CTIMES
元件・次系統・自動控制

2025 COMPUTEX
特別專刊
CHINESE-ENGLISH SPECIAL ISSUE

http://www.ctimes.com.tw

Edge AI
邊緣運算

COMPUTEX
TAIPEI

我們打造先進設施的用意：
在您需要時提供您所需的零件。

上百萬款零件任君挑選，就在 digikey.tw，
或來電 0080-185-4023

DigiKey

we get technical

STMicroelectronics

STM32

STM32U3 系列

最佳低功耗表現，兼具強大安全功能

有效延長電池壽命，減少開發成本及提供數據保護，適用於工業，醫療及消費型裝置。

STM32U3 系列微控制器提供最佳低功耗表現，同時兼具強大安全功能及良好的成本效益。

搭載Arm® Cortex®-M33核心，運行頻率 96MHz。STM32U3 提供多種封裝選擇: 32至100腳位數，包含LQFP、UFBGA、QFN、和WLCSP。

備有RAM 256KB及高達1MB快閃記憶體，可執行雙Bank模式，支援靈活的記憶體配置。

主要功能

- 首款採用近閥值電壓設計STM32產品
- 最佳功耗表現
- 功耗可低至10µA / MHz
- 關機模式下，功耗可低至200nA
- 停止模式下，功耗可低至1.6µA
- 強大的安全功能，保護敏感數據及關鍵任務執行程序
- 具備周邊：I3C 及 FDCAN IP
- 簡化PCB設計
- 搭配STM32Cube生態系體，提供簡便的開發流程
- 寬廣的溫度支援
  (-40 °C / +85° C 和 +105 °C)

主要應用

- **活動裝置：**
  相較於前一代產品，可延長電池壽命7倍
- **瓦斯及水表裝置：**
  相較於前一代產品，電池尺寸可縮小至4倍，有效支援迷你裝置開發，突破更多放置空間侷限
- **工業GPS追蹤裝置：**
  相較於前一代產品，追蹤效率提升至兩倍

意法半導體
TEL: (02)6603 2588
FAX: (02)6603 2599

代理商
伯東：(02)8772 8910 友尚： (02)2659 8168 艾睿：(02)7722 5168
文曄：(02)8226 9088 安富利： (02)2655 8688 益登：(02)2657 8811

www.st.com/stm32u3

COMPUTEX TAIPEI

Special Issue

社長 黃俊義 Wills Huang

編輯部/

副總編輯 籃貫銘 Korbin Lan

資深編輯 王岫晨 Steven Wang

陳復霞 Fuhsia Chen

陳念舜 Russell Chen

產業服務部/

經理 曾善美 Angelia Tseng

主任 翁家騏 Amy Weng

特助 劉家靖 Jason Liu

發行部/ 資訊管理部/

主任 孫桂芬 K.F. Sun

專員 何宗儒 Dave Ho

會計 林寶貴 Linda Lin

TEL：(02) 2585-5526

www.ctimes.com.tw

Contents

編輯室

2025年，全球科技產業正處於一場深刻的轉型浪潮之中。AI、HPC、自動駕駛、量子科技與5G/6G通訊等前沿技術，正以前所未有的速度重塑各行各業的運作模式。AI晶片的需求激增，預計2025年全球生成式AI晶片市場將突破1,500億美元，推動整體半導體產業持續成長 。在這場科技革命中，台灣不僅是參與者，更是關鍵推動者。

台灣的半導體優勢，源自完整產業生態系統，包括IC設計、製造、封裝與測試等緊密協作。在地緣政治日益複雜的背景下，台灣的半導體產業也成為全球關注的焦點。其在先進製程的領先地位，不僅支撐全球科技產業的發展，也成為維繫國際供應鏈穩定的關鍵力量 。

適用於 NVIDIA® Holoscan 的 PolarFire® FPGA 乙太網路感應器橋接器

靈活整合助力從邊緣到雲端的 AI 驅動感應器處理

這款適用於 NVIDIA® Holoscan 的 PolarFire® FPGA 乙太網路感應器橋接器為從邊緣到雲端的 AI 驅動感應器處理帶來了靈活整合能力。此解決方案透過一體化架構實現多協議支援、卓越能效、安全防護與可靠性，現已獲得 NVIDIA 官方認證，可無縫整合至其廣受歡迎的 Holoscan 生態系統。

主要特性

- 高效節能的協定轉換
- 高度安全
- 單粒子翻轉 (SEU) 免疫設計確保高可靠性

這款適用於 NVIDIA Holoscan 的 PolarFire FPGA 乙太網路感應器橋接器全面解決方案以卓越的技術實力和以客戶為中心的設計而著稱，絕對是您的放心之選。

聯繫信息
Microchip 台灣分公司
電郵：rtc.taipei@microchip.com
技術支援專線：0800-717-718
聯絡電話：• 新竹 (03) 577-8366 • 高雄 (07) 213-7830 • 台北 (02) 2508-8600

MICROCHIP

microchip.com/Ctimes-EthernetSensorBridge

COMPUTEX TAIPEI

從邊緣推理到異構運算 看AI的全方位進化

AI正在深刻改變我們的生活方式與產業結構。然而，隨著AI推動運算需求指數級增長，電力消耗、隱私與安全等挑戰也日益突出。未來，AI將更加個性化，從被動響應工具演變為主動建議的智慧助手。

文／王岫晨

在2025年，AI市場將持續高速發展，涵蓋從生成式AI、邊緣AI到工業自動化等多元應用領域。隨著運算需求的增加，邊緣運算的重要性日益凸顯，企業對低延遲、高效率的解決方案需求持續攀升，促使AI技術向終端設備深入滲透。

邊緣計算是指將數據處理和計算任務從雲端轉移到靠近數據源的設備上進行。這種方式能夠減少數據傳輸的延遲，提升系統的即時性和隱私保護。邊緣設備通常資源有限，因此需要高效且低功耗的AI處理器來支持複雜的AI任務。

滿足邊緣與終端設備需求

為了滿足邊緣與終端設備的需求，AI處理器需要具備以下特點：

1. **低功耗：**邊緣設備通常依賴電池供電，因此AI處理器必須在低功耗下運行。
2. **高效能：**AI任務如圖像識別、語音處理等需要高效的計算能力。
3. **小型化：**邊緣設備的空間有限，AI處理器需要高度集成且體積小巧。
4. **即時性：**許多邊緣應用（如自動駕駛、工業控制）要求即時響應，AI處理器必須能夠快速處理數據。

Arm物聯網事業部亞太區資深經理黃晏祥指出，邊緣運算跟雲端運算主要的差異，應該是在於應用的最佳化vs.產品的泛用性，邊緣運算為了更接近邊緣端的使用場景，通常會有功耗與佈建成本的局限性，因此在解決方案的選擇上，從效能與功耗的產品平衡，與多重的選擇都是最重要的。

傳統MCU主要用於控制任務，而AI處理器則專注於數據處理和模型推理。為了滿足邊緣設備的需求，現代MCU開始集成專用的AI加速器（如NPU神經網路處理單元），以實現高效的AI運算。

應對高性能運算挑戰

針對2025年的整體AI發展態勢，Arm認為半導體產業將逐漸採用AI輔助晶片設計工具，以應對晶片設計日益增長的複雜性和市場需求。AI在佈局規劃、配電和時序收斂等方面展現了強大的優化能力，不僅能夠提升晶片性能，還能顯著縮短晶片開發週期。這一技術的普及，為中小型企業開啟了進入專用晶片市場的大門，使其能夠更具競爭力地開發創新產品。

儘管AI無法取代人類工程師，但它正逐漸成為應對現代晶片設計挑戰的重要工具，尤其是在高能效AI加速器與邊緣設備設計領域的應用。AI輔助設計工具協助處理龐大數據分析和決策任務，讓工程師專注於創新設計，推動整體生產力提升。

平衡運算需求與能源消耗

在全球市場中，如何平衡日益增長的運算需求與能源消耗，成為政府、產業及社會重點關注事項。據統計，全球數據中心每年耗電量達460太瓦時（TWh），相當於德國全國的用電量。因此，企業無不積極尋找不降低性能並能有效減少能源消耗的方法。

實現高性能、高能效AI的關鍵在於硬體與軟體的協同設計。從硬體層面來看，底層處理器技術和CPU架構的持續改進，將為AI運算提供更高效的處理能力。同時，專用硬體的設計也在針對密集型AI工作負載進行優化，包括網路、儲存、安全以及數據管理等多個層面。

此外，創新的軟體解決方案也在推動AI應用的高效運行。這些軟體通過智能優化AI工作負載，使其在資源消耗減少的情況下，仍能保持甚至提高性能。硬體和軟體的雙向協同，將進一步提升數據中心的運營效率，減輕能源壓力，為AI時代的可持續發展奠定基礎。

AI 技術全方位進化

隨著AI技術的發展，未來數年將迎來AI推理、邊緣運算及智能應用的多方位突破，進一步推動AI在各領域的廣泛應用和可持續發展。

AI推理持續增長

AI推理技術正在快速增長，特別是在文本生成和摘要等日常應用領域。隨著更多支援AI的設備和服務問世，智慧手機和筆記型電腦如今可完成多數推理任務，帶來更快、更安全的用戶體驗。為支撐這一發展，設備需要更高效的處理能力、更低延遲和更佳的電源管理能力。例如Armv9架構的SVE2和SME2功能正是實現高效AI推理的關鍵。

邊緣AI與去中心化趨勢

邊緣AI日漸重要。未來更多AI任務將直接在邊緣設備上運行，減少對大型數據中心的依賴，不僅節省成本與電力，還提升了隱私與安全性。邊緣設備執行即時檢測，雲端負責深度分析，這種模式對智慧城市和工業物聯網等市場尤為重要。

小型語言模型（SLM）的快速發展也是邊緣AI的關鍵推動力。Llama、Gemma和Phi3等更緊湊的模型，能夠在資源有限的設備上運行，實現更高效能和更強隱私保護。SLM將成為語言交互、圖像處理和本地化決策的主力，進一步促進去中心化AI的普及。

異構運算與多模態AI的協同發展

異構運算成為AI時代的核心趨勢。單一硬體無法滿足所有AI工作負載需求，CPU與AI加速器的協同至關重要。例如，NVIDIA的Grace Blackwell超級晶片將基於Arm Neoverse架構的Grace CPU與Blackwell GPU相結合，實現了更靈活的AI運算能力。這類異構運算架構將在2025年迎來更廣泛的應用。

另一方面，多模態AI模型的興起也將帶來更接近人類感知的能力。這些模型結合文本、圖像、音訊和感測器數據，能更全面地理解場景，開啟了AI在視覺、聽覺和行為分析方面的新時代。

智能應用與AI代理

AI應用正變得更智慧、更個人化。從智能助理到個人醫生，未來的應用程式將主動為用戶提供建議，滿足個性化需求。與此同時，AI代理的廣泛使用正在改變產業格局，從客服支援到編碼助理，各領域將受益於AI的高效協作能力。

AI技術的全面進化不僅推動了運算能力與應用層面的突破，也加速了隱私與安全的創新需求，為未來智能化社會奠定了基石。

傳統應用與AI之間的平衡

MCU在傳統應用中主要負責控制和管理硬體設備，例如家電、工業自動化、汽車電子等。這些應用對MCU的要求包括：

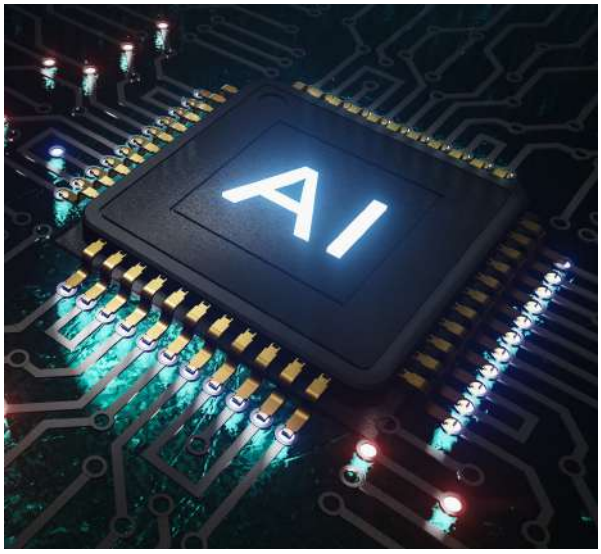
- **穩定性**：系統必須穩定可靠，能夠長時間運行。
- **即時性**：需要快速響應外部事件。
- **低功耗**：許多應用場景依賴電池供電，因此其功耗必須盡可能低。

隨著AI技術的普及，MCU需要支持更多的AI功能，例如語音識別、圖像處理、預測性維護等。這些功能對MCU的計算能力和存儲空間提出了更高的要求。然而，MCU的資源有限，如何在傳統應用與AI功能之間取得平衡成為一個關鍵問題。

處理器設計策略

為了在傳統應用與AI功能之間取得平衡，MCU的設計可以採取以下策略：

- **模組化設計**：將AI功能作為可選模塊，用戶可以根據需求選擇是否啟用。
- **資源分配優化**：通過動態資源分配，確保傳統控制任務和AI任務都能獲得足夠的資源。
- **多核架構**：採用多核設計，將傳統控制任務和AI任務分配到不同的核心上運行，避免資源競爭。
- **軟硬體協同設計**：通過優化硬體架構和軟體算法，提升MCU的整體性能。



圖一 為了提升AI運行效率，許多處理器開始整合專用的AI加速器。

為了提升AI模型的運行效率，許多MCU開始整合專用的AI加速器。這些加速器針對神經網路運算進行了優化，能夠大幅提升模型的推理速度。例如，Arm的Cortex-M系列MCU已經支持AI加速功能，能夠高效運行機器學習模型。

Arm 主任應用工程師林宜均指出，Arm 的 MCU 產品策略中結合了對傳統應用的深厚技術基礎，與對新興AI 技術發展的積極投入。產品不僅支援傳統的低功耗嵌入式應用，還加入了對 AI 推論的支援，尤其是在語音辨識、影像處理和感測器數據分析等場景。

Next-level wireless SoCs

Integrated multipurpose MCU

New ultra-low power radio

2x processing power

nRF 54 L SERIES

3x processing

MADE FOR DEVELOPERS

GET STARTED
nordicsemi.com/nRF54L15-DK

NORDIC® SEMICONDUCTOR

在MCU產品中導入AI功能的同時，必須兼顧傳統應用的需求，例如低功耗和即時性。例如Arm的Cortex-M 系列處理器，它在提供AI 處理能力的同時，也保持了低功耗的特性，適合應用於各種嵌入式裝置。

軟體可以釋放硬體的效能與潛力。Arm 除了提供硬體平台之外，也提供軟體庫和開發工具，協助開發人員進行軟體優化。

然而，隨著AI推動運算需求指數級增長，電力消耗、隱私與安全等挑戰也日益突出。業界正在透過創新硬體與軟體協同設計，實現高性能與高能效的平衡，並在邊緣與雲端之間靈活分配AI負載，以應對不同場景的需求。

微控制器在AI時代面臨著新的挑戰和機遇。透過整合AI處理器、優化資源分配、壓縮和量化AI模型，MCU能夠在傳統應用與AI功能之間取得平衡，並在邊緣與終端設備中發揮重要作用。隨著技術進步，MCU將在更多領域實現智能化。

結語

AI的快速進化，正在深刻改變我們的生活方式與產業結構。從邊緣推理的廣泛部署，到異構運算架構的強大整合，再到多模態AI模型的全面感知能力，這些突破不僅讓AI變得更加高效，也讓其應用範疇更加多元化。同時，小型語言模型（SLM）的發展，推動了去中心化AI的普及，特別是在智慧城市和工業物聯網等市場，帶來更智慧、更本地化的決策支持。

未來，AI將更加個性化，從被動響應工具演變為主動建議的智慧助手，甚至在多領域成為人類的工作與生活夥伴。隨著AI代理的應用拓展及更強大的隱私保護措施落實，AI正邁向更加互聯與智慧的發展階段，為社會與產業帶來深遠影響。只有在技術創新與倫理規範同步推進的基礎上，AI才能真正釋放其潛力，為全球創造更加可持續的未來。

u-blox精巧的高精準度GNSS解決方案 符合FCC AFC規格需求 協助廠商快速完成Wi-Fi 6E/7路由器開發設計

Wi-Fi 6E/7的問世增加了6GHz的頻段，使得資料傳輸速度大幅提昇。即使在非常密集和擁擠的網路環境中，例如體育場、大型商場或其他公共場所等，Wi-Fi 6E/7的設備也能夠提供更高的網路性能，同時支援更多的Wi-Fi用戶。

然而6GHz這個頻段與現行5G通訊的頻段相互重疊，因此很容易造成彼此干擾的問題。FCC為了解決這個問題，規範了6GHz的路由器必須具備AFC(自動頻率協調)功能，而Wi-Fi聯盟也提出對應的規格需求書，室外用的標準功率路由器需提供可靠的定位資訊，以完成AFC功能的規範。此外，標準功率設備也可能會用於室內環境，因此GNSS接收衛星訊號的靈敏度與天線的強化設計，也成為Wi-Fi 6E/7產品在開發中所面臨的重要課題。

u-blox推出的MIA-M10模組系列定位SiP模組，其尺寸為業界最小，僅4.5mm x 4.5mm，且內建SAW、LNA，可同時支援4個主要全球衛星系統。其保護級別(Protection Level)功能滿足FCC規定的95%置信度位置精準度。特點為不需任何外部元件，可減少線路設計時間和測試工作，是需要達成成本優化並能縮短上市時間的Wi-Fi 6E/7設計人員的理想選擇。如果裝置空間允許，MAX-M10模組系列是經濟實惠的方案。若需求量較大，UBX-M10050-KB晶片方案則為另一個極佳的選項。

若使用於大樓林立的城市，衛星信號易受干擾的地方，可考慮採用L1/L5雙頻段的MAX-F10S模組，以降低多重路徑干擾，提高都會區中的定位精確度。對裝置空間要求較小的場合，UBX-F10050-KB L1/L5雙頻段晶片方案是很好的選項。

u-blox高精準度、尺寸精巧與低功耗的定位解決方案，可滿足FCC對標準功率存取點的定位要求，提供公尺級精準度的解決方案。其晶片及相關模組和開發套件能簡化GNSS功能的設計，可協助廠商縮短開發週期並加速產品上市時間。

MIA-M10系列 u-blox M10標準精準度 GNSS SiP模組

用於微型資產追蹤設備的超低功耗GNSS模組

- 4.5x4.5mm超小晶片尺寸模組，不需外加任何元件
- 在不影響GNSS性能的情況下功耗低於25mW
- 優化的省電模式可將電池壽命延長一倍
- 可同時接收4個GNSS衛星訊號，達到最大位置可用性

MAX-M10系列 u-blox M10標準精準度 GNSS模組

適用於高效能資產追蹤裝置的超低功耗GNSS模組

- 功耗低於25mW，使用小型天線，也具備優異GNSS效能
- 可同時接收4種GNSS訊號（GPS、伽利略、GLONASS，北斗）
- 先進的偵測詐騙和干擾的功能

MAX-F10S u-blox F10 標準精準度GNSS模組

L1/L5雙頻GNSS接收器，可為都會環境實現公尺級準確度

- 有效緩解多重路徑效應，提高都會區準確度
- 具備與蜂巢式數據機並存的優異RF抗干擾能力
- 使用小型天線，也可得到經過驗證的卓越性能

UBX-M10050-KB u-blox M10標準精準度GNSS晶片

適用於高效能資產追蹤應用的超低功耗GNSS接收器

- 在不影響GNSS性能的情況下功耗低於15mW
- 可同時接收4個GNSS衛星訊號，達到最大位置可用性
- 使用小型天線，也可得到經過驗證的卓越性能
- 先進的偵測詐騙和干擾的功能

UBX-F10050-KB u-blox F10標準精準度 GNSS晶片

L1/L5雙頻GNSS接收器，可為都會環境實現公尺級準確度

- 有效緩解多重路徑效應，提高都會區準確度
- 保護等級技術，能以 95%信心度提供即時的定位準確度估算
- 使用小型天線，也可得到經過驗證的卓越性能
- 先進的偵測詐騙和干擾的功能

02-2657-1090 | info_tw@u-blox.com | www.u-blox.com

輕量化AI模型：在邊緣裝置實現高效能的策略

邊緣運算的浪潮正以前所未有的速度席捲而來，將智慧帶到離數據更近的地方。然而，如何在資源受限的邊緣裝置上部署高效能的AI模型，成為當前重要的技術挑戰，如何在不犧牲精度的前提下，巧妙地縮減模型體積、降低計算負擔，將是重要的一環。

文／籃貫銘

隨著人工智慧（AI）技術的飛速發展和物聯網（IoT）設備的普及，將 AI 能力部署到網路邊緣的需求日益增長。邊緣AI (Edge AI) 使得數據處理和決策能夠在靠近數據源頭的地方進行，進而帶來低延遲、低頻寬消耗和高可靠性等優勢。

然而，邊緣裝置通常面臨計算能力、記憶體容量和功耗等資源限制。為了在這些資源受限的設備上實現高效能的AI推理，輕量化AI模型及其相關優化技術應運而生。

什麼是輕量化AI模型？

輕量化AI模型是指經過優化，以實現高運算效能、低資源消耗，且更具成本效益的人工智慧解決方案。這些模型專門設計用於在處理能力有限的設備上運行，例如物聯網設備、嵌入式系統和邊緣裝置。其核心目標是在保持可接受的準確度的同時，顯著降低模型的複雜度和資源需求。

輕量化AI模型的關鍵特徵包括：

1. **降低的運算需求：**減少推理所需的運算操作（如乘法和加法）數量，使其能在處理能力有限的設備上更快運行。
2. **較低的記憶體佔用：**需要較少的儲存空間來存放模型參數和執行期間的中間數據，這對於記憶體容量有限的設備至關重要。
3. **高能源效率：**透過減少運算和記憶體存取，輕量化模型消耗更少的電力，延長電池壽命，適用於功耗敏感的邊緣應用。
4. **優化的性能：**雖然體積和運算量減小，但優化的輕量化模型仍能在特定任務上保持一定程度的準確度。

常見的輕量化AI形式包括部署在邊緣設備上的Edge AI、運行在微控制器上的TinyML，以及透過知識蒸餾創建的精簡模型。這些模型並非簡單的規則引擎，而是利用統計AI或機器學習，通過訓練數據學習模式並進行預測或決策。

邊緣運算的硬體加速方案

僅靠模型優化往往不足以滿足邊緣AI應用對性能和能效的嚴苛要求。硬體加速方案透過使用專門設計的處理單元來執行AI運算，能夠顯著提升推理速度並降低功耗。

ASIC

採用為特定應用或演算法量身定製的積體電路晶片。由於硬體是為特定任務設計的，ASIC通常能在該任務上提供最高的性能和最低的功耗。例如，AI ASIC在每瓦性能上可能比通用處理器高出數百甚至上千倍；此外，專用設計可以實現更