电 子 科 技 大 学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

专业学位硕士学位论文

MASTER THESIS FOR PROFESSIONAL DEGREE

论文题目 **基于连续吸引子神经网络的神经形态电路系统研究**

专业学位类别	工程硕士
学　　号	**201922140329**
作者姓名	赵琨鹏
指导教师	游宏志　　副教授
学　　院	生命科学与技术学院

分类号 _____ 密级 公开

UDC [注1] _____

学 位 论 文

基于连续吸引子神经网络的神经形态电路系统研究

（题名和副题名）

赵琨鹏

（作者姓名）

指导教师 游宏志 副教授

电子科技大学 成 都

申请学位级别 硕士 专业学位类别 工程硕士

专业学位领域 生物医学工程

提交论文日期 2022 年 3 月 20 日 论文答辩日期 2022 年 5 月 20 日

学位授予单位和日期 电子科技大学 2022 年 6 月

答辩委员会主席 郭大庆 教授

评阅人 _____

Neuromorphic Circuit System Based on Continuous Attractor Neural Network

A Master Thesis Submitted to
University of Electronic Science and Technology of China

Discipline **Biomedical Engineering**

Student ID **201922140329**

Author **Zhao Kunpeng**

Supervisor **A.P. You Hongzhi**

School **School of Life Science and Technology**

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

作者签名：_____ 日期：　　年　　月　　日

论文使用授权

本学位论文作者完全了解电子科技大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权电子科技大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后应遵守此规定）

作者签名：_____ 导师签名：_____

日期：　　年　　月　　日

摘 要

电子信息技术从上个世纪中叶开始迅猛发展，时至今日，传统体系架构下的电子计算机技术已经十分成熟。但若是将人类大脑视为一个信息处理系统，并将其与当前的电子计算机做比较时就会发现，人脑的各种认知功能与复杂思维能力依然是计算机无法轻易实现的。为了实现更加强大、更加高效的人工智能，一个很自然的想法便是从对于生物神经系统的相关研究中获取新的信息处理系统的设计灵感。神经形态电路便是这个思想的直接产物，该领域的目标是实现功能近似于生物神经系统的电路。另一方面，计算神经科学利用数学模型对生物神经系统进行定量的分析研究，其研究成果对于指导设计更加接近生物脑的神经形态电路系统具有重要意义。

在此背景下，本论文尝试以计算神经科学中提出的连续吸引子神经网络（Continuous Attractor Neural Networks, CANN）模型为理论基础，以现场可编程逻辑门阵列（Field Programmable Gate Array，FPGA）作为硬件平台，设计并实现了一组具有丰富的仿生物神经特性的神经形态电路系统。

论文按照研究的进展阶段，首先介绍了初步设计的具有 64 个神经元的 CANN 电路原型以及基于此电路进行的知觉决策和工作记忆仿真实验；然后介绍了规模更大且设计较为成熟的具有 512 个神经元的 CANN 电路系统以及基于此电路进行的静默式工作记忆和 T 型迷宫仿真实验；最后介绍了基于片上网络搭建的具有多个 CANN 核心的电路系统以及对该电路基本功能进行的测试结果。

本文中提出的神经形态电路系统具有对较为多样的神经机制进行模拟的功能，例如演化时间常数较大的非线性突触、突触的短时程可塑性以及神经元的发放频率适应性等神经机制，这使得该电路系统具有了丰富的计算功能。在各个阶段的电路设计完成后进行的仿真实验验证了电路的计算功能。该系列神经形态电路系统将有助于研究者对 CANN 模型进行进一步的研究，也将有助于把 CANN 模型相关的计算特性实际应用于潜在的应用领域。

关键词：神经形态电路，连续吸引子神经网络，现场可编程逻辑门阵列

ABSTRACT

Electronic information technology has developed rapidly since the middle of last century. Up to now, the electronic computer technology under the traditional architecture has been very mature. However, if the human brain is regarded as an information processing system and compared with the current electronic computer, it will be found that various cognitive functions and complex thinking abilities of the human brain are still not easily for the computer to realize. Therefore, in order to realize a more powerful and efficient artificial intelligence, a natural idea is to obtain the design inspiration of new information processing system from the research related to biological neural system. Neuromorphic circuit is the direct product of this idea. The goal of this field is to realize circuits whose function is similar to biological neural system. On the other hand, computational neuroscience uses mathematical models to conduct quantitative analysis and research on biological neural system, its research results are of great significance to guide the design of neuromorphic circuit system which is closer to biological brain.

In this context, based on the continuous attractor neural networks (CANN) model proposed in computational neuroscience, and taking field programmable gate array (FPGA) as the hardware platform, the thesis demonstrates an attempt of designing and implementing a series of neuromorphic circuit systems with rich biomimetic neural characteristics.

According to the research progress stage, the thesis first introduces the preliminarily designed CANN circuit prototype with 64 neurons and the perceptual decision-making and working memory simulation experiments based on this circuit; then it introduces a larger and more mature CANN circuit system with 512 neurons, as well as the silent working memory and T-maze simulation experiments based on this circuit; finally, the circuit system with multiple CANN cores based on network on chip and the test results of the basic functions of the circuit are introduced.

The circuit systems proposed in this thesis have the functions of simulating diverse neural mechanisms, such as nonlinear synapse with large evolution time constant, short-term plasticity of synapse and spike frequency adaptation of neurons, which make the circuit systems have rich computing functions. Those computational functions of the circuit are verified by simulation experiments after the circuit design in each stage is

completed. The series of neuromorphic circuit systems proposed in this thesis will not only be beneficial to researchers to do further study of CANN model, but also be beneficial to apply the relevant computational characteristics of the CANN model to potential application fields.

Keywords: Neuromorphic Circuit，CANN，FPGA

目 录

图目录

缩略词表

英文缩写	英文全称	中文全称
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazole-propionicacid	α-氨基-3-羟基-5-甲基-4-异恶唑丙酸
ANN	Artificial Neural Network	人工神经网络
BRAM	Block Random Access Memory	分块随机存取存储器
CANN	Continuous Attractor Neural Networks	连续吸引子神经网络
FEF	Frontal Eye Fields	额叶视野
FIFO	First Input First Output	先进先出储存器
FPGA	Field Programmable Gate Array	现场可编程逻辑门阵列
GABA	γ－Aminobutyric Acid	γ-氨基丁酸
LFSR	Linear Feedback Shift Register	线性反馈移位寄存器
LIF	Leaky Integrate-and-Fire	泄露的整合发放
LIP	Lateral Intraparietal Cortex	外侧顶叶内皮层
MT	Middle Temporal Area	中颞区域
NMDA	N-Methyl-D-Aspartic Acid	N-甲基-D-天冬氨酸
NoC	Network-on-Chip	片上网络
PFC	Prefrontal Cortex	前额叶皮质
RAM	Random Access Memory	随机存取存储器
SC	Superior Colliculus	上丘区域
SFA	Spike Frequency Adaptation	发放频率适应性
STD	Short-Term Depression	短时程抑制
STDP	Spike-Timing-Dependent Plasticity	发放时间依赖可塑性
STF	Short-Term Facilitation	短时程易化
STP	Short-Term Plasticity	短时程可塑性

英文缩写	英文全称	中文全称
VLSI	Very Large Scale Integration	超大规模集成电路
WTA	Winner-Take-All	赢者通吃机制

第一章 绪论

1.1 课题的研究背景及意义

生物神经系统，例如人的大脑，具有极其复杂的结构以及十分强大的功能，同时也具有极高的能量效率。人的大脑拥有约 850 亿个神经元，这些神经元又通过万亿个突触互相交流，然而这样一个极其复杂的系统的功率只有 20 瓦左右[1]。这意味着，通过研究神经系统，将有可能获得具有极高效率的计算系统的计算原理。

于是，对于神经形态电路系统的研究应运而生。神经形态电路系统是指模拟真实的生物神经系统的组织方式与功能的电路系统，通过采用模拟、数字或模数混合的超大规模集成电路（Very Large Scale Integration, VLSI）技术，并配合相关的软件来实现[2]。

神经形态电路系统具有广阔的应用前景：首先，它可以帮助研究人员更高效地探究神经系统的工作原理，具体来说，该类电路系统可以实时地模拟具有一定生物真实性的神经系统模型的运作，同时也可以方便地获取到系统运行结果的数据，这样的功能可以在一定程度上促进神经科学的研究。这是该技术在科学探索方面的应用前景。其次，该类系统也可以将从神经科学中获得的新知识转化为实际的电路系统，从而高能效地完成与认知相关的任务，这是它在工程实践方面的应用前景[3]。

另一方面，CANN 作为一种描述具有空间平移不变性的神经网络结构的数学模型，因其所具有的优良计算特性，已经被成功地运用于描述神经系统中简单连续特征的编码（如头朝向信息的编码）。若在基本的 CANN 的基础上，进一步在模型中加入一些对突触以及神经元所具有的独特生物学特性（如非线性突触，突触的短时程可塑性，神经元的发放频率适应性等）的模拟功能，则该类 CANN 模型就可以实现很多种类的认知任务，如基于赢者通吃（winner-take-all，WTA）机制的知觉决策，或者如文献[4]中提出的，利用 CANN 以及短时程可塑性对于工作记忆的建模的工作。

以 CANN 模型为理论基础设计的神经形态电路系统，将具有之前提到的（也就是神经形态电路系统所具有的）两个方面的应用前景：一方面，这样的电路可以高效地对 CANN 进行仿真，所以搭建这样的硬件平台将有利于促进对于 CANN 的进一步探索；另一方面，这样的电路可以将 CANN 已知的认知计算功能实时复现出来，而这样的计算功能具有现实的应用潜力，例如在清华大学的 Tianjic 芯片上实现的 CANN 被用于自动自行车的目标追踪[5]。

1.2 国内外研究现状

1.2.1 神经形态电路的代表性研究成果

关于神经形态电路系统的具有代表性的研究成果，大部分来自国外，所以接下来先介绍国外的研究成果。

世界上主要的神经形态工程中心之一是神经信息学研究所（Institute of Neuroinformatics，INI），该研究所于 1995 年由 Rodney Douglas 和 Kevan Martin 在苏黎世大学和苏黎世联邦理工学院建立。INI 的代表性工作包括开发神经形态视觉传感器[6]、硅耳蜗[7]，以及中等规模的神经形态处理器，例如可重构的在线学习发放神经形态处理器[8]（reconfigurable on-line learning spiking neuromorphic processor，ROLLS）和 cxQuad 芯片[9]。这些芯片使用亚阈值模拟电路，并已被用于实现深度发放神经网络。他们已经演示了一个带有 9 个 cxQuad 芯片和一个 ROLLS 芯片的电路板[9]：cxQuad 芯片用于实现分层卷积网络；ROLLS 芯片实现了深度网络的分类层。与运行在大型计算机集群上的标准深度网络相比，该系统表现出低延迟和极低的功耗。

BrainScaleS 神经形态系统是海德堡大学在欧洲联盟资助的一系列项目中开发出来的，包括 FACETS 项目和 BrainScaleS 项目。对 BrainScaleS 的持续支持来自欧盟的信通技术旗舰人脑项目。该项目的特点是有两个方面，一是使用高于阈值的模拟电路来实现神经元过程的物理模型，从而产生了更快的电路，运行速度是生物速度的 10 000 倍；二是使用晶圆规模的集成来提供大量的模拟神经元，这是很激进的设计方案，但结果是神经元可以非常有效地互连，可以适应 10 000 倍的加速[10,11]。该项目的应用包括：模拟那些在生物学上需要很长时间的应用，例如，长期的学习任务，模拟几年的儿童发展；涉及到非常大规模的参数搜索或高速批处理模式的操作的应用加速。

IBM 公司的 TrueNorth 芯片[12]也是一项代表性成果，它旨在解决从视觉（特别是使用基于事件的视觉传感器，如 INI 开发的传感器）到听觉和多感官融合等一系列问题。它为高维、嘈杂的感官数据提供了非常省电的实时处理方案。在 TrueNorth 上运行的实时物体识别等应用已经被证明是在极低的功率水平下运行的[12]。

Intel 公司的 Loihi 芯片[13]是一项较新的成果。Loihi 芯片中的所有逻辑都是数字的，功能确定性的，并以异步捆绑数据设计风格实现。这使得发放以事件驱动的方式生成、路由和接收。这种实现方式非常适合发放神经网络（spiking neural networks，SNN），因为 SNN 的基本特征是在空间和时间上都具有高度的活动稀疏

性。同时，每个 Loihi 核心包括一个可编程的学习引擎，学习引擎支持简单的发放时间依赖可塑性规则（spike-timing-dependent plasticity，STDP），也支持更复杂的规则，如三联 STDP、带有突触标签分配的强化学习，以及同时参考速率平均和尖峰时序轨迹的复杂规则。

国内的一项代表性工作是浙江大学与杭州电子科技大学合作研发的达尔文芯片[14]，硬件资源受限的嵌入式应用是该芯片的一个目标应用场景。该芯片包含 8 个物理实现了的神经元模块，而每一个神经元模块可以通过时分复用的方式模拟 256 个神经元，所以整个芯片可以对 2048 个神经元进行模拟。这些神经元通过数字逻辑的方式实现，并且它们之间可以进行任意的连接，理论上可以有 4194304 个突触。不过实际可以实现的最大突触数量，其实是受限于芯片外部用于存储突触信息的存储单元的容量的。

国内的另一项具有代表性的工作是清华大学的 Tianjic 芯片[5]。该芯片的一个特点是，其基本功能核心（FCore）可以单独配置为传统的以计算机科学为导向的人工神经网络（Artificial Neural Network，ANN）、或者以神经生物学为导向的 SNN，也可以配置为两种网络的混合，即，处理 ANN 输入并以 SNN 输出，或者处理 SNN 输入并产生 ANN 输出。FCores 的互联共用相同的路由模块，以统一的格式传输路由包。输入数据包可以根据配置形式将输出的数据包打包成 SNN 或 ANN 数据包，后续接收单元根据其配置将路由数据包解析成 SNN 或 ANN 格式。各种神经网络和混合编码方案可以自由集成，实现了包括 SNN 和 ANN 在内的多个网络之间的无缝通信。这种新型的神经形态架构，通过将跨范式模型和算法集成到一个平台上，提供了极大的灵活性；这项研究结果将有希望能够加速通用人工智能的发展，并有许多可能的现实世界应用（论文中展示了一个自动自行车的应用实例）。

1.2.2 CANN 模型的相关研究现状

CANN 相较于 ANN 来说，是一个更加接近真实的生物神经系统的模型。有很多的实验性研究和理论性研究的发现证明，CANN 可以模拟诸如感官知觉、认知功能以及运动控制等许多存在于生物神经系统中的计算机制，下面例举一些具体的相关研究成果。

在感官知觉方面，对于猫的初级视觉皮层（primary visual cortex）中的神经回路的研究发现，这些神经回路可以对视觉刺激中的纹理朝向信息进行连续编码[15,16]；在对于恒河猴脑部的中颞区域（middle temporal area，MT）的研究中，研究者发现该区域的神经回路可以对视觉刺激中的运动朝向信息进行连续性编码[17]。在认知功能方面，CANN 可以模拟多种神经回路的行为，例如在外侧顶叶内皮层

（lateral intraparietal cortex，LIP）和额叶视野（frontal eye fields，FEF）中发现的知觉决策回路[18]、前额叶皮质（prefrontal cortex，PFC）中的空间信息工作记忆回路[19]以及老鼠和果蝇的脑神经系统进行空间导航定位时使用的头朝向神经回路[20,21]。在运动控制方面，CANN 可以模拟猴类的上丘区域（superior colliculus，SC）中对眼动控制指令进行连续编码的神经回路[18]。

另一方面，学界也已经有不少与 CANN 相关的神经形态电路研究。对于之前提到的 ROLLS 芯片，通过利用它具有的局部兴奋性突触（连接到神经元自身以及与之邻近的其他神经元）以及全局抑制性突触，可以将该芯片的神经网络配置为一个具有较弱的 WTA 机制的神经网络[22]。另一个例子是一款叫做 HICANN-DLS 的芯片[23]，该芯片的设计灵感来源于生物物理学的相关研究。在该芯片的网络拓扑结构的基础上，通过配置自连接的、以及邻近神经元相互连接的兴奋性突触，以及全局性连接的抑制性突触，同样可以让该芯片的神经网络表现出 WTA 的计算机制。另外，通过在之前提到的 Intel 公司的 Loihi 芯片上配置路径整合网络，该芯片也可以高效地实现能够对连续变化的信息进行编码的多层 SNN[24]。

相较于上述的大部分神经形态系统来说，本研究中实现的 CANN 系统具有对于更加丰富的神经机制的模拟能力。本文将在后续章节逐步展示和说明这一点。

在 CANN 的理论研究方面，除了针对于单个 CANN 的研究，也有对于多个 CANN 之间进行信息整合的相关研究。多个 CANN 组成的系统具有相对于单个 CANN 而言更加强大的计算能力，如 Zhang 和 Wu 提出[25]，理论上，由多个 CANN 互联构成的网络将可以模拟大脑对于不同的感官输入信息的最佳整合。在他们的模型中，每个 CANN 对应一个感觉模块，各个 CANN 之间相互连接，连接强度控制着整合的程度。在相互作用的调节下，来自不同线索的信息在感觉模块之间进行交换，进而在每个局部处理器上实现了全局信息的整合，而不需要一个中心化的整合单元。本论文的研究内容也涉及到了具有多个 CANN 核心的电路系统的设计。

1.3 论文章节安排

本论文以 CANN 模型为主要理论基础，结合多种神经机制的数学模型，利用 FPGA 作为硬件平台，实现了具有较为丰富的仿生物神经特性的神经形态电路系统。论文分为五个章节，内容安排如下：

第一章中，主要介绍了神经形态电路以及 CANN 模型的研究背景以及意义，通过列举国内外具有代表性的相关研究成果，展示了在这两个领域中的最新研究进展，同时也反映了本研究所提出的具有丰富特性的 CANN 电路系统所具有的创新性。

第二章中，详细介绍了具有多样突触动力学特性的 CANN 电路系统的工作原理、主要模块的结构以及两个仿真实验的结果。本章首先讲解了电路系统中使用到的 CANN 模型以及非线性突触的相关特点；随后详细介绍了神经元模型、突触模型和循环连接通路的原理与电路实现方式，以及系统中的外围电路的功能与模块；最后展示了基于 CANN 以及突触的特性实现的知觉决策以及工作记忆任务的仿真实验结果。这个阶段的研究重点在于建立起 CANN 核心电路系统的大致框架，验证各个相关的理论与技术要点。

第三章中，在第二章的研究成果的基础上，对 CANN 电路系统进行了两个重要的改进工作：一是加入了对于短时程可塑性和发放频率适应性这两个新的神经机制的仿真功能，使得该电路系统具有了更加丰富的特性；二是优化了循环连接通路的算法设计及其电路实现，大幅度降低了电路系统对于硬件资源的使用量。从而，本研究在该阶段实现一个电路设计较为成熟且计算功能丰富的 CANN 核心电路系统。最后本章展示了静默式工作记忆和 T 型迷宫这两个仿真实验的结果，证明了加入新的神经机制后的 CANN 电路系统具有更加多样的计算功能。

第四章中，介绍了基于片上网络技术实现的具有多个 CANN 核心的电路系统。本章主要讲解了该系统中的片上网络模块的功能和使用方法以及负责仿真数据存储的数据核心的工作原理，并在最后展示了证明该系统中各个核心之间的数据交互符号预期设计目标的测试结果。该阶段的电路，理论上将可以实现更加复杂的仿真功能。

第五章中，对论文中介绍的所有研究内容进行了详细的梳理与总结，并对在本研究基础上进行的后期研究工作进行了展望。

第二章 具有多样突触动力学特性的 CANN 电路系统

本章将介绍一个基于 CANN 模型的、具有丰富突触动力学特性的神经形态电路系统的原型设计，以及利用该系统进行的两个验证性仿真实验（知觉决策实验以及工作记忆实验）。

本章首先通过简述 CANN 模型的相关背景知识、电路中实现的 CANN 模型的结构以及该系统中使用到的突触特性的相关知识，说明了该电路系统的模型理论基础。

然后展开说明了 CANN 电路系统中最为重要的几个功能模块的工作原理及其结构，包括神经元突触模块、突触电流整合模块以及循环连接通路模块。另外，也简单介绍了用于与 PC 机进行交互的外围电路中的主要模块的功能。

最后介绍并展示了用于验证该电路系统原型功能的两个仿真实验以及实验结果：通过知觉决策实验，验证了系统中的非线性突触和循环连续通路共同作用下实现的知觉信息的积累过程；通过工作记忆实验，验证了系统在恰当参数设置下维持神经元发放活动的功能。

2.1 CANN 电路系统的模型理论基础

2.1.1 CANN 模型概述

CANN 模型是一种对于神经信息进行表征的网络模型，该类型神经网络中的信息被神经元的特殊放电模式所表征，一种放电模式对应于网络的一个稳态（即吸引子）[26]。一般来说，可以通过刺激一部分特定的神经元并使它们发生相对较强的发放活动（即一种放电模式）的方式向 CANN 系统中注入的特定的神经信息。

相较于其他类型的吸引子网络（如 Hopfield 网络[27]），CANN 最突出的特点是神经元之间的连接强度具有平移不变性。这里的平移不变性是指，网络中的某个特定神经元与其他神经元之间的连接强度，只与这些神经元各自表征的信息差异大小相关，而与输入的信息本身无关。举例来说，如果被一个 CANN 表征的信息是一个角的角度大小（0 到 360 度），那么表征 0 度与 30 度的两个神经元之间的连接强度和表征 50 度与 80 度的两个神经元之间的连接强度是一致的。

CANN 模型的数学描述有很多种形式，下面展示一种连续的（非离散的）CANN 模型，通过对这个模型的动力学行为进行理论分析，可以得到 CANN 的典型动力学特性，而其他的 CANN 模型的特性也与之类似[28]：

$$\begin{cases} \tau \dfrac{\partial U(x,t)}{\partial t} = -U(x,t) + I^{ext}(x,t) + \rho \int_{x'} J(x,x') r(x',t) dx' \\ J(x,x') = \dfrac{J_0}{\sqrt{2\pi}a} \exp[\dfrac{-(x-x')^2}{2a^2}] \\ r(x,t) = \dfrac{[U(x,t)]_+^2}{1+k\rho \int [U(x',t)]_+^2 dx'} \end{cases} \tag{2-1}$$

式中，参数 τ 是突触的时间常数。变量 x 是该 CANN 模型编码的一个连续性的刺激输入，类似于之前例子中的角度信息，变量 x 的取值范围是$(-\pi, \pi]$。该模型中的神经元是进行了质点化抽象的神经元。也就是说，每一个特定的 x 数值都对应一个，甚至是很多个神经元。实际对应多少神经元，由反映神经元的密度的参数 ρ 来决定。可以把一个特定的 x 数值作为与之对应的那一群神经元的编号来使用，所以式中的 $U(x, t)$表示 t 时刻输入到偏好 x 的神经元的突触输入大小；$I^{ext}(x, t)$表示 t 时刻输入到偏好 x 的神经元的外部突触电流大小；$J(x, x`)$表示偏好 $x$ 和偏好 $x`$的神经元之间的连接强度，它是自变量为 x-$x`$的函数；$r(x, t)$表示 t 时刻偏好 x 的神经元的发放率，其中，$[U]_+=\max(U, 0)$。

通过理论推导可以发现，在没有外部输入且 $0<k<\rho J_0/(8/\sqrt{2\pi}a)$的条件下，上述模型具有如式(2-2)所示的稳定状态解[26]：

$$\begin{cases} \bar{U}(x|z) = U_0 \exp[-(x-z)^2/(4a)^2] \\ \bar{r}(x|z) = r_0 \exp[-(x-z)^2/(2a)^2] \end{cases} \tag{2-2}$$

式中，z 是决定这个稳态解的位置的自由变量。

也就是说，CANN 的网络结构使得它可以拥有一系列吸引子（即系统稳态），而非一个孤立的吸引子。由于这些吸引子的存在，CANN 便可以较为稳定地对具有连续变化性质的信息进行编码以及追踪响应[29]。

在本研究中使用的模型是一维离散的 CANN 模型，该模型的结构将在接下来的这个小节中进行介绍。

2.1.2 电路系统中的 CANN

在本文提出的 CANN 神经形态电路系统中，实际使用的 CANN 模型的细节设计思路，来源于一个在生物物理意义上具有一定真实性的皮层神经回路模型的一维版本，该模型从计算神经科学的角度出发，阐明了知觉决策以及工作记忆在细胞以及神经回路层面的基础[30-32]。

在本文所使用的 CANN 模型中，锥体细胞之间的关系可以被形象地想象为这样的结构：所有椎体细胞均匀分布在一个圆周上，它们所编码的信息的差异正比于它们在这个圆周上的位置差异（准确地说，应该是它们所在位置与圆心构成的圆心角角度）。同时，它们互相之间的连接强度是以它们位置差异为自变量的一维高斯函数。

另一方面，系统中还含有负责产生全局性抑制的中间神经元，这里所谓的“全局性”是指“不对输入的信息有偏好”。也就是说，每当一个椎体细胞发放时，这个发放活动也会经由中间神经元，产生一个输入到所有的椎体细胞的抑制性输入。这样的设计就使得整个系统同时具有了强度呈高斯型分布的局部兴奋性连接和全局性的抑制性连接。

上述结构的示意图如下：

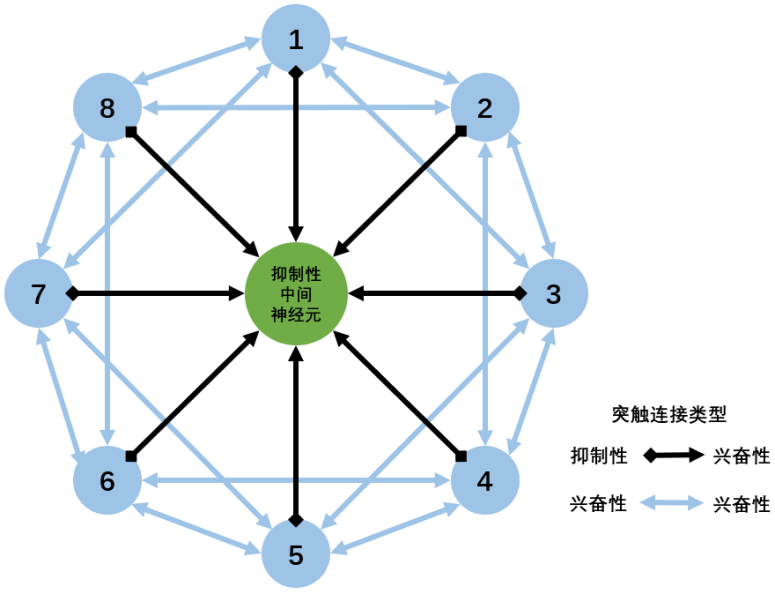

图 2-1 一个简单的 CANN 网络结构示意图

图 2-1 是一个只拥有 8 个椎体细胞的小规模 CANN 的网络结构示意图，其中带有数字编号的圆代表椎体细胞（也就是兴奋性神经元），中间绿色的圆表示抑制性的中间神经元。椎体细胞之间的连接都是兴奋性的，椎体细胞到中间神经元的连接也是兴奋性的，而中间神经元到椎体细胞的连接是抑制性的。椎体细胞之间的连接强度只与它们的对应编号的差异相关，也就是之前提到的“平移不变性”原则的体现。

在本章实际实现的电路中，CANN 系统含有 64 个锥体细胞，它们之间的局部兴奋性连接通过一个循环连接（recurrent connectivity）模块来实现。本文将在后续的部分中进一步仔细说明电路是如何实现这样的网络结构的。

2.1.3 突触特性概述

在真实的神经系统中，存在很多种类的兴奋性突触和抑制性突触[33-35]，多样的突触特性对于大脑实现其功能是至关重要的。具体来说，兴奋性突触按照突触递质的受体类型大致可以被分为 AMPA 型和 NMDA 型两类，类似地，抑制性突触可被分为 $GABA_A$ 型和 $GABA_B$ 型两类。AMPA 型和 $GABA_A$ 型突触在产生一个突触前发放后，会表现出突触电流快速升高然后再快速指数衰减的特点；而 NMDA 型和 $GABA_B$ 型突触的突触电流演化特点则是先缓慢上升然后在缓慢指数衰减[36,37]。

实验性研究和理论性研究发现，NMDA 型突触在决策[38,39]、工作记忆[30,36,40]以及神经迟滞[41]等现象的实现机制中扮演了十分重要的角色。而在最近的一项理论分析结合实验证据的研究中，研究者发现整个大脑皮层中的兴奋性突触和抑制性突触的时间常数呈现出一种层次关系，从这种层次关系中可以定性地获得非线性神经系统的新动力学特性[42]。

但对于大部分已有的神经形态电路来说，它们只含有发放后突触电流快速升高然后快速指数衰减的突触，也就是 AMPA 型和 $GABA_A$ 型突触，而没有对于具有较大的时间常数的 NMDA 类型的突触的模拟功能，这就导致这些电路系统缺乏真实神经系统所具有的丰富非线性动力学特性。

在本章所介绍的电路设计中，突触模块除了能够对 AMPA 型突触进行模拟外，还可以对 NMDA 型突触的动力学特性进行模拟。由于同时拥有循环连接通路与 NMDA 型突触，该电路系统在理论上就具有了对决策过程中进行的信息整合，以及对工作记忆维持神经元持续发放活动进行模拟的能力。这是因为在其他的一些研究中发现，循环连接通路中的 NMDA 型突触对于上述的这两类任务（决策过程和工作记忆）的实现是至关重要的[38-40]。

2.2 CANN 电路系统的工作原理和结构

本文描述的所有电路系统都是以型号为 ZYNQ-7100(xc7z100ffg2)的这个片上系统（System on Chip，SoC）作为平台实现的。该芯片的架构可以简单划分为两个部分：一个是 ARM 处理器部分，这个部分也被称作处理器端，或者 PS 端（processing system，PS）；另一个是可编程逻辑部分，也被称作 PL 端（programmable logic，PL），这个部分的功能与原本意义上的现场可编程逻辑门阵列（Field Programmable Gate Array，FPGA）是一致的。

实际的 CANN 电路系统的整体结构示意图如下：

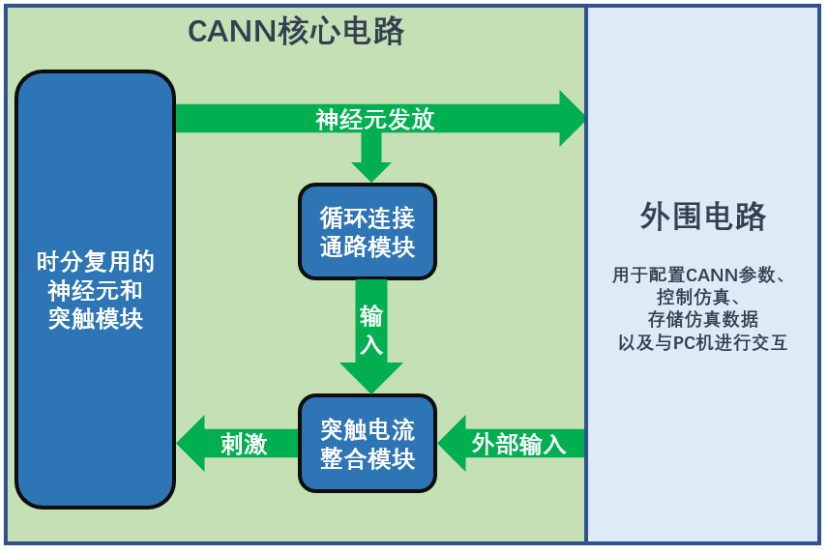

图 2-2 CANN 电路系统整体结构图

为了实现一个可用的 CANN 电路系统，上面提到的 PS 端和 PL 端这两个部分都被使用到了。

对于 CANN 核心电路，也就是实现 CANN 结构的电路，它是通过编写 Verilog 硬件描述语言代码，然后再利用与上述 SoC 配套的集成设计软件 vivado 生成的比特流文件，将代码描述的电路结构加载到 PL 端实现的。

而外围电路的情况稍微复杂一些，它既有在 PL 中的部分，也有在 PS 中的部分。具体来说，PS 端的部分主要负责整个 CANN 系统与 PC 机之间的控制指令交互以及数据交互，通过在与 SoC 配套的另外两个集成开发软件 vivado SDK（较老的版本）和 Vitis（较新的版本）中编写 C 语言代码来实现相关的功能。而外围电路在 PL 端的部分则主要是用于连接 CANN 核心电路与 PS 端，后续会对其中的子模块进行简单的介绍。

另一方面可以看到，CANN 核心电路的一个特点是，它只含有一个基于时分复用技术（time-division multiplexing，TDM）的神经元突触模块。也就是说，对于 CANN 中的所有兴奋性神经元的模拟计算并不是同时进行的，而是在不同的时钟周期内，按照神经元对应的编号从小到大依次计算。计算时所需要的相关变量以及参数等信息在 TDM 控制器的控制下，从神经元模块中负责数据存储的子模块中读出，然后再参与计算过程。这是对于神经元模块的工作原理的概述，在本节后续部分中将进一步详细介绍包括循环连接通路模块在内的 CANN 核心电路的工作原理。

2.2.1 神经元和突触的数学模型

神经元和突触作为基础性的计算单位，从根本上决定了整个神经形态电路系统的功能。实际上有很多种类的神经元模型是可以在硬件上实现的，例如 Hodgkin-Huxley 神经元模型[43]、Izhikevich 神经元模型[44]以及其他的双变量神经元模型[45]。

在本研究中使用的模型是泄露的整合发放（leaky integrate-and-fire，LIF）模型。该模型具有生物学意义上的合理性，同时，由于该模型并不十分复杂，所以它也具有出色的计算效率[46]。这是一个合理的折中方案。

神经元膜电位处于发放阈值之下的 LIF 模型的行为由式(2-3)描述：

$$\tau_m \frac{dV_m(t)}{dt} = -V_m + g_m I_{syn}(t) \tag{2-3}$$

式中，V_m 表示神经元的膜电位；I_{syn} 表示输入神经元的全部突触电流之和的大小，它由来自循环连接通路的电流（I_{rec}）和来自外部突触输入的电流（I_{ext}）累加得到；参数 τ_m 是反映 V_m 变化快慢的时间常数；参数 g_m 表示膜的增益，它反映 I_{syn} 对于 V_m 的影响大小。

而当膜电位 V_m 达到发放阈值电压 V_{thr} 后，神经元就产生一个发放。同时，V_m 将被重置到一个重置电位 V_{reset} 上，并在这个数值上维持一段长度为 τ_{ref} 的时间（这是对于神经元不应期的模拟）。

在实际的硬件实现上，对式(2-3)所描述过程的进行计算模拟的方法是数值计算中的欧拉法，需要将式(2-3)重新改写为如下形式：

$$\begin{cases} V_m(n+1) = \alpha_m V_m(n) + \beta_m I_{syn}(n) \\ \alpha_m = 1 - \Delta T / \tau_m, \beta_m = g_m \Delta T / \tau_m \end{cases} \tag{2-4}$$

式中，ΔT 表示欧拉法的计算步长，α_m 和 β_m 是用于欧拉法计算的固定参数。

另一方面，对于突触的模拟也是基于计算神经科学提出的模型来实现的。具体来说，本研究中包含两类突触模型，分别对应于之前提到的 AMPA（或 $GABA_A$）型突触和 NMDA（或 $GABA_B$）型突触。

对于 AMPA 型突触，其动力学特性可被式(2-5)描述[36,37]：

$$\frac{dS_A(t)}{dt} = -\frac{S_A(t)}{\tau_A} + \sum_k \delta(t - t_k) \tag{2-5}$$

式中，S_A 是 AMPA 型突触的门控变量，或者叫做突触变量，它表示突触中已经打开的离子通道所占比例；参数 τ_A 是反映 S_A 变化快慢的时间常数；δ 是单位脉冲函数，用来表示发放活动，当其自变量为 0 时，表示对应的时刻发生了一次发放。

对于 NMDA 型突触，其动力学特性可被式(2-6)描述[36,37]：

$$\begin{cases} \dfrac{dS_X(t)}{dt} = -\dfrac{S_X(t)}{\tau_X} + \sum_k \delta(t-t_k) \\ \dfrac{dS_N(t)}{dt} = -\dfrac{S_N(t)}{\tau_N} + [1-S_N(t)]\alpha S_X(t) \end{cases} \tag{2-6}$$

式中，S_N 是 NMDA 型突触的突触变量；S_X 是用于计算 S_N 的一个中间变量；参数 τ_X 和 τ_N 分别是 S_X 和 S_N 的时间常数；δ 是表示发放的单位脉冲函数。

上述的两个模型在硬件上的实现方式依然是欧拉法。为了方便设计，τ_X 和 τ_A 在数值上被设定为是一样的。因而，对于 S_A 和 S_X 的模拟计算可以同时在下面的式(2-7)中完成：

$$\begin{cases} S_A(n+1) = \alpha_A S_A(n) + \delta_D(n) \\ \alpha_A = 1 - \Delta T / \tau_A \end{cases} \tag{2-7}$$

式中，δ_D 是离散单位脉冲函数，它在发放产生的时刻数值为 1，其他时刻数值为 0；ΔT 表示欧拉法的计算步长；α_A 是用于欧拉法计算的固定参数。

而对于 S_N 的计算如式(2-8)所示:

$$\begin{cases} S_N(n+1) = \alpha_N S_N(n) + \beta_N [1-S_N(n)] S_A(n) \\ \alpha_N = 1 - \Delta T / \tau_N, \beta_N = \alpha \Delta T \end{cases} \tag{2-8}$$

式中，α_N 和 β_N 是用于欧拉法计算的固定参数。

2.2.2 神经元突触模块的详细介绍

基于上一小节中介绍的神经元与突触的模型以及与之对应的欧拉法近似计算公式，可以得到如下所示的神经元突触模块的**计算通路子模块**的结构设计，可以将这个模块划分为神经元模块和突触模块两个部分：

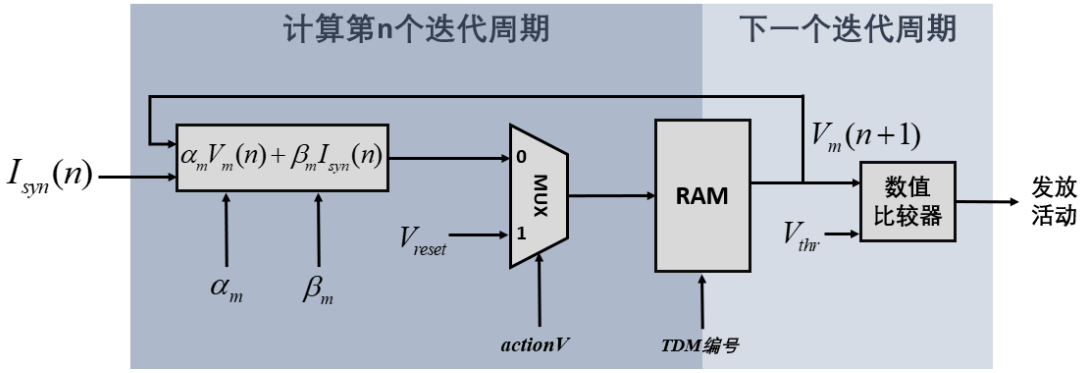

图 2-3 神经元模块的电路结构示意图

上图展示的是实际在电路中的神经元模块的结构。

需要注意的是，在两个迭代周期之间，有一个 RAM 模块作为过渡。在前面曾提到，神经元突触模块是基于 TDM 技术实现的，而这里的 RAM 模块便是实现 TDM 的基础。这个 RAM 模块的功能是，存储在模拟计算各个神经元变量的演化时所需的变量信息。当系统要进行第 i 号神经元迭代计算时，TDM 控制器就给出对应神经元的 TDM 编号信息（也就是 RAM 的地址信息），然后电路读出在上一个迭代周期计算出来并保存到 RAM 中的相关变量，送到迭代计算单位进行计算。

而对于整个 CANN 系统的模拟，就在 TDM 控制器的控制下，循环地按照从小到大的编号顺序遍历所有的神经元和突触。也就是说，TDM 控制器本质上只是一个很简单的计数器（它的代码实现也是这样的）。

数值比较器负责比较迭代计算出来的 V_m 与发放阈值 V_{thr} 的大小关系，当 V_m 超过 V_{thr} 时，就产生一个发放信号。

另一方面，由于突触模块也是基于 TDM 来实现的，所以突触模块的结构与神经元模块十分相似，如下图所示：

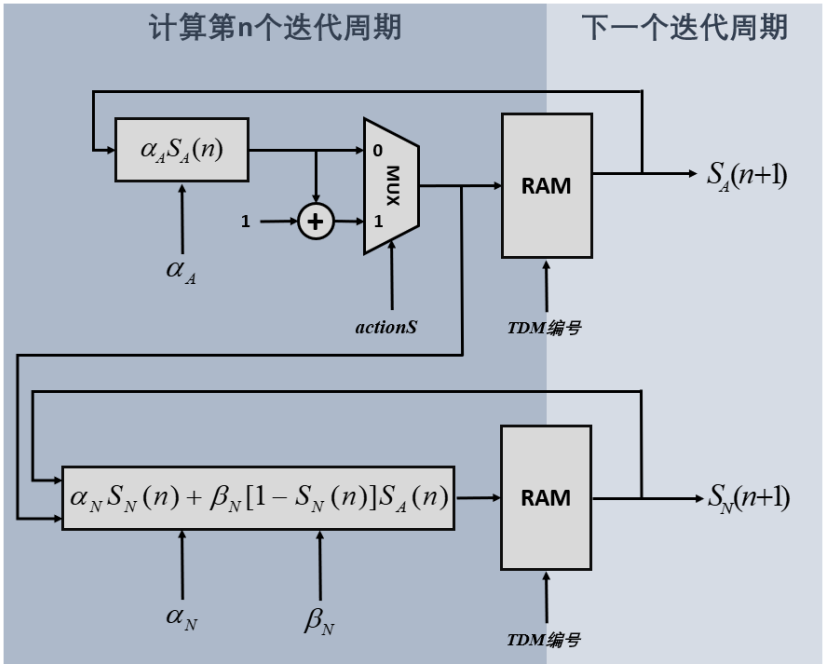

图 2-4 突触模块的电路结构示意图

在实际的代码实现中，上面展示的两个模块是在同一个子模块中的，也就是神经元突触模块的**计算通路子模块**。总体上来看，输入该子模块的数据主要包括：突触电流（I_{syn}）、TDM 编号以及各个用于欧拉法计算的参数。而该子模块输出的数据，即发放信息以及两个突触变量信息，被送到了外围电路中负责进行仿真数据保存的模块。

至于图 2-3 中的 actionV 信号和图 2-4 中的 actionS 信号，它们并不是从神经元突触模块的外部输入来的，而是来自神经元突触模块的另一个子模块（控制模块）的两个控制信号。

控制模块是一个很简单的状态机，它只有两个状态：计算状态和不应期状态。在计算状态下，计算通路模块就不停进行欧拉法迭代计算，直到产生一个发放后，上述的 actionV 和 actionS 信号变为 1，同时状态变为不应期状态；进入不应期状态后，actionV 和 actionS 信号变为 0，同时，一个不应期计数器开始从一个预设的数值开始，每一个周期进行一次减一操作，直到计数器归零，状态就再次进入计算状态。

但要注意的是，控制模块和计算通路模块一样，也是受到 TDM 控制器控制的。所以，上述状态机的状态变量和不应期计算器的计算数值，其实也是由 RAM 模块保存，然后在需要的时候再取出来使用的。不应期计数器的减一操作，并不是不停地在每一个时钟周期都减一，而是在对应的 TDM 编号到来的那个时钟周期才减一。所以严格来说，不应期计数器的计算周期长度是 TDM 控制器遍历了所用神经元一次的用时长度，而非时钟周期的长度。

2.2.3 突触电流的计算

本文中的电路系统可以对 AMPA 型和 NMDA 型两类突触进行模拟。具体来说，在循环连接通路中，也包含两个并行的通路，分别对应这两种突触模型。而在外部输入的突触中，包含有 4 组互相独立的 AMPA 型突触，其中 2 组与神经元（指的是之前提到椎体细胞，后面提到神经元时，若无特别说明，都是指椎体细胞）的连接是一对一的，即一个突触对应一个神经元；另外 2 组是高斯型连接，在后面会对此做进一步的说明。

所有这些突触的输入是在突触电流整合模块中进行累加后，再输入到神经元突触模块中的。下面分别介绍循环通路和外部输入的突触电流的计算方式。

循环通路的突触电流计算公式如下：

$$\begin{cases} I_{rec,R}^{i}(n+1)=\sum_{j=-(M-1)}^{M-1} W_R^{|j|} S_R^{i+j}(n)+W_R^{-}\sum_{j=0}^{N-1} S_R^{j}(n) \\ W_R^{|j|}=\dfrac{A_R}{\sqrt{2\pi}\sigma_R}\exp(-\dfrac{j^2}{2\sigma_R^2}) \end{cases} \tag{2-9}$$

式中，变量 S 的上标表示的是神经元编号，参数 W 的上标 j 表示两个神经元的编号差异，上标为-的 W 表示全局性抑制连接的强度。下标 R 从 A 和 N 两个中选取，分别用于表示 AMPA 型和 NMDA 型突触。W_R^j 表示两个神经元之间的循环通路连

接的强度，它的数值与两个神经元编码的信息的差异大小（即编号 j）有关。参数 N 表示整个 CANN 系统中的神经元数量，参数 M 则表示高斯型循环连接的实际作用范围。在 N 取值为 64 的条件下，M 取值为 8。参数 A_R 用于控制高斯型连接的整体强度，而参数 σ_R 控制高斯函数的波峰宽度。

在电路中，式(2-9)所描述的计算过程是通过卷积运算（S_R*W_R）来实现的，在下一小节中将对此详细说明。

另一方面，外部输入的突触电流计算公式如下：

$$I_{ext,R}^{i}(n+1)=\begin{cases} W_R S_R^i(n), R\in\{E_0,E_1\} \\ \sum_{j=-(M-1)}^{M-1} W_R^{|j|} S_R^{i+j}(n)+W_R^{-}\sum_{j=0}^{N-1} S_R^j(n), R\in\{E_2,E_3\} \end{cases} \tag{2-10}$$

式中，所有上标的意义和式(2-9)一致。下标 R 在 E_0、E_1、E_2 以及 E_3 中取。R 取前两个时，W 表示之前提到的一对一的外部突触的权重；而取后两个时，W 表示高斯型连接的外部突触的权重。外部突触的突触变量 S 的计算与式(2-7)一致。

2.2.4 循环连接通路模块的算法原理与结构

由于对于神经元和突触的模拟计算过程是受到 TDM 控制器控制的，所以计算循环连接通路的电流所需的突触变量 S（也就是 S_A 和 S_N）是逐个计算得到的，同时也是循环出现的。为了在硬件上实现循环通路电流的计算，循环通路的算法被设计为如下形式：

$$\begin{cases} C^j(n+1)=\begin{cases}(W^k+W^{-})S^i+T^{j-1}, j\in(N-M,M] \\ W^{-}S^i+T^{j-1}, otherwize\end{cases} \\ k=\min\{|j-1|, N+1-j\}\end{cases} \tag{2-11}$$

式中，T^{j-1} 在 TDM 编号为 0 时数值为 0，否则数值为 $C^{j-1}(n)$。变量 C 表示的是在电路中，对循环通路突触电流进行累加计算后得到的数据，其上标是神经元编号（取值范围从 0 到 N-1）。结合图 2-5 可以更方便且直观地解释这个算法。

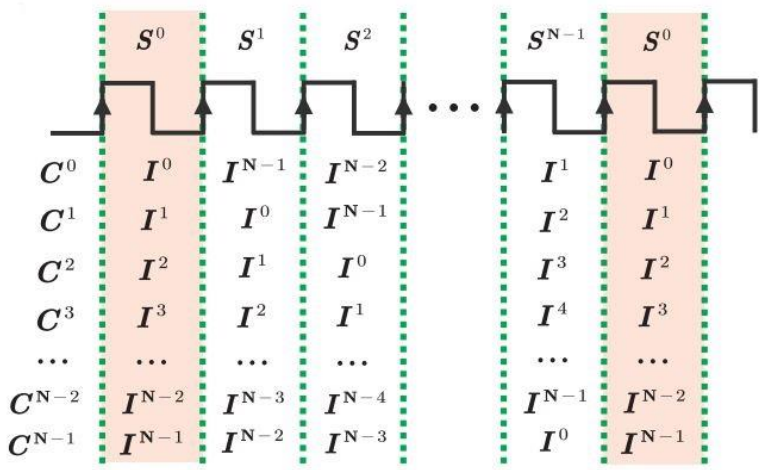

图 2-5 循环连接算法示意图[47]

在 TDM 编号为 0 的时钟周期内（即图中有背景阴影的部分），可以从神经元突触模块中读取到对应编号下的突触变量数据 S^0，同时，所有的 T^i 的数值都被重置为 0。在每一个时钟周期内，都可以读取到对应编号的突触变量 S^i，然后电路根据式(2-11)，计算出下一个时钟周期将要被存放到 C^i 位置的数据，并在时钟上升沿到来时，将计算出的数据写到对应的 C^i 中。

当运算过程来到 TDM 编号为 0 的时钟周期的上升沿时，也就是即将再一次进入阴影部分时，电路已经完成了对于所有编号的突触电流的计算，所以循环连接通路模块直接向后续的电流整合模块一次性地发送所有的计算结果。然后，电路进入下一个遍历所有神经元的循环过程中。

根据上述的计算过程，循环连接通路模块的电路结构被设计为如下的形式：

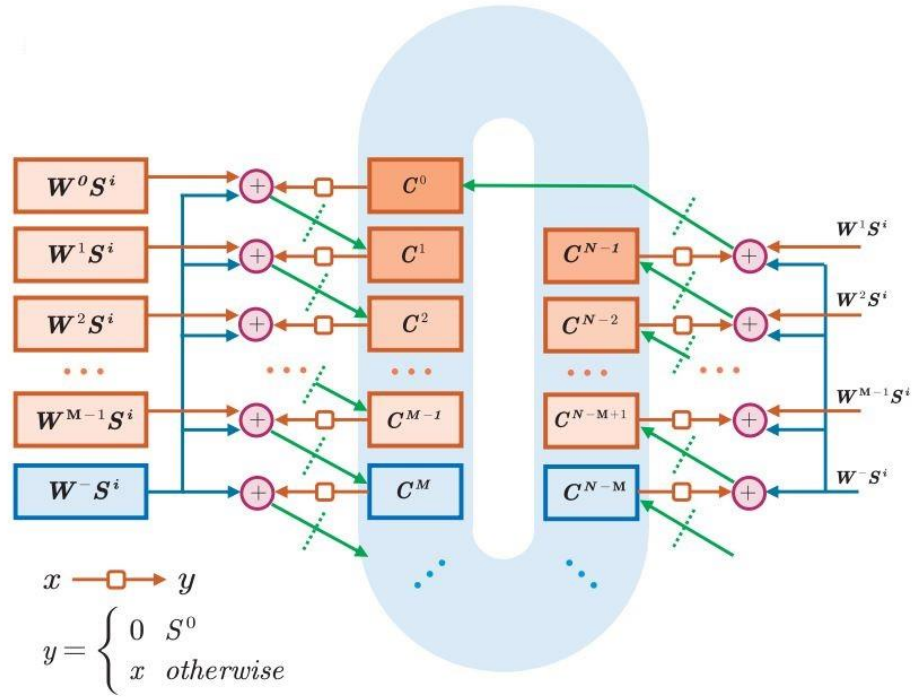

图 2-6 循环连接通路模块电路结构示意图[47]

在 CANN 中含有的神经元数量较少的时候，这样的电路结构消耗的硬件资源相对较少，是可以接受的。但随着神经元数量的提升，该结构将会消耗大量的硬件资源。所以，在第三章中，出于对更多神经元进行模拟的需求，本文将会提出对循环连接算法及其电路结构设计进行的改进。

2.2.5 外围电路概述

通过上面的介绍，实现 CANN 电路系统的各个关键模块（神经元突触模块、循环连接通路模块以及突触电流整合模块）的工作原理已经基本讲清楚了。下面再简单介绍一下外围电路部分。

之前已经提到过，外围电路包含在 PS 端和 PL 端的两个部分，下面先介绍 PS 端，然后在介绍 PL 端。

PS 端的设计内容主要是利用已有的 ARM 硬核，通过编写 C 语言实现对通信端口的控制和数据交互，进而来连接 PC 机与 PL 端。

PS 端与 PL 端是通过 GPIO 和 BRAM 作为中介，进而实现控制信号以及数据的交互的。PS 端和 PL 端都可以对 BRAM 进行读写操作，大部分系统配置参数仿真数据的交互是通过这个方式完成的。GPIO 的主要用途是从 PS 端向 PL 端写入一个 32 位的仿真控制参数，该参数包含有多个控制信号，如系统重置信号、仿真使能信号以及仿真数据存储使能信号等。

PS 端与 PC 机之间的交互则是利用 UART 串口完成的。在神经元数量较少的条件下，系统配置参数以及仿真产生的数据都比较少，UART 串口的数据传输速率是可以接受的。但在第三章中的新电路系统设计中，由于大幅度地增加了系统中的神经元数量，UART 串口的速率就显得很慢了，所以新的设计采用了其他通信方式。

另一方面，在 PL 端中，主要有这样几个模块：用于生成仿真时间信息的子模块、TDM 控制器、用于监控特定神经变量的监控模块、用于与 BRAM 进行交互的子模块以及用于模拟外部神经元发放进而生成输入到 CANN 的刺激信号的子模块。

产生仿真时间信息的模块和 TDM 控制器都很简单，都是计数器。在这里需要说明的是，生成仿真时间信息的操作与系统的时钟频率相关。为了让系统的仿真速度至少与真实的神经网络的运行速度一致，所以时钟频率 H 被设定为 $H=N(1000/\Delta T)=3.2$ MHz，其中的 $N=64$ 是神经元数量，$\Delta T=0.02$ ms 是欧拉法的时间步长。

PL 端中用于监控特定神经变量的监控模块会按照配置的相关参数来对仿真过程中的一个神经变量数据进行读取，并将读取到的数据传递给 BRAM 写入模块。具体来说，被监控的神经变量数据是在整个仿真过程中演化的某一个特定编号神经元的一个相关变量。被记录的变量类型从神经元膜电位、输入突触电流以及循环连接通路中的两类突触的突触变量中选择一个来记录，每当对于所用神经元的遍历完成 100 次后，该监控模块就记录一次被监控的变量的数值，并写入 BRAM 模块中。

PL 端中用于与 BRAM 进行数据交互的模块有两种，一种是用来读取 BRAM 中的数据的，另一种向 BRAM 中写入数据的。它们之中都包含一个根据 BRAM 的读或写操作的时序设计的状态机，它们在各自的状态机的控制下完成对 BRAM 的操作。

对于读取 BRAM 数据的模块，它接收一个来自 PS 端的 GPIO 控制信号。每当 PS 端向 BRAM 中写入了一个数据后，PS 端会拉高该 GPIO 信号，以告知该模块可以读取数据了。然后该模块在其状态机的控制下，完成对于数据的读取操作，并通过拉高 GPIO 的另一位数据的方式告知 PS 端读取完成，可以发送下一个数据。该模块读取出来的数据将被保存到 PL 端本地的存储单元，等待被调用。

对于向 BRAM 写入数据的模块，它会接收来自 CANN 核心电路部分的写入请求信号。每当需要向 BRAM 中写入数据时，这个请求信号就被拉高，该模块的状态机就开始进行写入数据的操作。

写入 BRAM 的仿真数据分为神经元发放数据和被监控的神经变量数据，这两类数据都是 32 位的。一个神经元发放数据由仿真时间信息和在该时刻发放了的神

经元的编号信息拼合而成，在神经元数量为 64 个的情况下，该数据用较高的 26 位来表示仿真时间信息，剩余 6 位表示神经元编号。而在第三章中，系统中含有的神经元数量被提升到了 512 个，在这个设计下，表示仿真时间信息的位数降为 23 位。而被监控的神经变量数据的相关信息则已经在前面提到，就不再重复了。

关于 PL 端中用于模拟外部神经元发放的子模块，或者叫做脉冲生成模块（Spike_Generator 模块），它的作用是根据从 BRAM 中读取来的配置参数信息来生成泊松发放序列，进而模拟来自 CANN 之外的神经系统的输入信号以及随机背景噪声。这个模块的工作原理是利用一个线性反馈移位寄存器（linear feedback shift register, LFSR）模块来生成伪随机数，然后让生成的伪随机数和一个用来控制发放率的配置参数比较大小，依据比较的结果来判断是否在当前的神经元编号对应的突触上产生发放活动。

2.3 仿真实验

为了测试上面介绍的 CANN 电路系统是否能够完成那些理论上它应该能够完成的认识计算任务，也就是之前提到的基于缓慢且非线性的动力学特性的复杂认识功能，本文使用在 FPGA 开发板上实现了的 CANN 电路系统，进行了对知觉决策[32]和工作记忆[30]这两项任务的模拟仿真。

仿真实验中所有的配置参数，包括 CANN 内部的网络参数（如神经元模型、突触模型以及循环连接通路中的权重信息）和输入到 CANN 的外部刺激相关的刺激协议参数（如输入到各个神经元的发放率配置参数）都是在 PC 机上使用 MATLAB 生成的定点数（Q12.20）。然后，这些数据再通过 UART 串口输入到开发板上。

下述的这些参数设置在这两个实验中是一致的，而其他的参数则在两个实验中并不一致：ΔT=0.02 ms，τ_m=8 ms，g_m=1，V_{thr}=0.5，V_{reset}=0 以及 τ_{ref}=2 ms。

2.3.1 知觉决策任务简介

在一些对猴类进行的视觉知觉决策实验中，神经生理学家们发现，猴子脑部的 LIP 区域的一些神经元对于特定的视觉目标（也就是猴子看到的在屏幕上出现的运动点）的运动方向具有偏好，并且这些神经元的活动与决策结果相关。

具体来说，这些实验中的视觉目标是这样的：一部分呈现给猴子的运动点的运动方向是一致的（向左或向右），其余运动点的运动方向是随机的。猴子需要判断运动方向一致的动点的运动方向，随机运动的点所占比例越大，这个任务就越难。而在猴子对这个的运动方向做出判断之前，LIP 的那些神经元的活动强度会出现一

个爬升过程，这被认为是神经系统进行的信息积累过程。在此过程中，偏好方向不同的神经元群之间会出现竞争关系，只会有一群神经元胜出（发放率达到一个阈值），而最终猴子做出的判断是与胜出神经元群的偏好方向是一致的[48,49]。

实验步骤的示意图如下：

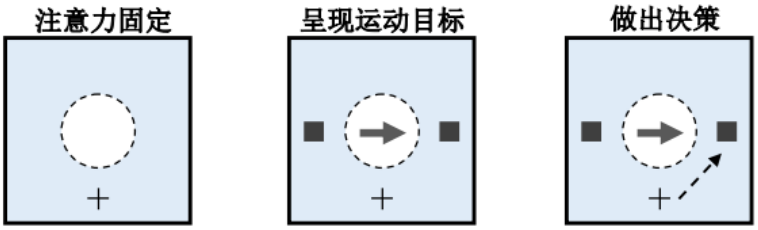

图 2-7 知觉决策任务实验流程示意图[47]

可以看到，实验的步骤分为三步：注意力固定、呈现运动目标以及做出决策。具体来说，猴子需要先将注意力集中的屏幕上一段时间（即，注视屏幕上的“+”标记一段时间），然后，屏幕会呈现上面提到的运动点和目标刺激（分别对应于图中的箭头和两个小方块），运动方向一致的动点是朝向两个目标刺激中的一个运动的。在猴子完成了对于运动点的运动方向的判断后，这一轮实验结束。需要进一步解释的一点是，猴子是通过一段时间的训练后，学会了用扫视活动（将视线移动到某一个目标刺激上）来表示它做出的选择是什么，实验人员记录的决策结果数据是猴子的扫视结果。

下面介绍该仿真实验的相关参数设置。

2.3.2 知觉决策仿真实验的设置

在电路上进行的知觉决策任务仿真实验，主要是对上述的实验中观测到的 LIP 区域神经元的活动的重现。

在这个实验中，CANN 网络参数被设置为：对于循环连接的非线性的突触（NMDA 型），τ_N=100 ms、A_N=4、σ_N=4.27 以及 W_N^- = -1.15；对于一对一的外部输入突触，W_{E0}=W_{EI}=1。每一个 CANN 中的神经元都从一对一的外部突触上接收到一个不相关的泊松发放序列作为背景噪声，其平均发放率为 ν^{Back}=1000 Hz。本实验中没有使用到循环连接中的 AMPA 型突触和高斯型连接的外部突触。

输入到 CANN 的刺激协议参考了[32]中的设置，输入的刺激由三部分组成：目标刺激（Target Input）、运动信息刺激（Motion Input）以及刺激控制信号（control signal）。刺激协议的示意图如下所示（没有明确画出刺激控制信号）：

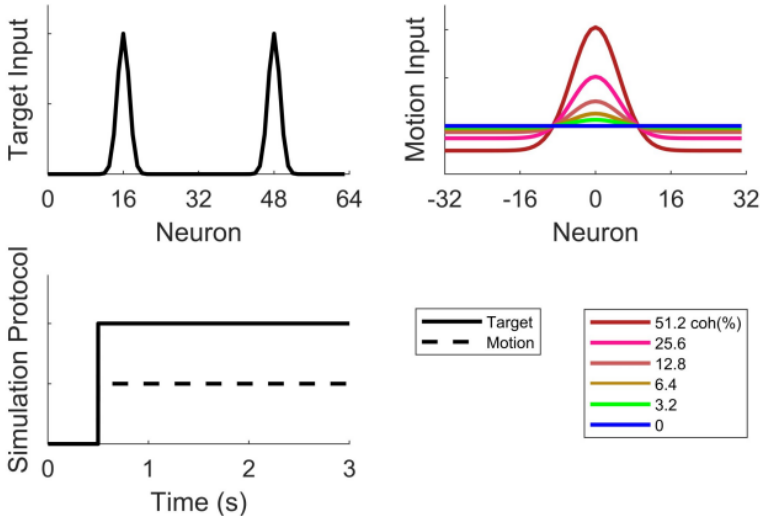

图 2-8 知觉决策刺激协议示意图[47]

具体来说，目标刺激 v_i^{Tar} 在 500 ms 的时刻会到达 CANN 网络，这个时长包含了长度为 200 ms 的神经系统的延时。输入到 CANN 系统中目标刺激参数通过式(2-12)计算得到：

$$v_i^{Tar} = v^{Tar}\sum_{k=1}^{2}\exp(-\frac{(i-\theta_{Tar}^{k})^2}{\sigma_{Tar}^2}) \tag{2-12}$$

式中，目标刺激的位置参数 θ_{Tar}^{k}取值为 16 或 48，分别对应在图 2-7 中展示的两个目标刺激（小方块）。目标刺激强度参数 v^{Tar}=300 Hz，宽度参数 σ_{Tar}=2.13。

对于运动信息刺激 v_i^{Mot}，它和目标刺激同时呈现，并同时到达 CANN，其计算方式如下：

$$v_i^{Mot} = v^{Mot}(1+coh\cdot(-1+5\exp(-\frac{(i-\theta_{Mot})^2}{\sigma_{Mot}^2})) \tag{2-13}$$

式中，coh（$0\leqslant coh\leqslant 1$）表示运动方向一致的动点所占的比例。运动方向参数 θ_{Mot} 取值为 16 或 48。其余参数设置为 v^{Mot}=100 Hz，σ_{Mot}=7.1。

至于刺激控制信号，它是输入到每一个神经元的互不相关的泊松发放序列，其平均发放率 $v^{Control}$=600 Hz。

系统是否做出了决策的判断条件是：参与竞争的两群神经元的发放是否到了阈值。在本实验中，这个阈值被设定为 20 Hz[32,38]。决策的反应用时（reaction time，RT）通过如下方式计算得到：

$$RT = t_D - t_{start} \tag{2-14}$$

式中，t_D 是做出决策的时刻（即达到阈值的时刻），t_{start}=300 ms 是考虑到了神经系统的 200 ms 延时的运动刺激信息呈现在屏幕上的时刻。

2.3.3 知觉决策仿真实验的结果

知觉决策仿真实验的结果可以通过分析下图中展示的数据来加以说明：

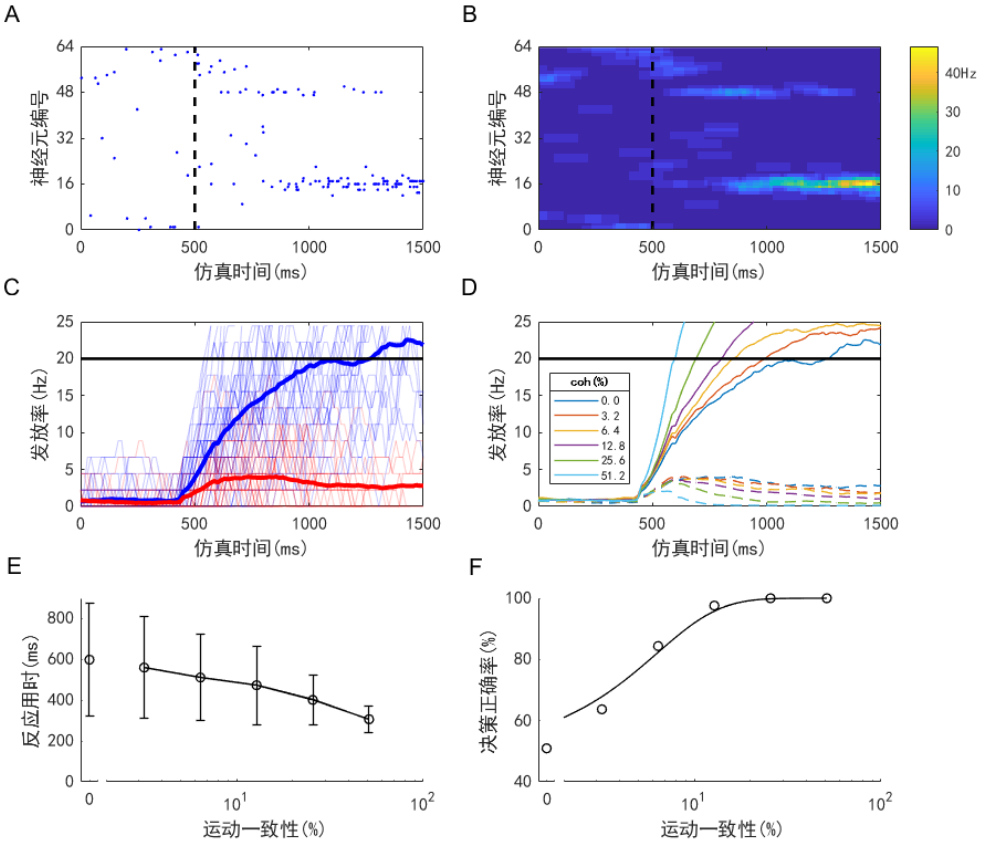

图 2-9 知觉决策仿真实验结果图

接下来分别介绍图 2-9 中的各个子图。

图 A 展示的是在动点运动一致性（指之前提到的 *coh* 参数）为 0 时，也就是所有点都在随机运动时，某一轮仿真实验的神经元发放数据。CANN 中的 64 个神经元按照它们的编号顺序排列，编号信息对应于它们偏好的信息。图中的每一个点的坐标对应一个发放事件，所谓发放事件就是哪一个神经元（纵坐标）在什么时候（横坐标）产生了一个发放。虚线表示运动刺激信息和目标刺激信息到的 CANN 的时刻。

图 B 是对图 A 进行平滑处理后得到的热力图。通过利用一个长度为 75 ms、滑动步长为 15 ms 的滑动窗来计算出在每一段时间内的神经元平均发放率，然后

再用热力图的方式来展示计算结果。这样就可以直观地看出发放率的演化过程，同时也就获得了判断决策是否产生的发放率信息。

结合图 A 和图 B 来分析一下整个仿真过程。从图 A 可以看到，在刺激信号到达 CANN 之前（虚线左侧），神经元的整体发放模式并没有表现出明显的偏好性，是在背景噪声的驱动下产生的随机发放。从图 B 可以看到，在刺激信号到达 CANN 之后（虚线右侧），在目标刺激信号的作用下，编号 16 和 48 的神经元出现了发放率的显著提升。然后，在编号 16 及其附近的神经元在全局抑制性连接和局部兴奋性连接的共同作用下，也就是在 WTA 机制的作用下，实现了对其他所用神经的发放活动的压制和自身发放活动的维持。

在图 C 中，上述的 WTA 机制的效果得到了更加明显的展示。在该图的背景中（即颜色较淡的那些较细曲线）包含在参数 *coh* 为 0 的条件下，随机挑选出的 100 轮决策结果正确的仿真实验中两个神经元（编号 16 和 48）的发放率数据。其中，胜出的神经元的发放率曲线颜色为蓝色，失败的神经元的曲线颜色为红色。而图中较粗的曲线则是对所有参数 *coh* 为 0 的仿真实验中，决策结果正确的实验轮次的数据的平均结果。也就是将所有选择了正确选项的轮次中，所有胜出的神经元的每一个时间窗口下计算得到的瞬时发放率做平均，得到较粗的蓝色曲线，对失败的神经与做相同的操作，得到较粗的红色曲线。图中的黑色水平线表示之前提到过的决策阈值。

图 D 是在不同的 *coh* 数值下，对决策结果正确的实验轮次的发放率数据取平均后的结果。可以直观地看到，参数 *coh* 数值越大，胜出的神经元的发放率就攀升得越快。这与在动物实验观察到的现象是一致的[48]。

在知觉决策任务中，做出正确决策的平均反应用时和决策的正确率是两个评判决策表现的重要指标。图 E 和图 F 分别展示了 CANN 系统的这两个指标。图 E 展示的是在不同 *coh* 数值下，CANN 做出正确决策的平均反应时间。可以看到，平均决策用时随着任务的难度降低而减少。图 F 展示的则是不同 *coh* 数值下的决策正确率数据，在所有动点都随机运动（即 *coh* 为 0）时，CANN 系统的决策结果是近乎随机的。而当运动一致性提高时，决策正确率也随之提高。

下图是知觉决策任务的动物实验结果图：

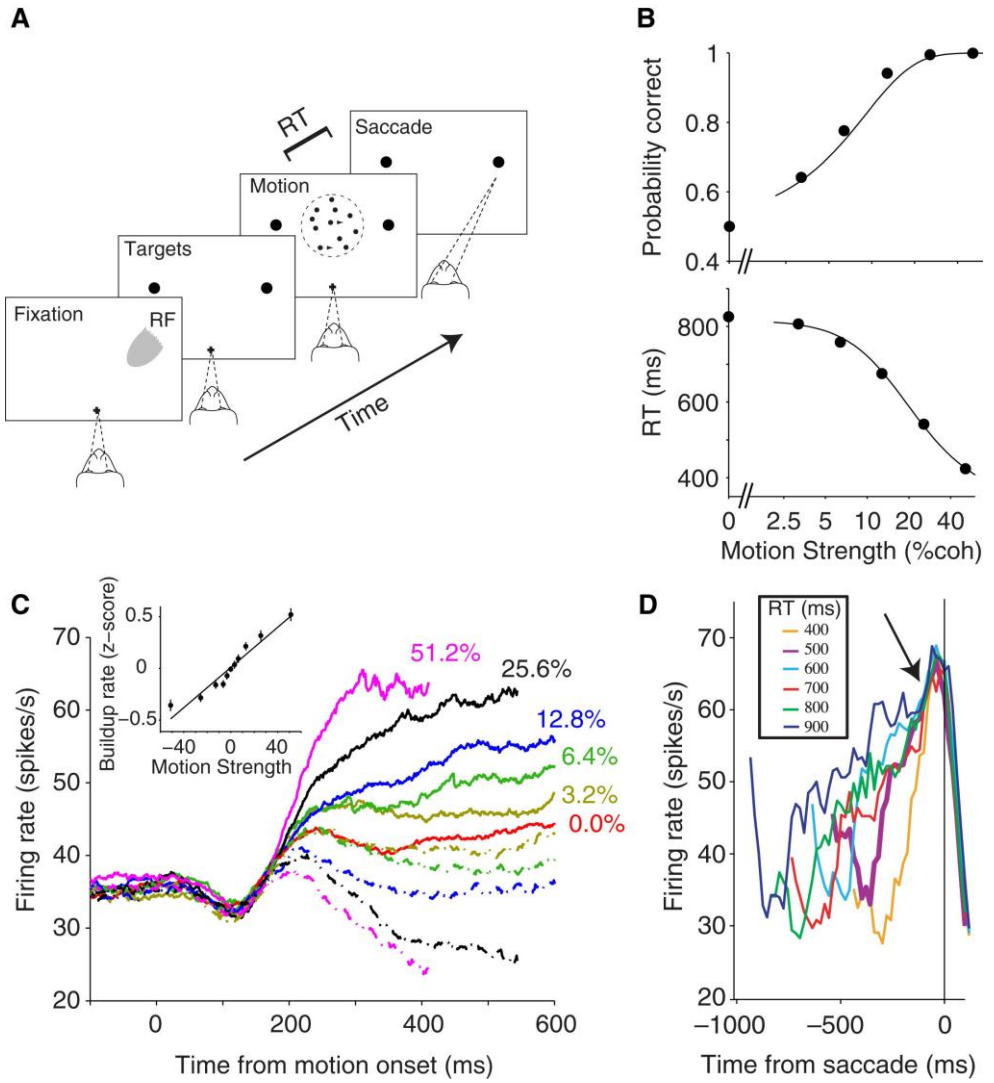

图 2-10 知觉决策任务动物实验结果图[49]

在图 2-10 中，图 A 是动物实验的流程示意图。图 B 展示了动物实验中得到的决策正确率以及反应用时数据。图 C 展示了在不同 *coh* 数值下，被监测的 LIP 区域的神经元的平均发放率演化趋势；图 C 中左上角的小图展示的是以 *coh* 数值（横坐标）为自变量的，在目标刺激出现后 200 ms 至 400 ms 之间的被测神经元发放率的变化趋势（纵坐标），横坐标为负的情况对应于被压制的神经元群的输入，纵坐标为正（负）表示发放率有提高（降低）的趋势。图 D 展示的是把不同条件下的发放率数据在被试猴子做出眼动的时刻进行对齐后的结果。

总体来说，仿真实验结果和在猴类动物上进行的实验研究结果[48-50]以及理论模型仿真结果[32,39,51]是吻合的，这就证明了该 CANN 电路系统初步地达到了设计的目标。

2.3.4 工作记忆任务的简介与仿真实验设置

在对猴类动物进行的延迟反应任务相关的实验研究中发现，猴类前额叶皮层记录到的神经元的持续性发放活动与工作记忆有关。所谓工作记忆，是指神经系统将输入信息维持并对这些信息进行操作的能力[52,53]。在这个仿真实验中，对神经元维持自身的持续发放活动进行了模拟。

实验的步骤如下图所示：

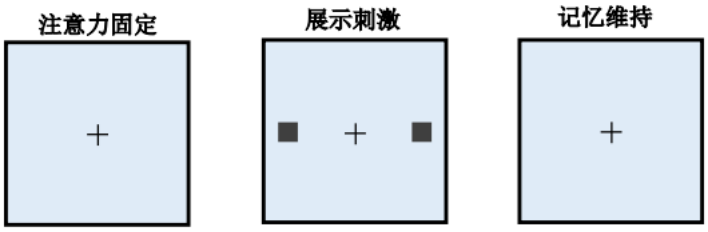

图 2-11 工作记忆任务流程示意图[47]

和知觉决策任务一样，工作记忆任务首先也需要被试进行注意力固定。然后，给被试呈现目标刺激（图中的小方块），刺激在出现 200 ms 后消失，从而被试进入对刺激信息的维持阶段（也就是工作记忆阶段）。在动物实验[52]和本文的仿真实验中，目标刺激的个数并不一定就只有两个。记忆的目标数量越多，记忆维持的难度就越大。

在该实验中，CANN 网络参数被设置为：对于循环连接的非线性的突触（NMDA 型），τ_N=100 ms、A_N=1.1、σ_N=0.71 以及 W_N^-=0；对于循环连接的线性的突触（AMPA 型），τ_A=5 ms、A_A=-2、σ_A=0.71 以及 W_A^-=-0.035；对于一对一的外部输入突触，W_{E0}=W_{E1}=0.1。每一个 CANN 中的神经元都从一对一的外部突触上接收到一个不相关的泊松发放序列作为背景噪声，平均发放率 ν^{Back}=1000 Hz。

输入到 CANN 的刺激协议参考了[30]中的设置，输入的刺激由两个部分组成：均匀分布的目标刺激信号（Cue Input）以及刺激控制信号（control signal）。刺激协议的示意图如下（没有明确画出刺激控制信号）：

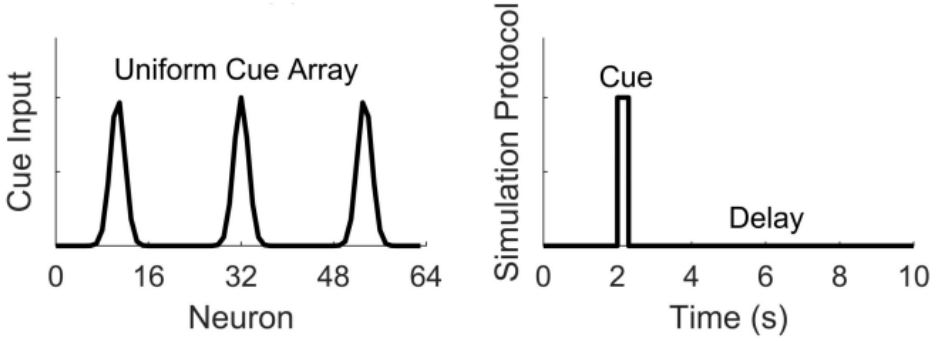

图 2-12 工作记忆刺激协议示意图[47]

具体来说，目标刺激信号在 2000 ms 时刻输入到 CANN 中，并在 2200 ms 时刻归零。其具体数值由式(2-15)计算得到：

$$v_i^{Cue} = v^{Cue} \sum_{k=1}^{N_{Cue}} \exp(-\frac{(i-\theta_{Cue}^k)^2}{\sigma_{Cue}^2}) \tag{2-15}$$

式中，目标刺激的位置参数 θ_{Cue}^k取值取决于目标刺激的数量 N_{Cue}，目标刺激均匀地分布在 CANN 的圆环上（见图 2-1 中的结构）。目标刺激强度参数 v^{Cue}=2000 Hz，单个目标刺激的宽度参数 σ_{Cue}=2.13。

而刺激控制信号与知觉决策仿真实验中一样，是输入到每一个神经元的互不相关的泊松发放序列，其平均发放率 $v^{Control}$=800 Hz。

2.3.5 工作记忆仿真实验的结果

和之前分析知觉决策仿真实验一样，还是以分析图片中展示的仿真数据的形式来说明仿真结果。

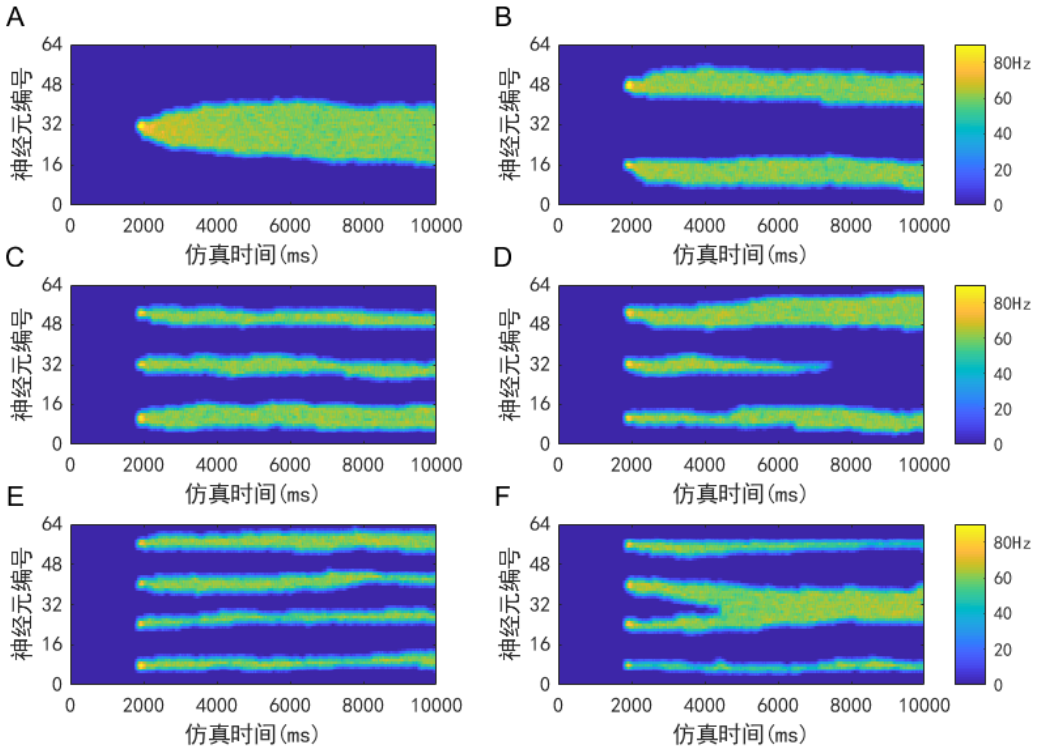

图 2-13 工作记忆仿真实验结果图

图 2-13 使用与图 2-9 中的图 B 一样的方式得到的热力图，展示了具有 1 个、2 个、3 个以及 4 个均匀分布在 CANN 圆环上的目标刺激输入的工作记忆任务仿真结果。可以看到，在目标刺激输入到 CANN 以后，被刺激神经元及其附近的神经元立即产生了响应，并在刺激消失后，继续持续自身的发放活动数秒。

对于刺激目标较少的情况（如图 A 和图 B 所示），被记忆的信息可以稳定地维持大约 8s，直到仿真实验结束，并且刺激目标只有 1 个或 2 个的绝大多数仿真实验轮次的结果都是这样的。

当刺激目标数量提升到 3 个或 4 个时，就会出现记忆信息没有被很好维持的情况，例如图 D 中出现的记忆信息丢失和图 F 中的记忆信息融合混淆。但在有的轮次下，记忆信息还是可以被维持到仿真结束的（图 C 和图 E）。而当刺激目标数量进一步提高到 5 个时，正确地维持记忆信息对于这个 CANN 系统来说就不能实现了。也就是说，在图 D 或图 F 中出现的情况，在每一个目标数为 5 的仿真轮次中都出现了。

上述的这些仿真实验结果与在发放神经网络模型中的仿真结果[30,54]以及在猴类动物上进行的实验结果[19]是一致的。下图是工作记忆的动物实验结果图：

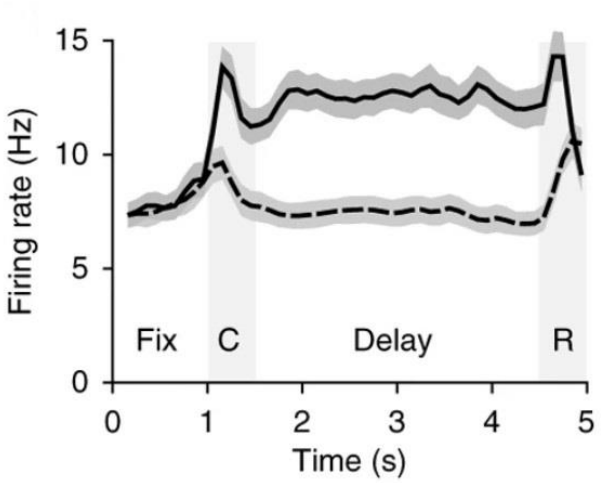

图 2-14 工作记忆动物实验结果图[19]

上图展示的是猴类动物实验中， PFC 区域的 204 个被测量神经元的发放率在输入不同目标刺激后的演化情况。实线对应目标刺激是被测神经元偏好的刺激的情况，而虚线对应目标刺激不是被测神经元偏好刺激的情况。被标记字母 C 的阴影部分表示刺激呈现的时间区间，被标记字母 R 的阴影部分表示要求被试动物回忆起记忆内容的时间区间。可以看到，被测的神经元在其偏好的目标刺激消失后依然维持了相对较高的发放率（即实线对应的情况相较于虚线对应的情况，在实线的情况下，神经元在 Delay 区间具有更高的发放率），基于电路实现的仿真实验较好地复现了这个过程。

2.4 本章小结

本章主要介绍了一个基于 CANN 模型构建的、具有多样突触动力学特性的神经形态系统的电路设计及其仿真实验应用。本章的具体内容是依据研究的进展顺序来划分的：

首先对该电路系统的设计理论基础进行了介绍，包括所使用的 CANN 模型以及线性突触和非线性突触的相关研究，通过这些内容，展示了该电路系统的设计理念；

然后较为详细地介绍了该 CANN 电路系统的各个重要组成模块的工作原理以及结构，包括该系统所使用的神经元和突触的数学模型、基于这些数学模型搭建的神经元突触模块、输入到神经元模块的突触电流的计算方法、基于卷积运算实现的循环连接通路模块以及用于仿真控制和数据交互的外围电路；

最后的部分介绍了使用该 CANN 电路系统进行的知觉决策和工作记忆任务的仿真实验，通过这些仿真实验，验证了该 CANN 电路具有的认知计算能力。

第三章 具有 STP 特性的 CANN 电路系统

上一章中的 CANN 系统是一个原型设计，目标在于初步实现模拟 CANN 的功能，验证整个架构的可行性。而本章将在原设计的基础上，探讨如何进一步实现神经元数量更大，功能更加丰富的 CANN 系统。

为了让 CANN 系统具有更加丰富的认知计算功能，新的设计中加入了对于短时程可塑性（Short-term plasticity，STP）和发放频率适应性（Spike frequency adaptation，SFA）两种新的神经机制的仿真功能。同时，由于原设计并未充分考虑循环连接算法的硬件实现所消耗的计算资源过多的问题，导致原设计的神经元数量可扩展性受限。所以，在本章中，系统改进的另一个重点在于循环连接算法的改进。CANN 电路系统的其他改动还包括：

1. 将系统中含有的神经元数量从 64 个提升至 512 个；
2. 提高了 TDM 控制器的时钟频率（从 3.2 MHz 提升到了 50 MHz，与之对应的，欧拉法的时间步长改为 0.0102 ms），以满足对拥有更多神经元的系统的仿真的实时性；
3. 使用 BRAM 来存储神经元和突触相关的各个变量，减少了对于 FPGA 逻辑资源（如 LUT 和 FF）的消耗；
4. 从原有的四组输入突触中去除了具有高斯型连接的那两组（因为其使用价值不大），只留下两组一对一的输入突触。其中代号为 E0 的那一组用于接收来自于脉冲生成模块（外围电路中的 Spike_Generator 模块）的输入，代号为 E1 的那一组则是为后续搭建多核心系统预留的，用于接收来自其他 CANN 核心的发放输入；
5. PC 机与 FPGA 开发板之间的通信方式由 UART 串口改为了基于 UDP 协议的网口通信，获得了较高的数据传输速度。

3.1 加入更多神经机制的神经元突触模块

3.1.1 新加入的 STP 和 SFA 机制的简介

STP 是指突触的传递效率在较短时间内（从数百毫秒至几分钟的量级）发生增强或减弱的现象，突触效率增强的 STP 被称为短时程易化（short-term facilitation，STF），而突触效率减弱的 STP 被称为短时程抑制（short-term depression，STD）。

与 STP 相关的研究发现，该机制对很多种类的神经认知功能的实现都有重要的意义，例如皮质增益控制[55]、听觉定位[56]、工作记忆[57]以及神经网络的信息存储[58,59]等神经机制或认知功能都与 STP 相关，STP 也被认为与中枢神经系统进行学习和记忆的生物学机制有关[60]。

另一方面，SFA 是指神经元在接收到一个方波刺激（或者说恒定刺激）时，其发放率响应由一开始时的快速升高转变为逐渐降低的现象。有很多种机制可以造成这一现象，但这些机制都有一个共同点：它们都包含某种对于兴奋性神经元的负反馈[61]。

相关研究发现，SFA 机制广泛存在于生物神经系统中，并和神经系统的一些认知功能和节律行为相关。例如，大蜗牛食管下神经节中的神经元就具有 SFA 机制[62]，这个发现被用来解释这些神经元的发放节律；前庭小脑中的浦肯野细胞（Purkinje cells of the vestibulocerebellum）被发现具有 SFA 机制[63]；初级听觉中间神经元的强度不变性的产生机制与 SFA 相关[64]；新皮质神经元的节律行为也与 SFA 机制相关[65]。

上述 STP 和 SFA 机制的加入，将会让 CANN 系统具有更加丰富的特性，进而可以实现更多的认知计算功能。

3.1.2 STP 机制的数学模型

本文所使用的 STP 机制的数学模型如式(3-1)所示，通过调节模型参数的方式，该 STP 数学模型可以实现对 STF 和 STD 这两种机制的模拟[60]：

$$\begin{cases} \dfrac{dx(t)}{dt} = \dfrac{1-x(t)}{\tau_D} - u(t)x(t)\delta(t-t_{sp}) \\ \dfrac{du(t)}{dt} = \dfrac{U-u(t)}{\tau_F} + U[1-u(t)]\delta(t-t_{sp}) \end{cases} \tag{3-1}$$

式中，变量 x 被用来描述与 STD 相关的过程（即突触中神经递质因发放活动而被消耗的过程），其含义是突触中剩余可使用的所有神经递质的量，参数 τ_D 是反映 x 变化快慢的时间常数。变量 u 被用来描述与 STF 相关的过程（即发放活动使钙离子进入突触前端，进而导致突触释放递质概率增大的过程），其含义是突触中将要被某一次发放活动使用的神经递质的量，它也反映了突触中残余的钙离子的量。参数 τ_F 是反映 u 变化快慢的时间常数。参数 U 是变量 u 的初始值，同时，这个参数也限制 u 在一次发放活动中的变化幅度。δ 是单位脉冲函数，用来表示发放活动，当其自变量为 0 时，表示对应的时刻发生了一次发放。

变量 x 和 u 都有一个初始的基准值，且它们都在一定范围内发生变化：变量 x 的基准值为 1，变化范围是 0 到 1 之间；而变量 u 的基准值为参数 U，U<1，变化范围为 U 到 1 之间。每当突触发生了一次发放后，该突触对应的 x 和 u 变量都会发生数值上的突变：x 减小而 u 增大。而当 x 和 u 变量偏离了基准值且没有发放的时候，它们会向着各自的基准值缓慢变化，变化速率与其对应的时间常数相关，时间常数越大，变化速率越小。通过调节两个时间常数的大小，可以让模型的行为更偏向于 STD（$\tau_D > \tau_F$）或 STF（$\tau_F > \tau_D$）。

为了在硬件上实现上述数学模型描述的 STP 机制，这里依然是使用欧拉法作为数值计算方法来进行对模型的近似模拟。将式(3-1)改写为如下形式：

$$\begin{cases} x(n+1) = \beta_x + \alpha_x x(n) + u(n)x(n)\delta_D(n) \\ \alpha_x = 1 - \dfrac{\Delta T}{\tau_D}, \beta_x = \dfrac{\Delta T}{\tau_D} \\ u(n+1) = \beta_u + \alpha_u x(n) + U[1-u(n)]\delta_D(n) \\ \alpha_u = 1 - \dfrac{\Delta T}{\tau_F}, \beta_u = \dfrac{\Delta T}{\tau_F} U \end{cases} \tag{3-2}$$

式中，δ_D 是离散单位脉冲函数，它在发放产生的时刻数值为 1，其他时刻数值为 0；参数 α_x、α_u、β_x 以及 β_u 是用于欧拉法迭代计算的常数，ΔT 是迭代时间步长。

最终，STP 相关的两个变量（u 和 x）的乘积作为突触发放时的突触变量改变量，增加到原有的突触模型中，如式(3-3)所示：

$$\begin{cases} S_A(n+1) = S_A(n) + u(n)x(n)\delta_D(n) \\ S_N(n+1) = S_N(n) + u(n)x(n)[1-S_N(n)]\delta_D(n) \end{cases} \tag{3-3}$$

式中，S_A 和 S_N 分别是 AMPA 型突触和 NMDA 型突触的突触变量。

3.1.3 SFA 机制的数学模型

本文使用了一种较为简单的 SFA 模型[61]，如式(3-4)所示：

$$\begin{cases} \dfrac{dg(t)}{dt} = -\dfrac{g(t)}{\tau_{SFA}} + \delta(t - t_{sp}) \\ \tau_m \dfrac{dV_m(t)}{dt} = -V_m(t) + g_m(t) I_{syn}(t) - w_{SFA} \cdot g(t) \end{cases} \tag{3-4}$$

式中，变量 g 用于描述负反馈的瞬时强度，参数 τ_{SFA} 是反映 g 变化快慢的时间常数，δ 是用于表示发放活动的单位脉冲函数，τ_m 是反映神经元的膜电位变化快慢的时间常数，变量 V_m 是神经元的膜电位，g_m 是神经元细胞膜的增益参数，变量 I_{syn} 是输入神经元的突触电流，参数 w_{SFA} 是用于调节 SFA 强度的权重变量。

SFA 机制在硬件上的实现方式依然是基于欧拉法的数值计算近似，将式(3-4)改写为用于欧拉法计算的公式，如式(3-5)所示：

$$\begin{cases} g(n+1)=\alpha_{SFA}g(n)+\delta_D(n) \\ \alpha_{SFA}=1-\dfrac{\Delta T}{\tau_{SFA}} \\ V_m(n+1)=\alpha_m V_m(n)+\beta_m I_{syn}(n)-w_{SFA}g(n) \\ \alpha_m=1-\dfrac{\Delta T}{\tau_m},\beta_m=g_m\dfrac{\Delta T}{\tau_m} \end{cases} \tag{3-5}$$

式中，δ_D 的意义与在式(3-2)中相同，还是用于表示发放活动的离散单位脉冲函数。参数 α_{SFA}、α_m 和 β_m 是用于欧拉法迭代计算的常数，ΔT 是迭代时间步长。

可以看到，新加入的 SFA 机制是在上一章的神经元模型的基础上改进得到的。当神经元产生一个发放后，反映 SFA 强度的变量 g 在乘了一个权重后，直接加入到了膜电位的计算过程中。这样就形成了一个负反馈：一个神经元越是发放，变量 g 就增长得越快，神经元也就越容易受到 SFA 机制的抑制，从而不再容易发放。

3.1.4 新的神经元突触模块的电路结构简介

对于新加入的 STP 和 SFA 机制的模拟功能是按照上述的两个欧拉法近似计算公式，式(3-3)和式(3-5)，利用与第二章中描述的神经元模块（图 2-3）和突触模块（图 2-4）的电路结构类似的设计方式来实现。也就是说，对于 STP 和 SFA 机制进行模拟计算的电路结构，依然是利用一个 RAM 模块（新模块在此处使用了 BRAM）作为两次迭代计算的中介的，同时，该 RAM 模块被 TDM 控制器控制。

由于这个新的设计和原本的设计十分类似，所以这里就不对新电路的结构进行更进一步的介绍和展示了。

3.2 使用更少资源的循环连接算法

上一章中的 CANN 系统使用了卷积计算作为实现高斯型循环连接的方式，这个方法的局限性在于：当 CANN 中的神经元数量增多，进而使得参与循环连接计算的神经元数量等比例增多时，该方法的硬件资源消耗量将会大幅度上升。这将限制该 CANN 系统的规模与未来的应用，所以有必要对循环连接算法进行优化。

3.2.1 对 NMDA 型突触变量计算方式的简化

为了实现新的循环连接算法，首先对 NMDA 型突触变量的计算过程进行了改进，简化了原先计算方法中对于该变量的非线性变化过程的模拟，如式(3-6)所示：

$$\frac{dS_N(t)}{dt} = -\frac{S_N(t)}{\tau_N} + [1 - S_N(t)]\alpha\delta(t) \tag{3-6}$$

式中各个变量、参数和函数的含义与式(2-6)相同。对应地，NMDA 型突触变量的欧拉法近似计算公式改写如式(3-7)所示：

$$S_N(n+1) = \alpha_N S_N(n) + [1 - S_N(n)]\alpha\delta_D(n) \tag{3-7}$$

式中各个变量、参数和函数的含义与式(2-6)和式(2-8)相同。

通过 MATLAB 的模型仿真测试发现，这样的改动并不会影响到 CANN 系统原有的认知计算功能。具体来说，通过对比在给予系统相同输入刺激信号的条件下，改动后和改动前的系统中的神经元发放活动，可以发现两者差别很小。同时，这样的改动也减少了对于计算资源的消耗，符合改进算法的初衷。

3.2.2 累加发放活动的循环连接算法

下面以 NMDA 型突触为例，介绍旧算法与新算法的差异。

之前的循环连接算法，可以归纳为“累加突触变量”的循环连接算法。为了方便突出新旧算法的差异，这里将上一章中的式(2-8)和式(2-9)合并到一个公式中，如式(3-8)所示：

$$\begin{cases} S_N^i(n+1) = S_N^i(n)\alpha_N + S_A^i(n)[1 - S_N^i(n)] \cdot \alpha \cdot \Delta T \cdot \delta_D^i(n) \\ I_{rec}^j(n+1) = \sum_i [w_i \bullet S_N^i(n+1)] + w_- \sum_k S_N^k(n+1) \end{cases} \tag{3-8}$$

式中，所有上标都表示神经元编号，例如，S_N^i 表示第 i 号神经元的 NMDA 型突触变量；I_{rec}^j 表示输入第 j 号神经元的循环连接电流（在这里只包含了来自于 NMDA 型突触的电流）。S_A^i 是计算 S_N^i 的非线性变化部分时使用的中间变量。i 表示所有与 j 号神经元有循环连接的神经元编号；k 表示所有的神经元编号，用于计算全局抑制电流。w_i 表示 i 号神经元所对应的高斯型循环连接强度，w_- 表示全局性抑制连接权重。

可以看到，最终输入到目标神经元的循环连接通路电流，是通过累加突触变量与权重的乘积的方式计算得到的。这个累加计算过程，正如之前章节所描述，是通过使用大量的乘法器，并行地计算各个参与了循环连接的神经元对电流的贡献，并不断累加这些神经元的贡献，然后在遍历完所有神经元后，一次性地在一个时钟周期内更新所有的循环连接电流。

新的循环连接算法针对旧算法使用乘法器资源过多的缺点，修改了计算的步骤，如式(3-9)所示：

$$\begin{cases} S_N^j(n+1) = S_N^j(n)\alpha_N + \sum_i \{[1-S_N^i(n)]\cdot w_i' \cdot \alpha \cdot \Delta T \cdot \delta_D^i(n)\} \\ I_{rec}^j(n+1) = W \cdot S_N^j(n+1) \end{cases} \tag{3-9}$$

式中，曾在式(3-8)中出现过的符号意义不变。需要强调的一点是，编号为 i 的神经元是**突触前神经元**，编号为 j 的神经元是**突触后神经元**。新出现的参数 w_i`表示 i 号神经元所对应的高斯型循环连接的相对强度，它与之前的参数 w_i 相差 W 倍，在实际的电路计算过程中，这样做造成了一定的精度损失，即参与计算的数值的低位被截断了。这样修改用于计算的连接强度的目的是为了尽可能减少所使用的计算资源。同时，用于刻画 NMDA 型突触非线性变化过程的中间变量 S_A^i 也被省略了（如上一小节所述），目的同样是减少算法所需的计算资源。

新算法最重要的改变在于，大幅度地减少了同时进行的乘法计算的数量。具体来说，对于第 j 号神经元的 NMDA 突触变量的计算，即式(3-9)的第一式子中，累加各个与第 j 号神经元有循环连接的神经元（即编号为 i 的神经元）对于第 j 号突触变量的改变的贡献的过程（即连加号所表示的部分），是经过数个时钟周期才完成的，并且是**由脉冲驱动的**。

也就是说，循环连接造成的第 j 号神经元突触变量的改变量的计算，不再是数个乘法模块同时进行计算。而是只有一个乘法模块，它在第 i 号神经元（也就是一个突触前神经元）产生发放后，依次计算出将要受到第 i 号神经元影响的各个突触后神经元（也就是一系列在公式中编号为 j 的神经元）的突触变量改变量，并将这些改变量保存到 BRAM 中等待被读取使用。若两个或更多的突触前神经元产生的发放的影响范围有重叠，则还需要累加那些重叠范围内的突触变量改变量，然后更新 BRAM 中保存的改变量的值。这些被保存到 BRAM 中的突触变量改变量接下来会被 TDM 控制器控制下的后续模块读取，累加到变量 S_N^j 上。在读取完成后，BRAM 中的改变量就被置零。这个计算过程示意图如图 3-1 所示：

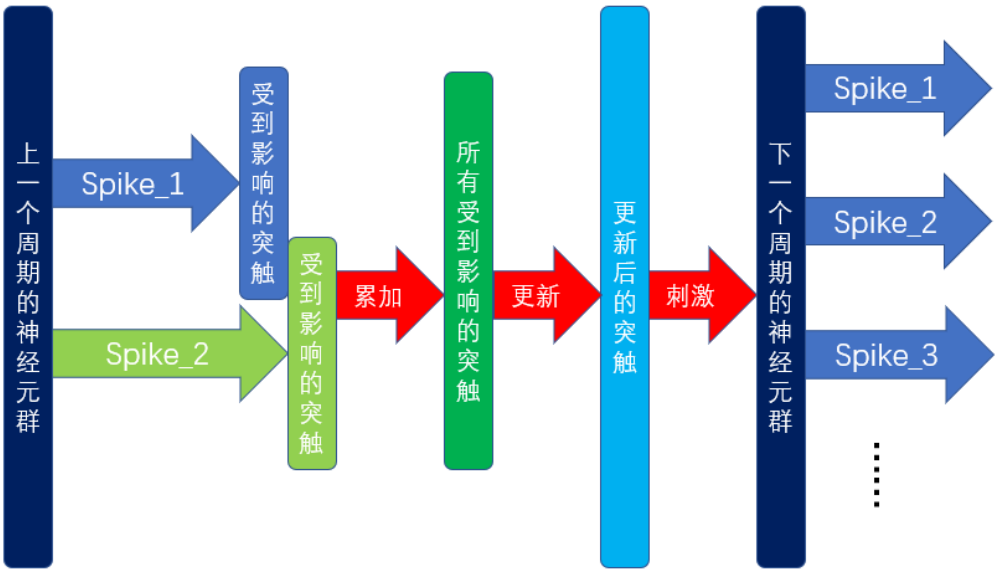

图 3-1 新循环连接算法示意图

另一方面，在计算输入到第 j 号神经元的循环连接电流时，即式(3-9)的第二式子中，只含有一个乘法计算。这个计算也由 TDM 控制器控制，也就是说，只需要一个计算模块来负责这个乘法计算即可。

同时，可以看到，这个式子中没有了抑制性的输入。这是因为对抑制性突触的设计进行了改动，不再像上一章节中描述的那样，分别使用 AMPA 和 NMDA 两种兴奋性突触变量来计算两种抑制性突触输入。在新的设计中，只含有一种突触变量演化方式与 AMPA 型突触模型一致的抑制性突触，它的计算过程独立于兴奋性突触的计算，计算方式与式(2-7)一致。

具体来说，新的抑制电流输入计算方式如下：

$$\begin{cases} S_{Inhb}^{i}(n+1)=\alpha_{Inhb}S_{Inhb}^{i}(n)+\delta_{D}(n) \\ \alpha_{Inhb}=1-\Delta T/\tau_{Inhb} \\ I_{Inhb}(n+1)=\sum_{i}W_{Inhb}S_{Inhb}^{i}(n+1) \end{cases} \tag{3-10}$$

式中，S_{Inhb}^{i}表示第 i 号神经元对应的抑制性突触变量，δ_D 依然是用于表示发放活动的离散单位脉冲函数，τ_{Inhb} 是抑制性突触的时间常数，ΔT 表示欧拉法的计算步长；α_{Inhb} 是用于欧拉法计算的固定参数，I_{Inhb} 是抑制性电流，W_{Inhb} 是控制抑制性电流大小的权重变量。

通过这样的改动，可以让抑制性突触的配置更加灵活。

3.3 新循环连接算法的硬件实现

3.3.1 新的循环通路整体架构

新的循环连接通路模块由三个子模块组成：FIFO 模块（用于连接神经元模块与循环连接通路中的后续模块）、权重发生器模块（用于将循环连接权重乘到输入的 STP 变量上）以及较为复杂的变量 P 发生器模块。

新的 CANN 系统中存在两个并行的、结构一致的循环连接通路，分别对应于 AMPA 型突触和 NMDA 型突触。

为了让循环通路中的计算速度可以满足神经元模块进行计算的需求，权重发生器模块以及变量 P 模块需要用到比神经元模块更快的时钟。其原因是，每当 CANN 系统中的神经元产生了一个发放，这个发放造成的影响将会作用到 50 个神经元上（如果系统中有 512 个神经元的话）。如果循环通路中用于计算发放造成影响的电路所使用的时钟与神经元时钟一致，那么在一个大循环周期内，也就是在欧拉法计算完全部 512 个神经元的一个时间步长所使用的时间内，循环通路模块最多只能计算完 10 个神经元发放。尽管短时间内出现这个数量的发放的可能性不是很大，但为了让系统具有一定的冗余性，这里还是将循环通路的时钟频率设置成为 100 MHz。

一个循环通路的示意图如图 3-2 所示：

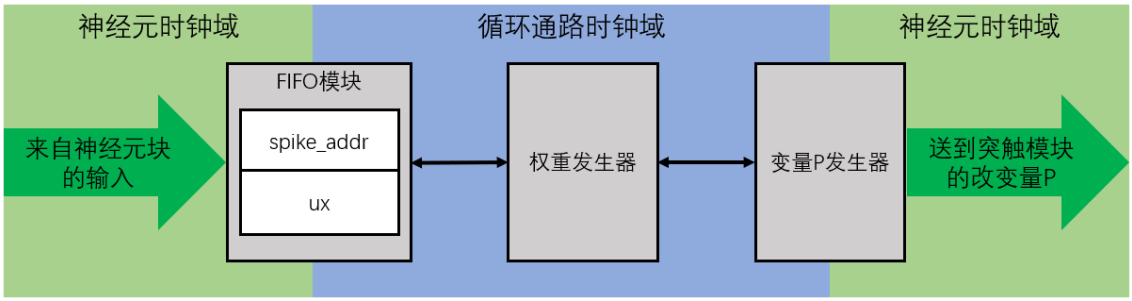

图 3-2 循环连接通路示意图

3.3.2 FIFO 模块以及权重发生器

FIFO 模块中包含两个直接调用 vivado 的 ip 核生成的异步 FIFO，分别用来存放产生了发放的神经元的编号信息（图 3-2 中的 spike_addr 部分）以及该发放神经元的 STP 变量（图 3-2 中的 ux 部分）。送入 spike_addr 部分对应的 FIFO 的有效数据位宽是 9 位（对应 512 个神经元），而 ux 部分是 23 位（包含了一个符号位），这些数据在写入对应 FIFO 中时都进行了补零的位数扩充操作。spike_addr 部分的 FIFO 数据位宽设置为了 16 位，而 ux 部分的数据位宽为 32 位，都是具有冗余性

的设计。FIFO 模块完成了从神经元模块到循环连接通路模块的跨时钟域数据传输，然后将数据交给权重发生器模块。

权重发生器模块的工作流程是：从 FIFO 模块中读取神经元编号信息，根据此信息，在不同时钟周期内，依次从小到大地计算出将要受到影响的神经元编号，然后在每一个时钟周期内向后续的变量 P 发生器发送一个计算出来的神经元编号，直到所有被影响到了的神经元被遍历了一遍，然后再读取下一个神经元编号（如果有的话）。在发送神经元编号的同一个时钟周期内，还要发送读取到的 STP 变量与对应编号的权重值的乘积，也就是变量 P。

上述的权重发生器的工作流程示意图如图 3-3 所示：

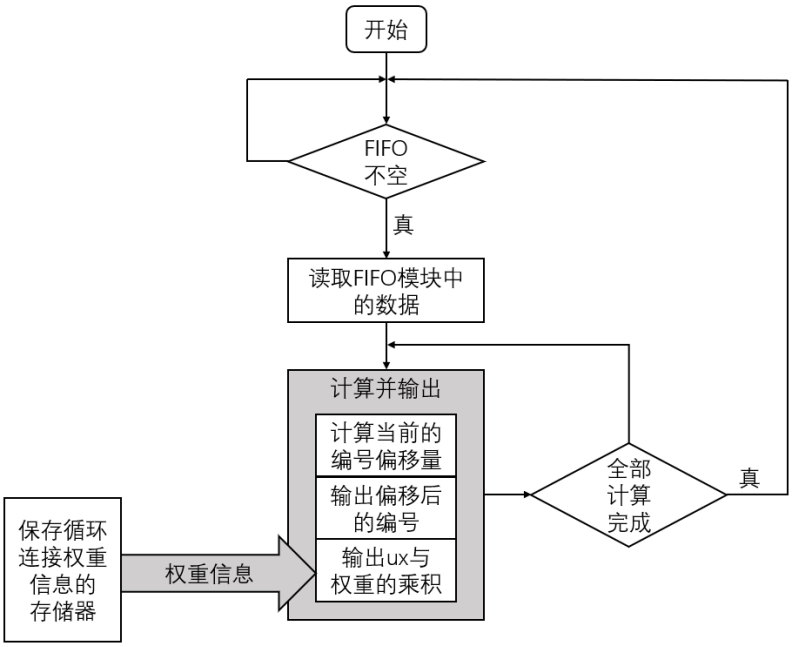

图 3-3 权重发生器的算法流程图

3.3.3 对变量 P 发生器的概述

变量 P 发生器的作用，正如之前在介绍新算法时提到的，是将突触变量改变量（即变量 P）保存到 BRAM 中，或者更新 BRAM 中已有的变量 P 数据（新的变量 P 来源于发放影响编号范围有重叠时需要进行的累加操作）。BRAM 中的数据将等待被后续的突触电流整合模块调用，然后被清空置零。

由于突触电流整合模块是一个在较慢的神经元时钟域下工作的被 TDM 控制器控制的模块，而变量 P 发生器接收到的数据是在较快的循环通路时钟域下的，这又涉及到跨时钟域的数据传输问题。另一方面，并不能对 BRAM 的一个地址同时进行读操作和写操作，这会让电路出现不稳定的状态。同时，也不能在同一个时

钟周期内，同时读取一个 BRAM 中两个不同地址的数据。而为了实现更新变量 P 数据的操作，又需要对数据不停地进行读出后再写入的操作。

出于上述的这些因素，变量 P 发生器被设计为如图 3-4 所示的结构：

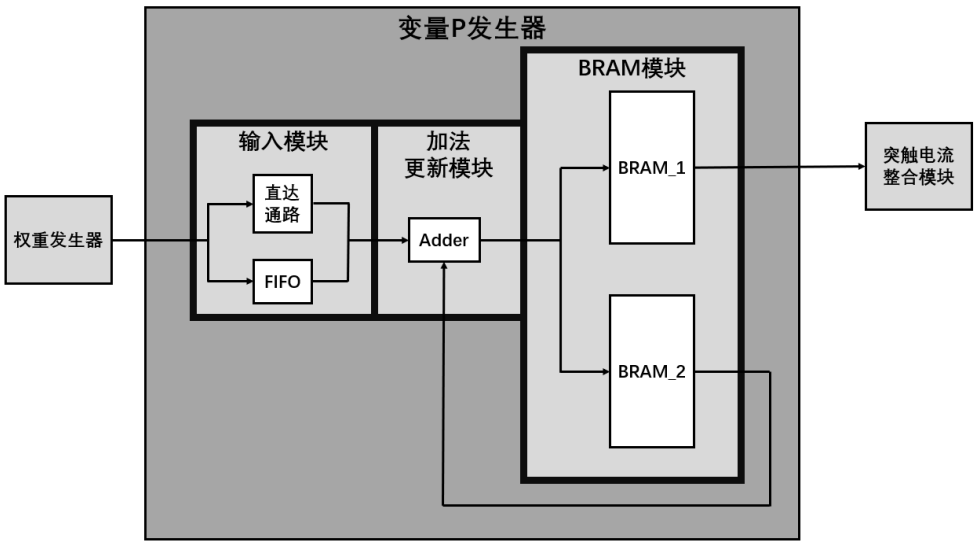

图 3-4 变量 P 发生器结构示意图

从图 3-4 中可以看到，变量 P 发生器有三个子模块：输入模块、加法更新模块以及 BRAM 模块。

输入模块中包含两个并行的数据通路，这是一个并不合理的设计。设计中加入直达通路的初衷是希望离散的单个数据能够不经过 FIFO 读写操作，直接到达后续模块，从而获得更快的数据传输速度。但这其实并没有必要，因为从上一模块输入的数据是一段连续的数据，这样的数据通过 FIFO 时并不会在整体上降低数据的传输效率。所以只需要一个 FIFO 就行了，而非两个通路。在设计这个子模块的时候没有意识到这个问题，这是一个可以进一步改进的细节。

加法更新模块的作用是对 BRAM 模块中存储的数据进行累加更新操作，同时，整个变量 P 发生器模块的控制也由这个模块完成。在下一小节中将详细说明这个作为变量 P 发生器中的核心模块的结构和工作原理。

BRAM 模块中包含有两个 BRAM 子模块，工作模式都是简单双口模式，这个结构（双 BRAM）是实现预期功能的一个关键。这两个 BRAM 子模块的功能分别是：BRAM_1 模块只负责进行跨时钟域的数据传递，当后续的突触电流整合模块需要编号为 j 的神经元的数据时，它就输出它所存储的对应地址上的数据；BRAM_2 模块则是对于 BRAM_1 模块的一个备份，它存储的数据与 BRAM_1 模块的数据完全一致。然后，当加法更新模块需要用到 BRAM 中存储的数据时，就直接读取 BRAM_2 模块中的数据。

这样的 BRAM 模块设计，使得在一个时钟周期内同时获得两个不同地址下的数据（也就是同时获取两个不同编号的神经元的数据）成为可能。同时，循环通路的计算过程中对于数据的读取操作与跨时钟域传输数据时的读取操作也被隔离开来，这大幅度地降低了数据读取操作的控制复杂性。

3.3.4 加法更新模块的结构与工作原理

加法更新模块的结构示意图如下：

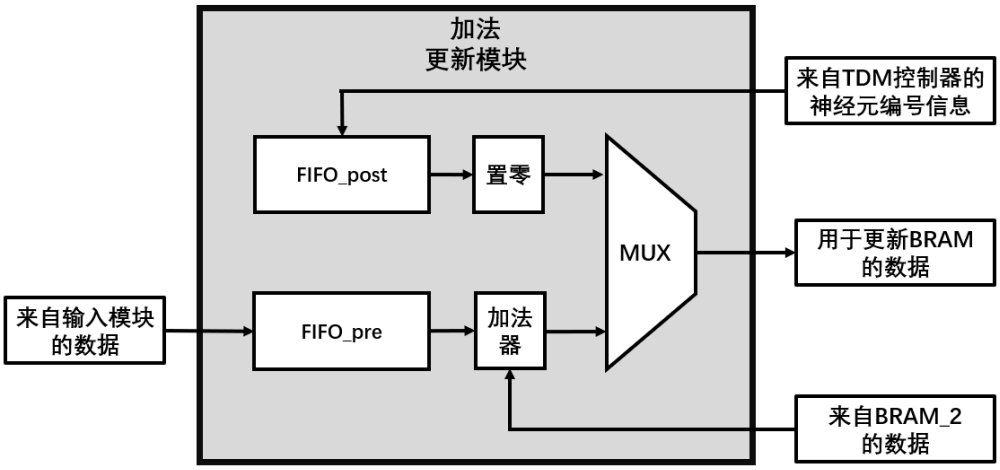

图 3-5 加法器更新模块结构示意图

如图 3-5 所示，加法更新模块的结构大致可以划分为两个 FIFO 模块和一个多路选择器（MUX）。输入到 MUX 的数据是由置零操作或加法操作得到的。

FIFO_pre 模块负责接收来自输入模块的神经元编号信息以及变量 P 信息；FIFO_post 模块则接收 TDM 控制输出的神经元编号，这个信息主要用于告知加法更新模块，当前突触电流整合模块正在读取的神经元的编号，进而避免加法更新模块在对 BRAM 进行写入操作时与突触电流整合模块的读操作发生冲突（即之前提到的，BRAM 模块不能对一个地址同时读写）。加法器负责在适当的时候计算累加后的变量 P，然后输出到 MUX；而置零模块负责将 TDM 发来的神经元编号对应地址上的变量 P 清空（即置零）。最后由 MUX 直接将接收到的数据送到正确的 BRAM 地址中。

整个加法更新模块的工作流程如图 3-6 所示：

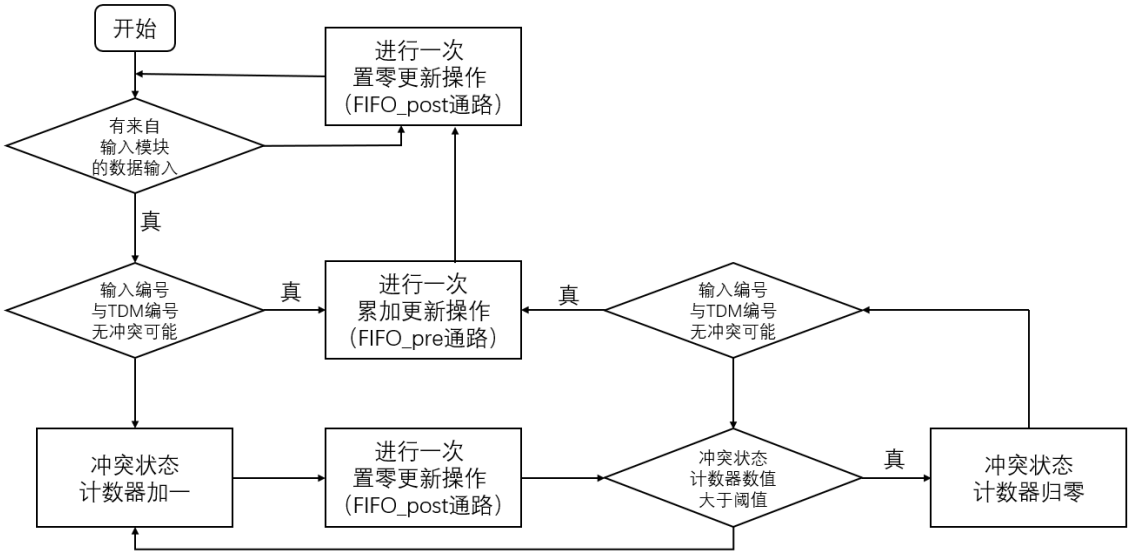

图 3-6 加法更新模块工作流程图

从图 3-6 中可以看到，在没有来自输入模块的输入时，加法更新模块会持续不断将 BRAM 中的数据置零，也就是刷新 BRAM 中的变量 P。

当有来自输入模块的输入时，首先判断输入数据中的神经元编号信息是否会与来自 TDM 控制器的神经元编号发生冲突，判断的标志是这两个神经元编号的差值是否小于 2，小于 2 表示有可能发生冲突。由于 BRAM 的读写时序导致的地址对齐等问题，实际代码中的操作其实比这里描述的要复杂一些，但思想是一致的。

当上述判断的结果是不可能发生冲突时，加法更新模块就开始交替地向 BRAM 模块输送 FIFO_pre 模块与 FIFO_post 模块各自对应通路产生的数据。

而当上述判断的结果是有可能发生冲突时，FIFO_pre 模块对应的通路就暂停工作一段时间，也就是累加更新操作暂停，只进行 FIFO_post 模块对应的置零操作。在 FIFO_pre 模块暂停期间，一个计数器从零开始，每过一个时钟周期就累加一次数值（加 1 操作）。累加到一个阈值（代码中设置为 5）以后归零，然后再次判断是否可能发生冲突。不会发生冲突的话，加法更新模块就重新回到交替输出两个通路的数据的工作模式，直到 FIFO_pre 模块中的数据被全部读出。

以上便是对于加法更新模块的结构与功能的说明。

3.3.5 改进前后的硬件资源消耗对比

下表展示的是分别使用新旧两个循环连接通路模块的 CANN 电路系统的硬件资源消耗情况。

表 3-1 新旧设计的 CANN 核心电路硬件资源消耗对比

设计版本	神经元数量	LUT	FF	DSP	BRAM
旧版	64	21301/277400 7.68%	35477/554800 6.39%	192/2020 9.50%	0/755 0%
旧版	256	84676/277400 30.52%	139920/554800 25.33%	576/2020 28.51%	0/755 0%
新版	512	4166/277400 1.50%	6774/554800 1.22%	82/2020 4.06%	18.5/755 2.45%

在表 3-1 的第一行中， LUT 表示查找表（Look Up Table），它是 FPGA 中的基础逻辑功能结构；FF 表示触发器（Flip Flop）；DSP 表示数字信号处理器（Digital Signal Processor）。表格右侧的数据则表示 CANN 核心电路所消耗的上述各类资源占 FPGA 上可用资源总量的比例。

需要注意的是，旧版设计中的神经元数量原本只有 64 个。为了使得新旧两个版本的资源消耗情况更具有可比性，这里通过修改代码中相关参数的方式，将旧版的神经元数量提升到了 256 个。可以看到，在新的设计下，CANN 核心电路的硬件资源消耗有了极大的降低。

3.4 基于 STP 机制的相关仿真实验

新加入的神经机制使得 CANN 系统可以进行更多种类的仿真任务，下面将展示基于 STP 机制的静默式工作记忆仿真实验和 T 型迷宫仿真实验。

下述的这些参数设置在两个实验中是一致的，而其他的参数则在两个实验中则不一致：ΔT=0.0102 ms，τ_m=8 ms，g_m=1，V_{thr}=0.5，V_{reset}=0 以及 τ_{ref}=2 ms。

3.4.1 静默式工作记忆仿真实验简介

基于新加入的 STP 机制，CANN 系统可以实现不表现为持续性发放的静默式工作记忆，被记忆的信息将被保留在 STP 突触变量 *ux* 的变化中。

具体来说，先选择出系统中的一部分神经，通过刺激系统中的这部分特定的神经元的方式，将信息注入到 CANN 系统中（就像第二章中的工作记忆仿真实验那样）。这些被刺激的神经元由于 STF 机制（即短时程易化机制）的影响，其突触效率相较于系统中的其他未被选中的神经元，将发生显著的提升，也就是它们的突触变量 *ux* 将显著的高于其他神经元的对应变量。然后撤除刺激，系统中的神经元将全都不再发生显著持续性发放活动。

但之前注入到系统中的刺激信号代表的信息其实并没有消失。在刺激信号消失一段较短的时间（1 s 左右）之后，给予系统中所有神经元一个稍强的刺激。这

个刺激的强度需要恰到好处，一方面要足够强，强到可以让之前被选中的神经元再次持续发放；但也不能太强，要让其他未被选中的神经元都无法持续性发放。由于这两组神经元之间存在突触效率的差异，这样的刺激强度是一定存在的。最终的结果就是，先前被注入到系统中的信息，在这个全局性的刺激信号（或者说回忆信号）的作用下，重新以持续性发放的形式在系统中显现。

3.4.2 静默式工作记忆仿真实验参数设置及结果

该实验中，CANN 网络参数被设置为：对于循环连接的非线性的突触（NMDA 型），τ_N=100 ms、A_N=0.65 以及 σ_N= 5.69；对于循环连接的线性的突触（AMPA 型），τ_A=2 ms、A_A=0.1 以及 σ_A= 17.07；对于一对一的外部输入突触，其权重值 W_{E0}= 0.35。每一个 CANN 中的神经元都从一对一的外部突触上接收到一个不相关的泊松发放序列作为背景噪声，平均发放率 ν^{Back}=250 Hz；抑制性突触相关参数为：τ_{Inhb}=5 ms，W_{Inhb}= -0.006；STP 机制的相关参数为：U=0.2，τ_D=150 ms，τ_F=1600 ms。

系统中只有 NMDA 型突触的 STP 机制被启用，AMPA 型突触的 STP 机制未被启用。

输入到 CANN 系统的刺激信号如下图所示：

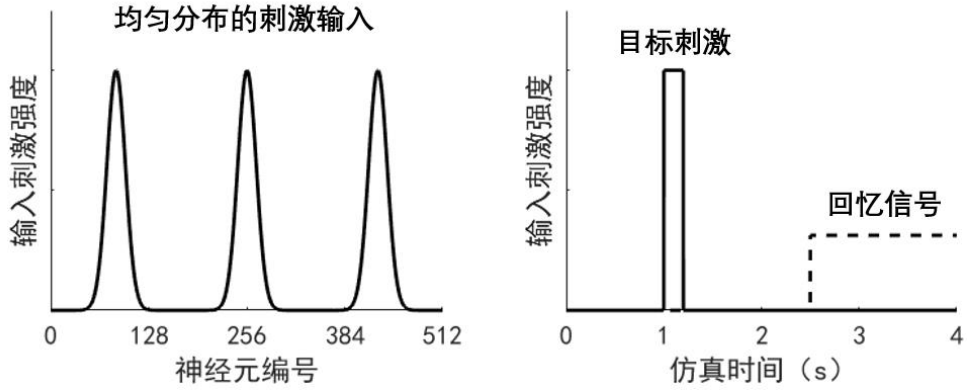

图 3-7 静默式工作记忆刺激协议示意图

其中，目标刺激的设置与第二章中工作记忆的刺激设置类似，同样是使用式(2-15)计算得到。其参数为：目标刺激强度参数 ν^{Cue}=400 Hz，单个目标刺激的宽度参数 σ_{Cue}=2.13。这个目标刺激只在仿真时间进行到 1000 ms 时出现，维持 200 ms 后消失。

在仿真进行到 2500 ms 的时刻，所有神经元都会收到一个回忆信号。该信号和背景噪声一样是各个神经元之间互不相关的泊松发放序列，其平均发放率 ν^{Recall}=125 Hz。

静默式工作记忆的仿真实验结果如下：

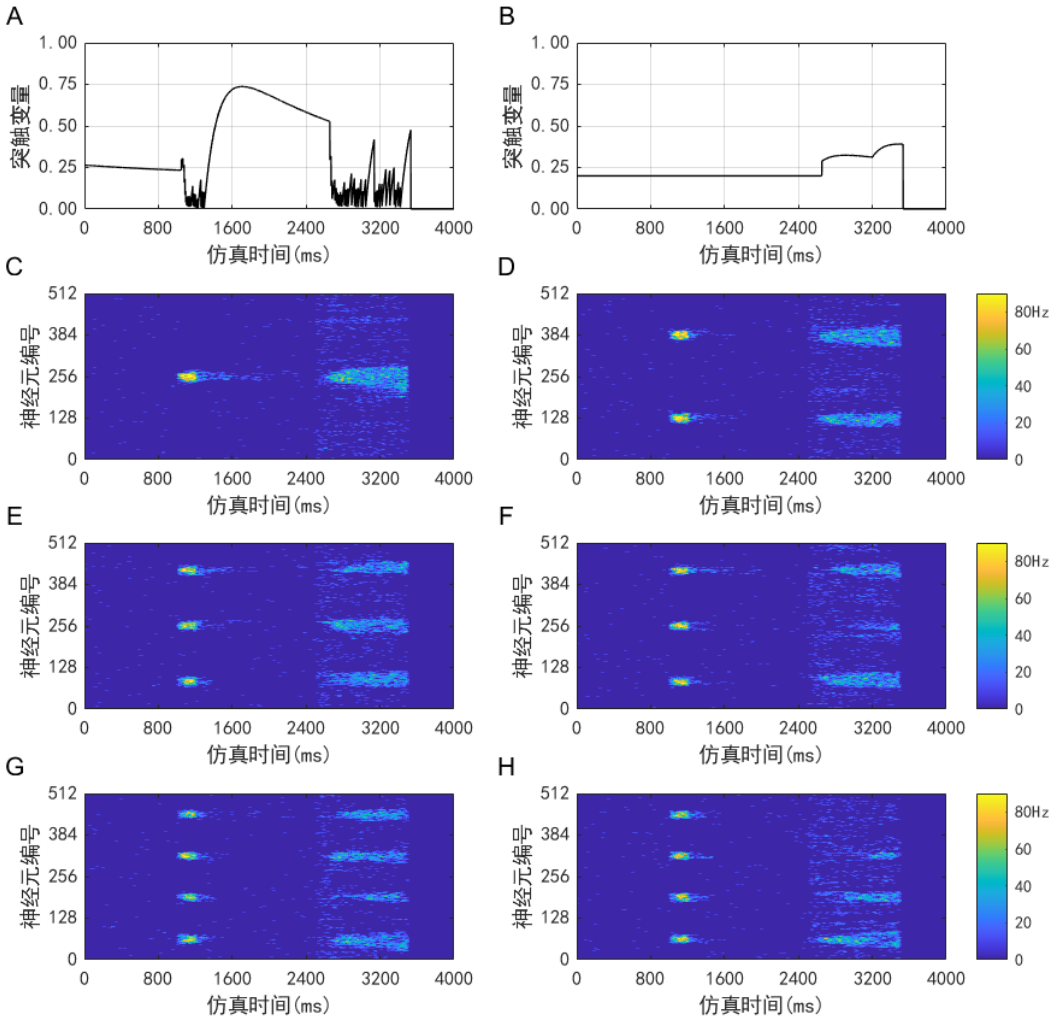

图 3-8 静默式工作记忆仿真结果图

其中，图 A 和图 B 分别是一个被目标信号刺激的神经元对应的 NMDA 型突触的 STP 变量 *ux* 和一个未被目标信号刺激的神经元对应的变量 *ux*。可以看到，正如上一小节描述的那样，在 STP 机制的作用下，被目标信号刺激的神经元的突触效率在刺激消失后显著高于未被目标信号刺激的神经元的突触效率，这个差异是静默式工作记忆仿真实验能够成功的原因。

系统对于输入的一个短暂目标刺激会先表现出较高发放率的快速响应，然后在目标刺激消失后，神经元也就不再较为频繁地发放了。直到系统接收到了回忆信号时，在变量 *ux* 的差异作用下，先前被刺激的那些神经元又会产生频率显著高于其他神经元的发放活动。图 C、图 D、图 E 以及图 G 分别展示了目标数量为 1、2、3 和 4 时，记忆信息成功恢复的例子，是静默式工作记忆的典型形式。而图 F

和图 H 则是记忆信息恢复失败的例子。另外可以看到，和第二章中的工作记忆仿真实验一样，目标数量越多，记忆的信息就越难以维持。

3.4.3 T 型迷宫仿真实验简介

在一项对老鼠进行的动物行为实验中，实验者研究了当老鼠在虚拟导航任务中积累视觉刺激信息时，后顶叶皮层是如何将新信息与正在进行的信息积累活动结合起来的[66]。在这个实验中，一只头部受到限制的老鼠在一个虚拟现实的 T 型迷宫中奔跑。老鼠将被提供 6 个视觉刺激线索，这些刺激信息都出现在迷宫的左侧墙壁或右侧墙壁的固定位置。为了得到奖励，老鼠需要在迷宫中分叉路口的地方选择刺激数量较多的一侧。两侧的视觉刺激数量差异越小，任务难度越大。

上述实验的示意图如下：

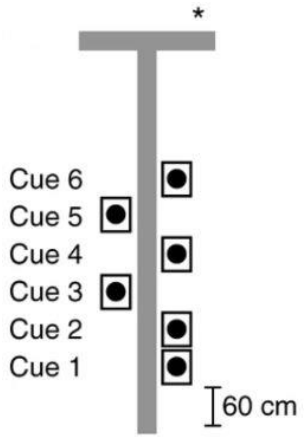

图 3-9 T 型迷宫动物实验示意图[66]

在图 3-9 中，左侧墙壁出现了 2 个刺激，右侧是 4 个刺激，所以被试老鼠需要在分叉路口选择右侧（即图中星号标记的一侧）才能获得奖励。

在这个实验的启发下，在此设计了一个仿真实验来模拟这个动物实验中的信息积累现象，以此来验证具有 STP 机制的 CANN 电路系统的认知计算能力。

3.4.4 T 型迷宫仿真实验参数设置及结果

该实验中，CANN 网络参数被设置为：对于循环连接的非线性的突触（NMDA 型），τ_N=100 ms、A_N=1 以及 σ_N= 17.07；对于一对一的外部输入突触，其权重值 W_{E0}= 0.4。每一个 CANN 中的神经元都从一对一的外部突触上接收到一个不相关的泊松发放序列作为背景噪声，平均发放率 ν^{Back}=150 Hz；抑制性突触相关参数为：τ_{Inhb}=5 ms， W_{Inhb}= -0.03；STP 机制的相关参数为：U=0.1，τ_D=75 ms，τ_F=2000 ms。

这个仿真实验只使用了 NMDA 型循环连接突触，并启用了其 STP 机制的仿真，没有使用到 AMPA 型突触。

T 型迷宫仿真实验的刺激协议如下图所示：

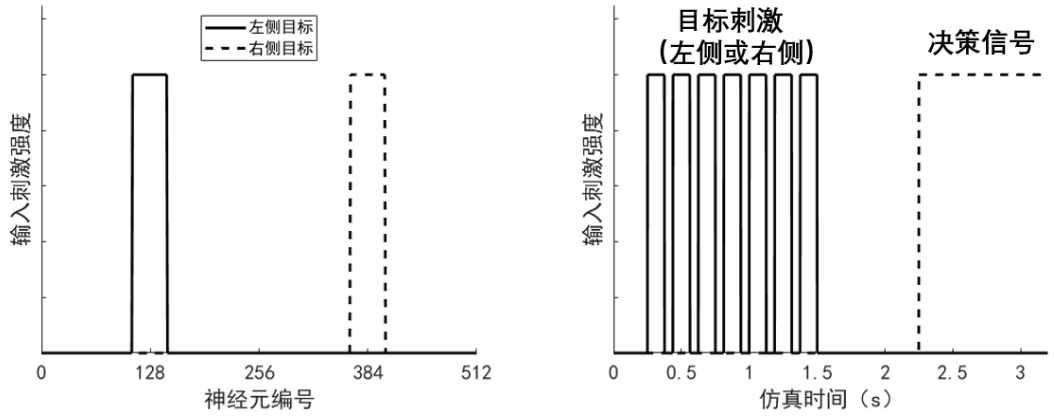

图 3-10 T 型迷宫刺激协议示意图

输入到 CANN 系统的刺激信号包含两个部分：目标刺激信号和决策信号。目标刺激信号分为左侧和右侧两类：左侧目标刺激影响范围是以 128 号神经为中心，半径为 20 个编号内的所有神经元，也就是 108 号到 148 号的所有神经元；类似的，右侧目标刺激的影响范围是 364 号到 404 号的所有神经元。在目标刺激影响范围内的神经元会收到平均发放率 ν^{Tar}=200 Hz 的泊松发放序列。

在仿真开始后 200 ms 后，依次向 CANN 系统输入 7 个目标刺激信号，随机分配各个目标刺激的类型（左侧或右侧）。每个目标刺激持续 100 ms，两个目标刺激之间间隔 50 ms。在最后一个目标刺激消失后，将有一段 600 ms 的延时，用来模拟被试老鼠在视觉刺激结束后向着分叉路口前进的过程。然后，在仿真时间到了 1800 ms 时，输入一个强度为 ν^{Dec}=200 Hz 的泊松发放序列作为决策信号，使得 CANN 系统以某一侧的神经元持续性发放的形式做出决策：在某一侧的神经元的发放率到达了阈值 30 Hz 后，则判定决策结果为这群神经元对应的那一侧。

下图是进行了 1100 轮上述仿真实验后得到的结果：

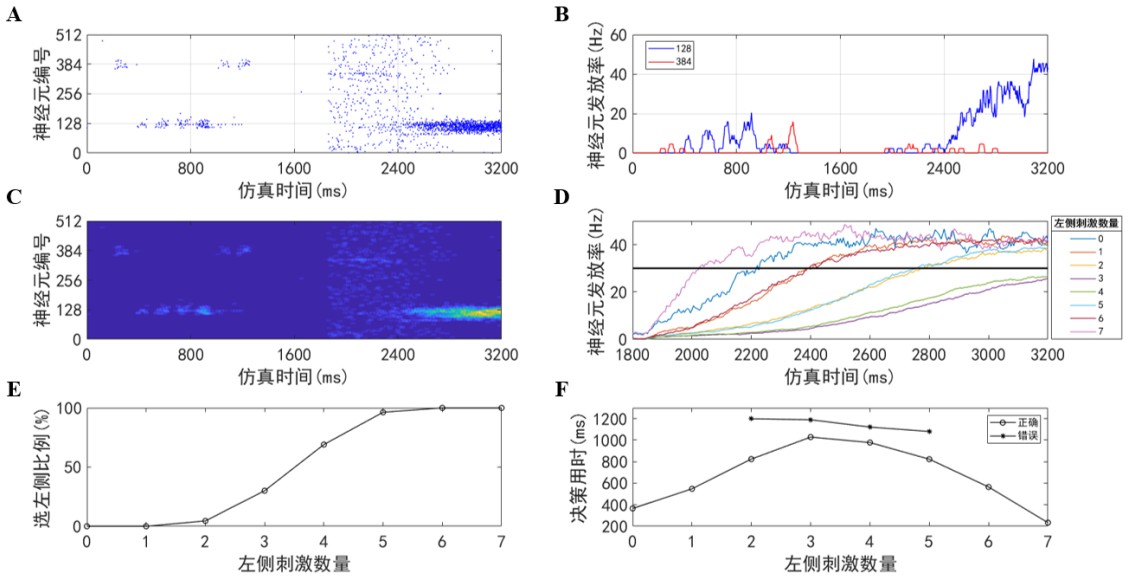

图 3-11 T 型迷宫仿真实验结果图

在每一轮测试中，系统会通过 STP 机制累积目标刺激信息的影响，然后在决策阶段（即仿真时间到了 1800 ms 以后的部分）根据先前积累的信息来做出决策。系统如果在决策阶段选择了左右两侧目标刺激数量较多的那一侧作为答案，那么决策结果就是正确的。图 A、图 B 和图 C 是一轮具有代表性决策结果正确的仿真实验中，神经元发放活动示意图。在此轮实验中，输入到系统的刺激类型依次是：1 次右侧、4 次左侧以及 2 次右侧。而最终系统做出来正确的决策，也就是选择了左侧。

将仿真实验的轮次按照不同的左侧目标刺激数量进行分类后，可以发现任务难度与决策用时的关系。图 D 展示了在上述分类规则下，决策结果正确的轮次中胜出的那一侧神经元的平均发放率。可以看到，左右目标刺激数量差异越大（即任务难度越低），正确决策的用时就越少。

在同样的分类规则下，还可以得到决策正确率与任务难度的关系。图 E 展示了在不同仿真轮次类型下（横坐标），系统决策结果为左侧的实验轮次的占比（纵坐标）。可以看到，左右目标刺激数量差异越大，系统越容易给出正确的决策结果。

还可以进一步对决策结果分别为正确或错误的实验轮次中决策用时的差异。图 F 展示了按照左侧目标刺激数量以及决策结果正确与否这两个分类标准对仿真实验进行分类后，每一类仿真实验的平均决策用时。可以看到，在相同的任务难度下，错误决策的平均用时略高于正确决策的平均用时；而正确决策的平均用时曲线也是与图 D 的数据一致。

上述仿真实验的决策结果数据与动物实验[66]中获得的数据是类似的。下图是动物实验中的决策结果数据：

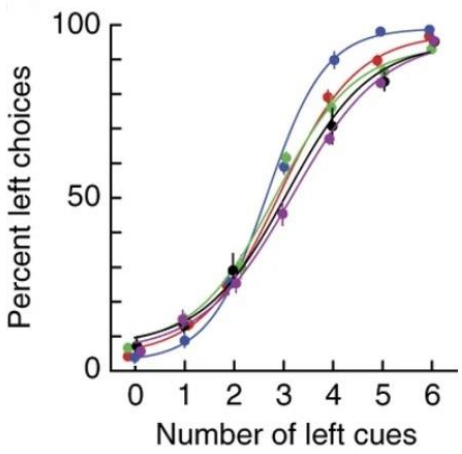

图 3-12 T 型迷宫动物实验决策结果图[66]

上图是 5 只被测试的小鼠在进行了 7 到 12 轮次的实验后得到的决策结果数据，该类型数据与仿真实验中得到的数据图 3-11 中图 E 的变化趋势是一致的。

尽管该动物实验中涉及到的神经机制未必就是由具有 STP 机制的 CANN 系统实现的，但这样的实验结果还是展示出了 CANN 电路系统具有的认知计算能力。

3.5 本章小结

本章主要介绍了在第二章中提出的 CANN 电路系统的基础上，加入了 STP 等新的神经机制的，硬件资源消耗更少的新 CANN 电路系统，并展示了在该系统上进行的静默式工作记忆以及 T 型迷宫仿真实验。

本章首先对新加入是 STP 机制和 SFA 机制进行了介绍，并讲解了这 2 种新的神经机制的数学模型以及电路实现方式。

然后介绍了新的循环连接算法，这个算法针对旧算法在 CANN 系统中的神经元数量增大后，硬件资源消耗过大的问题进行了改进。通过简化 NMDA 型突触变量的计算以及改变计算循环通路突触电流的算法等方式，新算法减少了实现循环通路所需的同时进行的乘法运算数量，从而减少了硬件资源的消耗。

在介绍了新算法的原理后，展示了实现该算法的新循环连接通路模块的结构及其各个子模块的工作原理。其中着重介绍了较为复杂的变量 P 发生器模块，以及该模块的核心子模块，加法更新模块。还展示了新旧算法对应的电路在综合为可在 FPGA 上实际运行的 CANN 核心模块电路后，对于 FPGA 的硬件资源的消耗情况。

最后展示了利用新的 CANN 电路系统进行的两个仿真实验。静默式工作记忆仿真实验初步验证了新加入的 STP 机制的认知计算能力；T 型迷宫仿真实验则是新的 CANN 系统可以进行的一项较为复杂的任务，综合地展示了该系统具有的计算功能。

第四章 具有多个 CANN 核心的电路系统

在第三章的新 CANN 系统的基础上，本章进一步搭建了一个基于片上网络技术（network-on-chip，NoC）的多核心 CANN 系统，该系统为实现更加复杂多样的仿真计算任务奠定了基础。

本章将主要介绍该多核心 CANN 电路系统中新增的两个重要模块，片上网络模块以及数据核心模块。对于片上网络模块，由于该模块是由他人设计实现的，所以对于该模块的介绍并不涉及该模块的结构与工作原理，只讲解该模块的使用方法与功能。对于数据核心模块，将主要介绍与之相关的仿真数据类型的设计以及控制该模块的状态机的设计。然后将会简单介绍 CANN 核心电路的新增模块以及其他的改动。

本章最后还将展示验证该多核心系统基本功能的测试以及测试数据。

4.1 多核心 CANN 系统的结构

多核心 CANN 系统主要由 7 个第三章中介绍的 CANN 核心模块、1 个数据核心模块、1 个片上网络模块以及外围电路构成。其结构示意图如下：

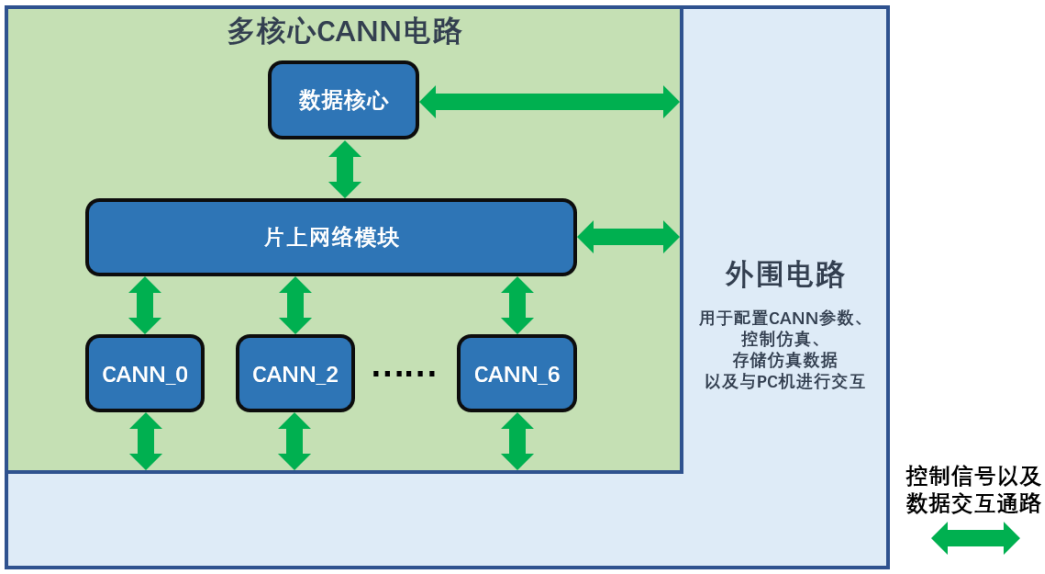

图 4-1 多核心 CANN 电路系统结构示意图

片上网络模块主要负责连接各个 CANN 模块以及数据核心模块，各个 CANN 模块之间的神经元可被设置为有编号偏移量的一对一连接（参见 4.1.3 小节的描述）；数据核心模块则用于接收来自各个 CANN 模块的神经元发放数据以及被监控的神

经变量数据，并把这些被接收到的数据打包，然后传输给外围电路中的 BRAM 模块，等待被 PC 机读取。

下面将分别介绍新加入的片上网络模块以及数据核心模块。

4.1.1 片上网络模块概述

多核心 CANN 系统所使用的片上网络模块是由我的师弟黄衍林同学设计实现的，本人对该模块的内部结构与工作原理并不十分了解，所以在此就只简单介绍一下该模块的使用方法以及它能够实现的功能。

该模块以数据包的方式来传输数据，每一个数据包用一个 last 信号作为结束的标记。一个数据包包含 1 个到任意多个数据，每个数据长度为 32 位。除去包头数据，数据包内的其他数据的 32 位都是有效数据。

数据包包头的数据格式如下表所示：

表 4-1 片上网络模块包头数据格式

数据位	31 至 28 位 （4 位）	27 至 24 位 （4 位）	23 至 16 位 （8 位）	15 至 0 位 （16 位）
数据含义	数据终点的 X 坐标	数据终点的 Y 坐标	私有片段 （不可读写）	有效的 传输数据

包头数据中的 X 坐标与 Y 坐标信息是用于表示该数据包需要被送到片上网络的哪一个位置的信息。对于在多核心 CANN 系统中实际使用的片上网络模块，它在 X 方向上可以容纳 8 个接入模块，在 Y 方向上可以容纳 2 个接入模块，所以一共是 16 个接入模块。7 个 CANN 核心模块分别被连接到了片上网络模块中的 Y 坐标为 0，X 坐标为 0 到 6 的接口上；而数据核心模块则使用了 Y 坐标为 1，X 坐标为 3 和 4 的两个接口；其余接口没有被使用。

与片上网络模块进行交互的主要交互信号如下表所示：

表 4-2 片上网络模块的相关交互信号

信号名称	信号方向 以及位宽	描述
localIfs_y_x_pop_valid	out 1	输出的握手信号
localIfs_y_x_pop_ready	in 1	输出的握手信号
localIfs_y_x_pop_payload_last	out 1	Last 信号
localIfs_y_x_pop_payload	out 32	输出数据
localIfs_y_x_push_valid	in 1	输入的握手信号
localIfs_y_x_push_ready	out 1	输入的握手信号
localIfs_y_x_push_payload_last	in 1	Last 信号

表 4-3 片上网络模块的相关交互信号（续）

信号名称	信号方向 以及位宽	描述
localIfs_y_x_push_payload	in 32	输入数据

在表 4-2 中，信号名称中的 x 和 y 字段用于填写接口的坐标信息，例如 localIf_2_2_pop_ready。每一个接口都可以进行读写操作。

数据写入片上网络模块的时序图如下：

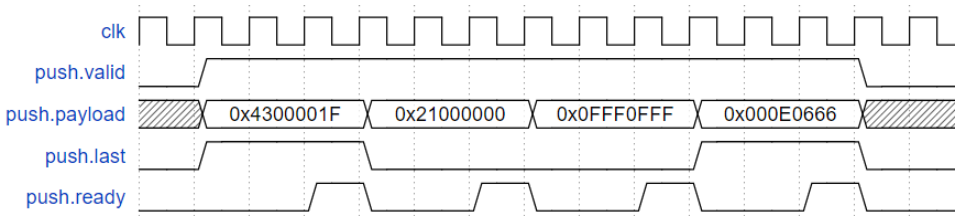

图 4-2 片上网络模块数据写入时序图

在图 4-2 中：第一个数据包的长度为 1，坐标为 x=4，y=3，数据是 16 位的 001F；第二个数据包长度为 3，坐标为 x=2，y=1，包头内的数据为 0，包内的两个数据为 32 位的 0FFF0FFF 和 000E0666。

在实际的电路系统中，使用了长度为 1 的数据包发送神经元发放数据，该数据包含神经元所在 CANN 核心的编号信息以及神经元本身的编号信息，共 12 位；使用长度为 2 的数据包发送被监控的神经变量数据，该数据在数据包的第二个数据中。

4.1.2 数据核心模块

之前提到，数据核心模块的作用是将接收到的仿真数据打包后，传输给外围电路中的 BRAM 模块。在负责数据打包的子模块中，使用了一个状态机来实现预期的功能。接下来首先讲解数据核心的结构，然后再说明数据打包子模块的工作流程。

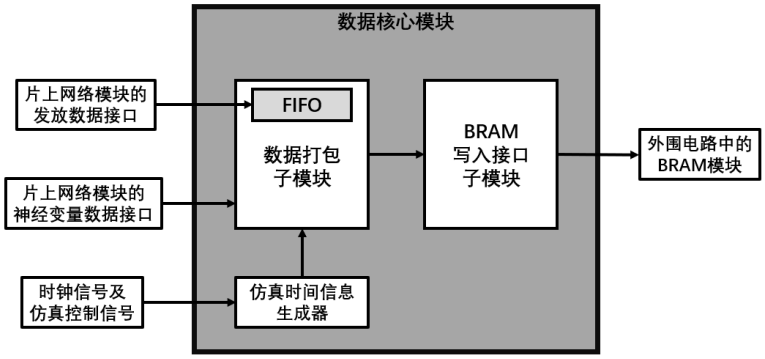

图 4-3 数据核心模块结构示意图

在数据核心模块中，仿真时间信息生成器的作用是给数据打包子模块提供当前的仿真时间信息，数据打包子模块会把该信息与来自片上网络模块的仿真数据信息进行组合，然后输送到 BRAM 写入接口子模块，传输到外围电路中。这里的 BRAM 写入接口子模块与 2.2.5 小节中介绍的具有同样功能的 BRAM 写入模块是工作原理完全相同的模块。可以看到，进入到数据打包子模块的发放数据会先进入一个 FIFO 中，这个 FIFO 的作用是暂时保存输入的发放信息，若在同一个仿真时刻内有多个发放到来，这个 FIFO 就可以与状态机配合，打包并输出发放数据。

在介绍数据打包子模块中的状态机之前，需要先说明一下多核心 CANN 系统的仿真数据格式是如何设计的。在先前的单核心 CANN 系统中，一个神经元发放数据和与之对应的仿真时间信息是包含在同一个 32 位的数据中的（如 2.2.5 小节所述）。由于需要 3 位数据来记录 CANN 核心的编号信息，如果使用与之前系统相同的策略来表示多核心 CANN 系统的发放数据，那么仿真时间信息就只剩下了 20 位。在欧拉法时间步长为 0.0102 ms 的条件下，系统能记录的仿真时间长度最多只有 10.7s，超过这个时间，仿真时间信息就会溢出。

为了让系统具有一定的冗余性，避免上述的溢出情况发生。多核心系统的仿真数据被重新设计为如下表所示的四类数据。

表 4-4 多核心 CANN 系统各个数据类型的数据格式

数据类型	31 至 28 位 （4 位）	27 至 24 位 （4 位）	23 至 12 位 （12 位）	11 至 0 位 （12 位）
单个发放信息 （B1 类）	标志位 （数值为 0xB）	发放数量信息 （数值为 1）	数值为 0	发放信息
两个发放信息 （B2 类）	标志位 （数值为 0xB）	发放数量信息 （数值为 2）	发放信息 1	发放信息 2
仿真时间信息 （A 类）	标志位 （数值为 0xA）	仿真时间信息（27 至 0 位）		
神经变量信息 （C 类）	标志位 （数值为 0xC）	神经变量信息（27 至 0 位）		

每当数据核心接收到发放信息或神经变量信息后，数据打包子模块首先从仿真时间信息生成器中读取一个仿真时间信息，将该信息组装成为表 4-4 中的 A 类数据的格式后，发送给 BRAM 写入模块，然后再处理接收到的仿真数据。这样的设计就可以避免仿真时间信息的溢出，每一个仿真数据的前面都是与之对应的仿真时间信息。

完整的数据打包状态机的状态转移方式描述如下：

- S0：IDLE 状态，等待仿真数据的到来。无论什么仿真数据到来时，状态都跳转到 S1。
- S1：读取仿真时间信息，并将该信息打包后发送给 BRAM 写入模块，然后状态跳转到 S2。
- S2：判断读取到的仿真数据是发放数据还是神经变量数据。若是发放数据，则状态跳转到 S4；若是神经变量数据，则状态停留在 S2，并等待从片上网络模块中读取有效的神经变量数据（神经变量数据包长度为 2），数据读取完成后状态跳转到 S3。
- S3：打包接收到的神经变量数据，并将其发送给 BRAM 写入模块，然后状态跳转到 S4。
- S4：进入这个状态后，先检查 FIFO 中保存的发放信息的数量：若没有发放信息，状态跳转到 S0；若有 2 个及以上数量的发放信息，则读取一个发放信息，并将这个数据保存下来，等待后续使用，状态跳转到 S5；若只有 1 个发放信息，则等待仿真时间信息发生跳变，若时间信息发生了跳变，则状态跳转到 S7，否则状态保持在 S4。
- S5：读取 FIFO 中的一个发放信息，然后状态跳转到 S6。
- S6：将在状态 S4 中读取保存下来的发放信息和在状态 S5 中读取出来的发放信息打包为一个包含有两个发放信息的 B2 类数据，将这个数据发送给 BRAM 写入模块，然后状态跳转到 S4。
- S7：将在状态 S4 中读取保存下来的发放信息打包为一个包含有 1 个发放信息的 B1 类数据，将这个数据发送给 BRAM 写入模块，然后判断是否有新的仿真数据输入，若有，状态跳转到 S1，否则状态跳转到 S0。

上述状态机的状态转移图如下所示：

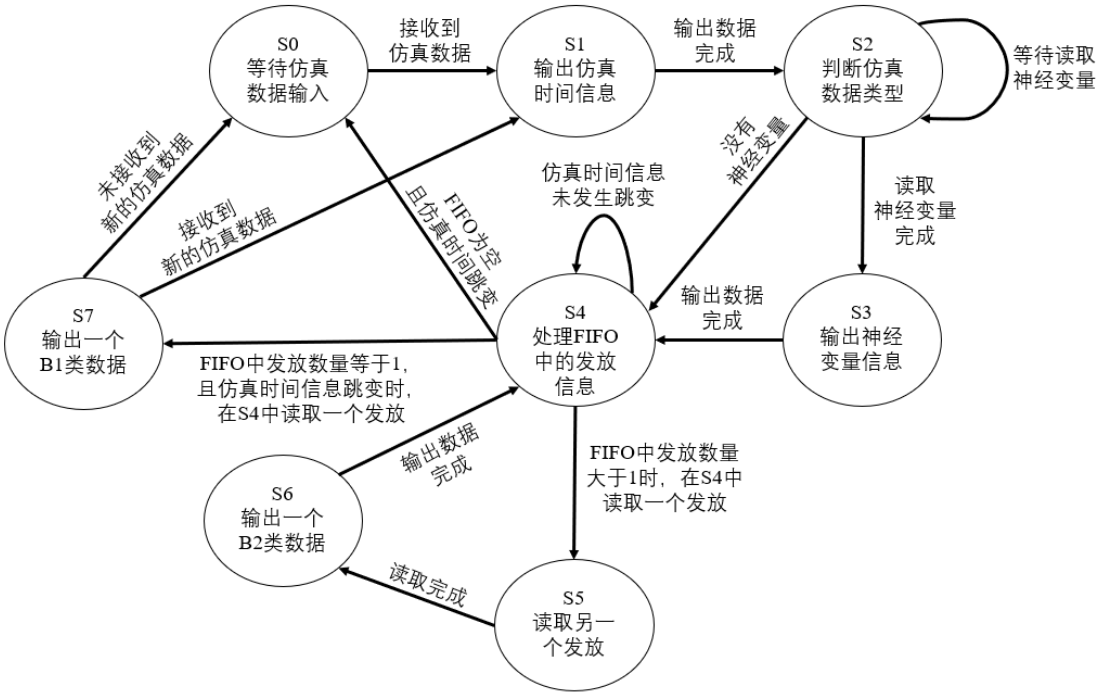

图 4-4 数据打包模块的状态转移图

4.1.3 CANN 核心模块概述

多核心 CANN 系统中的每一个 CANN 核心模块都包含一个第三章中的 CANN 核心电路，除此以外，还含有一些原本属于外围电路的模块，如脉冲生成模块和神经变量监控模块等，以及新加入的用于接收来自其他 CANN 核心的神经元发放刺激的发放接收器模块，和用于向其他 CANN 核心以及数据核心发送数据的数据发送器模块。

由于新加入的两个模块（发放接收器和数据发送器）都是对片上网络模块的接口进行对接的模块，功能并不复杂，就不再对这两个模块进行很详细的说明了。这里只说明两个值得注意的点：

首先，数据发送器模块在发送神经元的发放数据时，需要向数据核心以及与本 CANN 核心有连接的其他 CANN 核心发送发放数据，这个功能使用了一个较为简单的状态机来实现。

其次，由于各个 CANN 核心之间的神经元的连接（如果有的话）是一对一的有偏移量的连接，所以输送到其他 CANN 核心的发放数据需要再加上一个可配置的偏移量。例如，如果偏移量为 100，则发送核心的 30 号神经元就连接到目标核心中的 130 号神经元；而发送核心的 500 号神经元则是连接到目标核心中的 88 号神经元，这一点可以参考图 2-1 中所示的结构，所有神经元是分布在一个圆上的，编号偏移量的作用只是在转动这个圆。

4.2 验证多核心 CANN 系统基本功能的测试

为了验证各个 CANN 核心之间的一对一连接功能是否正常，本实验测试了从 0 号核心到其余所有核心的连接，以及 6 号核心到其余所有核心的连接。

由于从编号较小的核心连接到编号较大的核心的代码与相反方向（编号从大到小）的连接的代码有一定的差异，而且各个 CANN 核心的代码都是一致的，不同核心之间只是编号参数不一样，所以上述的这两个测试就可以验证所有的连接功能是否可以正常工作了。

下列这些参数设置对于本测试中的所有核心都是一样的：

ΔT=0.0102 ms，τ_m=8 ms，g_m=1，V_{thr}=0.5，V_{reset}=0，τ_{ref}=2 ms，τ_A=2 ms，τ_N=100 ms，$A_A = A_N$ =0.1，$\sigma_A = \sigma_N$=17.07，τ_{Inhb}=2 ms 以及 W_{Inhb}= -0.005。

对于作为起点的 CANN 核心，其接收发放发生器的输入的外部突触参数为：W_{E0}= 0.4 以及 τ_{E0}=2 ms。而作为终点的 CANN 核心，其接收来自其他核心的输入的外部突触参数为：W_{E1}= 0.6 以及 τ_{E0}=10 ms。

输入起点 CANN 核心的刺激信号只包含一个目标刺激，如下图所示：

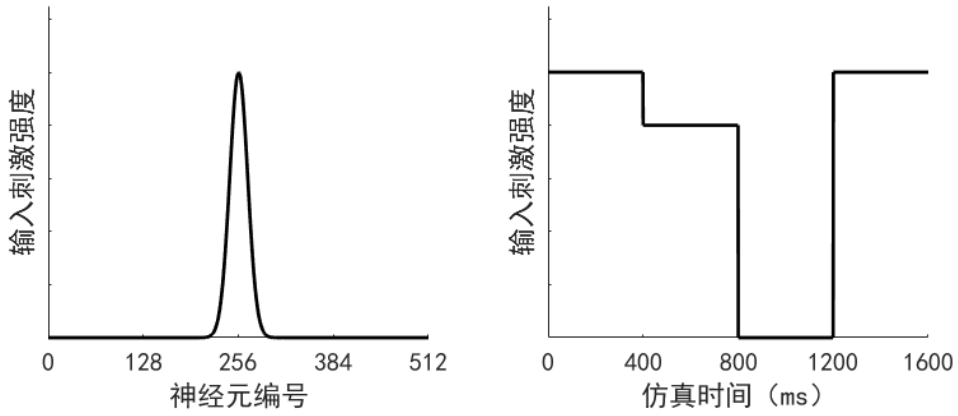

图 4-5 多核心系统测试刺激协议示意图

目标刺激信号的计算方式如式(2-12)所示，该信号只包含一个目标刺激，即 N_{Cue}=1。目标刺激强度参数 v^{Cue} 在 0 ms 到 400 ms 之间为 500 Hz；在 400 ms 到 800 ms 之间为 400 Hz；在 800 ms 到 1200 ms 之间为 0 Hz；在 1200 ms 到 1600 ms 之间为 500 Hz。

作为终点的 CANN 核心则只接收来自起点的 CANN 核心的输入。

下图是测试的结果：

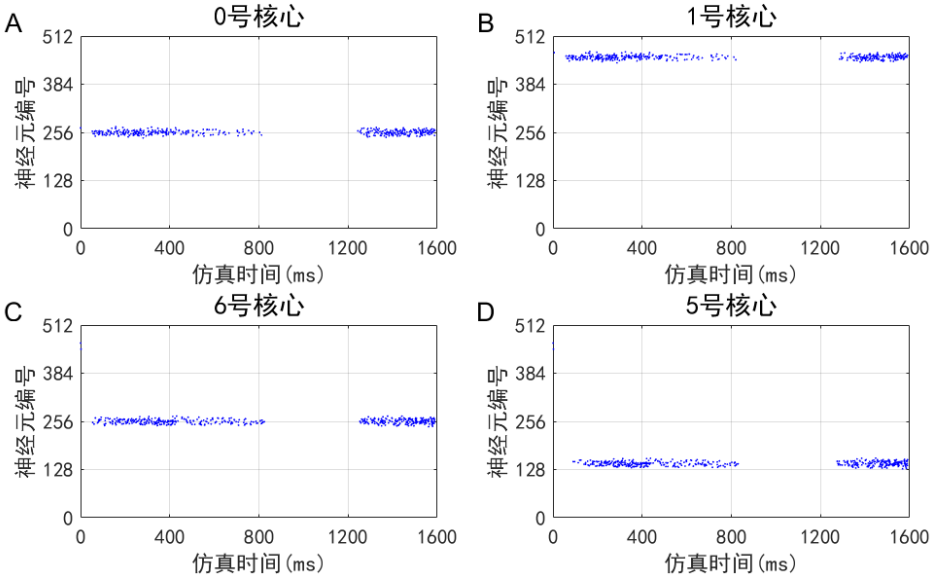

图 4-6 多核心系统测试结果图

图 A 和图 B 是从编号较小的核心向编号较大的核心发送发放的测试结果：0 号核心是起点，其余所有核心是终点。由于各个终点核心的参数设置是一致的，它们的发放模式也是完全一致的，所以就只展示了 1 号核心的数据。这个测试中，神经元编号偏移量为 200，所以 1 号核心中响应的神经元群在 456 号神经元附近。

图 C 和图 D 是相反方向的测试结果，出于与第一个测试相同的原因，同样只展示了作为终点的 5 号核心的数据。这个测试中，神经元编号偏移量为 400，所以 5 号核心中响应的神经元群在 144 号神经元附近。

上述的测试结果说明多核心 CANN 系统的各个核心之间的连接是符合预期设计目标的。

4.3 本章小结

本章主要介绍了多核心 CANN 电路系统的电路结构，以及在该系统中新增的两个主要子模块，片上网络模块和数据核心模块。利用片上网络模块， 7 个 CANN 核心模块以及数据核心模块实现了各模块之间的数据交互；而数据核心模块则整合了所有 7 个 CANN 核心电路的仿真数据，简化了外围电路读取仿真数据的操作。本章最后还展示了证明各个 CANN 核心之间的连接正常的测试数据。

第五章 总结与展望

5.1 研究总结

神经形态电路是一项被学者们认为具有广阔应用前景的前沿技术。基于该项技术，研究者可以实现对于神经系统的高效仿真模拟，从而促进研究者对于神经系统的相关模型的研究工作；同时该技术也让那些从神经系统中发现的优秀计算特性在投入到其他潜在的应用领域之前，可以进行充分的理论和工程技术细节完善。

本论文主要介绍了一项在 FPGA 上实现、基于 CANN 模型的、具有丰富仿生特性的神经形态电路系统的研究工作。文章内容按照研究进行的三个阶段进行划分为：一是初步搭建小规模 CANN 电路系统原型；二是搭建加入了 STP 等特性的硬件资源消耗较少的 CANN 电路系统；三是搭建具有多个 CANN 核心模块的电路系统。这三个阶段的研究重点各有不同：

第一阶段的重点在于建立起 CANN 核心电路系统的大致框架，这个阶段的设计并不追求资源利用率上的最优化以及系统功能的丰富程度。通过这个阶段的电路设计以及仿真实验，成功验证了诸如时分复用技术、CANN 模型的动力学特性、非线性突触特性、以及用于与 PC 机进行控制信号与数据交互的外围电路等理论与技术要点。

第二阶段的重点则是，在第一阶段实现了的 CANN 电路原型的基础上，加入更多神经特性的仿真功能，进行电路设计优化以减少硬件资源的消耗，从而实现一个电路设计较为成熟且计算功能丰富的 CANN 核心电路系统。这个阶段的系统由于 STP 机制的加入，已经可以进行对 T 型迷宫实验这样较为复杂的认知任务的模拟仿真。并且因为对循环连接算法及其电路实现的优化设计，同等参数规模的新电路系统的资源消耗量占 FPGA 可用资源的比例，相较于第一阶段的设计下降了一个数量级。

第三阶段则着重于搭建一个基于片上网络技术的具有多个 CANN 核心的电路系统，第二阶段中的电路资源优化设计是能够实现第三阶段的系统的前提。尽管现在并未在多核心系统进行认知任务仿真实验，但理论上多核心系统将可以实现更加复杂多样的任务。

通过这三个阶段的研究，最终实现了一个基于 CANN 模型的功能十分丰富且使用方便的神经形态电路系统。

5.2 后续研究展望

后续的研究至少可以在以下三个方面进行：

首先，可以在多核心 CANN 系统的基础上，利用该电路系统对多个 CANN 之间发生的相互作用进行仿真模拟，探索这类系统进行信息整合处理的规律。也可以结合其他对于具有多个 CANN 的模型的理论研究成果，尝试进行与社会生产实践活动较为贴切的应用型研究。

其次，可以进一步优化 CANN 核心电路的设计。在当前的设计下，系统中的所有配置参数、系统变量和仿真数据都是存储在 FPGA 芯片上的 BRAM 模块中的，但 BRAM 资源是十分有限的，这最终会限制可以系统模拟实现的 CANN 模型的规模。所以，可以通过利用 FPGA 开发板上的片外存储资源，并进行必要的算法以及电路设计优化，实现对具有更多神经元以及更多神经特性的 CANN 模型的模拟。

最后，可以尝试设计功能更加强大的神经变量监控模块（该模块在外围电路中）。在当前的设计下，该模块在每一次仿真实验中只能读取一个特定神经元的一个神经变量，这会使得该电路系统无法进行那些需要同时记录多个神经变量的仿真实验。由于在当前设计下，开发板与 PC 机之间的数据交互速率相较于第一阶段的设计有了极大的提升（从使用 UART 串口改为使用网口进行通信），所以理论上，系统已经具备对很多个神经变量进行读取并及时将这些数据传输到 PC 机的条件。

致　谢

时光飞逝，三年的研究生生涯即将告一段落，我将要离开多年来培养教育我的母校电子科技大学，然后踏入社会。回望在电子科大的求学经历，我很庆幸我能够遇到在学习和科研上给予了我重要帮助的老师和同学们，我也十分感激一直以来关心并支持我的家人和朋友们。在此，我向你们表达我最真挚的感谢。

首先我要感谢我的导师游宏志老师，在游老师的指导下，我从本科刚毕业时对于电路设计没有多少经验的状态中快速地成长起来，专业能力有了极大的提升。老师和我一起研究解决科研过程中遇到的难题的场景我依然记忆犹新，我由衷地感谢游老师的教导，我也很庆幸能够成为游老师的学生，谢谢老师的栽培。

我也要感谢教研室的李永杰老师、杨开富老师以及张显石老师，感谢各位老师对我读研期间的科研的支持和帮助。

然后我要感谢教研室的曾广、宋健、张隽睿、李志丹同学，黄衍林、程成师弟，梁思琴师妹，以及我的室友陈安成、何亚辉和黄惠泉，感谢各位在我的科研以及生活中给予我的关心和帮助。同时也感谢与我相识多年的朋友：冯宵林、姚雪梅以及马红跃，感谢你们对我的关心。

最后我要感谢我的父母对我的鼓励与支持，让我能够顺利地完成学业。

参考文献

[1] Steve, Furber. Large-scale neuromorphic computing systems[J]. Journal of Neural Engineering, 2016, 13(5): 51001-51001.

[2] Indiveri G, Chicca E, Douglas R J. Artificial Cognitive Systems: From VLSI Networks of Spiking Neurons to Neuromorphic Cognition[J]. Cognitive Computation, 2009, 1(2): 119-127.

[3] Zenke F, Bohté S, Clopath C, et al. Visualizing a joint future of neuroscience and neuromorphic engineering[J]. Neuron, 2021, 109(4): 571-575.

[4] Li J, Huang Q, Han Q, et al. Temporally coherent perturbation of neural dynamics during retention alters human multi-item working memory[J]. Prog Neurobiol, 2021, 201: 102023.

[5] Pei J, Deng L, Song S, et al. Towards artificial general intelligence with hybrid Tianjic chip architecture[J]. Nature, 2019, 572(7767): 106.

[6] Brandli C, Muller L, Delbruck T. Real-time, high-speed video decompression using a frame- and event-based DAVIS sensor[C]. 2014 IEEE International Symposium on Circuits and Systems (ISCAS), 2014: 686-689.

[7] Yang M, Chien C H, Delbruck T, et al. A 0.5V 55 uW 64X2-channel binaural silicon cochlea for event-driven stereo-audio sensing[J]. IEEE Journal of Solid-State Circuits, 2016: 1-16.

[8] Ning Q, Mostafa H, Corradi F, et al. A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses[J]. Frontiers in Neuroscience, 9.

[9] Indiveri G, Corradi F, Ning Q. Neuromorphic architectures for spiking deep neural networks[C]. 2015 IEEE International Electron Devices Meeting (IEDM), 2016.

[10] Scholze S, Eisenreich H, Hppner S, et al. A 32GBit/s communication SoC for a waferscale neuromorphic system[J]. Integration, 2012, 45(1): 61-75.

[11] Schemmel J, Brüderle D, Grübl A, et al. AWafer-Scale Neuromorphic Hardware System for Large-Scale Neural Modeling[C]. Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, 2010.

[12] Merolla P A, Arthur J V, Alvarez-Icaza R, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface[J]. Science, 2014, 345(6197): 668-673.

[13] Davies M, Srinivasa N, Lin T H, et al. Loihi: A Neuromorphic Manycore Processor with On-Chip Learning[J]. IEEE Micro, 2018: 82-99.

[14] Shen J, Ma D, Gu Z, et al. Darwin: a neuromorphic hardware co-processor based on Spiking Neural Networks[J]. Sciece China. Information Sciences, 2015, 59(2): 1-5.

[15] Ko H, Hofer S B, Pichler B, et al. Functional specificity of local synaptic connections in neocortical networks[J]. Nature, 2011, 473(7345): 87-91.

[16] Bosking W H, Zhang Y, Schofield B, et al. Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex[J]. The Journal of Neuroscience, 1997, 17(6): 2112-2127.

[17] Albright T D, Desimone R, Gross C G. Columnar organization of directionally selective cells in visual area MT of the macaque[J]. Journal of Neurophysiology, 1984, 51(1): 16-31.

[18] Kable J W, Glimcher P W. The Neurobiology of Decision: Consensus and Controversy[J]. Neuron, 2009, 63(6): 733-745.

[19] Wimmer K, Nykamp D Q, Constantinidis C, et al. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory[J]. Nature Neuroscience, 2014, 17(3): 431-439.

[20] Knierim J J, Zhang K. Attractor Dynamics of Spatially Correlated Neural Activity in the Limbic System[J]. Annual Review of Neuroscience, 2012, 35(1): 267-285.

[21] Kim S S, Rouault H, Druckmann S, et al. Ring attractor dynamics in the Drosophila central brain[J]. Science, 2017, 356(6340): 849-853.

[22] Kreiser R, Aathmani D, Qiao N, et al. Organizing Sequential Memory in a Neuromorphic Device Using Dynamic Neural Fields[J]. Frontiers in Neuroscience, 2018, 12.

[23] Aamir S A, Stradmann Y, Müller P, et al. An Accelerated LIF Neuronal Network Array for a Large-Scale Mixed-Signal Neuromorphic Architecture[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2018, 65(12): 4299-4312.

[24] Kreiser R, Renner A, Leite V R C, et al. An On-chip Spiking Neural Network for Estimation of the Head Pose of the iCub Robot[J]. Frontiers in Neuroscience, 2020, 14.

[25] Zhang W H, Wu S. Reciprocally Coupled Local Estimators Implement Bayesian Information Integration Distributively[J]. Advances in Neural Information Processing Systems, 2013: 19-27.

[26] Wu S, Wong K Y M, Fung C C A, et al. Continuous Attractor Neural Networks: Candidate of a Canonical Model for Neural Information Representation[J]. F1000Research, 2016, 5: F1000 Faculty Rev-156.

[27] Hopfield J J. Neural networks and physical systems with emergent collective computational abilities[J]. Proceedings of the National Academy of Sciences of the United States of America, 1982, 79(8): 2554-2558.

[28] Wu S, Hamaguchi K, Amari S-I. Dynamics and Computation of Continuous Attractors[J]. Neural Computation, 2008, 20(4): 994-1025.

[29] Fung C C A, Wong K Y M, Wu S. A Moving Bump in a Continuous Manifold: A Comprehensive Study of the Tracking Dynamics of Continuous Attractor Neural Networks[J]. Neural Computation, 2010, 22(3): 752-792.

[30] Wei Z, Wang X-J, Wang D-H. From Distributed Resources to Limited Slots in Multiple-Item Working Memory: A Spiking Network Model with Normalization[J]. The Journal of Neuroscience, 2012, 32(33): 11228-11240.

[31] Compte A, Brunel N, Goldman-Rakic P S, et al. Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model[J]. Cerebral Cortex, 2000, 10(9): 910-923.

[32] Furman M, Wang X-J. Similarity Effect and Optimal Control of Multiple-Choice Decision Making[J]. Neuron, 2008, 60(6): 1153-1168.

[33] Klausberger T, Somogyi P. Neuronal Diversity and Temporal Dynamics: The Unity of Hippocampal Circuit Operations[J]. Science, 2008, 321(5885): 53-57.

[34] Markram H, Toledo-Rodriguez M, Wang Y, et al. Interneurons of the neocortical inhibitory system[J]. Nature Reviews Neuroscience, 2004, 5(10): 793-807.

[35] Destexhe A, Mainen Z F, Sejnowski T J. Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism[J]. Journal of Computational Neuroscience, 1994, 1(3): 195-230.

[36] Wang X-J. Synaptic Basis of Cortical Persistent Activity: the Importance of NMDA Receptors to Working Memory[J]. The Journal of Neuroscience, 1999, 19(21): 9587-9603.

[37] Gerstner W, Kistler W M, Naud R, et al. Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition[M]. Cambridge: Cambridge University Press, 2014.

[38] Wong K-F, Wang X-J. A Recurrent Network Mechanism of Time Integration in Perceptual Decisions[J]. The Journal of Neuroscience, 2006, 26(4): 1314-1328.

[39] Wang X-J. Probabilistic Decision Making by Slow Reverberation in Cortical Circuits[J]. Neuron, 2002, 36(5): 955-968.

[40] Lisman J E, Fellous J-M, Wang X-J. A role for NMDA-receptor channels in working memory[J]. Nature Neuroscience, 1998, 1(4): 273-275.

[41] You H, Meng Y, Huan D, et al. The neural dynamics for hysteresis in visual perception[J]. Neurocomputing, 2011, 74(17): 3502-3508.

[42] Wang X-J. Macroscopic gradients of synaptic excitation and inhibition in the neocortex[J]. Nature Reviews Neuroscience, 2020, 21(3): 169-178.

[43] Saïghi S, Bornat Y, Tomas J, et al. A Library of Analog Operators Based on the Hodgkin-Huxley Formalism for the Design of Tunable, Real-Time, Silicon Neurons[J]. IEEE Transactions on Biomedical Circuits and Systems, 2011, 5(1): 3-19.

[44] Mizoguchi N, Nagamatsu Y, Aihara K, et al. A two-variable silicon neuron circuit based on the Izhikevich model[J]. Artificial Life and Robotics, 2011, 16(3): 383-388.

[45] Takemoto T, Kohno T, Aihara K. CIRCUIT IMPLEMENTATION AND DYNAMICS OF A TWO-DIMENSIONAL MOSFET NEURON MODEL[J]. International Journal of Bifurcation and Chaos, 2007, 17(02): 459-508.

[46] Tuckwell H C. Introduction to Theoretical Neurobiology: Volume 2: Nonlinear and Stochastic Theories[M]. 2. Cambridge: Cambridge University Press, 1988.

[47] You H, Zhao K. Neuromorphic Implementation of a Continuous Attractor Neural Network With Various Synaptic Dynamics[J]. IEEE Access, 2021, 9: 109224-109240.

[48] Shadlen M N, Newsome W T. Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey[J]. Journal of Neurophysiology, 2001, 86(4): 1916-1936.

[49] Shadlen M, Kiani R. Decision making as a window on cognition[J]. Neuron, 2013, 80(3): 791-806.

[50] Churchland A K, Kiani R, Shadlen M N. Decision-making with multiple alternatives[J]. Nature Neuroscience, 2008, 11(6): 693-702.

[51] You H, Wang D. Dynamics of Multiple-Choice Decision Making[J]. Neural Computation, 2013, 25(8): 2108-2145.

[52] Funahashi S, Bruce C J, Goldman-Rakic P S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex[J]. Journal of Neurophysiology, 1989, 61(2): 331-349.

[53] Wang X-J. Synaptic reverberation underlying mnemonic persistent activity[J]. Trends in Neurosciences, 2001, 24(8): 455-463.

[54] Edin F, Klingberg T, Johansson P, et al. Mechanism for top-down control of working memory capacity[J]. Proceedings of the National Academy of Sciences, 2009, 106(16): 6802-6807.

[55] Abbott, L. F, Varela, et al. Synaptic depression and cortical gain control[J]. Science, 1997.

[56] Cook D L, Schwindt P C, Grande L A, et al. Synaptic depression in the localization of sound[J]. Nature, 2003.

[57] Romani S, Amit D J, Mongillo G. Mean-field analysis of selective persistent activity in presence of short-term synaptic depression[J]. Journal of Computational Neuroscience, 2006, 20(2): 201-217.

[58] Bibitchkov D, Herrmann J M, Geisel T. Effects of short-time plasticity on the associative memory[J]. Neurocomputing, 2002, 44: 329-335.

[59] Mejias, Jorge, F., et al. Maximum Memory Capacity on Neural Networks with Short-Term Synaptic Depression and Facilitation[J]. Neural Computation, 2009.

[60] Mongillo G, Barak O, Tsodyks M. Synaptic Theory of Working Memory[J]. Science, 2008, 319(5869): 1543-1546.

[61] Gutkin B, Zeldenrust F. Spike frequency adaptation[J]. Scholarpedia, 2014, 9: 30643.

[62] Monticelli G. Adaptation in Helix pomatia neurons[J]. Comparative Biochemistry and Physiology Part A: Physiology, 1987, 88(1): 119-126.

[63] Kim C-H, Shin J J, Kim J, et al. Reduced spike frequency adaptation in Purkinje cells of the vestibulocerebellum[J]. Neuroscience Letters, 2013, 535: 45-50.

[64] Benda J, Hennig R M. Spike-frequency adaptation generates intensity invariance in a primary auditory interneuron[J]. Journal of Computational Neuroscience, 2008, 24(2): 113-136.

[65] Fuhrmann G, Markram H, Tsodyks M. Spike frequency adaptation and neocortical rhythms[J]. J Neurophysiol, 2002, 88(2): 761-70.

[66] Morcos A S, Harvey C D. History-dependent variability in population dynamics during evidence accumulation in cortex[J]. Nature Neuroscience, 2016, 19(12): 1672-1681.

攻读硕士学位期间取得的成果

[1] You H, **Zhao K**. Neuromorphic Implementation of a Continuous Attractor Neural Network With Various Synaptic Dynamics[J]. IEEE Access, 2021, 9: 109224-109240.