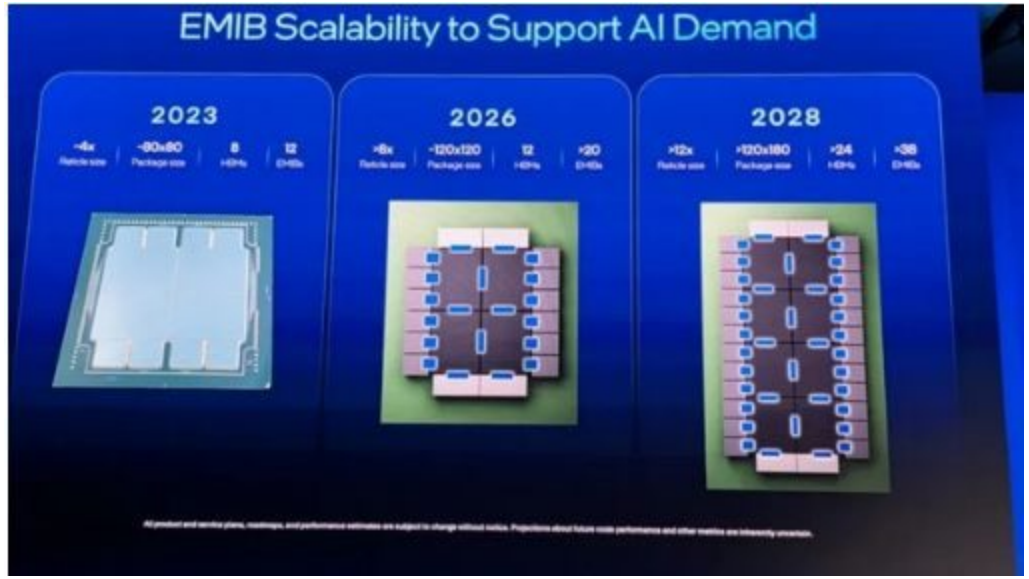
資料來源：本簡報自繪示意；各層廠商見對應章節

CoWoS 與 EMIB-T：v9 選 Intel 是因為尺寸

Humufish 約 10 倍光罩，決策當時 CoWoS-L 最大只到 9 倍多；EMIB-T 把小矽橋埋進載板



CoWoS-L 剖面：中央 SoC、左右 HBM；中間層為 RDL、局部矽中介層 LSI、IPD、TIV（群益×定錨，2026-07-08）



Intel EMIB 規模：2026 年 8 倍以上光罩、120×120mm、12 HBM；2028 年 12 倍以上、120×180mm、24 HBM 以上（引自 MS 2026-09-27）

CoWoS（台積電）

用整片矽（-S）或 RDL＋局部矽橋（-L）當中介層；成熟、良率高，v8 世代主力。台積電 CoWoS 路線約到 9.7 倍光罩，2028 年再提 14 倍。

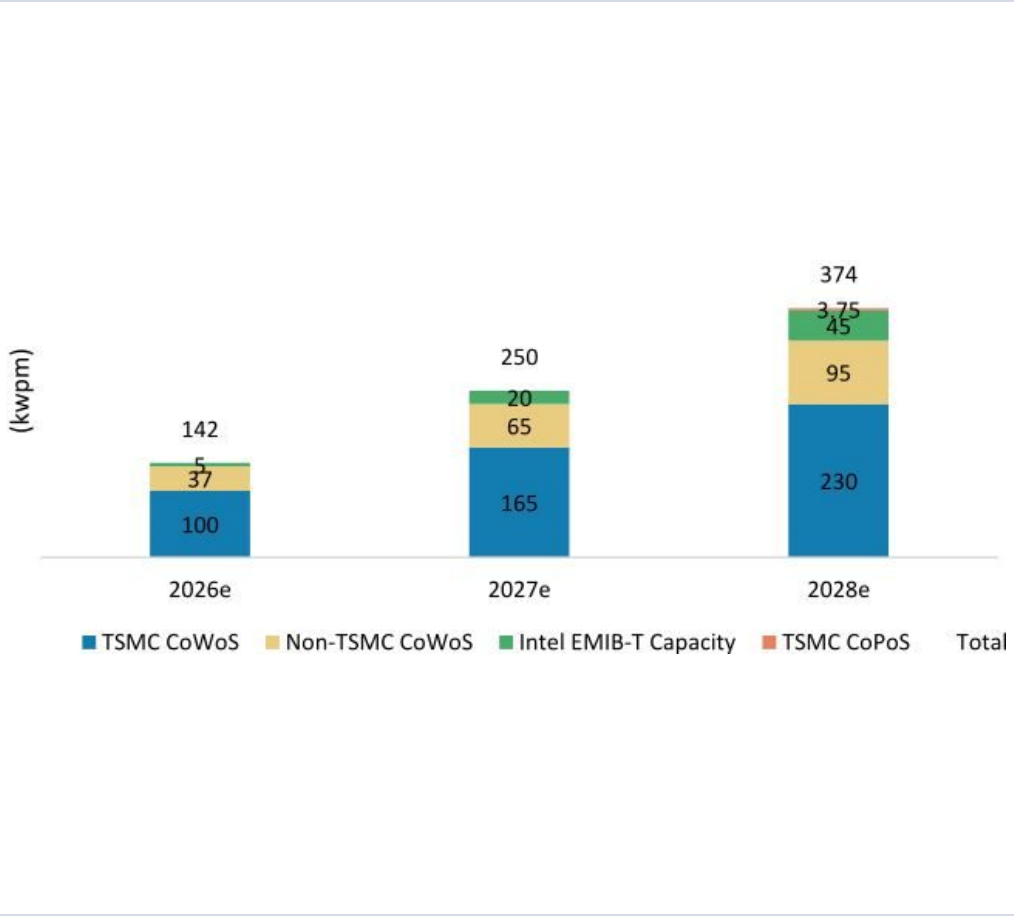
EMIB-T（Intel）

不用整片中介層，矽橋直接埋進 ABF 載板，橋內加 TSV 垂直供電。Nomura：v9 約 30 顆矽橋（Intel 自家產品最多 12—14 顆），規格偏離恐有良率挑戰。

資料來源：定錨產業講座；群益×定錨（2026-07-08）；Intel 路線圖（引自 MS 2026-09-27）；Nomura（2026-06-30）

2028 年最緊的是封裝：EMIB-T 能撐多少顆 v9 說法不一

MS 09-27 說要 CoWoS-L 補 50 萬顆；聯發科 10-01 澄清 v9 不需要 CoWoS



全球 2.5D 封裝月產能（千片 12 吋約當）：2026e 142 → 2027e 250 → 2028e 374；EMIB-T 5 → 20 → 45（MS，2026-09-27）

各家說法並列

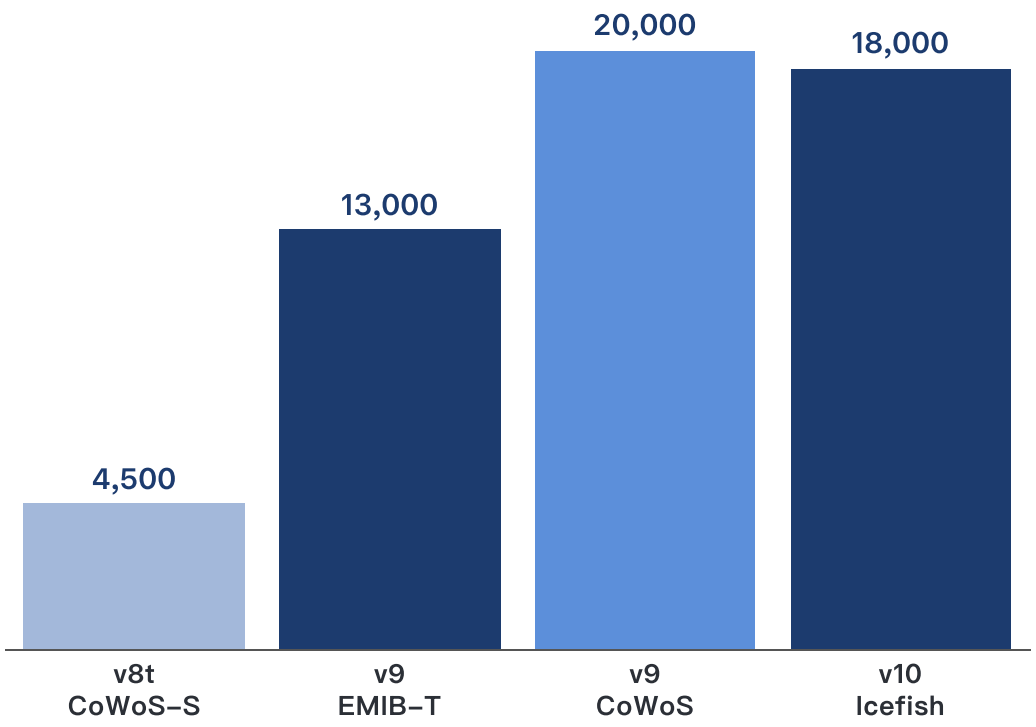
議題	說法
EMIB-T 可撐 v9	MS 09-27 約 250 萬顆（另寫約 200 萬）；MS 10-05【新】近 300 萬；UBS 09-23 >200 萬
是否要 CoWoS	MS 09-27：CoWoS-L 補 50 萬；聯發科 10-01【新】：v9 不需，討論 CoWoS 的是次世代
2028 v9 出貨	MS 300 萬；JPM 300 萬；CLSA、GFHK 240 萬；Aletheia 10-04【新】150 萬
EMIB-T 良率	JPM 約 60%；UBS Ibiden 試產約 50%；BofA 載板 30% / 封裝 80%（經同事筆記轉述）；定錨 20–50%
單價差	聯發科認列 ASP：EMIB-T 版 1.3 萬、CoWoS 版 2 萬美元（多 3 層光罩，Google 吸收）

資料來源：MS（09-27、10-05）、UBS（09-23、07-31）、BofA（09-26）、JPM（07-22）、Aletheia（10-04）、聯發科管理層經 MS 轉述（10-01）

聯發科：單顆認列金額反而上升，各家 TPU 營收模型並列

Google 收走運算晶粒，但 I/O 晶粒越做越大、封裝越來越難

每顆可認列 ASP（美元）



TPU 營收（億美元）

券商（日期）	2026E	2027E	2028E
MS（09-27）	—	135	515＋子系統 25
UBS（09-23）	21	180	435
Citi（09-29）	20	180	400
GS（07-31）	22	203	525
GFHK（06-30）	25	184	734
JPM（07-22）	22	129	418（含其他 ASIC；TPU 416）
華南（09-18）	22	160	417
Daiwa（07-31）	21	150	220

ASP 為 MS 口徑；JPM v8t 4,250、v9 12,000，GFHK Triggerfish 17,700。占聯發科營收：MS 2027 年 38% / 2028 年 72%；UBS 46% / 65%；GFHK 44% / 73%。聯發科 10-01【新】：4Q26 TPU 營收逾 20 億美元、已出貨、無遞延（回應經濟日報 9/29「延到 2027Q2」報導，當日股價 −7.1%）。

資料來源：MS（08-24、09-27）、JPM（07-22）、GFHK（06-30）、UBS（09-23）、Citi（09-29）、華南（09-18）、GS / BofA / Daiwa（07-31）

HBM：TPU 一年要吃掉近三成 HBM 位元

JPM：2027 年 ASIC 占 HBM 位元需求 48%，首度超過 NVIDIA 的 43%

每顆 TPU 的 HBM（各家並列）

世代	MS	JPM	Aletheia
v7 Ironwood	8×HBM3e 8Hi＝192GB	同左	—
v8i Sunfish	8×HBM3e 12Hi＝288GB	同左	同左
v8t Zebrafish	6×HBM3e 12Hi＝216GB	同左	—
v9 Humufish	12×HBM4e 12Hi＝576GB	12×HBM4 8Hi＝288GB	12×HBM4E 8Hi＝384GB
v10 Icefish	—	—	16×HBM4E（GFHK：HBM5）

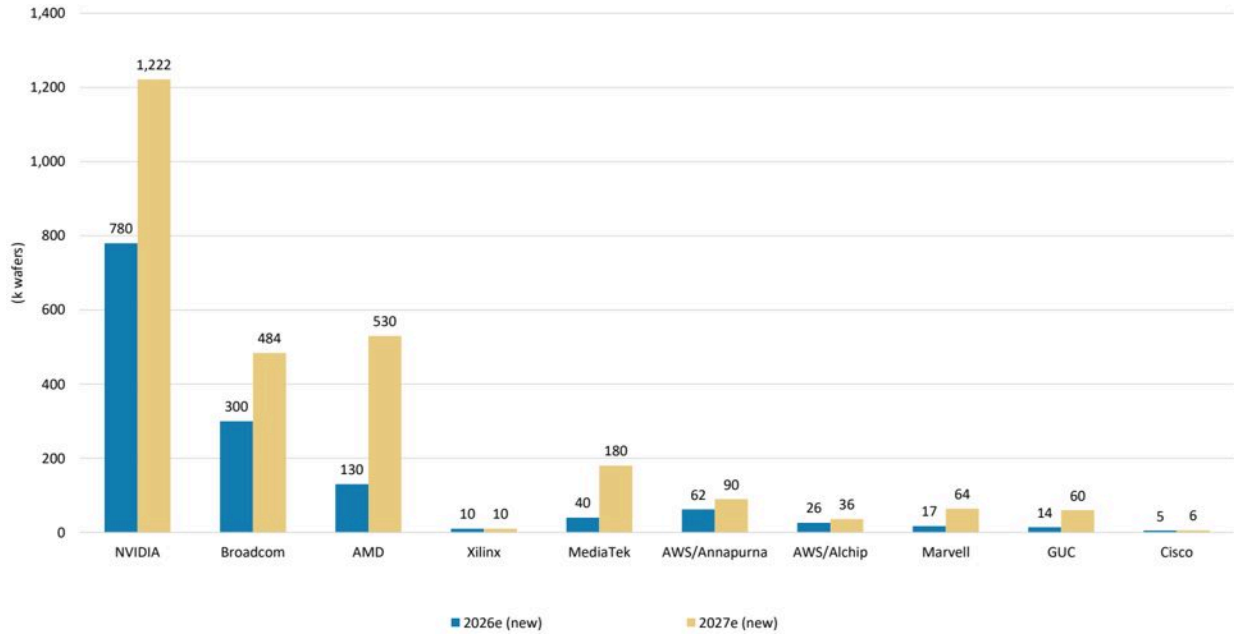
需求規模

- UBS：2027 年 HBM 總消耗 615 億 Gb（+91%）；Google TPU 192 億 Gb（前估 174）、NVIDIA 251 億
- MS：2027 年 AI HBM 需求約 486 億 Gb；TPU 四款約 175 億 Gb（自算）
- JPM：ASIC 系統單位 +102% vs NVIDIA +15%；HBM 消耗兩年 CAGR 63%
- 供應商：v7 為 SK 海力士 / 三星；v8 起三家都供（MS）
- TD Cowen：Google 自研運算 IP 並自行採購 HBM，瓶頸在載板與 HBM

客製 HBM：控制器與自訂邏輯搬進 HBM 底層晶片，ASIC 省出面積放運算核心；v10 的 16 顆 HBM 預期往此方向走（SemiAnalysis，原圖 Marvell）。

CoWoS 分配：Broadcom＋聯發科 2027 年拿到 66 萬片

聯發科 CoWoS 由 4 萬片跳到 18 萬片（+350%），隱含 v8t 約 360 萬顆



CoWoS 需求 2026e vs 2027e（千片）：NVIDIA 780→1,222、Broadcom 300→484、AMD 130→530、聯發科 40→180（MS，2026-06-23）

換算與產能

- Broadcom 2027 年 CoWoS-S：Ironwood＋Sunfish 34.3 萬片 ≈ 417 萬顆 TPU（另含 Meta、Tomahawk 等）
- 聯發科 18 萬片 CoWoS-S × 每片 20 顆 ≈ 360 萬顆 v8t
- 每片顆數：Ironwood 16、Sunfish 12、Zebrafish 20（MS）
- CoWoS-S 基本上只有 Google 在用（專家訪談 05-26）
- 台積電 CoWoS 年底月產能：2026 年 13 萬 → 2027 年 20 萬 → 2028 年 26 萬片（MS 10-05【新】）
- 日月光市占 2026 年 18% → 2028 年 29%（BofA 10-12【新】）

資料來源：MS（2027 CoWoS allocation）（2026-06-23）、MS（07-08）、Nomura（06-30）、BofA 台積電（10-12）

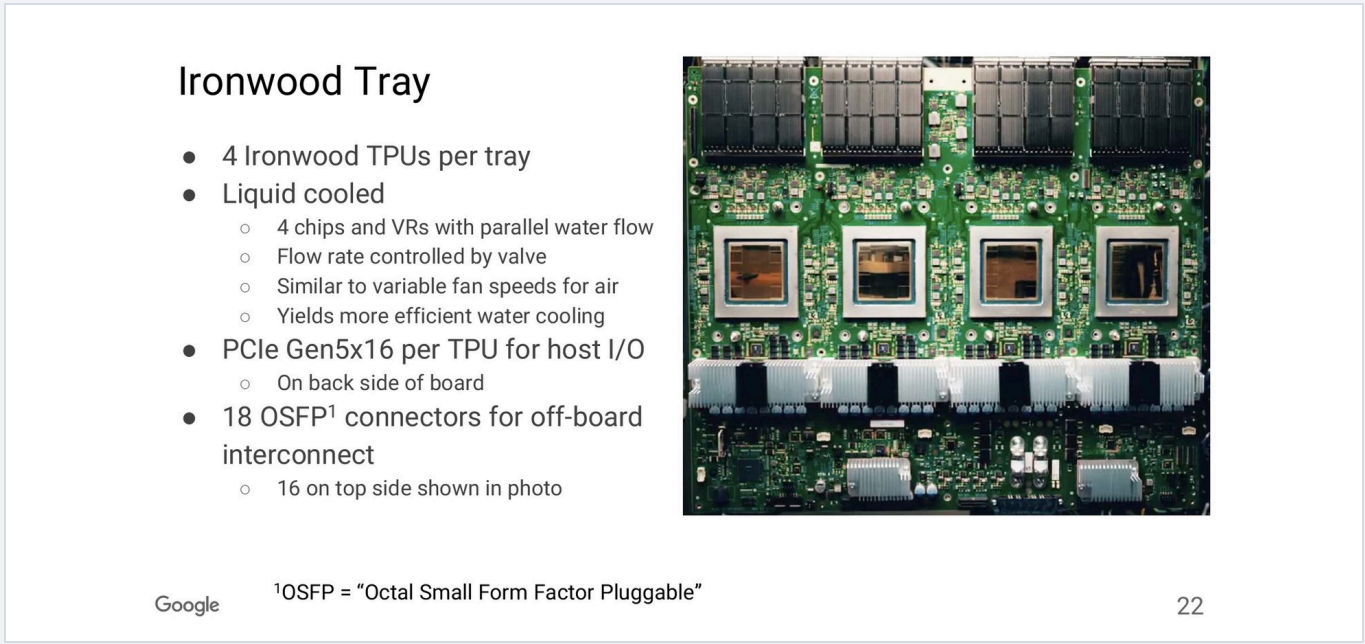
04

板子與托盤

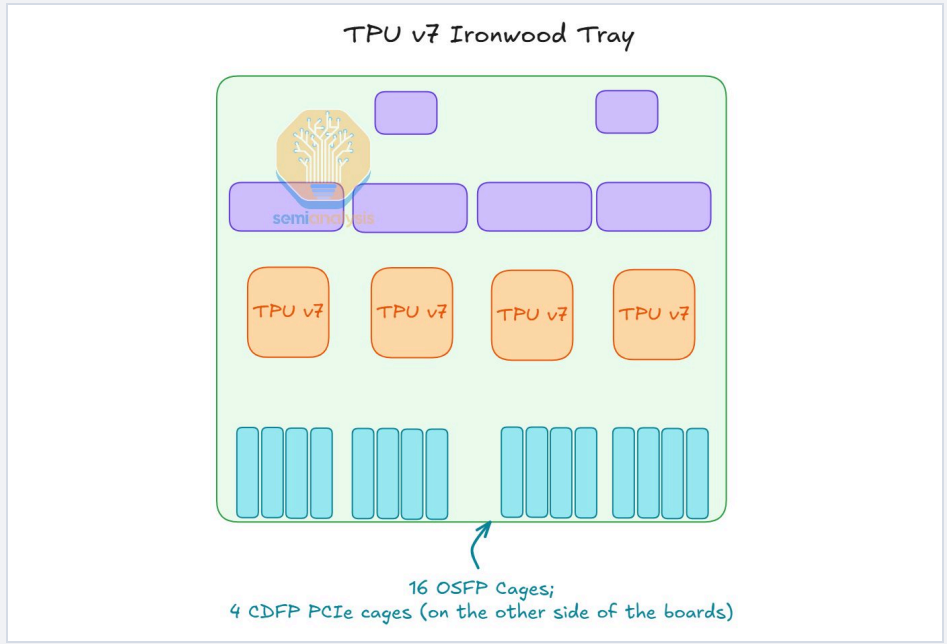
一個托盤 4 顆 TPU；v8 把供電、光模組一起整合進來，並改搭自家 Axion CPU

Ironwood 托盤：一塊主板、4 顆 TPU、整盤液冷

每顆 TPU 用 PCIe Gen5 x16 接 host（板子背面）；板外互連走 16–18 個 OSFP 光 / 銅接口



Google 原廠 Ironwood Tray 簡報：4 顆 TPU、液冷（晶片與 VR 並聯水路）、18 個 OSFP（照片可見 16 個）（Hot Chips 2025）



SemiAnalysis 托盤示意：4 顆 TPU v7、16 個 OSFP cage、4 個 CDFP PCIe cage 在板子另一面（2025-11-28）

板內 / 板外怎麼連

每顆 TPU 固定 2 個鄰居走 PCB 走線；另 4 個鄰居依在 cube 中的位置走 DAC 銅纜或光模組（SemiAnalysis）。

供電與冷卻

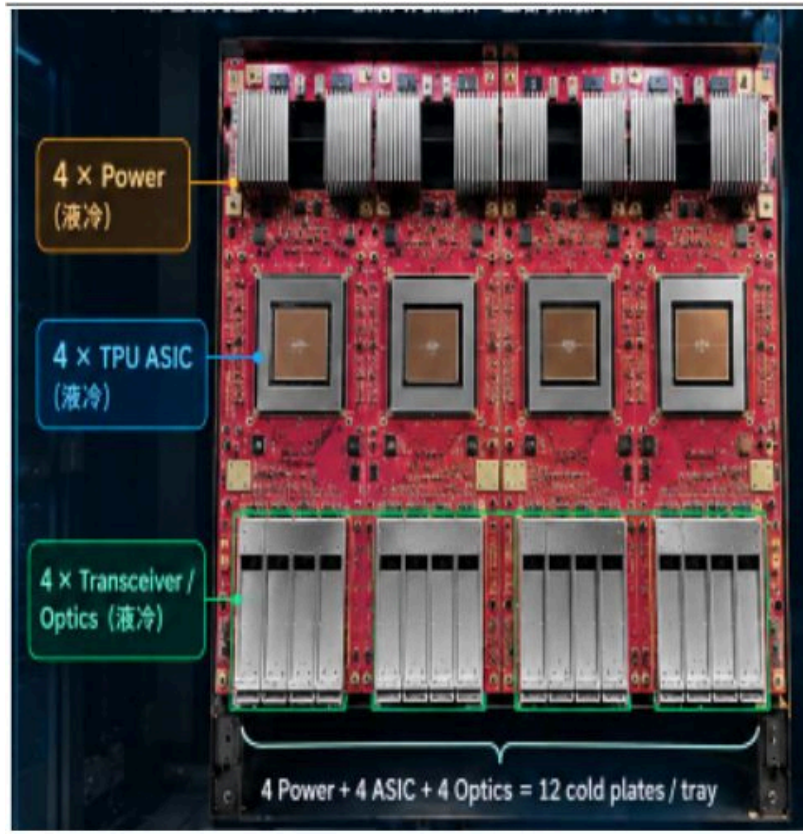
VRM 放在 PCB 背面（垂直供電），VRM 也要冷板；冷卻液流量由閥門主動控制，類似氣冷的可變風扇。

和 v4 比

TPU v4（2022）就是同一套：4 顆液冷封裝一塊板、前面板 PCIe＋16 個 OSFP、板內 4 條 ICI（ISCA 2023 論文）。

v8 托盤：4 顆 TPU、4 組供電、4 組光模組，全部上冷板

每盤 12 片冷板（v7 為 4 片）；8t 與 8i 托盤外觀相近，差在拓樸與記憶體



TPU v8 托盤俯視：4×Power、4×TPU ASIC、4×Transceiver / Optics 皆液冷，12 cold plates / tray（凱基，引 Google）

托盤上有什麼

元件	數量 / 說明	來源
TPU	4 顆（8t 或 8i）	Google
Host CPU	2 顆 Axion（4:2）	Google 圖、Aletheia、Nomura
NIC	4 張（Google 圖畫 1 NIC 配 2 CPU、4 TPU）	Aletheia / Google
光模組	4 組，液冷；1.6T	凱基
冷板	12 片：4 大（鍍金，TPU）＋8 小（供電、光模組等）	JPM、凱基
快接頭	每盤 2 顆 NSP1	JPM
熱插拔 MOSFET	每板 14 顆（v7 為 4 顆）	Edgewater
BMC	AST2700，配比約 70%	MS

Google 原廠 v8 文章只揭露拓樸與 Axion；PCB、供電、冷卻等板級細節都來自券商拆解。Nomura 另稱 v8 新增 Axion CPU 板（22L / M6）與 8i 的 all-to-all 交換板（22L）。

Host CPU：v7 以前用 x86，v8 起全面換成自家 Axion

TPU 本身不跑作業系統，資料前處理、排程、儲存都靠 host CPU；v8 把 CPU 數量加倍

各代 host CPU（官方規格）

世代	Host CPU	架構	配置
v4	—	x86	每 host 搭 4 顆 TPU（ISCA 論文）
v5p	Intel Sapphire Rapids	x86	4 晶片 VM：208 vCPU、448GB
v6e Trillium	AMD EPYC Genoa	x86	44—180 vCPU
v7 Ironwood	Intel Emerald Rapids	x86	224 vCPU、960GB；每節點 2 CPU 配 8 TPU
TPU 8t / 8i	**Google Axion**	**Arm**	**每 4 顆 TPU 配 2 顆 CPU**

為什麼要換、要加倍

- Google 原文：第八代 TPU 系統全面整合 Arm 架構 Axion，消除資料前處理造成的 host 瓶頸，讓 TPU 不會等資料
- Nomura 引 Google 副總裁 Amin Vahdat：每台伺服器實體 CPU 數加倍、改用 Axion
- 8t 另加 TPUDirect RDMA / Storage：HBM 直通網卡與儲存，繞過 host CPU 與 DRAM
- SemiAnalysis：RL 與 agent 工作量讓 CPU 重新變重要

機櫃裡的擺法（v7）

TPU 托盤與 CPU 托盤分開、交錯排列，用 PCIe DAC 銅纜連接；液冷櫃 16 TPU 托盤配 16 CPU 托盤（1:1），氣冷櫃 2:1（SemiAnalysis）。

口徑衝突

凱基寫 v7 每層「4 TPU＋4 Axion」，與官方規格頁（v7 用 Emerald Rapids）不符，疑把 v8 混寫；本簡報採官方口徑。

資料來源：Google Cloud TPU 規格頁；Google Cloud blog（2026-04-22）；SemiAnalysis〈CPUs are Back〉（2026-02-09）、〈TPUv7〉（2025-11-28）；Nomura（2026-06-30）

跟著 TPU 出貨的 Axion：2027 年約 400 萬顆

CPU:TPU 由 1:4 提高到 1:2；Google CPU 出貨 2026 年 125–130 萬 → 2027 年約 400 萬顆

1:4 → 1:2
CPU:TPU 配比（Agentic AI、深度推論）

125 → 400 萬顆
Google 自研 N3 CPU 出貨 2026 → 2027（法人，工商時報 10–08）

約 28 億美元
MS：創意 2027 年 Axion 相關營收（Axion2 300 萬＋Axion3 50 萬顆）

供應鏈

環節	廠商	說明
設計服務 / 晶圓物流	創意 3443	Axion 第 2、3 代；turnkey 毛利約 10%（MS）
晶圓	台積電	N3
封裝	FCBGA	下一代可能雙晶粒、仍用 FCBGA（野村）
FT 測試	京元電 2449	Advantest ATE＋鴻勁 ATC；Google CPU 占京元電營收 1.6%→5.7%（MS 10–08）
SLT	致茂 2360	Axion SLT handler（野村）
測試座	穎崴 6515	野村：預期參與 Axion 專案

怎麼看

- TPU 出貨越多，Axion 跟著同比例放大；這是 v8 才開始的新關係
- UBS：Google CPU 2027E 貢獻創意總營收 67%
- 傳聞：Google 與 AMD 探索在第 10 代 TPU 把 x86 CPU 與 TPU 同封裝（SemiAnalysis，經 INSIDE 2026–08–17；未證實）

資料來源：工商時報（2026–10–08）；UBS ASIC 與 CPU（2026–09–22）；MS（08–24、10–08）；華南（09–18）；野村半導體測試（07–24）

各家 AI 系統的 CPU 配比：TPU8 從 1:4 拉到 1:2

CPU 重新變重要：RL 與 agent 工作讓資料前處理、工具呼叫、狀態管理都吃 CPU

CPU：加速器配比（SemiAnalysis）

系統	CPU	配比
NVIDIA Vera Rubin	Vera（Arm，自研）	1 CPU：2 GPU
AMD MI455X	Venice（x86）	1：4
AWS Trainium3	Graviton5（Arm，自研）	1：4
Google TPU v7	Intel Emerald Rapids（x86）	2：8（＝1:4）
Google TPU 8t / 8i	**Axion（Arm，自研）**	**2：4（＝1:2）**

Axion 是什麼

- Google 自研 Arm 伺服器 CPU，2024 年發表、2025 年 GA；C4A 實例最多 72 顆 Neoverse V2 核心（SemiAnalysis）
- 雲端一般用途（C4A、N4A 實例）與 TPU host 共用同一條產品線
- TPU8 起成為所有 TPU 系統的 host（Google 官方）
- 下一代由創意負責第 2、3 代；N3 製程（工商時報、MS）

隨 TPU 出貨的 CPU 受惠：3443 創意｜2449 京元電（FT）｜2360 致茂（SLT）｜6515 穎崴｜5274 信驊（BMC）

資料來源：SemiAnalysis〈CPUs are Back〉（2026-02-09）；Google Cloud blog（2026-04-22）；工商時報（2026-10-08）

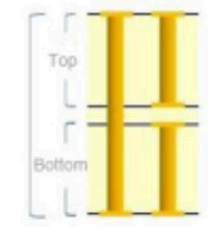
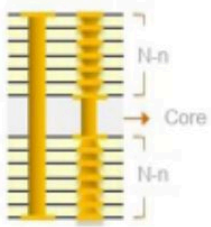
主板 PCB：層數從 24L 跳到 32–44L，v9 轉 HDI

N+M 混壓（訊號層 M8、電源層 M6）多次壓合良率可能掉 15–30%；HDI 層數少但更難做

各世代 PCB（各家並列）

世代	Aletheia	MS	福邦 10 月	凱基
v7 Ironwood	26L / M7	24L PTH / M7	22 / 24L	24L / M7 / HVLP2
v8i Sunfish	32L（15+2+15）N+M / M8＋M6	44L / M8＋M6	34L / M8	34–40L / M8+ / HVLP4
v8t Zebrafish	34L（16+2+16） / M8＋M6	40L / M8＋M6	34L / M8	同上
v9 Humufish	HDI 約 30–40L / M8–M9	HDI? / 測試 M9	6 壓 28L / M9＋M8	—
v10	HDI / M9–M10	M9?	—	—

FIGURE 11: HDI IS MORE ADVANCED THAN MULTI-LAM PCB

	Multi-lam	HDI
Process		
End Product	AI accelerators, switches, networking	for more advanced AI accelerators
Layers	30-50L+	20-40L
Strengths	Optimized for high layer counts (HLC), large panel sizes and power/signal separation (N+M)	Optimized for fine-line routing and high-density interconnects

多層板 N+M（30–50L+）與 HDI（20–40L）剖面比較（Aletheia，2026-10-04）

ASP 指數（Aletheia）

以 v5 / v6 為 1：v7 1.3、v8i 2.2、v8t 2.1、v9 2.5、v10 約 5–6；v5→v8 PCB ASP 約 2.5 倍。

交期變長

先進 AI PCB 從玻纖布到成板約 102 天，一般多層板 46 天；M9 鑽針壽命只有 100–200 孔（M6–M7 為 600–800）。

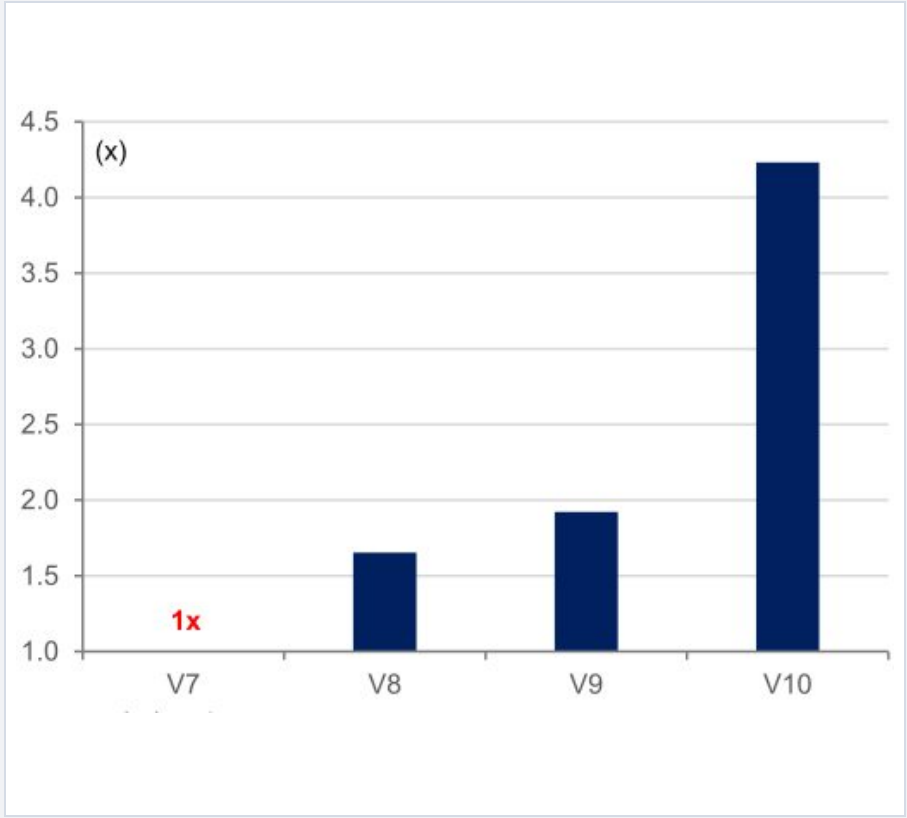
資料來源：Aletheia（2026-10-04）、MS CCL（2026-09-13）、福邦（2026-07、2026-10）、凱基（2026-09-30）

每櫃材料價值：PCB 兩倍、CCL 兩倍、銅箔兩倍半

MS 估 v9 再翻倍，Aletheia 估 v9 只再多 40–50%，分歧大

每櫃內含值（美元）

項目	v7	v8	v9	來源
PCB	19,474	37,861（+94%）	87,906（+132%）	MS
PCB	約 2 萬	3–4 萬	再 +40–50%	Aletheia
CCL	6,550	13,009（+99%）	32,288（+148%）	MS
CCL	約 7 千	1.2–1.8 萬	—	Aletheia
銅箔	1,783	4,457（+150%）	4,657（+4%）	MS
ABF 載板	80×85mm / 16L	90–100mm / 18–22L，ASP +40–50%	—	凱基



TPU 運算托盤 PCB 價值指數：v7＝1x、v8 約 1.65x、v9 約 1.9x、v10 約 4.2x（Aletheia，讀圖估計）

同表對照每櫃 PCB：NVIDIA GB300 37,491、VR200 91,227；Trainium T3 26,287、T4 84,770（MS）。Aletheia 的「每櫃」若以 8 托盤 / 32 顆計，與 MS 的 64 顆口徑不同，不可直接比。

PCB 與 CCL 供應商：名單各家說法不同

金像電在 GS、凱基是重要供應商，在福邦 7 月是備用；欣興在 v9 進入主供

PCB 主供（並列）

世代	Aletheia	福邦 10 月
v7	ISU、LCS、WUS、TTM	ISU / LCS / WUS / TTM / VGT
v8i	ISU、LCS、WUS、TTM、定穎	WUS / TTM / LCS / VGT
v8t	上列＋金像電（GCE）	ISU / WUS / TTM / VGT / 金像電
v9	上列＋欣興	ISU / VGT / WUS / TTM / 欣興

CCL

世代	等級	供應商
v7	M7	Panasonic（80–90%）、台光電
v8i	M8＋M6	Panasonic、台光電、南亞
v8t	M8＋M6	Panasonic、台光電、Doosan
v9	M8 / M9	Panasonic、台光電、Doosan

金像電 2368

GS：市占 2026 年 <5% → 2027 年 20–30%+；凱基：PCB 市占 20–25%、2027 年 TPU 營收貢獻 15–20%；福邦 7 月：v8 僅備用。

台光電 2383

凱基：v8 CCL 市占約 70%、2027 年 TPU 營收貢獻 20–25%；野村：Google TPU / CPU 板與交換器占其 AI 營收 2027F 58%；GS：TPU CCL 2026-09 開始出貨。

台燿 6274

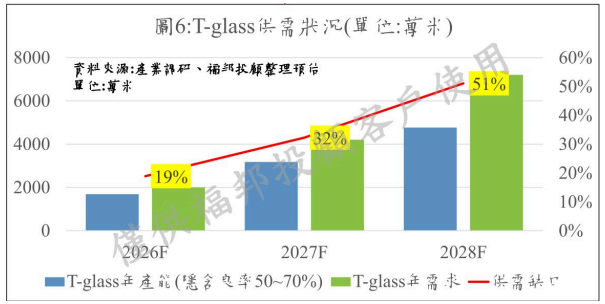
GS 07-17 稱 TPU CCL 2026-09 與台光電同時出貨；其他來源名單未列，證據偏弱。

資料來源：Aletheia（10-04）、福邦（2026-07、2026-10）、MS CCL Exhibit 37（09-13）、GS（07-17、08-11）、凱基（09-30）、野村（06-30）

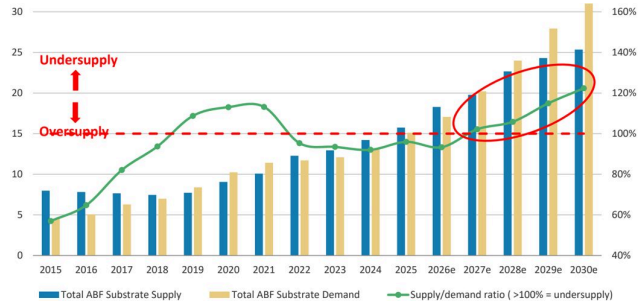
材料缺口：HVLP4 銅箔、T-glass、ABF 載板都不夠

M8 之後要同時升級低 Dk 玻纖布、HVLP 銅箔與樹脂；合格供應商少、驗證要 12—18 個月

材料	缺口 / 供需	TPU 相關	台廠
HVLP4 銅箔	MS：2027 年缺 31%、2028 年缺 20%；福邦：2026 年 27% → 2027 年 17%	v8 家族全用 HVLP4；2028 年 Google TPU 需求 1,100+ 噸 / 月	金居 8358（凱基：TPU 貢獻 5—10%）
T-glass 玻纖布	福邦：2026 年缺 19% → 2027 年 32% → 2028 年 51%	日東紡為 TPU 主要 T-glass 供應；聯發科幫 Google 爭取配額，v8t 因此上修	台玻
Low-Dk 二代布	福邦：缺 25%（2025）→ 33%（2026H2—27）	需求主要來自 VR200 與 Google v8t / v8i	台玻、富喬、南亞
ABF 載板	MS 供需比 2027 年 102% → 2030 年 122%（>100% 為缺貨）	EMIB-T 載板良率低、吃大量產能	欣興、南電、臻鼎
CCL	高階（M8+）合格只有台光電、Panasonic、台燿、Doosan、生益	CCL 每升一級 ASP 近乎翻倍	台光電、台燿、聯茂



T-glass 年產能與需求：缺口 19%→32%→51%（福邦，2026-10）



ABF 載板供需 2015—2030e，2027 年起供不應求（MS，2026-07-06）

僅供福邦投顧客戶使用

資料來源：MS CCL（2026-09-13）、MS 科技供應鏈（2026-07-06）、福邦 PCB（2026-10）

板上電源：TPU 是唯一真正量產垂直供電（VPD）的雲端業者

VRM 搬到晶片正下方的板背，TPU8 VRM 半導體內含值是 TPU7 的兩倍多

US$140
TPU8 VRM 半導體內含（TPU7 約 50–60）

4–6 個 × 24 相
每顆 TPU8 的 VPD 模組數與相數

4 → 14 顆
每板熱插拔 MOSFET（TPU7 → TPU8）

約 60 萬顆
每櫃 MLCC：供電 20 萬＋運算 40 萬

供應商

- VPD 合格三家：Infineon、Renesas、MPS；TPU7 原本 Infineon 獨供，缺貨後增補 MPS、Renesas
- MPS 是唯一同時賣 IC 與完整模組的，模組價約 IC 內含的 2 倍
- 熱插拔 MOSFET：AOSL、Infineon、Rohm
- Renesas 功率級產能擴 5 倍，1Q27 達每季 1 億顆

對其他零件的連動

- VPD 把電源元件放到板背，背面也要冷板；v9 晶片更大、正討論雙面冷板（JPM）
- 晶片背面放大量高容 MLCC（0201 10μF、0402 47μF、0603 100μF）；日系與三星為主（日電貿）
- 禾伸堂 3026：下游已從 GPU 擴散到 Google 等 TPU / ASIC 平台，無份額數字
- Meta 評估 gen4 MTIA 導入、NVIDIA 最早 Feynman 才用 VPD

資料來源：Edgewater Research 板上電源（2026-06-18）；SemiAnalysis（2025-11-28）；日電貿 call memo（2026-07-08）；JPM（2026-09-30）

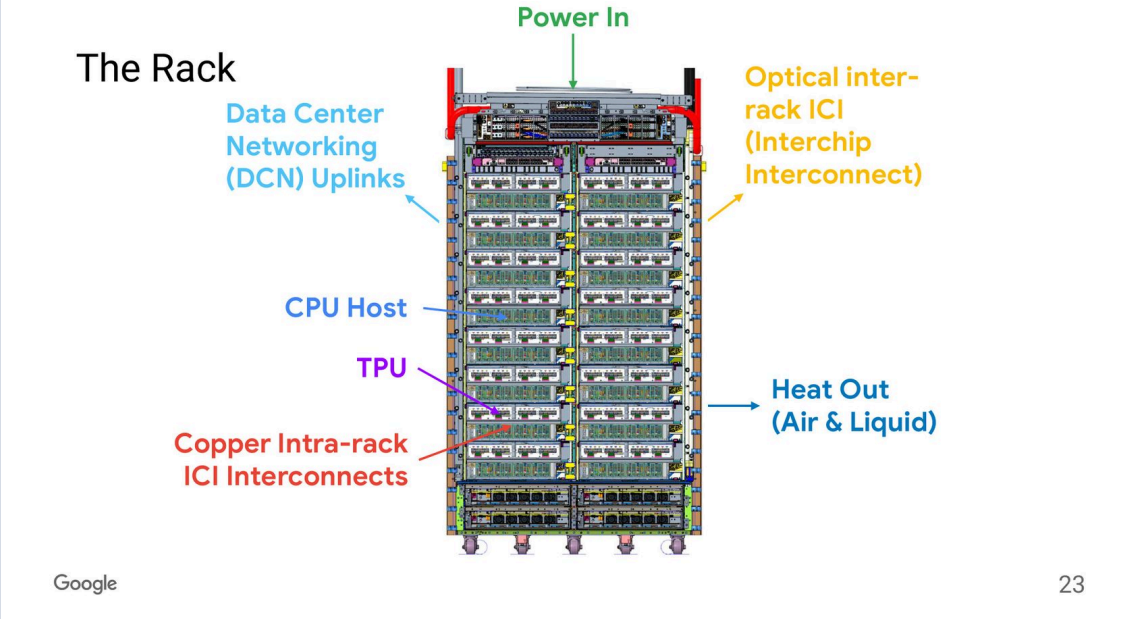
05

機櫃與網路

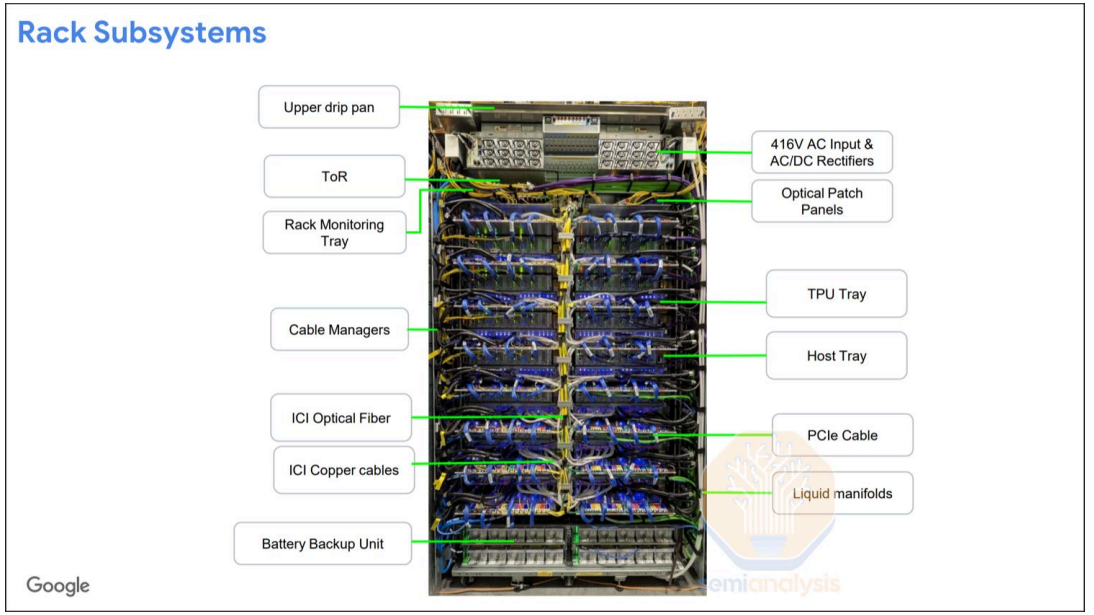
沒有交換器托盤：櫃內銅纜直連、櫃外光纖＋OCS，價值移到光與銅

Ironwood 機櫃：64 顆 TPU 就是一個 4×4×4 立方體

16 個 TPU 托盤＋16 個 CPU 托盤；頂部電源、底部電池，櫃內 ICI 全走銅纜



Google 原廠「The Rack」：Power In、DCN uplinks、TPU 與 CPU Host 交錯、銅纜櫃內 ICI、光纖櫃間 ICI、熱排出（Hot Chips 2025）



機櫃背面實景：ToR、416V AC 輸入與整流、光纖配線盤、TPU / Host 托盤、ICI 光纖與銅纜、PCIe 線、液冷歧管、BBU（Google 簡報，經 SemiAnalysis）

一櫃有什麼

16 TPU 托盤、16（液冷）或 8（氣冷）CPU 托盤、1 台 ToR、頂部 power shelf、底部 BBU（SemiAnalysis）。

功耗

9,216 顆約 10MW → 平均約 70kW / 櫃（推算）；UBS 估 v7x 約 90kW / 櫃。

價格

SemiAnalysis：40 萬顆 v7 約 100 億美元 → 約 160 萬美元 / 櫃（推算）；UBS 估 180 萬。

資料來源：Google Hot Chips 2025 第 23 頁（經 STH）；Google 簡報 Rack Subsystems（經 SemiAnalysis 2025-11-28）

TPU 機櫃和 NVIDIA NVL72 最大差別：沒有背板與交換器托盤

單晶片功耗較低、快接頭少一個數量級，但冷板片數多

項目	Google TPU（v7 / v8）	NVIDIA VR200 NVL72
每櫃運算托盤	16 個（v8 部分口徑 8 個）	18 個
每盤晶片	4 顆 TPU（CPU 另有托盤）	4 顆 GPU＋CPU 同盤
每櫃晶片	64 顆（v8 有 32 顆說）	72 顆 GPU
單晶片功耗	v7 約 1,000W；v8 約 800–1,300W	Rubin 約 2,300W
櫃內互連	無交換器托盤、無背板；托盤間外接銅纜（DAC）	NVLink 背板＋交換器托盤
櫃間互連	光纖＋OCS，組成 9,216 / 9,600 顆 pod	Ethernet / InfiniBand 交換器
每盤液冷（冷板＋QD＋歧管）	12 片冷板（4 大＋8 小），約 2,000–2,500 美元	5 片冷板，約 2,000–2,500 美元
每盤快接頭	2 顆 NSP1（單價為一般型 2–3 倍）	22 顆
整櫃液冷（不含 CDU）	約 3.8 萬（JPM）；4.5–5 萬（凱基）；5.18 萬（凱基奇鋐 09–21）	約 6.6 萬美元（JPM）
機櫃功耗	v7x 約 90kW；8t 約 90kW；v9 約 200kW（UBS）	—

資料來源：JPM 台灣散熱零組件 Table 6（2026-09-30）；SemiAnalysis（2025-11-28）；凱基（2026-09-30）；UBS AI 電力（2026-09-14）

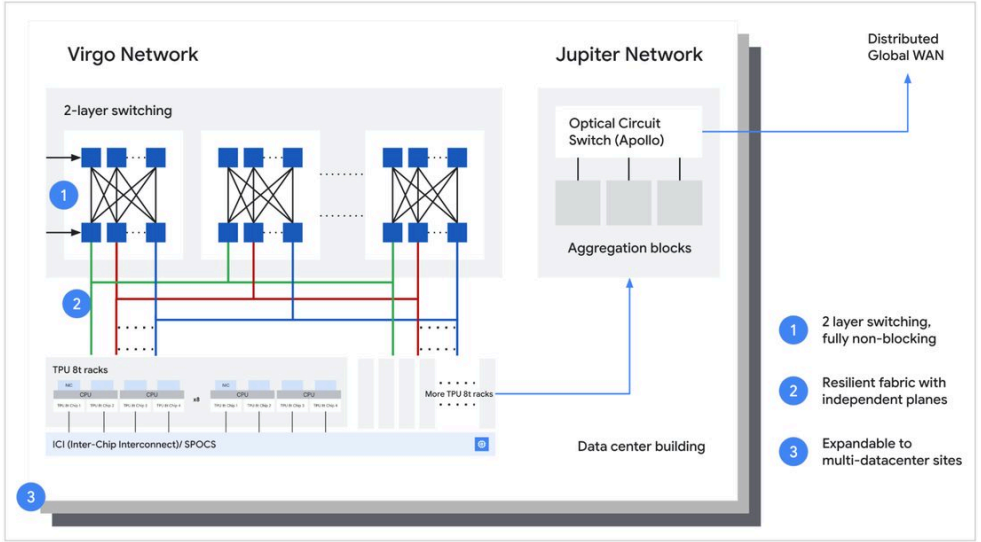
v8 每櫃到底幾顆？32 顆和 64 顆兩種口徑並存

引用任何「每櫃」數字前，要先確認分母是單一機櫃還是兩櫃一組

兩種口徑

口徑	來源	用在哪些數字
8 托盤 / 32 顆	Aletheia（每櫃由 64 降到 32）、UBS 8i、凱基 Boardfly 與 8t 網路推算（4×4×2）	Aletheia 每櫃 PCB / CCL；8t 約 300 櫃 = 9,600 顆
16 托盤 / 64 顆	凱基水冷與 PSU、JPM 水冷、UBS 內文、v7 原廠	整櫃液冷 3.8–5.2 萬、PSU 2.8–3 萬、QD 32 顆

凱基的 v8 機櫃照片是兩櫃並排一組，可能就是兩種口徑的來源（推論，原文未說明）。Google 原廠 8t 機櫃圖只畫「1 NIC＋2 CPU＋4 TPU」單元 ×8，未標每櫃托盤數。

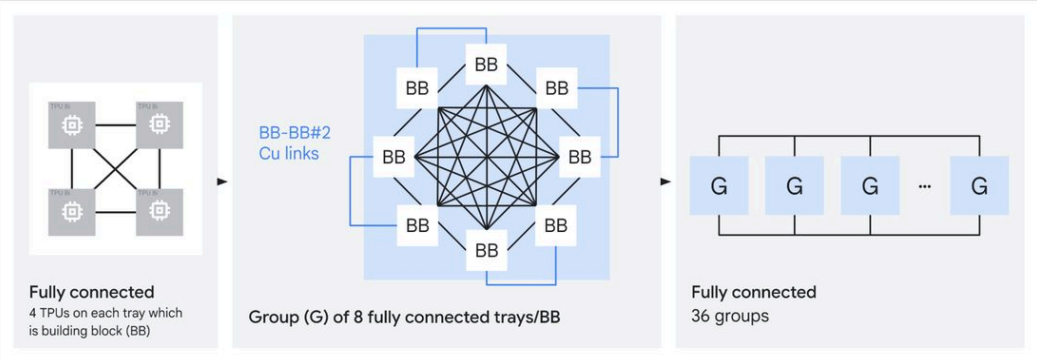


Google 原廠 TPU 8t 機櫃連線：每個單元 NIC＋CPU＋4 顆 8t，往上接 Virgo（東西向）與 Jupiter（南北向）（Google Cloud blog 2026-04-22）

資料來源：Aletheia（10–04、10–06）、UBS（09–14）、凱基（09–30）、JPM（09–30）、Google Cloud blog（2026–04–22）

拓樸：8t 沿用 3D Torus，8i 改 Boardfly 把跳數從 16 降到 7

MoE 與推理模型大量 all-to-all 通訊，延遲比頻寬重要；Google 稱 all-to-all 延遲最多改善 50%



Google 原廠 Boardfly 三階：4 顆 8i 全互連成 building block；8 個 BB 以銅纜全互連成 group；36 個 group 經 OCS 全互連（2026-04-22）

兩種拓樸

- **3D Torus（v7、8t）**：每顆晶片接上下左右前後 6 個鄰居；4×4×4＝64 顆一個 cube，cube 六面 96 條光纖接 OCS；9,216 顆＝144 cube
- **Boardfly（8i）**：托盤內 4 顆成環→8 塊板銅纜全互連（32 顆）→36 組經 OCS＝1,152 顆，任兩顆最多 7 跳（同規模 3D Torus 為 16 跳）
- Boardfly 也減少每櫃 ICI 銅纜（Aletheia）
- **下一步傳聞**：永豐論壇稱 2028 世代升到 6D Torus，每顆直連鄰居 6→12、光 I/O 推向 3.2T（其他來源未證）

從機櫃到資料中心：Virgo 東西向、Jupiter 南北向

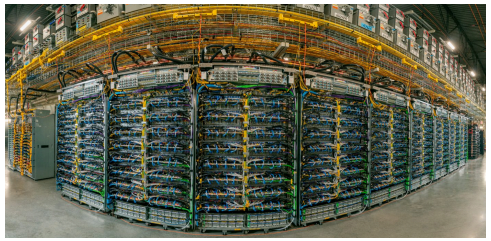
一個 Virgo 叢集 134,400 顆 8t、47 Pb/s；JAX＋Pathways 可組成 100 萬顆的邏輯訓練叢集



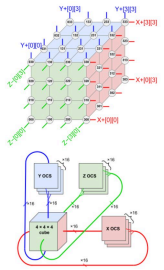
8i（左，OCS＋Boardfly，1,152 顆 / pod）與 8t（右，OCS＋3D Torus＋Virgo，9,600 顆 / pod、14 pod＝134,400 顆）（Aletheia，引 Google）

三層網路

層級	內容
ICI（scale-up）	晶片間直連；v7 9.6 → v8 19.2 Tb/s；櫃內銅、櫃間光＋OCS
Virgo（8t 東西向）	兩層無阻塞、200G SerDes 乙太網；134,400 顆、1.6 YottaFLOPS
Jupiter（南北向）	100G SerDes 乙太網＋Apollo OCS；接儲存、CPU 伺服器、全球 WAN
DCN 頻寬	每顆 100 → 400 Gbps



Ironwood superpod 一角：直接互連 9,216 顆 TPU 的機櫃群與大量光纖（Google Cloud blog 2025-11-06）



TPU v4 論文：4×4×4 cube 六面連到 X / Y / Z 三組 OCS（ISCA 2023）

網路硬體：TPU8 每顆內含值是 TPU7 的 2–5 倍

交換頻寬 5 倍、光纖 2 倍多；OCS 生命週期總台數 12 倍（換算每顆約 3.4 倍）

每顆 TPU 的網路用量（Aletheia）

項目	TPU7	TPU8	變化
乙太網交換頻寬	0.4 Tb	2.0 Tb	約 5 倍
光收發模組	1.8 顆	3.0 顆	+67%
光纖線纜	8.0 條	18.2 條	逾 2 倍
銅纜 DAC / ACC	1.5 條	1.9 條	200G/lane 單價更高
OCS（生命週期總台數）	20,833	258,542	總量 12 倍，每顆約 3.4 倍

假設 TPU7 生命週期 400 萬顆、TPU8 1,400–1,500 萬顆。

另一種算法：只算 scale-up（凱基）

	v7	v8i	v8t
每 pod 晶片	9,216	1,152	9,600
光纖＝光模組	13,824	1,440	19,200
OCS 台數	48	約 5	64
光模組 / 顆	1.50	1.25	2.00

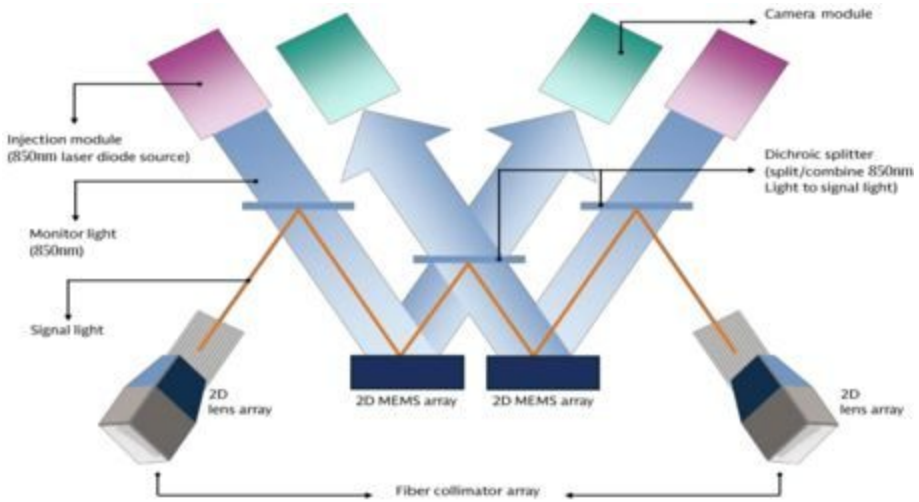
凱基只算 ICI，Aletheia 含 Virgo / Jupiter / DCN，兩者不可直接比。

Aletheia 三大受惠：Lumentum（光模組、雷射、OCS）、光聖 6442（光纖線纜，原文 EZconn）、Celestica（交換器與 OCS 代工）

資料來源：Aletheia〈AI Networking〉Figure 2（2026-10-06）；珩石研究筆記摘譯（10-08）；凱基 圖 18（09-30）

OCS：Google 獨有的光路交換，用鏡子反射光

不做光電轉換，800G 升到 3.2T 都不用換交換器；約 95% 的 OCS 由 Google 使用



MEMS 型 OCS 光路：光纖準直陣列 → 兩片 2D MEMS 鏡面陣列 → 目標埠；上方 850nm 監控光即時校正鏡面（GS Exhibit 40，Google Palomar OCS）



Lumentum R300 OCS 產品照（凱基 2026-09-30）

為什麼用 OCS

- 降延遲、省 SerDes 與 DSP，單埠功耗遠低於電交換器
- 可繞過故障節點、重組切片；Google 2015 年起 Apollo 專案自研

規模與供應

- 單台 5–20 萬美元（傳統交換器 1–10 萬）
- Lumentum OCS 在手訂單逾 4 億美元，預期 CY2027 貢獻 10 億美元；泰國產線排到 2027
- 代工：Fabrinet、Celestica；Google 自研＋外購 Lumentum

OCS 出貨時程：Google 自研十年，2026 年起外購放量

Lumentum 在手訂單逾 4 億美元、CY2027 目標 10 億美元；TPU8 生命週期 OCS 需求約 25.9 萬台

落地時間軸

時間	事件
2015	Google 啟動 Apollo 專案自研 OCS
約 2014—22	Jupiter 資料中心網路逐步導入 OCS（Google 2022 年說「過去八年」）
2020—23	TPU v4 首度用 OCS：48 台 Palomar（136×136 埠）連 4,096 顆；2023 年論文公開
2025	Ironwood superpod 以 OCS 連 9,216 顆；OCP 成立 OCS 專案（Google、Lumentum、Coherent、NVIDIA、Microsoft、iPronics）
2026 上半	Lumentum backlog 逾 4 億美元（2 月），多數 CY26H2 交付；Coherent OCS 已出貨 2—3 季、洽談客戶 10 家以上
2026 下半	TPU8 量產；Lumentum OCS 爬坡快於預期（MS 08–12），泰國產線滿載、交期排到 2027
2027	Lumentum 目標 OCS 營收 10 億美元；InnoLight 目標推出矽光 OCS

出貨量級

口徑	數字
TPU7 生命週期	約 2.08 萬台（Aletheia）
TPU8 生命週期	約 25.9 萬台（含 Jupiter 與 8i SPOCS）
每 pod（凱基，只算 ICI）	v7 48 台；8i 約 5 台；8t 約 64 台
使用者集中	約 95% 由 Google 使用（群益訪查）
單價	每台 5—20 萬美元（GS）

OCS 供應鏈：MEMS 是主流，台股直接受惠的是光纖線纜

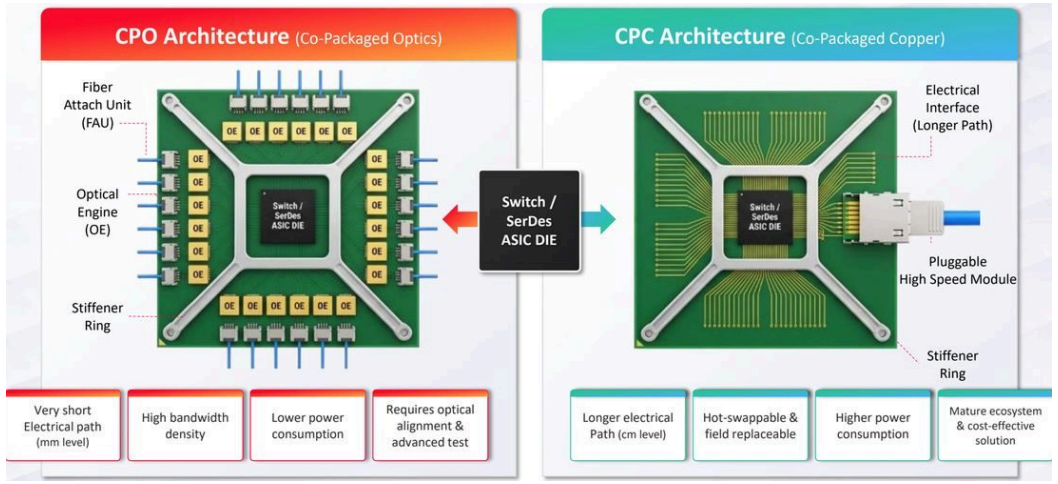
系統由 Google 自研＋Lumentum 外購；代工往 Fabrinet 移，台廠著墨在光纖與配線

環節	MEMS 路線（Google 主流）	其他路線
系統設計	Google 自研（Palomar / Apollo）、Lumentum（R300）、Calient	液晶：Coherent（約 64×64 埠）；壓電：Huber+Suhner；矽光：InnoLight、iPronics、Taclink
MEMS 鏡面	賽微電子、XDLK	—
監控雷射 / FAU	雷射：Lumentum、Coherent；FAU：Corning	光學元件：Optowide
光纖線纜	光聖 6442（EZconn；Aletheia 列 TPU8 三大受惠之一）	—
系統代工	昂納 AFR（展出 320×320 MEMS OCS）；Fabrinet（MEMS 型與其 MEMS mirror 製程相近，客戶要提高委外）；Celestica（EMS，規模小）	Luster；封裝設備：Robotechnik

Google 用 OCS 的前提是 WDM 光模組（單纖 4 或 8 波、2.4T），OCS 本身不需要光模組，但兩端伺服器仍要；矽光小埠數 OCS 仍屬題材階段（Omdia 10-03）。

CPO / NPO 不是 v8 的驅動力：可插拔＋OCS＋銅纜仍是主流

SemiAnalysis：未來幾年看不到 NPO 在 Google 的角色；NPO 最早搭 Tomahawk 7 與 TPU v10t



CPO（左）vs CPC（右）：CPO 毫米級電路徑、功耗低但壞了整顆報廢；CPC 公分級、可熱插拔、成本低（穎崴 CPO 論壇，2026-05-14）

Google 的光銅路線

- SemiAnalysis：v8i、v8t 的 scale-up 面板密度沒逼近極限，用 NPO 只增風險；Virgo 首批隨 v8t 部署、目前未用 XPO 或 NPO
- v8i 每顆 scale-up：2 條 1.6T ACC＋1 條 1.6T DAC＋1 個 1.6T 可插拔光模組
- Virgo 之後可能隨 Tomahawk 與 Humufish / Icefish 走 XPO（Arista 的 8 收發合一可插拔）
- 銅纜距離：200G/lane 約 2.5 公尺、400G 減半；AEC 傳聞為 Anthropic 要求（貿聯自稱一供代工、2028 年才顯著）
- 散熱連動：JPM 認為 CPO / NPO 與 1.6T / 3.2T 光模組液冷是散熱下一個驅動

資料來源：SemiAnalysis〈NPO 光互連接棒〉（2026-07-13）；穎崴 CPO 論壇（05-14）；群益 ECOC（10-06）；貿聯私訪（09-23）；JPM（09-30）

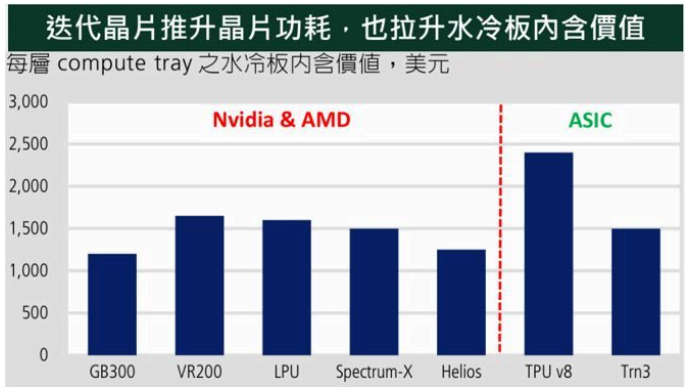
06

散熱與電源

功耗不高，冷板價值卻最高；電源往 10kW PSU 與 ±400V 走

TPU 單層冷板價值是各平台最高

價值來自冷板片數（每盤 12 片）與鍍金底座，不是功耗



每層運算托盤水冷板內含值：NVIDIA / AMD 約 1,200–1,650 美元，TPU v8 約 2,400 美元，Trainium3 約 1,500 美元（散熱產業研究簡報）

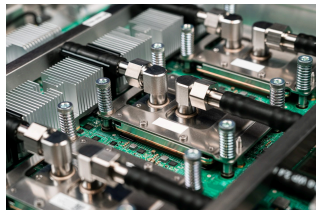
單層與單櫃（並列，分母都是 16 托盤）

來源	單層	單櫃
JPM（09–30）	2,000–2,500	約 3.8 萬
凱基（09–30）	2,500–3,000	4.5–5 萬
凱基奇鋐（09–21）	3,240（12×270）	5.18 萬
散熱簡報（06–30）	2,400；v9 約 3,500	3.8–4.8 萬（推算）

不含 CDU。對照 VR NVL72 約 6.6 萬、Trn3 Max 約 8.6 萬（JPM）。



液冷機櫃側面：不鏽鋼供回水管、快接頭與歧管、右側 CDU（凱基）



Ironwood 板上冷板特寫：不鏽鋼冷板、彎頭接頭與水管（Google Cloud blog 2025–11–06）

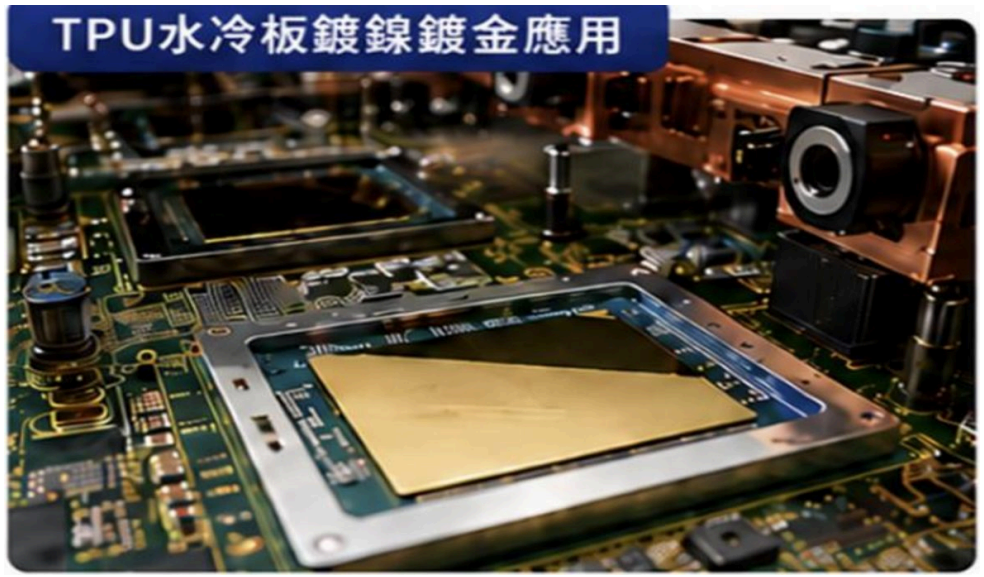


Ironwood superpod 第三代 CDU 與垂直歧管（Google Cloud blog 2025–11–06）

資料來源：散熱產業研究簡報（2026–06–30，圖源凱基）；JPM（09–30）；凱基奇鋐（09–21）

鍍金冷板：拿掉上蓋、錳片直接貼冷板，底座必須鍍金

銅鎳金三明治防止錳銅遷移脆化；每片多出約 65 美元，良率與執照門檻高



TPU 水冷板鍍鎳鍍金：冷板底座銅鎳金結構（匯鑽科 2026-07-21 法說簡報）

製程與經濟

- TPU 取消上蓋，晶片上直接放錳片接觸冷板，兩層介質變一層；風險是空焊與錳銅遷移形成合金脆化
- 冷板銅基材鍍鎳保護、鎳上鍍軟金防氧化，金層不可省；TPU 下一代仍鍍金
- Google 冷板：電鍍約 35＋金料約 30＝約 65 美元 / 片；NVIDIA 不帶料、約 20 美元
- 匯鑽科：TPU 鍍金冷板市場 2026 年 300 萬片以上、2027 年翻倍到約 700 萬片
- 產能：深圳 1,000＋4,500＝5,500 片 / 日；客戶預告 4Q26 需求 13,000–15,000 片 / 日
- 時程：7/21 說 8 月 PVT、9 月量產；9/15 說 EVT / DVT 完成、即將 PVT（說法前後不一）

相關：匯鑽科 8431（經奇鋐下單）｜奇鋐 3017｜時碩 4566（古河電工 TPU 水冷三年大單，年約 20 億元，媒體＋公司簡報，10-06【新】）

散熱份額：奇鋐份額有四種說法，並列不調和

v8t 冷板：奇鋐、古河；v8i：Cooler Master、Boyd（凱基）；Broadcom 側奇鋐仍在洽談

元件	TPU 規格與價值	台廠與份額（並列）
冷板	每盤 12 片；冷板 ASP 200–300 美元	奇鋐 3017：JPM 約 30% / Andrew 40% / 散熱簡報 >50% / 公司自述聯發科側約 50%、總 20–30%；Citi 10–07【新】：4Q26 TPU 專案量產
冷板設計	v8 冷板由古河電工設計	古河份額約 40%（Andrew），持有奇鋐 13.8%；Cooler Master 約 20%；Boyd 份額下降
鍍金	每片約 65 美元，2026–09 起量產	匯鑽科 8431（經奇鋐）
快接頭 QD	JPM：每盤 2 顆 NSP1、整櫃 32 顆、約 1,800 美元	富世達 6805：4Q26 起出貨；TPU QD 市占保守參考奇鋐約 4–5 成（call memo）
補強環	單顆 15–30 美元（v7 15）；v9 可能改可拆式上蓋，內含值或 10 倍（未確認）	健策 3653 約 80%、一詮 2486 約 20%（CSP ASIC）
CDU	v7 約 1.4MW、v8t 約 2MW，一台帶約 8 櫃；第五代 Project Deschutes	台達電 2308、高力 8996（TPU 專屬供應商來源未載）

口徑提醒：富世達 call memo 寫「每 Tray 32 顆 QD」與 JPM「每櫃 32 顆」差 16 倍；散熱簡報圖中「快接頭每櫃 4.5 萬美元」數值等於凱基整櫃液冷總值，疑為標錯；「均熱片每櫃 2.64 萬美元」與單顆 15–30 美元差 14 倍以上——這三個數字本簡報不採用。

補強環與上蓋：CSP ASIC 補強環健策約 80%、一詮約 20%

晶片越大越容易翹曲，市場由單片 Lid 轉向 Stiffener；v9 若改可拆式上蓋，內含值可能放大約 10 倍

TPU 用的是什麼

- v7 用 stiffener，單顆 ASP 約 15 美元；v8i / v8t 也用 stiffener 防翹曲，15–30 美元
- TPU 取消上蓋、鋁片直接貼冷板，所以封裝端主要是補強環，不是均熱蓋
- 對照：Rubin 1-piece lid 約 50 美元
- v9 晶片更大、功耗可能達 2,000W；若改「可拆式上蓋＋補強環」，內含值約 10 倍（JPM，未確認）
- Daiwa：可拆式上蓋未來 2–3 年由 GPU 與 ASIC 客戶同步採用

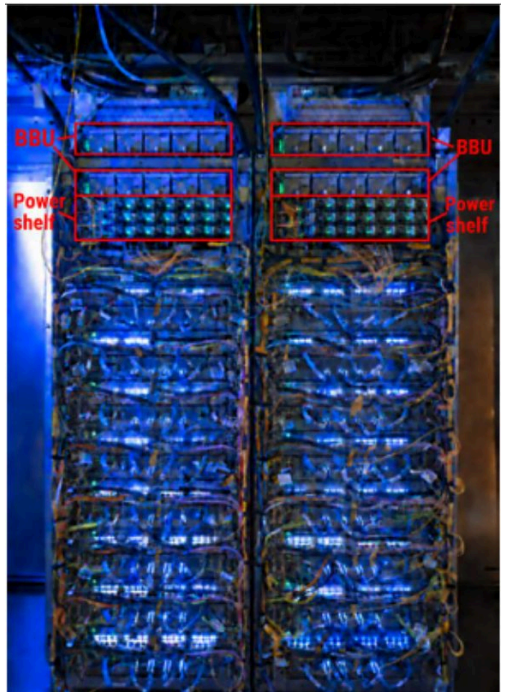
健策 3653 vs 一詮 2486

	健策	一詮
CSP ASIC 補強環市占	約 80%	約 20%
TPU 相關	Google TPU 欄列名（散熱簡報）；凱基：2027 TPU 營收 <10%	Stiffener 占散熱元件營收約 40%，主要客戶 Google TPU、AWS Trn2 / 3、Broadcom ASIC（中信）
其他產品	GPU lid、高規案領先	Grace CPU lid；Microsoft 高階 lid（鋁片＋鍍金，ASP 高 50%）；MCL 開發中
產能	—	2026、2027 年散熱產能各翻倍；稼動率 65–70%
營收目標	—	2026 散熱產值 36 億元以上、2027 營收 60 億元以上

JPM 09-29：一詮已拿到部分 CPU 與 ASIC stiffener 案，但 AI GPU lid 與 CPO ASIC stiffener 等高規案，健策差距仍大。

電源：PSU 從 5.5kW 升到 10–12kW，每櫃價值多 40–50%

v8 相容 ±400V DC，但 50V 仍是本代主流；BBU 在櫃頂、power shelf 在其下



TPU 雙櫃正面：紅框為頂部 BBU 與下方 power shelf（凱基，引 Google）

每櫃電源（並列）

項目	v7	v8
配置（凱基）	6 排 power shelf × 4 顆 5.5kW	8i：3 排 × 6 顆 10kW
配置（UBS 8t）	—	2 個 72kW shelf × 6 顆 12kW
PSU 單櫃價值	約 2 萬美元	2.8–3 萬（凱基）；2.62 萬（UBS）
台廠貢獻（凱基）	—	台達電 2027 年 TPU 營收占 20–25%；光寶 <10%

UBS 同表：Trn3 1.98 萬、MTIA 400 2.51 萬、Maia 200 2.15 萬、MI455 8.47 萬美元。

BBU

GS：BBU shelf ASP 2025 年 1.56 萬 → 2030 年 2.63 萬美元，電池占 BOM 80–90%；未見 Google 專屬數量。

HVDC 電源櫃

MS【新】：800VDC 電源櫃 BOM 為 PSU 50–55%、BBU 25–30%；每瓦內含值是櫃內 PSU 的 4–5 倍。

資料來源：凱基（2026-09-30）；UBS AI 電力 Figure 29（2026-09-14）；MS 整合電源 IPS（2026-10-06）；GS BBU（09-24）

±400V 高壓直流：Google 是推手，但時程說法不一

Google 2025 年在 OCP 推 ±400VDC（Mt Diablo），機櫃由 100kW 擴到 1MW

各家時程

來源	說法
Google（2025–04）	與 Meta、Microsoft 合作 ±400VDC；電源與電池移到側邊電源櫃，端到端效率約 +3%；2030 年前 ML 需 >500kW / 櫃
凱基（09–30）	台達電 ±400V 電源櫃「已開始出貨」、主供、市占高；Flex 第二、光寶 2026 年切入
Edgewater（06–18）	Google ±400VDC 出貨約 4Q27、綁 TPU v9 首批；之後轉 800V；>200–240kW / 櫃才必須電源櫃
UBS（09–14）	Google 與 Meta 比 NVIDIA 更早導入 ±400V；2028 年起 4 晶粒架構帶動電源櫃
MS（10–06）【新】	800VDC 電源櫃 4Q26 末小量、2027 年有限、2027 年底到 2028 年較廣導入
TD Cowen（09–29）	800V 架構遞延到約 2028 年後

機櫃功耗往上走

世代	kW / 櫃（UBS）
v5e	28
v6	40
v7x	90
v8i	50
v8t	90
v9	200

UBS 內文另寫 v8i 90kW，表內不一致。

相關：台達電 2308（±400V 主供）｜光寶科 2301（48V ASIC 電源櫃 2026 底量產、2027 轉 800V）

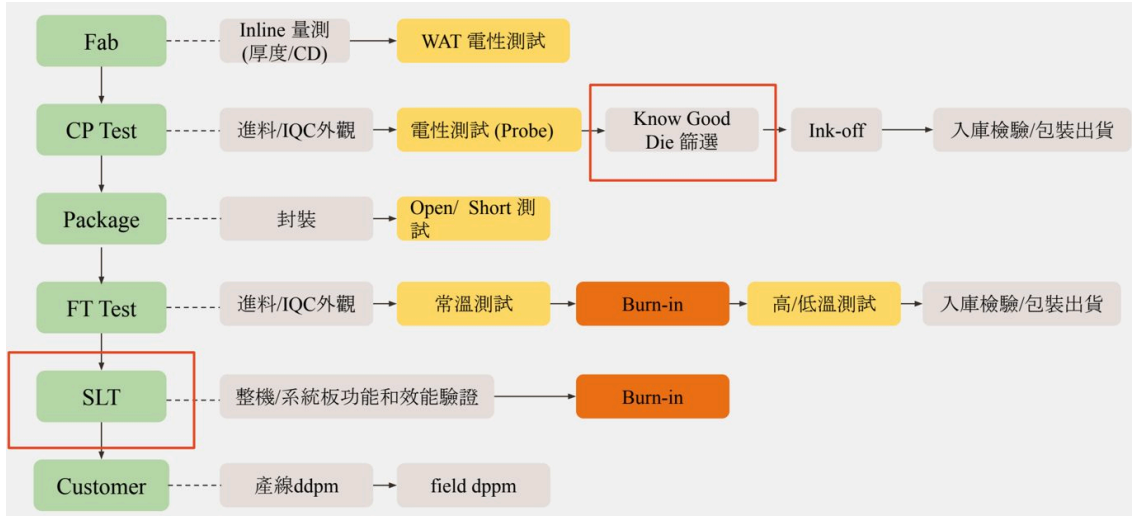
07

供應鏈與市場

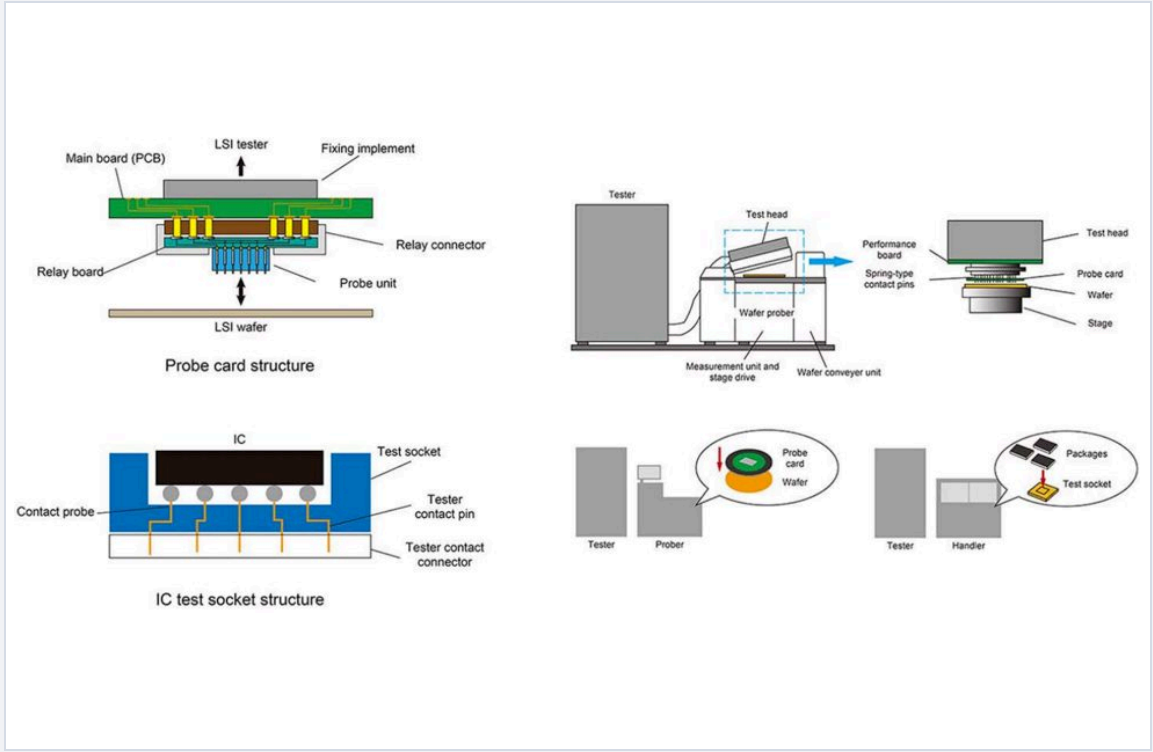
封測分兩條線；台股曝險一覽；市場多空與追蹤訊號

封測流程：CP → 封裝 → FT / 燒機 → SLT

AI 晶片燒機正由 GPU 擴到 TPU；SLT 用來抓最後約 0.5% 的失效



封測流程：Fab → CP（KGD 篩選）→ 封裝 → FT（常溫→燒機→高低溫）→ SLT → 客戶（測試產業研究簡報）



探針卡（CP）與測試座（FT）結構對照（野村 Fig.67，2026-07-24）

CP 晶圓測試
探針卡＋探測機＋ATE｜旺矽、精測

封裝
CoWoS / EMIB-T｜台積電、日月光

FT 成品測試
ATE＋分選機＋測試座｜京元電、鴻勁、穎崴

燒機
100°C 以上、數小時｜京元電

SLT 系統級
聯發科線才有｜鴻勁、致茂

資料來源：測試產業研究簡報（2026-09-12）；野村〈Advanced Semi Testing〉（2026-07-24）；MS / Teradyne

封測分兩條線：2026 年看京元電，2027 年看分流

Broadcom 線（v7 / v8i）沒有 SLT、2026 年就認列；聯發科線（v8t / v9）多一段 SLT、量 2027 年才出來

環節	Broadcom 線（v7 Ironwood / v8i Sunfish）	聯發科線（v8t Zebrafish / v9 Humufish）
2026 / 2027 出貨（UBS）	368 萬 / 500 萬顆	45 萬 / 400 萬顆
CP 探針卡	旺矽為主：v8 / v9 需求約 v7 的 3 倍（Aletheia）	v9 未定：TechnoProbe、FormFactor、Keystone、旺矽
FT 代測＋燒機	京元電主力：TPU 營收 2026E 26.5 億 → 2027E 63.5 億元（BofA）	京元電與日月光分食（UBS 因此下修京元電 EPS 11–15%）
FT 分選機	鴻勁（Google TPU / CPU 所有專案）	鴻勁
測試座	穎崴（野村供應鏈圖列名；產業報告寫「預期參與」）	同左，無分線資料
探針卡（其他）	精測（PoR 驗證中）	雍智：經聯發科承接 TPU 高低混針探針卡（新聞）
ATE	Advantest（Teradyne 認證中）	Advantest
SLT	無	鴻勁（研究簡報）；鴻勁＋致茂（野村）
封裝後段 oS	TPU oS 外包日月光 70–90%，單顆含量 150 → 180 美元（MS 10-08【新】）	同左

受惠：2449 京元電 | 3711 日月光 | 6223 旺矽 | 7769 鴻勁 | 6515 穎崴 | 2360 致茂 | 6510 精測 | 6683 雍智

資料來源：野村（07-24）、測試產業研究（09-12）、BofA（08-08）、UBS（07-01、08-07）、MS（10-08）、Aletheia（06-10、06-23）、JPM（07-12）

京元電與日月光：MS 10–08 更新 TPU 數字

京元電 TPU 占比下修，但加上 Google CPU 後合計不變；日月光 TPU 占先進封裝測試營收升到 18%

京元電 2449：TPU 占營收（三組數字）

來源	2026E	2027E
BofA（08–08）	5.5%	8.6%
MS（08–10）	7–8%	10%+
MS（10–08）TPU	5.7%	7.3%
MS（10–08）Google CPU	1.6%	5.7%
MS（10–08）Google 合計	**—**	**10–15%**

日月光 3711

項目	2026E	2027E
TPU 占 LEAP 營收	5 億美元（14%）	16 億美元（18%）
LEAP 營收	—	90.8 億美元（公司指引 >75 億）
TPU oS 外包比例	70–90%	同左

MS 京元電測試模型的 TPU 顆數為 2026E 368 萬、2027E 450 萬（只算京元電測的量，非總量）；日月光模型 2027E 693 萬。

資料來源：MS〈CPO 測試與封裝供應鏈〉（2026–10–08）【新】；BofA（08–08）；MS（08–10）；UBS（07–01）

設計趨勢（一）：晶片層——越拆越細、越做越大

Google 收回運算晶粒、晶粒數與光罩面積倍增、記憶體與 SRAM 加大、精度一路往低走

趨勢	v7 Ironwood	v8（8t / 8i）	v9 → v10	影響誰
產品線分化	訓練＋推論一顆	8t 訓練、8i 推論兩顆	v9t / v9t+ / v9i 三顆	設計服務商各拿不同線
設計權	Broadcom 全包	8t 運算晶粒 Google 自做	v10 交 KGD、周邊拆給多家（semi-COT）	聯發科、創意、Marvell
晶粒數 / 光罩	2 運算晶粒（雙 chiplet）	8i 約 2.5 倍光罩	v9 24 晶粒、9.5x → v10 46 晶粒、15x	台積電、Intel、載板
封裝	CoWoS-S	CoWoS-S / L	EMIB-T、面板級 → v10 首度 3D	欣興、Ibiden、力成
HBM	8×HBM3E 8Hi 192GB	6–8×12Hi，216–288GB	12×HBM4E → 16 顆；客製 base die	記憶體三雄
片上 SRAM	—	8t 128MB、8i 384MB	—	推論延遲、晶粒面積
精度	FP8	原生 FP4	—	對 GPU 成本比較
SerDes	112G（ICI）	200 / 224G	336G 特規（聯發科）→ 448G	聯發科 IP、PCB 材料
專用引擎	SparseCore	8t 加 LLM Decoder Engine；8i 以 CAE 取代 SparseCore	—	—
功耗	約 1,000W	約 800–1,300W	約 2,400–4,500W → 約 10,000W	散熱、供電規格

資料來源：GFHK（08-24）；MS（08-24、10-05）；Aletheia（10-04）；Google Cloud blog（2025-11-06、2026-04-22）；永豐金論壇（06-23）；SemiAnalysis（05-18）；JPM（07-01）

設計趨勢（二）：系統層——拓樸、CPU、散熱、供電一起換代

TPU 的升級越來越不是晶片本身，而是晶片外面的網路、冷卻與電

趨勢	v7 Ironwood	v8（8t / 8i）	v9 → v10	影響誰
拓樸	3D Torus＋OCS	8t Torus＋Virgo；8i Boardfly（跳數 16→7）	傳 6D Torus，鄰居 6→12	光模組、銅纜、OCS
Host CPU	x86，1:4	Axion，1:2；TPUDirect 繞過 host	傳 AMD CPU 同封裝（v10）	創意、京元電、致茂
網路內含值	基準	每顆 2–5 倍；DCN 100→400G	光 I/O 推向 3.2T	光聖、Lumentum、Celestica
光互連	可插拔＋OCS	可插拔＋OCS（1.6T）	XPO → NPO 最早 v10t	嘉基、嘉澤（推論）
散熱	Gen3 液冷，每盤 4 片冷板	Gen4、每盤 12 片、銅片＋鍍金	雙面冷板、可拆式上蓋	奇鋐、匯鑽科、健策
CDU	約 1.4MW	約 2MW	Project Deschutes（第五代）	台達電、高力（推論）
供電	VPD；PSU 5.5kW	VRM 內含 2.5 倍；PSU 10–12kW	±400VDC（v9）→ 800V	台達電、光寶科
機櫃功耗	約 90kW	50–90kW	約 200kW（v9）	電源櫃、液冷
主板 PCB	24L / M7	32–44L / M8＋M6（N+M）	HDI / M9 → M10	金像電、欣興、台光電

資料來源：Google Cloud blog（2026-04-22）；凱基（09-30）；Aletheia（10-06）；JPM（09-30）；Edgewater（06-18）；UBS（09-14）；SemiAnalysis（07-13）；永豐（09-16）；MS CCL（09-13）

Roadmap（一）：散熱——從冷板到雙面冷板、從 1.4MW 到 2MW CDU

功耗 1,000W → 約 10,000W；TPU 的散熱路線是「拿掉上蓋、冷板貼晶片、整盤液冷」

	v7 Ironwood（2025）	v8 8t / 8i（4Q26–2027）	v9 Humufish（2028）	v10 Icefish（2029）	狀態
晶片功耗	約 1,000W	8t <800W、8i 約 1,300W	約 2,400–4,500W	約 10,000W	券商估
機櫃功耗	約 90kW	50–90kW	約 200kW	—	UBS
液冷世代	Gen3	Gen4	—	—	已確認
CDU	約 1.4MW	約 2MW；一台帶約 8 櫃	第五代 Project Deschutes（2MW、500 GPM，已貢獻 OCP）	—	已發表
每盤冷板	4 片（只冷 TPU）	12 片（TPU＋供電＋光模組）	VPD 背面加冷板 → 雙面冷板（討論中）	—	券商估
導熱介面	—	取消上蓋、銦片 TIM＋冷板底座鍍金	TPU 下一代仍鍍金	—	公司說法
封裝結構件	補強環 15 美元	補強環 15–30 美元	可拆式上蓋＋補強環，內含值或 10 倍	—	未確認
每層冷板價值	約 3,000 美元	2,000–3,240 美元	約 3,500 美元	—	券商估
光模組散熱	氣冷	光模組上冷板（每盤 4 組）	1.6T / 3.2T 光模組液冷（OSFP / ELSFP）	—	JPM 預期

受惠：奇鋐（冷板）｜匯鑽科（鍍金）｜健策、一詮（補強環 / 上蓋）｜富世達（快接頭）｜時碩（古河管件）｜台達電（CDU）

Roadmap（二）：光通——可插拔＋OCS 走到 3.2T，CPO / NPO 排在最後

Google 的路線是「先把 OCS 與可插拔用到底」，與 NVIDIA 先推 CPO 不同

	v7 Ironwood（2025）	v8 8t / 8i（4Q26–2027）	v9 Humufish（2028）	v10 Icefish（2029）	狀態
scale-up 光模組	800G	1.6T（8i 每顆 1 個可插拔＋銅纜）	3.2T（GS：Google 2028 年導入）	—	券商估
每顆光模組數	1.8（全網路）	3.0（Aletheia）；scale-up 8t 2.0、8i 1.25（凱基）	6D Torus 傳聞：鄰居 6→12，光 I/O 3.2T	—	傳聞
OCS	ICI＋Jupiter	擴到 Jupiter 與 8i 機櫃間（SPOCS）；總台數 12 倍	延續；配 WDM 多波模組（2.4T）	—	已確認
scale-out	DCN 100G / 顆	Virgo：200G SerDes、每顆 400G；8t 每顆 800G	Virgo 延續到 Humufish、Icefish，頻寬再拉高	—	SemiAnalysis
光纖	8 條 / 顆	18.2 條 / 顆	—	—	Aletheia
XPO（可插拔 8 合 1）	—	未採用	可能隨 Tomahawk 7 導入	—	研究推估
NPO / CPO	—	未採用	「未來幾年看不到 NPO」	最早 v10 世代評估；GS 列 Google CPO scale-out 為「評估中」	研究推估

Celestica：已有 1 家 hyperscaler 的 Broadcom CPO 專案 2027 下半年量產（客戶未具名）。嘉基 / 嘉澤的 NPO 插座與 Google 的關係仍是推論。

受惠：光聖（光纖 / 配線）｜聯亞、華星光（雷射）｜Lumentum、Coherent（OCS、光模組）｜Celestica、Fabrinet（代工）

Roadmap（三）：傳輸——SerDes 翻倍、銅纜距離縮短、CPU 與儲存直通

電訊號速度每代翻倍，銅能跑的距離就減半；這決定哪些連線留銅、哪些改光

	v7 Ironwood（2025）	v8 8t / 8i（4Q26–2027）	v9 Humufish（2028）	v10 Icefish（2029）	狀態
ICI SerDes	112G（6 組 octal）	200 / 224G（6 組）	336G 特規（聯發科）	448G	部分確認
ICI 頻寬 / 顆	9.6 Tb/s	19.2 Tb/s	6D Torus 傳聞再翻倍	—	已確認 / 傳聞
銅纜距離	100G/lane	200G/lane 約 2.5 公尺	400G/lane 距離減半	—	產業訪查
櫃內銅連線	DAC	DAC / ACC；8i 每顆 2 條 1.6T ACC＋1 條 DAC	AEC（傳 Anthropic 要求；貿聯 2028 年顯著）	—	傳聞
Host 介面	PCIe Gen5 x16；x86	PCIe Gen5；Axion 1:2；TPUDirect RDMA / Storage 繞過 host	—	傳 CPU 與 TPU 同封裝（AMD）	已確認 / 傳聞
交換器	Jupiter 100G SerDes 乙太網	Virgo 200G SerDes 乙太網	Tomahawk 7（XPO）	—	研究推估
主板材料	M7 / HVLP2–3	M8＋M6 / HVLP4（N+M）	M9 / HDI	M9–M10	券商估

受惠：聯發科（SerDes IP）｜貿聯（AEC，弱）｜佳必琪（高速線，弱）｜信驊（BMC）｜台光電、金居、金像電（高速材料與 PCB）

資料來源：GFHK（08–24）；Google Hot Chips（2025）；Google Cloud blog（2026–04–22）；永豐金論壇（06–23）；Citi（09–29）；SemiAnalysis（07–13）；群益 ECOC（10–06）；貿聯私訪（09–23）；M...

目前各項缺口：從晶圓到雷射，幾乎每一層都在搶

缺口分兩種：卡住 TPU 出貨上限的（封裝、載板、HBM），和推高零件單價的（材料、雷射、設備）

環節	缺口數字	對 TPU 的影響	時間
N3 晶圓	JPM：2027 年 N3 供需惡化；台積電先進製程 2027 年漲 5–10%（MS）	v8 全在 N3	2027
CoWoS	Google 2027 年需 48 萬片（+115%）；Nomura：WoS 與小零件比 CoW 更緊	v8 產能分配	2026–27
EMIB-T	2028 年 40–45k 片 / 月，只夠 200–300 萬顆 v9；載板良率 20–60%	v9 出貨上限	2028
ABF 載板	MS 供需比 2027 年 102% → 2030 年 122%；福邦缺口 2027 年 26%、2028 年 46%	大尺寸、EMIB-T 吃產能	2027–30
HBM	MS：2027 年 DRAM 整體缺 17%；美光 2027 年 HBM 產能已售完	每顆 216–576GB	2027
T-glass	缺 19%（2026）→ 32%（2027）→ 51%（2028）	載板與 PCB 用料	2026–28
Low-Dk2 / Q-glass	Low-Dk2 缺 33%（2026H2–27）；Q-glass 6% → 25%	M8 / M9 CCL	2026–27
HVLP4 銅箔	MS：2027 年缺 31%、2028 年 20%（福邦 27% → 17%）	v8 全用 HVLP4	2026–28
矽電容	非台積電缺口 57%（2026）→ 16% → 11%（MS）	EMIB-T 內嵌	2026–28
雷射 / InP	EML、泵浦雷射缺口可能 >30%；InP 缺口 >70%、賣方市場到 2029+	光模組、OCS	2026–29
OCS	Lumentum 泰國產線滿載、交期排到 2027	TPU8 生命週期 OCS 總台數 12 倍	2026–27
鍍金冷板	客戶要 13,000–15,000 片 / 日，匯鑽科產能 5,500 片 / 日	v8 每顆 1 片	4Q26
PCB 設備	mSAP 設備交期 10–12 個月；鑽孔機最長 36 個月；M9 鑽針壽命只剩 100–200 孔	v9 HDI、交換器板	2026–28

資料來源：MS（06-02、07-06、09-13、10-05、05-21）；福邦 PCB（2026-10）；JPM（10-12）；Nomura（06-30）；玉山美光（10-01）；Lumentum 法說 memo（05-09）；Omdia（10-03）；匯鑽科（07-2...

缺口怎麼讀：卡量的看產能，漲價的看議價

卡 TPU 出貨上限的環節決定「量」，材料與零件缺口決定誰能「漲價」

卡住出貨上限（量）

- **EMIB-T 與 ABF 載板**：決定 2028 年 v9 是 150、250 還是 300 萬顆
- **CoWoS**：v8 的 2027 年量；台積電願不願意補 v9
- **HBM**：2027 年產能已賣完，TPU 與 NVIDIA 搶同一批 12-Hi
- **T-glass**：聯發科幫 Google 爭取日東紡配額，是 v8t 上修的理由之一

推高單價（價）

- **CCL**：2026 年至今漲 50–100%，後續每波 10–15%（Aletheia）
- **HVLP4、Low-Dk 玻纖**：合格供應商少、驗證 12–18 個月
- **雷射 / InP**：越上游越賺錢，基板毛利 >50%
- **VRM**：ASIC VRM 價格已漲至少 15%（Edgewater）

最快可能緩解的

- EMIB-T 載板良率已找到根因（廣發 10-13）
- HVLP4 2027 年供應商增加，福邦估缺口降到 17%
- 矽電容缺口 2027 年起收斂（57% → 16%）
- InP 6 吋 2028 年成熟後，光元件緩解（Omdia）

受惠於缺口的台廠：欣興、南電（載板）｜台光電、台燿（CCL）｜金居（HVLP4）｜台玻、富喬、南亞（玻纖）｜力積電、愛普（矽電容）｜聯亞（磊晶）｜匯鑽科（鍍金）

資料來源：本簡報整理；數字見前頁

單櫃 BOM 表（一）：晶片與板子

TPU v8 機櫃，16 托盤 / 64 顆口徑；單價或每櫃價值缺的只列用量

層	品項	每櫃用量	單價（美元）	每櫃價值（美元）	台廠
晶片	運算晶粒晶圓（台積電 N3）	64 顆	約 2,000 / 顆	約 12.8 萬（推算）	台積電
	設計服務認列（8t，含 I/O 晶粒與封裝整合）	64 顆	4,500	28.8 萬（推算）	聯發科
	HBM3E 12-Hi	8t 216GB / 8i 288GB × 64	—	—	（SK 海力士、三星、美光）
	oS 封裝外包	64 顆	150	9,600（推算）	日月光
	補強環	64 顆	15–30	960–1,920（推算）	健策、一詮
板子	主板 PCB（含 CCL、銅箔）	16 托盤＋CPU 板	—	37,861（CCL 13,009；銅箔 4,457）	金像電、欣興、台光電、金居
	Axion CPU	32 顆	約 700	22,400（推算）	創意
	VRM 半導體（VPD）	64 組	約 140 / 顆 TPU	8,960（推算）	—
	BMC	約 45 顆	約 25	約 1,120（推算）	信驊
	MLCC	約 60 萬顆	—	—	（村田、三星電機）

晶圓單價：MS 以 N3 晶圓 27,300 美元、v8i 每片 12 顆推算約 2,000 美元 / 顆；Nomura 台積電 2027F Google 營收 165 億 ÷ 830 萬顆亦約 2,000。「台積電晶圓」與「聯發科認列」可能部分重疊，不宜相加。

單櫃 BOM 表（二）：散熱、電源、網路與整櫃

網路類以 Aletheia 每顆 TPU 用量 × 64 換算；光模組取 1.6T 單價，屬上限

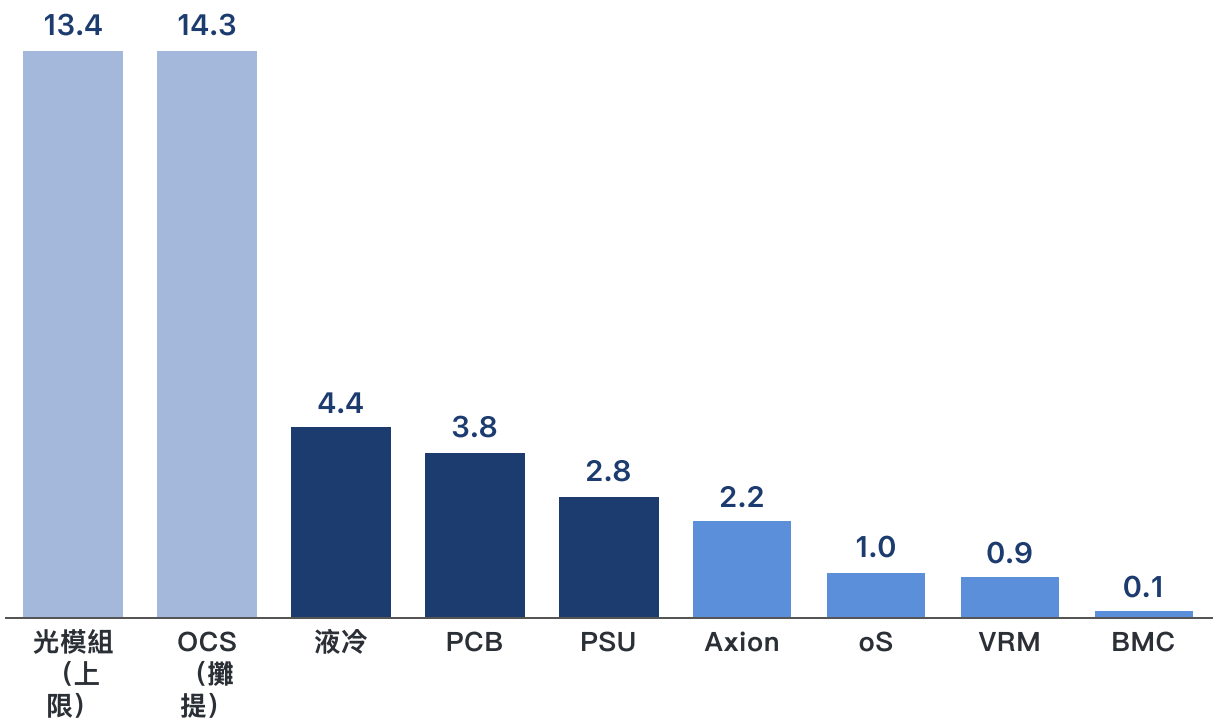
層	品項	每櫃用量	單價（美元）	每櫃價值（美元）	台廠
散熱	冷板＋快接頭＋歧管（不含 CDU）	192 片冷板、32 顆 QD	冷板 200–300	約 3.8 萬（JPM）～5 萬（凱基）	奇鋐、富世達
	其中鍍金	64 片	約 65	4,160（推算）	匯鑽科
	其中快接頭	32 顆 NSP1	約 56	約 1,800	富世達
電源	PSU / power shelf	18 顆 10kW	約 1,560	2.6–3 萬	台達電、光寶科
網路	光收發模組	約 192 顆（3.0 / 顆）	1.6T >700（2027）	≤13.4 萬（上限推算）	（Lumentum、InnoLight）
	光纖線纜 / 銅纜	約 1,165 條 / 約 122 條	—	—	光聖
	OCS（攤提）	約 1.1 台	5–20 萬	5.7–22.8 萬（推算）	（Lumentum）
整櫃	機櫃售價	—	—	v7 約 160 萬（SemiAnalysis）；v7x 180 萬、8i 210 萬（UBS）	Celestica、鴻海、英業達

OCS 為 Aletheia 生命週期總台數（25.9 萬）攤到每顆 TPU（約 0.018 台）× 64；光模組 1.6T 單價「2027 年 >700 美元」取自珩石 NPO 專題；800G 較低，所以 13.4 萬是上限。

BOM 價值拆分：晶片與 HBM 約占整櫃七到九成

非晶片零組件約 15–16 萬美元 / 櫃、網路最多約 36 萬；整櫃 160–210 萬美元，剩下的主要是晶片與 HBM

非晶片零組件（每櫃，萬美元，取區間中值）



用整櫃售價反推

項目	每櫃（萬美元）
整櫃售價	160–210
非晶片零組件（不含網路）	約 15–16
網路（OCS 攤提＋光模組上限）	約 6–36
剩餘：晶片＋HBM＋組裝毛利	**約 108–189**
換算每顆 TPU	**約 1.7–3.0 萬美元**

對照：SemiAnalysis 推算 v7 約 2.5 萬美元 / 顆（40 萬顆約 100 億美元）；聯發科只認列其中 4,500，台積電晶圓約 2,000，其餘多為 HBM、Google 自有運算晶粒價值與整機毛利。

讀法：台股能分到的「非晶片」價值每櫃約 15–16 萬美元（PCB、液冷、電源、Axion、封測、BMC），網路另計；晶片與 HBM 的錢主要在台積電、聯發科與記憶體廠

財模（二）：換算到每顆 TPU，再乘上 2027 年 800 萬顆

800 萬顆取各家 735–910 萬的中段；網路類單價券商未揭露，只列用量

每顆 TPU 內含值與 2027 年 TAM（推算）

品項	每顆 TPU（美元）	2027 TAM（億美元）	對照券商
OCS	約 890–3,560	71–285	每顆約 0.018 台 × 5–20 萬美元
液冷（不含 CDU）	約 594	約 48	—
主板 PCB	約 592	約 47	—
PSU	約 453	約 36	—
Axion CPU	約 350	約 28	MS：創意 2027 年 Axion 約 28 億 ✓
oS 封裝外包	150	約 12	MS：日月光 TPU 16 億（含測試）
VRM 半導體	140	約 11	—
鍍金	65	約 5.2	匯鑽科：2027 年約 700 萬片
補強環	15–30	1.2–2.4	—
快接頭	約 28	約 2.2	—
BMC	約 17.5	約 1.4	MS：信驊 TPU 占營收 14.4%

網路類：只有用量（每顆 TPU，Aletheia）

品項	TPU7	TPU8
光收發模組	1.8 顆	3.0 顆
光纖線纜	8.0 條	18.2 條
DAC / ACC 銅纜	1.5 條	1.9 條
乙太網交換頻寬	0.4 Tb	2.0 Tb

TPU8 每顆網路硬體內含值是 TPU7 的 2–5 倍（Aletheia）；光模組、光纖單價未揭露，不另推 TAM。聯發科檢查：v8t 300 萬顆 × 4,500 = 135 億美元，與 MS 2027 年 135 億一致 ✓

OCS 推算：Aletheia 生命週期 258,542 台 ÷ TPU8 1,400–1,500 萬顆 ≈ 每顆 0.018 台，再乘 GS 單價 5–20 萬美元；範圍很寬，僅供量級參考。

財模（三）：各世代單價怎麼變——誰的內含值放大最多

v7 → v8 → v9 的單位價值變化；放大倍數最高的是設計服務、PCB / CCL 與網路

品項	v7 Ironwood	v8（8t / 8i）	v9 Humufish	放大
聯發科每顆認列	—（Broadcom）	4,500（8t）	13,000（EMIB-T） / 20,000（CoWoS）	約 3—4 倍
每櫃 PCB（MS）	19,474	37,861	87,906	v7→v9 4.5 倍
每櫃 CCL（MS）	6,550	13,009	32,288	v7→v9 4.9 倍
每櫃銅箔（MS）	1,783	4,457	4,657	v8 後持平
ABF 載板	80×85mm / 16L	90—100mm / 18—22L，ASP +40—50%	EMIB-T 大尺寸	—
每層冷板	約 3,000（4 片）	2,000—3,240（12 片）	約 3,500	片數 3 倍
補強環 / 上蓋	15	15—30	可拆式上蓋，或約 10 倍（未確認）	最高 10 倍
VRM 半導體 / 顆	50—60	140	—	2.5 倍
PSU / 櫃	約 2 萬	2.8—3 萬	±400V 電源櫃（每瓦內含 4—5 倍）	1.4—1.5 倍
網路硬體 / 顆	基準	2—5 倍	—	2—5 倍
CPU：TPU	1:4（x86）	1:2（Axion）	—	CPU 數 2 倍

資料來源：MS（08-24、09-13、09-27）；凱基（09-30）；Edgewater（06-18）；JPM（09-30）；散熱產業簡報（06-30）；Aletheia（10-04、10-06）

台股曝險總表（一）：晶片、封測、載板材料

強＝券商具名份額或營收占比；中＝點名供應但無份額；弱＝潛在、列名或轉述

代號	公司	環節	可追蹤數字（並列）	強度
2454	聯發科	I/O＋封裝整合	TPU 占營收 2027 38% / 2028 72%（MS）；4Q26 TPU 營收 >20 億美元【新】	強
2330	台積電	代工＋CoWoS	TPU 占其 AI 營收 2027F 約 25%；Google 占 CoWoS 約 27%（野村）	強
3711	日月光	oS 外包、終測	TPU 占 LEAP 14% → 18%（MS 10–08【新】）	強
2449	京元電	FT＋燒機	TPU 5.5% → 8.6%（BofA）；Google 合計 2027 10–15%（MS【新】）	強
6223	旺矽	探針卡	Broadcom v8 / v9 需求約 v7 的 3 倍；無份額	強（方向）
7769	鴻勁	分選機、SLT handler	Google TPU / CPU 所有 FT 專案；無份額	強
6515	穎崴	測試座	野村圖列名；MS：5 月營收 +120% 含 TPU	中
6683	雍智	探針卡（高低混針）	經聯發科承接 TPU 專案，2H26 小量、2027 放量（經濟日報 07–18）；野村圖列「KSMT?」；v9 探針卡候選（Aletheia）	中～弱
2360	致茂	SLT	Axion CPU SLT 確認；TPU 為列名	中
3443	創意	Axion 設計服務	Google CPU 2027E 占營收 67%（UBS）	強（CPU）
3037	欣興	ABF、v9 HDI	凱基：2027 TPU 營收 <10%；v9 PCB 主供之一【新】	中
4958	臻鼎	ABF 二供、HDI	TPU 平台轉 HDI（Daiwa 10–07【新】）	弱～中
2368	金像電	主板 PCB	GS <5% → 20–30%；凱基市占 20–25%；福邦 v8 備用	中（分歧）
2383	台光電	CCL	v8 CCL 市占約 70%、2027 TPU 營收 20–25%（凱基）	強
8358	金居	HVLP4 銅箔	2027 TPU 營收 5–10%（凱基）	中
1303	南亞	Low-Dk2、CCL	Google 預訂約每月 100 萬單位（通路）；v8i CCL 列名【新】	中
6770 / 6531	力積電 / 愛普	矽電容 IPD	S-SiCap 占力積電營收 3% / 5% / 9%（MS）	中

資料來源：本簡報整理；個股來源見各章與 digest；【新】＝10–01 後有 TPU 新資訊

台股曝險總表（二）：散熱、電源、網路、組裝、潛在

凱基估 2027 年 TPU 營收貢獻：台達電、台光電 20–25%；金像電 15–20%；奇鋐 10–15%

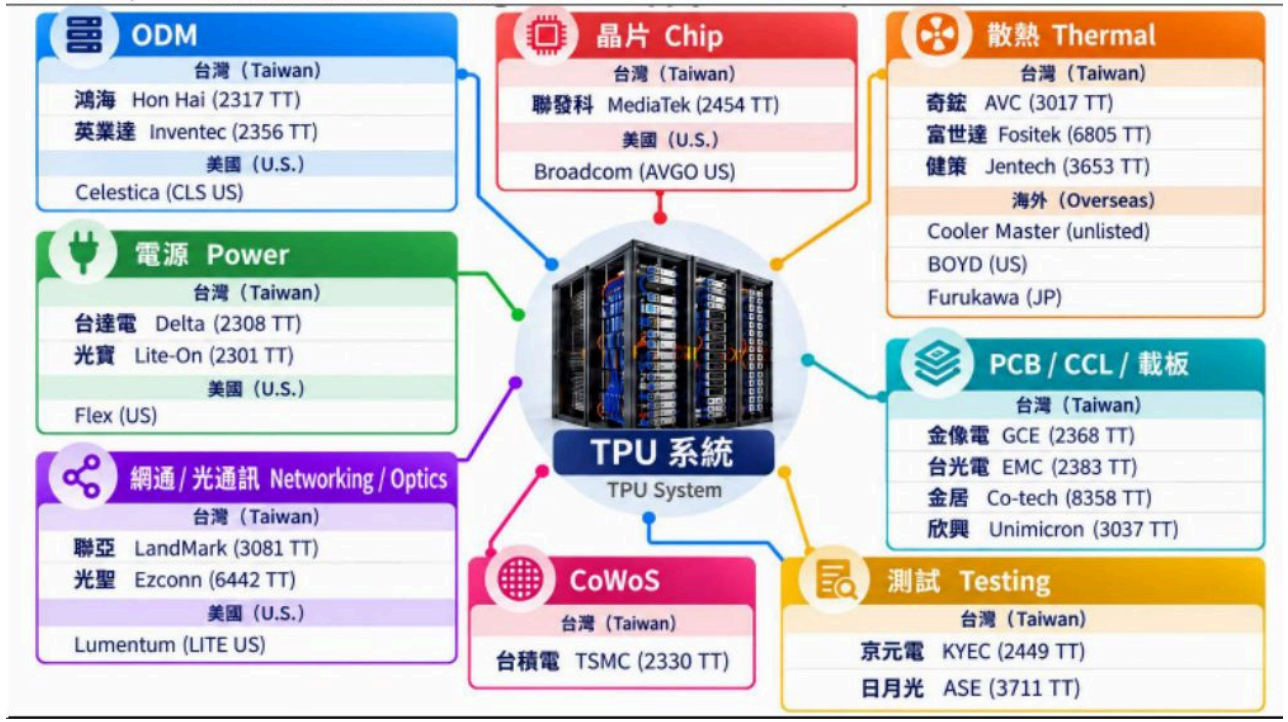
代號	公司	環節	可追蹤數字（並列）	強度
3017	奇鋐	冷板、歧管	份額 30% / 40% / >50%；2027 TPU 營收 10–15%（凱基）；4Q26 量產【新】	強
8431	匯鑽科	冷板鍍金	每片約 65 美元；TPU 鍍金冷板 2027 年約 700 萬片	中
6805	富世達	快接頭	4Q26 起出貨；TPU <10%（凱基）	中
3653	健策	補強環	CSP ASIC 補強環約 80%（JPM）；2027 TPU 營收 <10%（凱基）	中
2486	一詮	補強環	CSP ASIC 約 20%（JPM）；主要客戶含 Google TPU（中信）	中
4566	時碩	液冷管件（經古河）	古河 TPU 水冷三年大單，年約 20 億元【新】	中
2308	台達電	PSU、±400V	2027 TPU 營收 20–25%（凱基）	中
2301	光寶科	PSU	2026 年切入、<10%	弱～中
6442	光聖	光纖線纜	TPU8 每顆光纖 8 → 18.2 條，Aletheia 三大受惠【新】；凱基：雙位數貢獻	中
3081	聯亞	雷射磊晶	凱基：雙位數貢獻	弱～中
5274	信驊	BMC	TPU 占營收 4.1% → 14.4% → 9.4%（MS）	強（單一來源）
2317	鴻海	L10.5 / L11	v8 增加 TPU 層與 L11 份額；<10%	中
2356	英業達	L6、交換器	TPU L6 為主；<10%	中
2059	川湖	滑軌	MS、野村供應鏈圖列名	中～弱
8996	高力	CDU、RDHx 液冷	液冷占營收約 60%，客戶為美系 hyperscaler；庫內無來源指名 Google TPU	弱（推論）
3219	倚強科	主板功能測試	受邀參與 TPU 專案，無金額	弱
3665	貿聯	AEC	自稱一供代工、2028 年才顯著；中信稱難切入	弱
6715 / 3533	嘉基 / 嘉澤	NPO / CPC 插座	庫內無來源指向 Google；CTBC：AMD 平台原型測試、2H27 小量	弱（推論）

資料來源：本簡報整理；凱基（2026-09-30）等

凱基估 2027 年 TPU 營收貢獻：一張圖看台廠分工

Celestica 25–30% 最高；台股以台達電、台光電、金像電、奇鋐占比最高

圖 23：Google TPU 供應鏈關鍵零組件與台廠布局



Google TPU 供應鏈關鍵零組件與台廠布局：ODM、晶片、散熱、電源、PCB / CCL / 載板、網通光通訊、CoWoS、測試（凱基 2026-09-30）

2027 年 TPU 營收貢獻（凱基）

區間	公司
40–45%	聯發科約 40%、Broadcom 約 45%
20–30%	Celestica 25–30%、台達電 20–25%、台光電 20–25%
10–20%	金像電 15–20%、奇鋐 10–15%
5–10%	金居
<10%	鴻海、英業達、健策、富世達、欣興、光寶
雙位數以上	光聖、聯亞、Lumentum

資料來源：凱基〈電子硬體產業：Google TPU v8〉圖 23（2026-09-30）

市場說法（一）：TPU 對 GPU，便宜多少？

多方看每美元效能，空方看 FP4、disagg 與下一代 Rubin

−44%

Ironwood 全 3D Torus 配置 TCO vs GB200（Google 自身成本）

−30% / −41%

Google 加價外租後，每小時 TCO vs GB200 / GB300

US$0.181

InferenceX：Qwen3.5 FP8 每百萬 token；B200 0.222、B300 0.276

多方

- SemiAnalysis：TPU 約 15% MFU 就能和 GB300 在 30% MFU 打平，因 GPU 標稱 FLOPs 較灌水
- InferenceX：每美元效能最多領先 B200 50%、B300 96%
- OpenAI 尚未部署 TPU，就靠「威脅要買」換到約 30% 的 NVIDIA 折扣（SemiAnalysis）

空方 / 質疑

- TPUv7 沒有原生 FP4；disagg 下 GB300 中段延遲每美元約領先 30%（SemiAnalysis 自己的但書）
- 2026 底—2027 年 TPUv8 的 TCO 可能落後 Rubin（SemiAnalysis，經科技新報）
- 黃仁勳：TPU 成長「100% 是 Anthropic」，是個案不是趨勢
- NVIDIA：領先業界一整代，是唯一能跑所有模型的平台

資料來源：SemiAnalysis〈TPUv7〉（2025-11-28）、〈InferenceX〉（2026-09-07）；科技新報（2026-04-21）；黃仁勳訪談（經 Dave Friedman 2026-04-16）；NVIDIA X 聲明（經 T客邦 2025-11-29）

市場說法（二）：TPU 怎麼賣出去——從租雲到整櫃交付

Google 2Q26 首度把 TPU 系統交付客戶機房；外部客戶目前只有 Anthropic 有量化

客戶 / 管道	說法	狀態
Anthropic	官方：數 GW 次世代 TPU 2027 年起上線；WSJ：約 3.5 GW、Broadcom 合約到 2031，取決於 Anthropic 商業表現	已確認
Anthropic（v7）	SemiAnalysis：40 萬顆由 Broadcom 直賣整櫃（約 100 億美元）、60 萬顆經 GCP 租用；Fluidstack 建置、礦商 TeraWulf / Cipher 供機房	研究推估
Meta	已簽多年、數十億美元租 TPU；2027 年起直接採購放自家機房仍在洽談	報導
黑石合資 TPU 雲	黑石出資 50 億美元、首批 500MW、2027 年上線（「TPU 版 CoreWeave」）	已宣布
Fluidstack	估值 180 億美元；承建 Anthropic 約 500 億美元資料中心；Google 對部分租約提供擔保	報導
OpenAI	2025-06 否認大規模採用；2026 年無新證據	否認
SSI、xAI、Apple	SSI 為 Google 官方點名客戶；Apple 2024 年用 TPU 訓練 AFM；xAI 僅研究機構提及	片段

MS（07-23）：約 3.5 GW TPU 銷售集中 2027 年、另約 4 GW 在 2028 年，假設每瓦 20 美元、毛利約 20%；Google 存貨由 2025 年底 24 億升到 100 億美元，2027 年變現。MLQ 提醒：晶片與機房付款最終都仰賴 Anthropic 持續用量。

市場說法（三）：Broadcom vs 聯發科，份額終局三種看法

股價已對傳聞反應兩次：Broadcom 06–04 早盤跌逾 13%、聯發科 09–29 跌 7.1%

Broadcom 在 Google TPU 的占比

來源	2026	2027	2028
Macquarie（06–05）	約 95%	約 80%	約 65%
MS（07–14）	—	長期約 80%	—
台媒（經濟日報 06–04）	—	—	聯發科朝五成以上

Hock Tan 06–03：完全可以預見 Google 會有多元化的晶片供應來源。

聯發科的雜訊與澄清

- 09–29：傳客戶轉向有量產經驗的 Broadcom / Marvell，股價跌 7.1%；公司稱不實
- 09–30：美系外資稱市場過度放大 EMIB–T 載板瓶頸，維持 2027 / 28 年各 400 萬顆、營收 180 / 400 億美元
- 10–05 工商時報：v8 第 4 季量產、今年 45–50 萬顆、明年拚 300 萬顆，無遞延
- 10–13 廣發：v9 Google 端 Core A 重設計延約 3 週，tape–out 延到 2027–01、出貨或延到 2028Q1；v8t 穩健、v9 市占 100%
- 郭明錤：Google 採 semi–COT，連自己投片或由聯發科代投都向台積電詢價——「從好好先生變成錙銖必較的精算者」

v8 誰做 8t、誰做 8i，媒體說法一度互相矛盾（科技新報 / Wccftech 04–21 寫聯發科做 8i）；本簡報採 GFHK、MS、凱基等多家券商一致的「聯發科做 8t」。

市場說法（四）：封裝路線與 AMD 傳聞

EMIB-T 良率 90% 與 45–50% 是不同量測對象；AMD 參與 v10 只是傳聞

EMIB-T：各方說法

誰	說法
SemiAnalysis（07-01）	Humufish 改用 Intel EMIB-T，多年來首次；若良率或時程不如預期，不排除追加 CoWoS-L
郭明錤（05-04）	技術驗證良率 90%「正向也合理」；但從 90% 到 FCBGA 的 98%，難度可能更高
廣發（10-13）	EMIB-T 載板良率約 45–50%，已找到根因；v9 100% 用 EMIB-T
MS（07-16）	2nm TPU 量產仍須依賴台積電 CoWoS；EMIB 大規模應用尚未完全驗證
工商時報（10-05）	v9 規劃 CoWoS 與 EMIB-T 兩方案並行
MS（05-27）	利多三項：ABF 產能、EMIB 良率、矽電容與凸塊；點名欣興、力成、愛普（當日欣興、力成漲停）

AMD 與 v10

- SemiAnalysis 客戶報告：市場傳言 Google 與 AMD 就一項第 10 代 TPU 合作，可能把 CPU 核心與 TPU 同封裝、鎖定強化學習
- AMD 有 MI300A（CPU＋GPU 同封裝）、SoIC 與 CPU IP
- SemiAnalysis 自承可能有誤；AMD 直接負責整顆 v10 的可能性較低，較可能提供 IP 或封裝能力
- 廣發：v10 Broadcom 恐面臨 AMD（運算）與 Marvell（I/O）競爭

資料來源：SemiAnalysis（07-01，經鉅亨 07-03）；郭明錤（經科技新報 05-04）；廣發（10-13）；MS（經科技新報 07-16）；工商時報（05-27、10-05）；SemiAnalysis（經科技新報、INSIDE 08-17）

市場說法（五）：台股「TPU 概念股」被點名的順序

一年內媒體點名清單從晶片、PCB 擴到測試、載板、光通訊；純度與實際訂單差異大

時間	來源	點名
2025–11	工商時報	台積電、聯發科、台光電、金像電、創意、精成科、欣興、旺矽、精測、英業達、聯亞、光聖
2025–12	鉅亨（內資投顧）	台積電、聯發科、京元電、鴻勁、致茂、旺矽、穎崴、精測；稱聯發科「最被低估」
2026–04	經濟日報	光聖、聯亞（光通訊）、奇鋐（v7 起首度切入 Google 液冷）、台達電（電源主供）
2026–04	工商時報	京元電、精測、鴻勁因 FT / SLT 機台增購受惠
2026–05	工商時報（MS）	欣興、力成、愛普（EMIB–T 相關）
2026–08	經濟日報（MS）	v10 由聯發科協調、創意參與設計服務、世芯有機會；上調創意、世芯目標價
2026–10	MoneyDJ / 工商時報	TPU8 網通內含值 2—5 倍：光聖、Lumentum、Celestica；光聖 9 月營收年增 97%

質疑面

世芯仍是「潛在挑戰者」（工商時報 08–31）；設計服務由統包轉 COT 再降逾 2 成成本，毛利有壓力；聯發科是最純 TPU 股，也最怕傳聞。

讀法

媒體多半不分「已量產」與「送樣中」；本簡報以第 7 章曝險表的證據強度為準，不以點名次數排序。

資料來源：工商時報（2025–11–26、2026–04–27、05–27、08–31、10–09）；鉅亨（2025–11–25、12–01）；經濟日報（2026–04–20、08–26）；MoneyDJ（2026–10–09）

市場說法（六）：空方最在意的是 Google 的錢

需求強、backlog 5,140 億美元；但自由現金流首度轉負、靠舉債與增發撐 Capex

−59 億美元
2Q26 自由現金流，2004 年上市以來首次單季轉負

982 億美元
長期負債（2025 年底 465 億）；6 月增發淨額 496 億

3,750 億美元
MS 預估 2027 年 Capex（2028 年約 4,000 億）

空方論點

- 財報後 07–23 股價早盤跌約 7%；MS 質疑 GenAI 的 ROIC，下調 2027 / 28 年 EPS 7%
- 2Q26 淨利約 990 億美元是「其他收入」（Anthropic 估值、SpaceX IPO 帳面收益），核心 EPS 2.85 低於預期 2.89
- 向 SpaceX 租算力約每瓦 50 美元，3Q 起雲端利潤率承壓
- 廣發：2026 上半年大量 TPUv7 給外部客戶，Google 自身擴張落後
- Chanos：AI 基建「比網路泡沫更糟」（泛指 AI，不專指 TPU）

多方回應

- 雲端年增 82%，CEO：需求大於供給
- 存貨 100 億美元是 TPU 系統，2027 年變現
- 阮慕驊：疑慮成立但非泡沫，風險在回報時間差與融資，不是需求崩塌；未來 24 個月約 2,500 億 backlog 要兌現

資料來源：Alphabet 2Q26 財報（2026–07–22）；Tom's Hardware（2026–07）；經濟日報；MS（經鉅亨 07–23）；風傳媒阮慕驊（07–24）；鉅亨 Chanos（07–19）；廣發（10–13）

市場共識與分歧一覽

共識在方向，分歧在幅度、時程與口徑

議題	共識	分歧點
對 GPU 成本	FP8 推論下 TPUv7 每美元效能優於 B200 / B300	幅度看口徑；FP4、disagg 落後；TPUv8 對 Rubin 可能落後
外部客戶	Anthropic 是核心（2027 起數 GW）	其他客戶規模未證；Meta 購買仍在洽談；「100% Anthropic」
賣法	從租雲走向整櫃交付、擔保 neocloud、合資	擔保與融資結構是否等於表外風險
Broadcom vs 聯發科	Google 在多元化，Broadcom 自己承認	終局 80%（MS）vs 65%（Macquarie）vs 聯發科過半（台媒）
聯發科時程	v8 4Q26 量產，公司否認掉單	v9 是否延一季（廣發）vs 無遞延（工商時報）
封裝	v9 走 EMIB-T、CoWoS 為備案或並行	良率 90% vs 45–50%；大規模量產是否已驗證
CPU	TPU8 起 Axion 配比加倍	AMD 參與 v10 只是傳聞
台股	光通訊、CCL / PCB、散熱、測試、載板反覆被點名	純度與實際訂單差異大；聯發科情緒敏感
Google 財務	需求強、backlog 創高	現金流轉負、舉債增發、2027 Capex 3,750 億能否持續

資料來源：本簡報整理；各列出處見前六頁

要追的訊號與反證條件

時程硬訊號在 4Q26 v8 出貨與 v9 tape-out；分歧點在每櫃口徑、PCB 主供與封測分流

追蹤訊號

指標	為什麼重要
v8 4Q26 出貨（聯發科 4Q26 TPU 營收 >20 億美元）	京元電、奇鋐、台光電、金像電營收斜率
v9 tape-out 與 EMIB-T 載板良率	2028 年 v9 是 150、250 還是 300 萬顆
v8t 測試在京元電 / 日月光的分配	京元電 TPU 占比能否到 7–9%
Axion 出貨（2027 年約 400 萬顆）	創意、京元電 CPU 營收
Meta 是否直接採購 TPU	外部客戶是否不只 Anthropic
±400V 電源櫃導入時點	台達電、光寶科的 HVDC 機會

反證條件

- Google 2027 年 Capex 未如預期增加，或 Anthropic 部署延後
- EMIB-T 良率卡關且台積電不追加 CoWoS
- PCB 主供名單如福邦所列、金像電只是備用
- v10 採非可插拔 CPO，NPO / CPC 插座機會縮小

資料來源：本簡報整理