

# 大规模类脑计算系统 BiCoSS: 架构、实现及应用

杨双鸣<sup>1</sup> 郝新宇<sup>1</sup> 王江<sup>1</sup> 李会艳<sup>2</sup> 魏熙乐<sup>1</sup> 于海涛<sup>1</sup> 邓斌<sup>1</sup>

**摘要** 人脑具有协同多种认知功能的能力与极强的自主学习能力, 随着脑与神经科学的快速发展, 亟需计算结构模拟人脑的、性能更强大的计算平台进行人脑智能与认知行为机制的进一步探索. 受人脑神经机制的启发, 本文提出了基于神经认知计算架构的众核类脑计算系统 BiCoSS, 该系统以并行计算的现场可编程门阵列 (Field-programmable gate array, FPGA) 为核心处理器, 以地址事件表达的神经放电作为信息传递载体, 以具有认知计算功能的神经元作为信息处理单元, 实现了四百万神经元数量级大规模神经元网络认知行为的实时计算, 填补了从细胞动力学层面理解人脑认知功能的鸿沟. 实验结果从计算能力、计算效率、功耗、通信效率、可扩展性等方面显示了 BiCoSS 系统的优越性能. BiCoSS 通过人脑信息处理的计算架构以更贴近神经科学本质的模式实现了类脑智能; 同时, BiCoSS 为神经认知和类脑计算的研究和应用提供了新的有效手段.

**关键词** 类脑计算, 脉冲神经网络, 神经形态硬件, 类脑智能

**引用格式** 杨双鸣, 郝新宇, 王江, 李会艳, 魏熙乐, 于海涛, 邓斌. 大规模类脑计算系统 BiCoSS: 架构、实现及应用. 自动化学报, 2021, 47(9): 2154–2169

**DOI** 10.16383/j.aas.c190035

## Large-scale Brain-inspired Computing System BiCoSS: Its Architecture, Implementation and Application

YANG Shuang-Ming<sup>1</sup> HAO Xin-Yu<sup>1</sup> WANG Jiang<sup>1</sup> LI Hui-Yan<sup>2</sup> WEI Xi-Le<sup>1</sup> YU Hai-Tao<sup>1</sup> Deng Bin<sup>1</sup>

**Abstract** Human brain has the ability of integrating multiple cognitive functions and strong autonomous learning capability. With the rapid development of neuroscience, it is important and necessary to implement a brain simulation platform with higher performance that is inspired by brain structure to further explore the brain intelligent and mechanism of cognitive behaviors. Inspired by the mechanism of human neural system, a multi-core brain simulation system BiCoSS is presented in this paper, which is based on neurocognitive structure. The presented system uses parallel computing field-programmable gate array (FPGA) as core processor, address event representation (AER) neural spikes as carrier of information transmission, neuron with cognitive computing functions as information processing unit. It realizes the real-time computing of cognitive behaviors in a large-scale neural network with four million neurons, and bridges the gap from the cellular dynamics level to comprehend the human brain cognition functions. The superior performance of BiCoSS is shown in terms of computing power, computing efficiency, power consumption, communication efficiency and scalability. BiCoSS realizes brain-like intelligence based on the computing architecture of brain information processing that is closer to the essence of neuroscience, and provides new effective methods for the research and application of neural cognition and brain-like computing.

**Key words** Brain-like computing, spiking neural network, neuromorphic engineering, brain-like intelligence

**Citation** Yang Shuang-Ming, Hao Xin-Yu, Wang Jiang, Li Hui-Yan, Wei Xi-Le, Yu Hai-Tao, Deng Bin. Large-scale brain-inspired computing system BiCoSS: Its architecture, implementation and application. *Acta Automatica Sinica*, 2021, 47(9): 2154–2169

收稿日期 2019-01-14 录用日期 2019-06-06

Manuscript received January 14, 2019; accepted June 6, 2019

国家自然科学基金 (61871287, 61671320, 61601320, 61771330, 62071324, 62006170), 中国博士后科学基金 (2021T140510, 2020M680885), 天津市自然科学基金 (18JCZDJC32000) 资助

Supported by National Natural Science Foundation of China (61871287, 61671320, 61601320, 61771330, 62071324, 62006170), Cina Postdoctoral Science Foundation (2021T140510, 2020M680885), and Natural Science Foundation of Tianjin, China (18JCZDJC32000)

本文责任编辑 曾志刚

Recommended by Associate Editor ZENG Zhi-Gang

1. 天津大学电气自动化与信息工程学院 天津 300072 2. 天津职业技术师范大学自动化与电气工程学院 天津 300222

1. School of Electrical and Information Engineering, Tianjin

人脑是一个由数百种不同类型的上千亿个神经元组成的极为复杂的生物组织, 可处理视觉、听觉、语言、学习、推理、决策、规划等各类问题, 然而, 我们对脑功能相关的神经网络结构和神经信息处理机制的解析仍极不清楚<sup>[1-4]</sup>. 因此, 需要构建接近人脑计算能力与智能水平的类脑计算系统, 从而理解人脑的结构和功能, 理解认知、思维、意识和语言,

University, Tianjin 300072 2. School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin 300222

探索脑疾病的发病机制与治疗方案, 推动新一代类脑智能技术的发展. 近年来, 高性能类脑计算已经成为研究脑科学的必要手段, 为脑研究开启了全新的研究视角和研究模式<sup>[5-10]</sup>; 反过来脑科学的研究为高性能计算提供新思路、提出新要求, 催生新型计算模式的出现<sup>[11-13]</sup>. 人脑最简单的认知行为需要上百万个神经元的共同活动<sup>[5-7]</sup>, 大规模类脑计算能够将神经元电生理活动机制与认知行为联系起来, 从而理解人脑认知行为的信息处理机制<sup>[14-17]</sup>. 总体而言, 现有的大规模类脑计算与仿真研究包含三个关键问题:

1) 冯·诺依曼计算架构存在计算瓶颈, 与人脑计算模式存在本质区别, 不能实现真正意义上的类脑. 首先, 冯·诺依曼计算架构将高维信息转换成纯时间维度的一维信息进行处理, 无法满足处理实时复杂智能问题的需求; 其次, 其信息处理过程在物理分离的存储器和中央处理器内完成, 造成大量的能耗损失, 往复传输速率与信息处理速率的不匹配导致严重的存储墙效应, 限制了类脑智能的研究与开发.

2) 现有的类脑计算平台由于自身计算能力的限制, 尚未解决神经元模型复杂度与神经网络规模之间的矛盾, 由具备非线性动力学特性的生物神经元组成的大规模脑功能神经网络需要更强的并行计算能力.

3) 当前的类脑计算工作忽视了神经元的动力学行为与人脑认知功能之间的联系, 导致理解人脑信息处理机制的从细胞层次到认知层次信息的研究鸿沟.

近年来, 各国政府、高校以及研究机构均密切关注着类脑计算这一全新领域, 并从国家层面提出了一批重大研究项目. 瑞士洛桑联邦理工学院和 IBM 联合支持了“蓝脑计划”(Blue brain project), 该项目旨在通过超级计算机 Blue Gene 模拟大规模神经网络活动<sup>[18]</sup>. 2013 年欧盟再次提出“人类大脑工程”(Human brain project, HBP), 其中包括神经科学、医学与类脑计算等领域, 涵盖了计算神经科学、神经信息学、人脑战略数据、高性能计算平台、大脑模拟并计算及神经机器人平台等子项目<sup>[19]</sup>. 同时, 美国政府提出了“脑创新计划”(Brain initiative), 重点从微观层面探索神经元、神经回路与大脑功能之间的关系<sup>[20]</sup>. 我国政府提出了“一体两翼”的脑科学战略布局, 将促进中国的信息产业与工业、金融、国防等方面的迅猛发展<sup>[1-3]</sup>. 要实现真正、彻底的类脑智能还需要研发全新的、打破冯·诺依曼架构的计算系统<sup>[21-22]</sup>. 人脑智能来自于其超大规模、复

杂互联的神经网络, 对与人脑大规模神经网络的并行计算是实现类脑智能的关键<sup>[6]</sup>, 目前欧美等国均开展了该领域的深入研究. 英国曼彻斯特大学的 SpiNNaker 系统基于 ARM (Advanced RISC machine) 芯片构建了大规模类脑计算系统, 该系统借鉴了神经元脉冲放电模式, 以较少的物理连接快速传递尖峰脉冲<sup>[21]</sup>, 该项目目前已成为欧盟脑计划的重要组成部分. 目前, IBM 公司开发出新型类脑计算系统, 它采用 16 颗 TrueNorth 芯片, TrueNorth 芯片借鉴了神经元的脉冲放电原理及其信息传递机制, 实现了计算存储融合的非冯·诺依曼计算架构, 可执行低功耗的多目标学习任务与实时视频处理任务<sup>[6]</sup>. 高通公司也推出了类脑计算系统 NPU, 并应用于行为学习与机器人研发等领域<sup>[23]</sup>. 此外, 德国海德堡大学作为欧盟脑计划的核心成员之一, 提出了 BrainScaleS 类脑计算系统, 在类脑机制与高性能计算方面取得了重大的突破<sup>[24]</sup>. 国内在大规模类脑计算的研究领域仍处于萌芽期, 中国科学院提出的深度学习系统与寒武纪系统均面向人工神经网络<sup>[25-26]</sup>, 这些研究均极大地提升了人工神经网络与深度学习的计算效率, 降低计算功耗, 然而这些计算模式仍然基于人工神经元的频率编码, 网络节点无法模拟人脑神经元的复杂动力学行为, 无法从人脑神经元脉冲编码的放电模式上进行更深层次的类脑智能. 因此, 如何借鉴人脑神经网络的放电机制与动力学行为, 克服大规模类脑计算的上述三个关键问题, 提出脑启发的新型计算架构与计算模式, 实现脑启发的大规模实时计算平台, 是开发类脑智能、理解人类认知与实现类脑智能应用的重要手段.

本文从人脑神经系统信息处理机制出发, 采用并行计算的现场可编程门阵列 (Field programmable gate array, FPGA) 构建了基于神经认知架构的众核类脑计算系统 BiCoSS, 实现了四百万数量级神经元的大规模类脑计算, 从而更好地理解人脑工作机制与疾病机制, 开发新一代类脑智能. 本文的结构组织如下: 第 1 节总述本文所构建的系统架构及需要解决的关键问题; 第 2 节阐述系统计算使用的相关模型; 第 3 节和第 4 节阐述关键技术问题的具体解决方案; 第 5 节展示了系统的性能; 第 6 节讨论了 BiCoSS 系统的特色; 第 7 节是结论及展望.

## 1 系统架构

BiCoSS 系统包含 7 个子系统, 采用片上网络结构的改进树形结构连接, 极大地增强了系统的可扩展性. 以神经元为基本计算核, 以神经网络为基本计算单元, 每个计算单元中包含若干计算核,

计算核的数量取决于计算单元的最大计算能力. 如图 1 所示, 每个子系统包含 4 个神经网络单元和一个路由单元, 每个神经网络单元与路由单元均由 FPGA 实现. 改进树形结构的每个节点均采用四个计算单元通过混合互联结构实现, 避免了单纯使用树形拓扑结构造成的频繁的计算单元跨层级通信, 减少了不必要的传播延迟和资源消耗. 树形结构中每个节点的四个计算单元采用混合互连结构, 四个计算单元通过路由单元与改进树形结构中其他节点相连, 非顶层节点中每个计算单元可以与其余 15 个计算单元通过 1 个路由单元进行直接通信, 顶层与底层节点中每个计算单元可以与其他 7 个计算单元进行直接通信. 相比较之前工作的拓扑结构, 本结构更有利于进行众核并行计算系统的扩展.

BiCoSS 系统实物图见图 2, 在 BiCoSS 系统中, 每个计算单元上包含一片 FPGA、两片 SDRAM (Synchronous dynamic random-access memory) 和一个用于程序存储的串行配置芯片 EPCS128, SDRAM 用于存储网络计算的中间变量与网络参数, 串行配置芯片 EPCS128 用于在电源关闭时存储程序. 并行计算的多个计算核, 对于 LIF (Leaky integrate and fire) 或 Izhikevich 神经元模型, 每个神经网络单元使用包含 16 节点的蝴蝶胖树 (Butterfly fat tree, BFT) 片上网络结构, 采用片上网络技术增强了每个计算单元上众核系统的通信效率、系统可扩展性与灵活性. 此外, 系统留有 GPIO (General-purpose input/output) 通信端口用于外接摄像头、机械臂、数模转换模块、模数转换模块等输入输出设备与外界进行数据交互. 当外部输入模

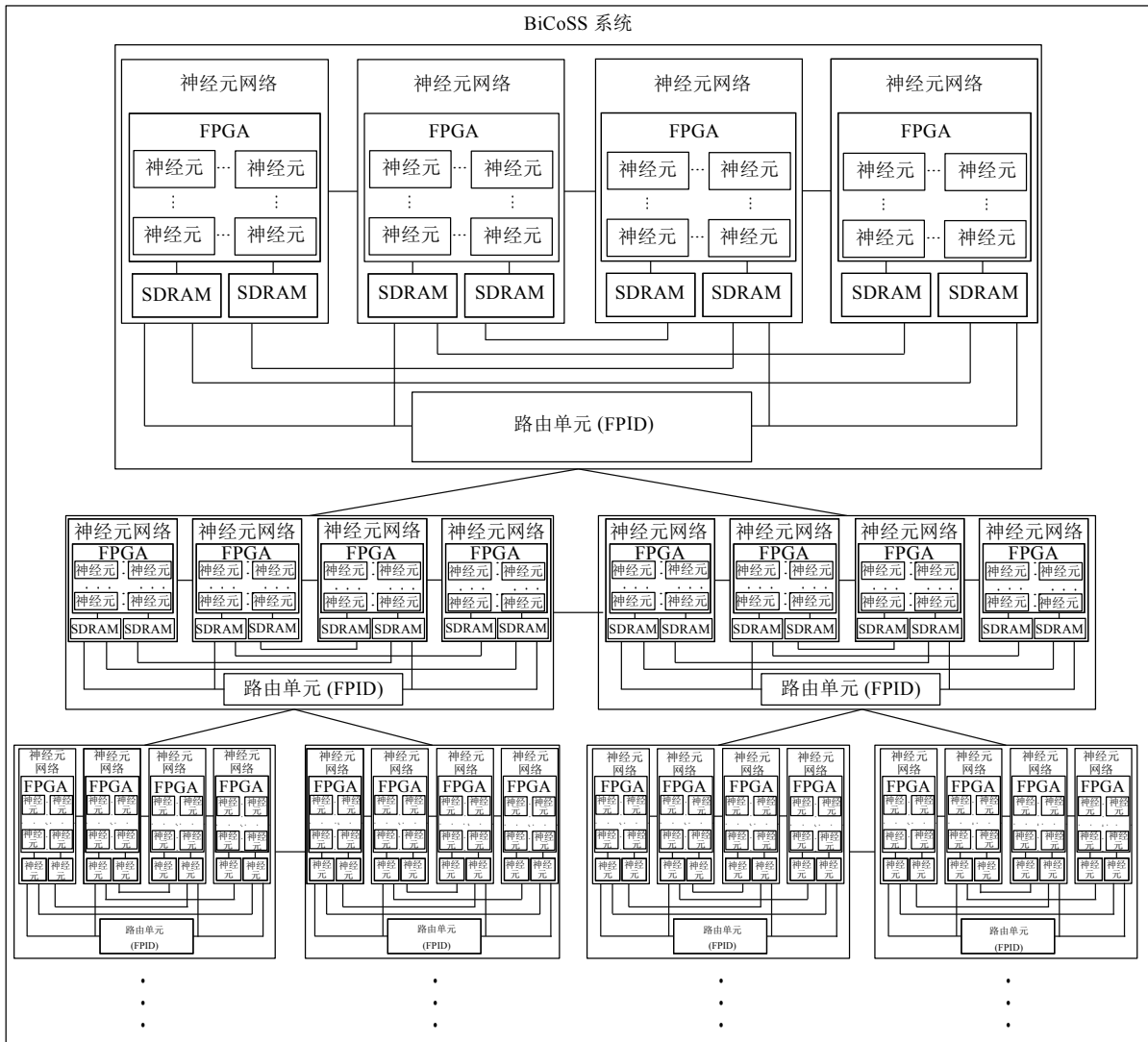


图 1 BiCoSS 系统架构

Fig.1 System architecture of BiCoSS

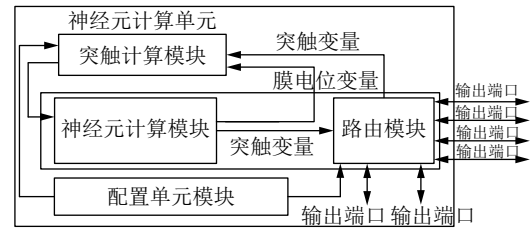


图 2 BiCoSS 系统实物图  
Fig. 2 Physical map of BiCoSS

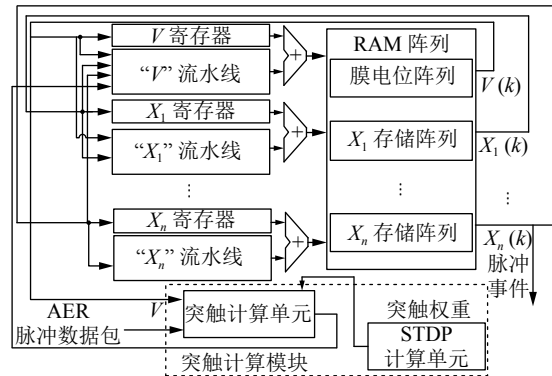
拟信号时, 通过模数转换模块将其转化为数字信号, 再通过通信端口传输到 FPGA 芯片的通用 IO (Input/output) 引脚上, 从而完成对外部信号的接收和处理; 外部数字信号则可直接通过通信端口传递. 同样, 如果需要输出模拟信号, FPGA 的输出信号可通过通用 IO 引脚传输到连接在通信端口的数模转换模块上, 转化为模拟信号进行输出; 数字信号可直接通过通信端口输出. 利用通信端口, 可以完成类脑感知与决策等实时应用开发.

BiCoSS 神经网络的详细计算架构如图 3 所示. 在神经网络计算中, BiCoSS 使用了神经元计算单元, 其中包括神经元计算模块、突触计算模块、配置单元模块、路由模块, 如图 3(a) 所示. 神经元计算模块输出突触变量到路由模块, 从而将突触变量信息传输到其他神经网络单元中; 输出膜电位变量到突触计算模块中进行突触计算, 配置单元模块用于配置路由模块的参数, 路由模块从片上网络临近节点接收外部地址事件表达 (Address event representation, AER) 脉冲数据包, 并且根据配置单元中的可编程路由表决定数据传输方向. 如图 3(b) 所示为神经元计算单元的详细计算架构. 对于神经元模型的变量  $V$ ,  $X_i$ ,  $i = 1, 2, \dots, n$ , 构建相应的神经元模型流水线计算模块以及随机存取存储器 (Random access memory, RAM) 存储阵列, 流水线计算模块实时更新神经元状态, RAM 阵列对神经元计算中间变量进行存储, 从而对神经网络进行时分复用操作, 在 BiCoSS 中对神经元计算模块进行 9000 次时分复用处理. STDP (Spike-timing-dependent plasticity) 计算单元将突触权重实时更新输出给突触计算单元中, 突触计算单元接收 AER 脉冲数据包与膜电位变量  $V$  进行计算, 神经元计算单元最终计算输出膜电位变量与突触变量.

图 4 所示为 BiCoSS 系统的神经元与神经突触



(a) 神经网络计算单元  
(a) Computing unit of the neural network



(b) 神经元计算单元详细架构  
(b) Detailed architecture of the neuron computing unit

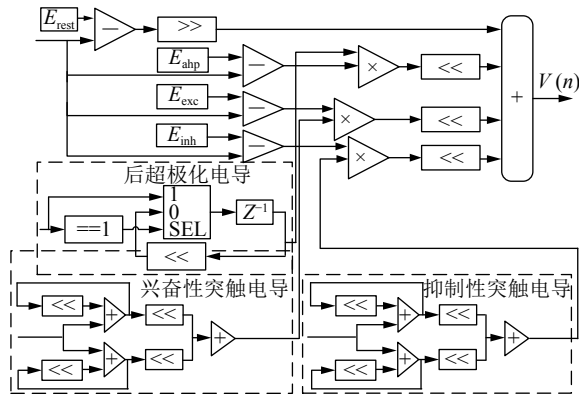
图 3 BiCoSS 系统神经网络计算架构

Fig. 3 Neural network computing architecture of BiCoSS system

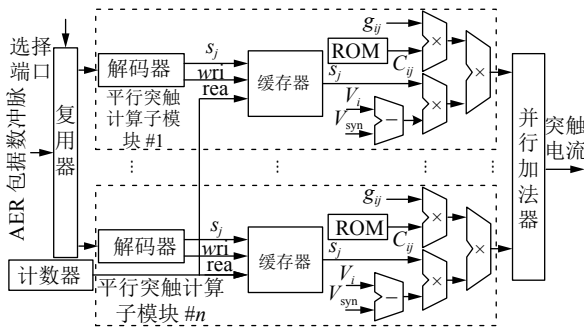
计算模块详细架构. 图 4(a) 所示为以 LIF 神经元模型为例的小脑神经元计算模块架构, 为了节约乘法器资源, 通过移位器实现常数乘法, 通过逻辑单元实现变量乘法. 如图 4(b) 所示为突触计算单元的详细计算架构, 其中包括  $n$  个平行突触计算子模块,  $rea$  表示缓存器的读地址, 其余变量与参数均在式 (1) 中定义. AER 脉冲数据包输入到复用器中, 通过常规计数器进行顺序选择. 解码器解码出 AER 脉冲数据包的突触变量信息  $s_i$  和时间标记信息, 时间标记信息用作缓存器的写地址  $wri$ . 在每个平行突触计算模块中, 网络连接矩阵被存储在只读存储器 (Read-only memory, ROM) 中, 并行加法器将  $n$  个平行突触计算模块的输出相加, 并将更新后的结果输出到神经元计算模块中进行计算.

## 2 类脑计算模型

大规模类脑计算的关键问题之一是选用哪种神经网络模型, 神经网络模型决定了网络的组成部分、各部分如何运行、这些部分之间如何相互作用. 受生物神经网络的启发, 神经网络的常见组成部分是神经元和神经突触, 如果大规模类脑计算的目的是利用类脑计算系统模拟人类大脑, 从而产生更高级的类脑智能, 同时理解人脑的工作机制, 那么生物学现实和合理的模型是必要的.



(a) 小脑神经元计算模块架构  
(a) Architecture of the cerebellar neuron computing module



(b) 突触计算单元详细架构  
(b) Detailed architecture of the synapse computing unit

图 4 BiCoSS 系统神经元与突触计算模块架构  
Fig. 4 Neuron and synapse computing architecture of BiCoSS system

2.1 神经网络与神经突触模型

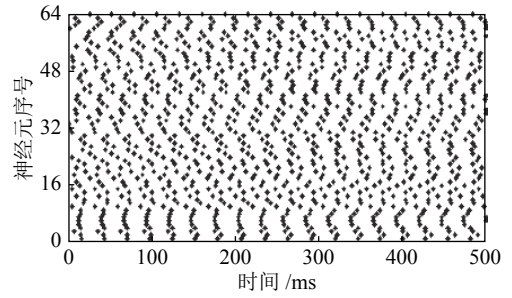
人脑的信息处理过程中,神经网络模型描述了不同神经元和突触如何相互连接以及相互作用. BiCoSS 系统根据人脑认知的具体计算任务,如图 5(a) 所示为神经网络放电栅图,其中横坐标为神经网络放电时间,纵轴为神经网络中产生放电行为的神经元序号,神经网络通过每个神经元的时间序列进行信息的传递与处理工作,从而构成人脑的认知行为.在突触模型方面, BiCoSS 系统采用针对人脑神经网络的生物启发突触,其表达式为

$$I_{ij}^{syn} = \sum_{j=1, j \neq i}^N g_{ij} C_{ij} s_j(t) (V_i - V_{syn}) \quad (1)$$

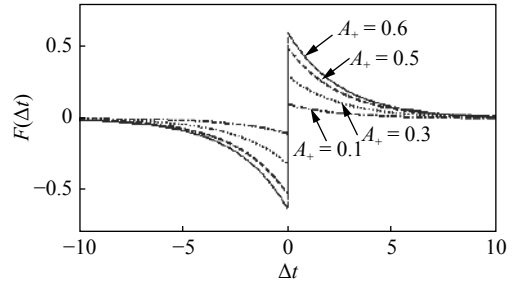
其中,  $C_{ij}$  表示连接矩阵,  $g_{ij}$  表示耦合强度.  $V_i$  表示突触前神经元和突触后神经元的膜电位,  $V_{syn}$  表示突触的反转电势,  $s_j$  为突触变量.

2.2 自学习算法

人脑的可塑性机制与大脑的学习有关,使神经



(a) 神经网络放电栅图  
(a) Firing raster plot of neural network



(b) 突触可塑性动力学特性  
(b) Dynamical characteristics of the synaptic plasticity

图 5 生物启发的神经网络放电行为与突触可塑性  
Fig. 5 Spiking activities of biologically inspired neuron model and STDP characteristics

元网络具有学习功能,即神经网络的突触权重随着时间变化. BiCoSS 系统采用在线的无监督学习,采用具有高度生物合理性的突触可塑性 (STDP) 的学习规则<sup>[27]</sup>,其表达式为

$$g_{ij} = g_{ij} + \Delta g_{ij} \quad (2)$$

$$\Delta g_{ij} = g_{ij} F(\Delta t) \quad (3)$$

$$F(\Delta t) = \begin{cases} A_+ \exp\left(\frac{-|\Delta t|}{\tau_+}\right), & \Delta t > 0 \\ -A_- \exp\left(\frac{-|\Delta t|}{\tau_-}\right), & \Delta t < 0 \\ 0, & \Delta t = 0 \end{cases} \quad (4)$$

其中,  $g_{ij}$  和  $\Delta g_{ij}$  分别表示某两个神经元之间的突触强度和突触强度变化值,  $\Delta t = t_i - t_j$  表示突触前后神经元放电的时间差,  $A_+$  和  $A_-$  限制了突触修正的最大值,  $\tau_+$  和  $\tau_-$  表示突触强度增强和减弱过程中的时间窗参数. 图 5(b) 所示为突触可塑性的动力学特性,当突触前神经元的放电活动先于突触后神经元时,所得曲线位于第 1 象限,表明突触耦合强度增加;反之,当突触前神经元后于突触后神经元放电时,所得曲线位于第 3 象限,表明突触耦合强度减弱.不同的 STDP 调节率会影响  $F$  函数的变化速度,调节率升高时  $F$  函数的变化趋势明显增加.

### 3 神经信息路由

人脑中神经网络信息的传递载体是神经元放电行为, 基于 AER 的神经信息处理方式来自生物神经网络进行信息加工的灵感<sup>[28]</sup>, 即神经元在产生动作电位后, 输出放电信号, 放电信号触发地址编码电路编码事件发生地址, 通过神经放电的峰-峰间期或放电频率来定义神经信号属性, 由放电事件主动触发, 未产生放电事件的神经元无输出, 从本质上避免了信息传递的冗余, 减少放电数据量, 减少系统开销与功耗, 提升系统计算效率. 在不发生严重竞争的情况下可以实时地输出放电事件, 从而实现实时计算系统. 对于同时发生的神经放电传递现象, 会引发数据传输冲突, 需要实现控制模块进行通信仲裁.

如图 6(a) 所示, BiCoSS 系统传输的放电信息使用 28 位数据包, 其中包含 8 位 AER 数据, 5 位芯片地址数据, 3 位层地址数据, 6 位节点地址以及 6 位时间标记. 5 位芯片地址用于编码 28 片负责计算神经网络的 FPGA 芯片, 3 位层地址 (包含 000, 001, 010, 011) 用于标记 BFT 结构的当前层数, 6 位节点地址 (包含 000000 ~ 001111) 用于表示编码 0 ~ 15 的神经元计算节点, 并可进一步扩展为 64 个神经元计算节点. 6 位时间标记用于决定 AER 数据包进行计算的时间, 其中, 6 bit 时间标记能够使 AER 数据延迟 0 ~ 63 个时钟周期进行计算. 该数据包表示方法可以实现 114 688 个神经元之间的任意连接. 如图 6(b) 所示, BiCoSS 系统的路由器包含 6 个输入端口与其他路由节点或神经计算单元相连接, 每个输入端口包含一个配有 6 个 FIFO (First input and first output) 的虚通道. 这些虚通道由虚通道仲裁器控制, 使用一个计数器去顺序选择每个虚通道内的 FIFO. 在神经放电信息路由过程中, 当第 1 层路由从神经计算单元中收到放电事件信息时, 放电信息包装器模块将打包过程初始化. 放电信息包装器在虚通道后面使用, 以便由神经计算单元输出的放电行为可以被处理成一个 AER 神经放电包. AER 神经放电包有 28 位, 由 5 部分组成, 分别为 AER 数据、芯片地址、层地址、节点地址与时间标记. 放电信息包装器使用配置处理器中存储的信息, 配置处理器能随时进行现场重复编程, 根据神经元数目与连接信息进行存储信息的修改. 配置处理器存储的数据包含芯片地址、层地址、节点地址以及时间标记芯片地址和层地址采用 3 位定点数, 节点地址选用 6 位定点数, 时间标记数据选用 6 位定点数. 输入的事件暂存在存储器中, 当时间标记到达时进行处理. 当前路由器相关地址编码如表 1 所示.

表 1 当前路由器相关地址编码

Table 1 The address coding of the current router

当前节点	子节点	邻居节点	父节点	负责神经计算单元
001, 0000	—	001, 0100	010, 0000	0000; 0001, 0010; 0011
001, 0100	—	001, 0100	010, 0000	0000; 0001, 0010; 0011
001, 1000	—	001, 0100	010, 0000	0000; 0001, 0010; 0011
001, 1100	—	001, 0100	010, 0000	0000; 0001, 0010; 0011
010, 0000	—	001, 0100	010, 0000	0000; 0001, 0010; 0011
010, 1000	—	001, 0100	010, 0000	0000; 0001, 0010; 0011

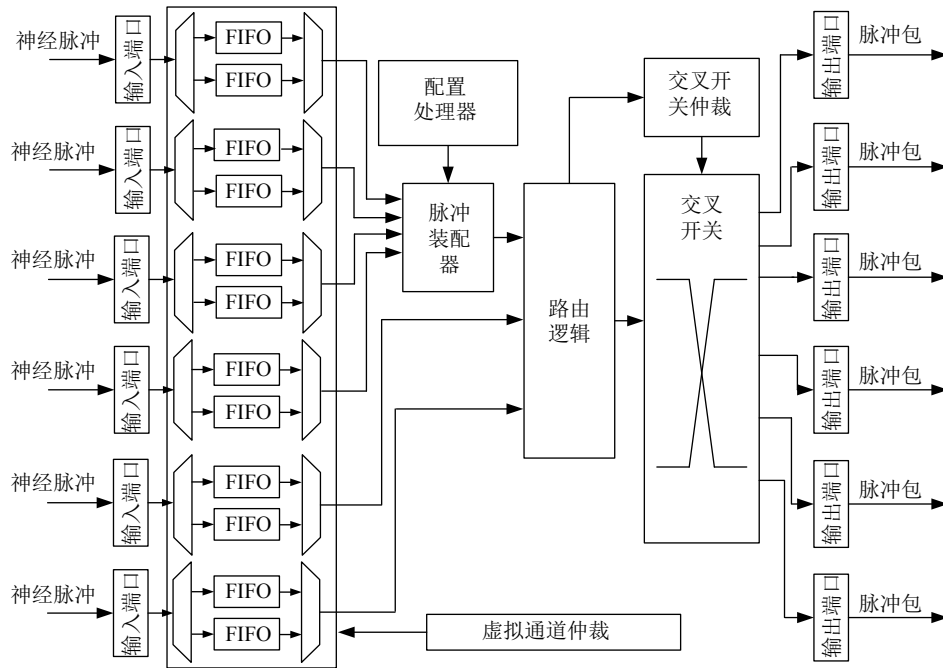
在神经放电事件被封装成 AER 神经放电包以后, 路由逻辑模块基于路由算法对 AER 神经放电包进行处理. 路由算法如图 6(c) 所示, 交叉互联模块由多路复用器实现, 由交叉互连仲裁器控制 BiCoSS 系统的路由算法为确定性的非最短路径长度路由算法. 路由器的节点编码与目的节点的编码相比较, 在算法中,  $dest$  表示了目的节点当前路由的节点地址, 由  $S(l, i)$  表示源节点和目的节点的地址分别由  $src$  和  $dest$  表示, 神经计算单元的数目为  $N$ . 首先输入源节点和目的节点的地址 ( $src, dest$ ) 以及当前路由器的地址 ( $l, i$ ), 如果  $l = 1$ , 则计算当前路由器的最低范围  $lowRange$  和最高范围  $highRange$  的值, 如果  $dest$  属于这个范围, 神经放电包将被传输到目的节点. 否则, 神经放电包将被传到当前路由节点任意父节点, 并重新开始路由算法. 如果  $dest$  属于交叉相连的路由器, 神经放电包通过交叉连接的链路传到相应路由节点, 并重新开始路由算法. 否则, 需要判断源节点与目的节点的关系, 如果存在  $dest - src < N/2$  和  $dest - src > -N/2$ , 神经放电包将传递到正确的邻居节点, 否则将传递到左邻居节点, 并重新开始路由算法, 继续执行.

### 4 实现

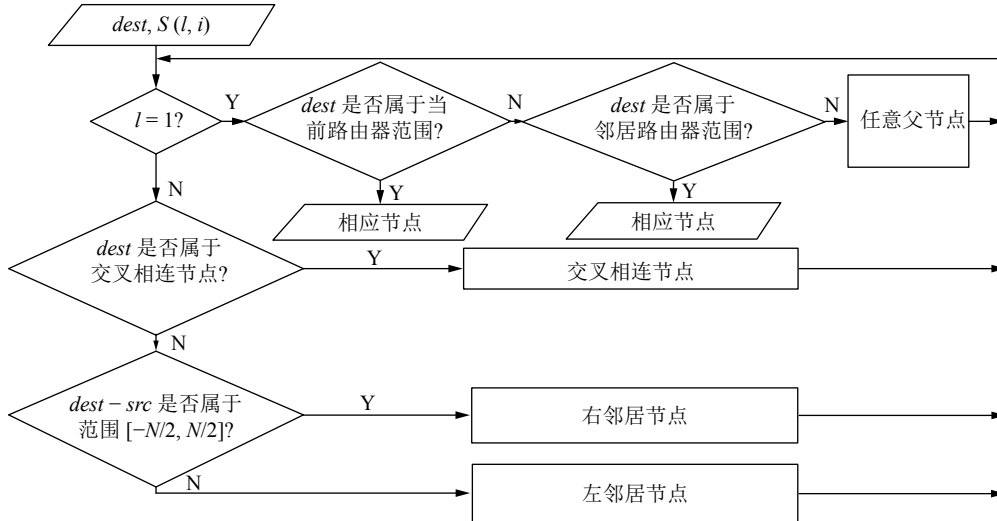
大规模类脑计算系统主要包含 3 种类型, 分别为数字电路实现、模拟电路实现以及数模混合电路实现. 模拟电路系统被认为是直接实现生物启发系统的首选方式, 具有高计算速度、低功耗等优点. 其缺点是计算精度比数字电路低, 软件指令与硬件动作之间的一对一通信能力较差, 一旦网络参数或结构发生变化则需要新一轮开发周期, 难以进行灵活配置. 数模混合实现方式通过定制计算系统对大规模神经网络进行模拟, 是对于外界环境比较敏感, 稳定性差, 应用开发十分困难, 可编程性差. 相比之下, 数字电路具有计算精度高、计算能力强和可重复配置的优势. 同时, 串行计算的数字电路 (如 DSP (Digital signal processing)、ARM) 不适用于神

AER数据	芯片地址	层地址	节点地址	时间标记
8位	5位	3位	6位	6位

(a) 数据包格式  
(a) Data packet format



(b) 路由器架构  
(b) Router architecture



(c) 路由算法  
(c) Routing algorithm

图 6 BiCoSS 系统的神经放电信息路由

Fig.6 Spike information routing of BiCoSS system

神经网络计算所需的并行性. 现场可编程门阵列 (FPGA) 具有并行计算能力, 包含丰富的逻辑计算单元与存储单元, 适合进行复杂算法与模型的实时计算.

因此, BiCoSS 系统采用英特尔公司的 Cyclone EP4CE115 系列 FPGA 芯片, 采用树形结构完成系统中芯片的级联与扩展, 每个芯片上采用网格型拓

扑结构进行片上系统实现, 片上与片间通信采用基于地址事件表达的神经放电信息路由方式, 从而提升计算效率, 节约计算开销与功耗, 增强系统可扩展性.

#### 4.1 硬件

BiCoSS 系统采用英特尔公司的 Cyclone IV 系列 FPGA 芯片 EP4CE115F29C7N, 该芯片具有 115 K 垂直排列的逻辑元件 (Logic elements, LEs)、4 Mbit 的嵌入式存储器和 528 个 IO 引脚. 每个芯片与两个 SDRAM 以及一个串行配置设备 EPCS128 相连, 被扩展为 6 层 PCB (Printed circuit board) 的核心系统. 模拟神经网络需要神经元状态信息、突触事件、网络耦合矩阵与强度等. 串行配置芯片 EPCS128 用于在电源关闭时存储程序. 否则, 每次重新上电时都要将所实现的大规模神经网络在 BiCoSS 系统中逐片上传. BiCoSS 的 7 个子系统通过 PCB 底板实现引脚之间的直接连接. 每个子系统上还留有额外的引脚并制成外部接口, 以便实现数模转换、视觉设备、机器人控制、系统扩展等功能.

#### 4.2 模型与算法

为了兼顾硬件资源消耗与精度, BiCoSS 系统选用欧拉法对神经元模型进行离散化, 以负责运动控制认知功能的小脑中 LIF 神经元模型欧拉法离散化为例, 所得差分方程为

$$V(n+1) = V(n) + \frac{\Delta T}{C} \times (g_{\text{leak}}(E_{\text{leak}} - V(n)) + g_{\text{ex:AMPA}}(E_{\text{ex}} - V(n)) + g_{\text{ex:NMDA}} \times (E_{\text{ex}} - V(n)) + g_{\text{inh}}(E_{\text{inh}} - V(n)) + g_{\text{ahp}}(E_{\text{ahp}} - V(n)) + I(n)) \quad (5)$$

式中,  $\Delta T$  是离散化采样时间,  $V(n)$  是当前时刻的变量迭代值,  $C$  为神经元电导. 参数  $g_{\text{leak}}$ ,  $g_{\text{ex:AMPA}}$ ,  $g_{\text{ex:NMDA}}$ ,  $g_{\text{inh}}$ ,  $g_{\text{ahp}}$  分别为漏电导、兴奋性 AMPA (Aminomethylphosphonic acid) 突触电导、兴奋性 NMDA (N-methyl-D-aspartate) 突触电导、抑制性突触电导、后超极化突触电导,  $E_{\text{leak}}$ ,  $E_{\text{ex}}$ ,  $E_{\text{inh}}$ ,  $E_{\text{ahp}}$  分别为漏电势、兴奋性电势、抑制性电势以及后超极化电势,  $I$  为神经元外部施加电流.

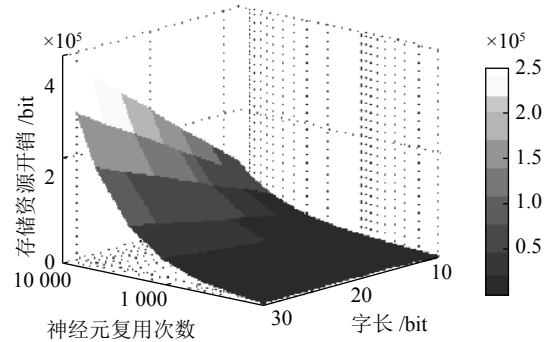
为了解决 FPGA 片上资源与网络规模之间的矛盾, BiCoSS 系统采用了分时复用技术, 应用同一个计算单元分时段计算多路数据. 对于神经网络仿真, 将神经元计算单元计算后得到的状态变量作为输出, 存储在片上 RAM 中. 在系统始终激励下, 向神经元计算单元依次输入每个神经元上一步的状态信息, 在分时复用的神经元进行状态更新后, 再进行下一步计算. 假设网络状态存储单元的 RAM

存储器的存储深度为  $N_{\text{neu}}$ , 网络中包含的神经元总数为  $N_{\text{total}}$ , 考虑时间复杂度有如下式成立:

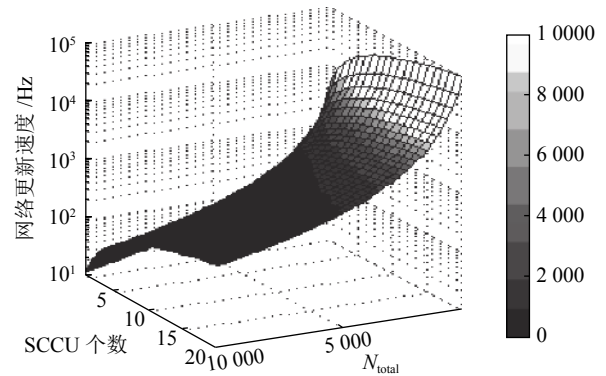
$$\begin{cases} N_{\text{total}} = n \times N_{\text{neu}} \\ N_{N_{\text{pipe}}} \times N_{\text{neu}} + N_{S_{\text{pipe}}} \times N_{\text{syn}} \leq \frac{1}{f_{\text{max}}} \\ N_{\text{syn}} = m \times N_{\text{neu}} \end{cases} \quad (6)$$

其中,  $f_{\text{max}}$  为 FPGA 可以实现的最大计算频率,  $N_{N_{\text{pipe}}}$  为神经元计算单元的流水线深度,  $N_{S_{\text{pipe}}}$  为突触电流计算单元的流水线深度,  $m$  为并行的突触电流计算单元数目,  $n$  为并行的神经元计算单元数目.

在类脑计算系统中, 硬件资源消耗随神经元功能性网络的网络规模增大而增大, 图 7(a) 直观体现片上存储资源开销与神经元计算位宽、神经元复用次数间的关系. 同时, 我们定义了网络更新速度与并行突触电流计算单元 (Synaptic current computation unit, SCCU) 个数、网络所有神经元数  $N_{\text{total}}$  之间的关系. 图 7(b) 显示网络更新速度会随 SCCU 数目的增加而增大, 同时增加 SCCU 有助于在确定



(a) 存储资源开销、神经元复用次数与字长关系  
(a) Relationship between memory resource cost, multiplexing number and bit



(b) 网络更新速度与 SCCU 个数、神经元数关系  
(b) Relationship between network updating speed, SCCU number and total neuron number

图 7 模型实现的性能分析

Fig. 7 Analysis performance of model implementation



网络连接拓扑情况下仿真更多神经元数目,网络中神经元数目的减少有利于提高网络的整体计算速度,这需要进行不同数量级的神经元功能性网络仿真以确定产生期望动力学行为的网络中最小神经元数目。

## 5 性能分析

本文通过 4 个方面测试与验证 BiCoSS 系统的性能,这 4 个方面分别为计算效率、功耗、通信效率与可扩展性,通过计算基于生物合理性的神经元模型与具有 STDP 学习机制的生物神经网络,从而测试 BiCoSS 系统的性能。

### 5.1 计算效率与功耗

为了更好地展示 BiCoSS 系统的计算效率,本文将 BiCoSS 系统与 3 个替代平台相比,3 个替代平台分别为 CPU, GPU 与多核总线平台。为了量化计算效率,我们使用代价函数  $Q$ ,其表达式描述如下:

$$Q = \frac{t_{\text{com}}}{t_{\text{bio}}} \quad (7)$$

$t_{\text{com}}$  与  $t_{\text{bio}}$  分别表示计算平台与真实生物神经网络计算同等尺度神经网络的的计算时间。CPU 计算平台使用 Intel 双核 2.4 GHz CPU 进行计算, GPU 计算平台采用 NVIDIA GTX 280 GPU,多核总线平台使用 EP3SE340 FPGA 芯片。仿真采用的是基于 LIF 神经元模型的生物小世界神经网络,如图 8(a) 所示,我们考察了  $10^3 \sim 4 \times 10^6$  的神经元数目范围,随着神经元数目的增长, BiCoSS 系统计算效率维持在更高的数值,而其余三种替代计算平台的计算效率均呈现指数下降的趋势,这是由于 BiCoSS 采用全并行的非冯·诺依曼计算架构,使计算效率得到了显著提升。

在系统功耗方面, BiCoSS 系统包含 35 片 Intel 公司的 EP4CE115 FPGA 芯片,消耗功耗为 10.419 W,系统的功率密度  $35.4 \text{ mW/cm}^2$ ,传统计算机 CPU 的功率密度为  $50 \sim 100 \text{ W/cm}^2$ 。

### 5.2 通信效率与可扩展性

在通信效率方面,如图 8(b) 所示, BiCoSS 系统的延时与传统线性结构与传统树形结构进行比较,当数据包注入率小于 0.18 (数据包/周期/节点) 时, BiCoSS 系统结构的负载延时远低于以 Neurogrid<sup>[22]</sup> 为代表的传统线性结构,低于以 HiAER<sup>[29]</sup> 为代表的传统树形结构,随着数据包注入速率的增加,平均延时略有升高,线形拓扑结构在 0.1 (数据包/周期/节点) 处出现拐点,传统树形结构与 BiCoSS 结构均在 0.18 (数据包/周期/节点) 处出现拐点,由

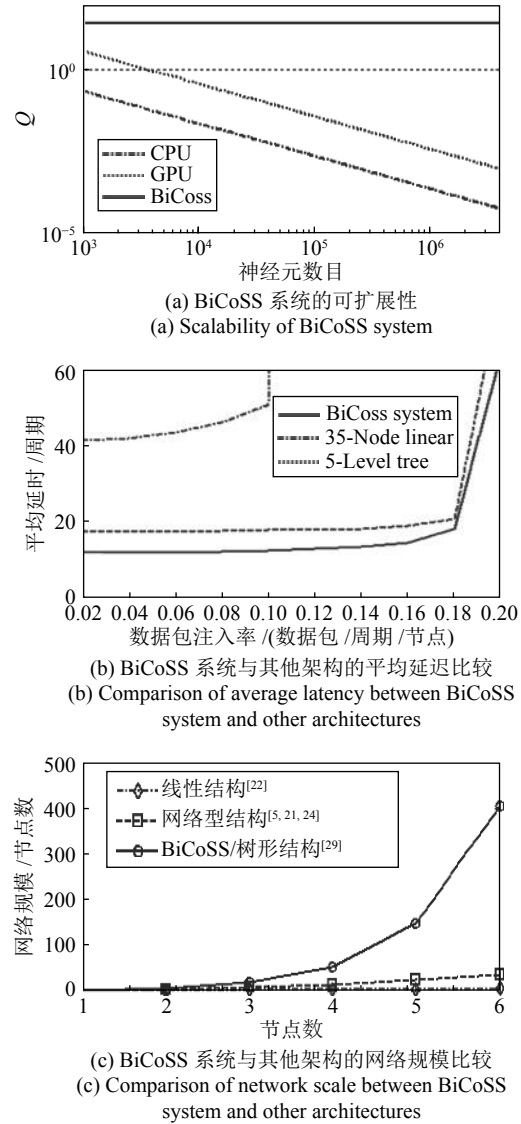


图 8 BiCoSS 系统性能分析

Fig.8 Performance analysis of BiCoSS system

于 BiCoSS 结构的节点连通度高于传统树形结构, BiCoSS 结构在拐点后的增幅略低于传统树形结构。

之前的神经形态学计算平台采用了不同的多片拓扑结构完成对于大规模生物神经网络的实现与仿真,典型的结构包括 Neurogrid 使用的线性结构<sup>[22]</sup>以及 BrainScaleS、Truenorth、SpiNNaker 使用的网格型结构<sup>[5, 21, 24]</sup>。在可扩展性方面,如图 8(c) 所示, BiCoSS 系统采用改进的树形结构,其可扩展性远高于传统线性结构与网格型结构,使得系统具有更强的可扩展性,更利于大规模人脑神经网络实现的扩展与网络规模的提升。

BiCoSS 系统神经网络单元使用 BFT 结构实现,具有良好的系统延迟特性。为了显示 BiCoSS 系统神经网络单元的延迟优越性,如图 9(a)~

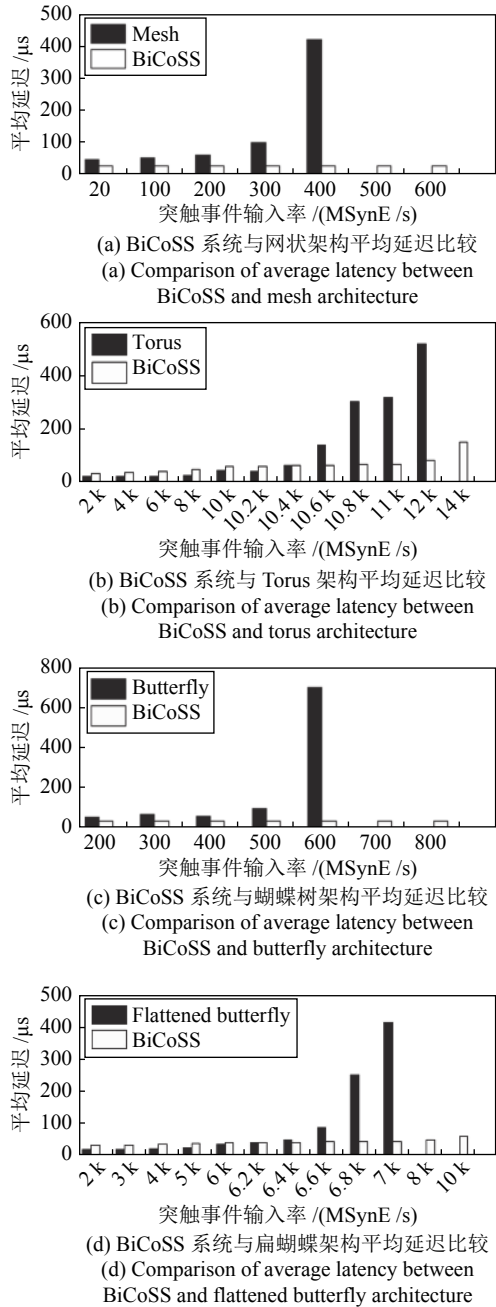


图 9 BiCoSS 系统神经网络单元平均延迟  
Fig.9 Average latency of neural network unit on BiCoSS

9(d) 所示, 比较了 BiCoSS 神经网络单元与传统片上网络 mesh 型、torus 型、Butterfly 型、Flattened Butterfly 型的平均延迟, 本文定义了突触事件输入率为兆突触事件数据包数/秒 (M synaptic events/s, MSynE/s). 由图 9 可以看出, 随着突触事件输入率的增大, BiCoSS 系统神经网络单元的平均延迟基本维持不变, 而另外三种片上网络结构的平均延迟随着突触事件输入率的增大具有明显增加, 表明了 BiCoSS 系统的神经网络单元具备在

较高突触事件输入率条件下的良好延迟特性.

### 5.3 系统性能比较

为了进一步揭示 BiCoSS 系统的性能优势, 如表 2 所示, 我们将 BiCoSS 与国际上主要的大规模类脑计算系统进行比较. 从实现方式上来看, BrainScaleS<sup>[24]</sup>、HiAER<sup>[29]</sup> 等系统采用了模拟电路的方式实现了大规模类脑计算系统, 缺点是计算精度比数字电路低, 软件指令与硬件动作之间的一对一通信能力较差, 一旦研究的神经网络参数或结构发生变化则需要新一轮开发周期, 难以进行灵活配置. 与此相对的是以 Truenorth<sup>[5]</sup>、SpiNNaker<sup>[21]</sup> 为代表的数字类脑计算系统与以 Neurogrid<sup>[22]</sup> 为代表的数模混合类脑计算系统. 数模混合系统结合了数字电路与模拟电路系统的优点, 缺点是对于外界环境比较敏感, 稳定性差, 数模转换产生延迟, 应用开发十分困难, 在可编程性方面输给了数字电路实现. 相比于模拟电路与数模混合电路, BiCoSS 采用数字实现的方式进行大规模类脑计算, 尽管数字电路功耗更高, 但其具有计算精度高, 计算能力强与可重复配置的优势. 神经元模型的复杂度决定了类脑计算的生理可信性, BiCoSS 系统采用数字电路可实现包括基于电导的 H-H 模型在内的任意模型, 以及基于生物启发的 STDP 机制的学习算法, 具有更强大的揭示人脑信息处理机制的能力. 在计算规模方面, BiCoSS 可以计算包含 100 M 神经元的大规模神经网络; 对于包含  $N$  个计算节点的 BiCoSS 系统, 其具有  $4eN$  的扩展性. BiCoSS 的缺点在于功耗为 10.419 W, 尽管远低于 SpiNNaker 的 49 W 功耗, 但相较于 Neurogrid (功耗为 2.7 W)、Truenorth (功耗为 63 mW) 等系统, BiCoSS 仍会消耗更高的功耗. 然而, 综合考虑网络规模、可实现模型的复杂度、生物启发的学习规则与系统可扩展性, 与国际上当前代表性的大规模类脑计算系统相比, BiCoSS 系统具有更突出的优势.

表 2 与当前代表性大规模类脑计算系统比较  
Table 2 The comparison with the state-of-the art large-scale brain-inspired computing systems

类脑计算系统	实现	模型	学习	规模	扩展性
BrainScaleS <sup>[24]</sup>	模拟	AIF	STDP	4 M	$N^2$
Truenorth <sup>[5]</sup>	数字	LIF	无	1 M	$N^2$
Neurogrid <sup>[22]</sup>	混合	QIF	无	1 M	$2N$
SpiNNaker <sup>[21]</sup>	数字	任意	STDP	1 B	$N^2$
LaCSNN <sup>[28]</sup>	数字	任意	STDP	1 M	$N^2$
BlueHive <sup>[30]</sup>	数字	任意	无	64 k	$N^2$
IFAT <sup>[31]</sup>	模拟	LIF	无	65 k	$2N$
HiAER <sup>[29]</sup>	模拟	LIF	无	1 M	$2N$
BiCoSS	数字	任意	STDP	4 M	$4.2N$

## 6 BiCoSS 的类脑智能与脑机制应用

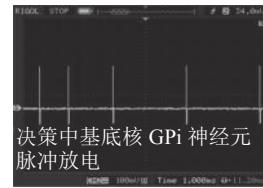
人脑一种简单的高层认知行为涉及上百万的神经元脉冲活动<sup>[5-7]</sup>, 大规模神经网络的计算有助于将神经网络动力学特性与人脑高层认知行为联系起来, 接近人脑的计算规模具有更强的生物可信性. BiCoSS 系统可以实时计算大规模人脑具有认知功能的神经网络, 一方面增加的网络规模可以更加接近人脑认知行为的计算规模, 从而准确地揭示人脑的大规模认知机制, 另一方面 BiCoSS 的实时计算性能可以用于与真实生物体的接口或人工智能设备中. 为了进一步研究人脑认知行为的机制, 从神经元水平分析了产生类脑认知的深层机制, BiCoSS 系统从类脑决策、类脑运动控制、神经疾病机制以及类脑视觉目标识别进一步探索脑机制. 在后面包括类脑决策、类脑运动控制以及运动障碍性疾病机制研究的类脑计算应用中, 我们均采用 BiCoSS 系统中包含的所有 35 片 FPGA, 搭建了 400 万神经元的大规模神经网络进行人脑认知机制的探索与实时计算. 图 10(a) 所示为基于 BiCoSS 系统的类脑计算平台实物图, 由 DE2 开发板实现的输入信号发生器, 用于产生输入的离散脉冲序列, 通过两位线传送到 BiCoSS 系统中, 用于模拟大规模神经网络的 BiCoSS 计算系统通过数模转换器将输出转换成模拟信号, 并在示波器显示输出波形. 图 9(b)~9(e) 显示了决策、运动控制、以及正常与运动障碍状态下, 基底核 GPi 神经元、小脑 GO 神经元、丘脑皮层神经元的实时输出离散脉冲信号, 在运动障碍性疾病状态下, 丘脑皮层神经元放电从尖峰放电变为了簇放电, 所显示的实时离散脉冲信号均来自 BiCoSS 大规模网络中某一特定核团内任意选取的待观测神经元.

### 6.1 类脑决策

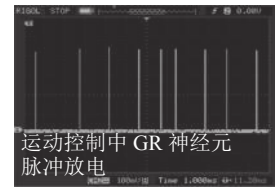
在各类功能性神经网络中, 认知决策功能相关的神经网络是当今研究的热点. 大脑决策是一个极为复杂的神经活动过程, 皮层-基底核-丘脑回路是人脑认知决策的重要网络, 不同的多巴胺水平会影响人脑在决策的时候采用不同的策略, 包括保守、冒险与随机<sup>[32]</sup>, 因此, BiCoSS 系统建立了负责人脑决策的大规模基底核神经网络. 该网络由纹状体 (Str), 苍白球外侧部 (GPe)、苍白球内侧部 (GPi)、丘脑底核 (STN) 四种核团组成, 其中纹状体 (Str) 被分为 D1 和 D2 两部分, 后三种核团各自为一部分, 每部分均由呈网格状排列的 100 万个神经元构成. 除纹状体以泊松序列模拟放电信息外, 其他核团的神经元均为带有离子通道电流的 Izhikevich 模型, 其神经元方程如下所示:



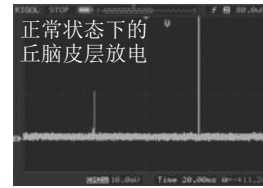
(a) 基于 BiCoSS 系统的类脑计算平台实物图  
(a) Brain-inspired computing platform based on BiCoSS system



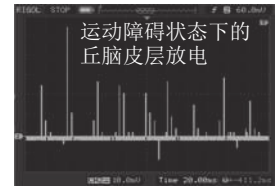
(b) 决策过程中的 GPi 神经元脉冲放电  
(b) Spiking activity of GPi neuron during decision making



(c) 运动控制过程中的 GR 神经元脉冲放电  
(c) Spiking activity of GR neuron during motor control



(d) 正常状态下的丘脑皮层放电  
(d) Spiking activity of thalamocortical neuron under the normal state



(e) 运动障碍状态下的丘脑皮层放电  
(e) Spiking activity of thalamocortical neuron under the movement disorder state

图 10 实验系统实物图与计算结果

Fig. 10 Experimental setup of BiCoSS system and computing results

$$\begin{cases} \frac{dV_{ij}^x}{dt} = 0.04(V_{ij}^x)^2 + 5V_{ij}^x - \\ \quad u_{ij}^x + I_{ij}^x + I_{ij}^{\text{Syn}}, & V_{ij}^x < V_{\text{peak}} \\ \frac{du_{ij}^x}{dt} = a(bV_{ij}^x - u_{ij}^x), & V_{ij}^x < V_{\text{peak}} \\ V_{ij}^x = c, & V_{ij}^x \geq V_{\text{peak}} \\ u_{ij}^x = u_{ij}^x + d, & V_{ij}^x \geq V_{\text{peak}} \end{cases} \quad (8)$$

其中,  $V_{ij}^x$  是膜电位,  $u_{ij}^x$  是膜恢复变量,  $I_{ij}^{\text{Syn}}$  是神经元接收的总突触电流,  $I_{ij}^x$  是神经元在  $x$  在位置坐标  $(i, j)$  处的外部电流,  $V_{\text{peak}}$  是神经元的放电阈值 (本模型中为 30 mV),  $x$  可以代表 GPe, GPi, STN 三种神经元中的任意一种. 突触电流公式如下所示:

$$\begin{cases} \tau_{Re} \frac{dh_{ij}^{x \rightarrow y}}{dt} = -h_{ij}^{x \rightarrow y}(t) + S_{ij}^x(t) \\ I_{ij}^{x \rightarrow y}(t) = W_{x \rightarrow y} h_{ij}^{x \rightarrow y}(t) (E_{Re} - V_{ij}^y(t)) \end{cases} \quad (9)$$

其中,  $\tau_{Re}$  是突触受体的衰减常数,  $E_{Re}$  是相关受体的突触电位,  $h_{ij}^{x \rightarrow y}(t)$  是神经元  $x$  在时间  $t$  的放电情况, 放电为 1, 不放电为 0,  $S_{ij}^x(t)$  是从神经元  $x$  到神经元  $y$  的突触电流的门控变量,  $W_{x \rightarrow y}$  是神经元  $x$  与神经元  $y$  连接的突触权重,  $V_{ij}^y(t)$  是神经元  $y$  在坐标位置  $(i, j)$  处的膜电位, Re 可以代表 AMPA, GABA ( $\gamma$ -aminobutyric acid), NMDA 三种突触受体中的任意一种. 不同类型细胞的具体参数值见表 3.

表 3 基底核模型中不同细胞的参数值  
Table 3 Parameter values of different cells in the basal ganglia model

参数	GPe	GPi	STN
$a$	0.1	0.1	0.005
$b$	0.2	0.2	0.265
$c$	-65	-65	-65
$d$	2	2	1.5
$I^x$ (nA)	10	10	30
$E_{AMPA}$ (mV)	0	0	0
$E_{NMDA}$ (mV)	0	0	0
$E_{GABA}$ (mV)	-60	-60	-60
$\tau_{AMPA}$ (ms)	6	6	6
$\tau_{NMDA}$ (ms)	160	160	160
$\tau_{GABA}$ (ms)	4	4	4
$W_{StrD2 \rightarrow GPe}$	0.8	—	—
$W_{StrD1 \rightarrow GPi}$	—	1	—
$W_{STN \rightarrow GPi}$	—	—	1.15

在基底核网络中, D1 纹状体 (Str) 和 D2 纹状体 (Str) 分别向苍白球内侧部 (GPi) 和苍白球外侧部 (GPe) 传递抑制性信号, 苍白球外侧部 (GPe) 向丘脑底核 (STN) 传递抑制性信号, 丘脑底核 (STN) 向苍白球内侧部 (GPi) 和苍白球外侧部 (GPe) 传递兴奋性信号, 基底核网络的最终输出由苍白球内侧部 (GPi) 给出. 在高多巴胺水平的条件下, 我们输入 BiCoSS 网络两个不同的刺激信号, 两个刺激信号分别为频率 4 Hz 和 8 Hz 的泊松序列. 我们从 100 万个 STN 和 GPi 神经元中各取 2 000 个神经元绘制放电栅图, 观测到 STN 和 GPi 核团的放电栅图如图 11(a) 和图 11(b) 所示, 可以看出, 在高多巴胺水平下, 基底核网络做出了相应的选择策略, 更强的 8 Hz 泊松序列刺激使得网络活动更快达到阈值, 因此基底核网络选择了 8 Hz 刺激作为最佳决策.

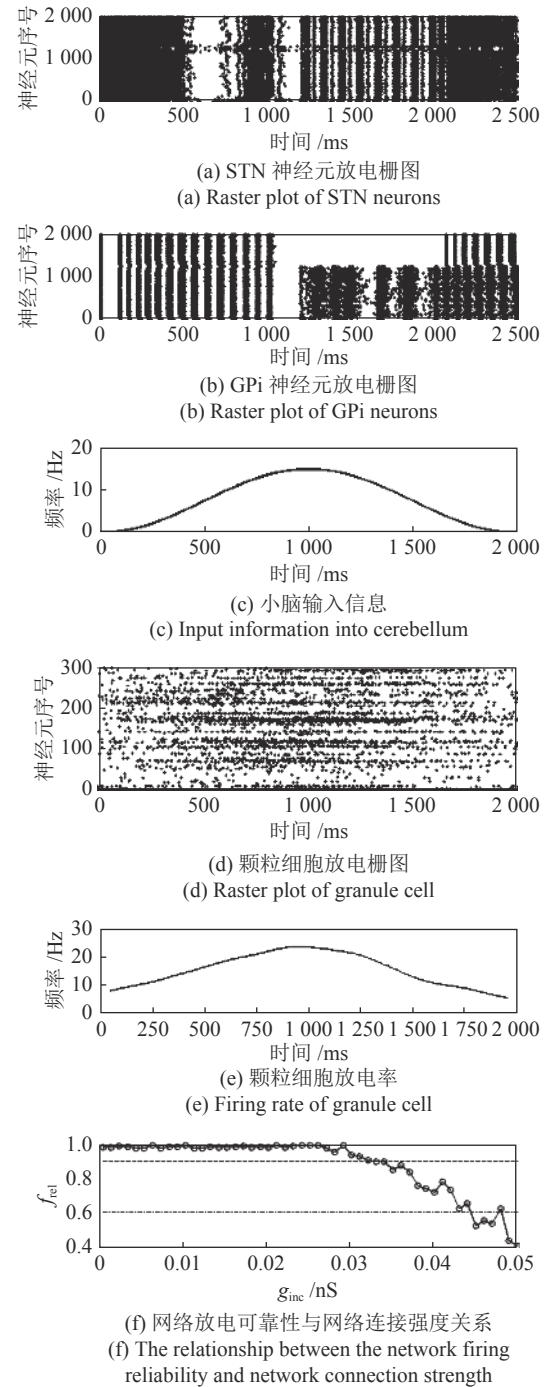


图 11 基于 BiCoSS 系统的认知计算

Fig. 11 Cognition computing based on BiCoSS system

## 6.2 类脑运动控制

小脑在人脑的运动控制功能中起着重要的作用, 包括维持躯体平衡、调节肌张力及协调运动等. BiCoSS 系统建立了大规模的小脑神经网络, 用于人脑运动控制机制的探索. 其中, 搭建了由 400 万个神经元构成的小脑模型, 其中包括 100 万个颗粒细胞 (Granule cell, GR), 100 万个高尔基细胞

(Golgi cell, GO), 50 万个浦肯野细胞 (Purkinje cell, PC), 50 万个篮状细胞 (Basket cell, BS), 50 万个下橄榄核细胞 (Inferior olive, IO) 和 50 万个前庭核细胞 (Vestibular nuclei, VN). 颗粒细胞 (GR) 和高尔基细胞 (GO) 均呈网格状排列. 所有的神经元细胞均为带有离子通道的 LIF 模型, 其统一的膜电位方程和突触电导方程如下所示.

$$C \frac{dV}{dt} = \begin{cases} -g_{\text{leak}}(V(t) - E_{\text{leak}}) - g_{\text{exc}}(t)(V(t) - E_{\text{exc}}) - g_{\text{ahp}}(t - t_0)(V(t) - E_{\text{ahp}}) - g_{\text{inh}}(t)(V(t) - E_{\text{inh}}) + I, & V \leq \theta \\ E_{\text{leak}}, & V > \theta \end{cases} \quad (10)$$

$$\begin{cases} g_{\text{ahp}}(t - t_0) = G_{\text{ahp}} \exp\left(-\frac{t - t_0}{\tau_{\text{ahp}}}\right) \\ g_{\text{exc}}(t) = G_{\text{exc1}} \exp\left(-\frac{t}{\tau_{\text{exc1}}}\right) + G_{\text{exc2}} \exp\left(-\frac{t}{\tau_{\text{exc2}}}\right) + G_{\text{exc3}} \exp\left(-\frac{t}{\tau_{\text{exc3}}}\right) \\ g_{\text{inh}}(t) = G_{\text{inh1}} \exp\left(-\frac{t}{\tau_{\text{inh1}}}\right) + G_{\text{exc2}} \exp\left(-\frac{t}{\tau_{\text{inh2}}}\right) \end{cases} \quad (11)$$

其中,  $C$  是神经元的膜电容,  $V$  是膜电位,  $g_x$  代表突触电导,  $E_x$  代表静息电位,  $G_x$  为最大突触电导,  $\tau_{\text{ahp}}$  是电导延迟时间,  $x$  可以是 {leak, exc, inh, ahp} 中的任意一个,  $\theta$  是膜电位阈值,  $t_0$  是神经元的上一个放电时刻.  $I$  是自发电流, 只存在于少数类型的细胞中. 不同类型细胞的具体参数值参见表 4.

在小脑网络结构中, 我们采用之前的研究<sup>[33]</sup>提出的具有解剖学意义的小脑网络结构进行计算. 其中, 外部输入通过苔状纤维 (Mossy fiber, MF) 传递给颗粒细胞 (GR), 颗粒细胞 (GR) 与高尔基细胞 (GO) 相互耦合, 将输入信息进行处理并通过平行纤维 (Parallel fiber, PF) 传递给浦肯野细胞 (PC) 和篮状细胞 (BS). 外部误差信号输入给下橄榄核细胞 (IO), 并转化为学习信号通过爬行纤维 (Climbing fiber, CF) 传递给浦肯野细胞 (PC), 通过长时程增强 (Long term potentiation, LTP) 和长时程抑制 (Long term depression, LTD) 改变其与颗粒细胞在 PF-PC 的突触强度, 这也是小脑学习的基础. 此外, 浦肯野细胞 (PC) 还接受来自篮状

表 4 小脑模型中不同细胞的参数值

Table 4 Parameter values of different cerebellar cells

	GR	GO	PC	BS	VN	IO
$\theta$ (mV)	-35	-52	-55	-55	-38	-50
$C$ (pF)	3.1	28	106	107	122.3	10
$G_{\text{leak}}$ (nS)	0.43	2.3	2.32	2.32	1.63	0.67
$E_{\text{leak}}$ (mV)	-58	-55	-68	-68	-56	-60
$G_{\text{exc1}}$ (nS)	0.1584	36.4	1	1	33	1
$G_{\text{exc2}}$ (nS)	0.0216	3.003	—	—	17	—
$G_{\text{exc3}}$ (nS)	—	6.097	—	—	—	—
$\tau_{\text{exc1}}$ (ms)	1.2	1.5	8.3	8.3	9.9	10
$\tau_{\text{exc2}}$ (ms)	52	31	—	—	30.6	—
$\tau_{\text{exc3}}$ (ms)	—	170	—	—	—	—
$E_{\text{exc}}$ (mV)	0	0	0	0	0	0
$G_{\text{inh1}}$ (nS)	0.012	—	1	—	30	0.18
$G_{\text{inh2}}$ (nS)	0.016	—	—	—	—	—
$\tau_{\text{inh1}}$ (ms)	7	—	10	—	42.3	10
$\tau_{\text{inh2}}$ (ms)	59	—	—	—	—	—
$E_{\text{inh}}$ (mV)	-82	—	-75	—	-88	-75
$G_{\text{ahp}}$ (nS)	1	20	100	1	50	1
$E_{\text{ahp}}$ (mV)	-82	-72.7	-70	-70	-70	-75
$\tau_{\text{ahp}}$ (ms)	5	5	5	2.5	2.5	10
$I$ (nA)	—	—	250	—	700	—

细胞 (BS) 的抑制信号, 之后将信息传递给前庭核细胞 (VN), 前庭核细胞 (VN) 的输出即为小脑的最终输出信号. 前庭动眼反射实验反映了当头部位置改变时, 视网膜成像保持稳定的神经响应过程. 在头部运动的过程中, 前庭会产生刺激并引起眼球的反射运动, 使得眼球不会随头部位置的改变而改变, 从而保证视网膜稳定成像. 图 11(c)~11(e) 显示了在头部运动时, 随头部移动的变化的前庭信息被编码为泊松序列输入给颗粒细胞, 颗粒层细胞中任意 300 个细胞的放电栅图与颗粒层细胞的放电率. 从图中可以看出, 颗粒细胞群的总体放电率随着刺激强度从小变大, 然后变小, 表明颗粒细胞可以将输入信号的幅度和时间信息传递给浦肯野细胞, 从而对眼球运动产生影响, 实现小脑的运动控制认知功能.

### 6.3 神经疾病机制研究

BiCoSS 系统的另一个重要应用是人脑精神疾病机制的仿真与分析. 为了探索运动障碍性疾病的发病机制, 基于 BiCoSS 系统实现了大规模皮层-基底核—丘脑皮层神经网络, 在此基础上探索运动障碍性疾病的发病与皮层-基底核-丘脑皮层神经网络耦合强度的关系. 皮层-基底核-丘脑皮层神经网络由 Izhikevich 神经元模型构建, 其中包含

GPe、GPi、STN、丘脑皮层 (Thalamocortical, TC) 四种核团, 神经元模型由下式表示:

$$\frac{dV}{dt} = 0.04V^2 + 5V + 140 - u + I \quad (12)$$

$$\frac{du}{dt} = a(bV - u) \quad (13)$$

电流  $I$  包含 3 部分, 即

$$I = I_{\text{app}} + I_{\text{syn}} + I_{\text{noise}} \quad (14)$$

其中,  $I_{\text{app}}$  为对不同种类核团的施加电流,  $I_{\text{noise}}$  为神经元固有噪声,  $I_{\text{syn}}$  为突触电流, 脉冲复位等式为

$$V \geq 30 \text{ mV} \Rightarrow \begin{cases} V \leftarrow c \\ u \leftarrow u + d \end{cases} \quad (15)$$

其参数值如表 5 所示. 突触电流计算式为

$$I_{ij}^{\text{syn}}(t) = g_{ij}s_j [V_{\text{syn}} - V_i(t)] \quad (16)$$

表 5 皮层-基底核-丘脑皮层模型中不同神经元的参数值  
Table 5 Parameter values of different cells in the cortico-basal ganglia-thalamocortical model

—	$a$	$b$	$c$	$d$	$I_{\text{app}}$ (pA)
GPe	0.005	0.585	-65	4	10
GPi	0.005	0.585	-65	4	10
STN	0.006	0.262	-65	2	5
TC	0.002	0.2	-65	2	0

突触连接权重  $g_{ij}$  如表 6 所示, 表 6 中  $g_{\text{inc}}$  为突触增加权重. 突触变量  $s_j$  由下式定义:

$$\frac{ds_j}{dt} = F(V_j)(1 - s_j) - \frac{s_j}{\tau_S} \quad (17)$$

其中,  $\tau_S$  是突触延迟时间参数, 恢复函数计算如下:

$$F(V_j) = \frac{1}{1 + e^{-V_j}} \quad (18)$$

表 6 皮层-基底核-丘脑皮层模型网络连接权重

Table 6 Parameter values of synaptic coupling weight in the cortico-basal ganglia-thalamocortical model

源节点 → 目的节点	突触连接权重 $g_{ij}$
GPe → GPe	$0.075 + g_{\text{inc}}$
GPe → STN	$0.025 + g_{\text{inc}}$
GPe → GPi	$0.015 + g_{\text{inc}}$
STN → GPe	$0.075 + g_{\text{inc}}$
STN → GPi	$0.01 + 5g_{\text{inc}}$
GPi → TC	$0.01 + 5g_{\text{inc}}$

感觉运动区对 TC 神经元的输入电流表示为

$$I = i_\chi \times H\left(\sin\frac{2\pi t}{\rho_\chi}\right) \times \left(1 - H\left(\sin\frac{2\pi(t + \delta_\chi)}{\rho_\chi}\right)\right) \quad (19)$$

其中,  $H$  为 Heaviside 函数, 参数值  $i_\chi = 30 \text{ pA}$ ,  $\delta_\chi = 3 \text{ ms}$ ,  $\rho_\chi = 25 \text{ ms}$ .

定义神经网络放电的可靠性为

$$f_{\text{rel}} = 1 - f_{\text{err}} \quad (20)$$

$$f_{\text{err}} = \frac{m}{m + e} \quad (21)$$

其中,  $f_{\text{err}}$  表示放电失败的百分比,  $m$  表示丘脑皮层输出簇放电或放电失败的次数,  $e$  表示丘脑皮层输出正常放电的次数. 较高的可靠性  $f_{\text{rel}}$  表示神经网络的放电属于正常水平, 过低的可靠性表示神经网络已处于运动障碍性疾病发病状态.

如图 11(f) 所示, 随着网络耦合强度的增加, 神经元放电的可靠性在一定范围内保持不变, 网络耦合强度的进一步增加导致网络放电可靠性迅速下降. 因此, 皮层-基底核-丘脑皮层神经网络中突触耦合强度的增加是运动障碍性疾病致病的重要因素之一.

## 7 结论与展望

本文开发了一个基于神经认知计算架构的大规模类脑计算系统 BiCoSS. 该系统能够实时地计算四百万神经元数量级的、与人脑认知功能相关的大规模神经网络以及基于生物启发学习机制的在线学习与训练. 相比传统类脑计算平台, 该系统具有更高的计算效率, 更低的计算功耗以及良好的扩展性, 适合进行具有认知功能的人脑大规模神经网络实时计算与仿真分析. 增加的网络规模可以更加接近人脑认知行为的计算规模, 从而准确地揭示人脑的大规模认知机制; 同时, BiCoSS 的实时计算性能可以用于与真实生物体的接口或人工智能设备中. 综合考虑网络规模、可实现模型的复杂度、生物启发的学习规则与系统可扩展性, 与国际上当前代表性的大规模类脑计算系统相比, BiCoSS 系统具有更突出的优势. 同时, BiCoSS 系统从类脑决策、类脑运动控制、神经疾病机制的脑认知任务方面进一步进行脑机制的计算与研究. 本研究完成了将类脑智能与机器智能进行融合的重要一步, 并可以作为进行脑机制实验的重要依据.

## References

- 1 Poo Mu-Ming, Xu Bo, Tan Tie-Niu. Brain science and brain-inspired intelligence technology-an overview. *Bulletin of Chinese Academy of Sciences*, 2016, **31**(7): 725-736 (蒲慕明, 徐波, 谭铁牛. 脑科学与类脑研究概述. 中国科学院院刊, 2016, **31**(7): 725-736)
- 2 Xu Bo, Liu Cheng-Lin, Zeng Yi. Research status and developments of brain-inspired intelligence. *Bulletin of Chinese Academy of Sciences*, 2016, **31**(7): 793-802 (徐波, 刘成林, 曾毅. 类脑智能研究现状与发展思考. 中国科学院院刊, 2016, **31**(7): 793-802)

- 3 Wang Li-Wei, Xu Li, Xu Ping, Yu Han-Chao, Kong Ming-Hui, Shen Yi, Zhang Yong-Qing. Brain science and brain-like intelligence research in Chinese Academy of Sciences. *Bulletin of Chinese Academy of Sciences*, 2016, **31**(7): 747–754  
(王力为, 许丽, 徐萍, 于汉超, 孔明辉, 沈毅, 张永清. 面向未来的中国科学院脑科学与类脑智能研究—强化基础研究, 推进深度融合. 中国科学院院刊, 2016, **31**(7): 747–754)
- 4 Park H J, Friston K. Structural and functional brain networks: from connections to cognition. *Science*, 2013, **342**(6158): 1238–1241
- 5 Merolla P A, Arthur J V, Alvarez-Icaza R, Cassidy A S, Sawada J, Akopyan F, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 2014, **345**(6197): 668–673
- 6 Zenke F, Bohté S M, Clopath C, Comşa I M, Göltz J, Maass W, et al. Visualizing a joint future of neuroscience and neuromorphic engineering. *Neuron*, 2021, **109**(4): 571–575
- 7 Yang S M, Wei X L, Wang J, Deng B, Liu C, Yu H T, Li H Y. Efficient hardware implementation of the subthalamic nucleus-external globus pallidus oscillation system and its dynamics investigation. *Neural Networks*, 2018, **94**: 220–238
- 8 Yang S M, Deng B, Li H Y, Liu C, Wang J, Yu H T, Qin Y M. FPGA implementation of hippocampal spiking network and its real-time simulation on dynamical neuromodulation of oscillations. *Neurocomputing*, 2018, **282**: 262–276
- 9 Wang Xiao-Feng, Yang Ya-Dong. Research on structure model of general intelligent system based on ecological evolution. *Acta Automatica Sinica*, 2020, **46**(5): 1017–1030  
(王晓峰, 杨亚东. 基于生态演化的通用智能系统结构模型研究. 自动化学报, 2020, **46**(5): 1017–1030)
- 10 Neckar A, Fok S, Benjamin B V, Stewart T C, Oza N N, Voelker A R, et al. Braindrop: A mixed-signal neuromorphic architecture with a dynamical systems-based programming model. *Proceedings of the IEEE*, 2019, **107**(1): 144–164
- 11 Zhang Hui, Wang Kun-Feng, Wang Fei-Yue. Advances and perspectives on applications of deep learning in visual object detection. *Acta Automatica Sinica*, 2017, **43**(8): 1289–1305  
(张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望. 自动化学报, 2017, **43**(8): 1289–1305)
- 12 Chen Wei-Hong, An Ji-Yao, Li Ren-Fa, Li Wan-Li. Review on deep-learning-based cognitive computing. *Acta Automatica Sinica*, 2017, **43**(11): 1886–1997  
(陈伟宏, 安吉尧, 李仁发, 李万里. 深度学习认知计算综述. 自动化学报, 2017, **43**(11): 1886–1997)
- 13 Hassabis D, Kumaran D, Summerfield C, Botvinick M. Neuroscience-inspired artificial intelligence. *Neuron*, 2017, **95**(2): 245–258
- 14 Thakur C S, Molin J L, Cauwenberghs G, Indiveri G, Kumar K, Qian N, et al. Large-scale neuromorphic spiking array processors: A quest to mimic the brain. *Frontiers in Neuroscience*, 2018, **12**: 891
- 15 Eliasmith C, Trujillo O. The use and abuse of large-scale brain models. *Current Opinion in Neurobiology*, 2014, **25**: 1–6
- 16 Izhikevich E M, Edelman G M. Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences*, 2008, **105**(9): 3593–3598
- 17 Yang S M, Deng B, Wang J, Li H Y, Lu M L, Che Y Q, et al. Scalable digital neuromorphic architecture for large-scale biophysically meaningful neural network with multi-compartment neurons. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, **31**(1): 148–162
- 18 Markram H. The blue brain project. *Nature Reviews Neuroscience*, 2006, **7**(2): 153
- 19 Markram H. The human brain project. *Scientific American*, 2012, **306**(6): 50–55
- 20 Insel T R, Landis S C, Collins F S. The NIH brain initiative. *Science*, 2013, **340**(6133): 687–688
- 21 Furber S B, Galluppi F, Temple S, Plana L A. The SpiNNaker Project. *Proceedings of the IEEE*, 2014, **102**(5): 652–665
- 22 Benjamin B V, Gao P, McQuinn E, Choudhary S, Chandrasekaram A R, Bussat J M. Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proceedings of the IEEE*, 2014, **102**(5): 699–716
- 23 Kumar S. Introducing qualcomm zeroth processors: Brain-inspired computing. *Qualcomm OnQ Blog*, 2013: 1–11
- 24 Petrovici M A, Vogginger B, Müller P, Breitwieser O, Lundqvist M, Muller L, et al. Characterization and compensation of network-level anomalies in mixed-signal neuromorphic modeling platforms. *PLoS One*, 2014, **9**(10): e108590
- 25 Chen T S, Du Z D, Sun N H, Wang J, Wu C Y, Chen Y J, Temam O. Diannao: A small-footprint high-throughput accelerator for ubiquitous machine learning. *ACM Sigplan Notices*, 2014, **49**(4): 269–284
- 26 Chen Y J, Luo T, Liu S L, Zhang S J, He L Q, Wang J, et al. Dadiannao: A machine learning supercomputer. In: Proceedings of the 47th Annual IEEE/ACM International Symposium on Microarchitecture, Cambridge, UK: IEEE, 2014. 609–622
- 27 Song S, Miller K D, Abbott L F. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 2000, **3**(9): 919–926
- 28 Yang S M, Wang J, Deng B, Liu C, Li H Y, Fietkiewicz C, et al. Real-time neuromorphic system for large-scale conductance-based spiking neural networks. *IEEE Transactions on Cybernetics*, 2019, **49**(7): 2490–2503
- 29 Park J, Yu T, Joshi S, Maier C, Cauwenberghs G. Hierarchical address event routing for reconfigurable large-scale neuromorphic systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, **28**(10): 2408–2422
- 30 Moore S W, Fox P J, Marsh S J T, Marketos A T, Mujumdar A. Bluehive — A field-programmable custom computing machine for extremescale real-time neural network simulation. In: Proceedings of the 20th IEEE International Symposium on Field-Programmable Custom Computing Machines. Toronto, ON, Canada: IEEE, 2012. 133–140
- 31 Yu T, Park J, Joshi S, Maier C, Cauwenberghs G. 65k-neuron integrate-and-fire array transceiver with address-event reconfigurable synaptic routing. In: Proceedings of the 2012 IEEE Bio-medical Circuits and Systems Conference (BioCAS), Hsinchu, China, 2012. 21–24
- 32 Zhao F F, Zeng Y, Wang G X, Bai J, Xu B. A brain-inspired decision making model based on top-down biasing of prefrontal cortex to basal ganglia and its application in autonomous UAV explorations. *Cognitive Computation*, 2018, **10**(2): 296–306
- 33 Yamazaki T, Tanaka S. A spiking network model for passage-of-time representation in the cerebellum. *European Journal of Neuroscience*, 2007, **26**(8): 2279–2292



杨双鸣 天津大学电气自动化与信息工程学院讲师。主要研究方向为类脑计算, 人工智能与深度学习。

E-mail: yangshuangming@tju.edu.cn  
(YANG Shuang-Ming Lecturer at the School of Electrical and Information Engineering, Tianjin University. His research interest covers brain-inspired computing, artificial intelligence, and deep learning.)



**郝新宇** 天津大学电气自动化与信息工程学院博士研究生. 主要研究方向为神经计算及 FPGA 实现.

E-mail: haoxy@tju.edu.cn

**(HAO Xin-Yu** Ph.D. candidate at the School of Electrical and Information Engineering, Tianjin Uni-

versity. His research interest covers neural computing and its FPGA implementation.)

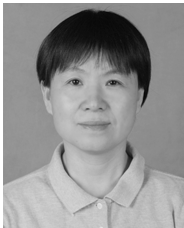


**王江** 天津大学电气自动化与信息工程学院教授. 主要研究方向为神经控制工程与神经科学.

E-mail: jiangwang@tju.edu.cn

**(WANG Jiang** Professor at the School of Electrical and Information Engineering, Tianjin Uni-

versity. His research interest covers neural control engineering and neuroscience.)



**李会艳** 天津职业技术师范大学自动化与电气工程学院教授. 主要研究方向为非线性系统与神经网络.

E-mail: lhy2740@126.com

**(LI Hui-Yan** Professor at the School of Automation and Electric-

al Engineering, Tianjin University of Technology and Education. Her research interest covers nonlinear systems and neural networks.)



**魏熙乐** 天津大学电气自动化与信息工程学院教授. 主要研究方向为神经控制工程与无创式脑调制技术.

E-mail: xilewei@tju.edu.cn

**(WEI Xi-Le** Professor at the School of Electrical and Information Engineering, Tianjin Uni-

versity. His research interest covers neural control engineering and noninvasive brain modulation technology.)

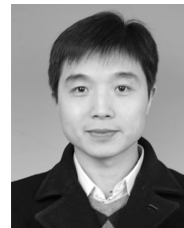


**于海涛** 天津大学电气自动化与信息工程学院副教授. 主要研究方向为神经系统建模与动力学分析.

E-mail: htyu@tju.edu.cn

**(YU Hai-Tao** Associate professor at the School of Electrical and In-

formation Engineering, Tianjin University. His research interest covers neural system modeling and dynamics analysis.)



**邓斌** 天津大学电气自动化与信息工程学院教授. 主要研究方向为神经计算及其非线性动力学分析. 本文通信作者. E-mail: dengbin@tju.edu.cn

**(DENG Bin** Professor at the School of Electrical and Information Engineering, Tianjin Uni-

versity. His research interest covers neural computing and nonlinear dynamics analysis. Corresponding author of this paper.)