

ITRI

Industrial Technology
Research Institute

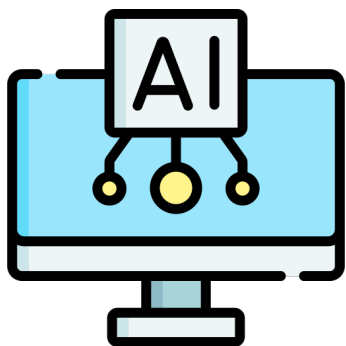
高速運算晶片開發與生成式AI應用

謝明得 技術長

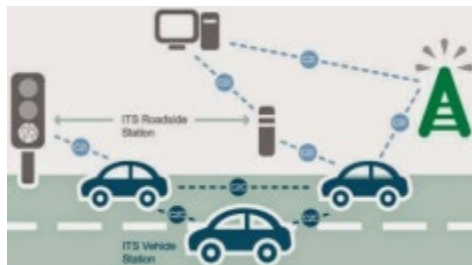
電光系統所

2024/12/20

市場趨勢-高速運算



AI PC



自駕車



Edge server



大型AI模型

2022 HPC營收佔43%
2025 將達55%



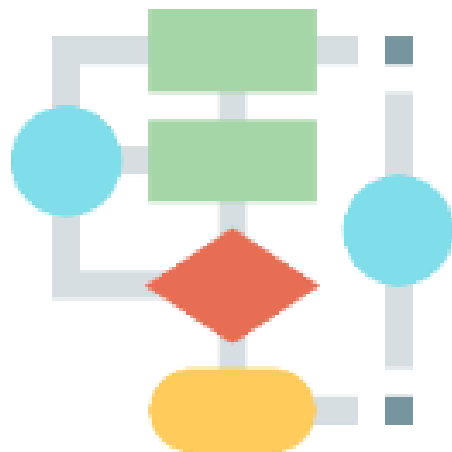
HPC是推動科學進步
最重要的工具之一



2030年HPC產值
將達651億美金



ITRI 高速運算發展策略



應用-LLM (Edge)

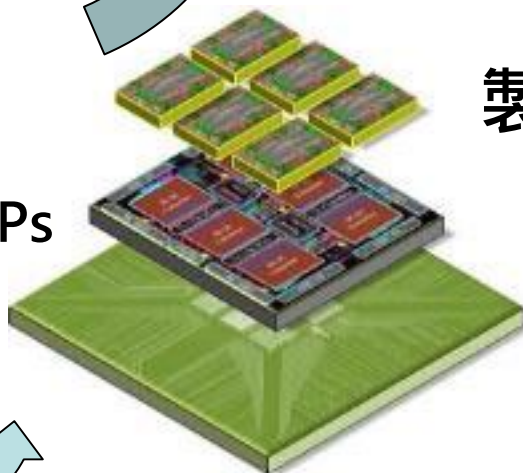
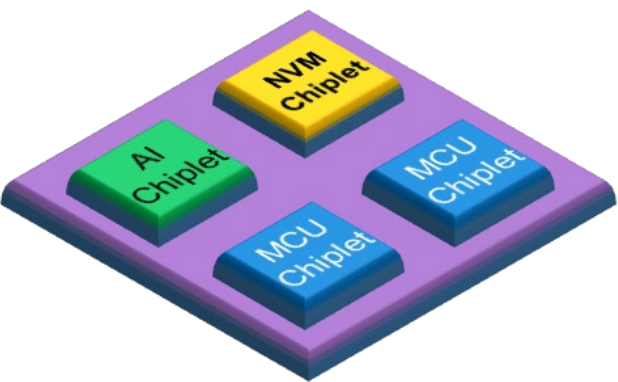
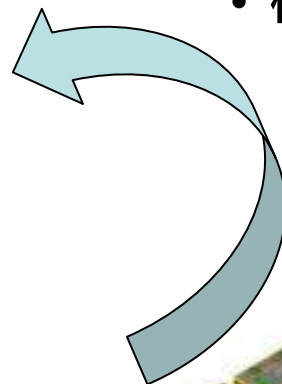
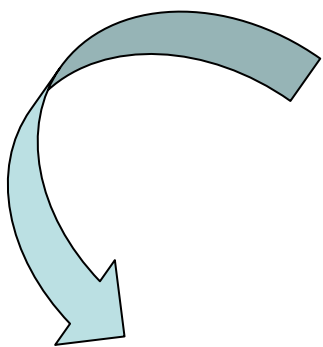
- 大型 AI 模型壓縮至1/20
- 模型平行化佈署推論速度提升30%

設計-Chiplet

- 異質可擴展算力 > 1Peta OPs
- 記憶體頻寬 > 40Gb/s

製造-多樣化異質整合

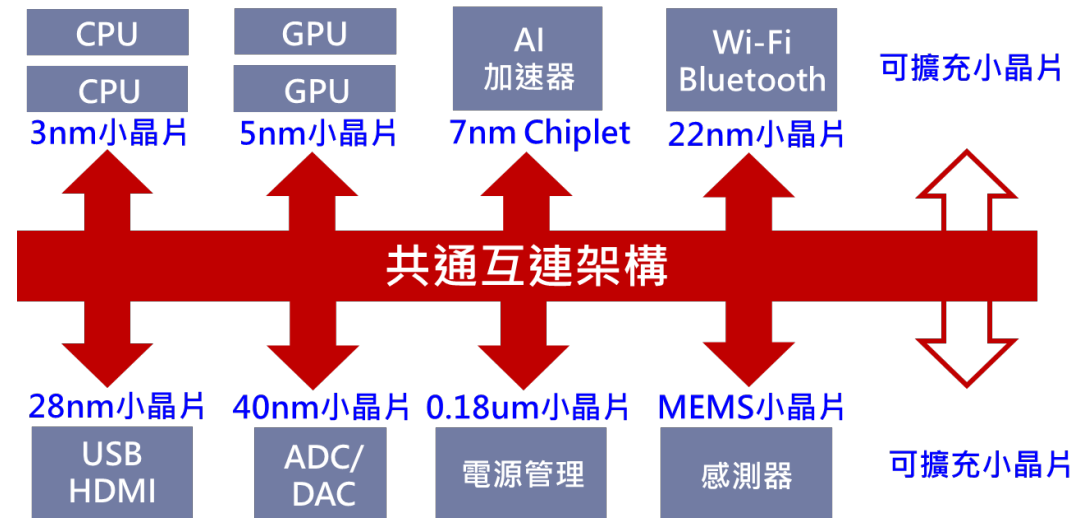
- 高密度 < 4 μ m 接點接合
- 整體製程溫度 < 200 °C



設計 - Chiplet互聯介面

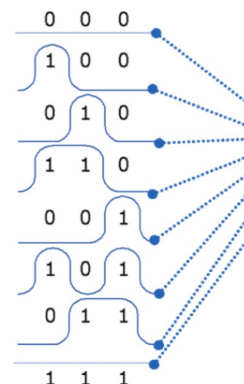
技術挑戰

- 構建Chiplet標準化、隨插即用互通介面
- 高速、高面積效率資料交換 (200 → 600 GB/s/mm²)
- 兼容先進/成熟製程，具通道補償、時序/通道自我校正

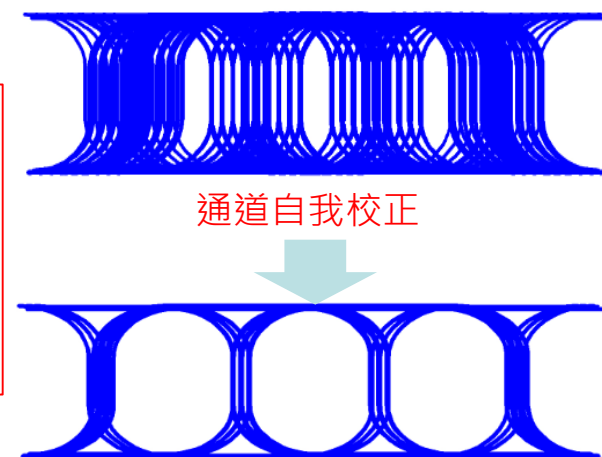


關鍵技術

- 延遲鎖相迴路數位化通道編碼、補償，適應性接收訊號補償
- 時脈降低($F_{ck} = \frac{1}{4} F_{data}$) 便於通道延展、核心邏輯電路整合同步、降低設計複雜度
- 可程式化輸出驅動級
- 彈性適應高性能(矽中介層)或低成本(載板)整合構裝



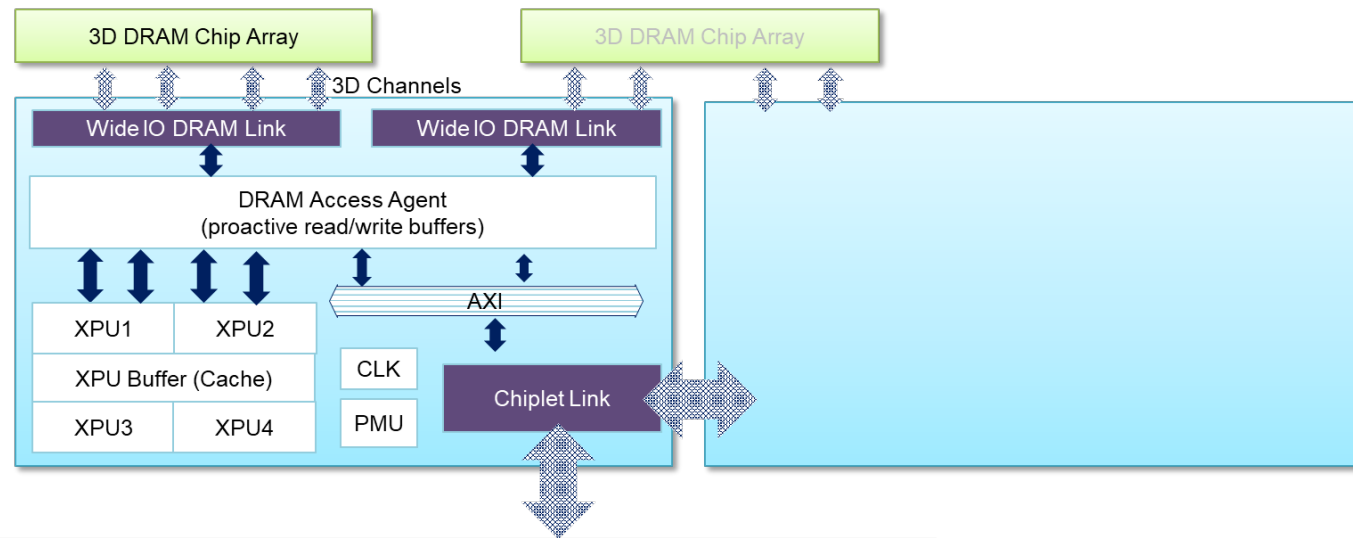
Channel /
Co-channel
Equalizer &
Timing
auto align



設計 - 高頻寬記憶體技術

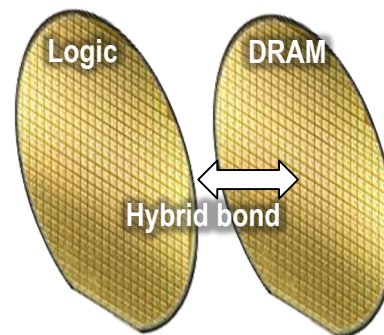
技術挑戰

- 運算記憶體(DRAM)一體化單元，構建個別Chiplet分散化記憶體
- 高速記憶體控制介面(12 → 40Gb/s)
- 晶片堆疊，實現高密度、低功耗(0.4pJ/b)、低延遲(55ns)

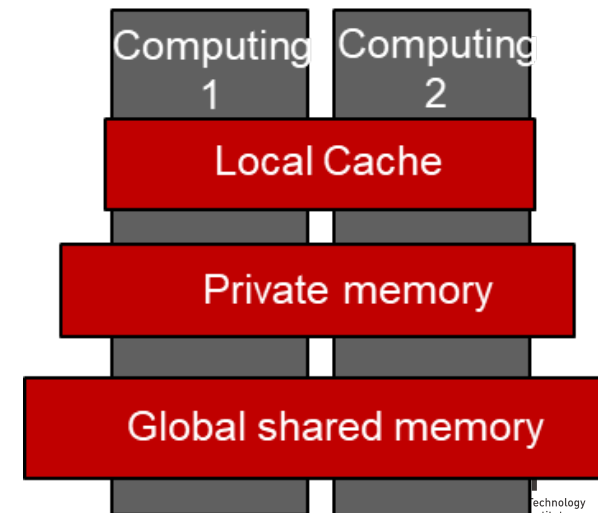


關鍵技術

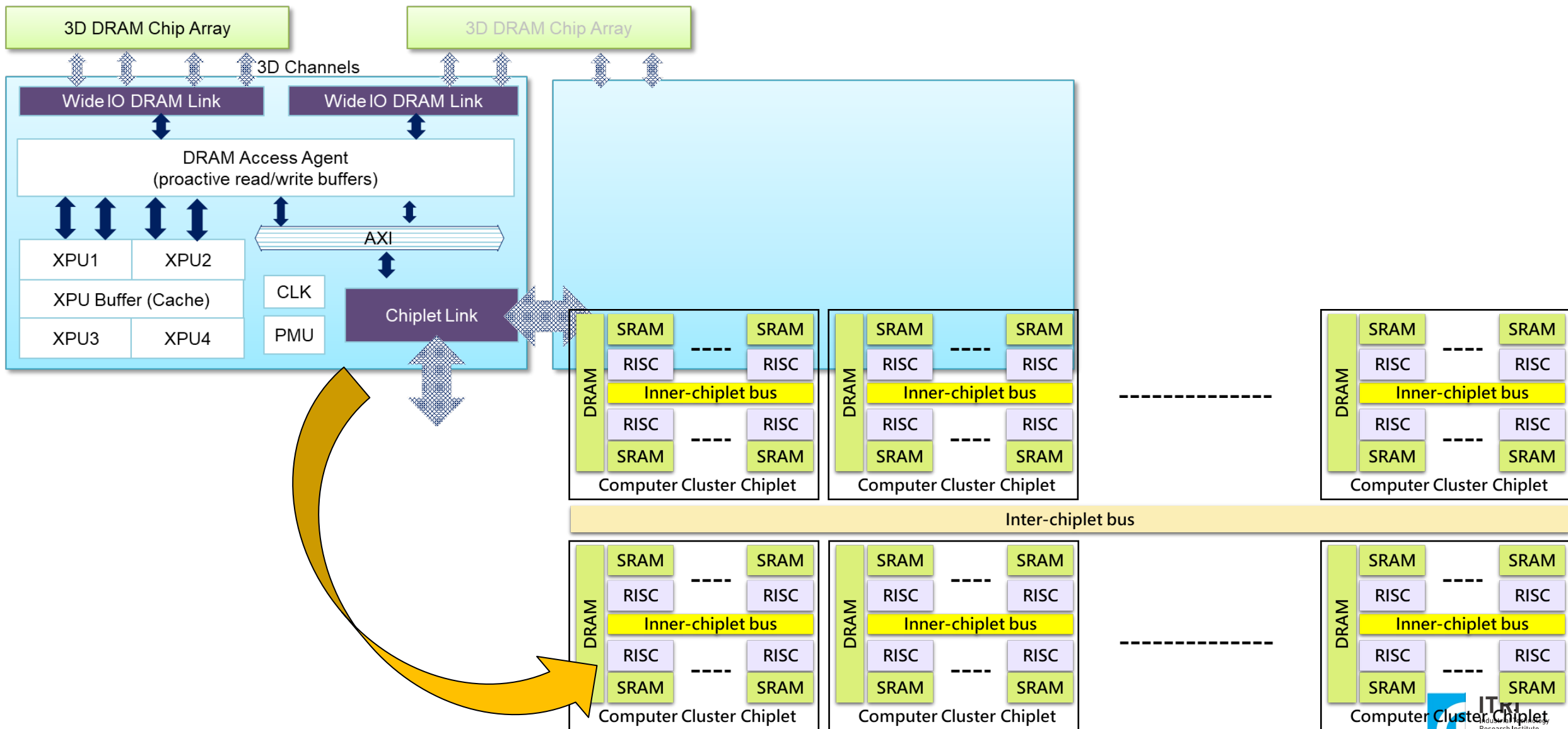
- 實體層：運算及記憶體，Wide IO晶圓堆疊整合，時序同步，時脈/電源閘控
- 管理層：階層式記憶體管理，邏輯化切割支援記憶體虛擬化管理



運算記憶體一體化



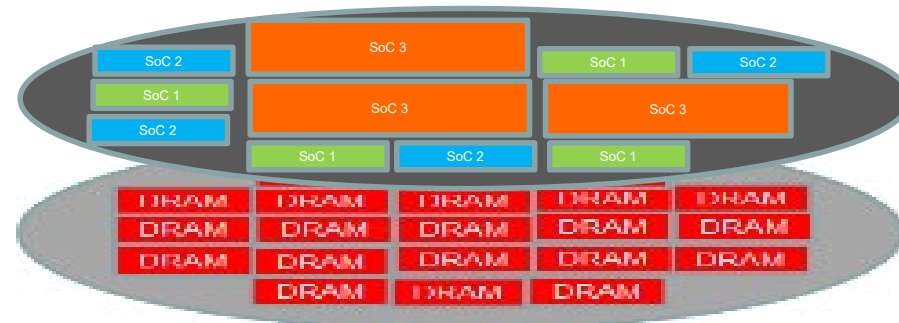
設計 – 從基礎核心運算小晶片擴展到多核運算架構



製造 – 重組晶圓低溫接合技術

技術挑戰

- Reconstituted Wafer低翹曲度
(Warp < 1000 μm @ 製程溫度 < 200 $^{\circ}\text{C}$)
- 絕緣材料/金屬共高度平面
(平整度 < 100nm)
- 材料低接合孔隙率
(孔隙率 < 10%)



多晶片重組晶圓

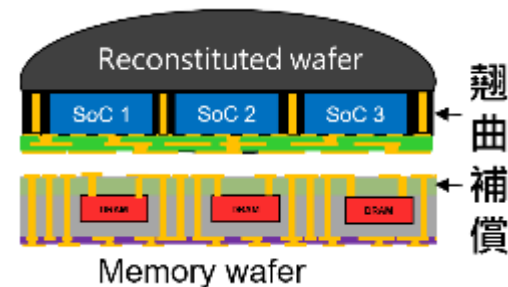
關鍵技術

- 多晶片Reconstituted Wafer技術
- 新型奈米級 CMP技術
- 晶圓級低溫有機材料對接技術



重組晶圓
翹曲過高

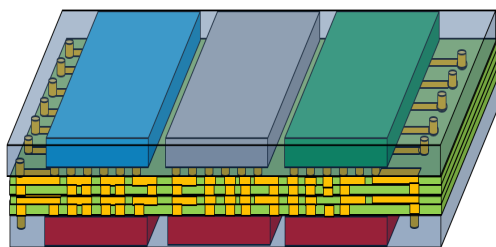
低熱積存
翹曲補償



製造 - 異質整合製造POC服務

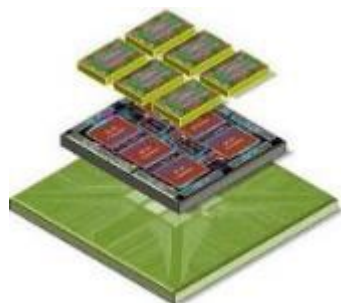
建置Reconstituted Wafer構裝設計ADK，提供POC設計驗證服務
與TSRI完成NDA簽署，提供Fan-in與Fan-out WLP shuttle service

通用小晶片



+

客製小晶片



客戶既有
小晶片

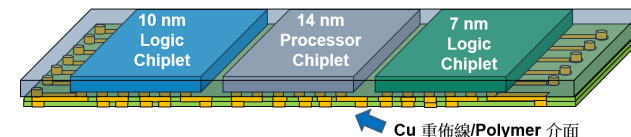
+

建立Reconstituted Wafer
構裝標準化設計
Chiplet材料構裝設計



ITRI Standard
Design Rule

Chiplet材料接合技術POC
平台



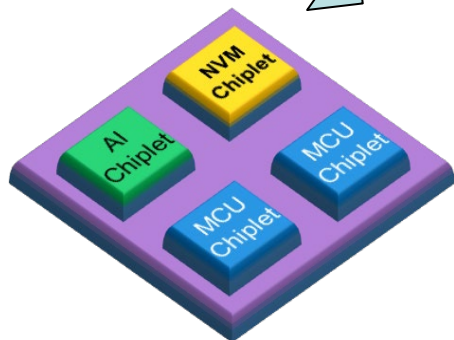
衍生合作方案規劃

應用 - LLM

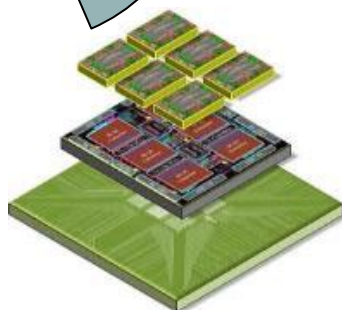
- 晶片系統架構探索
- 虛擬平台
- 設計輔助工具

- 客製化生成式AI模型開發
- 邊緣運算應用

設計 - Chiplet



製造 - 多樣化異質整合



- 運算Chiplet
- AD/DA、特定用途Chiplets



INNOVATING
A BETTER FUTURE



ITRI
Industrial Technology
Research Institute

衍生合作方案規劃

1. 由IP到系統

已具備IP：NPU(ITRI DLA), MCU (RISC-V 32)

合作方向：由IP建立 system level model，建立虛擬平台，架構探索及優化

2. IP及電路資源互補

已具備IP：65nm Power, PLL, SerDes, AFE, 及部分28nm基礎IP

晶創將投入：16nm IP

合作方向：<10nm IP,