



Dyna-Q演算法之硬體加速器晶片設計與實踐

Hardware Accelerator Chip Design and Implementation of the Dyna-Q Algorithm

組別：晶片系統組

指導教授：朱元三 教授

學生：游竣喆、黃士嚴

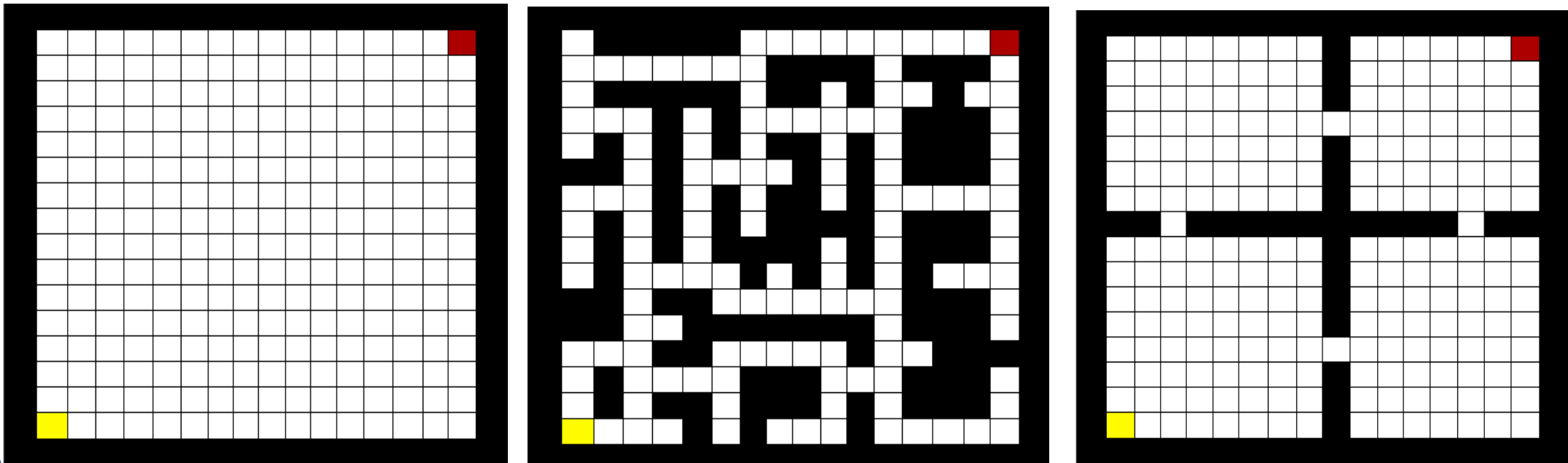
摘要

增強式學習的應用愈加廣泛，自動駕駛、無人機與機器人控制大量普及，其中路徑規劃成為關鍵技術之一，透過演算法讓系統能自主選擇最佳路徑。然而，增強式學習的過程在純軟體實現上耗時高，難以滿足即時的實際應用。

本專題採用Q-learning演算法，並加入 Dyna-Q 和經驗模型以縮短學習時間，在未知迷宮中能夠找出最佳路徑。

實驗地圖

本專題以不包含外牆的16*16格環境進行實驗模擬，使用以下三種不同環境，分別為空白地圖、多死路地圖及四個房間地圖，紀錄到達最佳路徑 3次和 5次。白色為可行走，黑色為障礙物，左下為起點，右上為終點。

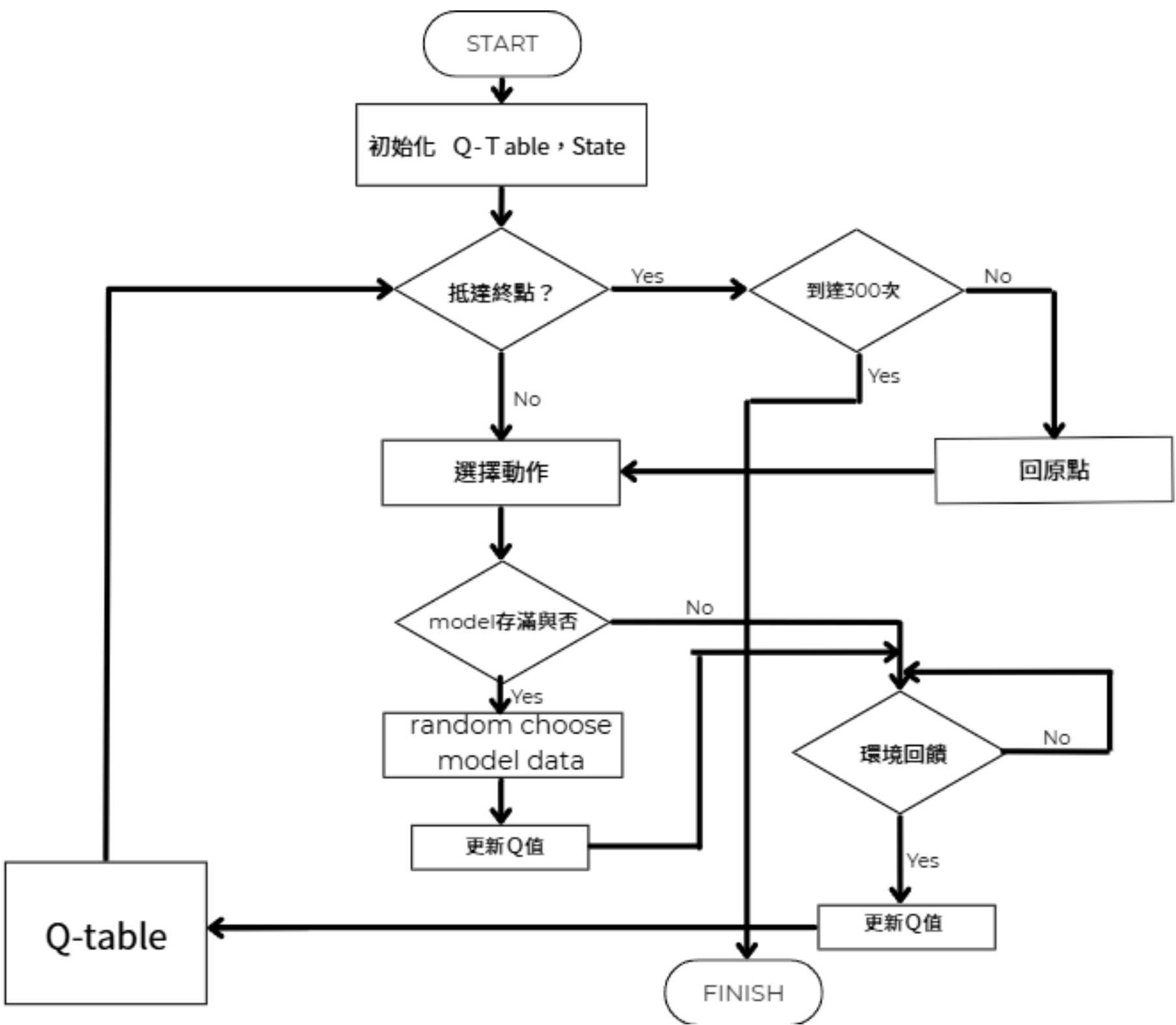


■ 空白地圖

■ 多死路迷宮

■ 四個房間地圖

軟體流程圖



實驗結果

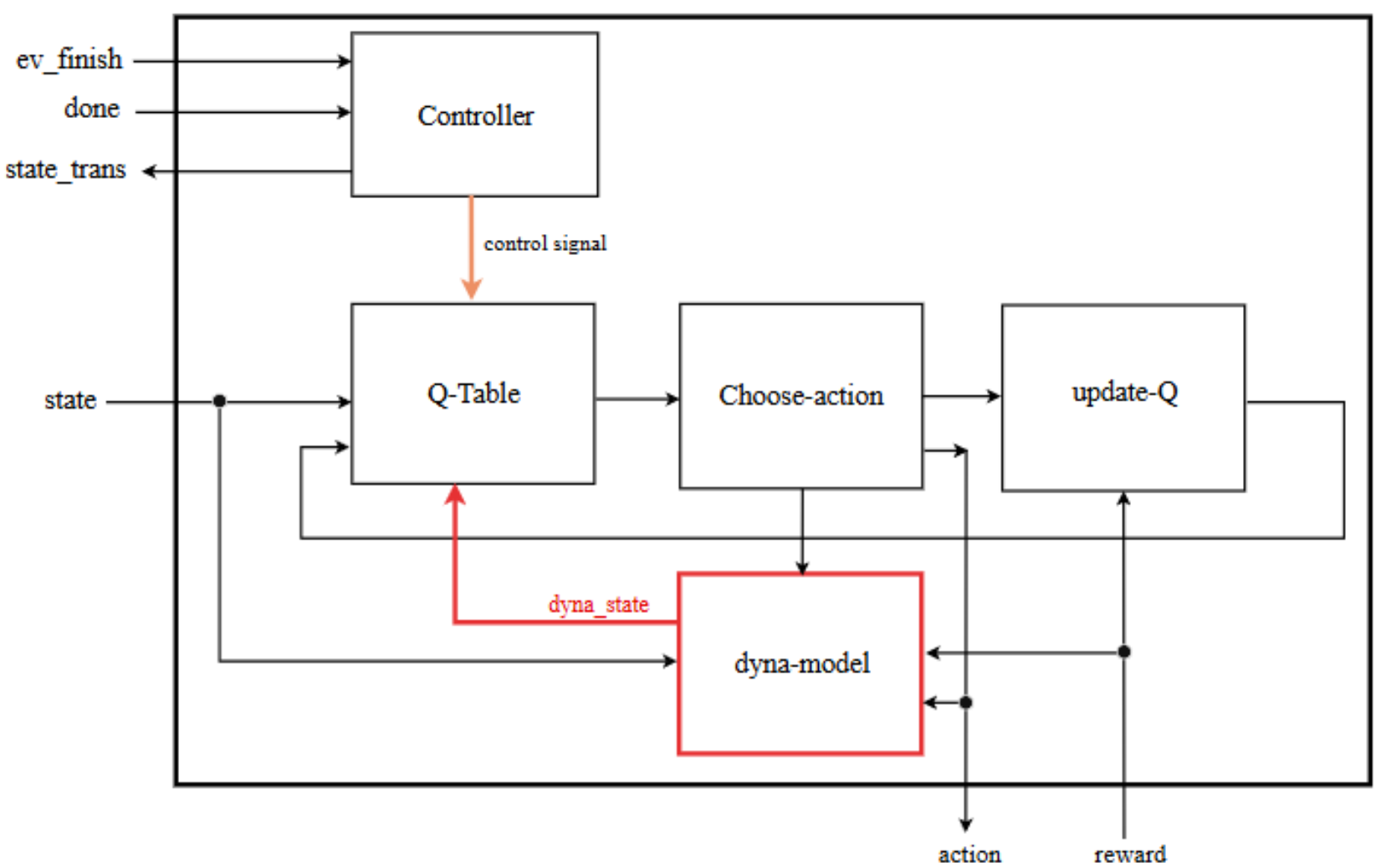
以下為成功到達到達最佳路徑 3次和 5次的硬體模擬時間和加速效果比較。
加上Dyna-Q經驗模型對比純Q-learning僅需多花1.4%的power，就有以下的加速效果。

地圖種類	空白地圖	多死路迷宮	四個房間地圖	地圖種類	空白地圖	多死路迷宮	四個房間地圖
到達最佳路徑時間				到達最佳路徑時間			
純Q-learning	0.3241ms	0.1854ms	0.3088ms	純Q-learning	0.3279ms	0.1884ms	0.3118ms
硬體加速時間				硬體加速時間			
Q-learning + Dyna-Q	0.1397ms	0.1267ms	0.1607ms	Q-learning + Dyna-Q	0.142ms	0.1289ms	0.1671ms
硬體加速時間				硬體加速時間			
純Q-learning 對比加上 Dyna-Q	56.9%	31.7%	47.9%	純Q-learning 對比加上 Dyna-Q	56.7%	31.6%	46.4%
硬體加速效果				硬體加速效果			

■ 訓練到達到達最佳路徑三次

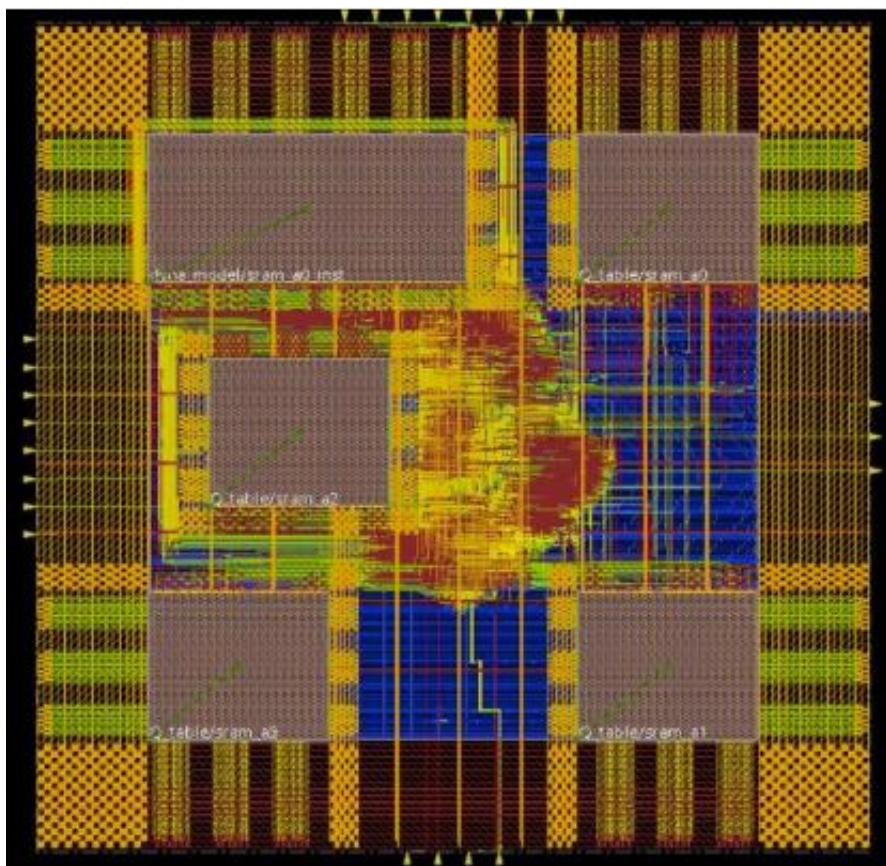
■ 訓練到達到達最佳路徑五次

硬體架構圖



實作結果

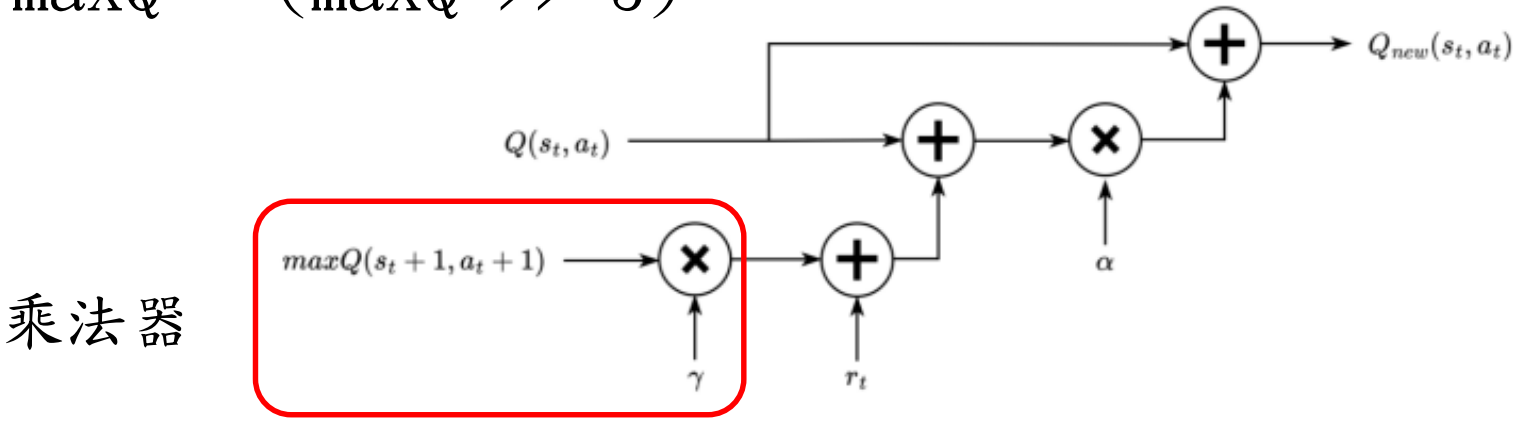
Technology	TSRI_U18
Frequency	156.25 MHz
Core Area	1449790.15 μm^2
Power	52.865 mW
Gate Count	46564



硬體優化

本專題利用減法取代乘法器中複數個加法器移位。

$$\begin{aligned} \gamma \text{ 值若為 } 0.875(10) &= 0.111(2) \\ \gamma \cdot \max Q &= \max Q - (\max Q \cdot 0.001(2)) \\ &= \max Q - (\max Q \gg 3) \end{aligned}$$



■ 更新Q值架構圖

參考文獻

- [1] Algorithm Optimization and Hardware Acceleration for Machine Learning Applications on Low-energy Systems, Jianchi Sun
- [2] 許博凱, “增強式學習系統加速器之晶片設計”, 2020
- [3] Towards Hardware Accelerated Reinforcement Learning for Application-Specific Robotic Control, Shengjia Shao; Jason Tsai; Michal Mysior; Wayne Luk; Thomas Chau; Alexander Warren
- [4] 曾彥智, 應用於3D路徑探索之增強式學習晶片設計, 2024
- [5] Data-driven inventory management for new products: A warm-start and adjusted Dyna-Q approach, Xinye Qiu, Longxiao Liu and Wenjie Huang
- [6] An efficient hardware implementation of reinforcement learning The q-learning algorithm, Spanò, Sergio; Cardarilli, Gian Carlo; Di Nunzio, Luca; Faz zolari, Rocco; Giardino, Daniele; Matta, Marco; Nannarelli, Alberto; Re, Marco