

# 神经计算机的研究与发展<sup>\*</sup>

Researchs and Advances in Neurocomputer

徐 宁 虞厥邦

(电子科技大学光电子技术系 成都 610054)

**Abstract** First, this paper introduces the relation and the distinction between Von Neumann computer and neurocomputer. Then the advances of neurocomputers are introduced, and the prospect of neuro-computers in the future is described.

**Keywords** Von neumann computer, Neurocomputer, Neural network

## 1 引言

从40年代冯·诺依曼发明基于串行符号处理的数字计算机以来,数字电子计算机已经取得了巨大的成功,但在诸如模式识别、人工智能等研究领域却碰到了极大的困难,人们不得不以更大的兴趣去研究以并行处理模式为特征的神经计算机<sup>[1]</sup>。80年代,在美国、日本和欧洲都掀起了一股研究神经网络理论和神经计算机的热潮。各先进国家相继投入巨额资金用于研究新型智能计算机,其研究重点主要是将神经网络原理用于图像处理、模式识别、语音综合及智能机器人控制等领域。1987年6月在美国召开了第一届神经网络国

际会议,并发起成立了国际神经网络学会(INNS),并且各大公司,如:IBM公司、日本富士通和NEC等,纷纷开始研制各种神经芯片,并相继推出了各种硬件和软件产品<sup>[2-7]</sup>。本文主要对神经计算机的发展进行了综述性的介绍,并对神经计算机的未来进行了展望。

## 2 冯·诺依曼计算机与神经计算机

### 2.1 冯·诺依曼计算机与神经计算机的差别<sup>[8-11]</sup>

与冯·诺依曼计算机不同,神经计算机的主要特征是具有快速并行处理的学习功能。表1列出了两者的主要差别。

表1 冯·诺依曼计算机与神经计算机的差别

	冯·诺依曼计算机	神经计算机
基本结构	采用的是冯·诺依曼体系结构	并行处理
	以精确的“0”和“1”的方式表示数字信息	能处理连续变化的模拟信息
	一般以CPU作为处理问题的核心	由大量简单的处理单元协同处理问题
运行模式	为寻求问题的精确解,常常遇到计算时间困难	在短时间内寻找好的,但不一定是最好的解
	注重算法,执行过程严格控制,结果可预料	注重变换,通过学习训练形成信息处理方法,结果难以预料
	由指定的存储器存储信息,能够轻易读出具有特定意义的数据信息	无特定的数据存储和数据存储器,以神经元之间的互连形式分布进行存储,可以通过部分信息而联想出事件的整体
功能	利用数字和逻辑进行二值逻辑判断	对模糊的、不完整或冗余,甚至矛盾的数据进行问题求解
	容错能力较弱	具有很强的容错能力
	擅长于科学与工程计算,过程控制与信息管理等领域	在视觉、语言、信息处理、机器人控制、模式识别等方面取得进展

### 2.2 冯·诺依曼计算机与神经计算机的联系<sup>[8-11]</sup>

在表1中列出了冯·诺依曼计算机与神经计算机之间的不同特点,但从现在技术发展的水平来看,它们

<sup>\*</sup>中国博士后资助项目。

之间又有密切的联系。

(1)冯·诺依曼计算机是目前各种神经网络的分析 and 开发的工具；

(2)目前绝大多数应用是根据神经网络的计算机模型编制而成的运行在传统计算机上的软件，这显然丧失了神经计算机固有的并行性，但也保留了许多其他优点，如非线性特性、容错能力以及联想记忆等功能，在一些实时性要求不高或处理信息量不大的场合，这种应用是有效的；

(3)以现有的计算机为核心，配置专用的高速处理芯片与插件，虚拟实现神经计算机的功能；

(4)实现全硬件的神经计算机，真正反映神经网络的各种机制，这是神经计算机的发展趋势。

冯·诺依曼计算机与神经计算机之间的联系决定了现阶段神经网络实现技术的方法和物质手段，而它们之间的区别决定了神经网络实现技术研究的内容和特点。

### 3 神经计算机的发展现状

#### 3.1 神经计算机的体系结构

当前的并行神经计算机系统都是由面向用户的前端和面向神经计算机的后端机组成。一般后端机与前端机之间的关系有三种：

(1)后端机通过总线作为前端机的存储器影象设备 如图 1 所示。这种方式可以使后端机和前端机结合得更加紧密，由于后端机的存储器空间是前端机存储器地址空间的一部分，所以前端机和后端机可以通过对存储器访问的方式实现信息的交互，这一功能可以易于通过存储器影象设备驱动软件实现。其缺点是必然引起神经计算机系统存储器总线的拥塞。后端机与前端机及其它一些总线上连接的设备如果频繁进行大量的信息交换，并行神经计算机系统的存储器总线将成为系统的瓶颈，这必然会降低系统的计算性能。

这种结构的神经计算机有 NHC 开发的 ANZA<sup>+</sup> 神经计算机系统。该系统的后端机通过与 IBM/PC 286/386 兼容的 ISA 总线协议与前端机进行通讯，后端机占有系统的一块存储空间，作为前端机的存储器影象总线资源。

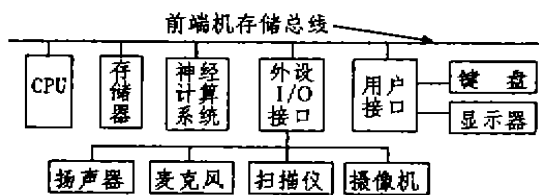


图 1

(2)后端机通过前端机的标准外设接口作为前端机的外部设备 此方案如图 2 所示。其优点在于后端机可以与任何具有后端机标准外设接口的前端机相连，具有很好的硬件移植性。但此方案也有缺点，在该方案中前端机和后端机之间的关系可以看成是神经网络应用主程序调用神经计算处理子程序的关系，这种调用关系的实现是和系统所采用的操作系统以及前端机的设备驱动器紧密相关的，所以硬件的移植必然会带来软件的修改问题，而修改软件往往比修改硬件更费事。

TRW 公司开发的 MARK-Ⅲ 和 MARK-Ⅳ 并行神经计算机系统就是采用这种方案设计的，两种机器都是通过 DRV-11 外设接口与 VAX 相连。

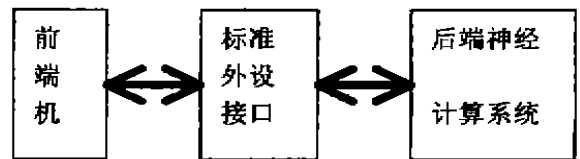


图 2

(3)后端机作为协处理器 如图 3 所示。它是将后端机和其他外设、协处理器、文件服务器等设备通过高带宽局域网连接在一起。其优点在于，后端机可以通过高速局域网访问其它网上设备，而不必利用前端机的存储总线，这有效地降低了前端机的负载，同时大大增加了后端机直接访问信息的类型。这种方案可以通过分时的方式使一台后端机同时为多台前端机服务。由于软件的开发是基于网络的高级层次，因此软件开发比第二种方案简单。同时，由于网络提供的多层次的信息传输协议，此方案大大提高了后端机的接口能力。

HNC 公司开发的 ANZA<sup>+</sup>VME 并行神经计算机系统采用此方案，它是在 SUN3 或 SUN4 工作站的文件服务器上插上一块或多块 ANZA<sup>+</sup>VME 板，通过局域网，任何 SUN 工作站都可以访问这些插板，因此，多个用户可同时共享同一神经计算机系统。

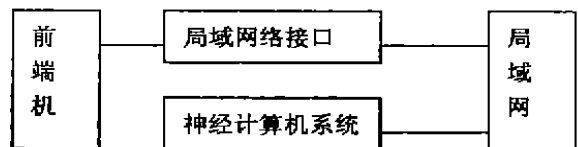


图 3

总之，在并行神经计算机系统设计中，若前端机系统比较简单，把后端机作为前端机的存储器影象设备是最好的选择；当前端机系统较复杂，一般选择协处理器方案较好。

### 3.2 几种并行神经计算机系统介绍

3.2.1 MARK-III 并行神经计算机系统 它是 TRW 公司于 1984 年~1985 年期间进行研制。它具有完善的神经网络软件环境,在主机上利用该软件开发环境进行神经网络描述。当完成了神经网络描述之后,系统所提供的编译器将描述编译成可由八个处理单元执行并可利用八个可执行代码和数据代码。它将神经计算应用程序分割成为计算量大大致相等的八个部分,这样有利于 MARK-III 并行神经计算机系统的八个处理单元负载均衡。

在 MARK-III 神经计算机中,VME 总线提供了 32 位的数据通道,其中 16 位为 16 位处理单元所拥有,输入/输出接口也是 16 位设备,MARK-III 神经计算机最多可模拟  $8 \times 10^3$  个神经元和  $4.8 \times 10^4$  条连接,该神经计算机是单用户系统,最高执行速度为每秒  $3.8 \times 10^5$  条连接。

MARK-III 神经计算机研制出来以后进行了产品化,已销售出神经网络计算模拟和神经网络应用。

3.2.2 MARK-IV 并行神经计算机系统 采用流水方式工作,其神经计算能力是最大可模拟  $2.5 \times 10^3$  个神经元和  $5.5 \times 10^5$  条连接,其学习过程的计算速度可达每秒  $5 \times 10^6$  条连接。以流水处理方式来实现神经元的计算,可以在一个单独的处理单元内并行实现神经计算。一是处理单元内部的并行,二是处理单元之间的并行。

MARK-IV 基于大多数神经网络模型进行分析和描述,使得它能模拟绝大多数神经网络模型。受到当时神经网络理论的发展限制,MARK-IV 神经计算机系统设计时主要考虑的是如何有效地实现神经感知机、盒中脑状态、时空模式识别等神经网络模型。

MARK-IV 神经计算机的主存由三个大容量具有连续地址的存储器组成,每个存储器可以存储  $5.5 \times 10^6$  个 24 位字。可实现 262144 个神经元和  $5.5 \times 10^5$  条连接。

#### 3.2.3 TI NETSIM 并行神经计算机系统

NETSIM 是由 TI 公司英国分公司和剑桥大学共同研制的一种基于并行处理器阵列的神经计算机。它由一组神经网络仿真器卡所组成,仿真器卡的物理互连拓扑结构是三维立方体。由一台 IBM-PC 主机作为前端机,主要完成输入输出等任务。仿真器卡包括一个标准的 80188 微处理器,两块专用芯片,一个计算单元与通信处理器,以及用于存储连接权、程序和 BIOS 的三个存储器。计算单元主要用作 80188 微处理器的后端向量处理器。

NETSIM 是一种通用并行神经计算机。它可模拟大部分神经网络模型,如 Hopfield 和 BP 模型。

#### 3.2.4 KD-GP<sup>2</sup>N<sup>2</sup>S<sup>2</sup> 并行神经计算机系统

1)KD-GP<sup>2</sup>N<sup>2</sup>S<sup>2</sup> 硬件系统结构。它由两部分组成:①一台 IBM PC/386 作为宿主;②四片 T800 构成 Transputer 阵列。其中 PC 机的主要功能是文件管理。Transputer 阵列的主要功能是高速、并行地运行各个神经网络模拟软件,是系统的核心。

2)系统的软件构成。整个系统软件可划分为两个组成部分:①运行在 PC 机内的 PC 机服务管理软件;②运行在 Transputer 阵列内的 Transputer 模拟软件。

#### 3.2.5 Systolic 并行神经计算机系统

1)硬件体系结构。如图 4 所示,其中:HOST 主要负责与用户进行交互、用户程序的编译和加载。处理单元 PE 将根据从 GCU 发来的宏指令运行相应的用户程序,它是完成计算的实体。GCU 全局控制部件是系统的核心。它从 HOST 获取用户程序,并根据用户程序发出宏指令 MI,从而控制各个 PE 的执行;它还控制与 PE 相连的数据通路,向 PE 发送数据和加载程序,并从各个 PE 上收集结果。

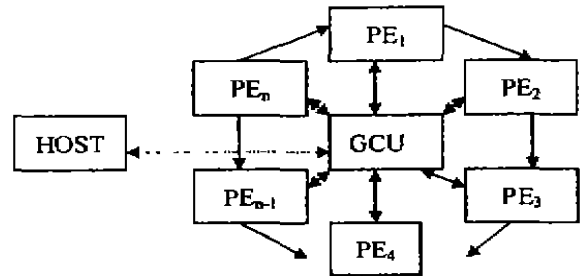


图 4

2)软件体系结构。软件主要包括以下部分:①HOST 上的集成环境;②神经网络描述语言编译器;③GCU 和 PE 上的监控程序。

## 4 未来的计算机<sup>[9,10]</sup>

### 4.1 光神经计算机

光神经计算机的研究是将光物理学和光技术应用到神经网络理论,并探讨用光学或光电混合器件实现神经网络硬件系统的方法,通过光技术和神经网络结合,形成一种新的“光+连接”的信息处理世界。

光神经计算机具有三个重要特征:(1)光具有并行性,它与神经计算吻合;(2)光波的传播无交叉失真,传播容量大;(3)可能实现超高速运算。其中第一点最重要,由于神经元数目的增加,使神经元之间的连接数目呈指数增加,用硅 VLSI 实现<sup>[12]</sup>有困难,而用光技术并利用其空间分布技术,可实现神经元之间的大量连接。

### 4.2 分子计算机

(下转第 15 页)

表7 NHPCC-NOW 上 PSRS 算法运行时间 ( $E'=0.43$ )

处理器数	1	2	4	8	16
执行时间(s)	1.89	1.89	46.67	521.99	880.41

表8 NHPCC-NOW 上 PSRS 算法可扩放性表示 ( $E'=0.43$ )

Time scale ( $E, (P, P')$ )	1	2	4	8	16
1	1.0000	1.0000	0.0405	0.0036	0.0036
2		1.0000	0.0405	0.0036	0.0021
4			1.0000	0.0894	0.0530
8				1.0000	0.5929
16					1.0000

参考文献

- 1 Kumar V, Gupta A. Analyzing Scalability of Parallel Algorithms and Architectures In: Proc. of the 1991 Int Conf. on SuperComputing, Cologne, Germany, June 1991
- 2 Zhang X D, Yan Y, He K. Latency Metric: An Experimental Method for Measuring and Evaluating Parallel Program and Architecture Scalability. Journal of Parallel and Distributed Computing, 1994, 22: 392~410

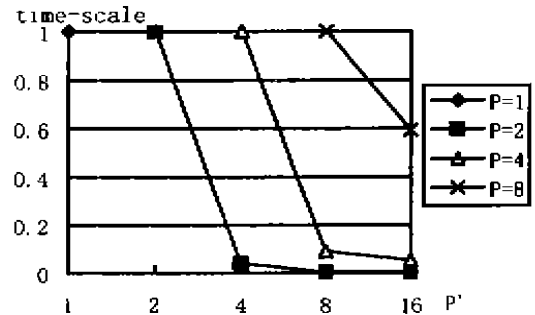


图5 NHPCC-NOW 上 PSRS 算法可扩放性曲线图 ( $E'=0.43$ )

- 3 Sun X H, Rover D T. Scalability of Parallel Algorithm-Machine Combinations. IEEE Trans. Parallel Distributed Systems, 1994, 5(6): 599~613
- 4 曙光-1000 软件系统技术报告. 国家智能计算机研究开发中心, 1995
- 5 Message Passing Interface Forum. MPI: A message-passing interface standard. International Journal of Supercomputer Applications, 1994, 8(3/4)
- 6 陈国良. 并行算法设计与分析. 北京: 高等教育出版社, 1994
- 7 Shi H, Schaeffer J. Parallel Sorting by Regular Sampling. Parallel and Distributed Computing, 1992, 14(4): 316~372
- 8 Ji Yongchang Cluster of NHPCC: [Technique report] CS of USTC, 1997

(上接第 11 页)

分子计算机的基本计算方法是把一种物理识别作为运算的技术, 其特点如下:

(1) 分子计算由蛋白质和它们所综合的系统结构实现, 即不需要象数字电子计算机那样编程;

(2) 分子活动的大规模并行特征可以提高计算效率, 故比电子开关慢五个数量级的单分子速度不会影响整体的运算效率;

(3) 分子计算机属于一种基质感触识别, 因此它依照几何的方法可以实现与上下文有关的自适应信息输入, 而无需按位串行处理。

分子计算机的实现途径可分为两种:

(1) 分子计算机受目前电子计算机的启发, 从分子到电子线路设计与制作出发, 形成一种体积小、速度快、存储量大而成本低的数字分子计算机;

(2) 分子计算机模仿生物的图像信息处理功能, 即快速的模式识别、自组织和自学习能力。

光神经计算机和分子计算机的研究程度及其成果是无法和电子计算机相比拟的, 然而它所发现的新颖计算理论与信息处理的结构体系已引起了生物学家、化学家、物理学家以及计算机学家的极大兴趣, 这会对智能计算机的研究具有极其深远的意义。

参考文献

- 1 Nielsen R H. Neurocomputing; Picking the Human Brain. IEEE Spectrum, Mar. 1988. 36~41
- 2 Perkov N. Systolic simulation of multilayer, feed forward neural networks. Parallel Processing in Neural Systems and Computers. In: R. Eckmiller, G. Hartmann and G. Hauske, eds. Elsevier Science Publishers B. V (North-Holland) 1990. 303~306
- 3 Duncan R. A survey of parallel computer architectures. IEEE computers, 1990, 23(Feb.): 5~16
- 4 Hecht-Niesen R. Performance limits of optical, electrooptical, and electronic neurocomputers. SPIE optical and hybrid computing, 1986, 634
- 5 Kuczewsk R, Myers M, Crawford W. Neurocomputer workstation and processors; Approaches and Applications. IEEE Intl. Conf. Neural Networks, 1988, 3: 487~500
- 6 Hecht-Niesen R. Neurocomputer Interfaces and Performance Measures. IEEE Symp. Circuit and System, 1989. 74~77
- 7 Hecht-Niesen R. Neurocomputing. Addison-Wesley Publishing Company, Inc., 1990
- 8 庄镇全, 王煦法, 王东生. 神经网络与神经计算机. 电子科学技术, 1990(4~11)
- 9 庄镇全, 王煦法, 王东生. 神经网络与神经计算机. 科学出版社, 1992
- 10 戴葵. 神经网络实现技术. 国防科学大学出版社, 1998
- 11 王汝笠, 章明, 周斌. 第六代计算机——人工神经网络计算机. 科学技术出版社, 1992
- 12 Simon Y F, Lisa R A, Yoshiyasu T. Analog Components for the VLSI of Neural Networks. IEEE Circuits and Devices, July 1990