



三星電子成功研發業界首款搭載 AI 處理效能的高頻寬記憶體

全新架構將提升逾兩倍系統效能，同時降低 70% 以上耗電量

全球半導體技術龍頭三星電子宣布，成功開發業界首款整合 AI 人工智慧處理能力的高頻寬記憶體（HBM）— HBM-PIM。全新記憶體處理（PIM）架構賦予高效能記憶體 AI 運算能力，可加速資料中心大規模數據運作、高效能運算（HPC）系統以及 AI 行動應用的處理速度。

三星電子記憶體產品企劃部資深副總裁 Kwangil Park 指出：「三星研發的 HBM-PIM 為業界首款可程式化的 PIM 解決方案，專為 HPC、訓練與推論等多樣由 AI 驅動的工作量身打造。我們將以此突破性技術為基礎，進一步與 AI 解決方案提供者合作，聯手開發更先進的 PIM 應用。」

阿岡國家實驗室電腦運算暨環境與生命科學副主任 Rick Stevens 表示：「非常高興見證三星成功解決 HPC 與 AI 運算所面臨的記憶體頻寬 / 電力挑戰。HBM-PIM 的效能，及在 AI 應用類別中展現的功率增益表現令人印象深刻。期待透過雙方合作，針對其他阿岡國家實驗室感興趣的議題進行效能評估。」

迄今多數運算系統係以馮紐曼架構為基礎，該架構採用獨立處理器與記憶體單元傳送數百萬筆複雜的資料處理工作。此循序處理方式需將資料來回移動，在處理持續增加的資料時，會導致系統效能降低。

反之，HBM-PIM 於每個記憶體庫（儲存子單元）內加入 DRAM 優化 AI 引擎，直接提供資料儲存位置所需的處理能力，不僅達到並行處理，亦同時大幅降低資料移動。當套用三星現有 HBM2 Aquabolt 解決方案時，新架構便能展現超過兩倍的系統效能，同時降低逾 70% 的耗電量。HBM-PIM 亦無須修改任何軟硬體，遂能加快現有系統的整合速度。

三星針對 HBM-PIM 的研究論文經獲選並發表於 2 月 22 日的國際固態電路研討會（ISSCC）。目前三星已提供 HBM-PIM 予 AI 解決方案領域的領導合作夥伴，並透過 AI 加速器進行測試，預計將於今年上半年完成驗證。