

Boletim Técnico da Escola Politécnica da USP
Departamento de Engenharia de Sistemas
Eletrônicos

ISSN 1517-3542

BT/PSI/0009

Redes Neurais em VLSI

Antonio Ramirez Hidalgo
Francisco J. Ramirez Fernandez

São Paulo – 2000

FICHA CATALOGRÁFICA

Ramirez Hidalgo, Antonio

Redes neurais em VLSI / A. Ramirez Hidalgo, F.J. Ramirez Fernandez. – São Paulo : EPUSP, 2000.

p. – (Boletim Técnico da Escola Politécnica da USP, Departamento de Engenharia de Sistemas Eletrônicos, BT/PSI/0009)

1. Redes neurais (Inteligência artificial) 2. Circuitos integrados VLSI I. Ramirez Fernandez, Francisco Javier II. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos III. Título IV. Série

ISSN 1517-3542

CDD 006.3
621.395

REDES NEURAIS EM VLSI

ANTONIO RAMIREZ HIDALGO

FRANCISCO JAVIER RAMIREZ FERNANDEZ

Laboratório de Microeletrônica LME - EPUSP.

Caixa Postal 8174, CEP 01065 - 970 - São Paulo - SP - Brasil.

Fax: (011) 818 - 5585

e-mail: ahidalgo@lme.usp.br

jramirez@lme.usp.br

Abstract

In this work we present the design of artificial neural networks in VLSI. A synaptic array of two layers has been designed with multipliers and capacitors that storage the synaptic weights. These synapses circuits were designed with the technology CMOS 0.6 μ m of the AMS. This neural network will be used in the implementation of a GAS IDENTIFICATION SYSTEM, but not is restricted only to this application. The neural network can be trained for another type of applications.

1. Introdução

A pesquisa e atividades industriais no campo da instrumentação tem se preocupado geralmente com o projeto de, e experimentação com, novos sensores, novos materiais, novas tecnologias e novos métodos para analisar dados, processar sinais e imagens, identificação e sistemas de controle.

Em instrumentos tradicionais, incluídos os baseados em técnicas complexas para o processamento digital de sinais, as operações executadas sobre os sinais de entrada usualmente seguem estritamente algoritmos determinísticos. Inversamente, as técnicas de inteligência artificial abandonam tais formas de abordagem, permitindo assim tratar com procedimentos de medida de alto nível de abstração, como o processamento não algorítmico complexo, auto-validação de resultados de medição, e seleção automática dos muitos procedimentos de medidas para condições ambientais dadas.

As redes neurais artificiais realizam um tratamento inteligente dos dados, a qual geralmente é capaz de capturar o comportamento de um sistema sem precisar necessariamente de um conhecimento prévio. A solução desejada de um problema é aprendida através de exemplos em lugar de ser definida por instruções de um algoritmo. Os paradigmas computacionais adaptativos são técnicas adequadas para descrever a manipulação de entradas não algorítmicas tais como as que reproduzem o não determinístico, fuzzificação (fuzziness), a adaptabilidade, a habilidade de generalização, e o paralelismo típico de raciocínio humano. O uso de tecnologias adaptativas na área de instrumentação tem várias aplicações na indústria e na pesquisa como é o análise e manipulação de dados, em sistemas de identificação a predição e em complexos procedimentos de medidas até a instrumentação inteligente. Sua limitada complexidade computacional com respeito a outras aproximações tradicionais, é geralmente adequado para suportar aplicações de tempo real e **permite realizações de circuitos integrados analógicos e digitais**. A crescente pesquisa que se está realizando com as redes neurais artificiais e as diferentes aplicações que elas estão tendo em uma variedade de áreas científicas e comerciais nos levam a procurar soluções de maior desempenho para atender as diferentes tarefas propostas [1]. Baseados nestas condições e depois de fazer várias pesquisas bibliográficas chegamos à conclusão que se poderia melhorar este desempenho mediante um hardware dedicado, para aproveitar o inerente paralelismo massivo destas [2]. As redes neurais estão sendo utilizadas para o reconhecimento de padrões, para otimização e catalogação de dados com maior ênfase em software, já que as simulações no computador são mais fáceis de serem feitas que a implementação em hardware [3]. Nesta pesquisa se projetará e implementará em VLSI um tipo de arquitetura de rede neural artificial (ANN) e seu respectivo algoritmo de aprendizagem, o qual será implementado com circuitos analógicos e também com circuitos digitais.

As redes neurais dificilmente serão sistemas de uso geral, devido a que elas são treinadas para uma aplicação específica, mas podemos projetar circuitos individuais que quando são montados juntos formam um sistema de uma aplicação só. O circuito de redes neurais tem que ser projetado de forma que trabalhe de acordo com o algoritmo de aprendizagem. As partes que formarão as redes neurais são as sinapses e neurônios. Estas células sinápticas determinam a precisão, a área de silício e o consumo de potência do circuito integrado [4]. A presente pesquisa trata sobre a implementação em hardware das redes neurais, com o algoritmo de treinamento de retropropagação usando a tecnologia CMOS de 0.6 μm da AMS. O circuito da pesquisa será usado em um sistema identificador de gases, mas poderia ser usado para outros tipos de sistemas.

2. Objetivos

O objetivo desta pesquisa é melhorar e expandir as capacidades de sistemas de instrumentação tradicional e introduzir novos instrumentos para novas aplicações. A estratégia adotada nesta pesquisa está centralizada na consolidação de uma primeira versão de instrumentação portátil capaz de efetuar o reconhecimento de uma substância usando exclusivamente o aroma que dela emana. Usando experiências passadas, foi definida uma proposta para atuar em todos os aspectos técnicos envolvidos desde a geração dos sinais nos sensores até a configuração do sistema de reconhecimento de padrões com as redes neurais artificiais implementadas em um circuito integrado de aplicação específica. Cientes da complexidade do escopo global desta proposta e os riscos inerentes à sua implementação, optamos por estabelecer uma abordagem minimizando a redundância mas enfatizando a confiabilidade na funcionalidade do sistema. Nesse sentido o sistema foi idealizado conciliando a implementação parcial de neurônios e sinapses em VLSI harmonizados com as interfaces de conexão aos sensores em eletrônica discreta.

3. Projeto do Sistema

Se apresenta o projeto de um sistema identificador de gases formado por três grandes blocos, o primeiro bloco é a matriz de sensores, o segundo é o adaptador de sinal que vai acondicionar os sinais dos sensores e o terceiro bloco é a rede neural com alguns circuitos periféricos. O sistema é baseado na rede neural, portanto, se aplicará um procedimento para o projeto desta [5]. Com os dados obtidos pelos sensores de diferentes amostras formadas por aguardentes comerciais brasileiros se treinará a rede neural definindo a arquitetura, o algoritmo de aprendizagem e identificando os correspondentes pesos sinápticos. Este treinamento será feito com o programa comercial "NeuralWork" [6]. A matriz de sensores que é formada pela associação de sensores químicos da marca TAGUCHI. Depois de ter suficientemente bem definidos os parâmetros da rede neural se procede a projetar o segundo e terceiro bloco do sistema. O segundo bloco têm circuitos para o acondicionamento dos sinais dos sensores para a rede neural, seguidores de tensão, chaves analógicas, circuito comparadores de janela e circuitos que fazem a divisão de tensão. O terceiro bloco está composto pela matriz sináptica (Circuito Integrado), os neurônios, conversores de corrente a tensão e um circuito de refrescamento baseado em um FPGA ou um microcontrolador que tenha uma memória FLASH assim com também um conversor D/A para converter os dados binários, dos pesos, em sinais analógicos que possam ser armazenados nos capacitores das sinapses.

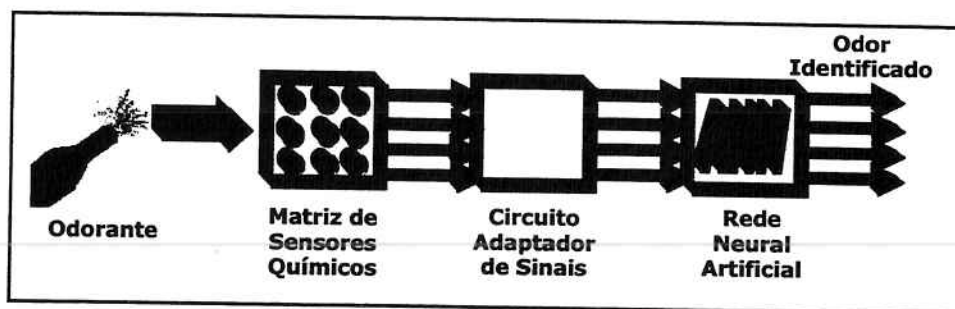


Figura 1. Diagrama Geral do Sistema Identificador de Gases

O sistema identificador de gases que foi projetado servirá para fazer medidas automáticas que identificarão os gases que entram nele em concordância com o treinamento da rede neural é formado por três grandes blocos os quais são mostrados na Figura 1. Este é um sistema de detecção e classificação automática de odores, vapores e gases.

4. Procedimento Geral para Implementar Redes Neurais Artificiais

Um diagrama em blocos é apresentado mostrando o procedimento para a implementação das redes neurais para uma aplicação específica.

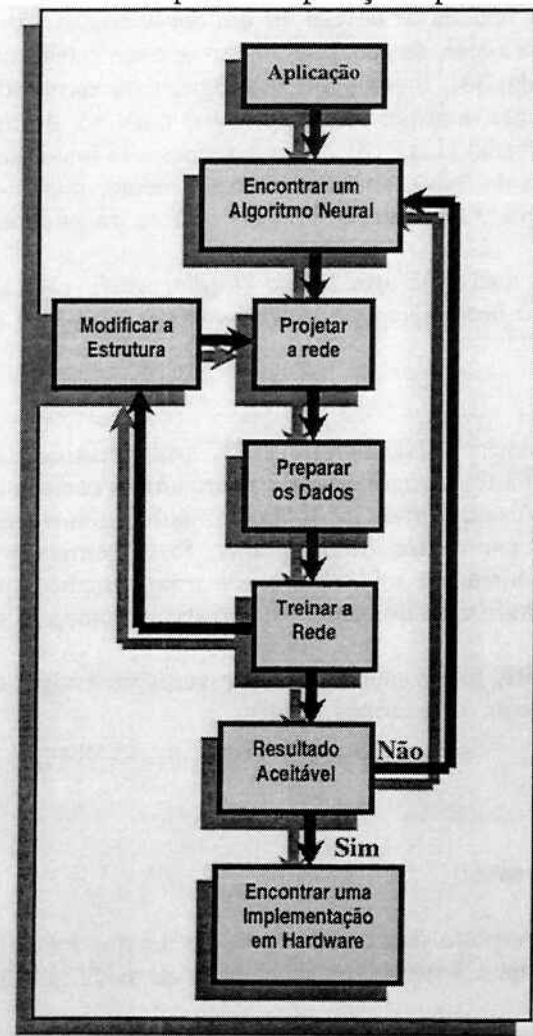


Figura 2. Diagrama em Blocos mostrando o procedimento geral para a implementação de redes neurais artificiais.

Na Figura 2 podemos observar que existem vários processos de adaptação e retroalimentação. A mais óbvia é o ajuste dos pesos (incluído no processo de treinamento). Outro é o ajuste fino da estrutura em termos do número de camadas e elementos de processamento em cada uma. No procedimento, a implementação em hardware é o último passo. Cada bloco é definido a continuação:

Aplicação

O uso das redes neurais tem se estendido a muitas áreas da ciência e indústria, portanto as redes neurais podem ser aplicadas nas seguintes áreas [7]:

Deteção de padrões robustos (espacial, temporal e espaço - temporal), Filtragem de sinais, Segmentação de dados, Compressão de dados, Controle adaptativo, Otimização.

Encontrar Um Algoritmo Neural

O algoritmo neural a ser utilizado para treinar a rede será de "retroalimentação" (backpropagation), com uma arquitetura "direta" (feedforward) que é regular, e apropriada para a implementação em VLSI.

Projetar A Rede

A rede terá como topologia base 3 neurônios de entrada, 2 neurônios na camada escondida e 3 neurônios na camada de saída. Se escolheu esta topologia com o critério de ter uma rede mínima que possa identificar alguns compostos. A determinação desta topologia vai depender das simulações que se façam com o programa "NeuralWorks". Vão ser testadas diferentes topologias, para determinar a rede que melhor resultado apresente.

Preparar Os Dados

Os dados obtidos nos testes de laboratório são os valores iniciais e finais de tensão que os sensores têm nas suas saídas, ou seja são dois valores de tensão para

cada sensor (4 sensores). O valor inicial é obtido antes de introduzir a amostra e o valor final se registra logo depois de um tempo predeterminado. Os tipos de aguardentes e álcoois que serviram como amostras foram : Aguardente 3 Fazendas, Aguardente Chave de Ouro, Aguardente Pirassununga, Aguardente Pitu, Álcool de Arroz e Pisco. As amostras tiveram diferentes volumes de ar saturado (5, 10, 15, 20, 25 ml).

Antes de fazer a simulação os dados coletados passam por uma avaliação estatística para selecionar os aguardentes e sensores que serão usados.

Treinar A Rede

A rede será treinada com o programa NeuralWorks, se utilizará o algoritmo de "retropropagação" com todos os valores programados na configuração do programa. Estas

configurações serão mudadas convenientemente para treinar a rede de tal forma que nos proporcione o melhor resultado possível.

Modificar Estrutura

A estrutura (topologia) será modificada até que se obtenha o melhor resultado. Portanto se mudarão o número de camadas da rede e o número de elementos de processamento (neurônios).

Resultado Aceitável

Se depois de modificar a estrutura, se obtém uma rede que tenha o melhor resultado de acordo aos requerimentos do sistema, então se procederá à implementação do Hardware da Rede Neural.

Implementação Em Hardware Da Rede Neural

Para a implementação em hardware de redes neurais se devem ter em consideração vários critérios. Por exemplo, o tipo de sinal com o qual opera a rede, de acordo com isso se pode categorizar a rede como uma implementação com sinais pulsadas [8], digitais [9], analógicas ou misturadas (mixed-mode)[10]. Dependendo do fluxo da informação, a arquitetura (topologia) pode ser do tipo direta (feedforward) [11] ou de retroalimentação (feedback) [12], [13]. Outra consideração importante é o processo de treinamento, desde o ponto de vista da implementação do treinamento, podem-se considerar dois tipos: implementação dentro do integrado (on-chip) [14], [15] ou fora do integrado (off-chip) [16], [17].

A rede neural a ser implementada utilizará uma topologia direta (feedforward), com um algoritmo de retroalimentação (backpropagation) e a implementação do treinamento será fora do integrado.

5. Treinamento da Rede Neural

O treinamento é efetuado com o programa comercial **NEURALWORK**, para identificar os pesos sinápticos da rede neural. A arquitetura da rede básica inicial possuía 3 neurônios na camada de entrada, 2 neurônios na camada escondida e 3 neurônios na camada de saída. Fazendo o treinamento para uma série de arquiteturas e mudando vários parâmetros do programa, foi identificada a arquitetura, opção que proporcionava uma resposta ótima, ou seja, valores de peso sináptico que apresentam uma maior porcentagem de acerto na identificação dos compostos (esta porcentagem de acerto é tomada do **NEURALWORK**).

Os parâmetros utilizados pelo **NEURALWORK** foram mudados várias vezes no estágio de treinamento da rede, ao final esses parâmetros ficaram com os seguintes valores:

Relação de aprendizagem = 0,9

Momentum = 0,6

Algoritmo = Retropropagação (backpropagation)

Arquitetura = Direta (Feedforward)

Função de Transferência = Tanh (camada escondida e saída)

Iterações = 50000

A arquitetura encontrada que forneceu uma resposta ótima (94.44 % de acerto) foi de 3 neurônios de entrada, 4 neurônios na camada escondida e 3 neurônios na camada de saída (3-4-3), como mostrada na figura 3.

Os coeficientes dos pesos sinápticos encontrados com o auxílio do NeuralWork são arranjados em duas matrizes, que representam os pesos sinápticos encontrados na primeira e segunda camada. A matriz de pesos sinápticos da primeira camada (desde a camada de entrada até a camada escondida) é:

$$\begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ s_{b1} & s_{b2} & s_{b3} & s_{b4} \end{bmatrix} = \begin{bmatrix} -5,6971 & -1,9012 & -3,6582 & -1,7243 \\ -9 & -6,7045 & -7,8557 & 2,6619 \\ -8,2175 & 3,4997 & 8,3499 & 3,3856 \\ 5,5348 & 3,0462 & 2,2486 & -1,9373 \end{bmatrix}$$

A matriz de pesos sinápticos da segunda camada (desde a camada escondida até a camada de saída) é:

$$\begin{bmatrix} s_{45} & s_{46} & s_{47} \\ s_{55} & s_{56} & s_{57} \\ s_{65} & s_{66} & s_{67} \\ s_{75} & s_{76} & s_{77} \\ s_{b5} & s_{b6} & s_{b7} \end{bmatrix} = \begin{bmatrix} -1,2459 & -0,1661 & 1,9362 \\ -0,8139 & -0,9255 & -0,4867 \\ 1,5408 & -0,9818 & 0,4058 \\ -0,5203 & -1,1839 & 0,5032 \\ -0,2809 & 0,9023 & 2,0541 \end{bmatrix}$$

Onde, $s_{11}, s_{12}, \dots, s_{34}, s_{45}, s_{46}, \dots, s_{77}$, são os coeficientes dos pesos sinápticos entre os neurônios das diferentes camadas e $s_{b1}, s_{b2}, \dots, s_{b7}$, são os coeficientes dos pesos sinápticos entre o neurônio de polarização e os neurônios da camada escondida e de saída.

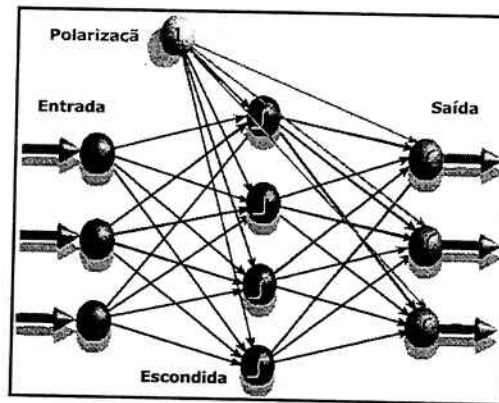


Figura 3. Arquitetura da rede Neural a ser treinada

6. Partes do Sistema

Baseados nos resultados anteriores se procedeu a fazer o projeto do sistema tendo em consideração os seguintes aspectos:

- A matriz de sensores é composta por sensores da marca TAGUCHI, de óxido de estanho. Este tipo de sensores são utilizados no sistema identificador de gases, devido ao seu baixo custo, possuem uma larga faixa de sensibilidade e são pouco seletivos, no sistema para o qual serão utilizados a seletividade é tarefa de uma etapa posterior (neste caso as redes neurais).
- Uma Unidade de Entrada cuja função é acondicionar o sinal dos sensores para o posterior processamento dos dados pela rede neural. O circuito acondicionador do sinal terá que proporcionar os valores normalizados para a rede neural, esses valores foram obtidos dividindo o valor final pelo valor inicial.
- A Rede Neural implementada em Hardware (Matriz Sináptica em CI e Neurônios com circuitos discretos).
- Uma unidade de Controle para o refrescamento da matriz de pesos sinápticos, composta por um microcontrolador com memória *flash* incorporada, um conversor D/A de 12 bits, um amplificador operacional e dois decodificadores de 3 para 8 linhas. A rede neural é um sistema formado por sinapses e neurônios, onde a matriz sináptica é formada por capacitores e multiplicadores analógicos. Os capacitores da matriz sináptica necessitam ser carregados por uma tensão pré-determinada (peso sináptico) portanto, foi projetado um sistema que deve ter um funcionamento muito semelhante a um sistema de refrescamento de memórias dinâmicas, que controle os decodificadores e o conversor D/A.

A continuação apresentamos os diagramas esquemáticos das partes do sistema, a rede neural será apresentada depois.

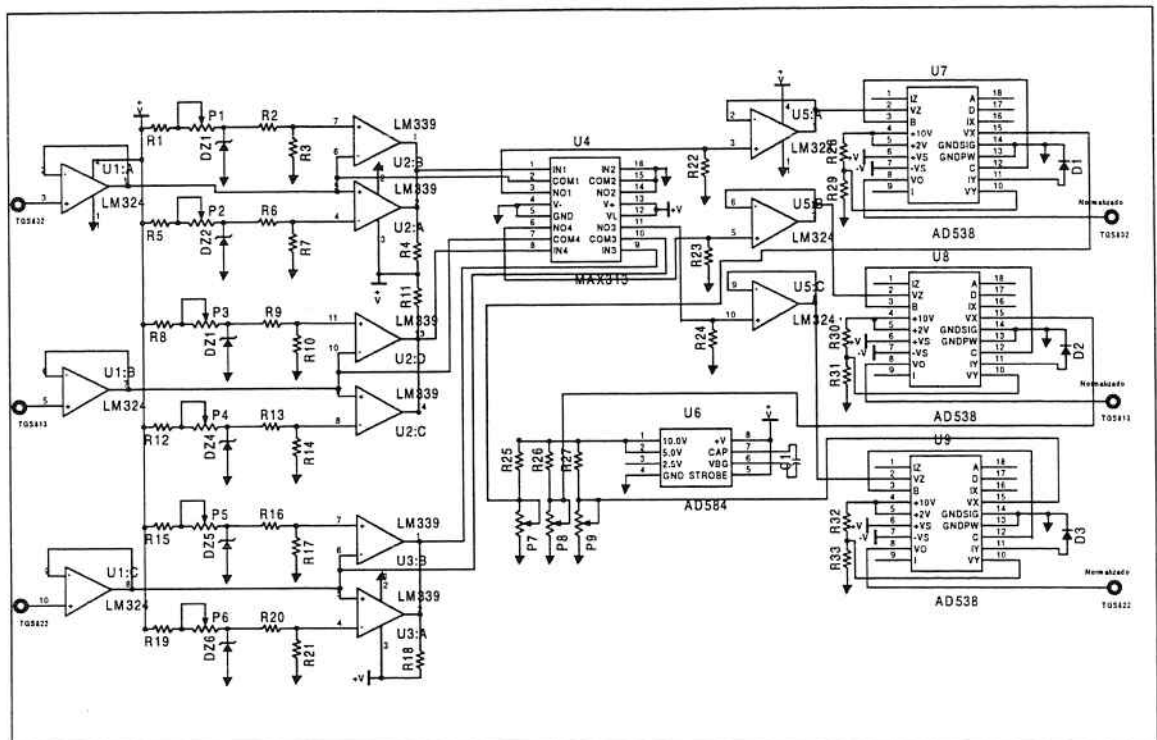


Figura 4. Esquemático do circuito de entrada ao sistema.

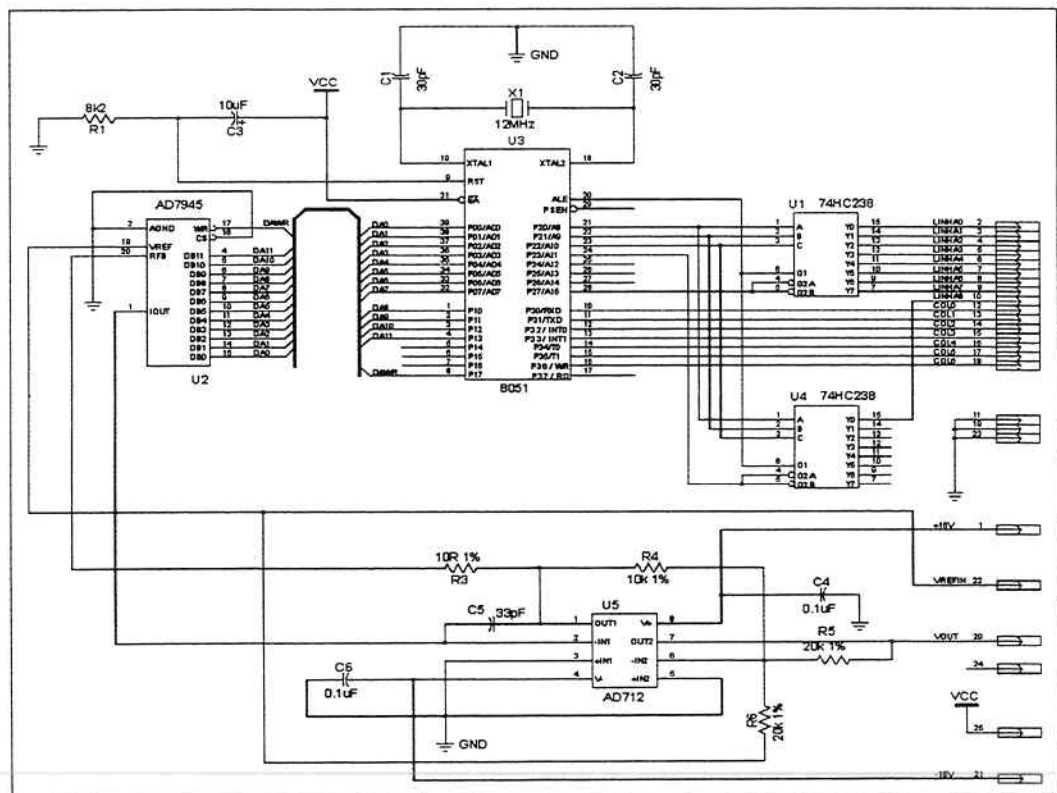


Figura 5. Diagrama elétrico do circuito de controle para o refresco das sinapses.

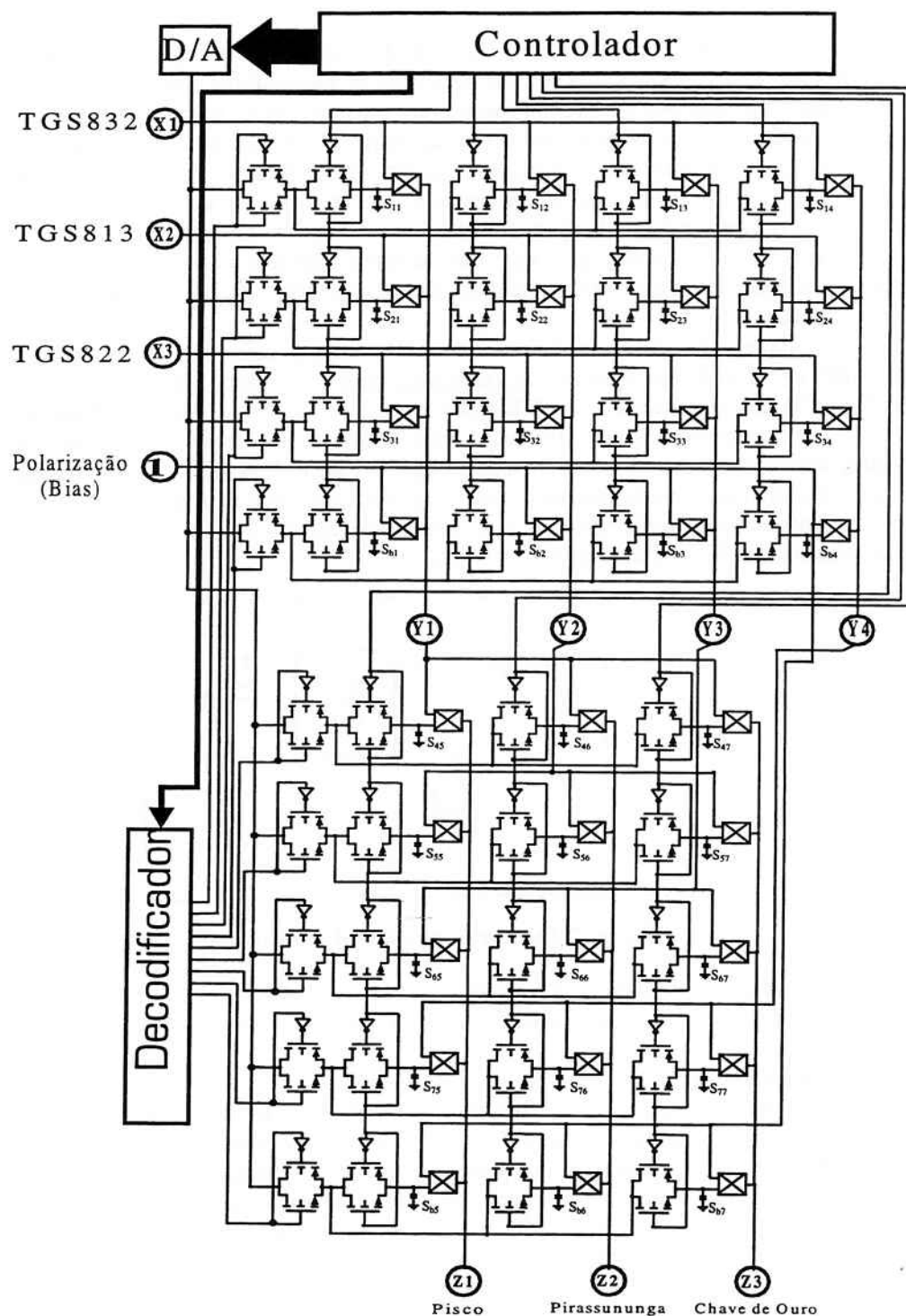


Figura 6 Diagrama do sistema formado pela rede neural e blocos do circuito de controle

A figura 6 nos mostra a matriz sináptica e os circuitos que fazem o refresh dos capacitores (figura 5) que armazenam os pesos sinápticos. Os blocos, capacitores e multiplicadores, que formam essa matriz serão projetados em VLSI, mas os neurônios serão projetados com circuitos discretos nos itens próximos.

7. Projeto dos blocos da matriz sináptica

A rede neural a ser projetada é formada de blocos especiais de circuitos que simulam sinapses e neurônios. Estes blocos a sua vez são formados de outros blocos menores, por exemplo, a sinapse eletrônica é formada por um multiplicador e um capacitor, onde o multiplicador realiza a operação de multiplicação da tensão proporcionada por algum neurônio e a tensão que representa o peso sináptico armazenada no capacitor. Os neurônios são de tipos diferentes, temos os neurônios de entrada que são simples “buffers”, os neurônios da camada escondida mostram na sua saída uma função de transferência igual a tanh (tangente hiperbólica) e também tem um buffer para fornecer os sinais necessários para as sinapses que saiam dele, e depois temos o neurônio de saída que tem as mesmas funções do neurônio da camada escondida além de nos proporcionar a identificação do sinal que está na sua entrada.

A matriz sináptica e as chaves de passo foram projetadas com a tecnologia CMOS de 0,6 μm da AMS (Austria Mikro Systeme International), para a simulação elétrica dos circuitos se utilizou o software Mentor Graphics e o PSPICE e para a edição do leiaute se utilizou o Mentor Graphics.

Os neurônios foram projetados com circuitos discretos, amplificadores operacionais e multiplicadores. Os neurônios da camada escondida e de saída têm uma função de transferência igual a “tanh” do sinal de entrada, portanto, esta função foi aproximada a uma equação linear e polinomial respectivamente para poder projetar os circuitos que cumprissem essas condições. Na figura 7 se vê um diagrama em blocos da sinapse eletrônica.

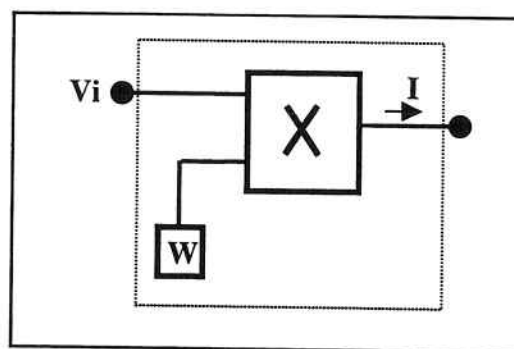


Figura 7. Diagrama em blocos de uma sinapse artificial.

Onde V_i é o sinal de entrada (vinda de algum neurônio), o bloco X é o multiplicador, o bloco W é o valor da tensão do peso sináptico armazenado no capacitor e I é a corrente de saída.

7.1 O Multiplicador

O circuito multiplicador elaborado neste trabalho obedece a lei quadrática de um transistor MOS em saturação. Em relação aos outros multiplicadores o resultado proposto tem várias vantagens, a área de silício e a potência são pequenas, as entradas e saídas a este são simples (single-ended) evitando assim que os sinais que entram sejam antes preprocessados economizando área de silício. É por isso que este tipo de multiplicador é bastante prático para ser usado na implementação de redes neurais artificiais. O princípio básico de operação deste multiplicador é baseado na seguinte identidade:

$$(V_1 + V_2)^2 - V_1^2 - V_2^2 = 2V_1V_2 \quad (1)$$

Na equação (5.1) o quadrado é alcançado usando a relação quadrática entre a tensão de porta-fonte e a corrente de dreno do transistor PMOS em saturação. O multiplicador a ser projetado consiste de 2 espelhos de corrente (NMOS) e cinco transistores PMOS, M1 - M5, os quais tem a mesma relação de valores geométricos. O diagrama do multiplicador é mostrado na figura 8. Os 5 transistores estão na região de saturação, e suas fontes e substratos são ligados juntos para diminuir o efeito de corpo.

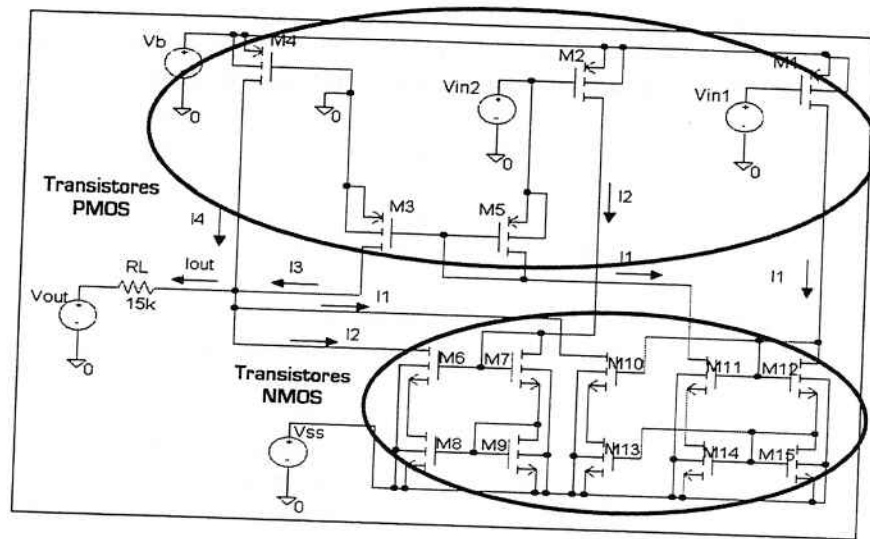


Figura 8. Diagrama esquemático do multiplicador a ser implementado.

A corrente de saída deste multiplicador é dada pela seguinte equação:

$$I_{OUT} = 2K V_{in1} V_{in2} \quad (2)$$

Das medidas experimentais realizadas com este multiplicador se obteve o gráfico mostrado na Figura 9.

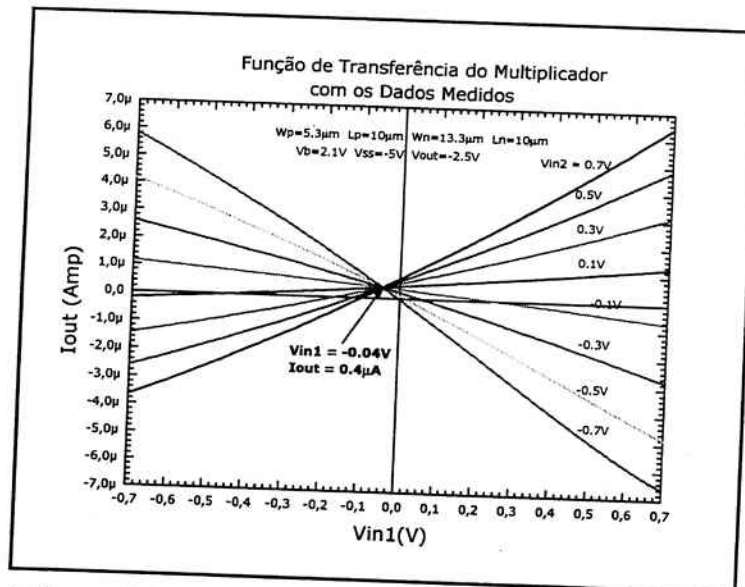


Figura 9. Curvas de Saída do Multiplicador PMOS com os Dados Experimentais.

Na figura 9. podemos ver que a curva esta deslocada, o ponto onde todas as curvas se cruzam está localizado em $V_{in1} = -0.04V$ e $I_{out} = 0.4 \mu A$. As não linearidades e imperfeições na saída do multiplicador se devem principalmente à limitação que impõe a tecnologia com respeito dos V_T 's, degradação da mobilidade, o não casamento dos componentes analógicos o qual é influenciado pelo leiaute, onde se tem variações no processo de fabricação e pela modulação de comprimento de canal (λ) que não foi considerado nos cálculos.

7.2 O Capacitor

Para iniciar o projeto podemos considerar que, o valor do capacitor não é muito importante se uma fonte de tensão é usada para carregar este a um valor desejado [10]. Se uma fonte de corrente é usada, seu valor diretamente vai decidir a carga armazenada e assim a tensão através do capacitor. O valor da tensão no capacitor é influenciado pela fuga de carga e a injeção de carga.

A fuga de carga elétrica é a carga no capacitor que pode vaziar através dos caminhos de fuga parasitas e pode causar um erro na armazenagem dos pesos. Os dois caminhos de fuga são a junção de polarização inversa e a resistência OFF da chave de acesso (CMOS). Um valor típico de fuga (1 pA) [10] e uma capacitância de 1 pF levam a uma queda de tensão de $1 \mu\text{V}/\mu\text{S}$. Portanto, a tensão do capacitor deve ser restaurado uma vez cada 10 ms se um erro de não mais de 10 mV é permitido. O cálculo do valor do capacitor será encontrado em base ao seguinte diagrama mostrado na figura 10.

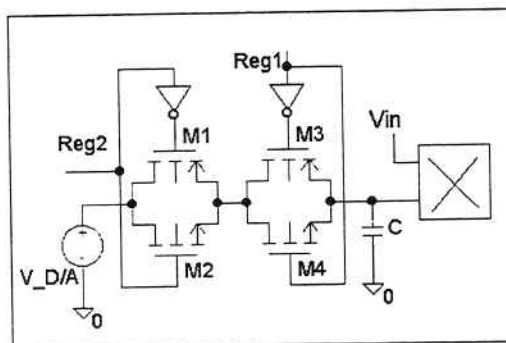


Figura 10. Diagrama para o cálculo do capacitor.

Na figura 10 se mostram os pinos por onde são fornecidos os sinais que vem do microcontrolador e decodificador, estes são representados por Reg1 e Reg2 respectivamente, estes sinais terão tempos de duração suficientes para carregar o capacitor e níveis adequados para que as chaves (transmission gate) conduzam ou deixem de conduzir o sinal de $V_{D/A}$ (tensão analógica fornecida pelo conversor D/A). O capacitor C, armazena o peso sináptico determinado pelo software (NeuralWorks), este está ligado à porta de um transistor PMOS do bloco multiplicador representado pelo símbolo X. Segundo a tecnologia AMS de $0.6 \mu\text{m}$ se tem que o capacitor feito com poly1 e poly2 tem as seguintes dimensões:

Tabela 1. Dimensões para encontrar o valor do capacitor, segundo a AMS.

Parâmetros	Símbolo	Min	Typ	Max	Units
Poly2 to Poly1	CPOX	0.78	0.86	0.96	$\text{ff}/\mu\text{m}^2$

A área do capacitor será calculada com os dados da tabela 1 e se tomará o que melhor se adequar ao projeto. Tendo-se em conta que o capacitor que armazena os pesos sinápticos é 1 pF , temos:

Para CPOX_{MIN} : $A_1 = (1 \times 10^{-12}) / (0.78 \times 10^{-15}) = 1282.02 \mu\text{m}^2$ mais ou menos $36 \mu\text{m}/\text{lado}$.

Para CPOX_{MAX} : $A_2 = (1 \times 10^{-12}) / (0.96 \times 10^{-15}) = 1041.66 \mu\text{m}^2$... mais ou menos $32.27 \mu\text{m}/\text{lado}$.

Portanto, vai ser selecionado o capacitor com **$36 \mu\text{m}/\text{lado}$** . As chaves analógicas (transmission gates) terão dimensões mínimas ($W_p = W_n = 0.8 \mu\text{m}$ e $L_p = L_n = 0.6 \mu\text{m}$) para manter assim a injeção de carga em um mínimo e reduzir o "feedthrough".

O diagrama esquemático, o leiaute do circuito da sinapse e as chaves de habilitação feito com o Mentor Graphics e o leiaute do CI real são mostrados na figura 11.

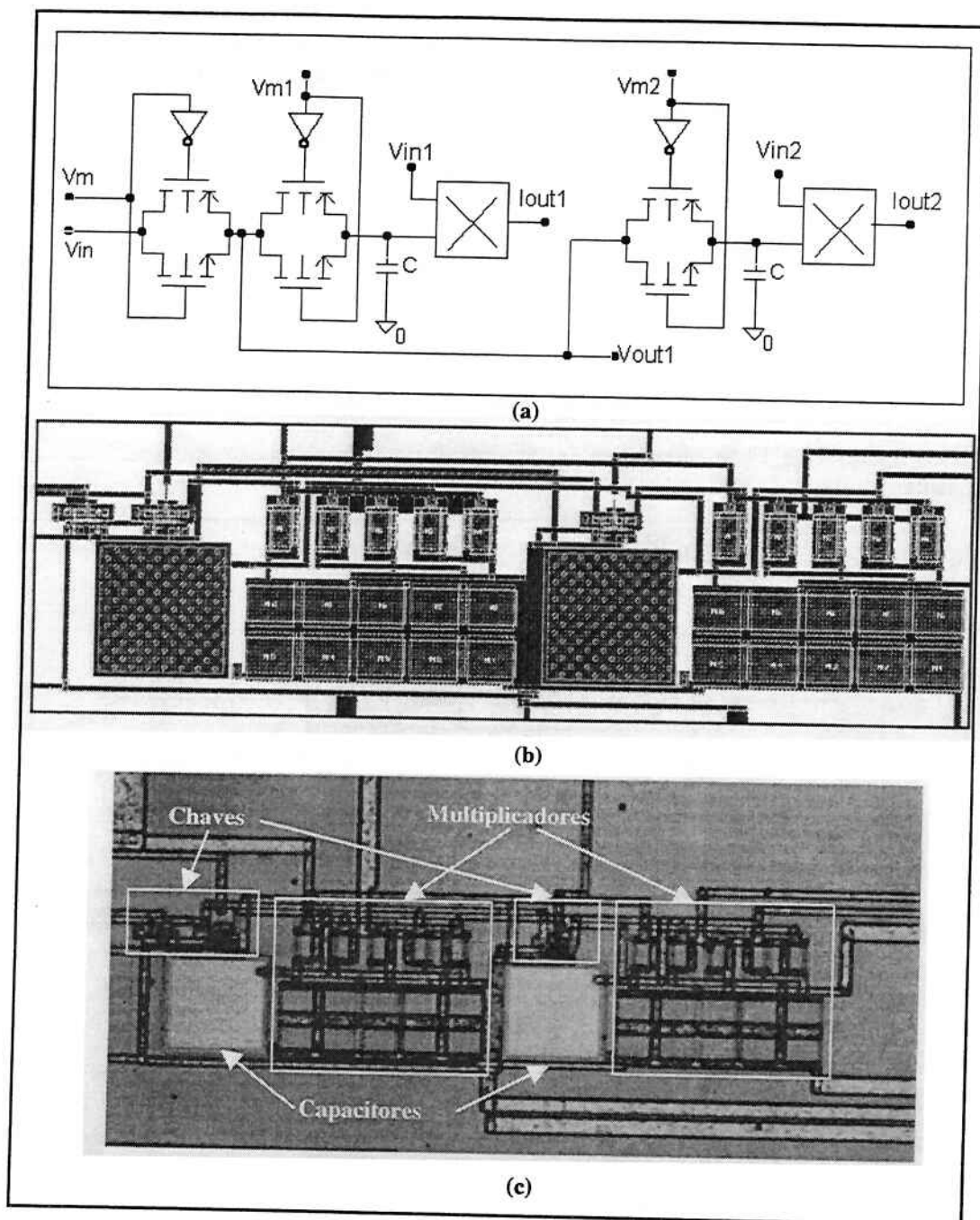


Figura 11. Diagrama (a) elétrico, (b) leiaute com Mentor, (c) leiaute real, da primeira e segunda sinapse

8. Projeto dos Neurônios

O neurônio tem duas partes; uma executa a somatória das correntes e a outra fornece a função de transferência ($\tanh$). As saídas das sinapses são correntes, portanto a somatória é executada ligando todas as saídas das sinapses em um só ponto. A corrente total que deixa o nó comum será a soma de todas as saídas sinápticas. A função de transferência se consegue implementando um dispositivo não linear que toma a somatória das correntes e na sua saída entrega uma tensão. Esta tensão de saída é a tensão do neurônio. Desta forma a saída do neurônio pode alimentar as entradas das sinapses na seguinte camada.

Como se apresentou na rede neural se tem três camadas de neurônios (entrada, escondida e saída). Os neurônios da camada de entrada são simples "buffers", devido a que o único que se precisa

é que os sinais sejam os mesmos quando eles passarem através desta camada. Os neurônios da camada escondida e de saída precisam ter o conversor de corrente a tensão e o elemento não linear que proporcione a função de transferência. Tomando em consideração estas afirmações só serão projetados os neurônios da camada escondida e os neurônios da camada de saída. Na figura 12 podemos ver o diagrama em blocos de um neurônio (camada escondida ou camada de saída).

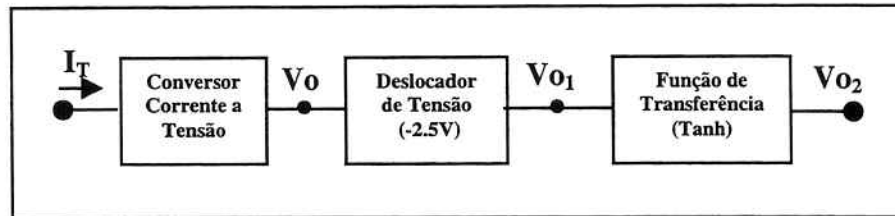


Figura 12. Diagrama em blocos do neurônio

Nas figuras 13 e 14 se apresentam os circuitos dos neurônios da camada escondida e da camada de saída.

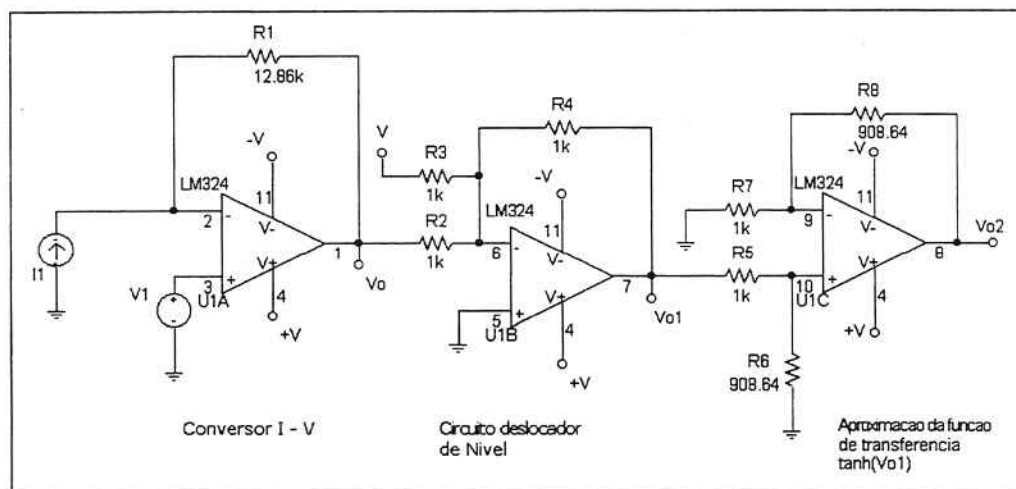


Figura 13. Esquema elétrico dos neurônios da camada escondida.

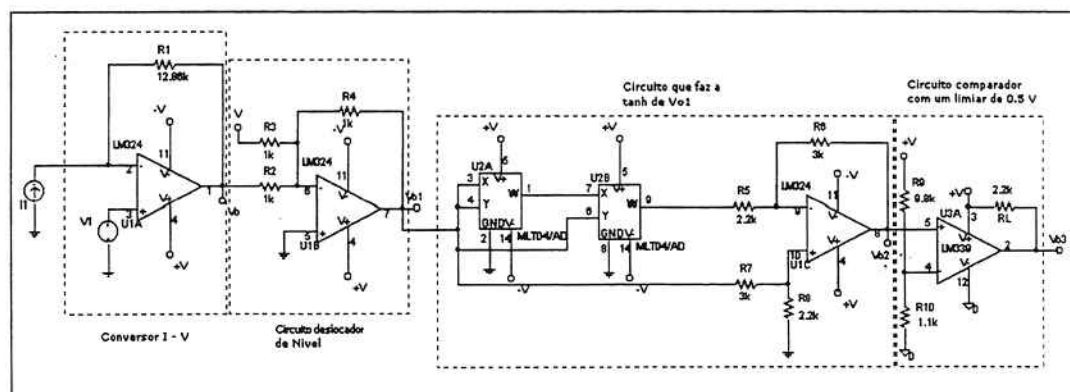


Figura 14. Circuito completo do neurônio da camada de saída.

9. Conclusões

Nesta pesquisa usou-se as técnicas adaptativas para o campo da instrumentação, estas sem dúvida são desafiadoras, devido a que capturam o comportamento de complexos sistemas dinâmicos não lineares e a solução desejada é aprendida mediante exemplos sem ter que ser definida por algum algoritmo. As técnicas mais representativas são as redes neurais e a lógica nebulosa (Fuzzy Logic). A complexidade computacional das técnicas adaptativas são limitadas razão pela qual é importante contribuirmos na realização de soluções com circuitos integrados analógicos e digitais.

Se projetou um circuito acondicionador do sinais de saída do sensor para fornecer à rede neural os valores apropriados para que ele possa identificar o padrão entrante. Este circuito conta com amplificadores operacionais, comparadores, chaves analógicas e CI's divisores de tensão fáceis de encontrar no mercado. O tempo de propagação do sinal desde a entrada à saída será de aproximadamente 5 μ s.

A matriz sináptica tem como um dos blocos principais o capacitor, devido a que este tem fugas de corrente será refresco cada certo tempo para não perder a carga (tensão) depositada nele, portanto se projetou um circuito que controlasse o rfresco dos capacitores da matriz sináptica, este está composto por um microcontrolador ATMEL 8051, um conversor D/A de 12 bits e um amplificador operacional duplo para obter os valores bipolares (positivo ou negativo) que serão carregados no capacitor.

Outro bloco pertencente à matriz sináptica é o multiplicador que em comparação com outros mencionados na bibliografia tem duas entradas simples e uma saída só, o consumo de potência deste é aproximadamente de 0.325 mW e ocupa uma área de silício reduzida, $63 \mu\text{m} \times 83.4 \mu\text{m} = 5254.2 \mu\text{m}^2$. A matriz sináptica projetada consumirá uma potência de 10 mW, e a sua área será de 0.203 mm^2 aproximadamente.

Os neurônios foram implementados com CI's comerciais que nos deram a flexibilidade de poder realizar circuitos com função de transferência de saída igual a tanh, o que era um pouco complicado devido às diferentes dificuldades apresentadas para conseguir a função tanh.

Com a pesquisa das redes neurais aplicadas a este sistema projetado se pode conseguir uma série de instrumentos com uma variedade de aplicações os quais dependerão de como se treinar a rede neural, assim também poderiam ser implementados instrumentos inteligentes portáteis e de baixo custo. A pesquisa também nos demonstra que a utilização de técnicas adaptativas em muitos casos solucionam problemas que não tem uma solução algorítmica ou o algoritmo é muito difícil de ser implementado.

Esta pesquisa consolida uma primeira versão de instrumentação portátil capaz de efetuar o reconhecimento de uma substância usando exclusivamente o aroma que dela emana. Se trabalhou nos aspectos técnicos envolvidos desde a geração dos sinais nos sensores até a configuração do sistema de reconhecimento de padrões com as redes neurais artificiais implementadas em um circuito integrado de aplicação específica. Portanto, se conciliou a implementação parcial de neurônios e sinapses em VLSI harmonizados com as interfaces de conexão aos sensores em eletrônica discreta.

Referências Bibliográficas

- [1] Donlin M, Child J., Is Neural computing the key to artificial intelligence?., *Computer Design.*, October 1992, pp. 87 – 104.
- [2] Kramer A, H., Array-Based Analog Computation., *IEEE Micro.*, October 1996., pp. 20 – 29.
- [3] Goser K, Hilleringmann U, Schumacher K., VLSI Technologies for Artificial Neural Networks., *IEEE Micro.*, Dec 1989, pp. 28 – 44.
- [4] Alianna J. Maren., Craig T. Harston., Robert M. Pap., *Handbook of Neural Computing Applications.*, Editora Academic Press Inc., 1^a. edição., 1990.
- [5] Sied M. Fakhraie, Kenneth C. Smith, *VLSI – Compatible Implementations for Artificial Neural Networks*, Kluwer Academic Publishers, 1^a edição, 1997.
- [6] NeuralWorks, manuals.
- [7] Alianna J. Maren., Craig T. Harston., Robert M. Pap., *Handbook of Neural Computing Applications.*, Editora Academic Press Inc., 1^a. edição., 1990.
- [8] Graf, H.P., Jackel, L.D., "Analog Electronic Neural Network Circuits," *IEEE Circuits and Devices Magazine*, vol. 5, no. 4, July 1989, pp. 44-49.
- [9] Atlas L, E., Suzuki Y.; "Digital systems for artificial neural networks," *IEEE Circuits and Devices Magazine*, pp. 20-24, May 1989.
- [10] Satyanarayana S.; Tsividis Y, P.; Graf H, P.; "A reconfigurable VLSI neural networks," *IEEE J. Solid-State Circuits*, vol. 27, pp. 67-81, Jan 1992.
- [11] J.B. Lont, and W. Guggenbuhl, "Analog CMOS implementation of a multilayer perceptron with nonlinear synapses," *IEEE Trans. Neural Networks*, vol. 3, no. 6, pp. 1444-1455, June 1992.
- [12] J. Alspector, R. Allen, A. Jayakumar, "Relaxation networks for large supervised learning problems," *Advances in Neural Information Processing Systems 3, San Mateo, CA: Morgan Kaufmann, 1991.*
- [13] Y. Arima, K. Mashito, K Okada, T Yamada, et all , "A 336- neuron, 28K-synapse, self-learning neural network chip with branch-neuron-unit architecture," *IEEE J. Solid State Circuits*, vol. 26, no. 11, pp. 1637-1644, Nov. 1991.
- [14] H. C. Card, C. R. Schneider, and W. R. Moore, "Hebbian plasticity in MOS synapses," *IEE Proc. Part F*, vol. 138, no. 1, pp. 13 -16, Feb. 1991.
- [15] T. Morie, and Y Amemiya, "An all-analog expandable neural network LSI with on-chip backpropagation learning," *IEEE J. Solid State Circuits*, vol. 29, no. 9, Sept. 1994.
- [16] M. Holler, S. Tam, H. Castro, R Benson, "An electrically trainable artificial neural network (ETANN) with 10240 floating - gates synapses," *Proc. 1989 Int. Joint Conf. Neural Networks*, vol. 2, pp. 191-196, Washington DC, June 1989.
- [17] H. P. Graf, and D. Henderson, "A reconfigurable CMOS neural network, " *ISSCC Dig. Tech. Papers*, pp 144-5, Feb. 1990.

BOLETINS TÉCNICOS - TEXTOS PUBLICADOS

- BT/PSI/0001 – Observabilidade Topológica de Osawa em Redes não Lineares – ARMANDO HANDAYA, FLÁVIO A. M. CIPPARRONE
- BT/PSI/0002 – Desenvolvimento de uma Microbalança de Quartzo para Detectar Gases – ROBERTO CHURA CHAMBI, FRANCISCO JAVIER RAMIREZ FERNANDEZ
- BT/PSI/0003 – Sistema para Desenvolvimento de Sensores Inteligentes – ANTONIO CARLOS GASPARETTI, FRANCISCO JAVIER RAMIREZ FERNANDEZ
- BT/PSI/0004 – A 1.6GHz Dual Modulus Prescaler Using the Extended True Single-Phase Clock CMOS Circuit Technique (E-TSPC) – JOÃO NAVARRO SOARES JÚNIOR, WILHELMUS ADRIANUS M. VAN NOIJE
- BT/PSI/0005 – Modelamento em Linguagem VHDL de uma Unidade de Policiamento para Redes Locais ATM – ÉDSON TAKESHI NAKAMURA, MARIUS STRUM
- BT/PSI/0006 – Otimização das Operações Coletivas para um Aglomerado de 8 Computadores usando uma Rede Ethernet 10 Mbps baseada em Hub – MARTHA TORRES, SERGIO TAKEO KOFUJI
- BT/PSI/0007 – Short Temporal Coherence Optical Source With External Fiber Optics Cavity – CARMEM LÚCIA BARBOSA, JOSÉ KEBLER DA CUNHA PINTO
- BT/PSI/0008 – Hydrogenated Carbon Films Used as Mask in Wafer Processing With Integrated Circuits: Post-Processing – JUAN M. JARAMILLO O., RONALDO D. MANSANO, EDGAR CHARRY R.

