

高效能影片解碼平行可規劃 計算機架構

報告人：謝政宏

大綱

- 簡介
- 影像解碼平行化分析
- 系統設計
- 實驗結果
- 結論

大綱

- 簡介
- 影像解碼平行化分析
- 系統設計
- 實驗結果
- 結論

影像編解碼

- 影像編解碼器主要用來壓縮影片容量，以解決空間不足與頻寬限制的問題。
- 方便編解碼器的發展與通用性，制定了編解碼器的標準。
- 常用的影像編解碼標準
 - 國際電信聯盟：H.261、H.262、H.263、H.264以及H.265。
 - 國際標準組織：MPEG-1、MPEG-2和MPEG-4。
 - 中國編碼標準：AVS。

目標與設計考量

● 目標

- 滿足多規格的影像解碼功能。
- 能廣泛應用在電腦與行動裝置上。
- 提高效能與降低能源消耗。

● 設計考量

- 影像的編解碼需要龐大的計算量，連續撥放代表要求計算速度與即時性。
- 分析影片解碼的演算法以實行平行處理。
- 在行動裝置上，硬體資源和能源消耗有更嚴格的規範。
- 設計創新計算機架構，以符合高性能的需求。

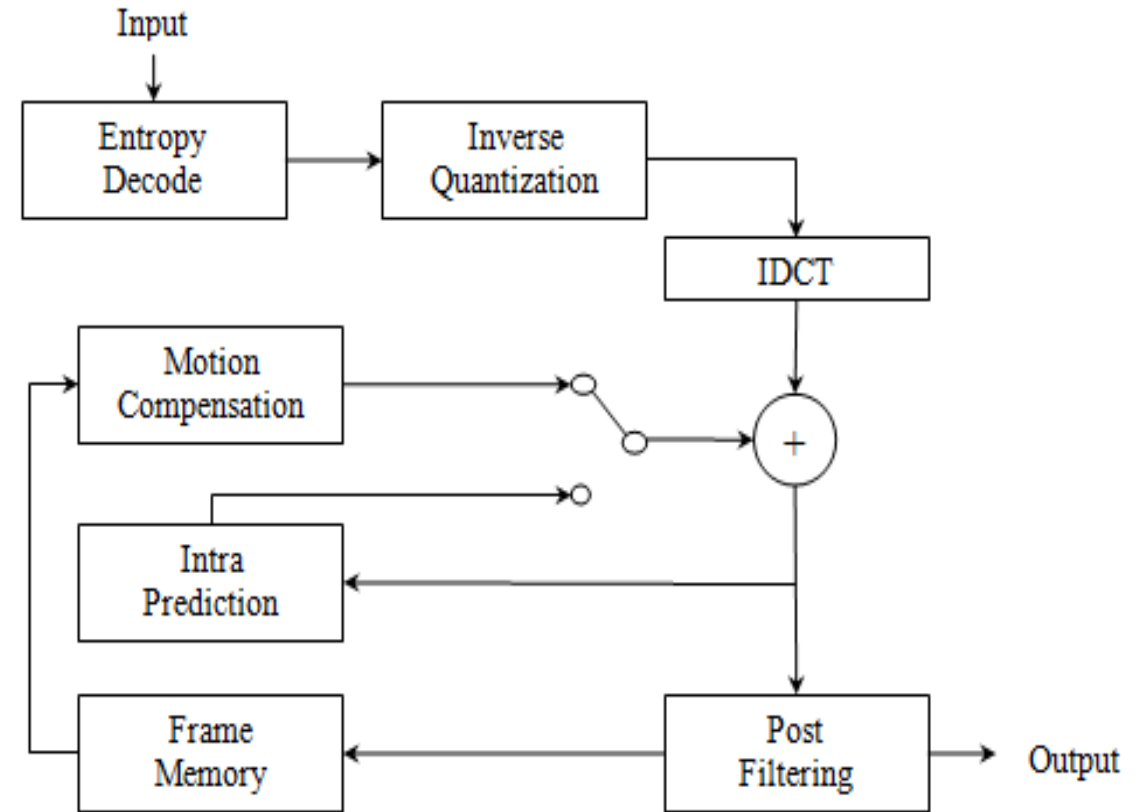
大綱

- 簡介
- 影像解碼平行化分析
- 系統設計
- 實驗結果
- 結論

通用的影片解碼架構

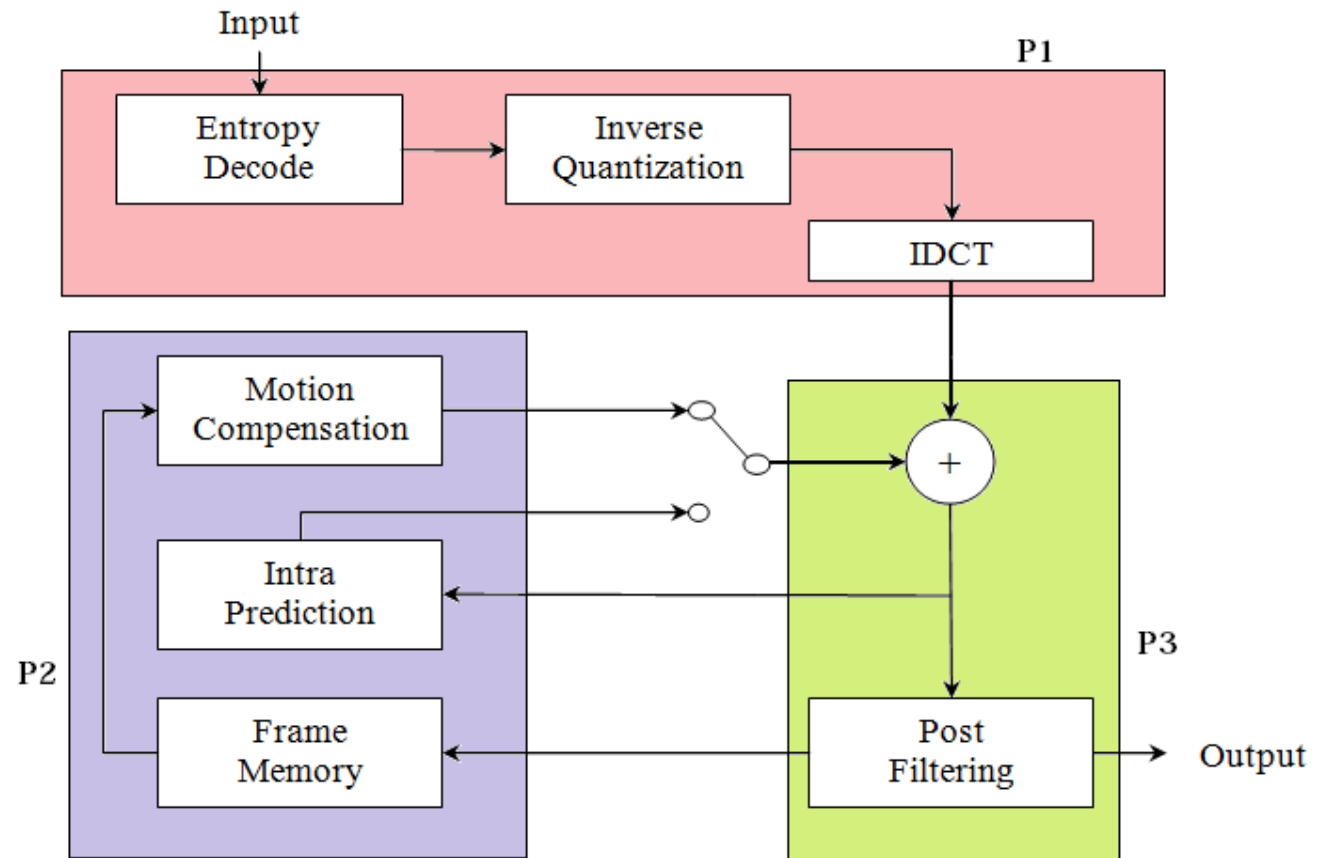
● 解碼任務的分類

- 位元等級的算術運算（熵編碼法和反量化）。
- 基於區塊Macroblock(MB)的字級運算占75%以上的計算量(反離散餘弦變換，運動補償，畫面內預測，重建，和去區塊濾波器)。
- 2-D圖像數據讀/寫(影格內存訪問操作)。
- 為了加速解碼的速度，下面介紹幾種影像解碼的平行處理方式。



任務分區的平行處理

- 此平行化類似管線化(Pipeline)的概念，主要是讓讓各個核心處理器執行影片解碼中不同的函式計算。
 - P1負責計算熵解碼(Entropy Decoding)、反量化(Inverse-Quantization)及反離散餘弦變換(IDCT)。
 - P2負責處理運動補償(Motion Compensation)及內部預測(Intra Prediction)。
 - P3負責濾波處理(post filtering)。
 - 因為每一分區的解碼計算量並不相等，造成每個區域將有若干的等待時間。



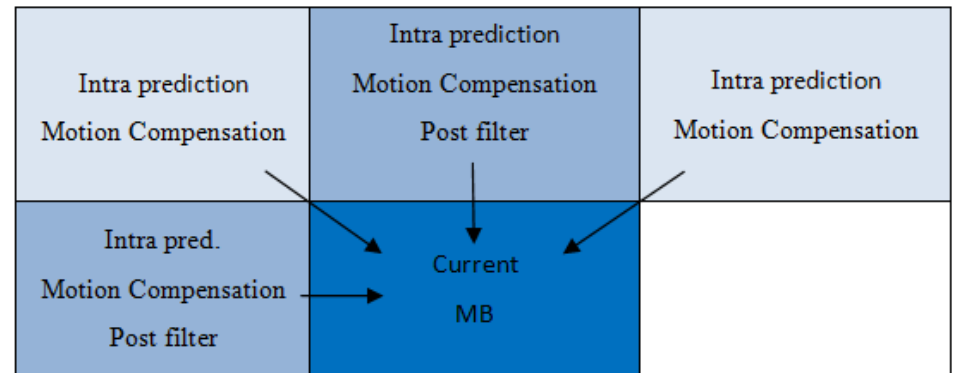
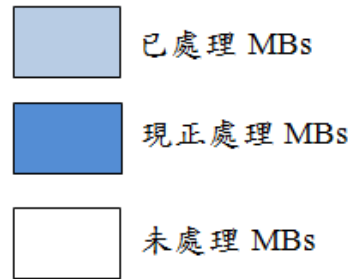
片級別(Slice-level)的平行處理

- 在影片的編碼標準中，每個frame可以分成一個或是多個Slice，而Slice由多個區塊(Macroblock)組成。slice間並不相依，我們可以依Slice的數目進行平行處理。
- 片級別(Slice-level)的平行的缺點
 - 片Slice的數目由編碼時決定，一般不會太多。
沒有延展性平行度不足。
 - 在處理區塊濾波的時候，將會超過Slice的邊界，這會也會降低平行處理的速度。

Slice 1
Slice 2
Slice 3
Slice 4

2D-Wave 平行處理

MB(0,0) T1	MB(1,0) T2	MB(2,0) T3	MB(3,0) T4	MB(4,0) T5
MB(0,1) T3	MB(1,1) T4	MB(2,1) T5	MB(3,1) T6	MB(4,1) T7
MB(0,2) T5	MB(1,2) T6	MB(2,2) T7	MB(3,2) T8	MB(4,2) T9
MB(0,3) T7	MB(1,3) T8	MB(2,3) T9	MB(3,3) T10	MB(4,3) T11
MB(0,4) T9	MB(1,4) T10	MB(2,4) T11	MB(3,4) T12	MB(4,4) T13



- 解碼定義的數據結構中，最小單位為區塊(Macroblock)，簡稱為MB，這方法就是將影像切割成區塊再做平行處理。
- 為了將這些區塊(MB)做平行處理我們首先要考慮區塊(MB)之間的相依關係。
- 當畫面解析度越大最大可平行處理的區塊(MBs)越多。

大綱

- 簡介
- 影像解碼平行化分析
- 系統設計
- 實驗結果
- 結論

系統設計

- 提出一個高效能的平行可規劃計算機架構。
- 提出一個降低晶片面積的線路交換網路連接路由方案。
- 加入低功率模式的能源管理系統。

系統設計

- 提出一個高效能的平行可規劃計算機架構。
- 提出一個降低晶片面積的線路交換網路連接路由方案。
- 加入低功率模式的能源管理系統。

GPP vs. ASIC

- 通用處理器(GPP)

- 容易使用而且靈活性高
- 效能不符合要求

- 特殊應用積體電路(ASIC)

- 在特定的應用上提供最高的性能。
- 電路的結構不能更動，缺乏靈活性。

- 解決方案

- 應用特定的DSP或FPGA與通用處理器（GPP）配合加快計算速度和提高系統的能源使用效率。—已可規劃計算機實現。

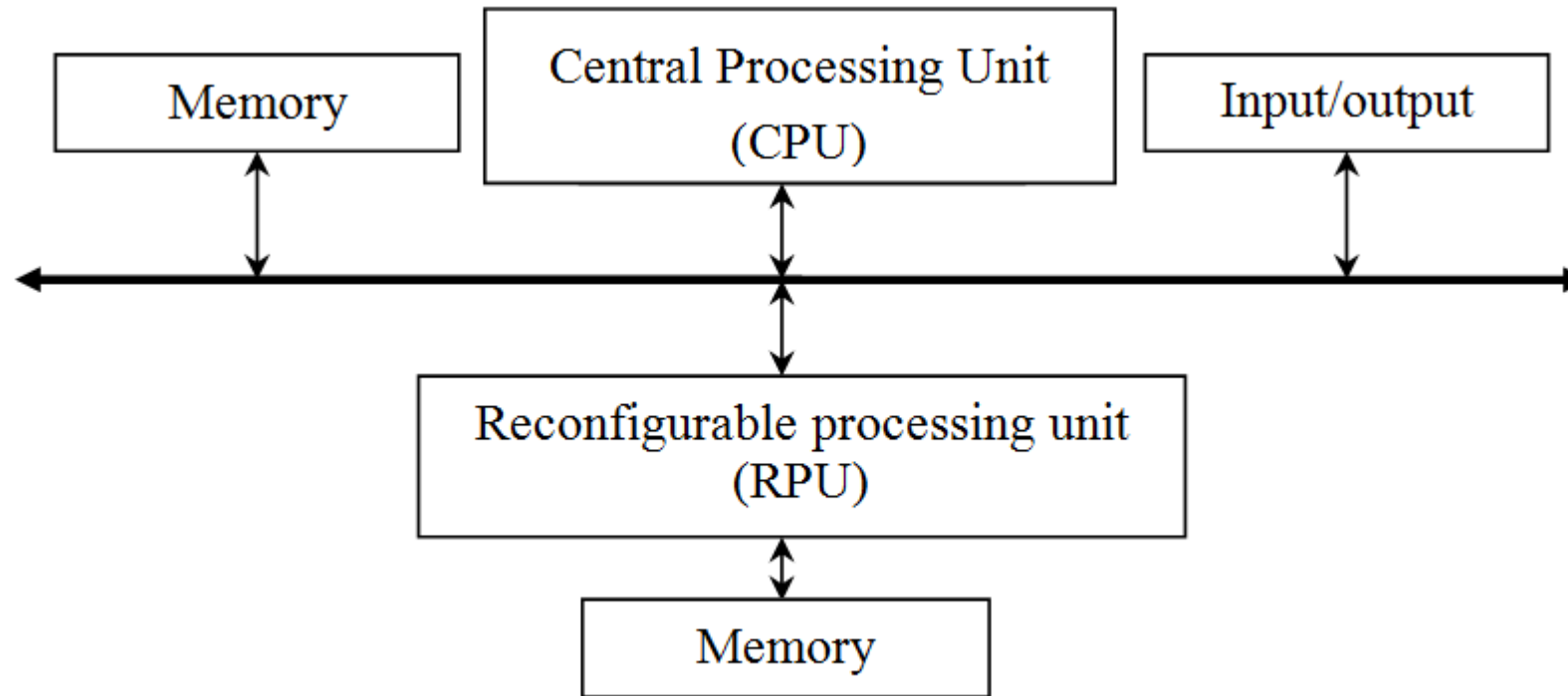
可規劃計算

● 可規劃計算(Reconfigurable Computing,RC)

- 以現場可程式邏輯閘陣列(Field Programmable Gate Array,FPGA)為主。
- 依照需求變更邏輯閘功能。
- 擁有接近特殊應用積體電路(ASIC)的性能與通用處理器(GPP)靈活性和容易使用的優點。

Computing Architecture	Flexibility	Performance
General Purpose Processor (GPP)	Yes	Slow
Application-specific integrated circuit (ASIC)	No	Fastest
Reconfigurable Computing(RC)	Yes	Fast

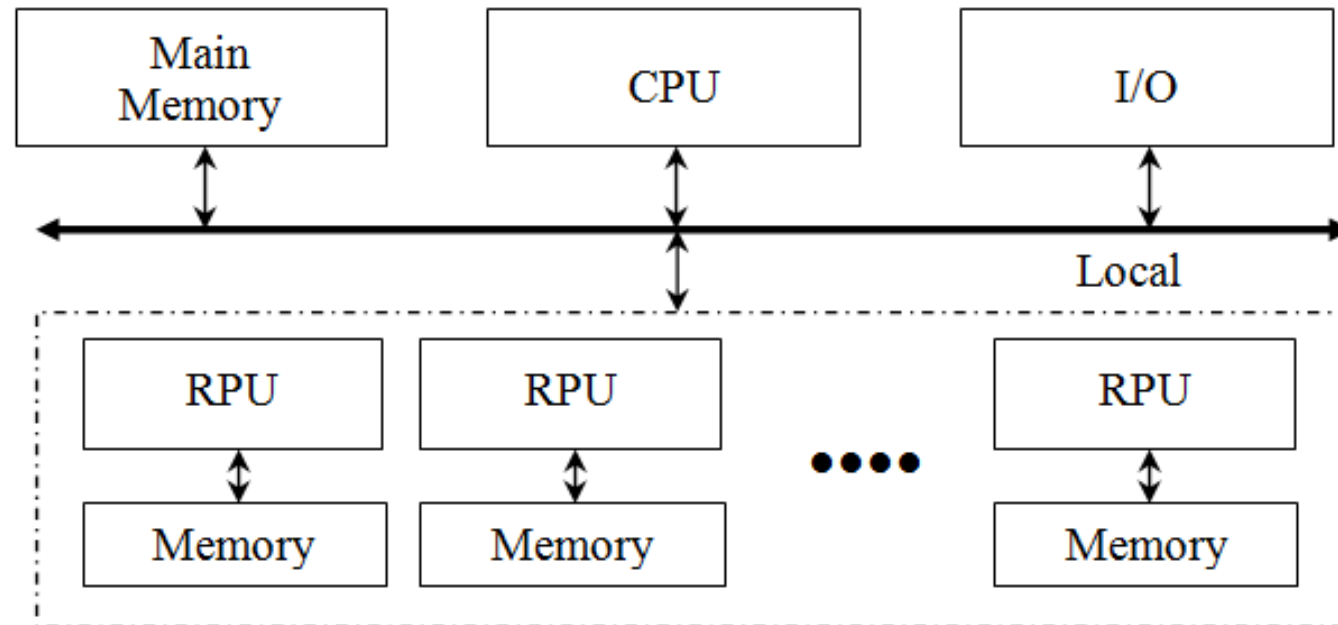
傳統可規劃計算機架構



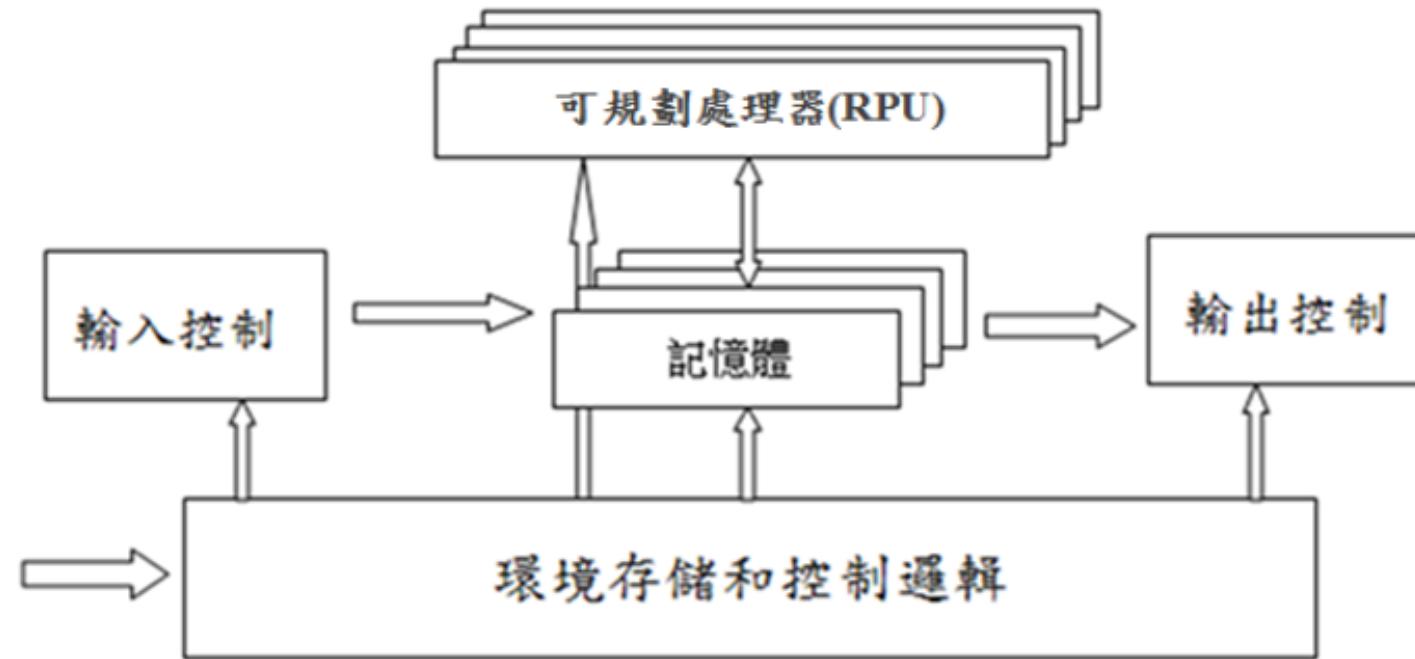
- 傳統的可規劃計算機無法進行多工處理，為了進一步的提升效能，我們希望將可規劃計算機結合平行處理技術。

平行可規劃計算架構

- 多指令多資料流(multiple instruction, multiple data, MIMD)的平行處理技術。
- 配置各自記憶體以改善效能。



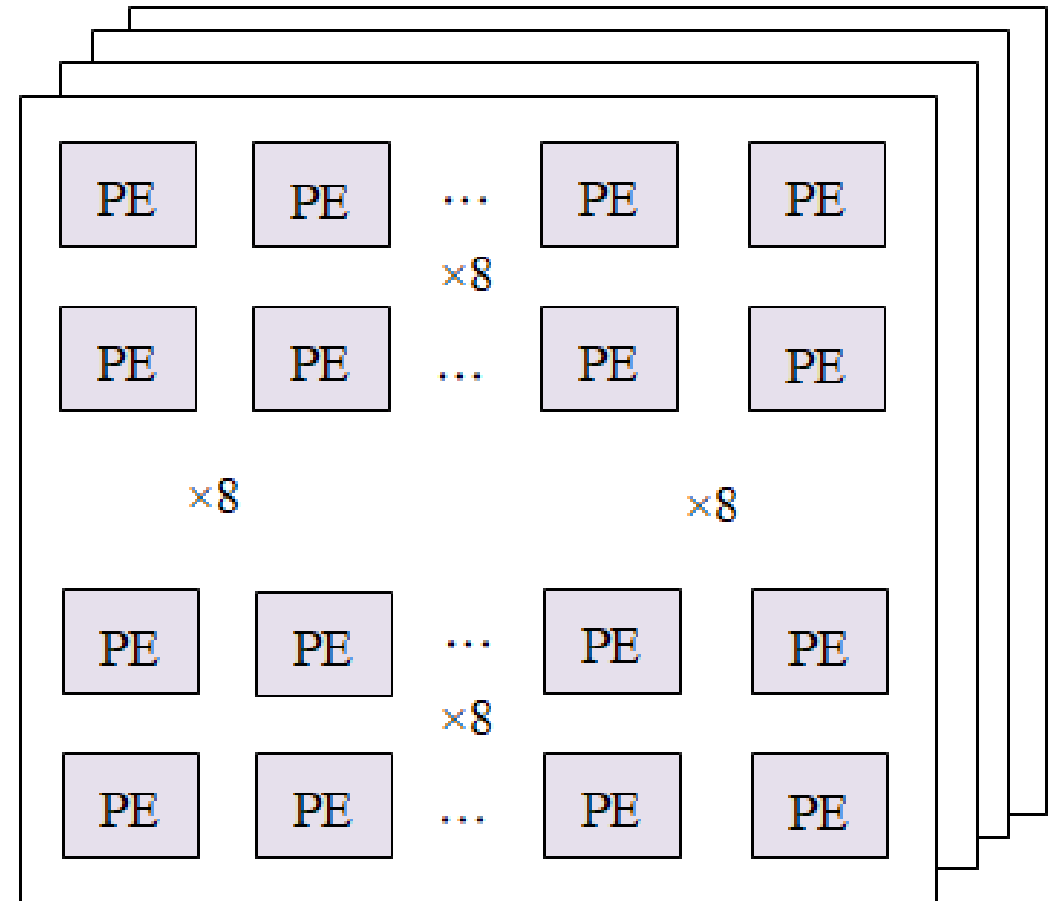
影片解碼平行可規劃計算機架構



- 可規劃處理器，輸入控制邏輯，輸出控制邏輯，內部緩衝內存組成的數據處理路徑。
- 環境存儲和控制邏輯形成的配置路徑。

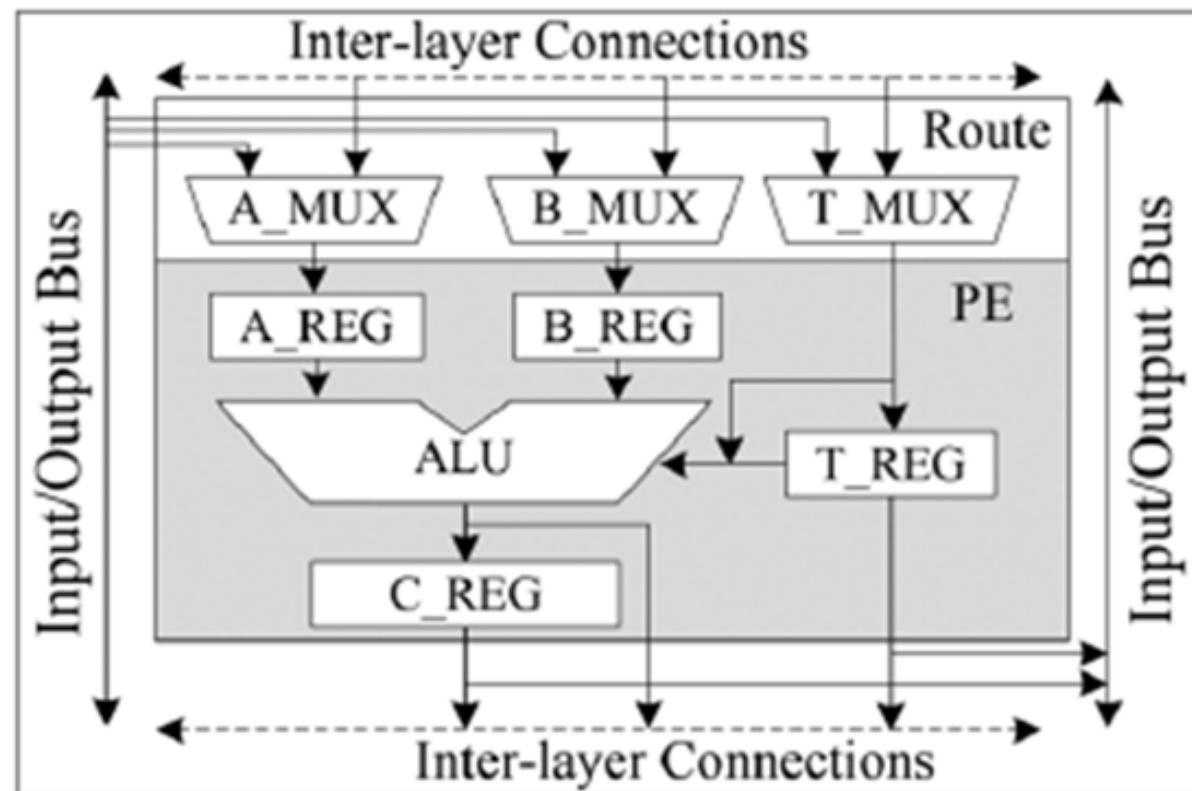
RPU結構

- 一個RPU結構由四組可重配置單元陣列 (PEA)組成。
- 每組PEA包含8x8可重構PEs。
- 每個PEA可以獨立工作，從而提供更高級的平行處理，以改善RPU吞吐量。
- 需要維持低的功耗時某些PEA也可以關閉。



Processing Element(PE)架構圖

- 每個PE包含
 - 一個算術邏輯單元 (ALU)
 - 一組輸入/輸出 (A_REG、B_REG、C_REG)
 - 臨時結果暫存器 (T_REG)

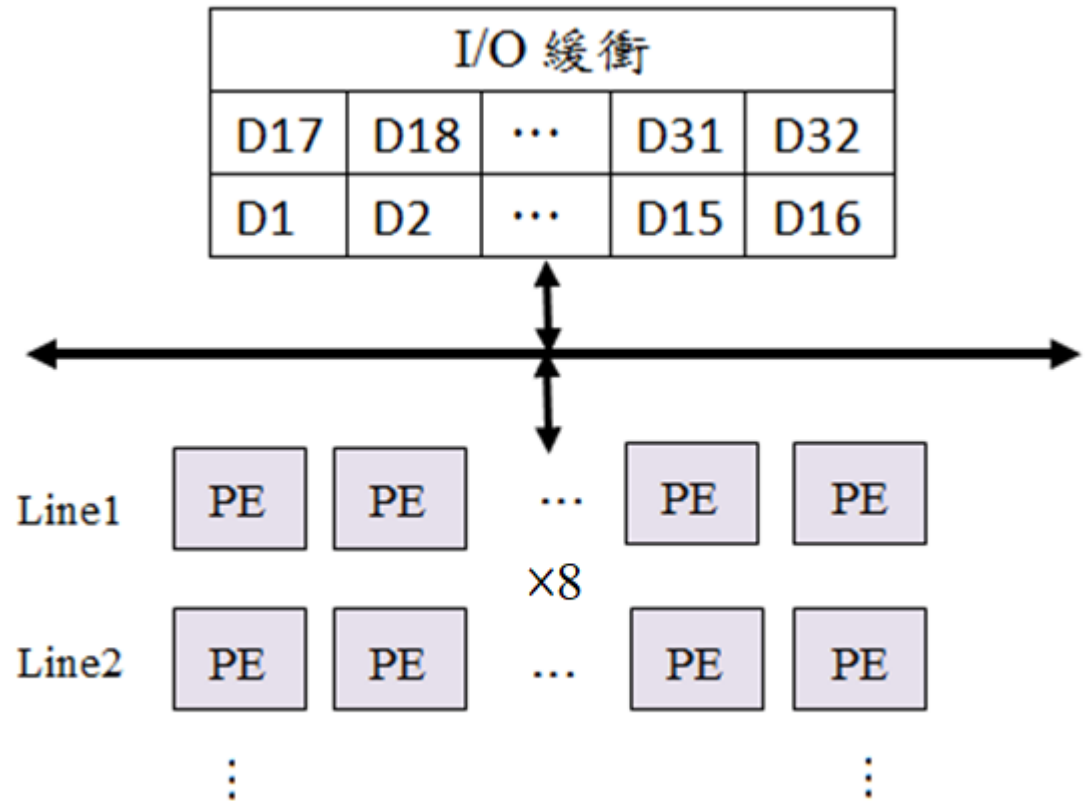


系統設計

- 提出一個低功耗的平行可規劃計算機架構。
- 提出一個降低晶片面積的線路交換網路連接路由方案。
- 加入低功率模式的能源管理系統。

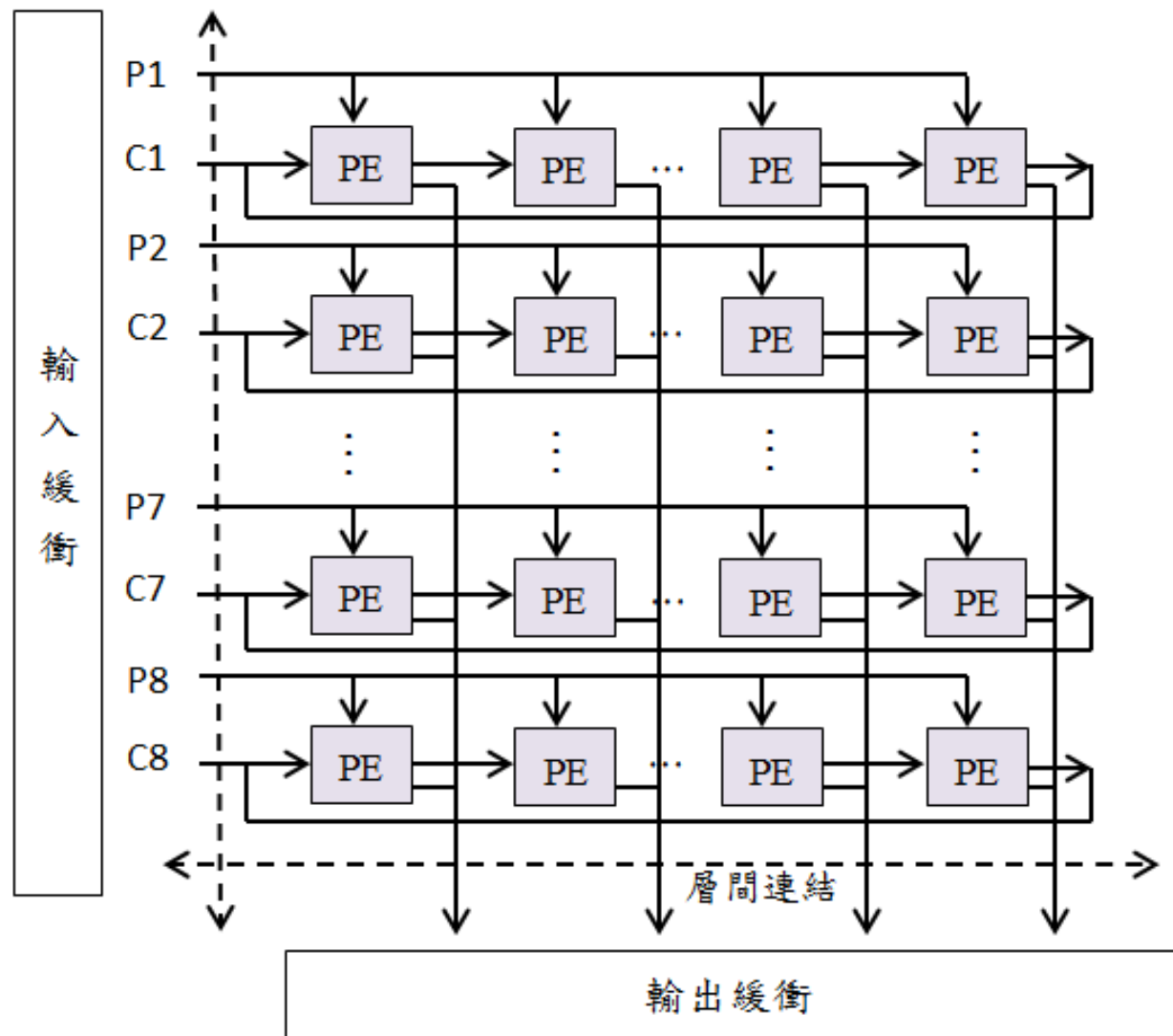
輸入輸出連接方案

- 解碼過程期間，將所取來源數據流通常是離散的，因此需要進行緩衝。
- 我們設計了兩個256位元寬的輸入/輸出緩衝於每個RPU處理。
- 比起傳統的每個PE連接I/O，只讓特定的PE的線連接到I/O



PE間連接架構

- 每個PE可以連接到在相鄰的前部和後部線，最後將連接到第一線，以支持疊代操作。
- 層間的互連讓RPU中每個PE能互通。

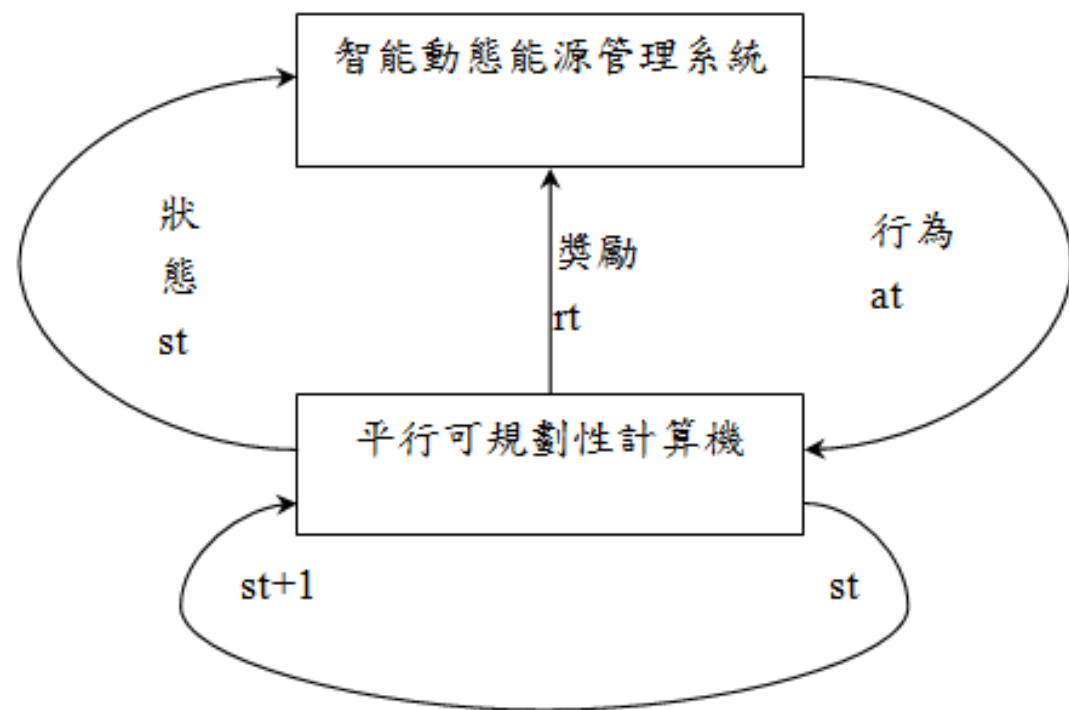


系統設計

- 提出一個高效能的平行可規劃計算機架構。
- 提出一個降低晶片面積的線路交換網路連接路由方案。
- 加入低功率模式的能源管理系統。

智能動態能源管理

- 代理檢查系統休眠時間 st (省電時間)，如果系統已經超過所閒置時間，代理獎勵當前策略 rt 。
- 反之，懲罰電流控制策略，並修正改變下一策略 at 。
- 最後，通過強化學習的方法預測最合適的政策。

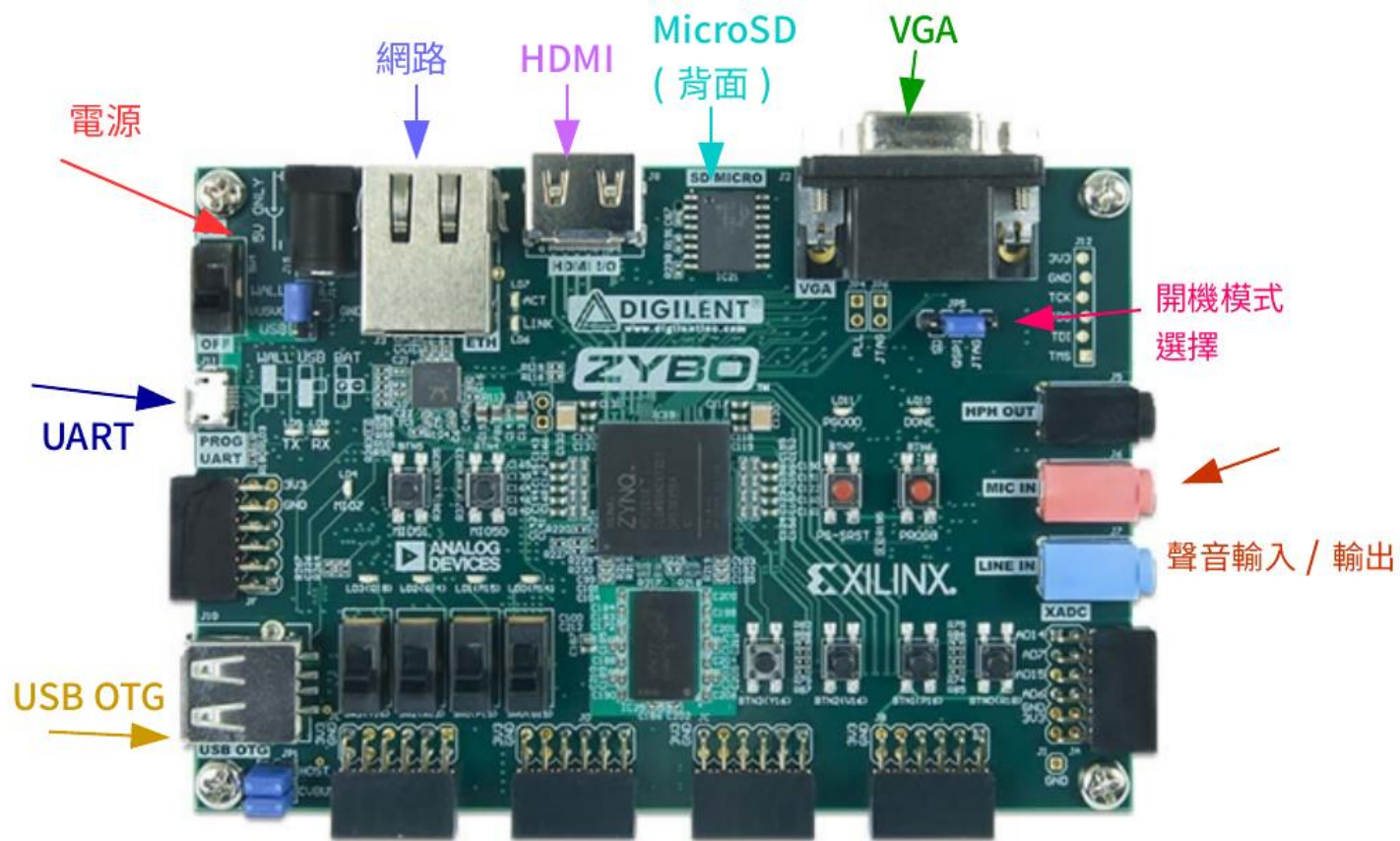


大綱

- 簡介
- 影像解碼平行化分析
- 系統設計
- 實驗結果
- 結論

實驗環境

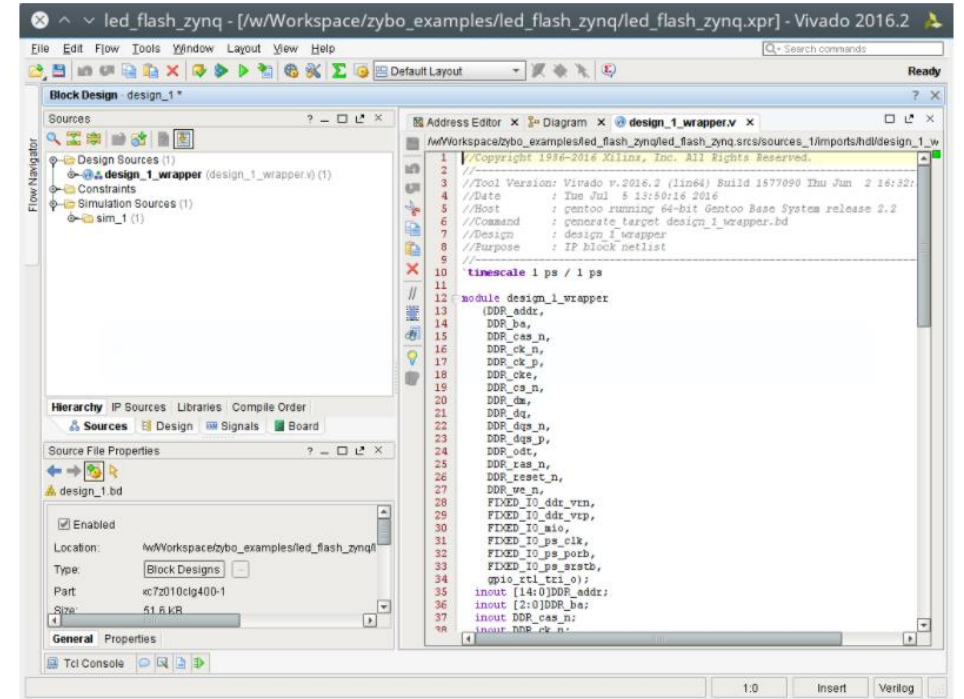
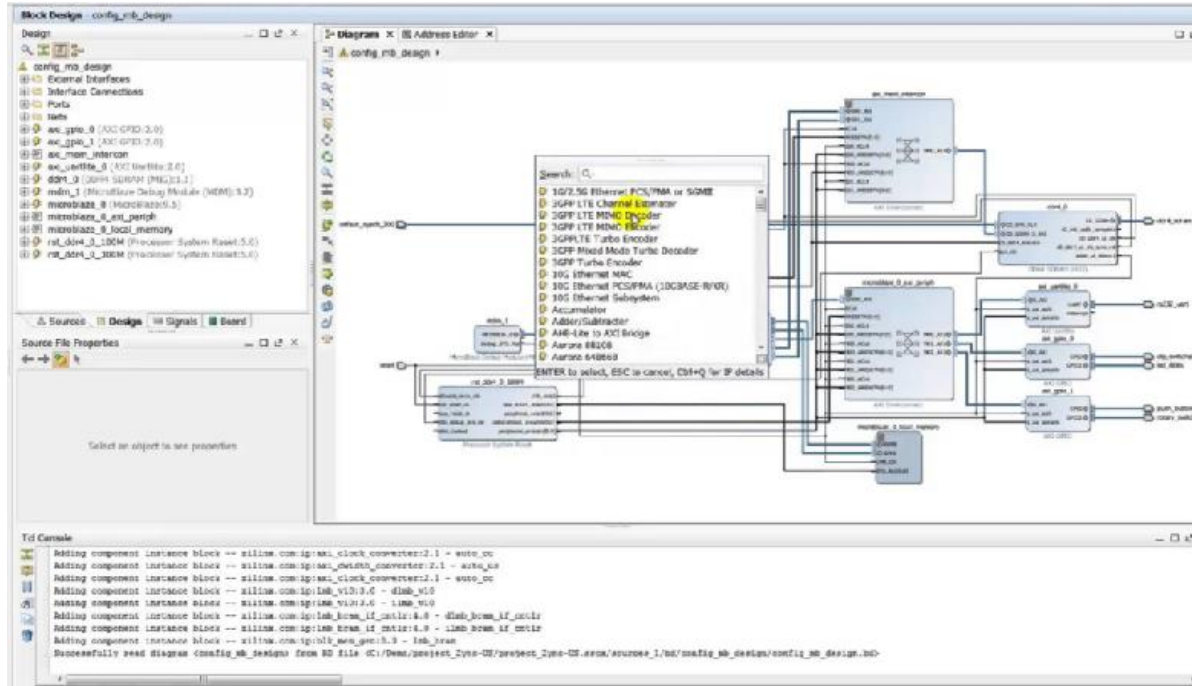
Zedboard



• ZYNQ7000系列

- 具有 FPGA 同時又包含了 ARM Cortex-A9 雙核心的開發板
- 28奈米的低功耗可程式邏輯
- 豐富的 OS、中介軟體、堆疊、加速器，以及 IP 生態系統
- 多重等級的硬體和軟體保全
- 整合提供實際全可編程平台
- 透過最佳化架構的系統層級效能
- 可提供最低系統功率的架構
- 最高彈性且可擴充的平台，提供多次重複使用和最佳 TTM
- 業界領先的設計工具、C/C++、OpenCL 設計概念

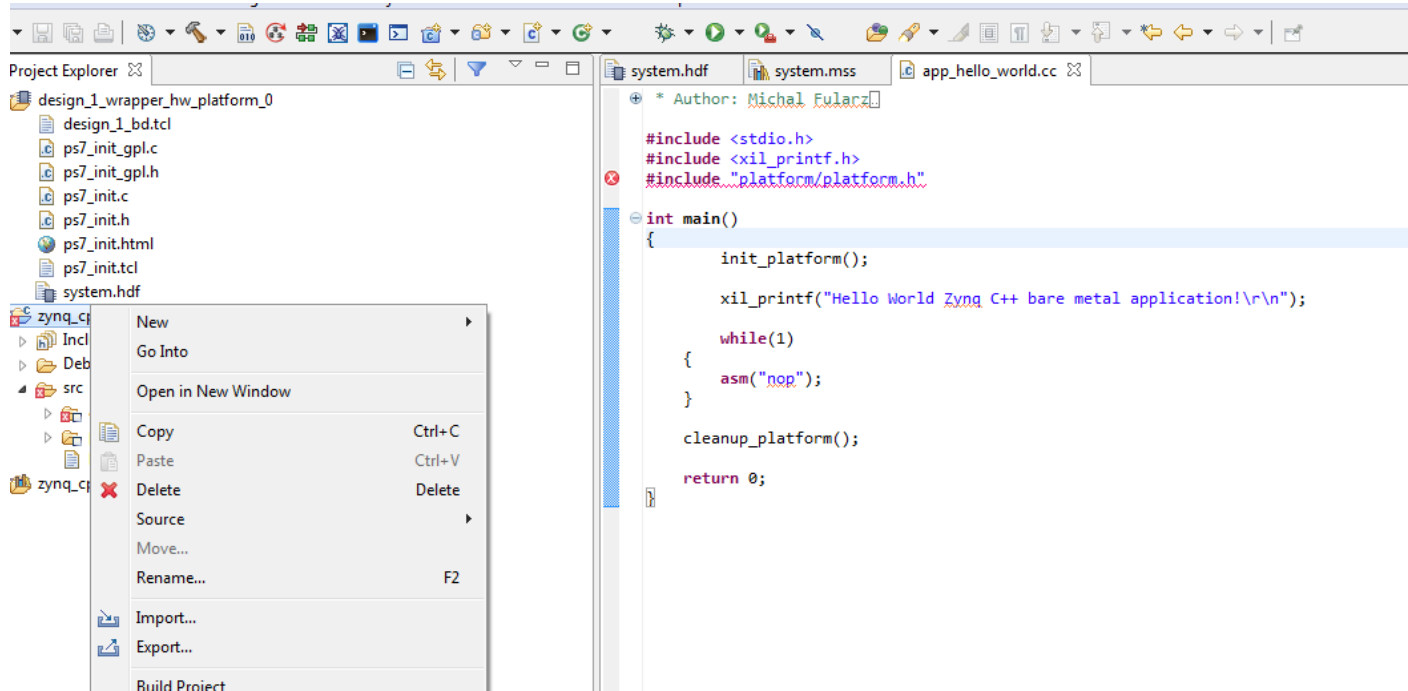
Vivado 套件



支援 Verilog 和 HDL。從 C、RTL 到 HDL，由高至低層，都可在單一套件的 GUI 環境中合成。

支援手動式的佈線規劃或分區。

Vivado 套件



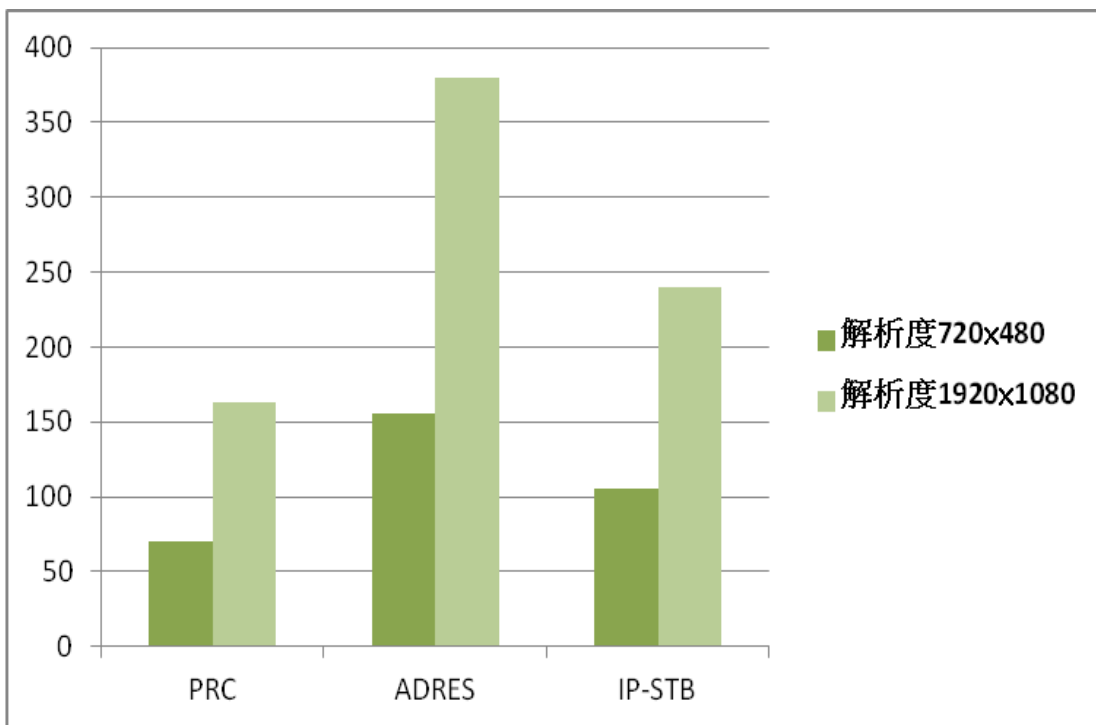
- 軟/硬體整合開發，支援高階到低階的設計。
- C程式碼能夠自動轉成RTL碼，並對應FPGA的邏輯，當輸出報告時便能立即檢查功耗預算或是晶片尺寸是否符合需求。當功耗或尺寸過大時，可立即將C程式碼轉為平行處理。

運算速度比較

	解析度 720x480	解析度 1920x1080
RPU	41100(MBs/s)	82300(MBs/s)
4*RPU	87850(MBs/s)	232910(MBs/s)
8*RPU	91550(MBs/s)	244560(MBs/s)

從實驗結果可以看出使用越多的RPU能夠有效的增加，使用四個RPU和使用八個的效能差距已經不明顯，因此在設計我們的平行可規劃計算機架構時連結了四個RPU形成我們的運算中心。

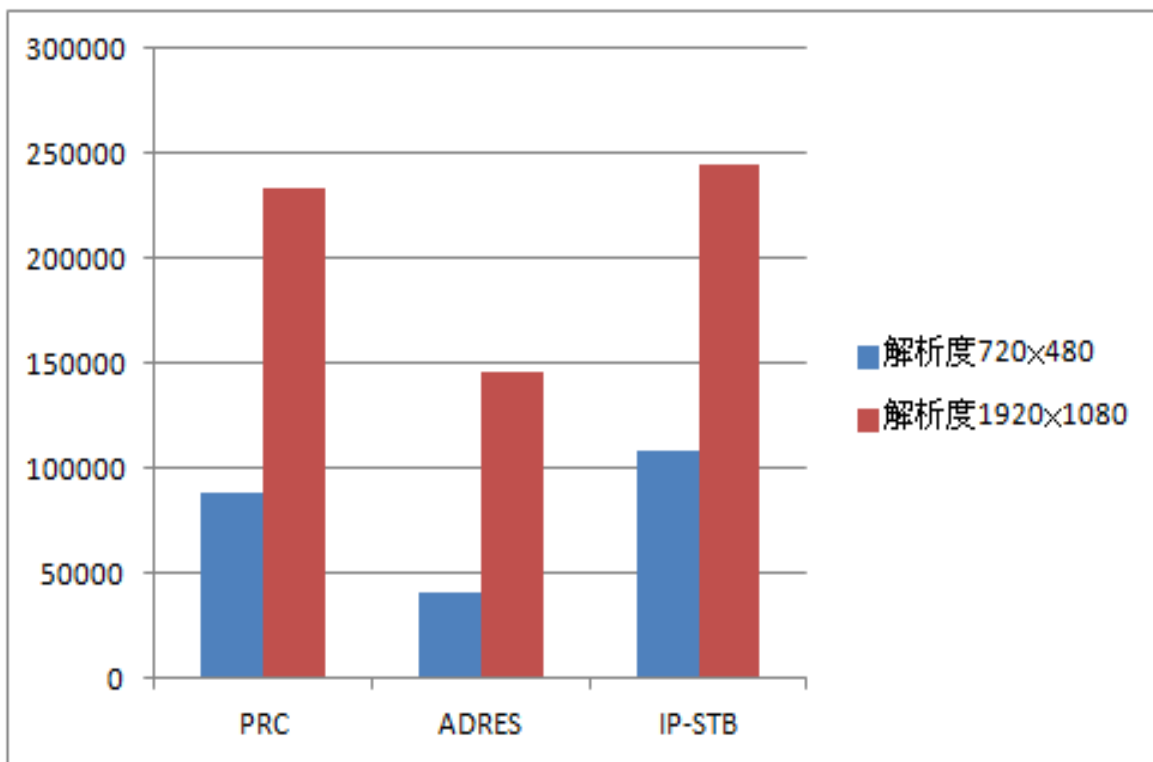
能源消耗比較



	<i>PRC</i>	<i>ADRES</i>	<i>IP-STB</i>
解析度 720x480	70mW	156mW	105mW
解析度 1920x1080	163mW	380mW	240mW

我們的平行可規劃系統(PRC)和稱為ADRES的影片解碼可規劃計算機架構(Mei, Bingfeng, et al, 2008)，與一個影片解碼專用的 DSP晶片 IP-STB (Pescador, Fernando, et al, 2008)做比較。從圖表可以看出我們的系統所耗的能源是三者之中消耗最少的。

效能比較



	<i>PRC</i>	<i>ADRES</i>	<i>IP-STB</i>
解析度 720x480	87850(MBs/s)	40500(MBs/s)	107892(MBs/s)
解析度 1920x1080	232910(MBs/s)	146200(MBs/s)	244684(MBs/s)

效能比較方面我們的系統明顯比起ADRES的可規劃計算機架構要好，雖然說專用晶片(DSP)的效能不管在哪一個解析度都是最好的，但我們的系統效能的其實差距不大。

結論

- 提出的架構能夠確實的提高效率，在節省能源消耗方面有不錯的成效，雖說計算能力比專用的DSP系統略低但也相去不遠。
- 所提出的架構在計算能力、能源消耗等方面的數據有所提升，證明所提出的計算機架構確實為可行的方案。
- 本架構即使在各方面都有所優化，但以整體架構的最佳化設計而言，仍然有很大研究的空間。