

# 輝達 2026 年 CoWoS 需求觀察：Rubin 與 Blackwell 的產能配置調整

獨家產業報告 Isaiah Research

2026-01-03 | 收藏 | 半導體 | 每週報告



## 為什麼重要

輝達對台積電 **CoWoS**（先進封裝）的需求量被視為 AI 產業的領先指標。近期 2026 年需求預估下修，表面看似需求放緩，實則是輝達面對 **Rubin** 架構物理極限與 **HBM4**（第四代高頻寬記憶體）整合挑戰時的風險控管。理解這一策略性「產能置換」，能幫助讀者分辨短期的技術磨合與長期的成長趨勢，避免因數據波動誤判 AI 市場前景。

## 三大重點（Key Viewpoints）

- CoWoS 總量修正，策略以退為進** 2026 年需求由 70 萬片下修至約 64 萬片。這並非訂單流失，而是輝達為應對新架構技術瓶頸，主動減少高風險新品、轉向增加成熟產品供應的避險策略。
- Rubin 架構受挫，散熱與記憶體成瓶頸** Rubin 因運算密度過高面臨散熱餘裕不足，且 HBM4 驗證進度落後，導致其量產時程被迫推遲。GR100 需求預估大幅削減，反映了先進封裝正面臨物理極限挑戰。
- Blackwell Ultra 緊急補位，維持出貨動能** 輝達將重心移向製程成熟的 **GB300**（**Blackwell Ultra**），需求量逆勢上修約 10 萬片。此舉確保了供應鏈穩定與良率，以「空間換取時間」來解決下一代技術難題。

# 一、輝達 2026 年 CoWoS 調整預測

近期市場對於輝達 ( Nvidia ) 未來的產品藍圖 ( Roadmap ) 出現了不少雜音，特別是關於下一代旗艦晶片 Rubin 架構與現行 Blackwell 架構之間的產能消長。根據我們最新的供應鏈訪查與數據推估，輝達在 2026 年的 CoWoS ( Chip on Wafer on Substrate，基板上晶圓上晶片封裝 ) 需求總量可能出現下修，從原先預估的 **700,000 ~ 720,000 片 ( kpcs )** 區間，調整至約 **630,000 ~ 650,000 片** 左右的水準。然而，我們必須強調，這個數字的下滑**並不代表終端 AI 市場需求的疲軟**，而更像是輝達在面對新技術驗證挑戰時，所採取的一種「以退為進」的風險控管策略。簡單來說，這是因為 Rubin 架構在後段製程面臨技術瓶頸，導致輝達決定暫時增加成熟度較高的 Blackwell 產能來填補空缺，形成了一種「產能置換」的過渡現象。

## 二、Rubin 架構的技術逆風：散熱與 HBM4 的雙重挑戰

市場原先對代號 GR100 的 Rubin 晶片寄予厚望，但目前的跡象顯示，其放量時間點可能將推遲至 2026 年較晚的時間點。這並非市場不買單，而是生產端的「後段製程驗證」 ( Backend Qualification ) 尚未完全過關。

### 1. 散熱管理與結構極限的拉扯

Rubin 的設計大幅推升了運算密度，這導致「熱餘裕」 ( Thermal Margins ) 變得相當緊繃。在 CoWoS 封裝環節中，中介層 ( Interposer ) 與重分佈層 ( RDL, Re-distribution Layer ) 的結構複雜度，似乎已經被推到了極限，超出了先前 Hopper 或 Blackwell 世代所驗證過的範圍。當晶片運作功率大幅提升，如何在堆疊結構中有效散熱成為了一大難題，這可能是導致 GR100 需求預估量從原本的 390,000 ~ 410,000 片，大幅下修至 250,000 ~ 270,000 片 區間的主因。

### 2. GR150 的延遲登場

至於更受矚目的高階版本 GR150，目前的觀察顯示其在 2026 年前三季幾乎沒有顯著的量能，可能要等到該年第四季才會出現約 4,000 ~ 5,000 片 的極少量產出。這種新舊世代交替的「斷層」，並非正常的產品迭代節奏，這進一步佐證了 Rubin 系列在量產準備度上，尚未達到輝達內部的嚴格標準。

## 三、Blackwell 的戰略補位：相對成熟製程的避風港

面對 Rubin 的技術卡關，輝達並未坐以待斃，而是迅速啟動了緩解措施（Mitigation Strategy），將重心暫時移回製程相對成熟的 Blackwell 架構，特別是 GB300 系列。

### 1. GB300 需求逆勢上修

我們觀察到，隨著 Rubin 的產能預期下修，代號 GB300（Blackwell Ultra）的需求量出現了顯著的增長。原本預估 2026 年的需求約在 215,000 ~ 235,000 片，目前已上修至 315,000 ~ 335,000 片左右。這多出來的約 10 萬片 CoWoS 產能需求，正是為了填補 Rubin 延遲所留下的市場空缺。

### 2. 以空間換取時間的策略

這並非偶然的市場需求波動，而是輝達深思熟慮後的戰略調整。Blackwell 的後段封裝流程已經相當成熟，良率與穩定性皆有保障。透過「由 Rubin 轉向 Blackwell」的產能置換，輝達可以在維持整體系統出貨動能的同時，爭取更多時間來解決 Rubin 在散熱與封裝結構上的技術難題。雖然這導致 2026 年整體的 CoWoS 晶圓需求總量看起來是下降的，但這其實是一種「寧缺勿濫」的選擇——寧可提供穩定可靠的舊架構升級版，也不願貿然將尚未完全成熟的新產品推向市場。

## 四、 HBM4 整合難題：記憶體大廠的驗證進度

除了晶片本身的封裝結構外，第四代高頻寬記憶體（HBM4, High Bandwidth Memory 4）的整合也是 Rubin 難以快速放量的關鍵制約因素。

### 1. 介面與訊號的挑戰

HBM4 帶來了更寬的傳輸介面與更高的訊號速度，這直接導致了功率密度的增加。在 CoWoS 封裝階段，這會顯著放大對良率的敏感度（Yield Sensitivity）。換句話說，只要製程稍有偏差，配合 HBM4 的封裝良率就可能大幅下滑。

### 2. 供應商驗證進度落後

根據我們對供應鏈的觀察，主要記憶體供應商的進度似乎比預期稍慢。SK 海力士（SK hynix）在 11 月時，針對 HBM4 的迴路測試（Loopback）與功率特性分析（Power-characterization）進度顯得較為緩慢；而三星電子（Samsung Electronics）的情況可能也相去不遠，

其 HBM4 與 Rubin 的堆疊驗證推測直至 12 月初才正式展開。這些上游關鍵零組件的驗證延遲，都進一步限制了 Rubin 在 2026 年大規模量產的可能性。

## 五、 結論與未來展望

總結來看，輝達在 2026 年的 CoWoS 需求預估下修，反映的並非 AI 需求的冷卻，而是半導體先進封裝技術正面臨物理極限的挑戰。

我們預期，在散熱管理與 HBM4 整合風險完全排除之前，Rubin 架構將難以重回放量的主流隊伍。在此之前，輝達採取向 Blackwell 傾斜的產能置換策略是相當理性的。值得注意的是，Blackwell 架構在中介層上的晶片封裝密度 ( Gross packages per interposer ) 大約是 Rubin 的兩倍，這意味著即便 CoWoS 晶圓總片數下降，實際產出的終端運算單元數量未必會同比例減少。

展望 2026 年，可持續關注台積電 CoWoS 產能的實際分配情況，以及記憶體大廠在 HBM4 驗證上的突破進度。這段技術磨合期或許會持續數個季度，但隨著 Blackwell Ultra ( GB300 ) 的補位，整體 AI 伺服器的出貨動能預計仍將維持在相對高檔的水準。

2026 年輝達 CoWoS 先進封裝產能需求預估調整 (單位：片數)

產品代號 (架構)	原始預估需求 (片)	最新修正預估 (片)	趨勢變化	關鍵因素解析
GR100 (Rubin)	390k ~ 410k	250k ~ 270k	大幅下修	散熱瓶頸、HBM4 驗證未過，導致量產推遲。
GR150 (Rubin Ultra)	(極少)	4k ~ 5k	遞延	預計僅在 2026 Q4 出現少量產出，技術門檻更高。
GB300 (Blackwell Ultra)	215k ~ 235k	315k ~ 335k	顯著上修	作為 Rubin 的替代方案，製程成熟，填補產能空缺。
其他/舊款	90k ~ 110k	50k ~ 70k	持平/微調	Vera 等有一定程度下修。
總計 CoWoS 需求	700k ~ 720k	630k ~ 650k	總量下修	策略性置換導致總片數下降，但旨在降低生產風險。