

AI 晶片散熱全景

散熱介面 (TIM) × 冷卻方式 × 關鍵元件 × 製程流程 × CSP / ASIC 採用模式

報告日期：2026-07-22 | 資料基礎：內部散熱產業簡報 (2026-06-30，彙整凱基 / 元大 / Morgan Stanley / JPM / 玉山)、SemiAnalysis (2026-05-20 Vera Rubin、2026-07-02 ECTC 2026)、富邦投顧 (2026-07-21)、Daiwa (2026-06-16)、Goldman Sachs (2026-07-13)、Indium Corporation 官網 (2026-07-22 查證) 等，經整理改寫。
本報告僅供產業研究參考，非投資建議。

本報告回答五個問題

1. 熱從晶片到冷卻塔怎麼走？每一層「介面」(TIM1 / TIM2 / TIM1.5) 是什麼、用什麼材料？
2. 目前主流的散熱方式是哪些？氣冷、水對氣、水對水、浸沒式各支援多大功耗？
3. 水冷板、Manifold、快接頭、CDU 這些關鍵元件怎麼做出來的？完整製程流程為何？
4. 未來往哪走？液態金屬 TIM、一片式 (one-piece) 蓋板 / 冷板、MCCP、MCL、晶片內微流道。
5. 每一家 CSP / ASIC (NVIDIA、AMD、Google、AWS、Meta、Microsoft、Groq) 各採用什麼散熱模式？

核心結論 (TL;DR)

- **功耗曲線碾壓一切**：NVIDIA 晶片 TDP 從 Ampere 400W 一路升到 Rubin 2,300W、Rubin Ultra 3,600W、Feynman 4,400W；機櫃功耗 GB200 198kW → VR200 444kW (100% 液冷、無風扇)。散熱從配角變成 AI 基建的先決條件。
- **介面材料同步升級**：TIM1 由散熱膏升級到銅片 / 石墨烯；TIM2 走向液態金屬 (銅)，連動 Lid 與水冷板「鍍金防蝕」新工序——材料每升一級，設備與表面處理供應鏈就多一層價值。惟 **VR 端鍍金水冷板已確定延後 (2026-07-22 供應鏈查核)**，NVIDIA 端時程需重估；TPU 端銅片 + 鍍金持續。
- **路線圖三段式**：現行 Direct-to-Chip 冷板 (~2,000W) → 微通道 (MCCP 2,000W+ 已量產 / MCL 3,000W+ 測試中) → 晶片內微流道 (5,000W+，TSMC 實測 4–5.3kW)。
- **ASIC 是第二成長曲線**：2027 年 ASIC 出貨為 2024 年 3 倍以上；Google TPU (銅片 TIM2 + 鍍金冷板)、AWS Trainium (雙面 Teton max 冷板) 散熱客製化程度高於 GPU，單 tray 散熱價值 TPU v8 (\$2,400) 反而是全場最高。

目錄

1. 為什麼散熱成了 AI 基建的核心瓶頸
2. 熱的路徑：從電晶體到冷卻塔（散熱疊構總覽）
3. 散熱介面 TIM 深度解析：TIM1 / TIM2 / TIM1.5、材料光譜、銅片兩種製程、液態金屬
4. 散熱方式全光譜：氣冷 → 水對氣 → 水對水 → 浸沒式 → 微通道
5. 關鍵元件與完整製程流程：水冷板、Manifold / QD、CDU
6. 未來發展方向：一片式設計、MCCP vs MCL、兩相液冷、晶片內微流道
7. 哪家 CSP / ASIC 採用什麼模式：完整對照表
8. 台股供應鏈映射與投資判讀
9. 風險、驗證清單與資料來源

1. 為什麼散熱成了 AI 基建的核心瓶頸

一顆 AI 晶片本質上是一個面積約 50—110 mm 見方、卻要吞吐上千瓦電力的「電熱爐」。熱設計功耗（TDP，Thermal Design Power）代表散熱系統必須持續帶走的熱量：家用吹風機約 1,200W——也就是說，一顆 Rubin GPU（2,300W）發的熱接近兩支吹風機全速運轉，而且集中在比手掌還小的封裝上，熱通量（單位面積熱流）達到每平方公分上百瓦的等級。傳統氣冷靠風扇吹鰭片，物理極限約在 1,000W；超過之後只能改用比熱與導熱能力遠高於空氣的液體。

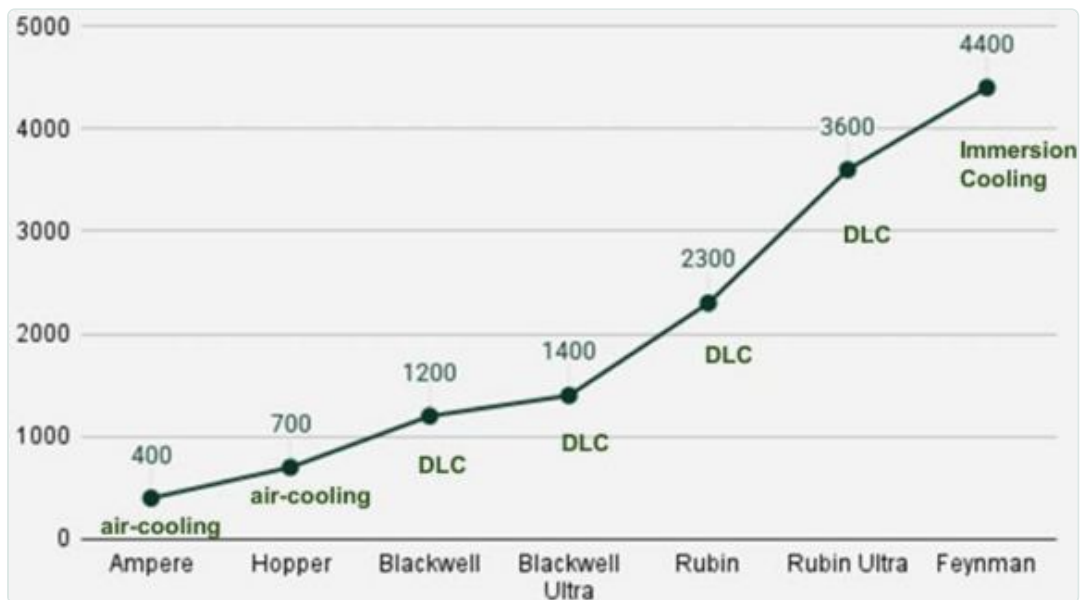


圖 1 | NVIDIA 歷代晶片 TDP 與對應散熱方式。Ampere 400W / Hopper 700W 仍可氣冷（air-cooling）；Blackwell 1,200W 起全面進入 DLC（Direct Liquid Cooling 直接液冷）；Rubin Ultra 3,600W、Feynman 4,400W 已標注浸沒式冷卻（Immersion Cooling）等更激進方案。來源：內部散熱產業簡報（2026-06-30，玉山投顧彙整）。

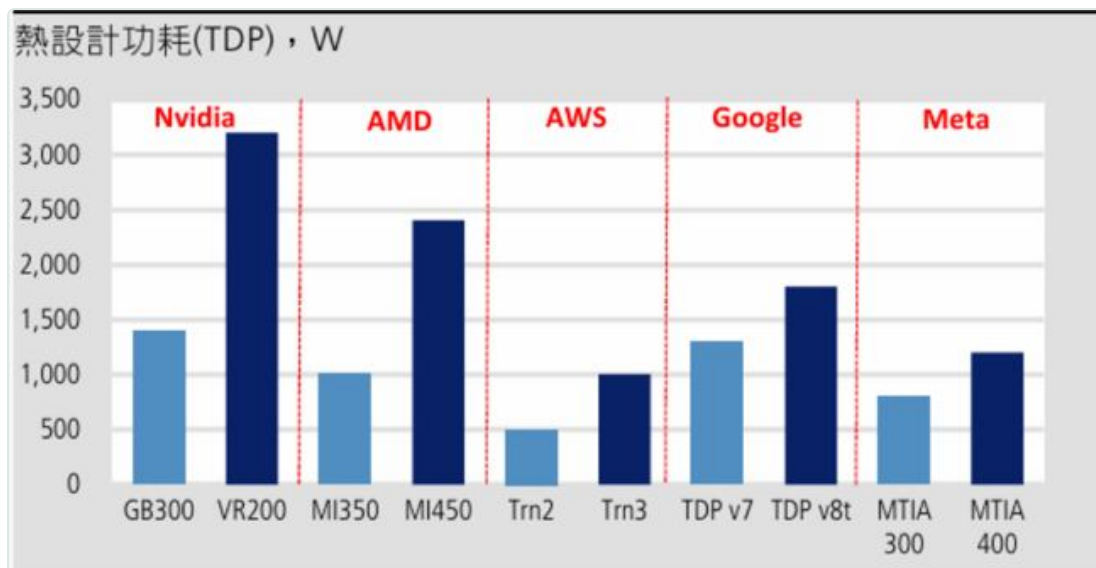
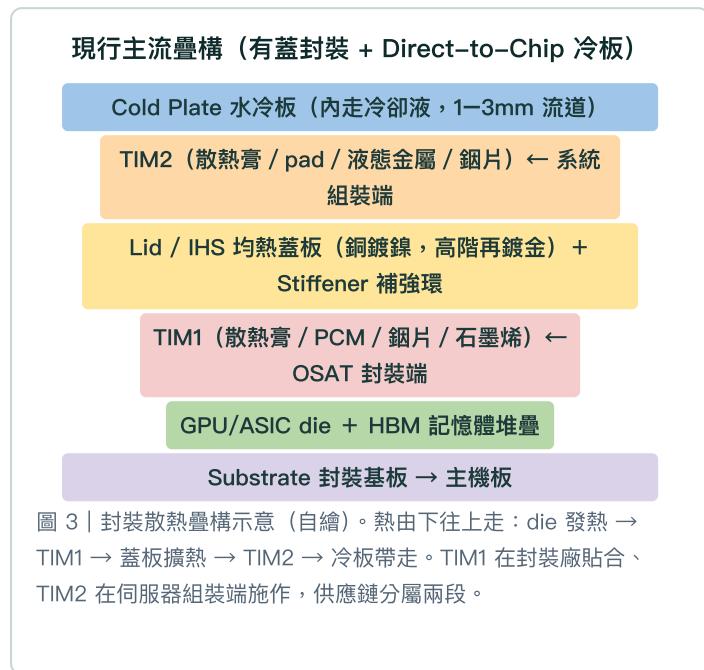


圖 2 | 五大廠 AI 晶片 TDP 對照 (W)。NVIDIA GB300 約 1,400W、VR200 (Rubin) 約 3,200W 為全場最高；AMD MI450 約 2,400W；AWS Trainium 3 約 1,000W；Google TPU v7 約 1,300W、TPU v8t 約 1,800W；Meta MTIA 300 / 400 約 800 / 1,200W。ASIC 功耗普遍低於 GPU，但迭代速度快、客製化散熱設計更多。來源：內部散熱產業簡報（2026-06-30，凱基彙整）。

晶片功耗只是第一層。機櫃層級的變化更劇烈：GB200 NVL72 整櫃約 198kW、液冷占 80%；VR200 NVL144 整櫃 444kW (+124%)、**100% 液冷、全機無風扇**，連網卡 (CX9)、配電板、收發模組、交換機都納入液冷回路。散熱不再是「加個模組」，而是整個機櫃、機房、電力與用水規劃的前提——Goldman Sachs 估計 2026—2035 年全球 56% 的新建資料中心產能位於高溫、高濕或缺水風險地區，Direct-to-Chip 液冷與浸沒式是唯一能同時壓低 PUE（電力使用效率，可達 ~1.2—1.4）與 WUE（用水效率，近零）的白空間方案，並預估 2027 年 AI 訓練伺服器的 Direct-to-Chip 液冷滲透率達 40%。

2. 熱的路徑：從電晶體到冷卻塔

理解散熱產業的關鍵是把「熱的旅程」拆成兩段：**封裝內**（die 到蓋板，OSAT / 封測廠負責）與**系統端**（蓋板到冷卻塔，散熱模組廠與機電廠負責）。每經過一個固體對固體的交界面，就需要一層 TIM（Thermal Interface Material，導熱介面材料）填補微觀空隙、排除空氣（空氣導熱係數僅 $0.026 \text{ W/m}\cdot\text{K}$ ，是最糟的熱阻來源）。



三個常用名詞：

- **TIM1**：die 與蓋板（Lid / IHS）之間，封裝端製程，講究厚度均勻性（BLT，Bond Line Thickness，越薄熱阻越低）、壓合力與翹曲控制。
- **TIM2**：蓋板與水冷板 / 散熱器之間，系統端施作，講究可重工性（維修可拆）、抗 pump-out（熱循環把材料擠出）與腐蝕相容性。
- **TIM1.5**：把蓋板拆掉或做成可拆式後，die（或可拆蓋）與冷板之間只剩一層介面，介於傳統 TIM1 / TIM2 之間，故稱 1.5——這是「Removable Lid / 無蓋化」路線的核心賣點：少一層介面、少一份熱阻。

系統端之後的路徑：冷板 → 機內分歧管（Inner Manifold）→ 機櫃分歧管（Rack Manifold，含快接頭 QD 與漏液偵測）→ CDU（冷卻液分配單元，液對液熱交換）→ 設施水 → 冷卻塔排到大氣。



圖 4 | 熱路徑全鏈（自繪）。封裝內每層介面吃掉的溫差稱為「熱阻預算」；功耗越高，每層介面能容忍的熱阻越小，材料與製程就被迫升級。

3. 散熱介面（TIM）深度解析

3.1 材料光譜：從矽膠膏到液態金屬

TIM 的演進邏輯很單純：功耗越高 → 熱阻預算越緊 → 從「有機材料」往「金屬」升級。代價是金屬類 TIM 帶來新的工程問題——氧化、腐蝕、電性短路風險與更嚴苛的貼合製程。

材料	導熱係數（約）	優點	限制 / 風險	常見位置
散熱膏 / gel	4–10 W/m·K	成熟、便宜、可填補粗糙面	pump-out（熱循環擠出）、老化乾裂	TIM1 / TIM2 主流
導熱墊片 pad / gap filler	1–10 W/m·K	容忍公差、組裝簡單	熱阻偏高	TIM2、板級（VRM 等）
相變材料 PCM	3–8 W/m·K	到相變溫度後貼合變好、量產一致性佳	熱循環可靠度、滲出	TIM1 / TIM2
石墨 / 石墨烯片	面內極高、垂直方向受限	輕薄、面內擴熱佳	垂直導熱與接觸面需設計	TIM1 輔助（GB300 均熱片維持石墨烯封裝）
銦片（Indium foil）	~82 W/m·K	軟金屬可塑性佳、熱阻業界最低段	成本高、氧化與壓合條件嚴、需鍍層配套	高階 TIM1 / TIM2（Google TPU v8 TIM2）
液態金屬（鎳基 / 銦系）	複合型 5.7–10（SPIL 實測）；純液金更高	高熱通量下表現最好	腐蝕銅（需鍍金阻隔）、滲漏、導電短路風險	高階 TIM2（NVIDIA Rubin 規劃液態金屬銦 TIM2）

導熱係數僅是材料本體性質；系統熱阻還取決於 BLT 厚度與兩側接觸熱阻——「數字高不等於系統熱阻低」，這是評估 TIM 供應商時最常見的誤區。

3.2 銦片的兩種使用模式：焊接式 sTIM vs 壓合式（2026-07-22 官網查證）

銦片不是單一製程，而是兩條完全不同的工程路線。判讀設備與鍍層供應鏈前，必須先分清楚是哪一種：

項目	焊接式 sTIM（solder TIM）	壓合式（compressible TIM，如 Indium Corp Heat-Spring）
接合機制	迴焊（reflow）加熱到銦熔點 156.6°C 以上，熔融銦與金屬化層潤濕，形成冶金接合	常溫下靠鎖固壓力使軟銦塑性變形填縫；不熔融、不進迴焊爐
金屬化需求	必要：die 背面 BSM 金屬化（Ti/Ni/Au 類）+ 蓋板內側鍍鎳金，否則熔銦不潤濕	不需金屬化；burn-in 用款（HSK）覆薄鋁層防沾黏
助焊 / 氣氛	需助焊劑或真空迴焊，控制銦氧化與空洞（void）	不需
可重工性	焊死不可重工，開蓋易傷 die	可拆、可重複使用
熱阻	業界最低（廠商稱 lowest total thermal resistance）	低（優於膏類），與壓力相關
典型位置	高階 TIM1 / TIM1.5（Intel 高階 CPU sTIM 為經典案例）	burn-in / 測試座、bare-die 壓貼冷板；亦有 no-reflow 直接當 TIM1 的產品

關鍵澄清：焊接 vs 壓合是「同一介面二擇一」，但焊接路線裡也有壓合工序

對同一個介面而言，最終接合機制只能選一種（焊死或壓住）。但焊接路線在進迴焊爐之前，仍需要點膠（助焊劑）、鋼片精準置片、熱壓頭預壓合（讓鋼片與接觸面貼實）、AOI 檢測——貼合 / 壓合設備與迴焊爐是上下游關係，不是替代關係。台股 TIM1 貼合設備商（竄騰）做的正是這段「爐前工段」，其設備強調精度、壓力與氧化管理，與鋼焊接製程特徵一致。



圖 7 | 鋼片 sTIM 焊接路線完整製程流程（自繪，依 Indium Corporation 官網製程說明與設備商公開資訊整理）。純壓合路線則跳過 STEP 1 / 2 / 5，以鎖固機構的持續壓力取代冶金接合。

3.3 液態金屬 TIM2：性能天花板與腐蝕代價

液態金屬（鎳基合金或鋼系）是目前 TIM2 的性能天花板，但它會腐蝕銅——這正是「鍍金」成為散熱供應鏈新關鍵字的原因。SemiAnalysis 對 Vera Rubin NVL72 的拆解指出：Rubin 封裝的均熱蓋是雙片 lid 模組 + stiffener 補強，蓋板表面電鍍金，目的就是防止蓋板與冷板之間的液態金屬（鋼）TIM2 腐蝕；冷板與 GPU 接觸面同樣鍍金防止銅被腐蝕。Google TPU v8 則因 TIM2 改用鋼片，水冷板底座需「鍍鎳 + 鍍金」阻隔層，防止銅—鋼生成脆化的介金屬化合物（IMC）。

可靠度是這條路線的最大變數：液態金屬有滲漏與導電短路風險、自動化組裝難度高。Morgan Stanley 原即提示 VR200 的金鍍層設計尚未最終確認、有下修可能；2026-07-22 供應鏈查核確認：VR 世代的鍍金水冷板已確定延後——防液金腐蝕的鍍金配套時程後移，暗示液態金屬 TIM2 在 VR200 的導入時程或材料選擇同步生變（改回傳統 TIM 或後移至更高階平台，待確認）。注意這是 VR（NVIDIA）端的變化；Google TPU 端的鋼片 TIM2 + 鍍鎳金冷板規格不受影響，兩條鍍金邏輯需分開追蹤。

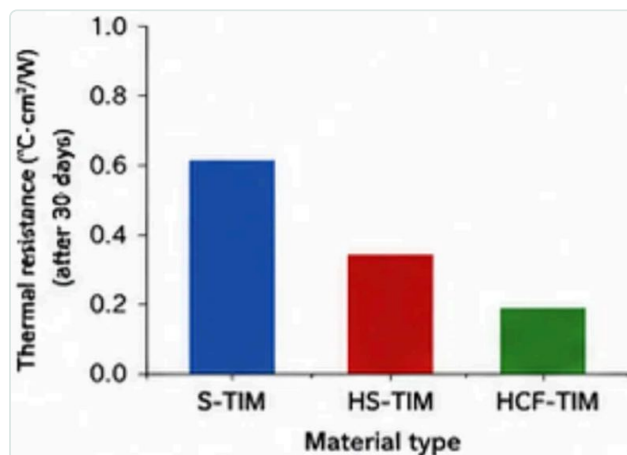


圖 8 | SPIL（日月光旗下）三種 TIM 30 天後熱阻實測（ECTC 2026）。S-TIM（一般矽膠）0.61、HS-TIM（鎳基液金 + 矽膠，5.7 W/m·K）0.34、HCF-TIM（液金 + 碳纖維，10 W/m·K）0.19 °C·cm²/W。可靠度分化明顯：150°C 烘烤 1,000 小時後 HCF-TIM 維持 95% 覆蓋率、HS-TIM 掉到 75%——液金複合材料的「基質」決定壽命。來源：SPIL, ECTC 2026 © SemiAnalysis（2026-07-02）。

4. 散熱方式全光譜：目前的主流與分級

把所有散熱方式放在同一把「TDP 尺」上，就能看懂技術替代的順序。目前的主流是水對水 Direct-to-Chip (DtC) 冷板——NVIDIA GB200/GB300 世代的標準配置；水對氣 (RDHx 後門熱交換器) 是既有機房改建的過渡方案；浸沒式與微通道則是 2,000W 以上的下一棒。

方案	可支援 TDP	採用時間	原理與定位	終端應用
一般氣冷	500—1,000W	已成熟	風扇 + 熱管 / 均熱片 + 鰭片；成本最低、噪音大	傳統伺服器
水對氣 (RDHx)	800—1,200W	過渡期	機櫃後門裝風水熱交換器，吸收排出熱風；不需大改伺服器	既有資料中心改建
水對水 (DtC 冷板) = 現行主流	1,200—1,500W	2025—2030	冷板直貼晶片、二次側水路經 CDU 與設施水交換；可移除機櫃風扇	DGX H100 / GB200 NVL72
單相浸沒式	1,500—1,800W	2027 以後	整機浸入介電液，泵浦強制循環	AI 訓練叢集、高密度邊緣 DC
兩相浸沒式	1,800—2,000W	2028 以後	低沸點氟化液沸騰汽化吸熱、槽頂冷凝；受 PFAS 監管與 3M 停產衝擊	Hyperscale AI 超級叢集
微通道冷板 MCCP	2,000W+	2026 已量產	冷板流道從 1—3mm 縮到微米級 (80—300μm)	Rubin Ultra 平台、各 ASIC
微通道冷卻蓋 MCL	3,000W+	2H26 開始採用	微流道做進封裝蓋板本身，Lid 與冷板合一	Rubin Ultra → Feynman 世代

來源：內部散熱產業簡報（2026-06-30，玉山投顧彙整）。浸沒式補充：3M 已於 2025 年底停產 Novec/Fluorinert 氟化液，歐盟 REACH 預計 2027 年立法限制 PFAS，兩相浸沒的冷卻液斷鏈風險推動業界轉向單相或冷板路線。

驅動液冷滲透率的除了晶片功耗，還有宏觀的「電力一用水」權衡：液冷出水溫度可達 45—60°C（氣冷僅 25—35°C），廢熱可回收；NVIDIA 揭露 Rubin 世代在合宜氣候下可 100% 液冷、冷卻液溫度上看 45°C，近零用水且大減冰水機依賴（GS，2026-07-13）。

5. 關鍵元件與完整製程流程

5.1 水冷板（Cold Plate）——供應鏈價值最高的零件

水冷板直接貼附熱源，是液冷系統中離晶片最近、ASP 最高的零件（一般 200—400 美元），必須與晶片廠及伺服器平台共同開發。結構上分三部分：底板（Base Plate，貼晶片面）、內部流道（Flow Channel）、頂蓋（Top Cover），外加流體接頭與固定支架。

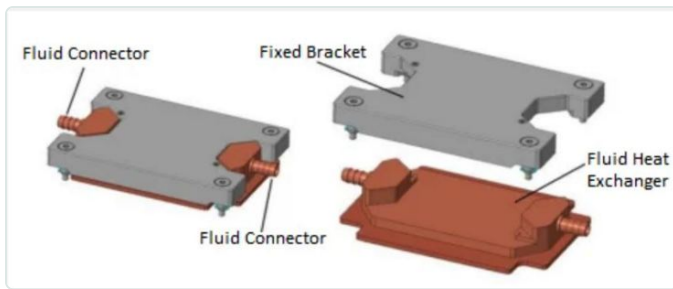


圖 9 | 水冷板外部構造拆解：Fluid Connector（流體接頭）、Fixed Bracket（固定支架）、Fluid Heat Exchanger（流體熱交換器本體）。來源：內部散熱產業簡報（FiberMall）。

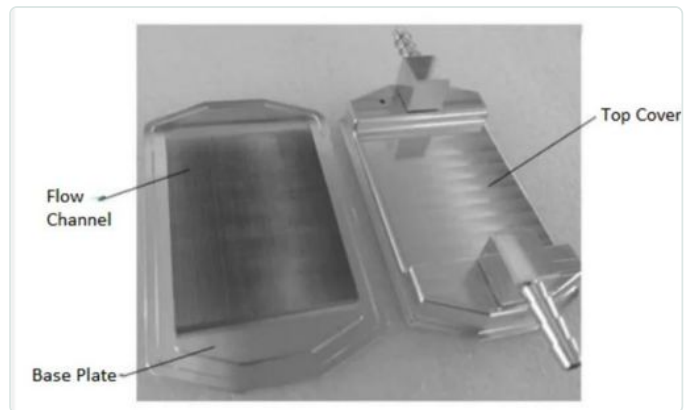


圖 10 | 水冷板實物零件照：左為底板上的 Flow Channel（流道，密齒狀鰭片），右為 Top Cover（頂蓋）。鰭片越密、接觸面積越大，換熱越好，但流阻與堵塞風險也上升。來源：同上。

完整製程流程——一片高階水冷板的生產可拆成五段：

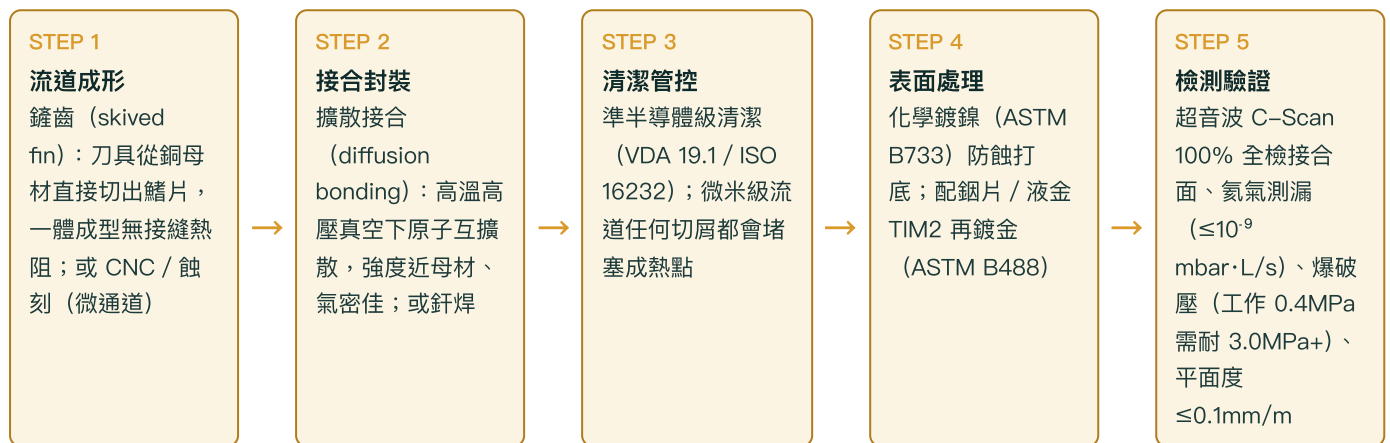


圖 11 | 水冷板完整製程流程（自繪，依內部散熱產業簡報與 Boyd/ASTM 公開技術文件整理）。平面度直接決定 TIM 厚度與熱阻；檢測環節（測漏、爆破壓）是液冷「不能漏」的品質底線。

5.2 Manifold 分歧管與 QD 快接頭——可維運性的關鍵

Manifold 負責把冷卻液分流到各節點冷板，分成機內（Inner）與機櫃（Rack）兩級；QD（Quick Disconnect 快接頭）讓維修人員可以把 tray 熱插拔而不噴液。設計重點：流量均勻、耐壓、低漏液、維修便利。NVIDIA 自訂更高規格的 NVQD，供應商須通過雙重驗證——這是快接頭供應商（富世達、雙鴻，外商 Stäubli、CPC）的門檻所在。VR200 世代 QD 用量從 GB300 的 360 組增至 450 組 / 櫃，並新增機內分歧管取代風扇（每層約 \$1,000）。

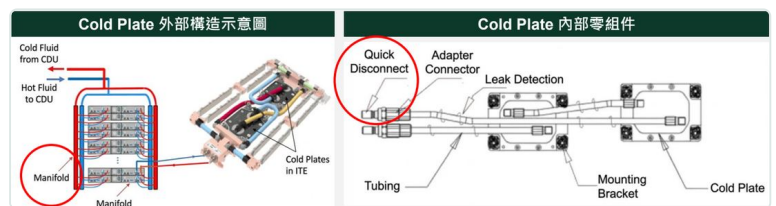


圖 12 | Manifold 與冷板迴路零組件分解。左：機櫃 Manifold 從 CDU 接冷熱水並分配到各層；右：Quick Disconnect、Adapter Connector、Leak Detection（漏液偵測線）、Tubing、Mounting Bracket 與 Cold Plate 的組合關係。來源：內部散熱產業簡報（FiberMall）。

5.3 CDU 冷卻液分配單元——液冷系統的中樞

CDU（Coolant Distribution Unit）整合泵浦、板式熱交換器（PHE）、感測、過濾與控制，負責隔離並橋接伺服器側冷卻液與設施水，決定機櫃功率密度上限與系統可靠度。規格正在快速跳級：高力 300kW in-rack CDU（支援 VR NVL72）、雙鴻 in-row CDU 達 2MW、台達電展出 3MW in-row 並規劃年底 6.8MW 版本；

GB200 世代的 Sidecar CDU（約 \$50,000 / 櫃）台達電市佔約 50%。兩相 CDU（高力 4U 5kW）開始露出，是下一階段觀察點。

6. 未來發展方向：封裝級散熱的四條路

6.1 路線圖總覽：把冷卻液一步步推向晶片

所有未來方案的邏輯是同一句話：縮短熱路徑、消滅介面。現行 DtC 要穿過 TIM1+Lid+TIM2 三層才到冷板；MCCP 強化冷板本身；MCL 把流道做進蓋板（消滅 TIM2）；晶片內微流道直接讓水在矽裡流（消滅全部介面）。



圖 13 | 五種方案的效率—熱阻定位（泡泡大小=成本）。X 軸熱阻 (°C/W)、Y 軸冷卻效率 (W/cm²)：Microfluid（晶片內微流道）效率最高、熱阻最低但成本最高；MCCP 是現在量產可行的甜蜜點；MLCP / 傳統水冷 / 氣冷依序退後。來源：內部散熱產業簡報（2026-06-30）。

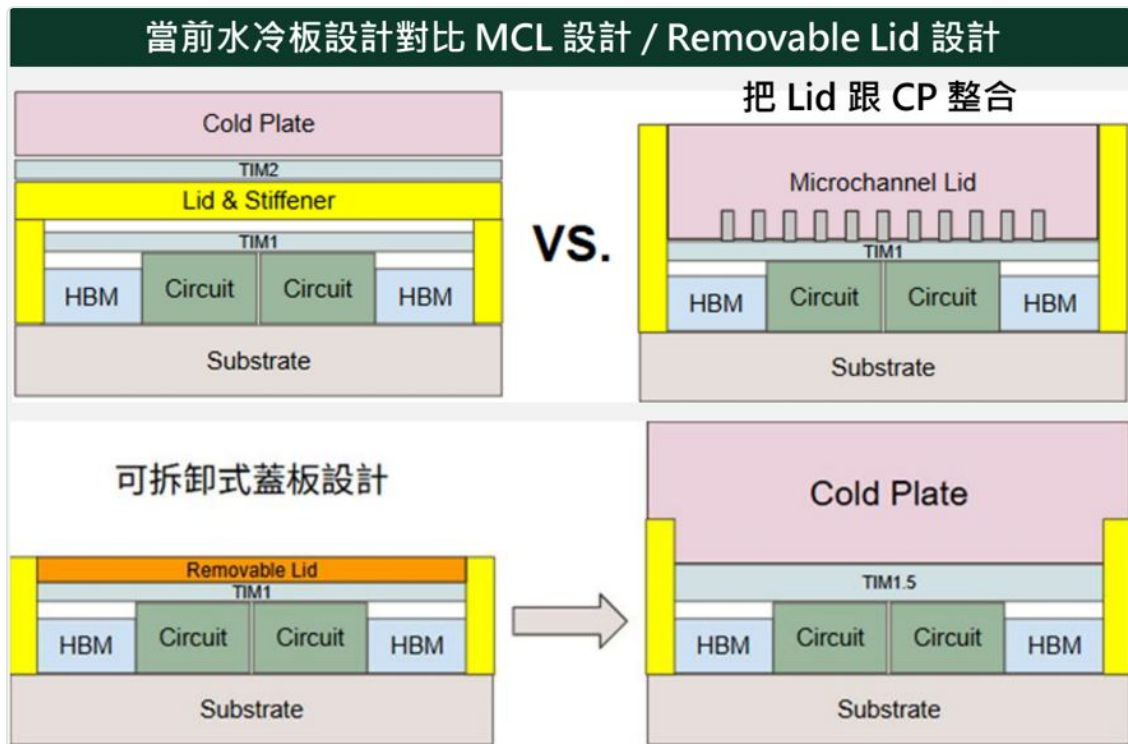


圖 14 | 三種疊構的正面對決。左上：現行設計（Cold Plate / TIM2 / Lid&Stiffener / TIM1 / die）；右上：MCL 設計——微流道直接做進蓋板，只剩 TIM1；下排：可拆卸式蓋板（Removable Lid）設計——SoC 與 Cold Plate 之間收斂成單層 TIM1.5。來源：內部散熱產業簡報（2026-06-30，健策方案說明）。

6.2 「一片式」的兩個層次：one-piece lid 與一片式冷板

- **一片式蓋板（one-piece lid）：**封裝端把多件式蓋板改為一體成型，施壓均勻、翹曲控制更好。NVIDIA 曾在 2026 年 5 月把 Rubin 設計改回 one-piece lid，對蓋板供應商（健策）形成急單與 ASP 上行；下一代仍可能走 removable dual lid 或 MCL——規格反覆本身就是蓋板廠的重定價機會。Rubin 量產版封裝為雙片 lid 模組 + stiffener + 表面鍍金（SemiAnalysis）。
- **一片式冷板（single-module cold plate）：**系統端把整個運算板（Strata board：2 顆 Rubin GPU + 1 顆 Vera CPU + SOCAMM 記憶體 + VRM 電源）用單一冷板模組整片覆蓋，取代一顆晶片一片冷板的做法——冷板從「零件」變成「板級模組」，組裝層級從 L10 前移到 L6，冷板廠更深地綁進系統製造。來源：SemiAnalysis VR NVL72 拆解（2026-05-20）。

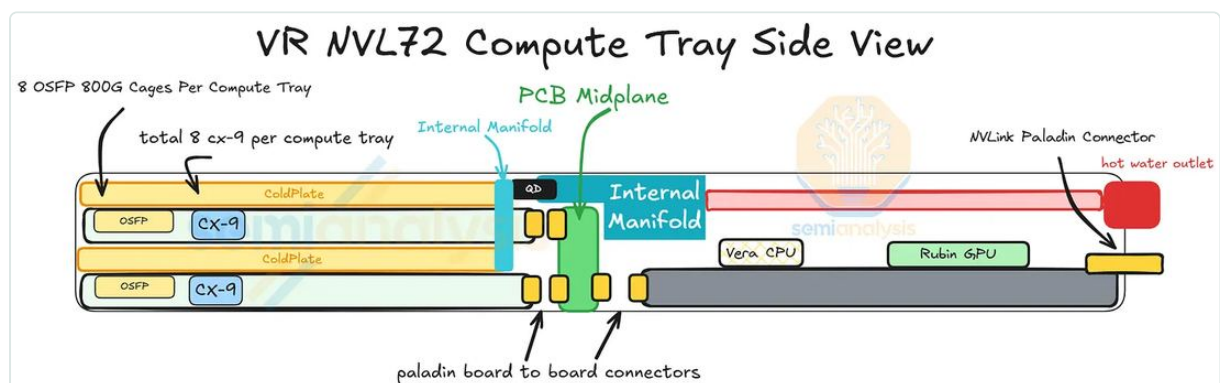


圖 15 | VR NVL72 Compute Tray 側視圖。上下兩片大型 ColdPlate 覆蓋整個運算板，中央 Internal Manifold（機內歧管）配送冷卻液，右側為 Rubin GPU + Vera CPU 位置；每 tray 8 組 CX-9 網卡同樣液冷。來源：SemiAnalysis（2026-05-20）。

VR NVL72 Coolant Flow Path

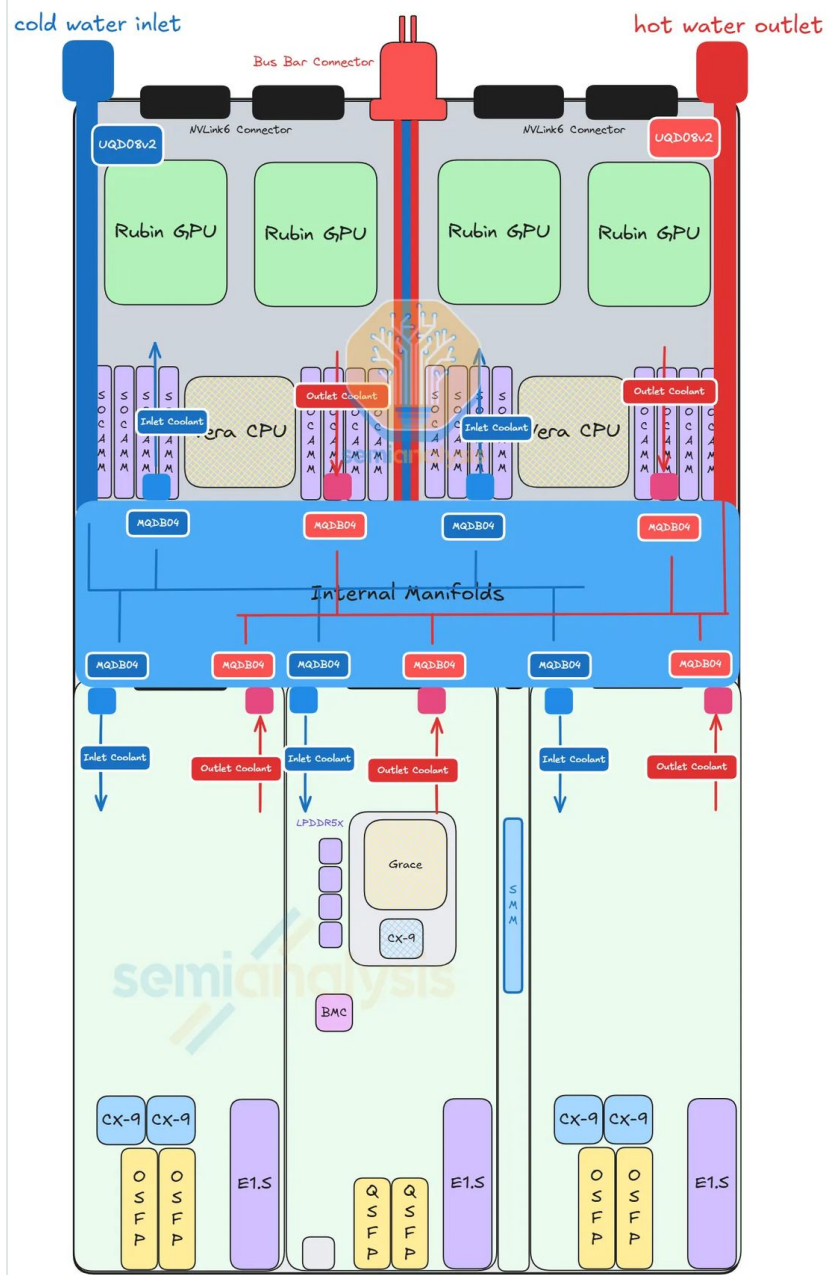


圖 16 | VR NVL72 tray 內冷卻液流路。冷水（藍）自 UQD08v2 快接頭進入，經 Internal Manifolds 以 MQDB04 接頭分配到 4 顆 Rubin GPU、2 顆 Vera CPU、SOCAMM 與下層 Grace / 網路板，熱水（紅）匯流排出。整個 tray 無風扇、100% 液冷。來源：SemiAnalysis (2026-05-20)。

6.3 MCCP vs MCL：微通道世代的路線之爭

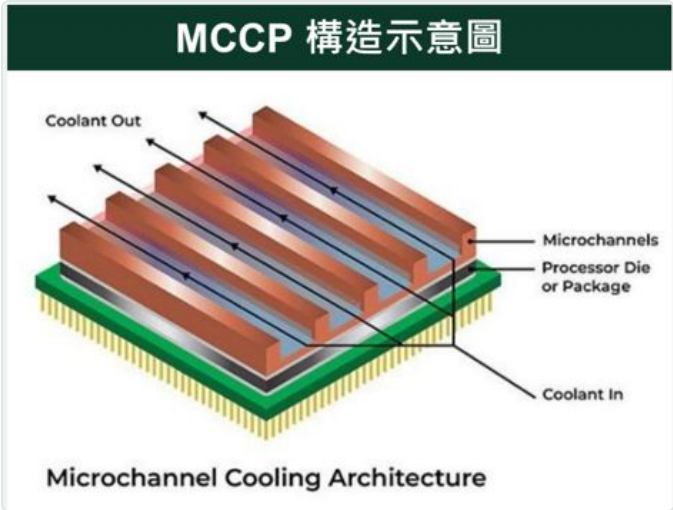


圖 17 | MCCP 微通道冷卻架構。冷卻液在微米級通道（80—300µm；VR 世代通道間距從 150µm 縮至 100µm）中流過 Processor Die 上方，接觸面積大增。仍屬系統端方案：疊構維持 TIM1+Lid+TIM2。來源：內部散熱產業簡報（2026-06-30）。



圖 18 | 健策 MCL 產品架構（Computex 2026 展示）。Explode view：Hose Plate / Bolster / MC Lid / Substrate / Back Plate；MCL Fin Base 厚度與間距皆 100–150µm。來源：富邦投顧（2026-07-21），引自健策展示。

比較	MCCP（奇鋌主推）	MCL（健策主導研發）
位置	冷板（系統端）	封裝蓋板本身
熱路徑	TIM1+Lid+TIM2 三層保留	只剩 TIM1（Lid 與冷板合一）
可支援 TDP	2,000W+	3,000W+
狀態	2026 已量產；AMD MI455 每 tray 配 5 組	單體測試已過關；2H26 起用（Rubin Ultra）、2027+ Feynman 世代主流預期
優缺點	沿用既有封裝與機構、良率穩；效果次於 MCL	熱性能最好、單價門檻最高；漏液＝整組報廢風險；奇鋌不看好此路線
目前瓶頸	清潔度管控、降本	測試設備開發、灌水測試（水壓/流速/水量）標準良品定義、系統層級搭配

補充：Kyber 直立式 blade（0.5U）場景下，MCL 採側面水路可把高度壓進 0.5U、解熱 3,000W+，空間表現最具優勢；同場競爭方案為兩相液冷 TPLC（奇鋌、Cooler Master 開發主軸）——迴路灌冷媒利用汽化潛熱吸熱，解決直立式單相水冷受重力導致流速不均的痛點。另注意 Morgan Stanley 判斷 Kyber 可能不會在 Rubin Ultra 世代大規模部署。來源：富邦投顧（2026-07-21）、MS（2026-06-07）。

6.4 終極路線：晶片內微流道（In-Chip Microfluidics）

把冷卻流道直接做在矽上，理論支援 5,000W+。兩大陣營的實測數據已經出爐（ECTC 2026）：

- TSMC micropillar 直達矽冷卻：在 CoWoS-R 平台上讓冷卻液直接流過 SoC die 背面的微柱（micropillar）。實測（40°C 去離子水）：傳統帶蓋冷板 1.9—2.3kW 就飽和（TIM 成瓶頸）、無蓋 2.5—3.0kW，micropillar 4kW@4LPM、5.3kW@8LPM，且全載具均勻散熱超過 5kW；測試載具通過 MSL4、無氬漏。此方案可完全移除 TIM1。
- Microsoft × Corintis：直線微通道直接蝕刻進 GPU 矽，用 NVIDIA GH200 實機驗證——GPU 熱阻降 51—60%、整體封裝熱阻 —50%；6 個月約 4,370 次觀測僅 9 次潛在堵塞、無可量測矽侵蝕。另以 AI 設計「仿

生葉脈型」流道，散熱效率達傳統冷板 3 倍、GPU 溫升最多降 65%。

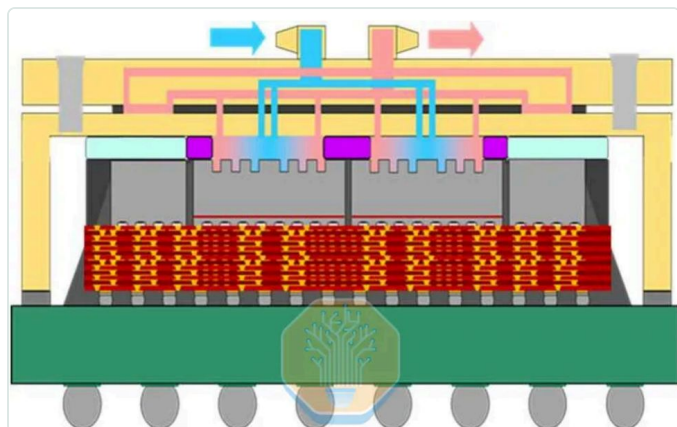


圖 19 | TSMC 直達矽冷卻封裝剖面。冷卻液（藍入紅出）流經 die 背面齒狀 micropillar，下方為 HBM 堆疊與基板。來源：TSMC, ECTC 2026 © SemiAnalysis。

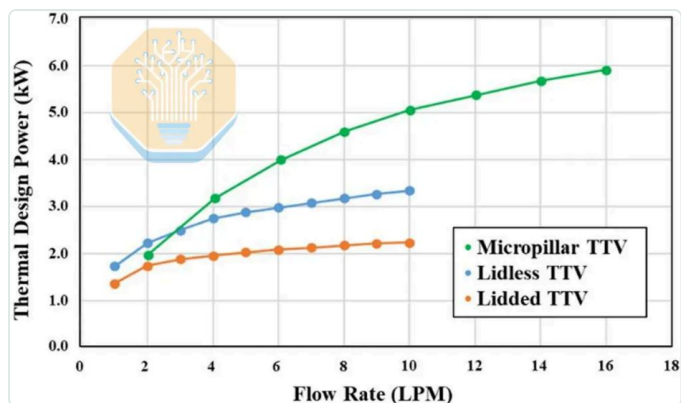


圖 20 | 三種封裝的散熱能力對流量曲線。Lidded（帶蓋）在 4LPM 後飽和於 ~2.2kW、Lidless ~3.3kW；Micropillar 持續上升至 ~5.9kW@16LPM——介面層就是天花板，拆掉介面天花板就消失。來源：TSMC, ECTC 2026 © SemiAnalysis。

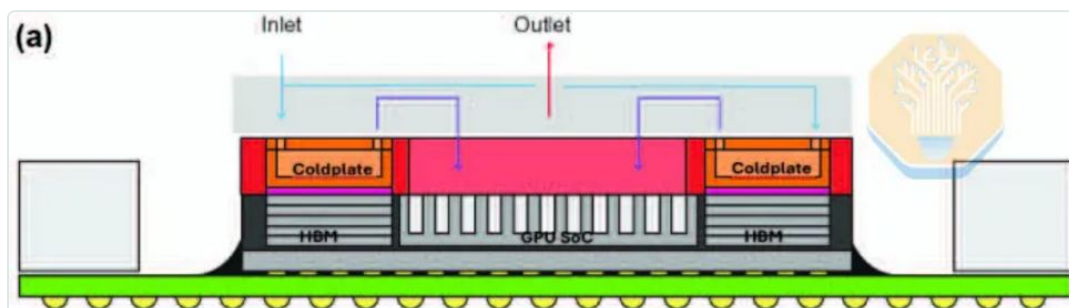


圖 21 | Microsoft GH200 實機微流道冷卻組裝剖面。中央 GPU SoC 走晶片內微通道（Inlet/Outlet 直通矽），兩側 HBM 仍用傳統 Coldplate + TIM——混合架構顯示過渡期的實務取捨。來源：Microsoft, ECTC 2026 © SemiAnalysis。

四條路線一句話總結

MCCP = 把冷板做細（2,000W+，已量產）；MCL = 把冷板做進蓋子（3,000W+，測試中）；Removable Lid / TIM1.5 = 把蓋子做成可拆（過渡方案）；In-Chip = 把水送進矽裡（5,000W+，2028 以後）。風險與價值都隨「離晶片的距離」縮短而放大。

7. 哪家 CSP / ASIC 採用什麼模式：完整對照

晶片 / 平台	TDP	封裝端 (Lid / TIM)	系統端 (冷板 / 機櫃)	放量時點與備註
NVIDIA GB300 NVL72	~1,400W	有蓋封裝；均熱片維持石墨烯封裝與沙漏形結構	水對水 DtC；135 片冷板+360 組 QD / 櫃；液冷占 85%；整櫃 <160kW	量產中；2Q26 起讓位給 VR200
NVIDIA VR200 NVL144 (Rubin)	2,300~3,200W 級	雙片 lid 模組+stiffener +表面鍍金；TIM2 原規劃液態金屬（鋼）——鍍金延後（右欄）後導入時程 / 材料待確認	100% 液冷無風扇（444kW / 櫃）；一片式冷板覆蓋整個 Strata 板；MCCP 通道 150→100μm；冷板鍍金已確定延後（2026-07-22 供應鏈查核）；234 片冷板+450 組 QD +機內分歧管；CX9 網卡 / 交換機全液冷	3Q26 液冷 readiness 量產目標；液冷價值 \$63,900 / 櫃（凱基）
NVIDIA Rubin Ultra	2,500—2,800W（簡報標 3,600W 級）	stiffener+雙鍍金 Lid（ASP 預測 \$80 / \$150）；MCL 或 MCCP 可能部分測試採用	奇鋌研究「鑄金水冷板」；供應商集中至 3—4 家	2027；MCL 2H26 起在此平台開始採用
NVIDIA Feynman	~4,400W	MCL 預期成為 3kW+ 世代主流（健策研發、奇鋌不看好）	浸沒式 / 兩相 TPLC 進入討論範圍	2027 以後
AMD MI450	~2,400W	stiffener+雙鍍金 Lid（同 Rubin Ultra 方案）	DtC 冷板；Helios 機櫃	4Q26 出貨（初期約 30 萬顆）
AMD MI455 (Helios)	—	—	每層 Compute Tray 配 5 組 MCCP+水冷分歧管——MCCP 最早的大規模採用者	2027 出貨上修至 150—160 萬顆；Helios 機櫃 3,200→11,000 櫃
Google TPU v7 (Ironwood)	~1,300W	Stiffener 防翹曲（ASP \$15）	每 tray 4 片 ASIC 冷板×16 層 = 64 片 / 櫃	量產中
Google TPU v8i/v8t (Sunfish/Zebrafish)	1,300—1,600W（v8t 至 ~1,800W）	Stiffener（\$15—30）；TIM2 改用鋼片——冷板底座需鍍鎳+鍍金防銅鋼脆化（IMC）	每 tray 4 大+8 小冷板、CPU 維持氣冷；單 tray 液冷零組件約 \$3,000、冷板價值 \$2,400（全平台最高）；冷板 ASP \$200—300	4Q26 機櫃放量；奇鋌市佔預估 >50%（凱基假設）
Google TPU v9 (Humufish)	—	Stiffener 延續	晶片+電源模組+Transceiver 一起納入散熱設計，單 tray 液冷價值假設 \$3,500	2H27 起
AWS Trainium 2	~500W	—	每 tray 4 片 ASIC 冷板；單 tray 冷板價值 \$1,500—2,000	量產中
AWS Trainium 3 / 4	Trn3 ~1,000—1,200W	—	氣冷+水冷雙軌；附加 Teton max 冷板：基板上加一片、雙面接觸（晶片+電源），結構複雜；單 tray 冷板價值 \$2,000—3,000	Trn3 機櫃 2Q26 未放量（領先 GB/VR 轉換）；水冷 4Q26 出貨、初期滲透率 10—20%；奇鋌市佔預估 >50%

晶片 / 平台	TDP	封裝端 (Lid / TIM)	系統端 (冷板 / 機櫃)	放量時點與備註
Meta MTIA 2/3/3.5	MTIA 300/400 約 800/1,200W	Stiffener (假設, \$15)	每 tray 4 顆 MTIA+1 CPU、5 片冷板；每櫃 8–10 tray	2026 起逐代放量
Microsoft Maia 200/300	—	Stiffener (假設)	DtC 冷板 (假設同 MTIA 配置)；另以 GH200 實機驗證晶片內微流道 (熱阻 –50%)，是 CSP 中最積極押注 in-chip 路線者	Maia 200 於 1H26
Groq LPU (LPX rack)	—	—	每櫃 256 片冷板+832 組 QD——冷板/QD 用量全場最高；液冷價值 \$80,640 / 櫃、QD 價值 \$88k / 櫃	Groq 3 LPX compute tray 3Q26 量產；每 8 座 VR NVL72 搭 5 座 LPX rack

來源：內部散熱產業簡報（2026-06-30，彙整凱基 / 元大 / MS / JPM）、SemiAnalysis（2026-05-20）、MS Computex 快報（2026-06-07）、SemiAnalysis ECTC 2026（2026-07-02）。「奇鈹市佔 >50%」等為券商彙整假設，待法說與供應鏈訪查驗證。

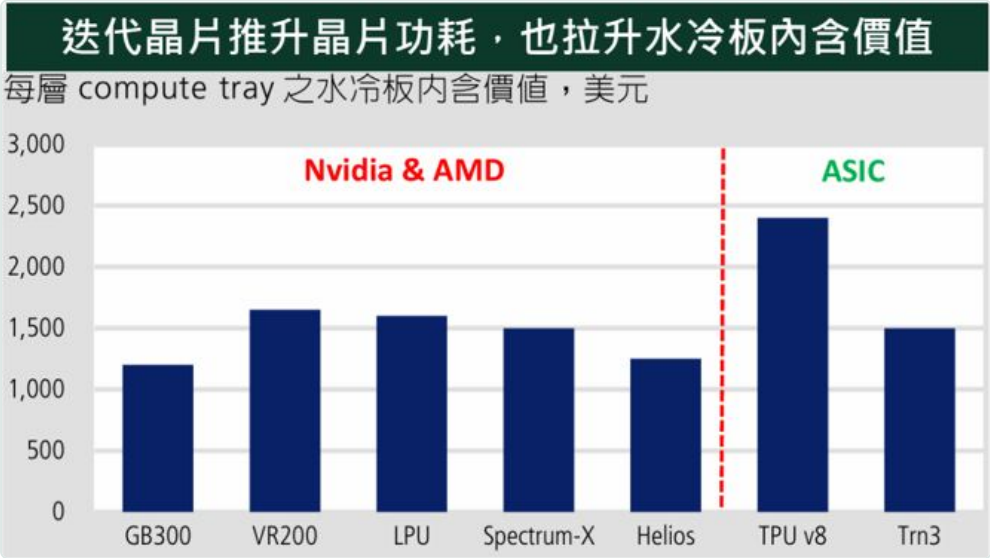


圖 22 | 每層 Compute Tray 水冷板內含價值 (美元)。GB300 \$1,200 → VR200 ~\$1,650；ASIC 陣營中 TPU v8 達 \$2,400 全場最高（銅片 TIM2+鍍金底座+4 大 8 小冷板配置堆高價值），Trn3 \$1,500。ASIC 的散熱客製化程度不輸 GPU。來源：內部散熱產業簡報（凱基）。



圖 23 | 均熱片內含價值 (每機架, 千美元)。GB300 66.0 → VR200 105.6；MI450 103.7 與 VR200 並列最高；Trainium 3 39.6、TPU v8 26.4。封裝端蓋板價值升級集中在 GPU 陣營。來源：凱基投顧。



圖 24 | 快接頭內含價值 (每機架, 千美元)。GB300 28.4 → VR200 54.9 (近翻倍)；ASIC 端 TPU v8 45.0、Trainium 3-liquid 40.2、LPU 88.0 全場最高。模組化維修架構讓 QD 量價齊升。來源：凱基投顧。

零組件廠商的營收放量時間將較組裝廠提早約一個月

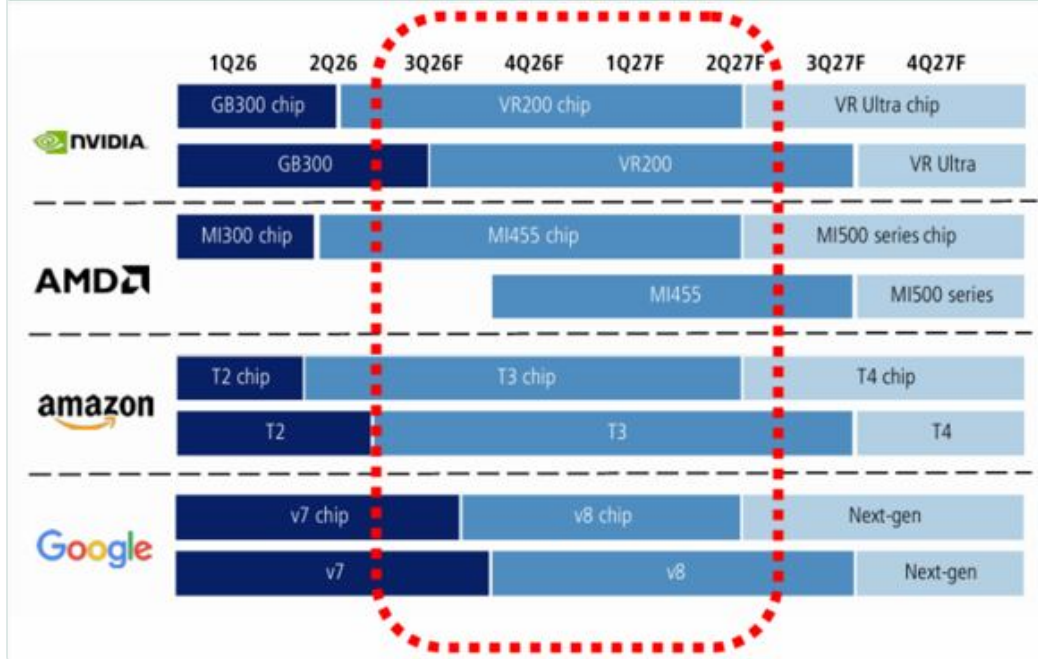


圖 25 | 各廠晶片與機櫃放量甘特圖（1Q26—4Q27F）。紅框標出 3Q26F—2Q27F 的密集放量窗口：VR200、MI455、Trainium 3、TPU v8 幾乎同時進入 ramp；零組件廠營收放量較組裝廠提早約一個月，是追蹤月營收的領先指標。來源：凱基投顧。

8. 台股供應鏈映射與投資判讀

Segment	Company	Ticker	GPU		ASIC	
			NVIDIA GB	AMD Helios	AWS Trainium / Google TPU	
Thermal	AVC	3017 TT				
	Auras	3324 TT				
	Fositek	6805 TT				
	Jentech	3653 TT				

圖 26 | 台廠散熱公司 × GPU / ASIC 客戶矩陣。奇鋌（AVC, 3017）全平台覆蓋（NVIDIA GB + AMD Helios + AWS Trainium + Google TPU）；健策（Jentech, 3653）同樣四平台；富世達（Fositek, 6805）NVIDIA + AWS + Google；雙鴻（Auras, 3324）NVIDIA + AWS。來源：凱基投顧。

環節	公司	角色與最新狀態
冷板 / Manifold / 整機散熱	奇鋌 (3017)	台廠液冷龍頭、NVIDIA First Design Partner；1Q26 營收 490 億（YoY +110%）、散熱占 63.6%；MCCP 主推者（擴散接合工法）；ASIC 冷板市佔預估 >50%（TPU v8/v9、Trn3/4）；研究鑄金水冷板應對 Rubin Ultra；冷板產能年底 5 倍
IHS / Lid / MCL (封裝端蓋板)	健策 (3653)	均熱片 / 蓋板純度最高標的；one-piece lid 設計變更急單受害者；Removable Lid（TIM1.5）與 MCL 主導研發（MCL 單體測試過關、Kyber 0.5U 側水路方案領先）；5 月營收 23.13 億創高（YoY +37.7%）
CDU / RDHx	高力 (8996)	1Q26 液冷占營收 60%；RDHx 主力 + 300kW in-rack CDU（支援 VR NVL72）+ 自製 PHE 板式熱交換器；4U 5kW 兩相 CDU 展出
冷板 / CDU / QD 多元模組	雙鴻 (3324)	液冷 total solution；in-row CDU 達 2MW；CPU/DIMM 雙冷模組；嘗試 TIM1+heat spreader 整包往封裝端滲透
QD 快接頭	富世達 (6805)	NVQD 雙重驗證門檻受害者；VR200+Trainium 3 專案 ramp，3Q26 營收預期雙位數 QoQ 成長（Daiwa）
CDU / 電源 + 散熱統包	台達電 (2308)	GB200 Sidecar CDU 市佔約 50%；3MW in-row CDU 展出、年底規劃 6.8MW；電源 + 散熱系統級整合
TIM1 貼合設備	竑騰 (7751)	Chip-TIM1-Lid 段點膠 / 置片 / 壓合 / AOI 設備（sTIM 焊接路線的爐前工段）；銅片設備 2H26 出貨年增雙位數；通用平台換治具可切換 Stiffener / 均熱片 / 銅片 / 石墨烯
表面處理 / 鍍金	匯鑽科 (8431)	TPU 冷板鍍金外包環節（經奇鋌 / 全億大，2026-09 量產）；電鍍 + 金料約佔冷板 ASP 中 ~US\$65；良率與執照壁壘高（中國對手良率僅 15-20%）
焊接材料	昇貿 (3305)	錫膏 / 預型片等焊接材料；與 TIM 材料需分開判讀，列觀察
UQD 新進者	新日興 (3376)	液冷 manifold + 通用快接頭 UQD 新切入；2H26 小量、2027 放量

投資判讀三句話

- 價值升級 > 出貨量成長：機櫃液冷價值 GB200 \$38,340 → VR200 \$63,900 → LPU \$80,640；均熱片、QD 內含價值逐代近翻倍——散熱族群營收成長會快於機櫃出貨成長。
- ASIC 是第二曲線：機架出貨 2026F→2027F：AWS 47→75 千櫃、Google 60→120 千櫃（vs NVIDIA 65→70）；2027 ASIC 出貨為 2024 的 3 倍以上，且 TPU 鋼片 + 鍍金、Trainium 雙面冷板的客製化讓單位價值更高。
- 材料每升一級，供應鏈多一層：TIM 從膏到鋼到液金，牽動貼合設備（竑騰）、鍍金表面處理（匯鑽科 / 健策）、鑄金冷板（奇鋐）——判讀時先確認「焊接或壓合、鍍金存續與否」再對應受惠者。

9. 風險、驗證清單與資料來源

9.1 主要風險

- 規格反覆（已部分兌現）：VR 鍍金水冷板已於 2026-07-22 確認延後；蓋板 one-piece / dual lid / MCL 路線仍可能變動——對高估值散熱股（健策、竑騰前瞻 PE 曾達 73x / 40x）是雙面刃。
- 液態金屬可靠度：滲漏、腐蝕、自動化組裝良率未經大規模量產驗證；VR 鍍金延後即為此風險的體現——若最終退回傳統 TIM，NVIDIA 端鍍金 / 液金邏輯將放緩（TPU 端鋼片 + 鍍金不受影響）。
- MCL 漏液 = 整組報廢：封裝級方案風險集中；測試標準（灌水測試良品定義）未定，導入時程可能後移；Kyber 大規模部署時點亦有變數。
- PFAS 監管：兩相浸沒的氟化液斷鏈（3M 停產、歐盟 2027 立法），限制其放量。
- 供應商集中與競價：NVIDIA VR300 世代冷板供應商收斂至 3-4 家；VR200 冷板 ASP 已因標準化下滑（\$280→\$150）。

9.2 驗證清單（追什麼、為什麼）

追蹤指標	為什麼重要
VR 冷板鍍金延後的落點：後移至 Rubin Ultra？TIM2 改用何種材料？	鍍金延後（2026-07-22 確認）後，NVIDIA 端表面處理與液金供應鏈的時程重估基準
Rubin IHS / 鍍金規格定案（one-piece / dual lid / MCL）	健策 ASP 與竑騰設備/耗材 mix 的直接決定因子
奇鋐 3Q26 新 GPU 冷板模組量產、冷板產能 5x 進度	VR 放量與單機價值 +30-40% 兌現
竑騰鋼片設備 2H26 出貨年增雙位數；其服務製程屬迴焊式或純壓合式	TIM1 升級到鋼片的最直接量化證據；決定 BSM / 鍍金連動需求
TPU v8 / Trainium 3 水冷出貨滲透率（4Q26 起）	2027「ASIC 量 > GPU」第二波量能的先行指標
液態金屬 TIM2 量產良率 / 可靠度	鍍金規格與材料路線存續前提
MCL 灌水測試標準與 Rubin Ultra 測試名單	封裝級散熱路線之爭的勝負手
匯鑽科 2026-09 鍍金量產進度	高門檻外包環節的實績驗證

9.3 資料來源

- ① 內部散熱產業簡報（2026-06-30，彙整凱基投顧、元大投顧、Morgan Stanley、JPM、玉山投顧數據）——散熱方式分級、元件介紹、機櫃液冷價值、ASIC 出貨與散熱方案、放量時程。
- ② SemiAnalysis《Vera Rubin NVL72 拆解》（2026-05-20）——Rubin 雙 lid + stiffener + 鍍金、液態金屬鋼 TIM2、一片式 Strata 冷

板、MCCP 100 μ m、tray 冷卻流路。

- ③ SemiAnalysis《ECTC 2026 先進封裝》(2026-07-02) ——TSMC micropillar 直達矽實測、Microsoft GH200 微流道、SPIL 液態金屬複合 TIM。
- ④ Morgan Stanley Computex 快報 (2026-06-07) ——台達 3MW/6.8MW CDU、Kyber 部署判斷、液冷 readiness 3Q26。
- ⑤ 富邦投顧 (2026-07-21) ——Kyber 0.5U、MCL 側面水路與測試瓶頸、TPLC 兩相液冷、健策 MCL 架構圖。
- ⑥ Daiwa (2026-06-16 / 05-29 / 06-06) ——熱管理 2H26 首選、健策 one-piece lid 急單與 removable dual lid/MCL 線索。
- ⑦ Goldman Sachs《資料中心冷卻與電力》(2026-07-13) ——液冷滲透率 40% (2027)、PUE/WUE 權衡、廢熱回收。
- ⑧ Indium Corporation 官網 (2026-07-22 查證) ——sTIM 迴焊製程與金屬化需求、Heat-Spring 壓合式 TIM 規格。
- ⑨ Microsoft Research / Corintis 公開資料 (2025-09) ——仿生葉脈微流道、3 倍散熱效率。
- ⑩ 供應鏈查核 (2026-07-22) ——VR 世代鍍金水冷板確定延後。

內文標注「假設 / 預估」者為券商彙整推估，非公司正式指引。本報告由 LLM 整理改寫，僅供研究參考，非投資建議。