



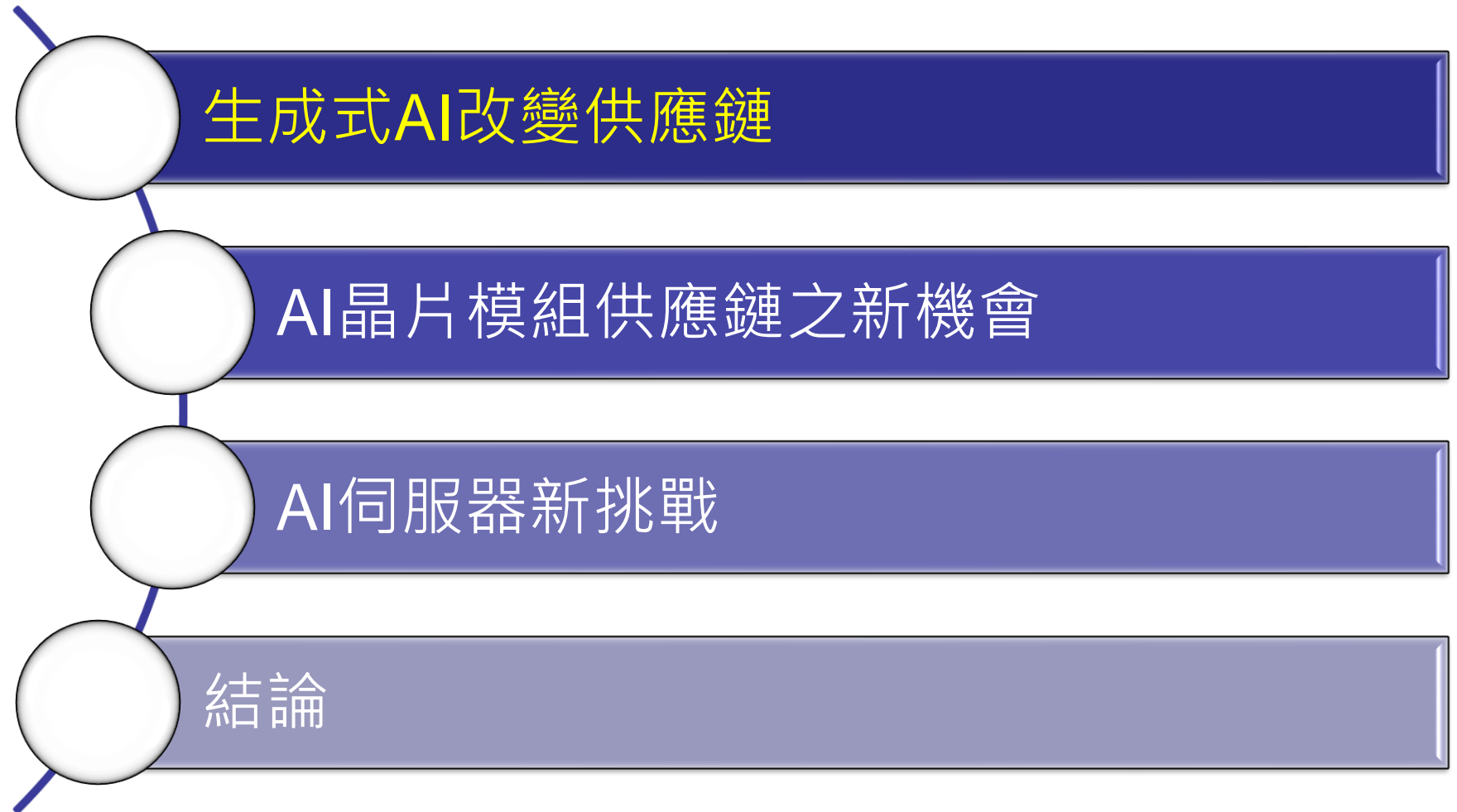
# ChatGPT創造封裝供應鏈 新產值

電光系統所異質整合技術組 (EOSL)

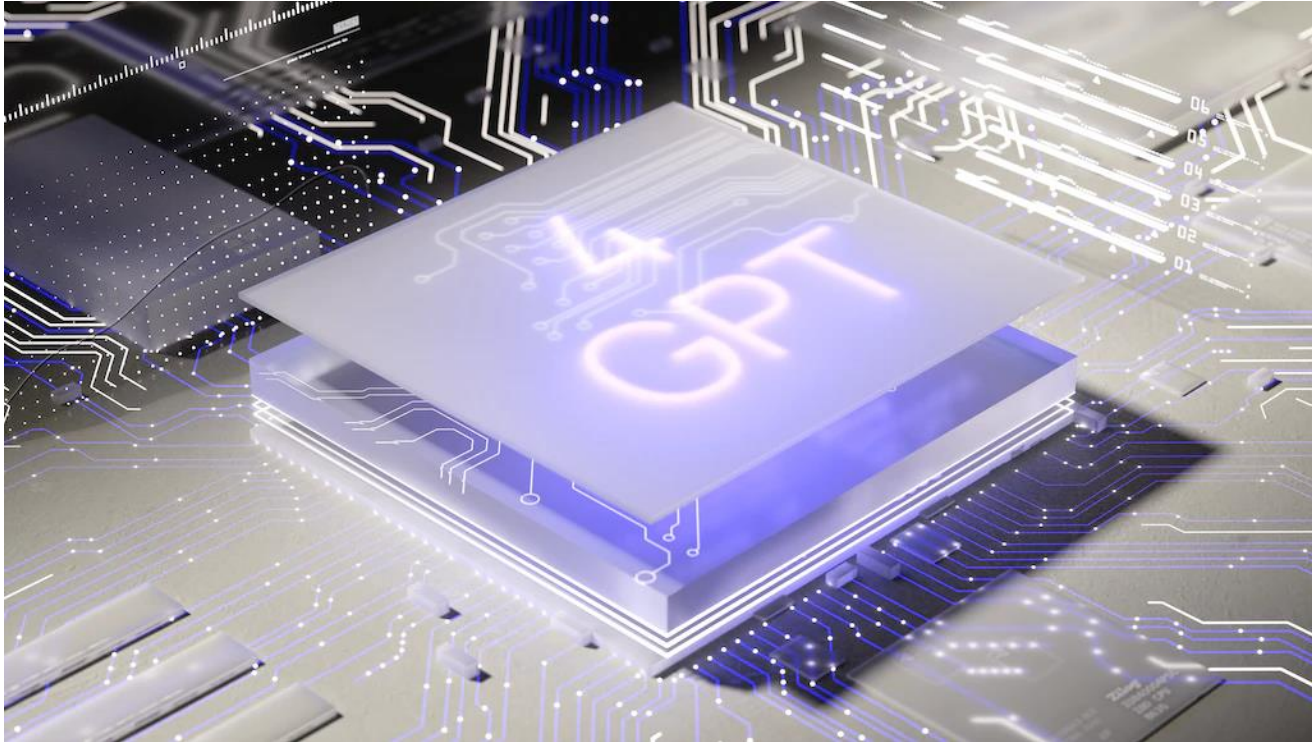
王欽宏 組長

2023/6/2

# 大綱



# ChatGPT帶來機會

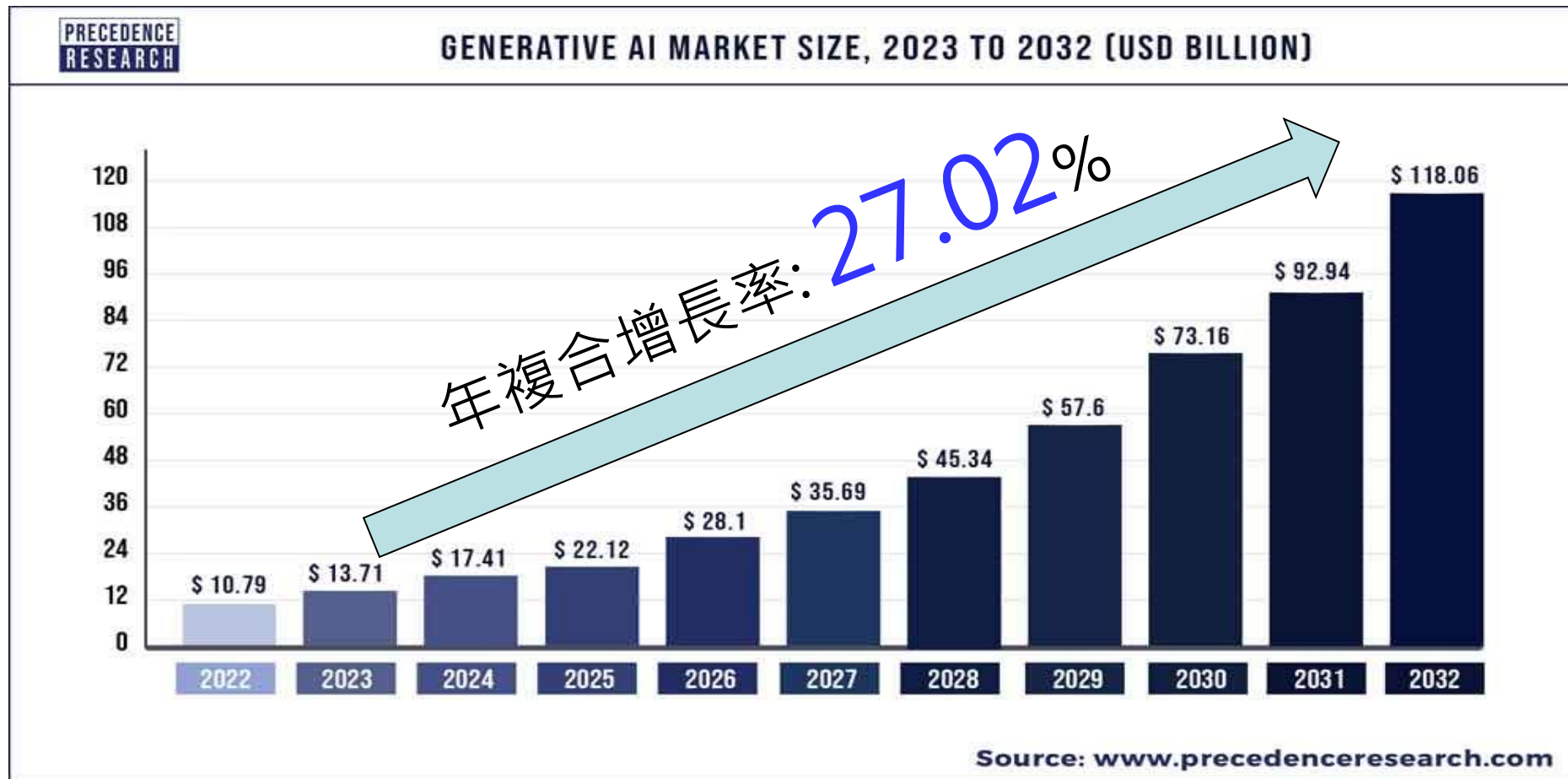


ChatGPT 是 OpenAI 開發的創新人工智能聊天機器人

- ✓ ChatGPT 2023/02 就積累了超過**1 億用戶**(成立兩個月)
- ✓ ChatGPT 在全球 **162** 個國家/地區可用
- ✓ 每月運行 ChatGPT 的成本超過**300 萬**美元

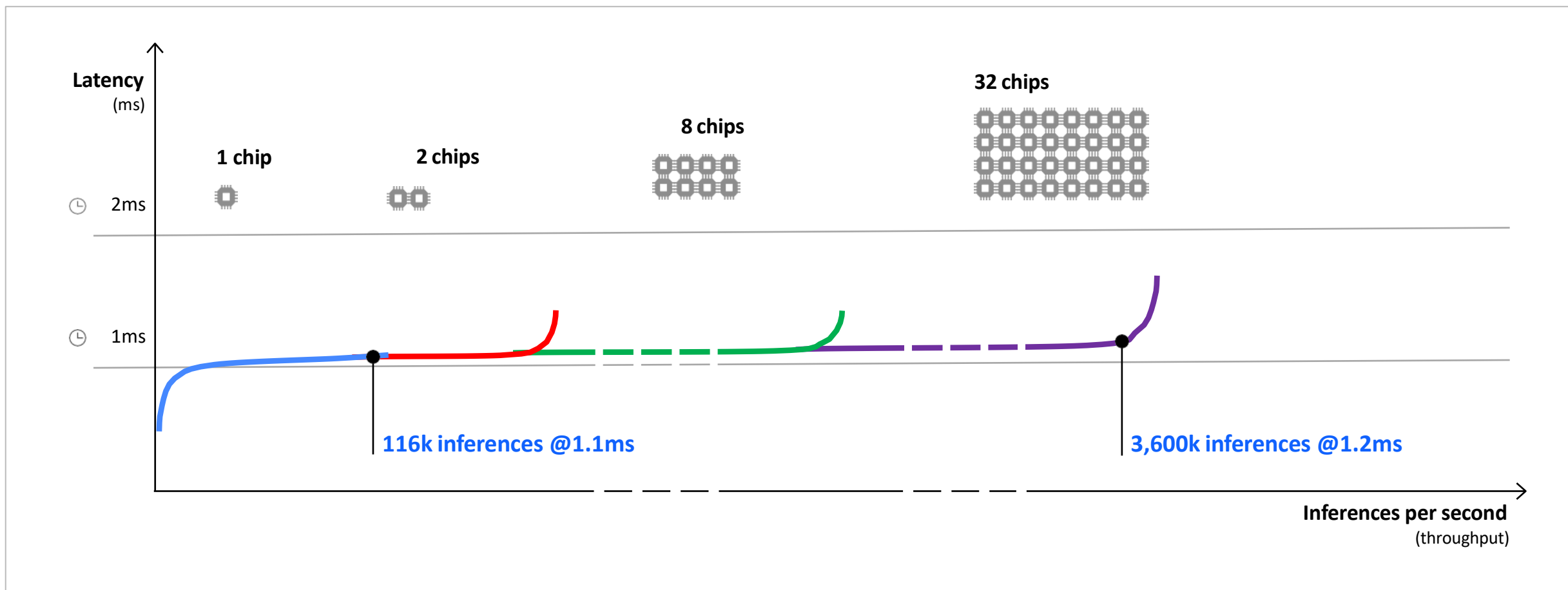
# 生成式AI應用市場未來產值預估

- ✓ ChatGPT是由生成式AI開發出來的產品，運用超大型AI算法進行運算
- ✓ 全球生成式AI應用市場預估2023年為137.1億美元，到2032年將達到約1180.6億美元



# AI Accelerator Performance Trend

➤ RNN multi-layer model for Credit Card Fraud – proxy model developed with global bank



# 超大型AI帶動晶片模組產業鏈改變

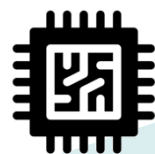
AI演算法

BERT  
340M

➤ AI參數指數級成長 **千倍**

GPT-3  
175B

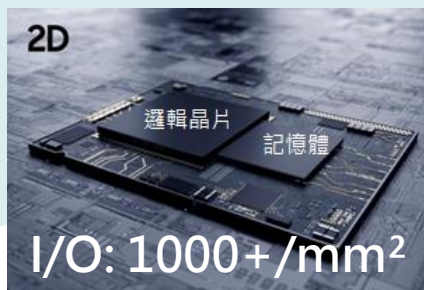
AI運算晶片



➤ GPU運算力增加 **百倍**

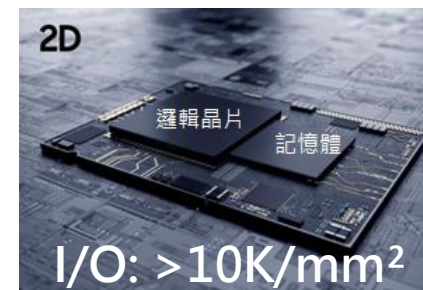


AI異質封裝



➤ I/O傳輸點增加 **數十倍**

➤ 運算瓦數 **近千瓦**



2019

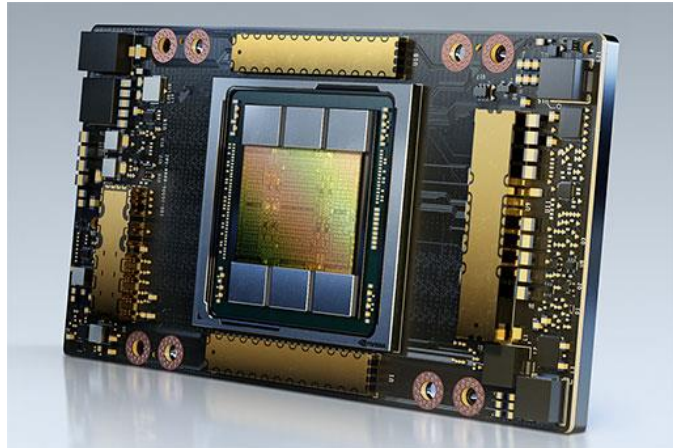
2023~



# 生成式AI需要多核心高速運算力

- ✓ 先進半導體技術實現高算力的AI晶片
- ✓ 高算力多核心AI晶片模組需要更多晶片數量

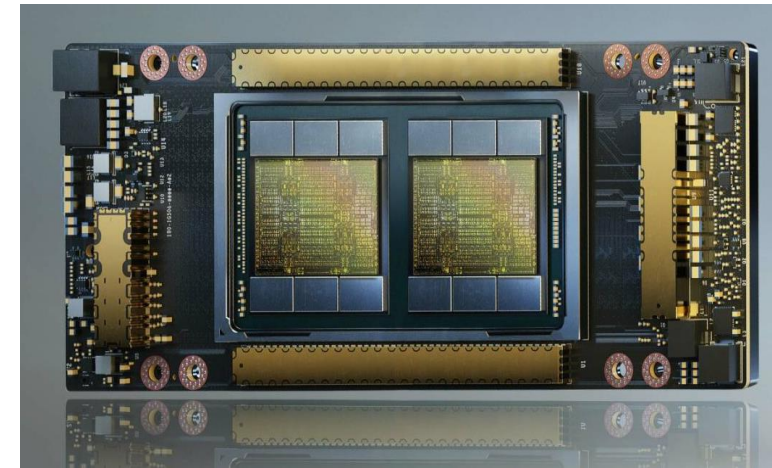
## Nvidia A100



7<sub>nm</sub>  
54.2億電晶體  
6<sub>HBM</sub>  
< 500 W

Source: Nvidia

## Nvidia H100

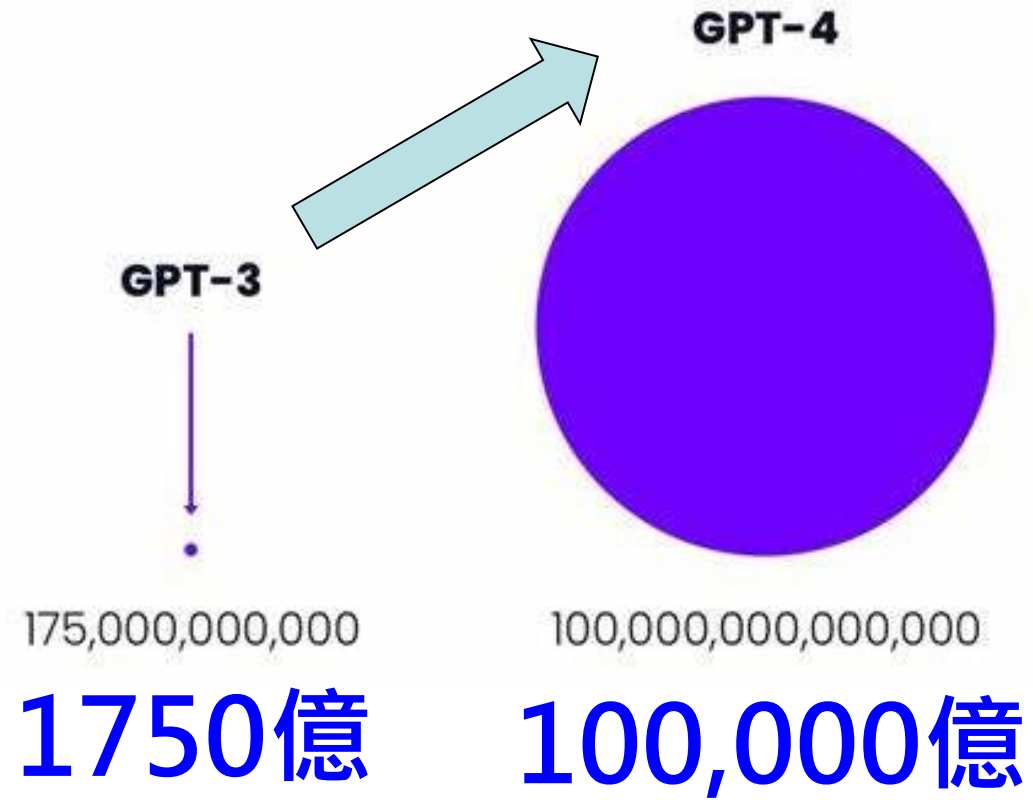


4<sub>nm</sub>  
80億電晶體  
12<sub>HBM</sub>  
> 600 W

# 生成式AI未來運作挑戰?

未來硬體面臨共同設計挑戰

更高算力、更大頻寬、更省電

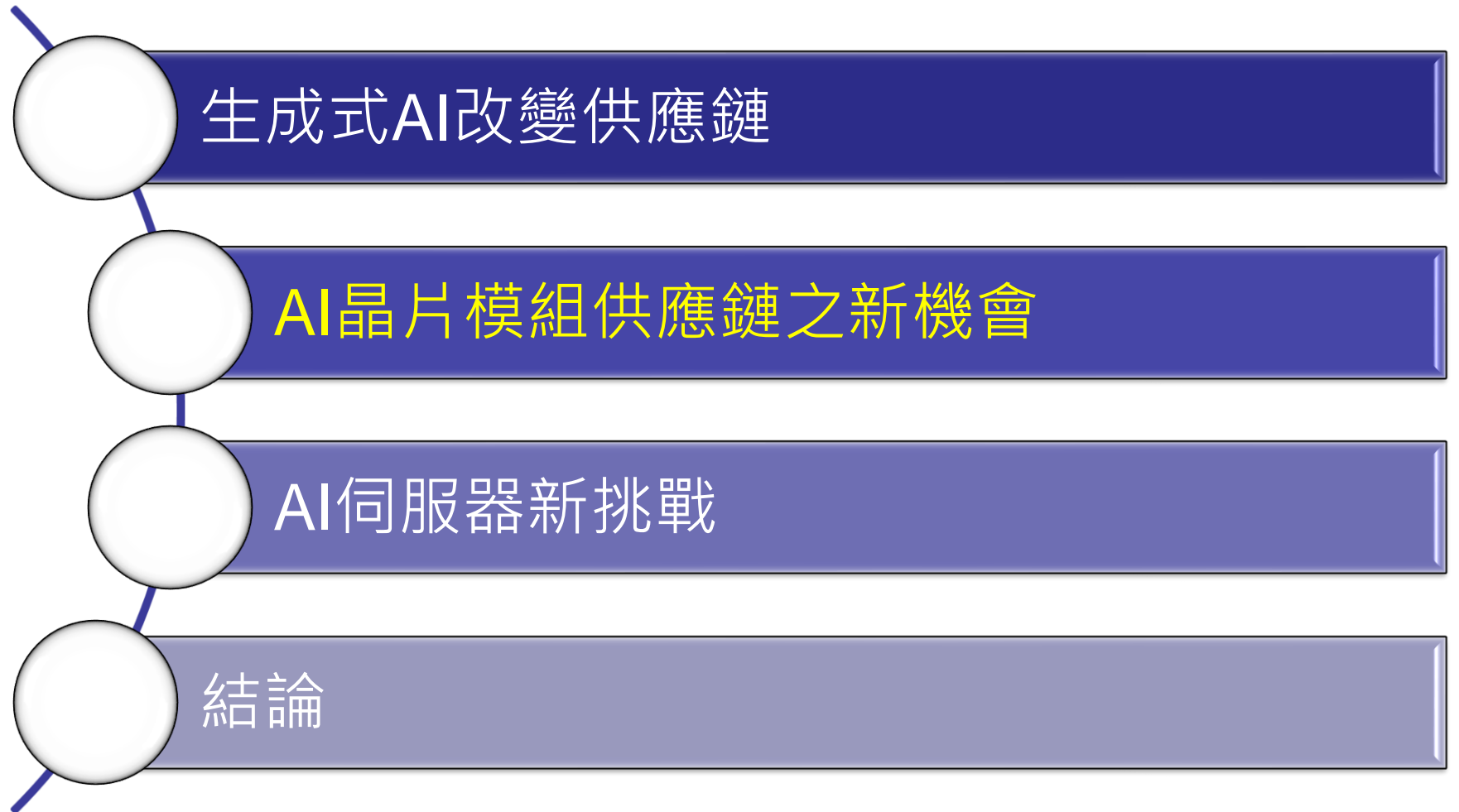


超過萬億級參數

數十億筆數據/微秒

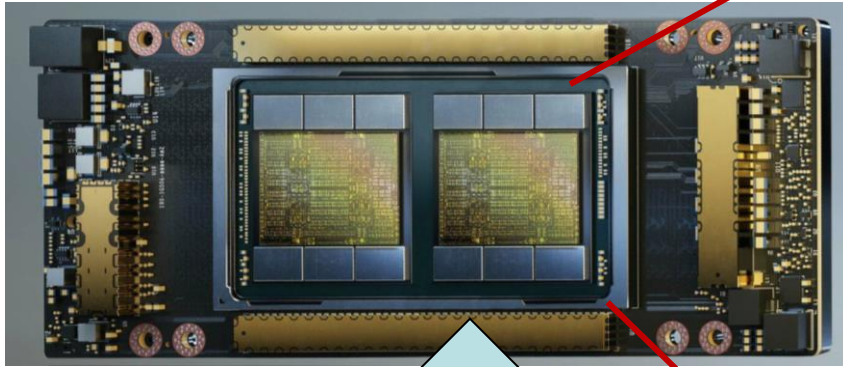


# 大綱

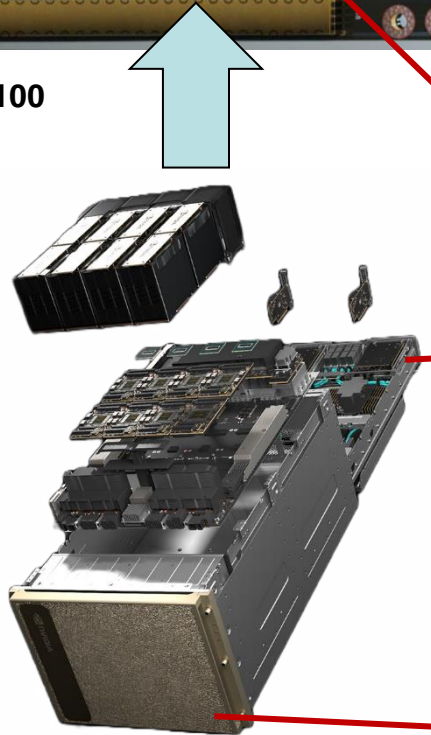


# 生成式AI運算伺服器之台灣產業鏈

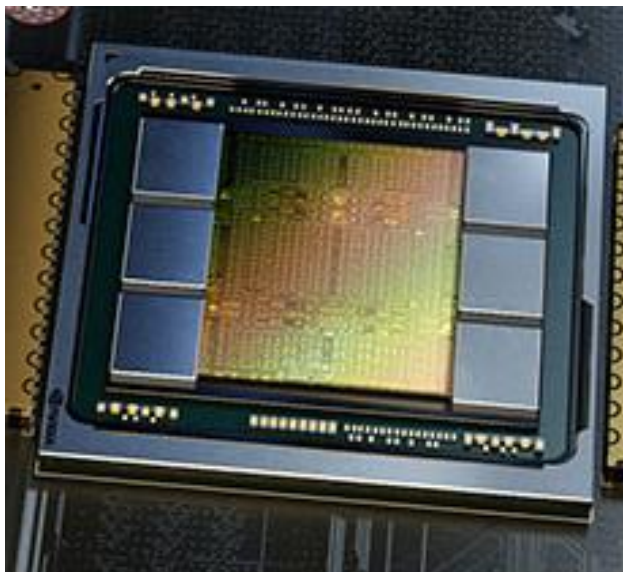
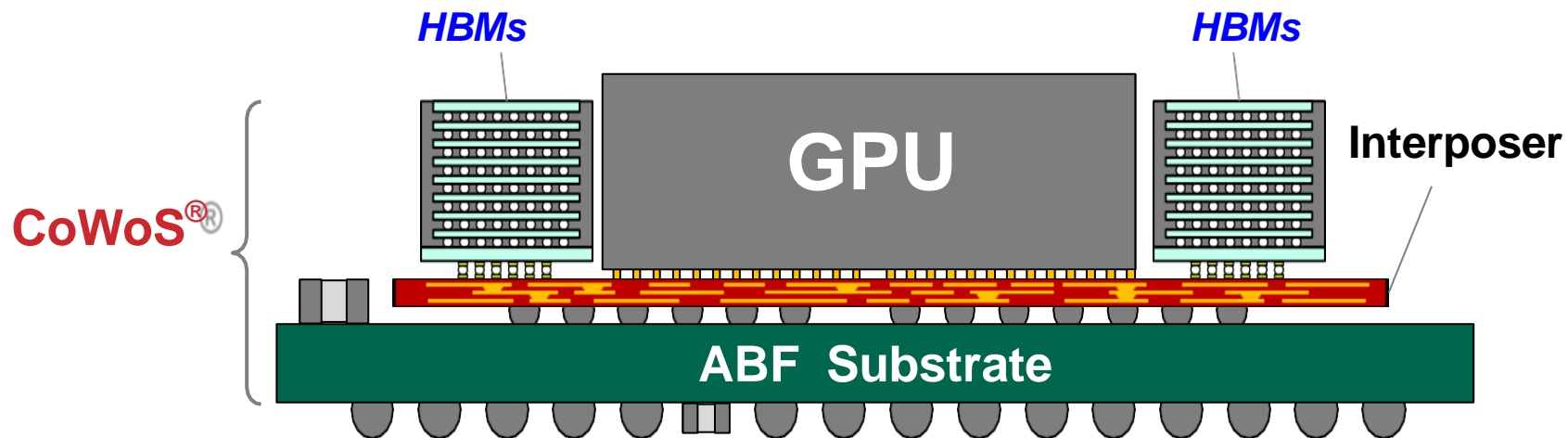
## AI Computing Module



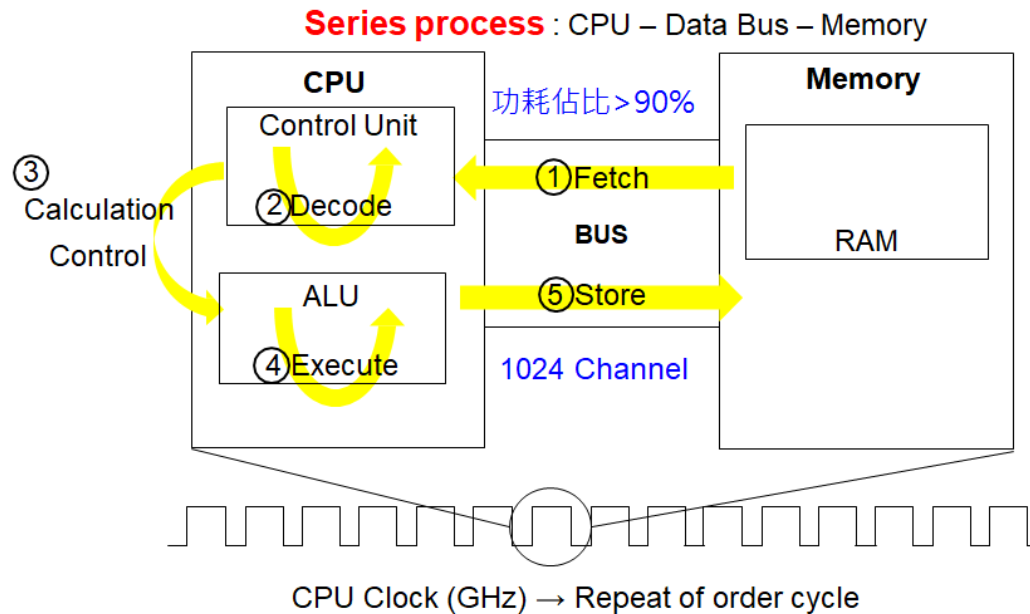
Source: Nvidia H100



# AI晶片模組架構與Von-Neumann架構運算



Source: <https://3dfabric.tsmc.com/english/dedicatedFoundry/technology>



AI 晶片運算  
Von-Neumann架構

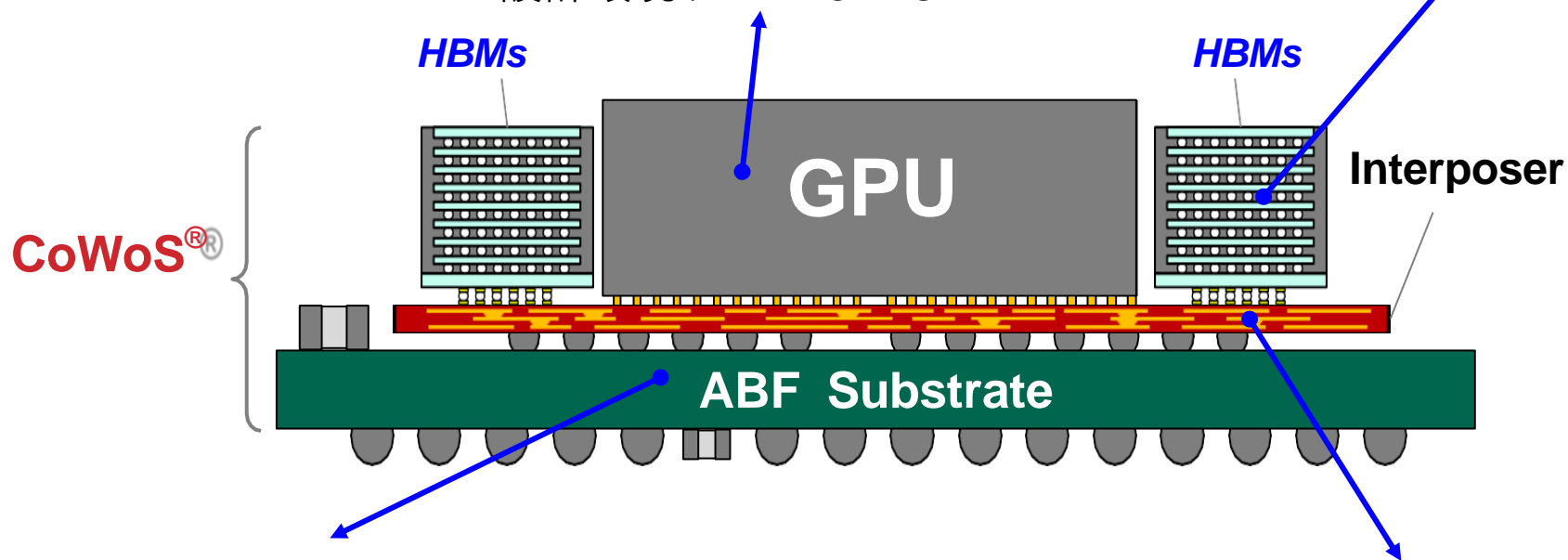
# 生成式AI晶片模組挑戰

高運算力、低耗能

- ✓ 節點技術：<5nm
- ✓ Data rate：UCIe 1.0→ UCIe2.0
- ✓ 設計環境：PDK for <5nm

大儲存容量、高速資訊傳輸

- ✓ HBM：8H
- ✓ Data rate：7.2Gbps@HBM3



大面積、高密度、高速傳輸

- ✓ 載板面積：> 90mmx90mm
- ✓ 導線細線距：< 5 um

大面積、高密度、高速傳輸

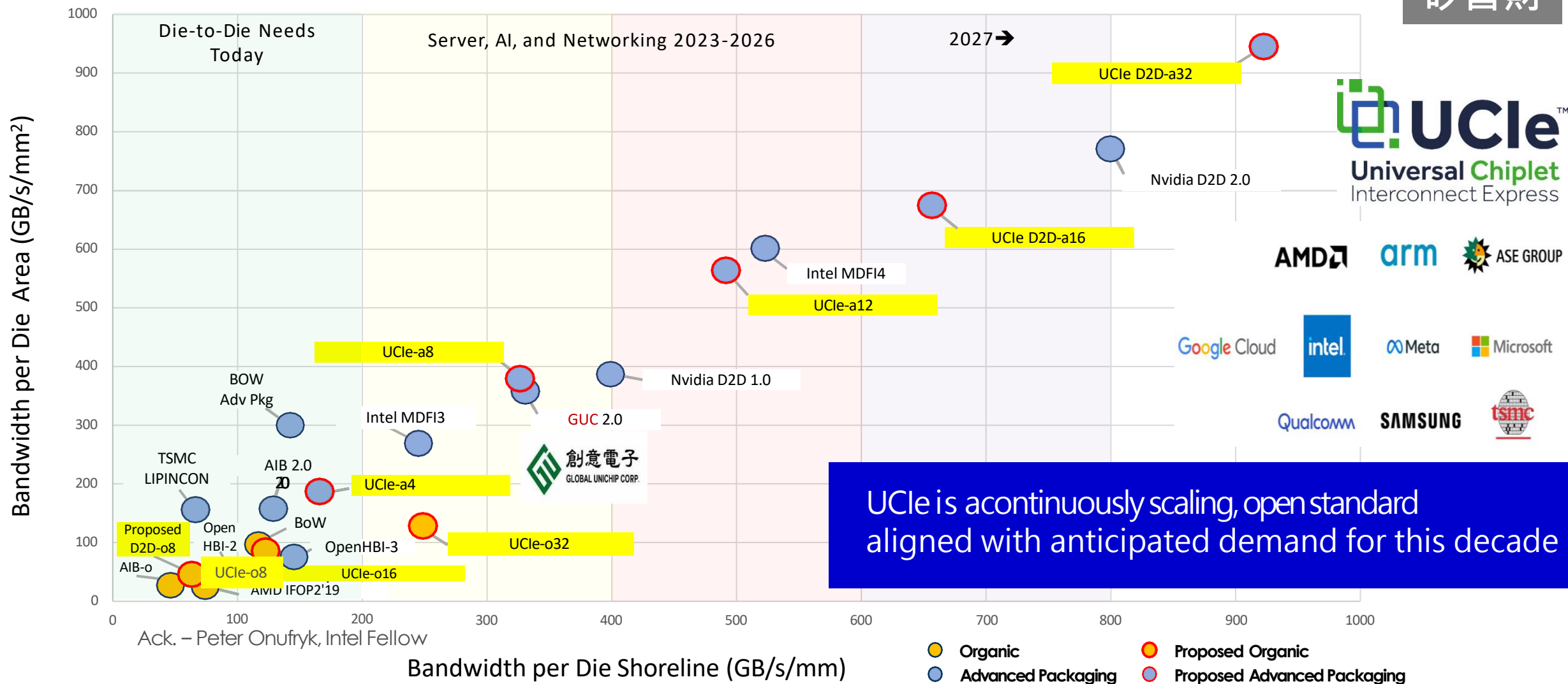
- ✓ CoWoS面積：~6 reticle
- ✓ 導線層數：> 5 layer
- ✓ Data rate：UCIe 1.0→ UCIe2.0

# UCle and Interconnect Competitive Landscape

✓ 掌握設計規畫提供高算力AI晶片設計服務

限閱  
RESTRICTED

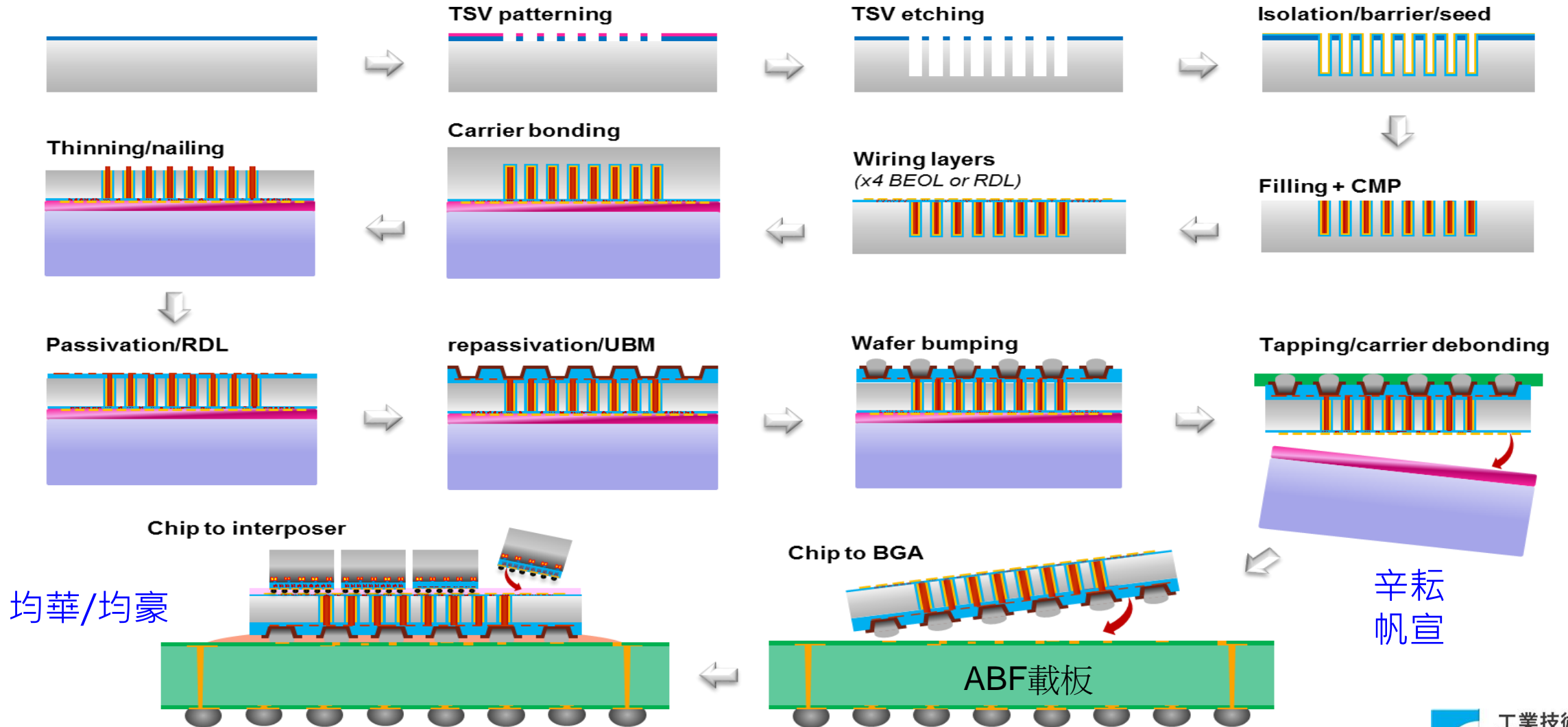
矽智財



Source: Intel 2022 VLSI-TSA

工業技術研究院  
Industrial Technology  
Research Institute

# 2.5D Silicon Interposers製造流程





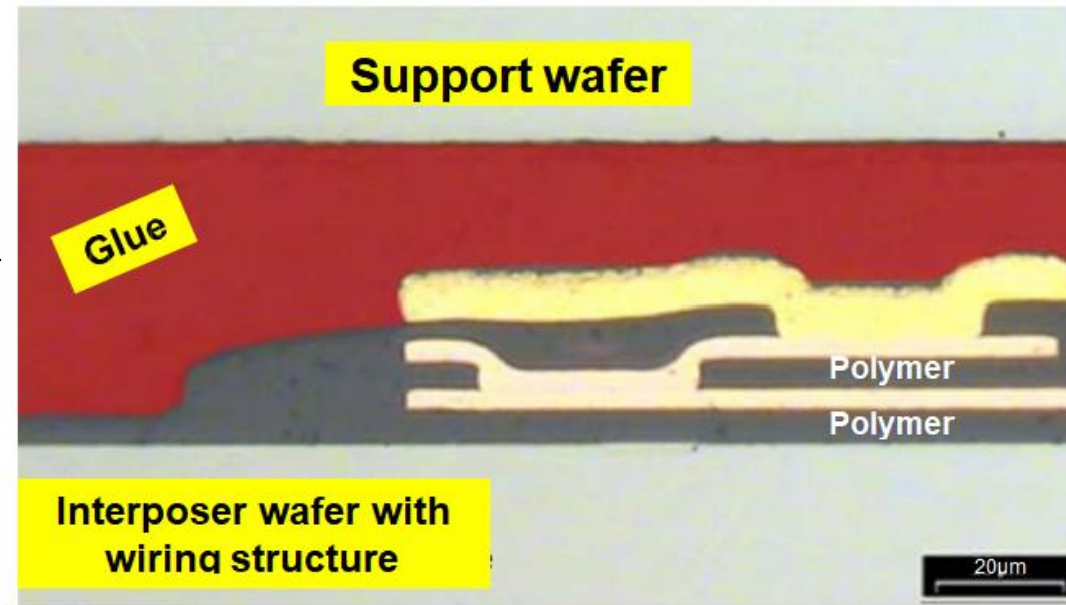
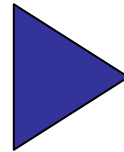
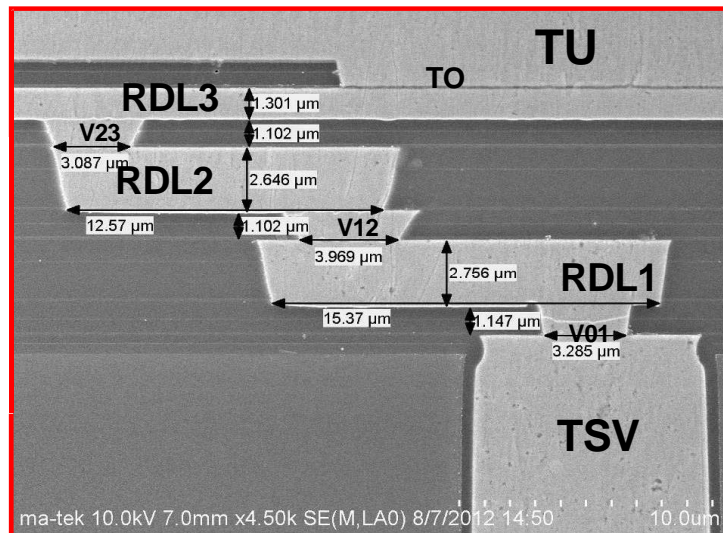
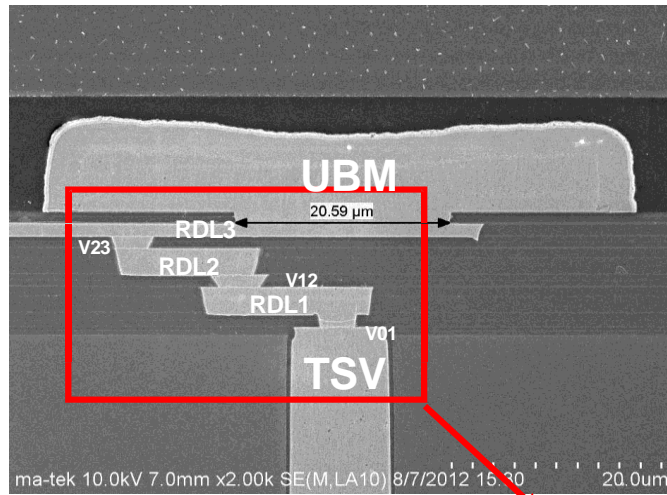
# Si 中介層的挑戰

無機介電材料：氧化矽

Si 中介層高速傳輸：32Gbps

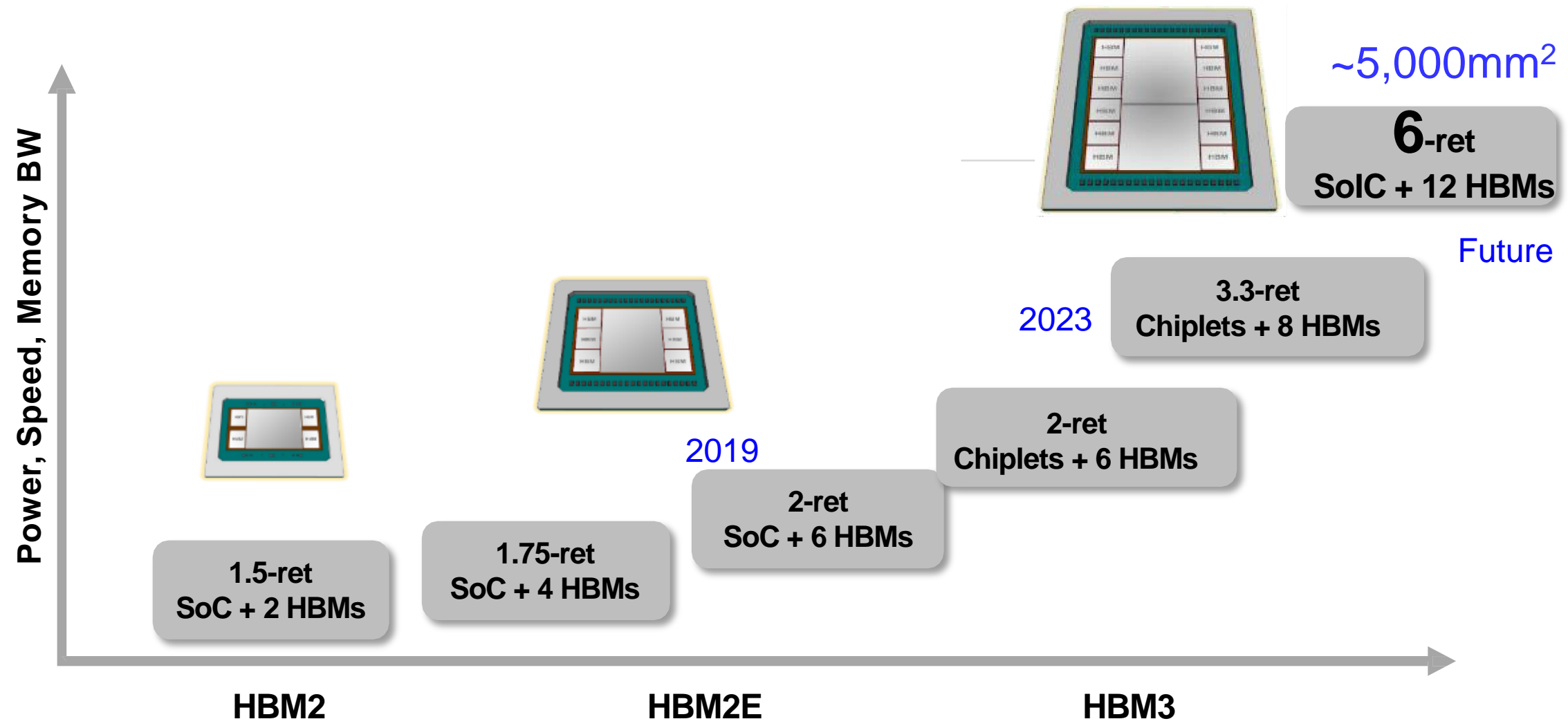
材料Dk：PI 比 氧化矽低

有無機介電材料：PI

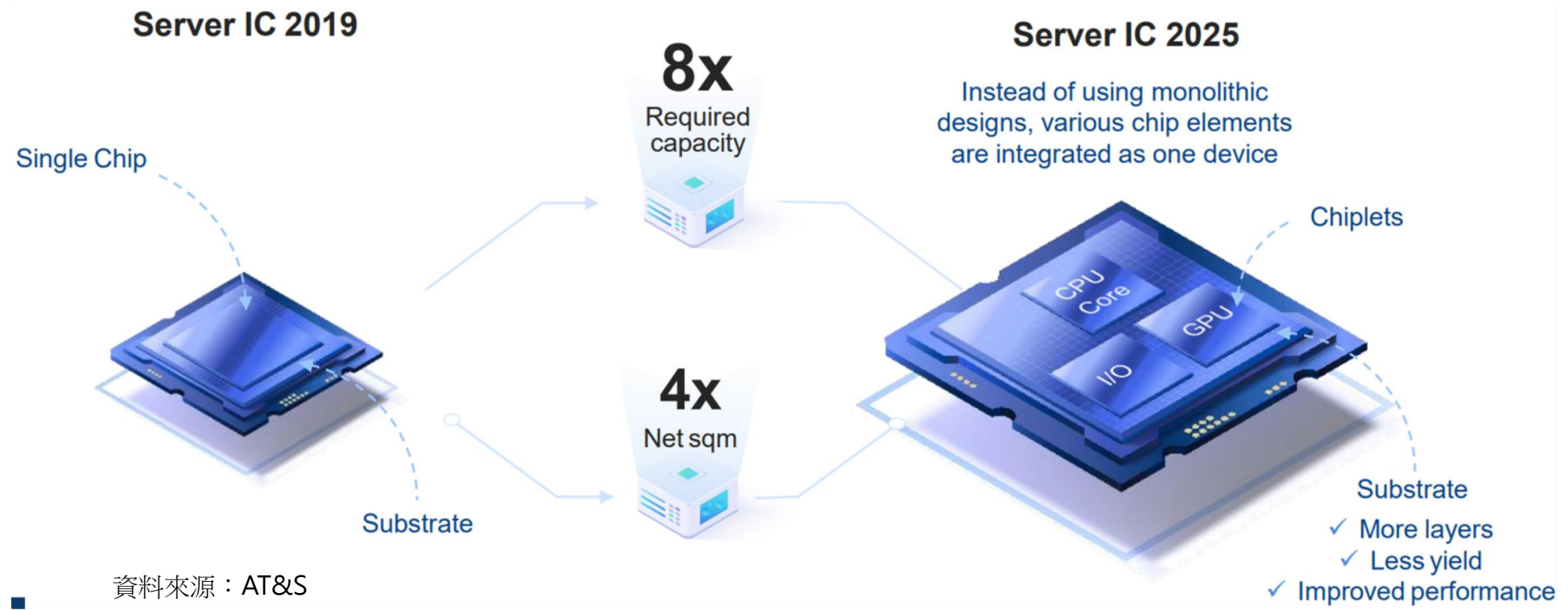


國內材料商  
?

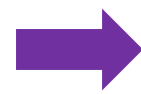
# CoWoS基板發展趨勢



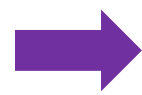
# 算力需求與封裝升級 ABF載板需求擴大



為滿足超大型AI運算需求，異質整合封裝技術是未來的發展趨勢，將來自不同FAB、不同製程節點與不同屬性的晶片整合成SOC-like晶片，提供IC設計業者更高的運算效能。

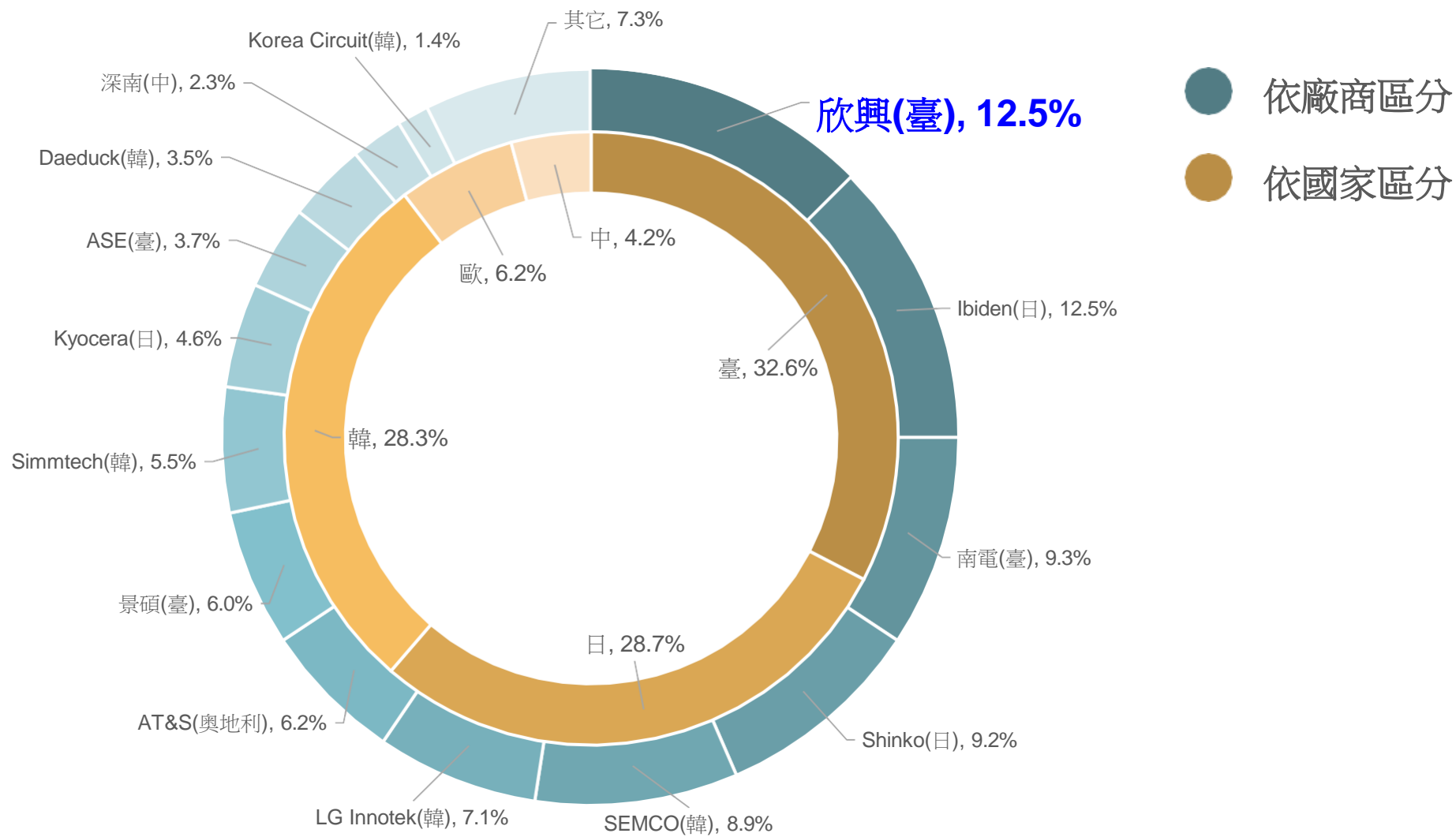


- ABF載板面積增加
- ABF載板層數增加
- 線路密度更密集
- 製作難度提高



- 翹曲
- 細線路
- 微導孔

# 2021全球載板市占率表現

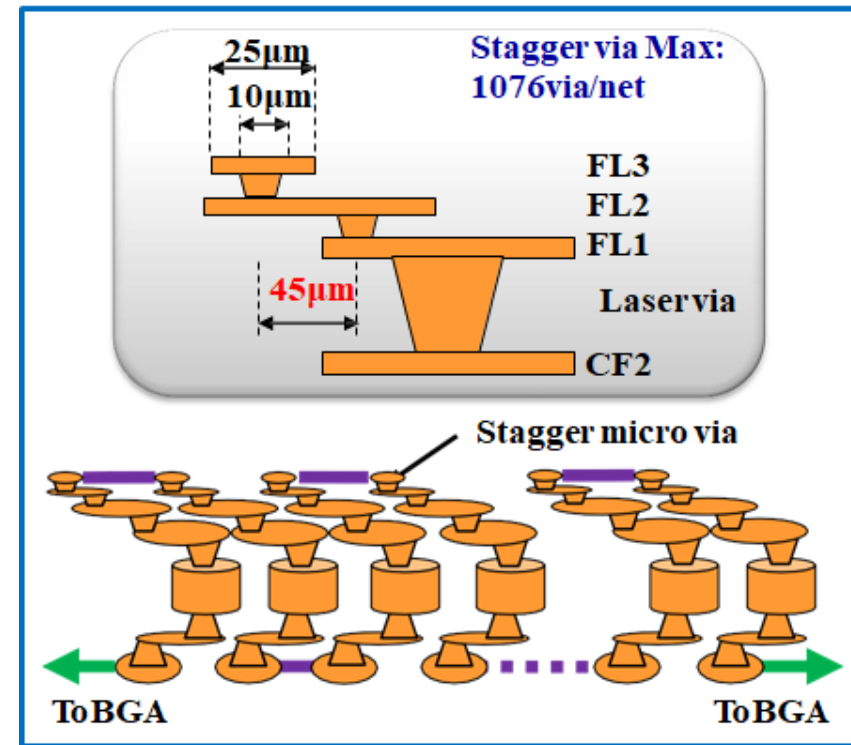
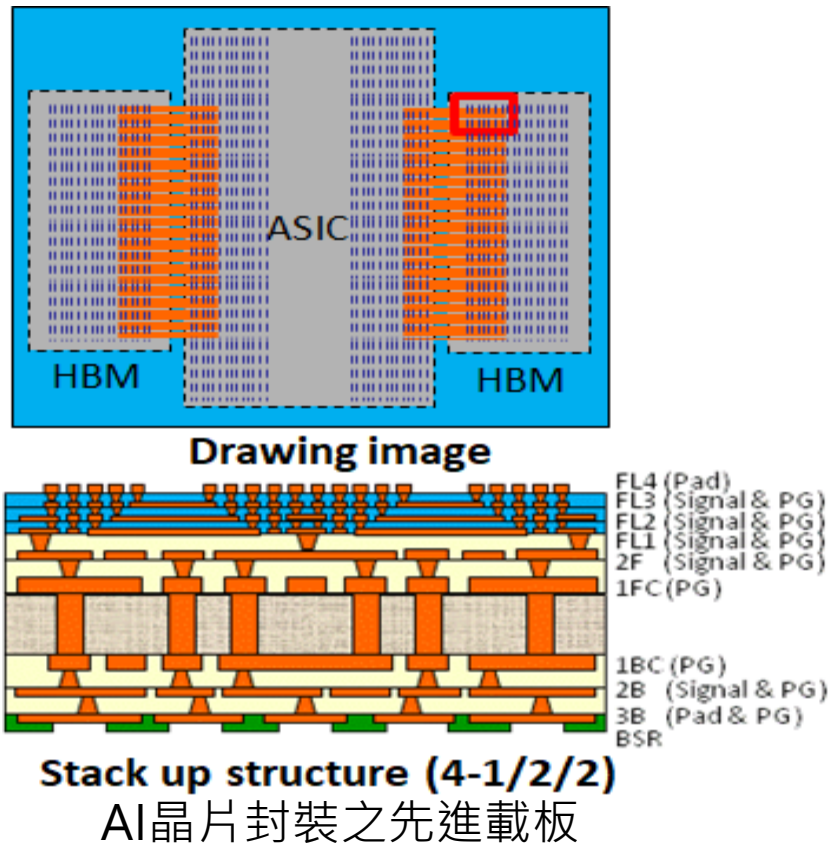


資料來源：工研院產科國際所



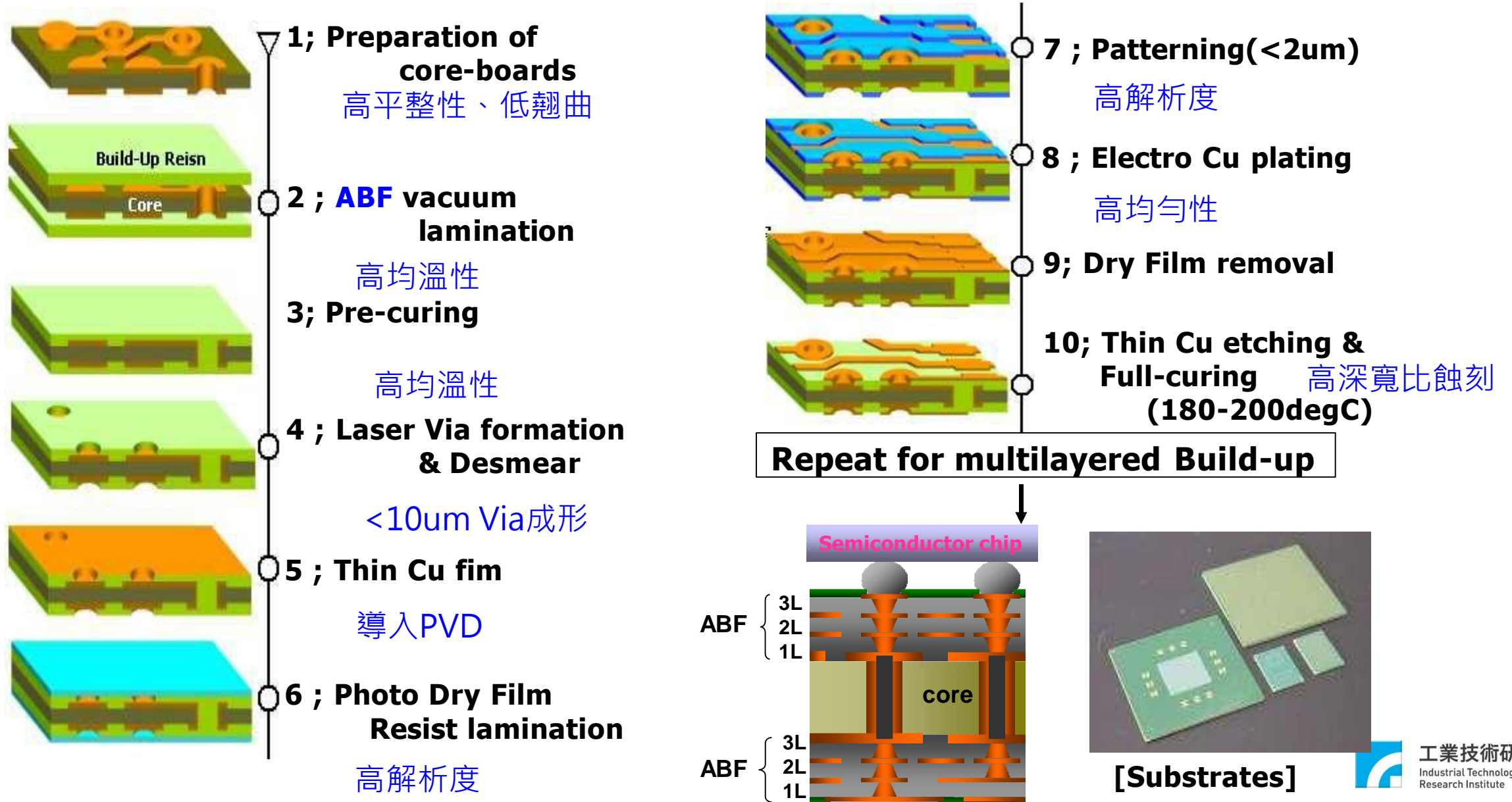
# 下世代ABF載板技術發展挑戰

- AI晶片封裝之先進載板必須提供**高密度導線(1024I/O)架構**，滿足快速AI運算需求
- 先進載板經計算需**10um Via** 才能提供提供超過1024 Via供系統運用
- 先進載板絕緣層 loss tangent 需 < 0.005，才能滿足AI運算高速傳輸，**ABF是最佳選擇**。



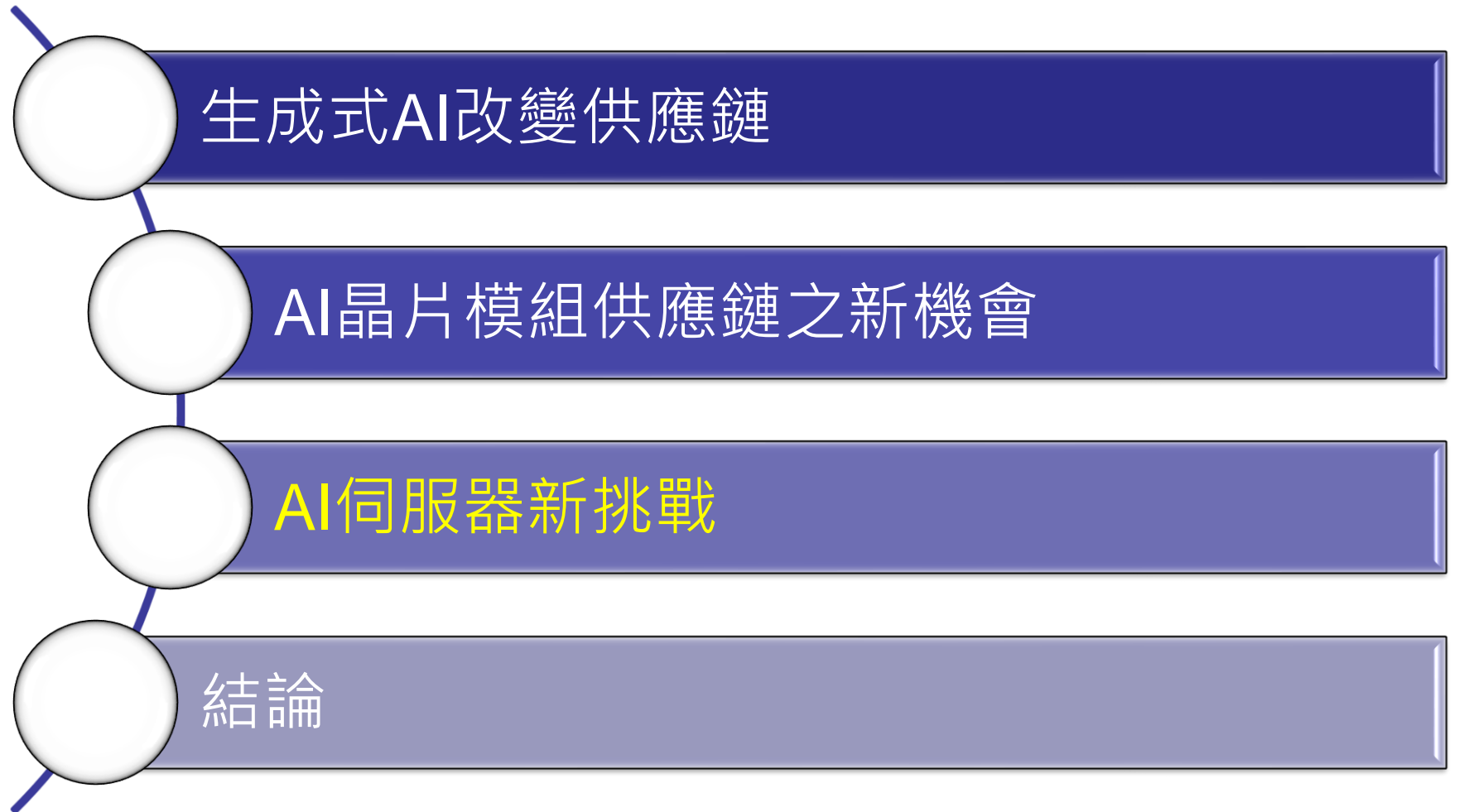
AI晶片先進載板需10um Via

# 下世代ABF載板製程挑戰

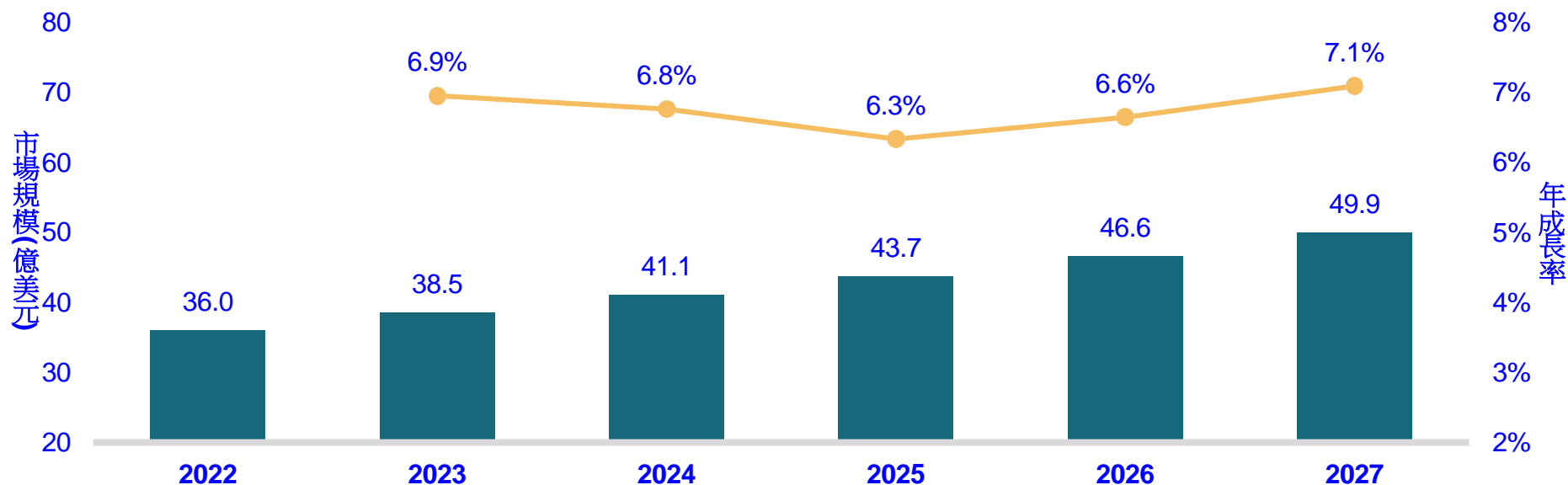




# 大綱



# 全球HPC市場規模預估



CAGR (2022~2027)

**6.7** %

2027全球預估產值

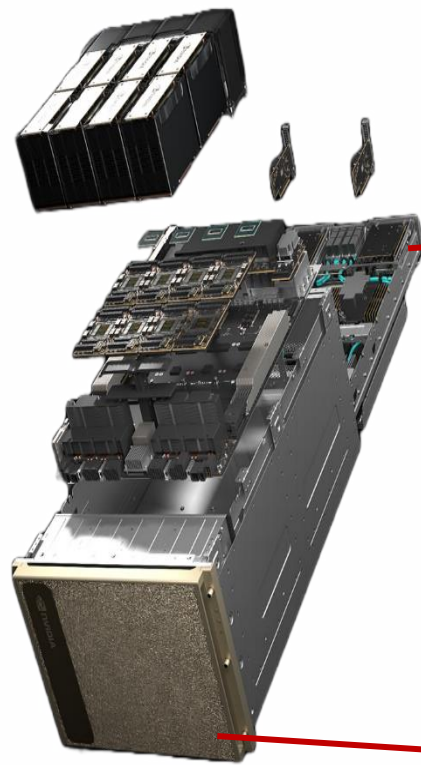
**49.9** 億美元

主要成長動能

1. AI
2. Machine Learning
3. Big Data

# 生成式AI運算伺服器最大挑戰

Thermal: 600 W  $\rightarrow$  **>1000 W**



伺服器其他元件

信驊

PCB

金像電、台光電

散熱

健策、奇鋳、雙鴻、廣運

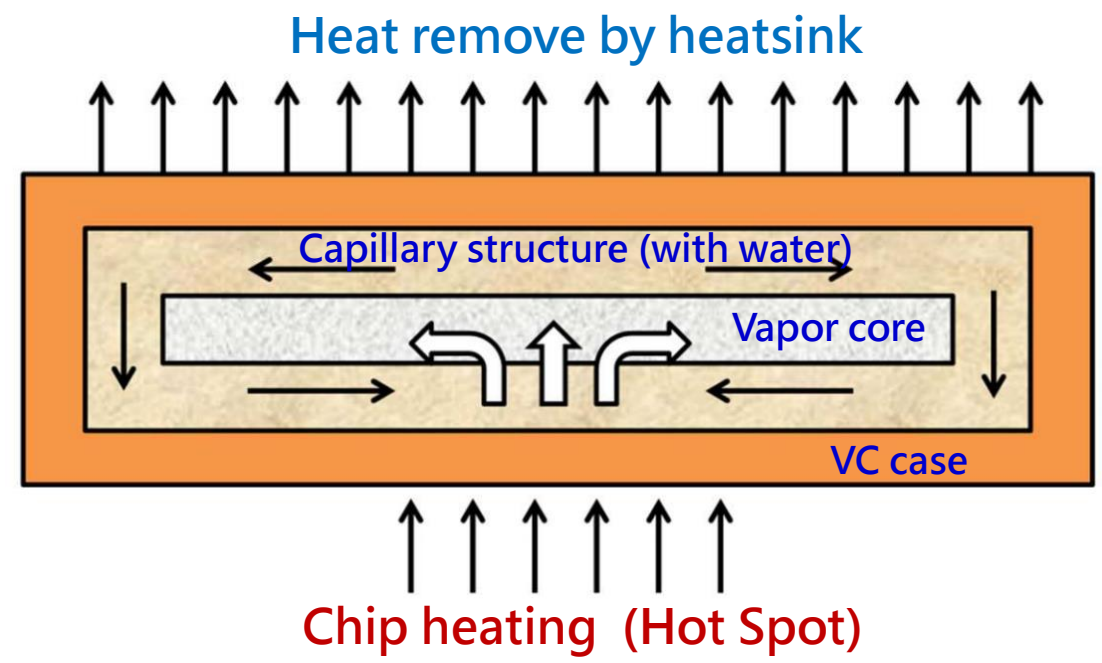
雲端伺服器

緯穎、廣達、技嘉、鴻海、緯創

# VC (Vapor Chamber)

An excellent heat spreading device

## VC 元件架構



VC is a closed hollow shell, the inner cavity is evacuated, and the capillary structure contains water.

## VC 運作原理

- VC 腔內近似真空狀態，液態水可在50°C 左右開始沸騰，成為蒸氣。
- 水的蒸發沸騰可吸收大量的熱，產生50°C 的蒸氣。
- 50°C 的蒸氣流動至 TGP 外圍，因散熱再凝結成 50°C 液態水，蒸氣凝結可釋放大量熱量至外界。
- 50°C 的液態水利用毛細作用再流回加熱區，並完成一個循環。
- **因此，TGP 可維持極佳的均溫性，提升散熱效果。**

# 雙相液冷系統

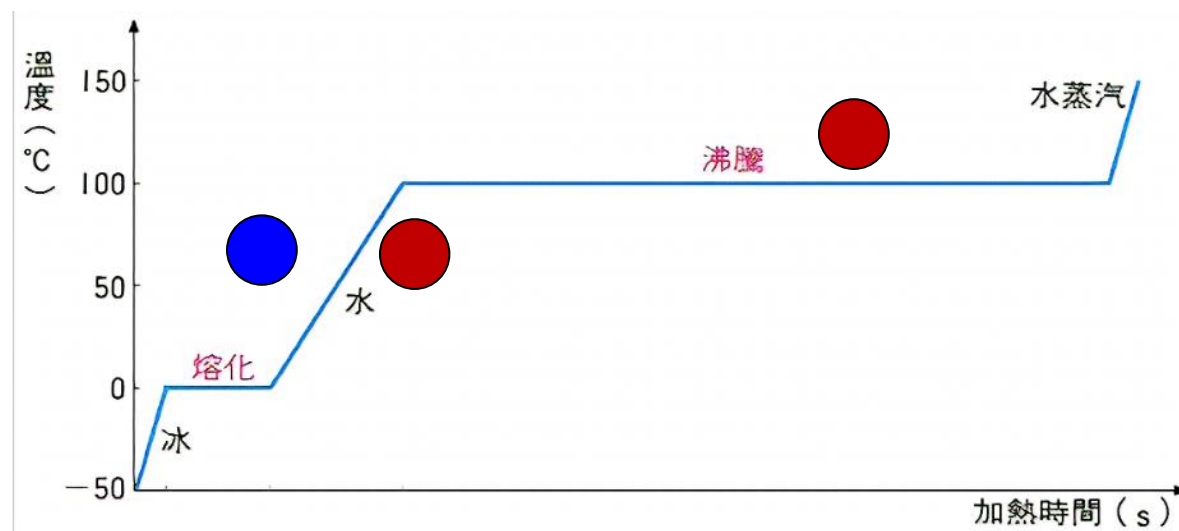
✓ 雙相液冷系統實現超高算力的AI晶片伺服器的解決方案

## 雙相液冷系統

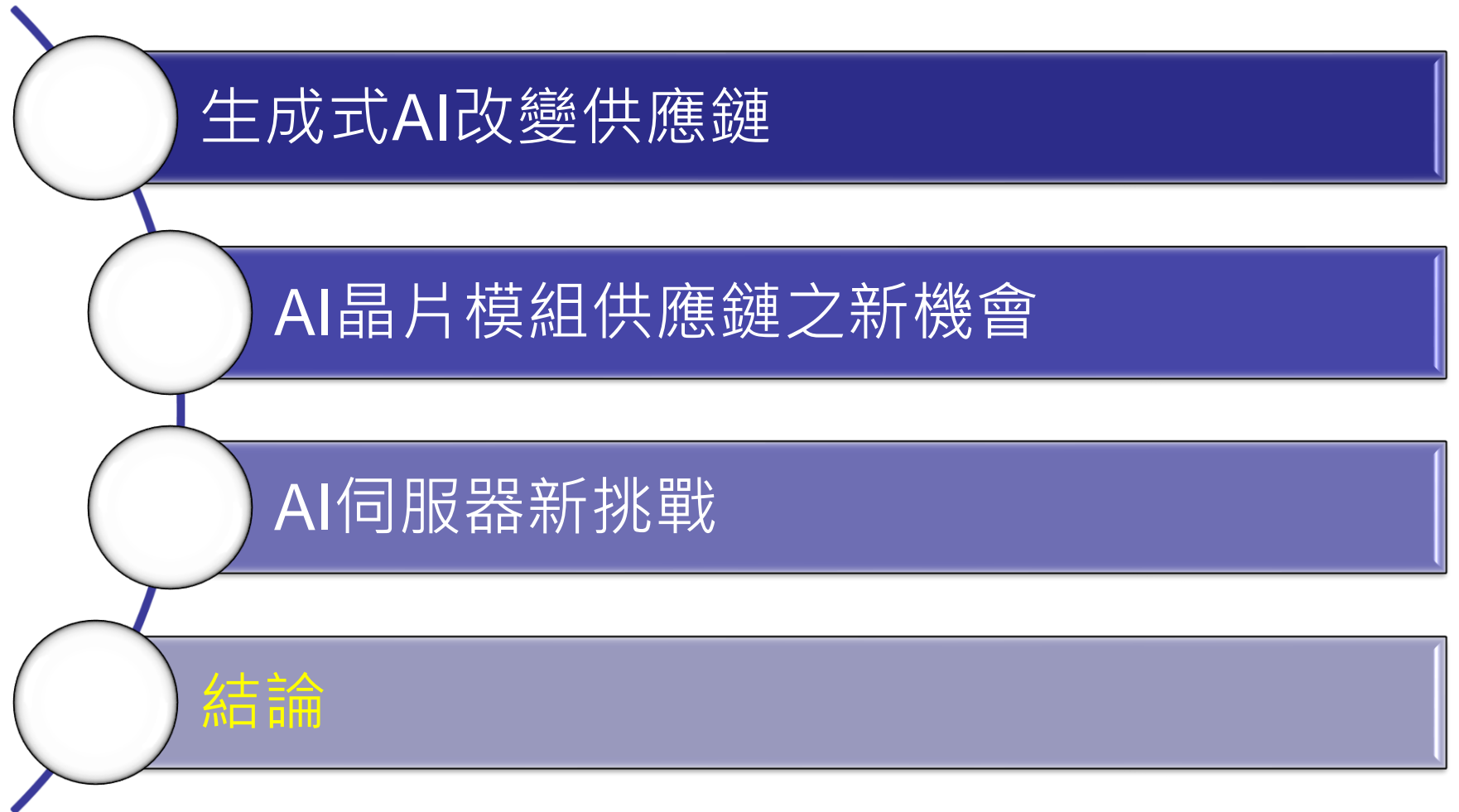


## 雙向液冷運作原理

- 運用流動液體吸熱轉換至氣體之雙向行為創造單位面積最大熱力。
- 單相為20-40W/cm<sup>2</sup> ● 雙相是200W/cm<sup>2</sup> ●



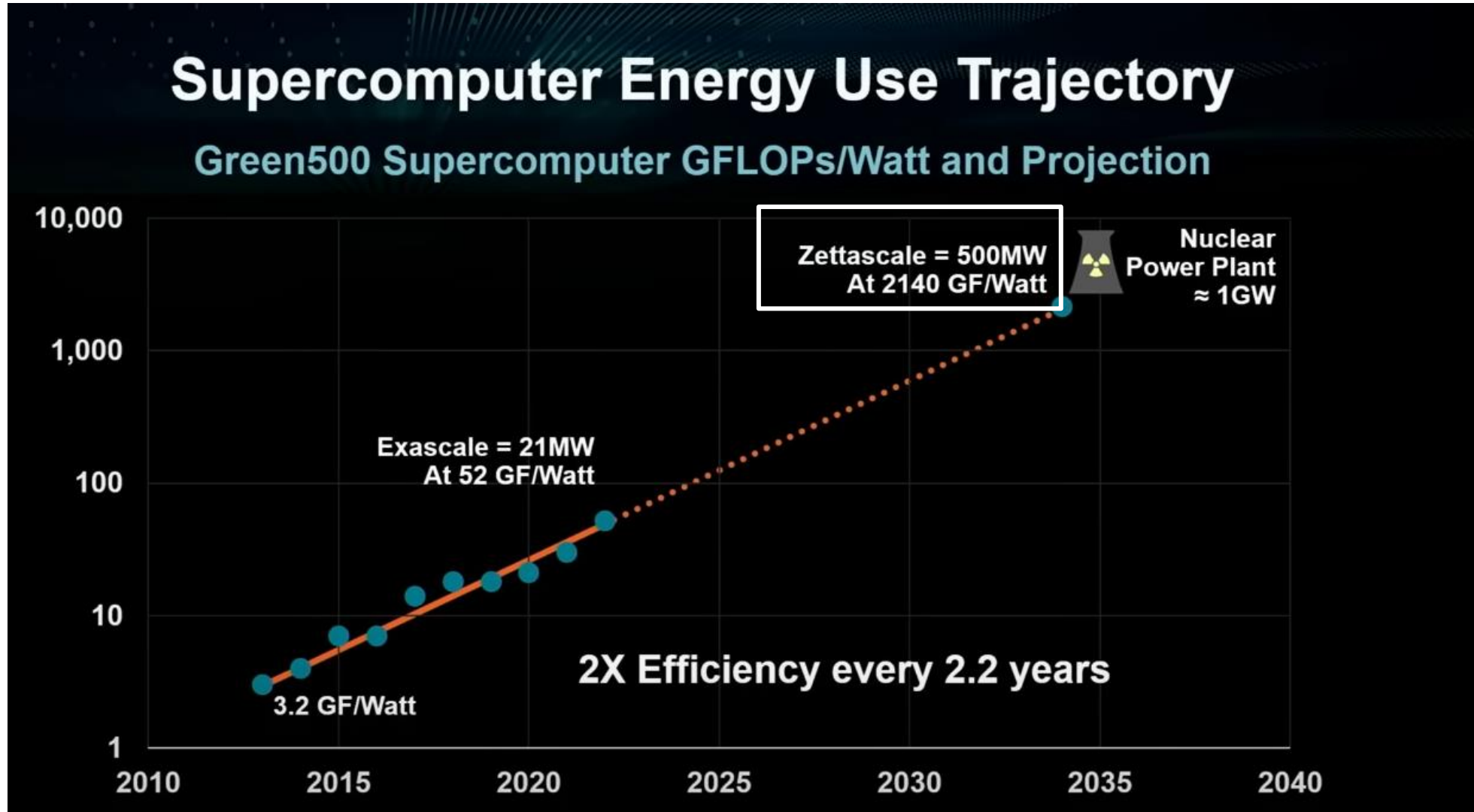
# 大綱





# 生成式AI雲端伺服器未來趨勢

- Zettascale的AI運算系統預計於2035到來，供應鏈需再有新技術產品



- 生成式AI引領互動式AI蓬勃發展，創新許多新應用
- 生成式AI帶動新AI運算模組硬體需求與技術研發
- AI運算模組供應鏈將迎來百億新商機
- Zettascale的AI運算系統預計於2035到來，擁有新技術將可掌握商機

