

國立臺灣大學 電機資訊學院 資訊工程學研究所、資訊網路與多媒體研究所
數位、人工智慧與新興運算架構
Computing Architecture for Digital, AI and Emerging Computing
作業 (2023/10/06 前繳交)

姓名:林東甫

學號:R12631055

(1) 若你打算採用 Nvidia H100 SXM 來組成超級電腦來提供 AI 訓練，使用的運算精度為 FP16。而經過查詢規格，H100 SXM 的 FP16 Tensor 核心運算力為 1,979 Tera FLOP/s。

若你需要的計算力為 1 Exa-FLOP/s。先不算及整個系統、網路與記憶體存取的性能損耗，單純考慮計算力。請問你至少需要採購幾個 H100 SXM，才能提供這樣等級的算力？

已知 $1 \text{ Exa-FLOP/s} = 1,000,000 \text{ Tera-FLOP/s}$ ， $1,000,000 / 1,979 = 505.3$ (小數點第一位後捨去)

Ans: 至少需要 506 個 H100 SXM

(2) 若你在設計一個 IC，因為晶片面積成本受限，你可能有兩個設計方案

(A) 大核 CPU + 小核 CPU

(B) 大核 CPU+ 具有運算功能的 GPU

請問 (A)(B)兩個方案，各有怎樣的優缺點？各自適合怎樣的應用領域，各舉一個例子說明。

(A) 大核 CPU + 小核 CPU：

優點：

1. 適合多任務處理：小核 CPU 可以處理輕量級任務，而大核 CPU 可以處理較大且複雜的任務，使該方案適合多任務處理的應用。
2. 較低功耗與溫度：小核 CPU 可以在處理簡單任務時使用，而不需要動用大核 CPU 因此便可以實現較低的功耗與溫度，有助於節省電池電量與延長其壽命。
3. 對各種類型任務的處理能力都有一定水準，因為 CPU 本來就是為了多用途目標任務而開發的，大核搭配小核可以使其針更多不同的任務種類更加有效率。

缺點：

1. 有限的圖形處理性能：因為沒有專門的 GPU，對於大量同質性任務例如在圖形密集型應用中只能依靠有限的核心，而除了圖形密集型應用，近年來流行需要大量浮點式計算資源，諸如 AI 深度學習應用這類任務中的性能可能較差，經常會需要額外再搭配 GPU。
2. 與典型多核心相仿，對於單執行緒的任務，若沒有專門的系統程式設計支援，其表現效能並不好。

適用應用領域範例：個人電腦 PC，大型伺服器或工作站，這類設備經常需要處理多種類型任務。

(B) 大核 CPU + 具有運算功能的 GPU：

優點：

1. 強大的圖形處理能力：GPU 在處理圖形和浮點式運算方面表現出色，適用於今日相當流行的影音、遊戲、3D 等圖形密集型應用，而在這類型任務需求不斷增加的當前，GPU 的重要性也與日俱增。
2. 適合大量同質性任務諸如深度學習和科學計算：在英偉達主導的強大框架生態系 CUDA 之下，GPU 能夠加速深度學習模型訓練和科學計算，今天 GPU 已經是人工智慧研究領域的主流。

缺點：

1. 較高的功耗：GPU 通常消耗較多的電力，影響電池續航時間，並且因為主要使用大核來進行多樣性的任務，因此整體的功耗與溫度都會在比較高的狀態。
2. 對於多用途任務的效率差，因為只能靠大核來進行，相較於多個小核，整體計算成本也比較高昂。

適用應用領域範例：智慧型手機、平板電腦或遊戲主機、深度學習訓練伺服器、大型科學計算集群。

(3) 目前在 Digital 領域，有很多新領域的技術在進展，包括 IC 運算、超級電腦、5G/6G 通訊、衛星通訊、新太空經濟、AR/VR/Metaverse、分散式計算。就你的看法，這些新技術的進展，能否結合成新的應用方式？就你的觀點，請用 1~2 頁的篇幅，提出未來可能出現的技術或應用，描述其優點與缺點，與對世界的影響範圍，並討論要讓這樣的應用誕生，必須要有的先決條件 (例如需要那些技術,商業模式或法規)。

我認為可以考慮在太空中建置超級電腦，也許可以像國際太空站或通信衛星、觀測衛星一樣在近地軌道環繞，或甚至是在深太空或月球上建置，端看需要的應用場景而定。台灣本身在晶片與零組件與電腦系統領域擁有從設計到製造測試的完整產業鏈，也是全世界少數擁有衛星與探空火箭科技的國家，若能採行並成功建置太空超級電腦，未來可能可以實現的應用場景諸如：

1. 科學研究與太空任務支援：隨著火箭發射成本降低與 5G/6G 通訊技術的進步，可以預期未來會誕生非常多在太空中進行的工程與科學任務，如建造太空站、維修衛星等，對於宇宙學、氣象學和地球科學也會更進一步加速科學研究進展，而如果要將這些工作都傳輸回在地表上超級電腦，不僅要考慮通訊延遲的問題，也會因為大筆資料的傳輸而佔據大量的太空頻寬，因此太空超級電腦可以直接用於就近處理大量的科學數據、太空任務支援等，用於支援太空任務的計算需求，其中也包括工程或支援任務諸如衛星控制、導航、軌道計算等。
2. 緊急救援與即時性任務應對：同上，太空超級電腦可以用於各種災害監測和緊急支援，不再需要將資料傳輸回地表再運算，而可以提供重要的實時數據處理能力，等於是把雲端服務擴展到了宇宙。

但同時也存在一些技術和成本上的限制。要使這樣的應用誕生，先決條件主要取決於以下因素：

1. 技術挑戰：在太空中建立超級電腦需要克服諸多技術挑戰，因為太空的環境風險因素包括太空輻射、太陽耀斑、亞原子粒子、微流星體等，以及若採用太陽能可能要面對不穩定的電力和冷卻等因素，因此需要在硬體條件上克服諸如在太空中的電子元件的耐輻射性、熱控制、能源供應、通信和維護等。太空環境的極端條件都是對技術上的高度要求。
2. 成本與政策法令環境：雖然 SPACEX 的可回收火箭的成功導致發射貨物的成本大為降低，但跟一般類型之運輸相比，太空任務的成本仍然極高，除了發射成本以外，政策與相關規定也是一大問題，2019 年晉陞太空科技公司在台東縣達仁鄉南田村達仁橋下建蓋火箭發射場，由於事前未徵詢部落意見及獲得同意，致使當時首發試射中斷，後因程序不符公聽會要求，遭達仁鄉公所否決。這個案例也凸顯出台灣在太空任務與太空產業相關的法規與政策環境仍然相對不成熟。
3. 可維護性：在太空中的超級電腦可能需要進行遠程維護和修復，這需要高度可靠的自主運作和遠程監控系統，而人工智能在這方面的研究，也許也能在這方面有所幫助。

若能夠克服以上種種條件才有可能可以實現，可以參考 2018 年 NASA 與 HPE 所合作的 Spaceborne 計畫，其採用 HPE Apollo 40 系統，具備 32 個核心及大量的固態硬碟與記憶體，可提供 1 teraFLOP 的運算能力，背誦到了國際太空站，進行了整整一年的測試與實驗，足以證明它的可行性，若未來可以進一步對它進行擴充，可以預期有更多的應用服務，不過在大方向的上優缺點仍然會存在，諸如：

優點：

1. 無重力環境：在太空中，沒有地心引力也沒有大氣層，這意味著超級電腦可以更有效率地進行冷卻，並允許新型硬件架構和冷卻技術的應用提高其性能。
2. 就近提供高性能計算：太空超級電腦可以提供計算能力，用於處理大規模的數據和複雜的模擬，幫助可見的未來中將會產生的各種太空科學和工程任務。
3. 零干擾通信：位於太空中的超級電腦可以實現零干擾通信，不受地球上的無線電波干擾，也不會佔用到近地大氣層內的高空頻寬。

缺點：

1. 高昂的成本：太空任務成本非常高昂，包括發射、建設和維護費用等，可能成為這項工程的門檻。
2. 可維護性：太空中的超級電腦可能需要高度自主可維護性，只要是需要人力解決的軟硬體故障和問題，都可能是不小的挑戰。