

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE  
Centro de Engenharia Elétrica e Informática  
Programa de Pós-Graduação em Ciência da Computação

Dissertação de Mestrado

Utilização de Aritmética Bit-serial para Redução de Consumo de  
Energia

Roberto Medeiros de Faria

Orientador

Elmar Uwe Kurt Melcher

Campina Grande – PB

Dezembro – 2014

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE  
Centro de Engenharia Elétrica e Informática  
Programa de Pós-Graduação em Ciência da Computação

## **Utilização de Aritmética Bit-serial para Redução de Consumo de Energia**

**Roberto Medeiros de Faria**

Dissertação submetida à Coordenação do Curso de Pós-graduação em Informática da Universidade Federal de Campina Grande, como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Orientador

Elmar Uwe Kurt Melcher

**Área de concentração:** Ciência da Computação.

**Linha de pesquisa:** Redes de Computadores e Sistemas Distribuídos.

Campina Grande – PB

Dezembro – 2014

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

F224u	<p>Faria, Roberto Medeiros de. Utilização de aritmética bit-serial para redução de consumo de energia / Roberto Medeiros de Faria. – Campina Grande, 2014. 82 f. : il. color.</p> <p>Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2014.</p> <p>"Orientação: Prof. Elmar Uwe Kurt Melcher". Referências.</p> <p>1. Redução de Consumo de Energia. 2. Aritmética Bit-Serial. 3. Processamento Digital de Sinais. 4. Standard Cells. 5. IP-Core. I. Melcher, Elmar Uwe Kurt. II. Título.</p> <p style="text-align: right;">CDU 004.031.6 (043)</p>
-------	---

**"UTILIZAÇÃO DE ARITMÉTICA BIT-SERIAL PARA REDUÇÃO DE CONSUMO DE ENERGIA"**

**ROBERTO MEDEIROS DE FARIA**

**DISSERTAÇÃO APROVADA EM 10/12/2014**



**ELMAR UWE KURT MELCHER, Dr., UFCG**  
**Orientador(a)**



**JOSÉ ANTÔNIO GOMES DE LIMA, D.Sc, UFPB**  
**Examinador(a)**



**ANDREY ELÍSIO MONTEIRO BRITO, Dr., UFCG**  
**Examinador(a)**

**CAMPINA GRANDE - PB**

## Agradecimentos

Agradeço:

a **Deus** pelo dom da vida e por tudo que dela recebi;

aos meus pais, **Edson** e **Nízia**, por terem sido coadjuvantes nessa tarefa de Deus, serem a base do meu caráter e por me mostrarem sempre o caminho certo;

às minhas três segundas-mães, minhas tias **Neuza**, **Nilza** e **Maria da Piedade**, que, com seu amor, sempre tiveram uma presença especial, significativa e construtiva na minha vida;

à minha esposa **Lúcia**, pela paciência, dedicação e amor, que sempre dispensou a mim, e, em especial, nesse período de trabalho;

às minhas filhas **Aline** e **Andréa**, razão maior para eu querer ir mais longe;

à **minha família** que, sempre unida, é o meu porto seguro e meu mar da tranquilidade;

ao professor e amigo **Elmar** pelas orientações com segurança e paciência para que eu chegasse até aqui;

à professora e amiga **Joseana** pela presteza e boa vontade nos ensinamentos e esclarecimentos de dúvidas;

em especial, à professora e amiga **Joseluce**, por ter, voluntariamente e generosamente, suportado a minha carga de trabalho na reta final e pelo incentivo durante toda caminhada;

ao pessoal do **LAD** pelo companheirismo e bons momentos na caminhada;

aos que fazem a **COPIN** pela dedicação e ajuda a nós alunos;

aos que fazem o **DSC** pelo companheirismo e especialmente àqueles que tiveram sua carga de trabalho aumentada para que eu me afastasse para capacitação;

aos **colegas** de curso pela partilha de conhecimentos e bons momentos; e

a **todos** que direta ou indiretamente contribuíram, oraram ou torceram, para esta realização.

## Resumo

Hoje, uma das maiores preocupações, senão a maior, da indústria de semicondutores é o desenvolvimento de chips com baixo consumo de energia. Existem vários fenômenos físicos causadores de consumo de energia em circuitos CMOS e várias técnicas que reduzem o consumo de energia de um *chip*. O objetivo principal desta pesquisa de mestrado foi investigar o quanto o consumo de energia estática em circuitos CMOS pode ser reduzido por meio do emprego de aritmética bit-serial em substituição à aritmética bit-paralela. A pesquisa está focada em circuitos construídos a partir de *standard cells* (células padrão), com aplicação em processamento de sinais, e para os quais o principal requisito não é o alto desempenho computacional, mas o baixo consumo de energia. A metodologia foi aplicada em um estudo de caso, utilizando-se para isto, simulações com o IP core SPVR. O SPVR é um verificador de identidade vocal implementado em um circuito dedicado capaz de ter desempenho suficiente para funcionar em tempo real, mesmo empregando um sinal de *clock* lento. Foi constatado na pesquisa, que o uso de aritmética bit-serial, em termos de diminuição de consumo estático, é vantajoso para somadores e circuitos de pequena complexidade. Porém, para sistemas de maior complexidade, esta substituição só é vantajosa em situações específicas de grande número de operações aritméticas e baixo uso de armazenamento em registradores paralelos. No caso inverso, as vantagens se perdem, porque embora haja diminuição de consumo estático, há um crescimento muito grande de consumo dinâmico.

## Abstract

Today, one of the biggest concerns, if not the largest, for the semiconductor industry is the development of chips with low power consumption. There are several physical phenomena that cause power consumption in CMOS circuits and various techniques that reduce the energy consumption of a chip. The main objective of this masters research was to investigate how the static power consumption in CMOS circuits can be reduced through the use of bit-serial arithmetic in place of bit-parallel arithmetic. The research is focused on circuits built from standard cells, with application to signal processing, and for which the main requirement is not the high computing performance, but the low power consumption. The methodology was applied in a case study, using for this, simulations with the SPVR IP core. The SPVR is a vocal identity checker implemented in a dedicated circuit able to have enough performance to run in real time, even employing a slow clock signal. It has been found in research that the use of bit-serial arithmetic, in terms of reduction of static consumption, is advantageous to adders and small circuit complexity. However, for more complex systems, this substitution is only advantageous in specific situations of large number of arithmetic operations and low storage usage in parallel registers. In the reverse case, the advantages are lost, because although there are static consumption decrease, there is a very large dynamic consumption growth.

# Sumário

1	Introdução.....	15
1.1	Estrutura do Documento.....	16
2	Fundamentação Teórica.....	18
2.1	Tecnologia CMOS.....	18
2.2	Tipos de Fonte de Consumo de Energia em Circuitos CMOS.....	19
2.3	Consumo de Energia.....	19
2.3.1	Consumo Dinâmico.....	20
2.3.2	Consumo Estático.....	20
2.3.3	Tendências de Consumo.....	21
2.3.4	Técnicas para Redução de Consumo.....	23
2.3.5	Fatores de Impacto da Redução do Consumo.....	25
2.4	Aritméticas Bit-paralela e Bit-serial.....	25
3	Objetivos da Pesquisa.....	29
3.1	Objetivo Geral.....	29
3.2	Objetivos Específicos.....	29
4	Pesquisas Relacionadas.....	31
4.1	Técnicas para Redução de Consumo de Energia de um Circuito CMOS....	31
4.1.1	Técnicas Low Power Aplicáveis no Nível de Transistores.....	31
4.1.2	Técnicas Low Power Aplicáveis no Nível de Circuito.....	33
4.1.3	Técnicas Low Power Aplicáveis no Nível de Sistema.....	37
4.2	Resumo das Pesquisas Relacionadas.....	39
5	Metodologia da Pesquisa.....	43
5.1	Caminhos da Pesquisa.....	43



6 Apresentação e Análise dos Resultados.....	50
6.1 Atividade A – Recuperação dos Dados Obtidos da Substituição de Aritmética Bit-paralela por Bit-serial no Nível de Transistores.....	50
Atividade B – Redução de Consumo com Uso de Aritmética Bit-serial em Projetos Baseados em Standard Cells.....	51
6.2 Atividade F – Uso de Aritmética Bit-serial para Redução de Consumo com Tecnologia Submicrônica vs. Nanométricas.....	56
6.3 Atividade H – Aplicação de Aritmética Bit-serial para Redução de Consumo em um Sistema de Maior Complexidade.....	60
6.3.1 O Handshake dos Módulos Seriais do SPVR.....	60
6.3.2 A Serialização dos Módulos com a Aritmética Bit-serial.....	60
6.3.2.1 O Módulo Pré-ênfase (PE).....	63
6.3.2.2 O Módulo de Janelamento (WIN).....	66
6.3.2.3 O Módulo Análise de Codificação Preditiva Linear (LPCA).....	67
6.3.3 Análise dos Resultados.....	73
7 Considerações Finais.....	75
7.1 Contribuições da Pesquisa.....	75
7.2 Sugestões para Pesquisas.....	76
8 Referências Bibliográficas.....	77

## Lista de Figuras

Figura 2.1 – Tendência do consumo de energia dos circuitos CMOS

Figura 2.2 – Popularidade das técnicas de *Low Power*

Figura 2.3 – Fator de impacto das decisões sobre a redução do consumo

Figura 2.4 – Somador bit-paralelo

Figura 2.5 – Somador bit-serial.

Figura 5.1 – Arquitetura do SPVR

Figura 5.2 – Caminhos da Pesquisa

Figura 6.1 – Consumo do Modelo de Transistores com Palavras de 12 bits

Figura 6.2 – Consumo do Modelo de Transistores com Palavras de 24 bits

Figura 6.3 – Comparação dos resultados com palavras de 12 bits

Figura 6.4 – Comparação dos resultados com palavras de 24 bits

Figura 6.5 – Consumo do Modelo de *Standars Cells* com tecnologia 130 nm vs. 28 nm com Palavras de 12 bits

Figura 6.6 – Consumo do Modelo de *Standars Cells* com tecnologia 130 nm vs. 28 nm com Palavras de 24 bits

Figura 6.7 – Fluxo de Dados e Operações para o Cálculo da Autocorrelação

Figura 6.8 – Fluxo de Dados e Operações para o Cálculo dos Coeficientes LPCs

Figura 6.9 – Blocos de Operações Repetidas no Cálculo dos Coeficientes LPCs

## Lista de Tabelas

Tabela 4.1 – Técnicas *Low Power* aplicáveis no nível de transistor

Tabela 4.2 – Técnicas *Low Power* aplicáveis no nível de circuito

Tabela 4.3 – Técnicas *Low Power* aplicáveis no nível de sistema

Tabela 6.1 – Consumo do Modelo de Transistores com palavras de 12 bits

Tabela 6.2 – Consumo do Modelo de Transistores com palavras de 24 bits

Tabela 6.3 – Consumo do Modelo de Transistores 130 nm vs. *Standard Cells* 130 nm com palavras de 12 bits

Tabela 6.4 – Consumo do Modelo de Transistores 130 nm vs. *Standard Cells* 130 nm com palavras de 24 bits

Tabela 6.5 – Consumo do Modelo de *Standars Cells* com tecnologia 130 nm vs. 28 nm com Palavras de 12 bits

Tabela 6.6 – Consumo do Modelo de *Standars Cells* com tecnologia 130 nm vs. 28 nm com Palavras de 24 bits

Tabela 6.7 – Consumo do PE (com saturação)

Tabela 6.8 – Consumo do PE (sem saturação)

Tabela 6.9 – Consumo do WIN

Tabela 6.10 – Consumo do LPCA (com um *clock*)

Tabela 6.11 – Consumo do LPCA (com dois *clocks*)

## Lista de Quadros

Quadro 6.1 – Pseudo-código para o Cálculo dos Coeficientes Autocorrelacionados

Quadro 6.2 – Pseudo-código para o Cálculo dos Coeficientes LPCs

## Lista de Abreviaturas

AMD	Advanced Micro Devices, Inc.
ASIC	Application-Specific Integrated Circuits
Brazil-IP	Brazil-Intellectual Property
CI	Circuito Integrado
CMOS	Complementary Metal-Oxide-Semiconductor
CMP	Circuits Multi-Projets
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DC	Direct Current
DM	Decision Maker
DSTN	Distributed Sleep Transistors Network
EDA	Electronic Design Automation
ESL	Electronic System-Level
FA	Full-Adder
GPU	Graphics Processing Unit
GND	Ground
IEEE-SA	Institute of Electrical and Electronics Engineers Standard Association
IP core	Intellectual Property core
$I_{sub}$	Subthreshold Current
ITRS	International Technology Roadmap for Semiconductors
IVC	Input Vector Control
LAD	Laberatório de Arquiteturas Dedicadas
LPC	Linear Prediction Coding
LPCA	Linear Prediction Coding Analysis
LSB	Least Significant Bit
MLV	Minimum Leakage Vector
MSB	Most Significant Bit
MTCMOS	Multi Threshold CMOS
nMOS	N-type Metal-Oxide-Semiconductor
NoC	Network-on-Chip

PE	Preemphasis
PFI	Power Forward Initiative
PM	Pattern Matching
pMOS	P-type Metal-Oxide-Semiconductor
RTL	Register Transfer Language
SCCMOS	Super Cutoff CMOS
SoC	System-on-Chip
SPICE	Simulation Program with Integrated Circuit Emphasis
SPVR	Speaker Verification System
DVFS	Dynamic Voltage and Frequency Scaling
TDP	Thermal Design Power
TLM	Transaction-Level Modeling
UFCG	Universidade Federal de Campina Grande
UMC	United Microelectronics Corporation
VD	Voice Detector
$V_{dd}$	Power Supply/Drain Voltage
$V_t$	Threshold Voltage
$V_{th}$	Threshold Voltage
WIN	Windowing
XOR	Exclusive OR
ZPG	Zigzag Power-Gated

## 1 Introdução

Os circuitos integrados (CIs) ou *chips* são amplamente utilizados, nos dias de hoje, em dispositivos eletrônicos de telecomunicações, de informática, na indústria automotiva, etc. As indústrias de produtos eletrônicos, nos últimos dez anos, têm utilizado dispositivos conhecidos como SoC (*System-on-Chip*), que são sistemas complexos colocados em um único *chip*. Os SoCs, atualmente, são utilizados em todos os dispositivos portáteis do tipo telefones celulares, câmeras fotográficas, filmadoras digitais, entre outros.

Os SoCs são compostos por IP *cores*. Um IP *core* (núcleo de Propriedade Intelectual) é um módulo de circuito digital desenvolvido para executar uma tarefa específica em um determinado sistema. Os IP *cores* são projetados para serem componentes programáveis (processadores) ou não, tais como os dispositivos de entrada e saída. Em um SoC, os IP *cores* são interligados para se comunicarem entre si, por estruturas de comunicação, como por exemplo, um barramento compartilhado ou uma NoC (*Network-on-Chip*) (KEATING *et al.*, 2007).

Hoje, consegue-se colocar mais de 7 bilhões de transistores num CI. A GPU (*Graphics Processing Unit*) GTX110 das placas de vídeo GeForce GTX TITAN e GeForce GTX 780, ambas da NVIDIA®, é um exemplo em que tal fato ocorre (CHIAPPETTA, 2013; KIRSCH, 2013). Com este nível de integração em pequenos dispositivos eletrônicos, deve-se considerar dois fatores: o aumento da complexidade no fluxo de desenvolvimento e o aumento do consumo energético dos circuitos CMOS (*complementary metal-oxide-semiconductor*), devido ao grande número de funcionalidades que os dispositivos apresentam. As TDPs (*Thermal Design Power*) da NVIDIA® GeForce GTX TITAN e da GTX 780, de 250 W cada, são exemplos. O fato de que a placa AMD™ HD 7970, cuja GPU, denominada Thaiti, contém 4,3 bilhões de transistores e possui o mesmo valor de TDP (SMITH, 2013), é um indicador de que o consumo de energia seria o principal fator limitante para a complexidade de circuitos integrados. Em consequência desse fator, houve nos últimos anos significativas mudanças no processo de desenvolvimento de *chips* (KEATING *et al.*, 2007).

Hoje, uma das principais preocupações, senão a maior, da indústria de semicondutores é o desenvolvimento de *chips* com baixo consumo de energia (PFI, 2009). Existem vários fenômenos físicos causadores de consumo de energia em

circuitos CMOS e várias técnicas que reduzem o consumo de energia, chamadas de *Low Power*. Estas técnicas podem ser aplicadas nas diversas fases do fluxo de desenvolvimento de um *chip*.

Nesta pesquisa, se fez uso de uma técnica para a redução de consumo de energia em circuitos CMOS, com ênfase na redução do consumo de energia estática, que faz a substituição de operadores de aritmética bit-paralela por operadores de aritmética bit-serial, que foi utilizada em vários trabalhos publicados (NILSSON; TORKELSON, 1997; ÅSTRÖM; NILSSON; TORKELSON, 1999, NILSSON, 2006, 2007, 2008, 2009, 2010).

Inicialmente, foi aplicada a técnica de substituição nos somadores trabalhados por Peter Nilsson no artigo "*A Methodology for Arithmetic Reduction of the Static Power Consumption Verified on Filter Architectures*" (Nilsson, 2008), porém, desta feita, utilizando o modelo no nível de *standard cell* (células padrão), com a mesma tecnologia (130 nm). Esperava-se obter resultados de redução de consumo de energia estática, semelhantes aos apresentados no artigo, que utilizou um modelo no nível de transístores. Em seguida, a técnica foi aplicada aos mesmos somadores com uma tecnologia abaixo dos 50 nm (28 nm), como sugeria o artigo citado, a fim de verificar se haveria uma redução maior de consumo de energia estática. Por último, foi verificado a redução de consumo aplicando-se a técnica em um sistema mais complexo. O sistema no qual aplicou-se a técnica foi o SPVR (*Speaker Verification System*) – um de verificador de identidade vocal. Todos os resultados foram comparados para verificação da validade da aplicação desta técnica para redução de consumo de energia estática.

## **1.1 Estrutura do Documento**

As demais seções desta dissertação estão estruturadas da seguinte maneira: na Seção 2 é apresentada a fundamentação teórica da pesquisa; na Seção 3 são apresentados os objetivos da pesquisa, geral e específicos; na Seção 4 é feito um estudo das pesquisas relacionadas com esta, em que é feita a apresentação das técnicas de *Low Power*, de interesse da pesquisa, agrupadas pelos níveis de abstração pelos quais passam o fluxo de projeto de um circuito; a metodologia utilizada para a realização dessa dissertação de mestrado é apresentada na Seção 5; na Seção 6 é feita apresentação e análise dos resultados, organizadas por atividade, tal qual foi



definida na metodologia; as considerações finais e sugestões para novas pesquisas aparecem na Seção 7; e, por último, na Seção 8, são listadas as referências bibliográficas utilizadas na pesquisa.

## 2 Fundamentação Teórica

### 2.1 Tecnologia CMOS

Os transistores CMOS (*Complementary Metal-Oxide-Semiconductor*) são os componentes elementares básicos de quase todos os circuitos digitais. Transistores são chaves controladas eletricamente que ligam ou desligam quando uma tensão ou corrente é aplicada a um terminal de controle (HARRIS; HARRIS, 2007). Os computadores modernos usam transistores porque eles são baratos, pequenos e confiáveis.

A tecnologia CMOS, que é predominante no processo de fabricação de um SoC (*System-on-Chip*) e de um ASIC (*Application-Specific Integrated Circuits*), utiliza dois tipos de transistores: nMOS e pMOS. Uma porta lógica CMOS corresponde a uma topologia de porta lógica formada por duas redes lógicas independentes e complementares, formadas por estes dois tipos de transistores (MATTOS; ROSA JR; PILLA, 2009).

Um transistor CMOS é fabricado sobre uma base de silício, chamada de substrato. No transistor nMOS a região da base é um substrato de silício tipo "p" e, no pMOS, a base é um substrato tipo "n". Duas regiões fortemente dopadas são adicionadas ao substrato, com os tipos invertidos, isto é, para o substrato "p" a dopagem é do tipo "n" e vice-versa, originando o *drain* (dreno) e o *source* (fonte). Uma camada muito fina de dióxido de silício isolante é depositada sobre a região do substrato entre o *drain* e o *source*. Sobre a camada isolante é depositada uma camada de polissilício condutor, que é denominada de *gate* (porta).

O funcionamento do transistor depende da aplicação de um potencial elétrico na *gate*. Para um transistor nMOS, caso seja aplicada uma tensão zero na *gate*, o dispositivo se comportará como um circuito aberto, resultando numa alta impedância entre o *drain* e o *source*. Caso seja aplicada na *gate* uma tensão positiva superior à tensão de *threshold*  $V_{th}$  (tensão de limiar), será induzida a criação de um canal condutor ligando a região do *drain* com a região do *source*, passando, o dispositivo, a se comportar como um circuito fechado. Para um transistor pMOS, o funcionamento é ao contrário, ou seja, o circuito é fechado quando se aplica uma tensão zero na *gate*.

## 2.2 Tipos de Fonte de Consumo de Energia em Circuitos CMOS

Existem quatro tipos de fontes de consumo de energia num circuito CMOS. Em função da tecnologia de fabricação, estas fontes podem ser mais ou menos significativas no cálculo do consumo de energia total de um circuito CMOS (PEDRAM; VAISHNAV, 1997; JOSHI; BHAVSAR, 2012; JONVANOVIĆ; JEVTIĆ, 2011; SHAULY, 2012).

A seguir, são listados os quatro tipos de fontes de consumo de energia em um circuito CMOS (PEDRAM; VAISHNAV, 1997).

- corrente de fuga (*leakage current*), que é causada por dois componentes: 1) a corrente nos díodos parasitas (*parasitic diodes*) de polarização inversa, formada entre as difusões do *source* (fonte) e do *drain* (dreno) e a região de substrato do transistor CMOS; e 2) a corrente *subthreshold* (sub-limiar), que decorre da inversão da carga que existe nas tensões da *gate* (porta), abaixo da tensão de *threshold* (limiar). A corrente de fuga é determinada principalmente pela tecnologia de fabricação;
- corrente de tunelamento, que atravessa o isolamento da *gate*;
- corrente de curto-circuito (*short-circuit current*), que é devido ao caminho DC entre as trilhas de alimentação durante as transições de saída;
- corrente de capacitância (*capacitance current*), que flui para carregar e descarregar as cargas capacitivas durante as mudanças (chaveamento – *switching*) de estados lógicos.

## 2.3 Consumo de Energia

O consumo de energia elétrica total de um circuito CMOS é uma função da atividade de transição dos estados (ligado e desligado), da capacitância, da tensão de funcionamento e da tecnologia de fabricação do transistor em si. Existem dois tipos de consumo de energia em um circuito CMOS: o estático e o dinâmico. A potência total  $P_t$  é o somatório da potência dinâmica  $P_d$  e da estática  $P_e$ , conforme a Equação 2.1 (KEATING *et al.*, 2007).

$$P_t = P_d + P_e \quad (2.1)$$

### 2.3.1 Consumo Dinâmico

A consumo dinâmico é decorrente da soma de dois fatores: da corrente de capacitância (*capacitance current*), durante o chaveamento (*switching*), e da corrente de curto-circuito (*short-circuit current*). A energia de chaveamento é dissipada durante a carga ou descarga das capacitâncias internas do transistor e das capacitâncias dos fios de interconexão. A potência de curto-circuito é a energia dissipada por um curto-circuito, que ocorre durante um curto período de tempo, na conexão entre o *drain* e o *source*, durante o chaveamento entre os estados do transistor (HENZLER, 2007).

A Potência de chaveamento é calculada pela Equação 2.2 (CHABINI, 2007):

$$P_d \approx AFC_{ef}V_{dd}^2, \quad (2.2)$$

em que  $A$  é o fator de atividade de chaveamento,  $F$  é a frequência de *clock*,  $C_{ef}$  é a capacitância efetiva e  $V_{dd}$  é a tensão de alimentação. O produto  $AF$  corresponde à frequência média de chaveamento e o produto  $C_{ef}V_{dd}^2$  corresponde à energia dissipada para cada chaveamento.

### 2.3.2 Consumo Estático

O consumo estático corresponde à energia dissipada quando o circuito não está chaveando, ou seja, quando o circuito está inativo. Existem diversas fontes de dissipação do consumo estático, porém, a maior parte é resultante da fuga (*leakage*) de corrente entre o *drain* e o *source*, corrente de *subthreshold*, e por meio do dióxido de silício (*gate oxide leakage*). Note-se que, ao contrário da corrente de *subthreshold*, que ocorre quando o transistor não está conduzindo, a corrente por meio do dióxido de silício existe decorrente de uma diferença de potencial, ou seja, ocorre independentemente do estado do transistor (MATTOS; ROSA JR; PILLA, 2009). Na Seção 2.2, já foram listadas as correntes de fuga em um circuito CMOS.

A potência estática  $P_e$  é calculada pela Equação 2.3 (BUTZEN et al., 2010):

$$P_e = V_{dd}I_{leakage}, \quad (2.3)$$

em que  $I_{leakage}$  é o somatório das correntes de fuga e  $V_{dd}$  é a tensão de

alimentação.

A corrente de fuga total pode ser aproximada pela soma das correntes de *subthreshold* e *gate oxide leakage*. Estas correntes são calculadas pelas Equações 2.4 e 2.5, a seguir (BUTZEN et al., 2010):

$$I_s = I_0 e^{V_{gs} - V_{th} / nV_T} [1 - e^{-(V_{ds} / V_T)}], \quad (2.4)$$

$$I_g = WLA \left( \frac{V_{ox}}{T_{ox}} \right)^2 \exp\left( \frac{-B(1 - (1 - (V_{ox} / \phi_{ox}))^{3/2})}{V_{ox} / t_{ox}} \right), \quad (2.5)$$

em que  $I_s$  = corrente de subthreshold,  $I_g$  = corrente *gate oxide leakage*,

$I_0 = (W \mu_0 C_{ox} V_T^2 e^{1.8}) / L$ ,  $V_t$  = tensão térmica,  $V_{th}$  = tensão de *threshold*,  $V_{ds}$  = tensão *drain-source*,  $V_{gs}$  = tensão *gate-source*,  $W$  = largura do transistor,  $L$  = comprimento do transistor,  $C_{ox}$  = capacitância do óxido da *gate*,  $\mu_0$  = mobilidade de portadores de carga,  $n$  = *subthreshold swing coefficient*,  $V_{ox}$  = potencial por meio do óxido,  $t_{ox}$  = espessura do óxido,  $A = q^3 / 16 \pi^2 h \phi_{ox}$ ,  $B = 4 \pi \sqrt{2 m_{ox}} \phi_{ox}^{3/2} / 3 h q$ ,  $m_{ox}$  = massa efetiva da partícula em tunelamento,  $\phi_{ox}$  = barreira de potencial para o tunelamento,  $h$  = Constante de Planck e  $q$  = carga do elétron.

### 2.3.3 Tendências de Consumo

Com a evolução da tecnologia CMOS, a tensão de  $V_{dd}$  vem sendo reduzida para diminuir a dissipação da potência estática e a tensão de *threshold* vem sendo reduzida para garantir a eficiência de velocidade de comutação dos transistores. Conseqüentemente, de acordo com as equações apresentadas, diminuindo a tensão  $V_{th}$ , as correntes de fuga crescem exponencialmente, tornando-se uma das principais preocupações nos projetos de circuitos CMOS em escala nanométrica (BUTZEN et al., 2008).

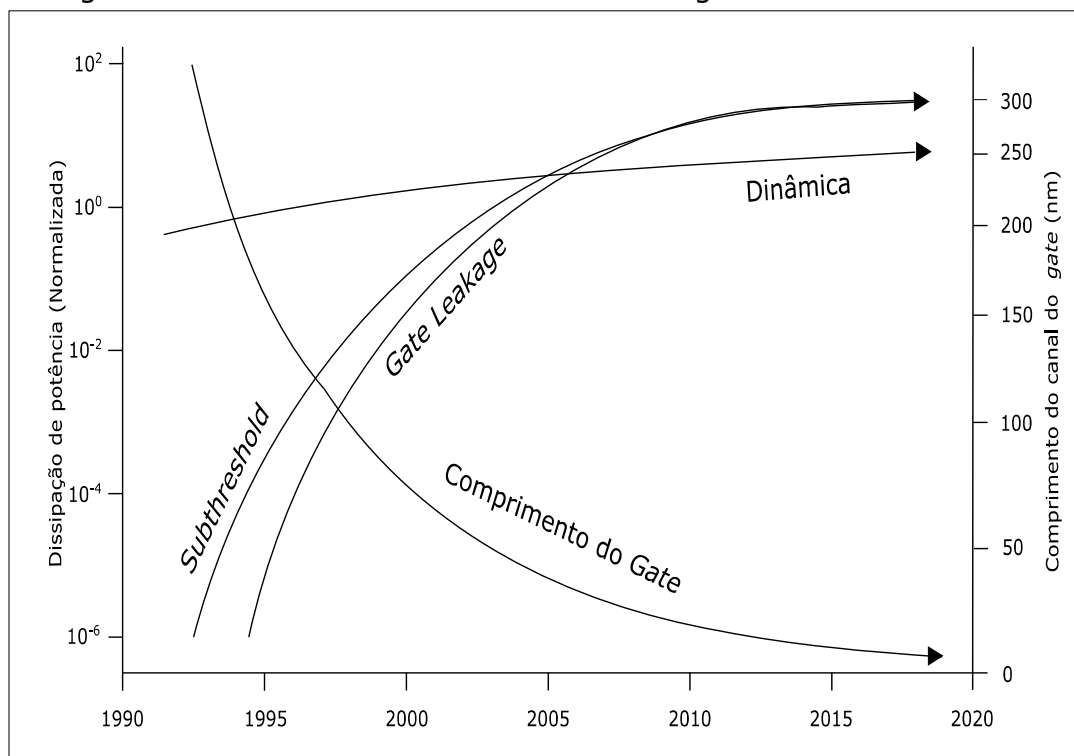
Nas tecnologias de fabricação acima de 90 nm, o consumo de energia dinâmica é uma importante restrição no projeto de circuitos CMOS (IEEE-SA, 2009). Para reduzir o consumo de energia dinâmica, técnicas baseadas no gerenciamento do sinal de *clock*, que normalmente são automatizadas por meio de ferramentas EDA (*Electronic Design Automation*), tornaram-se comuns (PFI, 2009). Com tecnologias de processo inferiores a 90 nm, o consumo de energia estática vem a ser uma restrição

dominante no projeto, pois, mesmo quando o circuito está inativo, ocorre um consumo considerável devido às correntes de fuga (IEEE-SA, 2009). Com as novas nanotecnologias de fabricação dos circuitos CMOS, a relação entre as energias dinâmica e estática teve uma considerável diminuição. Em pesquisas já antigas, pesquisadores têm relatado que em tecnologias de fabricação menores que 90 nm, o consumo da energia estática se aproxima, e normalmente supera, o da dinâmica (IEEE-SA, 2009; JADCHERLA *et al.*, 2009; KAWA, 2008).

A principal fonte de informações sobre o estado atual da tecnologia de semicondutores em geral e, em particular da nanotecnologia CMOS, é o ITRS (*International Technology Roadmap for Semiconductors*) (ITRS, 2014). O ITRS é um consórcio internacional composto por grandes fabricantes de semicondutores, fornecedores de equipamentos e instituições acadêmicas. Este consórcio tem por objetivo identificar os principais desafios tecnológicos da indústria de semicondutores para os próximos anos, por meio de um relatório bianual, contendo estimativas dos principais desafios da indústria de semicondutores (MOHANTY *et al.*, 2008).

Na Figura 2.1 é apresentado um gráfico, em escala logarítmica, com a previsão da relação entre o consumo da energia estática e a dinâmica nos circuitos

Figura 2.1 - Tendência do consumo de energia dos circuitos CMOS.



Fonte: Figura adaptada de (MOHANTY *et al.*, 2008).

CMOS, com base nos dados da ITRS. O gráfico apresenta as principais fontes de dissipação de energia em circuitos integrados CMOS, a partir do qual é possível observar que as correntes de *subthreshold* e *gate leakage*, hoje, já são dominantes no somatório da energia total.

Acompanhando a tecnologia CMOS, novas técnicas foram desenvolvidas para gerenciar o consumo de energia estática, por exemplo, técnicas de desligamento automático para um submódulo do circuito ou técnicas com base na variação da tensão de *threshold*  $V_{th}$  dos transistores. Quando se desliga um submódulo de um circuito, conforme a Equação 2.3, a tensão de alimentação  $V_{dd}$  será igual a zero e, por conseguinte, a potência estática dissipada pelo submódulo também será igual a zero. A Equação 2.4 mostra que, para as técnicas com base na tensão de *threshold*  $V_{th}$  dos transistores, a corrente de *subthreshold* é inversamente proporcional à tensão  $V_{th}$ . Portanto, quanto maior for a tensão de  $V_{th}$ , menor será a corrente de *subthreshold*, o que resulta numa menor velocidade de chaveamento dos estados dos transistores (GIRARD; NICOLICI; WEN, 2009), e o desafio, é encontrar um valor para  $V_{th}$  que não comprometa a velocidade de chaveamento e reduza a dissipação de potência estática dos transistores.

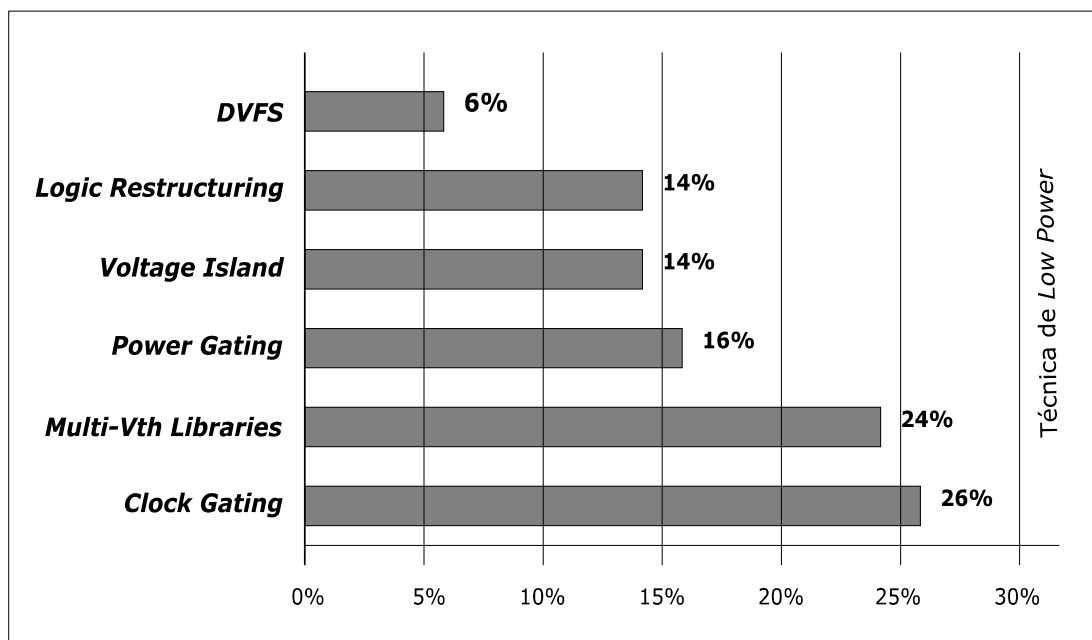
### 2.3.4 Técnicas para Redução de Consumo

Para reduzir o consumo de energia dos circuitos CMOS existem, e continuam sendo desenvolvidas, várias técnicas chamadas de técnicas de *Low Power* (KEATING *et al.*, 2007; SREENIVASULU *et al.*, 2012). Dessas técnicas, algumas são fáceis de implementar, outras possuem um alto impacto com relação à frequência de operação, à área ocupada no CI (GAMMIE *et al.*, 2010; KAWA, 2008) e às dificuldades para sua implementação. A Figura 2.2 contém as técnicas de *Low Power* preferidas e mais utilizadas por 115 empresas pesquisadas no ano de 2007, dentro dos segmentos das telecomunicações, dos dispositivos eletrônicos portáteis e dos equipamentos de redes (SINGH, 2008).

Dentre as principais técnicas aplicadas para desenvolver circuitos CMOS de baixo consumo, destacam-se:

- *Clock Gating* – interrupção do sinal de *clock* para os circuitos sequenciais, *latches* e *flip-flops*, que não mudarão de valor naquele ciclo de *clock*.

Figura 2.2 - Popularidade das técnicas de *Low Power*.



Fonte: Figura adaptada de (SINGH, 2008).

- *Voltage Islands* – divisão do circuito em submódulos funcionais alimentados com diferentes níveis de tensão.
- *Multi- $V_{th}$*  – utiliza portas *standard cell* (células padrão) com diferentes níveis de  $V_{th}$  no mesmo circuito. As *standard cells* com baixa  $V_{th}$  apresentarão menor atraso e maior consumo de energia estática e, as *standard cells* com alta  $V_{th}$  apresentarão maior atraso e menor consumo de energia estática.
- *Power Gating* – suspende a alimentação de submódulos funcionais do circuito, quando eles não estão operacionalmente ativos. É uma técnica sofisticada, com implicações de tempo e alta complexidade de implementação. Essa técnica proporciona tipicamente uma maior redução no consumo de energia estática quando comparada às outras técnicas.
- *Logic Restructuring* – algumas portas lógicas são remapeadas para uma combinação de outras portas lógicas equivalentes, que consumam menos energia. Esta técnica aplica-se durante o processo de síntese lógica, utilizando como critério de escolha o nível de atividade nas entradas das portas lógicas.
- *Dynamic Voltage and Frequency Scaling* (DVFS) – varia o nível de tensão de



alimentação e/ou frequência de módulos funcionais de acordo com as exigências de desempenho em operação.

Essas técnicas podem ser utilizadas isoladamente ou em conjunto, e, exigem funcionalidades adicionais no circuito, tais como controladores de gerenciamento de energia, células de isolamento de domínios de energia e/ou registradores para retenção de estados e valores lógicos (CHEN; LIN, 2008; HENZLER, 2007; JADCHERLA *et al.*, 2009).

### **2.3.5 Fatores de Impacto da Redução do Consumo**

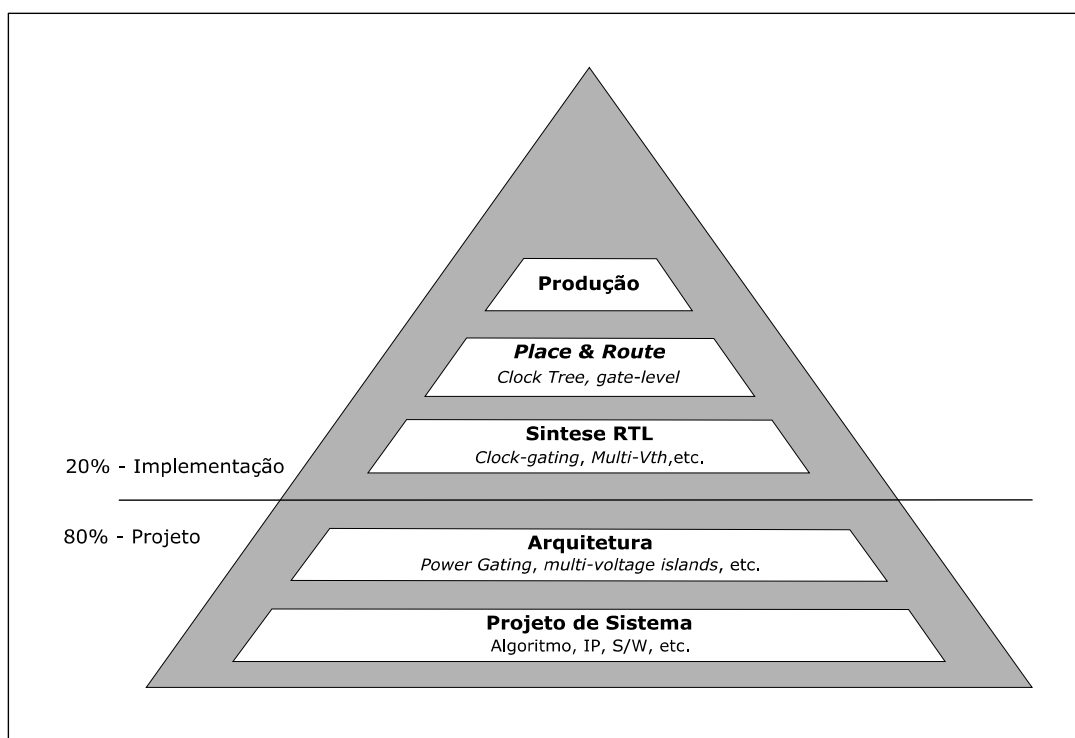
A escolha das técnicas para reduzir o consumo de energia nos circuitos CMOS impacta em benefícios e penalidades no desenvolvimento de um circuito. As técnicas mais eficientes para reduzir o consumo de energia nos circuitos CMOS operam com desligamento e/ou redução do nível da tensão e pelo controle do sinal de *clock* em determinados submódulos do circuito (CHEN; LIN, 2009; KAPOOR *et al.*, 2009; KUWAHARA, 2010). As decisões sobre as restrições de consumo de energia de um circuito CMOS têm um fator de impacto superior às decisões tomadas na implementação RTL quando tomadas no início do fluxo de projeto, principalmente na concepção e na definição da arquitetura do sistema (JADCHERLA *et al.*, 2009; PFI, 2009; XU; YONG, 2009).

Na Figura 2.3 é apresentado um diagrama que expressa a taxa de impacto das decisões nas duas grandes fases do fluxo de desenvolvimento de um circuito CMOS, projeto e implementação, com sugestões para os momentos em que se pode obter mais eficiência na redução do consumo e minimizar os problemas gerados a partir da aplicação de algumas técnicas, em função das etapas do fluxo de desenvolvimento (MATHUR, 2013).

## **2.4 Aritméticas Bit-paralela e Bit-serial**

A proposta de substituição de aritmética bit-paralela por aritmética bit-serial para diminuição de área de silício e/ou redução de consumo de energia em circuitos digitais não é nova (LYON, 1776; BERLEKAMP, 1982; NILSSON; TORKELSON, 1997; WANHAMMAR, 1999; NILSSON, 2006, 2007, 2008, NILSSON, 2009, 2010), porém, mostrou-se, ao longo do tempo, pouco explorada. Talvez porque, essa técnica não se apresenta como uma solução de uso geral, ou seja, as vantagens podem variar de

Figura 2.3 - Fator de impacto das decisões sobre a redução do consumo.



Fonte: Figura adaptada de (MATHUR, 2013).

acordo com a aplicação, dependendo do maior ou menor número de somadores, tamanho de palavras e/ou vazão do circuito e, também, porque a redução de área de silício não é hoje em dia uma preocupação tão grande quanto a redução de consumo de energia. Ela mostra-se vantajosa em determinadas situações onde as palavras a serem operadas são grandes e a vazão não é muito alta (NILSSON, 2008). Estas condições ideais serão especificadas, de maneira mais acurada, mais adiante.

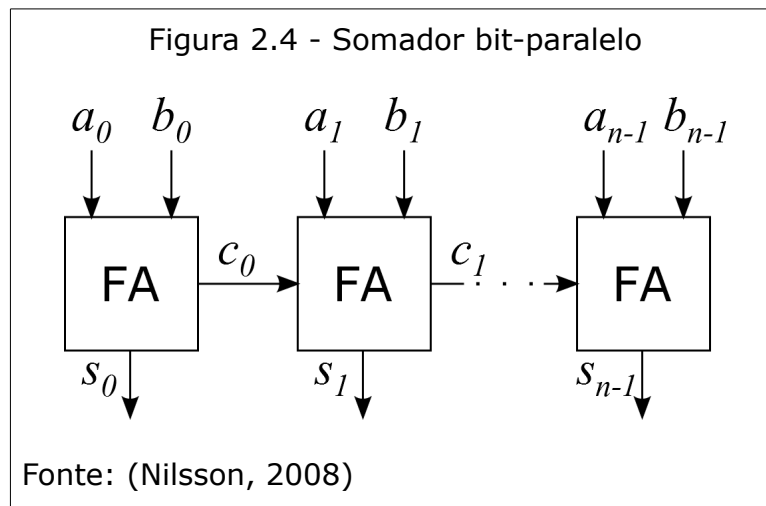
Peter Nilsson em vários artigos (NILSSON; TORKELOSON, 1997; NILSSON, 2006, 2007, 2008, 2009, 2010) defendeu a substituição de aritmética bit-paralela por aritmética bit-serial como meio de reduzir o consumo de energia em circuitos digitais, com ênfase na redução de consumo de energia estática. Nilsson propõe a utilização de aritmética bit-serial para a redução de consumo de energia, principalmente de energia estática, em circuitos que utilizam somadores.

A aritmética bit-paralela utiliza um diferente *full-adder* para realizar a soma de cada par de bits numa palavra, ou seja, para somar duas palavras com  $n$  bits, é necessário um somador com  $n$  *full-adders* paralelos conectados, onde o *carry* de saída do primeiro *full-adder* é entrada para o segundo e assim por diante. O *carry* de

entrada do somador do par de bits menos significativos (*LSB*) é igual a 0 (WANHAMMAR, 1999; NILSSON, 2008).

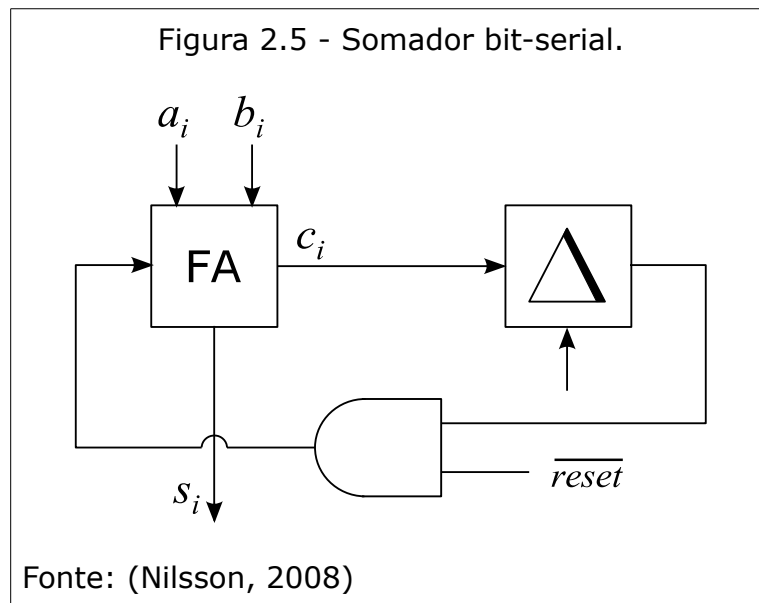
A aritmética bit-serial, para realizar a soma de duas palavras, em vez de utilizar um diferente *full-adder* para somar cada par de bit, usa um mesmo *full-adder* para realizar a soma de todos os pares de bits das palavras que estão sendo somadas,  $n$  vezes. Isso implica em acrescentar um registrador para guardar a saída do *carry* para que seja entrada do *full-adder* na soma do próximo par de bits e uma porta AND para tornar 0 o *carry* de entrada na soma do par de bits menos significativo. Na aritmética bit-serial o primeiro par de bit a ser somado é o par *LSB* (WANHAMMAR, 1999; NILSSON, 2008). É importante ressaltar que com a utilização da aritmética bit-serial, temos menos transístores nos somadores, conseqüentemente, temos menos transístores no *chip* e, por conseguinte, menor ocupação de área de *chip* e menos consumo de energia estática.

A Figura 2.4 ilustra o conceito de aritmética bit-paralela, utilizando-se de um somador de transporte em cascata. No somador paralelo, todos os bits,  $a_i$  e  $b_i$ , estão disponíveis para o somador ao mesmo tempo. Para se obter os bits  $s_i$  na soma de saída, os bits da saída de transporte,  $c_i$  (bits "carry out" ou bits "vai-um"), têm que se propagar por todas as células somadoras FAs (*full-adders*) no somador (WANHAMMAR, 1999; NILSSON, 2008).



A Figura 2.5 mostra o somador bit-serial correspondente. Os bits,  $a_i$  e  $b_i$ , são fornecidos em série, com os bits menos significativos primeiro, para formar a soma de saída correspondente, também de forma serial. O bit da saída de transporte  $c_i$  ("carry out" ou "vai-um") é atrasado um ciclo de *clock* para ser adicionado com os próximos

bits de entrada mais significativos (WANHAMMAR, 1999; NILSON, 2008).



## **3 Objetivos da Pesquisa**

### **3.1 Objetivo Geral**

O objetivo principal desta pesquisa de mestrado é investigar o quanto o consumo de energia estática em circuitos CMOS pode ser reduzido, por meio do emprego de operadores aritméticos bit-serial (LYON, 1976; BERLEKAMP, 1982; WHANHAMMAR, 1999; NILSSON, 2008). A pesquisa está focada em circuitos construídos a partir de *standard cells* (células padrão), com aplicação em processamento de sinais, para os quais o principal requisito não é o alto desempenho computacional, mas o baixo consumo de energia.

A metodologia foi aplicada em um estudo de caso, utilizando-se para isto simulações com o IP core SPVR (*Speaker Verification System*). O SPVR é um verificador de identidade vocal implementado em um circuito dedicado capaz de ter desempenho suficiente para funcionar em tempo real mesmo empregando um sinal de *clock* lento, ou seja, um sinal cujo período é mais de uma ordem de grandeza maior do que o tempo de atraso dos seus circuitos combinacionais.

Os resultados obtidos foram avaliados e comparados com os resultados obtidos por Nilsson, mostrados no artigo "*A Methodology for Arithmetic Reduction of the Static Power Consumption Verified on Filter Architectures*" (NILSSON, 2008), para verificar a validade das conclusões apresentadas no mesmo.

### **3.2 Objetivos Específicos**

Para atingir o objetivo geral, os objetivos listados a seguir foram alcançados:

1) Mostrou-se que a comparação de consumo entre um circuito que emprega aritmética bit-serial e um circuito que emprega aritmética bit-paralela pode ser feita a partir de simulação lógica no nível de *standard cells* (células padrão);

2) Avaliou-se a vantagem, em termos de consumo de energia ao se usar aritmética bit-serial, aumenta quando se usa uma tecnologia de fabricação mais avançada, com transistores menores;

3) Avaliou-se a redução de consumo de energia, obtida na aplicação de

aritmética bit-serial num circuito mais complexo, comparada com a redução de consumo alcançado com um circuito mais simples.

## 4 Pesquisas Relacionadas

Nesta seção, são apresentadas pesquisas relacionadas com a economia de energia em projetos de circuitos digitais, aplicando as técnicas de *Low Power*, com maior foco nas técnicas de redução de consumo de energia estática.

### 4.1 Técnicas para Redução de Consumo de Energia de um Circuito CMOS

Para diminuição do consumo de energia de circuitos integrados CMOS, tem-se várias técnicas, chamadas de técnicas de *Low Power*, que podem ser aplicadas de várias formas, de acordo com os vários estágios de desenvolvimento de um circuito integrado. Essas técnicas têm complexidade de aplicação variada e a economia que elas geram são influenciadas por vários fatores, dependendo de sua natureza e do estágio de desenvolvimento no qual são aplicadas, tendo assim taxas de redução de consumo de energia também variadas. Há técnicas que conseguem diminuir apenas um tipo de consumo de energia, dinâmico ou estático, e outras que atuam na diminuição de ambos os tipos de consumo. Tendo em vista o objetivo da pesquisa, serão mostradas a seguir, algumas técnicas de *Low Power*, agrupadas por níveis de abstração do fluxo de projeto de circuito: de transistores, de circuito e de sistemas.

#### 4.1.1 Técnicas *Low Power* Aplicáveis no Nível de Transistores

As técnicas aplicadas no nível de transistores atuam sobre a construção (materiais utilizados, dimensões e processo de fabricação) dos mesmos, bem como sobre as características da energia para o seu funcionamento (tensão e corrente).

Shauly apresenta a técnica *threshold voltage reduction* (redução de voltagem de limiar) como a maneira mais simples de deixar mais rápido os transistores e reduzir a propagação do atraso (SHAULY, 2012). No entanto, a redução de  $V_{th}$  (*threshold voltage*) significa um incremento exponencial de  $I_{sub}$  (*subthreshold current*). Pelo uso de valores muito altos de  $V_{th}$  para os caminhos não críticos, a corrente de fuga pode ser reduzida em 2 a 3 ordens de grandeza. O artigo também mostra que os valores de  $V_{th}$ , para plataformas de 250 nm até 65 nm, decrescem para as tecnologias de 250 nm até 90 nm e que a partir daí não eram mais reduzidos.

Uma das técnicas utilizadas para redução do consumo de energia devido às correntes de fuga é a utilização de transistores com  $V_{th}$  diferentes, de modo que caminhos não críticos contêm transistores com maior  $V_{th}$ . Santos, em seu artigo "An Optimization Mechanism Intended for Static Power Reduction Using Dual- $V_{th}$  Technique" (SANTOS et al., 2012), comenta variações da técnica e apresenta um algoritmo, baseado em heurística, de *dual- $V_{th}$* , para substituição de transistores que utilizam menor  $V_{th}$ , por transistores que utilizam maior  $V_{th}$ , para redução da corrente de fuga. Como limitação, o algoritmo se aplica apenas a circuitos combinacionais.

Em um artigo, Joshi e Bhavisar afirmam que o tamanho do dispositivo tem diminuído por um fator  $k = 0,7$  por geração de tecnologia, mas que o consumo de energia não está seguindo a mesma tendência e que, para diminuição do consumo, foram desenvolvidas técnicas de uso de múltiplas *threshold voltage*  $V_{th}$  e de tensão de alimentação  $V_{dd}$  também variável (JOSHI; BHAVSAR, 2012). No artigo, é dito que na técnica de múltiplas  $V_{dd}$ , os sub-circuitos de temporização crítica são alimentados por uma tensão alta e que a maioria dos outros sub-circuitos são alimentados por uma tensão baixa. É dito também que o principal desafio, para implementar a técnica descrita acima, é projetar escalonadores de níveis energia eficientes.

No artigo já citado de Joshi e Bhavisar (JOSHI; BHAVSAR, 2012), também é analisada a técnica de diminuição de consumo utilizando *multi- $V_{th}$*  e as dificuldades dela decorrentes, afirmando que em circuitos com vários limiares de tensão são usados dois transistores de diferentes  $V_{th}$ . Transistores de limiares baixos, que são rápidos e dão mais fuga, são empregados nos sub-circuitos de velocidade crítica. Por outro lado, os transistores de limiares altos que apresentam fuga baixa de *subthreshold*, mas que são mais lentos e, portanto, são empregados nos caminhos não-críticos do *chip*. No entanto, como mais transistores estão se tornando críticos no tempo, circuitos de múltiplos limiares não são eficazes.

No artigo "Leakage Power Reduction in CMOS Circuits using Leakage Control Transistor Technique in Nanoscale Technology" (DILIP; PRASAD; BHAVANI, 2012), os autores mostram que os circuitos com limiares variáveis superam o problema acima, controlando dinamicamente a tensão de *threshold* do transistor, por polarização do substrato.



#### 4.1.2 Técnicas *Low Power* Aplicáveis no Nível de Circuito

Joshi e Bhavisar (JOSHI; BHAVSAR, 2012) mostram que a otimização de circuitos envolve o projeto da biblioteca de células e o dimensionamento de transistores, para obter-se baixo consumo de energia. As células mais críticas em uma biblioteca digital são os componentes sequenciais, ou sejam, *latches* e *flip-flops*. Eles são encontrados em muitos circuitos de *pipeline* e estão ligados à rede de fiação mais ativa no *chip*, isto é, o *clock*. O *Clock Driver*, por causa da grande carga capacitiva da rede de distribuição de *clock*, contribui fortemente para o consumo total de energia de um *chip*. *Flip-flops*, projetados para baixo consumo de energia, visam reduzir a carga capacitiva do *clock* e reduzir a potência do sinal de *clock* interno, quando é alternado (chaveado). O dimensionamento de transistores é também utilizado para minimizar o consumo de energia nos circuitos lógicos combinacionais. Aumentando o tamanho do transistor, variando a relação W/L (largura/altura), aumenta a velocidade, mas, também dá origem a um aumento da dissipação de energia. Por isso, deve-se usar transistores de menor porte, que irão satisfazer às restrições de atraso. Economia de energia também pode ser alcançada usando estilos de implementação de lógica não convencionais, utilizando lógica de transistor de passagem e lógica de transmissão, o que pode reduzir o número de transistores necessários para função lógica usada em unidades aritméticas.

Kathuria, Ayoubkahan e Noor, fazem uma revisão das cinco técnicas existentes de *Clock Gating*, fazendo observações sobre a simulação de cada técnica, apontando as falhas que põem em risco o correto funcionamento do sistema (KATHURIA; AYOUBKHAN; NOOR, 2011). A técnica de *Clock Gating*, para otimizar o consumo de dispositivos, inibe o chaveamento do *clock* de um bloco sequencial de um sub-circuito, quando a função do sub-circuito requer que os valores armazenados no bloco sequencial sejam mantidos inalterados. *Clock Gating* é uma técnica que pode ser usada para controlar a energia dissipada pela rede de *clock*. O artigo propõe ainda uma nova técnica que vence os problemas localizados nas técnicas anteriores. Os autores, no trabalho, não fazem nenhuma comparação que permita quantificar a redução de consumo de cada técnica já apresentada, nem a redução de consumo a ser obtida pela nova proposta. O artigo carece também da apresentação de resultados obtidos sobre um sistema de maior complexidade.

No ano de 1997, Nilsson e Torkelson, no artigo "A Custom Digital Intermediate

*Frequency Filter for the American Mobile Telephone System*" (NILSSON; TORKELSON, 1997), propuseram um filtro digital para sistemas de comunicações móveis de frequência intermediária como alternativa para os filtros analógicos, que eram utilizados na maioria dos receptores heteródinos da época. As aritméticas bit-paralela e bit-serial foram aplicadas no algoritmo do filtro. A aplicação da aritmética bit-serial, em relação a aplicação da bit-paralela, mostrou a metade da área utilizada pelo circuito e que o consumo de energia mostrado, também foi quase a metade. Por conseguinte, os autores concluíram que aritmética bit-serial poderia ser útil para projetos de *Low Power*.

Em "*Power Reduction in Custom CMOS Digital Filter Structures*" (ÅSTRÖM; NILSSON; TORKELSON, 1999), os autores, por meio de duas implementações de filtros, mostraram que a potência e a velocidade poderiam ser melhoradas se o esforço de otimização fosse feito na redução dos comprimentos dos coeficientes do filtro. Ambos os filtros foram implementados com o uso da aritmética bit-serial de coeficientes fixos. As medições mostraram um aumento de velocidade de oito vezes, com a metade do consumo de energia e apenas 30% da área para os coeficientes do filtro otimizado. Neste artigo, os autores também mostraram que na implementação bit-serial a perda de velocidade foi de apenas 33%.

No artigo "*Arithmetic Reduction of the Static Power Consumption in Nanoscale CMOS*" (NILSSON, 2006), o autor apresenta uma técnica para diminuir o maior problema do projeto de circuitos CMOS de nano-escala, que é o consumo de energia estática. A técnica utiliza a aritmética bit-serial em substituição da bit-paralela, para a redução do consumo de energia. Foi feita uma comparação entre dois filtros, um usando aritmética bit-serial e outro utilizando aritmética bit-paralela, mostrando por simulação SPICE que o consumo de energia estática decai com o uso da arquitetura bit-serial e decai também quando os tamanhos das palavras aumentam. Em tecnologias de nano-escala, as fugas em transistores é um dos principais contribuintes para o consumo de energia e a diminuição do número de transistores, pelo uso de aritmética bit-serial, leva à redução do consumo. Técnicas atuais para a redução de consumo de energia estática no nível de arquitetura, como a paralelização, *pipelining* e *Clock Gating* têm efeito limitado. Métodos de baixo nível, como os transistores do sono, são aplicados predominantemente, ao modo de repouso (*standby*). O autor conclui que novas metodologias são, portanto, necessárias para reduzir o consumo de energia, em especial no modo ativo de operação. O trabalho demonstra, ainda, que

esta metodologia aritmética reduz o consumo de energia estática substancialmente. Uma redução de até 20% é declarada. O artigo necessitava fazer mais validações, utilizando aplicações mais significativas e de maior complexidade.

Nilsson, em seu artigo "*Arithmetic and Architectural Design to Reduce Leakage in Nano-Scale Digital Circuits*" (NILSSON, 2007), afirma que em tecnologias de escala nanométricas a energia estática, ou seja, o vazamento ou fuga, é uma importante contribuição para o consumo total de energia num circuito. O artigo discute, ainda, metodologias de redução das energias estática e dinâmica no nível de arquitetura e aritmética. Também nele, são abordadas técnicas para reduzir o consumo de energia estática em aplicações digitais para tecnologias CMOS de nano-escala. Em conclusão, é apontada a redução do consumo de energia estática de 79%, usando aritmética bit-serial, em vez de bit-paralela.

Em 2008, no artigo "*A Methodology for Arithmetic Reduction of the Static Power Consumption Verified on Filter Architectures*", Peter Nilsson apresentou uma metodologia para redução de consumo de energia estática, em filtros, de até 37%, por meio da utilização de arquitetura bit-serial e de até 7% na aritmética no filtro (NILSSON, 2008). A metodologia alcança uma redução de consumo de energia estática, no nível de aritmética, substituindo arquitetura bit-paralela por bit-serial, em somadores. No artigo, é mostrado também, que a redução de consumo de energia estática, também é fortemente dependente do tamanho da palavra utilizada. A redução é dependente também da razão entre a aritmética e os registradores utilizados. O autor concluiu, por fim, que nas arquiteturas dominadas pela aritmética consegue-se maior redução. O trabalho necessita de uma análise mais detalhada da redução de velocidade do circuito, pela aplicação da metodologia. Este artigo fornecerá, com seus resultados apresentados, os elementos de comparação para a avaliação dos resultados a serem obtidos na pesquisa de mestrado ora apresentada.

No artigo "*Architectures and Arithmetic for Low Static Power Consumption in Nanoscale CMOS*" (NILSSON, 2009), o autor foca na redução do vazamento (fuga) de energia no nível de arquitetura e aritmética. Uma metodologia para a redução considerável do consumo de energia estática é mostrada. As simulações são feitas no nível de transistores usando SPICE, com uma tecnologia CMOS de 130 nm. Com base nos resultados de simulações, o consumo de energia estática é estimado e comparado para as diferentes arquiteturas de filtro. Três diferentes tipos de arquiteturas, ou seja,

bit-paralela, dígito-serial, e bit-serial são usadas para demonstrar a metodologia. Com base nos resultados da simulação, a energia estática em um filtro de Hilbert e um filtro IIR (*Infinite Impulse Response*) de terceira ordem, filtro de meia banda, são investigados e os resultados comparados, utilizando as três diferentes arquiteturas. O artigo mostra que uma redução substancial do consumo de energia estática pode ser alcançada quando são utilizadas aritmética bit-serial e dígito-serial. O artigo também mostra que a relação de energia é fortemente dependente do comprimento de palavra usada, isto é, a redução é maior para maiores comprimentos de palavras. Arquiteturas nas quais a aritmética domina apresentará uma maior redução no consumo de energia estática. A redução de energia estática pela metade é alcançada para o filtro de Hilbert bit-serial e de quase um terço no filtro de meia banda. Considerando só para a parte aritmética, uma relação potência de 0,11 e 0,06 é apontada, respectivamente, para os filtros. A conclusão geral é que é uma boa ideia mudar para aritmética bit-serial, para arquiteturas de baixa e média velocidades, quando são utilizadas tecnologias menos densas, tais como a tecnologia de 130 nm usada nesta pesquisa. Para as tecnologias mais densas, como de 65 nm abaixo, vai valer a pena mudar para aritmética bit-serial, em arquiteturas de maior velocidade, já que o consumo de energia dinâmica não é dominante.

Em "*Approaching Green Electronics: Power Efficient Arithmetic in Nano-scale CMOS*" (NILSSON, 2010) é apresentado o mesmo problema do artigo anterior, apenas mudando o enfoque para dispositivos eletrônicos verdes, tendo em vista a necessidade que ora se apresenta, de que os dispositivos eletrônicos devem consumir o mínimo possível de energia. As conclusões, basicamente, são as mesmas do artigo apresentado anteriormente, não trazendo nada de novo em contribuições.

Nos trabalhos pesquisados, Nilsson usa sempre simulação SPICE no nível de transistores, embora simuladores no nível de *standard cells* também sejam capazes de fornecer estimativas de consumo de energia. Mesmo Nilsson mencionando que a redução de consumo pela aplicação da técnica bit-serial será significativa em tecnologias menores do que 50 nm, os estudos de caso são feitos com tecnologia de 130 nm e os circuitos para os estudos de caso são relativamente de baixa complexidade.

### 4.1.3 Técnicas *Low Power* Aplicáveis no Nível de Sistema

Rani e Malakan, no ano de 2010, fizeram um levantamento de várias técnicas de *Power Gating* aplicáveis para a redução de consumo de energia no modo ativo de operação, apoiados na ideia de desligar blocos do circuito que estão em estado de espera (*standby*) (RANI; MALARKKAN, 2010). Ao serem desligados, os blocos deixam de consumir energia estática. A técnica *Multi Threshold CMOS* (MTCMOS) permite que um bloco lógico seja ligado, ou desligado durante o modo de espera, usando dois transistores com voltagem de limiar (*voltage threshold*) alta, chamados de transistores do sono. Esses transistores são colocados entre o bloco lógico e a alimentação ( $V_{dd}$ ) e entre o bloco lógico e o terra ( $GND$ ). Outra técnica apresentada utiliza transistores empilhados para desativar blocos do circuito. Uma terceira técnica apresentada utiliza conjuntamente as duas anteriores para desativação dos blocos em estado de espera. A técnica chamada de VCLEARIT implementa um autocontrole de redução de energia, usando a probabilidade do sinal para ativar ou desativar um bloco, com redução de até 61% da energia estática dissipada. Como alternativa à primeira técnica, MTCMOS, é apresentada a técnica *Super Cutoff CMOS* (SCCMOS), que usa uma voltagem negativa para desligar um transistor do sono nMOS, diminuindo exponencialmente a corrente sub-limiar (*subthreshold*). Janaki e Malakkan apresentam, ainda, outra abordagem de *Power Gating*, chamada de *Input Vector Control* (IVC), para a redução da corrente de fuga. IVC tira vantagem do efeito da pilha de transistores, para aplicar o vetor de fuga mínima, *minimum leakage vector* (MLV), para as entradas primárias do circuito, durante o modo de espera. É apresentada, também, uma Rede de Transistores do Sono Distribuída (DSTN) para a redução de energia. Transistores do sono são eficientes para reduzir a energia dinâmica e as fugas. O projeto baseado em *cluster*, foi apresentado para reduzir a área de transistores do sono, agrupando portas para minimizar a corrente de chaveamento (*switching*) simultâneo por *cluster*, inserindo um transistor do sono por *cluster*. Por último, o trabalho apresentou um projeto semi-personalizado de circuitos *Power-Gated Zigzag* em elementos de células padronizadas (*Semicustom Design of Zigzag Power-Gated Circuits in Standard Cell Elements* - ZPG). ZPG resolveu o atraso na ativação da alimentação padrão, mas a sua exigência para ambas as chaves de corrente nMOS e pMOS, em zigue-zague, requer redes de alimentação complexas, limitando a aplicação em projetos personalizados. Depois de analisarem várias técnicas de redução de fuga, os autores concluíram que existe uma correlação entre

as três métricas de performance: fuga de potência, a potência e atraso de propagação. Se uma métrica é otimizada, leva a um comprometimento das outras métricas. Concluíram também, que a técnica *Super Cutoff CMOS* proporciona uma economia eficiente de fuga de energia e Empilhamento Forçado de Transistores é um esquema muito eficaz de economia de energia. No entanto, se o atraso de propagação é o principal critério, recomenda-se circuitos baseados em um único transistor do sono. A abordagem baseada em pilha de transistores do sono é adequada para uma operação mais rápida do circuito.

LACKEY e seus colegas, em um artigo no ano 2002, mostraram como reduzir o consumo estático e dinâmico de energia com o uso de *Voltage Island*, que é uma arquitetura de sistema e uma metodologia de implementação de *chip* (LACKEY et al., 2002). *Voltage Island* é aplicada em função das características individuais dos vários blocos funcionais que compõem o sistema, aplicando a menor voltagem, suficiente ao funcionamento correto de cada bloco. O artigo apresenta uma visão global sobre *Voltage Island*, mostrando as possibilidades de uso da técnica, por meio de vários cenários favoráveis à sua aplicação. O trabalho discute ainda a questão de se operar com várias fontes de energia e com vários níveis de tensão, definindo um fluxo de projeto para o desenvolvimento de um SoC (*System-on-Chip*) que utiliza a técnica de *Voltage Island*. Somado a isso, o trabalho fornece recomendações para a execução do particionamento dos blocos do sistema, ou seja, a definição das ilhas de voltagens, tecendo considerações sobre o *timing* na síntese e o *timing* estático. O artigo orienta, ainda, como desenvolver o projeto para atender questões de testabilidade e testes de fabricação, de planejamento físico e implementação de estruturas. O artigo, porém, deixa de apresentar qualquer avaliação quantitativa quanto a redução de consumo alcançada pela técnica, limitando-se a qualificar a redução e consumo como “dramática e significativa”.

Silveira, em sua tese de doutorado (SILVEIRA, 2012), apresentou uma abordagem que permite a verificação de circuitos integrados que usam a técnica de *Power Gating*, de uma forma mais eficiente e ágil no nível ESL, utilizando *SystemC*. A pesquisa também propiciou o desenvolvimento de uma ferramenta de simulação para sistemas descritos no nível ESL e RTL, com base em *SystemC-TLM*, a ser utilizada na verificação funcional da técnica de *Power Gating* aplicada ao circuito. Outra contribuição desta pesquisa ocorre na etapa do projeto arquitetural, quando é possível realizar a análise da arquitetura do projeto que usa a técnica *Power Gating* e avaliar o

impacto da aplicação da técnica ao sistema, permitindo com isso distribuir as funcionalidades do sistema entre os submódulos, obtendo, assim, a maximização da economia de energia, com base na aplicação de *Power Gating*. A ferramenta, ainda, permite realizar o confronto dos resultados do nível ESL com os do nível RTL. O novo simulador carece de uma forma de estimar o consumo de energia, mas que Silveira deixou como sugestão para pesquisas futuras. A técnica proposta nessa pesquisa foi validada por meio de quatro estudos de casos que demonstraram a efetiva aplicação, simulação e verificação da técnica de *Power Gating* nos níveis ESL e RTL, permitindo comparar se os modelos são logicamente equivalentes.

## 4.2 Resumo das Pesquisas Relacionadas

A Tabela 4.1, a Tabela 4.2 e a Tabela 4.3, a seguir, apresentam uma síntese das pesquisas relacionadas aqui apresentadas.

Tabela 4.1 - <b>Técnicas <i>Low Power</i> aplicáveis no nível de transistor</b>			
Referência	Técnica	Abordagem	Resultados e/ou restrições
(SHAULY, 2012)	<i>Threshold voltage reduction</i>	Deixar mais rápido os transistores e reduzir a propagação do atraso.	Incremento exponencial de $I_{sub}$ . Corrente de fuga pode ser reduzida em 2 a 3 ordens de grandeza. $V_{th}$ só decresce para as tecnologias de 250 nm até 90 nm.
(SANTOS et al., 2012)	Transistores com $V_{th}$ diferentes	Algoritmo de <i>dual-<math>V_{th}</math></i> , baseado em heurística, que substitui transistores com menor $V_{th}$ , por transistores com maior $V_{th}$ .	Caminhos não críticos com transistores com maior $V_{th}$ Redução da corrente de fuga. Aplica-se apenas a circuitos combinacionais.
(JOSHI; BHAVSAR, 2012)	Múltiplas $V_{th}$ e tensão de alimentação $V_{dd}$ variável	Os sub-circuitos de temporização crítica são alimentados por uma tensão alta. Transistores de $V_{th}$ baixas (rápidos e com mais fuga) são empregados nos sub-circuitos de velocidade crítica.	O desafio é projetar escalonadores de níveis energia eficientes. Mais transistores estão se tornando críticos no tempo, circuitos de múltiplos limiares não são eficazes.

(DILIP; PRASAD; BHAVANI, 2012)	$V_{th}$ variável	Controla dinamicamente a $V_{th}$ do transistor, por polarização do substrato.	Os circuitos com $V_{th}$ variável superam o problema anterior.
---	-------------------	--	---

Tabela 4.2 - **Técnicas *Low Power* aplicáveis no nível de circuito**

<b>Referência</b>	<b>Técnica</b>	<b>Abordagem</b>	<b>Resultados e/ou restrições</b>
(JOSHI; BHAVSAR, 2012)	Projeto da biblioteca de células e o dimensionamento de transistores	Reduzir a carga capacitiva do <i>clock</i> e reduzir o consumo do sinal de <i>clock</i> interno, quando é alternado. Dimensionamento de transistores para minimizar o consumo de energia nos circuitos lógicos combinacionais.	As células mais críticas em uma biblioteca digital são os componentes sequenciais. Aumentando o tamanho do transistor, variando a relação W/L (largura/altura), aumenta a velocidade, mas, também dá origem a um aumento da dissipação de energia.
(KATHURIA; AYOUBKHAN; NOOR, 2011)	<i>Clock Gating</i>	Revisão das cinco técnicas existentes de <i>Clock Gating</i> . Uso de <i>Clock Gating</i> para controlar a energia dissipada pela rede de <i>clock</i> .	Aponta falhas que põem em risco o correto funcionamento do sistema. <i>Clock Gating</i> para controlar a energia dissipada pela rede de <i>clock</i> . Não quantifica a redução de consumo.
(NILSSON, 2008)	Substituição arquitetura bit-paralela por bit-serial	Redução de consumo de energia estática, no nível de aritmética, substituindo a arquitetura bit-paralela por bit-serial, em somadores.	Redução de consumo de energia estática, em filtros, de até 37% e até 7% na aritmética no filtro. A redução é dependente da razão entre a aritmética e os registradores, e do tamanho da palavra utilizada.
(NILSSON; TORKE-SON, 1997)	Substituição arquitetura bit-paralela por bit-serial	A aplicação da aritmética bit-serial em filtro para diminuir a área utilizada e o consumo de energia.	Redução de metade da área utilizada pelo circuito. Consumo de energia foi quase a metade. Aritmética bit-serial poderia ser útil para projetos de <i>Low Power</i> .
(ÅSTRÖM; NILSSON; TORKE-SON, 1999)	Substituição da aritmética bit-paralela por bit-serial	Diminuição do consumo de energia estática, substituindo a aritmética bit-paralela por bit-serial, na implementação de dois filtros.	Aumento de velocidade de oito vezes e metade do consumo de energia. 30% da área para os coeficientes do filtro otimizado. Na implementação bit-serial a perda de velocidade foi de apenas 33%.



(NILSSON, 2006)	Substituição da aritmética bit-paralela por bit-serial	Aritmética bit-serial em substituição da bit-paralela, para a redução do consumo de energia em tecnologia de nanoescala.	Com o uso da arquitetura bit-serial decaí o consumo de energia estática e decaí também com o aumento dos tamanhos das palavras. Diminuição do número de transistores pelo uso de bit-serial leva à redução do consumo. Redução de consumo de até 20%.
(NILSSON, 2007)	Substituição da aritmética bit-paralela por bit-serial	Redução do consumo de energia estática em tecnologias nanoescala, usando aritmética bit-serial, em vez de bit-paralela para redução de vazamento.	Redução do consumo de energia estática de 79%.
(NILSSON, 2009)	Substituição da aritmética bit-paralela por bit-serial	Redução do consumo de energia estática em um transformador de Hilbert e um filtro de meia banda, usando aritmética bit-serial.	É uma boa ideia usar aritmética bit-serial para arquiteturas de baixa e média velocidades, com tecnologias de 130 nm e mais densas, já que o consumo de energia dinâmica não é dominante.
(NILSSON, 2010)	Substituição da aritmética bit-paralela por bit-serial	Mesmo problema do artigo anterior, apenas mudando o enfoque para dispositivos eletrônicos verdes.	As conclusões, basicamente, são as mesmas do artigo anterior, não trazendo nada de novo em contribuições.

Tabela 4.3 - **Técnicas *Low Power* aplicáveis no nível de sistema**

<b>Referência</b>	<b>Técnica</b>	<b>Abordagem</b>	<b>Resultados e/ou restrições</b>
(RANI; MALARKKAN, 2010)	<i>Power Gating</i>	Várias técnicas de <i>Power Gating</i> aplicáveis para a redução de consumo de energia no modo ativo de operação, desligando blocos do circuito que estão em estado de espera.	Desligados, os blocos deixam de consumir energia estática. Redução de até 61% da energia estática consumida.
(LACKEY et al., 2002)	Voltage Island	Aplicação da menor voltagem, suficiente ao funcionamento correto de cada bloco.	Possibilidades de uso da técnica. Operação com várias fontes de energia e com vários níveis de tensão. Recomendações para o desenvolvimento do projeto.

(SILVEIRA, 2012)	Verificação de circuitos integrados que usam a técnica de <i>Power Gating</i>	Verificação de circuitos que usam <i>Power Gating</i> , com uma ferramenta de simulação para sistemas descritos no nível ESL e RTL, com base em <i>SystemC-TLM</i> .	Forma eficiente e ágil no nível ESL. Validação por meio de quatro estudos de casos. Compara, se os modelos são logicamente equivalentes. Avalia o impacto da aplicação da técnica.
------------------	---	--	---

## 5 Metodologia da Pesquisa

Nesta seção, é apresentado como está organizada a pesquisa, os critérios utilizados para a seleção e o sequenciamento das atividades que a compõe. Para isso, utilizou-se como elementos balizadores os objetivos da pesquisa aqui apresentada e as técnicas que se mostram como tendências atuais no desenvolvimento de circuitos digitais.

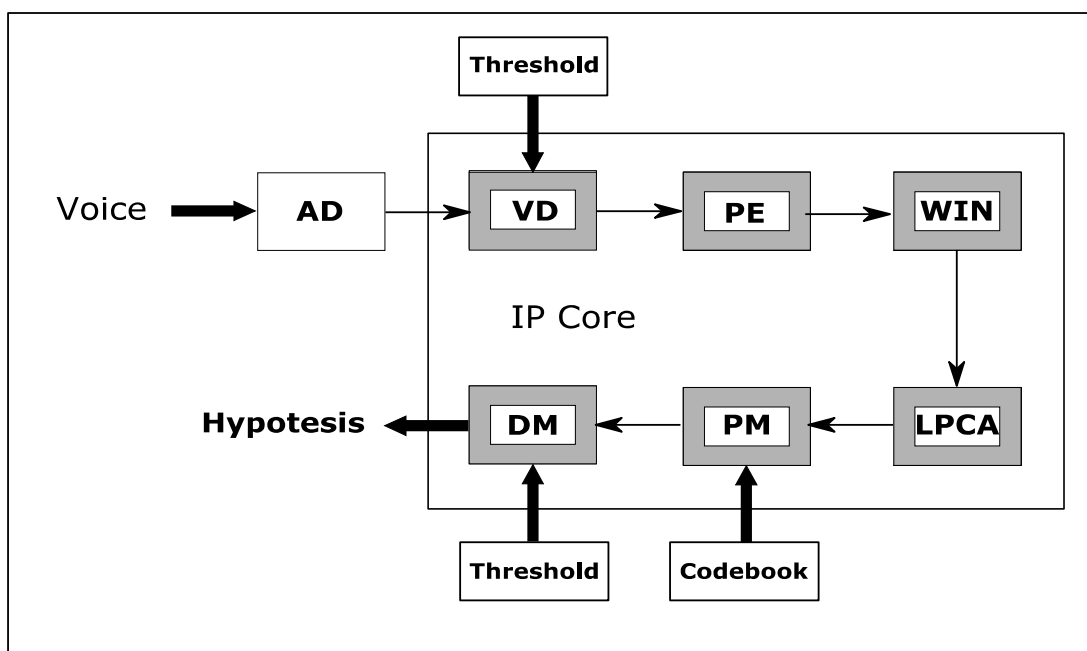
### 5.1 Caminhos da Pesquisa

O ponto de partida para essa pesquisa foi a redução de consumo de energia estática resultante da substituição de aritmética bit-paralela pela aritmética bit-serial relatada por Peter Nilsson no artigo "*A Methodology for Arithmetic Reduction of the Static Power Consumption Verified on Filter Architectures*" (Nilsson, 2008), que utilizou uma aplicação de pequena complexidade, um filtro de meia banda de terceira ordem e somadores de palavra bit-paralelo e bit-serial, utilizando modelo de transistores e tecnologia de 130 nm da UMC. O ponto final que se objetivava atingir nessa pesquisa era a verificação da redução de consumo estático e as vantagens e desvantagens dessa substituição de aritméticas, quando aplicada a um sistema de maior complexidade, utilizando o modelo de *standard cells* com uma tecnologia abaixo de 50 nm, no caso, a tecnologia de 28 nm por meio do uso, nas simulações, das bibliotecas de *standard cells* CMOS028FDSOI da STIMicroelectronics. O sistema escolhido para aplicar esse experimento foi o SPVR (*Speaker Verification System*).

O SPVR é um sistema que realiza o reconhecimento automático de identidade vocal de um locutor e foi escolhido para essa pesquisa por ser um sistema de processamento digital de sinais que não exige alto desempenho computacional, ou seja, pode trabalhar com um *clock* mais lento, e porque possui uma grande quantidade de somadores necessários para o processamento de sinais na verificação da identidade vocal do locutor, em tempo real. O SPVR foi inicialmente implementado em software. Posteriormente, deu origem a um IP *core* (circuito digital) que foi implementado em hardware por meio da fabricação de um circuito integrado (*chip*). O projeto do SPVR foi desenvolvido pelo LAD/UFCG (Laboratório de Arquiteturas Dedicadas da UFCG), no âmbito do Brazil-IP, projeto do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq do Ministério da Ciência, Tecnologia e Inovação - MCTI. O objetivo maior do Brazil-IP é fornecer treinamento especializado

no projeto de IP cores para alunos de universidades brasileiras.

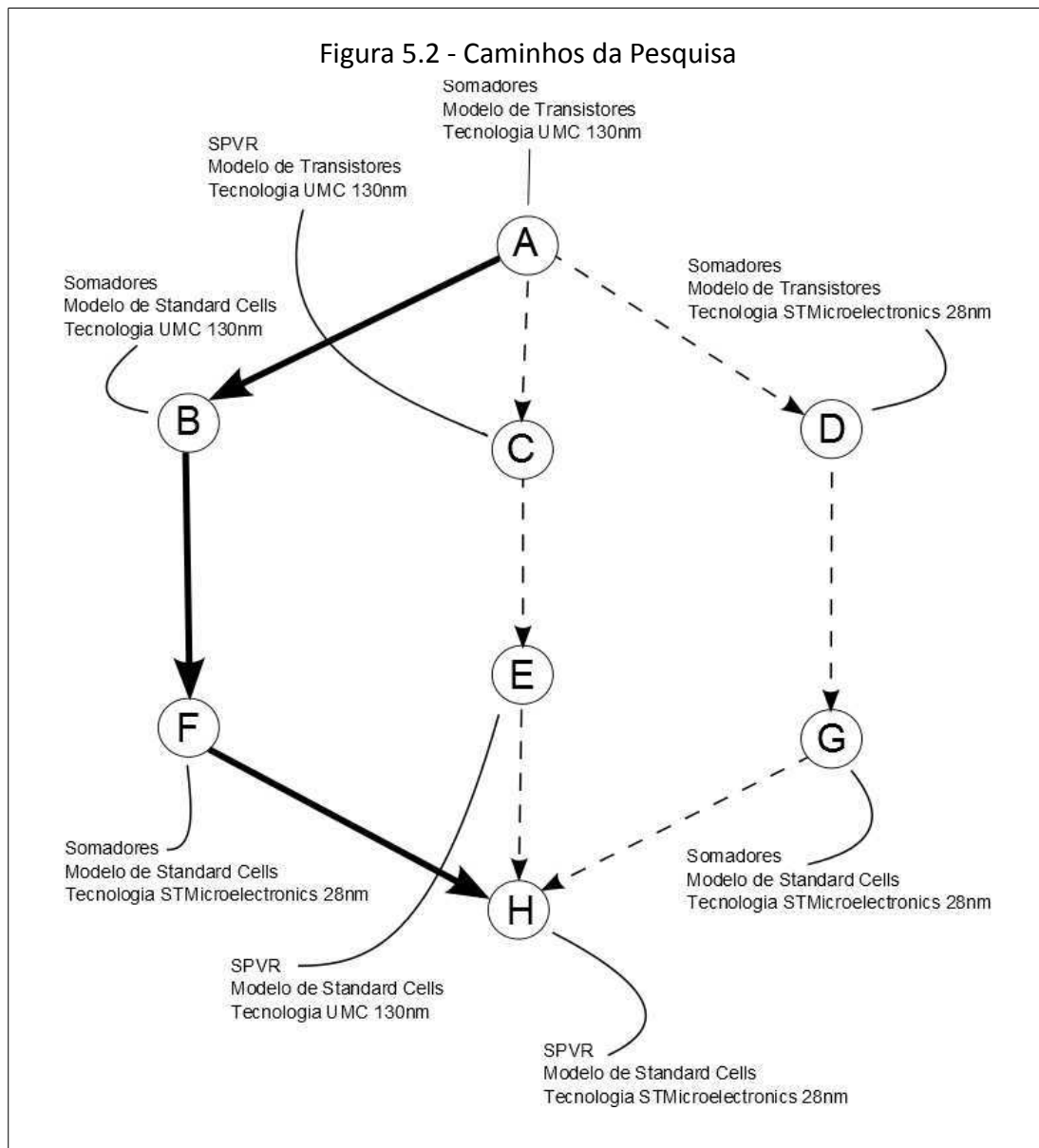
Figura 5.1 - Arquitetura do SPVR



Fonte: Figura adaptada de (FECHINE et al., 2010)

O SPVR é dividido em seis blocos, conforme é mostrado no artigo "SPVR: An IP core for Real-Time Speaker Verification" (FECHINE et al., 2010), e a Figura 5.1 mostra a arquitetura do IP core, na qual os blocos: VD-Voice Detector (detector de voz) elimina amostras de áudio de ruído ou silêncio; PE-Preemphasis (pré-ênfase) é responsável pela atenuação dos componentes de baixa frequência do sinal de voz; WIN-Windowing (janelamento) divide o sinal de voz em blocos ou frames; LPCA-Linear Prediction Coding Analysis (análise de codificação preditiva linear) aplica uma das técnicas mais simples de análise de fala e extrai as suas características; PM-Pattern Matching (casamento de padrões) é responsável pela análise dos coeficientes LPC com relação a um conjunto de vetores ideais, chamado de codebook, que é uma representação do trato vocal do locutor; e, DM-Decision Maker (tomador de decisões), que é responsável pela decisão de reconhecer ou não o locutor, com base na média das distâncias recebidos do PM e os limiares das distâncias (codebook).

Para se chegar ao ponto final, iniciando do ponto de partida, foi observado que poder-se-ia percorrer vários caminhos com a pesquisa, como é mostrado no grafo da Figura 5.2. Nesse grafo, o ponto de partida é assinalado com a letra A e o ponto final com a letra H. A partir do ponto inicial (A) verificou-se que havia várias opções de caminhos para a pesquisa, considerando as variantes que apresentavam-se:



- a) a complexidade da aplicação – com somadores de pequena complexidade ou com um sistema de maior complexidade, o SPVR;
- b) o modelo do circuito – de transistores ou de *standard cells*; e,
- c) a tecnologia – de 130 nm da UMC ou de 28 nm da STMicroelectronics.

Diante dessas três variáveis e das duas opções que se apresentavam para cada uma, vários caminhos apresentaram-se como possíveis no grafo. Porém, para direcionar melhor a pesquisa, algumas regras foram estabelecidas para se chegar à definição do caminho metodológico para essa pesquisa de mestrado. A primeira regra a ser seguida foi que só seria mudada uma única opção de uma única variável entre etapas ou atividades contíguas do caminho a ser escolhido, para que se pudesse avaliar o efeito de uma mudança, sem interferências de outras. A segunda regra foi

que não haveria retrocessos nem de complexidade, nem de modelos e nem de tecnologia, ou seja: a mudança de complexidade seria sempre da menor para a maior; o modelo só poderia evoluir no sentido de transistores para *standard cells* e a tecnologia da menos densa (130 nm) para a mais densa (28 nm). Foi decidido também, por questões de predominância de utilização (tendência atual) no desenvolvimento de projetos de circuitos digitais, que os experimentos seriam feitos unicamente utilizando o modelo de *standard cells*. Outra decisão tomada foi que seriam implementados os somadores, bit-paralelo e bit-serial, nas duas tecnologias consideradas. Essas decisões foram tomadas para atender os objetivos da pesquisa e porque desejava-se avaliar a redução do consumo quando variava-se a complexidade da aplicação, o modelo do circuito e a tecnologia de construção do circuito. Com todos estes fatores em consideração, o grafo, já mostrado (Figura 5.2), apresentava três diferentes caminhos. Esses três caminhos estão representados, a seguir, pela sequência de letras das atividades neles incluídas, como aparecem da direita para a esquerda no grafo:

- a) A – D – G – H
- b) A – C – E – H
- c) A – B – F – H

O primeiro caminho (a) apresenta uma sequência que fere a decisão de só realizar experimentos com *standard cells*. O segundo caminho (b), pelo mesmo motivo do caminho a, não poderia ser realizado. Assim sendo, observando-se o grafo, optou-se pelo último caminho, o c. O caminho c efetua: na atividade A, a recuperação dos consumos resultantes das simulações com somadores bit-paralelo e bit-serial, com modelo de transistores e tecnologia de 130 nm, relatados pelo autor em (Nilsson, 2008); na atividade B, comparações dos resultados obtidos na atividade A, com os consumos obtidos na simulação dos mesmos somadores, porém, desta feita, usando o modelo de *standard cells* e tecnologia de 130 nm; na atividade F, usando apenas o modelo de *standard cells*, foram comparados os consumos obtidos da simulação com os somadores da atividade anterior (B), com os resultados de consumo dos mesmos somadores, com a tecnologia de 28 nm; e, por último, na atividade H, as comparações dos consumos, utilizando as aritméticas bit-paralela e bit-serial, no sistema SPVR, de maior complexidade, com modelo de *standard cells* e tecnologia de 28 nm.

A seguir, serão detalhadas as atividades A, B, F e H, que definem o caminho escolhido para a pesquisa:

- **Atividade A:** Foi a recuperação, por observação gráfica, dos valores dos consumos estáticos e totais dos somadores, utilizando aritméticas bit-paralela e bit-serial, com vazões de até 500 Mega palavras/s para palavras de 12 bits e de até 250 Mega palavras/s para palavras de 24 bits, com modelo de transistores e tecnologia de 130 nm, a partir de medidas nos gráficos, foram extraídos, por meio de proporções, os valores de consumo estimados por Peter Nilsson no artigo "*A Methodology for Arithmetic Reduction of the Static Power Consumption Verified on Filter Architectures*" (NILSSON, 2008), que é o artigo-referência para esta pesquisa. Os gráficos que aparecem no artigo, de onde foram recuperados os valores, foram reconstruídos utilizando planilhas eletrônicas Calc do LibreOffice, alimentadas com os valores de consumos recuperados.
- **Atividade B:** Foram simulados os dois circuitos somadores trabalhados na Atividade A, utilizando, desta feita, o modelo de *standard cells* com as aritméticas *bit-paralela* e *bit-serial*, fazendo uso das mesmas vazões utilizadas no trabalho-referência. Nessa etapa, foi escrito em SystemVerilog o RTL para os somadores utilizados e descritos por Peter Nilsson. Desta vez, porém, foi utilizada a simulação no nível de *standard cells* com a tecnologia de 130 nm, que foi a tecnologia originalmente empregada nos experimentos relatados pelo autor no artigo-referência (NILSSON, 2008) e para que pudessem ser feitas as mesmas avaliações e comparações dos resultados, usando as abordagens de aritméticas bit-paralela e bit-serial. Para essas simulações, foi utilizada a biblioteca CMOS 130 nm da UMC (United Microelectronics Corporation), distribuída pela EUROPRACTICE IC Service, que foi o mesmo fornecedor utilizado por Nilsson em seus experimentos. A ferramenta de simulação foi o Design Compiler da Synopsys, que possibilita estimar os consumos estático e dinâmico do circuito simulado. Nessa atividade, ainda, foram geradas tabelas com os resultados obtidos das simulações e as comparações de consumo, em relação aos dados obtidos no artigo-referência, inclusive, utilizando-se de gráficos comparativos gerados pela ferramenta Calc do LibreOffice.

- **Atividade F:** Nesta etapa foram reutilizados os códigos RTL implementados na atividade anterior (B) para os somadores bit-paralelo e bit-serial, simulados com o modelo *standard cells* e uma tecnologia igual ou menor que 50 nm, como foi sugerido por Peter Nilsson no citado artigo-referência (NILSSON, 2008), para obtenção de redução de consumo estático. Por disponibilidade, foi escolhida a tecnologia de 28 nm, para serem feitas as mesmas avaliações e comparações de consumos que foram feitas na atividade anterior, B, usando as abordagens de aritméticas bit-paralela e bit-serial. Para isso, foi utilizada a biblioteca de *standard cells* CMOS 28 nm da STMicroelectronics, distribuída pela CMP (Circuits Multi-Projets®). Os somadores tiveram seus consumos avaliados e comparados aos resultados obtidos na atividade anterior. Tabelas com os resultados obtidos das simulações e gráficos comparativos foram produzidos, dando origem a um artigo. O software utilizado para as simulações e traçado dos gráficos foram os mesmos da atividade anterior.
- **Atividade H:** Nesta atividade foram feitas as simulações de módulos do sistema SPVR (*Speaker Verification System*), utilizando o modelo de *standard cells* e tecnologia de 28 nm, com aritméticas bit-paralela e bit-serial. Nessa etapa foram reescritos com aritmética bit-serial, o código RTL de três módulos do SPVR, o PE (Pré-ênfase), o WIN (Janelamento) e o LPCA (Análise de Codificação Preditiva Linear), que foram simulados com a biblioteca de *standard cells* com a tecnologia de 28 nm, que foi a tecnologia, abaixo dos 50 nm, escolhida para realizar as simulações na atividade anterior. Estes módulos foram escolhidos por serem mais representativos em termos de utilização de operações aritméticas. Na realidade, o módulo LPCA, por si só seria suficiente para uma boa demonstração da diferença de consumo entre a abordagem com a aritmética bit-paralela e bit-serial pois, este módulo calcula a autocorrelação entre amostras contendo a representação das características vocais do locutor e, a partir de um vetor com esses coeficientes, calcula 12 coeficientes LPCs que representam um conjunto de 220 amostras. A execução desses cálculos envolve as quatro operações aritméticas básicas e mais um inversor de sinal, utilizando complemento de dois. Além destas operações, foram implementados conversores paralelo-serial e serial-paralelo, já que nem todas as operações



podem ser feitas com todos os operandos serializados. Nesta fase, os resultados de consumo das simulações foram coletados, analisados e comparados, individualmente, por módulos. A simulação das duas versões do SPVR, utilizando aritmética bit-paralela e bit-serial, com o modelo *standard cells* e tecnologia de 28 nm gerou um relatório com a análise dos resultados de consumo e as vantagens e desvantagens identificadas no processo de serialização.

Essa sequência de atividades aqui descritas, como já foi dito, foi o roteiro de todo o trabalho de experimentação, coleta, comparação e análise dos resultados, que, por isso, foi chamado de "Caminhos da Pesquisa". Adiante, serão apresentados mais detalhes dos experimentos, os resultados obtidos em cada atividade, bem como os dados serão comparados, analisados e discutidos, justificando as conclusões que serão apresentadas.

## 6 Apresentação e Análise dos Resultados

Nesta seção, são apresentados os resultados obtidos em cada fase definida na metodologia da pesquisa. O sequenciamento das atividades da metodologia foi anteriormente justificado para o alcance dos objetivos da pesquisa e alinhamento com as técnicas que se mostram tendências no desenvolvimento de circuitos digitais. Assim sendo, para cada atividade serão mostrados e analisados os resultados obtidos. De acordo com os objetivos específicos de cada atividade, a forma de apresentação dos resultados variam para que possa ser feita a análise da forma mais esclarecedora. As atividades são identificadas por letras maiúsculas no grafo que mostra o caminho escolhido para a pesquisa (Figura 5.2).

### 6.1 Atividade A – Recuperação dos Dados Obtidos da Substituição de Aritmética Bit-paralela por Bit-serial no Nível de Transistores

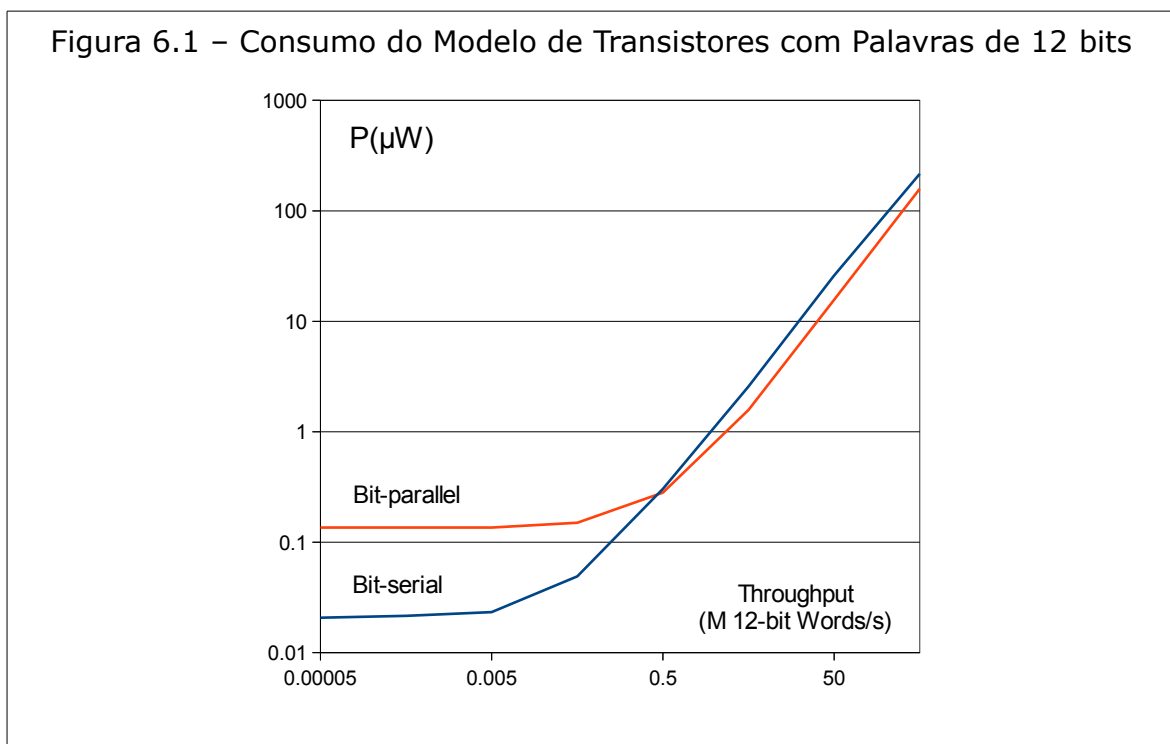
Nesta atividade, não foram realizados experimentos. O objetivo desta fase era coletar os dados obtidos por Peter Nilsson no seu artigo "A Methodology for Arithmetic Reduction of the Static Power Consumption Verified on Filter Architectures" (Nilsson, 2008), que é o artigo-referência para esta pesquisa.

A coleta destes dados foi feita por observação gráfica, já que o artigo não apresenta todos os dados de forma numérica, sendo a maioria deles apresentados de forma gráfica. O artigo apresenta dois gráficos com os consumos totais de energia de somadores utilizando as aritméticas bit-paralela e bit-serial. O primeiro gráfico

Vazão em milhão de palavras de 12 bits	Potência em $\mu\text{W}$		Medida do Gráfico em cm	
	Bit-serial	Bit-parallel	Bit-serial	Bit-parallel
0,000005	0,02	0,13	0,905	3,312
0,00005	0,02	0,14	0,925	3,325
0,0005	0,02	0,14	0,975	3,325
0,005	0,02	0,14	1,075	3,325
0,05	0,05	0,15	2,025	3,450
0,5	0,30	0,28	4,350	4,250
5	2,57	1,58	7,075	6,450
50	26,04	15,64	10,025	9,375
500	216,57	158,24	12,725	12,325

apresenta vazões de até 500 Mega palavras/s, com palavras de 12 bits. O segundo gráfico apresenta vazões de até 250 Mega palavras/s, com palavras de 24 bits. Assim sendo, após ampliar os gráficos e obter os valores proporcionais das medidas, foram construídas duas tabelas com os valores medidos, um para palavras de 12 bits (Tabela 6.1) e outra para palavras de 24 bits (Tabela 6.2).

A partir da Tabela 6.1 foi recuperado o gráfico correspondente ao do trabalho referência, ou seja, o gráfico de consumo para palavras de 12 bits, com vazão de até 500 Mega palavras/s, com modelo de transistores e tecnologia de 130 nm (Figura 6.1).



A partir da Tabela 6.2 também foi recuperado o gráfico correspondente, ou seja, o gráfico de consumo para palavras de 24 bits, com vazão de até 250 Mega palavras/s, com modelo de transistores e tecnologia de 130 nm (Figura 6.2).

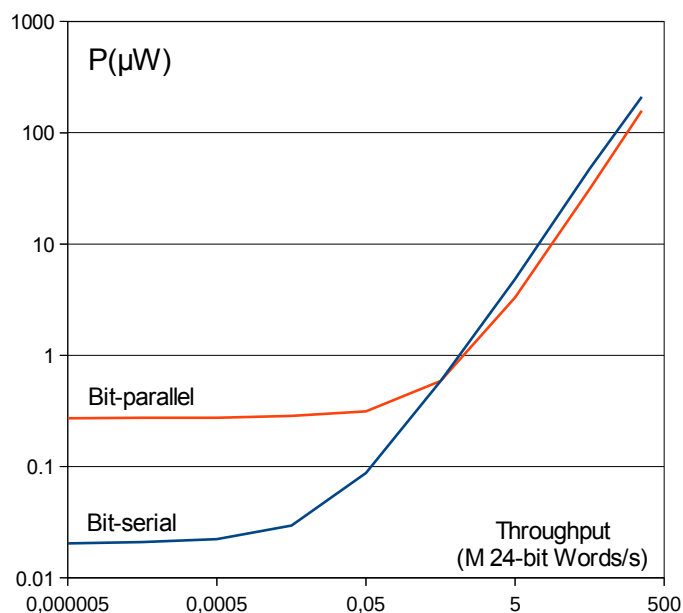
### **Atividade B – Redução de Consumo com Uso de Aritmética Bit-serial em Projetos Baseados em *Standard Cells***

Com o contínuo aumento da automação dos projetos de circuitos digitais, usando abordagens parcialmente personalizadas, o projeto personalizado e desenvolvido manualmente de forma específica foi perdendo espaço. Uma variedade de abordagens

Tabela 6.2 - Consumo do Modelo de Transistores com Palavras de 24 bits

Vazão em milhão de palavras de 24 bits	Potência em $\mu\text{W}$		Medida do Gráfico em cm	
	Bit-serial	Bit-paralelo	Bit-serial	Bit-paralelo
0,000005	0,02	0,27	0,933	4,340
0,00005	0,02	0,27	0,975	4,350
0,0005	0,02	0,27	1,050	4,350
0,005	0,03	0,28	1,425	4,400
0,05	0,09	0,31	2,850	4,525
0,5	0,59	0,59	5,350	5,350
5	4,85	3,32	8,125	7,625
50	47,61	31,32	11,125	10,575
250	210,05	157,89	13,075	12,700

Figura 6.2 - Consumo do Modelo de Transistores com Palavras de 24 bits



foi introduzida, com uso de ferramentas que diminuiriam o trabalho e o tempo aplicados pelos projetistas no desenvolvimento do projeto. Uma das abordagens que vai ao encontro da diminuição de tempo e aumento da automação no processo de desenvolvimento de circuitos digitais é a metodologia de projetos baseado em células padrão (*standard cells*). A ideia por trás desta abordagem é reduzir o esforço de implementação pelo reuso de uma limitada biblioteca de células (KESSLE; GANESA, 1985; RABEY; CHANDRAKASAN; NICOLIĆ, 2003).

A abordagem de *standard cells* padroniza o projeto no nível de portas lógicas.

Nessa abordagem é fornecida uma biblioteca contendo uma larga seleção de portas lógicas pré-caracterizadas. Essas portas prontas são acessíveis ao projetista, sendo selecionadas e instanciadas automaticamente pela ferramenta, na síntese do projeto. Antes, as bibliotecas de *standard cells* simplificavam o processo do projeto, em detrimento da otimização da área, do desempenho e do consumo de energia. Hoje as bibliotecas incluem muitas versões de cada porta, dimensionadas para diferentes níveis de tensões, desempenho e consumo, que são selecionadas pela ferramenta de síntese, a partir dos requisitos impostos ao projeto.

O objetivo dessa atividade foi verificar se o comportamento dos dados, resultantes no trabalho-referência, se conservariam utilizando *standard cells*, fazendo uso dos somadores descritos pelo autor no artigo-referência (Nilsson, 2008), com aritméticas bit-serial e bit-paralela. Nessa fase, foram feitas as sínteses e as estimativas de consumo, utilizando os mesmos tamanhos de palavras (12 e 24 bits) e as mesmas vazões, até 500 Mega palavras/s para palavras de 12 bits e até 250 Mega palavras/s para palavras de 24 bits. As sínteses foram realizadas utilizando bibliotecas de *standard cells* com caracterização dos modelos para condições típicas (temperatura de 25 °C) com tecnologia de 130 nm, fornecida pela UMC, a mesma tecnologia relatada no trabalho de Nilsson. Isso, para manter a compatibilidade com o trabalho que estamos tomando por referência, no que diz respeito a tecnologia, fornecedor e condições.

A partir dos consumos obtidos nas sínteses com o modelo de *standard cells*, juntamente com os consumos recuperados do modelo de transistores, foram construídas a Tabela 6.3, para palavras de 12 bit e a Tabela 6.4 para palavras de 24 bits. A partir dessas tabelas, foram traçados os gráficos da Figura 6.3 e da Figura 6.4. O primeiro gráfico mostra os resultados obtidos para palavras de 12 bits e o segundo, para palavras de 24 bits.

Foram encontradas diferenças significativas de valores de consumo de energia entre os dados resultantes do trabalho referenciado do Nilsson e o consumo obtido na síntese utilizando *standard cells*. Com *standard cells* as reduções de consumo foram mais acentuadas do que as obtidas por Nilsson nas mesmas condições: com o uso dos dois tamanhos de palavras (12 e 24 bits) e aplicando as duas aritméticas, bit-paralela e bit-serial.

Tabela 6.3 – Consumo do Modelo de Transistores 130 nm vs. Standard Cells  
130 nm com Palavras de 12 bits

Vazão em milhão de palavras de 12 bits	Potência em $\mu\text{W}$			
	Transistores 130nm		<i>Standard Cells 130nm</i>	
	Bit-serial	Bit-parallel	Bit-serial	Bit-parallel
0,000005	0,02	0,13	0,002	0,01
0,000050	0,02	0,14	0,002	0,01
0,000500	0,02	0,14	0,002	0,01
0,005000	0,02	0,14	0,003	0,01
0,050000	0,05	0,15	0,01	0,01
0,500000	0,30	0,28	0,14	0,09
5,000000	2,57	1,58	1,34	0,84
20,833333			5,56	3,49
50,000000	26,04	15,64	13,35	8,37
185,185185			49,45	31,00
390,625000				65,38
500,000000	216,57	158,24		

Os resultados apresentados nos gráficos da Figura 6.3 e da Figura 6.4, utilizando *standard cells* não chegam até a vazão máxima utilizada por Nilsson (até 500 Mega palavras/s para palavras de 12 bits e até 250 Mega palavras/s para palavras

Tabela 6.4 – Consumo do Modelo de Transistores 130 nm vs. Standard Cells  
130 nm com palavras de 24 bits

Vazão em milhão de palavras de 24 bits	Potência em $\mu\text{W}$			
	Transistores 130nm		<i>Standard Cells 130nm</i>	
	Bit-serial	Bit-parallel	Bit-serial	Bit-parallel
0,000005	0,02	0,27	0,002	0,13
0,000050	0,02	0,27	0,002	0,13
0,000500	0,02	0,27	0,002	0,13
0,005000	0,03	0,28	0,004	0,13
0,050000	0,09	0,31	0,03	0,16
0,500000	0,59	0,59	0,27	0,48
5,000000	4,85	3,32	2,67	3,71
10,416667			5,56	7,60
50,000000	47,61	31,32	26,70	36,01
92,592593			49,45	66,58
190,839695				137,10
250,000000	210,05	157,89		

Figura 6.3 – Comparação dos resultados com palavras de 12 bits

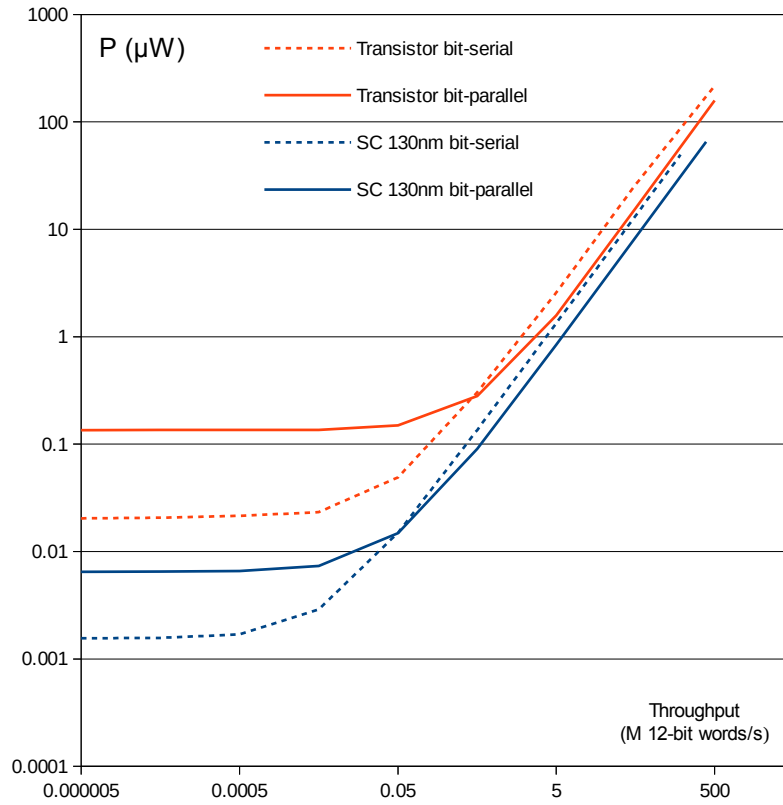
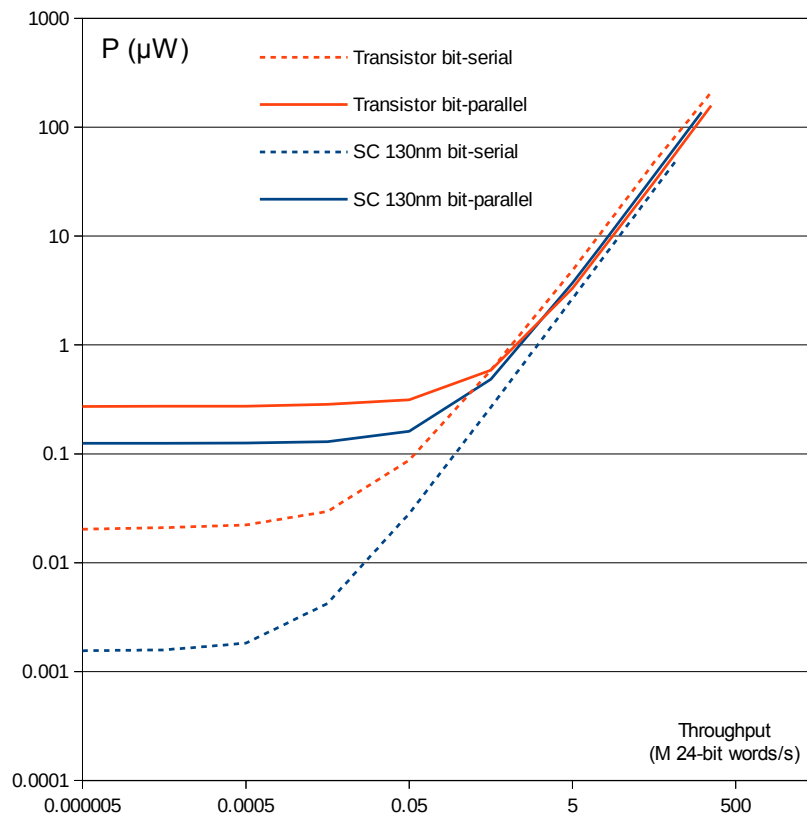


Figura 6.4 – Comparação dos resultados com palavras de 24 bits



de 24 bits), porque os atrasos das *standard cells* foram considerados e não permitiriam vazões até este limite.

O modelo de *standard cells* apresenta redução do consumo de energia estática em todas as situações consideradas com as duas aritméticas (bit-serial e bit-paralela) e os dois tamanhos de palavras (12 e 24 bits), como pode ser observado nos gráficos apresentados.

No modelo de *standard cells* com palavras de 12 bits, semelhantemente às curvas do modelo de transistores com as duas aritméticas, continua a haver um cruzamento das linhas, indicando que, com maiores vazões, o consumo com a aritmética bit-serial ultrapassa o consumo total com a aritmética bit-paralela. Porém, com *standard cells*, o cruzamento, que indica consumos iguais para as duas aritméticas, ocorre com uma vazão menor. Com o modelo de *standard cells* e tamanho de palavra de 24 bits, não há cruzamento, mantendo-se o consumo total, com o uso da aritmética bit-serial, sempre menor, muito embora haja uma tendência de equalização dos consumos totais para vazões mais altas.

Observando os gráficos, é visto que para vazões menores – parte direita dos gráficos – onde o consumo de energia estática é predominante, os pontos da curva de consumo total da aritmética bit-serial está sempre abaixo dos pontos da linha da curva de consumo da bit-paralela. No modelo de *standard cells*, palavras de 12 bits e com a menor vazão, o consumo total com a aritmética bit-serial é 0,002  $\mu\text{W}$  e com a aritmética bit-paralela é 0,006  $\mu\text{W}$ , ou seja, o consumo total com a aritmética bit-serial é 3,0 vezes menor. Já na mesma situação, com palavra de 24 bits, o consumo total com a aritmética bit-serial é 0,002  $\mu\text{W}$  e com a aritmética bit-paralela é 0,125  $\mu\text{W}$ , ou seja, o consumo total com a aritmética bit-serial é 62,5 vezes menor.

## **6.2 Atividade F – Uso de Aritmética Bit-serial para Redução de Consumo com Tecnologia Submicrônica vs. Nanométricas**

No artigo-referência (Nilsson, 2008), o autor afirma que, com tecnologias mais densas abaixo de 65 nm, haveria o domínio do consumo da energia estática sobre o consumo de energia dinâmica e que, utilizando a técnica de substituição de aritmética bit-paralela por aritmética bit-serial, haveria uma redução de consumo estático. Nesta fase da pesquisa confrontou-se os resultados de consumo das duas versões dos



somadores (bit-paralelo e bit-serial) utilizados na atividade anterior. As estimativas de consumo foram feitas utilizando bibliotecas de *standard cells* de 130 nm e 28 nm. A biblioteca de 130 nm foi a mesma utilizada na atividade anterior (B), fabricada pela UMC. As bibliotecas de 28 nm foram as bibliotecas C28SOI\_SC\_8\_CORE\_LL e a C28SOI\_SC\_8\_COREPBP16\_LL (referenciadas no texto e nas tabelas como PBP0 e PBP16) da tecnologia CMOS028FDSOI, fabricada pela STIMicroelectronics. A razão de ter-se escolhido duas bibliotecas de 28 nm é que a tecnologia selecionada de 28 nm oferecia várias opções de bibliotecas. Assim, como o objetivo era reduzir consumo de energia nas simulações, foram escolhidas duas bibliotecas que implementam “*Low Power*” e “*Low Vt*” (“\_LL” na referência da biblioteca), pois geram circuitos com menor consumo energético. A diferença entre elas é que a primeira gera circuitos com um consumo estático maior, a PBP0 (*poly bias* de 0 nm), pois, seus transistores têm a largura do canal padrão (28 nm), gerando assim uma fuga maior entre o *drain* e o *source em relação a* segunda, a PBP16 (*poly bias* de 16 nm), que tem a largura do canal com um adicional de 16 nm, dando um total de 44 nm (FLATRESSE, 2014) e gera circuitos com o consumo estático menor. Para todas as bibliotecas, tanto a de 130 nm como as de 28 nm foram utilizadas versões com condições típicas, com temperatura ambiente (25 °C).

Os experimentos foram executados sob as mesmas condições da atividade anterior, ou seja, foram feitas simulações com somadores, utilizando as aritméticas bit-paralela e bit serial com tecnologia de 28 nm com “*poly bias*” de 0 e 16 nm (FLATRESSE, 2014), e, tal qual no experimento anterior, usou-se duas vazões – até 500 Mega palavras/s para palavras de 12 bits e até 250 Mega palavras/s para palavras de 24 bits.

Os resultados obtidos com as simulações na atividade anterior (B) para os dois tipos de somadores simulados com a biblioteca de 130 nm são apresentados pelas Tabela 6.5, para palavras de 12 bits e pela Tabela 6.6, para palavras de 24 bits, juntamente, com os resultados das simulações com as duas bibliotecas de *standard cells* (PBP0 e PBP16) de tecnologia de 28 nm. As comparações dos valores dos consumos resultantes das simulações são apresentadas, por meio de gráficos, na Figura 6.5, para palavras de 12 bits e na Figura 6.6 para palavras de 24 bits.

Observando os gráficos, claramente, nota-se que não há uma regra geral das relações entre as tecnologias e a aplicação das aritméticas bit-paralelo e bit-serial. No

Tabela 6.5 - Consumo do Modelo de Standard Cells com Tecnologia 130 nm vs. 28 nm com Palavras de 12 bits

Vazão em milhão de palavras de 12 bits	Potência em $\mu$ W					
	<i>Standard Cells 130nm</i>		<i>Standard Cells 28nm PBP0</i>		<i>Standard Cells 28nm PBP16</i>	
	Bit-serial	Bit-parallel	Bit-serial	Bit-parallel	Bit-serial	Bit-parallel
0,000005	0,00	0,01	0,03	0,11	0,0004	0,002
0,000050	0,00	0,01	0,03	0,11	0,0004	0,002
0,000500	0,00	0,01	0,03	0,11	0,0004	0,002
0,005000	0,00	0,01	0,03	0,11	0,001	0,002
0,050000	0,01	0,01	0,03	0,11	0,002	0,002
0,500000	0,14	0,09	0,04	0,11	0,02	0,004
5,000000	1,34	0,84	0,20	0,13	0,19	0,03
20,833333	5,56	3,49	0,77	0,18	0,80	0,12
50,000000	13,35	8,37	1,81	0,29	1,91	0,28
185,185185	49,45	31,00	6,64	0,78	7,08	1,03
378,787879		63,39	13,56	1,48	14,49	2,11
390,625000		65,38	13,98	1,52		2,18
500,000000			17,88	1,91		2,79
595,238095			21,29	2,26		3,32
961,538462				3,58		5,36
1515,151515				5,57		

gráfico para as palavras de 12 bits, os menores consumos foram obtidos com a biblioteca PBP16 e o mesmo aconteceu para as palavras 24 bits. Observando os resultados obtidos com as demais bibliotecas, a de 130nm e a PBP0, vê-se que, em algumas situações, a biblioteca PBP0, que é de 28 nm, produz consumos maiores que a biblioteca de 130 nm. Este fato contradiz a afirmação que encontra-se no artigo-

Tabela 6.6 - Consumo do Modelo de Standard Cells com tecnologia 130 nm vs. 28 nm com Palavras de 24 bits

Vazão em milhão de palavras de 24 bits	Potência em $\mu$ W					
	<i>Standard Cells 130nm</i>		<i>Standard Cells 28nm PBP0</i>		<i>Standard Cells 28nm PBP16</i>	
	Bit-serial	Bit-parallel	Bit-serial	Bit-parallel	Bit-serial	Bit-parallel
0,000005	0,00	0,13	0,03	0,22	0,0004	0,003
0,000050	0,00	0,13	0,03	0,22	0,0004	0,003
0,000500	0,00	0,13	0,03	0,22	0,0004	0,003
0,005000	0,00	0,13	0,03	0,22	0,001	0,003
0,050000	0,03	0,16	0,03	0,22	0,004	0,005
0,500000	0,27	0,48	0,06	0,23	0,04	0,02
5,000000	2,67	3,71	0,38	0,29	0,38	0,13
10,416667	5,56	7,60	0,77	0,37	0,80	0,26
50,000000	26,70	36,01	3,60	0,94	3,83	1,23
92,592593	49,45	66,58	6,64	1,55	7,08	2,27
189,393939		136,06	13,56	2,94	14,49	4,64
190,839695		137,10	13,66	2,96		4,67
250,000000			17,88	3,81		6,12
297,619048			21,29	4,49		7,28
478,468900				7,09		11,71
769,230769				11,26		

Figura 6.5 - Consumo do Modelo de Standard Cells com tecnologia 130 nm vs. 28 nm com Palavras de 12 bits

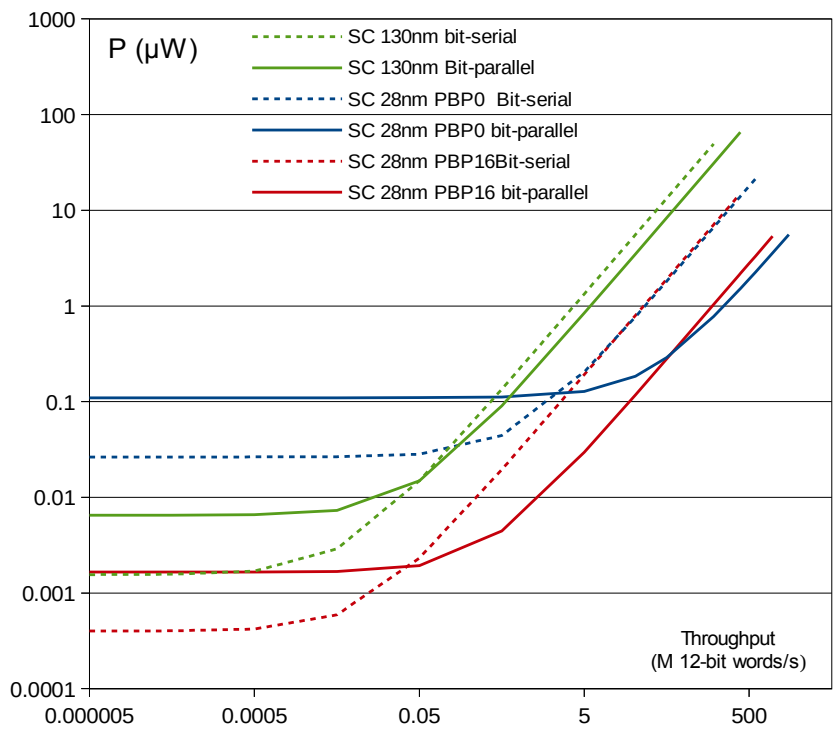
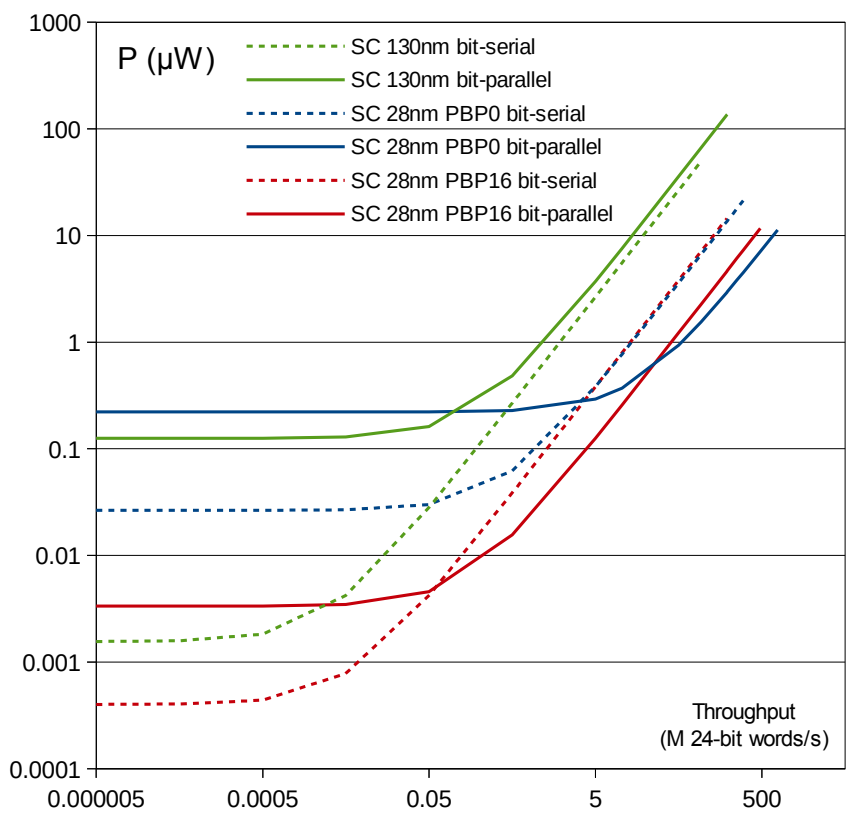


Figura 6.6 - Consumo do Modelo de Standard Cells com tecnologia 130 nm vs. 28 nm com Palavras de 24 bits



referência (Nilsson, 2008), que diz que com tecnologias mais densas que 50 nm o consumo estático sempre será menor que o de tecnologias abaixo dos 50 nm.

### **6.3 Atividade H – Aplicação de Aritmética Bit-serial para Redução de Consumo em um Sistema de Maior Complexidade**

Nesta atividade, o objetivo foi aplicar aritmética bit-serial ao SPVR (*Speaker Verification System*), para obter, assim, redução do consumo de energia estática e, conseqüentemente, redução do consumo total do sistema. Nessa fase da pesquisa concentraram-se os maiores desafios pois, a princípio não se tinha clareza do que poderia ser ou não serializado na seqüência do fluxo de execução das operações e armazenamento de dados. Descobriu-se ao longo do processo de serialização que a tarefa tinha um custo maior do que se previa, como será detalhado adiante. Foram feitas as simulações e estimativas de consumo com a ferramenta Design Compiler da Synopsys.

#### **6.3.1 O Handshake dos Módulos Seriais do SPVR**

A primeira tarefa que precisou ser realizada foi a adaptação do *handshake* (protocolo para interconexão e gerenciamento de blocos funcionais em um IP core ou SoC) do SPVR. O SPVR na sua versão original, utilizando aritmética bit-paralela, faz uso do protocolo AMBA<sup>®</sup> AXI (ARM, 2004). Como na nova versão dos módulos do SPVR que utilizariam aritmética bit-serial, as palavras de dados, que pudessem ser serializadas, fluiriam como uma cadeia de bits, tendo como primeiro bit, o do início da palavra, o *LSB (Least Significant Bit)*. No *handshake* adaptado, para indicar o início de cada palavra, seria levantado (valor 1) o sinal *start*, que seria abaixado (valor 0) logo que o LSB fosse consumido pelo bloco, ou módulo, receptor. Para o módulo receptor indicar que estava pronto para receber uma palavra serializada, levantaria o sinal *ready* (valor 1), que após começar a consumir a palavra serializada seria baixado (valor 0) pelo receptor. Uma palavra serializada, após ser iniciado o seu envio, seria transmitida até o seu final, sem interrupções.

#### **6.3.2 A Serialização dos Módulos com a Aritmética Bit-serial**

Num sistema utilizando aritmética bit-serial, para que uma operação completamente serial, como é o caso da soma, fosse executada, assim que os *LSBs* dos operandos de entrada fossem recebidos, começaria a operação soma serial e seria,

simultaneamente, dado início à saída com *LSB* da palavra resultante da soma (WANHAMMAR, 1999). Melhor explicando, cada par de bits dos operandos de entrada, um bit do primeiro operando com um bit do segundo operando, que possuem mesma posição relativa ao *LSB*, são somados a cada entrada do par, usando um *full-adder* (somador completo), já mostrado na Figura 2.5. Esse mesmo somador será usado para somar o par de bits da vez, mais o *carry in*, que é o *carry out* ("vai-um") resultante da soma do par de bits anterior. O primeiro *carry in*, que deve ser somado ao par *LSB*, deve ser zero. A soma cujos operandos têm sinal, também torna-se simples, pois um operando com sinal negativo é representado por seu complemento de dois e a soma com eles se dá da mesma forma. Caso a cadeia de bits de um operando termine primeiro, um operando tem tamanho de palavra menor, seu bit de sinal deve ser estendido até alcançar a quantidade de bits da palavra maior.

A subtração serial com sinal processa-se de forma semelhante à soma, já que subtrair o subtraendo do minuendo é o mesmo que somar o minuendo com o complemento de dois do subtraendo (WANHAMMAR, 1999).

Pelo apresentado na subtração, vê-se que é necessário ter um bloco, ou módulo, inversor de sinal, para o caso em que o operador seja serial e positivo e vá participar da operação com o sinal negativo. Esse módulo, na realidade, é um módulo que executa e envia para a saída, serialmente (bit a bit), o complemento de dois da entrada.

Infelizmente, quando trata-se de serialização de operações, "nem tudo são flores", como nos casos da soma e da subtração. A multiplicação serial é possível e existe algoritmo para isto, porém, exige um circuito muito complexo, extenso e que usa muitos *flip-flops* do tipo D (WANHAMMAR, 1999), o que leva a utilização de muitos transistores e, conseqüentemente, a um consumo de energia estática maior. Para a aplicação de operações seriais na multiplicação, optou-se por utilizar um multiplicador serial/paralelo. Nesse multiplicador, o multiplicando é serial, o operando multiplicador é paralelo e, na medida em que cada bit do multiplicando entra, multiplica cada bit do operando multiplicador. Os resultados das multiplicações parciais bit a bit são acumuladas e deslocadas para direita, dando saída a um bit do produto a cada deslocamento, com o bit *LSB* vindo em primeiro lugar.

A operação de divisão não é possível executar serialmente. Pois os algoritmos

que se tem, são variações daquele que aprendemos na escola e sempre produz o bit *MSB (Most Significant Bit)* em primeiro lugar. Isso vai de encontro ao fluxo de dados seriais do sistema SPVR, que sempre envia uma palavra serial com o LSB na frente. Por isso, a operação de divisão recebe os operandos dividendo e divisor paralelos e produz um quociente serial que tem sua sequência de bits invertida, para colocar o LSB na frente.

Pelo que foi apresentado para a serialização de operações aritméticas, nota-se que os operandos variam no seu tipo, ou seja, podem ser seriais e/ou paralelos dependendo da operação a ser executada. Daí, dependendo da sequência das operações nos blocos do SPVR, foi necessário criar blocos de conversão paralelo/serial e serial/paralelo. Outro fato que leva à demanda desses conversores é que as palavras dos dados intermediários são armazenadas em registradores paralelos.

Como já foi dito, aplicou-se aritmética bit-serial em três módulos do SPVR, o PE-*Preemphasis* (pré-ênfase), o WIN-*Windowing* (janelamento) e LPCA-*Linear Prediction Coding Analysis* (análise de codificação preditiva linear). Estes módulos, pela ordem, apresentam níveis de complexidade e tamanhos crescentes.

Tendo as versões bit-paralela e bit-serial dos três módulos operacionais, foram esses módulos simulados, utilizando a ferramenta *Design Compiler* da *Synopsys*, que fornece as estimativas do consumo energético do módulo simulado. O *clock* utilizado para a versão com aritmética bit-serial foi aproximadamente 30 vezes maior que o *clock* utilizado para a versão com aritmética bit-paralela, com o intuito de alcançar a mesma vazão nas simulações. O fator 30 foi escolhido porque as maiores palavras com que se estava lidando na versão bit-serial chegavam a 31 bits e, como na aritmética bit-serial, as operações são feitas bit a bit e cada operação com um bit é executada num ciclo de *clock*. Diferentemente, a aritmética bit-paralela opera palavras inteiras, em um único ciclo de *clock*. O *clock* que foi utilizado para a versão bit-paralela foi de 750 kHz e para a versão serial de 22.500 kHz, ou seja, 22,5 MHz. A tecnologia escolhida de 28 nm, oferecia várias opções de bibliotecas. Foram escolhidas bibliotecas que implementam “*Low-power*” e “*Low Vt*” (“\_LL” na referência da biblioteca) pois, oferecem menor consumo energético. Foram utilizadas nestas simulações, duas versões de bibliotecas de *standard cells* de tecnologia de 28 nm com “*Low-power*” e “*Low Vt*”: a biblioteca C28SOI\_SC\_8\_CORE\_LL (referenciada como PBP0) e a C28SOI\_SC\_8\_COREPBP16\_LL (referenciada como PBP16). Como já foi

dito, a diferença entre elas é que a primeira gera circuitos com um consumo estático maior, pois os transistores têm a largura do canal padrão (28 nm), gerando assim uma fuga maior entre o *drain* e o *source*. A segunda, de forma inversa, tem a largura do canal com um adicional de 16 nm, dando um total de 44 nm (FLATRESSE, 2014) e gera circuitos com o consumo estático menor.

### 6.3.2.1 O Módulo Pré-ênfase (PE)

O pré-processamento das amostras de voz tem a finalidade de prepará-las para a extração de seus parâmetros e/ou características. O pré-processamento, entre outras medidas, inclui a Pré-ênfase (PE).

Durante esta etapa, o sinal de voz passa através de um filtro de resposta de impulso infinito (FIR), filtro de primeira ordem conhecido como filtro de pré-ênfase. O objetivo desse filtro é atenuar os componentes do sinal de baixa frequência, precauções contra a estabilidade numérica (aritmética de precisão finita) e também minimizar os efeitos dos conjuntos de radiação de lábios e variação da área da glote.

Como o sistema seria implementado em hardware, ele fez o uso de operações de ponto fixo com número fracionário, simplificações de funções matemáticas, substituição de operações de divisão e multiplicação por deslocamento, paralelismo, entre outros (ALVES et al., 2009). A seguir tem-se a Equação 6.1 que define o filtro de pré-ênfase:

Filtro de Pré-ênfase ( $s$ ,  $\alpha$ ):

$$S_{pre}(n) = s(n) - \alpha s(n-1) \quad (6.1)$$

Em que:

$S_{pre}$  é o sinal de voz depois da Pré-ênfase;

$s$  é o sinal de voz vindo do VD;

$\alpha$  é o coeficiente de pré-ênfase. Na maioria dos casos, usa-se o valor 0,95 para *alfa*, mas, para simplificar a implementação do filtro de pré-ênfase, *alfa* foi substituído por 5/16, que corresponde a 0,9375:

$$S_{pre}(n) = s(n) - \frac{15}{16}s(n-1) = s(n) - s(n-1) + (s(n-1) \gg 4) \quad (6.2)$$

No módulo PE, o menor e mais simples dos três módulos selecionados do SPVR para a serialização, quase todas as operações puderam ser feitas serialmente, sem necessidade de armazenamento intermediário, de acordo com a última equação após o sinal de igualdade na Equação 6.2.

O módulo PE serial, que foi implementado para esta pesquisa, recebe serialmente as amostras vindas do módulo VD (detector de voz) e envia as amostras pré-enfatizadas para o módulo WIN (janelamento). Para isso, faz uso de três módulos com entradas e saídas seriais: o *delayer\_shift()*, que recebe o valor de uma amostra  $s(n)$ , retornando o valor da amostra anterior  $s(n-1)$  e  $s(n)$  deslocada 4 bits para direita que equivale ao valor da amostra ter sido dividido por 16; o *serial\_adder()*, que é instanciado duas vezes, uma para executar a soma e a outra para subtração (como foi explicado na Seção 6.3.2); e, por último, o módulo *saturation()*, que executa ajustes, de valores que ultrapassam o menor e o maior valores admissíveis, para esses limites.

A Tabela 6.7, apresenta os resultados de consumos obtidos da simulação das versões bit-paralela e bit-serial do PE.

Consumo do Módulo PE-Pré-ênfase (com saturação) com Modelo de Standard Cells e Tecnologia de 28 nm ( $\mu W$ )						
Tipos de Consumo	Aritmética Bit-paralela		Aritmética Bit-serial		Redução de Consumo por Mudança de Aritmética	
	Bibliotecas SC 28nm		Bibliotecas SC 28nm			
	PBP0	PBP16	PBP0	PBP16	PBP0	PBP16
Estático	1,67	0,02	1,11	0,02	33,68%	34,63%
Dinâmico	0,17	0,18	4,00	4,27	-2268,15%	-2261,77%
Total	1,84	0,20	5,11	4,29	-177,30%	-2001,86%

Como era esperado, houve uma redução de consumo de energia estática bastante significativa, da ordem de 1/3 do consumo, porém, houve um grande crescimento, não esperado, do consumo dinâmico. Investigadas as razões para este crescimento espantoso de consumo dinâmico, descobriu-se que a origem desse



grande crescimento estava na utilização de registradores e o módulo que mais fazia uso de registradores no PE, era o *saturation()*. O processo de eliminar as saturações decorrentes da aplicação do filtro exige que as amostras sejam paralelizadas para haver as comparações, pois, o fluxo de dados seriais leva o LSB sempre na frente da palavra e as comparações exigem que sejam feita a partir do MSB. Para certificar-se disso, o PE foi simulado sem o módulo de saturação e assim, obteve-se resultados menores do consumo dinâmico.

A Tabela 6.8 apresenta esses resultados, obtidos sem a saturação, o que demonstra a influência do uso dos registradores no consumo dinâmico. Porém esse uso de registradores que impacta num crescimento do consumo dinâmico, também ocorre na localização do ponto de deslocamento para que seja operada a divisão, bem como, para o armazenamento da palavra anterior  $s(n-1)$  (vide Equação 6.2). Assim sendo, como a saturação e as demais operações que utilizam-se de registradores paralelos não podem ser evitadas por necessidade de funcionalidade, a luta contra o consumo dinâmico é praticamente invencível, dependendo das operações necessárias para executar as funcionalidades especificadas para o módulo serial.

Tabela 6.8 – Consumo do PE (sem saturação)

Consumo do Módulo PE-Pré-ênfase (sem saturação) com Modelo de Standard Cells e Tecnologia de 28 nm ( $\mu W$ )						
Tipos de Consumo	Aritmética Bit-paralelo		Aritmética Bit-serial		Redução de Consumo por Mudança de Aritmética	
	Bibliotecas SC 28nm		Bibliotecas SC 28nm			
	PBP0	PBP16	PBP0	PBP16	PBP0	PBP16
Estático	1,67	0,02	0,58	0,00	65,20%	96,54%
Dinâmico	0,17	0,18	2,06	0,16	-1117,29%	11,27%
Total	1,84	0,20	2,64	0,16	-43,18%	20,92%

Como foi dito na Seção 3.1, o SPVR foi implementado em um circuito dedicado capaz de ter desempenho suficiente para funcionar em tempo real mesmo empregando um sinal de *clock* lento. O módulo PE recebe amostras vindas do módulo VD (*Voice Detector* – detector de voz). Uma amostra é o menor pedaço de um sinal de voz digital, aqui representado por valores de 16 bits a uma taxa de frequência de 11.025 Hz (TEIXEIRA, 2009). Assim sendo, foi observado que o módulo PE poderia trabalhar com um *clock* proporcional à frequência de amostragem, ou seja, poder-se-

ia trabalhar com um *clock* mais lento que o de 750 kHz originalmente utilizado, um *clock* de 11,025 kHz para a versão paralela e, por conseguinte, 330,75 kHz para versão serial, utilizando o fator de multiplicação 30 como foi justificado na Seção 6.3.2. Foram feitas simulações com esses *clocks* mais lentos com as versões paralela e serial do módulo PE. As Tabelas 6.9 e 6.10 mostram, respectivamente, o consumo do módulo PE, nas versões paralela e serial, com saturação, simuladas com os *clocks* mais lentos.

### **6.3.2.2 O Módulo de Janelamento (WIN)**

O módulo de janelamento WIN (Windowing) recebe o sinal de voz, em amostras pré-enfatizadas e divide-as em blocos (*frames* ou janelas), que são formados por uma função de janelamento.

O WIN utiliza o janelamento de Hamming, priorizando as características espectrais do centro da onda e dá peso mais baixo para as transições das extremidades. Para evitar a perda de informações importantes, localizadas nestas extremidades, durante a segmentação do sinal em blocos, é realizada uma sobreposição de quadros de 50%. Assim, uma nova janela se forma com a metade final da janela anterior e a metade inicial da janela atual. Essa nova janela contém, no centro, informações que seriam ponderadas, antes, com valores baixos, cobrindo as áreas não enfatizadas nas extremidades da janela (PAIVA, 2009).

Os valores da Janela de Hamming armazenados são acessados e multiplicam as amostras pré-enfatizadas (saída do módulo de PE). As amostras que passam por este processo são submetidas a uma superposição de 50%, conforme mostrado no pseudo-código do algoritmo de sobreposição usando 220 amostras, no documento "*Windowing Specification*" (PAIVA, 2009), para evitar a perda de informações importantes. O resultado, as amostras janeladas, são enviadas ao módulo LPCA.

O janelamento bit-serial não foi implementado com sobreposição. Em vez disso, optou-se por fazer a estimativa do consumo do módulo, duplicando o consumo estimado pela ferramenta de simulação.

Para executar o janelamento, a versão bit-serial do WIN faz uso de três módulos: o *rom\_win\_serial()*, que recebe a posição da amostra a ser multiplicada, pesquisa numa tabela valor da Janela de Hamming correspondente e retorna o valor

encontrado; o *sp\_multiplier()*, que multiplica o valor encontrado da Janela de Hamming pela amostra pré-enfatizada, bit a bit, gerando a amostra janelada; e o *truncate\_sample()*, que recebe serialmente a amostra janelada e a trunca para o seu tamanho normal (16 bits), permanecendo com seu formato serial.

As duas versões do WIN, com a aritmética bit-paralela e com a aritmética bit-serial foram simuladas e os valores de consumo estimados são apresentados na Tabela 6.9. Neste módulo, a redução do consumo estático foi muito significativa, chegando até a 94,06%, com a biblioteca PBP16. Isto confirma a afirmação no artigo-referência (Nilsson, 2008), que quando há predominância de operações aritméticas sobre uso de armazenamento em registradores, a redução de consumo se acentuará.

Tabela 6.9 – Consumo do WIN

Consumo Módulo de Janelamento (WIN) com Modelo de Standard Cells com Tecnologia de 28 nm e Consumo Bit-serial Duplicado( $\mu$ W)						
Tipos de Consumo	Aritmética Bit-paralela		Aritmética Bit-serial		Redução de Consumo por Mudança de Aritmética	
	Bibliotecas SC 28nm		Bibliotecas SC 28nm			
	PBP0	PBP16	PBP0	PBP16	PBP0	PBP16
Estático	57,81	0,75	3,24	0,04	94,40%	94,06%
Dinâmico	6,46	6,90	8,76	9,35	-35,64%	-35,49%
Total	64,27	7,65	11,99	9,40	81,34%	-22,79%

### 6.3.2.3 O Módulo Análise de Codificação Preditiva Linear (LPCA)

A ideia principal do LPCA (Análise de Codificação Preditiva Linear) consiste em que uma amostra de voz pode ser aproximada como uma combinação das amostras passadas, dentro de um *frame* de amostras que é gerado por um processo de janelamento. Essa aproximação é feita por meio da minimização da soma dos quadrados das diferenças entre as amostras reais e as amostras preditivas, obtidas de uma combinação linear das amostras anteriores. Assim, um conjunto de doze 12 coeficientes é calculado por este módulo para cada *frame* recebido. Este conjunto de coeficientes é exclusivo para cada pessoa. Assim, esse conjunto pode ser usado para determinar a pessoa a que pertence aquele *frame* de voz. O método mais utilizado é o método de autocorrelação, para gerar os coeficientes de autocorrelação, com o algoritmo de Levinson-Durbin, para gerar os coeficientes LPCs (TEIXEIRA, 2009).

Reforçando, uma amostra é o menor pedaço de um sinal de voz digital, aqui representado por valores de 16 bits a uma taxa de frequência de 11.025 Hz. Um *frame*, ou janela, é uma sequência de 220 amostras geradas pelo processo de janelamento.

O LPCA tem uma função de calcular um conjunto de 12 coeficientes LPCs para cada *frame* de amostras que recebe da entrada e colocar esses coeficientes no módulo de saída. Chama-se os coeficientes de  $a_1, a_2, a_3, \dots, a_{12}$ , e a maneira usada aqui para encontrar estes coeficientes é a aplicação do algoritmo de Levinson-Durbin para os coeficientes de autocorrelação.

Os coeficientes de autocorrelação são calculados pela expressão:

$$\forall k=0, \dots, K : r(k) = \sum_{n=k}^{N-1} s(n) \cdot s(n-k) \quad (6.3)$$

Em que:

$s(n)$  é a  $n$ -ésima amostra do quadro;

$K$  é o número de coeficientes (12); e,

$N$  é o número de amostras em um quadro ou o tamanho da janela (220).

No Quadro 6.1 é apresentado, em pseudo-código, o algoritmo para o cálculo dos coeficientes autocorrelacionados, onde *Sample* tem o valor 220 e *coefVectSize* tem o valor 12.

Quadro 6.1 – Pseudo-código para o Cálculo dos Coeficientes Autocorrelacionados

```

PROCEDURE Autocorrelation(
  VAR Sample      : INTEGER;
  VAR coefVectSize : INTEGER;
  VAR s           : ARRAY [0..Sample - 1] OF REAL;
  VAR r           : ARRAY [0..coefVectSize] OF REAL;
)

VAR
  k, n : INTEGER;

BEGIN {Autocorrelation}
  FOR n := 0 TO Sample - 1 DO
    FOR k := 0 TO min(n, coefVectSize) DO
      r[k] := r[k] + s[n] * s[n - k];
    END;
  END;

```

O algoritmo de Levinson-Durbin, em pseudo-código, para o cálculo dos coeficientes LPCs, é apresentado no Quadro 6.2, com o valor de *coefVectSize* igual a 12.

Quadro 6.2 – Pseudo-código para o Cálculo dos Coeficientes LPCs

```

PROCEDURE LevinsonDurbin(
VAR coefVectSize : INTEGER; {Predictor order}
VAR r              : ARRAY [0..coefVectSize] OF REAL; {Autocorrelation
array}
VAR a              : ARRAY [0..coefVectSize] OF REAL; {Predictor}
                                {coefficients}
VAR lpca_pm_coeflpc : ARRAY [0..(coefVectSize - 1)] OF REL; {Output}
VAR Alfa          : REAL; {Error variance}
)

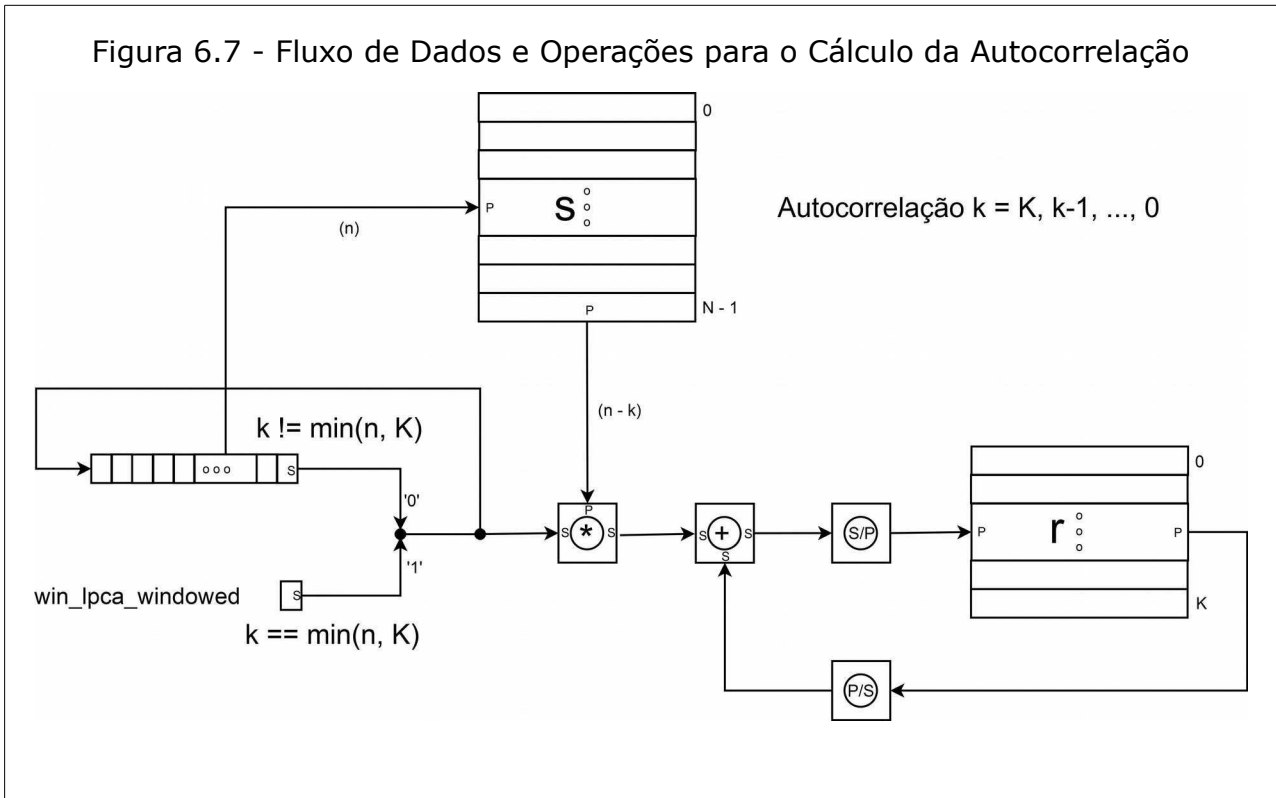
VAR {For LevinsonDurbin}
m, k      : INTEGER;
b         : ARRAY [1..coefVectSize - 1] OF REAL;
km, Alfam, t : REAL;

BEGIN {LevinsonDurbin}
a[0] := 1; {Initialise auxiliary element not part of final result}
Alfam := r[0];
FOR m := 1 TO coefVectSize DO BEGIN
t := r[m];
FOR k := 1 TO m - 1 DO
t := t + r[m - k] * a[k];
km := -t/Alfam;
FOR k := 1 TO m - 1 DO
b[k] := a[m - k];
FOR k := 1 TO m - 1 DO
a[k] := a[k] + km * b[k];
a[m] := km;
Alfam := Alfam * (1 - a[m] * a[m]);
END; { FOR m... }
Alfa := Alfam;
lpca_pm_coeflpc := a[1..coefVectSize];
END; {LevinsonDurbin}

```

Muito embora a versão bit-paralela do LPCA tivesse sido implementada num único módulo, resolveu-se implementar o LPCA bit-serial dividido em dois módulos: o *autocorrelation()*, para o cálculo dos coeficientes de autocorrelação; e o *lpc\_caculation()*, para o cálculo dos coeficientes LPCs. Em seguida, fez-se diagramas com o fluxo das operações necessárias para implementação de cada módulo. A Figura 6.7 mostra o diagrama para o módulo *autocorrelation()*. Identifica-se no diagrama o fluxo de dados e as operações que nesse caso foram necessárias ser implementadas. Os seguintes módulos implementaram as operações

Figura 6.7 - Fluxo de Dados e Operações para o Cálculo da Autocorrelação



necessárias: o *sp\_multiplier\_signed()*, que executa uma multiplicação com o multiplicando serial, o multiplicador paralelo e o produto serial; o *parallel\_2\_serial()*, que converte uma palavra paralela numa palavra serial, ou seja, uma cadeia de bits; o *serial\_adder()*, que implementa uma soma serial, todos os operandos de entrada e saída são seriais; e o *serial\_2\_parallel()*, que converte uma palavra serial, ou seja, uma cadeia de bits, em uma palavra paralela. A seguir foi feito um diagrama mostrando o fluxo de dados e a sequência de operações do *lpc\_calculation()*, para  $m$  assumindo os valores de 1 até 2. Esse diagrama é mostrado na Figura 6.8.

A partir desse diagrama, observou-se que alguns blocos (sequência de operações), eram repetidos e que mudavam apenas os dados de entrada e de saída. Então, o próximo passo foi criar um outro diagrama contendo apenas os blocos com os tipos diferentes de sequências de operações e os possíveis dados que poderiam entrar ou sair desses blocos (Figura 6.9).

Conforme é mostrado na Figura 6.9, conseguiu-se reduzir o fluxo de operações do cálculo dos LPCs a apenas dois blocos de sequência de operações, os blocos Tipo A e Tipo B. O bloco Tipo A, tem uma porta XOR que em conjunto com o *carry in* da

Figura 6.8 – Fluxo de Dados e Operações para o Cálculo dos Coeficientes LPCs para  $m$  assumindo valores de 1 até 2

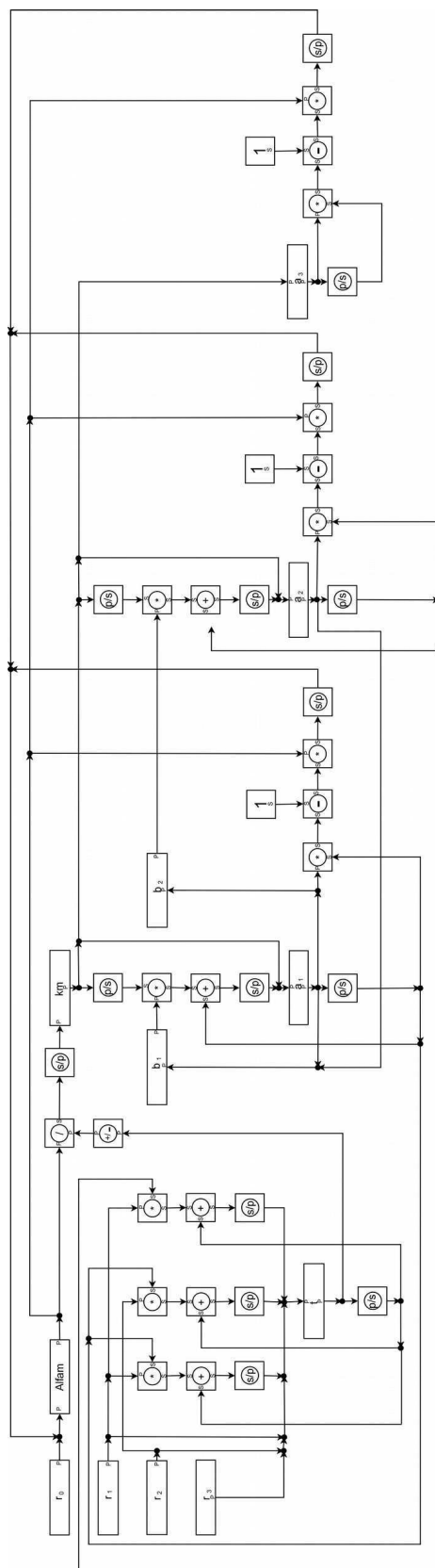
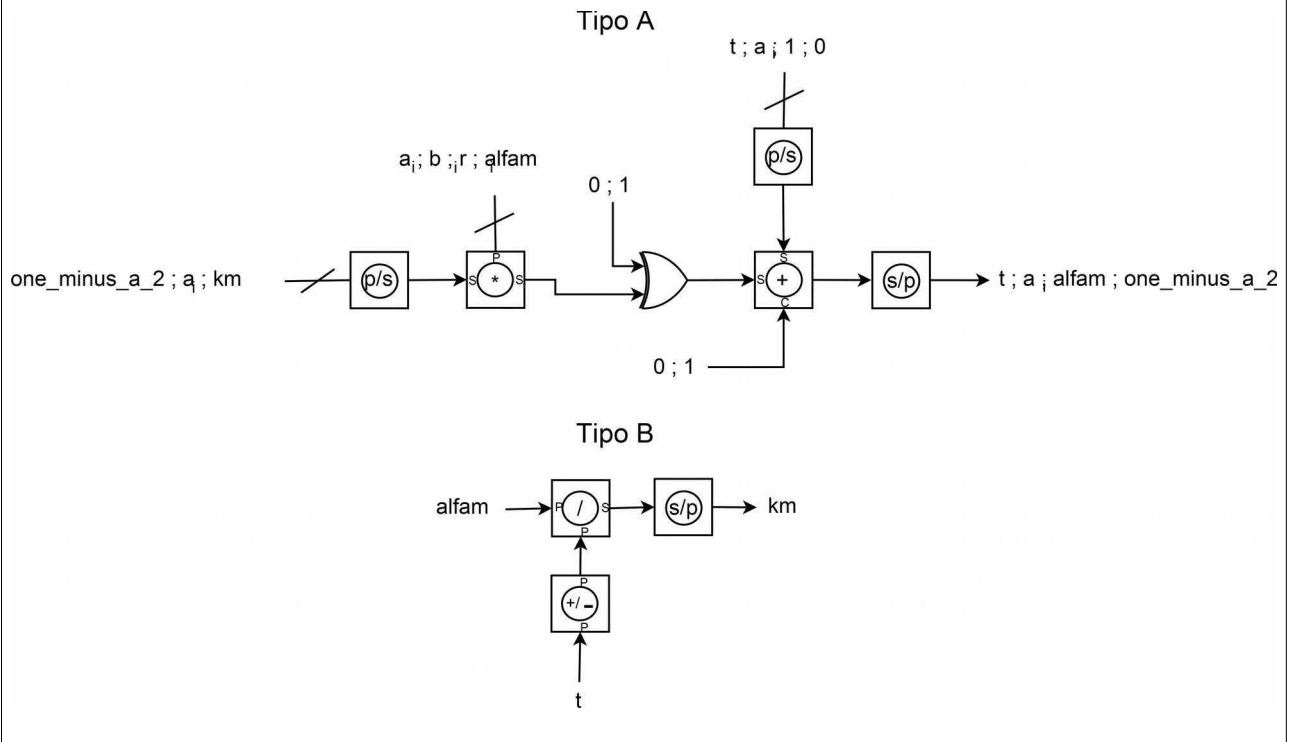


Figura 6.9 – Sequências de Operações Repetidas no Cálculo dos Coeficientes LPCs



operação de soma, pode executar uma soma ou uma subtração. Esse bloco, como é mostrado no diagrama, pode executar com múltiplas entradas e saídas. Para isso, basta enviar os dados que serão operados para as entradas adequadas e receber os dados da saída também na palavra adequada, selecionando-as por meio de um multiplexador e um demultiplexador. O bloco Tipo B diferencia-se bastante do bloco Tipo A, pois, instancia duas operações não contidas no bloco maior: o inversor de sinal e a divisão. A vantagem da redução de blocos a serem instanciados é porque quanto menos operações nos blocos, menor será o número de transistores no circuito e, conseqüentemente, menor será o consumo de energia estática.

Feitas as simulações das versões do LPCA com aritmética bit-paralela e com aritmética bit-serial, com a ferramenta Design Compiler da Synopsys, como era esperado, houve uma diminuição do consumo estático. Porém, o aumento do consumo dinâmico tornou-se muito acentuado como mostra a Tabela 6.10.

Ao ser verificado o que estava contribuindo para um consumo dinâmico tão alto, localizou-se o consumo acentuado nas operações com registros. Tendo em vista a implementação bit serial do LPCA utilizar um único *clock*, aproximadamente 30 vezes maior que o *clock* utilizado na versão bit-paralela, as operações com registradores, que são paralelos, não necessitavam um *clock* tão alto, que gerava



Tabela 6.10 – Consumo do LPCA (com um clock)

Consumo do Módulo de Cálculo dos LPCs (LPCA) com Modelo de Standard Cells, Tecnologia de 28 nm e Um Clock ( $\mu$ W)						
Tipos de Consumo	Aritmética Bit-paralela		Aritmética Bit-serial		Redução de Consumo por Mudança de Aritmética	
	Bibliotecas SC 28nm		Bibliotecas SC 28nm			
	PBP0	PBP16	PBP0	PBP16	PBP0	PBP16
Estático	72,83	1,07	44,67	0,62	38,66%	42,27%
Dinâmico	2,89	3,09	135,24	143,85	-4574,44%	-4553,92%
Total	75,72	4,16	179,91	144,47	-137,60%	-3374,80%

mais consumo dinâmico. A solução que apresentou-se foi a utilização de dois *clocks*: um *clock* para as operações aritméticas seriais e combinacionais, com o valor de 22.500 khz e um segundo *clock* com o valor de 750 kHz para as operações de armazenamento com registradores. Com essa nova abordagem, diminuiu-se o consumo dinâmico para valores mais baixos, como é mostrado na Tabela 6.11.

Tabela 6.11 – Consumo do LPCA (com dois clocks)

Consumo do Módulo de Cálculo dos LPCs (LPCA) com Modelo de Standard Cells, Tecnologia de 28 nm e Dois Clocks ( $\mu$ W)						
Tipos de Consumo	Aritmética Bit-paralela		Aritmética Bit-serial		Redução de Consumo por Mudança de Aritmética	
	Bibliotecas SC 28nm		Bibliotecas SC 28nm			
	PBP0	PBP16	PBP0	PBP16	PBP0	PBP16
Estático	72,83	1,07	43,98	0,61	39,61%	42,35%
Dinâmico	2,89	3,09	75,39	80,58	-2506,02%	-2506,89%
Total	75,72	4,16	119,37	81,19	-57,65%	-1853,00%

Mais uma vez ficou patente que na versão com aritmética bit-paralela, o “grande vilão” que propicia o aumento do consumo dinâmico são as operações de armazenamento em registradores, muitas vezes inevitáveis.

### 6.3.3 Análise dos Resultados

O objetivo principal da pesquisa era investigar a viabilidade de substituir aritmética

bit-paralela por aritmética bit-serial como forma de reduzir o consumo estático, sem aumentar o consumo dinâmico, como preconizado no artigo-referência "*A Methodology for Arithmetic Reduction of the Static Power Consumption Verified on Filter Architectures*" (Nilsson, 2008). O que foi descoberto com esta pesquisa é que isso é uma verdade parcial, ou seja, isto só ocorre em situações em que tem-se muitas operações aritméticas com poucos armazenamentos em registradores. Pois, quando a situação se inverte, isto é, quando se tem muito armazenamento em registradores e, relativamente, menos operações aritméticas, o consumo dinâmico se eleva a níveis que tornam o consumo total muitas vezes maior que quando se utiliza aritmética bit-paralela. Isso funciona bem para aplicações pouco complexas, como somadores ou filtros de pequena complexidade.

Outra afirmativa encontrada no artigo-referência (Nilsson, 2008), que também mostrou-se uma verdade parcial, é que com tecnologias mais densas que 50 nm o consumo de energia estática sempre será menor que com tecnologias menos densas que 50 nm. Existem outros fatores da tecnologia que fazem diferença e contradizem esta afirmação, p.e., a técnica de alargar o canal do transistor com "poly biasing".

## 7 Considerações Finais

Na Figura 2.1 foram mostrados alguns gráficos com as expectativas de tendências do consumo de energia dos circuitos CMOS, com base no ocorrido até 2008, levando em consideração os lançamentos, no mercado, de tecnologias já existentes de fabricação de circuitos CMOS e de novas tecnologias de fabricação ainda por serem lançadas (MOHANTY et al., 2008). Aqui são colocados alguns comentários a respeito de confirmações e contradições ocorridos a partir de 2008 até o presente. A previsão de lançamento da tecnologia de fabricação de circuitos de 14 nm, que é a mais densa que temos no mercado, foi confirmada como aparece no gráfico, pois seu lançamento comercial se deu nesse ano de 2014. Já o lançamento da tecnologia de 28 nm, usada nesta pesquisa, estava prevista para o ano de 2007, porém, só chegou ao mercado no ano de 2010 (GLOBALFOUNDRIES, 2010; DIXON-WARREN, 2012; PR NEWSWIRE, 2014), contradizendo as expectativas. Quanto à previsão do comportamento do aumento e diminuição do consumo estático e dinâmico dessa tecnologia de fabricação, foi visto que, tal qual esperava-se, o consumo estático tornou-se mais impactante para o consumo total que o consumo dinâmico, no entanto, surgiu com essa tecnologia a possibilidade de uso de *poly biasing* para mitigar o problema do aumento do consumo estático nessas tecnologias mais densas. Deve ser levado também em consideração, que a natureza do circuito (p.ex., volume de operações aritméticas vs. volume de operações de armazenamento) que vai influir na relação entre o consumo estático e o dinâmico.

Um aspecto que deve ser considerado com muito cuidado, na hora de se decidir por desenvolver aplicações bit-serial, é o esforço adicional que surge na hora de serializar um fluxo de execução. Esse esforço adicional, traduz-se em mais horas de projeto e desenvolvimento. Medidas adicionais podem ser tomadas, tais como, administração de múltiplos *clocks*, redução de operações que se repetem, minimização dos armazenamentos em registradores, podem acarretar custos à aplicação. Vale avaliar o compromisso entre custo adicional de desenvolvimento e redução do consumo de energia com as aplicações dessas medidas.

### 7.1 Contribuições da Pesquisa

Diante da importância e necessidade de desenvolver-se circuitos digitais dentro de

uma limitação especificada de uso de energia, é comum utilizar-se de várias técnicas de *Low Power* combinadas entre si. Substituição de aritmética bit-paralela por aritmética bit-serial pode ser uma dessas técnicas combinadas, desde que existam as condições favoráveis, discutidas na Seção 6.3.3.

## **7.2 Sugestões para Pesquisas**

Implementar completamente outro sistema de complexidade semelhante ou maior, como por exemplo, como já foi sugerido, serializar os sistemas de compressão sem perdas de dados de ECG (eletrocardiograma) e EEG (eletroencefalograma), desenvolvidos por equipes da UFPB. Outra pesquisa interessante, seria definir uma metodologia para serialização de sistemas, utilizando aritmética bit-serial no lugar de aritmética bit-paralela.

## 8 Referências Bibliográficas

ALVES, J. V. G. et al. **Pre-Emphasis Specification**. Brazil-IP/LAD/DSC/UGCG. Campina Grande, 2009.

ARM. **AMBA AXI Protocol Specification**, ARM Limited, 2004.

ÅSTRÖM, P.; NILSSON, P.; TORKELOSON, M. **Power Reduction in Custom CMOS Digital Filter Structures**. *Analog Integrated Circuits and Signal Processing*, v. 18, n. 1, p. 97-105, jan. 1999.

BERLEKAMP, E. R. **Bit-Serial Reed Solomon Encoders**, *IEEE Transactions on Information Theory*, v. 28, n. 6, pp. 869-874, 1982.

BUTZEN, P. F. et al. **Subthreshold and Gate Leakage Estimation in Complex Gates**, *International Workshop on Logic & Synthesis, Anais, Lake Tahoe, CA, USA: 2008*.

BUTZEN, P. F. et al. **Standby power consumption estimation by interacting leakage current mechanisms in nanoscaled CMOS digital circuits**. *Microelectronics Journal*, v. 41, n. 4, pp. 247-255, 2010.

CHABINI, N. **A Heuristic for Reducing Dynamic Power Dissipation in Clocked Sequential Designs**. In: *17th International Workshop on Power and Timing Modeling, Optimization and Simulation, Berlin, 2007. Anais...* Berlin: Springer-Verlag, 2007.

CHEN, S.; LIN, J. **Experiences of low power design implementation and verification**. In: *Proceedings of the 2008 Asia and South Pacific Design Automation Conference, Seoul, 2008. Anais...* Seoul: IEEE, p. 742-747.

CHEN, S.; LIN, J. **Implementation and verification practices of DVFS and power gating**. In: *2009 International Symposium on VLSI Design, Automation and Test, Hsinchu, 2009. Anais...* Hsinchu: ITRI, p. 19-22.

CHIAPPETTA, M. **NVIDIA GeForce GTX Titan Powered Maingear SHIFT** Disponível em: <http://hothardware.com/Reviews/NVIDIA-GeForce-GTX-Titan-Powered-Maingear-SHIFT-Debuts/>. Acesso em: 22 mai. 2013.

DILIP, B.; PRASAD, P. S.; BHAVANI, R. S. G. **Leakage Power Reduction in CMOS Circuits using Leakage Control Transistor Technique in Nanoscale Technology**. International Journal of Electronics Signals and Systems, v. 2, n. , p. 72-77, 2012.

DIXON-WARREN, S. **A Review of TSMC 28 nm Process Technology**. Chipworks, 2012. Disponível em: <<http://www.chipworks.com/en/technical-competitive-analysis/resources/blog/a-review-of-tsmc-28-nm-process-technology/>>. Acesso em: 12 jan. 2015.

FECHINE et al. **SPVR: An IP core for Real-Time Speaker Verification**. In: Conference Proceedings of IP Based Electronics System Conference & Exhibition, 2010, Grenoble. *Anais...* Grenoble: Conference Proceedings.

FLATRESSE, P. **FDSOI Design & Migration Methodology**. Disponível em: <[http://cmp.imag.fr/documents/doc/UTBB-FDSOI%20Design%20and%20Migration%20Methodology\\_.pdf](http://cmp.imag.fr/documents/doc/UTBB-FDSOI%20Design%20and%20Migration%20Methodology_.pdf)>. Acesso em: 29 nov. 2014.

GAMMIE, G. et al. **SmartReflex Power and Performance Management Technologies for 90 nm, 65 nm, and 45 nm Mobile Application Processors**. Proceedings of the IEEE, v. 98, n. 2, p. 144-159, fev. 2010.

GIRARD, P.; NICOLICI, N.; WEN, X. **Power-Aware Testing and Test Strategies for Low Power Devices**. Boston, MA: Springer Verlag, 2009.

GLOBALFOUNDRIES. **At GTC 2010, GLOBALFOUNDRIES Lays Out Path for Sustained Technology Leadership at 28nm and Beyond**. 2010 Press Releases, 2010. Disponível em: <<http://www.globalfoundries.com/newsroom/press-releases/2010-press-releases/2014/02/28/at-gtc-2010-globalfoundries-lays-out-path-for-sustained-technology-leadership-at-28nm-and-beyond>>. Acesso em: 12 jan. 2015.

HARRIS, D. M. ; HARRIS, S. L. **Digital design and computer architecture**. 1st ed. , EUA: Elsevier Inc., 2007, 592 p.

HENZLER, S. **Power Management of Digital Circuits in Deep Sub-Micron CMOS Technologies**. Munchen, Germany: Springer, 2007.

IEEE-SA. **IEEE Std 1801-2009: IEEE Standard for Design and Verification of Low Power Integrated Circuits**. New York, NY, USA: The Institute of Electrical and Electronics Engineers, Inc., 2009.

ITRS. **International Technology Roadmap for Semiconductors**. Disponível em: <<http://www.itrs.net/reports.html>>. Acesso em: 21 nov. 2014.

JADCHERLA, S. et al. **Verification Methodology Manual for Low Power**. Mountain View, CA, USA: Synopsys, 2009.

JONVANOVIĆ, B.; JEVTIĆ, M. **Static and Dynamic Power Consumption of Arithmetic Circuits in Modern Technologies**. In: 55th Conference for Electronics, Telecommunications, Computers, Automation, and Nuclear Engineering, 55, 2011, Banja Vrućica, *Anais...* Banja Vrućica, p. 1-4.

JOSHI, R. M.; BHAVSAR, S. **Designing Low Power Circuits: A Review**. International Journal of Electronics and Computer Science Engineering, v. 1, n. 4 , p. 2593-2603, 2012.

KAPOOR, B. et al. **Impact of SoC power management techniques on verification and testing**. 10th International Symposium on Quality of Electronic Design, 10, 2009. *Anais...* San Jose: IEEE, p. 692-695.

KATHURIA, J.; AYOUBKHAN, M.; NOOR, A. **A Review of Clock Gating Techniques**. MIT International Journal of Electronics and Computer Engennering, v. 1, n. 2, p. 106-114, ago. 2011.

KAWA, J. **Low Power and Power Management for CMOS-An EDA Perspective**. IEEE Transactions on Electron Devices, v. 55, n. 1, p. 186-196, jan. 2008.

KEATING, M. et al. **Low Power Methodology Manual: For System-on-Chip Design**, New York, NY: Springer, 2007.

KESSLE, A. J.; GANESA, A. Standard cell VLSI design: A tutorial, IEEE Circuits and Devices Magazine, v. 1, p.17-34, jan. 1985.

KIRSCH, N. **NVIDIA GeForce GTX 780 Video Card Review**. Disponível em: <<http://www.legitreviews.com/article/2201/1/>>. Acesso em: 28 mai. 2013.

KUWAHARA, M. **Design and verification methods of toshiba's wireless LAN baseband**. 15th Asia and South Pacific Design Automation Conference, 15, 2010. *Anais...* Taipei: IEEE, p.457-463.

LACKEY, D. E. et al. **Managing Power and Performance for System-on-Chip Designs using Voltage Islands**. In: IEEE/ACM International Conference on Computer Aided Design, 20, 2002, San Jose, *Anais...*, San Jose, pp. 195-202.

LYON, R. F. **Two's Complement Pipeline Multipliers**. IEEE Transactions on Communications, v. 24, n. 4, pp. 418-425, abr. 1976.

MATHUR, A. **Sequential Transformations for Low Power and CPF**. Disponível em: <[http://www.si2.org/events\\_dir/2008/lpc2008/calypso.pdf](http://www.si2.org/events_dir/2008/lpc2008/calypso.pdf)>. Acesso em: 21 maio 2013.

MATTOS, J.; ROSA JR, L.; PILLA, M. (Org.). **Desafios e Avanços em Computação - o estado da arte**. Pelotas, RS: Ed. da Universidade Federal de Pelotas, 2009. p. 260.

MOHANTY, S. P. et al. **Low-Power High-Level Synthesis for Nanoscale CMOS Circuits**. Boston, MA: Springer US, 2008.

NILSSON, P.; TORKEKELSON, M. **A Custom Digital Intermediate Frequency Filter for the American Mobile Telephone System**. IEEE Journal of Solid-state Circuits, v. 32, n. 6, p. 806-815, jun. 1997.

NILSSON, P. **Arithmetic Reduction of the Static Power Consumption in Nanoscale CMOS**. In: 13th IEEE International Conference on Electronics, Circuits, and Systems. 13, 2006, Nice, *Anais...* Nice: IEEE, 2006, p.656-659.

NILSSON, P. **Arithmetic and Architectural Design to Reduce Leakage in Nano-Scale Digital Circuits**. In: 18th European Conference on Circuit Theory and Design, 18, 2007, Seville, *Anais...* Seville: IEEE, p. 372-375.

NILSSON, P. **A Methodology for Arithmetic Reduction of the Static Power Consumption Verified on Filter Architectures**. In: IEEE 20th International Conference on Microelectronics, 20, 2008, *Anais...* Sharjah, 2008. p. 14-17.

NILSSON, P. **Architectures and Arithmetic for Low Static Power Consumption**



in **Nanoscale CMOS**, VLSI Design, Hindawi Publishing Corporation, v. 2009, p. 1-10, 2009.

NILSSON, P. **Approaching Green Electronics: Power Efficient Arithmetic in Nano-scale CMOS**. In: First International Conference on Green Circuits and Systems, 1, 2010, Shanghai, Anais... Shanghai: IEEE, p. 230-235.

PAIVA, A. C. C. **Windowing Specification**. Brazil-IP/LAD/DSC/UFCG. Campina Grande, 2009.

PEDRAM, M.; VAISHNAV, H. **Power Optimization in VLSI Layout: A Survey**. Journal of VLSI Signal Processing Systems, v. 15, n. 3, p. 221-232, mar. 1997.

PFI. **A Practical Guide to Low-Power Design: User Experience with CPF**. California, USA: Power Forward Initiative, 2009.

PR NEWSWIRE. **Qualcomm and TSMC Collaborating on 28nm Process Technology**. Disponível em: <<http://www.prnewswire.com/news-releases/qualcomm-and-tsmc-collaborating-on-28nm-process-technology-80980432.html>>. Acesso em: 21 nov. 2014.

RABEY, J. M.; CHANDRAKASAN, A.; NICOLIĆ, B. **Digital Integrated Circuits – A design Perspective**, 2nd edition, Upper Saddle River, NJ: Pearson Education Inc., 2003.

RANI, A. J.; MALARKKAN, S. **A Survey on Static Power Optimization in VLSI**. International Journal of Computer Science and Information Security, v. 8, n. 9, p. 344-348, dez. 2010.

SANTOS, R. P. et al. **An Optimization Mechanism Intended for Static Power Reduction Using Dual- $V_{th}$  Technique**. Journal of Electrical and Computer Engineering, v. 2012, n. 1, p. 1-12, jan. 2012.

SINGH, A. **Understanding RTL power reduction techniques**. Disponível em: <<http://www.scdsource.com/article.php?id=50>>. Acesso em 25 set. 2008.

SHAULY, E. **CMOS Leakage and Power Reduction in Transistors and Circuits: Process and Layout Considerations**. Journal of Low Power Electronics and

Applications, v. 2, n. 1, p. 1-29, mar. 2012.

SILVEIRA, G. S. **Uma Abordagem para Suporte à Verificação Funcional no Nível de Sistema Aplicada a Circuitos Digitais que Empregam a Técnica Power Gating**, Agosto de 2012, 135 f. Tese (Doutorado em Ciências no Domínio da Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande, Campina Grande, 2012.

SMITH, R. **AMD Radeon HD 7990 Review: 7990 Gets Official**. Disponível em: <<http://www.anandtech.com/show/6915/amd-radeon-hd-7990-review-7990-gets-official>>. Acesso em: 29 mai. 2013.

SREENIVASULU, P. et al. **LOW POWER DESIGN TECHNIQUES OF CMOS DIGITAL CIRCUITS**. Journal of Electronics and Communication Engineering & Technology, v. 3, n. 2, p. 199-208, jul.-set. 2012.

TEIXEIRA Jr., A. G.; CAVALCANTE, V. S. **LPCA Specification**. Brazil-IP/LAD/DSC/UFCG. Campina Grande, 2009.

XU, X.; YONG, H. **Method to Design Low-Power Embedded Processor for Wireless Endoscope** 2009 3rd International Conference on Bioinformatics and Biomedical Engineering, 3, 2009, Beijing. *Anais...Beijing*: IEEE, p. 1-4.

WANHAMMAR, L. **DSP Integrated Circuits**, USA: Academic Press, 1999.