

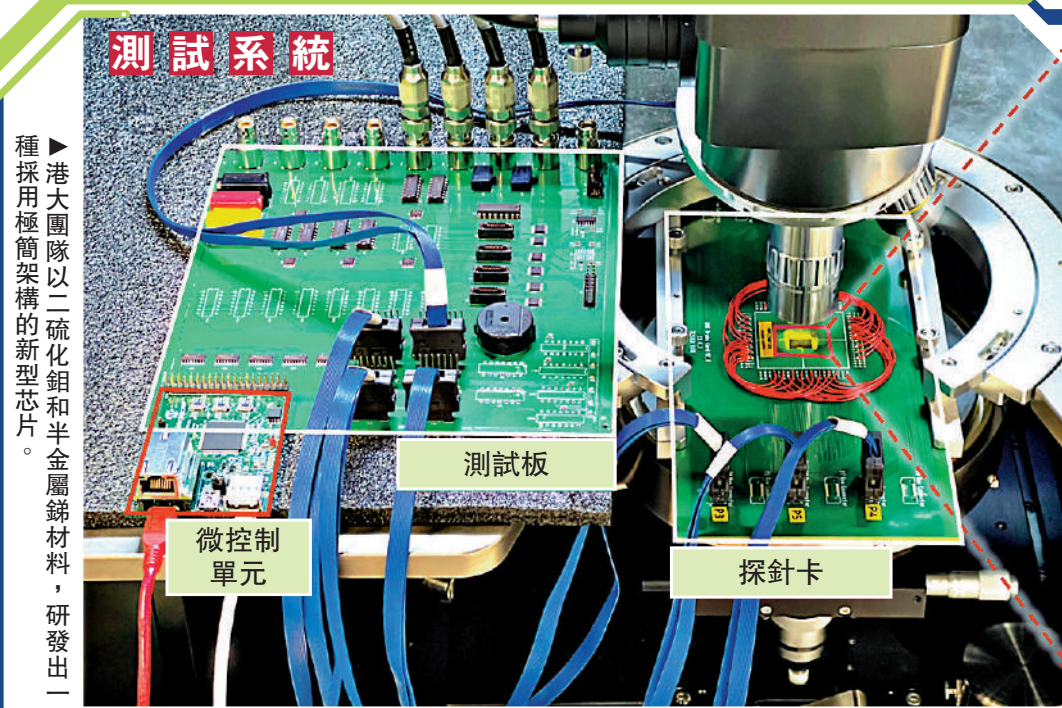
新皇崗口岸 本月將四次大演練

A9



A7 聯招今放榜 港大取錄七超級狀元

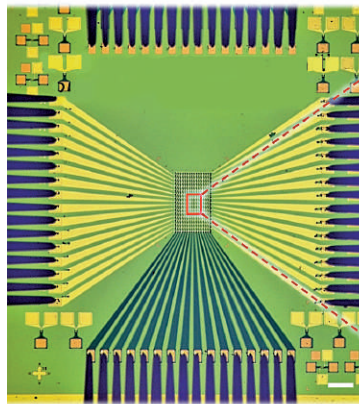
A3 港零售銷售漲4.6% 連續14個月上升



►港大團隊以二硫化鉬和半金屬銻材料，研發出一種採用極簡架構的新型芯片。

晶片架構

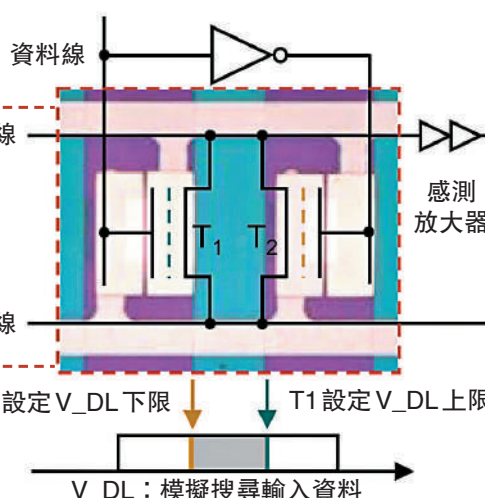
利用原子級超薄二維二硫化鉬（MoS₂）閃存構建的模擬內容尋址存儲器（CAM）陣列晶片



包含 256 個 MoS₂ 快閃記憶體
的 8 × 16 模擬 CAM 陣列

電路圖

MoS₂ 模擬 CAM 電路單元



新物料製成 比CPU快1億倍

港大研原子級芯片 重塑AI算力極限

重大突破



在人工智能時代，我們的手機每次進行人臉識別或語音助手喚醒，背後都涉及海量數據在記憶體與處理器之間「來回搬運」。這種搬運不僅造成明顯延遲，更讓電池電量急速消耗，此乃困擾計算機界數十年的「馮·諾依曼瓶頸」。

香港大學工程學院李燦教授領導的研究團隊，近日在國際頂級學術期刊《自然——納米技術》（Nature Nanotechnology）上發表一項突破性成果：他們利用比頭髮絲還細數萬倍的二維材料二硫化鉬（MoS₂），研發出一種採用極簡架構的新型芯片，即原子級「模擬存內搜索」芯片，能夠直接在數據存儲的位置完成搜索。其處理速度比標準CPU快了約一億倍，並在多個數據集測試中獲得了極高的準確率。團隊表示，這項技術為下一代高性能、低功耗AI硬件提供了清晰的發展藍圖；至於應用場景，團隊舉例智能手機的面部識別技術，便可以直接在裝置上瞬間完成，而毋須將私隱數據上傳至雲端。

大公報記者 苑向芹、盛德文

二維材料較髮絲細數萬倍

據了解，這款以比頭髮絲還細數萬倍的二維材料打造的芯片，核心是一種名為「模擬內容尋址存儲器」（CAM）的新型記憶體。傳統的同類產品基於靜態隨機存取存儲器（SRAM），每個存儲單元需要多達16個晶體管，導致芯片體積龐大、功耗極高。港大團隊的模擬CAM僅需兩個晶體管即可構建一個單元，面積大幅縮小，還能直接處理更接近真實世界的連續模擬信號。

能耗低創紀錄「超光速」搜索

將二維材料連接到電路時，會遇到一道限制電子流動的「肖特基勢壘」，它就像公路上的收費站，嚴重拖慢電子速度。港大團隊使用半金屬銻（Sb）作為接觸電極，成功拆除了這道屏障，創造了一條暢通無阻、極低電阻的通道。由此產生的器件不僅速度極快，而且效率驚人：這款芯片的單次搜索耗電創下了低於0.1飛焦耳（fJ）能量的世界紀錄，延遲僅為36皮秒（Picosecond，時間單位，等於一兆分之一秒）。這是什麼概念？光在真空中一皮秒僅能走0.3毫米；這意味着搜索完成時，光甚至還沒來得及跨越芯片本身。

團隊進一步展示了該技術的進階潛力。透過三維異質整合技術，他們

在單個存儲單元內垂直堆疊了不同類型的二維材料，使所需面積減半，功耗進一步降低。

在執行機器學習分類任務時，新型模擬CAM的處理速度更比標準CPU快了約一億倍，並在多個數據集測試中獲得了極高的準確率。

李燦教授解釋道：「從應用層面來看，搜索操作具有極大價值。作為大語言模型（LLMs）基礎的『注意力機制』，本質上就是一個搜索過程——從壓縮的知識庫中提取資訊。我們正在探索如何利用這種高性能硬件，來實現大型AI模型的搜索機制。」他還特別提到這項技術對隱私保護的意義：「例如智能手機的面部識別技術，可以直接在裝置上瞬間完成，而毋須將私隱數據上傳至雲端。」

至於從科研到落地的進程，李燦

教授坦言，大規模商業化仍需克服封裝方面的工程挑戰，但這項研究已成功驗證了該原理，為下一代高性能、低功耗AI硬件提供了清晰的發展藍圖。

為下一代AI硬件提供發展藍圖

據悉，這項突破是團隊通力合作的成果，論文共同第一作者高國雲博士主導了器件及陣列的製造，聞博先生則負責陣列測量。



▲港大李燦教授領導的研究團隊，在「存算一體」領域取得重大突破。

「馮·諾依曼瓶頸」是什麼？



馮·諾依曼瓶頸是指電腦中央處理器（CPU）速度與記憶體資料傳輸速度不符的現象，即CPU的運算速度遠快於記憶體的資料傳輸速度，導致處理器大部分時間都在空等資料，浪費運算能力。

◀新技術可應用在不同場景，例如以面部識別技術解鎖手機，可直接在裝置瞬間完成。



傳統架構 耗時耗能

分開放置：處理器負責計算，記憶體負責存儲，形成結構性缺陷。

通道共用：指令和資料共用同一條傳輸路徑，CPU在執行時必須輪流讀取指令和數據，無法同時進出。

速度落差：晶片技術進步極快，但記憶體讀寫速度跟不上，造成嚴重塞車。

解決方法

改進架構：現代電腦會使用快取記憶體（Cache）來緩解速度差，或是採用「哈佛架構」將指令與資料分開走不同通道。

存算一體：將運算單元直接嵌入儲存單元中，實現資料儲存與處理的融合，在記憶體裏完成運算工作。港大最新研究就是這方面的技術。

非馮·諾依曼架構：發展如神經形態運算、量子電腦及光子電腦等新型運算模式，模仿人腦「存算一體」的分散式處理機制。

在記憶體運算 省時節能

存算一體

香港大學李燦教授研究團隊日前在國際頂級學術期刊《自然——納米技術》（Nature Nanotechnology）上發表論文，指出隨着人工智能與智慧裝置快速發展，傳統晶片發展和運算遇到了一大瓶頸：運算資料總要在「處理器」和「記憶體」之間跑來跑去，不但耗時且非常耗電。為打破此限制，港大的研究團隊將目標轉向「存算一體」技術，讓記憶體直接具備計算與搜尋功能。有專家認為，「存算一體」技術正處在從實驗室走向產業化的爆發前夜，全球競爭激烈。香港在基礎研究和特定技術點上表現突出，而這項技術本身被視為計算架構的「範式革命」，影響意義深遠。

全球攻堅技術產業化

根據網上資料，「存算一體」已不僅是學術概念，而是一個快速增長的市場，正處於從驗證到落地的關鍵階段。香港在「存算一體」的全球版圖中，扮

演着重要的基礎研究和關鍵技術突破者的角色，尤其在憶阻器應用和模擬計算領域特色鮮明。香港理工大學、香港大學等高校團隊都有相關研究。

發展「存算一體」芯片技術的意義主要包括四點：一、終結「數據搬運」時代：它從根本上解決了困擾電腦界數十年的馮·諾依曼瓶頸。數據「存在哪裏，就在哪裏算」，避免了在處理器和記憶體間的無效搬運；二、能效與速度的指數級提升：由於省去了搬運能耗，能效得到極大提升，有研究稱AI計算中數據搬運成本可達計算的100倍。同時，延遲也大幅降低；三、解鎖全新應用場景：超高能效使邊緣AI成為可能，如手機端側AI、可穿戴設備和自動駕駛等，毋須再依賴雲端；四、催生新器件與新生態：「存算一體」的發展極大推動了憶阻器等新型器件的研發。同時，全球首個RISC-V「存算一體」標準已啟動研製，預示着一個新軟體生態的誕生。

責任編輯：嚴偉豪 美術編輯：莫家威