

行政院國家科學委員會專題研究計畫 期中進度報告

子計畫四：用於低功率處理器之暫存器組織及運算資料產生

(2/3)

計畫類別：整合型計畫

計畫編號：NSC93-2220-E-009-017-

執行期間：93年08月01日至94年07月31日

執行單位：國立交通大學電子工程學系暨電子研究所

計畫主持人：劉志尉

計畫參與人員：林泰吉、歐士豪、劉晉宏、黃維聖、郭羽庭

報告類型：完整報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 94 年 5 月 31 日

低功率系統之設計及自動化

子計畫四：用於低功率處理器之暫存器組織及運算資料產生(2/3)

計畫編號：NSC93-2220-E-009-017

執行期限：93 年 8 月 1 日至 94 年 7 月 31 日

主持人：劉志尉 國立交通大學電子工程學系

參與人員：林泰吉、歐士豪、劉晉宏、黃維聖、郭羽庭

一、中文摘要

本計畫採用暫存器組叢集 (clustering) 及 banking 等技術探討高效能、低功率處理器中運算資料的產生及處理器中暫存器組的架構設計。本計畫前半年的成果包含了：一、分散式暫存器組，及基於該暫存器組所設計之低功率數位訊號處理器(programmable digital signal processor; DSP processor)[1][2]。此數位訊號處理器之特色即是模組化(modular)的微架構(micro-architecture)，可輕易抽換，並隨意搭配具不同延遲時間之功能模組。此數位訊號處理器已經由 CIC 下線到聯電製作，最高操作頻率為 314MHz，功率消耗僅為 52mW；二、低功率管線化設計：on-demand pipelining，其可在系統處理能力改變時，動態降低功率消耗。將此技巧應用至 16 位元乘法器上，可節省達 70% 的能量消耗。

關鍵詞：低功率、數位訊號處理器、暫存器組織

Abstract

This project investigates the clustering and banking techniques to explore energy-efficient register organization and some effective data generation schemes for high-performance and low-power processors. Our achievements include: (1) distributed register file and a DSP processor design based it. The important feature of the DSP is its modular micro-architecture, which simplifies the plug-ins of functional units of various latencies. The proposed DSP core has been fabricated in the UMC 0.18um CMOS technology, and it can operate at 314MHz with 52mW power consumption; (2) low-power pipelining technique, the on-demand pipelining, which dynamically reduce the energy dissipation while the system does not achieve its peak rate.

In our experiment, a 16-bit multiplier with the proposed on-demand pipelining can save 70% energy dissipation.

二、計畫緣由與目的

在電腦、通訊、消費性電子等 3C 領域，可攜式行動裝置的發展越來越受重視。這類行動裝置的電力來源通常都仰賴電池來支應。然而受限於電池有限的電力容量，可攜式行動裝置的工作時間或是待機時間也因此受到限制，為了要延長裝置的工作或是待機時間，除了增加電池的容量外，減少功率/能量消耗則是唯一途徑。以現今電池功率的容量大約是 1.0~1.5 amp-hr 而言，如果要滿足大約六天或一星期的 working hours (約 150hrs)，則該可攜式行動裝置只允許 consume 10mA 左右。換言之，假設所使用的 supply voltage 是 1.2V、工作頻率 300MHz，該可攜式行動裝置的設計規格大約是 $12\text{mW}/300\text{MHz} = 0.04\text{mW}/\text{MHz}$ 。目前(2003 年) state-of-the-art 的產品規格大約是 $\sim 1\text{mW}/\text{MHz}$ ，因此仍需要約 25 倍的 overall performance improvement。

另一方面，通訊標準的制訂以及通訊系統所提供的服務品質，隨著使用者的要求，雖稱不上日新月異，但變化速度已越來越快，因此未來的通訊相關產品，除了必須提供高效能 (performance) 外，具備 configurable 以及 scalability (或是 extensibility) 的功能可以大大延長通訊產品的 time-in-market。所謂 configurable 功能，近年來在 software-defined radio 或 B3G 應用常被提及，係指透過 parameters 的控制(如採用 register setting 或變更 memory context 方式來達成)，可改變組件功能性(functionality)或是組件間的結構性(structure)，以達到提供系統產品的 flexibility 能力。scalability 或是 extensibility，則是透過 architecture 規劃與設計成一模組形式，使得未來如有 performance 增加的需求時，透過增加

模組的方式即可容易達成所需。在設計的同時，設計者需考量容易加入或移走運算或動作之同時性(concurrencies)。可能方式有下列數種(亦可同時採用之)：(1) data parallelism，即 SIMD 式的 multi-word，sub-words 處理；(2) 採多個 functional units，如用 VLIW (static scheduling, issue) 或 superscalar (dynamic scheduling, issue) 方式；(3) 使用 attached coprocessor 方式。

綜觀而言，low power design 必是當前國內外學術界、產業界努力的研究課題，也是國科會工程處近年來推動的重要研究主題之一。值得注意的是，高效能與低功率損耗在 SoC 的設計發展來看，往往是設計者必須相互 trade-off 的兩大因素。未來的 SoC 產品，已漸漸朝向 low power and enough performance with scalability 之設計走向。

現今多媒體及通訊系統動輒需要數至數百個 GOPS (giga operations per second) 的運算能力 (例如，滿足 QCIF、15 frames/s 之 motion estimation 運算，通常約需 5 GOPS)，而目前的半導體製程技術，可輕易地整合數十至數百個運算單元、或十數個高平行度的功能模組在單一的處理器中，當其工作在數百 MHz 至 GHz 的時脈下可達到此運算需求。但功能模組之間的資料交換及資料暫存卻是極大的挑戰。如何利用有限的頻寬來安排資料在平行的功能模組間順暢流動是主要的設計問題。傳統 RISC 處理器將記憶體搬移抽離計算動作已大幅度降低此問題的複雜度，但其架構中負責功能模組間資料互換及資料暫存之集中式 (centralized) 暫存器組 (register file) 的延展性極差，已經成為高效能處理器主要的設計瓶頸。本子計畫主要為探討高效能、低功率處理器中運算資料的產生及處理器中暫存器組的架構設計，而如何設計一個高效能且低功率的暫存器組即是現今處理器設計成功與否的最大關鍵。

三、研究方法及成果

近年來市面上一些電子產品對於多媒體應用或是無線網路功能的支援日益豐富[3]，對於資料運算的需求也越來越高，並且在產品上市時間 (time-to-market) 及產品生命期 (time-in-market) 的壓力下，平台式的設計

(platform-based design) 以及可程式化 (programmable) 或是可重新設定 (reconfigurable) 的設計越來越被大家所認同及接受。可程式化的設計可以容許在設計非常末期，甚至產品已經到顧客手中還可以進行修改，解決設計上的瑕疵。而 platform-based design 則可以依照先前的設計，經由恰當的修改，延伸出各式各樣的新產品，降低 NRE cost。多處理器核心架構是目前一個常見的多媒體系統平台。而在這類的系統中的運算需求，可大致分為兩類：以控制為主(control-oriented)的運算及以資料計算(data-intensive)的運算。以常見的無線通訊手機為例，控制為主的運算包括有人機介面、作業系統及通信協定的處理等；資料處理為主的運算則包括有基頻處理(baseband processing)、數值轉換(transformation)，甚至聲音、影像的處理等等。TI OMAP[4]是一個最有名的範例。它主要是包含一個 RISC 處理器，和一個 DSP。RISC 用來實現 user interface, protocol stack 等工作[5]，並負責整個系統的 coordination。DSP 則是用來處理大量的訊號資料處理經由適當的工作分配及動作協調，它可以極有效率地展現單一處理器無法達到的 performance。

但目前市面上的多核心處理器平台都有一個共同的問題：他們都是由現成的處理器直接拼湊起來，這些處理器不是為多核心架構所設計，他們當初沒有考慮到其他處理器的存在，而是以單獨使用為原則來設計，因此這些處理器彼此之間會存在著重複的功能，也就是存在累贅的元件及設計，以致過於浪費。

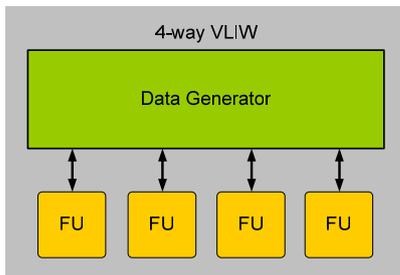
因此，我們針對多核心(multi-core)之媒體系統晶片架構，分析應用程式之特性，開始設計一高效率的訊號處理資料路徑(datapath)及一模組化且不受功能單元時間延遲的處理器架構，可輕易抽換，並隨意搭配具有不同延遲時間之功能模組，只需以軟體工具做簡單的設定，完全不會影響其他的硬體區塊。

此晶片已經經由 CIC 下線到聯電製作，在 0.18um CMOS 的製程下，最高操作頻率為 314MHz，功率消耗僅為 52mW。

另外我們也發展了一個管線化設計技巧稱為 on-demand pipelining，可增強系統對總處理能力改變的感知能力。將此技巧應用至一個 16 位元的乘法器發現可節省達 70% 的能量消耗。

(1) 分散式暫存器組織及相對應之 DSP 微架構

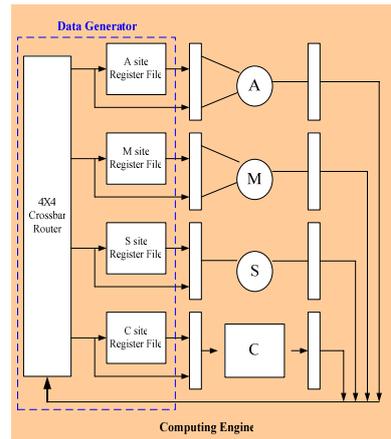
圖一是我所提出的一個 4-way VLIW 數位訊號處理器的示意圖[6][7]。指令長度為 64 位元，一共有 4 個 issue slot，分別為 addition slot，multiplication slot，shift slot 及 control slot。圖一中資料產生器負責提供每個 cycle 所有功能單元所需的運算資料，並且收集每個 cycle 由所有功能單元傳回來的運算結果做進一步的處理或輸出。而這些動作都是由軟體在編譯(compile)時，進行靜態的分析及排程。由圖一可看出我們所設計之處理器架構非常的模組化，也就是資料路徑(data-path)非常乾淨，不會像一般傳統處理器內部因為一些 forwarding path 而產生複雜的資料路徑。可輕易抽換，並隨意搭配具有不同延遲時間之功能模組，只需以軟體工具做簡單的設定，完全不會影響其他的硬體區塊。這在目前由於製程的快速進步導致整個 wire/interconnection overhead 愈來愈大的情形下，是一個非常可行的解決方案。



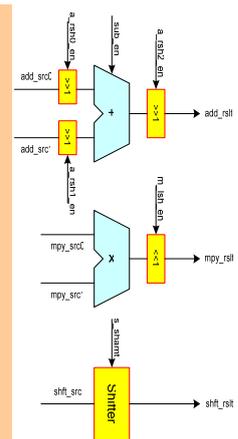
圖一. 4-way VLIW 示意圖

事實上，整個處理器包含了一個運算引擎 (computing engine)，一個控制介面單元 (host interface unit)，兩個皆為 8KByte 的指令及資料記憶體還有一個 AMBA interface。圖二、三分別為運算引擎架構及其內部的算術單元。運算引擎內除了 4 個功能單元分別為：加法器 (A site)，乘法器 (M site)，移位器 (S site) 及控制單元 (C site) 外，另外由每個功能單元前面的暫存器和一個 4-by-4 crossbar router 組成資料產生器。A site 提供有號數加減法的運算；M site 提供常數乘法 (constant multiplication) 運算；S site 提供向右 7 位向左 8 位的算術位移 (arithmetic shift)；而 C site 除了提供邏輯運算 (logic operation) 和記憶體的存取外，還支援了簡易的指令流程控制像是分支指令 (branch instruction) 或重複執行指令 (repeat

instruction)。我們額外在校加法器輸入端各配置了一位元之小數點對準裝置 (aligner) 及在輸出端配置了一位元之正規化裝置 (normalizer)，乘法器則是在輸出端配置了一位元之正規化裝置，相當於浮點單元中執行對準及正規的移位器的縮減形式，可增加我們定點運算的精確度。其中所有的移位控制訊號，以及各個運算單元之控制、暫存器檔案 (register file) 位址、crossbar router 控制訊號等，皆由提供的軟體產生之指令所控制，並儲存至指令記憶體。



圖二. 運算引擎架構

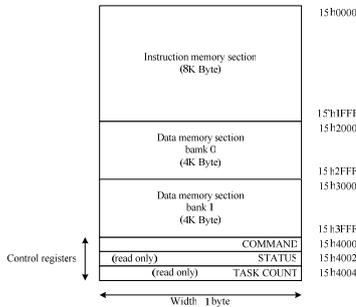


圖三. 算術單元

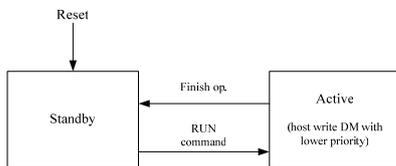
因為所提出的處理器整合進 AMBA 介面，為了方便與其他 host 處理器溝通方便，我們將此數位訊號處理器設計成具有一般標準 SRAM 的介面。圖四為處理器在 AMBA 上的記憶體對應 (memory map)。host 處理器可將與此數位訊號處理器的溝通視為存取一塊記憶體，經過系統匯流排 (system bus) 可以透過控制介面單元來編輯 (configure) 資料記憶體和指令記憶體，除此之外，也可以編輯控制暫存器 (control registers) 的方式來命令數位訊號處理器的動作。內建的系統控制器 (system controller) 將會以寫入控制暫存器的資料來切換數位訊號處理器運作的模式。

整個處理器的運作模式如圖五所示，擁有兩種工作模式，分別為待命 (standby) 模式和活動 (active) 模式。在系統經過 reset 後，數位訊號處理器將會處在待命模式，在這個模式中，host 處理器可以存取數位訊號處理器的資料記憶體和指令記憶體的方式來編輯數位訊號處理器的功能。當編輯完成後，host 處理器將以寫入控制暫存器 (control registers) 的方式啟動數位訊號處理器，而系統控制器 (system

controller)將以此指揮數位訊號處理器進入活動模式(active mode)並開始做運算動作。當數位訊號處理器完成指令，會發出結束執行的訊息並且跳回待命模式，此時 host 處理器可以改編(re-configure)數位訊號處理器，或再一次的命令數位訊號處理器執行指令。

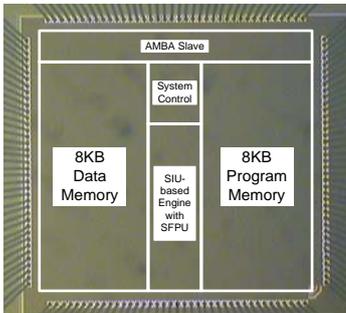


圖四. 記憶體對應情形



圖五. 運作模式

我們將提出的 DSP 處理器核心，以 CIC 提供的 UMC 0.18um CMOS 製程實現。core size 為 $1.5 \times 1.5 \text{mm}^2$ ，最高操作頻率為 314MHz，而功率消耗為 52mW。圖六為晶片實照。

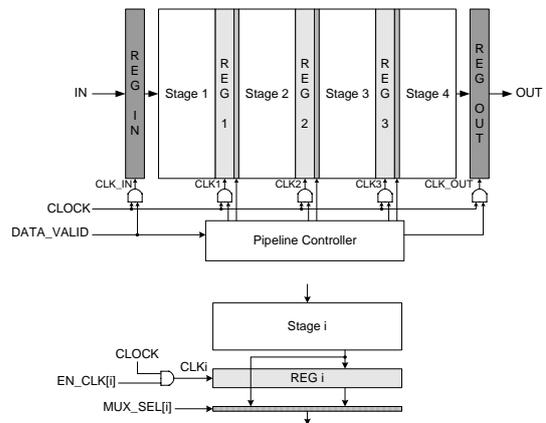


圖六. 晶片實照

(2) 低功率管線化設計

硬體架構的設計必須以最惡劣的運作環境作為設計考量，故傳統靜態式低功率電路設計的技巧在即時運算的環境裡效率並不好。功率(能量)消耗感知的系統能夠在不同的情況下調整功率的消耗，因而日益受到重視。在本計畫中，我們另外提出了一個新穎的管線化設計技巧，可增強對總處理能力改變的感知能力。在這個管線化設計技巧中，管線暫存器只有在需要時才會被使用。當一筆資料與下一筆資料

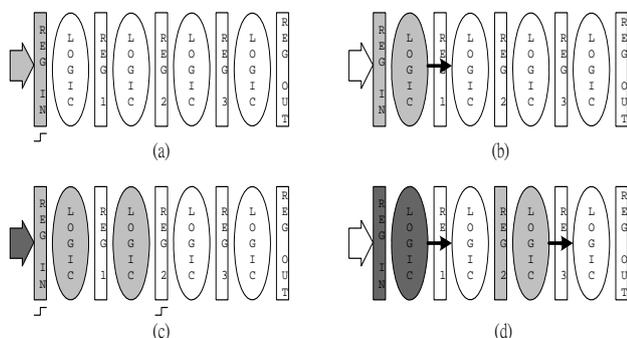
間沒有競跑現象且管線間 glitch 不嚴重時，此筆資料會藉由直接跨越暫存器的方式到達下一級管線；被跨越而未使用的暫存器會以時脈閘門的技巧來減低功率消耗。相較於傳統時脈閘門技巧，當總處理量未達滿載時，我們所提 on-demand 管線化資料路徑消除了所有管線暫存器多餘的動作。然而，我們的方法在所有的工作模式裡都有相同的輸出輸入延遲，這樣的特性有效的降低了系統整合的難度。我們還實現了一個使用 on-demand pipeline 資料路徑的 16 位元乘法器。在對於 JPEG 中，de-quantize 的模擬裡，相較於一般的管線化設計，我們的技巧可節省達 70% 的能量消耗。



圖七. Architecture of 4-stage on-demand registered pipelined datapath

圖七為我們所提出的 on-demand pipelined datapath 的架構圖，每級 pipeline stage 都使用了彼此獨立的 clock 並且每一級都設計了一個多工器(multiplexer)來產生讓資料直接通過到下一級的路徑。圖中的 pipeline controller 除了 clock 和 reset 訊號外，唯一的輸入訊號為 DATA_VALID。這個 1-bit 的訊號是用來指示目前輸入的資料是否需要處理。整個 pipeline controller 的工作就是控制每級 pipeline register 的 clock enable 訊號 EN_CLK[i] 及多工器的選擇訊號 MUL_SEL[i]。圖八我就以一個簡單的例子說明我們所提出的 on-demand pipeline 如何有效的降低能量消耗。假設在 time slot0 一有效的資料，也就是需要運算的資料出現在輸入端，則暫存器 REG_IN 必須開啟並將此資料儲存下來，如圖八(a)所示。為了方便說明，我們將此資料取名為 D1。當繼續來到圖八(b)time slot1 的時候，假設此時輸入端並沒有有效的資料進來，則 D1 繼續通過 REG 1 而並

不會開啟 *REG 1*。當再前進到圖八(c)的 time slot3 時，假設另一有效的資料 D2 出現在輸入端，則此時我們必須開啟 *REG 2* 用來儲存 D1，同時 *REG IN* 也會開啟儲存 D2。這樣一來就能確保整個電路的正確性。在圖八(d)time slot3 的時候，由於沒有有效的資料進來，所以 D1 和 D2 則分別直接通過 *REG 1* 和 *REG 3*。



圖八. (a)time slot0. (b)time slot1. (c)time slot2. (d)time slot3.

四、結論與討論

本計劃已順利完成各項預期工作項目。研究成果正陸續整理投稿於國際會議和期刊中，其中已發表之著作包含參考文獻[8][9][10][11]。

五、參考文獻

1. E. A. Lee, "Programmable DSP Architectures, Part I," *IEEE Acoustics, Speech and Signal Processing Magazine*, October 1988
2. E. A. Lee, "Programmable DSP Architectures, Part II," *IEEE Acoustics, Speech and Signal Processing Magazine*, January 1989
3. Steve Furber, *ARM System-on-Chip Architecture*, Addison Wesley, 2000
4. *OMAP5910 Dual Core Processor – Technical Reference Manual*, Texas Instruments, Jan 2003
5. J. L. Hennessy and D. A. Patterson, *Computer Architecture – A Quantitative Approach*, 3rd Edition, Morgan Kaufmann, 2002
6. K. K. Parhi, *VLSI Digital Signal Processing Systems – Design and Implementation*, John Wiley & Sons, 1999
7. P. Lapsley, J. Bier, and E. A. Lee, *DSP Processor Fundamentals – Architectures and Features*, IEEE Press, 1996
8. Tay-Jyi Lin, Hung-Yueh Lin, Chie-Min Chao, Chih-Wei Liu, and Chein-Wei Jen, "A compact DSP core with static floating-point arithmetic," *JVSP*, 2004
9. Tay-Jyi Lin, Chen-Chia Lee, Chih-Wei Liu and Chin-Wei Jen, "A Novel Register Organization for VLIW Digital Signal Processors," *IEEE VLSI-DAT*, 2005
10. Wei-Sheng Huang, Tay-Jyi Lin, Shih-Hao Ou, Chih-Wei Liu, and Chein-Wei Jen, "Pipelining

technique for energy-aware datapaths," *ISCAS*, 2005

11. Yu-Ting Kuo, Tay-Jyi Lin, Chih-Wei Liu, and Chein-Wei Jen, "Architecture for area-efficient 2-D transform in H.264/AVC," *ICME*, 2005