

CHIP-J Semiconductor Study Group 活動報告

2026 年 7 月 26 日 by T. Debun

活動概要

日期：2026 年 7 月 24 日（四） 19:00～21:00

地點：Niterra North America（Mountain View, CA）

費用：免費

本次讀書會邀請了兩位講者進行分享。Katayama 以「CPU 雜談」為題，以及 Eric 以「晶片開發的民主化」為題進行演講。每場演講結束後都進行了熱烈的問答環節，參加者之間的討論也十分活躍。

演講 1：CPU 雜談（Katayama）

Katayama 以其多年 CPU 產品定義的經驗為基礎，從 CPU 的歷史到最新技術進行了廣泛的介紹。

首先從 Intel 位於亞利桑那州 Ocotillo 園區的介紹開始，說明了 Intel 的「Tick-Tock」產品開發週期。Tick 引入新的製程節點，Tock 引入新的架構。這個有規律的開發節奏是為了配合 PC 廠商每年 1 月在 CES 發表新產品，Intel 於 9 月底出貨 CPU。這個週期在 SkyLake 之前運作順暢，但因 10nm 製程轉換受阻，從 KabyLake 到 RocketLake 連續五代產品都停留在 14nm。

Katayama 分享了兩款「幻之 CPU」的故事。CannonLake 是 Intel 首款 10nm 產品，但因 GPU 良率不佳，僅以停用 GPU 的狀態出貨給 Apple。Lakefield 採用了 ARM 的 big.LITTLE 概念的混合架構，使用 Foveros 3D 堆疊封裝技術，但 Lakefield 僅存活了 2020 至 2021 年。然而，其混合架構和 Foveros 技術在 AlderLake 之後得到延續。

從 1971 年的 Intel 4004（約 2000 個電晶體）到最新的 PantherLake（30 億個電晶體、5 晶片組成），CPU 的演進歷程令人驚嘆。此外，CPU 的概念其實源自日本的 Busicom 計算機公司向 Intel 提案並派遣工程師共同開發，可以說「CPU 幾乎是日本發明的」。

關於 GPU，Katayama 說明了其原本是用於顯示渲染的處理器，包括影音處理、4K 渲染（每秒 30 至 120 次計算 829 萬畫素的顏色）、多螢幕支援和數位看板等用途。GPU 之所以適合 AI，是因為其大規模平行處理能力（原本設計用於畫素計算）與深度學習的矩陣運算完美契合。他也區分了 NVIDIA 的 SIMT 方式與 Intel/AMD 的 SIMD 方式。

良率與分級（Binning）部分，以 AlderLake-N 產品定義為例進行了詳細說明。高階產品的良率較低，而停用部分功能的產品良率則大幅提升。「良率取決於產品如何定義」是核心觀點，這是產品企劃特有的視角，與製造端的觀點有所不同。

最後介紹了小晶片（Chiplet）技術，涵蓋 SoC、MCM、MCP、EMIB、CoWoS、Foveros 等封裝技術的差異，說明將晶片分割成更小的區塊如何提升良率並降低成本。

演講 1 問答重點

Q：像 Cerebras 這樣的大型晶圓級晶片與小晶片封裝方式，在 AI 應用上哪個更有優勢？

A：大型晶片很有吸引力，但一旦出現缺陷，商業風險很高。NVIDIA 未來很可能也會走向小晶片化，用多個小晶片實現同樣的功能，再用虛擬矽片連接。

Q：資料中心用 CPU 與筆記型電腦用 CPU 在壽命設計上有何不同？

A：壽命取決於特定 IP 的訊號通過次數和使用溫度條件等因素所形成的故障曲線。降低頻率可延長壽命，提高頻率則縮短壽命。資料中心產品雖然 24 小時運作但採用熱插拔方式，因此產品定義上的壽命與 PC 相近（約 5 年），但整體耐久性要求則高得多。

演講 2：晶片開發的民主化（Eric）

Eric 曾在多倫多大學研究 FPGA 虛擬化技術，致力於透過雲端 FPGA 和 Kubernetes 實現硬體開發的民主化。

演講從軟體與硬體開發環境的差距切入。軟體開發人人可做、成本低、數天到數月即可完成。硬體開發則需要數億到近十億美元的投資，耗時數年，小學生設計晶片更是不可能的事。雖然這種差距源於軟體處理「資訊」而硬體需要「物理製造」，但 Eric 提出質疑：真的無法改變嗎？

在後摩爾定律時代，應用程式＝軟體＋硬體的觀念日益普及。超大規模業者紛紛自行開發 AI 晶片、影片轉碼晶片和網路安全晶片。十年前只要 CPU 的資料中心業者，如今積極引進專用硬體。

軟體開發透過敏捷開發和 CI/CD 方法論實現快速的用戶回饋循環。但硬體開發完全缺乏這一環節。FPGA 設計約需一年、ASIC 製造需兩到三年、整合又需數月，開發者要等三到四年才能獲得第一次用戶回饋。

Chiptip Technology の做法是利用 Kubernetes 管理的雲端 FPGA 建立小型迭代循環。開發者在雲端 FPGA 上實現設計、公開部署讓真實用戶使用並獲取回饋，快速迭代。在獲得充分信心後才進入 ASIC 製造，避免重複進行昂貴的 ASIC 開發週期。

技術上的關鍵差異在於，Chiptip 將 FPGA 直接連接到網路，而 AWS 的架構中資料必須透過 CPU 經 PCIe 傳輸。這使延遲降低約 30 倍，適合大規模硬體應用的開發。

FPGA 雲端服務預計於 2026 年底推出，合作夥伴包括 Emotion X（日本新創）和 QSD（美國新創）。應用合作夥伴涵蓋網路安全和 5G/6G 通訊企業。2028 年還計劃推出 FPGA 轉 ASIC 的遷移支援服務。

演講 2 問答重點

Q：與傳統 FPGA 模擬加速服務（Cadence 等）有何不同？

A：傳統服務僅用於硬體模擬，終端用戶無法使用。Chiptip 的做法是將應用部署到雲端讓實際用戶使用，實現關鍵的回饋循環，這是根本性的差異。

Q：網路以外的 I/O（類比介面等）如何支援？

A：目前平台支援 PCIe 和網路介面。因此目標應用以網路導向為主，如網路安全和 5G/6G 通訊。

Q：如果業務成功，如何防止 Google 或 AWS 進入同一市場？

A：約 10 年前在多倫多大學研究期間取得了雲端 FPGA 虛擬化的基本專利，可作為進入障壁。不過也承認最終被大型企業收購的可能性。