

國科會工程處

113 年度「高效能晶片關鍵技術與創新應用計畫」

分項一：高運算力晶片

壹、計畫背景及目的

隨著人工智慧與各式前瞻應用的開展，未來下一世代的電子產品與技術的關鍵在於高效能晶片系統，而製作與設計高效能晶片所需的相關技術為重中之重。因此，本工作項目主軸在於強化我國學術界所研發之矽基半導體晶片與系統的運算效能，結合 2D/3D 記憶體의 整合設計與先進製程技術，能夠達到並進一步超越高效能運算(High Performance Computing, HPC)的指標。國內學術界雖然受到經費及人力資源的限制，因此無法使用業界半導體高階製程晶片，但在各領域均有傑出人才與充沛經驗，計畫團隊仍可發展相關技術領域，透過本計畫蓄積充沛研發能量與相關技術成果，以期結合邏輯與記憶體의 整合設計，同時開發並使用先進封裝技術，建立製作高效能晶片系統所需的異質整合平台研發能量，能為國內持續提供高效能的運算晶片做出貢獻。

本工作項目將規劃五年為階段性推動目標，計畫內容包括系統架構、運算與高速傳輸關鍵設計、軟硬體技術整合、異質整合與實體驗證等領域。計畫年度目標的制定可依照計畫團隊之專長及預計執行進度以模擬、技術開發、軟硬體整合、測試載具及實體驗證漸次展現。

關於本計畫的預期關鍵成果，初期包括高階模擬分析技術的開發、使用測試載具實質展示異質整合平台使用於高效能運算系統所需的 2.5D 關鍵技術；計畫執行中期進行運算及高速傳輸關鍵晶片下線與系統模擬；預期在計畫結案時能有 HPC 雛形系統展示，而針對大型語言模型(>10B Model)的計算力能夠達到每秒 1 POPS 以上，滿足大型語言模型高維度與超大權重參數的運算需求。

希望透過本計畫的執行，掌握關鍵的運算及高速傳輸關鍵電路設計，與實現異質整合平台，並能夠有效的跨領域整合所需技術，完成軟硬體的協同整合，依序以模擬或是實體驗證的形式，完成 HPC 系統架構規劃、HPC 系統軟硬體的偕同驗證、高效能運算的半導體系統晶片、HPC 系統應用的開發，並展示能

處理大型語言模型的高效能運算解決方案。

貳、 研究議題範疇

原則上以 16/7 奈米製程評估高算力晶片方案，研究主軸包含但不限於下列領域：

1. 高算力晶片架構與電路設計：採用同質(單態樣運算核心)、異質運算(多態樣核心，如 CPU+GPU+AI Engine)的整合設計，包含內建處理器(ARM, RISC-V, ...)以及客製化的硬體加速器(AI Accelerators)，並具有可延展性的實現方案，針對 8bit(定點或浮點運算)的大型語言模型(>10B Model)的運算需求，達成全程 POPS 的算力指標。
2. 整合高頻寬記憶體的高算力晶片模組：探討不同記憶體模組(DDR4/5, HBM, AIM, ...等)及軟硬體整合，有效提升整體運算效能，並以 MLPerf 的評比指標展現系統效能。
3. 超低節能的資料傳輸解決方案：透過異質整合的堆疊技術，包含晶圓層級的封裝與小晶片的可程式化先進封裝，滿足大型語言模型運算所需求的超大頻寬與節能傳輸需求。
4. AI 邊緣應用之完整解決方案：為了推動半導體產業與 AI 符合在未來生活各面向之需求，提出 AI 晶片邊緣應用及設計的完整解決方案，包括 edge inferencing、edge learning、驅動軟體、低耗能作業系統，進一步帶動晶片設計研發產業創新與社會發展。

參、 計畫重點

隨著新興應用的蓬勃發展，高運算技術受到高度的重視，本分項的研發目標希望由系統層面考量，來提升整體的運算力、能源效率及記憶體頻寬。技術開發的項目，除了引進 FinFET 製程外，電路及架構設計的創新，以及 2D、2.5D、3D 異質整合技術等，都是發揮最大效能及算力之關鍵技術，能由不同層面切入來突破現有的瓶頸和限制，達到預期的技術指標。除此之外，也期待研發團隊能強化產學與國際合作，加速前瞻科研發展，甚至於促成創新創業、生活應用，以提升國內晶片設計產業的發展及競爭力。

分項二：高頻、高功率電路與模組

壹、計畫背景及目的

高效能的運算需要有巨量資料的連結及傳遞，因此高速的通訊將是未來高性能晶片驅動的重要方向之一。5G 通訊是現在進行式，所提供的頻寬及通訊品質仍然快速進步當中，但隨著萬物聯網，感測資料的爆炸性成長，更高資料量及低反應時間是關鍵技術。巨量的通訊需要更高頻的無線通訊，由於美國聯邦通訊委員會(FCC)已有計畫開放 95 GHz 到 3THz 範圍內的 THz 頻譜執照，這股技術趨勢將使無線通訊的中心頻率超越 100GHz 並往太赫茲(Tera Hertz, THz)邁進，同時提高 THz 通訊在經濟層面的可行性。另外，隨著半導體 FINFET 前瞻製程及毫米波電路設計進步，軟體無線電(Software-defined radio)將在未來通訊晶片中扮演重要角色，此種射頻電路彈性，因應使用者的服務需求與射頻環境而調整無線電系統。

而當大量的資料傳遞到雲端的高效能運算中心，高速運算晶片需要諸多的小晶片支援，如多核心 GPU 及寬頻的記憶體，運算晶片的資料吞吐量是極為巨大的，晶片間的有線連結也將由目前幾十 Gbps 進展超過 1Tbps 的傳輸量，要達到快速通訊及高效能的運算，有效率的晶片連結技術也勢必發展在追求極致的通訊，能量效率也必需同時達成，因此利用不同的半導體製程(如與矽光子的異質整合)結合創新高頻、高速電路，來完成複合式的通訊系統，也將是研發重點。

目前環境永續淨零碳排為全球相當矚目的議題，而高效率高功率及小型化的功率轉換系統在其中扮演一個非常重要的角色。面對一些新興應用，例如智慧電網，電動車，B5G/6G 小型基地台(small cells)，超大規模(hyperscale)資料中心，無人載具及無線功率傳輸(wireless power transfer)等，全球持續投入高功率密度及高操作頻率的小型化功率轉換系統的研發。本計畫將聚焦於利用新興化合物半導體電路與模組及異質整合技術，開發下一世代的高效率及微小化的功率轉換系統。

貳、研究議題範疇

採用非 CMOS 製程(如 SiC, Ga₂O₃, GaN, InP, Silicon Photonics, …等)，探討

高頻、高功率電路設計與模組的研發，建議的研究主軸包含但不限於下列領域：

1. 高頻電路與模組：應用於下世代行動通訊(含低軌道衛星)的前端毫米波積體電路(MMIC)，如功率放大器，低雜訊放大器，射頻開關等。
2. 高功率電路與模組：高壓、高電流功率電路與模組，應用如電動運輸載具及智慧電網等；低壓、高操作頻率及高功率密度電路與模組，應用於資料中心與 AI 伺服器的電源供應；以及前述電路與系統之先進散熱技術與封裝。
3. 矽光子技術：超高速超低功耗光連結技術、異質光電整合技術、超高頻寬驅動及接收電路、光電晶片先進封裝與測試等，以應用於 AI 晶片連結、AI 高速運算、資料中心。

參、計畫重點

非 CMOS 製程高頻、高功率電路與模組應用之落實，在高頻方面以符合 W-band 甚至於 D-band 的操作頻率為目標；在高功率方面以 3300V 高壓之系統應用為主。而低壓、高操作頻率及高功率密度的 AI 晶片供電應用，應達成輸出功率、功率密度、效率分別大於 50 W、2500W/in³ 與 90% 為開發之技術指標。此外，矽光子應用的開發重點，則以異質積體化先進封裝技術，以達到矽光子光收發機單通道 200Gb/s (功耗<5pJ/bit；低延遲<100ns)，且具有 WDM 技術，可在兩個波長以上進行傳輸，以突破現有門檻之應用目標。

分項三：前瞻技術挑戰

壹、計畫背景及目的

現階段所有的資料驅動應用，除了強化高算力與高傳輸的晶片設計外，如何有效降低(一)運算的功耗，達成 fJ/b(每位元運算低於 10^{-15} Joule)等級的節能運算，透過創新的演算模型、新型態運算架構、與電路設計的探討，提供下世代節能運算的解決方案，已成為生成式 AI 邁向應用普及化的重大挑戰議題。(二)資料感測的時間與所需求的功耗(pJ/bit)，提供高解析、高可靠的原始感測資料，滿足多元應用的需求，亦成為感測晶片能否帶動創新應用服務的關鍵議題。以具有深度資訊的影像感測晶片為例，除了是無人機、ADAS、自駕車等應用中不可或缺的關鍵技術外，在許多創新產業及應用中，也需要使用此類型的感測晶片與裝置。因此，在本項目的推動中，主要鼓勵研究團隊採用矽基製程，探討創新應用與服務所需求的關鍵技術與晶片，透過異質整合堆疊與小晶片封裝技術，針對各類場域所需求的超低能耗運算、高時間/空間解析度、低延遲感測時間、高節能效率的運算與感測晶片，結合軟硬體設計技術，達成關鍵晶片的技術指標與模組的雛型系統，也具體呈現創新設計方案的價值與潛在落地應用的效益及影響力。

貳、研究議題範疇

本分項建議的研究主軸包含但不限於下列領域：

1. 以 CMOS 製程，探討超低功耗的演算法、系統架構、以及電路實現方案，達成能耗低於 fJ/b 的推論功能。
2. 高解析、低延遲感測晶片：以具有 SPAD 元件或類似的高壓製程，完成高速與低延遲的深度影像感測晶片與模組應用，達成之技術指標規格如下：(a) 影像數據量(含深度)： $> 1\text{Tb/s}$ 、(b)深度資訊： $< 1\text{cm}$ 、(c)能源效率： $< 1\text{pJ/b}$ 。
3. 以異質整合方式結合特殊應用感測晶片與高算力晶片、系統軟體，引領產業應用與服務創新者。

參、計畫重點

本分項將針對下世代運算的超低功耗運算與極低延遲、低能耗的智慧感測解決方案，研發關鍵晶片與軟硬體系統異質整合技術，並鼓勵學研團隊提出落地應用案例與雛形展示系統，透過具有國際競爭力的技術指標，包括有效降低運算的功耗、提升感測敏感度、降低感測延遲時間、提升感測能源效率等，展現此分項關鍵晶片的特色與價值創造。