

國立清華大學 電機工程學系  
實作專題研究成果報告

Spiking Neural Network Processor  
Application in Vision Obstacle Avoidance

神經群仿生神經網路處理器  
於電腦視覺避障之應用

專題領域：系統領域

組 別：B467

指導教授：鄭桂忠 教授

組員姓名：戴佑丞、朱豐蔚、戴光佑

研究期間：113年1月29日至113年11月15日止，共10.5個月

## Abstract

With the widely use of IoT applications in nowadays technological, artificial intelligence has become an indispensable part of our daily lives. However, as the complexity of neural network architectures keeps increasing, the excessive computational demands and power consumption bring significant challenges. In recent years, with the development of systems neuroscience, neural network connection maps can be observed at the cell scale, which enables the realization of spiking neural network(SNN) architectures for biological-like neural behaviors and has become a best-choice for power consumption consideration. In this project, we used the 3rd generation neural network - population-based neuromorphic SNN processor, which uses biological nerve spikes as communication between neurons and reaches a better overall performance in terms of power and accuracy compared with the traditional deep learning neural network architecture.

In this project, we implemented(based on the neural group unit) the FlowDep-SNN processing accelerator designed by cooperated laboratory team. Integer quadratic integrate-and-fire(I-QIF) is used as the neural architecture, which is the basis for realizing the bionic spiking behavior. By observing the changes and quantification of membrane potential, its linear equation is designed for the purpose of reducing computational complexity and required power consumption without losing most of the characteristics of neurons.

Input signals such as velocity, angular velocity, optical flow will be triggered when computer vision detects obstacle avoidance targets. Thus, we designed a 5-layers SNN to construct the connection mechanism of the neural groups and process the input signals to obtain the final obstacle avoidance decision-making result correctly. On the software side, Pytorch is used as learning framework to simulate behavior and generate spike patterns for measurements. On the hardware side, RTL code is used to realize the I-QIF behavior and significantly reduce the hardware area by the repeated use of neuron hardware. Finally, the hardware architecture was implemented and export as the PYNQ-Z1 FPGA IP to complete the FlowDep-SNN processing accelerator for obstacle avoidance, which is featured with its 0.1 msec/per decision response time, 92% accuracy, low hardware area and low power consumption.

## 摘要

在現今的科技發展趨勢中，物聯網的應用越來越普及，人工智慧已成為其中不可或缺的一部分，因此機器學習與深度學習的運用充斥在我們的生活中。然而，隨著想要解決的問題越龐大，所需要使用的神經模型越趨複雜，過量的運算與耗能也成為了必須面對解決的棘手問題。近年來，隨著系統神經科學的發展，可觀察至以細胞尺度為單位的神經網路連結圖譜，這使得類生物神經行為的脈衝神經網路架構得以實現，並成為了在功耗表現上的最佳選擇。本研究採用第三代神經網路—SNN 神經群仿生神經網路，其使用類生物神經脈衝作為神經元之間的溝通，比起傳統的深度學習神經網路架構有功率及準確率綜合表現上的顯著優勢。

在研究中，我們以神經群為單位，實作出實驗室合作團隊所設計、訓練的 FlowDep-SNN 神經網路處理加速器。在神經元的設計上，採用二次方程累積放電脈衝(I-QIF)為基本的架構，是實現仿生架構的基礎，並透過觀察膜電位的變化與量化，設計其線性方程式以模擬神經元突觸的基本行為。膜電位的量化及線性方程式化的目的是在不失去神經元的大部分特性下，去降低運算的複雜度與所需的功耗。

電腦視覺偵測到避障目標會有速度、角速度、光流等輸入訊號，本研究採用一共五層的神經網路架構來構建神經群的連結機制，使其能夠正確處理輸入以得到最終的避障決策結果。在學習框架上使用 Pytorch 來模擬神經網路行為並生成神經網路的脈衝測資，硬體端則以 RTL code 實現神經元膜電位的累積、漏電與觸發，並透過重複使用神經元硬體以大幅減少硬體面積，最終實作輸出為 PYNQ-Z1 FPGA IP，完成反應速度0.1 msec/per decision、避障準確率92%、低硬體面積、低功耗的避障 FlowDep-SNN 神經網路處理加速器。

# 1. Introduction

## 1-1. Research Background and Motivation

本研究將重心放在邊緣裝置的微型自動車輛避障問題上，這類微型車輛在近年來的應用領域持續加深加廣[1]，由於較低的成本以及較小的體積，微型自動車輛能夠在有限的道路空間做到運輸、救難、探測等工作。然而，對於這類體積不大的微型自動化裝置，受限於其成本和物理尺寸大小上的限制，傳統的電腦視覺感測方式如雷射、雷達對於深度的判斷會有所失準、缺失[2]。此外，以往基於深度學習的人工智慧算法雖然準確，但對於這類邊緣化的裝置在功耗、算力上是非常龐大的負擔。

為了解決這個問題，一種受自然界昆蟲判斷動態環境所啟發的新型演算法(FlowDep)被設計了出來[2]，其透過低解析度但高效的光流訊號偵測機制，結合了IMU(定位資訊)來估計每個影像像素的深度，使其適合在微型邊緣化裝置上有所應用。FlowDep 以及其在人工神經網路的變體(FlowDep-ANN)能在準確率一樣的情況下達到比傳統 DNN 神經網路快5.8到114.7倍的效果。

綜上所述，由於 FlowDep 的光流算法加上日漸成熟的 ANN 神經網路發展，運算速度與準確率在微型自動車輛上已經能有一定的成效與規模，但其功耗對於微型裝置這類仰賴充電電池等邊緣化裝置仍然為一大硬傷。

近年來，隨著系統神經科學的發展，可觀察至以細胞尺度為單位的神經網路連結圖譜，這使得類生物神經行為的第三代神經網路—SNN 脈衝神經網路架構得以實現，並成為了在功耗表現上的最佳選擇[3][4][5]。其使用類生物神經脈衝作為神經元之間的溝通，比起傳統的深度學習神經網路架構有功率及準確率綜合表現上的顯著優勢。

有基於此，國立清華大學鄭桂忠教授的實驗團隊近期選擇以 SNN 去重新實作 FlowDep-ANN 的架構，以此去改善在這類微型裝置避障功能上功耗的難解問題。本研究參與計畫其中，負責實作已經前端訓練好的神經網路加速處理器，並嘗試在實作過程中透過改善設計細節去發揮 SNN 在計算面積、功耗的雙重優勢，為自動化避障車輛提供神經網路選擇上的全新解方。

## 1-2. Research Process

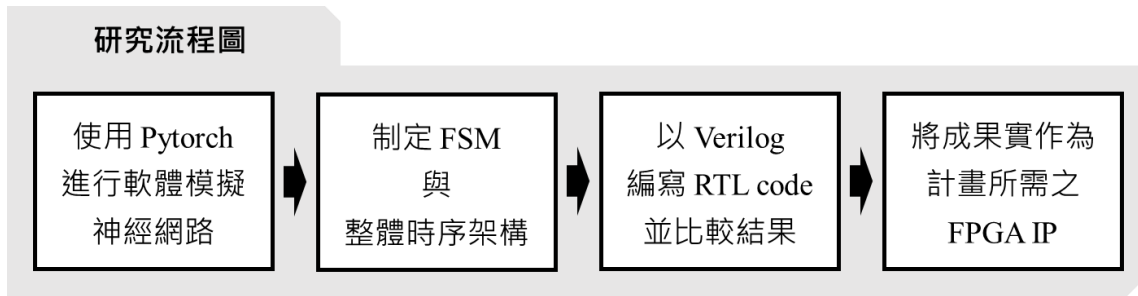


Fig. 1 研究流程圖

本研究架構如 Fig. 1 所示，目的是接續實驗室計畫的前端演算法設計，將其實作為高效能表現的 FPGA IP。第一步我們先使用學習框架的軟體 Pytorch 進行整個神經網路的行為模擬，將輸入速度、角速度、光流訊號進行初步軟體端的處理，並得到正確的目標輸出結果(以供第三步中去比較驗證 Verilog 硬體端的輸出結果)。

在第二步中，最核心的目的就是透過安排 SRAM 溝通時序的 pipeline、增減硬體單位數量，並壓低運算所需的 cycle 數，使神經群仿生神經網路處理器能達到邊緣化自動裝置在高反應速度、低計算面積、低功率三個指標上的綜合要求。

最後以 Verilog 實現整個 SNN 神經網路架構，並將成果實作輸出為 PYNQ-Z1 FPGA IP，提供並完善實驗室計畫，延續其前端的算法設計與後端的反應實測，完成電腦視覺避障決策之判斷。

## 2. Research Methodology

### 2-1. FlowDep-SNN Layer Architecture

實驗室合作團隊參考 FlowDep-ANN 的架構[2]，在維持其避障決策準確率 92% 的情況下，將其調整為適合 SNN 的神經網路架構，整個 FlowDep-SNN 的神經網路架構一共有五層：Rotation、De-Rotate、Scalar、Velocity、Depth。

參考自然界小體積昆蟲對目標的感測[2]，光流訊號與經 Rotation layer 初步處理的角速度訊號累加後，會一起經過 De-Rotate layer 對轉向、角度做消弭，透過神經元 4 to 1 的 Scalar layer 增加權重影響後，搭配速度訊號來幫助系統了解避障目標的所處深度，再進一步透過深度判斷的次數累加來做最後是否躲避的決策，以上即形成了整個神經網路的處理邏輯。

最後的 spike\_cnt 輸出為 Depth 這層神經網路中每個神經元在 100 個時間步 (decision time window) 累積下來的脈衝總數 (Depth spike 的次數)，此即為電腦視覺避障的決策結果，使用 decision time window 的設計也能夠讓硬體省下時時刻刻檢測決策的功耗，僅在每 100 個時間步進行決策即可。

## 2-2. System Design

### 2-2-1. SNN Accelerator Hardware Architecture

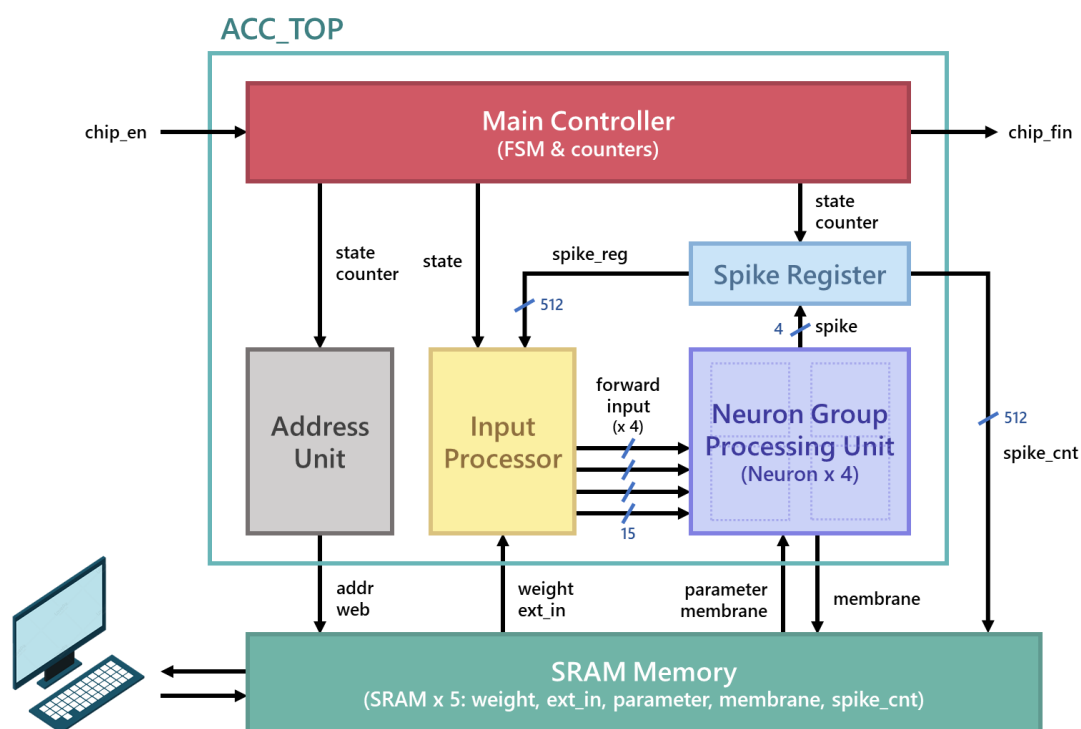


Fig. 2 SNN 神經群仿生神經網路處理器架構圖

本研究之神經群仿生神經網路處理器整體架構如 Fig. 3 所示，在加速器左方代表輸入包含起始訊號以及速度、角速度、光流等預先放入 SRAM 的輸入訊號，在加速器右方代表輸出包含結束訊號以及會寫入 SRAM 的脈衝數量，而加速器內外主要可以分為五大區塊：

1. 主控制器(Main Controller)
2. 輸入訊號處理(Input Processor)
3. 神經元組(Neuron Group)
4. 地址計算模組(Address Unit)
5. 記憶體(SRAM Memory)

## 2-2-2. Neuron Group – I-QIF Model

本研究所設計之神經元組為四個神經元構成一組，神經元所採用的模型為 Integer Quadratic Integrate-and-Fire neuron(簡稱 I-QIF)，是以傳統使用的 Quadratic Integrate-and-Fire neuron(簡稱 QIF)改良後的模型，利用整數的參數去近似原本是二次方程式的膜電位變化，達到簡化運算以及提高效率的效果 [6][7]。原傳統 QIF 的膜電位變化如下(式(1))：

$$C \frac{dV}{dt} = f(V) = K(V_r - V)(V - V_t) + I(t) \quad (1)$$

可看出上方式子中  $f(V)$  為  $V$ (神經元膜電位)的二次方程式， $V_r$ 代表 $V_{reset}$ 、 $V_t$ 代表 $V_{threshold}$ ， $I(t)$ 代表輸入值。簡化過後的 I-QIF neuron 模型膜電位變化如下(式(2))：

$$\frac{\Delta V}{\Delta t} = f(V) = \begin{cases} a(V_r - V) + I(t), & V < \frac{bV_t + aV_r}{a + b} \\ b(V - V_t) + I(t), & V \geq \frac{bV_t + aV_r}{a + b} \end{cases} \quad (2)$$

$$V \rightarrow V_{reset} \text{ when } V > V_{max}$$

在式(2)中， $V_r$ 代表 $V_{reset}$ 、 $V_t$ 代表 $V_{threshold}$ ， $a$ 和 $b$ 分別為整數之參數， $I(t)$ 代表輸入值，在進行累加時，會先將該神經元膜電位與  $\frac{bV_t + aV_r}{a + b}$  做比較，若小於此值將模擬上方  $a$  參數之累加，反之則是下方  $b$  參數之累加，當膜電位值大於 $V_{max}$ ，神經元便會發射出一個脈衝(spike)並且同時將該神經元的膜電位還原成 $V_{reset}$ ，接著再繼續做下一次的累加與發射。

### 2-2-3. Main Controller

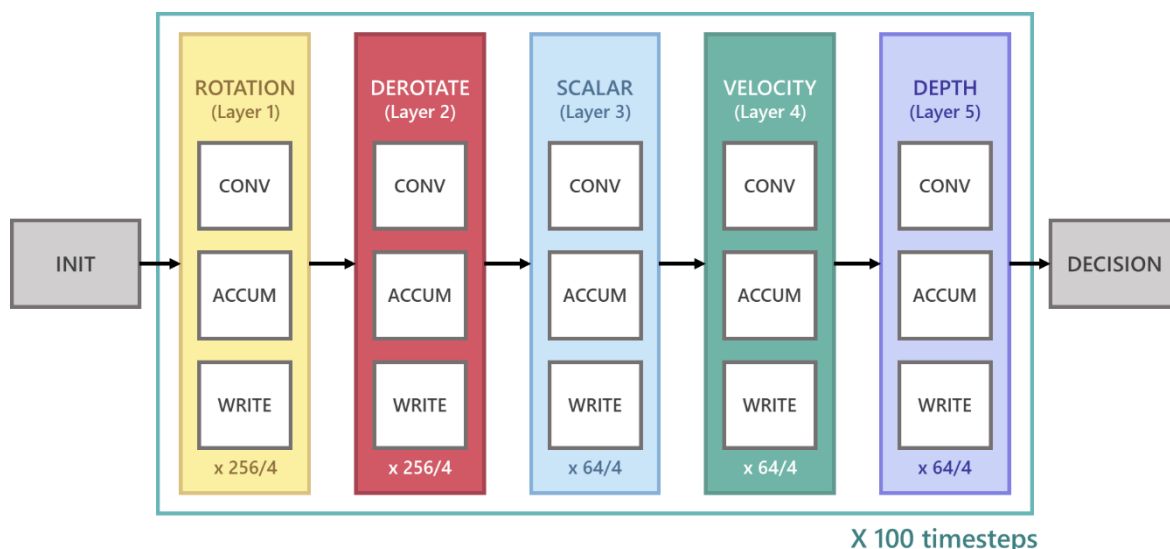


Fig. 3 FSM 架構圖

在五層神經網路分別在做計算時，每層神經網路皆是以 CONV、ACCUM、WRITE 三個狀態做循環 pipeline 運作直到該層神經網路之所有神經元皆累積放電完成一次，CONV 是進行前處理的運算，由輸入訊號處理器(Input Processor)負責；ACCUM 是神經元模擬累加與發射過程，由神經元組(Neuron Group)模組負責；最後 WRITE 是將神經元產生的脈衝存入暫存器 spike register 中儲存下來，同時也把更新後的神經元資料存回 SRAM 中。三個子狀態的切換充分利用了與總共 5 個 SRAM 的溝通時間，在保證功能正確的情況下，能夠減少每個時間步所需要的 cycle 數量，進而增加系統在避障上的反應速度。

五層神經網路的計算順序為 Rotation、De-Rotate、Scalar、Velocity、Depth，做完一層的累加與發射才進行到下一層，以確保整體神經網路的連結邏輯順序正確。

### 2-2-4. Input Processor

此區塊是直接接收 input 端資料的模組，全部運算也都是以四個神經元為一組進行，功用為把該層神經網路收到的輸入做前處理，輸出即可以傳送給下一區直接進行累積放電脈衝的運算。

在 Rotation 和 Velocity 兩層神經網路中不需要做前處理(已由軟體處理完成)，所以分別將處理完成的輸入訊號(速度與角速度)直接傳給後方的 Neuron Group 單元。而剩下的三層神經網路 De-Rotate、Scalar、Depth 皆會在此步驟先與權重進行二維的卷積運算才會進入到後續神經群累積、放電的過程。

### 3. Experimental Results

在時間表現上，總共五層的 SNN 神經網路架構能夠實現每 54320 cycles/per decision 的表現，實作於頻率 525 MHz 的 PYNQ-Z1 FPGA 開發板上後，即可以得到 0.1 msec/per decision 的表現(ISO 22737 低速自動駕駛車輛系統-系統要求、性能要求和性能測試規範之建議反應時間為 100-200 msec/per decision)，這意味實作於微型自動車輛裝置上能有快速且足夠的避障反應時間，能夠勝任各項微型裝置任務。

此外，與論文中 FlowDep 的總避障反應時間(含 sensor 感測輸入時間與做出相對應反應的時間)35 msec/per decision 和 FlowDep-ANN 的總避障反應時間 27 msec/per decision 做比較[2]，本研究的 FlowDep-SNN 在處理輸入做決策的加速器上能達到不亞於 FlowDep 和 FlowDep-ANN 的速度表現，展現了 FlowDep-SNN 實作自動化避障裝置的可行性。

而在避障準確率的結果上，由實驗室合作團隊所研擬建置的 FlowDep-SNN 能夠維持住同等 FlowDep 和 FlowDep-ANN 在避障準確率上的表現[2]，達到約莫 92%的避障準確率。結合上述，本研究在避障反應時間與準確率上的表現皆表明 FlowDep-SNN 能夠被使用於微型自動化裝置上，發揮其在面積與速度表現上的雙重優勢。

在硬體面積的表現上，在本研究中採用神經群且重複的硬體使用邏輯，使得整套神經網路的運作下來，只需 4 組神經元的硬體面積就能完成所有的脈衝決策計算。此外，SNN 相較於 ANN 的優勢就是在於神經層之間傳遞的訊號為 0、1 的 1-bit 神經脈衝訊號，這使得相較於大多數 ANN 等 DNN 中使用的 8-bit data 有較小的硬體乘加法計算與存放面積，在面積的表現上會優於 ANN 架構。

在功耗的表現上，FlowDep-SNN 能在維持避障準確率的情況下達到更小的功耗，這也是因為脈衝訊號在乘加法計算、傳遞上的低硬體要求，加上 SNN 的 event-driven 特性，即 SNN 只有在有收到脈衝輸入才會進行運算，這比起每時每刻都在計算的傳統 DNN 架構，有在功耗上的顯著優勢。

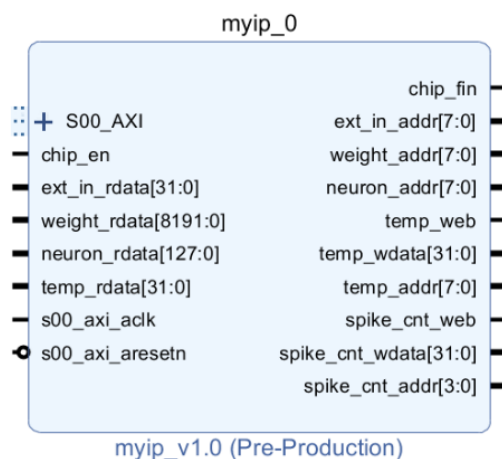


Fig. 4 輸出 FPGA IP 供實驗室合作團隊使用

## 4. Conclusion

本研究成功以硬體實現了基於神經群仿生脈衝神經網路(SNN)的 FlowDep-SNN 架構，針對微型邊緣化裝置的電腦視覺避障問題提供了一個具創新性且高效的解決方案，解決了傳統電腦視覺算法和深度學習算法在微型裝置上因算力、功耗和硬體面積限制所面臨的挑戰，並將脈衝神經網路的低功耗和面積優勢引入 FlowDep 算法。

實驗結果顯示，本研究之 FlowDep-SNN 處理加速器在以下幾個方面具有顯著優勢：

### 1. 快速的避障反應時間：

FlowDep-SNN 的決策反應時間(decision time window)為 0.1 msec/per decision，這表明其在實際應用中能提供足夠快速且安全的反應能力，適用於微型自動化裝置的多樣化任務需求。

### 2. 準確率穩定表現：

由實驗室合作團隊訓練完成的 FlowDep-SNN 神經網路算法能夠維持與 FlowDep 和 FlowDep-ANN 相當的 92% 避障準確率，展現了實用性與可靠性的平衡。

### 3. 硬體面積優勢：

本研究採用神經群重複使用神經元硬體的架構設計，使整體運算僅需少量硬體資源，加上神經脈衝的 1-bit 脈衝訊號特性，顯著降低了硬體乘法計算與存儲需求，充分滿足微型裝置對面積緊縮的要求。

### 4. 顯著的低功耗表現：

FlowDep-SNN 因 SNN 事件驅動(event-driven)的計算特性，僅在脈衝訊號出現時啟動運算、僅在 decision time window 時做出決策，因此降低了許多不必要的功耗，並在長時間運作中展現出顯著優於傳統 DNN 的能源使用效率。

綜上所述，FlowDep-SNN 結合了 SNN 的低功耗、低面積與快速運算優勢，為資源受限的微型邊緣化裝置在電腦視覺避障問題上提供了一個切實可行且高效的解決方案。未來，本研究可進一步探索 SNN 在多感測融合、動態場景預測以及系統整合中的潛力，以促進自動化邊緣裝置的多元應用發展。

## 5. Reference

- [1] Christensen, H., Paz, D., Zhang, H. et al. Autonomous vehicles for micro-mobility. *Auton. Intell. Syst.* 1, 11 (2021).
- [2] Chen-Fu Yeh, Chao-Yang Tang, Tsu-Chiao Chen, et al. FlowDep - An efficient and optical-flow-based algorithm of obstacle detection for autonomous mini-vehicles. *TechRxiv*. April 03, 2024.
- [3] Nguyen, Duy-Anh, Xuan-Tu Tran, and Francesca Iacopi. 2021. "A Review of Algorithms and Hardware Implementations for Spiking Neural Networks" *Journal of Low Power Electronics and Applications* 11, no. 2: 23.
- [4] C. Frenkel, M. Lefebvre, J. -D. Legat and D. Bol, "A 0.086-mm<sup>2</sup> 12.7-pJ/SOP 64k-Synapse 256-Neuron Online-Learning Digital Spiking Neuromorphic Processor in 28-nm CMOS," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 1, pp. 145-158, Feb. 2019.
- [5] Qiao, Chunming, Shaogang Hu, Junjie Wang, C Zhang, Tupei Chen, Ning Ning, Qi Yu and Yang Liu. "A Neuromorphic-Hardware Oriented Bio-Plausible Online-Learning Spiking Neural Network Model." *IEEE Access* 7 (2019): 71730-71740.
- [6] W. -C. Wu et al., "Integer Quadratic Integrate-and-Fire (IQIF): A Neuron Model for Digital Neuromorphic Systems," *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, Washington DC, DC, USA, 2021, pp. 1-4.
- [7] Z. -W. Yeh et al., "POPPINS: A Population-Based Digital Spiking Neuromorphic Processor with Integer Quadratic Integrate-and-Fire Neurons," *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, Daegu, Korea, 2021, pp. 1-5.

## 6. Review and Reflection

在專題之前，我們對神經網路的了解都是比較卷積、數學的，但 SNN 完全打破了我們的認知，我們都未曾想過可以用 0、1 的訊號去模擬、甚至做出如此生物神經性質的脈衝、膜電位行為電路。透過了解了 SNN，我們也是認識了傳統計算機的全新應用，並會開始思考電路的更多元可能性，往更加生物、仿真的未來邁進。

專題研究不只是帶來新知識，更重要的是在過程之中，我們養成了面對問題、拆解問題、解決問題這樣的研究精神，並從每週的會議報告養成規律、團結、謹慎的習慣，這些都是為我們未來的研究打下深厚基礎的重要養分。

最後要感謝鄭桂忠教授以及邱至鵬助教、胡育嘉助教、黃瑋琪助教這一年的指導，提供我們軟、硬體上的豐富資源，幫助我們一步一步從認識 SNN、構建硬體，到最後做出能夠應用於電腦視覺避障的 SNN 神經網路處理加速器。