

《各種處理器(PU)概念彙整》

作者：屏東大學會計系 周國華老師

2021/6/17 原文刊於 FB 社群：半導體產業討論區(元件物理與製程、IC Design 與 EDA)

CPU: 中央處理器，英文: Central Processing Unit。CPU 是電腦的心臟，其內部構造包含控制單元(control unit)、算術邏輯單元(arithmetic logic unit)、暫存器(register)及快取記憶體(cache)，它們之間透過匯流排(bus)來傳遞資料及指令。CPU 是為通用計算而設計，相較於 GPU，CPU 的核心數量較少。目前主流的 CPU 晶片架構有 x86、ARM 及 RISC-V 三種。過去曾經叱吒風雲的 MIPS 架構已經逐漸退出主流，轉向靠攏 RISC-V 陣營。

GPU: 圖形處理器，英文: Graphics Processing Unit。GPU 具有數百或數千個核心，經過最佳化，可並列執行大量計算。GPU 這個詞彙最早是由 NVIDIA 公司在 1999 年提出，它的功能最初與名字一致，是專門用於繪製圖像和處理圖元資料的特定晶片，後來漸漸加入很多其它功能，包含物理模擬、超級電腦運算、AI 運算..等。NVIDIA 公司的高階 GPU 晶片被大量用在全球超級電腦系統中，讓 NVIDIA 在超級電腦領域具有接近主宰的地位。

DPU: 資料處理器，英文: Data Processing Unit。DPU 是一種系統單晶片(SoC)，內容包含多核 CPU、高效能 NIC (網路介面控制器)及可程式化的資料加速引擎。DPU 使用 MIMD (多指令流、多資料流)架構，這讓它和使用 SIMD (單指令流、多資料流)架構的 GPU 有所區隔。最近幾年大量投入資料中心晶片業務的 NVIDIA 也是 DPU 概念最重要的推廣者，NVIDIA 認為專責資料處理的 DPU 能夠和負責通用運算的 CPU、專攻加速運算的 GPU 平起平坐，形成三足鼎立的態勢，實現資料中心規模運算(Datacenter-Scale Computing)的願景。

TPU: 張量處理器，英文: Tensor Processing Unit。TPU 是 Google 開發的專用積體電路(ASIC)，是 Google 為其 TensorFlow 框架設計的晶片，專門用於加速機器學習，目前已發展到第四代。TPU 被用在 Google 街景、Google 相簿的大量圖像及文字辨識上，也用在 AlphaGo、AlphaZero..等 AI 產品中。TPU 晶片並未對外販售，但企業用戶可透過購買 Google 的雲端運算服務使用到 TPU 的功能。

VPU: 視覺處理器，英文: Vision Processing Unit。VPU 是被 Intel 併購的 Movidius 公司開發的專注於影像分析的處理器，可用在需要進行快速智慧辨識的自動化機器上。Intel 有獨立販售的 VPU 產品線，但市場通常會把它們歸類為 GPU 的一種產品。

NPU: 有兩種概念，其一為網路處理器，英文: Network Processing Unit；其二為神經網路處理器，英文: Neural-network Processing Unit。NPU 比較不具備獨立的產品概念，其功能通常被包含在內含 CPU 的 SoC 晶片裡。例如蘋果公司這幾年推出的 A 系列仿生晶片，擁有強大的神經網路處理功能，能夠在手機上實現每秒數億次的運算，讓手機可以獨立進行 AI 影像辨識。

除了以上介紹的幾種 PU 概念，市場上也有很多其他名稱的 PU，通常是 ASIC 產品，例如：APU (加速處理器，Accelerated Processing Unit)，BPU (大腦處理器，Brain Processing Unit)，另外一種 DPU (深度學習處理器，Deep learning Processing Unit)，FPU (浮點處理器，Floating Processing Unit)，HPU (全息處理器，Holographic Processing Unit)，IPU (人工智慧處理器，Intelligence Processing Unit)，RPU (無線電處理器，Radio Processing Unit)...

附圖表：(以下資料分別擷取自蘋果、Intel 及輝達公司網站內容)

