一种可扩展BP在片学习神经网络芯片

卢 纯,石秉学,陈 卢
(清华大学微电子学研究所,北京 100084)

摘 要: 基于0.6μm标准CMOS工艺,设计并实现了一种可扩展BP在片学习神经网络芯片.该芯片包含8个神经元和64个突触.提出了一种新颖的可扩展拓扑结构,使得利用该芯片构成完整的神经网络系统时,不需附加额外的神经元误差计算芯片;将 L 个芯片层叠起来就可以得到一个 L 层的神经网络.该芯片采用模拟电路,利用电容进行电荷存储,在片学习本身可用于权重刷新以保证权重值的正确性.奇偶校验实验证明了该神经网络芯片具有在片学习的能力.

关键词: 神经网络;在片学习;CMOS模拟集成电路;误差反传算法
中图分类号: TP18 **文献标识码:** A **文章编号:** 0372-2112 (2002) 09-1270-04

An Expandable BP On-Chip Learning Neural Network Chip

LU Chun, SHI Bing-xue, CHEN Lu
(*Institute of Microelectronics, Tsinghua University, Beijing* 100084)

Abstract: An expandable BP on-chip learning neural network chip is designed and fabricated with a standard 0.6 μm CMOS technology. It includes 8 neurons and 64 synapses. A novel expanable topology is proposed so that no additional neuron error computation chip is needed to construct a whole neural network system. A neural network with L-layers can be composed by cascading L such chip. Analog circuits are used in this chip and capacitors are adopted to store weight values. The on-chip learning itself can be used as a refreshable tool to keep weight values right. The experiment of the parity check demonstrates the on-chip learning ability of the neural network chip.

Key words: neural networks; on-chip learning; CMOS analogue integrated circuits; Back-Propagation algorithm (BP)

1 引言

近年来,随着微电子技术突飞猛进的发展,VLSI技术被公认为是最能体现人工神经网络特点及目前最有效实现人工神经网络的方式.由于模拟电路在面积、速度、功耗、上都存在有优势,且更能发挥神经网络在并行性、实时性、分布性、容错性等方面的特点;而在片学习与离片学习和环片学习两种方式相比,又具有全并行、自适应、异步实时运算等方面的优势;所以,研究在片学习神经网络的模拟VLSI实现方式具有重大的意义,一些论文已从事了这方面的研究[1-9].BP算法是在片学习最常用的一种算法.它是已有成熟理论基础、应用广泛的一种梯度下降学习算法.它的优点在于所有的权重都是并行调整的,权重的调整只需一些局部的信息,其VLSI实现只需标准CMOS工艺.本文采用BP算法进行在片学习,设计并实现了一种新颖的可扩展BP在片学习神经网络.

不同场合需要不同的神经网络规模,这就要求神经网络芯片具有可扩展性.论文[2]提出了一种可扩展的在片学习神经网络芯片.但利用它的芯片搭建一个完整的在片学习系统时,需要附加专门的,不同于其它各层的输出层神经元误差计算芯片.而用本文提出的芯片构成系统时,不需附加其它芯片;将 L 个芯片层叠起来就可以得到一个 L 层的神经网络.

本论文提出的芯片主要由神经元阵列、突触阵列和误差发生器阵列构成,本文将对这些基本单元分别进行阐述.

2 芯片系统结构

芯片具有规整的网络拓扑结构,如图1所示.它包含一个 8×1 神经元"N"阵列、一个 8×8 突触"S"阵列和一个 8×1 误差发生器"E"阵列.图中,X_{in} 是网络输入信号,T 是目标信号,X_{out} 是网络输出信号,E_{out} 是输出到前一层的神经元误差信号,E_{in} 是从后一层输入的神经元误差信号,CFG 是配置信号(输出层设为1,其它层设为0).

神经元"N"的结构如图2(a)所示,它将输入电流 s 按 S 形函数的非线性关系转换为输出电压 x.神经元的另外两个输出端 x_1、x_2 接到下级乘法器的差分输入端,利用中心差分法输出 S 形激活函数的近似导数.图中的 R_M 为MOS管形成的压控电阻,V_c 为其控制信号,通过调整 V_c 的值,可以实现 S 形函数增益因子的编程.I_b 为偏置电流,改变 I_b 可改变 S 形

收稿日期:2002-03-20;修回日期:2002-06-08
基金项目:国家自然科学基金(No.69636030)

函数的阈值.

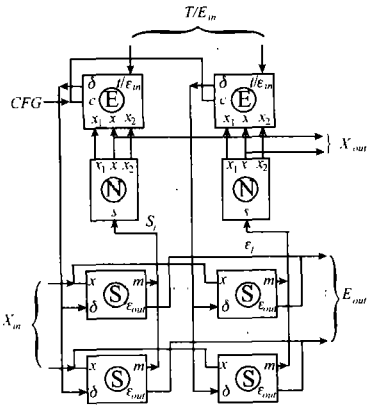


图 1　芯片拓扑结构

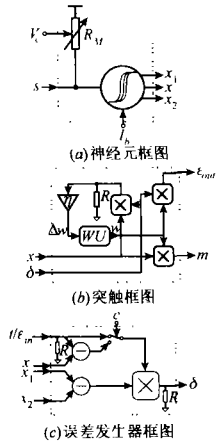


图 2　各模块结构示意图

图 2(*b*)为突触模块"*S*"的结构框图.它完成 4 项功能:(1)将权重值 w 与输入信号 x 相乘,产生信号 m,将若干 m 连起来就得到加权和信号 $s_j = \sum_i w_{ij}x_i$;(2)将权重值 w 与权重误差信号 δ 相乘,得到 ε_{out},ε_{out} 为一个电流信号,许多这样的 ε_{out} 相连,就得到神经元误差信号 $\varepsilon_i = \sum_j w_{ij}\delta_{ij}$;(3)将输入信号 x 与权重误差信号 δ 相乘得到的积放大 η 倍后得到权重更新值 Δw;(4)利用权重更新电路 *WU* 更新权重值.

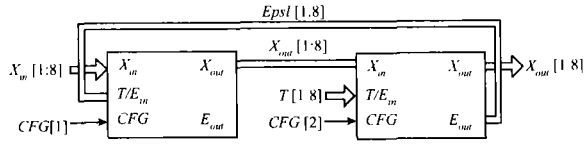


图 3　芯片的多芯片配置

误差发生器模块"E"的结构如图 2(*c*)所示.它主要用于产生权重误差信号 δ.t/ε_{in} 是一个复用端口.输出层神经元误差的计算方法与其它层不同,因此不少文献[2]在设计可扩展神经网络芯片时,不得不设计额外的输出层神经元误差计算芯片.本论文则通过设置一个控制信号 c 的值来控制电路进行不同的权重误差运算,从而无需添加额外的输出层神经元误差计算芯片,就可构成一个完整的 BP 在片学习神经网络系统.当该误差发生器模块位于输出层时,$c = 1$,目标值 t 输入到 t/ε_{in} 端口,信号 δ 由 $(t - x)$ 与 $(x_1 - x_2)$ 相乘得到;当该误差发生器模块位于其它层时,$c = 0$,神经元误差信号 ε_{in} 输入到 t/ε_{in} 端口,信号 δ 由 ε_{in} 与 $(x_1 - x_2)$ 相乘得到.这样用相同的硬件结构,不同的控制信号就可完成不同权重误差的计算.

大规模神经网络可由若干这样的芯片按一定的拓扑结构构成.图 3 为一个多芯片配置示意图.它是一个拓扑结构为 8 - 8 的在片学习系统.$X_{in}[1:8]$ 代表 $X_{in}[1]$、$X_{in}[2]$、…、$X_{in}[8]$ 一组共 8 个输入信号,$T[1:8]$ 是目标信号,$X_{out}[1:8]$ 是输出信号.*CFG* [1]是第一层的配置信号,等于 0;*CFG* [2]是输出层的配置信号,等于 1.*Epsl* [1:8]是神经元误差反传信号.X_{out1} [1:8]是第一层的神经元输出信号,也是第二层的输入信号.总的来说,一个 L 层(不包括输入层,因为输入层只是一些输入节点的集合,无需进行加权和、非线性转换等运算)的神经网络,需要 L 个芯片,每层的神经元个数最多为 8 个.

3　单元电路设计

3.1　神经元模块

神经元电路[7-9]如图 4 所示.它产生可编程 *S* 型激活函数及其导数.V_{dd} 为 0.9V 电压源,V_{ss} 为 - 2.5*V* 电压源.V_N、V_P、V_{ref1}、V_{ref2}、V_d、V_{d1} 和 V_{d2} 都是固定的偏置电压.其中,$V_{d2} - V_d = V_d - V_{d1} = \Delta V$,$\Delta V$ 是一个小的正电压.虚框中的 N 管 M_1 和 P 管 M_2 工作于线性状态,组成线性可调电阻 R_M,V_N 和 V_P 为其控制信号.N 管 M_{11}、M_{12} 组成的差分输入放大器和 P 管 M_3、M_4 组成的负载一起实现 *S* 型函数非线性 *I-V* 变换.

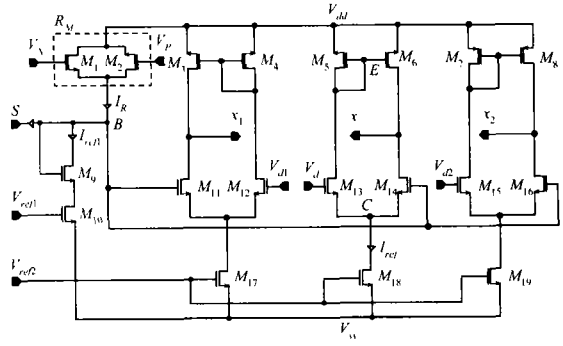


图 4　神经元电路图

设 x_j^l 是神经网络第 l 层第 j 个神经元的输出,k 是当前迭代次数,那么,神经元的功能可由下式表示:

$$x_j^l(k) = f(s_j^l(k)) \tag{1}$$

其中,双曲正切 *S* 型函数,$f(s) = \dfrac{e^{\alpha(s+\theta)} - e^{-\alpha(s+\theta)}}{e^{\alpha(s+\theta)} + e^{-\alpha(s+\theta)}}$,$\alpha$ 为增益因子,θ 为阈值.改变 V_N、V_P 的值起到调节增益因子大小的作用.$(V_N - V_P)$ 越大,R_M 越小,V_B 随输入电流变化的斜率越小,x 上升越缓慢,增益因子 α 越小.另外,改变 V_{ref1} 或 V_d 可以调节阈值.增大 V_{ref1} 或 V_d,转移曲线及其导数曲线向左平移;反之,则向右平移.

其它两组差分对也用来实现非线性 $I - V$ 变换,由于它们的固定电平端 V_d、V_{d1} 和 V_{d2} 存在微小差异,它们的输出电平也就存在微小差异.利用这些差异,运用中心差分法,转移函数的导数函数可由 $(x_1 - x_2)$ 得到.

在 VLSI 神经网络中,一个神经元总是跟很多的突触相连.在这种重负载的情况下,神经元的瞬态特性很重要,它影响到整个神经网络的速度.如图 2 中所示推挽输出级的应用使该神经元在充放电过程中都具有较强的驱动能力[8].

3.2　突触模块

突触模块的电路如图 5 所示.图中"M"是模拟乘法器电路,"Amp"是简单的二级运放,"*Rdw*"和"*Rw*"是阻值不同的由 MOS 管形成的电阻,"M_0"为 NMOS 管,"C_M"是作为存储单元的电容."*Amp*"、"C_M"、"M_0"和"*Rw*"构成权重的存储和迭代单元,完成对权重 w 的初始化、更新和存储.x 为前级神经元的输出,乘法器 I1 将 x 与 w 相乘,得到 m.不同突触的 m 连到 s_j 上(如图 1 所示),完成加权和功能,如下式所示:

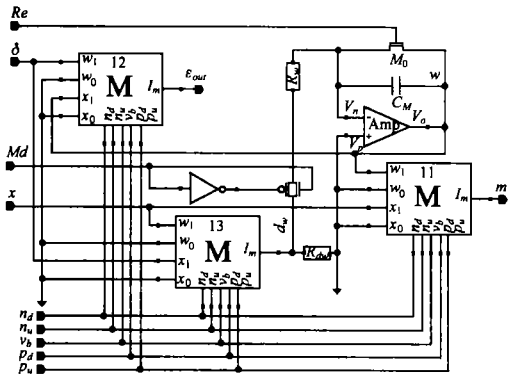


图 5 突触模块电路图

$$s_j^l(k) = \sum_i w_{ij}^l(k) x_i^{l-1}(k) \tag{2}$$

δ 为输入的神经元误差,它与 w 通过 I_2 相乘,得到 ε_{out},若干突触的 ε_{out}相连,完成下式的运算功能:

$$\varepsilon_{i,r}^l(k) = \sum_j w_{ij}^l(k) \delta_{j,r}^{l+1}(k), \qquad 1 \leqslant l < L \tag{3}$$

δ 与 x 通过乘法器 I3 相乘,得到权重改变量 dw,即完成了以下运算:

$$dw_{ij}^l(k+1) = \delta_j^l(k) x_i^l(k) \tag{4}$$

当 *Md* 信号为高电平时,权重迭代单元会根据 dw 的值根据下式进行实时的权重迭代,芯片的在片学习处于训练阶段:

$$w = -\frac{1}{R_w C_M} \int_0^t dw dt \tag{5}$$

当 *Md* 信号为低电平时,权重停止迭代,在片学习处于测试阶段.给权重赋初值时,芯片外部有一个控制信号 S(控制若干模拟二选一开关的开关状态)被置为高电平,使得此时输入的 x 不是测试样本的值,而是某预设值 x_0,同时 δ 也是某预设值 δ_0;设权重赋初值的时间为 Δt,只要选择适当的 x_0 和 δ_0,权重就会被置成想要的初值 w_0.权重赋完初值后,S 由高电平变为低电平,使得 x 输入训练样本的值,δ 输入由误差发生器得到的权重误差值,在片学习系统开始进行权重迭代.

突触中的乘法器是神经网络中重复利用次数最多的基本单元.该芯片采用了一种改进型的 Gilbert 乘法器,它相对于原始 Gilbert 乘法器的好处在于它的零点偏差可通过零点补偿电路得到校正,由于芯片在流水过程中有一些不可控的因素影响乘法器零点,而且乘法器的零点偏差对于在片学习是负面性的.所以这种改进虽然增加了电路复杂度,却有利于提高在片学习的收敛率.具体电路形式和电路分析,可参照我们以前的论文[9].

3.3 误差发生器模块

误差发生器模块电路如图 6 所示.如果它位于输出层,控制信号 *CFG* 将被置 1,目标值 t 通过端口 t/ε_{in}传输进来,δ 由 $(t-x)$与(x_1-x_2)相乘得到;否则,*CFG* 清为 0,ε_{in}从前面一层传输进端口 t/ε_{in},δ 由 ε_{in}与(x_1-x_2)相乘得到.所以一个完整的在片学习系统无需额外的输出层就可实现.如前面 3.1 中解释的那样,(x_1-x_2)可近似看作是 S 型函数的导数 $f'(s)$,所以误差发生器模块的功能可由下式表示:

$$\delta_j^l(k) = \begin{cases} f'(s_j^l(k))(t_j - x_j^l(k)), & l = L \\ f'(s_j^l(k))\varepsilon_j^{l+1}(k), & 1 \leqslant l < L \end{cases} \tag{6}$$

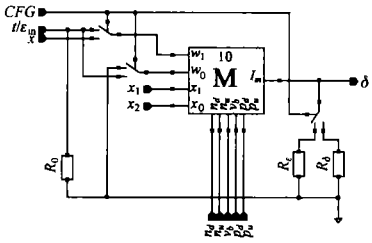
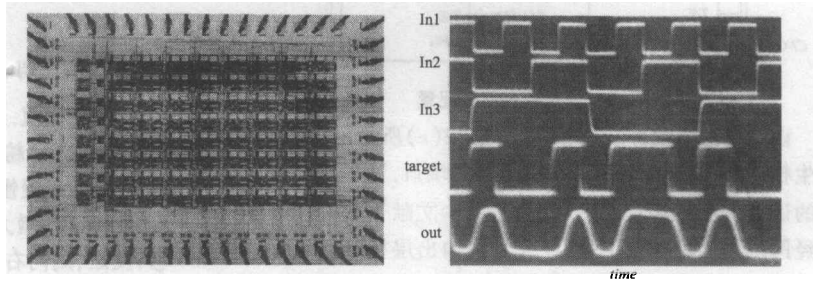


图 6 误差发生器电路图



(a)芯片照片 (b)3 位奇偶校验波形

图 7 实验结果

4 实验结果

采用 0.6μm 标准 CMOS 工艺,对该芯片进行了流水,芯片照片如图 7(*a*)所示.测试表明,神经元产生的非线性转换与理想 S 型函数[7]的误差不超过 5%,电路产生的导数函数与理想导数函数间的误差不超过 7%.乘法器在输入[-1V,1V]范围内的非线性度不超过 6%,其动态输出范围是[-18μA,18μA].对该神经网络芯片做了奇偶校验实验.图 7(*b*)为 3 位奇偶校验实验的波形.该神经网络的计算速率可达 130MCPS,可在 2ms 内完成 3 位奇偶校验的学习过程.而测量表明用于存储权重的电容的电荷泄漏速度约为 10^{-3}V/s.所以学习电路本身可作为权重刷新的工具来保证权重值的正确性.由电荷泄漏速度以及芯片的权重取值范围(-1V,1V)可知,若每隔 1s 刷新一次,权重精度可达 2^{-11},若每隔 10s 刷新一次,权重精度可达 2^{-8}.

5 结论

采用 0.6μm 标准 CMOS 工艺,硬件实现了一种可扩展 BP 在片学习神经网络芯片.它采用一种新颖的可扩展拓扑结构,使得利用该芯片构建完整的神经网络系统时,不需附加额外的神经元误差计算芯片;将 L 个芯片层叠起来就可以得到一个 L 层的神经网络.该芯片主要由神经元阵列、突触阵列和误差发生器阵列构成.神经元单元完成 S 型函数 *I-V* 非线性转换,并同时产生其导数;它具有良好的 DC 特性、瞬态特性和可编程性.突触单元初始化、刷新和存储权重;得到加权和信号和神经元误差信号.误差发生器产生权重误差信号.测试结果表明这些单元性能良好.该芯片采用模拟电路,利用电容进行电荷存储,在片学习本身可用于进行权重刷新来保证权重

值的正确性.3 位奇偶校验实验结果证明了该神经网络芯片具有在片学习能力.

参考文献：

[1] 卢纯,石秉学,陈卢.一种学习速率自适应的可编程片上学习 BP 神经网络电路系统的设计[J].电子学报,2001(5):701 - 703.

[2] C Lu, B X Shi, L Chen. An on-chip BP learning neural network with ideal neuron characteristics and learning rate adaptation [J]. Analog Integrated Circuits and Signal Processing, 2002(31):55 - 62.

[3] C Lu, B X Shi, L Chen. Circuit design of on-chip BP learning neural network with programmable neuron characteristics [J]. Chinese Journal of Semiconductors, 2000, 21(12):1164 - 1169.

[4] T Morie, Y Amemiya. An all-analog expandable neural-network LSI with on-chip back-propagation learning [J]. IEEE Trans. Neural Networks, 1996, 7(2):346 - 361.

[5] M Valle et al. An analog VLSI neural network with on-chip back propagation learning [J]. Analog Integrated Circuits and Signal Processing, 1996(9):231 - 245.

[6] Y Berg, et al. An analog feed-forward neural network with on-chip learning [J]. Analog Integrated Circuits and Signal Processing, 1996 (9):65 - 75.

[7] C Lu, B Shi. Circuit design of an adjustable neuron activation function and its derivative [J]. Electron, Lett, 2000, 36(6):553 - 555.

[8] C Lu, B X Shi, L Chen. Push-pull output neuron circuit [J]. Electron, Lett, 2001, 37(25):1531 - 1533.

[9] C Lu, B X Shi, L Chen. Hardware realization of building blocks for artificial neural networks [A]. ICSICT-2001 [C]. 上海:人民邮电出版社, 2001. 123 - 126.

作者简介：

卢　纯　女,1975 年 12 月出生,籍贯江苏,1997 年取得清华大学电子工程系微电子专业工学学士学位.同年获得直接攻读清华大学博士研究生资格.现为清华大学博士研究生,从事人工神经网络的 VLSI 实现和模拟集成电路研究.

石秉学　男,1936 年 2 月出生,教授,博士生导师,从事人工神经网络、模糊逻辑系统及其 VLSI 实现,RF 电路,模拟和数/模混合集成电路与系统研究.