

半導體產業

-景氣將呈現緩慢復甦

陳俐妍

06/14/2023

群益投顧

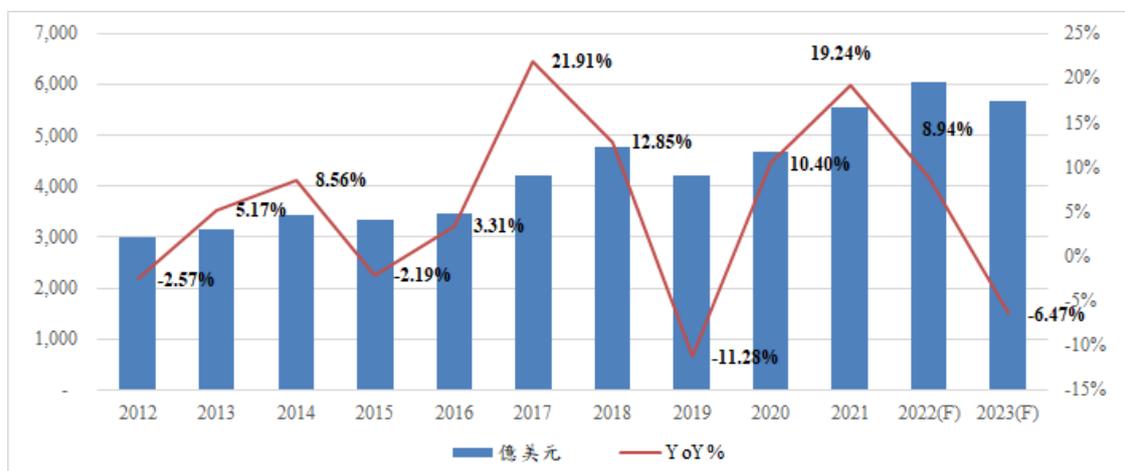


結論

- 總體環境負面因素未除，市場買氣不好，從系統廠到半導體晶片供應鏈者均調整庫存中，原預期調整庫存時間至1H23，但目前總體經濟不佳，庫存調整時間延長至3Q23，預估2023年半導體YoY-6.47%。
- 開始有少部分產品有急單與庫存回補需求，不過供應鏈仍積極把控庫存，IC設計業者也僅向晶圓代工廠維持低投片量。雖然3Q23客戶開始有新品推出，但3Q23晶圓代工和封測回升幅度將低於預期，期待4Q23。
- AI和車用將驅動半導體未來3~5年成長，高階製程需求不減，且晶圓廠/IDM 為延續摩爾定律，朝向2.5D、後段3D 及前段3D 技術研發。由於2.5D/3D 封裝大多採用高階先進製程晶片，因此晶圓廠較封測廠短期內有較大研發優勢，但封測廠技術亦逐步追趕中。
- 2023年半導體景氣不佳，所有半導體廠獲利皆下修，雖然產業接近落底，但底部要持續多長才是真正重點，惟有加強競爭力，成為產業前幾大，才能穩定生存。觀察重點以龍頭股台積電(2330 TT)和日月光(3711 TT)為主。亦可注意AI概念股創意(3443 TT)等個股。

2023年半導體負成長

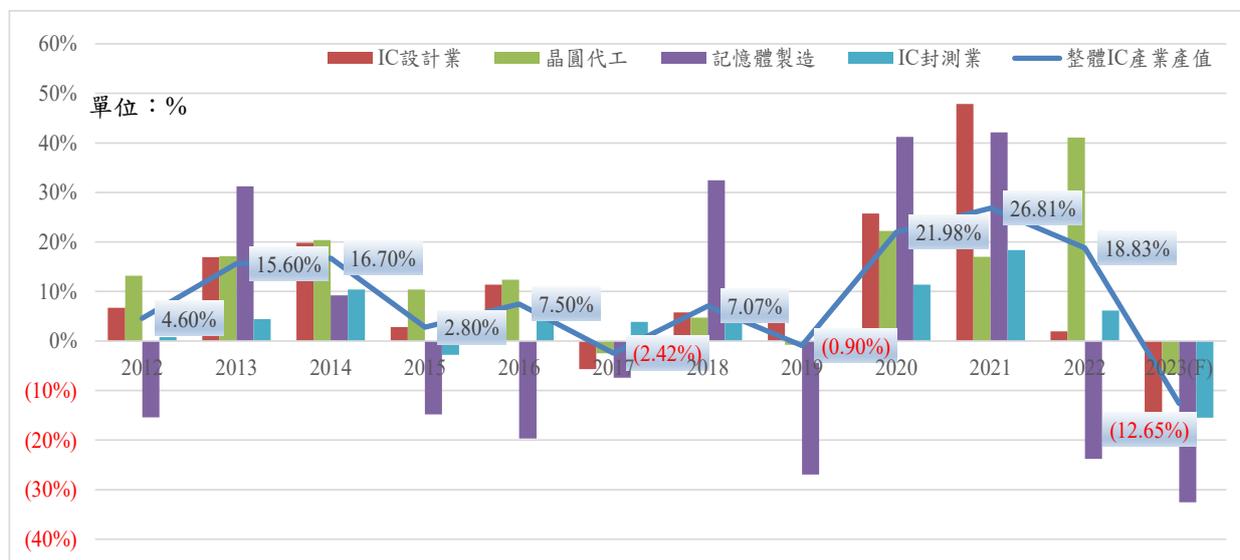
- 庫存去化與記憶體供過求將延續，特別是記憶體報價較2022年大幅下滑，預估2023年半導體YoY-6.47%。
- 目前終端需求不明，各家研究機構預估數差異大，但整體對2023年半導體前景較保守。



資料來源：IEK、WSTS、群益投顧預估彙整

台灣半導體2023年YoY-13%

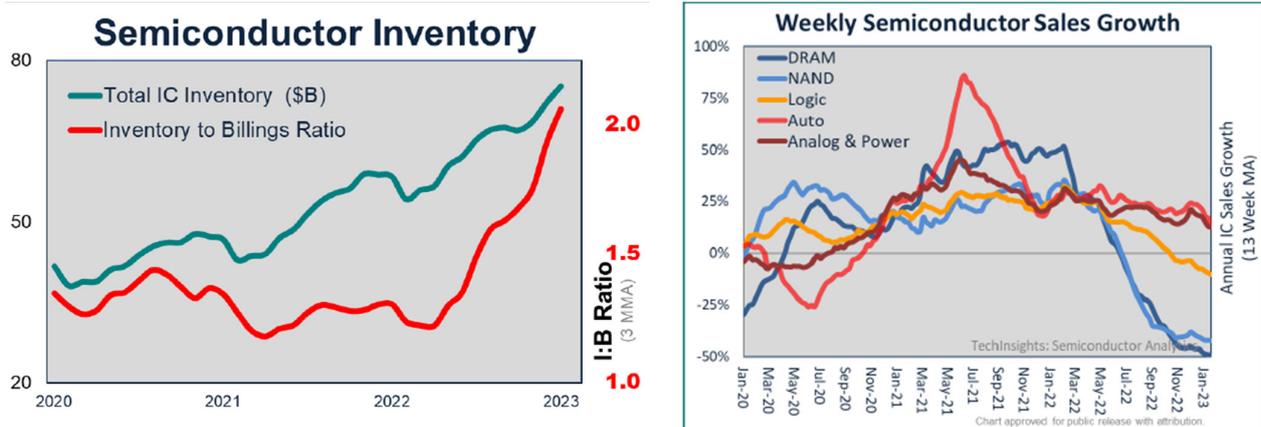
- 半導體產業進入庫存調整階段，IC設計與記憶體產業面臨需求滑落、供過於求的困境，分別YoY-20.3%、YoY-32.5%，晶圓代工也從持平成長下修至衰退，IC封測需求YoY-15.5%，預期台灣半導體2023年產值YoY-12.6%。



資料來源：MIC、群益投顧預估彙整

2H22~1H23 IC設計商庫存調整明顯

- 之前缺晶片，使廠商拉高庫存，再加上通膨等因素，使半導體庫存金額創新高。
- 隨著晶片ASP下滑，2H22晶片業者提列庫存損失，毛利率明顯受到影響，甚至4Q22~1Q23有廠商呈單季虧損，IC設計獲利下滑先行，上游製造廠商稼動率隨後下滑。

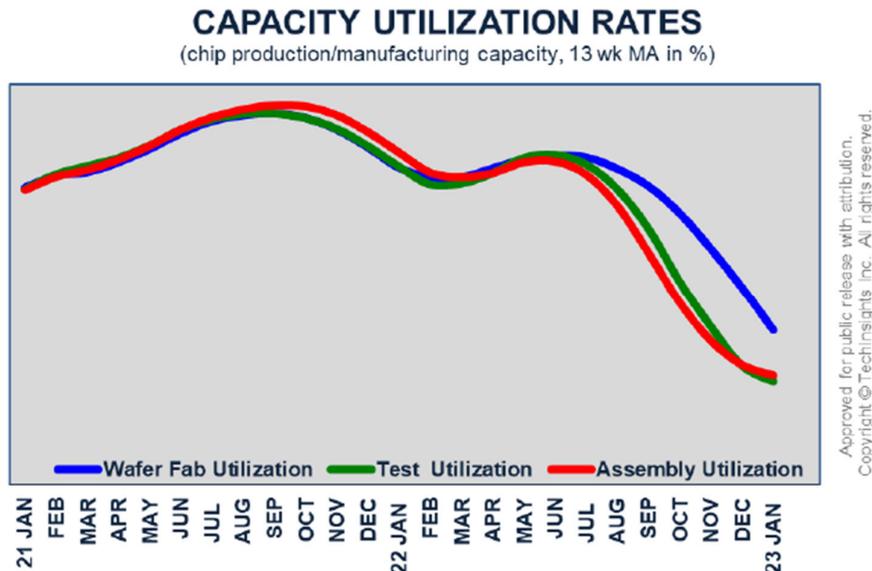


資料來源：Techinsights、群益投顧預估彙整

Capital Care 群益關心您

晶圓製造端產能利用率鬆動

- 晶圓製造端產能利用率4Q22開始鬆動，1Q23~2Q23稼動率下滑最明顯，目前雖樂觀看待2H23，但3Q23反彈幅度低於原先預期QoQ+20%。各晶圓廠開始下修2023年展望，目前期望4Q23可以大幅回溫。

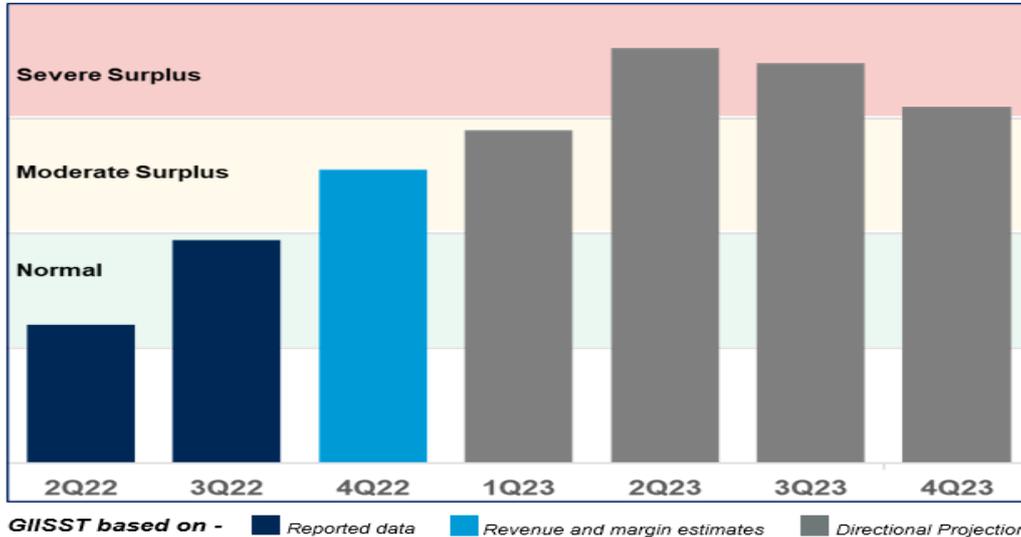


資料來源：Techinsights、群益投顧預估彙整

Capital Care 群益關心您

庫存時間調整較預期長

- 目前雖大多數晶片類別的庫存有改善，但經濟疲軟，原預期中國市場回溫落空，消費市場的半導體需求繼續惡化。非消費市場--包括網路、伺服器、存儲--在2023年很有可能出現類似趨勢。因此庫存時間延長，將調整至3Q23。

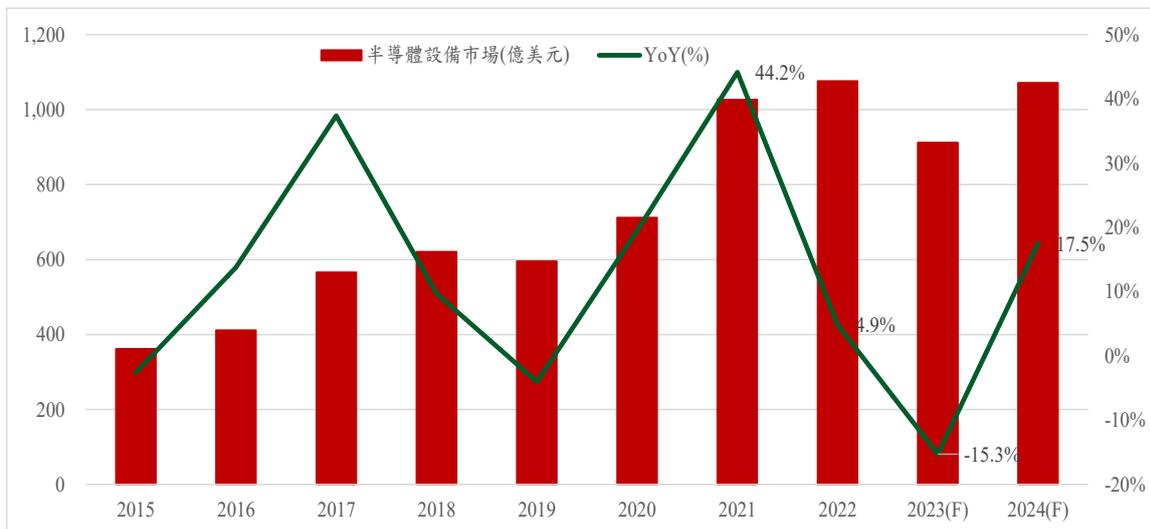


資料來源：Gartner、群益投顧預估彙整

Capital Care 群益關心您

2023年半導體資本支出YoY-15%

- 據SEMI，2023年半導體設備支出從2022年創紀錄1,076億美元下滑912億美元，YoY-15.3%，來自記憶體資本支出YoY-25%；其次則是美國對中國大陸半導體制裁，將使大陸半導體資本支出YoY-30%以上。

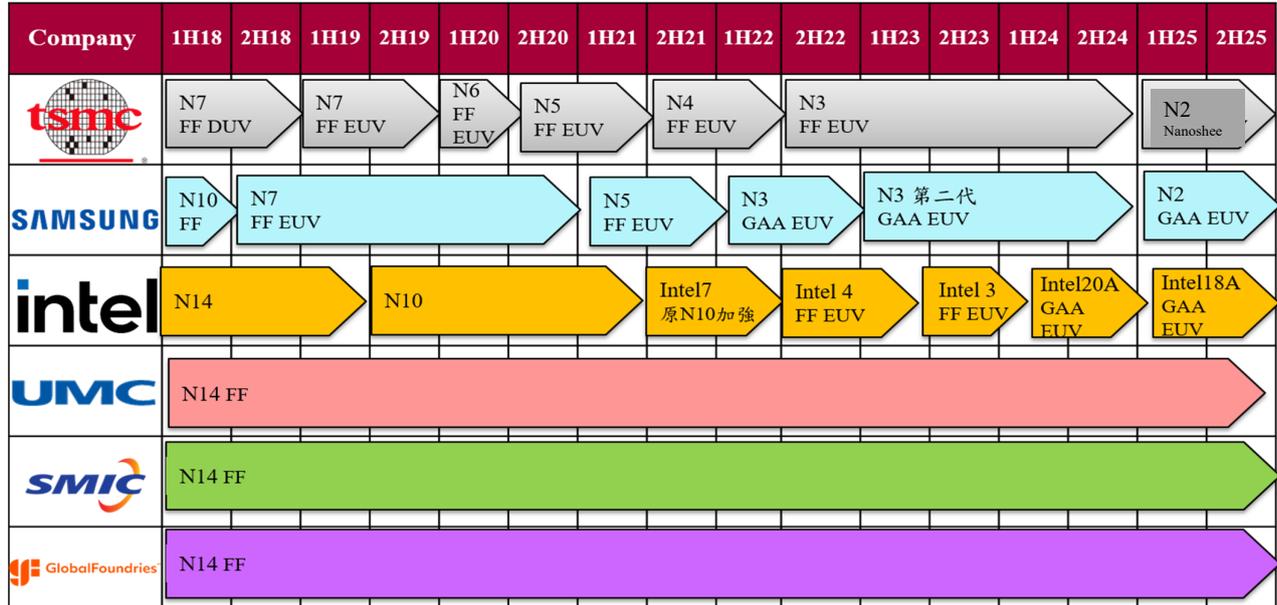


資料來源：SEMI、MIC、群益投顧預估彙整

Capital Care 群益關心您

競爭不因景氣差而延後

- N5以下製程，目前僅TSMC和Samsung能提供，Intel 緊追在後。台積電在2H22量產N3晶片，Samsung表示儘管目前先進製程落後台積電，不過N3就採用GAA製程，未來N2為關鍵轉折。



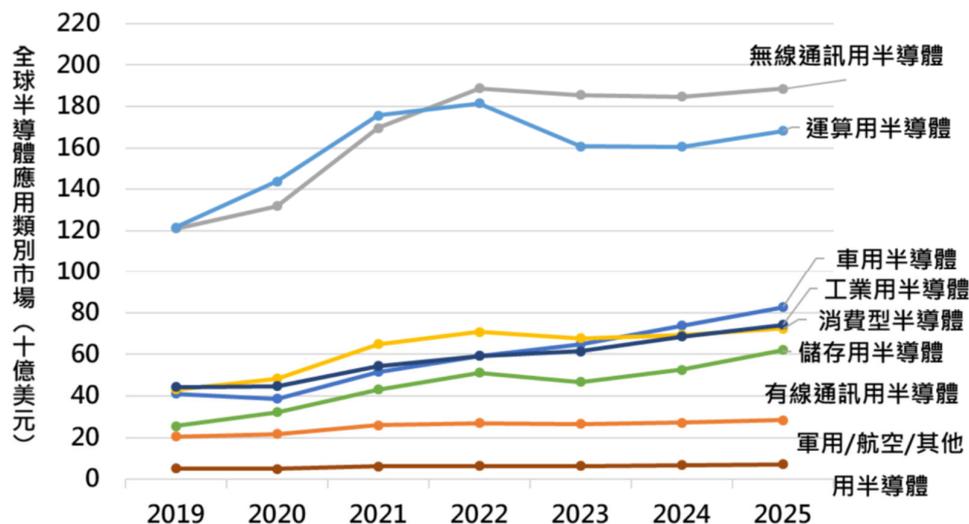
資料來源：群益投顧預估彙整

Capital Care 群益關心您

車用晶片走向SOC

車用晶片為半導體另一成長動能

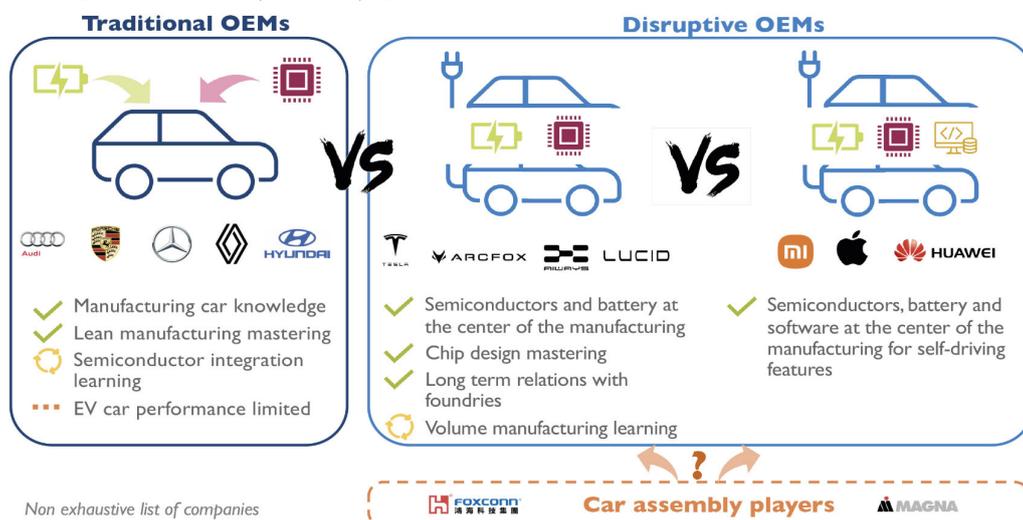
- 汽車使用半導體增多，是推升車用晶片成長的一大動力。不僅傳統燃油車的半導體含量持續增加，電動車的半導體含量是燃油車的3至5倍，隨著電動車滲透率提升，平均每輛車使用半導體的需求將強勁增長。車用半導體有可能成為半導體第三大應用。



資料來源：IEK、群益投顧預估彙整

汽車產業競爭變化中

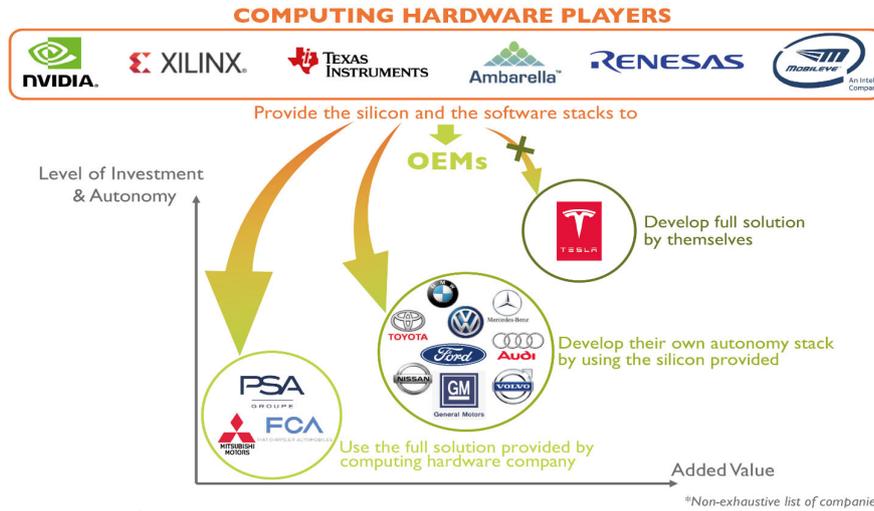
- 過去100多年，都是由OEM主導市場發展，但此局面正在轉變中。隨著電動車及自駕技術發展，OEM及Tier 1汽車零組件供應商投入，蔚來汽車、小鵬汽車以及Lucid Motors等新進電動車OEM廠商加入競局，並帶動半導體和消費性電子產業廠商搶進。如鴻海2021年積極與各傳統車廠、電動車新創企業等結盟，加速進軍電動車產業。



資料來源：yole、群益投顧預估彙整

More players, more chips

- Google, Amazon, and Baidu等大型科技公司開始開發汽車。這些公司完全是供應鏈的新成員。尤其是中電動汽車初創企業，走非傳統的供應鏈。像Tesla專業整車廠進行垂直整合，將整個傳統的供應鏈排除在外。
- 2020~2022年整車廠因晶片短缺，陷入減產、停工窘境，使整車廠開始改變供應鏈策略，除要求供應商增加晶片庫存外，也嘗試多種新興的供應鏈模式。



資料來源：yole、群益投顧預估彙整

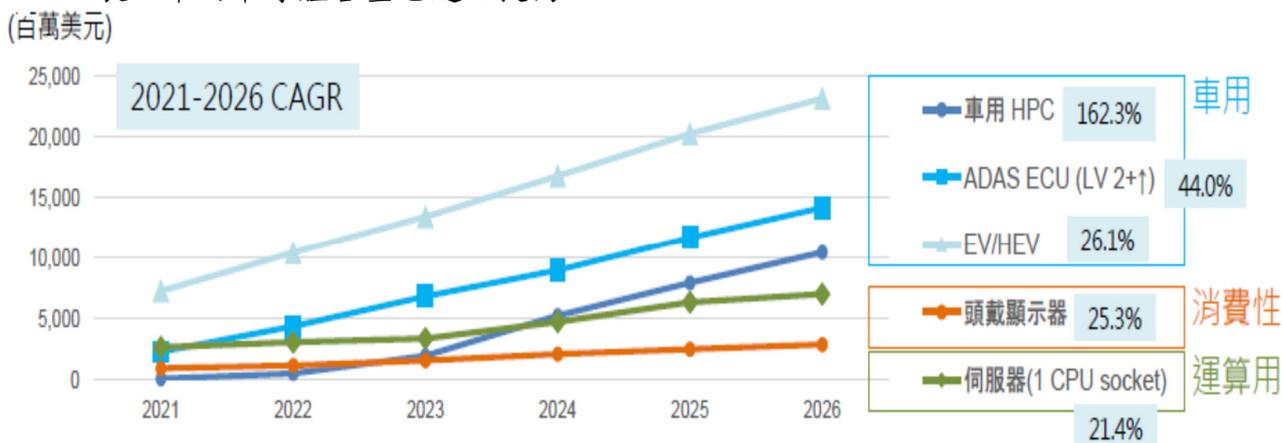
車用智慧晶片二大方向

自動駕駛	智慧座艙
<ul style="list-style-type: none"> ● 雷達、LiDAR、GPS及電腦視覺等技術來感測環境，以ADAS與車聯網等技術做為基礎，其中LiDAR、高精度地圖以及AI深度學習演算法與運算速度為重要關鍵。 ● 障礙物感測技術、動態定位技術、智慧決策與車輛動態控制技術。 	<ul style="list-style-type: none"> ● 在各國法規限制下，讓智慧座艙商業化進度領先自動駕駛，結合車輛攝影機、感測器、麥克風等元件收集來的資訊，透過情景感知能力提升駕駛的安全及體驗。 ● 整合多個螢幕顯示(中控、儀錶、抬頭等)、駕駛人監控、車聯網、娛樂系統及部分輔助駕駛功能技術。

資料來源：IEK、EE times、群益投顧預估彙整

車用 HPC 成長動能大

- 電動車的晶片需求與傳統油車最大的差別在使用先進製程晶片的比例，燃油車只需5% 先進製程晶片，電動車先進製程晶片比例則大幅提升到50%。
- 車用系統未來發展的方向，將會從原先充滿ECU的架構，整合為單一HPC來處理整輛車的資訊。
- 車輛電子化與智慧化趨勢不變，帶動ADAS、車用HPC高性能運算、EV/HEV成長，車內半導體含量也隨之提高。



資料來源：IEK、群益投顧預估彙整

Capital Care 群益關心您
台北 · 香港 · 上海

智慧座艙SoC

- 主流的智慧座艙SoC中，高通旗下820A與8155晶片獲得最多的市佔率，其次為Renesas與NXP。
- 智慧座艙SoC支援Apple和Google的手機汽車互動平台(CarPlay和Android Auto)。用於車內多台顯示器互動、全車語音交互、手勢控制、人臉識別等需求。

公司	華為		高通				NXP	Renesas	
產品	麒麟710A	麒麟990A	820A	8155	8195	8295	LMX8	R-CAR M3	R-CAR H3
CPU 運算力 (DMIPS)	NA	NA	45k	105k	150k	200k	29k	28k	43k
GPU 運算力 (GFLOPS)	NA	NA	588	1142	2100	3000	128	76	288
製程(nm)	14	28	14	7	7	5	16	28	16
量產時間	2020	2022	2019	2020	2022	2023	2019	2018	
合作車廠	北汽、AITO		Audi、賓士、通用、蔚來、小鵬、集度、吉利、長城、上汽、理想等				福特、榮威等	豐田、日產、福斯等	

資料來源：各業者、Wikichip、DIGITIMES Research整理、群益投顧預估彙整

Capital Care 群益關心您
台北 · 香港 · 上海

全球ADAS SoC Mobileye市佔過半

- L2級ADAS對運算力需求較低(約10TOPS)，Mobileye即可滿足，同時憑藉自動駕駛領域先行者優勢，迅速搶奪市場。目前Mobileye市佔率過半，其後依序為Xilinx、NVIDIA、Tesla。ADAS等級向L2以上邁進，ADAS SoC晶片運算力需求將逐步提高，NVIDIA Orin有較強的運算力，運算力/功耗比在眾晶片中脫穎而出，助NVIDIA提升市佔率。

產品	Tesla		NVIDIA			Mobileye			
	FSD 1.0	FSD 2.0	Xavier	Orin	Thor	EyeQ5	EyeQ6L	EyeQ6H	Eye Ultra
運算力 (TOPS)	72	216	30	200	2,000	24	5	34	176
功耗(W)	36	90	30	60	NA	10	3	13	NA
運算力/功耗	2	2.4	1	3.3	NA	2.4	2	2.7	NA
製程(nm)	14	7	12	7	5	7	7	7	5
量產時間	2019	2023	2020	2022	2025	2021	2023	2024	2025
合作車廠	自供		賓士、富豪、蓮花、廣汽、蔚來、小鵬、理想等			Audi、BMW、豐田、通用、蔚來、小鵬、理想、長城、吉利等			

資料來源：各業者、Wikichip、DIGITIMES Research整理、群益投顧預估彙整

中國ADAS SoC

- 中國車廠預先搭載高運算力硬體，以保證後續軟體更新時不受硬體性能限制。
- 黑芝麻在產品研發速度與投入程度僅次於地平線，然地平線可提供更豐富的軟體開發工具組，更容易獲得中國整車廠青睞。華為受限於美國管制無法在台積電投片，市場影響力逐漸式微。

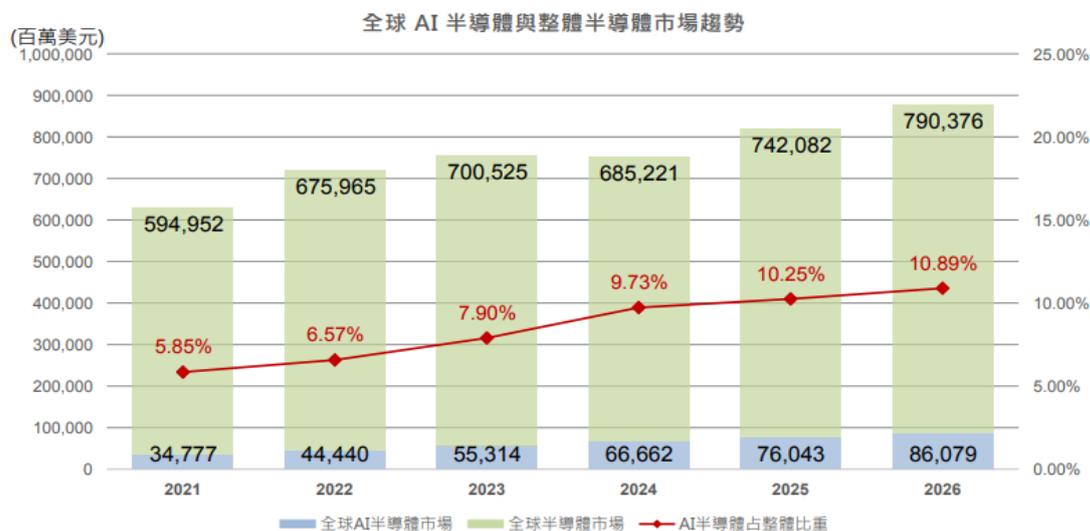
公司	華為			地平線			黑芝麻		
	Ascend 310	Ascend 910	Ascend 610	Journey3	Journey5	Journey6	A1000	A1000 pro	A2000
運算力 (TOPS)	16	512	200	5	128	400	58	106	250
功耗(W)	8	310	60	3	30	NA	18	25	NA
運算力/功耗	2	1.7	3.3	1.7	4.3	NA	3.2	4.2	NA
製程(nm)	12	7	7	16	7	7	16	16	7
量產時間	2018	NA	NA	2020	2022	NA	2020	2022	NA
合作車廠	北汽、廣汽、長城、長安			上汽、比亞迪、理想、長安、東風、哪吒等			上汽、一汽、東風、江淮等		

資料來源：各業者、DIGITIMES Research整理、群益投顧預估彙整

AI晶片商機大

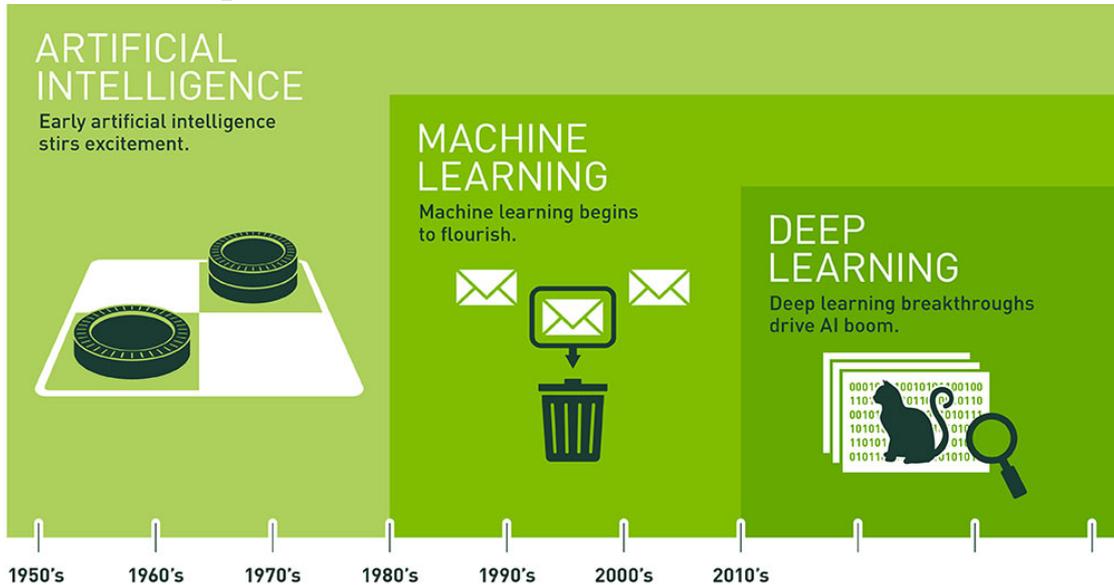
AI晶片市場快速成長

- 預估2026 年全球 AI 半導體 市場預估超過 860 億美元，佔全球半導體產值近 11%。
- 全球 AI 半導體 市場 2021~2026 CAGR: 19.9%
- 全球半導體市場 2021~2026 CAGR: 5.8%。



人工智慧可以追溯到1950年代

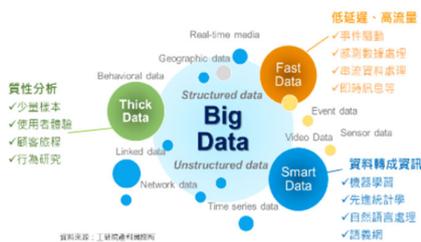
- 人工智慧 (AI) 是讓機器展現人類的智慧。機器學習是透過從過往的資料和經驗中學習，也是narrow AI，從資料中得到複雜的函數(或樣本)來學習以創造演算法或規則，並利用它來做預測。深度學習（深度神經網路）是讓電腦可以自行分析資料找出「特徵值」。



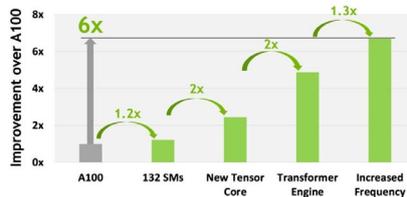
資料來源：NVIDIA、群益投顧預估彙整

數據、算力、算法發展使AI逐步實現

DATA

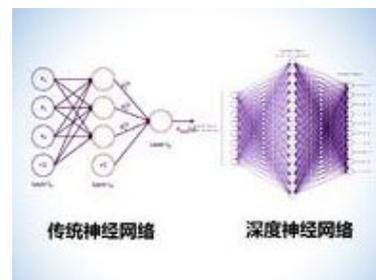


Computer power



晶片效能提升，運算速度大幅成長。此外硬體成本也在下降中。

Algorithm



深度學習DNN、RNN、CNN、到GAN.....不停的有新的演算模式發明，且正確率提升。

資料來源：nVidia、百度、群益投顧預估彙整

AI模型每2~3年級數成長

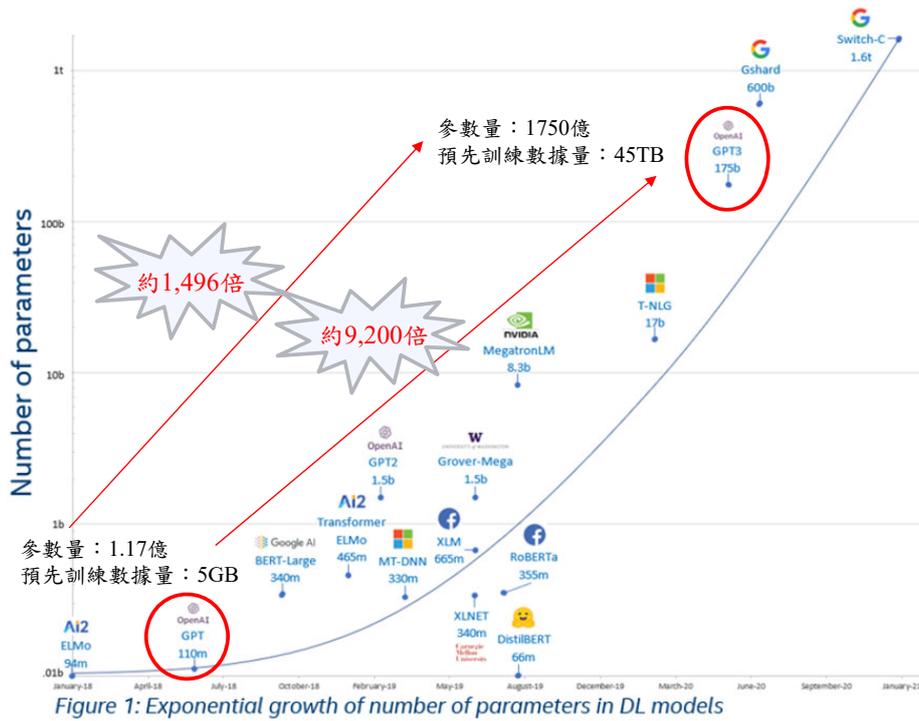
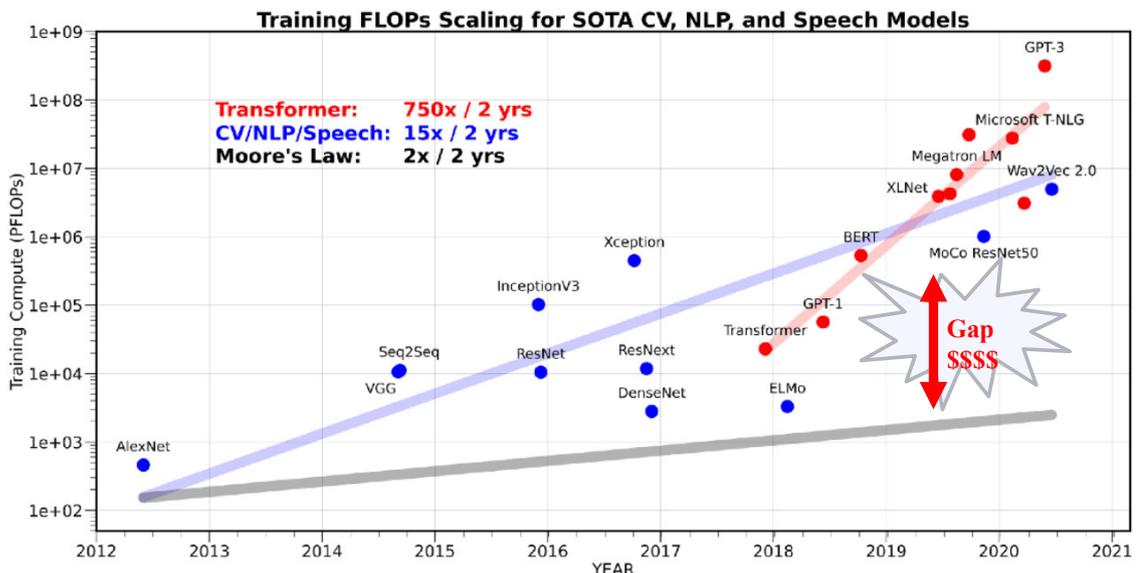


Figure 1: Exponential growth of number of parameters in DL models

資料來源：網管人、smarter.ai、群益投顧預估彙整

晶片效能提升落後AI模型的發展

- 計算機視覺(CV)，自然語言處理(NLP)和語音辨識領域以大約每兩年翻15倍數的速度在增長。Transformer類的模型運算量的增長則更大，約為每兩年翻750倍。但晶片效能提升落後AI模型的發展。



資料來源：onflow、群益投顧預估彙整

半導體受惠算力需求增加

實際總算力

$$= \boxed{\text{單一處理器的性能}} \times \boxed{\text{處理器的數量}} \times \boxed{\text{利用率}}$$

- § **提升單晶片性能**：先進製程、Chiplet封裝及架構/微架構創新。(受惠者為先進製程、封裝，以及封裝材料和設備等)。
- § **晶片數量規模**：首先要考慮的不是簡單的建設資料中心，買更多的伺服器，而是要考慮晶片是否能夠支援和適合。例如，受AI演算法快速多變和演算法眾多，晶片要具擴展性、可編程性、靈活性或通用性等。(ASIC、DSA等受惠)
- § **提升算力利用率**：考慮的重心不是單一晶片的利用率，而是整體資源利用率，就是把所有的計算資源連成一個大的計算資源池，然後可以非常靈活的資源切分、組合、分配和回收。挑戰就是如何把種類繁多的異質的計算資源彙集和溝通到一個資源池。(高速傳輸介面需求增加)

資料來源：群益投顧預估彙整

Capital Care 群益關心您
台北 · 香港 · 上海

先進製程使AI晶片更節能

- 雖然AI晶片應用於資料中心和其他HPC時，並沒有傳統行動設備方面的限制(如電池壽命、便攜性)；但其AI晶片仍需更小，更密集、更節能。
- 效率也可想成一個晶片的平均成本，除生產成本外也要考量運作成本，即晶片運作時的能源耗費。

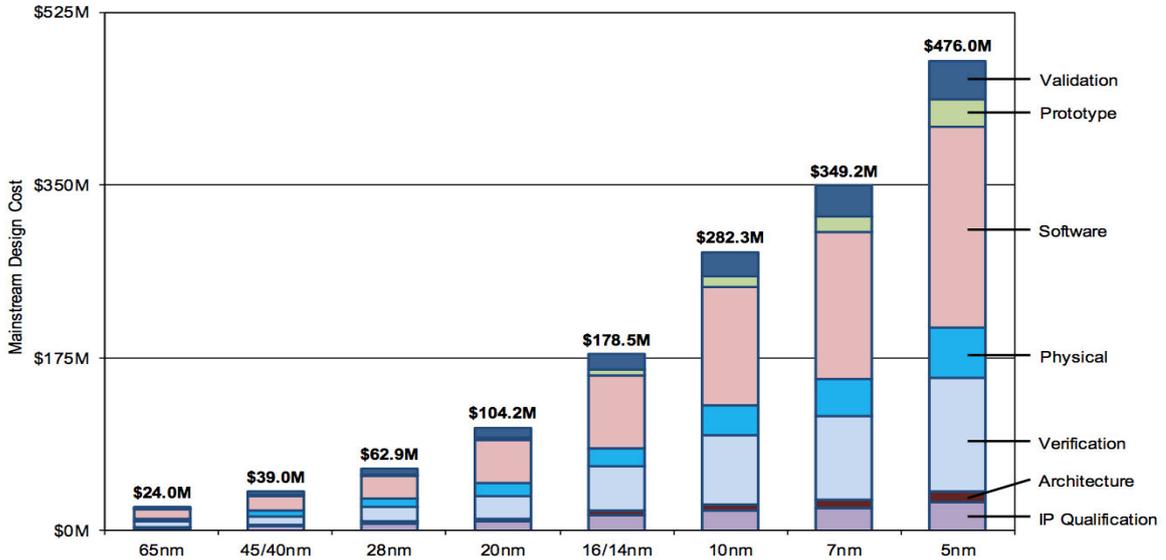


資料來源：工研院產科國際所、群益投顧預估彙整

Capital Care 群益關心您
台北 · 香港 · 上海

先進製程設計成本倍數成長

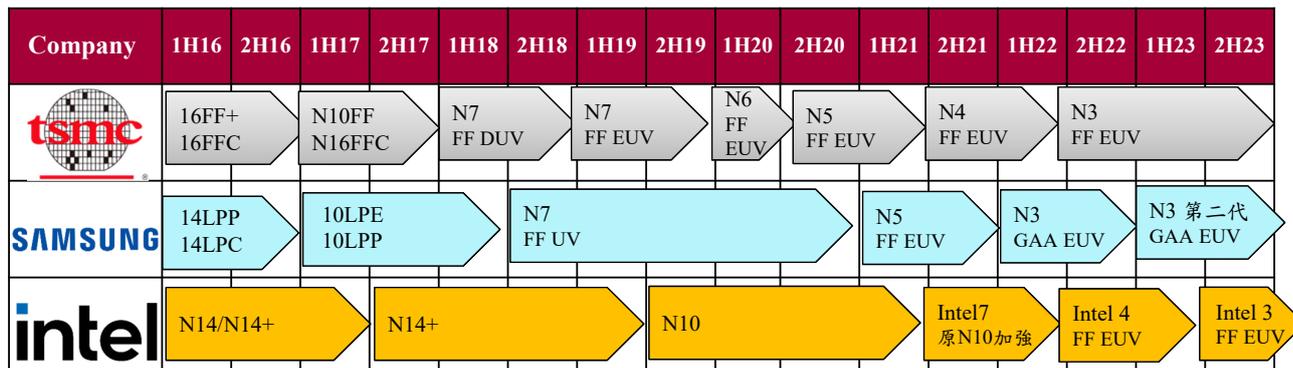
- 28nm IC設計成本為\$62.9 M，7nm成本增加至\$349.2M，5nm製程設計成本為\$476.0M。
- 最大的部分是Software，Verification和Validation。



資料來源：International Business Strategies、群益投顧預估彙整

先進製程晶圓ASP高

製程	N28	N16/N12	N7	N5	N4	N3
美元/wafer	3,000	5,000	10,000	16,000	18,000~20,000	20,000+



資料來源：群益投顧預估彙整

台積電N2 預計 2025 年小量投產

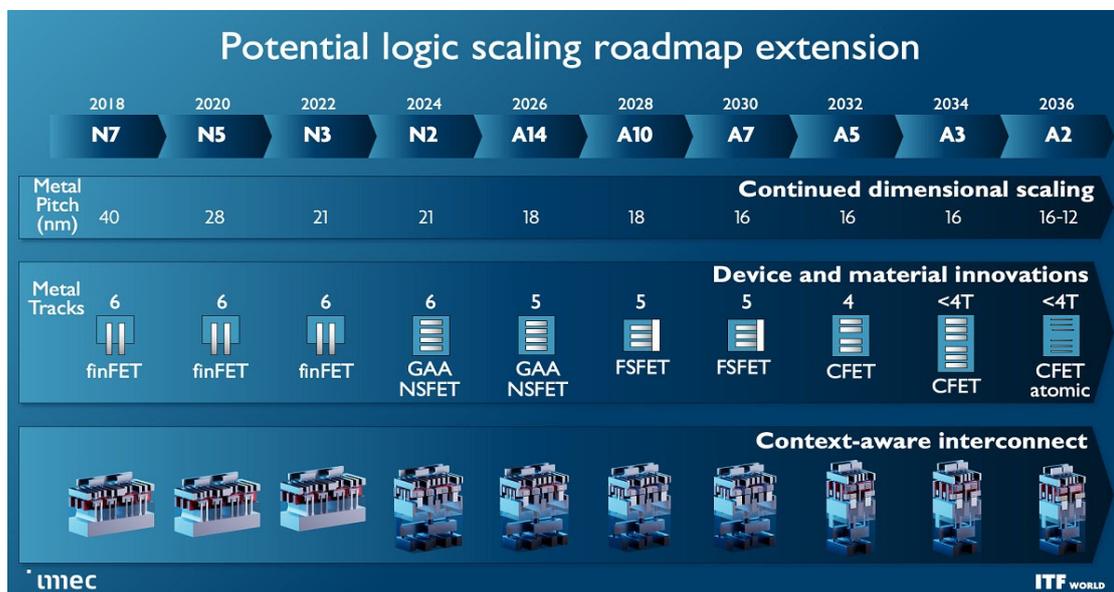
- 先進製程技術從N10推進至N2，晶片效能在約十年間以15%的年複合成長率增加。
- 台積電先進製程技術產能年複合成長率在2019年至2023年間將超過40%。



資料來源：TSMC、群益投顧預估彙整

FinFET 電晶體將於N3達盡頭

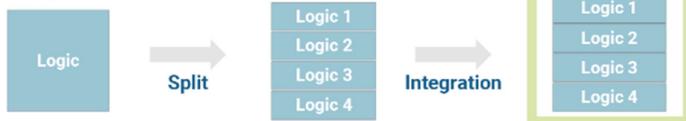
- 比利時微電子研究中心 (IMEC) 發表N1以下製程藍圖，FinFET 電晶體將於N3達盡頭，然轉換到 Gate All Around (GAA) 電晶體，2024 年進入量產，之後還有FSFET 和 CFET 等。



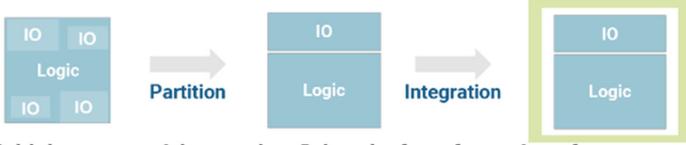
資料來源：IMEC、群益投顧預估彙整

More than Moore

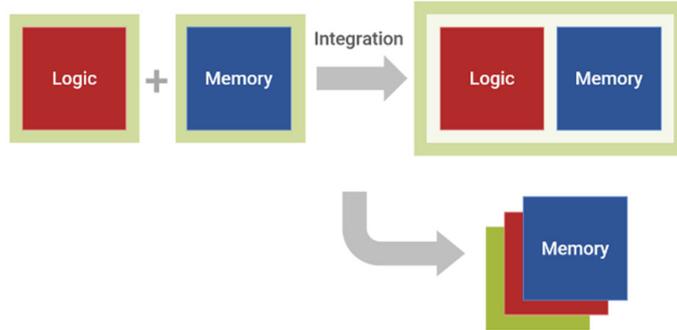
1. Chip split & integration: Driven by cost and yield



2. Chip partition & integration: Driven by cost and technology optimization



3. Multiple systems & integration: Driven by form factor & performance

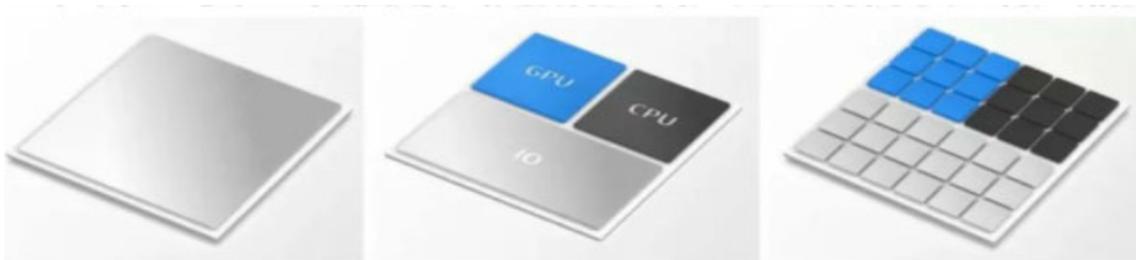


- 摩爾定律逐漸出現瓶頸，晶圓廠/IDM 為延續摩爾定律，朝向2.5D、後段3D 及前段3D 技術研發。

資料來源：TSMC、群益投顧預估彙整

SoC、SiP和Chiplet比較

- SoC拆解成GPU、CPU、IO晶片，通過SiP技術將它們集成在一個封裝內；通過Chiplet技術，小區塊擁有單獨的IP，並可重複使用，據特定客戶獨特需求定製的產品。



SoC

- 在 SoC層面驗證
- 3~4年開發時間
- 晶片中發現數百個缺陷
- 無法重複使用

SiP

- 在 SiP層面驗證
- 2~3年開發時間
- 晶片中發現數十個缺陷
- 部份可重複使用

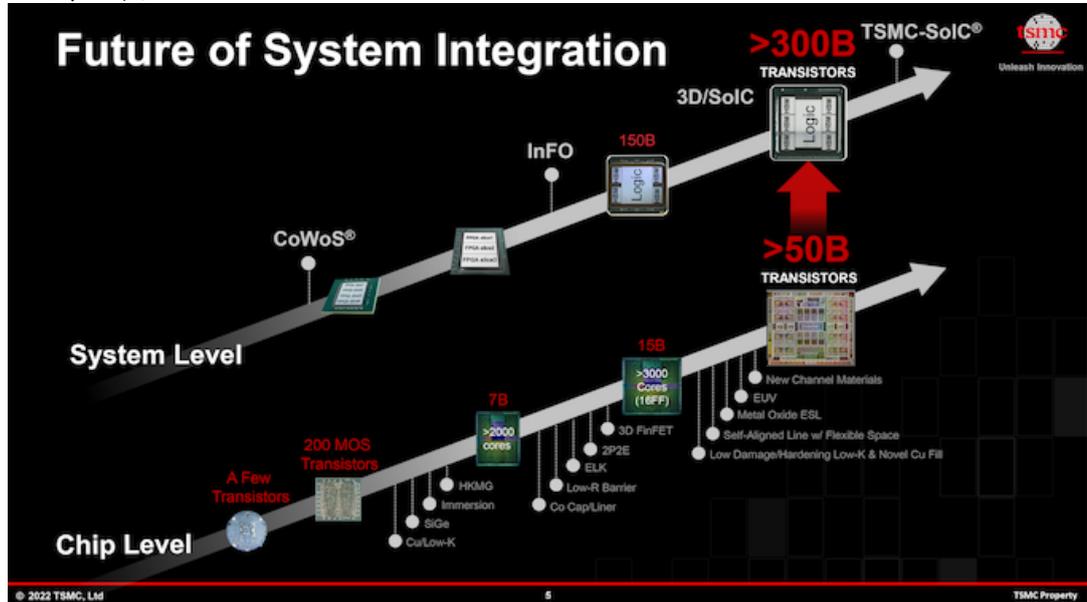
Chiplet

- 單獨IP集成Chiplet
- 1~2年開發時間
- 晶片中發現<10個缺陷
- 大都可重複使用

資料來源：群益投顧預估彙整

multi-tile Chiplet designs 比例增加

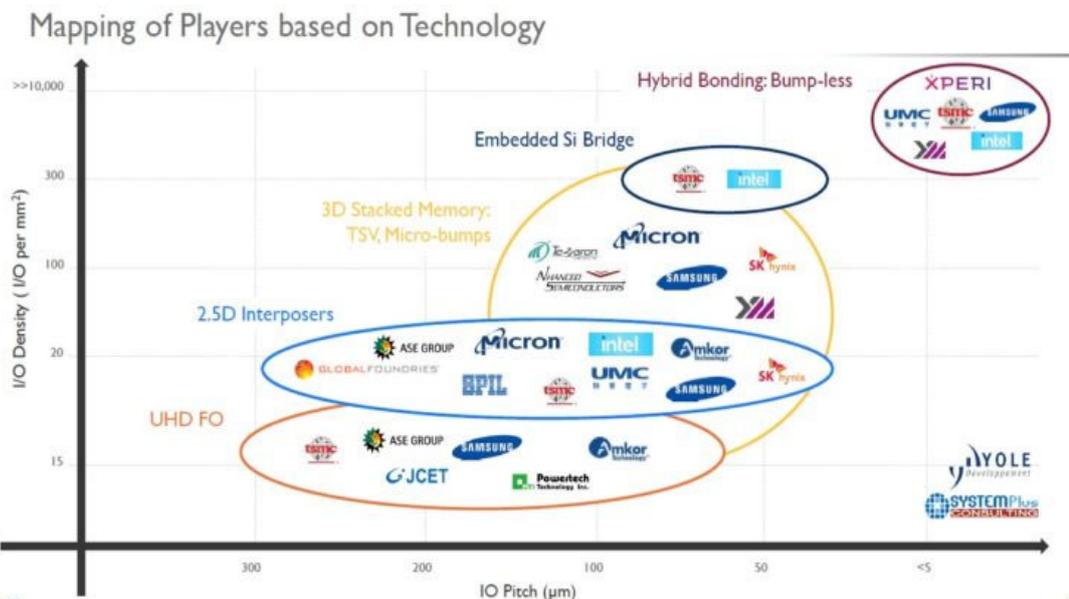
- 提高處理器計算能力的方法之一是增加其transistor count；AI/HPC 運算晶片採用multi-tile chiplet designs比例增加，Intel Ponte Vecchio GPU採用47個tile。



資料來源：TSMC、群益投顧預估彙整

IDM和封測廠都有相關技術

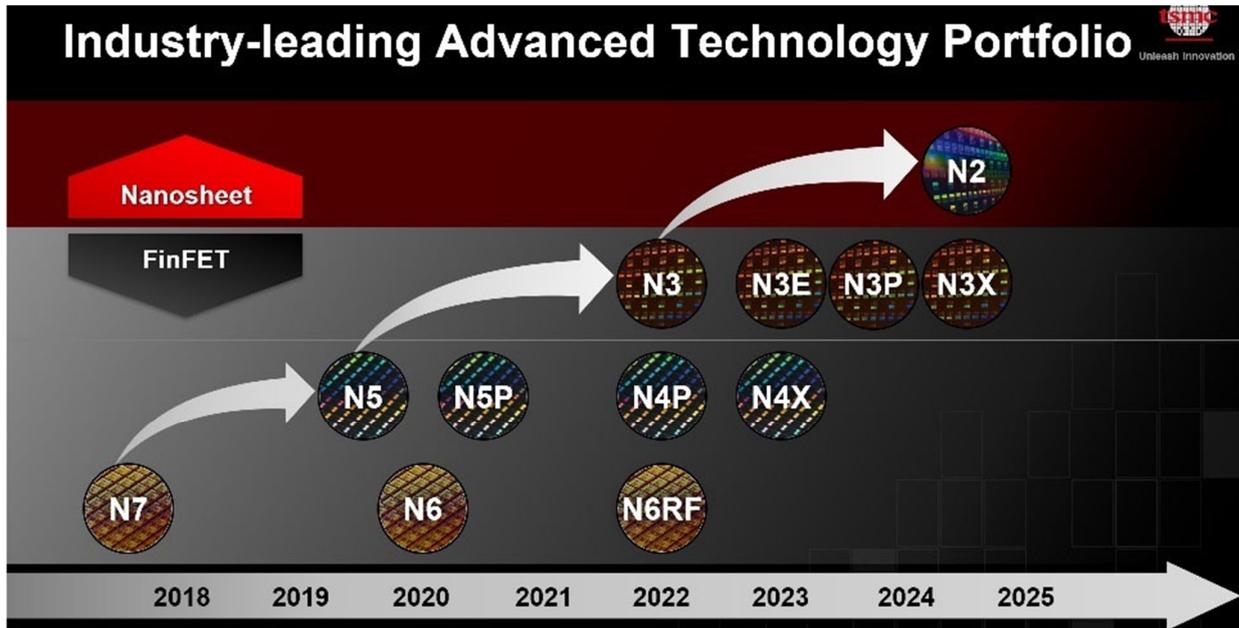
- 2.5/3D封裝相關技術發展多年，台積電2012年就推出CoWoS，很多廠商也都宣布有相關技術，但並沒有對外提供相關服務，如：你無法請GlobalFoundries幫忙代工silicon interposers，不過可看出這是未來重要的趨勢，各廠積極加入。



資料來源：Yole、群益投顧預估彙整

先進封裝加速晶片上市

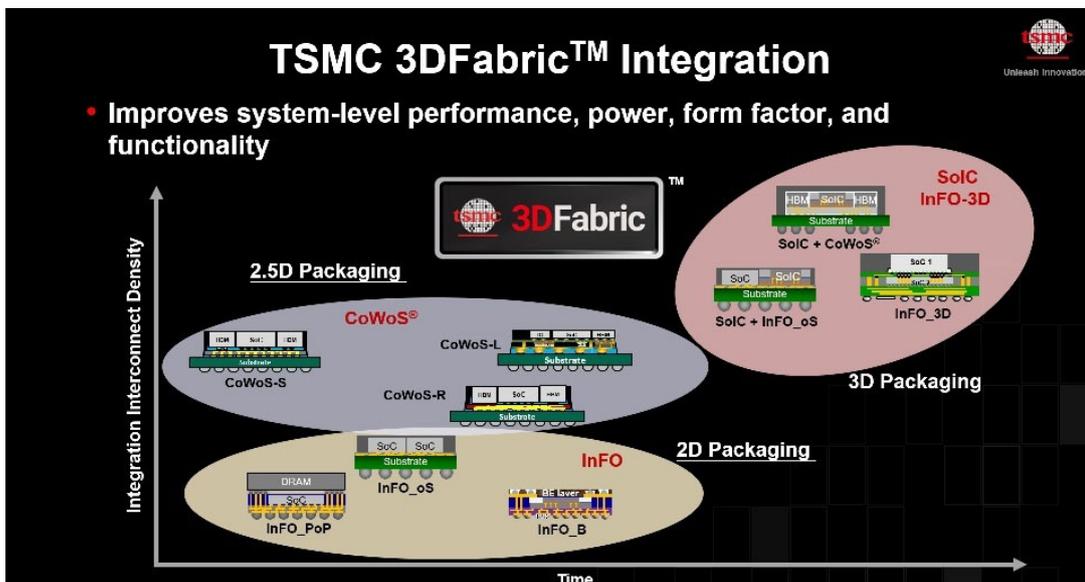
- 台積電從7nm晶片的量產到N7-on-N7封裝走了4年，5nm到N5-on-N5間用了3年，而3nm和N3-on-N3之間預計只有12個月。



資料來源：TSMC、群益投顧預估彙整

TSMC 3DFabric

- 台積電3D Fabric 系統整合技術包含各種先進的3D矽堆疊和先進封裝技術，以支援廣泛的次世代產品。



資料來源：TSMC、群益投顧預估彙整

TSMC 3DFabric聯盟

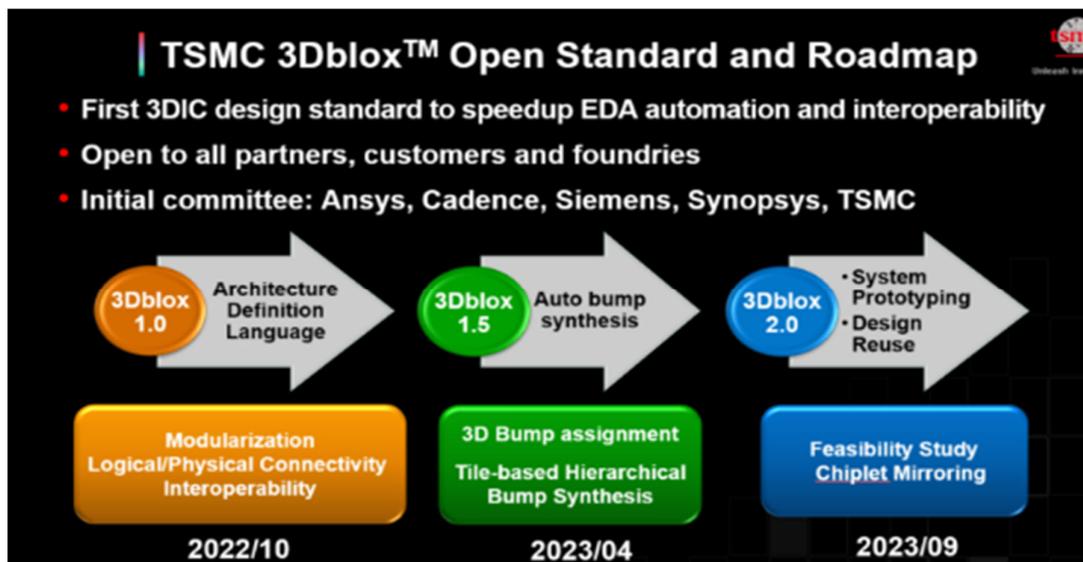
- 台積電宣佈成立新的3DFabric™聯盟，這是繼IP、EDA、DCA、Cloud和VCA聯盟之後的第6個OIP聯盟。
- 積電3D Fabric聯盟參加的合作夥伴包括Advantest、世芯、Alphawave、Amkor、Ansys、Arm、日月光、Cadence、創意、IBIDEN、美光、Samsung Memory、西門子、Silicon Creations、矽品、SK hynix、新思科技、Teradyne、欣興等。

EDA	IP	DCA/ VCA	Memory	OSAT	Substrate	Testing
取得台積電3DFabric技術進行EDA開發與升級，優化EDA工具及設計流程，提升3D IC設計效率	開發符合晶片對晶片介面標準，及台積電3DFabric技術的3D IC，為客戶提供經驗證的IP解決方案	透過設計中心聯盟(DCA)及價值鏈聚合(VCA)設計服務在3DFabric技術及與客戶合作，與台積電達路線圖共識，提升設計、IP整合、生產的服務	及早進行技術合作，定義規格並與台積電協調工程和技术標準，將縮短未來高頻寬記憶體(HBM)世代產品的上市時間	支援台積電生產品質及技術要求，持續改善各種技術與生產支援，和台積電合作滿足客戶的生產需求	及早進行技術合作與開發，滿足3DFabric技術的未來要求，提升載板材料的品質、可靠性及新基板的整合性	為台積電3DFabric技術開發測試與壓力研究方法提供可靠性及品質要求，協助客戶快速推出具差異化的產品
Ansys cadence SIEMENS SYNOPTIS	Alpha arm cadence Silicon CREATIONS SYNOPTIS	aichip GUC	Micron SAMSUNG Memory SK hynix	Amkor Technology ASE EPL	IBIDEN Unimicron 欣興電子	ADVANTEST TERADYNE

資料來源：TSMC、MIC、群益投顧預估彙整

3Dblox 業界首個3D IC設計標準

- 台積電3Dblox是業界首個3D IC設計標準，以加快EDA自動化和互通性。3Dblox開放標準、實現記憶體和台積電邏輯間的緊密合作，以及將基板和測試合作夥伴納入生態系統。台積電正在開發3Dblox 2.0，以實現系統原型開發和設計重用。



資料來源：TSMC、群益投顧預估彙整

3Dblox模組化設計

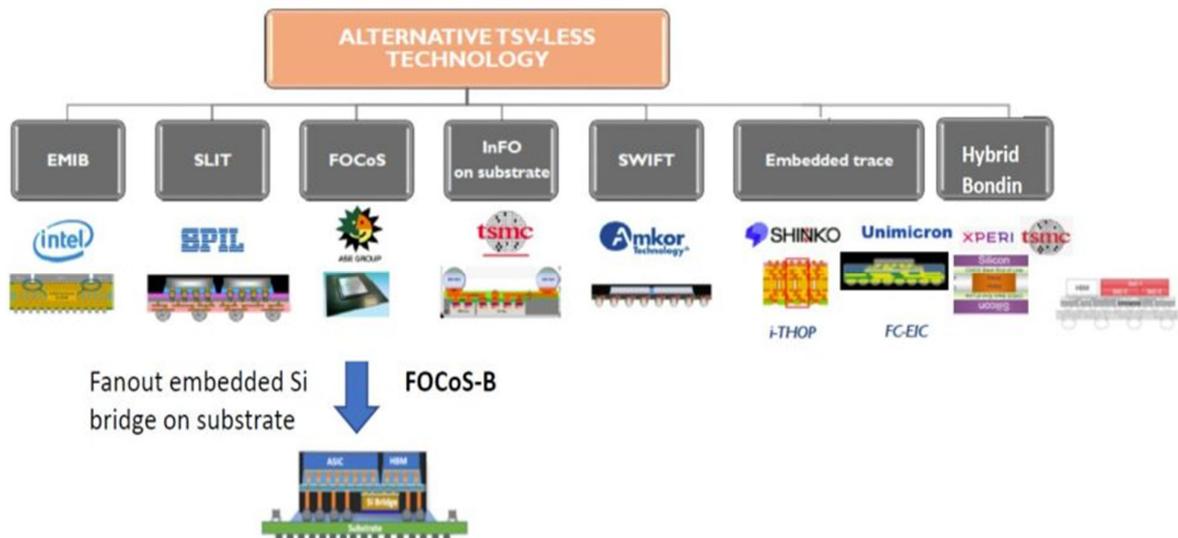
- 3Dblox旨在使複雜的2.5D和3D系統由上而下的模組化設計更加容易，同時也促進晶片重複使用。作為設計資料的標準化介面格式，3Dblox讓台積電的客戶能更容易充分利用台積電3DFabric技術下的許多技術方案，包括CoWoS、InFO、TSMC-SoIC等。



資料來源：TSMC、群益投顧預估彙整

封測廠2.5D/3D業務將往上

- 封測廠發展2.5D/3D的技術至少7~8年以上，各家的解決方案不同，但在台積電開始提出3D IC設計標準後，整個供應鏈有相對應的參考，封測廠有機會用優勢性價比，取得相關業務。



資料來源：Yole、群益投顧預估彙整

UCIe將成為Die2Die主流傳輸標準

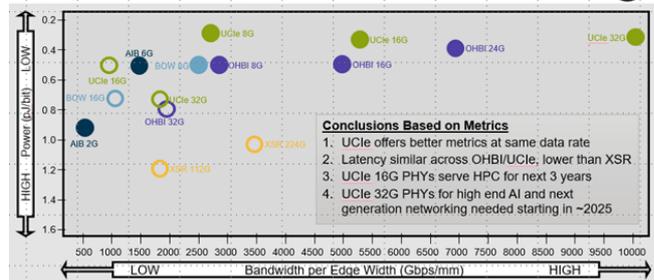
- 互通性和相容性取決於供應鏈的許多層次和環節達成協議。以往零散的要求導致大量的解決方案。標準化使專業分工容易，且可以降低成本。

Standards	XSR	BOW	OpenHBI	UCIe
IO PHY	Yes	Yes	Yes	Yes
Max Data Rate	112G	8G/16G	16G /32G	32G
Channel Pin/Lane	16Tx+16Rx	32 (16Tx+16Rx) /Slice	42 (Bi-Dire)/ DWord	16Tx+16Rx or 64Tx+64Rx
IO Swing	0.75Vpp	0.75V	0.4V	0.4V or 0.7V
RX Termination	Required	Optional	No	Optional
IO Direction	Uni	Uni	Bi	Uni
Pad Cap	0.13pF	0.2pF	0.35pF	0.25pF
Noise Reduction	Scrambling	DBI	DBI	Scrambling
Redundant Pin/Lane	No	1/Slice	2/DWord	4/64 Pins
Logical PHY	Yes	No	Yes	Yes
Training	Yes	No	Yes	Yes
Initialization	Yes	No	Yes	Yes
Sideband	No	No	Yes	Yes
Link Control	No	No	No	Yes
System Interface (PHY)	64bit@1.75G Hz	Direct @ 8GHz/16GHz	Geared @ 2GHz	RDI @ 2GHz
Packaging	Standard	Standard	No	Standard
	Advance	Advance	Advance	Advance
Protocol (Controller)	No	No	No	Yes

- Chiplet D-to-D主流傳輸標準包括XSR、BOW、OpenHBI、UCIe。綠色代表技術優勢，紅色代表劣勢。UCIe具有完整磊晶之間介面堆疊的標準，其他標準大多僅關注在特定層。且UCIe支援2D、2.5D和橋接封裝，預計未來還會支援3D封裝。

Summary of 5 Main D2D Standards

Key Figures of Merit: Technical (Bandwidth, Power, Latency) & Cost



資料來源：Yole、群益投顧預估彙整

Die-to-Die介面矽智財IP技術

- IP和IC設計服務公司也積極開發Die-to-Die介面矽智財IP技術。

GUC 2.5D and 3D Multi-die Advanced Packaging Technology Platform

GUC
The Advanced ASIC Leader

Platform Highlights

- HBM3 PHY + Controller IP
- Die-to-die IP (GLink-2.5D)
- Die-on-die IP (GLink-3D)
- High speed IP integration (112G SerDes, PCIe-5, GDDR6)
- Advanced Packaging Technology (CoWoS / InFO / 3D-SoIC)

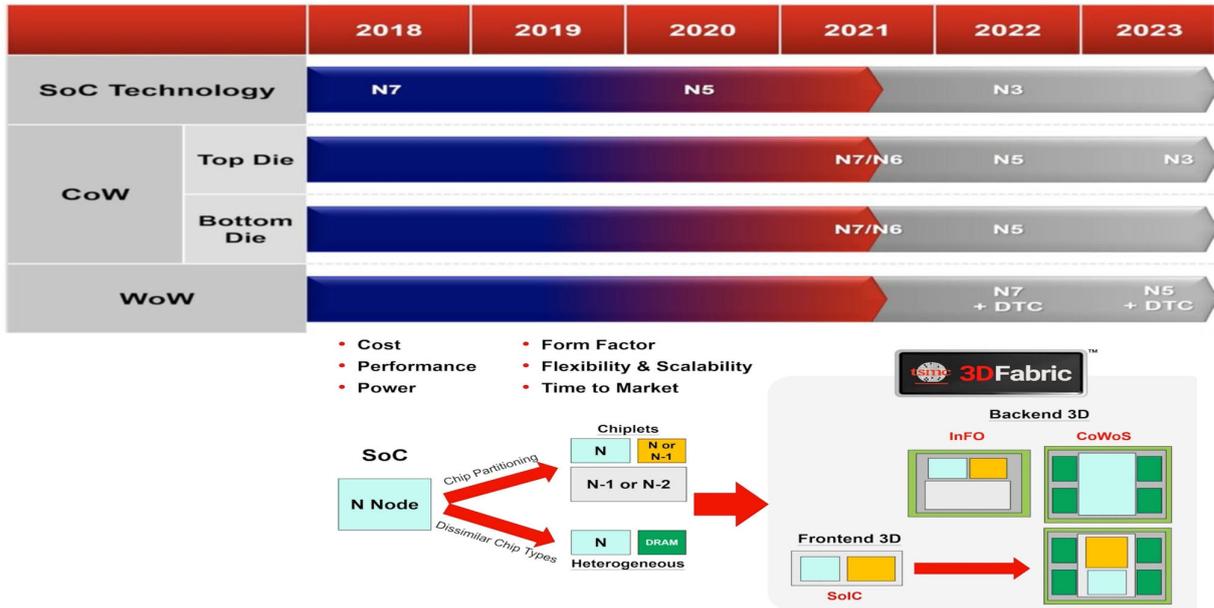
Service Scope

- SoC and ASIC design to production
- Interposer and RDL design
- SI/PI/IR/THM simulation
- Package and substrate design

資料來源：GUC、群益投顧預估彙整

SoIC 2022 年小量投產

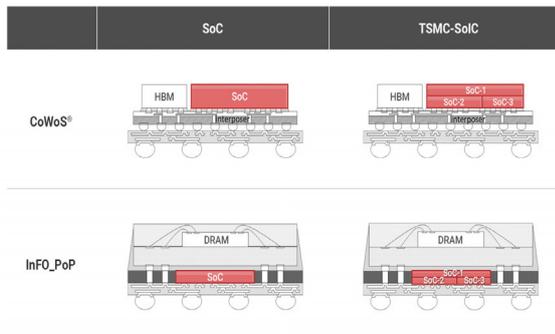
- 台積電2022年開始SoIC晶片堆疊製造，計劃2026年將產能擴大到20倍以上。竹南擁有第一座3D Fabric全自動化工廠，將先進測試及SoIC和InFO/CoWoS運作整合在一起，2023年開始3D Fabric全面運作。



資料來源：TSMC、群益投顧預估彙整

SoIC和InFO/CoWoS整合運作

- Front-end 3D：SoIC和InFO/CoWoS運作整合將同質或異構chipllets都整合到一個SoC-like的晶片中，使晶片面積更小和更薄。外觀上，SoIC就像普通的SoC，但嵌入所需的異質整合功能。本質就是在做一顆SoC晶片，基本上全部都在晶圓廠完成。
- Back-end 3D：前端封裝完成的SoIC晶片，搭配立體封裝技術advanced WLSI，如CoWoS和InFO。相關後端封裝技術是其他封測廠商積極跨入的領域。



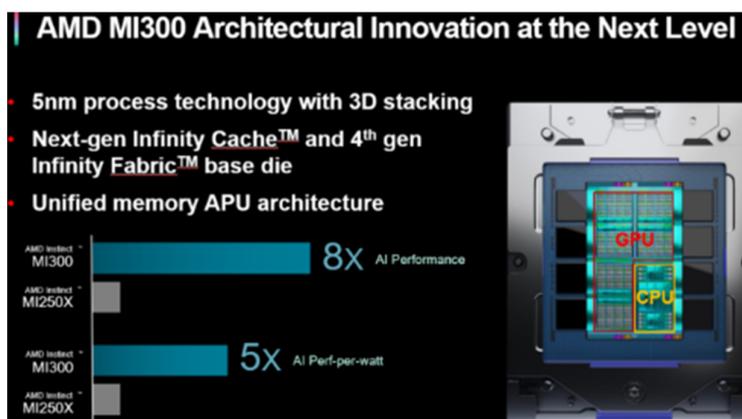
Technology	2.5D	3D-IC	SoIC
Structure cross-section	SoC1, SoC2, μbump, BEOL, Interposer	SoC1, μbump, SoC2	SoC1, SoIC Bond, SoC2
Interconnect	μbump + BEOL	μbump	SoIC bond
Bump Density	1.0X	1.0X	16.0X
Speed	0.01X	1.0X	11.9X
Bandwidth Density	0.01X	1.0X	191.0X
Power Efficiency (Energy/bit)	22.9X	1.0X	0.05X



資料來源：TSMC、ISSCC 2021、群益投顧預估彙整

AMD採用SoIC

- AMD早在2015年就推出首款HBM設計，2017年多晶片模組(MCM)，2019年推出Chiplet處理器，並在2022年推出首款採用3D V-Cache設計的晶片上3D封裝。
- AMD採用SoIC-X技術將N5 GPU和CPU堆疊於底層晶片，並整合在CoWoS封裝。Instinct MI300，用SoIC-X封裝技術組合13個晶片，整合N5和N6的CPU與GPU，以及堆棧式顯示記憶體，整體結構包含128GB顯示記憶體和1,460億個晶體管。
- MI300比前代MI250加速處理品有8倍的AI性能和5倍的功率性能提升，可將超大型AI模型的訓練時間從幾個月減少到幾周。未來將應用於美國新一代200億億次的El Capitan超級電腦。



資料來源：TSMC、群益投顧預估彙整

Capital Care 群益關心您
台北 · 香港 · 上海

MI 300A vs MI300X

- MI300A - 6 XCDs (Up To 228 CUs), 3 CCDs (Up To 24 Zen 4 Cores), 8 HBM3 Stacks (128 GB)。
- MI300X - 8 XCDs (Up To 304 CUs), 0 CCDs (Up To 0 Zen 4 Cores), 12 HBM3 Stacks (192 GB) in 4Q23。
- MI300A 是 AMD 首度將 APU 產品擴大到 AI 與 HPC 領域，採用 AMD CDNA3 加速架構；相較消費級 APU 產品，Instinct MI300A 採用 CPU + GPU 統一記憶體架構，能存取彼此運算的結果，共有 128GB HBM。
- AMD 稱 Instinct MI300X 提供的 HBM 密度最高是 NVIDIA AI 晶片 H100 的 2.4 倍，其 HBM 頻寬最高是 H100 的 1.6 倍。高達 192GB 的 HBM3，等同單一 Instinct MI300X 加速卡即有超越兩張 NVIDIA H100 的記憶體容量(註：即便是雙卡連接的 H100 NVL 總記憶體也僅有 188GB)。

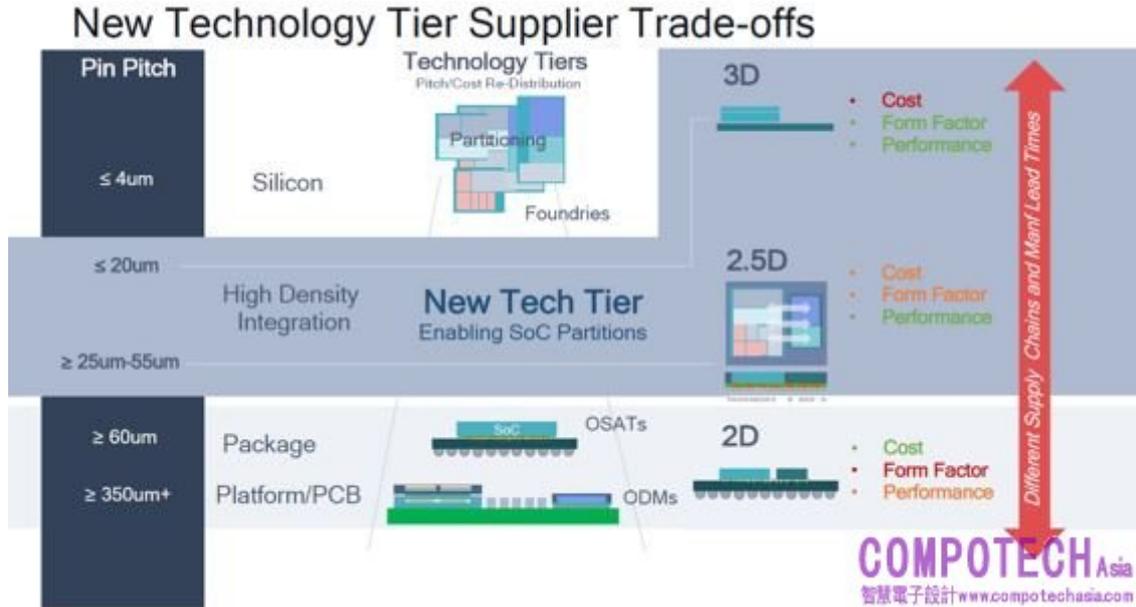


資料來源：AMD、群益投顧預估彙整

Capital Care 群益關心您
台北 · 香港 · 上海

沒有完美的解決方案

- 終端應用設計將依產品生命、功率、面積、封裝成本佔比而定。各廠需要在效能、速度、良率、和成本等做出選擇。



資料來源：智慧電子設計、Qualcomm、群益投顧預估彙整

AI晶片隨應用場景而不同

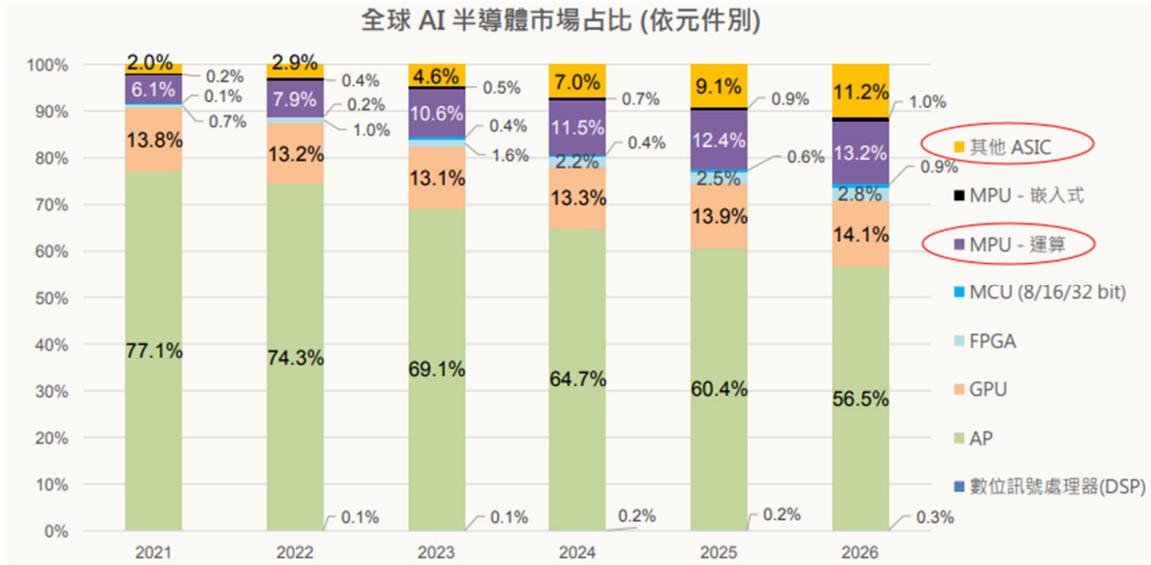
- AI晶片，若以技術架構來看，包括GPGPU、全訂製化ASIC、半訂製化FPGA，以及神經形態晶片或類腦晶片。
- 根據機器學習演算法的步驟，可分為訓練(training)和推論(inference)，在訓練的環節中，因為涉及需處理海量訓練資料及複雜的神經網路架構，因此需要GPU。推論環節是讓訓練好的模型，利用新的資料數據去推斷、預測出各種結論，運算量相對較少，因此CPU、GPU、FPGA和ASIC都能發揮作用。
- AI晶片如果以應用來分類，可分為雲端(伺服器)與終端(如行動裝置、無人自駕車、智慧家電...等)應用，前者需要高運算效能的IC；後者則須在一定的運算力之外，具備低功耗、低成本及低延遲等特性，甚至使用微控制器(MCU)即可。

	訓練速度	推論速度	推論準確度
CPU	以CPU為基準		最高 (98-99.7%) 🏆
GPU	10-1000倍 🏆	1-100倍	最高 (98-99.7%) 🏆
FPGA	-	10-100倍	高 (95-99%)
ASIC	10-1000倍 🏆	10-1000倍 🏆	較低 (90-98%)

資料來源：工研院產科國際所、群益投顧預估彙整

AP晶片為AI半導體最大宗

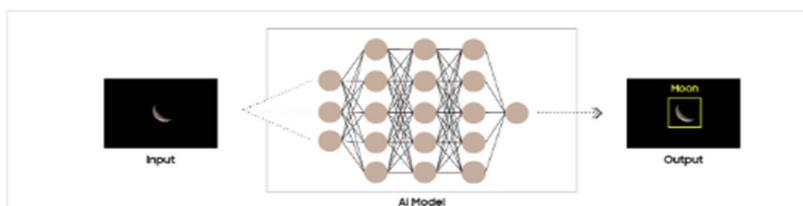
- 涵蓋AI運算功能的AP晶片佔AI半導體最大宗，但多媒體AP衰退，使佔比減少。
- 現在主流AI晶片大多為ASSP，但非主要執行AI運算，如手機 AP裡影像AI的演算法，是針對拍攝的照片進行處理，還有應用於自動駕駛設備中的處理器等等。



資料來源：工研院產科國際所、群益投顧預估彙整

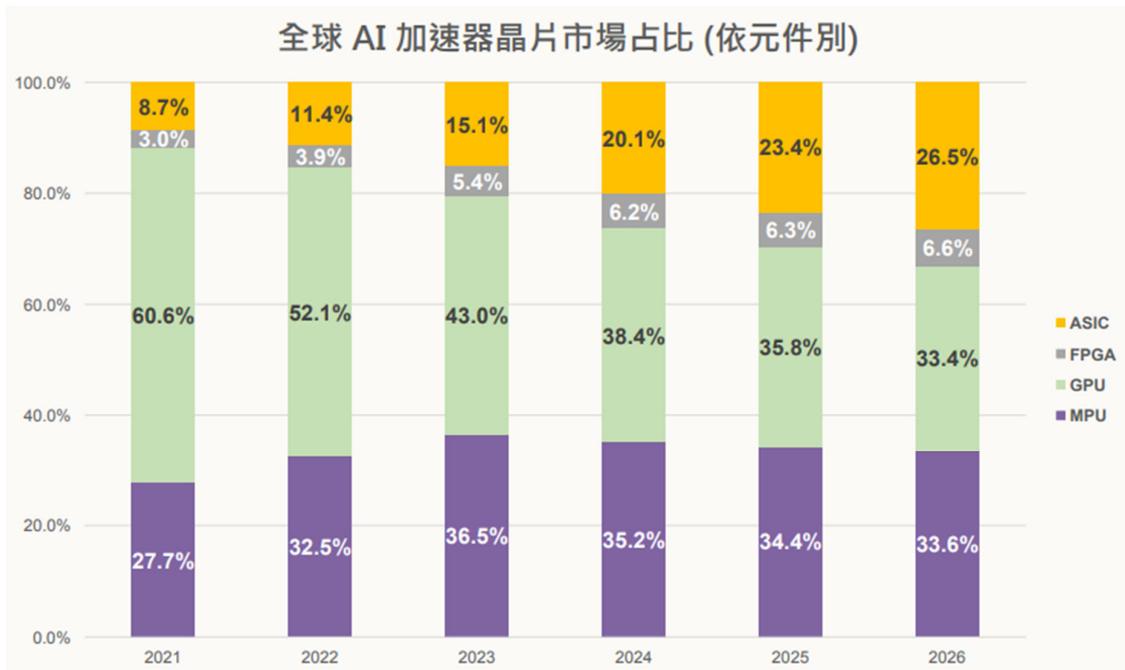
Samsung AI 繪月

- 獲得初步影像後，Galaxy 手機 AI 深度模型將會涉入影像增強處理，但Samsung並非單純使用固定的月亮照片進行合成，而是利用Samsung蒐集到的大量月相照片、從滿月到新月等資料進行 AI 模型訓練，再利用 AI 增強技術把細節透過 AI 增強方式補上拍攝的照片。
- Samsung 的 AI 如果判斷拍攝的不是月亮，就不會進行這些處理，避免把一盞昏黃的街燈，當作月亮進行AI影像處理。



資料來源：Samsung、群益投顧預估彙整

AI 加速器以 ASIC 成長最大

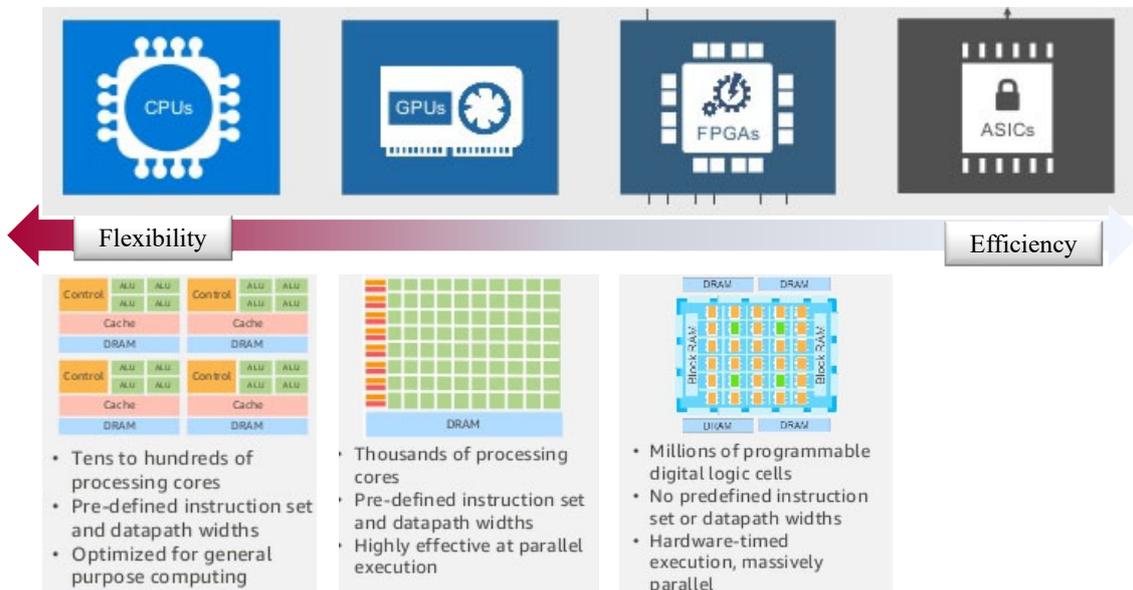


資料來源：工研院產科國際所、群益投顧預估彙整

Capital Care 群益關心您

深度學習 Training 為 GPU 天下

- GPU 比 CPU 擁有更多的運算器 (Arithmetic Logical Unit)，只需要進行高速運算而不需要邏輯判斷，其海量資料並行運算的能力與深度學習需求不謀而合。目前 GPU 市場 NVIDIA 市佔率達 70%，AMD 為輔。

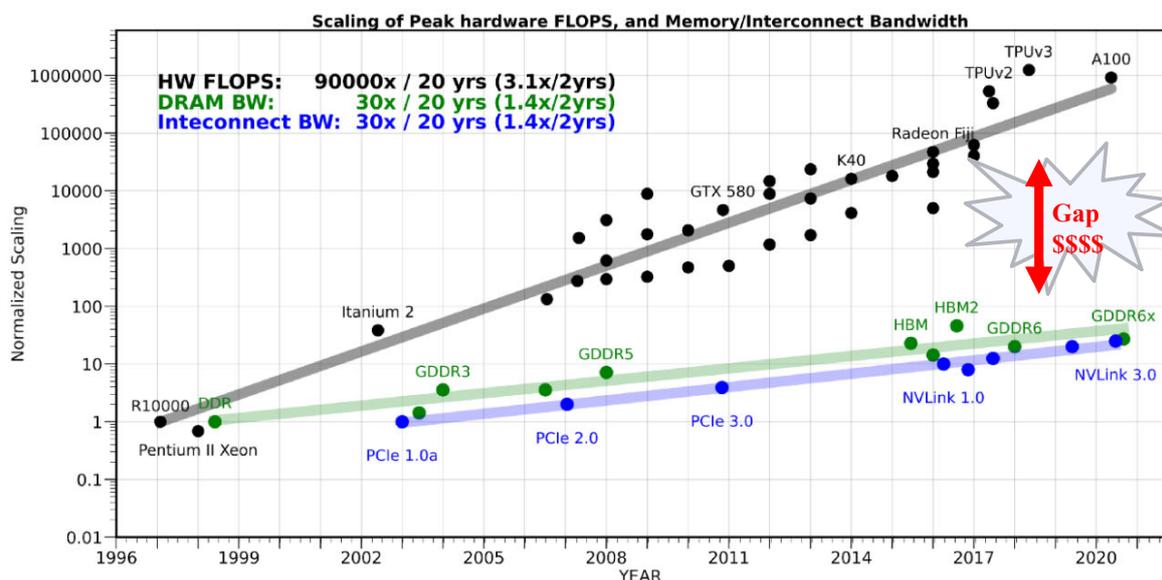


資料來源：Microsoft、AWS、群益投顧預估彙整

Capital Care 群益關心您

記憶體頻寬限制

- 晶片內部、晶片間之間的通信，已成為不少 AI 應用的瓶頸，過去20年硬體的峰值計算能力增加90,000倍，但記憶體/硬體互連頻寬卻只提高30倍。和單設備的記憶體牆問題類似，擴展 AI 硬體之間的網路頻寬的技術難題同樣還未被攻破。



資料來源：onflow、群益投顧預估彙整

Capital Care 群益關心您

Memory Wall限制運算效能

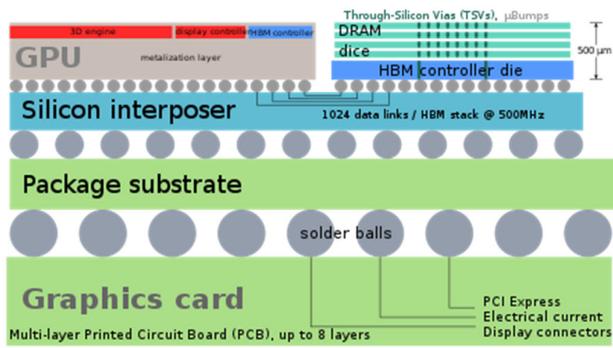
- 大型語言模型(LLM)僅用於模型權重的記憶體就需要 100 GB 以上，百度和 Meta Inference則需要10 TB以上的記憶體來儲存其巨量嵌入表。導致大型模型訓練超過40%時間都在等資料傳輸，並沒有用在矩陣乘法運算，也就是記憶體撞牆(Memory Wall)。
- Nvidia 2016 年 P100(Pascal)和最新 H100(Hopper)相比，FP16 性能提升 46 倍 (21.2TFlops → 989.5TFlops)，但記憶體容量只增長 5 倍(16GB→80GB)，記憶體理論頻寬更僅增約 3 倍(720GB→2TB/s，表示每個 Flops 分到的記憶體頻寬越來越少)。
- 當初富士通設計 A64FX 處理器，其實可選擇 1,024 位元長的運算資料寬度，而非後來的 512 位元。也就是本來可推出一台有「兩倍 Flops」超級電腦，但選擇不這樣做，就是因為沒有足夠頻寬，就無法達到真正效能。這也反映富岳擁有較高峰值性能百分比(實際性能 R_{max} ÷ 理論性能 R_{peak})。

項目	Top 500排名	單一節點理論浮點運算效能(FP64)	節點網路理論頻寬	每TFlops分配頻寬	峰值性能百分比 R_{max}/R_{peak}
Fugaku 富岳 (A64FX)	2	3.3TFlops	400Gb/s	133Gb/s	82%
Summit (Power9) V100	5	49TFlops	200Gb/s	4Gb/s	74%

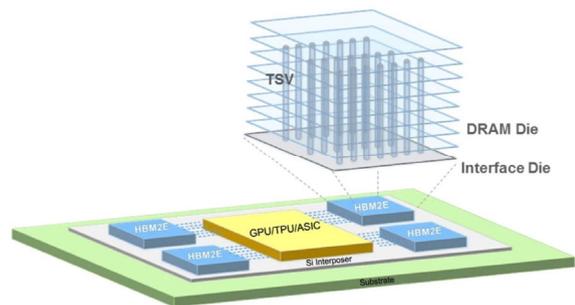
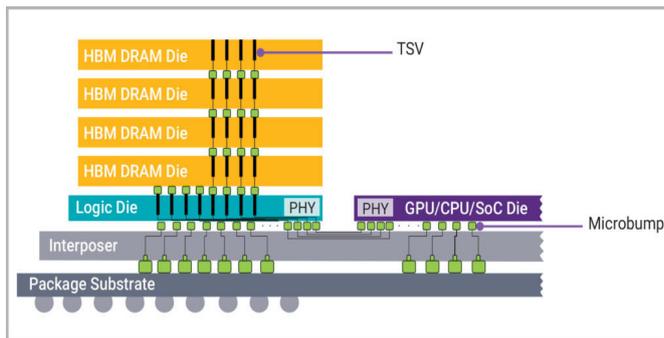
資料來源：TechNews、群益投顧預估彙整

Capital Care 群益關心您

HBM提供更高的頻寬但價格高昂

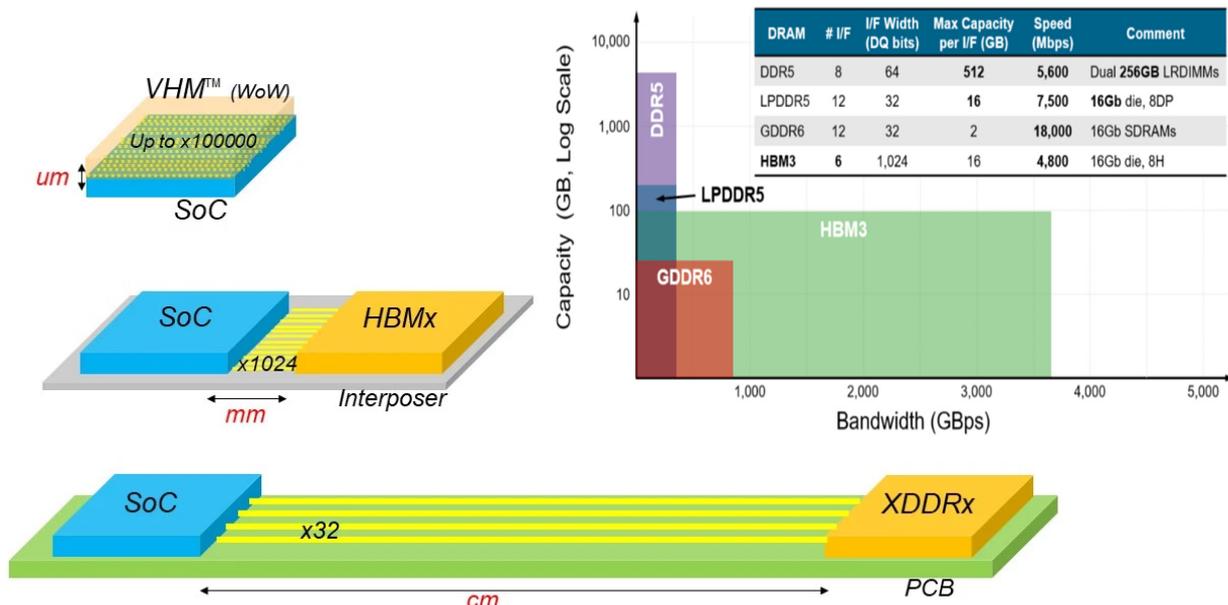


- HBM選擇3D堆疊，介面變得更寬。DDR的介面位元寬只有64位元，而HBM2E透過DRAM堆疊的方式就將位元寬提升到1,024位元。
- HBM3原本價格大約30美元每GB，2023年的價格更是大漲。目前標準型DDR4 16Gb顆粒約3美元。



資料來源：Synopsys、Micron、維基、群益投顧預估彙整

VHM等更高頻寬解決方案出現



資料來源：AP、Micron、群益投顧預估彙整

SK Hynix在HBM市佔率50%

- 據TrendForce，2022年三大原廠HBM市佔率分別SK Hynix 50%、Samsung約40%、Micron約10%。SK Hynix為目前唯一量產新世代HBM3，Samsung、Micron預計在2023年底至2024年初量產。
- 目前NVIDIA所定義DL/ML型AI伺服器平均每台均搭載4張或8張高階顯卡，搭配兩顆主流型號的x86伺服器CPU。Server DRAM配置約500~600GB，AI伺服器在單條模組上則採64~128GB，平均容量可達1.2~1.7TB。AI伺服器多增加GPGPU的使用，因此以NVIDIA A100 80GB配置4或8張採計，HBM用量約為320-640GB。

Table 1: Annual Growth Rate Forecast for Global AI Server Shipment Volume from 2022~2027

2022	2023(E)	2024(F)	2025(F)	2026(F)	2027(F)
9.0%	15.4%	10.0%	12.7%	11.3%	15.0%

Source: TrendForce, Apr., 2023

Table 2: Difference in Average Capacity Between General and AI Servers

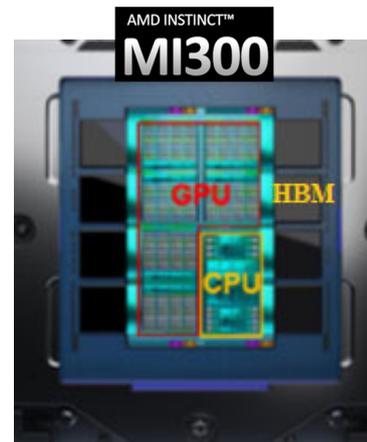
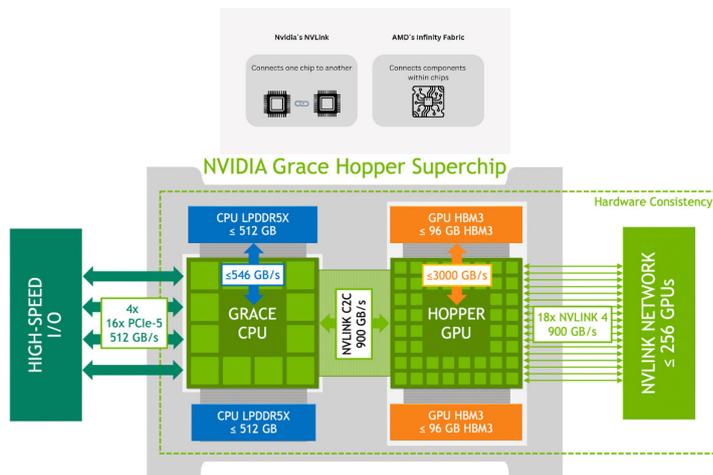
	Server	AI Server	Future AI Server
Server DRAM Content	500~600GB	1.2~1.7TB	2.2~2.7TB
Server SSD Content	4.1TB	4.1TB	8TB
HBM Usage	-	320~640GB	512~1024GB

Source: TrendForce, Apr., 2023

資料來源：TrendForce、群益投顧預估彙整

GPU 各有奇招提升效率

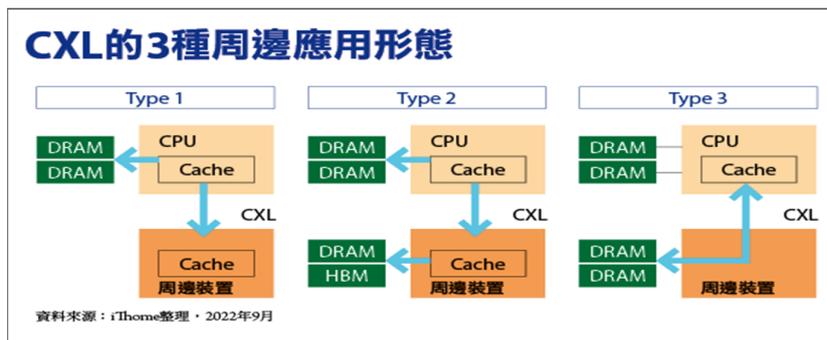
- NVIDIA和AMD都是採用HBM3，但架構不同，最後呈現的結果也不同。
- AMD Instinct MI300A 為CPU與GPU共享主記憶體，採用Cheplet封裝。
- 不同於Intel和AMD採用Chiplet架構的做法，NVIDIA首款GPU+CPU組合—Grace Hopper Superchip還是單晶片的方式。NVLink C2C提供高達900 GB/s的總頻寬，這比加速系統中常用的x16 PCIe Gen5通道的頻寬高7倍。



資料來源：NVIDIA、AMD、群益投顧預估彙整

CXL調度伺服器整機DRAM用量

- 規模越大的資料中心，卻有更嚴重的記憶體閒置問題，Google 報告指出，資料中心 DRAM 利用率只有 40%，微軟也曾表示，當 Azure 所有 CPU 核心都配給虛擬機器後，仍有 25% 記憶體容量沒有 CPU 用到，處閒置狀態。對 Google 和微軟這種大型雲端服務業者來說，這麼高記憶體閒置率，形同數10億美元損失。
- 跨越 CPU 與周邊裝置的藩籬，將記憶體虛擬成全部運算單元共用的共享池 Memory Pooling，大幅提升記憶體資源的應用彈性與效率。1H24 CXL1.1和 CXL2.0將有產品推出。Jensen Huang喜歡說“資料中心是新的計算單位”。
- CXL controller廠商有Montage、Marvell與Microchip等。



資料來源：iThome、群益投顧預估彙整

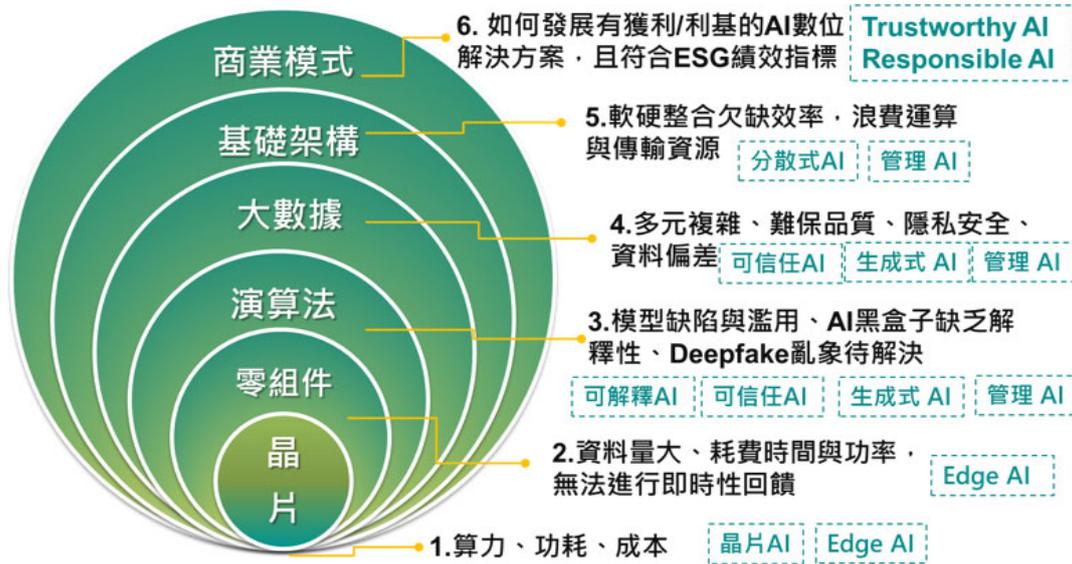
複雜算力系統須軟硬體融合

類型	算力	通用性	優點	缺點
CPU	x1	100	極度靈活可程式設計。	效率最低，性能已經瓶頸。
GPU	x10+	10	相比CPU的性能優勢，相比DSA/ASIC的通用靈活性，應用領域更加廣泛。	效率相比DSA/ASIC不高，性能即將到頂。
ASIC	x100+	1	極致性能	靈活性最低，完全定制的方案，用戶難以差異化，難以大規模。
DSA	x100+	5	接近於ASIC的極致性能，並且具有一定的可程式設計能力。	同一領域，應用的演算法依然多種多樣，並且快速反覆運算，應用層（如AI）DSA難以大規模落地。
CASH	x1000+	100	超異構：CPU+GPU+DSA+其他；更大規模，更大算力。 軟硬體融合架構：CPU軟體一樣靈活可程式設計，DSA/ASIC硬體一樣極致的性能。 開放的平臺及生態：開放、標準的程式設計模型和訪問介面，融合主流開源軟體。	

資料來源：群益投顧預估彙整

AI跨域多元

- AI跨域多元，從晶片、演算法、大數據、基礎架構及商業模式等，AI都面臨著不同層次的問題與挑戰，但也隱藏著AI技術創新機會或潛力商機。



資料來源：工研院產科國際所、群益投顧預估彙整

美國出口管制重擊中國AI發展

- 10/07/2022美國商務部工業與安全局宣佈針對中國出口先進晶片的管制新規。聲明規定，滿足輸入輸出(I/O)雙向傳送速率高於 600GB/s，同時每次操作的比特長度乘以 TOPS 計算出的處理性能合計為 4800 或更多算力的產品，將無法出口至中國。
- 若未來這個標準沒有改變，將會限制中國未來AI產業發展。

ECCN	內容
3A090 (晶片)	GPU /CPU /TPU、可程式設計邏輯裝置(FPLD)、ASIC等各類晶片，滿足輸入輸出(I/O)雙向傳送速率高於 600GB/s，同時每次操作的比特長度乘以 TOPS 計算出的處理性能合計為 4800或更多算力的產品。若未來這個標準沒有改變，將會限制中國未來晶片開發。
4A090 (電腦)	裝有超過3A090技術指標晶片的電腦、電子元件及含有積體電路的部件。 技術說明：電腦包括“數位電腦”、“混合電腦”和“類比計算機”。 超級電腦的定義是在 41,600 立方英尺或者更小的體積內，FP64（雙精度）理論計算能力是在 100 petaFLOPS（每秒千萬億次浮點運算） 或者以上，FP32（單精確度）在 200 petaFLOPS 或者以上浮點算力的超級電腦，被定義為對中國管控的相關電腦。

資料來源：BIS、群益投顧預估彙整

美國禁令使中國付出高價跑AI模型

NVIDIA A100 TENSOR CORE GPU SPECIFICATIONS (SXM4 AND PCIe FORM FACTORS)

	A100 80GB PCIe	A100 80GB SXM
FP64	9.7 TFLOPS	
FP64 Tensor Core	19.5 TFLOPS	
FP32	19.5 TFLOPS	
Tensor Float 32 (TF32)	154 TFLOPS 312 TFLOPS*	
BFLOAT16 Tensor Core	312 TFLOPS 624 TFLOPS*	
FP16 Tensor Core	312 TFLOPS 624 TFLOPS*	
INT8 Tensor Core	624 TOPS 1248 TOPS*	
GPU Memory	80GB HBM2e	80GB HBM2e
GPU Memory Bandwidth	1,935GB/s	2,039GB/s
Max Thermal Design Power (TDP)	300W	400W***
Multi-Instance GPU	Up to 7 MIGs @ 10GB	Up to 7 MIGs @ 10GB
Form Factor	PCIe dual-slot air cooled or single-slot liquid cooled	SXM
Interconnect	NVIDIA® NVLink® Bridge for 2 GPUs: 600GB/s** PCIe Gen4: 64GB/s	NVLink: 600GB/s PCIe Gen4: 64GB/s
Server Options	Partner and NVIDIA-Certified Systems™ with 1-8 GPUs	NVIDIA HGX™ A100-Partner and NVIDIA-Certified Systems with 4, 8, or 16 GPUs NVIDIA DGX™ A100 with 8 GPUs

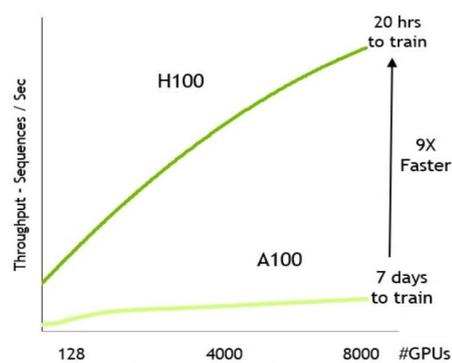
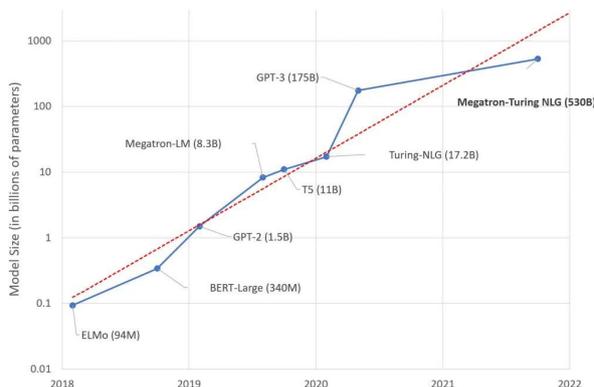
* With sparsity
** SXM4 GPUs via HGX A100 server boards; PCIe GPUs via NVLink Bridge for up to two GPUs
*** 400W TDP for standard configuration. HGX A100-800B CTS (Custom Thermal Solution) SKU can support TOPs up to 500W

- 美國禁令將速率限制為每秒 600 GB 及以上，A800 只是在 A100 的基礎上，將 NVLink 高速互連匯流排的頻寬從 600GB/s 降低到 400GB/s，僅此而已。
- 市場上並沒有 NVIDIA GPU 的直接替代品。因此許多中國企業嘗試合併 3 或 4 片較低階晶片(含 A800 與 H800)。若美企訓練大型語言運算模式需用 1,000 片 H100，中國企業就需 3,000 片以上的 H800 才能達成類似效果。
- 儘管如此，H800 晶片要多花 10% 到 30% 時間來執行 AI 工作，且成本增加一倍，儘管如此，還是強過目前大陸企業生產的晶片。騰訊 04/2023 曾估算，使用 H800 來訓練旗下的 AI 系統，時間能縮短一半以上，從 11 天大幅減少至 4 天。

資料來源：NVIDIA、群益投顧預估彙整

降規非常久之計

- 以 Nvidia 為例，H100 用於具 3950 億參數的 Mixture of Experts 網絡，訓練時間將從 A100 的 7 天減少到 H100 的 20 小時。若是用於 5300 億參數的 Megatron-530B，H100 的表現超過 A100 多達 30 倍。
- 雖然可以用數量較多的較低階晶片達到同樣效能，但每 1~2 年就會有更高階的晶片出現，效能會更佳，而舊款晶片將會逐步淘汰，因此降規非長久之計。



Mixture of Experts (395B) Training vs A100

資料來源：NVIDIA、群益投顧預估彙整

中國試圖提升晶片使用效率

Microsoft | Source

The supercomputer developed for OpenAI is a single system with more than 285,000 CPU cores, 10,000 GPUs and 400 gigabits per second of network connectivity for each GPU server. Compared with other machines listed on the TOP500 supercomputers in the world, it ranks in the top five, Microsoft says. Hosted in Azure, the supercomputer also benefits from all the capabilities of a robust modern cloud infrastructure, including rapid deployment, sustainable datacenters and access to Azure services.



資料來源：NVIDIA、群益投顧預估彙整

- 在 Azure 為 OpenAI 準備的訓練研發平台上，CPU 與 GPU 的數量比接近 1:2。GPU 為 1 萬個 V100。
- 意即至少要 1 萬片 A100 晶片，才能鍛鍊 ChatGPT 這種規模等級的 AI。據傳中國只有 4~5 萬片 A100。
- 中國在尋求透過各種軟體技術來幫助訓練大型 AI 模型的運算強度，無需額外 H100，訓練大型語言模型功能，使晶片效率提升，降低高運算晶片的用量。
- 積極發展中國本土 GPU。

中國客戶是商機也是危機

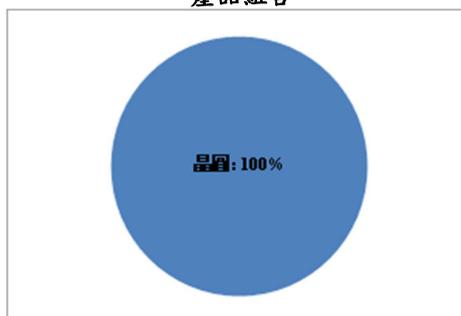
- 美中科技戰持續延燒，對台廠是「兩面刃」，部分廠商確實享有大陸「去美化」或美國「去中化」帶來的轉單效益，但只要政策一變，立馬受傷。
- 這次新禁令特別針對高算力晶片，IP 設計服務公司現在開始要接未來 2~3 年的案子，若算力限制沒有改變，將會影響手上 NRE 訂單，及現在 NRE 接單，進而影響未來 ASIC 投片量。

公司	觀察重點
創意	<ul style="list-style-type: none"> ● 5/3nm NRE 認列進度 ● 5nm 量產速度
世芯-KY	<ul style="list-style-type: none"> ● 中國客戶佔營收比重高 ● 5nm NRE 認列進度
智原	<ul style="list-style-type: none"> ● 中國新基建有項目，包括 5G 基建、特高壓電網、大數據中心以及人工智慧
晶心科	<ul style="list-style-type: none"> ● RISC-V 推廣情況 ● 中國導入 RISC-V 速度
M31	<ul style="list-style-type: none"> ● 中國廠商使用的意願 ● 拓展大陸以外的客戶 ● 高速介面 IP 滲透率

資料來源：群益投顧預估彙整

- 經濟環境疲弱和市場需求不振，庫存可能要到3Q23才能達到健康的水準。產能利用率下滑和及電費增影響，毛利率將會受到影響。
- 2023年成長動能來自HPC和Automotive，以及N5/N3製程營收貢獻上揚。台積電高階製程市佔率高，競爭力優於同業。

產品組合



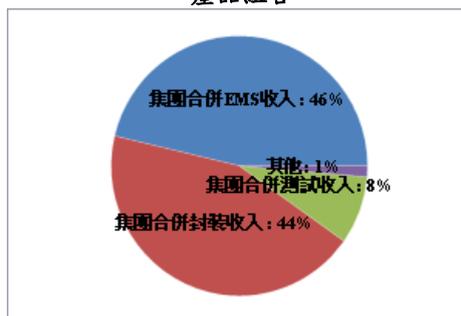
單位：百萬元	2022	2023F	2024F	1Q22	2Q22	3Q22	4Q22	1Q23	2Q23F	3Q23F	4Q23F	1Q24F
營業收入淨額	2,263,891	2,168,794	2,468,017	491,076	534,141	613,143	625,532	508,633	485,642	547,375	627,144	568,363
營業毛利淨額	1,348,355	1,232,590	1,417,825	273,203	315,468	370,499	389,185	286,500	264,218	308,521	373,350	320,412
營業利益	1,121,279	991,841	1,154,981	223,790	262,124	310,324	325,041	231,238	202,056	251,047	307,500	259,881
稅後純益	1,016,530	856,869	993,906	202,733	237,027	280,866	295,904	206,987	172,402	214,449	263,032	223,730
稅後EPS(元)	39.20	33.04	38.33	7.82	9.14	10.83	11.41	7.98	6.65	8.27	10.14	8.63
毛利率(%)	59.56%	56.83%	57.45%	55.63%	59.06%	60.43%	62.22%	56.33%	54.41%	56.36%	59.53%	56.37%
營業利率率(%)	49.53%	45.73%	46.80%	45.57%	49.07%	50.61%	51.96%	45.46%	41.61%	45.86%	49.03%	45.72%
稅後純益率(%)	44.90%	39.51%	40.27%	41.28%	44.38%	45.81%	47.30%	40.69%	35.50%	39.18%	41.94%	39.36%
營業收入YoY/QoQ(%)	42.61%	-4.20%	13.80%	12.07%	8.77%	14.79%	2.02%	-18.69%	-4.52%	12.71%	14.57%	-9.37%
稅後純益YoY/QoQ(%)	70.40%	-15.71%	15.99%	21.96%	16.92%	18.50%	5.35%	-30.05%	-16.71%	24.39%	22.65%	-14.94%

註：稅後純益係指本期淨利歸屬於母公司業主；EPS預估值以股本2593.21億元計算。

資料來源：群益投顧彙整

- 日月光2023年與客戶將回到正常的價格談判區間模式，也就是價格將會有壓力，回到以往每年價格都會調整模式。
- 客戶將在3Q23開始拉貨與回補庫存，預期將是全面的復甦，期望2H23 ATM整體稼動率從現階段60%回升至80%。

產品組合



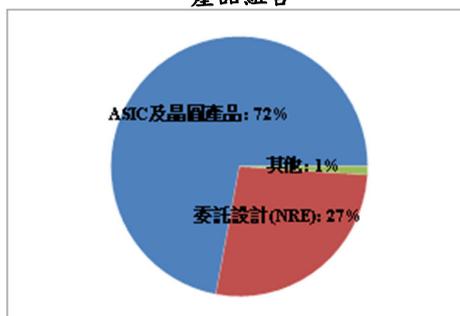
單位：百萬元	2022	2023F	2024F	1Q22	2Q22	3Q22	4Q22	1Q23	2Q23F	3Q23F	4Q23F	1Q24F
營業收入淨額	670,873	598,029	615,395	144,391	160,439	188,626	177,417	130,891	134,288	161,123	171,728	141,747
營業毛利淨額	134,930	102,606	118,604	28,471	34,388	37,972	34,099	19,339	19,847	30,550	32,869	25,997
營業利益	80,176	49,620	62,295	16,113	20,606	23,683	19,774	7,696	7,962	16,291	17,671	13,027
稅後純益	62,090	38,156	47,746	12,907	15,988	17,465	15,730	5,817	5,932	12,679	13,728	10,199
稅後EPS(元)	14.20	8.72	10.92	2.95	3.66	3.99	3.60	1.33	1.36	2.90	3.14	2.33
毛利率(%)	20.11%	17.16%	19.27%	19.72%	21.43%	20.13%	19.22%	14.78%	14.78%	18.96%	19.14%	18.34%
營業利率率(%)	11.95%	8.30%	10.12%	11.16%	12.84%	12.56%	11.15%	5.88%	5.93%	10.11%	10.29%	9.19%
稅後純益率(%)	9.26%	6.38%	7.76%	8.94%	9.97%	9.26%	8.87%	4.44%	4.42%	7.87%	7.99%	7.20%
營業收入YoY/QoQ(%)	17.70%	-10.86%	2.90%	-16.51%	11.11%	17.57%	-5.94%	-26.22%	2.59%	19.98%	6.58%	-17.46%
稅後純益YoY/QoQ(%)	-2.84%	-38.55%	25.13%	-58.25%	23.87%	9.23%	-9.93%	-63.02%	1.97%	113.75%	8.28%	-25.71%

註：稅後純益係指本期淨利歸屬於母公司業主；EPS預估值以股本437.34億元計算。

資料來源：群益投顧彙整

- 預期創意2023上下半年營收比例45:55，維持2023年營運目標不變，營收YoY雙位數成長其中Turnkey成長幅度大於NRE。
- 創意2023年主要成長動能為Networking、SSD 創意2022年新增NRE開案明顯增加，其中50%以上是AI及HPC相關應用，採用台積電N7及N5等先進晶圓製程，及CoWoS等先進封裝技術，將挹注2023年成長動能。

產品組合



單位：百萬元	2022	2023F	2024F	1Q22	2Q22	3Q22	4Q22	1Q23	2Q23F	3Q23F	4Q23F	1Q24F
營業收入淨額	24,040	30,378	29,570	4,514	5,381	6,062	8,083	6,529	6,551	8,011	9,287	6,587
營業毛利淨額	8,335	9,869	9,584	1,628	1,893	2,114	2,700	2,084	2,190	2,616	2,979	2,248
營業利益	4,099	5,105	4,442	633	831	1,146	1,488	1,103	1,174	1,334	1,493	1,069
稅後純益	3,710	4,349	3,792	545	760	1,058	1,346	934	1,002	1,139	1,274	912
稅後EPS(元)	27.69	32.46	28.30	4.07	5.68	7.90	10.05	6.97	7.48	8.50	9.51	6.80
毛利率(%)	34.67%	32.49%	32.41%	36.06%	35.18%	34.88%	33.40%	31.91%	33.43%	32.66%	32.08%	34.13%
營業利益率(%)	17.05%	16.80%	15.02%	14.02%	15.45%	18.91%	18.41%	16.89%	17.93%	16.66%	16.08%	16.23%
稅後純益率(%)	15.43%	14.32%	12.82%	12.08%	14.13%	17.46%	16.66%	14.31%	15.30%	14.22%	13.71%	13.84%
營業收入YoY/QoQ(%)	59.12%	26.37%	-2.66%	-8.06%	19.19%	12.66%	33.34%	-19.23%	0.34%	22.29%	15.93%	-29.08%
稅後純益YoY/QoQ(%)	154.11%	17.22%	-12.82%	5.90%	39.45%	39.15%	27.23%	-30.62%	7.30%	13.65%	11.81%	-28.43%

註：稅後純益係指本期淨利歸屬於母公司業主；EPS預估值以股本13.40億元計算。

資料來源：群益投顧彙整

Capital Care 群益關心您
台北 · 香港 · 上海

群益投顧

69

投資評等及免責聲明

投資評等說明

評等	定義
強力買進(Strong Buy)	首次評等潛在上漲空間≥35%
買進(Buy)	15%≤首次評等潛在上漲空間<35%
區間操作(Trading Buy)	5%≤首次評等潛在上漲空間<15%
中立(Neutral)	無法由基本面給予投資評等 預期近期股價將處於盤整 建議降低持股

免責聲明

本研究報告僅提供予特定人之客戶作為參考資料「非經同意不得轉載」。我們並不確保此資訊的完整性與正確性，投資人應了解，報告中有關未來預測之陳述可能不會實現，因而不應被依賴。而且此報告並非根據特定投資目的或依預定對象之財務狀況所撰寫出來的，因此，此研究報告的目的，既非對投資人於買賣證券、選擇權、期貨或其他證券相關之衍生性商品提供詢價服務，亦非作為進行交易的要約。投資人應注意到相關證券之價值及收益，可能會有無預警地上升或下降，產生投資回報金額可能比原始投資來得少的情形。

Capital Care 群益關心您
台北 · 香港 · 上海

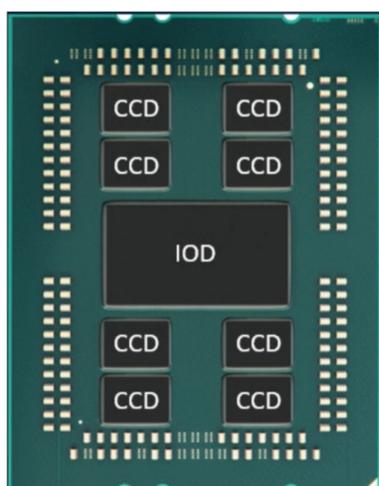
70

Appedix

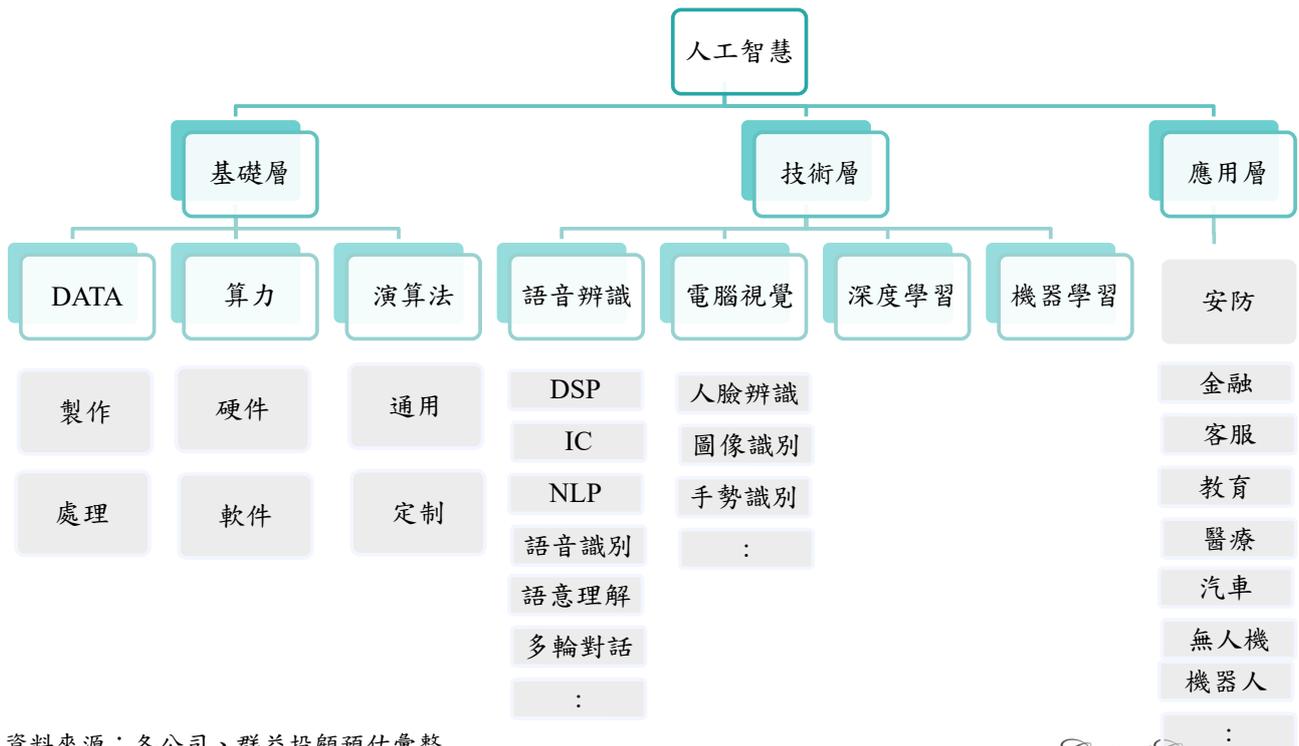


AMD EPYC 讓 Chiplet 進入主流

- AMD EPYC Rome處理器由9個晶片設計組成，也稱為MCM(多晶片模組)。9個晶片包括8個CCD(計算核心晶片)和一個IOD(輸入/輸出晶片)。每個CCD由2個CCX(計算核心複合體)組成，這些CCX擁有4個Zen 2核心，分別有各自L2快取和共享L3快取。所有8個CCD都使用Infinity Fabric連接到I/O Die。



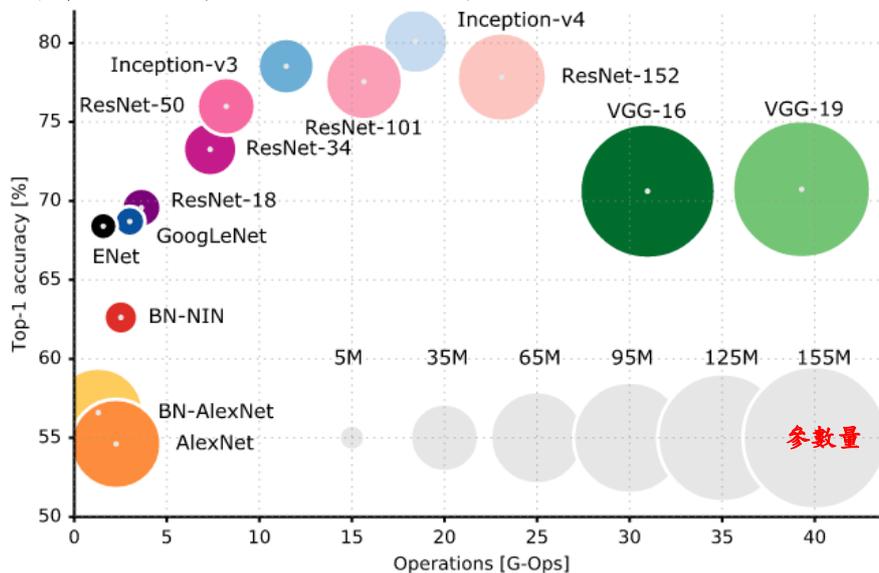
- IOD (Input / Output die) memory/PCIE
- CCD (Compute Core dies) compute cores only
- CCX (Compute Core complexes)
- Infinity Fabric定義AMD內部SoC IP區塊的通用控制方式，AMD可以更快在單一晶粒上「融合」不同的IP (如 Ryzen APU和Vega)，亦可「黏合」不同的晶粒 (像EPYC) 和打造多插座處理器環境



資料來源：各公司、群益投顧預估彙整

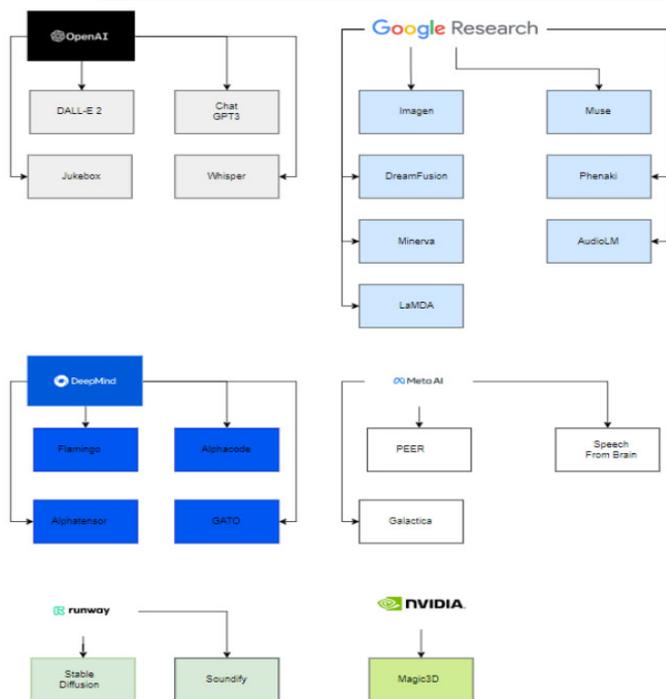
深度學習網路結構或演算法

- 橫軸是計算量，每個乘法或加法就算一次運算(Operations, Ops)，算法從數千萬次到數百億次的計算才能推論一筆資料，通常計算量和參數量(圓圈直徑)有直接關係，參數量可能從數百萬到數億個。但正確率和參數量及計算量不一定有直接關連，AI工程師希望找到計算量越低、正確率越高的模型。



資料來源：Piotr Migdał - blog、OmniXRI AI、群益投顧預估彙整

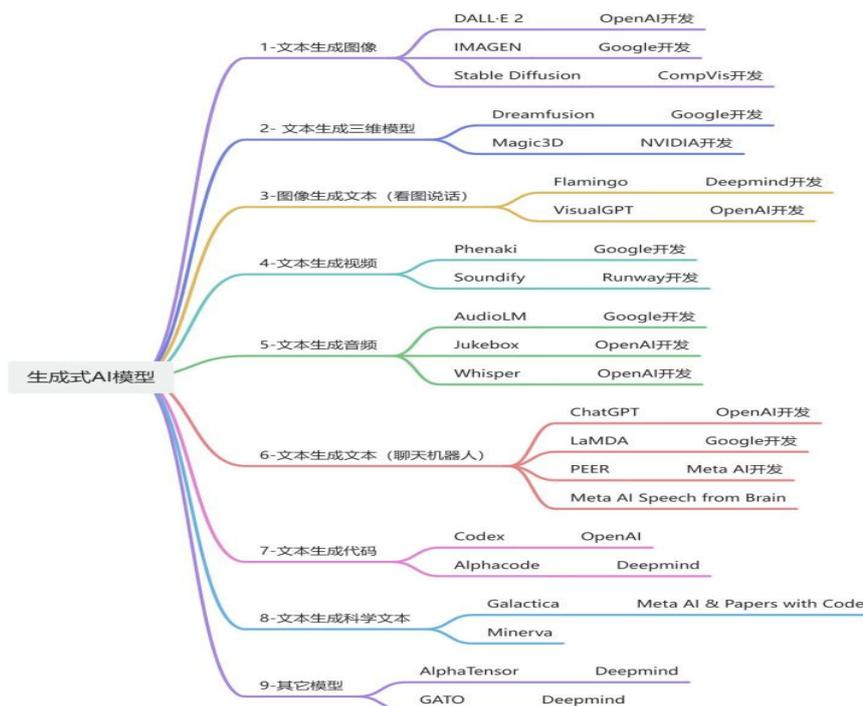
生成式AI模型投資高



- 目前只有6個組織在部署這些模型大模型(OpenAI、Google、DeepMind、Meta、runway和Nvidia)
- 為能運行模型的參數，必須擁有巨大的計算能力，以及在資料科學和資料工程方面高度熟練和經驗豐富的團隊。
- 高昂投入，讓普通創業公司難以為繼，因此參與者都是科技巨頭或研究單位。

資料來源：arxiv-vanity.com、群益投顧預估彙整

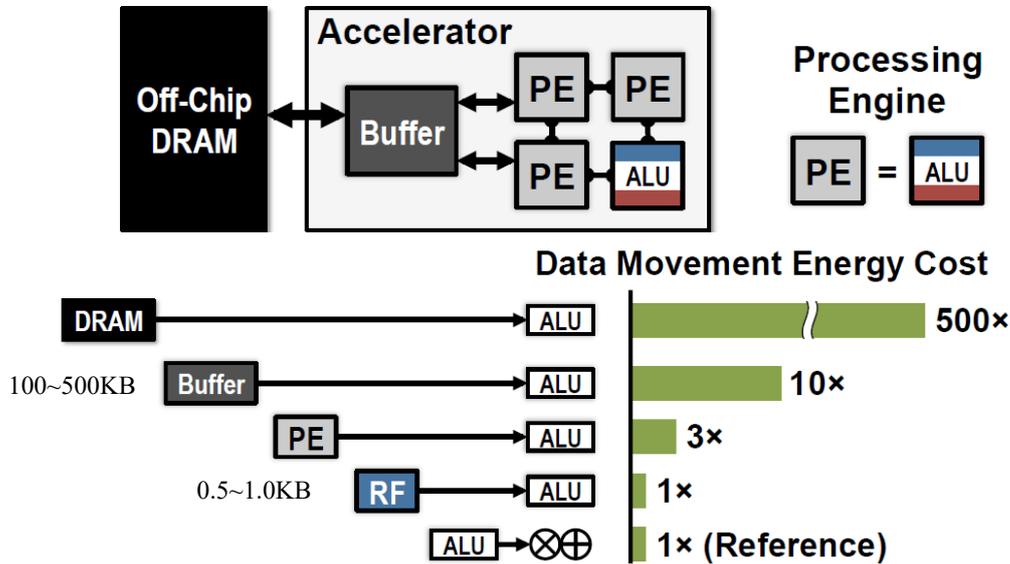
生成式AI模型9大類別



資料來源：XPU-進擊的晶片、群益投顧預估彙整

降低功耗

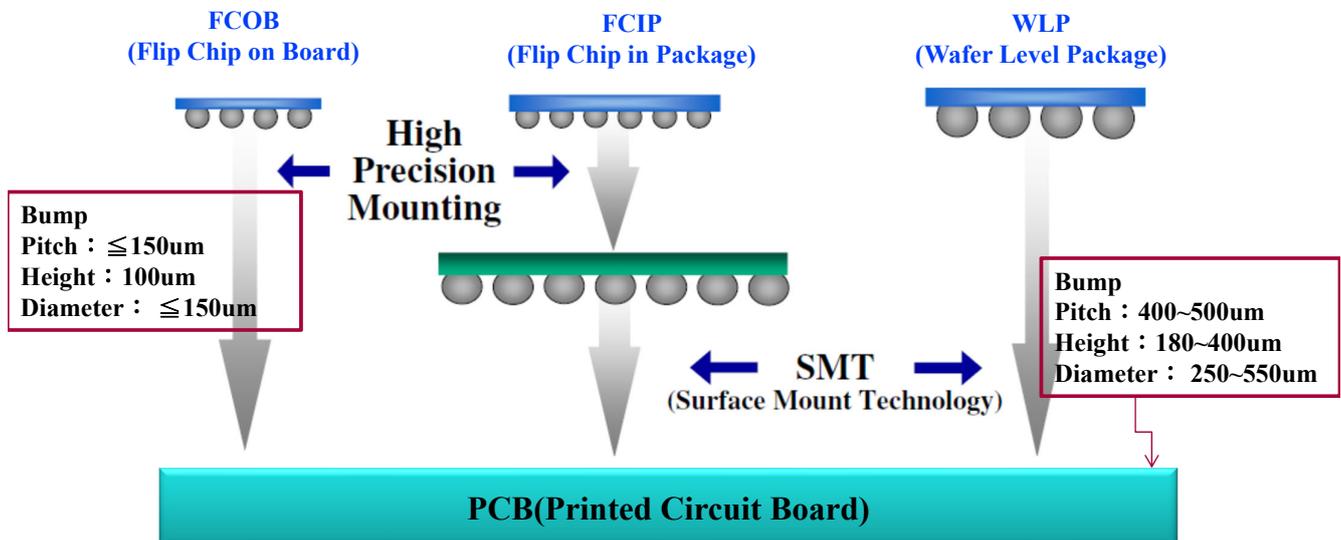
- DRAM存在兩個問題，一是頻寬不足，二是功耗過大。由於需要高頻驅動IO，DRAM的訪問能耗通常是單位運算的500倍。



資料來源：IEEE、群益投顧預估彙整

晶圓廠在晶圓級封裝較有優勢

- 3D堆疊封裝的難度在於對設備的精密度要求很高，TSMC有現成的晶圓級設備，透過適當改裝和DOE就可適用於3D堆疊封裝，同時憑藉其多年的晶圓級晶片代工經驗，比起封測廠商做3D晶片封裝的成功係數較高。使用矽中介層整合，因較簡化製程(不需TSV及u-bump等技術)，故可達降低成本目的。



資料來源：E&G Technology Partners、群益投顧彙整