

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**INFLUÊNCIA DAS INTERCONEXÕES SOBRE O
DESEMPENHO DE CIRCUITOS INTEGRADOS
NANOELETRÔNICOS BASEADOS EM TRANSISTORES
MONO-ELÉTRON**

VÍTOR GOUVÊA ANDREZO CARNEIRO

ORIENTADOR: JOSÉ CAMARGO DA COSTA

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA

PUBLICAÇÃO: PPGENE.DM – 320/07

BRASÍLIA/DF: DEZEMBRO – 2007

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**INFLUÊNCIA DAS INTERCONEXÕES SOBRE O DESEMPENHO DE
CIRCUITOS INTEGRADOS NANOELETRÔNICOS BASEADOS EM
TRANSISTORES MONO-ELÉTRON**

VÍTOR GOUVÊA ANDREZO CARNEIRO

**DISSERTAÇÃO SUBMETIDA AO DEPARTAMENTO DE
ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA
UNIVERSIDADE DE BRASÍLIA COMO PARTE DOS REQUISITOS
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM
ENGENHARIA ELÉTRICA.**

APROVADA POR:

**Prof. José Camargo da Costa, Docteur (ENE-UnB)
(Orientador)**

**Prof. Alexandre Ricardo Soares Romariz, PhD (ENE-UnB)
(Examinador Interno)**

**Prof. João Antonio Martino, Doutor (LSI-USP)
(Examinador Externo)**

BRASÍLIA/DF, 10 DE DEZEMBRO DE 2007

FICHA CATALOGRÁFICA

CARNEIRO, VÍTOR GOUVÊA ANDREZO

Influência das Interconexões sobre o Desempenho de Circuitos Nanoeletrônicos Baseados em Transistores Mono-Elétron [Distrito Federal] 2007.

xv, 99p., 210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2007).

Dissertação de Mestrado – Universidade de Brasília. Faculdade de Tecnologia. Departamento de Engenharia Elétrica.

1. Nanoeletrônica

2. Interconexão

3. Transistor mono-elétron

4. Desempenho de Circuitos Integrados

I. ENE/FT/UnB

II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

CARNEIRO, V. G. A. (2007). Influência das Interconexões sobre o Desempenho de Circuitos Integrados Baseados em Transistores Mono-Elétron. Dissertação de Mestrado em Engenharia Elétrica, Publicação PPGENE.DM-320/07, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 99p.

CESSÃO DE DIREITOS

AUTOR: Vítor Gouvêa Andrezo Carneiro.

TÍTULO: Influência das Interconexões sobre o Desempenho de Circuitos Integrados Baseados em Transistores Mono-Elétron.

GRAU: Mestre

ANO: 2007

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte dessa dissertação de mestrado pode ser reproduzida sem autorização por escrito do autor.

Vítor Gouvêa Andrezo Carneiro
SMU, ROS Bloco B Apto 207
70.630-000 Brasília – DF – Brasil.

AGRADECIMENTOS

Gostaria de agradecer ao meu orientador, José Camargo, por quem tenho profunda admiração e respeito. Ele não se limitou a me orientar somente na parte acadêmica, mas sempre aparecia com sábios conselhos nesta minha nova jornada na cidade de Brasília.

À minha co-orientadora Janaína, que foi mais do que uma amiga. Nossas conversas de laboratório me esclareceram mais coisas do que você imagina.

Aos professores que aceitaram participar da minha banca: Alexandre Romariz e João Martino. Suas observações e comentários contribuíram bastante para a finalização da minha dissertação.

Aos professores do Instituto Militar de Engenharia: Maj. Alexandre Mendonça e TC Marcelo Perez. Agradeço pelos ensinamentos que me incentivaram a aprofundar na área da eletrônica e pela recomendação ao mestrado na Universidade de Brasília.

Aos chefes do 7º Centro de Telemática de Área, do Exército Brasileiro: TC Assad, TC Real e TC Wolski. Agradeço pelo incentivo e apoio na realização deste trabalho.

Aos meus colegas dos laboratórios LTSD e LPCI, com quem compartilhei novidades e experiências. Sem nossas conversas, mesmo que distraídas, teria sido difícil manter o foco.

Ao PADCT – CNPq (Instituto do Milênio – NAMITEC) pelo apoio institucional e financeiro ao laboratório.

Aos meus pais, Filipe e Maria Inês; e irmãos, Fábio, Lúcia e Nádia. Seu apoio e carinho, mesmo que a distância, me fizeram manter no caminho.

Agradeço a Deus por ter realizado mais este sonho...

RESUMO

INFLUÊNCIA DAS INTERCONEXÕES SOBRE O DESEMPENHO DE CIRCUITOS INTEGRADOS NANOELETRÔNICOS BASEADOS EM TRANSISTORES MONO-ELÉTRON

Autor: Vítor Gouvêa Andrezo Carneiro.

Orientador: José Camargo da Costa.

Programa de Pós-graduação em Engenharia Elétrica.

Brasília, dezembro de 2007.

As tecnologias básicas adotadas atualmente pela indústria semicondutora para a fabricação de memórias e processadores podem alcançar certos limites que fazem com que novas tecnologias tenham que ser estudadas e desenvolvidas. Os transistores mono-elétron, como outros dispositivos em escala nanométrica, parecem ser uma opção próspera para implementações GSI ou TSI no futuro.

O desenvolvimento de arquiteturas de processador GSI e TSI, baseados em dispositivos nanoeletrônicos, está sendo feita atualmente. A abordagem adotada compreende a implementação de um núcleo de processamento extremamente paralelo e distribuído, construído com dispositivos nanoeletrônicos, organizados em células. A investigação destas arquiteturas considerou, até o presente, a utilização de interconexões ideais. Os limites das interconexões potencialmente ameaçam desacelerar ou parar o progresso histórico da indústria semicondutora.

Neste trabalho, o desempenho elétrico de associações conhecidas de subcircuitos nanoeletrônicos básicos são estudados com a ajuda de um modelo de interconexão cujos parâmetros podem ser mudados. Os circuitos das associações são simulados com parâmetros do modelo variando de uma interconexão ideal até os piores casos e seus comportamentos dinâmicos são analisados. O objetivo deste estudo é determinar a influência das interconexões sobre o comportamento dos circuitos e estabelecer limites relacionados a interconexões para suas funcionalidade.

As possibilidades de implementação usando novas tecnologias de interconexão, como nanotubos de carbono, são também apresentadas.

ABSTRACT

INTERCONNECTIONS INFLUENCE UPON NANO-ELECTRONIC INTEGRATED CIRCUITS PERFORMANCE BASED ON SINGLE-ELECTRON TRANSISTOR

Author: Vítor Gouvêa Andrezo Carneiro.

Supervisor: José Camargo da Costa.

Programa de Pós-graduação em Engenharia Elétrica.

Brasília, December of 2007.

The basic technologies presently adopted by the semiconductor industry for memory and processor fabrication can attain certain limits which make that new technologies have to be studied and developed. Single-electron transistors, like other nanoscale devices, seem to be a promising option for GSI or TSI implementations in the future.

The development of GSI and TSI processor architectures, based upon nanoelectronic devices, is currently being done. The adopted approach comprises the implementation of a massive parallel and distributed processing core, built with nanoelectronic devices, organized in cells. The investigation on these architectures has considered, up to now, the employment of ideal interconnections. Interconnection limits potentially threaten to decelerate or halt the historical progression of the semiconductor industry.

In this work, the electrical performance of known associations of basic nanoelectronic subcircuits are studied with the help of an interconnection model whose parameters can be changed. The association circuits are simulated with model parameters varying from an ideal interconnection to worst cases and their dynamic behavior are analyzed. The goal of this study is to determine the interconnection's influence upon the circuit behavior and to establish interconnection-related limits for its functionality.

The implementation possibilities using new interconnection technologies, like carbon nanotubes, are also presented.

SUMÁRIO

1 – INTRODUÇÃO.....	16
1.1 – TECNOLOGIA DE FABRICAÇÃO ATUAL.....	17
1.2 – DISPOSITIVOS NANOELETRÔNICOS.....	18
1.3 – INTERCONEXÕES.....	19
1.4 – PROCESSADORES.....	19
1.5 – OBJETIVO DA DISSERTAÇÃO.....	21
2 – FUNDAMENTAÇÃO TEÓRICA.....	22
2.1 – NANOTECNOLOGIA.....	22
2.2 – NANOELETRÔNICA.....	23
2.3 – CARACTERIZAÇÃO E FABRICAÇÃO DE NANOESTRUTURAS.....	23
2.3.1 – Introdução.....	23
2.3.2 – Métodos de nanofabricação <i>top-down</i>	24
2.3.2.1 – Extensão da fotolitografia.....	24
2.3.2.2 – Litografia por feixe de elétrons (EBL).....	24
2.3.2.3 – Litografia por varredura de sonda (SPL).....	25
2.3.3 – Métodos de nanofabricação <i>bottom-up</i>	25
2.3.3.1 – Crescimento de cristais.....	25
2.3.3.2 – Métodos químicos.....	26
2.3.3.3 – Auto-montagem.....	26
2.4 – EXEMPLOS DE DISPOSITIVOS NANOELETRÔNICOS.....	26
2.5 – DISPOSITIVOS BASEADOS EM TUNELAMENTO MONO-ELÉTRON.....	27
2.5.1 – Tunelamento.....	28
2.5.2 – Teoria ortodoxa do tunelamento mono-elétron.....	29
2.5.3 – Efeito de carregamento.....	30
2.5.4 – Bloqueio de Coulomb.....	32
2.5.5 – Transistor mono-elétron.....	33
2.5.5.1 – Equações características.....	34
2.5.5.2 – Corrente no transistor mono-elétron.....	35
2.5.6 – Co-tunelamento.....	36
2.5.7 – Cargas de desvio.....	37

2.6 – PROCESSADOR NANOELETRÔNICO.....	37
2.6.1 – Características desejáveis de um processador nanoeletrônico.....	39
2.7 – COMPONENTES DE UM PROCESSADOR NANOELETRÔNICO.....	40
2.7.1 – Rede neural <i>Winner-Take-All</i> (WTA).....	40
2.7.7.1 – Implementação de uma rede WTA mono-elétron.....	41
2.7.2 – Circuito inversor.....	42
2.7.3 – Circuito somador decimal.....	44
3 – METODOLOGIA EMPREGADA.....	45
3.1 – INTRODUÇÃO.....	45
3.2 – TIPOS DE SIMULAÇÃO.....	45
3.3 – PARÂMETROS CONCENTRADOS DE INTERCONEXÃO.....	46
3.4 – ESTRATÉGIAS PARA OBTENÇÃO DE RESULTADOS.....	47
3.4.1 – Modelagem dos parâmetros de interconexão para as simulações.....	47
3.4.2 – Extrapolação dos parâmetros de interconexão.....	48
4 – DESENVOLVIMENTO E RESULTADOS.....	49
4.1 – INTRODUÇÃO.....	49
4.2 – REDE SET-WTA COM DOIS NEURÔNIOS.....	49
4.2.1 – Procedimentos iniciais.....	50
4.2.2 – Simulações com interconexão não-ideal.....	51
4.2.3 – Diferenças entre os métodos de simulação.....	53
4.3 – REDE SET-WTA COM QUATRO NEURÔNIOS.....	55
4.3.1 – Procedimentos iniciais.....	55
4.3.2 – Simulações com interconexão não-ideal.....	57
4.3.3 – Discussão sobre retardo e velocidade.....	58
4.4 – CIRCUITO INVERSOR COM SOMADOR.....	59
4.4.1 – Procedimentos iniciais.....	60
4.4.2 – Simulações com interconexões não-ideais.....	62
4.4.3 – Discussão sobre retardo e velocidade.....	65
4.4.3.1 – Tipos de simulação do SIMON.....	65
4.4.3.2 – Capacitância da interconexão.....	65
4.4.3.3 – Resistência da interconexão.....	66
4.4.3.4 – Comprimento da interconexão.....	67

4.4.3.5 – Interconexões mais realistas e frequência de operação.....	67
4.5 – BONS PROCEDIMENTOS PARA SIMULAÇÕES.....	69
5 – CONCLUSÕES E RECOMENDAÇÕES.....	70
REFERÊNCIAS BIBLIOGRÁFICAS.....	72
APÊNDICES.....	76
A - RESISTÊNCIA DE TUNELAMENTO.....	77
B - ROTINA DO MATLAB.....	78
C – ARTIGO SUBMETIDO NA ÁREA DE NANOELETRÔNICA.....	82
D – 1º ARTIGO SUBMETIDO NA ÁREA DE MICROELETRÔNICA.....	85
E – 2º ARTIGO SUBMETIDO NA ÁREA DE MICROELETRÔNICA.....	91

LISTA DE TABELAS

Tabela 2.1 - Comparação entre dispositivos microeletrônicos e nanoeletrônicos.....	39
Tabela 2.2 - Comparação entre categorias de arquitetura paralela.....	39
Tabela 3.1 - Valores típicos para a resistência e capacitância por unidade de comprimentos para nanofios de cobre e nanotubos de carbono.....	47
Tabela 3.2 - Valores típicos para a resistência e capacitância para nanofios de cobre e nanotubos de carbono com comprimento de 7,11 μm	47
Tabela 4.1 - Valores dos parâmetros da rede SET-WTA com dois neurônios.....	50
Tabela 4.2 - Valores dos parâmetros da rede SET-WTA com quatro neurônios.....	56
Tabela 4.3 - Dados da análise dos efeitos causados pelo modelo de interconexão não-ideal sobre a rede WTA com quatro neurônios.....	59
Tabela 4.4 - Valores dos parâmetros do circuito inversor com somador.....	61
Tabela 4.5 - Dados da análise dos efeitos causados pelo modelo de interconexão não-ideal sobre o circuito inversor com somador.....	68

LISTA DE FIGURAS

Figura 1.1 - Tendência de evolução das dimensões mínimas características.....	16
Figura 1.2 - Tendência de evolução do número de transistores por chip (Intel).....	16
Figura 1.3 - Tendência de evolução da frequência de operação.....	17
Figura 1.4 - Arquitetura global do processador nanoeletrônico.....	20
Figura 2.1 - Eletrodos A e B separados por uma ilha.....	27
Figura 2.2 - Junção-túnel.....	28
Figura 2.3 - Tunelamento através de uma barreira de potencial. Os pontos A e B correspondem aos pontos críticos clássicos.....	28
Figura 2.4 - Caixa mono-elétron.....	32
Figura 2.5 - Diagramas de energia.....	32
Figura 2.6 - Característica do bloqueio de Coulomb.....	33
Figura 2.7 - Transistor mono-elétron.....	34
Figura 2.8 - Característica $I \times V_g$ do transistor mono-elétron.....	34
Figura 2.9 - Tipos de co-tunelamento.....	36
Figura 2.10 - Rede WTA com conexões diretas das entradas para os neurônios (linha cheia) e conexões inibitórias laterais (linhas tracejadas).....	41
Figura 2.11 - Rede SET-WTA de 2 neurônios com inibição lateral e interconexão ideal...42	
Figura 2.12 - Circuito inversor com SET.....	43
Figura 2.13 - Tensões de entrada V_{in} e saída V_{out} do inversor a 77 K.....	43
Figura 2.14 - Circuito somador decimal com SET.....	44
Figura 2.15 - Tensões de entrada V_{in} e saída V_{out} do somador a 77 K.....	44
Figura 4.1 - Saídas v_1 e v_2 do SET-WTA com dois neurônios e interconexões ideais para simulações no SIMON e no MatLab.....	50
Figura 4.2 - Rede SET-WTA de 2 neurônios com inibição lateral e interconexão não-ideal.....	51
Figura 4.3 - Saídas v_1 e v_2 do SET-WTA com dois neurônios e interconexões não-ideais para simulações no SIMON e no MatLab.....	52
Figura 4.4 - Diferença (valores simulados no SIMON menos valores simulados no53	

MatLab) entre as saídas v_1 e v_2 do SET-WTA com dois neurônios e interconexões ideais.....	
Figura 4.5 - Diferença (valores simulados no SIMON menos valores simulados no MatLab) entre as saídas v_1 e v_2 do SET-WTA com dois neurônios e interconexões não-ideais.....	53
Figura 4.6 - Diferença entre simulações com interconexões ideal e não-ideal (valores simulados no MatLab) para as saídas v_1 e v_2 do SET-WTA com dois neurônios.....	54
Figura 4.7 - Rede SET-WTA de 4 neurônios com inibição lateral e interconexão ideal...55	
Figura 4.8 - Saídas v_1, v_2, v_3 e v_4 do SET-WTA com quatro neurônios e interconexões ideais.....	57
Figura 4.9 - Saídas v_1, v_2, v_3 e v_4 do SET-WTA com quatro neurônios e interconexões não-ideais.....	58
Figura 4.10 - Circuito inversor interconectado a somador de duas entradas com interconexão ideal.....	60
Figura 4.11 - Entradas V_{in1} e V_{in2} do inversor e do somador, respectivamente.....	61
Figura 4.12 - Saídas V_{out1} e Q_{out} do inversor e do somador, respectivamente, com interconexão ideal.....	61
Figura 4.13 - Cálculo de $-C_g \times (V_{out1} + V_{in2})$	62
Figura 4.14 - Parte da tela do editor SIMON com o circuito inversor interconectado ao somador de duas entradas com interconexão não-ideal.....	63
Figura 4.15 - Parte da tela do editor SIMON com o circuito inversor interconectado ao somador de duas entradas com interconexão não-ideal (ordem dos componentes invertida).....	64
Figura 4.16 - Saídas V_{out1} e Q_{out} do inversor e do somador, respectivamente, com interconexão não-ideal.....	64
Figura 4.17 - Saídas V_{out1} e Q_{out} do inversor e do somador, respectivamente, com interconexão não-ideal no limite de C_{con}	66
Figura 4.18 - Saídas V_{out1} e Q_{out} do inversor e do somador, respectivamente, com interconexão não-ideal no limite de R_{con}	67

LISTA DE SÍMBOLOS, NOMECLATURA E ABREVIACÕES

AFM: Atomic Force Microscopy.

C_{con} : capacitância da interconexão.

C_j : capacitância de carregamento da junção-túnel.

C_{Σ} : capacitância resultante da ilha.

CCD: Charge Coupled Device

CMOS: Complementary Metal - Oxide - Semiconductor.

D_i : densidade de estados inicial.

D_f : densidade de estados final.

e : carga elementar de um elétron.

EBL: Electron-Beam Lithography.

E_C : energia eletrostática.

ΔE : variação na energia devido ao tunelamento de um elétron.

f : frequência.

GIGA: 10^9 .

GSI: Giga Scale Integration.

h : constante de Planck.

IL: inibição lateral.

k_B : constante de Boltzmann.

L_{con} : comprimento da interconexão.

MOS: Metal - Oxide – Semiconductor.

$|P_T|^2$: probabilidade de transmissão.

Q_0 : carga de desvio.

R_{con} : resistência da interconexão.

R_K : resistência quântica de tunelamento.

R_j : resistência de tunelamento da junção-túnel.

SEM: Scanning Electron Microscopy.

STM: Scanning Tunneling Microscopy.

SIA: Semiconductor Industry Association.

SPL: Scanned Probe Lithography.

T : temperatura de operação.

t : tempo de retardo.

t_{con} : tempo de retardo causado pela interconexão.

TEM: Transmission Electron Microscopy.

TERA: 10^{12} .

TSI: Tera Scale Integration.

V_C : tensão de bloqueio de Coulomb.

VLSI: Very-Large-Scale Integration.

WTA: Winner-Take-All.

Γ_j : taxa de tunelamento na junção j .

1 - INTRODUÇÃO

Atualmente, a indústria semicondutora é capaz de fornecer produtos como microprocessadores operando a 3 GHz ou frequências superiores, e com mais de 100 milhões de transistores, além de chips de memória de 1 GB ou maiores. Esse rápido progresso tecnológico foi previsto em 1965 por Gordon Moore, autor da lei de Moore. A lei de Moore prevê que a cada 18 meses a capacidade de processamento dos processadores dobra. Até a presente data, esta previsão tem se mantido às custas da diminuição das dimensões características dos transistores que os compõem. A evolução dos processadores em relação às dimensões mínimas características, ao número de transistores por chip e à frequência de operação está ilustrada, respectivamente, nas Figuras 1.1, 1.2 e 1.3 [4].

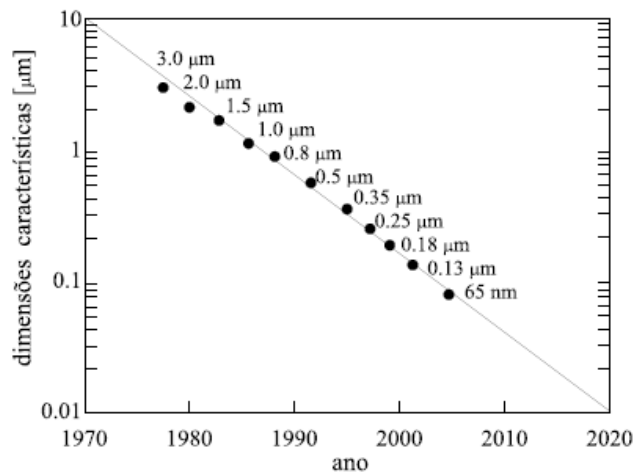


Figura 1.1 – Tendência de evolução das dimensões mínimas características. [4]

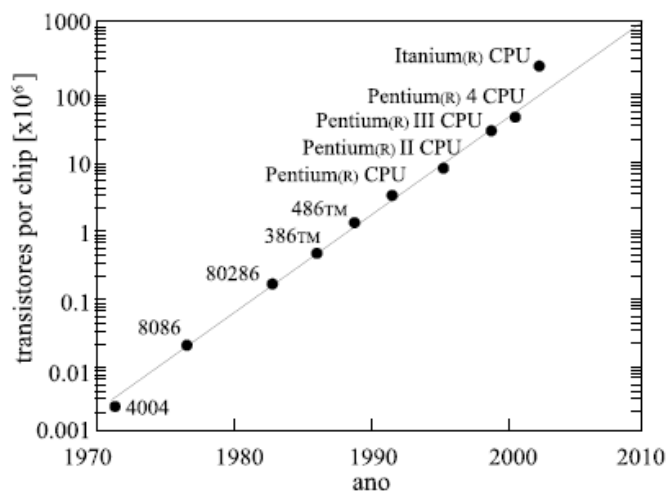


Figura 1.2 – Tendência de evolução do número de transistores por chip (Intel). [4]

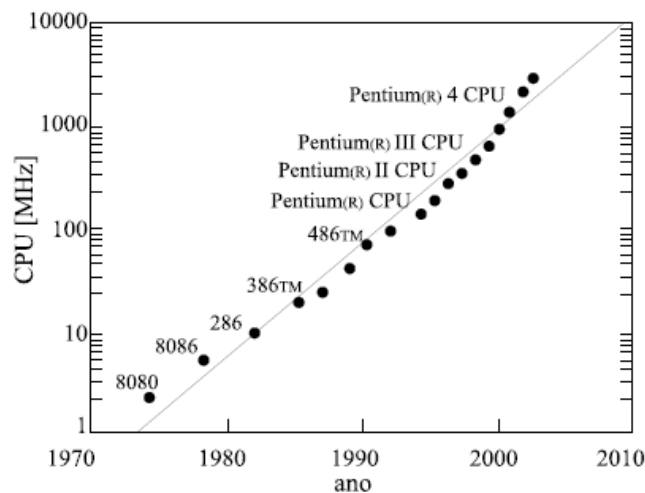


Figura 1.3 – Tendência de evolução da frequência de operação. [4]

1.1 - TECNOLOGIA DE FABRICAÇÃO ATUAL

Nos últimos 40 anos, o transistor MOS se tornou o bloco de construção básico da maioria dos dispositivos computacionais [42]. O crescimento gradual de sua popularidade está relacionado à sua capacidade de prosseguir com a diminuição, também gradual, de sua dimensão característica, conforme as previsões citadas anteriormente. Atualmente, essa dimensão está abaixo de 100 nm [42]. Entretanto, as leis da mecânica quântica e as limitações das tecnologias de fabricação disponíveis deverão, em breve, impedir a diminuição da dimensão característica do transistor MOS. Portanto, várias alternativas para a construção de circuitos de alta densidade estão sendo pesquisadas. Esses dispositivos, cujas dimensões são menores que dezenas de nanômetros são chamados de nanodispositivos [42].

A tecnologia MOS irá, eventualmente, atingir limites fundamentais e não será capaz de prover os aumentos esperados de desempenho. Em consequência desses limites, o tipo de evolução tecnológica que foi a base do desenvolvimento dos circuitos integrados nos últimos 40 anos não será mais capaz de sustentar a taxa de crescimento demonstrada até os dias de hoje. Uma nova tecnologia de circuitos integrados será necessária para substituir os transistores MOS como propulsora da evolução dos circuitos eletrônicos [4,44].

1.2 - DISPOSITIVOS NANOELETRÔNICOS

Para continuar a miniaturização dos circuitos eletrônicos até uma escala nanométrica, ou até mesmo uma escala molecular, várias alternativas ao transistor MOS têm sido investigadas. Esses novos dispositivos eletrônicos em escala nanométrica (nanoeletrônicos) podem executar chaveamentos e amplificações como os transistores atuais [19]. Todavia, diferente dos transistores MOS, que operam baseados no movimento de massas de elétrons em materiais densos, os novos dispositivos levam em consideração os fenômenos quânticos que emergem em escala nanométrica, incluindo a discretização dos elétrons [19].

Existem duas categorias de dispositivos nanoeletrônicos emergentes:

- dispositivos de efeito quântico e dispositivos mono-elétron; e
- dispositivos eletrônicos moleculares.

Ambas as categorias citadas fazem uso dos vários efeitos quânticos que dominam a dinâmica do elétron em escala nanométrica.

A fabricação de dispositivos de efeito quântico e mono-elétron é a alternativa seguida por muitos grupos que pesquisam novas tecnologias de dispositivos. Os novos dispositivos são desenvolvidos usando os mesmos semicondutores que os transistores atuais [19].

Nanodispositivos tendem a ser uma alternativa para o desenvolvimento de circuitos integrados em escala GIGA (10^9) e até mesmo TERA (10^{12}) com dimensões e performance acima das projeções do *roadmap* da Associação de Indústrias Semicondutoras (SIA) [44, 6]. De acordo com essas projeções, um processador CMOS com aproximadamente 10^9 dispositivos ativos deverá consumir até 180 W quando operando a uma frequência de 1 GHz em temperatura ambiente. Dispositivos mono-elétron [21, 31], por exemplo, por apresentarem baixo consumo, controle de corrente e baixo ruído, devem permitir a realização de chips contendo uma quantidade de dispositivos maior do que a indicada pelo *roadmap*, mas, ainda assim, respeitando as restrições de área e consumo de energia. Como resultado, um processador TSI (*Tera-Scale Integration*) pode ser um desafio realizável no futuro.

1.3 - INTERCONEXÕES

O sistema de interconexões para este processador TSI deverá ser o responsável por carregar a tensão de alimentação para cada transistor, fornecer um sinal de *clock* com baixo retardo, para sincronização dos circuitos dinâmicos, além de distribuir os sinais de dados e controle através do chip [8].

A miniaturização das interconexões, ao contrário do que acontece com os transistores, não aumenta o seu desempenho. Alguns problemas que podem ser facilmente enumerados com a miniaturização das interconexões são:

- degradação da resistividade, que pode aumentar bastante;
- dificuldades no controle da razão de aspecto das vias, que é a razão entre a altura e a largura do material das interconexões;
- dificuldades no controle de planaridade;
- problemas de confiabilidade devido a estresses elétricos, térmicos e mecânicos em um sistema de interconexões multinível; e
- aumento da capacitância de acoplamento entre vias que ficarão cada vez mais próximas.

Por este motivo, os limites físicos das interconexões ameaçam, potencialmente, desacelerar ou até parar o progresso histórico que vem sendo alcançado pela indústria de semicondutores nos últimos 40 anos [8].

1.4 - PROCESSADORES

Questões como área, dissipação e velocidade incentivam o projeto de processadores em escala GIGA ou TERA usando dispositivos nanoeletrônicos. No caso dos dispositivos baseados em transistores mono-elétron, adotados neste trabalho, algumas características produzidas pelo comportamento quântico comprometem sua funcionalidade. Entre essas características estão as cargas de desvio e o co-tunelamento. O conceito de um sistema nanoeletrônico baseado em arquiteturas de redes neurais é uma possível solução para esses problemas [7].

Tal sistema, em consequência de sua redundância, apresenta as seguintes qualidades:

- são robustos contra flutuações locais devido às cargas de desvio e ao co-tunelamento; e
- sua arquitetura paralela permite processamento local da informação.

Uma possível arquitetura global de sistema para um processador nanoeletrônico em escala GIGA ou TERA teria, como proposto em [7]:

- um núcleo nanoeletrônico com mais de 10^{12} elementos de processamento, responsável pelas principais tarefas de processamento de dados;
- um processador de controle com dispositivos convencionais, responsável pelas tarefas de controle e de comunicação e, eventualmente, aprendizagem da rede neural usada;
- unidades de entrada e saída de dados; e
- facilidades internas de interface compatíveis com a organização das redes neurais e com multiplexação.

Um esquema é mostrado na Figura 1.4.

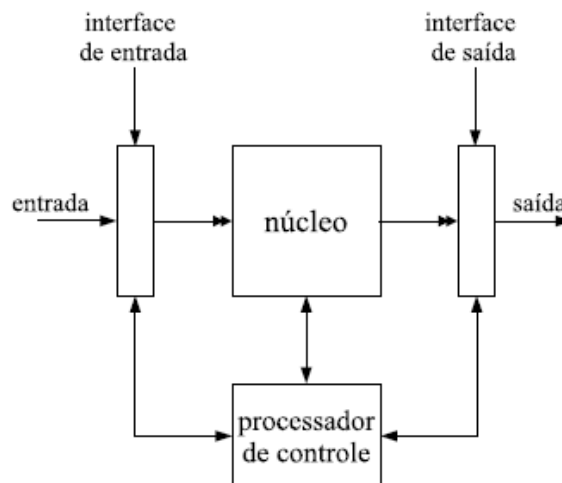


Figura 1.4 – Arquitetura global do processador nanoeletrônico. [7]

A investigação desta arquitetura tem considerado, até o presente momento, a utilização de interconexões ideais para interligação entre os diversos dispositivos nanoeletrônicos, ou

seja, a saída de uma célula ou dispositivo chega sem interferência alguma na entrada do próximo dispositivo.

1.5 - OBJETIVO DA DISSERTAÇÃO

Neste trabalho, os desempenhos elétricos de associações conhecidas de subcircuitos nanoeletrônicos são estudados com o auxílio de modelos de interconexões cujos parâmetros podem ser alterados. As associações são, então, simuladas com parâmetros variando de uma situação quase-ideal até situações de pior caso e seus comportamentos dinâmicos são analisados.

O objetivo deste trabalho é determinar a influência das interconexões sobre o comportamento de diversos circuitos e estabelecer limites para suas funcionalidades relacionados com as interconexões.

No Capítulo 2, a fundamentação teórica de dispositivos nanoeletrônicos e do funcionamento dos circuitos propostos neste trabalho é apresentada.

No Capítulo 3, a metodologia adotada para desenvolver este trabalho é descrita em detalhes.

No Capítulo 4, as simulações dos circuitos propostos com modelos de interconexão e os resultados obtidos são apresentados.

No Capítulo 5, as conclusões e recomendações para trabalhos futuros são feitas.

2 - FUNDAMENTAÇÃO TEÓRICA

2.1 - NANOTECNOLOGIA

A nanotecnologia é a área do conhecimento que mais rapidamente tem crescido nos últimos anos. Espera-se que a mesma deva liderar a fusão das ciências naturais como física, química, biologia e medicina já que todos os efeitos serão eventualmente dependentes de dimensões nanométricas [43].

Na escala nanométrica, as propriedades físicas, químicas e biológicas dos materiais diferem de formas fundamentais e de grande valor das propriedades de átomos individuais e moléculas ou materiais de maior volume [39]. Os novos resultados físicos e descobertas neste campo foram alcançados principalmente pelo desenvolvimento de novas técnicas de medição, métodos de fabricação e novos materiais.

Nanotecnologia é a compreensão e controle da matéria nos níveis atômicos, molecular e supra-molecular, que compreendem dimensões em torno de 1 a 100 nanômetros, para criar materiais, dispositivos e sistemas com novas propriedades fundamentais e funcionalidades, que possam ser exploradas para possibilitar novas aplicações. Aqui está uma lista de ordens de grandeza, para que se possa ter uma perspectiva das dimensões envolvidas:

- 10^{-15} m = 1 fm: raio típico de um núcleo atômico;
- 10^{-10} m = 1 Å: raio típico de um átomo;
- 10^{-9} m = 1 nm: diâmetro de um nanotubo de carbono de parede única;
- 10^{-8} m: limites da litografia por feixe de elétrons, tamanho dos grãos em típicos filmes de Au evaporado, raio de Bohr de doadores em semicondutores;
- 10^{-7} m: comprimento de onda da luz ultravioleta profunda, limite da fotolitografia, comprimento de onda de Fermi em semicondutores 2D típicos;
- 10^{-6} m = 1 μm: comprimento de onda da luz infravermelha;
- 10^{-4} m = 0,1 mm: diâmetro de um cabelo humano espesso, comprimento de onda de uma radiação na faixa de THz.

2.2 - NANOELETRÔNICA

Na área da eletrônica, a nanoeletrônica não é simplesmente a diminuição das dimensões mínimas dos dispositivos microeletrônicos para valores na ordem de nanômetros. A nanoeletrônica leva a uma transição dos transistores familiares para novos tipos de dispositivos e arquiteturas de circuito [43], que consideram as propriedades quânticas dos materiais.

No nível dos dispositivos, os transistores em silício podem dar lugar a novos materiais, como moléculas orgânicas ou nanofios inorgânicos. No nível das interconexões, a microeletrônica usa fios longos e grossos, enquanto a nanoeletrônica busca o uso de nano fios pequenos. Finalmente, novas arquiteturas serão necessárias para fazer uso de estruturas simples e localmente conectadas que são imperfeitas e que são compostas de dispositivos cujo desempenho varia largamente [34].

A nanoeletrônica, em combinação com novos paradigmas computacionais, terá um papel importante em áreas onde a tecnologia tradicional alcançou seu limite. Considerando aspectos econômicos, essa influência será possivelmente fortalecida se a nanoeletrônica oferecer uma melhor relação custo/eficiência na produção de chips complexos [20].

2.3 - CARACTERIZAÇÃO E FABRICAÇÃO DE NANOESTRUTURAS

Esta seção tem o objetivo de apresentar algumas formas de caracterização e fabricação de nanoestruturas que podem ser utilizadas nos dispositivos aqui propostos. A apresentação formal de uma proposta de metodologia para fabricação de circuitos nanoeletrônicos foge ao escopo deste trabalho.

2.3.1 - Introdução

A tarefa de fazer estruturas na escala nanométrica é complicada, entre outras coisas, pela dificuldade de se ver o que está acontecendo nestas dimensões tão extremas. Alguns métodos de investigação da escala nanométrica são:

- Microscopia por Varredura de Elétrons (SEM);
- Microscopia por Transmissão de Elétrons (TEM);
- Microscopia por Força Atômica (AFM);
- Microscopia por Varredura de Tunelamento (STM).

As técnicas de fabricação podem ser divididas em duas categorias: *top-down* e *bottom-up*. As técnicas *top-down* de nanofabricação são, geralmente, extensões de métodos de microfabricação.

2.3.2 - Métodos de nanofabricação *top-down*

2.3.2.1 - Extensão da fotolitografia

O método de fabricação padrão usado para fazer quase todos os dispositivos microeletrônicos de hoje em dia é a fotolitografia. Este método usa essencialmente substâncias químicas ativadas por luz para definir um modelo que será transferido para o substrato.

O modelo desejado é definido originalmente em uma máscara, freqüentemente em um tamanho maior que o modelo final desejado. Luz ultravioleta é projetada através da máscara e estruturas ópticas são usadas para reduzir e projetar a imagem da máscara no substrato. Essa exposição define o limite da fotolitografia convencional por causa dos limites causados pela difração nas estruturas ópticas de projeção.

A fotolitografia pode ser usada para produzir características abaixo de 100 nanômetros, mas a dificuldade é alta.

2.3.2.2 - Litografia por feixe de elétrons (EBL)

Este é o método mais comum e é uma analogia direta à fotolitografia. A maioria dos sistemas de litografia por feixe de elétrons existentes são baseados em MVE modificada, na qual computadores são usados para direcionar o feixe de elétrons sobre a superfície da amostra.

A litografia por feixe de elétrons pode produzir rotineiramente linhas de 20 a 30 nm em PMMA. Características menores são possíveis somente sob certas circunstâncias. O maior problema com este método é a velocidade em larga escala, pois requer que o feixe de elétrons desenhe uma linha de cada vez, enquanto a fotolitografia imprime uma face completa do substrato de uma única vez.

2.3.2.3 - Litografia por varredura de sonda (SPL)

Este tipo de litografia usa métodos de microscopia por varredura de sonda, como MFA e MVT. Uma abordagem é usar a sonda ou ponta presente nos microscópios para posicionar átomos individuais, um de cada vez, e criar nanoestruturas. Outra abordagem é modificar quimicamente a superfície do substrato com a sonda. Este método sofre das mesmas dificuldades de velocidade em larga escala que a litografia por feixe de elétrons.

2.3.3 - Métodos de nanofabricação *bottom-up*

Muitos acreditam que, para ir da microeletrônica para a nanoeletrônica, será necessária uma mudança das técnicas *top-down* para abordagens *bottom-up* [27]. A esperança é que as nanoestruturas possam ser construídas molécula por molécula, de forma controlada e automática de “baixo para cima” [34].

Os métodos de nanofabricação *bottom-up* são baseados nas seguintes técnicas:

2.3.3.1 - Crescimento de cristais

Nesta técnica, a aproximação de materiais com diferentes regiões de banda proibida pode resultar em transferência de cargas e no encurvamento de bandas, próximo a região de interface entre os materiais. Isto pode ser controlado por sistemas de crescimento de cristais para certos materiais, como os compostos III-V. Com as condições corretas de crescimento e para certos materiais, é possível criar cristais que crescem, literalmente, uma camada atômica de cada vez.

2.3.3.2 - Métodos químicos

A química de síntese se preocupa com a produção de estruturas menores do que 10 nm, ou seja, com dimensões de moléculas. Para a maior parte das aplicações químicas, o interesse está na produção de grandes quantidades (em torno de 10^{22}) dessas moléculas. Já para a física, o interesse está no exame de quantidades muito menores de tais sistemas, normalmente, com algum grau de isolamento entre eles.

Um grande número de pesquisadores vê os métodos de síntese química e seus materiais como um eventual substituto para os processos e elementos ativos na indústria microeletrônica. Em princípio, a idéia de fazer trilhões de estruturas complicadas e virtualmente livres de falhas, em paralelo, torna a idéia da eletrônica molecular bastante atrativa.

2.3.3.3 - Auto-montagem

Esta técnica está muito relacionada com a síntese química, citada anteriormente. Sob certas circunstâncias, é possível tirar vantagem do mesmo tipo de forças energéticas e estatísticas que causam a ordem cristalina em sólidos e ter a formação espontânea de matrizes de nanoestruturas altamente ordenadas.

2.4 - EXEMPLOS DE DISPOSITIVOS NANOELETRÔNICOS

Os problemas que apareceram com a diminuição das dimensões características da tecnologia CMOS levaram à procura de outros tipo de tecnologia [3, 21, 29, 44], como já foi discutido no Capítulo 1. Atualmente, existem muitas idéias promissoras para a implementação de dispositivos nanoeletrônicos. Algumas sugestões de dispositivos baseados em silício são: dispositivos de tunelamento ressonante [5], transistores de tunelamento de camada dupla [17], transistores MOS de barreira Schottky [32, 47] e dispositivos mono-elétron [29, 31]. As mais atuais são os diodos de tunelamento ressonante e os transistores mono-elétron, que surgem como uma promessa potencial para o futuro. Na próxima seção, este último dispositivo é apresentado.

2.5 - DISPOSITIVOS BASEADOS EM TUNELAMENTO MONO-ELÉTRON

Um tratamento mono-elétron de um sistema é adotado quando é possível controlar o movimento e a posição de um único ou de um pequeno grupo de elétrons em um dispositivo [1, 13, 20, 21, 29, 30, 31].

Por exemplo, suponha-se um experimento de transporte de cargas em que uma diferença de tensão é aplicada entre dois eletrodos A e B separados por uma barreira isolante, ilustrados na Figura 2.1. No meio do isolante encontra-se um terceiro eletrodo, chamado de ilha por estar cercado por isolantes de todos os lados. Para se moverem do eletrodo A para o eletrodo B os elétrons devem passar através da ilha. Assume-se que o transporte de elétrons através das barreiras de potencial entre o eletrodo A e a ilha e entre a ilha e o eletrodo B ocorre por tunelamento mono-elétron. Esse fenômeno é tão rápido que pode ser considerado que os elétrons atravessam as camadas isolantes um de cada vez [21].

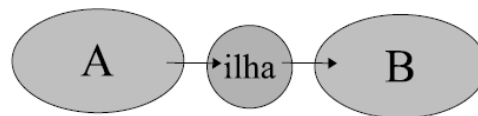


Figura 2.1 – Eletrodos A e B separados por uma ilha. [21]

O ponto principal do tunelamento mono-elétron é que, durante sua jornada entre os eletrodos A e B, o elétron, necessariamente, faz com que a carga da ilha varie de e , que é a carga fundamental do elétron. Essa é uma pequena transferência de carga se forem considerados dispositivos eletrônicos comuns. Durante a operação de um dispositivo de acoplamento de cargas (CCD), por exemplo, um pacote de carga é formado por, aproximadamente, 10^6 elétrons. Todavia, se a ilha for pequena o bastante, a variação de energia potencial imposta devido à presença de um elétron em excesso pode ser grande o suficiente para influenciar as probabilidades de tunelamento.

Existe um dispositivo de estado sólido através do qual as cargas fluem de maneira discreta, como descrito acima. Esse dispositivo, conhecido como junção-túnel e ilustrado na Figura 2.2, consiste em dois eletrodos metálicos separados por um isolante tão fino que permite que os elétrons o atravessem por tunelamento [13, 21].

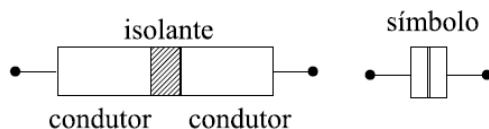


Figura 2.2 – Junção-túnel.

A resistência de tunelamento da junção R_j é uma quantidade fenomenológica que é definida quando uma diferença de tensão é aplicada aos dois terminais de uma junção-túnel. Sua definição está no Apêndice A. É um parâmetro macroscópico da junção que depende da área e da espessura da barreira isolante. Além de permitir o efeito de tunelamento, a junção apresenta uma capacitância de carregamento C_j , que é seu outro parâmetro macroscópico [13, 21, 31].

2.5.1 - Tunelamento

O tunelamento é, provavelmente, o fenômeno de transporte associado à transmissão quântica mais estudado [13]. Geralmente, o termo tunelamento refere-se ao transporte de partículas através de uma região classicamente proibida, ou seja, uma região na qual a energia total de uma partícula pontual clássica é menor que a energia potencial da região. Essa idéia está ilustrada na Figura 2.3, onde uma partícula de energia E incide sobre uma barreira potencial de formato arbitrário de altura $V_0 > E$.

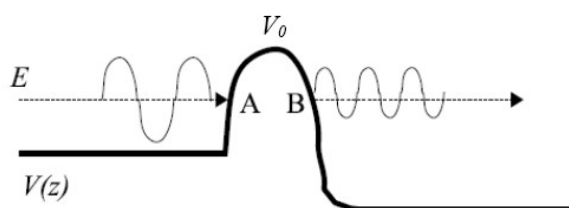


Figura 2.3 – Tunelamento através de uma barreira de potencial. Os pontos A e B correspondem aos pontos críticos clássicos. ([19] modificado)

Na mecânica clássica, uma partícula é completamente refletida nos chamados pontos críticos clássicos, ilustrados pelos pontos A e B da Figura 2.3, ou seja, os pontos onde a energia total se iguala a energia potencial. Na teoria eletromagnética, a solução da equação de onda deve satisfazer algumas condições de contorno na interface abrupta entre dois dielétricos de diferente permissividade, o que leva uma certa parte da onda incidente a ser

refletida e outra parte a ser transmitida. Do mesmo modo, na mecânica quântica, a função de onda e sua derivada normal devem ser contínuas através da fronteira formada entre duas regiões de diferentes energias potenciais, o que similarmente leva a reflexão e transmissão de ondas de probabilidade na região de contorno. A Figura 2.3 mostra a função de onda associada à partícula incidente à esquerda e à direita da barreira. Como o quadrado da função de onda representa a densidade de probabilidade de encontrar uma partícula em uma dada região do espaço, segue-se que quanticamente uma partícula incidente em uma barreira de potencial tem uma probabilidade finita de tunelar através da mesma e aparecer do outro lado.

Historicamente, o fenômeno do tunelamento foi reconhecido logo após os fundamentos da teoria quântica serem estabelecidos [13]. Já nos anos 60, foram desenvolvidas atividades relacionadas à medida de tunelamentos entre supercondutores e metais e entre os próprios supercondutores separados por camadas isolantes finas, que revelaram uma evidência extraordinária da densidade supercondutora de estados e o *gap* supercondutor associado. Isso levou ao desenvolvimento de várias teorias de tunelamento.

O modelo teórico mais utilizado para analisar o transporte de uma única carga é o modelo conhecido como teoria ortodoxa do tunelamento mono-elétron, na qual uma equação cinética é derivada da função distribuição descrevendo o estado de carregamento de uma junção ou de um sistema de junções. No limite semiclássico, essa teoria se mostrou extremamente valiosa na análise das propriedades de transporte de junções-túnel metálicas.

A fim de estabelecer uma equação cinética baseada no tunelamento para dentro e para fora de uma ilha, é geralmente mais conveniente descrever o tunelamento em termos de taxas de transição usando a teoria da perturbação em vez de em termos da probabilidade de tunelamento. A próxima seção descreve a teoria ortodoxa através dos seus postulados.

2.5.2 - Teoria ortodoxa do tunelamento mono-elétron

Em 1987, Averin e Likharev [29, 30, 31] formularam a teoria ortodoxa do tunelamento mono-elétron, que descrevia quantitativamente efeitos de carregamento importantes, como o bloqueio de Coulomb. A teoria ortodoxa faz as seguintes afirmativas:

1. Modelo de dimensão zero: as dimensões das ilhas são desprezíveis;
2. O evento de tunelamento é instantâneo;
3. A redistribuição das cargas após o tunelamento também é instantânea; e
4. O espectro de energia em condutores e ilhas é considerado contínuo: a quantização da energia do elétron é ignorada dentro dos condutores.

Apesar das limitações listadas acima, a teoria ortodoxa apresenta resultados compatíveis com todos os dados experimentais de sistemas com condutores metálicos [49]. Alguns experimentos indicaram alguns fenômenos que não são levados em conta pela teoria ortodoxa, como o co-tunelamento [29]. O resultado principal de todas essas condições assumidas pela teoria ortodoxa é que a taxa em que o tunelamento ocorre depende fortemente da mudança na energia livre causada pelo tunelamento [41].

2.5.3 - Efeito de carregamento

Os dispositivos mono-elétron de tunelamento baseiam-se em um fenômeno que ocorre quando elétrons estão prestes a entrar em um material condutor de dimensões muito pequenas isolado eletricamente, chamado efeito de carregamento [1, 18, 19, 40]. Quando esse material condutor, ou ilha, é extremamente pequeno, o potencial eletrostático (V_C) aumenta significativamente mesmo que apenas um elétron entre nessa ilha.

Por exemplo, para uma ilha de dimensões nanométricas, com uma capacitância equivalente C_Σ de 1 aF o aumento no potencial será de e/C_Σ que, considerando $e = 1,6 \times 10^{-19} C$, resulta em 160 mV. Esse valor é muito maior que o ruído térmico em temperatura ambiente que é 25,9 mV [40]. A repulsão de Coulomb evitará que outros elétrons entrem na ilha, a menos que o potencial dela seja intencionalmente diminuído por uma fonte de tensão de polarização externa. Se o potencial da ilha for diminuído gradualmente, outros elétrons poderão entrar e sair, um a um, dissipando uma potência desprezível.

Os requisitos básicos para que ocorram efeitos de carregamento são descritos a seguir. O primeiro considera o princípio da incerteza de Heisenberg dado pela equação:

$$\Delta E \Delta t > h \quad (2.1)$$

onde ΔE é a incerteza quanto a energia, Δt é a incerteza quanto ao tempo e h é a constante de Planck. No caso da junção-túnel, o tempo característico para flutuações de carga será dado pela constante de tempo associada à capacitância de carregamento da junção C_j e à resistência de tunelamento R_j , mostrado na Equação 2.2:

$$\Delta t \approx R_j C_j \quad (2.2)$$

A variação de energia associada a um elétron em excesso é dada pela Equação 2.3.

$$\Delta E = e^2/C_j \quad (2.3)$$

Combinando as Equações 2.1, 2.2 e 2.3, resulta a equação abaixo, que impõe uma condição para a resistência de tunelamento.

$$R_j > h/e^2 = R_K = 25,8 \text{ k}\Omega \quad (2.4)$$

A resistência R_K fornece o valor de limiar a partir do qual a natureza corpuscular do elétron tem predominância sobre a natureza ondulatória [21].

O segundo requisito exige que, para a operação em uma temperatura T , as dimensões da ilha devam ser tais que a energia eletrostática E_C associada seja muito maior que as flutuações térmicas existentes àquela temperatura, como explicitado na Equação 2.5:

$$E_C \gg k_B T \quad (2.5)$$

, onde k_B é a contante de Boltzmann.

As condições impostas pelas Equações 2.4 e 2.5 asseguram que o transporte de elétrons da ilha e para a ilha é controlado pelos efeitos de carregamento. Com o uso de tensões externas, a energia de carregamento das ilhas pode ser seqüencialmente diminuída ou aumentada de modo a manipular o fluxo dos elétrons [21]

Assim, se um condutor comum for interrompido por uma junção-túnel, a carga elétrica irá se mover através do sistema de dois modos: contínua e discretamente. Na medida em que a

carga flui continuamente através do condutor, a mesma irá se acumular na superfície do eletrodo em contato com a camada isolante (o eletrodo adjacente terá carga igual mas de sinal contrário) [21]. A carga Q se acumulará na junção-túnel até que seja energeticamente favorável para um elétron sofrer tunelamento através da barreira isolante. Quando isso ocorrer, a junção-túnel será descarregada em uma carga elementar e [41].

2.5.4 - Bloqueio de Coulomb

Em um circuito mono-elétron com apenas uma ilha é possível verificar os efeitos de carregamento descritos acima. O circuito mono-elétron mais simples com uma ilha está ilustrado na Figura 2.4. É o circuito da caixa mono-elétron [21].

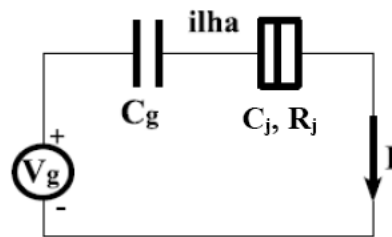
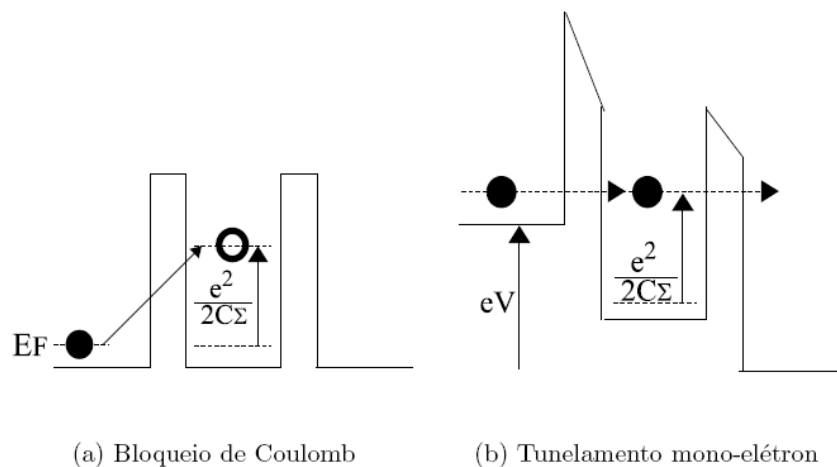


Figura 2.4 – Caixa mono-elétron.

Considerando que há um elétron na ilha, que é a região condutora entre C_g e a junção-túnel, a energia eletrostática E_C dela aumenta em $e^2/2C_\Sigma$, onde $C_\Sigma = C_j + C_g$. Nessa situação, ilustrada pelo diagrama de energia da Figura 2.5(a), um outro elétron só conseguirá tunelar para dentro da ilha se tiver energia o suficiente para ocupar o primeiro nível de energia vazio, ou seja, se tiver energia maior que E_C [1, 21].



(a) Bloqueio de Coulomb

(b) Tunelamento mono-elétron

Figura 2.5 – Diagramas de energia.

Quando uma tensão V_g é aplicada ao circuito, quedas de tensão ocorrerão em C_g e na junção, como está ilustrado no diagrama de energia da Figura 2.5(b). Se essa tensão V_g alcançar um dado valor de limiar, irá fornecer energia suficiente para que um elétron ocupe um dos níveis vazios da ilha, e haverá corrente fluindo pelo circuito. Esse fenômeno de suspensão do fluxo de elétrons é chamado de bloqueio de Coulomb [2, 9] e a tensão de limiar é a tensão de bloqueio de Coulomb (V_C) que é dada por e/C_Σ .

A curva característica do circuito, mostrando a região em que ocorre o bloqueio de Coulomb, é mostrada na Figura 2.6.

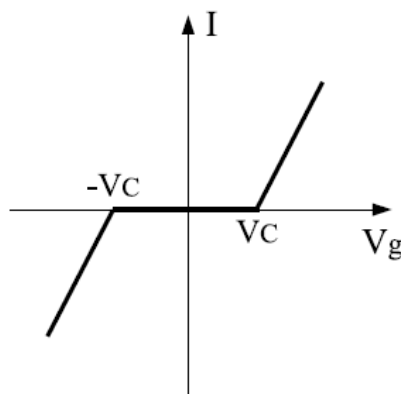


Figura 2.6 – Característica do bloqueio de Coulomb.

2.5.5 - Transistor mono-elétron

O transistor mono-elétron consiste em duas junções-túnel em série formando uma ilha. Nessa ilha, a energia eletrostática é controlada por uma tensão de porta V_g através de uma capacitância C_g . O circuito do transistor mono-elétron está ilustrado na Figura 2.7 [15, 21, 29, 31].

A ilha do transistor mono-elétron pode ser carregada por tunelamento através de uma junção e pode ser descarregada, também por tunelamento, através da outra junção. Esse fluxo controlado de cargas resulta em uma corrente [21].

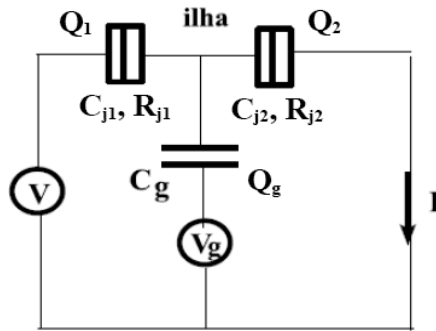


Figura 2.7 – Transistor mono-elétron.

A característica corrente versus tensão ($I \times V_g$) do transistor mono-elétron apresenta um efeito denominado oscilações de Coulomb. Esse efeito pode ser observado se a tensão V for mantida em um valor constante $V \ll e/C_\Sigma$, onde C_Σ é a capacitância equivalente da ilha $C_\Sigma = C_{j1} + C_{j2} + C_g$. Se a tensão V_g for aumentada gradualmente, começando em $V_g = 0$, quando essa tensão atingir o valor $V_g = e/2C_g$ dará condições para um elétron passar pela ilha, resultando em corrente. Quando esse elétron tunelar para fora da ilha, e a energia eletrostática dela voltar ao valor mínimo, nada acontecerá até que V_g atinja um valor que permita que outro elétron tunele que será $V_g = 3 (e/2C_g)$. Conseqüentemente, apenas em valores múltiplos ímpares da tensão $e/2C_g$ haverá corrente fluindo no circuito [21]. A Figura 2.8 mostra o efeito das oscilações de Coulomb na característica $I \times V_g$ do transistor mono-elétron.

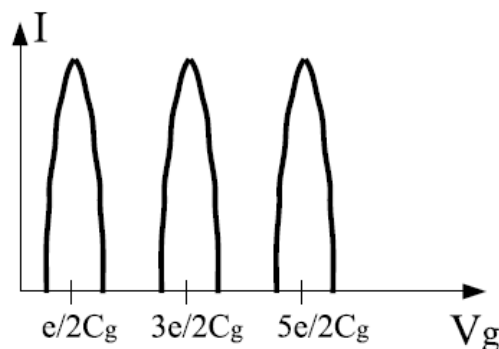


Figura 2.8 – Característica $I \times V_g$ do transistor mono-elétron.

2.5.5.1 - Equações características

A tensão de porta V_g controla a quantidade de elétrons n presentes na ilha na medida em que introduz uma carga $Q_g = C_g V_g$. Então, a carga da ilha Q será:

$$Q = Q_2 - Q_1 - Q_g = ne \quad (2.6)$$

onde Q_1 é a carga na junção 1 e Q_2 é a carga na junção 2. Aplicando a lei de Kirchhoff nas duas malhas do circuito, pode-se chegar a expressões para as cargas Q_1 , Q_2 e Q_g que estão nas Equações 2.7, 2.8 e 2.9, respectivamente [13].

$$Q_1 = \frac{C_{j1}}{C_\Sigma} \left[\left(C_{j2} + \frac{C_g}{2} \right) V + C_g V_g + ne \right] \quad (2.7)$$

$$Q_2 = \frac{-C_{j2}}{C_\Sigma} \left[- \left(C_{j1} + \frac{C_g}{2} \right) V + C_g V_g + ne \right] \quad (2.8)$$

$$Q_g = \frac{-C_g}{C_\Sigma} \left[\frac{1}{2} (C_{j2} - C_{j1}) V - (C_{j1} + C_{j2}) V_g + ne \right] \quad (2.9)$$

A expressão para energia eletrostática da ilha está na Equação 2.10.

$$E_{C(ilha)} = \frac{Q_1^2}{2C_{j1}} + \frac{Q_2^2}{2C_{j2}} + \frac{Q_g^2}{2C_g} \quad (2.10)$$

2.5.5.2 - Corrente no transistor mono-elétron

No transistor mono-elétron, a expressão para a corrente de tunelamento pode ser dada pela Equação 2.11 [21].

$$I(t) = e \frac{\Gamma_{j1}^+(0) \Gamma_{j2}^-(1)}{\Gamma_{j1}^+(0) + \Gamma_{j2}^-(1)} \quad (2.11)$$

onde e é $1,6 \times 10^{-19} C$ e $\Gamma_{j1}^\pm(n)$ é a taxa de tunelamento da junção 1 do transistor e $\Gamma_{j2}^\pm(n)$ é a taxa de tunelamento da junção 2. O sinal + indica que a carga está tunelando para dentro da ilha e o sinal - indica que ela está saindo [21, 38].

A taxa de tunelamento pode ser obtida da expressão [21, 38]:

$$\Gamma_{jl}^{\pm}(n) = \frac{\Delta E_j(n)}{e^2 R_j \left[1 - \exp\left(\frac{\Delta E_j(n)}{k_B T}\right) \right]} \quad (2.12)$$

onde ΔE_j é a variação da energia livre do sistema devido a um tunelamento, n é o número de elétrons livres na ilha ($n = 0$ ou $n = 1$) [21, 38], k_B é a constante de Boltzmann e T é a temperatura. A variação da energia livre do sistema, por sua vez, dependerá da diferença entre a energia eletrostática antes e depois de ocorrido um tunelamento. Então, para as duas junções do transistor [38]:

$$\Delta E_1^{\pm} = \frac{e}{C_{\Sigma}} \left[\frac{e}{2} \pm (ne + C_g V_g) \mp \left(C_{j2} + \frac{C_g}{2} \right) V \right] \quad (2.13)$$

$$\Delta E_2^{\pm} = \frac{e}{C_{\Sigma}} \left[\frac{e}{2} \pm (ne + C_g V_g) \pm \left(C_{j1} + \frac{C_g}{2} \right) V \right] \quad (2.14)$$

2.5.6 - Co-tunelamento

Quando um circuito apresenta mais que uma junção-túnel, podem ocorrer eventos de co-tunelamento. O co-tunelamento consiste no tunelamento de mais de um elétron através de barreiras túnel distintas ao mesmo tempo [49]. Geralmente, um circuito com N junções-túnel apresentará um co-tunelamento até de ordem N [41]. A Figura 2.9 mostra os tipos de eventos de co-tunelamento possíveis.

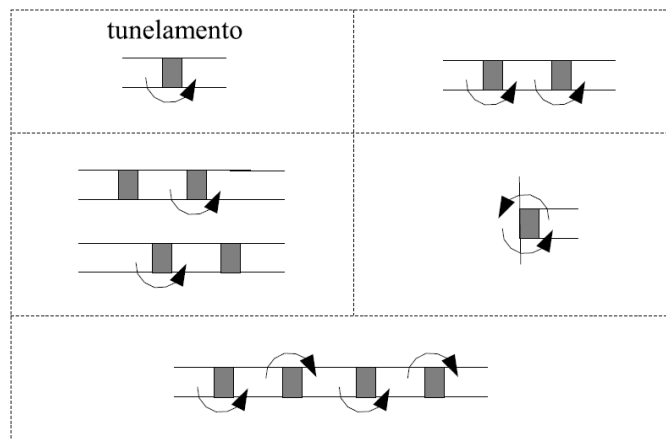


Figura 2.9 – Tipos de co-tunelamento.

Esse tipo de processo é uma fonte de erros em dispositivos mono-elétron. Especialmente em dispositivos lógicos que se apóiam na presença ou ausência de um único ou um pequeno número de elétrons.

2.5.7 - Cargas de desvio

Outra limitação dos circuitos mono-elétron são as cargas aleatórias de desvio [29]. Supondo que uma impureza ionizada Q_0 fique presa próxima à ilha do transistor mono-elétron, a uma distância comparável ao seu tamanho. Esse íon irá polarizar a ilha, criando em sua superfície uma carga imagem $-Q_0$ na ordem de grandeza de e . Esse valor pode ser suficiente para tirar o transistor do bloqueio de Coulomb [29].

Uma estimativa otimista prevê que um em cada mil dispositivos terá flutuações devido a cargas de desvio consideráveis, na ordem de $|Q_0| < 0,1e$. Valores maiores são inaceitáveis para uma aplicação VLSI [29].

A concepção de circuitos nanoeletrônicos, especialmente os mono-elétron, deve levar em consideração, além das propriedades e características apresentadas, as limitações de operação, como o co-tunelamento e as cargas de desvio. Na próxima seção a implementação de um processador nanoeletrônico será discutida.

2.6 - PROCESSADOR NANOELETRÔNICO

De modo similar ao progresso da microeletrônica nos últimos anos, o sucesso comercial da nanoeletrônica depende da disponibilidade de conceitos de sistemas adequados. Ou seja, os circuitos nanoeletrônicos devem ser desenvolvidos considerando o ponto de vista dos dispositivos e o ponto de vista dos sistemas [20].

O sucesso da microeletrônica nas últimas décadas foi baseado na diminuição de escala dos transistores de efeito de campo MOS e do resultante aumento de densidade nos chips de memória e lógicos [28]. Segundo a Associação de Indústrias Semicondutoras (SIA), em 2012, um chip CMOS (Complementary Metal-Oxide-Semiconductor) terá, aproximadamente, 10^{12} transistores e executará operações em velocidades de até 15 GHz

[44]. Ao final desse período, dispositivos com medidas características de, aproximadamente, 10 nm devem estar comercialmente disponíveis [28].

Para alcançar esse desempenho, algumas dificuldades deverão ser superadas [3, 44]. Entre elas, as maiores são [44]:

1. aumento do consumo de energia;
2. novas arquiteturas que superem os problemas das interconexões; e
3. os custos crescentes da litografia e da fabricação.

Além dessas questões, alguns problema físicos, como efeitos parasitários, dificuldade no controle da tensão de limiar, tunelamento e aumento da resistência de interconexão se tornam muito significantes na medida que as dimensões são reduzidas. Esses problemas, por sua vez, afetam a maneira como os dispositivos podem ser conectados [3]. Desse modo, o que passa a definir o desempenho das novas arquiteturas de processadores deixa de ser o desempenho da lógica e passam a ser as limitações das interconexões [3, 6]. Portanto, o foco da arquitetura em escala nano tende a passar do processamento para a comunicação.

A clássica arquitetura de von-Neumann, que separa claramente o conceito de arquitetura lógica do computador da sua implementação física, é de difícil implementação nesta nova tecnologia com dispositivos nanoeletrônicos [3, 6]. Seria complicado implementar uma arquitetura com interconexões globais – como é o caso de von-Neumann – utilizando dispositivos que apresentam problemas até mesmo na conexão de dois vizinhos.

A Tabela 2.1 mostra as principais diferenças entre os dispositivos microeletrônicos e os nanoeletrônicos. Como as dimensões mínimas são reduzidas de algumas centenas de nanômetros para alguns nanômetros, um cálculo simples pode mostrar que o número de dispositivos em um único chip de tamanho médio irá aumentar de 1 milhão para 10 bilhões. Entretanto, os fenômenos físicos dos quais os dispositivos atuais dependem tendem a deixar de ocorrer com a diminuição das dimensões. Assim, características dos dispositivos como, por exemplo, consumo de energia, capacidade de acionamento e características de transferência serão diferentes entre os dispositivos atuais e de uma classe de dispositivo nanoeletrônico para outra [14].

Tabela 2.1 – Comparação entre dispositivos microeletrônicos e nanoeletrônicos. [14]

Parâmetro	Microeletrônica	Nanoeletrônica
Dimensão mínima	50 – 1000 nm	5 – 10 nm
Consumo	1 μ W	1pW
Ganho de tensão	100	≈ 1
Estados estáveis	2	vários
<i>Clock</i> (GHz)	> 1	1000

Em consequência da baixa capacidade de acionamento dos dispositivos nanoeletrônicos, que levaria a problemas no caso de interconexões longas e na interface com o mundo exterior, uma arquitetura altamente regular, localmente interconectada e com interface periférica paralela forneceria combinação quase perfeita com as propriedades de vários nanodispositivos [14]. Algumas alternativas de computadores altamente paralelos já foram propostas e construídas. Entre elas existem as arquiteturas de entradas de dados paralela, a arquitetura de funções paralelas, que são baseadas na arquitetura de von-Neumann, e a arquitetura cognitiva, que é uma proposta bem diferente de arquitetura conexionista [14]. As características dessas categorias estão resumidas na Tabela 2.2.

Tabela 2.2 – Comparação entre categorias de arquitetura paralela. [14]

Parâmetro	Von-Neumann		Conexionista
	Dados paralelos	Funções paralelas	Cognitiva
Grau de paralelismo	alto	baixo	alto
Complexidade do processador	baixa	alta	média
Complexidade das interconexões	baixa	alta	alta/baixa
Quantidade de interfaces	baixa	alta	baixa
extensibilidade	alta	baixa	baixa

2.6.1 - Características desejáveis de um processador nanoeletrônico

A partir das especificações apresentadas anteriormente, algumas características desejáveis para o projeto de um processador nanoeletrônico em escala GIGA ou TERA [37] podem ser discutidas e são apresentadas a seguir:

1. conexões curtas: para minimizar os problemas de atraso na propagação da informação que os dispositivos nanoeletrônicos apresentam [3, 6, 8];
2. capacidade local de processamento: resultante das conexões de atraso minimizado, o que simplifica a arquitetura do processador; e
3. tolerante a falhas e a defeitos: é a maneira mais econômica de construir chips com bilhões ou trilhões de dispositivos, já que uma quantidade significativa deles será defeituosa [25].

Dentre as categorias de arquiteturas apresentadas na Tabela 2.2 as redes neurais artificiais (cognitivas), além de apresentarem redundância [38], são bastante robustas a flutuações locais e sua arquitetura paralela processa localmente a informação [7, 33, 48].

Redes neurais derivadas da neurobiologia e adaptadas à eletrônica possuem características como processamento paralelo, variação no tempo contínua e interação global entre os elementos da rede. Algumas aplicações dessas redes incluem otimização, programação linear e não-linear, memória associativa, reconhecimento de padrões e visão computacional [18].

2.7 - COMPONENTES DE UM PROCESSADOR NANOELETRÔNICO

Dentre os possíveis dispositivos que podem ser implementados dentro de um processador nanoeletrônico, como o descrito anteriormente, estão uma rede neural do tipo *Winner-Take-All* (WTA), um circuito inversor e um circuito somador decimal. Estes são os três tipos de dispositivos avaliados dentro deste trabalho e o seu funcionamento encontra-se descrito a seguir.

2.7.1 - Rede neural *Winner-Take-All* (WTA)

Na arquitetura de rede WTA [11, 12, 26], as entradas são diretamente conectadas às unidades de saída, que podem estar ligadas entre si por meio de conexões laterais inibitórias. A unidade de saída com maior valor, em resposta à entrada, terá mais chance de vencer a disputa com as outras unidades. A saída da unidade vencedora fica cada vez mais forte e seu efeito inibidor sobre as outras saídas torna-se dominante [12, 26]. Com o tempo, todas as outras unidades de saída ficarão inativas, exceto a vencedora. Portanto, a rede

neural básica para o aprendizado por competição é constituída de duas camadas: uma camada de entrada e uma camada competitiva. A Figura 2.10 mostra uma rede WTA simples com N entradas e N neurônios.

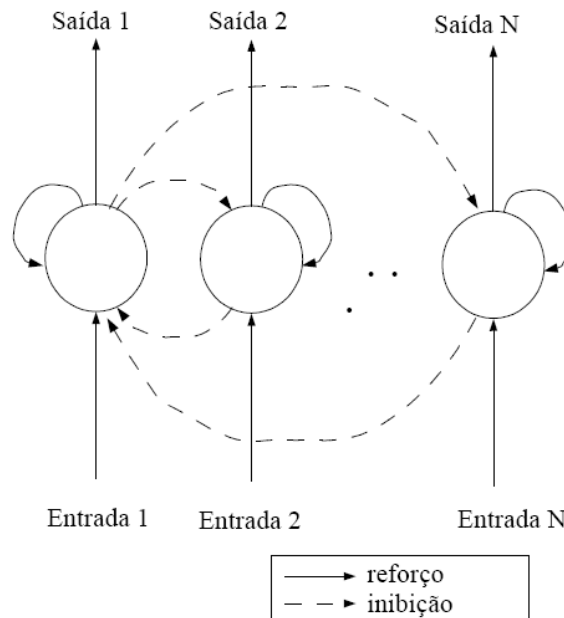


Figura 2.10 – Rede WTA com conexões diretas das entradas para os neurônios (linha cheia) e conexões inibitórias laterais (linhas tracejadas).

As conexões inibitórias laterais – ou inibição lateral – ocorre quando cada unidade de saída da camada competitiva inibe somente as outras unidades que estão mais próximas. O resultado é que a saída da rede destaca algumas das fronteiras e áreas de transição que aparecem nos padrões de entrada.

As redes *winner-take-all* são amplamente utilizadas para tomada de decisões, reconhecimento de padrões, redes com aprendizagem competitiva e redes neurais auto-organizáveis [11]. Além disso, têm sido usadas em várias aplicações de processamento de sinais, incluindo a extração de características de imagens, inibição não-linear, rede de Hamming, aplicações de processamento de imagens e compressão de vídeo [11].

2.7.1.1 - Implementação de uma rede WTA mono-elétron

No caso da implementação mono-elétron de uma rede WTA apresentada aqui, a idéia básica é demonstrar a factibilidade desse tipo de rede, em equivalência com o circuito de

tecnologia MOS. O funcionamento de uma rede WTA pode ser descrito a partir da conexão de dois neurônios, como mostra a Figura 2.11 [22].

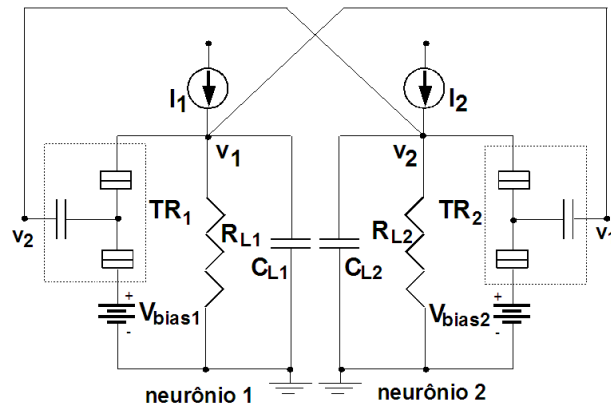


Figura 2.11 – Rede SET-WTA de 2 neurônios com inibição lateral e interconexão ideal.

Cada neurônio da Figura 2.11, portanto, terá um transistor mono-elétron TR, uma capacitância de carga C_L e uma resistência de carga R_L . Suponha-se que as correntes de entrada são tais que $I_1 < I_2$. Quando as entradas forem aplicadas ao circuito, no neurônio 1 o transistor TR_1 receberá em sua porta uma tensão $v_2(t)$ que fará com que ele conduza a corrente, e diminua a sua tensão de saída $v_1(t)$ a cada instante de tempo. Já o neurônio 2 receberá em sua porta uma tensão $v_1(t) < v_2(t)$ cada vez menor, impossibilitando seu transistor TR_2 de operar, e acumulando toda a tensão no capacitor C_{L1} e no resistor R_{L2} .

Para uma rede mono-elétron formada por N neurônios, sendo $N > 2$, o funcionamento é análogo. Quando as entradas são aplicadas à rede, o neurônio que recebe a maior corrente terá, conseqüentemente, a maior tensão de saída (ver Figura 2.11). Essa tensão, que estará conectada às portas (gates) de transistores dos outros neurônios, operará como uma conexão inibitória, fazendo com que a corrente desses transistores aumente, diminuindo as suas tensões de saída. De maneira contrária, no neurônio de maior corrente, as entradas de inibição estarão recebendo tensões cada vez menores, fazendo com que seus transistores conduzam cada vez menos e que a tensão de saída aumente até atingir um limite definido pela constante de tempo do circuito e pela tensão de polarização.

2.7.2 - Circuito inversor

O circuito inversor, como o próprio nome diz e similarmente ao encontrado em implementações CMOS, emite na sua saída um sinal digital invertido ao sinal de entrada.

Sendo assim, sua saída V_{out} possuirá um nível de tensão alto (nível lógico 1) quando o sinal de entrada V_{in} possuir um nível de tensão baixo (nível lógico 0). Da mesma forma, quando a entrada V_{in} possuir um nível de tensão alto (nível lógico 1), sua saída V_{out} possuirá um nível de tensão baixo (nível lógico 0). A implementação mono-elétron de um circuito inversor é apresentada na Figura 2.12 [23].

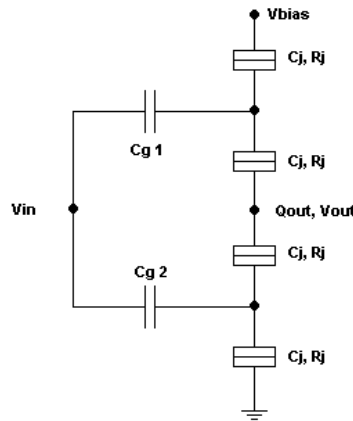


Figura 2.12 – Circuito inversor com SET.

Seu funcionamento pode ser observado na Figura 2.13 [23], a seguir, onde são mostrados sinais de entrada e saída para um circuito inversor com SET, como o da Figura 2.12, simulado a 77 K, que é a temperatura do nitrogênio líquido.

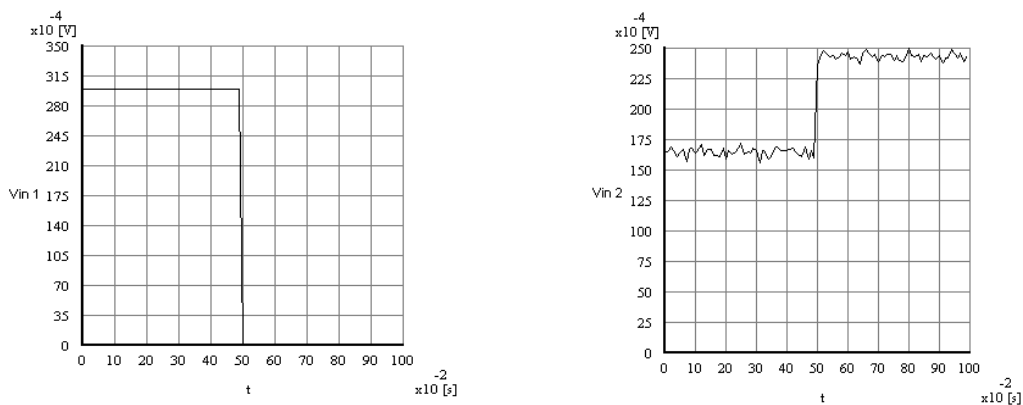


Figura 2.13 – Tensões de entrada V_{in} e saída V_{out} do inversor a 77 K.

Como pode ser observado, o nível lógico 0 na saída do inversor apresenta uma tensão residual diferente de 0 V. Além disso, ocorre uma perda de tensão entre entrada e saída para o nível lógico 1.

2.7.3 - Circuito somador decimal

O circuito somador decimal, também como esperado, realiza a soma dos sinais de tensão nas entradas. Esta soma, porém, é apresentada através da carga do sinal de saída segundo a Equação 2.15 [10]. O termo decimal é utilizado por esta soma não ser binária, como nos somadores digitais.

$$V_{g1} + V_{g2} = q/C_g \quad (2.15)$$

A implementação mono-elétron de um circuito somador decimal é apresentada na Figura 2.14 [23].

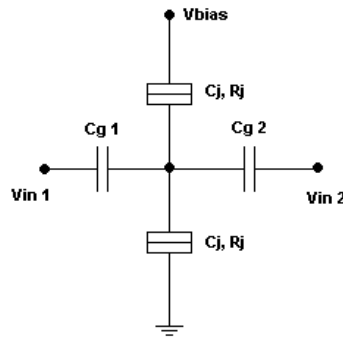


Figura 2.14 – Circuito somador decimal com SET.

Seu funcionamento pode ser observado na Figura 2.15 [23], a seguir, onde são mostrados sinais de entrada e saída para um circuito somador decimal com SET, como o da Figura 2.14, simulado a 77 K.

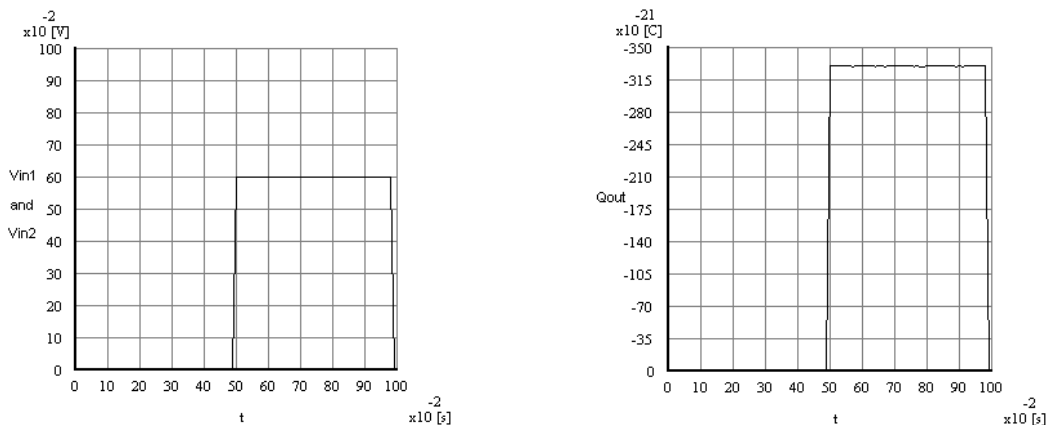


Figura 2.15 – Tensões de entrada V_{in} e saída V_{out} do somador a 77 K.

3 - METODOLOGIA EMPREGADA

3.1 - INTRODUÇÃO

Neste trabalho, procura-se levar em conta os efeitos de interconexões que não sejam ideais sobre o comportamento de circuitos com dispositivos monoelétron. Sendo assim, a metodologia apresentada neste capítulo visa, inicialmente, reproduzir resultados anteriores de simulações de subcircuitos existentes com interconexões ideais.

Após a confirmação desses resultados, tem-se como objetivo simular novamente os mesmos subcircuitos interligados, dessa vez, com certos efeitos das interconexões sendo modelados através de parâmetros concentrados.

3.2 - TIPOS DE SIMULAÇÃO

O comportamento de simuladores de circuitos monoelétron baseados em métodos estocásticos ao se alterar o ambiente de impedância [21] entre os subcircuitos não é intuitivo e torna-se de difícil predição. Sendo assim, para confirmar os resultados obtidos, dois tipos de simulações foram inicialmente realizadas:

1. Com a ferramenta SIMON (Simulador de Nanoestruturas), simulador conhecido e largamente aceito que utiliza o método de Monte-Carlo e de Master Equation [50];
e
2. Com a ferramenta MatLab [36], através de rotina em que o funcionamento do SET é aproximado por função quase-analítica [35].

A rotina usada para as simulações através do MatLab se encontra no Apêndice B.

Além disto, a utilização de ambas ferramentas visava também saber se havia coerência entre os dois tipos de simulação, inclusive porque os detalhes dos procedimentos de cálculo utilizados no SIMON não são completamente conhecidos.

Após as simulações dos primeiros subcircuitos interconectados, a coerência entre ambas as simulações pôde ser observada. Apesar disto, a ferramenta SIMON simulou melhor o comportamento destes subcircuitos, como o retardo RC, sobre a influência de algumas características de interconexões reais. No capítulo sobre discussão dos resultados, estas diferenças poderão ser mais bem observadas e esclarecidas. Isto deveu-se à simplicidade do algoritmo quase-analítico [35] adotado na ferramenta MatLab, que implementava somente o comportamento de um transistor mono-elétron, não levando em conta a influência das outras junções-túnel existentes nos subcircuitos adjacentes sobre o funcionamento global do sistema.

Como a criação de uma ferramenta aperfeiçoada que una a precisão de uma ferramenta baseada em métodos estocásticos para o cálculo de circuitos baseados em fenômenos de tunelamento mono-eletrônico à rapidez de uma ferramenta que faça cálculos de circuitos baseados nas Leis de Kirchhoff fugia ao escopo deste trabalho, passou-se a adotar somente a simulação na ferramenta SIMON.

3.3 - PARÂMETROS CONCENTRADOS DE INTERCONEXÃO

Como primeira aproximação, foi considerada a resistividade intrínseca ρ do material usado na interconexão, bem como a capacitância de acoplamento C_{con} entre duas trilhas muito próximas. Sob estas considerações, o modelo adotado para a interconexão foi um circuito RC simples que pode ser facilmente modelado.

Neste ponto, tornou-se necessário obter valores baseados em implementações reais existentes para interconexões. De [45], valores típicos para a resistência e capacitância por unidade de comprimento puderam ser obtidos para interconexões realizadas através de nanofios de cobre (Cu) com as seguintes características:

1. 15 nm de largura;
2. 15 nm de altura;
3. 15 nm de separação entre cada fio na mesma camada;
4. Uma camada de separação de 3 nm entre duas camadas de nanofios; e
5. Óxido de silício (SiO_2) (constante dielétrica de 3,9) ao redor de todos os fios.

De [46] também foi possível obter valores típicos para interconexões realizadas através de nanotubos de carbono (CNT) com as seguintes características:

1. Um único nanotubo de carbono isolado com parede única;
2. Contatos perfeitos com os dispositivos;
3. 1 nm de diâmetro; e
4. 1 μm de separação de um plano de terra;

Estes valores típicos são apresentados na Tabela 3.1, a seguir:

Tabela 3.1 – Valores típicos para a resistência e capacitância por unidade de comprimentos para nanofios de cobre e nanotubos de carbono. [45,46]

Tipo de interconexão	Resistência por unidade de comprimento (L)	Capacitância por unidade de comprimento (L)
Nanofio de cobre	355 $\Omega/\mu\text{m}$	280 aF/ μm
Nanotubo de carbono	6,45 k Ω , para $L \leq 1 \mu\text{m}$ 6,45 k $\Omega/\mu\text{m}$, para $L > 1 \mu\text{m}$	28 aF/ μm

3.4 - ESTRATÉGIAS PARA OBTENÇÃO DE RESULTADOS

3.4.1 - Modelagem dos parâmetros de interconexão para as simulações

Para a obtenção dos resultados finais de resistência e capacitância das interconexões, foram adotados, inicialmente, os valores arredondados descritos na Tabela 3.2 para um comprimento de interconexão de $L_{\text{con}} = 7,11 \mu\text{m}$. Este valor foi adotado como um comprimento máximo razoável para uma interconexão global dentro de um processador nanoeletrônico, como descrito em [45] e do observado no ITRS *roadmap* [44].

Tabela 3.2 – Valores típicos para a resistência e capacitância para nanofios de cobre e nanotubos de carbono com comprimento de 7,11 μm .

Tipo de interconexão	Resistência	Capacitância
Nanofio de cobre	2,5 k Ω	2 fF
Nanotubo de carbono	50 k Ω	200 aF

Com estes valores em mente, foi utilizado o maior valor dentre os dois para a resistência e para a capacitância, ou seja, $R_{con} = R_{CNT} = 50 \text{ k}\Omega$ e $C_{con} = C_{cu} = 2 \text{ fF}$, pois este procedimento gera a maior constante de retardo RC. Observou-se, então, se os circuitos simulados com estes parâmetros ainda mantinham suas funcionalidades e, em caso negativo, o motivo pelo qual esta funcionalidade era perdida.

Procedimentos simples, como a inserção de capacitâncias de carga nas entradas ou saídas dos subcircuitos que seriam interconectados, redução da temperatura de simulação ou mudança na ordem dos componentes da interconexão foram adotados, quando necessários, para manter a funcionalidade dos circuitos tentando minimizar o tempo de simulação e mantendo a generalidade do circuito inicialmente proposto.

3.4.2 - Extrapolação dos parâmetros de interconexão

A partir do momento em que a manutenção da funcionalidade dos circuitos simulados foi observada, extrapolou-se o valor das resistências e capacitâncias de cada interconexão modelada até a perda dessa funcionalidade. Este resultado foi levado em conta para o cálculo do comprimento máximo da interconexão que interligava cada um dos subcircuitos simulados.

Finalmente, tentou-se obter um resultado geral de todos os circuitos simulados válido para interconexões baseadas em diferentes materiais. Este resultado geral apresenta um comprimento máximo que pode ser utilizado na interligação de todos os subcircuitos simulados sem que estes percam suas funcionalidades, tanto para interconexões baseadas em cobre quanto em nanotubos de carbono. Este resultado também pode ser levado em conta para futuras simulações com diferentes subcircuitos ou materiais de interconexão com o objetivo de tentar manter este comprimento máximo obtido sem perder a funcionalidade dos circuitos.

Além disto, os bons procedimentos para manter a funcionalidade dos circuitos tentando minimizar o tempo de simulação e mantendo a generalidade do circuito inicialmente proposto foram sintetizados e estão apresentados no capítulo sobre discussão dos resultados.

4 - DESENVOLVIMENTO E RESULTADOS

4.1 - INTRODUÇÃO

Neste capítulo, serão apresentados os resultados e procedimentos adotados para 3 diferentes tipos de circuitos:

1. Rede *Winner-Take-All* (WTA) com dois neurônios, composta por dois sub-circuitos idênticos (neurônios) interligados;
2. Rede *Winner-Take-All* (WTA) com quatro neurônios, composta por quatro sub-circuitos idênticos (neurônios) interligados; e
3. Inversor com Somador, composto por dois subcircuitos diferentes (inversor e somador) interligados.

O funcionamento destes circuitos já foi descrito no capítulo de fundamentação teórica.

Além disto, como dito no capítulo sobre a metodologia empregada neste trabalho, o primeiro passo para cada um destes circuitos foi sua simulação com sucesso, admitindo-se interconexões ideais entre os seus sub-circuitos componentes, para confirmar seu funcionamento de acordo com o esperado em [22 e 24] para as redes WTA e [23] para o circuito inversor com somador. Esta confirmação inicial poderá ser observada dentro dos resultados alcançados para cada circuito.

Após esta confirmação, os resultados novos com interconexões não-ideais serão apresentados e seu funcionamento sob a influência dessas interconexões será observado e comparado.

4.2 - REDE SET-WTA COM DOIS NEURÔNIOS

Neste circuito, por ter sido o primeiro a ser testado, foram utilizados os dois tipos de simulação (SIMON e MatLab), como explicado no capítulo sobre metodologia empregada.

4.2.1 - Procedimentos iniciais

Para a comparação inicial de resultados, utilizou-se um circuito de uma rede construída com transistores mono-elétron (SET-WTA) como o da Figura 2.11. A simulação deste circuito foi realizada com as interconexões entre os neurônios consideradas ideais. Os resultados, tanto para a simulação com o SIMON, quanto com o MatLab se encontram na Figura 4.1. Os parâmetros usados nas duas simulações são idênticos e podem ser observados na Tabela 4.1 e também na rotina do MatLab do Apêndice B.

Tabela 4.1 – Valores dos parâmetros da rede SET-WTA com dois neurônios.

Parâmetro	Valor
C_{gate1} e C_{gate2}	1 aF
$C_{junção1}$ e $C_{junção2}$	0,1 aF
$R_{junção1}$ e $R_{junção2}$	1 M Ω
C_{L1} e C_{L2}	1 fF
R_{L1} e R_{L2}	10 M Ω
V_{bias1} e V_{bias2}	-50 mV
I_1	4 nA
I_2	2 nA
Temperatura	3 K
Tipo de Simulação	Transiente

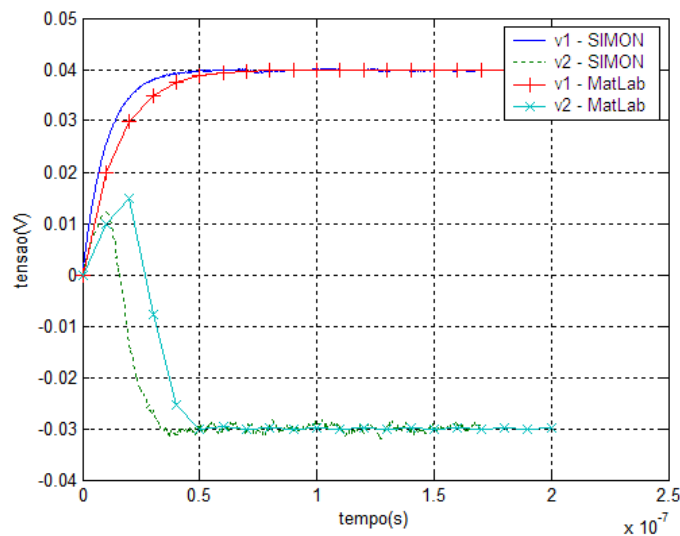


Figura 4.1 – Saídas v_1 e v_2 do SET-WTA com dois neurônios e interconexões ideais para simulações no SIMON e no MatLab.

Pelos resultados apresentados na Figura 4.1, é possível chegar às seguintes conclusões:

1. Ambos métodos de simulação (SIMON e MatLab) alcançaram o comportamento esperado para a rede WTA com dois neurônios, conseguindo decidir qual foi o neurônio vencedor;
2. Ambos métodos de simulação chegaram aos mesmos níveis finais de tensão na saída de cada neurônio; e
3. A fase de tomada de decisão da rede WTA foi menor quando simulada pela ferramenta SIMON, ou seja, o tempo de transição para o valor final das tensões na saída de cada neurônio foi menor na simulação do SIMON.

Neste momento, vale ser comentado que a temperatura de 3 K foi escolhida para manter a compatibilidade com as simulações observadas anteriormente na literatura [22 e 24] e por ser uma temperatura baixa, que não comprometeria os resultados das primeiras simulações.

4.2.2 - Simulações com interconexão não-ideal

Para a simulação inicial com um modelo de interconexão não-ideal, foi usado o circuito da Figura 4.2. O local escolhido para a utilização do modelo de interconexão foi na ligação entre os dois neurônios por esta poder vir a ser a interconexão com maior comprimento, e pelo fato de já existirem as resistências de tunelamento $R_{\text{junção}}$ dos SETs dos neurônios e as resistências de carga R_L dentro dos próprios neurônios.

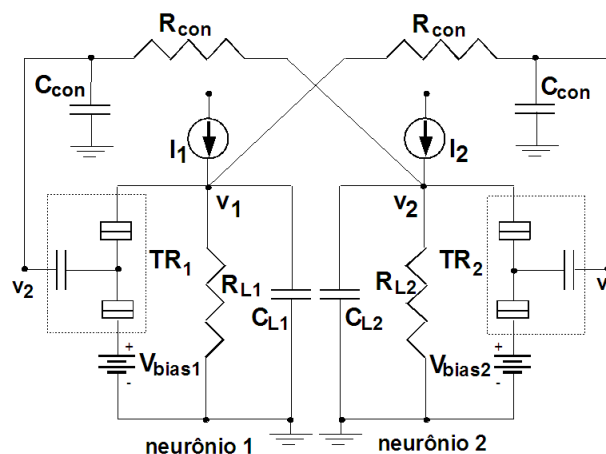


Figura 4.2 – Rede SET-WTA de 2 neurônios com inibição lateral e interconexão não-ideal.

Com este modelo, foram realizadas diversas simulações, tanto no MatLab, quanto no

SIMON. Na Figura 4.3, a seguir, são apresentados alguns resultados para os valores $R_{con} = 50 \text{ k}\Omega$ e $C_{con} = 2 \text{ fF}$, que foram os valores predeterminados para os parâmetros das primeiras simulações com interconexões não-ideais, de acordo com o item 3.4.1 da metodologia deste trabalho. Todos os outros parâmetros da simulação com interconexão ideal forma mantidos.

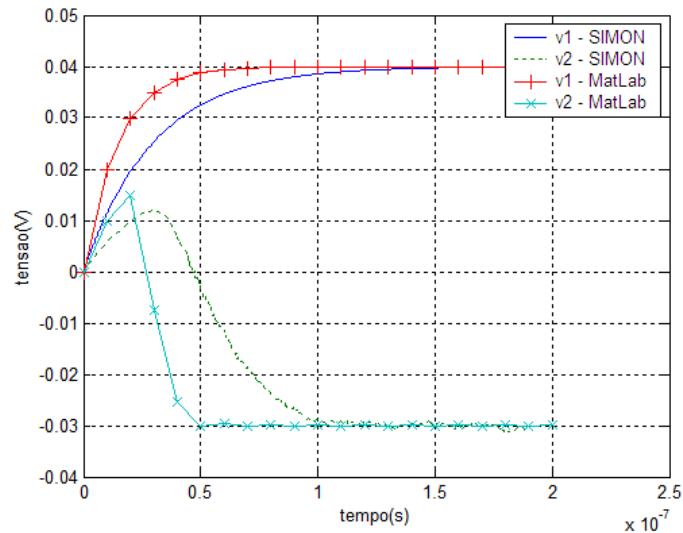


Figura 4.3 – Saídas v_1 e v_2 do SET-WTA com dois neurônios e interconexões não-ideais para simulações no SIMON e no MatLab.

Pelos resultados apresentados na Figura 4.3, é possível chegar às seguintes conclusões:

1. Ambos métodos de simulação (SIMON e MatLab) alcançaram o comportamento esperado para a rede WTA com dois neurônios, conseguindo decidir qual foi o neurônio vencedor;
2. Ambos métodos de simulação chegaram aos mesmos níveis finais de tensão na saída de cada neurônio, que foram os mesmo valores obtidos para a simulação com interconexão ideal;
3. A fase de tomada de decisão da rede WTA foi menor quando simulada pela ferramenta MatLab, ou seja, o tempo de transição para o valor final das tensões na saída de cada neurônio foi menor na simulação do MatLab; e
4. Na simulação do SIMON, enquanto que a fase de tomada de decisão da rede levou em torno de 50 ns (Figura 4.1), com a interconexão ideal sendo usada, esta mesma fase levou cerca de 100 ns, com o modelo de interconexão não-ideal sendo usado. Isto dá um retardo de cerca de 50 ns, para o modelo de interconexão e os valores dos parâmetros utilizados.

4.2.3 - Diferenças entre os métodos de simulação

Como observado anteriormente nas Figuras 4.1 e 4.3, algumas diferenças puderam ser notadas entre os dois métodos de simulação realizados: com a ferramenta SIMON e com a ferramenta MatLab. Esta seção destina-se à análise e discussão acerca destas diferenças.

Inicialmente, serão apresentadas as diferenças entre os dois métodos de simulação tanto para a simulação com interconexão ideal quanto para não-ideal. A Figura 4.4, a seguir, apresenta o erro (diferença entre os valores simulados com o SIMON e com o MatLab) para as tensões na saída de cada neurônio simuladas considerando interconexões ideais.

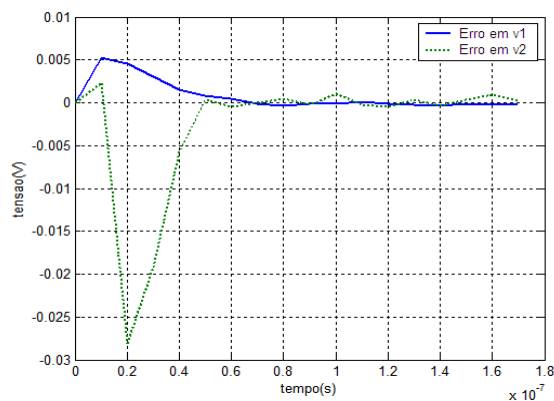


Figure 4.4 – Diferença (valores simulados no SIMON menos valores simulados no MatLab) entre as saídas v_1 e v_2 do SET-WTA com dois neurônios e interconexões ideais.

Já a Figura 4.5 apresenta o erro (diferença entre os valores simulados com o SIMON e com o MatLab) para as tensões na saída de cada neurônio simuladas considerando interconexões não-ideais.

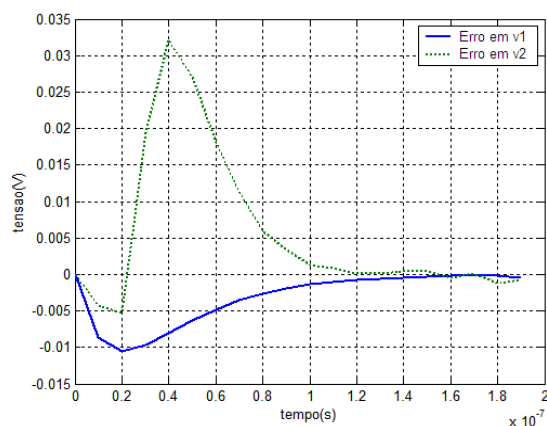


Figure 4.5 – Diferença (valores simulados no SIMON menos valores simulados no MatLab) entre as saídas v_1 e v_2 do SET-WTA com dois neurônios e interconexões não-ideais.

Pelos valores observados nestas duas figuras, é possível fazer as seguintes observações:

1. Como citado anteriormente, para as simulações com interconexão ideal, a fase de tomada de decisão da rede WTA simulada pelo MatLab encontra-se atrasada em relação à simulada pelo SIMON, o que pode ser observado pelos valores positivos para o erro de v_1 e negativos para o erro de v_2 , durante esta fase de transição;
2. Também como citado anteriormente, para as simulações com interconexão não-ideal, a fase de tomada de decisão da rede WTA simulada pelo MatLab encontra-se adiantada em relação à simulada pelo SIMON, o que pode ser observado pelos valores negativos para o erro de v_1 e positivos para o erro de v_2 , durante esta fase de transição; e
3. O erro entre os valores de saída de tensão se aproxima de zero ao final da simulação, o que indica que ambas se aproximam de valores similares de tensão.

Neste ponto, é interessante notar que as simulações com a ferramenta SIMON apresentaram um retardo na definição de seus valores finais de saída dos neurônios, quando a interconexão não-ideal foi usada, o que é característico da influência da constante de retardo RC, causada pelos parâmetros de interconexão modelados. A Figura 4.6 mostra a diferença entre os valores simulados com o MatLab para interconexão não-ideal e ideal.

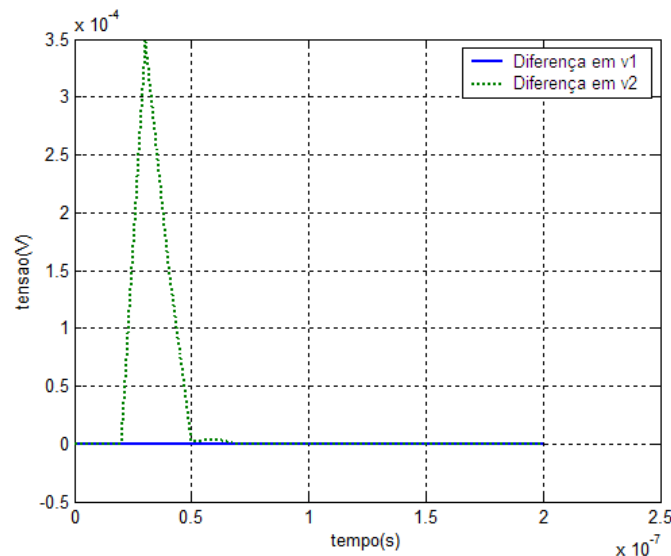


Figura 4.6 – Diferença entre simulações com interconexões ideal e não-ideal (valores simulados no MatLab) para as saídas v_1 e v_2 do SET-WTA com dois neurônios.

Como pode ser observado, a diferença entre os valores simulados é nula para v_1 e quase nula (valor máximo de 0,35 mV para uma tensão em torno de -7,4 mV) para v_2 . Sendo

assim, é possível dizer que a simulação com o MatLab não avaliou bem a influência dos parâmetros da interconexão sobre o funcionamento do circuito. Como dito anteriormente, isto deveu-se à simplicidade do algoritmo quase-analítico adotado na ferramenta MatLab [35], que implementava somente o comportamento de um transistor mono-elétron, não levando em conta a influência das outras junções-túnel existentes e da própria interconexão sobre as junções-túnel dos subcircuitos adjacentes.

Sendo assim, a partir deste momento, passou-se a adotar somente a simulação na ferramenta SIMON.

4.3 - REDE SET-WTA COM QUATRO NEURÔNIOS

Neste momento, resolveu-se testar uma rede WTA de quatro neurônios com inibição lateral para verificar a generalidade dos resultados obtidos anteriormente para um circuito um pouco mais complexo, mas que deve funcionar da mesma forma. Este circuito também foi escolhido para que fosse possível testar a influência de mais de uma interconexão não-ideal simultaneamente.

4.3.1 - Procedimentos iniciais

Para a comparação inicial de resultados, utilizou-se um circuito de uma rede construída com transistores mono-elétron (SET-WTA) como o da Figura 4.7 [24].

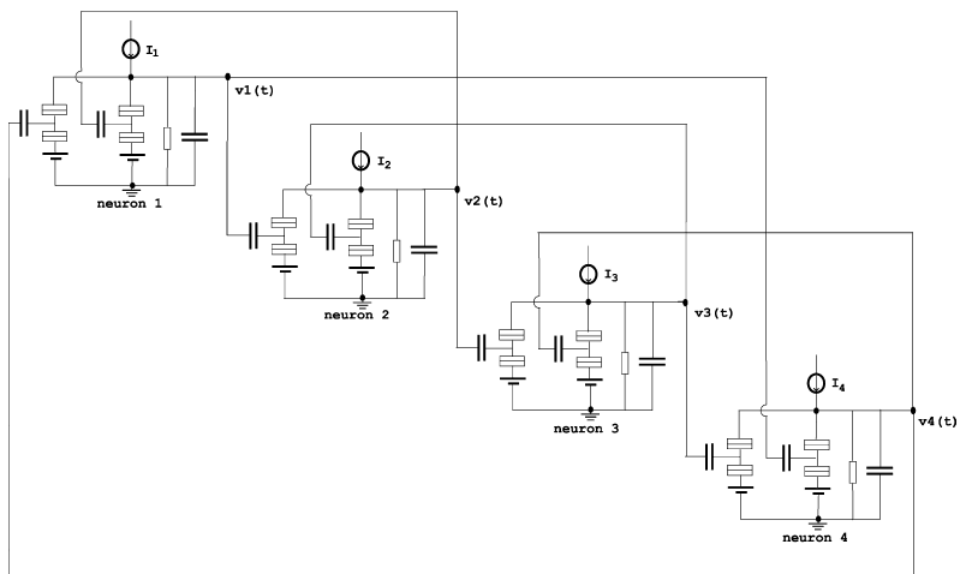


Figura 4.7 – Rede SET-WTA de 4 neurônios com inibição lateral e interconexão ideal.

A simulação deste circuito foi realizada com as interconexões entre os neurônios consideradas ideais. Os parâmetros usados na simulação podem ser observados na Tabela 4.2, onde os valores são válidos para todos os neurônios.

Tabela 4.2 – Valores dos parâmetros da rede SET-WTA com quatro neurônios.

Parâmetro	Valor
C_{gate}	1 aF
$C_{junção}$	0,1 aF
$R_{junção}$	1 M Ω
C_L	1 fF
R_L	10 M Ω
V_{bias}	-50 mV
I_1	5,5 nA
I_2	10,5 nA
I_3	5,5 nA
I_4	-0,5 nA
Temperatura	77 K
Ordem de co-tunelamento	2
Distribuição de cargas de desvio	5%
Tipo de Simulação	Transiente

Pelo observado na tabela, alguns parâmetros a mais foram utilizados nas simulações do WTA com 4 neurônios para demonstrar o seu funcionamento em condições mais restritas, assim como em [31]. Estas condições são:

1. Temperatura de 77 K (nitrogênio líquido), que é mais alta do que a temperatura das simulações anteriores;
2. Co-tunelamento de ordem 2, ou seja, podem acontecer até dois tunelamentos simultaneamente; e
3. Distribuição de cargas de desvio, aleatoriamente, com valores entre $-0,05e$ e $0,05e$.

Os resultados da simulação no SIMON encontram-se na Figura 4.8.

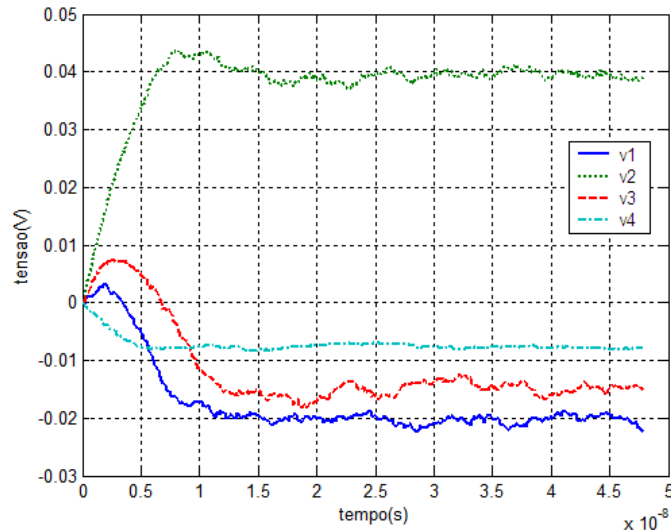


Figura 4.8 – Saídas v_1 , v_2 , v_3 e v_4 do SET-WTA com quatro neurônios e interconexões ideais.

As seguintes observações devem ser feitas a respeito do observado nesta figura:

1. A simulação alcançou o comportamento esperado para a rede WTA com quatro neurônios, conseguindo decidir qual foi o neurônio vencedor;
2. O nível de tensão final da saída do neurônio 4 acabou sendo maior que o dos neurônios 1 e 3, apesar daquele possuir uma corrente de entrada menor, porque estes sofreram uma maior inibição lateral do neurônio 2;
3. Os níveis de tensão finais das saídas dos neurônios 1 e 3 foram ligeiramente diferentes para esta simulação, provavelmente, por conta da influência das cargas de desvio ou co-tunelamento, pois, em outras simulações em que estes efeitos não foram levados em conta, estes níveis de tensão ficaram bem próximos.

4.3.2 - Simulações com interconexões não-ideais

Para esta simulação foi adotado o mesmo modelo de interconexão não-ideal já descrito anteriormente. Novamente, o local escolhido para a utilização do modelo de interconexão foi na ligação entre cada um dos quatro neurônios, com os valores dos parâmetros das interconexões sendo $R_{con} = 50 \text{ k}\Omega$ e $C_{con} = 2 \text{ fF}$. Todos os outros parâmetros da simulação com interconexão ideal foram mantidos. Na Figura 4.9 são apresentados os resultados desta nova simulação.

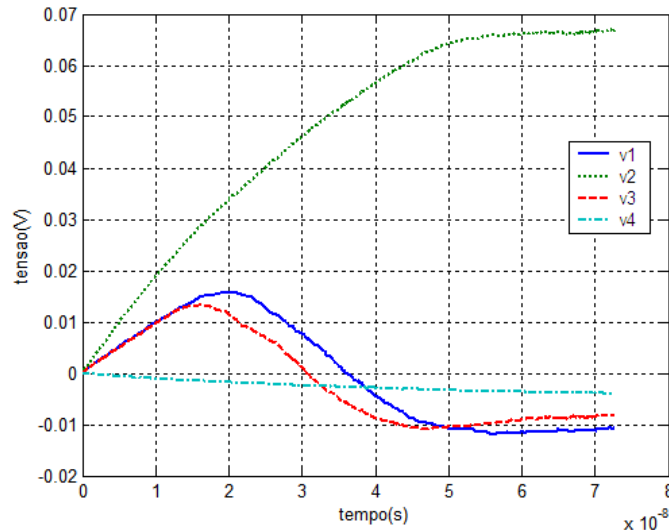


Figura 4.9 – Saídas v_1 , v_2 , v_3 e v_4 do SET-WTA com quatro neurônios e interconexões não-ideais.

Pelos resultados apresentados na Figura 4.10, é possível chegar às seguintes conclusões:

1. A simulação alcançou o comportamento esperado para a rede WTA com quatro neurônios, conseguindo decidir qual foi o neurônio vencedor;
2. Os níveis de tensão finais na saída de cada neurônio foram diferentes em relação aos obtidos com interconexões ideais;
3. Enquanto que a fase de tomada de decisão da rede levou em torno de 15 ns (Figura 4.8), com a interconexão ideal sendo usada, esta mesma fase levou cerca de 60 ns, com o modelo de interconexão não-ideal sendo usado. Isto dá um retardo de cerca de 45 ns, para o modelo de interconexão e os valores dos parâmetros utilizados.

4.3.3 - Discussão sobre retardo e velocidade

Dos resultados obtidos para a rede WTA com interconexões ideais, temos que a menor diferença entre o nível de tensão final da saída do neurônio vencedor para os demais neurônios foi de, aproximadamente, $40 \text{ mV} - (-8 \text{ mV}) = 48 \text{ mV}$, que foi a diferença entre os neurônios 2 e 4, como pode ser visto na Figura 4.8. Admitamos que um módulo ligado na saída da rede WTA com quatro neurônios consiga identificar o neurônio vencedor quando 80% dessa diferença seja alcançado para a diferença entre todos os neurônios, ou seja, quando uma diferença de $38,4 \text{ mV}$ entre as tensões dos neurônios seja alcançada.

Para interconexões ideais, essa diferença é alcançada em $t = 6,4$ ns, o que resulta em uma frequência de operação de $f = 1/t = 156,25$ Mhz.

Já para interconexões não-ideais, essa diferença é alcançada em $t = 29,8$ ns. Isto resulta em um retardo de $t_{con} = 23,4$ ns observado através das simulações quando o modelo proposto de interconexão não-ideal é usado. Esse retardo limita a utilização da rede WTA segundo o modelo proposto a uma frequência de $f = 1/t = 33,56$ MHz. Neste momento, vale lembrar que estes valores de retardo e frequência são para a interconexão com comprimento de $L_{con} = 7,11$ μ m, que foi o valor proposto para o comprimento máximo de uma interconexão global e que não devem ser os valores encontrados em dispositivos mais realistas.

Para dispositivos mais realistas, um comprimento típico de interconexão seria de $L_{con} = 100$ nm, que resulta nos seguintes parâmetros, segundo a Tabela 3.1 e adotando-se os mesmos procedimentos anteriores: $R_{con} = 6,45$ k Ω e $C_{con} = 28$ aF. A rede WTA foi simulada novamente para estes parâmetros resultando em $t = 6,485$ ns, $t_{con} = 6,485$ ns - $6,4$ ns = 85 ps, e $f = 1/t = 154,2$ MHz.

Os resultados da análise dos efeitos das interconexões não-ideais sobre o funcionamento da rede WTA com quatro neurônios se encontram resumidos na Tabela 4.3, a seguir.

Tabela 4.3 – Dados da análise dos efeitos causados pelo modelo de interconexão não-ideal sobre a rede WTA com quatro neurônios.

L_{con}	R_{con}	C_{con}	RC	t	t_{con}	f
Ideal	0 Ω	0 F	0 s	6,4 ns	0 s	156,25 MHz
100 nm	6,45 k Ω	28 aF	180,6 fs	6,485 ns	85 ps	154,2 MHz
7,11 μ m	50 k Ω	2 fF	0,1 ns	29,8 ns	23,4 ns	33,56 MHz

Estes resultados para as frequências de operação f são, aparentemente, baixos devido ao tempo de decisão inerente da rede WTA, como pode ser observado pelo valor para interconexões ideais.

4.4 - CIRCUITO INVERSOR COM SOMADOR

As redes WTA são circuitos tipicamente analógicos que, como na maioria das redes neurais, após uma fase de decisão, mantém uma saída constante. As análises realizadas com essas redes foram análises estáticas onde, após a apresentação de uma corrente de entrada fixa, verificou-se o comportamento da tensão de saída de cada neurônio.

Neste momento, resolveu-se testar um circuito tipicamente digital, como o composto por uma porta inversora e uma porta somadora decimal, onde o comportamento dinâmico destes subcircuitos interligados pôde ser analisado através de uma entrada binária variando com o tempo. Como dito anteriormente, o funcionamento destes subcircuitos isolados pode ser observado no capítulo de fundamentação teórica.

4.4.1 - Procedimentos iniciais

Para comparação inicial de resultados com os já observados na literatura [23], utilizou-se um circuito com uma porta inversora ligada a uma porta somadora de duas entradas como o da Figura 4.10 [23].

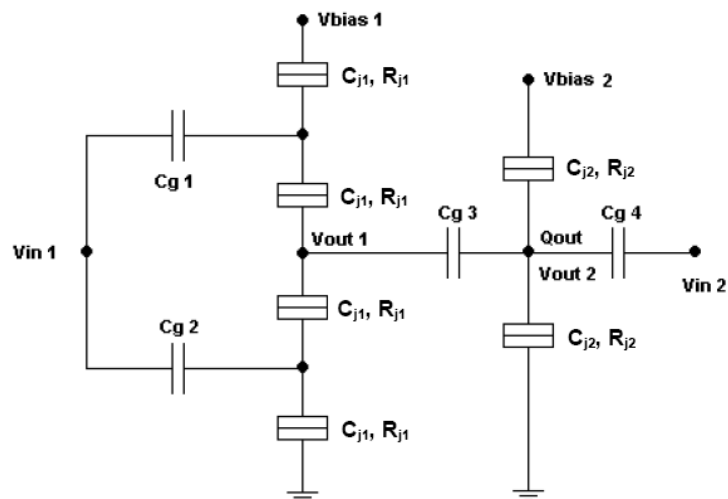


Figura 4.10 – Circuito inversor interconectado a somador de duas entradas com interconexão ideal.

A simulação deste circuito foi realizada primeiramente, como para os outros circuitos, com as interconexões entre as portas lógicas consideradas ideais. Os parâmetros usados na simulação podem ser observados na Tabela 4.4.

Tabela 4.4 – Valores dos parâmetros do circuito inversor com somador.

Parâmetro	Valor
C_{g1} e C_{g2}	2 aF
C_{g3} e C_{g4}	5,2 aF
C_{j1} e C_{j2}	0,05 aF
R_{j1} e R_{j2}	5 M Ω
V_{bias1}	40 mV
V_{bias2}	10 μ V
Temperatura	77 K
Ordem de co-tunelamento	2
Distribuição de cargas de desvio	5%
Tipo de Simulação	Estática

Além disso, a entrada do circuito inversor V_{in1} e a entrada independente do somador V_{in2} estão indicadas na Figura 4.11.

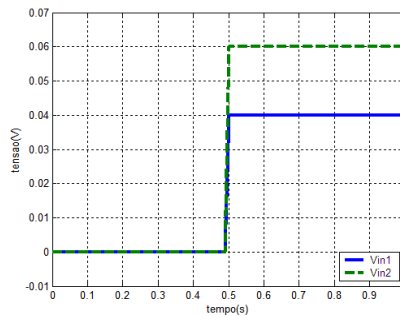


Figura 4.11 – Entradas V_{in1} e V_{in2} do inversor e do somador, respectivamente.

Os resultados da simulação no SIMON encontram-se na Figura 4.12. À esquerda é possível observar a tensão de saída do inversor, que serve como uma das entradas da porta somadora, e à direita é possível observar a carga de saída do somador.

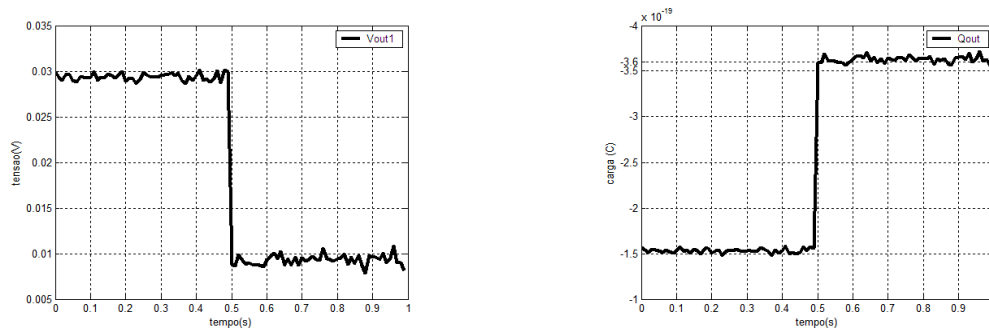


Figura 4.12 – Saídas V_{out1} e Q_{out} do inversor e do somador, respectivamente, com interconexão ideal.

As seguintes observações devem ser feitas a respeito do observado nesta figura:

1. A simulação alcançou o comportamento esperado na saída do inversor, pois o nível lógico 0 da entrada ($V_{in1} = 0$ V) foi levado a 30 mV (nível lógico 1 da saída) e o nível lógico 1 ($V_{in1} = 40$ mV) foi levado a cerca de 10 mV (nível lógico 0 da saída);
2. Houve uma variação de cerca de 10 mV entre os níveis lógicos da entrada e da saída do inversor;
3. A simulação também alcançou o comportamento esperado na saída do somador.

Lembrar que, para o somador decimal, vale a Equação 4.1, a seguir, que é a utilização da equação 2.15 especificamente para o circuito da Figura 4.10.

$$q = -C_g \times (V_{out1} + V_{in2}) \quad (4.1)$$

Utilizando o produto da direita da Equação 4.1, com C_g sendo 5,2 aF e as tensões V_{out1} e V_{in2} dadas nas Figuras 4.11 e 4.12, temos os valores apresentados na Figura 4.13, o que confirma a obtenção dos valores esperados para a carga de saída do somador.

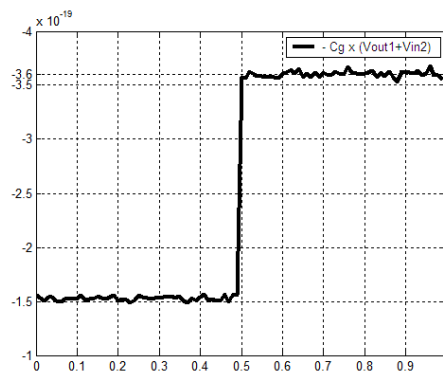


Figura 4.13 – Cálculo de $-C_g \times (V_{out1} + V_{in2})$.

4.4.2 - Simulações com interconexões não-ideais

Para estas simulações foi adotado o mesmo modelo de interconexão não-ideal já descrito anteriormente. O local escolhido para a utilização do modelo de interconexão foi na ligação entre o bloco inversor e o bloco somador, novamente, por esta poder vir a ser a interconexão com maior comprimento. Mais uma vez, os valores dos parâmetros das interconexões usados foram $R_{con} = 50$ k Ω e $C_{con} = 2$ fF. Todos os outros parâmetros da simulação com interconexão ideal forma mantidos.

Inicialmente, tentou-se simular o circuito como descrito na Figura 4.14, a seguir. Esta figura é de uma tela típica de um circuito na ferramenta SIMON.

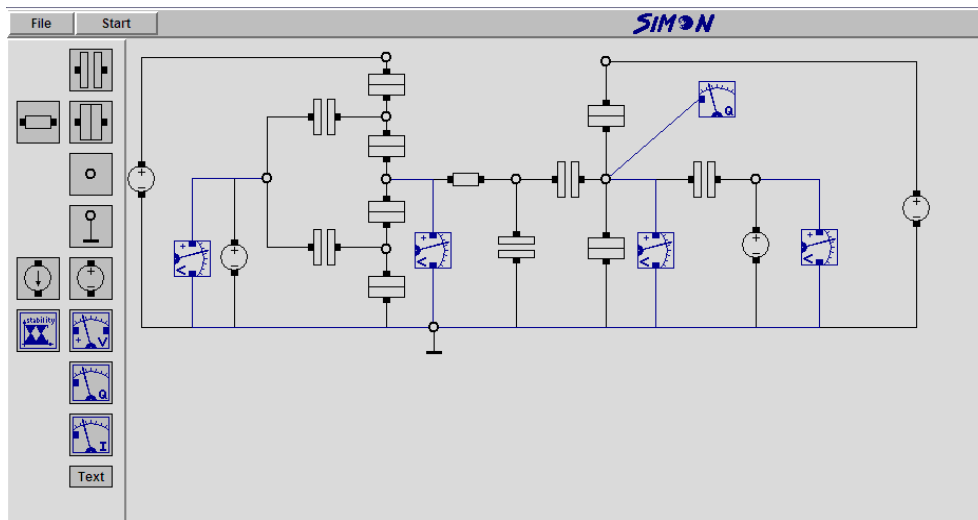


Figura 4.14 – Parte da tela do editor SIMON com o circuito inversor interconectado ao somador de duas entradas com interconexão não-ideal.

Esta simulação, porém, não obteve bons resultados, pois o inversor parou de funcionar. Provavelmente, isto aconteceu por causa da diferença entre as regras global e local de simulação causada pela mudança do ambiente de impedância entre o estágio inversor e o somador.

De acordo com [51], usando-se um capacitor de carga na saída da porta inversora com capacitância cerca de 10 vezes maior que uma capacitância típica da junção-túnel, os sub-circuitos interconectados passam a funcionar como se estivessem isolados e sua interação pode ser calculada baseada nas Leis de Kirchhoff. Baseado nesta idéia, pensou-se na necessidade de adequação do ambiente de impedância entre os sub-circuitos para que eles pudessem continuar funcionando como num ambiente com interconexões ideais. Isto foi conseguido simplesmente com a alteração da ordem dos componentes que simulavam os parâmetros concentrados da interconexão não-ideal, como na Figura 4.15.

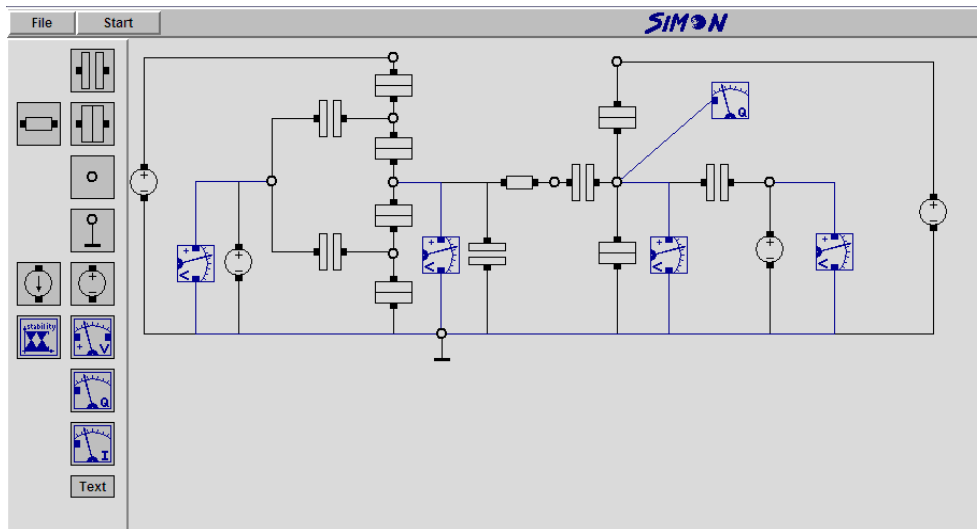


Figura 4.15 – Parte da tela do editor SIMON com o circuito inversor interconectado ao somador de duas entradas com interconexão não-ideal (ordem dos componentes invertida).

Com esta mudança simples, ambos sub-circuitos voltaram a funcionar e a simulação foi realizada bem mais rapidamente. Os resultados tanto da tensão de saída do inversor quanto da carga de saída do somador dessa nova simulação estão na Figura 4.16.

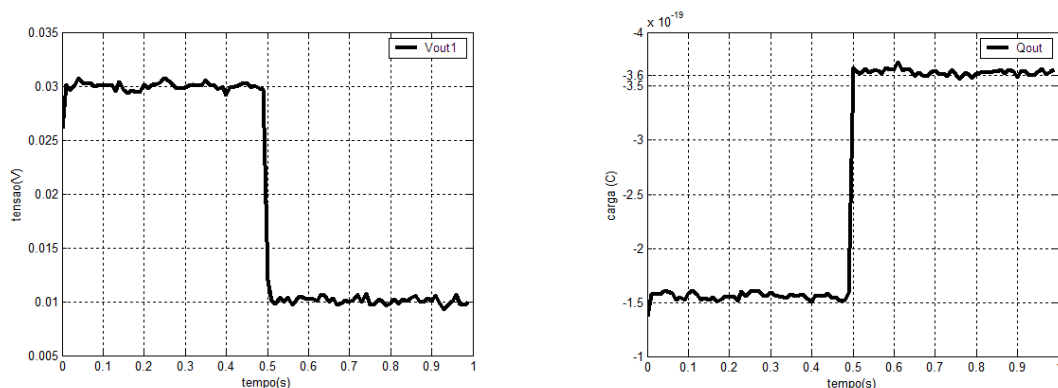


Figura 4.16 – Saídas V_{out1} e Q_{out} do inversor e do somador, respectivamente, com interconexão não-ideal.

As seguintes observações devem ser feitas a respeito do observado nesta figura:

1. A simulação alcançou o comportamento esperado na saída do inversor, com os mesmos níveis de tensão da simulação com interconexão ideal;
2. A simulação também alcançou o comportamento esperado na saída do somador, com o mesmo valor de carga da simulação com interconexão ideal; e
3. Houve um leve retardo no início de cada transição no sinal da saída. O valor deste retardo não pôde ser computado pela simulação estática.

4.4.3 - Discussão sobre retardo e velocidade

A partir do momento em que o circuito do inversor com o somador passou a funcionar com a interconexão não-ideal, iniciou-se a fase de testes dos limites dos parâmetros concentrados desta interconexão. Para os testes a seguir, deve ser lembrado que o sinal de entrada escolhido para as simulações possui dois bits com uma duração total de 1 s, ou seja, uma velocidade bem baixa de 2 bps.

4.4.3.1 - Tipos de simulação do SIMON

Como pode ser observado nos dados das Tabelas 4.1, 4.2 e 4.4, as simulações para o circuito inversor com somador foram do tipo estática, ao contrário das simulações realizadas para as redes WTA. Isto foi feito para manter a compatibilidade com os dados obtidos em outros trabalhos [23].

Este tipo de simulação, porém, apresenta um aspecto mais qualitativo do comportamento do circuito, na medida em que a saída do circuito é proporcional à entrada, ou seja, se for colocada na entrada do circuito um sinal de 2 Mbp, ao invés de 2 bps, a saída será a mesma proporcionalmente à duração do sinal. Deste modo, ao se alterar os parâmetros da interconexão, é possível observar uma alteração do sinal da saída em relação ao simulado para interconexão ideal e fazer uma análise desta alteração proporcionalmente à duração do sinal. Não é possível, porém, se observar qualquer tipo de duração do retardo em segundos.

A simulação transiente, porém, para circuitos mais complexos como os propostos neste trabalho, demanda muito tempo de simulação e pode gerar arquivos de dezenas a centenas de megabytes para cada um dos medidores de tensão, corrente ou carga dentro do circuito. Por este motivo, muitas vezes, a simulação estacionária é preferível.

4.4.3.2 - Capacitância da interconexão

Primeiramente, foi testada a capacitância da interconexão não-ideal. Para capacitâncias muito pequenas, da ordem de 10^{-30} F, o inversor deixa de funcionar, mesmo com resistências também pequenas. O limite superior desta capacitância também foi testado e o

circuito permaneceu funcionando mesmo para capacitâncias muito altas, da ordem de $C_{con} = 10$ pF. Com esta ordem de grandeza, porém, o retardo proporcionado pela capacitância da interconexão limita a velocidade do sinal dentro do circuito.

Para que os níveis de tensão e carga que identificam a saída binária do circuito interconectado possam ser mantidos por tempo suficiente para que seu nível lógico ainda possa ser identificado, o limite superior para a capacitância da interconexão foi de $C_{con} = 0,1$ pF. Com este valor para a capacitância de interconexão e mantendo o valor da resistência da interconexão $R_{con} = 50$ k Ω , foram obtidos os resultados mostrados na Figura 4.17, onde pode ser identificada claramente a influência do retardo RC causado pela interconexão não-ideal, tanto na saída do inversor quanto do somador.

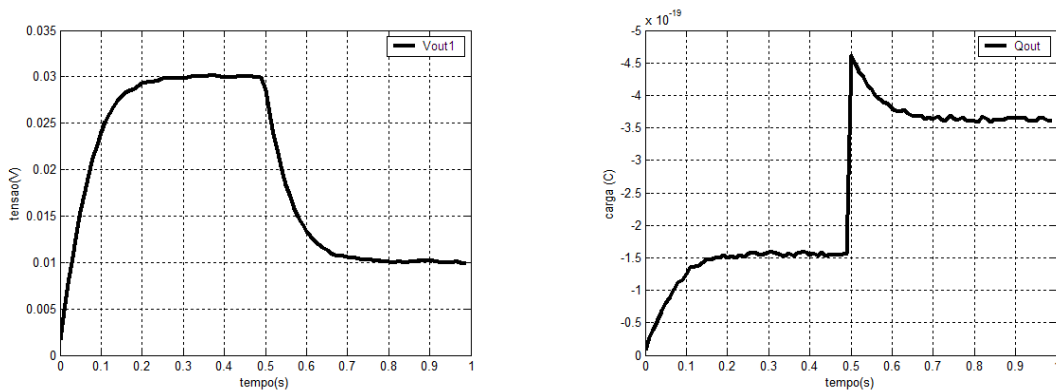


Figura 4.17 – Saídas V_{out1} e Q_{out} do inversor e do somador, respectivamente, com interconexão não-ideal no limite de C_{con} .

4.4.3.3 - Resistência da interconexão

Neste momento, passou-se a testar o limite da resistência da interconexão não-ideal voltando o valor de capacitância da interconexão ao valor inicial, ou seja, $C_{con} = 2$ fF. O circuito permaneceu funcionando mesmo para resistências muito altas, da ordem de $R_{con} = 1$ T Ω . Para este valor, apesar do inversor continuar funcionando normalmente sem demonstrar nenhum tipo de retardo RC, o retardo na saída do módulo somador passou a ser muito alto para o sinal de entrada proposto, como pode ser observado na Figura 4.18.

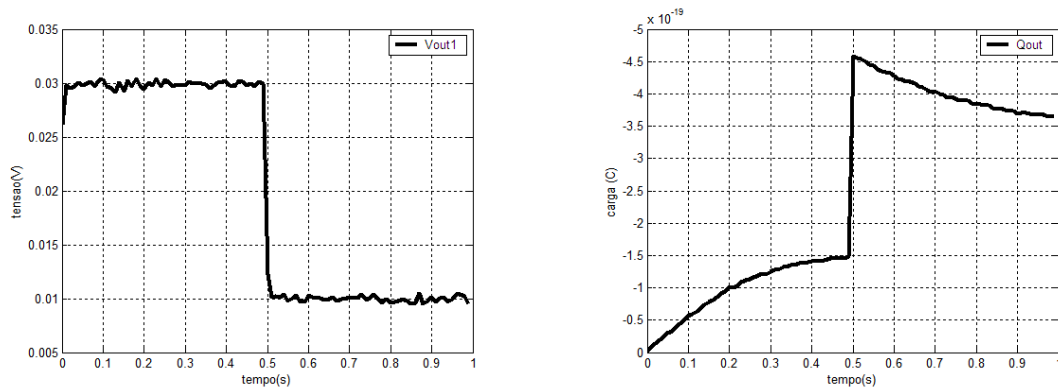


Figura 4.18 – Saídas V_{out1} e Q_{out} do inversor e do somador, respectivamente, com interconexão não-ideal no limite de R_{con} .

4.4.3.4 - Comprimento da interconexão

Baseado na Tabela 3.1, temos que o limite inferior do comprimento da interconexão não-ideal do nosso modelo para que uma determinada constante RC não seja ultrapassada é dada pela Equação 4.2.

$$L_{con}(\mu m) = 0,744 \times 10^6 \times \sqrt{Constante RC} \quad (4.2)$$

4.4.3.5 - Interconexões mais realistas e frequência de operação

Neste momento, deve ser lembrado que os resultados obtidos anteriormente foram para condições limites da capacitância e da resistência da interconexão com valores que não são muito realistas, como:

1. Comprimento de interconexão (L_{con}) muito grande, como 52,61 μm e 332,73 μm , para dispositivos da ordem de dezenas de nanômetros;
2. Resistência de interconexão (R_{con}) muito grande, como 1 T Ω , onde esta deixa de se comportar como uma interconexão;
3. Frequências de operação (f_{con}) muito baixas para dispositivos nanométricos.

Sendo assim, resolveu-se fazer novas simulações com um sinal de entrada de frequência mais alta e do tipo transiente para o circuito inversor com somador utilizando o modelo proposto de interconexão não-ideal. Isto foi feito para os comprimentos de 100 nm e 7,11

μm , que foi o valor proposto para o comprimento máximo de uma interconexão global e que teve que ser simulado novamente para a nova frequência do sinal de entrada.

Dos resultados obtidos para o circuito inversor com somador com interconexões ideais, temos que diferença em módulo entre o níveis lógicos 1 e 0 na saída do circuito foi de, aproximadamente, $3,6 \times 10^{-19} \text{ C} - 1,5 \times 10^{-19} \text{ C} = 2,1 \times 10^{-19} \text{ C}$, como pode ser visto na Figura 4.13.

Admitamos que um módulo ligado na saída do circuito interconectado consiga identificar o nível lógico quando este estiver dentro de 20% dessa diferença acima ou abaixo do nível lógico original. Ou seja, para o nível lógico 0, quando uma carga de $-1,5 \times 10^{-19} \text{ C} + 0,42 \times 10^{-19} \text{ C} = -1,08 \times 10^{-19} \text{ C}$ for alcançada, ou $V_{\text{out1}} = 1,08 \times 10^{-19} / C_g = 20,77 \text{ mV}$.

Para o comprimento de $L_{\text{con}} = 100 \text{ nm}$, foi observado que o tempo de transição do sinal de saída foi de $t = t_{\text{passo}} = 0,64 \text{ ns}$. Como não foi possível obter um valor desse tempo para a interconexão ideal através de simulação do tipo transiente, esse valor foi tomado como referência para os outros valores. Sendo assim: $t_{\text{con}} = 0 \text{ s}$ e $f = f_{\text{original}} = 1/t = 1,56 \text{ GHz}$. Já para o comprimento de $L_{\text{con}} = 7,11 \mu\text{m}$, foram obtidos os valores: $t = 50,32 \text{ ns}$, $t_{\text{con}} = 49,68 \text{ ns}$ e $f = 1/t = 19,87 \text{ MHz}$.

Dos valores acima, também é possível calcular o restante dos valores para os limites observados nas simulações estáticas para a capacitância e para a resistência da interconexão não-ideal.

Sendo assim, baseado na equação 4.2 e nas observações ao longo desta seção, os resultados da análise dos efeitos das interconexões não-ideais sobre o funcionamento do circuito inversor com somador se encontram resumidos na Tabela 4.5, a seguir.

Tabela 4.5 – Dados da análise dos efeitos causados pelo modelo de interconexão não-ideal sobre o circuito inversor com somador.

L_{con}	R_{con}	C_{con}	RC	t	t_{con}	f
100 nm	6,45 k Ω	28 aF	180,6 fs	0,64 ns	0	1,56 GHz
7,11 μm	50 k Ω	2 fF	0,1 ns	50,32 ns	49,68 ns	19,87 MHz
52,61 μm	50 k Ω	0,1 pF	5 ns	-	-	-
332,73 μm	1 T Ω	2 fF	2 ms	-	-	-

Nestes resultados, já foi possível observar uma frequência de operação f acima de 1 GHz para uma interconexão de 100 nm. Isto, porém, está longe do previsto na Tabela 2.1 para os dispositivos nanoeletrônicos, que seria um *clock* de 1 THz.

Neste momento torna-se oportuno, então, lembrar que a velocidade de chaveamento dos dispositivos estudados não é o único fator que dita a velocidade do processador proposto. O número de interconexões existente na arquitetura proposta, baseada em redes neurais, aumenta muito a velocidade de processamento das informações dentro do núcleo nanoeletrônico.

4.5 - BONS PROCEDIMENTOS PARA SIMULAÇÕES

Como observado nas simulações anteriores, principalmente para o circuito inversor com somador, algumas boas práticas para a simulação dos circuito são as seguintes:

1. Inclusão de uma capacitância de carga (cerca de 10 vezes maior que a capacitância típica das junções-túnel usadas no circuito) na saída de cada módulo a ser interconectado. Este procedimento, apesar de poder aumentar o retardo do sinal dentro do circuito simulado, tende a diminuir o tempo de simulação e o tamanho dos arquivos simulados, como indicado em [51];
2. Teste dos circuitos inicialmente usando simulações do tipo estáticas para verificar o seu funcionamento;
3. Para obtenção de relações temporais dos circuitos, as simulações transientes são necessárias. Para isto, porém, é necessário um maior entendimento do circuito a ser simulado, como a frequência do sinal para a qual o circuito responderá como esperado; e
4. Após a verificação do perfeito funcionamento do circuito, as capacitâncias de carga inicialmente colocadas podem ser retiradas para verificar se o retardo do sinal dentro do circuito diminui sem afetar seu funcionamento.

5 - CONCLUSÕES E RECOMENDAÇÕES

Como dito no começo deste trabalho, o objetivo desta dissertação era determinar a influência das interconexões sobre o comportamento de diversos circuitos e estabelecer limites para suas funcionalidades, relacionados com as interconexões.

Para que este objetivo fosse alcançado, inicialmente foi feita uma comparação entre dois tipos de simulação: uma com a ferramenta SIMON e outra com a ferramenta MatLab. Isto foi feito porque comportamento de simuladores de circuitos monoelétron baseados em métodos estocásticos ao se alterar o ambiente de impedância não é intuitivo e torna-se de difícil predição. Verificou-se, então, que os resultados obtidos com a ferramenta SIMON foram os mais pertinentes.

Três tipos de circuitos que implementam subcircuitos interligados por conexões ideais – rede WTA com dois neurônios, rede WTA com quatro neurônios e circuito inversor com somador – foram simulados e seus desempenhos comparados com outros resultados encontrados na literatura. Todos os circuitos funcionaram como era esperado.

Um modelo de interconexão não-ideal baseado em parâmetros concentrados foi proposto. Este modelo levou em consideração o pior caso entre dois tipos de interconexão: nanofios de cobre e nanotubos de carbono. Este fato fez com que os resultados alcançados pudessem ser válidos para ambos tipos de interconexão.

Os mesmos três circuitos citados anteriormente foram simulados novamente utilizando o modelo de interconexão proposto e funcionaram conforme esperado apresentando retardos que puderam ser medidos, pelo menos, para dois comprimentos de interconexão diferentes: 100 nm e 7,11 μm . Este último valor foi um comprimento máximo proposto para uma interconexão global dentro do processador nanoeletrônico.

Os retardos medidos e as respectivas frequências de operação calculadas servem como previsão do comportamento dinâmico destes e de outros circuitos, para interconexões em escalas nanométricas. Apesar das frequências de operação medidas terem apresentado valores abaixo do esperado, estes valores são extremamente válidos para a compreensão da

dinâmica do funcionamento de subcircuitos interconectados. Os trabalhos futuros recomendados deverão apresentar resultados mais afinados com as expectativas.

Além disto, bons procedimentos a serem utilizados para simulação de circuitos nanoeletrônicos baseados em dispositivos mono-elétron com a ferramenta SIMON puderam ser observados.

Os seguintes trabalhos são recomendados para o futuro:

1. Simulação dos mesmos circuitos propostos para comprimentos de interconexão diferentes e verificação da compatibilidade dos retardos obtidos com os aqui observados;
2. Simulação dos mesmos circuitos propostos com modelo de interconexão mais complexo que avalie uma maior quantidade de efeitos dentro das interconexões, como a condutância entre duas vias, e que avalie uma maior quantidade de materiais além dos nanofios de cobre e nanotubos de carbono;
3. Simulação de novos circuitos com o novo modelo de interconexão proposto para extrapolação do comportamento de qualquer circuito; e
4. Criação de ferramenta aperfeiçoada de simulação que una a precisão de uma ferramenta baseada em métodos estocásticos à rapidez de uma ferramenta que faça cálculos de circuitos baseados nas Leis de Kirchhoff e que já faça previsão automática do comportamento dinâmico do circuito e inserção ou retirada de capacitâncias de carga quando houver uma maior necessidade de velocidade ou de precisão na simulação.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Ahmed, H. e Nakazato, K., “Single-electron devices”, *Microelectronics Engineering*, 32(1), 297–315, 1996.
- [2] Averin, D. e Likharev, K. Coulomb blockade of tunneling, and coherent oscillations in small tunnel junctions. *Journal of Low Temperature Physics*, 62(2):345–372, 1986.
- [3] Beckett, P. e Jennings, A., “Towards nanocomputer architecture”, *Proceedings of the seventh Asia-Pacific Conference on Computer systems architecture*, Vitória, Austrália, 2002.
- [4] Bohr, M. T., “Nanotechnology Goals and Challenges for Electronic Applications”, *IEEE Transactions on Nanotechnology*, 1(1), 56–62, 2002.
- [5] Capasso, F., Sen, S., Beltram, F., Vangurlekar, L. L. A. S., Smith, P., Shah, N. J., Malikand, R. J., e Cho, A. Y., “Quantum Functional Devices: Resonant-Tunneling Transistors, Circuits with Reduced Complexity, and Multiple-Valued Logic”, *IEEE Transactions on Electron Devices*, 36(10), 2065–2082, 1989.
- [6] Carruthers, J., Hammerstrom, D., Colwell, B., Bourianoff, G., e Zhirnov, V., “Computer architectures for nanoscale devices”, *Semiconductor Industry Association, Relatório Técnico*, 2003.
- [7] Costa, J. C., Hoekstra, J., Goossens, M., Verhoeven, C., e Roermund, H. M. V., “Considerations about Nanoelectronic GSI Processors”, *Analog Integrated Circuits and Signal Processing*, 24, 59–71, 2000.
- [8] Davis, J. A., Venkatesan, R., Kaloyeros, A., Beylansky, M., Souri, S. J., Banerjee, K., Saraswat, K. C., Rahman, A., Reif, R. e Meindl, J. D., “Interconnect Limits on Gigascale Integration (GSI) in the 21st Century”, *Proceedings of the IEEE*, 89, 305-324, 2001.
- [9] Durrani, Z. A. K. Coulomb blockade, single-electron transistors and circuits in silicon. *Physica E*, 17:572–578, 2003.
- [10] Fahmy, H., e Ismail, K., “Analysis of a single-electron decimal adder”, *Applied Physics Letters*, 70(19), 2613-2615, 1997.
- [11] Fang, Y., Cohen, M. A., e Kincaid, T. G., “Dynamics of a winner-take-all neural network”, *Neural Networks*, 9(7), 1141–1154, 1997.
- [12] Fausett, L., “Fundamentals of Neural Networks”, Prentice-Hall, New Jersey, E.U.A., 1994.

- [13] Ferry, D. K., “Transport in nanostructures”, volume 294. Cambridge University Press, E.U.A., 1991.
- [14] Fountain, T. J., Duff, M. J. B., Crawley, D. G., Tomlinson, C. D., e Moffat, C. D., “The use of nanoelectronic devices in highly parallel computing systems”, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 6(1), 31–38, 1998.
- [15] Fulton, T. A. e Dolan, G. D. Observation of single-electron charging effects in small tunnel junctions. *Physics Review Letters*, 59:109–112, 1987.
- [16] Geerligs, L. J. e Mooij, J. E., “Charging effects and turnstile clocking of single electrons in small tunnel junctions”, *Granular Nanoelectronics*, 1, 393–411, 1991.
- [17] Gepert, L., “Quantum transistors: toward nanoelectronics”, *IEEE Spectrum*, 36(9), 46–51, 2000.
- [18] Gerousis, C., Goodnick, S. M., e Porod, W., “Towards nanoelectronics cellular neural networks”, *International Journal of Circuit Theory and Applications*, 28, 523–535, 2000.
- [19] Goldhaber-Gordon, D. Montemerlo, M. S., Love, J. C., Opiteck, G. J., e Ellenbogen, J. C., “Overview of Nanoelectronic Devices”, *Proceedings of the IEEE*, 85(4), 521–540, 1997.
- [20] Goser, K., Pacha, C., Kanstein, A., e Rossman, M., “Aspects of systems and circuits for nanoelectronics”, *Proceedings of the IEEE*, 85(4), 558–573, 1997.
- [21] Grabert, H. e Devoret, M. H., editors, “Single Charge Tunneling - Coulomb Blockade Phenomena in Nanostructures”, *NATO ASI Series B: Physics*, Plenum Press, New York, 1992.
- [22] Guimarães, J. G., “Arquiteturas de redes neurais nanoeletrônicas para processadores em escala giga ou tera”, Tese de doutorado. Universidade de Brasília, Brasília, Brasil, 2005.
- [23] Guimarães, J. G., Carmo, H. C., e Costa, J. C., “Basic subcircuits with single- electron tunneling devices”, *Proceedings of the 17th Symposium on Technology and devices SBMICRO2002*, Porto Alegre, Brasil, 2002.
- [24] Guimarães, J. G., Carmo, H. C., Costa, J. C., “Single-electron winner-take-all network”, *Microelectronics Journal*, 35, 173-178, 2004.
- [25] Han, J. e Jonker, P., “A system architecture solution for unreliable nanoelectronic devices”, *IEEE Transactions on Nanotechnology*, 1(4), 201–208, 2002.
- [26] Haykin, S., “Neural Networks - a comprehensive foundation”, Prentice-Hall, New Jersey, E.U.A., 1994.

- [27] Lieber, C. M., “The Incredible Shrinking Circuit”, *Scientific American*, 285(3), 59-64, 2001.
- [28] Likharev, K. K., “Electronics below 10 nm”, *In Symposium and Summer School on Nano and Giga Challenges in Microelectronics.*, Elsevier, Amsterdam, 2002.
- [29] Likharev, K. K., “Single-electron devices and their applications”, *Proceedings of the IEEE*, 87, 606–632, 1999.
- [30] Likharev, K. K., “Single-electron transistors electrostatic analogs of the DC SQUIDS”, *IEEE Transactions on Magnetics*, 23, 1142–1145, 1987.
- [31] Likharev, K. K., Bakhvalov, N. S., Kazacha, G. S., e Serdyukova, S. I., “Single-Electron Tunnel Junction Array: An Electrostatic Analog of the Josephson Transmission Line”, *IEEE Transactions on Magnetics*, 25, 1436–1439, 1989.
- [32] Lin, H. C., Wang, M. F., Hou, F. J., Lin, H. N., Lu, C. Y., Liu, J. T., e Huang, C. Y., “High-performance p-channel schottky barrier SOI FinFET featuring self-aligned P₁Si source/drain and electrical junctions”, *IEEE Electron Devices Letters*, 24(2), 102–104, 2003.
- [33] Lyshevski, L., “Neuroscience - neuroarchitectronics - nanocomputers - and – nanotechnology”, *In Proceedings of the 2002 2nd IEEE Conference on Nanotechnology IEEE NANO2002*, Washington, E.U.A., 2002.
- [34] Lundstrom, M., “Is nanoelectronics the future of microelectronics?”, *Proceedings of the 2002 International Symposium on Low Power Electronics*, Monterey, E.U.A., 2002.
- [35] Mahapatra, S., Ionescu, A. M., e Banerjee, K., “A Quasi-Analytical SET Model for Few Electron Circuit Simulation”, *IEEE Electron. Lett.*, 23(6), 366-368, 2002.
- [36] MathWorks, T. Matlab 6.5, 2003.
- [37] Meindl, J. D., Chen, Q., e Davis, J. A., “Limits on Silicon Nanoelectronics for Terascale Integration”, *Science*, 293(5537), 2044–2049, 2001.
- [38] Nakashima, H. e Uozumi, K. Zero-bias current in a single-electron transistor with identical trapezoidal tunneling barriers for finite gate voltages. *Journal of Physics:Condensed Matter*, 10, 1998.
- [39] National Nanotechnology Initiative, <http://www.nano.gov>
- [40] Ono, Y., Fujiwara, A., Nishiguchi, K., Inokawa, H., e Takahashi, Y. Manipulation and detection of single electrons for future information processing. *Journal of Applied Physics*, 97(031101):1–9, 2005.
- [41] Pothier, H., Lafarge, P., Urbina, C., Esteve, D., e Devoret, M. H., “Single-electron

- pump based on charging effects”, *Europhysics Letters*, 17(3), 249–251, 1992.
- [42] Raja, T., Agraval, V. D. e Bushnell, M. L., “A Tutorial on the Emerging Nanotechnology Devices”, *Proceedings of the 17th International Conference on VLSI Design (VLSID04)*, Mumbai, Índia, 2004.
- [43] Risch, L., “Nanoelectronics: The key technology of the 21st century”, *Siemens-Review, R & D Special*, 32–35, 1996.
- [44] Semiconductor Industry Association, “International Technology Roadmap for Semiconductors”, 2005.
- [45] Snider, G. S., e Williams, R. S., “Nano/CMOS architectures using a field-programmable nanowire interconnect”, *Nanotechnology*, 18(3), 1-11, 2007.
- [46] Srivastava, N., e Banerjee, K., “Performance Analysis of Carbon Nanotube Interconnects for VLSI Applications”, *International Conference on Computer-Aided Design IEEE/ACM*, 2005.
- [47] Tucker, J. R., Wang, C., e Carney, P. A., “Silicon field-effect transistor based on quantum tunneling”, *Applied Physics Letters*, 65(5), 618–620, 1994.
- [48] Turel, O. e Likharev, K., “Crossnets: possible neuromorphic networks based on nanoscale components”, *International Journal of Circuit Theory and Applications*, 31, 37–54, 2003.
- [49] Verbrugh, S., “Development of a Single-Electron Turnstile as a Current Standard”, *Tese de Doutorado*, Delft University of Technology, Delft, Holanda, 1995.
- [50] Wasshuber, C., Kosina, H., e Selberherr, S., “SIMON - a Simulator for Single-Electron Tunnel Devices and Circuits.”, *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 16(9), 937–944, 1997.
- [51] Yu, Y. S., Hwang S. W., e Ahn, D. D., “Macromodeling of Single-Electron Transistors for Efficient Circuit Simulation”, *IEEE Transactions on Electron Devices*, 46(8), 1667-1671, 1999.

APÊNDICES

A - RESISTÊNCIA DE TUNELAMENTO

A resistência de tunelamento da junção R_j é uma quantidade fenomenológica definida quando uma diferença de tensão V é aplicada aos dois terminais de uma junção-túnel. A expressão para R_j está na Equação A.1.

$$R_j = \frac{\hbar}{2\pi e^2 |P_T|^2 D_i D_f} \quad (\text{A.1})$$

onde $\hbar = h/2\pi$, $|P_T|^2$ é a probabilidade de transmissão, D_i é a densidade de estados inicial e D_f é a densidade de estados final.

A probabilidade de transmissão é definida como a razão entre a amplitude provável da frente de onda transmitida após o choque com uma barreira de potencial, B_3 , e a amplitude provável da frente de onda antes do choque, B_1 . $|P_T|^2$ está na Equação A.2.

$$|P_T|^2 = \frac{|B_3|^2}{|B_1|^2} \quad (\text{A.2})$$

A taxa de tunelamento de elétrons através da barreira Γ , imposta pela junção, é proporcional a V : $\Gamma = V/eR_T$.

B - ROTINA DO MATLAB

```
% definicao dos parametros do circuito:
Cg=1e-18;
C=1e-19;
Vbias=-0.05;
R=1e6;
Rt=1e7;
Ct=1e-15;
Rcon=1e3;
Ccon=100e-18;
Csum=2*C+Cg;
Rsum=2*R;
I1=4e-9;
I2=2e-9;
% definicao das constantes:
e=1.6e-19;
kB=1.38e-23;
% parametros de simulacao:
T=3;
n=10;
% valores iniciais das tensoes:
v1(1)=0;
v2(1)=0;
vout1(1)=0;
vout2(1)=0;
% definicao do intervalo de tempo e do tempo total:
dt=1e-8;
tfinal=2e-7;
t(1)=dt;
i=1;
% equacoes iniciais:
VT=kB*T/e;
```

```

Vsum=e/(2*Csum);
Ioff=(VT/(2*Rsum))*log(1+exp(1-(Vsum/(n*VT))));
while (t(i) <= tfinal)
    i=i+1;
    t(i)=t(i-1)+dt;

    % para o neuronio 1:
    vds1(i-1)=v1(i-1)-Vbias;
    vgs1(i-1)=vout2(i-1)-Vbias;
    vilha1(i-1)=vds1(i-1)*(C/Csum)+vgs1(i-1)*(Cg/Csum);
    if (vilha1(i-1)>max([Vsum abs(vds1(i-1)-vilha1(i-1))]))
        while (abs((vilha1(i-1))>Vsum))
            vilha1(i-1)=vilha1(i-1)-2*Vsum*sign(vds1(i-1));
        end
        if (abs(vds1(i-1)-vilha1(i-1))>Vsum)
            ID1(i-1)=(vds1(i-1)-vilha1(i-1)-Vsum*sign(vds1(i-1)))/R;
            IS1(i-1)=(vilha1(i-1)+Vsum*sign(vds1(i-1)))/R;
            IDS1tun(i-1)=(IS1(i-1)*ID1(i-1))/(IS1(i-1)+ID1(i-1));
            IDS1temp(i-1)=(vds1(i-1)/(2*Rsum)-IDS1tun(i-1))*2*VT/Vsum;
        else
            IDS1tun(i-1)=0;
            IDS1temp(i-1)=Ioff;
        end
    else
        if (abs(vds1(i-1)-vilha1(i-1))>Vsum)
            while (abs(vds1(i-1)-vilha1(i-1))>Vsum)
                vilha1(i-1)=vilha1(i-1)+2*Vsum*sign(vds1(i-1));
            end
            if (abs(vilha1(i-1))>Vsum)
                ID1(i-1)=(vilha1(i-1)-Vsum*sign(vds1(i-1)))/R;
                IS1(i-1)=(vds1(i-1)-vilha1(i-1)+Vsum*sign(vds1(i-1)))/R;
                IDS1tun(i-1)=(IS1(i-1)*ID1(i-1))/(IS1(i-1)+ID1(i-1));
                IDS1temp(i-1)=(vds1(i-1)/(2*Rsum)-IDS1tun(i-1))*2*VT/Vsum;
            end
        end
    end
end

```

```

else
    IDS1tun(i-1)=0;
    IDS1temp(i-1)=Ioff;
end
else
    IDS1tun(i-1)=0;
    IDS1temp(i-1)=Ioff;
end
end
IT1(i-1)=IDS1tun(i-1)+IDS1temp(i-1);
v1(i)=(Rt/(1+Rt*Ct/(dt)))*(I1-IT1(i-1)+Ct*v1(i-1)/(dt));
vout1(i)=(1/(1+Rcon*Ccon/(dt)))*(v1(i)+Rcon*Ccon*vout1(i-1)/(dt));

% para o neuronio 2:
vds2(i-1)=v2(i-1)-Vbias;
vgs2(i-1)=vout1(i-1)-Vbias;
vilha2(i-1)=vds2(i-1)*(C/Csum)+vgs2(i-1)*(Cg/Csum);
if (vilha2(i-1)>max([Vsum abs(vds2(i-1)-vilha2(i-1))]))
    while (abs((vilha2(i-1))>Vsum))
        vilha2(i-1)=vilha2(i-1)-2*Vsum*sign(vds2(i-1));
    end
    if (abs(vds2(i-1)-vilha2(i-1))>Vsum)
        ID2(i-1)=(vds2(i-1)-vilha2(i-1)-Vsum*sign(vds2(i-1)))/R;
        IS2(i-1)=(vilha2(i-1)+Vsum*sign(vds2(i-1)))/R;
        IDS2tun(i-1)=(IS2(i-1)*ID2(i-1))/(IS2(i-1)+ID2(i-1));
        IDS2temp(i-1)=(vds2(i-1)/(2*Rsum)-IDS2tun(i-1))*2*VT/Vsum;
    else
        IDS2tun(i-1)=0;
        IDS2temp(i-1)=Ioff;
    end
end
else
    if (abs(vds2(i-1)-vilha2(i-1))>Vsum)
        while (abs(vds2(i-1)-vilha2(i-1))>Vsum)
            vilha2(i-1)=vilha2(i-1)+2*Vsum*sign(vds2(i-1));

```



```

end
if (abs(vilha2(i-1))>Vsum)
    ID2(i-1)=(vilha2(i-1)-Vsum*sign(vds2(i-1)))/R;
    IS2(i-1)=(vds2(i-1)-vilha2(i-1)+Vsum*sign(vds2(i-1)))/R;
    IDS2tun(i-1)=(IS2(i-1)*ID2(i-1))/(IS2(i-1)+ID2(i-1));
    IDS2temp(i-1)=(vds2(i-1)/(2*Rsum)-IDS2tun(i-
1))*2*VT/Vsum;
else
    IDS2tun(i-1)=0;
    IDS2temp(i-1)=Ioff;
end
else
    IDS2tun(i-1)=0;
    IDS2temp(i-1)=Ioff;
end
end
IT2(i-1)=IDS2tun(i-1)+IDS2temp(i-1);
v2(i)=(Rt/(1+Rt*Ct/(dt)))*(I2-IT2(i-1)+Ct*v2(i-1)/(dt));
vout2(i)=(1/(1+Rcon*Ccon/(dt)))*(v2(i)+Rcon*Ccon*vout2(i-1)/(dt));
end
g=max(size(v1));
h=1:1:g;
plot(t,v1,t,v2);

```



Considerations about interconnection limits on basic subcircuits with single-electron tunneling devices

V. G. A. Carneiro ^a, J. G. Guimarães ^{a*} and J. C. da Costa ^a

^(a) Departamento de Engenharia Elétrica, Universidade de Brasília, Campus Universitário, CP 4386, Brasília-DF, 70904-970, Brazil

Elsevier use only: Received date here; revised date here; accepted date here

Abstract

Interconnection limits seem to be a potential problem to the evolution of the semiconductor industry, especially in the nanoscale. In this work, the electrical performance of basic cells is studied with the help of a simple interconnection model, whose parameters can be changed. Our goal, with this study, is to determine the interconnection's influence upon the circuit behavior and to establish interconnection-related limits for its functionality. An extrapolation to more complex circuit topologies is also discussed. Finally, the implementation possibilities using new interconnection technologies, like carbon nanotubes, are presented. © 2007 Elsevier Science. All rights reserved.

PACS: 73.20Dx; 71.50+t; 71.10+X; 71.70Ej

Keywords: GSI/TSI Processor, Nanoelectronic, Single-electron transistor, Winner-take-all, Interconnection, Carbon nanotube

1. Introduction

The basic technologies presently adopted by the semiconductor industry for memory and processor fabrication should attain limits imposed by the laws of physics around the year 2010 [1]. Single-electron transistors, like other nanoscale devices, seem to be a promising option for GSI or TSI implementations in the future.

The development of GSI and TSI processor architectures, based upon nanoelectronic devices, like the one in Figure 1 [1], is being carried out in our research group. The adopted approach comprises the implementation of a massive parallel-distributed processing core, built with nanoelectronic devices, organized in cells.

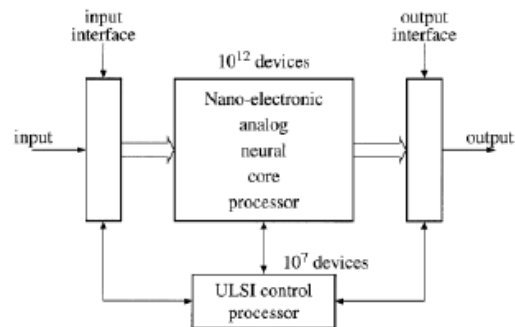


Figure 1. GSI processor's block diagram.

* Corresponding author. E-mail addresses: janaina@ene.unb.br (J.G. Guimarães); camargo@ene.unb.br (J.C. da Costa).

The investigation on these architectures has considered, up to now, the employment of ideal interconnections. But interconnection limits potentially threaten to decelerate or halt the historical progression of the semiconductor industry [2].

In this work, the electrical performance of a winner-take-all neuron circuit, built with single-electron transistors (SET-WTA) [3], like the one shown in Figure 2 is studied. The cell in Figure 2 has two neurons with a lateral inhibition configuration.

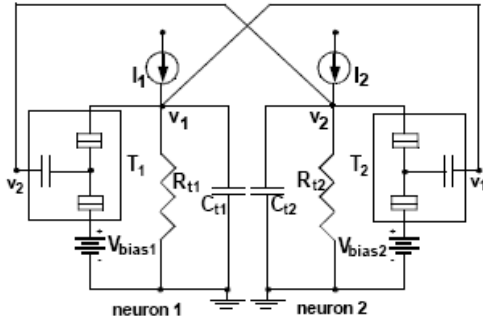


Figure 2. Two-neuron SET-WTA network with lateral inhibition.

In order to obtain the electrical performance of this circuit, we have performed two kinds of simulation. In one of them, we used SIMON (Simulation of Nanostructures), a known and largely accepted simulator that uses Monte-Carlo method [4]. In the other, a quasi-analytical SET model [5] was implemented in MatLab [6] to confirm our results.

2. WTA Operation and First Simulations

Each neuron has a primary current input, which brings the data information to the network, and secondary voltage input units, which come from the neighbour. These secondary connections provide the stimulation or inhibition features.

The negative reinforcement from the winner to its neighbour suppresses its output, increasing the activation of the neuron with a higher primary input

current and decreasing the activation of the neuron with a lower primary input current [3].

The first simulations were performed without interconnection concerns. In both simulations (SIMON and MatLab) we acquired similar results, which can be seen in Figure 3.

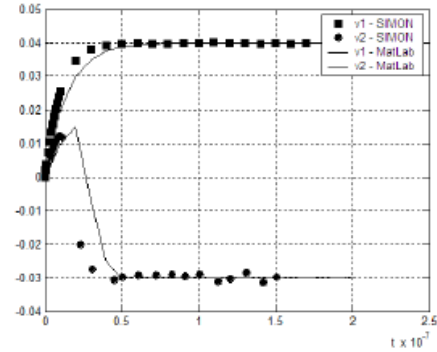


Figure 3. Outputs v_1 and v_2 of the two-neuron SET-WTA with ideal interconnections for simulations in Matlab and SIMON. The simulation parameters are $I_1 = 4$ nA, $I_2 = 2$ nA, $V_{bias} = -50$ mV, $R_{t1} = R_{t2} = 10$ M Ω , $C_{t1} = C_{t2} = 1$ fF. Both SETs have $R = 1$ M Ω , $C = 0.1$ aF and $C_g = 1$ aF and the temperature are 3 Kelvin.

3. Interconnection Model and Results

To a first approximation, we considered the influences of the intrinsic resistivity ρ and the coupling capacitance between adjacent wires. Under these considerations, our interconnection model is a simple RC circuit. How the tunneling resistances inside the SET are very high, the best point to simulate the interconnection influences is between each neuron, especially because in larger WTA networks, the distance between neurons can be larger. So, the entire WTA-SET circuit with interconnection parameters can be modeled like the one in Figure 4.

From [7], typical values of resistance and capacitance per unit length can be taken for copper (Cu) nanowires with 15 nm of width and a 3 nm thick SiO_2 switching layer between them. Also from [7], a typical nanowire length of 7.11 μm was assumed, yielding approximately for this configuration a maximum nanowire resistance and capacitance of 2.5 k Ω and 2 fF, respectively.

Besides these typical values for Cu, parameter values for carbon nanotube (CNT) interconnections can be found in the literature (8-12). For the same maximum connection length, it is estimated that $R_{CNT} = 46 \text{ k}\Omega$ and $C_{CNT} = 200 \text{ aF}$.

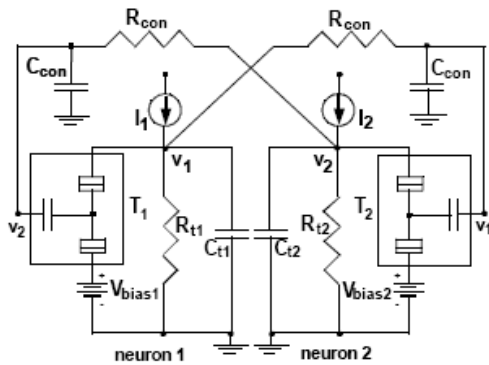


Figure 4. Two-neuron SET-WTA network with interconnection model.

With these values in mind, the simulation of the circuit of Figure 4 was performed with the following values for the resistance and capacitance: $R_{con} = R_{CNT} \sim 50 \text{ k}\Omega$ e $C_{con} = C_{Cu} = 2 \text{ fF}$. The results can be seen in Figure 5.

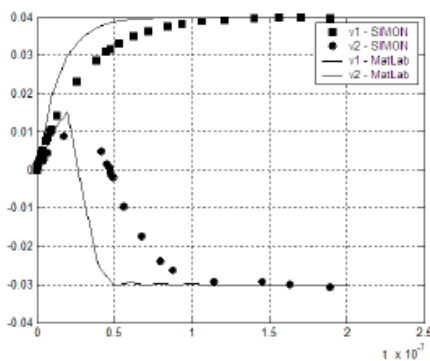


Figure 5. Outputs v_1 and v_2 of the two-neuron SET-WTA with realistic interconnections for simulations in Matlab and SIMON. The simulation parameters are the same as in Figure 3.

4. Conclusion and Next Studies

As can be seen, the results were a little bit different. The RC constant influence on the circuit was starker in SIMON. This is due to approximations used in the quasi-analytical SET model for MatLab. In SIMON, the interactions between interconnection parameter and SETs are considered, while this is not done in our MatLab model.

Despite this fact, both methods predict well the behavior of both SET-WTAs and we can observe the influence of some interconnection parameters with real values taken from the literature. We shall persist in the use of other kinds of simulation like this in MatLab, because simulation times are much better.

Next studies include improvements in simulation and interconnection models, to achieve more accuracy, extrapolation of parameters to achieve interconnection length limits and simulations with association of other basic nanoelectronic subcircuits.

References

- [1] J. C. da Costa, J. Hoekstra, M. J. Goossens, C. J. M. Verhoeven and A. H. M. V. Roermund, *Journal of Analog Integrated Circuits and Signal Processing*, 24, (2000) 59.
- [2] J. A. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S. J. Souri, K. Banerjee, K. C. Saraswat, A. Rahman, R. Reif and J. D. Meindl, *Proc. IEEE*, 89 (2001) 305.
- [3] J. G. Guimarães, H. C. do Carmo, J. C. da Costa, *Microelectronics Journal*, 35 (2004) 173
- [4] C. Wasshuber, H. Kosina, S. Selberherr, *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, 16 (1997) 937.
- [5] S. Mahapatra, A. M. Ionescu, K. Banerjee, *IEEE Electron. Lett.*, 23(6) (2002) 366.
- [6] MathWorks, *T. Matlab 6.5*, 2003.
- [7] G. S. Snider, R. S. Williams, *Nanotechnology*, 18 (2007) 1.
- [8] P. L. McEuen, et al., *IEEE Trans. Nanotechnology*, 1(1) (2002) 78
- [9] P. J. Burke, *IEEE Trans. Nanotechnology*, 1(3) (2002) 129
- [10] S. Datta, *Nanotechnology*, 15 (2004) S433
- [11] J.-Y. Park, et al., *Nano Letters*, 4(3) (2004) 517
- [12] S. Datta, Cambridge University Press, Cambridge (1995)

Design of a Reconfigurable SoC-Based Fruit Monitoring System

G. S. Beserra, L. T. Maia, G. M. Araujo, J. V. Pimentel, F. Schlag, M. L. Graciano Jr.,
D. M. Amaral, V. G. A. Carneiro, J. E. G. Medeiros, W. A. Araujo, J. C. da Costa
Universidade de Brasilia – Brasilia - Brazil
{gbeserra, letmaia, schlag, camargo}@umb.br

Abstract

In this paper, a strategy based upon a reconfigurable System-on-Chip (rSoC) system is proposed as a new solution to evaluate product condition during vegetable postharvest handling procedures. The system is composed by: 1) local monitoring modules (pseudo-fruits), each one containing an rSoC with an embedded image sensor; 2) a set of environmental sensors (e.g. temperature, humidity, and impact); 3) a communication and local supervision module, for data and control information exchange with terrestrial or satellite based links. The system is organized through a pseudo-fruit network inside the cargo container. Data gathered by the pseudo-fruits can be sent to external supervision centers through the communication module, allowing preventive or sanitary actions during the cargo displacement. In this work, a study of suitable architectures for this application is carried out.

1. Introduction

The increasing food export activity in global scale presents various challenges to producers. In the case of perishable goods such as fresh fruits, the maintenance of adequate environmental transport conditions is a crucial goal. Tracking a fruit load throughout its whole export chain is an essential requirement to assure the product quality to its end consumer. The strategy presented in this work adopts wireless sensor networks distributed inside the fruit packages in a cargo container to monitor fruit samples and the container environmental conditions in order to evaluate the cargo conditions along its transport chain. A communication and local supervision module (also attached to the transport vehicle) integrates the data gathered by the sensor network and provides status reports via terrestrial monitoring stations placed along the transport route or via the satellite link also used for the transporter global positioning tracking. The embedded supervision unit may also generate alarm signals whenever the cargo might be under hazardous transport

conditions. The solution proposed in this work uses sensor nodes based upon System on Chip (SoC) [1] approach, in order to optimize cost x performance trade-offs. A wireless SoC is typically composed by several modules, and their integration is a complex task. Issues like IP modules integration, power consumption, electromagnetic compatibility (EMC), data management, design and test issues and system performance should be tackled together, trying to optimize the design for a given application. An rSoC includes reconfigurable structures, providing a flexible way to use a variety of logic, memory, and mixed-mode signal cores in a single chip with additional advantages, such as real-time response and adaptability. Because of that, it is well-suited to be used in conjunction with different sensors to assemble a pseudo-fruit that can be dispatched in the midst of fruit containers used to deliver these products to intermediate and end consumers. That autonomous device can collect environmental data as well as IR spectral responses from fruits, in order to evaluate their state during transportation. Due to its reconfigurability, the rSoC can also be used to monitor various kinds of fruit and many other food export products.

2. Monitoring System Specifications

Figure 1 shows how the system works. The fruit transportation can cause serious damage, like mechanical compression and temperature rise, increasing the losses. One or more local monitoring units can be included inside the container, composing a wireless network. They will replace actual fruits and collect environmental data. The rSoC microprocessor will send the processed data through the RF transceiver to the communication module, that will gather the information and, if needed, transmit it to terrestrial stations or to the GPS transponder located in the vehicle for satellite communication.

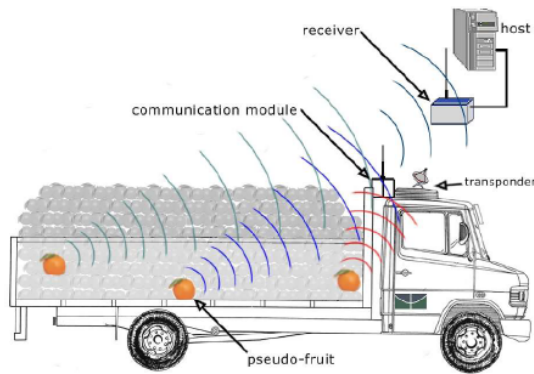


Figure 1. Fruit monitoring system

The monitoring module is responsible for data acquisition, processing and communication. It is composed by a rSoC with an integrated APS (Active Pixel Sensor) matrix operating in IR range that will be responsible for the acquisition of the fruits spectral responses. The rSoC also has a RF transceiver operating in 900 MHz to send data to the communication module, a 32-bit RISC microprocessor, memory, reconfigurable structures working as a co-processor, and analog and digital interfaces.

Since the main goal of the pseudo-fruit is to collect environmental data, some discrete sensors will be included on it. They will measure important parameters, such as temperature, humidity and impact, to be evaluated and sent to external supervision centers.

The communication and local supervision module consists of an embedded microprocessor and an RF transceiver for terrestrial and satellite communications. This module will gather all the information sent by the pseudo-fruits and can use the vehicle battery as power supply. After processing it, data will be sent to the terrestrial supervision stations via an omnidirectional antenna or to the vehicle GPS transponder via modem. Eventually, an actuator can be added (e.g., an alarm to warn the driver about critical conditions for the fruits or low battery in the monitoring module).

3. Design Considerations

System design begins with specifying the required functionality. The most common way to achieve the precision in specification is to consider the system as a collection of simpler subsystems and methods for composing these subsystems (objects) to create the required functionality [2]. The modules and their interconnections can be modeled in a behavioral level and simulated using MATLAB functional blocks, as it was shown in [3]. The system structure can also be expressed using UML (Unified Modeling Language) [4].

The next step is to transform the system functionality into an architecture [2]. Since there will be analog and digital instances in the same chip, two different design flows will be adopted. The first one will be used to implement the analog and RF sections, and it consists of a DFT, full-custom design methodology, using both top-down and bottom-up approaches. The second one will be applied on the digital section design, which will be implemented using Electronic Design Automation (EDA) for synthesis, place and route, timing analysis and verification.

Because the system will contain modules operating at different frequencies, there are several challenges related to electromagnetic compatibility (EMC) issues. Because of that, the methodology aimed towards EMC at chip level. It shall be implemented in association with power, delay and area minimization strategies [4, 5].

A hardware/software co-design will be used to allocate tasks and define the interface functions. Because component reuse is an important key to gain productivity, the design flow will be based on reusable IP.

The system will be fabricated using CMOS technology. Considering the APS matrix and the RF transceiver as critical parts of the chip, 0.18um technology seems to be adequate because it allows a smaller area and higher frequency operation with a 1.8V power supply. Low power consumption is a critical requirement to increase the operating time.

4. rSoC Description

The SoC design will be IP-based, adherent to VSIA standards and using high-performance bus architectures [6]. The diagram of the rSoC is shown in Figure 2. The blocks are described in the following subsections.

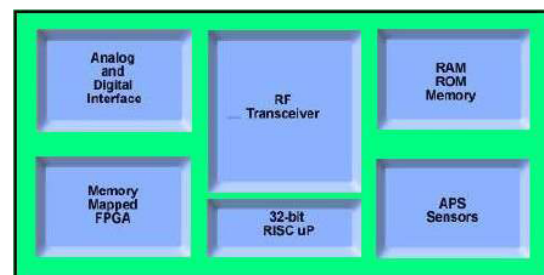


Figure 2. rSoC architecture

Table 1 shows the estimated area of the complete chip. A CQFP 204 package will be used.

Table 1. Estimated area

Component	Estimated Area (mm ²)
PA (Power Amplifier)	10
Memory:	
- Instruction cache and Data cache	9
- Configuration cache	2
Microprocessor Core	2
Configurable Blocks	16
Interfaces	5
Core (total)	44
PADS	26
Chip	70

4.1. Sensors and Interfaces

The data retrieved by the measuring devices must be accurate. Because of that, the sensors must be carefully chosen among the available devices.

The specifications for each sensor will vary according to the desired use. For example, in quality control monitoring, knowing if the fruit has been kept under certain conditions during handling procedures may be enough. To analyze the transportation, the variation of the sensed characteristics with time should be recorded. This will influence the choice of the devices, since each use may have different requirements for certain characteristics, such as time response.

4.1.1. Humidity

A common way of measuring humidity using microelectronics is through capacitance. The value of the capacitance between two conducting plates depends on the relative permittivity (ϵ_r) of the dielectric – which changes with moisture. Capacitive humidity sensors are small and their properties are nearly independent of temperature changes. For use in vegetable tracking applications using pseudo-fruits, small size is desirable, but miniaturization is not a requirement, since the sensors will be placed inside a sphere or on its surface. Thus, it is possible to use commercial capacitive sensors, of which there are plenty available at low cost [7], already packed with pre-processing interfaces.

One drawback of capacitive humidity sensors is that their response time is not very fast – typically, 30 to 60s for a 63% change in the relative humidity (RH) [8]. For the given application, though, they are still adequate. A different type of humidity sensor with a better time response is the resistive sensors, but these sensors present various disadvantages in other aspects: high temperature dependencies, hysteresis, nonlinear (exponential) response, etc. MEMS (micro-electro-machine systems) offer even better response time, but at the cost of a much higher complexity.

4.1.3. Temperature

A ZMD temperature sensor IC, model TSic 301 [9], is proposed as a possible choice to the system temperature sensor. This sensor employs a high precision bandgap reference with PTAT output; a low power precision ADC and a on-chip DSP core with EEPROM to precisely calibrate the output temperature signal. It is a low cost precision device and its characteristics are appropriate to the project. Some important features are supply voltage between 3V and 5V, fast data measurement, no calibration by customer needed and CMOS technology.

4.1.4. Impact

Among the available devices, the LIS2L02AQ3 [10] is a high performance, low-power, analog output 2-axis linear accelerometer. It is provided by STMicroelectronics and includes a sensing element and an IC interface, with important features such as supply voltage between 2.4V and 5.25V, low power consumption and embedded self-test and power down.

4.1.5. Optical sensor

Because it can be produced using CMOS process technology, the chosen image sensor is the APS (Active Pixel Sensor). Since the application environment is dark, the operating range is the infrared.

As shown in Figure 3, a standard active pixel has a reset transistor (M1), a readout transistor (M2) and a row select transistor (M3). The photodiode (PD) has a better blue and green response than the photogate (PG), which makes it a better choice for color APS matrix [11]. In order to get color images, the integrated color pixel (ICP) arrangement is a new alternative to current pixel design that employs color filter arrays (CFA), as shown in [12].

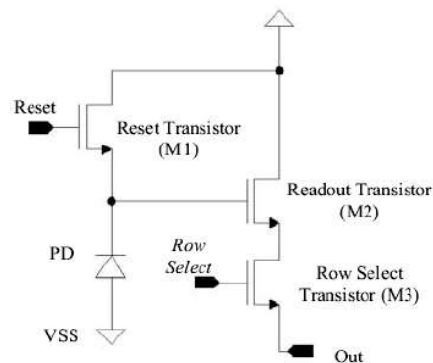


Figure 3. Pixel architecture

The sensors can be arranged in a matrix, with additional logic circuits for row and column selection, amplifiers, A/D converter to digitalize the image acquired by the sensors, and a buffer to store the data before they are sent to the microprocessor. Considering that a 256-color video application needs a frame rate of 30 Hz, it would be possible to improve the results by increasing the frame rate to 90 Hz.

The system reconfigurability can optimize the APS operation by improving the APS signal processing efficiency and changing the execution of a specific instruction by software, if needed.

4.2. The RF Transceiver

An RF transceiver is included in each rSoC to allow communication among the units. In the monitoring system, in which the pseudo-fruit is inside the truck, the collected data have to be transmitted to a repeater located on the truck. Then, the signal can be transmitted to the target device. An RF transceiver designed in [1] is well suited to this system. The block diagram is shown in Figure 4.

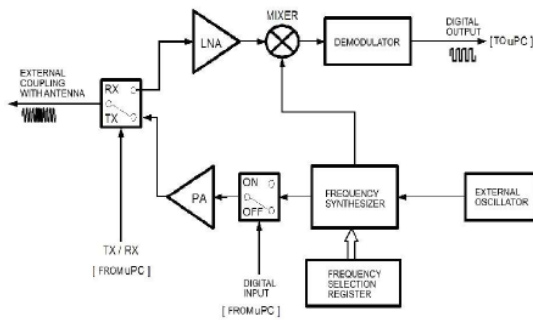


Figure 4. RF transceiver blocks

The transceiver operates at ISM band (915-927,75 MHz) and uses OOK modulation. The transmission is half-duplex and uses TDM (Time Division Multiplex). The first stage of the receiver is a LNA with a sensitivity of -100 dB, gain of 30 dB and noise figure in 2 – 5 dB range. The PA output stage has a maximum transmission power of 10 mW. The estimated transceiver area is 14 mm². A spread-spectrum transceiver operating at 2.4 GHz is being analyzed.

4.3. Microprocessor

A 32-bit RISC (Reduced Instruction Set Computer) architecture, with a dedicated multiplication hardware circuit, fixed-point arithmetic operations, memory hierarchy and efficient pipeline was chosen because its fixed size instructions increase the system performance and decrease the datapath complexity.

Considering those characteristics, there are some core options available, but the choice will be limited to one of the following architectures. One of them is the IP-suite for SoC XiRisc microprocessor [13], a 32-bit RISC HDL model with configurable instruction set architecture and datapath. It has been implemented in some SoC architectures using 0.18µm technology and tested with DSP algorithms [14]. The other one is the MOLEN polymorphic CCM (Custom Computing Machine) processor based on co-processor architectural paradigm that has been synthesized and prototyped into FPGA platform. Its concept has been validated using MPEG-2 application [15].

4.4. Memory

A static RAM will be used to store the microprocessor instructions and data because it does not need refresh and it works faster than dynamic RAM. In [14], a 64 KB instruction cache and a 64 KB data cache were implemented in 0.18µm technology with a total area of 9mm².

A ROM memory will be included in order to store boot and test routines. It is also possible to use an external flash memory to increase the storage size and allow bigger routines.

4.5. Reconfigurable Blocks

Since the hardware reconfigurability is one of the main characteristics of the system, an integrated memory mapped FPGA will be included, i. e., some memory addresses will be used to access the FPGA.

The FPGA can configure specific routines that usually demand successive calculations, like treatment, zoom, compression and edge recognition of the images acquired by the APS matrix, and also frequency selection of the RF transceiver. These routines will work as co-processors in the FPGA, which increases the processing speed and reduces the computational load on the main processor core. The parameters of each co-processor will be saved on the memory, and a specific hardware to select the co-processors and to access the memory in order to improve the parallel processing among the microprocessor and the co-processors must be implemented. Finally, the microprocessor must have several specific instructions to handle the FPGA configuration.

It is also possible to use a pipelined run-time configurable datapath (PiCo gate array, pGA or PiCoGA) as a configurable unit integrated in the processor core. It consists of an array of rows, each one representing a possible stage of a customized pipeline. This approach was used in [14], in which each PiCoGA row implemented in 0.18µm technology occupies 0.6 mm². The average power was 16mW/row.

4.6. Operating System

In embedded systems, it is very useful to adopt RTOS (Real-Time Operating Systems) because of its concurrent execution and real-time response, two important characteristics for the increasing amount of applications in which time constraints are critical [16]. In our system, the RTOS will also be used to enhance the hardware reconfigurability.

Among the several available RTOS, eCos (Embedded Configurable Operating System) seems to be a good choice. It is currently developed by eCosCentric Limited [17] with a free version, open-source code. It consists of a real-time kernel with several components, such as TCP/IP protocol, flash file systems, POSIX (Portable Operating System Interface for Unix) compatibility, drivers for many devices, a graphical interface, and a portable HAL (Hardware Abstraction Layer), a layer between hardware and software that allows portability to different architectures (16, 32 and 64-bit) by managing hardware devices. eCos was chosen because it offers many advantages, such as flexibility to be adapted to many applications and hardware architectures. It is also free of royalties and has an excellent real-time performance, with several scaling algorithms and a mechanism for management of interruptions and exceptions. Finally, it can be programmed using C language.

The goal is to use eCos in the chip to customize the system in order to reach the best real-time performance and to optimize the area occupied by the hardware. With a graphical tool, the programmer will be able to configure the system using hardware and software libraries and compilation packages [18].

5. Application and Discussion

The system diagram can be seen in Figure 5. The goal is to monitor the cargo environmental conditions in order to evaluate the mechanical damage to product [19], and hazardous humidity and temperature effects. By analyzing those data, preventive actions can be taken to minimize potential losses.

The pseudo-fruits included inside the cargo will work as a wireless network.

First of all, the sensors must be activated and the gathered data processed by the rSoC microprocessor. In the communication module, the received data can be processed and sent to a terrestrial supervision center using an RF transceiver or eventually, via modem, to a GPS transponder that is commonly used as a tracking device in trucks. In this case, a commercial protocol can be used in the communication between the transponder and the satellite, and also between the satellite and the supervision center. Each supervision center must have a computer with a reliable and dedicated software to store and manage the collected and processed data.

The values acquired by the sensors must be analyzed and compared with the optimum values, for example the expected spectral, temperature and humidity range, in order to evaluate if the fruits will be in an adequate consumption condition or not.

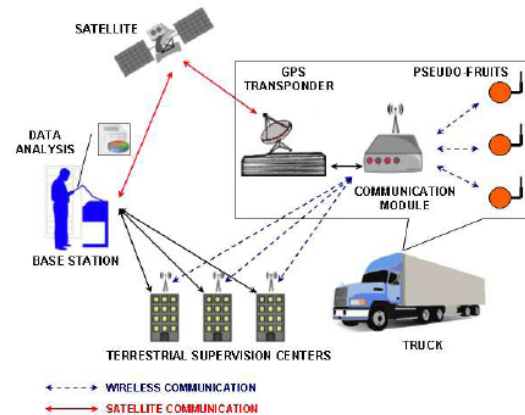


Figure 5. System diagram

The commercial available systems [19] usually focus on registering impact on the fruits. The proposed system uses more instruments to measure other parameters, like temperature and humidity, and provides flexibility due to the rSoC device present on it.

Energy saving strategies will be adopted to optimize the system's operating time. In order to increase the pseudo-fruit battery duration, eventually the acquired data can be only saved in the rSoC memory and sent to be processed in the communication module. Another possibility is to work in standby after data acquisition and transmission.

The communication inside the container can be made using the rSoC transceiver and a free protocol like SNAP (Scaleable Node Address Protocol), that presents advantages like error detection, broadcast messages, scalable header, etc.

6. Conclusion

In this work a system solution for fresh fruit monitoring along global export chains was presented. It employs wireless rSoC-based sensor networks to carry out monitoring and alert tasks, allowing a prompt and efficient surveillance of the status of such perishable goods during their transportation. Presently the overall system specifications and rSoC design are under construction and the first rSoC building blocks shall be sent for prototyping by the last quarter 2007. The approach adopted to the fruit tracking application can easily be extended to many other major logistics problems at global scale. Due to the adopted

reconfigurable approach, several different applications may be implemented with only minor hardware modifications and a favorable cost x efficiency tradeoff.

7. Acknowledgments

The authors would like to acknowledge CNPq and PADCT/Millennium Institute (Brazilian government agencies) for providing the sponsorship.

8. References

- [1] J. C. da Costa *et al.*, "CMOS SoC for Irrigation Control", In: Proceedings of IEEE International SoC Conference, pp. 51-54, 2005.
- [2] Rajsuman, R., "System-on-a-Chip: Design and Test", Artech House, 2000.
- [3] M. D. Melo; J. C. Costa; R. P. Jacobi, Modelagem Funcional de um Processador RISC 16 Bits em Sistema em Chip para Aplicações Sem Fio". In: X Workshop Iberchip, 2004.
- [4] N.S. Voros, S. Tsasakou, V. Mariatos, M. Birbas, A. Birbas and A. Andritsou, "System-on-a-Chip Methodology for Telecom Applications", Int. Symp on Signal, Systems and Electronics, URSI, pp. 321-325, 1998.
- [5] D. E. Lackey, "Applying Placement-based Synthesis for n-time System-on-a-Chip Design", IEEE Custom Integrated Circuits Conference, pp. 121-124, 2000.
- [6] <http://www.eetimes.com/story/OEG20000515S0026>
- [7] Fruett, F., "Sensores Integrados em Silício – Sensores de umidade", UNICAMP – FEEC – DSIF.
- [8] <http://www.sensorsmag.com/articles/0701/54/main.shtml>
- [9] www.zmd.biz
- [10] www.st.com
- [11] S. Djaja, G. H. Chapman, D. Y. H. Cheung, Y. Audet, "Implementation and Testing of Fault-Tolerant Photodiode-based Active Pixel Sensor (APS)". In: IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, 2003.
- [12] P. B. Catrysse, D. A. Wandell, "Integrated Color Pixels in 0.18um Complementary Metal Oxide Semiconductor", JOSA A, Vol. 20, n. 12, December, 2003.
- [13] <http://xirisc.deis.unibo.it/>
- [14] A. Lodi *et al.*, "A VLIW Processor with Reconfigurable Instruction Set for Embedded Applications". In: IEEE Journal of Solid-State Circuits, Vol. 38, No. 11, pp. 1876-1886, November 2003
- [15] <http://ce.et.tudelft.nl/MOLEN/Prototype/>
- [16] Farines, J. M., Fraga, J. S., Oliveira, R. S., "Sistemas de Tempo Real", Florianópolis, 2000.
- [17] <http://ecos.sourceforge.org/>
- [18] Massa, A. J., "Embedded Software Development with eCos", Prentice Hall, 2003.
- [19] A. F. Bollen, "Technological innovations in sensors for assessment of postharvest mechanical handling systems", Supply Chain Systems Group, Lincoln Ventures Ltd. Ruakura Research Centre, Hamilton, New Zealand, 2006.

E - 2º ARTIGO SUBMETIDO NA ÁREA DE MICROELETRÔNICA

Re-configurable architectures and applications

Design of a Reconfigurable SoC with RTOS for Animal and Vegetable Tracking

*Gilmar S. Beserra, Fabricio Schlag, Leticia T. Maia, Genival M. Araujo,
Joao V. Pimentel, Vitor G. A. Carneiro, Dino M. Amaral, Jose E. G. Medeiros,
Wagner A. Araujo, Janaina G. Guimaraes, Jose C. da Costa*

Universidade de Brasilia – Brasilia - Brasil

*Departamento de Engenharia Elétrica
C.P. 4386 - Brasília - DF – 70904-970 - BRASIL
Tel: 55-61-33072308
Fax: 55-61-32746651*

{gbeserra, schlag, letmaia, camargo}@unb.br

DESIGN OF A RECONFIGURABLE SOC WITH RTOS FOR ANIMAL AND VEGETABLE TRACKING

Abstract This paper presents the description of a rSoC (Reconfigurable System-on-Chip) that is being currently developed in our institution, aiming animal and vegetable tracking applications. The proposed rSoC contains a 32-bit RISC microprocessor, a memory-mapped FPGA, a RF transceiver and an APS (Active Pixel Sensor) matrix whose function will consist basically on image acquisition of animals and plants. A Real-Time Operating System (RTOS) will be adapted to the chosen architecture in order to allow the user to configure the system according to the application.

Keywords: rSoC, RTOS, APS, CMOS

Introduction

Pastoral and environmental applications have been offering a wide range to the development of new systems, especially for animal and vegetal tracking and control. Since the need to improve the quality of the meat is increasing in order to avoid diseases, it is crucial to have an efficient system to control and manage livestock raising. Similarly, by tracking food products at all stages of production, like storage, processing, and distribution, it is possible to prevent plague dissemination, helping consumers, farmers, and others. Some prototypes have been designed [1], but were usually implemented in System-on-Board technology, which leads to bigger and more expensive products. A good way to improve those systems is to implement all functionalities in a single chip.

A System-on-Chip (SoC) has many subsystems that are interconnected to perform specific functions for end users in a single chip. Despite the high complexity, there are many advantages, such as reduced size, low cost and power consumption, and increased performance. A Reconfigurable System-on-Chip (rSoC) is a SoC that provides a flexible medium to implement custom ASICs, with additional advantages like real-time response and adaptability (the project can be updated) [2].

Our institution has participated in the development of a wireless SoC dedicated to irrigation control on crops. The goal is to optimize water management while keeping humidity at an optimum level. The SoC has a RISC microprocessor, memory, an RF transceiver, digital and A/D interfaces. A full version

was implemented, complying with area and power consumption restrictions [3].

1. Design Methodology

Because of its high complexity, the rSoC design is divided into many steps, as it is suggested in [4]. Once the application is already defined, the next step is the system specification. The modules and their interconnections can be modeled and simulated using MATLAB functional blocks to describe the system behavior. This approach was used in [5] to describe the 16-bit RISC microprocessor of the previous SoC in a behavioral level, showing MATLAB as one more tool to validate mixed analog and digital design. The system structure can also be expressed using UML (Unified Modeling Language) [6].

A hardware/software co-design will be used to allocate tasks and define the interface functions. Since there will be analog and digital instances in the same chip, two different design flows will be adopted. The first one will be used to implement the analog and RF sections, and it consists of a DFT, full-custom design methodology, using both top-down and bottom-up approaches. The second one will be applied on the digital section design, which will be implemented using Electronic Design Automation (EDA) for synthesis, place and route, timing analysis and verification. In both cases, component reuse is an important key to gain productivity.

Finally, the system will be fabricated using CMOS technology, provided that it offers low cost, high density and low power. Considering the APS matrix and the RF transceiver as critical parts of the chip, $0.18 \mu m$ technology seems to be adequate because it allows a smaller area and higher frequency operation with a 1.8V power supply. It also supports the microprocessor clock needed to handle the real-time data acquisition.

2. System Description

The system architecture is described in the next subsections. The SoC design will be IP-based, adherent to VSIA standards and using high-performance bus architectures [7]. The estimated area of the complete chip is 70 mm^2 (including the pads), as shown in Table 1, and a 208-pin CQFP package will be used.

Microprocessor

A 32-bit RISC (Reduced Instruction Set Computer) architecture, with a dedicated multiplication hardware circuit, fixed-point arithmetic operations, mem-

Table 1. Estimated area

Component	Estimated Area (mm ²)
PA	320
Memory:	
- Instruction cache and Data cache	9
- Configuration cache	2
Microprocessor Core	2
Configurable Blocks	16
Interfaces	5
Core (total)	44
PADS	26
Chip	70

ory hierarchy and efficient pipeline was chosen because its fixed size instructions increase the system performance and decrease the datapath complexity.

Considering those characteristics, there are some core options available, but the choice will be limited to one of the following architectures. One of them is the IP-suite for SoC XiRisc microprocessor [8], a 32-bit RISC HDL model with configurable instruction set architecture and datapath. It has been implemented in some SoC architectures using 0.18 μm technology and tested with DSP algorithms [9]. The other one is the MOLEN polymorphic CCM (Custom Computing Machine) processor based on co-processor architectural paradigm that has been synthesized and prototyped into FPGA platform. Its concept has been validated using MPEG-2 application [10].

Memory

A static RAM will be used to store the microprocessor instructions and data because it does not need refresh and it works faster than dynamic RAM. In [9], a 64 KB instruction cache and a 64 KB data cache were implemented in 0.18 μm technology with a total area of 9mm², and the same structures can be used in the chip.

A ROM memory will be included in order to store boot and test routines. It is also possible to use an external flash memory to increase the storage size and allow bigger routines.

256x256 APS (Active Pixel Sensor) Matrix

This block is composed by APS sensors arranged in a 256x256 matrix, logic circuits for row and column selection, amplifiers, A/D converter to digitalize the image acquired by the sensors, and a buffer to store the image before it is sent to the microprocessor. Considering a 256-color video application with a

frame rate of 30 Hz, we would need a processing speed of nearly 2 MHz. A compression algorithm like MPEG can reduce the video size, but it increases the image processing.

As shown in Figure 1, a standard active pixel has a reset transistor (M1), a readout transistor (M2) and a row select transistor (M3). The photodiode (PD) has a better blue and green response than the photogate (PG), what makes it a better choice for color APS matrix [11]. In order to get color images, the integrated color pixel (ICP) arrangement is a new alternative to current pixel design that employs color filter arrays (CFA), as shown in [12].

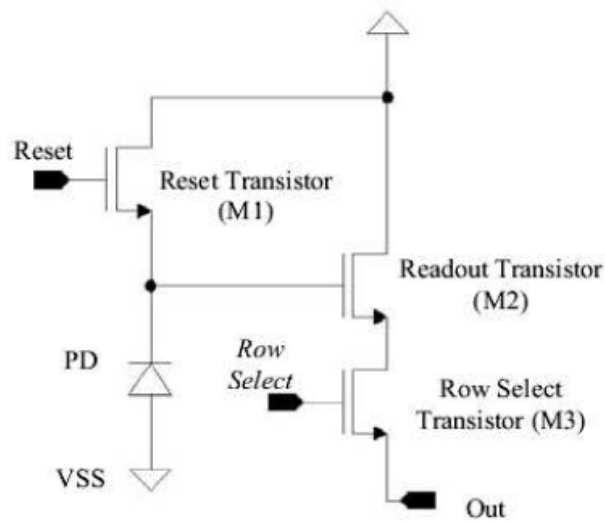


Figure 1. Standard Active Pixel

Memory-Mapped FPGA (Field Programmable Gate Array)

Since the hardware reconfigurability is one of the main characteristics of the system, an integrated memory mapped FPGA will be included, i. e., some memory addresses will be used to access the FPGA.

The FPGA can configure specific routines that usually demand successive calculations, like treatment, zoom, compression and edge recognition of the images acquired by the APS matrix, and also frequency selection of the RF transceptor. These routines will work as co-processors in the FPGA, which increases the processing speed and reduces the computational load on the main processor core. The parameters of each co-processor will be saved on the memory, and a specific hardware to select the co-processors and to access the memory in order to improve the parallel processing among the microprocessor and the co-processors must be implemented. Finally, the microprocessor must have several specific instructions to handle the FPGA configuration.

It is also possible to use a pipelined run-time configurable datapath (PiCo gate array, pGA or PiCoGA) as a configurable unit integrated in the processor core. It consists of an array of rows, each one representing a possible stage of a customized pipeline. This approach was used in [9], in which each PiCoGA row implemented in 0.18 μm technology occupies 0.6 mm². The average power was 16mW/row.

Integrated RF Transceiver

Each chip has an RF transceiver [3] which allows the communication among the units. During transmission, the processor selects the operating frequency using the Frequency Selection Register, so that the frequency originated from the External Oscillator is converted to the desired one. At the end of the Synthesizer, the Buffer switches the signal according to the microprocessor control, allowing the Power Amplifier (PA) to amplify the modulated signal to the power used in the transmission. During reception, the Low Noise Amplifier boosts the incoming signal before the action of the Demodulator, in which the signal is mixed with a single frequency signal, bringing it to the base-band. The received digital data is then sent to the microprocessor.

The available transceiver operates at ISM band (915-927,75 MHz) and uses OOK modulation. The transmission is half-duplex and uses TDM (Time Division Multiplex). The first stage of the receiver is a LNA (Low Noise Amplifier) with a sensitivity of -100 dB, gain of 30 dB and noise figure in 2 – 5 dB range. The PA (Power Amplifier) output stage has a maximum transmission power of 10 mW. The estimated transceiver area is 14 mm². A spread-spectrum transceiver operating at 2.4 GHz is being designed.

Analog and Digital Interfaces

In order to properly process the electric signals in the system, an analog-to-digital converter (ADC) will be used. The ADC will be a $\Sigma \Delta$ type, with a resolution of 10 bits, thus allowing for the conversion of each pixel defined by three sets (i.e., colors) of values in the 0 – 255 range, at a 1 μs conversion rate. The analog interface will consist of circuitry responsible for conditioning the analog signals to the AD conversion. The signal-to-noise ratio (SNR) in the analog circuitry will be kept higher than 60dB, following a rule of thumb of 6dB per bit of conversion in the ADC.

3. RTOS (Real-Time Operating System)

The operating system for a reconfigurable chip has the hardware performance optimization and lower cost as important purposes. It must offer concurrent execution and real-time response, and it must be flexible and fault-tolerant.

In addition to the flexibility provided by the hardware, the system reconfigurability will be enhanced by using RTOS [13]. A good RTOS alternative is eCos (Embedded Configurable Operating System), an open-source real-time embedded operating system ported to a variety of architectures. eCos has an open-source application named RedBoot, that is managed and restarted by the HAL (Hardware Abstraction Layer), a layer between hardware and software that allows portability to different architectures (16, 32 and 64-bit) by managing hardware devices.

eCos can be used in real-time applications and can be programmed using C language. Since it is configurable and has only a dedicated code to execute the system, its size varies from kilobytes to megabytes, depending on the application [14].

In eCos RTOS implementation, the software can be divided in two parts, allowing its reuse. One of them is a HdS (Hardware-dependent Software), and the other one is independent. They are separated by HAL [15]. Some HdS examples are codes for booting the system, context switching, and configuring and accessing hardware devices, such as MMU (Memory Management Unit), bus, interfaces and timers. The drivers to peripheral devices are mostly hardware-dependent, and because of that they are typically offered at HAL level. In short, HAL allows the concurrent design of the hardware and a minimal dedicated RTOS according to the embedded application needs [16].

The goal is to use eCos in the chip in order to allow the programmer to use a graphical tool to configure the system through hardware and software libraries and compilation packages, so that the operating system can be customized and adapted to the system.

4. Some Results and Discussion

A simulation was performed in a standard PC platform, as shown in Figure 2. The RedBoot application was used to start and verify the system. Considering that eCos can be transferred to the target platform using floppy disks, ROM, serial interface RS232 or Ethernet, a floppy disk was used in this case. With this procedure, it was possible to verify the available documentation and to prepare the environment to run eCos.

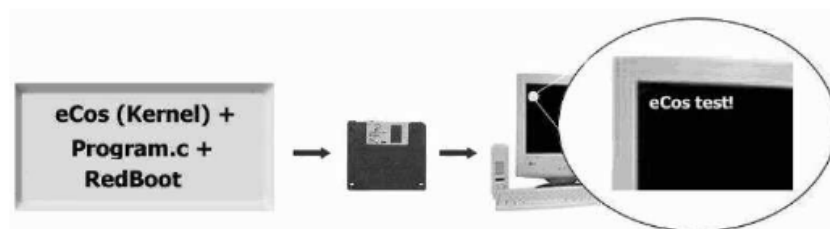


Figure 2. eCos simulation in a Pentium II 400 MHz

In order to finish the rSoC's specification, some applications have been taken into account. A possible one is livestock tracking, as shown in Figure 3. The movement of the animals can be detected by a sensor that will activate a RFID system. The RFID supplies a tag located on the body of the animal, and this tag sends information (e.g. identification) via a wireless link to the nearest rSoC. With many sensors properly positioned (e.g. near the water or salt supplying points), it is possible to monitor the animal's displacement and to analyze its dynamic behavior from the images acquired by the rSoC.

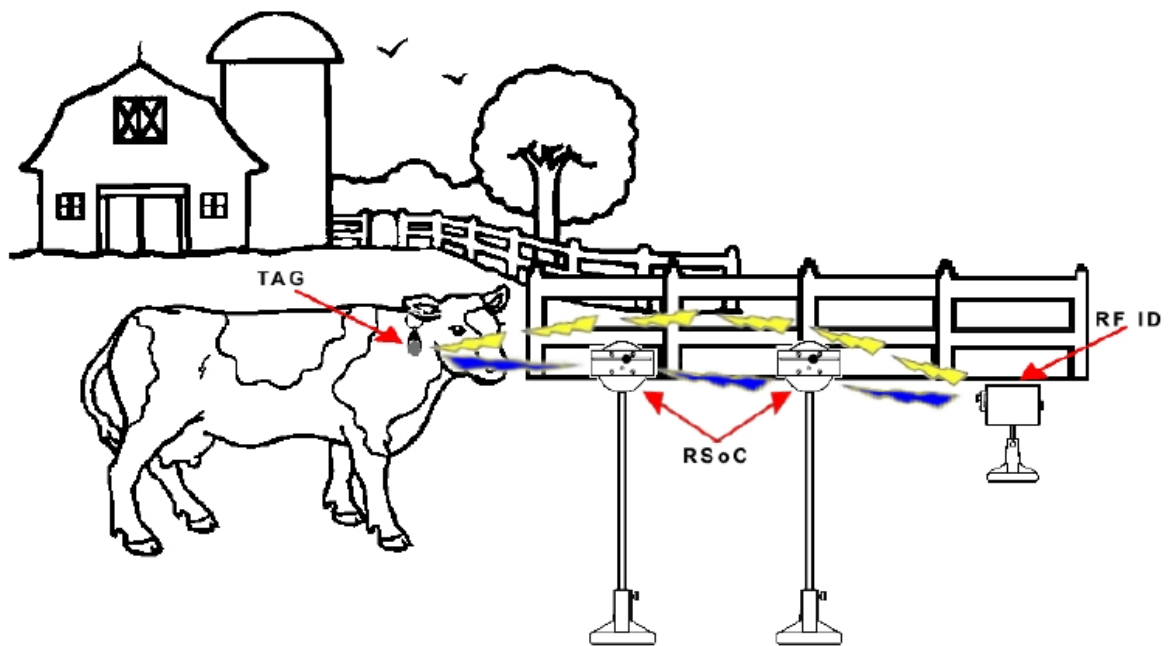


Figure 3. Animal tracking

Another application is fruit monitoring in commercial distribution networks. An rSoC could be used in conjunction with different sensors to assemble an artificial fruit that would be dispatched in the middle of fruit containers used to deliver these products to intermediate and end consumers. That autonomous device will collect environmental data as well as IR spectral responses from fruits, in order to evaluate the transport conditions and the fruit state during transportation.

Acknowledgments

The authors would like to acknowledge CNPq and PADCT/Millennium Institute (Brazilian government agencies) for financial support.

References

- [1] www.biochip.com.br

- [2] Bergmann, N.W., and Williams, J. (2003). "The Egret platform for reconfigurable system on chip," Proceedings of the 2003 IEEE International Conference on Field-Programmable Technology 340 - 343.
- [3] Ommited for blind review.
- [4] Bergmann, N. W. (2002). "Enabling Technologies for Reconfigurable System-on-Chip," Proceedings of the IEEE International Conference on Field-Programmable Technology 360 - 363.
- [5] Ommited for blind review.
- [6] Voros, N. S., Tsasakov, S., and Mariatos, V. (1998). "System-on-a-chip Methodology for Telecom Applications," Proceedings of the URSI IEEE International Symposium on Signals, Systems, and Electronics (ISSSE 98) 321 - 325.
- [7] <http://www.eetimes.com/story/OEG20000515S0026>
- [8] <http://xirisc.deis.unibo.it/>
- [9] Lodi A., Toma M., Campi F., Cappelli A., Canegallo R., and Guerrieri R. (2003). "A VLIW Processor With Reconfigurable Instruction Set for Embedded Applications," IEEE Journal of Solid-State Circuits 38:1876 - 1886.
- [10] <http://ce.et.tudelft.nl/MOLEN/>
- [11] Djaja, S., Chapman, G. H., Cheung, D. Y. H., and Audet, Y. (2003). "Implementation and Testing of Fault-Tolerant Photodiode-based Active Pixel Sensor (APS)," Proceedings of the 18th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems 53 - 60.
- [12] Catrysse, P. B., and Wandell, D. A. (2003). "Integrated Color Pixels in 0.18um Complementary Metal Oxide Semiconductor," J. Opt. Soc. Am. A 20:2293 - 2306.
- [13] <http://doi.ieeecomputersociety.org/10.1109/IPDPS.2003.1213320>
- [14] <http://ecos.sourceforge.org>
- [15] Yoo, S., and Jerraya, A. A. (2003). "Introduction to Hardware Abstraction Layers for SoC.," Proceedings of the Design, Automation and Test in Europe Conference and Exhibition (DATE03) 10336 - 10337.
- [16] <http://www.eetimes.com/story/OEG20020301S0104>