

國立清華大學 電機工程學系

實作專題研究成果摘要

Architecture of Address Event  
Representation based Spiking Neural  
Network

位址事件表示突波神經網路架構

專題領域：系統領域

組 別：A300

指導教授：鄭桂忠

組員姓名：蔡昀蓁 厲彥剛

研究期間：111 年 8 月 11 日至 112 年 5 月止，共 10 個月

## 摘要

機器學習是近年最熱門的話題之一，有關神經網路的研究結果不斷的推陳出新，而在現存的多樣神經網路中，我們選擇了突波神經網路進行實作。突波神經網路具有事件驅動特性，只有在神經脈衝發生時會進行權重的傳遞，有如生物的神經系統。由於事件驅動的特性，將事件發生位址和資料類型進行包裝進行傳輸的位址事件表示（AER）非常適合應用在突波神經網路的資料傳遞上，我們選擇了兩種 AER 架構進行比較，分別是階層位址事件表示（Hierarchical AER）和環狀位址事件表示（Token ring AER）。

階層位址事件表示利用中繼神經元分別將輸入層、中間層、以及輸出層的神經元分成三層，並透過中繼神經元把神經衝動訊號轉成 AER 封包傳到神經網路的下一層，或把神經網路上一層傳來的 AER 轉為權重並輸入神經元，階層位址事件表示能夠迅速處理時間相近的突出輸入；環狀位址事件表示的神經元透過取得在環中跑動的 token 取得輸出許可，將 AER 轉為權重加總之後輸入下一層，遇到時間相近的突波輸入時會需要額外的排序時間。

實驗結果符合我們一開始的推論，由於環狀位址事件表示需要依序處理突波的行為，在全連接且輸入中度密集的情況下環狀位址事件表示需要花上比階層位址事件多出幾百倍的時間。不過若在非全連接的情況我們推測，隨著輸入密集程度的下降，環狀位址事件表示的速度將會貼近階層位址事件表示。

## 第一章、研究概要

機器學習的基本想法是利用電腦模擬人類的智慧，讓電腦能夠擁有和人類相似的思維和行為。科學家透過研究生物的神經系統，試圖用電腦程式模擬神經元的行為，並以此為單位連接無數個神經元，建立各式各樣的神經網路模型。包括人工神經網路（Artificial Neural Network，ANN）、深度神經網路（Deep Neural Network，DNN）、較主流的卷積神經網路（Convolutional Neural Network，CNN），以及這次的研究目標，突波神經網路（Spiking Neural Network，SNN）。

相較於其他神經網路，我們研究的突波神經網路具有以下四個優勢：

- 一、訊號二極化（Binarization）：突波訊號的傳遞只有0、1兩種可能，代表突波（spike）的發生與否。因此在後端的累加運算上只用到加法而不需要乘法運算，在硬體和時間上都能省下大量資源。
- 二、突波和時間相關性（Spike-Time correlation）：突波訊號的發生是時間的函數，可以保存時間維度的資訊，因此突波神經網路很適合連續訊號與影像的應用。

三、突波稀疏性 (Sparsity)：突波的發生具有稀疏性，大量的神經元同時僅會有幾個神經元發出突波訊號，使得傳輸頻寬需求大幅降低。

四、事件驅動 (Event-driven)：因神經系統只有在神經元累積電位超過閾值產生突波訊號時驅動，因此突波神經網路具有事件驅動的特性，可以節省消耗的功率。

綜合以上優點，突波神經網路具有低功耗、低延遲的特性，且相較其他神經網路，突波神經網路的行為模式和生物神經系統十分相近，為相當具有潛力的神經網路架構。

當我們決定著手研究這個主題後開始閱覽相關論文，並發現其中一個研究方向：「位址事件表示」(Address event representation, 簡稱AER)。由於神經網路的規模可以大至有上百萬顆神經元，若將所有神經元點對點兩兩相接，導線的規模將過於龐大，為了簡化接線問題，AER在突波發生的時候，會找到突波發生的神經元，並用「位址」紀錄下來，再根據突波的目的地讀取權重，藉此減少接線的數量。

由於對 AER 的妙用感興趣，我們接下來讀了幾篇和 AER 相關的論文，並從中選出兩篇不同的 AER 架構。在此研究中我們將用 verilog 把兩者的模型架構實作出來，並比較兩者的表現。

## 第二章、研究主題

### 一、突波神經網路 (Spiking Neural Network, 簡稱 SNN)

首先，神經網路主要指的是，透過由神經元為單位組成之神經網路架構，在計算機上模擬實際生物體中的神經運作。而突波神經網路，顧名思義，即是藉由神經元與其他神經元之間傳輸的「突波」驅動整個神經網路的運作。

再來，針對各神經元之間的結構細說。在生物體內，神經元之間透過「突觸」(synapse) 連接對方，並且相互之間有著「刺激性」、或是「抑制性」。以上敘述，在計算機中，則是在各個突觸上，賦予其「權重」(weight)，並且在運算中，將前一層級的神經元所產生之突波乘上各自突觸上所記載的權重，最後將其加總，輸入下一層級的神經元。以下為其方程式：

$$y_i = \sum_j w_{i,j} \times s_j$$

$w_{i,j}$  為編號  $i$  的神經元與編號  $j$  的神經元之間的突觸權重， $s_j$  為編號  $j$  的神經元是否發出突波訊號。

接著，縮小焦點至神經元。突觸累加完的訊號輸入神經元之後，會被進一步累加至神經元內部的膜電位（membrane potential），一旦累加的膜電位超過此神經元的閾值（threshold voltage），即判斷產生突波，輸入到下一層級的神經元。

## 二、位址事件表示（Address Event Representation，簡稱 AER）

位址事件表示是一種傳遞訊號的處理方式，在 AER 中，訊號不只代表某個參數的變量，而是將訊號編譯為一串二進制編碼，並送到共同的輸出端，再由接收端解碼，而這串 AER 編碼中包含了訊號來源、訊號類型、訊號目的地等資訊，常被用於處理感測器的數據，如聽覺、視覺訊號和無人機系統。

不同於一般計算機對訊號的時脈同步的處理方式，位址事件表示是由事件驅動的，也就是訊號只在事件發生時被傳輸，而不是以固定的週期連續取樣。這種訊號的傳遞方式優勢在於可以減少功率與延遲，故在事件稀疏的系統上有很好的效能表現。儘管 AER 擁有多種優點，它還是存在明顯的短板，例如在遇到密集或同時的事件發生時，容易出現延遲、訊號時序錯誤等問題。

## 第三章、研究方法

本次專題的研究目的為，對「階層位址事件表示以及環狀位址事件表示在突波神經網路上的差異」進行分析，因此，我們選擇透過實作到應用端，觀察兩種位址事件表示在突波神經網路架構上所展現出的結果，再做比較。

應用端的部分，我們將利用 MNIST 手寫數字的資料集，輸入突波神經網路，讓它輸出對此數字的辨識。因為 MNIST 手寫數字是 16 x 16 pixels 的圖片，所以令輸入端的神經元數量為 256，同時，判斷出的數字是從 0 到 9，因此，輸出端的神經元數量為 10，進一步，在輸入層和輸出層中間加入兩層神經元數量為 128 的中間層（Hidden Layer）。除此之外，設定此神經網路屬於全聯接網路（Fully Connected Network，簡稱 FC network）。最後，透過 Python 使用開源機器學習資料庫 Torch，針對此突波神經網路進行模擬，訓練出各神經元之間的權重。

以下將分別解說兩種位址事件表示架構，以及實作兩種架構的方法：

### 一、階層位址事件表示（Hierarchical AER）

階層位址事件表示的架構，正如其名，將突波神經網路的神經元分組（partition），並且以階層的方式，鏈接各個神經元。

在本次專題，分組的步驟中，我們設定在「L1」內存在 16 個神經元，因此，總共會有 16 個輸入層的「L1」，8 個第一中間層的「L1」，8 個第二中間層的「L1」，與唯一的例外 1 個僅存有 10 個神經元的輸出層「L1」；接著，令輸入層、第一中間層、第二中間層、輸出層，各自形成自己的「L2」；最後，所有「L2」皆存在於「L3」之下。

在連接的部分裡，我們設定每個階層之間都會有 4 個中繼神經元存在。最終，就能夠形成 Fig 1.的突波神經網路架構。

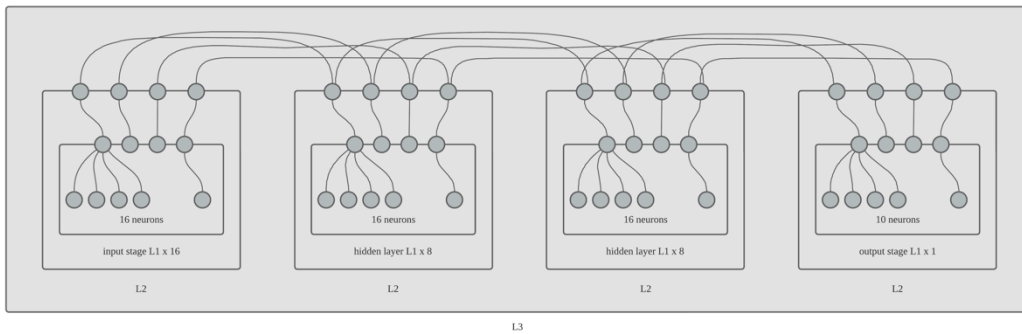


Fig 1. 本次專題設計之 hierarchical AER SNN 之架構

更進一步，關於中繼神經元，它是一個假想的神經元，只是以和神經元相似的功能「接收訊號，再傳輸訊號」運作。我們將中繼神經元分為兩種功能，分別是 Spike-to-AER RN、以及 AER-to-Weight RN。以下將各別解釋：

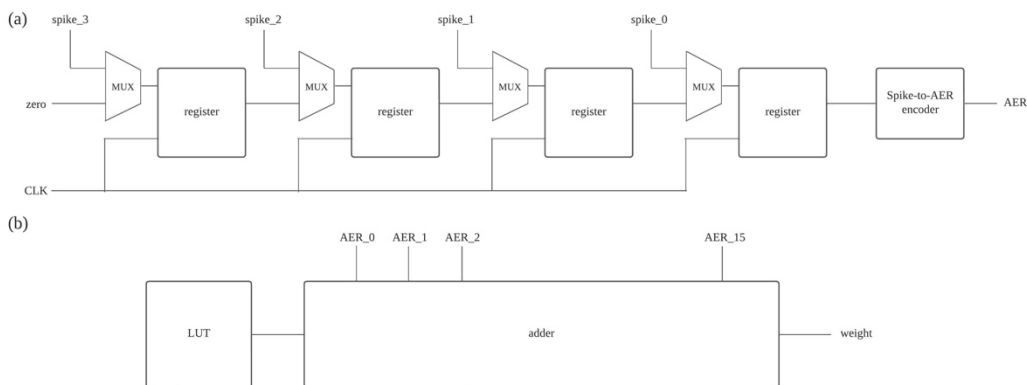


Fig 2. (a) Spike-to-AER RN 之硬體架構 (b) AER-to-Weight RN 之硬體架構

首先，Spike-to-AER RN。它的硬體架構如 Fig 2. (a)，它的功能是将一個時間步 (Timestep) 中的突波轉換為位址事件表示的形式，並且一個時間步中可以輸出 4 個輸入神經元的位址事件表示，因此，才會設計層與層之間的中繼神經元數量為 4 個。

第二，AER-to-Weight RN。它的硬體架構如 Fig 2. (b)，它的功能是将一個時間步 (Timestep) 中的位址事件表示，透過查詢查找表 (Look-up Table，簡稱 LUT)，轉變成權重，並且一個時間步中可以對位址事件表示所代表的權重進行累加，得到下一層神經元的輸入訊號。

## 二、環狀位址事件表示 (Token ring AER)

顧名思義，環狀階層即是在權狀環 (token ring) 的基礎上再加入分層的結構，形成有多個環的樹狀網路。

權狀環 (token ring) 最初是應用於電腦網路的连接協定，在一個區域網路中，每個工作站透過和兩個相鄰的工作站相連形成一個環狀網路。環形網路內存在一個許可令牌 (token)，在環形網路內以單一方向傳遞，需要上傳資料的工作站必須要求 token，只有在拿到 token 的時候才可以把資料上傳，如此一來網路上的工作站就能依序輪流取得輸出的權限，資料的上傳不會發生衝突。在此篇設計的 token ring AER 中，我們用 ring server 擔任 ring 中工作站的角色。

而環狀階層即是把多個權狀環以樹狀結構連接起來，所以有 higher level ring 和 leaf ring。

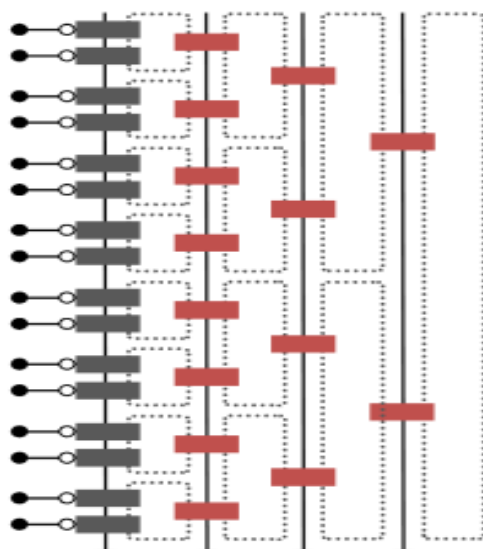


Fig 3. Token ring AER 架構<sup>[4]</sup>

如 Fig 3.，位於 leaf ring 的灰色 server 稱為 leaf server (簡稱 Lserver)，位於高層 ring 的紅色 server 為 higher server (簡稱 Hserver)，每一個 Lserver 都連接著一個神經元，當神經元產生突波時，連接該神經元的 Lserver 便會開始要求 token，token 在環狀階層中傳遞和索取的規則如下：

- (1) Lserver 有 token，直接輸出訊號；反之，則向下一個 Lserver 索取 token。

- (2) 被要求傳遞 token 的 Lserver 若無 token，則向下個 Lserver 索取；若有則向索取者傳遞 token。
- (3) 若索取傳遞到 leaf ring 的最後一個 Lserver 仍無法取得 token，向上一層的 Hserver 索取。
- (4) Higher level ring 在一個環裡遵循相同的規則。
- (5) Hserver 具有變數紀錄 token 的位置，token 的位置有三種可能：
  - (i) Hserver 擁有 token。
  - (ii) 下層的 rings 擁有 token，向下層的 ring 索取 token。
  - (iii) Hserver 和下層的 ring 都沒有 token，往下一個 Hserver 索取 token。
- (6) 要求輸出的 Lserver 獲取 token 後，token 最終停留在該 Lserver 等待下一次的事件發生。

由此可知，不管在 higher ring 還是 leaf ring，token 都是以順時針方向傳遞的，且索取 token 的訊號會持續直到接收到 token 為止。

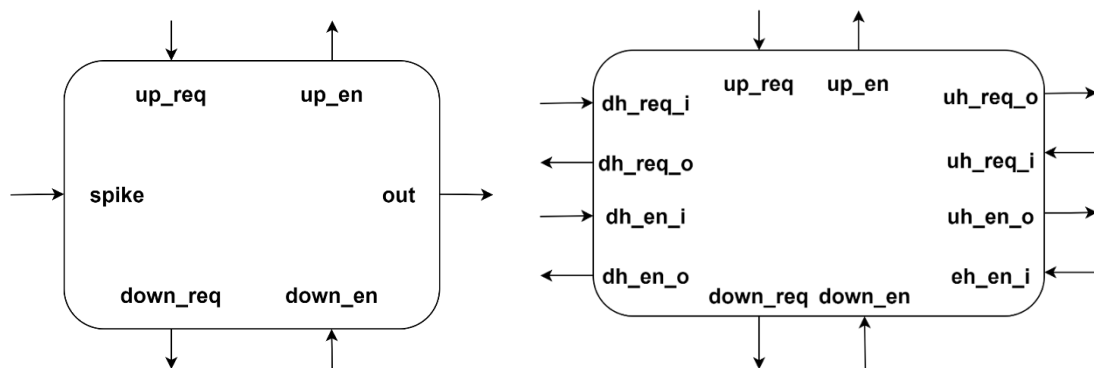


Fig 4. (a) Lserver 方塊圖 (b) Hserver 方塊圖

Req 表示 token 的索取，en 表示 token 的傳出和傳入，token 藉由 req 和 en 的信息流在 ring 中傳遞。Fig 4. (b) 中可見 Hserver 中多了層級間傳遞的訊號，dh 類訊號表示和低階層間往來的訊號，uh 類訊號則負責與高階層之間的通訊。

由於 token ring 一次只能讓一個節點輸出的性質，若允許多個 Lserver 同時要求 token，將會發生 output 時序和 token 索取流的錯亂，因此當多個神經元同時發生衝動時，需要簡單的排序機制讓神經衝動可以依序輪流進入 AER 內進行傳輸許可。

在我們設計的網路中，Lserver 依照順序進行編號，當一個 Lserver 拿到 token 時，整排 Lserver 的輸出會形成一段位址，擁有 token 的 Lserver 輸出 1，其餘的 Lserver 全部輸出 0。輸出端在收到這段編碼後進行解碼，從 LUT 中調

取對應權重，並將一個時間步內同時發生的所有神經衝動的權重累加，累加完成後輸出給下一層神經元接收。

## 第四章、研究結果

Fig 5. (a) 以及 Fig 5. (b) 分別是階層位址事件表示和環狀位址事件表示的判斷數字 7 的結果：

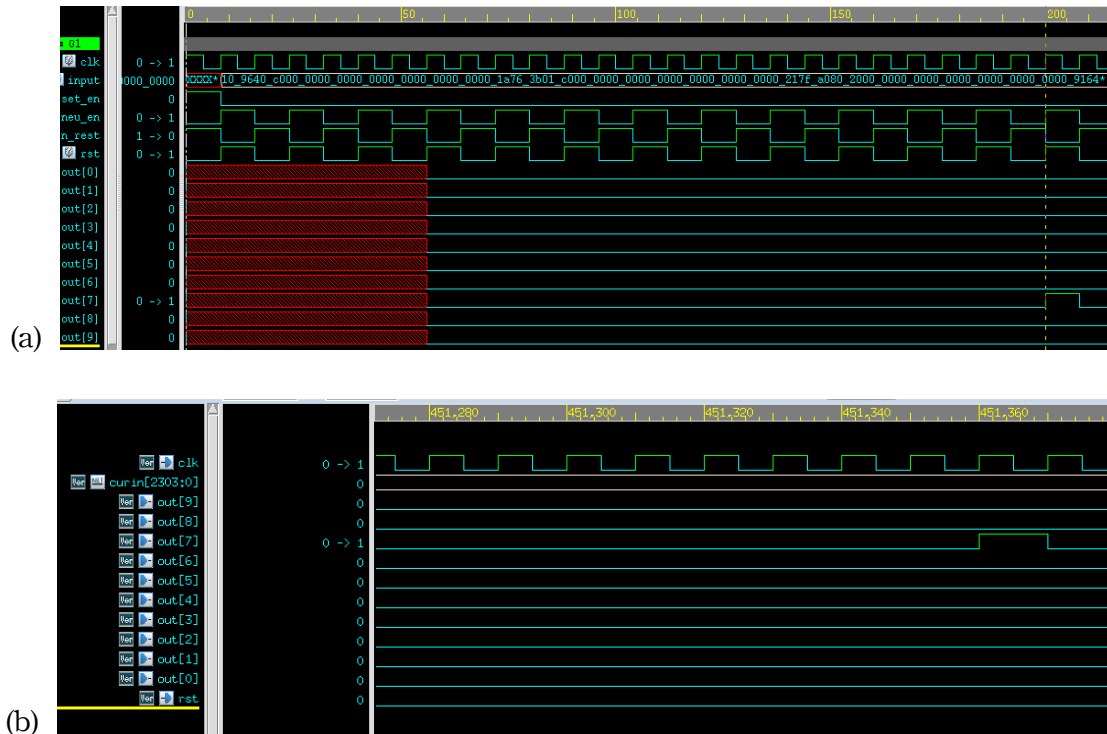


Fig 5. (a) 階層位址事件表示架構產生之結果 (b) 環狀位址事件表示架構產生之結果

根據 Fig 5. (a) 以及 Fig 5. (b)，能夠發現階層位址事件表示所使用的時間明顯較環狀位址事件表示所採用的時間少。因此，對於此次專題的目的，「使用階層位址事件表示」是一個較好的選擇。

進一步地，推測產生兩種結果的原因。首先，對於環狀位址事件表示的架構來說，由於 token 的操作是一次傳送一個神經元的輸出訊號至下一層的一個神經元，並且一次操作所需使用的時間可能是在 10 至 100 多個 clock 的範圍，因此，可想而知，在一個 256-to-128-to-128-to-10 的突波神經網路架構下，處理完所有訊號傳輸所耗時間必定相當長；而對階層位址事件表示的架構而言，因為對一層中所有的神經元所需要的計算量皆是相同的，在設計中經過優化，能夠同時處理較多的資料，不會像是環狀位址事件表示處理一個訊號需要耗費多個時間步，所以總消耗時間一定會比較短。

除此之外，我們還思考了如果是在其他的設計背景下，這兩者之間會有何差別，第一個情況是，若設計架構不屬於 FC network；第二個情況是，若輸入訊號間隔較稀疏。以下將逐一探討：

第一種情況，若設計架構不屬於 FC network。首先，對階層位址事件表示來說，由於設計上是經過優化，較適合處理具有相同計算量的神經元所構成的架構，所以當被套用至不屬於 FC network 的神經網路時，有可能需要針對不同計算量的神經元進行一些微調，才能夠實際應用，如此，設計方法則變得傾向於 full-custom design，這樣就不屬於數位電路設計的特色；而就環狀位址事件表示而言，它是一個通用型的架構，所以可以直接應用在不屬於 FC network 的神經網路上，這在設計的角度上，是一個較友善的架構。因此，綜上所述，若將兩種位址事件表示架構套用至非 FC network，選擇環狀位址事件表示會是一個設計上較友善的選擇。

第二種情況，若是輸入訊號的間隔較稀疏，而此間隔的範圍在環狀位址事件表示架構的 token 一次操作的範圍，或以上。第一，對環狀位址事件表示而言，已知處理每個訊號所需的時間相當大，輸入訊號的時間間隔如果和此相當，那麼最終產生訊號的時間則是會和此次專題設計的結果相當；第二，對階層位址事件表示來說，在稀疏的輸入訊號影響下，會使的處理完一個訊號的時間和環狀位址事件表示處理的時間相近。因此，綜上所述，在輸入訊號的時間間隔較稀疏的情況下，推測兩種位址事件表示架構的表現相似。

## 第五章、總結

專題以兩種位址事件表示架構應用在固定規格之突波神經網路，並對兩者表現進行比較，總結以下幾點結論：

- 階層位址事件表示 (Hierarchical AER) 對於處理全連接、輸入訊號密集的神經網路具有絕對的優勢。
- 環狀位址事件表示 (Token ring AER) 面對相對密集的突波輸入時會顯得效能低落，若輸入訊號較稀疏則會有不錯的表現。
- 環狀位址事件表示 (Token ring AER) 具有通用性，可以直接應用在各種連接情況的神經網路而不需做修正。
- 在大型的神經網路中，如多核心仿生神經網路晶片，由於神經元的數量非常龐大，同一時間必定有不少的輸出突波，因此環狀位址事件表示 (Token ring AER) 不適合用於大規模的神經網路。

在資訊爆炸的時代，我們都希望大量的信息能夠被快速地處理，而神經網路的功能性和神經元數量呈高度相關，可靠的神經網路都具有一定的規模，因此必須依賴一套快速且適應性強的訊息傳遞方式。根據此篇的實作結果，若能在階層位址事件表示（Hierarchical AER）的基礎上進行適應性的改良，將來會是一項有潛力的訊息傳遞技術。

## 第六章、心得感想

經過這一年的專題實作，我們學到了許多東西，在第一學期時，我們以閱讀各種突波神經網路的文獻為主，並且需要針對每一篇文獻在每兩個禮拜一次的會議中，分享自己的想法，我認為這個過程讓我養成了，能夠在閱讀文獻時，同時思考該篇論文所提出電路的優缺點。在這一學期的訓練中，讓我在翻閱文獻時，可以用更快的速度理解該篇文獻所述，更加深了我對文獻的理解。

而第二學期時，我們則是開始使用 verilog，實作使用 AER 架構的 SNN。在實作的過程中，更能夠清楚地看出 implement 的架構的優缺點，這也讓我發現，唯有實際實作過後，才能夠真正的瞭解一樣東西。而經過第二學期後，我也學到了，在建構大型的硬體架構時，需要非常有效率的使用 verilog，且熟悉 verilog 的語法，才能夠有效地達成目標。

在這兩個充實的學期後，雖然實作出的 SNN 不是最完整的，但是也讓我們充分地了解到 SNN 這個新興的神經網路的可能性，以及 AER 對於大型 SNN 的重要性，也希望如果未來有機會，可以繼續深入了解它，並且把在這次實作專題中不夠好的地方，更加的改進。