

Population-based Neuromorphic Spiking Neural Network Processor

神經群仿生神經網路處理器

專業領域：系統組

組別: B259

指導教授：鄭桂忠 教授

組員姓名：王語卉、柯佩妤

Abstract

隨著人工智慧的發展，各式各樣的神經網路是近期熱門的研究主題。在突波神經網路中，模擬了神經元接受刺激、超過閾值產生突波的過程。隨著神經網路的規模越來越大，神經元的連接變得相當複雜。運用突波神經網路的四個優勢：兩極化、與時間相關、稀疏性、事件驅動，我們可以用一塊架構化簡複雜的接線，稱作「位址事件表示」。位址事件表示接收了前一級神經元傳送出來的突波之後，會產生一組位址用來表示產生突波的神經元。根據這組位址，我們再去取得相對應的權重，以及突波要傳遞的方向，藉此化簡複雜的接線。

在這份研究中，我們選擇兩種位址事件表示進行實作，分別是階層位址事件表示，以及環狀位址事件表示。階層位址事件表示藉由有效的分層，減少每個訊號輸出的扇出，而能夠迅速取得輸入突波的位址，善於處理時間相近的突波輸入。環狀位址事件表示藉由許可在不同階層的環中流動，只要取得許可，端點便可以輸出位址訊號。環狀分層的設計，使得此位址事件表示能夠用同一個架構處理距離相近的輸入，也能處理距離較遠的輸入，兩者的表現都能有相當的優勢。

實作的結果讓我們發現，若要讓位址事件表示有更好的表現，則可以優化輸入到輸出的時間。另外，位址事件表示在大規模的神經網路系統中，能夠發揮更好的效果。而因為製程的進步，走線的耗能比例越來越高。為了減少接線，位址事件表示在一個系統的重要性漸趨上升。

Introduction

第一章、背景與動機

隨著人工智慧蓬勃發展，神經網路的發展與應用成為近期熱門的研究議題。當人類對分子生物學的研究程度越來越深入，便發現人體的神經元可以在密度非常高的條件下，消耗非常少的功率。因此科學家從神經元著手，模擬生物體神經元的行為，並以此為單位，創建各式各樣的神經網路系統，如：卷積神經網路、深度神經網路……等，以達到不同目的。

在眾多神經網路當中，我們以突波神經網路（spiking neuron network）作為研究目標，研究他的架構、模型、運作方式等等。突波神經網路有兩極化（binarization）、與時間相關（time-correlation）、稀疏性（sparsity）以及事件驅動（event-driven）等特性。

在閱讀有關突波神經網路的文獻時我們發現，其中幾篇文獻的突波神經網路晶片裡，有包含一塊區域稱做「位址事件表示」（Address event representation，後面以 AER 簡稱）。隨著神經網路的規模越來越大，動輒上千萬顆神經元，神經元點對點的傳輸需要相當的傳輸頻寬。因此，為了簡化複雜的點對點傳輸，我們使用 AER 這塊模組來處理突波訊息的傳遞。前面有提及突波神經網路的四個優勢，因此，AER 可以在突波發生的時候，根據突波發生的來源，而輸出一組「位址」代表突波的發生，進而根據這組位址去存取相對應的權重，以及決定突波的去向。

因為對 AER 運作的機制與模式感到興趣，於是我們讀了幾篇與 AER 相關的論文做研究，並挑出兩篇描述不同 AER 運作模式的論文做比較。在這次專題中，我們主要是想要將兩篇論文中的兩種 AER 用 verilog 實作出來，實際模擬 AER 運作的形式，並比較兩者的表現與效率等等。

第二章、研究主題

一、突波神經網路（spiking neuron network，簡稱 SNN）

突波神經網路主要是模擬生物體神經元的行為，以神經元為單位所建立出來的神經網路架構。在生物體中，不同神經元之間會用「突觸」

（synapse）作為連結。突觸會收集所有前一級神經元所產生的突波，各自乘上相對應的權重之後做累加，再加上偏壓（bias）之後傳送給下一級的神經元作為輸入。以方程式表示為：

$$y_j = \sum_i w_{i,j} \times x_i + b_j$$

因此，下一級的神經元會接收到突觸做完累加之後的結果。
在神經網路中，根據神經元的連接方式，我們可以將神經元分層。有外部輸入端的那層稱作 input layer，有外部輸出端的稱作 output layer，其餘則稱之為 hidden layer。

二、位址事件表示（address event representation，簡稱 AER）

在巨大的神經網路中，面臨複雜的神經元連接，我們可以用 AER 這個模組來取代。他運作的方式是，將每個神經元的輸出端接到 AER，成為 AER 的輸入。AER 根據突波的來源，從一條共用的輸出線（shared output bus）輸出一組「位址」（address），用來表示這個位址的神經元產生了突波，進而取得相對應的權重資訊，以及要送往下一級的方向。如此一來，在大規模的神經網路中，便可以免除複雜的接線，權重與連接也有了更高的彈性。AER 可以用編碼與解碼的概念解釋，將所有產生的突波編碼成位址，再傳送到下一級成為輸入。

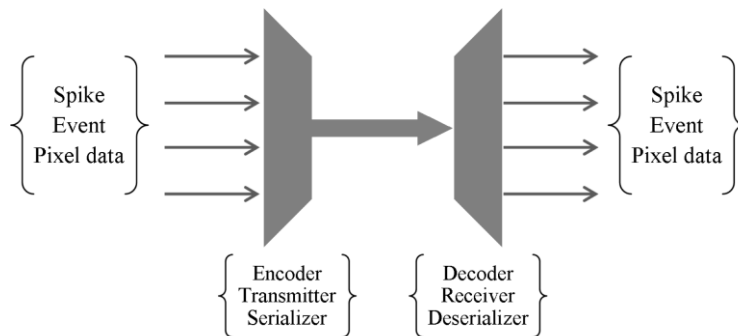


Fig. 1 AER 編碼與解碼示意圖。除了運用在神經網路中，也能運用在圖片處理

因為突波神經網路有著兩極性（binarization），只需要判斷是否產生突波，不用記錄其他資訊，所以可以用一個位址去表示。加上突波神經網路的稀疏性（sparsity），使得 AER 有足夠的頻寬可以用一條共用通道傳送位址。而時間相關（time-correlation）與事件驅動（event-driven）的特性，會隨著突波的發生一併繼承在位址資訊中。

接著，以下介紹我們這次主要研究的兩種 AER。

1. 階層位址事件表示（Hierarchical AER）

將神經元分群成好幾個核心（core），當位於不同核心的神經元想要進行突波傳遞時，會利用 L1 和 L2 等共享匯流排（Shared-bus communication）進行資料傳遞。核心又分成幾個簇（cluster），在同個簇

裡利用 L1 溝通，若在不同的簇裡，則利用 L2，當層級距離越遠的神經元要進行傳遞時，所需要花費的時間也就越長，以此達到分群分層的方式。

L1和 L2利用遮罩（mask bit）的方式決定要跟哪個核心進行溝通，所以在同一層時，神經元間即可用一組二進位數字溝通。當前一個神經元要傳遞突波給不同簇的神經元時，就需要先將訊息存在中繼連接（relay connection），再將其訊息傳遞給所有的目標神經元。因為層級越高造成傳遞時間越長，我們必須先利用演算法等將神經元適當的分組，降低到達高層級的頻率。

而階層位址事件表示分層的特性讓神經元在傳遞突波資訊時，能夠平行運作。相較於單純平面的傳遞減少需要的 cycle。

2. 環狀位址事件表示（Token ring AER）

利用兩種環（ring）分別為由 leave servers(Lserver)和 higher server(Hserver)組成，進行神經元分群分層。每個神經元接上一個 Lserver，而四個 Lserver 被分為一組環，上面接著一個 Hserver。當這四個 Lserver 要跟其他 Lserver 進行溝通時，則會利用上層的環傳遞。當神經元想要輸出突波的地點和時間訊息到共享匯流排時，需要先取得許可（token）。為了取得許可，該神經元的 Lserver 會向下一個 Lserver 拿取許可，若下一個 Lserver 沒有，再繼續向下一個 Lserver 要求。若位於同一環的 Lserver 都沒有許可，就會向連接此環的 Hserver 索要，若該 Hserver 沒有，則向下一個 Hserver 索要，以此形成階層環狀的輸出方式。

所有的 Lserver 都會連接到 Lcounter，用來儲存許可所在 Lserver 的「相對位置」，第一層的環有 4 個 Lserver，所以 Lcounter 的只會有 0、1、2、3 四種情況。而所有的 Hserver 會連接 Hcounter，與剛剛介紹的 Lcounter 的功能一樣，只是改成紀錄許可所在 Hserver 的相對位置。利用 Lcounter 和 Hcounter，我們就可以得到許可存在的「絕對位置」。

第三章、 研究方法

我們根據上一章所提到的階層位址事件表示和環狀位址事件表示進行 verilog 硬體實作，並比較其不同之處。

一、階層位址事件表示（Hierarchical AER）

我們建立了 1 對 20 的神經元列，同時規定輸入神經元的突波一次只

能傳給一個神經元，所以可以模擬出 AER bandwidth 的限制，表現突波輸出共需要的時間。

為了模擬階層位址事件表示，我們將 20 個目標神經元分成 4 組，每組有 5 個神經元，每組的上層有一個虛擬的中繼神經元（neuron）傳遞訊號。所以輸入神經元會先傳遞訊號給中繼神經元，再由中繼神經元傳遞給該組的神經元群。

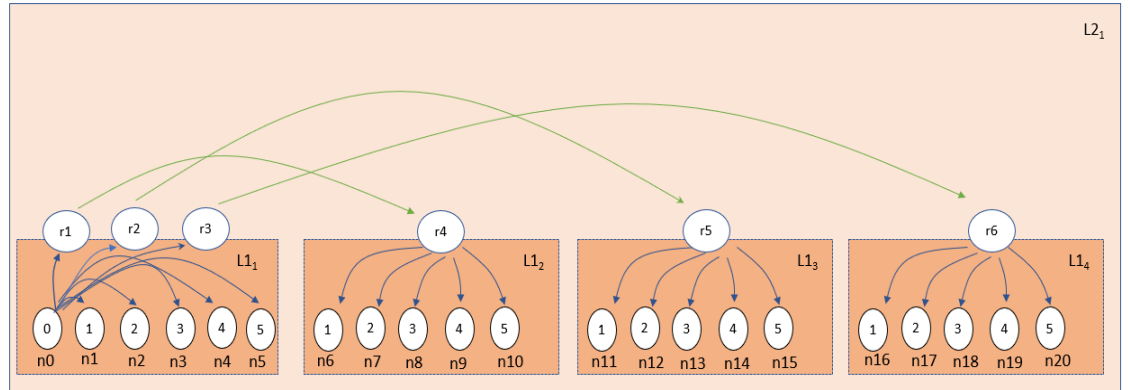


Fig. 2 本組階層位址事件表示實施架構。為 1 對 20 神經元，將 20 個神經元評分

如 Fig2.，r1、r2、r3、r4、r5、r6 為 6 顆中繼神經元，而灰色圓圈代表輸入和目標神經元，n1 代表第一顆神經元，箭頭為傳遞訊號的方向。第 20 顆神經元會是最後一個接收到突波的神經元（根據我們的設定），因此整個訊號傳遞的過程結束於第 20 顆神經元接收到訊號。完成時間的比較會由第 20 顆神經元決定。

二、環狀位址事件表示（Token ring AER）

階層以四組 Lserver 為一環，高階層以三組 Hserver 為一環，而低階層與高階層之間則用了四條線來傳遞資訊。如圖所示：

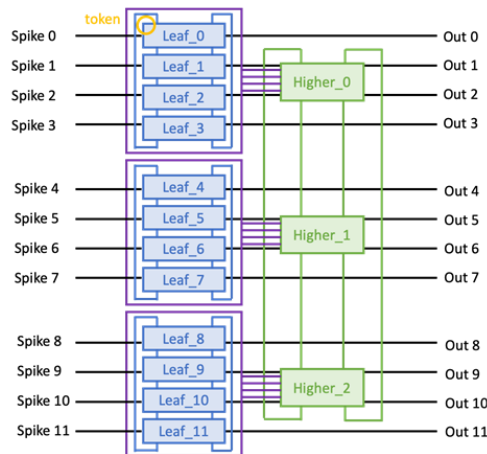
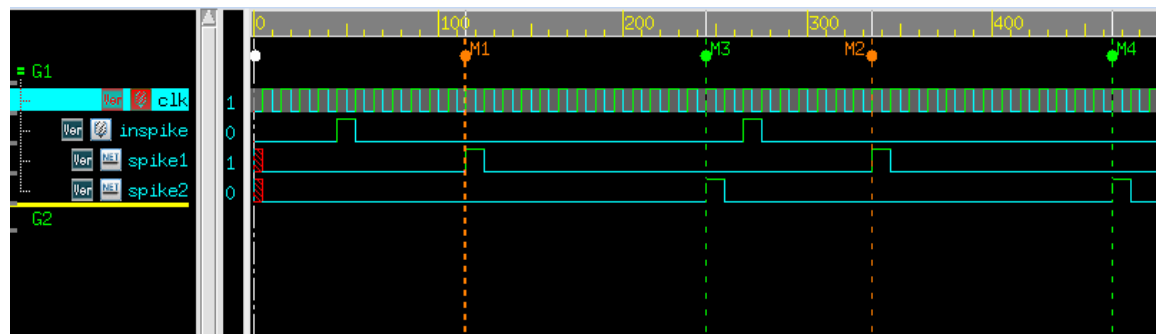


Fig. 3 Hierarchical Token Ring AER block

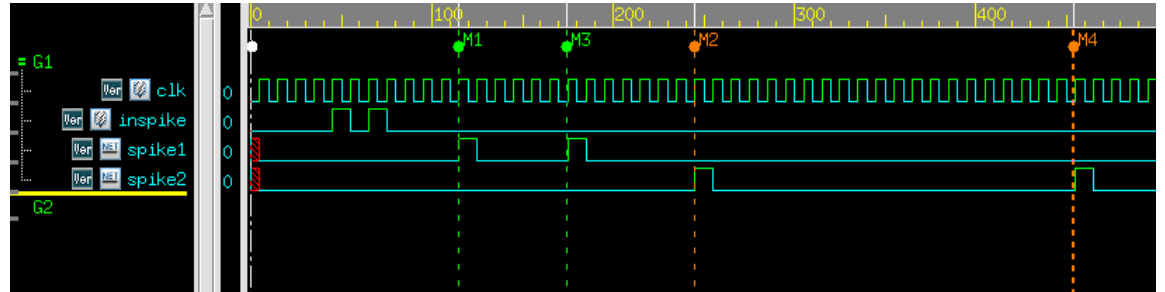
其中，spike 訊號為接收前一級神經元輸出的訊號，out 訊號為 Lserver 獲取許可（token）之後的輸出結果。許可（token）在初始設定時，會在 Lserver 的 Leaf_0 位置。在這個系統裡，因為一個環的數量不多，因此設計過程我們省略了 Lcounter 與 Hcounter，而是用其他訊號判別許可（token）的位置。

第四章、 研究結果

一、 階層位址事件表示（Hierarchical AER）



(a)



(b)

Fig. 4 成果波型圖 (a) 輸入突波時間差大，可以發現不做處理的 AER 和以階層位址事件表示（HiAER）的兩個輸出突波時間差相似 (b) 輸入突波時間差小，則不做處理的 AER 和以階層位址事件表示（HiAER）的兩個輸出突波時間差差異大，表現出分層分群的優勢。但兩種情況都可發現以階層位址事件表示處理的 AER 輸出突波的時間較快

根據 Fig. 4，我們可得階層位址事件表示的表現情況。

Inspike 為編號0的神經元，當 inspike 為一，代表編號0的神經元要傳遞訊號。spike1為階層位址事件表示編號20的神經元，當他為1時，代表接收到編號0傳遞的訊號。spike2為不做任何處理的1對20神經元傳遞，編號20的神經元，當他為1時，代表接收到編號0傳遞的訊號。

當輸入突波時間差大如圖 a，可以發現不做處理的 AER 和以階層位址事件表示（HiAER）的兩個輸出突波時間差相似。但當輸入突波時間差小，則不做處理的 AER 和以階層位址事件表示（HiAER）的兩個輸出突波時間差差異大，表現出分層分群的優勢。但兩種情況都可發現以階層位址事件表示處理的 AER 輸出突波的時間較快。

二、環狀位址事件表示（Token ring AER）

我們檢測的方式是：對不同 Lserver 輸入突波訊號（spike），看相對應的輸出（out）是否為正確的位址。從結果可以得證，當輸入突波之後，經過一些運作時間，可以確實輸出接收到突波的位址。從波形圖可以應證我們的設計：許可（token）在一個時間點，只會被一個 server 所持有，而根據不同 server 的索取，許可在環狀結構中相互傳遞。

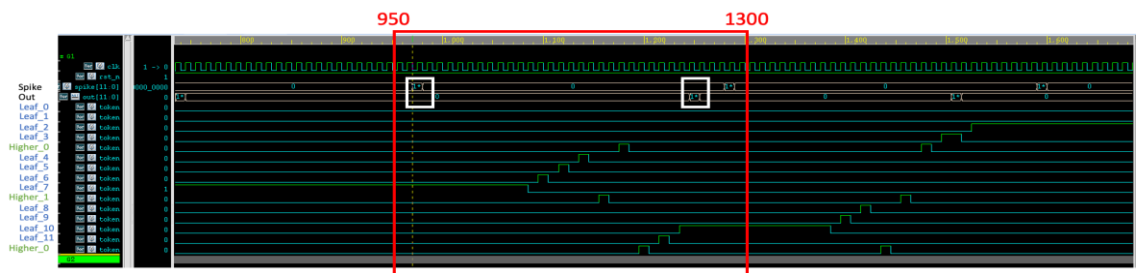


Fig. 5 白色方框為輸入與輸出。 波形為每個 server 所擁有 token 的情形。

第五章、總結

經過實作和閱讀多篇論文之後，我們有五個結論：

一、階層位址事件表示（Hierarchical AER）面對輸入突波時間差短，以及扇出（fanout）龐大的神經網路具有優勢。

二、環狀位址事件表示（Token ring AER）是個通用性結構，不論輸入突波情況皆具有一定的優勢

三、兩種 AER 都需要一段時間（cycle）才能取得位址。因此若要改良 AER 的話，我們認為可以思考更節省時間的方式實作 AER。

四、AER 的發想是源自於巨大的神經網路，但為了驗證文獻的內容，我們選擇從小規模的神經網路著手。因此我們認為若要明顯發揮 AER 的功能，需要先建立相當規模的神經網路，再進行 AER 連接，方能進行較具參考價值的比較。

五、AER 可以節省複雜的走線與連接。而在先進製程中，走線（wire）的功耗隨製成的進步而相對越來越多。因此我們認為，為了節省行動裝置的耗能，AER 的研究是未來數位電路發展中漸趨重要的一環。

第六章、心得感想

從一年前開始閱讀大量文獻，我們對突波神經網路的架構越來越熟悉，除了了解突波神經網路、神經元模型之外，更了解了對我們來說是全新的一塊領域：AER。在閱讀文獻的過程中，因為數位晶片的內容都非常龐大，很難透過少少幾頁文獻便深入理解完整的運作過程。因此從書面轉為硬體語言的這段過程我們認為是最困難的部分。除了跟夥伴討論溝通，也很感謝林郁軒學姐一直從旁給予支援與幫助，讓我們得以在最後得到成功的結果。

雖然專題的僅實現了非常小部分的成果，但這次的實作經驗讓我們深入了解突波神經網路以及 AER 對於神經網路的重要性，我們覺得收穫非常豐富。未來如果有機會繼續實作類似的主題，我認為我們可以試著把規模變大，也可以進一步探討不同區塊的面積以及耗能，優化現在的模型。

除了在突波神經網路部分學到很多知識之外，我們也學習到如何讀懂文獻並做摘要，以及如何有效討論與分工。很感謝這次專題的經驗，也感謝夥伴、學長姐以及指導教授。