

# 一种模拟集成电路 Hamming 神经网络及其应用\*

李斌桥 李志坚 石秉学

(清华大学微电子学研究所 北京 100084)

**摘要** 采用 P 阵 CMOS 工艺设计并制作了一种模拟型的 Hamming 神经网络集成电路。测试结果表明, 网络中的宏单元, 模板匹配运算电路, 可很好地实现容差匹配的功能。将芯片用作识别器, 识别手写体阿拉伯数字的实验结果表明, 可在 500ns 内完成一次识别运算; 对所采集的数据集得到了较好的识别效果。识别测试结果还表明, 识别效果对控制参量阈值电压  $V_T$  的变化较敏感。

EEACC: 1295, 1290, 1160

## 1 引言

自 Hopfield 建立新的网络结构, 以及许多新的学习和训练算法出现以来, 近十多年中, 人工神经网络一直是一个热门的研究领域<sup>[1,2]</sup>。用这种方法解决问题的特点是, 我们无须对要求解的问题提出精确的数学模型, 只要利用现有的神经网络模型和相应地学习、训练算法, 按所要达到的要求对网络进行训练, 就可得到问题的解。神经网络理论已成为解决许多人工智能问题的极有潜力的方法。但基于软件的神经网络系统不能够满足许多要求实时处理的应用场合, 因此设计与制作 VLSI 神经网络系统便成为一个重要的研究课题。VLSI 神经网络分为数字电路和模拟电路两种形式。数字电路结构的神经网络实际为某种特殊结构的数字计算机, 由于采用成熟的数字电路设计与制作技术, 电路规模可以做得较大, 甚至可做到圆片(wafer)级的, 包含上千个神经元的网络芯片<sup>[3]</sup>。与数字电路相比, 模拟集成电路中, 器件特性利用充分, 可以直接将某种神经网络的结构映射到电路中, 实现高速的并行运算, 是一种很有潜力的方法。

由于许多神经网络中神经元间的连接密度高, 并且要求连接权值可灵活调节, 因而用

\* 国家自然科学基金资助项目和高校博士点基金资助项目

李斌桥 男, 1956 年生, 博士, 天津大学电子工程系副教授, 从事人工神经网络及模糊逻辑系统的集成电路实现、ASIC 设计、半导体器件的计算机模拟研究

李志坚 男, 1928 年生, 教授, 中国科学院院士, 从事微电子学领域内的研究工作

石秉学 男, 1936 年生, 教授, 从事人工神经网络及模糊逻辑系统的集成电路实现、模拟集成电路研究

1994 年 10 月 24 日收到初稿, 1995 年 3 月 13 日收到修改稿

VLSI 技术制作包含神经元数目较多的网络是比较困难的。Hamming 神经网络的结构简单，并且在许多应用中，连接权值可预先设定后固化到电路中，因而是集成电路神经网络中的一种优选结构<sup>[4]</sup>。我们设计了一种结构简单、便于制作的 Hamming 神经网络模拟集成电路模型，并采用 P 阵 CMOS 工艺制作了电路，成功地用于字符识别的实验中。

## 2 电路设计

典型的 Hamming 神经网络为二层结构，如图 1 所示<sup>[2]</sup>。在模式识别应用中，网络中的

第一层起计算输入模式同所有模板匹配距离的作用。第二层是一个求大网络。这是一种结构较简单、训练方便的网络。

图 1 中，最上边一层的方框点是输入象素点，第一层中每个神经元的输入同所有象素点的输入相连，在每个神经元的这一组连接权中存入一个识别对象的模板，如对于字符识别，则存入一个字形模板。神经元的作用就是计算输入字符同其连接权中所存模板的匹配距离，匹配程度由神经元的输出表示，输出为一模拟量，其与匹配距离之间为某种单调递增或递减的函数关系。

第二层是一个求大网络，也称 Winner-Take-All(WTA) 网络，网络的输出端数同输入端数相等，并为一一对应的关系。该网络的作用是找出输入端中输入值最大的一端，并在相应的输出端输出状态 1，而其它端的输出为状态 0，在识别应用中，该层的作用是进行最后的识别判决。

根据上述 Hamming 神经网络结构及实现的算法，我们设计了一集成电路实现方案。该电路的输入和输出都为数字信号，内部是模拟电路，并以电压方式工作。

图 2 是网络第一层中神经元及其连接的电路模型。这是由一系列上拉-下拉 MOS 管对组成的模板匹配运算电路，电路中所存模板反映在输入端的连接方式上，存 0 端直接同输入相连，存 1 端与输入端隔一反相器，在图 2 的示例中，模板中存有 1010…1，当以高低电平表示的输入信号同模板完全匹配时，所有 NMOS 管截止，PMOS 管导通，神经元的输出为最大电压  $V_{DD}$ ，若有一端不匹配，一个 NMOS 管导通，一个 PMOS 截止，将输出电压拉下去一些，不匹配端数越多输出电压越低。设每个导通的 PMOS 管的电阻为  $r_p$ ，导通 NMOS 管的电阻为  $r_n$ ，总输入端数目为  $n$ ，神经元的传输函数，即输出电压与不匹配点数的关系可表示为

$$V_o = \frac{(n - i)r_n}{(n - i)r_n + i r_p} V_{DD} \quad (1)$$

若  $r_n$  和  $r_p$  都为常数，并取  $r_n = r_p$ ，则有

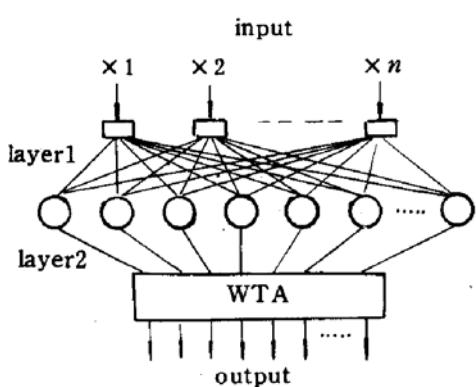


图 1 Hamming 神经网络

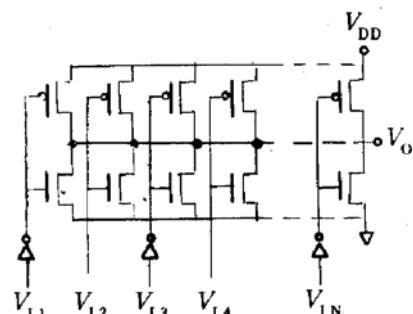


图 2 模板匹配运算电路

$$V_o = \left(1 - \frac{i}{n}\right) V_{DD} \quad (2)$$

这种情况下,输出电压同不匹配点数  $i$  是一种线性递减的函数关系。但实际上对于 MOS 管,在不同的漏源电压下,管子将工作在非饱和区、饱和区等不同状态,导通电阻不是常数,因此输出电压与不匹配点数不是一种简单的单调递减函数关系。在实际应用中,网络所要存入的模板是由训练确定的,这部分工作可由软件来完成,当模板选定好后,可据此设计专用电路。

设计了一种 Hopfield 型求大网络,实现该 Hamming 神经网络中第二层的功能,电路示于图 3。虚线框内就是一个神经元的电路,其中,上面的一个加有固定栅偏压  $V_p$  的 PMOS 管用作负载,下面的各个 NMOS 管的栅极对应神经元的输入端,并且分别接到其它各神经元的输出端。神经元间的这种相互连接就构成了一种强抑制连接权,因为当某一神经元的输出电压较高时,由这种连接形成的负反馈作用就促使其它神经元的输出电压下降。每个神经元的输出到其自身的输入端并未设增强型的连接权,但当其它输出端的电位降低时,最大输出端的电位就会自然抬高,因此使最大输出端电位增强的作用是自然存在的。求大网络的工作过程是,在时钟信号  $clk2$  的控制下输入待比较的信号,对网络各输出端点置初值,之后网络在负反馈机制的作用下,由不稳定态达到稳定态,此时初始值最高的输出端输出高电平,其它端输出低电平。由时钟信号  $clk1$  控制的 NMOS 开关管,其作用是在输入待比较信号之前,先将各输出端节点的电位拉平,以便在同一起点上置初值,这样可提高电路的分辨精度。

若用作识别器,该网络实现的算法可用下述方程描述。

$$V_{max} = \text{MAX} \{f(d_k), k = 1, 2, \dots, p\} \quad (3)$$

$$d_k = \sum_{j=1}^m \sum_{i=1}^n |M_k(i, j) - X(i, j)|$$

$$V_o = \begin{cases} V_{DD} & V_{max} > V_T \\ 0 & V_{max} \leq V_T \end{cases} \quad (4)$$

式中  $p$  是电路中所存模板数,  $M_k(i, j)$  和  $X(i, j)$  分别是点阵模板和输入的待识模式点阵;  $d$  即为输入模式同模板的匹配距离,其值愈小,匹配程度愈好,相应模板匹配运算电路的输出电压就愈高; $V_o$  是求大网络的输出电压; $V_T$  是控制误识率的阈值电压; $f(x)$  是模板匹配运算电路的传输函数。

### 3 宏单元电路的测试结果

我们采用中国科学院微电子中心  $3\mu\text{m}$  级 P 阵 CMOS 工艺制作了上述的 Hamming 神

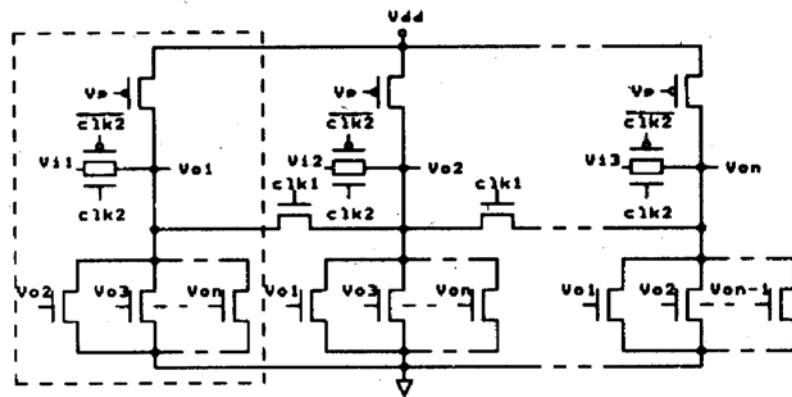


图 3 求大网络电路

经网络和其中的各种宏单元电路，详细研究了电路特性。

为研究网络第一层中匹配运算电路的工作特性，制作了含有十个输入信号端的实验电路，为降低功耗，电路中 MOS 管设计为栅长大于栅宽的结构，栅长为  $20\mu\text{m}$ ，N 管栅宽为  $10\mu\text{m}$ ，P 管栅宽为  $15\mu\text{m}$ 。图 4 是实验测试结果，横坐标是不匹配点数，左边纵坐标是输出电压，右边纵坐标是功耗电流。实测结果表明，随不匹配点数的增加，输出电压逐渐降低，但在前后两段电压下降幅度较缓，在中间段电压下降较快。输出电压的这一变化规律是由于，在不同的匹配点数下，电路中的 PMOS 和 NMOS 管将工作在饱和及非饱和各不同状态下。在输出电压较高时 ( $V_o > V_{DD} - V_{TN}$ ,  $V_{TN}$  为 NMOS 管的开启电压)，导通的 PMOS 管和 NMOS 管分别处于非饱和与饱和工作状态。输出电压较低时 ( $V_o < -V_{TP}$ ,  $V_{TP}$  为 PMOS 管的开启电压)，导通的 PMOS 管和 NMOS 管分别处于饱和与非饱和工作状态。输出电压为中间值时 ( $-V_{TP} < V_o < V_{DD} - V_{TN}$ )，导通的 PMOS 管和 NMOS 管都处于非饱和工作状态，这时由于两管的导通电阻值相当，匹配点数的变化，就会引起输出电压的较大改变。输出电压与不匹配点数的这种单调递减函数关系符合对网络第一层中神经元功能的要求。对第二层的求大网络也单独进行了研究，结果已在另文介绍<sup>[5]</sup>。

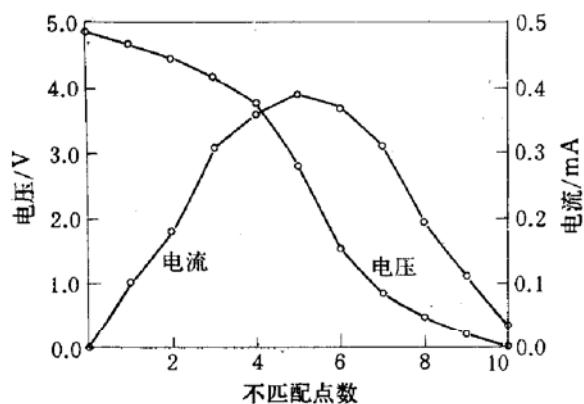


图 4 输出电压和功耗电流与不匹配点数关系的测量结果

NMOS 管分别处于饱和与非饱和工作状态。输出电压为中间值时 ( $-V_{TP} < V_o < V_{DD} - V_{TN}$ )，导通的 PMOS 管和 NMOS 管都处于非饱和工作状态，这时由于两管的导通电阻值相当，匹配点数的变化，就会引起输出电压的较大改变。输出电压与不匹配点数的这种单调递减函数关系符合对网络第一层中神经元功能的要求。对第二层的求大网络也单独进行了研究，结果已在另文介绍<sup>[5]</sup>。

## 4 神经网络芯片的测试和应用实验

为测试电路的特性，我们将该 Hamming 神经网络芯片用作识别器，识别手写体阿拉伯数字。我们采集了十个人以较规则的方式书写的 1000 个阿拉伯数字，其中每人写 100 个。对原始图象数据经细化和压缩变换处理后<sup>[6]</sup>，将每个样本用  $5 \times 5$  的二值点阵图象表示出。然后将样本等分为两部分，一部分用作训练数据集，另一部分用作测试数据集。在训练数据集中，用统计比较的方法为每一个目标模式确定一个字型模板，图 5(见图版 I)即为选定的模板，据此设计出具有固定模板的 Hamming 神经网络。

Hamming 神经网络芯片中共有十个模板匹配运算单元，每个都有 25 个输入端，一个输出端，每个模板存图 5 中的一个字符。芯片中的 WTA 网络包含十一个神经元，对应十一个输入端，其中十个输入端来自十个匹配运算网络的输出，另一个输入端用作阈值电压输入，控制识别率和误识率。WTA 网络共有十个输出端，对应 0~9 十个数字的识别判决输出。为隔离内部电路和提高输出驱动能力，每个输出端加有一个反相器，所以，如代表“2”的反相器输出端输出为低电平，则识别结果为字符 2，若所有十端的输出都为高电平，则表示拒识。版图设计中，MOS 管的最小栅长为  $5\mu\text{m}$ ，图 6(见图版 I)是芯片照片，管芯面积为  $5.99 \times 4.94\text{mm}^2$ 。

### 4.1 电路特性的测试

对电路的工作特性进行了测试。方法是，在输入端输入一个待识字符，观测输出端的工

作波形。图 7(见图版 I)是低电平输出,即识别输出端的工作波形,输出波形反映了芯片中求大网络由瞬态过渡到稳态的工作过程。在输出波形的低电平段,电路处于稳态,这时的输出表明识别结果。clk2 时钟信号的作用已在前面介绍。实际测试中,完成一次识别运算的时间约为 500ns,但这并不是芯片的最高工作速度,若减小时钟信号的脉宽,工作速度可进一步提高。芯片的功耗约为 20mW。

#### 4.2 识别测试

对所采集的手写体阿拉伯数字集进行了识别测试。方法是,在不同的阈值电压下,测量系统对所用数据集的识别率、误识率和拒识率三项指标。图 8(见图版 I)是对测试集的识别测试结果。表 1 为阈值电压取 3V 时,对测试集中各类样本的识别结果。表中的第一列为输入样本,第二列为样本数目,第三、四、五列分别是拒识率、误识率和识别率,最后一列为可靠率(识别率/[识别率+误识率]),最下面一行是总识别结果。

表 1 测试数据集的识别结果

| 输入样本  | 样本数目 | 拒识率 / % | 误识率 / % | 识别率 / % | 可靠率 / % |
|-------|------|---------|---------|---------|---------|
| 0     | 50   | 0       | 0       | 100     | 100     |
| 1     | 50   | 0       | 0       | 100     | 100     |
| 2     | 50   | 14      | 4       | 82      | 95.35   |
| 3     | 50   | 26      | 4       | 70      | 94.59   |
| 4     | 50   | 42      | 4       | 54      | 93.10   |
| 5     | 50   | 10      | 6       | 84      | 93.33   |
| 6     | 50   | 14      | 14      | 72      | 83.72   |
| 7     | 50   | 18      | 2       | 80      | 97.56   |
| 8     | 50   | 28      | 12      | 60      | 83.33   |
| 9     | 50   | 10      | 4       | 86      | 95.56   |
| Total | 500  | 16.2    | 5       | 78.8    | 94.03   |

识别测试结果可归纳为如下两点:

- 1) 阈值电压取值较低时,识别率高,拒识率低,但误识率也高,如当阈值电压为 2V 时,得到的识别率、误识率和拒识率分别为 88.2%、10.4% 和 1.4%。
- 2) 当阈值电压取较高值时,识别率和误识率都下降,拒识率上升,如当阈值电压为 3.5V 时,得到的识别率、误识率和拒识率分别为 66.8%、0.8% 和 32.4%。此外,识别率和误识率并不是等比例下降的,误识率下降的幅度要快得多,这表明识别的可靠性增加了,如对于阈值电压为 2V 和 3.5V 的两种情况,可靠率分别为 89.5% 和 98.8%。

## 5 总结

我们设计了一种 Hamming 神经网络的集成电路模型,采用 P 阵 CMOS 工艺制作了这种电路的实验芯片。电路具有结构简单,易于大规模集成实现的特点。对网络中的宏单元,模板匹配运算电路的测试结果表明,该单元电路可很好地实现容差匹配的功能。将芯片用作识别器,识别手写体阿拉伯数字的实验结果表明,该 Hamming 神经网络芯片可在 500ns 内完成一次识别运算;对所采集的数据集得到了较好的识别效果。识别测试结果还表明,识别效果对控制参量的变化较敏感,当阈值电压较低时,识别率高,拒识率低,但误识率也高;当阈值电压较高时,识别率和误识率都下降,拒识率上升,识别的可靠性增加。

## 参 考 文 献

- [1] J. J. Hopfield and D. W. Tank, Biol. Cybern., 1985, **52**:141~152.
- [2] R. P. Lippmann, IEEE Assp Magazine, 1987, **4**(2):4~22.
- [3] M. Yasunaga, N. Masuda, M. Yagyu *et al.*, IEEE J. Solid-State Circuits, 1993, **28**(2):106~114.
- [4] U. Cilingiroglu, IEEE J. Solid-State Circuits, 1993, **28**(1):59~67.
- [5] 李斌桥, 李志坚, 石秉学, “一种以 Hopfield 神经网络方式工作的多端比较器电路”, 第八届全国半导体集成电路硅材料学术会议论文集, 杭州: 1993, 369~370.
- [6] B. Q. Li, Z. J. Li and B. X. Shi, The VLSI Implementation of an one-pass thinning and a compressing algorithms, Proc. ICNNSP'93, Guangzhou, China, 1993, 970~973.

## An Analog Hamming Neural Network Integrated Circuit and Its Application

Li Binqiao, Li Zhijian and Shi Bingxue

*(Institute of Microelectronics, Tsinghua University, Beijing 100084)*

Received 24 October 1994, revised manuscript received 13 March 1995

**Abstract** An analog Hamming neural network integrated circuit has been designed and fabricated in P-Well CMOS processes. The experimental results show that the macro circuit cell, template matching calculation circuit, can realize the tolerance matching function as expected. One recognition calculation can be finished in 500ns when the neural network chip used as a classifier. For an application example of handwritten numeral recognition, satisfactory recognition results have been obtained. The recognition experiment also shows that the recognition effect depends tightly on the controlling parameter, threshold voltage  $V_T$ .

**EEACC:** 1295, 1290, 1160