

**"Fluxo de Geração e Caracterização de Bibliotecas de Células
Padrão Para Aplicações de Ultra-baixo Consumo"**

Michael Lopes de Oliveira

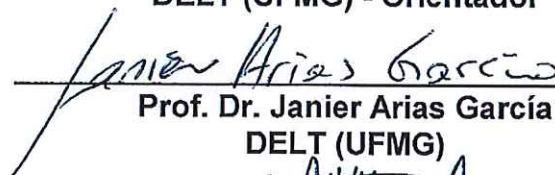
Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Mestre em Engenharia Elétrica.

Aprovada em 21 de fevereiro de 2019.

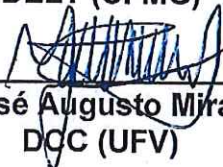
Por:



Prof. Dr. Frank Sill Torres
DELT (UFMG) - Orientador



Prof. Dr. Janier Arias García
DELT (UFMG)



Prof. Dr. José Augusto Miranda Nacif
DCC (UFV)

Michael Lopes de Oliveira

**Fluxo de Geração e Caracterização de
Bibliotecas de Células Padrão para Aplicações
de Ultra-Baixo Consumo**

Belo Horizonte - Brasil

2019

Michael Lopes de Oliveira

Fluxo de Geração e Caracterização de Bibliotecas de Células Padrão para Aplicações de Ultra-Baixo Consumo

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do Título de Mestre em Engenharia Elétrica.

Universidade Federal de Minas Gerais - UFMG

Escola de Engenharia

Programa de Pós-Graduação em Engenharia Elétrica

Belo Horizonte - Brasil

2019

Agradecimentos

Gostaria de agradecer em primeiro lugar a Deus, pois toda honra e glória é dada a Ele. Sem Ele nada disso seria possível.

Também gostaria de agradecer minha esposa Jordana por todo carinho, dedicação e cuidado comigo durante todo esse percurso. A minha mãe Mara, por todo apoio dado a mim nesse caminhada.

Agradeço ao meu orientador, Prof Frank Sill Torres, por todos os conselhos, correções e apoio durante esse trabalho. Muito obrigado pela paciência e por sempre acreditar no meu potencial.

Aos membros que compõem a banca avaliadora, professores Janier Arias García e José Augusto Miranda Nacif, pela disponibilidade de avaliar esse trabalho.

Agradeço também a todos os amigos membros do OptMA Lab por toda ajuda e pelos momentos de descontração durante essa jornada.

“Eu não terei medo.
O medo mata a mente.
O medo é a pequena morte que leva à aniquilação total.
Enfrentarei meu medo.
Permitirei que passe por cima e através de mim.
E, quando tiver passado, voltarei o olho interior para ver seu rastro.
Para onde o medo foi, nada haverá.”
- Litânia do Medo
(Duna - Frank Herbert)

Resumo

Diversos trabalhos no campo da microeletrônica possuem recursos energéticos extremamente limitados. Aplicações biomédicas, de identificação por rádio frequência (RFID) e redes de sensores sem fio podem ser citadas como exemplos desses sistemas. Nessas aplicações, o foco principal é o consumo de energia, em troca da performance geral do sistema. Dessa forma, têm-se a necessidade de designs com potências ultra-baixas (do inglês *Ultra-low Power - ULP*), que possibilitam esse processamento com um ultra-baixo consumo de energia. Uma das técnicas mais comuns para esse fim, é o uso de tensões de alimentação abaixo do limiar do transistor, devido a relação quadrática desta com o consumo de energia. Tendo em mente que essa metodologia precisa de um novo design dos circuitos existentes, novos fluxos de design flexíveis devem ser criados, permitindo aos projetistas gerarem bibliotecas de células padrão onde possam escolher a melhor relação entre o consumo de energia e a performance da aplicação destinada. Ao contrário da literatura, que aborda a manipulação dos parâmetros de forma separada, o presente trabalho tem como função apresentar um fluxo único, capaz de realizar a completa geração de uma biblioteca de células padrão para aplicações de potências ultra-baixas. O fluxo comporta a escolha da melhor tensão de alimentação e dos parâmetros iniciais para o dimensionamento, o design, a caracterização e verificação de todas as células que compõem a biblioteca. A validação do fluxo proposto foi realizada através de simulações realizadas com um processador ARM de código aberto, além de diversos circuitos de teste das plataformas ITC'99 e EPFEL em uma tecnologia de fabricação de $130nm$. O fluxo foi aplicado para gerar uma biblioteca para as tensões de $0,25V$ e $0,35V$ e os resultados dos testes indicam que o uso do fluxo habilita uma redução na potência dissipada em cerca de 98% para a primeira e de cerca de 95% para a última, enquanto que a performance degrada em, aproximadamente, um fator 100 e 35, respectivamente.

Palavras-chave: Consumo ultra-baixo, Biblioteca de células padrão, Design Sub-limiar

Abstract

Many applications in the microelectronic field possess strong limited energy resources. Biomedical applications, RFID and wireless sensors networks can be used as example of those systems. In these applications, the main focus is on energy consumption in exchange for the overall system performance. Thus, there is a need for ultra-low power (ULP) designs, that enable processing with ultra-low energy consumption. One of the most common techniques toward this end is the use of supply voltages lower than the transistor threshold value, due to its quadratic relationship with the energy consumption. Having in mind that this methodology needs a redesign of existing circuits, there is a need to new flexible design flows, allowing the designer to generate standard cell libraries where he can choose the best tradeoff between the energy consumption and the performance of the destined application. In contrast to the state-of-art, which treats each parameter separately, the present work proposes a unique flow, capable of the complete generation of standard cell libraries for ultra-low power applications. The flow comprises the choose of the best supply voltage, the initial sizing parameters, the design, characterization and the verification of all cells that compose the library. The proposed flow was validated through simulations with an ARM-based processor and several testbench circuits from the ITC'99 and EPFEL sets, using a $130nm$ technology. The flow was applied to generate a library for the supply voltages of $0.25V$ and $0.35V$ and results indicate that the flow can enable a reduction in the power dissipation of about 98% for the first and 95% for the last, while the performance degrades in a factor of 100 and 35, respectively.

Keywords: Ultra-low power, Standard cell library, Subthreshold Design

Lista de ilustrações

Figura 1 – O Capacitor MOS. Fonte: (HU, 2010)	20
Figura 2 – Modos de operação do capacitor MOS (a) Acumulação, (b) Depleção e (c) Inversão. Fonte: (WESTE; HARRIS, 2011)	22
Figura 3 – Estrutura dos dois tipos de transistores MOS. Fonte:(AGARWAL, 2005)	22
Figura 4 – Curva I-V característica de um transistor MOS. Fonte:(VLSI, 2013) .	25
Figura 5 – Curva I_d versus V_{gs} para o transistor MOS. Fonte: Adaptado de (RABAEY; CHANDRAKASAN; NIKOLIC, 2002)	26
Figura 6 – Velocidade de portadores versus campo elétrico. Fonte: Adaptado de (JACOBONI et al., 1977)	28
Figura 8 – Atraso de propagação de uma porta lógica	32
Figura 9 – Atraso versus tensão de alimentação em um oscilador em anel. Fonte: Adaptado de (KANITAR, 2009)	34
Figura 10 – Leiaute versus formato abstrato de uma célula NAND2. Fonte: Adaptado de (KULKARNI, 2012)	38
Figura 11 – Estrutura geral de uma memória SRAM. Fonte: Adaptado de (BRITO, 2014)	39
Figura 12 – Estrutura de uma memória SRAM para região sub-limiar. Fonte: Adaptado de (KIM et al., 2008)	40
Figura 13 – Visão geral do fluxo proposto para o desenvolvimento de uma biblioteca de células padrão	44
Figura 14 – Comparação do atraso máximo pela energia consumida para os circuitos de Inversor e NAND2	54
Figura 15 – Atraso do caminho crítico x Potência total para a síntese 1	59
Figura 16 – Atraso do caminho crítico x Potência total para a síntese 2	60
Figura 17 – Atraso do caminho crítico x Potência total para a síntese 3	62
Figura 18 – Atraso do caminho crítico x Potência total para a síntese 4	64
Figura 19 – Estrutura de testes da célula de memória SRAM	65
Figura 20 – AMBER23 após fluxo de EDA	67

Lista de tabelas

Tabela 1 – Tabela de dimensionamento da biblioteca	56
Tabela 2 – Resultados de área e atraso para a síntese 1	58
Tabela 3 – Resultados de dissipação de potência da síntese 1	58
Tabela 4 – Resultados de área e atraso para a síntese 2	59
Tabela 5 – Resultados de dissipação de potência da síntese 2	60
Tabela 6 – Resultados de área e atraso para a síntese 3	61
Tabela 7 – Resultados de dissipação de potência da síntese 3	62
Tabela 8 – Resultados de área e atraso para a síntese 4	63
Tabela 9 – Resultados de dissipação de potência da síntese 4	63
Tabela 10 – Consumo médio de energia por célula para a síntese 3	64
Tabela 11 – Resultados para a memória SRAM	66

Índice de algoritmos

1	Definição de V_{DD} e parâmetros iniciais de dimensionamento	49
---	--	----

Lista de abreviaturas e siglas

CI	Circuito Integrado
CLK	<i>Clock</i>
CMOS	<i>Complementary Metal-Oxide-Semiconductor</i>
DIBL	<i>Drai-Induced Barrier Lowering</i>
DRAM	<i>Dynamic Random Access Memory</i>
DRC	<i>Design Rules Check</i>
EDA	<i>Electronic Design Automation</i>
FET	Transistor de efeito de campo
FO4	<i>Fanout-of-4</i>
GHz	GigaHertz
HDL	<i>Hardware Description Language</i>
LVS	<i>Layout versus Schematic</i>
MHz	MegaHertz
MOS	<i>Metal-oxide-Semiconductor</i>
NMOS	Transistor MOSFET com canal N
PMOS	Transistor MOSFET com canal P
RAM	<i>Random Access Memory</i>
RBL	<i>Read Bitline</i>
RWL	<i>Read Wordline</i>
RFID	<i>Radio Frequency Identification</i>
ROM	<i>Read-Only Memory</i>
RSCE	Reverse Short-Channel Effect
SCE	Short-Channel Effect

SRAM	Static Random Access Memory
ULP	<i>Ultra-low Power</i>
VLSI	Very Large Scale Integration
WSN	Wireless Sensor Network
WBL	<i>Write Bitline</i>
WBLB	<i>Write Bitline Bar</i>
WWL	<i>Write Wordline</i>

Sumário

1	Introdução	15
1.1	Contexto	16
1.2	Objetivos	17
1.3	Organização do trabalho	18
2	Revisão Bibliográfica	19
2.1	Estrutura do transistor MOS	19
2.1.1	Capacitor MOS	19
2.1.2	Transistor MOS	22
2.1.2.1	Modos de Operação	22
2.1.2.2	Características I-V ideais do Transistor MOS	23
2.1.2.3	Características I-V não-ideais do Transistor MOS	27
2.2	Lógica CMOS	30
2.3	Energia e atraso em circuitos integrados	31
2.3.1	Atraso em portas lógicas	32
2.3.2	Dissipação de Potência em transistores MOS	34
2.3.3	Energia em circuitos integrados	36
2.4	Bibliotecas de células Padrão	37
2.5	Memórias SRAM	38
2.6	Estado da Arte	41
3	Metodologia	43
3.1	Fluxo para geração de biblioteca de células padrão	44
3.1.1	Exploração da tecnologia	45
3.1.2	Design das células e caracterização	50
3.2	Conclusão à metodologia	52
4	Simulações e Resultados	53
4.1	Aplicação do Fluxo	54
4.2	Resultados das Simulações	57
5	Conclusão e trabalhos futuros	69
5.1	Trabalhos Futuros	70
	Referências	71

1 Introdução

Em um passado recente, grande parte das pesquisas no campo da microeletrônica foi voltada para o desenvolvimento de equipamentos com performances cada vez melhores. Entretanto, recentes avanços tecnológicos possibilitaram o surgimento de aplicações cujo foco é a redução do consumo de energia, em detrimento da melhora dessa performance geral do sistema. Aplicações no campo de Internet-das-coisas (do inglês *Internet-of-Things - IoT*) e dispositivos embarcados têm em comum o fato de possuírem recursos energéticos extremamente limitados. Isso é devido ao fato de estarem localizados em lugares pouco acessíveis, serem alimentados por baterias (POPOVIC, 2013) ou fazerem uso de fontes energéticas que geram pequenas quantidades de energia elétrica como luz do sol, vibração, indução térmica e eletromagnética (COLOMER-FARRARONS et al., 2011). Como exemplos dessas aplicações podem ser citadas as redes de sensores sem fio, sistemas biomédicos e dispositivos de identificação por rádio frequência (do inglês *radio frequency identification - RFID*).

Nessas aplicações, comumente chamadas de aplicações de ultra-baixo consumo (do inglês *Ultra-low Power - ULP*), o foco principal é reduzir o consumo total de energia em troca da performance geral do sistema. Para cumprir esse objetivo, tensões de alimentação abaixo da tensão de limiar dos transistores são usadas, tendo em mente seu impacto quadrático no consumo de energia. Entretanto, essa metodologia de design requer um total redesenho dos circuitos existentes.

Enquanto que em blocos de circuito analógico esse redesenho é feito quase inteiramente de forma manual, blocos digitais podem ser realizados através da metodologia de células padrão. Nessa forma, após uma etapa de síntese, as funções lógicas são mapeadas em uma biblioteca de células padrão, as quais são automaticamente roteadas em circuitos através das ferramentas de design comerciais. Entretanto, as células padrão existentes atualmente não podem ser usadas nesse regime, devido ao fato de que circuitos CMOS se comportam de forma diferente quando trabalham na região sub-limiar. Tem-se dessa forma a necessidade de se criar novos fluxos para o design de bibliotecas de células para aplicações ULP.

Esse trabalho tem como objetivo apresentar um fluxo completo para a geração de uma biblioteca de células de ultra-baixo consumo de energia. Esse fluxo pode ser utilizado de forma independente ou como ferramenta auxiliar para quaisquer outras técnicas de geração de bibliotecas encontradas na literatura. O fluxo proposto faz a seleção da melhor tensão de alimentação para o projeto no qual será utilizada, bem como define os parâmetros iniciais para o dimensionamento das células da biblioteca. Além disso, o fluxo considera a

caracterização e a verificação da biblioteca criada. Como prova de conceito, uma biblioteca foi gerada para as tensões de alimentação de 0,25V e 0,35V, sendo realizados diversos testes e simulações com as mesmas.

Na seção seguinte o contexto no qual esse trabalho está sendo realizado será apresentado, mostrando quais os principais campos de pesquisa nos quais sistemas ULP estão inseridos e qual a direção que está sendo tomada pelo estado-da-arte. Os objetivos serão mostrados na seção seguinte, enquanto que a última seção mostrará a organização desse documento e a ordem que os assuntos serão discutidos.

1.1 Contexto

O advento de aplicações no campo de IoT e dispositivos portáteis trouxe a tona a necessidade de se avançar nas pesquisas de sistemas com ultra-baixo consumo de energia. Os primeiros trabalhos datam da década de 70, nos quais os pesquisadores buscavam modelar o comportamento das correntes sub-limiar e entender o funcionamento do circuito como um todo, quando o mesmo opera com tensões abaixo do limiar (BARRON, 1972); (TROUTMAN; CHAKRAVARTI, 1973); (TROUTMAN, 1975); (MASUHARA; ETOH; NAGATA, 1974); (VITTOZ; FELLRATH, 1977).

Até meados dos anos 90, o design de circuitos operando na região sub-limiar foi praticamente ignorado (WANG; CALHOUN; CHANDRAKASAN, 2006). Entretanto, com o crescimento do uso de dispositivos portáteis e embarcados, houve uma necessidade de se limitar o consumo de energia, de forma a prolongar a usabilidade desses tipos de dispositivos.

Atualmente, existe um crescente conjunto de aplicações nas quais os recursos energéticos são fatores limitante. Esses sistemas geralmente possuem um consumo de energia muito baixo, mas possuem altos valores de atraso e baixa performance. Dentre essas aplicações, pode-se destacar as redes de micro sensores, as aplicações de identificação por frequência de rádio e dispositivos biomédicos.

Redes de sensores se referem a hardwares que provêm funcionalidades de sensoria-mento, computação e comunicação, podendo ser com ou sem fio (WANG; CALHOUN; CHANDRAKASAN, 2006). As redes sem fio, conhecidas pela sigla WSN (do inglês *wireless sensor network*), são as mais comuns atualmente e são compostas por milhares de nós que captam e processam os dados e os disponibiliza para o usuário final. As aplicações para WSNs podem incluir monitoração ambiental, de hábitos e para construção cívica (BIAGIONI; BRIDGES, 2002); (SCHWIEBERT; GUPTA; WEINMANN, 2001); (CHINTALAPUDI; JOHNSON; GOVINDAN, 2005).

Identificação por rádio frequência é uma tecnologia que identifica automaticamente

algum objeto através de uma tarja afixada ao mesmo. Essas tarjas são capazes de transmitir e receber dados de forma sem fio usando frequências de rádio. Juntamente com a antena para comunicação, as tarjas possuem um pequeno circuito lógico de processamento digital. As tarjas podem ser do tipo passiva ou ativa. As ativas possuem uma bateria para comunicações com distâncias maiores e processamento estendido enquanto que as passivas se comunicam com o leitor fazendo a modulação da carga que ele percebe. Nessa última, a energia necessária para seu funcionamento vêm do aproveitamento da potência vinda do leitor, captada através da antena. Essa característica faz com que o consumo seja extremamente baixo. Por outro lado, também faz com que a distância entre o objeto e o leitor seja muito reduzida além de requerer um processamento lógico simplificado (WEINSTEIN, 2005).

Várias soluções são descritas na literatura visando a geração de circuitos com ultra-baixo consumo de energia. Diversas técnicas visam dimensionar os transistores com objetivo de maximizar a redução de energia e mitigar a redução de performance. Como exemplo, podem ser citadas a otimização multi-objetivo (VOHRMANN et al., 2015) e o aproveitamento do efeito reverso de canal curto (RSCE) (KIM et al., 2007). Em outros trabalhos, o foco é o estudo dos efeitos da redução da tensão de alimentação e o impacto na geração de bibliotecas de células padrão para uso geral (GAO et al., 2015) ou para uso específico (IEONG et al., 2013). Essas técnicas serão mostradas nesse trabalho e será proposto um fluxo que pode ser usado em conjunto com esses trabalhos ou de forma independente para a geração de uma biblioteca de células para aplicações com ultra-baixo consumo energético.

Como principal contribuição desse trabalho, pode-se destacar a criação de uma ferramenta de auxílio para o projetista para a construção de bibliotecas de células padrão para aplicações de baixo consumo e que una a análise e manipulação da tensão de alimentação e dos parâmetros iniciais de dimensionamento em uma única solução. De acordo com os resultados que serão apresentados nas próximas seções, tem-se que uma biblioteca gerada utilizando o fluxo proposto consegue uma redução média de 96% na dissipação de potência.

1.2 **Objetivos**

Os principais objetivos desse trabalho podem ser definidos da seguinte forma:

- Realizar uma revisão bibliográfica capaz de criar uma base para o entendimento das técnicas utilizadas no estado-da-arte e da técnica proposta desse trabalho;
- Elaborar um fluxo que tenha como finalidade a geração e caracterização de uma biblioteca com ultra-baixo consumo de energia;

- Realizar um paralelo entre fluxo proposto e os trabalhos apresentados no estado-da-arte, mostrando as vantagens de se utilizar os dados gerados pelo fluxo;
- Gerar e caracterizar uma biblioteca de células padrão utilizando o fluxo para duas tensões de alimentação distintas;
- Testar a biblioteca gerada, aplicando as células na síntese de diversos circuitos de teste e um microprocessador de código aberto e quantificar os resultados;

1.3 Organização do trabalho

A organização desse texto é realizada da seguinte forma:

- **Capítulo 2 - Revisão Bibliográfica:** Fundamentação teórica necessária para o entendimento de todo o trabalho. Aqui serão apresentados os princípios de funcionamento de um transistor MOS, seguindo para a lógica CMOS. Também serão apresentados os conceitos de dissipação de energia e atraso em circuitos integrados com foco em sistemas de ultra-baixo consumo, a metodologia de células padrão e o estado-da-arte;
- **Capítulo 3 - Metodologia:** Uma visão detalhada do fluxo proposto é apresentada nesse capítulo. Além disso, enquanto cada etapa é descrita, um paralelo com o estado-da-arte é realizado, com o intuito de mostrar as vantagens de se usar o fluxo;
- **Capítulo 4 - Simulações e Resultados:** Nesse capítulo é apresentado o resultado da aplicação do fluxo proposto ao se gerar uma biblioteca de células padrão para duas tensões de alimentação distintas. As células geradas são aplicadas na síntese de vários circuitos de teste e os resultados são apresentados nessa etapa;
- **Capítulo 5 - Conclusão e trabalhos futuros:** Esse capítulo finaliza a dissertação, discutindo os resultados encontrados e as vantagens e desvantagens de se usar o fluxo proposto;

2 Revisão Bibliográfica

Esse capítulo tem como objetivo fazer uma revisão nos conceitos teóricos relacionados à construção e operação de um transistor MOS e suas principais características. Além disso, serão realizadas considerações sobre os parâmetros de energia e atraso nesses dispositivos.

Inicialmente, o capacitor MOS é apresentado na seção 2.1.1. Seus conceitos são ampliados para que a base teórica do transistor MOS seja mostrada na seção 2.1.2. A lógica CMOS é mostrada e exemplificada na seção 2.2. Os principais fatores relacionados ao consumo energético e o atraso serão mostrados na seção 2.3. Nessa seção, também são apresentadas as vantagens e desvantagens de se operar um circuito na região sub-limiar. Por fim, uma discussão sobre a metodologia em células padrão é mostrada na seção 2.4

2.1 Estrutura do transistor MOS

O transistor MOS (Metal Oxide Semiconductor) é um tipo de transistor de efeito de campo (FET - do inglês *Field Effect Transistor*), composto por um terminal de fonte (*source*), um terminal de dreno (*drain*) e um terminal de porta (*gate*) isolado dos demais. Nesse tipo de dispositivo, a corrente conduzida entre os terminais de fonte e dreno é controlada por uma tensão aplicada ao terminal de porta. O transistor MOS pode ser de dois tipos, dependendo de quais são os portadores majoritários que o compõem: nMOS quando os portadores são os elétrons e pMOS quando são as lacunas.

2.1.1 Capacitor MOS

Para entender a operação de um transistor MOS, pode-se fazer uma análise de uma estrutura isolada com uma porta e um corpo, mas sem fonte e dreno. Nessa estrutura a porta de polisilício, o isolante e o corpo formam uma estrutura de placas paralelas chamada capacitor MOS, conforme mostrado na Figura 1.

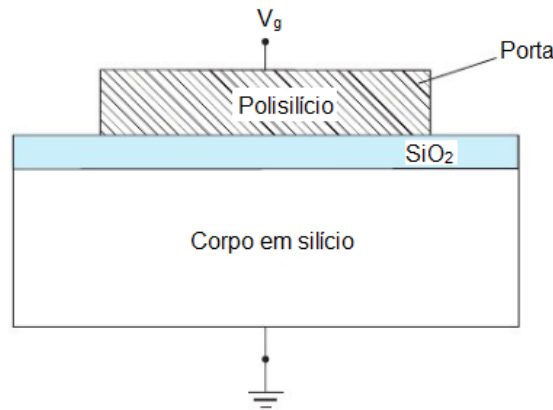


Figura 1 – O Capacitor MOS. Fonte: (HU, 2010)

O capacitor MOS pode funcionar de três maneiras diferentes de acordo com o nível de tensão V_g aplicada na porta. Considerando o corpo do tipo p , cujos portadores majoritários são as lacunas, o modo de acumulação ocorre quando uma tensão negativa é aplicada. A carga negativa faz com que as lacunas sejam atraídas à interface óxido-semicondutor, formando uma região de acumulação de portadores majoritários, como mostrado na Figura 2(a). Essa camada de acumulação pode ser considerada como um capacitor de placas paralelas, sendo que um dos eletrodos é essa camada e o outro é o eletrodo superior formado pela porta. Nessas condições, contanto que ocorra um contato ôhmico direto entre o substrato e a região de acumulação, a capacitância máxima C_{max} da estrutura MOS é igual a capacitância do óxido C_{ox} e pode ser dada pela Equação 2.1.

$$C_{max} = C_{ox} = (\epsilon_o \times \epsilon_{ox} \times A)/t_{ox} \quad (2.1)$$

onde ϵ_o é a permissividade no vácuo, ϵ_{ox} é a permissividade do óxido, t_{ox} é a espessura do óxido e A é a área do eletrodo superior (NICOLLIAN; BREWS, 1982).

A medida em que a tensão negativa aplicada ao terminal de porta diminui, a camada de acumulação reduz devido a perda de intensidade do campo elétrico no óxido. Esse efeito ocorre até a tensão de porta (V_g) ser igual a zero. Nesse momento, a camada de acumulação é extinta e as concentrações de portadores majoritários se assemelham às daquelas do corpo do substrato.

Ao se aplicar uma pequena tensão positiva na porta, até um valor limite chamado de tensão de limiar V_{th} , as lacunas móveis presentes na superfície são deslocadas em direção ao substrato. Dessa forma, uma região de depleção é formada pelos íons das impurezas aceitadoras, que compensa o campo elétrico aplicado. A Figura 2(b) mostra o modo de depleção

A largura W_d dessa região de depleção é proporcional à tensão V_g aplicada no

terminal de porta. O valor de W_d é dado pela Equação 2.2.

$$W_d = \sqrt{[(2 \times \epsilon_{si} \times V_g)/(q \times N)]} \quad (2.2)$$

onde ϵ_{si} é a permissividade elétrica do silício, q é a carga do elétron e N é a concentração de dopantes no silício.

Nesse modo, a capacitância da região de depleção é associada em série com a capacitância do óxido do terminal de porta. Então, a capacitância total da estrutura MOS pode ser obtida através da Equação 2.3 (SWART, 2002).

$$C_t = \sqrt{[(1/C_{ox}) + (1/C_{dep})]} \quad (2.3)$$

onde C_t é a capacitância total, C_{ox} é a capacitância do óxido e C_{dep} é a capacitância da região de depleção, que é dada pela Equação 2.4

$$C_{dep} = \epsilon_{si}/W_d \quad (2.4)$$

Quando o valor da tensão aplicada na porta ultrapassa o valor da tensão de limiar V_{th} , as concentrações de portadores minoritário e majoritários se igualam com os níveis intrínsecos do semiconductor. Os portadores minoritários, que para o caso do corpo tipo p são os elétrons, são atraídos para a superfície do substrato, formando uma camada de inversão do tipo n na interface óxido-semiconductor. Essa condição é chamada de modo de inversão forte e está mostrado na Figura 2(c). A tensão de limiar depende da concentração de dopantes no corpo do semiconductor. Esse valor pode ser obtido usando a Equação 2.5

$$V_{th} = 2\phi_B + \frac{\sqrt{qN^2\epsilon_s 2\phi_B}}{C_{ox}} \quad (2.5)$$

onde $2\phi_B$ é dado pela Equação 2.6

$$2\phi_B = 2\frac{kT}{q} \ln \frac{N}{n_i} \quad (2.6)$$

sendo $\frac{kT}{q}$ igual ao valor da tensão termal, com k igual a constante de Boltzmann, T a temperatura e q a carga do elétron e n_i é a concentração de portadores intrínseca do silício (HU, 2010).

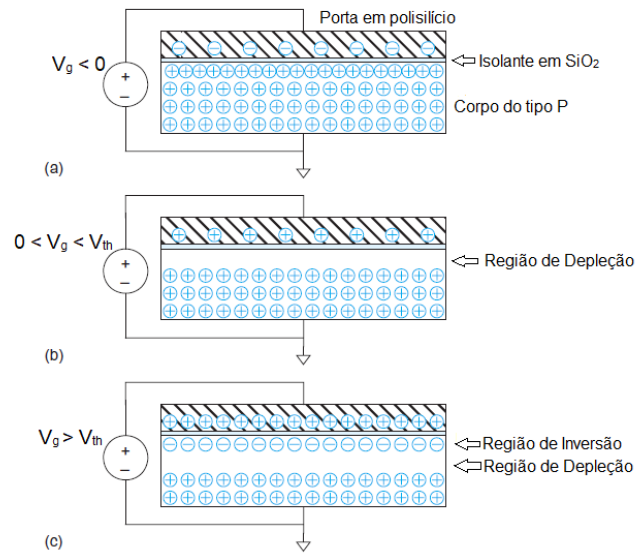


Figura 2 – Modos de operação do capacitor MOS (a) Acumulação, (b) Depleção e (c) Inversão. Fonte: (WESTE; HARRIS, 2011)

2.1.2 Transistor MOS

2.1.2.1 Modos de Operação

Um transistor MOS é formado ao se adicionar terminais de fonte e dreno à estrutura MOS estudada na seção anterior. Para substratos do tipo p , adiciona-se duas regiões do tipo n para a fonte e para o dreno. Além disso, o corpo da estrutura é conectado ao potencial de terra. Essa é a estrutura que forma um transistor do tipo n ou, como é mais conhecido, nMOS. Para o transistor do tipo p , ou pMOS, temos um estrutura contrária à do nMOS, com o corpo sendo feito de material do tipo n e a fonte e o dreno do tipo p . Nesse tipo de transistor, o corpo é usualmente conectado a um terminal com um potencial mais alto, geralmente a tensão de alimentação V_{dd} . A Figura 3 mostra a estrutura dos dois tipos de transistores MOS.

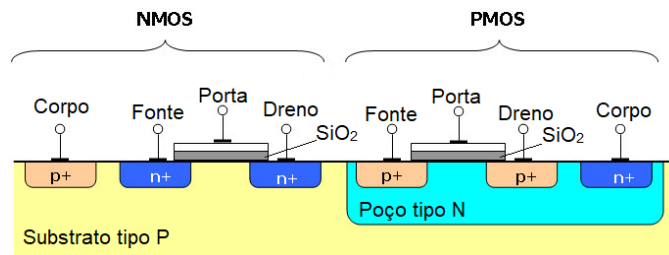


Figura 3 – Estrutura dos dois tipos de transistores MOS. Fonte:(AGARWAL, 2005)

O transistor MOS possui três diferentes modos de operação, que dependem do valor da tensão aplicada entre os terminais de porta e fonte (V_{gs}). A análise apresentada

considera um transistor do tipo nMOS, sendo que para o pMOS, o funcionamento é o oposto do apresentado.

Quando o valor de V_{gs} é menor do que o valor da tensão de limiar V_{th} , os terminais de fonte e dreno possuem elétrons livres, entretanto o corpo possui apenas lacunas livres. Considerando o corpo está conectado ao terra, as junções PN entre o corpo e a fonte ou o dreno estão polarizadas reversamente. Nessas condições, pode-se dizer que o transistor está desligado e esse modo de operação é chamado de *corte*.

Elevando a tensão entre porta e fonte para um valor acima da tensão de limiar, tem-se o modo *linear* de operação do transistor. Nessa situação, uma região de inversão com elétrons livres é formada, criando um caminho condutivo que conecta a fonte e o dreno. Esse caminho é conhecido como o canal do transistor. O número de portadores majoritários e a condutividade desse canal aumenta com a tensão aplicada no canal. A diferença de potencial entre o dreno e a fonte, nessas condições é $V_{ds} = V_{gs} - V_{gd}$.

No modo linear, a corrente flui através do canal entre dreno e fonte (I_{ds}). Seu valor é diretamente proporcional à tensão entre porta e fonte (V_{gs}) e a tensão entre dreno e fonte (V_{ds}), atuando de forma semelhante a um resistor linear (WESTE; HARRIS, 2011).

Considerando a tensão entre porta e fonte maior que a tensão de limiar ($V_{gs} > V_{th}$) e a tensão entre dreno e fonte maior que a diferença entre a tensão de porta e a tensão de limiar ($V_{ds} > V_{gs} - V_{th}$), tem-se o transistor em modo de saturação. Nesse modo, o transistor ainda fica ligado entretanto, o canal não está mais em inversão na região próxima ao dreno. A criação dessa região é chamada de pinçamento (*pinch-off*). A condução nesse modo é feita pelo movimento dos elétrons sob a influência da tensão positiva no dreno. Quando os elétrons chegam ao final do canal, eles são injetados na região de depleção formada e em seguida, acelerados para o dreno. Dessa forma, a corrente I_{ds} é controlada apenas pela tensão aplicada na porta, não sendo influenciada pelo dreno.

2.1.2.2 Características I-V ideais do Transistor MOS

Como dito anteriormente, o transistor possui três modos de operação. Ao derivar os modelos apresentados por Shockley (1952), Cobbold (1970) e C.Sah (1964), pode-se relacionar as correntes e as tensões para cada um desses três modos de operação.

Esse modelo assume que o canal é longo o bastante ($> 2\mu m$) para que o campo elétrico lateral seja relativamente pequeno e é conhecido como o modelo de primeira ordem de Shockley (SHOCKLEY, 1952).

Para o modo de corte, o modelo assume que a corrente através um transistor desligado é zero. Quando a tensão de porta é maior que a tensão de limiar, e o transistor entra no modo linear, a porta atrai os elétrons para formar um canal de condução. Os elétrons migram da fonte para o dreno em uma razão proporcional ao campo elétrico

entre essas regiões. Nesse modo, a estrutura MOS se assemelha a um capacitor de placas paralelas. Dessa forma, pode-se dizer que a carga Q_{chan} no canal é dada por $Q_{chan} = CV$.

Considerando a porta tendo um comprimento L , uma largura W e t_{ox} como sendo a espessura do óxido, a capacitância C_g pode ser dada pela Equação 2.7.

$$C_g = C_{ox}WL \quad (2.7)$$

onde $C_{ox} = \epsilon_{ox}/t_{ox}$ e ϵ_{ox} é a permitividade do óxido.

Já a tensão V está referenciada ao canal. Considerando a tensão de fonte como V_s e a tensão de dreno como V_d , tem-se que $V = V_{gc} - V_{th} = (V_{gs} - V_{ds}/2) - V_{th}$.

Em transistores nMOS, a carga é transportada pelos elétrons. Esses são acelerados no canal pelo campo elétrico lateral E_{lat} entre a fonte e o dreno, que pode ser encontrado dividindo a tensão V_{ds} pelo comprimento L do canal. A velocidade v dos portadores é proporcional a esse campo elétrico lateral, e pode ser dada pela Equação 2.8

$$v = \mu E_{lat} \quad (2.8)$$

onde o parâmetro μ é uma constante de proporcionalidade chamada de mobilidade. Um valor típico para a mobilidade em transistores nMOS com campos elétricos fracos é entre $500 - 700 \text{cm}^2/Vs$ (WESTE; HARRIS, 2011).

O tempo necessário para os portadores cruzarem o canal é igual ao comprimento do mesmo, dividido pela velocidade dos portadores, ou seja, $t = L/v$. Assim, a corrente entre a fonte e o dreno (I_{ds}) é o total de cargas no canal dividido pelo tempo necessário para cruzá-lo. Então, tem-se a Equação 2.9.

$$\begin{aligned} I_{ds} &= \frac{Q_{chan}}{L/v} \\ &= \mu C_{ox} \frac{W}{L} (V_{gs} - V_{th} - V_{ds}/2) V_{ds} \\ &= \beta (V_{gs} - V_{th} - V_{ds}/2) V_{ds} \end{aligned} \quad (2.9)$$

onde $\beta = \mu C_{ox} \frac{W}{L}$

Considerando V_{dsat} a tensão de saturação do dreno, cujo valor pode ser obtido por $V_{dsat} = V_{gs} - V_{th}$, quando $V_{gd} < V_{th}$ e $V_{ds} > V_{dsat}$, tem-se a formação de uma região de pinçamento próxima ao dreno. Nessa região, o transistor está no modo de saturação e a corrente I_{ds} não aumenta com a tensão de dreno (V_{ds}). Dessa forma, pode-se definir a Equação 2.10 para a corrente de saturação em um transistor MOS, que seja independente

da tensão V_{ds} :

$$\begin{aligned} I_{ds} &= \beta (V_{gs} - V_{th} - V_{dsat}/2) v_{dsat} \\ &= \frac{\beta}{2} (V_{gs} - V_{th})^2 \end{aligned} \quad (2.10)$$

Assim, a Equação 2.11 mostra o comportamento da corrente I_{ds} para as três regiões de operação do transistor.

$$I_{ds} = \begin{cases} 0, & \text{Corte} \\ \beta (V_{gs} - V_{th} - V_{ds}/2) V_{ds}, & \text{Linear} \\ \frac{\beta}{2} (V_{gs} - V_{th})^2, & \text{Saturação} \end{cases} \quad (2.11)$$

A Figura 4, mostra como o aumento da tensão V_{gs} aumenta o valor da corrente I_{ds} até o transistor entrar no modo de saturação. Além disso, são mostradas os três diferentes modos de funcionamento, de acordo com o valor de V_{ds} aplicado.

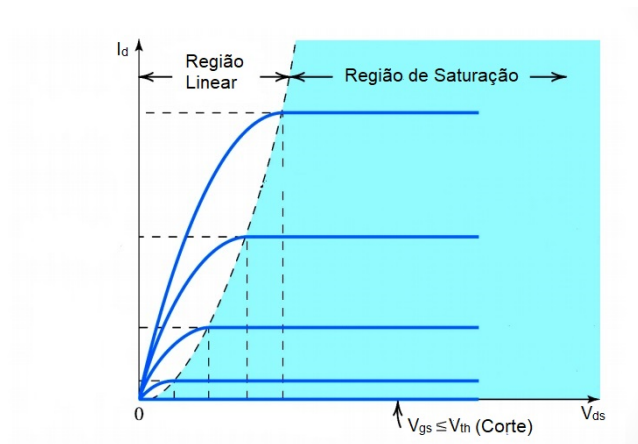


Figura 4 – Curva I-V característica de um transistor MOS. Fonte:(VLSI, 2013)

Uma outra maneira de se caracterizar o funcionamento de um transistor MOS é relacionado a sua tensão de alimentação V_{DD} . Um dispositivo MOS pode operar em duas regiões distintas de acordo com o valor do V_{DD} em relação à tensão de limiar V_{th} . Considere o gráfico mostrado na Figura 5, onde a corrente de dreno do transistor I_d é plotada contra a tensão entre a porta e a fonte V_{gs} em uma escala logarítmica.

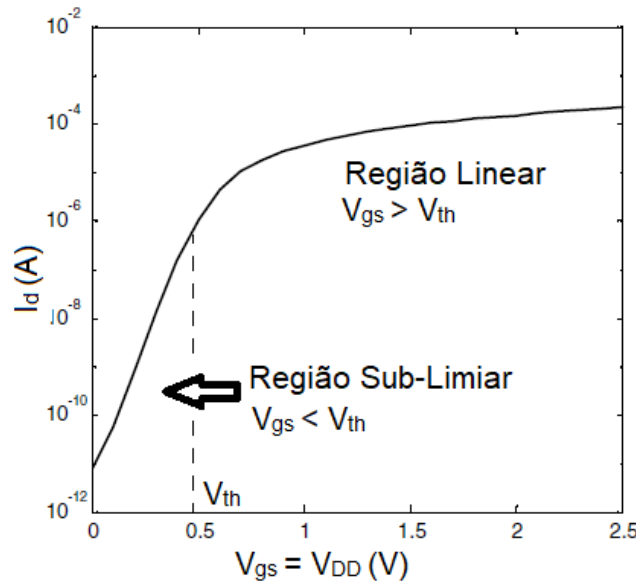


Figura 5 – Curva I_d versus V_{gs} para o transistor MOS. Fonte: Adaptado de (RABAEY; CHANDRAKASAN; NIKOLIC, 2002)

Considerando $V_{gs} = V_{DD}$, pode-se perceber, através do gráfico, as duas regiões de operação do transistor MOS, relacionados com o valor da tensão V_{gs} . Dispositivos que trabalham com valores de V_{gs} acima da tensão de limiar V_{th} operam na região linear, também conhecida como região de inversão forte ou *superthreshold*. Para operações com V_{gs} abaixo de V_{th} , tem-se a região sub-limiar, também conhecida como região de inversão fraca ou *subthreshold*. Nesse contexto, as correntes dominantes no dispositivo são a corrente sub-limiar I_{sub} , a corrente de tunelamento através do óxido da porta I_{gate} e a corrente de tunelamento entre as junções dreno-corpo e fonte-corpo I_{dio} .

A corrente sub-limiar I_{sub} ocorre entre os terminais de dreno e de fonte quando o transistor está operando na região de inversão fraca. Nessa região de operação, a concentração de portadores minoritários é menor e o canal entre os terminais de dreno e fonte não possuem um campo elétrico horizontal. Dessa forma, um campo elétrico longitudinal é criado devido a tensão entre dreno a fonte V_{ds} fazendo com que os portadores se movam por difusão entre esses terminais. A corrente sub-limiar pode ser definida pela Equação 2.12 (BUTZEN; RIBAS, 2007).

$$I_{sub} = I_0 e^{\left(\frac{V_{gs}-V_{th}}{nV_T}\right)} \left(1 - e^{-\frac{V_{ds}}{V_T}}\right) \quad (2.12)$$

Onde V_T denota a tensão termal $V_T = \frac{kT}{q}$, com k sendo a constante de Boltzmann, T a temperatura em Kelvin e q a carga elementar do elétron. I_0 é dado pela Equação 2.13.

$$I_0 = \frac{W}{L} \mu_0 C_{ox} V_T^2 e^{1.8} \quad (2.13)$$

Com W e L representando o comprimento e a largura do transistor respectivamente, C_{ox} a capacitância do óxido, μ_0 sendo a mobilidade dos portadores e n é a inclinação sub-limiar. Esse último descreve o quanto a tensão de porta V_g deve reduzir para que a corrente de vazamento diminua em uma ordem de magnitude. A Equação 2.14 descreve a inclinação sub-limiar, cujo valor é aproximadamente 100 mV/década, em temperatura ambiente.

$$n = \frac{dV_g}{d(\log I_d)} \quad (2.14)$$

A corrente de tunelamento do óxido da porta I_{gate} é referido como o tunelamento de elétrons ou lacunas do corpo do transistor e da região de encontro entre a fonte e o dreno, através da barreira de potencial do óxido da porta. Esse efeito é relacionado ao conceito de capacitor MOS, mostrado nas seções anteriores.

O transistor MOS possui duas junções pn . Uma entre o terminal de dreno e o corpo e outra entre o terminal de fonte e o corpo. Essas junções são, usualmente, polarizadas reversamente, o que causa a corrente de tunelamento entre essas junções I_{dio} . Essa corrente é uma função entre a área da junção e a concentração de dopantes.

2.1.2.3 Características I-V não-ideais do Transistor MOS

Os conceitos apresentados até o momento são válidos para dispositivos que possuem o canal entre os terminais de fonte e dreno muito grande ($> 2\mu m$) e que trabalham com campos elétricos mais fracos. Entretanto, a maioria das tecnologias atuais possuem canais mais estreitos ($< 1\mu m$), o que faz com que apareçam os chamados efeitos I-V não-ideais (HU, 2010).

Como visto pela Equação 2.8, a velocidade v dos portadores é proporcional ao campo elétrico lateral E_{lat} e a mobilidade μ . Entretanto, sob campos elétricos fortes, a velocidade dos portadores para de crescer linearmente com o campo elétrico, chegando a um limite máximo de velocidade conhecido como v_{sat} . Esse efeito é chamado de saturação de velocidade e faz com que se tenha valores mais baixos que o esperado para a corrente I_{ds} quando aplicados valores mais elevados de V_{gs} (WESTE; HARRIS, 2011).

Tensões elevadas na porta de um transistor tem como consequência um aumento no campo elétrico vertical E_{ver} , visto que $E_{ver} = V_{gs}/t_{ox}$. Esse comportamento faz com que os portadores sejam atraídos para as arestas do canal, causando colisões com a interface do óxido, tornando os portadores mais lentos. Esse efeito é chamado de degradação da mobilidade (WESTE; HARRIS, 2011).

Sob essas condições, para se modelar a mobilidade μ para transistores $nMOS$ e $pMOS$, pode-se usar um valor de mobilidade efetivo μ_{eff} , proposto por Chen et al. (1996),

mostrado na Equação 2.15.

$$\mu_{eff-nmos} = \frac{540}{1 + \left(\frac{V_{gs}+V_{th}}{0,54 \cdot t_{ox}}\right)^{1,85}} \quad \mu_{eff-pmos} = \frac{185}{1 + \frac{|V_{gs}+V_{th}|}{0,338 \cdot t_{ox}}} \quad (2.15)$$

A Figura 6 mostra a relação entre a velocidade dos portadores e o valor do campo elétrico. Nesse gráfico, as inclinações μ_n e μ_p representam a mobilidade do nMOS e do pMOS respectivamente e v_{sat-n} e v_{sat-p} os limites máximos de velocidades para o nMOS e o pMOS. A linha contínua no gráfico representa os valores medidos empiricamente enquanto que a linha tracejada representa os valores aproximados através da mobilidade efetiva.

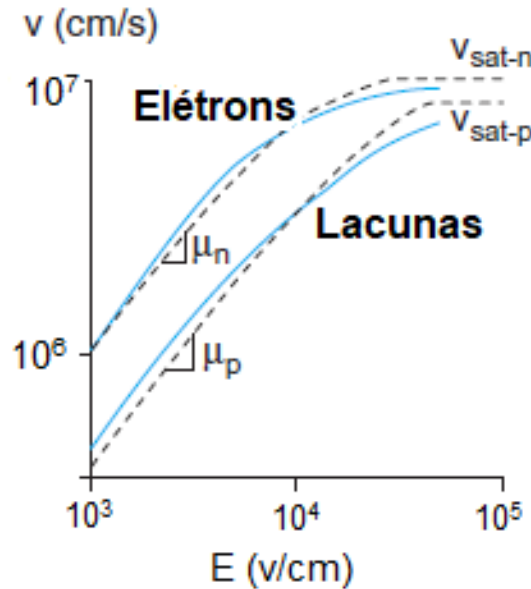


Figura 6 – Velocidade de portadores versus campo elétrico. Fonte: Adaptado de (JACOBONI et al., 1977)

Como pode ser percebido pelo gráfico, em campos elétricos mais fracos o comportamento da velocidade dos portadores cresce linearmente com o campo. Entretanto, para campos maiores do que um valor crítico E_{crit} a velocidade dos portadores chega ao limite v_{sat} , cujo valor é de cerca de 10^7 cm/s para o nMOS e de $8 \times 10^6 \text{ cm/s}$ para o pMOS. Uma aproximação para o cálculo da velocidade dos portadores pode ser vista na Equação 2.16 (TOH; KO; MEYER, 1988).

$$v = \begin{cases} \frac{\mu_{eff} E}{1 + \frac{E}{E_{crit}}}, & E < E_{crit} \\ v_{sat}, & E \geq E_{crit} \end{cases} \quad (2.16)$$

Com o campo elétrico crítico dado por $E_{crit} = \frac{2v_{sat}}{\mu_{eff}}$.

A tensão crítica V_{crit} é o valor da tensão entre dreno e fonte V_{ds} no qual o campo crítico é alcançado. Seu valor depende do comprimento do canal de forma que $V_{crit} = E_{crit}L$.

Tendo em mente os conceitos de saturação de velocidade e de degradação de mobilidade, a Equação 2.11 pode ser modificada em termos dessas grandezas, como mostrada na Equação 2.17

$$I_{ds} = \begin{cases} \frac{\mu_{eff}E}{1 + \frac{E}{E_{crit}}} C_{ox} \frac{W}{L} (V_{gs} - V_{th} - V_{ds}/2) V_{ds}, & \text{Linear} \\ C_{ox} W (V_{gs} - V_{th} - V_{dsat}) v_{sat}, & \text{Saturação} \end{cases} \quad (2.17)$$

$$\text{Com } V_{ds} = V_{dsat} = \frac{(V_{gs} - V_{th})V_{crit}}{(V_{gs} - V_{th}) + V_{crit}}.$$

Um outro fenômeno que ocorre em dispositivos canal estreito é aquele relacionado à diminuição do valor da tensão de limiar V_{th} a medida que o comprimento do canal se torna menor. Esse comportamento é conhecido como efeito de canal curto (SCE - do inglês *short-channel effect*) (KIM et al., 2007).

Esse efeito usualmente é indesejado, visto que a tensão de limiar deve manter-se invariável a medida que as dimensões do transistor reduzem. Quanto menor o comprimento do canal, mais crítico se torna o SCE, visto que seu efeito piora devido ao aumento do fenômeno de diminuição da barreira de potencial da junção canal-fonte (DIBL - do inglês *drain-induced barrier lowering*) (TROUTMAN, 1979). Esse efeito está relacionado à proximidade entre as regiões de depleção de fonte e de dreno e ao aumento da tensão entre dreno e fonte V_{ds} , fazendo com que a corrente I_{ds} aumente proporcionalmente com o valor de V_{ds} . A Equação 2.18 mostra como a tensão de limiar é afetada ao se levar em conta esses fenômenos.

$$V_{th} = V_{th0} - \eta V_{ds} \quad (2.18)$$

Onde V_{th0} é o valor da tensão de limiar padrão do dispositivo e η é o coeficiente DIBL, tipicamente expresso em mV/V .

Uma maneira de compensar o comportamento causado pelo SCE, é o acréscimo de componentes dopantes nas fronteiras dreno-corpo e fonte-corpo do transistor, chamados de implantes HALO. Esses implantes tem como função diminuir o controle que o dreno tem no canal, fazendo com que haja um estreitamento na largura da camada de depleção. Entretanto, como um subproduto desses implantes, a tensão de limiar V_{th} passa a diminuir com o aumento do comprimento do canal. Esse efeito é conhecido como SCE reverso (RSCE - do inglês *reverse short-channel effect*) (LU; SUNG, 1989).

O RSCE também pode impactar na corrente I_{ds} do dispositivo. Em transistores

que operam na região de inversão forte, a máxima corrente ocorre quando o valor do comprimento do canal é o mínimo para a tecnologia de fabricação usada. Isso é devido ao fato de o efeito na corrente causado pela maximização da relação W/L entre a largura e o comprimento do canal ser mais forte que os efeitos do RSCE. Entretanto, quando se trabalha na região sub-limiar, os efeitos do RSCE passam a prover um acréscimo exponencial para a corrente, devido ao fato de se haver um menor V_{th} . Dessa forma, para se obter uma máxima corrente em dispositivos que operem na região sub-limiar, deve-se usar valores maiores que o mínimo para o comprimento do canal (KIM et al., 2007). Como a corrente ainda é dependente da razão W/L , o aumento do comprimento do canal deve ser feito até certo limite, visto que, acima de um determinado valor, os efeitos do RSCE passam a ser menores.

Pode-se calcular o valor do comprimento do canal que consegue fornecer a corrente máxima devido aos efeitos causados pelo RSCE. Inicialmente, obtêm-se o valor da tensão de limiar quando o dispositivo está sob o RSCE. A Equação 2.19 mostra esse cálculo.

$$V_{th} = V_{th0} + K_1 \left(\sqrt{1 + \frac{K_2}{L}} - 1 \right) \sqrt{\phi_s} \quad (2.19)$$

Onde V_{th0} é a tensão de limiar padrão, K_1 e K_2 são parâmetros da tecnologia utilizada e são números positivos e ϕ_s é o potencial de superfície.

Em seguida, utilizando a Equação 2.19 e tomando a derivada da Equação 2.12, de forma que $\frac{\partial I_{ds}}{\partial L} = 0$, tem-se a relação que fornece o valor do comprimento do canal que fornece a corrente máxima, mostrado na Equação 2.20.

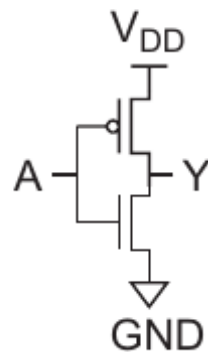
$$L = \frac{-K_2 + \sqrt{K_2^2 - 4K_3}}{2} \quad (2.20)$$

$$\text{Com } K_3 = -\frac{K_1^2 \phi_s}{n^2 V_T^2} K_2.$$

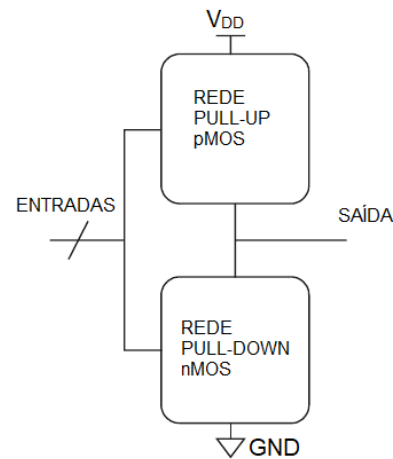
O efeito RSCE não é uma preocupação em circuitos que operam na região de inversão forte, visto que nesse contexto, somente o SCE é dominante. Em dispositivos que operam na região sub-limiar, os valores de V_{ds} são menores, levando a um reduzido efeito DIBL. Isso faz com que o SCE possa ser desprezado em transistores que operem nessa região e que somente o RSCE seja considerado (PAUL, 2005).

2.2 Lógica CMOS

Para se compreender a lógica CMOS de uma maneira simplificada pode-se usar um Inversor CMOS como o apresentado na Figura 7a, composto por uma entrada A , uma saída Y e as entradas de energia V_{DD} e GND .



(a) Inversor CMOS



(b) Estrutura geral de uma porta lógica CMOS estática

Quando a entrada A é igual a 0, o transistor nMOS não está conduzindo, enquanto que o transistor pMOS conduz. Dessa forma, a saída Y é levada a 1 devido ao fato do pMOS estar conectado em V_{DD} . De forma contrária, quando a entrada é levada a 1, o nMOS passa a conduzir enquanto que o pMOS deixa de fazê-lo, levando a saída ao nível lógico 0.

O inversor apresentado na Figura 7a é um exemplo de uma porta lógica CMOS estática, ou porta lógica CMOS complementar. Em geral, uma porta lógica CMOS estática possui uma rede de *pull-down* composta de transistores nMOS conectando a saída ao nível lógico 0 e uma rede de *pull-up* composta por transistores pMOS ligando a saída ao nível lógico 1. Uma estrutura geral de uma porta lógica CMOS é mostrada na Figura 7b

No exemplo do inversor, as redes de *pull-up* e *pull-down* possuem apenas um transistor cada. Para se gerar portas lógicas mais complexas, mais transistores devem ser usados em cada uma das redes. Dessa forma, dois ou mais transistores conectados em série, só estarão conduzindo caso todos os transistores estejam ligados. Para transistores em paralelo, basta apenas um estar ligado para que se possa considerar que todos estejam conduzindo.

2.3 Energia e atraso em circuitos integrados

Esta seção tem como função descrever o comportamento de circuitos CMOS com relação ao consumo de energia e ao atraso. Inicialmente, será apresentada como é a dissipação de potência e quais as são as diferenças de acordo com a região de operação do circuito. Em seguida, o modelo RC de atraso será mostrado, descrevendo como essa grandeza pode ser afetada pela variação da tensão de alimentação. Por fim, serão mostrados os conceitos sobre o consumo de energia e as relações entre energia e o atraso como uma

métrica para o design de sistemas eletrônicos.

2.3.1 Atraso em portas lógicas

O atraso de propagação em uma porta lógica t_p denota o tempo máximo que o sinal de saída dessa porta muda após uma mudança no sinal de entrada. Essa grandeza é expressa em segundos e, usualmente, é medida do momento em que o sinal de entrada atinge 50% do seu valor até o momento em que o sinal de saída atinge 50% do seu valor máximo, como mostrado pela Figura 8.

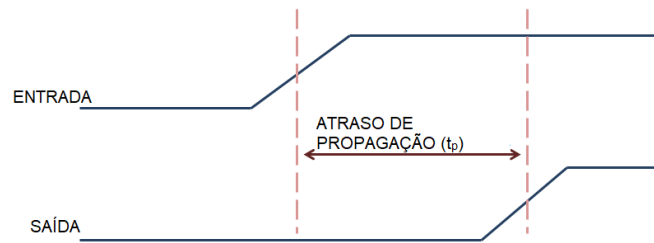


Figura 8 – Atraso de propagação de uma porta lógica

Uma forma simplificada de se calcular o atraso é usando um modelo de atraso RC. Esse modelo aproxima as características não-lineares do transistor com uma capacitância e resistência médias de acordo com o chaveamento da porta. Esse modelo consegue uma boa estimativa para o atraso, apesar de possuir limitações relacionadas ao prever o comportamento analógico da porta (WESTE; HARRIS, 2011). Dessa forma, o atraso de propagação pode ser aproximado pela Equação 2.21.

$$t_p = RC \quad (2.21)$$

Onde C é a capacitância total no nó de saída e R é a resistência efetiva.

O modelo de atraso RC trata o transistor como uma chave ligada em série com uma resistência. Assim, a resistência efetiva é a razão de V_{ds} para I_{ds} média através do chaveamento em um intervalo de interesse.

O modelo RC pode ser usado para se obter o dimensionamento de portas lógicas. Inicialmente, obtêm-se os valores iniciais das dimensões dos transistores que compõem um inversor. Esses valores devem fazer com que a razão entre os tempos de subida t_{rise} e de descida t_{fall} seja igual a $t_{rise}/t_{fall} = 1$ (STANGHERLIN, 2013). Os tempos de subida e descida se referem ao tempo máximo no qual a entrada ou a saída levam para ir de 10% até 90% de seu valor máximo, respectivamente. Em seguida, todas as outras células devem ter seus respectivos tamanhos de pMOS e nMOS definidos de forma a se obter uma resistência efetiva igual à do inversor, de forma a minimizar o atraso de cada célula (WESTE; HARRIS, 2011).

Para se obter uma estimativa de como o atraso irá se comportar com as mudanças da tensão de alimentação, pode-se alterar a Equação 2.21 de forma que $t_p = C_L V_{DD} / I_{ds}$. Usando uma aproximação de primeira ordem para o valor de I_{ds} , tem-se que o atraso possa ser estimado segundo a Equação 2.22 (WESTE; HARRIS, 2011).

$$t_p = \beta \frac{C_L V_{DD}}{(V_{DD} - V_{th})^\gamma} \quad (2.22)$$

Onde β é um parâmetro que depende da tecnologia de fabricação e γ é um índice que representa a saturação de velocidade.

Entretanto, essa relação só é válida enquanto $V_{DD} > V_{th}$. Em circuitos que operam na região de inversão fraca, o mecanismo dominante de condução é a corrente sub-limiar, mostrada na Equação 2.12. Nessa situação, o atraso passa a ser dependente dessa corrente sub-limiar, como mostrado na Equação 2.23.

$$t_{p,sub} = \beta \frac{C_L V_{DD}}{I_{sub} 10^{\frac{V_{DD}}{n}}} \quad (2.23)$$

Utilizando as relações mostradas nas Equações 2.12 e 2.13, pode-se modificar a Equação 2.23 para mostrar como o atraso se relaciona com a largura W e o comprimento L dos transistores, como mostrado na Equação 2.24.

$$t_{p,sub} = \beta \frac{C_L V_{DD}}{\left(\frac{W}{L} \mu_0 C_{ox} V_T^2 e^{1,8}\right) e^{\left(\frac{V_{gs} - V_{th}}{n V_T}\right)}} 10^{\frac{V_{DD}}{n}} \quad (2.24)$$

Ao se analisar as Equações 2.22 e 2.23, pode-se perceber o comportamento do atraso em relação à tensão de alimentação nas regiões de inversão forte e fraca, respectivamente. Quando se opera na inversão forte, o atraso tem uma relação linear com o V_{DD} . Entretanto, para operações na região sub-limiar, o atraso aumenta de forma exponencial com a redução da tensão de alimentação. Para exemplificar esse comportamento pode-se utilizar o experimento realizado por Kanitar (2009) mostrado na Figura 9, que mostra a relação entre a tensão de alimentação e o atraso para um oscilador em anel baseado em uma cadeia de inversores, em tecnologia de 65nm.

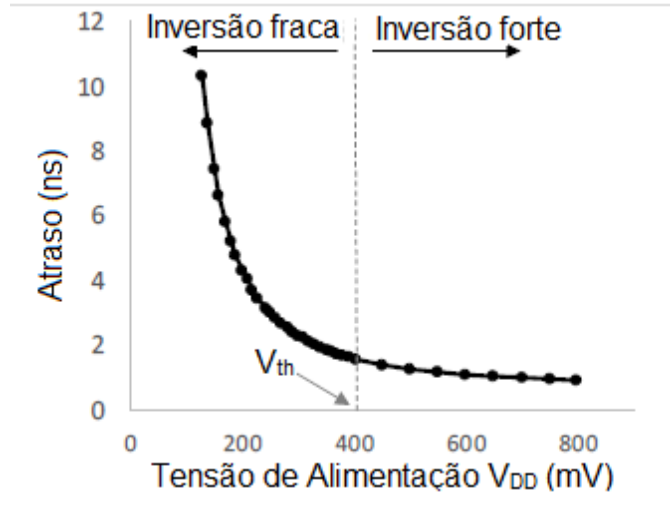


Figura 9 – Atraso versus tensão de alimentação em um oscilador em anel. Fonte: Adaptado de (KANITAR, 2009)

2.3.2 Dissipação de Potência em transistores MOS

Em circuitos eletrônicos, a potência instantânea $P(t)$ consumida ou fornecida por um determinado circuito é definida pela corrente que circula através do dispositivo, multiplicada pela tensão sobre ele, como mostrado pela Equação 2.25.

$$P(t) = I(t)V(t) \quad (2.25)$$

A potência total pode ser dividida em potência dinâmica P_{dyn} e pelo vazamento P_{sta} , como mostrado na Equação 2.26.

$$P(t) = P_{dyn}(t) + P_{sta}(t) \quad (2.26)$$

A potência dinâmica pode também ser dividida em duas outras componentes, sendo elas a potência de chaveamento P_{sw} e a potência de curto-circuito P_{sc} , como mostrado na Equação 2.27

$$P_{dyn} = P_{sw} + P_{sc} \quad (2.27)$$

Em circuitos atuais, a maior parcela da dissipação de energia é devida a potência de chaveamento. Essa, depende principalmente da frequência de chaveamentos e pode ser dada pela Equação 2.28.

$$P_{sw} = \alpha f_{clk} C_L V_{DD}^2 \quad (2.28)$$

Onde, α é o fator de atividade, f_{clk} é a frequência de clock, C_L é a capacitância de carga e V_{DD} é a tensão de alimentação. O fator de atividade α é a probabilidade de um

nó do circuito transitar do nível lógico 0 para o 1, visto que esse é o único momento no qual o circuito consome energia (WESTE; HARRIS, 2011).

A potência de curto-circuito ocorre quando tanto a rede de *pull-up* quanto a rede de *pull-down* estão parcialmente conduzindo no momento em que há uma mudança nas entradas. Nesse estágio, a corrente flui por um caminho direto entre os terminais de V_{DD} e GND . A Equação 2.29 mostra como a potência de curto-circuito pode ser calculada.

$$P_{sc} = W(V_{DD} - 2V_{th})^3 \tau f \quad (2.29)$$

Onde W é uma constante e depende do tamanho do transistor e dos parâmetros da tecnologia de fabricação utilizada, V_{th} é a tensão de gatilho do transistor, τ é o tempo de subida ou tempo de descida do sinal de entrada e f é a frequência do *clock*. A potência de curto-circuito representa cerca de 10% do total da potência dinâmica, sendo usualmente desprezada (HANSON et al., 2006).

O vazamento P_{sta} é a potência consumida mesmo quando o circuito não está chaveando e é resultante das correntes parasitas do transistor. As principais fontes do vazamento são a corrente sub-limiar, a corrente através do óxido da porta e a corrente de tunelamento entre as junções. Esses dois últimos geralmente são desprezados pois os seus valores, usualmente, são muito inferiores ao valor da corrente sub-limiar (BUTZEN; RIBAS, 2007). O vazamento pode ser aproximado pela relação mostrada na Equação 2.30.

$$P_{sta} = V_{DD}(I_{sub} + I_{gate} + I_{dio}) \quad (2.30)$$

Desprezando I_{gate} e I_{dio} , e utilizando a relação para I_{sub} , mostrada na Equação 2.12, tem-se a Equação 2.31, que descreve o vazamento para circuitos que operam na região sub-limiar.

$$P_{sta} \approx V_{DD} \left(I_0 e^{\left(\frac{V_{gs} - V_{th}}{nV_T} \right)} (1 - e^{-\frac{V_{ds}}{V_T}}) \right) \quad (2.31)$$

Atualmente, a maioria dos designs convencionais utilizam um valor de V_{DD} acima do limiar V_{th} , fazendo com que o sistema opere na região de inversão forte, ou *superthreshold*. Nesse contexto, o mecanismo principal para o consumo de energia é a potência de chaveamento, mostrada na Equação 2.28. Por outro lado, em circuitos operando na região sub-limiar, o vazamento passa a ter maior dominância no consumo total do sistema. Dessa forma, a potência devida à corrente sub-limiar passa a ser o principal mecanismo para o consumo energético total do sistema.

A variação da tensão de alimentação V_{DD} afeta tanto a potência dinâmica quanto o vazamento. De acordo com as Equações 2.28 e 2.31, ela possui uma relação quadrática com a potência de chaveamento e uma relação linear com o vazamento. Dessa forma, reduzir o

V_{DD} é a forma mais efetiva de minimizar o consumo de energia de um circuito eletrônico (CHANDRAKASAN; SHENG; BRODERSEN, 1992). Entretanto, essa redução também traz algumas penalidades como, por exemplo, uma maior sensibilidade dos transistores às variações de processo, ambiente e envelhecimento (WESTE; HARRIS, 2011). Um outro fator negativo é um deterioramento na performance geral do sistema, devido ao aumento do atraso de propagação.

2.3.3 Energia em circuitos integrados

Como mostrado pela Equação 2.25, a potência instantânea em um circuito é determinada pela relação entre sua tensão de alimentação e a corrente que flui pelo dispositivo em um determinado instante de tempo. Dessa forma, a energia E consumida ou fornecida a um dispositivo pode ser descrita como a integral dessa potência instantânea, durante um período de tempo t , de acordo com a Equação 2.32

$$E = \int_0^T P(t) dt \quad (2.32)$$

Da mesma forma como a potência, a energia consumida também pode ser dividida em energia dinâmica E_{din} e energia devido ao vazamento E_{sta} , de forma que $E_{total} = E_{din} + E_{sta}$.

Considerando o inversor da Figura 7a conectado a uma carga capacitiva C_L e a relação mostrada na Equação 2.32, a energia dinâmica E_{din} em cada ciclo é dada pela Equação 2.33 (WESTE; HARRIS, 2011).

$$E_{din} = \int_0^\infty I(t)V_{DD}dt = \int_0^\infty C \frac{dV}{dt} V_{DD} dt = CV_{DD} \int_0^{V_{DD}} dV = CV_{DD}^2 \quad (2.33)$$

No circuito do inversor, apenas um dos transistores está ativo por vez em cada ciclo. Isso faz com que apenas metade da energia fornecida pela fonte de alimentação seja armazenada no capacitor de carga. A outra metade é dissipada por calor. Dessa forma, a energia dinâmica de uma porta lógica é dependente apenas da capacitância de carga conectada a ela.

A energia devida ao vazamento depende apenas do tempo no qual a corrente sub-limiar flui pelos transistores. Assim, essa energia passa a ter relação direta com o atraso do dispositivo, como mostrado na Equação 2.34 (WESTE; HARRIS, 2011).

$$E_{sta} = V_{DD} I_{sub} t_{p,sub} \quad (2.34)$$

Na maior parte dos circuitos atuais que operam na região linear, a energia dinâmica é o mecanismo dominante no consumo total do sistemas, representando cerca de 90% do total. Dessa forma, a redução da tensão de alimentação passa a ser a melhor estratégia

para se reduzir o consumo total. Essa relação é válida até o limite do regime de operação do circuito. Em sistemas operando na região sub-limiar, deve-se atentar para técnicas de redução do atraso, visto que nessa região o vazamento passa a ser o mecanismo dominante no consumo geral de energia do sistema (BUTZEN; RIBAS, 2007).

Tendo em mente as características apresentadas até então, para se desenvolver circuitos que operem no regime sub-limiar, o projetista tem que levar em conta a escolha apropriada da tensão de alimentação que será usada no projeto, visto que o atraso tende a aumentar de forma exponencial com a redução da tensão de alimentação V_{DD} . Além disso, o correto dimensionamento dos transistores também deve ser realizado, já que essa grandeza tem impacto direto no atraso, como mostrado na Equação 2.24. Dessa forma, o projetista deve dispor de ferramentas que o auxiliem na escolha desses parâmetros, como aquele apresentado nesse trabalho.

2.4 Bibliotecas de células Padrão

Uma biblioteca de células padrão é um conjunto de células combinacionais e sequenciais construídas para serem usadas em fluxos de design eletrônico automatizado (do inglês *Electronic Design Automation - EDA*). Isso somente é possível devido ao fato de todas as células seguirem um rigoroso padrão de implementação. Esse fator auxilia na síntese de projetos que utilizem um alto nível de abstração, além de fazer com que as células possam ser combinadas, posicionadas e roteadas livremente (OLIVEIRA, 2017).

Uma biblioteca de células padrão típica contém um conjunto de diferentes formatos de cada célula, como esquemático, leiaute, lógico e abstrato, assim como informações sobre atrasos, potência e ruído.

O leiaute é a implementação física de uma célula padrão e deve seguir algumas definições para que as principais características dessa metodologia possam ser alcançadas. Os trilhos onde percorrem a tensão de alimentação V_{DD} e o terminal de terra GND devem ser alinhados horizontalmente nas partes superiores e inferiores da célula, respectivamente. Os transistores pMOS devem ser colocados na parte superior, enquanto que os transistores nMOS devem ser alinhados na parte inferior. Esse alinhamento deve ser seguido para que as células possam ser conectadas pelo encontro dos trilhos de alimentação e pelo poço do tipo n dos transistores pMOS (WESTE; HARRIS, 2011).

Uma outra característica do leiaute de uma célula padrão é que todos os componentes da célula estão alinhados em uma grade de roteamento padrão. As dimensões das células que compõem uma biblioteca devem seguir regras de design específicas de acordo com as grades horizontais e verticais. A altura de todas as células deve ser exatamente a mesma, sendo esse valor múltiplo da distância entre os trilhos da grade horizontal. O valor da largura das células pode ser variável, entretanto, ele deve ser um múltiplo da grade

vertical (C.SAINT; SAINT, 2002).

O formato abstrato é uma versão simplificada do leiaute da célula e é usado pelas ferramentas automáticas de roteamento. Nesse formato, são mostrados apenas os trilhos de alimentação, as camadas de metal e os pinos de cada célula. As diferenças entre o leiaute e o formato abstrato são mostrados na Figura 10.

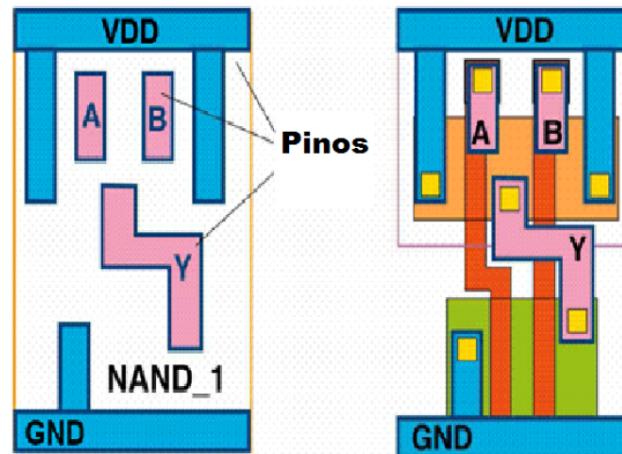


Figura 10 – Leiaute versus formato abstrato de uma célula NAND2. Fonte: Adaptado de (KULKARNI, 2012)

O formato lógico de uma célula padrão é composto pela descrição lógica de cada células utilizando uma linguagem de descrição de hardware, como VHDL ou Verilog. Nessa descrição devem aparecer todos os pinos que compõem a célula, especificando se os mesmos são pinos de entrada ou saída de sinais. Uma outra característica do formato lógico é a especificação do atraso entre cada um dos pinos da célula (OLIVEIRA, 2017).

A caracterização é o processo no qual todas as informações são extraídas das células. A caracterização se inicia com a extração das resistências e capacitâncias parasitas a partir do leiaute da célula. De posse desses dados, são realizadas simulações exaustivas do modelo da célula, sob várias condições de operação distintas e com alterações nos parâmetros de processo. Ao final da caracterização, as informações relacionadas ao consumo de energia e ao atraso de cada uma das células são compilados e fornecidos em um único arquivo, que é utilizado pelas ferramentas de síntese digital de circuitos.

2.5 Memórias SRAM

As memórias são parte integrante de grande parte do design de aplicações em circuitos integrados. As memórias podem ser divididas em memórias de acesso aleatório (RAM - do inglês *random-access memory*) ou memórias apenas de leitura (ROM - do inglês *read-only memory*). As memórias RAM também poder ser divididas em memórias estáticas

(SRAM) e memórias dinâmicas (DRAM). Esse trabalho irá focar nas memórias SRAM, devido ao fato de serem mais comumente usadas em projetos de circuitos integrados.

A memória SRAM é assim chamada devido ao fato de o dado armazenado na mesma poder ser retido indefinidamente, enquanto ela ainda estiver alimentada com a tensão V_{DD} . De forma análoga, as memórias DRAM necessitam que o dado seja recarregado a cada determinado período de tempo afim de evitar que o dado armazenado seja perdido, em um processo chamado refrescagem.

A Figura 11 mostra a estrutura geral de uma célula de memória SRAM. Essa célula é o componente mais básico desse tipo de memória e é responsável pelo armazenamento de um bit de dado, seja ele 0 ou 1.

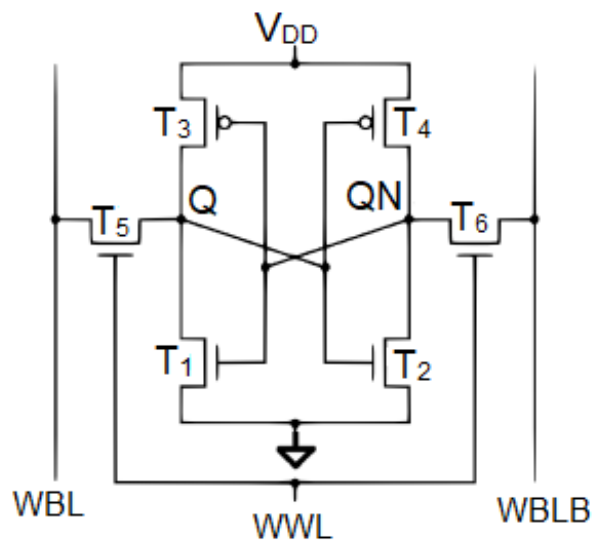


Figura 11 – Estrutura geral de uma memória SRAM. Fonte: Adaptado de (BRITO, 2014)

Essa estrutura é composta por dois inversores ($T_1 - T_3$ e $T_4 - T_2$) ligados entre si em uma configuração de acoplamento cruzado, no qual a saída de um é ligada a entrada do outro. Esse tipo de ligação tem como função manter o dado que está armazenado na célula a medida que sua carga vai sendo reduzida. Os transistores restantes (T_5 e T_6) são usados para controlar a entrada e saída dos dados na memória e são conectadas as linhas de endereço WWL (*Write Wordline*). Os dados chegam até a memória vindos das linhas de bit WBL (*Write Bitline*) e sua complementar $WBLB$ (*Write Bitline Bar*).

A estrutura mostrada na Figura 11 é útil em grande parte das aplicações que operam na região *superthreshold* dos transistores. Entretanto, para sistemas operando na região sub-limiar, esse tipo de configuração já não é a mais recomendada, visto que essa configuração não cumpre os requerimentos necessários de densidade e de rendimento. Esse comportamento é devido ao fato dessa configuração possuir uma baixa margem de ruído

estático (SNM), uma baixa capacidade de escrita, número limitado de células por cada linha de bit e baixa margem de sensibilidade na linha (KIM et al., 2008).

Para mitigar esses problemas, a estrutura básica da memória SRAM, utilizando 6 transistores, pode ser expandida para uma que se adapte a operações na região sub-limiar. A Figura 12 mostra uma configuração que utiliza 10 transistores e que pode ser usada para esse tipo de operação.

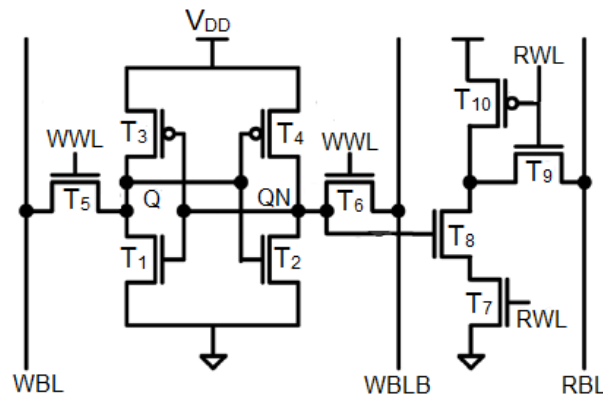


Figura 12 – Estrutura de uma memória SRAM para região sub-limiar. Fonte: Adaptado de (KIM et al., 2008)

Essa célula consiste na mesma estrutura com os inversores em configuração de acoplamento cruzado da célula tradicional acrescido de um circuito adicional, composto por 4 transistores, responsável pela operação de leitura dos dados armazenados. Além disso, essa configuração possui linhas de bit individuais para escrita (WBL ()) e $WBLB$) e leitura (RBL - *Read Bitline*) e uma linha de endereço específica para a leitura (RWL - *Read Wordline*).

A operação de escrita para esse tipo de célula e para a tradicional é a mesma. Inicialmente, as linhas de bit devem ser pré-carregadas com o valor da tensão de alimentação para que o atraso seja reduzido nessa operação (S.SEBASTINSURESH et al., 2013). Em seguida, o dado a ser escrito é carregado na linha de bit e um pulso é dado na linha de endereço WWL fazendo com que o dado entre no circuito de inversores e ali fique armazenado.

Para a leitura do dado armazenado, a linha de bit de leitura RBL também deve ser pré-carregada com o valor da tensão de alimentação. Quando a leitura é habilitada, colocando a linha de endereço $RWL = 1$, a linha de bit é descarregada condicionalmente através dos transistores T_7 , T_8 e T_9 dependendo do dado armazenado na memória. O restante da memória fica desacoplada da linha de bit de leitura, o que faz com que a margem SNM seja melhorada. No momento em que a leitura é desabilitada ($RWL = 0$), o transistor T_{10} é ativado, fazendo com que o vazamento da linha de bit seja descarregado,

independente do dado armazenado na memória. Esse efeito faz com que uma maior quantidade de células seja conectada a uma única linha de bit (KIM et al., 2008).

2.6 Estado da Arte

Essa seção tem como função mostrar as pesquisas e trabalhos mais recentes sobre circuitos e aplicações que operam na região sub-limiar. Serão discutidas nessa seção, trabalhos que desenvolvem bibliotecas de células padrão que operem nessa região, além das técnicas mais recentes para se obter um ultra-baixo consumo de energia em circuitos integrados.

Para a geração de bibliotecas que operem na região sub-limiar, várias diretrizes devem ser observadas, visto que os critérios de design para essa região são diferentes para a operação na região linear. Com isso em mente, o trabalho de Timarchi e Alioto (2015) faz uma análise das principais características que devem ser observadas ao projetar esse tipo de biblioteca. Um estudo de caso é realizado em tecnologia de $180nm$, para aplicar os princípios apresentados.

Trabalhos como o apresentado por Gao et al. (2015) realizam a geração da biblioteca para o uso em qualquer projeto que deva ter um ultra-baixo consumo de energia. Nesse estudo, os autores conseguiram uma redução de cerca de 40% para o consumo dinâmico e 97% para o vazamento, quando comparada com uma biblioteca tradicional. Entretanto, a geração das células pode ser realizada tendo em mente um fim específico. Em Jeong et al. (2013) o foco dos trabalhos são aplicações biomédicas. Os autores estudam a redução da tensão de alimentação e os efeitos do dimensionamento em uma tecnologia de $180nm$, além de gerar duas bibliotecas com tensões de $0,6V$ e $0,45V$. Os resultados encontrados mostraram uma redução de fator 18 comparada com uma biblioteca operando com $V_{DD} = 1,8V$.

Diversas técnicas podem ser empregadas para se obter os melhores desempenhos em sistemas que operam na região sub-limiar. O maior foco dos trabalhos nessa área é a tentativa de maximizar a performance dos sistemas, enquanto ainda se mantém um baixo consumo de energia.

Uma das formas de se obter esses resultados é a aplicação de técnicas para se obter o melhor dimensionamento dos transistores. Uma dessas técnicas é a otimização multi-objetivo (VOHRMANN et al., 2015), que leva em conta a margem de ruído, o consumo e o atraso nos cálculos de otimização para o dimensionamento, obtendo uma maior robustez e melhor desempenho de cada célula. Testes utilizando essa técnica demonstram ganhos em energia em um fator de 9 vezes, operando a uma tensão de alimentação de $250mV$, quando comparados com uma biblioteca comercial que opera com $V_{DD} = 1,2V$.

Uma outra maneira de se abordar o dimensionamento dos transistores é a aplicação de técnicas que visam mitigar ou aproveitar os efeitos das características não-ideais dos transistores. Os efeitos causados pelo RSCE, ao se utilizar transistores com maiores comprimentos, podem ser utilizados para a geração de bibliotecas completas (KIM et al., 2007) ou para a construção de memórias SRAM (KIM et al., 2008). A utilização desse efeito gera ganhos de cerca de 13% para o atraso e de 50% para o consumo de energia. Ganhos semelhantes também podem ser alcançados ao se dimensionar levando em conta fenômenos que são significantes em circuitos operando na região sub-limiar, como por exemplo o efeito inverso de largura estreita (INWE - do inglês *inverse narrow width effect*) (ZHOU et al., 2014). Esse efeito ocorre em transistores que possuem largura estreita e causa relações não-proporcionais entre a corrente de dreno e a largura do transistor. Como consequências, pode-se observar um aumento no atraso e no consumo de energia do sistema (XINFU et al., 2008).

Além disso, os avanços nas pesquisas possibilitaram a implementação de processadores que operam na região sub-limiar para aplicações em redes ubíquas de sensores e computação. Nesse contexto, podem ser explorados cenários de aplicações com alto fluxo de dados e de computação intensa para sistemas de multimídia ou processadores de baixo custo e baixo fluxo de dados para sistemas de Internet-das-Coisas (NING, 2013).

3 Metodologia

Nesse capítulo será apresentada a metodologia desenvolvida para a geração de uma biblioteca de células padrão que operem na região sub-limiar. Essa metodologia é composta por uma sequência de etapas pre-determinadas que, ao final, irá fornecer uma biblioteca totalmente caracterizada e pronta para o uso em ferramentas de design eletrônico automatizado.

Como visto no capítulo anterior, os trabalhos encontrados na literatura visam técnicas que procuram mitigar os efeitos negativos e explorar os efeitos positivos de se trabalhar na região sub-limiar. Os estudos de [Kim et al. \(2007\)](#) e [Zhou et al. \(2014\)](#) buscam realizar de novas formas o dimensionamento dos transistores. Entretanto, esse estudo é realizado para uma determinada tensão de alimentação, que não é alterada em nenhuma etapa. Ao final, os autores realizam apenas a caracterização das células para outros possíveis valores e não o redimensionamento. Esse artifício deve ser usado de forma cuidadosa, visto que a alteração do V_{DD} pode ter impacto direto nos valores de atraso e consumo, conforme visto nas seções anteriores. Já nos trabalhos de [Gao et al. \(2015\)](#) e [Jeong et al. \(2013\)](#) são construídas várias bibliotecas, com diferentes tensões de alimentação, com os autores escolhendo aquelas que tem os melhores resultados. Essa tarefa poderia ser realizada de forma mais simples e mais rápida se os autores já soubessem de antemão qual, ou quais, valores de V_{DD} gerariam os melhores resultados.

Baseando-se nessa demanda, uma intensa pesquisa na literatura foi realizada, visando encontrar soluções que englobem, ao mesmo tempo, a análise e a manipulação desses dois parâmetros. O fluxo proposto nesse trabalho e que será mostrado nesse capítulo, foi motivado pelo resultado infrutífero dessa análise da literatura. A principal funcionalidade do método proposto é servir como um auxílio ao pesquisador, fornecendo maneiras de se encontrar o melhor valor de V_{DD} para seu projeto antes da aplicação das técnicas propostas, além de entregar os melhores parâmetros iniciais para o dimensionamento e de auxiliar na caracterização e verificação das células criadas. É importante frisar que o fluxo proposto nesse trabalho não substitui as técnicas já existentes, mas fornece ferramentas que podem ser utilizadas em concomitância com estas. Apesar disso, o fluxo também pode ser aplicado de forma independente, pois possui maneiras de entregar a biblioteca totalmente funcional sem aplicação de outras ferramentas.

Esse capítulo está dividido da seguinte forma: inicialmente, o fluxo será apresentado de uma forma geral, além dos principais parâmetros de entrada que o compõem. Em seguida, as seções que fazem parte do fluxo serão mostradas, detalhando cada uma das etapas. Por fim, os dados e arquivos de saída do fluxo serão detalhados, mostrando como

esses podem ser usados em outros fluxos de EDA.

3.1 Fluxo para geração de biblioteca de células padrão

O fluxo proposto é dividido em duas seções. A primeira parte tem como objetivo a estimativa do V_{DD} e dos parâmetros iniciais para o dimensionamento das células, enquanto que a segunda parte é relacionada à construção das células e sua caracterização.

Essa divisão em duas etapas é importante, pois visa deixar uma separação entre os diferentes processos que serão revistos pelo fluxo. Ao final da primeira metade do fluxo, serão definidos importantes parâmetros que são fundamentais para o restante dos trabalhos e para a utilização da biblioteca que será gerada, tais como a tensão de alimentação, o comprimento e a largura dos transistores que será utilizado. Já ao termino da segunda etapa, a biblioteca estará totalmente caracterizada e verificada, podendo ser usada em qualquer fluxo de design automatizado.

A Figura 13 mostra uma visão geral do fluxo. O mesmo começa com os dados da tecnologia escolhida e as restrições de design e termina com uma biblioteca totalmente caracterizada. Nas etapas mostradas no lado esquerdo, com a cor rosa, são mostrados os passos relacionados à exploração da tecnologia, enquanto que no lado direito, em azul, os passos mostrados focam na construção das células e na caracterização.

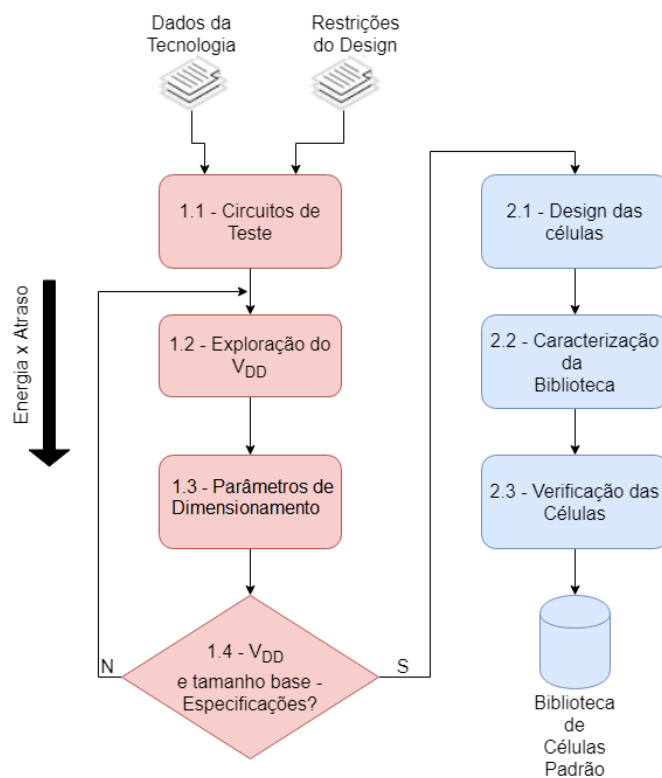


Figura 13 – Visão geral do fluxo proposto para o desenvolvimento de uma biblioteca de células padrão

Os principais dados de tecnologia que devem ser utilizados como entrada do fluxo são os valores da tensão de alimentação e de limiar padrão da tecnologia, os tamanhos máximos do comprimento e largura dos transistores e os valores das distâncias padrão entre os trilhos das grades vertical e horizontal para o leiaute das células.

Com relação às restrições de design, devem ser observadas os parâmetros relacionados com a performance do sistema, como a frequência de operação máxima e mínima necessária para o correto funcionamento, bem como as limitações relacionadas ao consumo energético geral. Essas restrições devem ser observadas e seguidas durante todo o fluxo, visto que são elas que irão pautar as decisões que o designer irá tomar durante toda sua execução.

É importante notar que o fluxo proposto pode ser utilizado em qualquer tecnologia de fabricação que o projetista escolher. Entretanto, deve ser levado em conta as variações dos parâmetros de cada tecnologia, como as tensões de alimentação e de limiar padrão, além dos valores mínimos para a largura e comprimento dos transistores.

3.1.1 Exploração da tecnologia

Essa é a primeira seção que compõe o fluxo. A partir dos dados de entrada, citados anteriormente, a tensão de alimentação que será usada durante todo o restante do projeto será determinada. Além disso, o comprimento e largura iniciais dos transistores também serão definidos nessa etapa.

A primeira etapa do fluxo, chamada de *1.1 - Circuitos de teste*, tem como função determinar um conjunto de circuitos de teste, que são selecionados de forma a servir como base para todas as etapas subsequentes. Esses circuitos devem ser capazes de testar um conjunto de células lógicas básicas, além de habilitar a variação dos sinais de entrada e saída e da tensão de alimentação geral do sistema. Além disso, é necessário que hajam cargas de entrada e saída apropriadas, de forma a tentar emular o comportamento real do sistema. Uma alternativa para o teste de células combinacionais, é a utilização de uma configuração de *fanout-of-4* (FO4) da célula que está sendo analisada. A configuração de FO4 quer dizer que a saída da célula em análise está conectada em 4 cópias da mesma célula, sendo que os tamanhos dos transistores para todas as cópias deve ser o mesmo (HARRIS et al., 2015). Embora essa configuração seja a mais utilizada, podem ser utilizadas outras versões de acordo com a necessidade do projeto que está sendo realizado.

Em circuitos sequencias, pode-se usar células básicas desse tipo para se realizar os testes. Como exemplo, pode-se utilizar um circuito composto por um flip-flop juntamente com um conjunto de células combinacionais, necessárias para habilitar a manipulação dos sinais de entrada e *clock*. Assim, aplica-se um bit na entrada do flip-flop e mede-se o tempo necessário para esse sinal ser lido na saída, a partir do pulso de *clock*. Dessa forma,

pode-se encontrar o atraso correspondente dessa célula. Nessa mesma operação, pode-se também avaliar o consumo de energia necessário para célula realizar essa operação.

Assim, apesar dos circuitos recomendados aqui, pode-se utilizar outras configurações para os circuitos de teste, de acordo com os requisitos do projeto no qual está sendo feita a avaliação. No entanto, a única restrição para a utilização de outras configurações é que se tenha total controle dos sinais de entrada e que os sinais de saída possam ser lidos sem maiores interferências. Isso se deve ao fato de que os parâmetros que serão escolhidos nos próximos passos do fluxo dependem diretamente desses sinais.

A escolha do circuito de teste apropriado é necessária em cada rodada do fluxo e o circuito escolhido pode variar de acordo com a tecnologia escolhida. Isso se deve ao fato de que cada uma tem a sua particularidade e os seus parâmetros de teste. Dessa forma, um circuito usado em um teste pode não servir como parâmetro para outra tecnologia. Além disso, esses circuitos devem ser simples o suficiente para que os testes não tenham elevado gasto de tempo, mas complexos o bastante para que sejam representativos para a projeto.

O próximo passo do fluxo, chamado *1.2 - Exploração do V_{DD}* , se refere a relação entre a tensão de alimentação e a performance para a tecnologia escolhida. Nessa etapa, a tensão de alimentação em cada circuito de teste é variada, com passos de 10%, a partir do menor valor identificado para a tecnologia e que resulte em saídas corretas $V_{DD,min}$. Em seguida, os valores máximos para a dissipação de energia por transição e o atraso de transição são extraídos. Aqui, os tamanhos dos transistores não são alterados. Entretanto, devidos aos efeitos do SCE e RSCE, explicados na seção 2.1.2.3 do capítulo 2, o dimensionamento tem impacto direto na tensão de limiar dos transistores. Dessa forma, após a definição dos parâmetros iniciais dos transistores, ao final da primeira etapa do fluxo, valores diferentes para o $V_{DD,min}$ poderão ser observados. Apesar disso, esses fenômenos podem ser ignorados nesse passo do fluxo, pois os valores encontrados aqui podem fornecer uma ideia razoável da relação entre o consumo de energia e o atraso. Os efeitos causados pelo RSCE serão importantes no passo seguinte do fluxo, principalmente na definição dos parâmetros iniciais para o comprimento dos transistores. Como será visto adiante, essa etapa do fluxo poderá ser realizada novamente, após a definição dos parâmetros de dimensionamento, com intuito de fornecer um valor mais preciso para o V_{DD}

Baseado nos dados encontrados nesse passo, o fluxo irá retornar qual tensão de alimentação poderá ser usada para o restante do processo e para a biblioteca gerada, apoiando-se na escolha do projetista relacionada aos parâmetros iniciais de consumo e atraso. Essa escolha é motivada pelo fato de que as relações extraídas dos circuitos de teste podem ser levadas à designs maiores. Isso significa que pode-se fazer uma aproximação do ganho no consumo de energia relativo à perda de performance para a aplicação final.

A etapa *1.3 - Parâmetros de dimensionamento* é focada nos parâmetros relacionados

ao dimensionamento inicial e que serão aplicados como valores iniciais durante as etapas subsequentes de desenvolvimento das células. Esses parâmetros definem o comprimento da porta L e as larguras mínimas para os dispositivos nMOS $W_{nmos,min}$ e pMOS $W_{pmos,min}$, além das relações entre os tamanhos do transistor em estruturas simples.

A análise é iniciada com a seleção do comprimento dos transistores. Mantendo-se fixo o valor da tensão de alimentação e da largura, o comprimento do nMOS e pMOS é setado para o valor mínimo da tecnologia. Então, são extraídas e guardadas as informações sobre o atraso e da dissipação de energia máximos. O processo é repetido a medida que se incrementa o comprimento dos transistores, em passos de 10%, até o limite de três vezes o valor mínimo. Esse valor é assim definido, pois os efeitos da manipulação do comprimento dos transistores, devido ao RSCE, são desprezíveis para valores que excedam esse limite (KIM et al., 2007). Assim, baseado nos requisitos iniciais definidos pelo projetista para o consumo de energia e atraso máximo, o fluxo irá retornar o melhor valor de comprimento para os transistores e que será usado em toda biblioteca.

Em seguida, a largura de todos os dispositivos também é colocada no valor mínimo definido pela tecnologia. Então, a largura do pMOS é aumentada até que os valores para o atraso de subida t_{rise} e descida t_{fall} sejam iguais, com uma margem de $\pm 10\%$, até o limite de seis vezes o tamanho mínimo do pMOS (STANGHERLIN, 2013). Caso essa relação não seja alcançada, o valor da largura do nMOS é acrescida em 10% e os testes são repetidos. O limite máximo de seis vezes o valor do pMOS pode ser maior ou menor, dependendo das limitações quanto à área disponível para projeto.

Os acréscimos de 10% em cada passo se referem à precisão que o projetista terá na definição dos parâmetros iniciais. Esse valor foi escolhido devido ao fato de fornecer valores razoáveis em um determinado número de simulações. O tamanho desse passo poderá ser aumentado ou reduzido de acordo com os requisitos definidos pelo projetista. Entretanto, deve ser levada em conta que a redução desse valor, apesar de fornecer dados mais precisos, aumenta o tempo gasto para a conclusão total do fluxo, visto que mais simulações e testes serão necessários.

É importante notar que a escolha do V_{DD} deve ser feita antes da escolha dos parâmetros iniciais do dimensionamento. Devido ao fato da última possuir um grande número de simulações, possuir o valor da tensão de alimentação definido faz com diversos cenários não tenham a necessidade de serem testados, o que gera um ganho na velocidade da aplicação do fluxo. Entretanto, esses dois passos do fluxo podem ser realizados de forma iterativa. Depois do projetista definir os parâmetros iniciais para o dimensionamento, ele pode retornar à etapa de definição de V_{DD} e aplicá-lo novamente, utilizando os valores de dimensionamento encontrados, de forma a se obter os melhores resultados possíveis.

Ao se realizar a escolha do V_{DD} e dos parâmetros iniciais de dimensionamento a priori, o fluxo nos traz uma melhoria em relação aos trabalhos do estado-da-arte. Como essas

escolhas são baseadas nos requisitos definidos para cada projeto e nas decisões do projetista, tem-se que todas as decisões subsequentes serão tomadas levando em consideração uma grandeza que já está definida de acordo com o trabalho que está sendo realizado e que o desempenho das demais técnicas que serão utilizadas pelo projetista será maximizado.

Um resumo de toda a primeira parte do fluxo pode ser visto através do Algoritmo 1. Nesse algoritmo, E_{req} e E_{sim} correspondem aos requisitos iniciais de consumo de energia e os simulados pelos testes do fluxo, respectivamente. A mesma relação também é válida para os parâmetros d_{req} e d_{sim} , respectivamente. É importante salientar que os valores retornados pelo fluxo serão escolhidos pelo projetista a partir dos requisitos iniciais propostos e da análise dos dados gerados.

Algoritmo 1: Definição de V_{DD} e parâmetros iniciais de dimensionamento

```

início
   $V_{DD} \leftarrow V_{DD,min}$ 
  enquanto  $V_{DD} < V_{DD,max}$  faça
    Simulação com  $V_{DD}$ 
    se  $E_{sim} \neq E_{req}$  e  $d_{sim} \neq d_{req}$  então
       $V_{DD} \leftarrow V_{DDmin} + 10\%$ 
    senão
      | retorna  $V_{DD}$ 
    fim
  fim
fim
   $L \leftarrow L_{min}$ 
  enquanto  $L \leq 3(L_{min})$  faça
    Simulação com  $L$ 
    se  $E_{sim} \neq E_{req}$  e  $d_{sim} \neq d_{req}$  então
       $L \leftarrow L + 10\%$ 
    senão
      | retorna  $L$ 
    fim
  fim
fim
   $W_{pmos} \leftarrow W_{pmos,min}$ 
   $W_{nmos} \leftarrow W_{nmos,min}$ 
  enquanto  $t_{rise} \neq t_{fall}$  faça
    Simulação com  $W$ 
    se  $t_{rise} \neq t_{fall}$  então
       $W_{pmos} \leftarrow W_{pmos-min} + 10\%$ 
      se  $W_{pmos} = 6(W_{pmos-min})$  então
        |  $W_{nmos} \leftarrow W_{nmos} + 10\%$ 
        |  $W_{pmos} \leftarrow W_{pmos-min}$ 
      fim
    senão
      | retorna  $W_{pmos}, W_{nmos}$ 
    fim
  fim
fim
fim

```

3.1.2 Design das células e caracterização

A segunda seção do fluxo descreve a criação das células e da caracterização das mesmas. Como dados de entrada, tem-se os valores de tensão de alimentação encontrados na seção anterior. No final dessa sequência de etapas, a biblioteca estará finalizada.

A primeira etapa dessa seção, *2.1 - Design das células*, se refere a geração dos esquemáticos e do leiaute das células que irão compor a biblioteca final. Além dos dados de dimensionamento extraídos da seção anterior, os dados relacionados aos trilhos das grades vertical e horizontal para o leiaute das células também são usados nessa etapa.

Em primeiro lugar, devem ser gerados os esquemáticos de todas as células que compõem a biblioteca. Para se realizar o dimensionamento, deve-se usar como ponto de partida os valores mínimos encontrados na seção anterior. Esse dimensionamento pode ser realizado de forma manual ou automatizada, com o auxílio de diversas ferramentas. Além disso, é nessa etapa que a grande parte das técnicas apresentadas no estado-da-arte do desenvolvimento de circuitos de ultra-baixo consumo de energia são aplicadas. Como mostrado na seção 2.6 do capítulo anterior, recentes técnicas para aplicações sub-limiares se referem a diferentes maneiras de se analisar e dimensionar as células que compõem as bibliotecas. Como o fluxo proposto retorna os parâmetros iniciais de dimensionamento baseado nos requisitos do projeto que se está trabalhando, a aplicação das técnicas propostas torna-se mais simples e rápido, visto que o projetista já terá em mãos dados pertinentes ao seu sistema. Assim, várias etapas de testes e análises podem ser evitadas, gerando ganhos no tempo e na eficiência das técnicas desenvolvidas.

Em seguida, é realizada a criação dos leiautes das células. Esses leiautes devem seguir de forma rigorosa todas as restrições e regras específicas de cada tecnologia, como o tamanho das grades vertical e horizontal, a quantidade de camadas de metal, entre outros.

Depois de finalizados os esquemáticos e os leiautes, devem ser realizados testes de DRC (do inglês *Design Rules Check*) e de LVS (do inglês *Layout versus schematic*). O DRC é um teste que verifica se o leiaute respeita as regras estabelecidas pela fábrica para a tecnologia escolhida. Já o LVS verifica se as camadas projetadas no leiaute equivalem ao dispositivo representado pelo esquemático. Nessa etapa também são extraídas as informações sobre capacitâncias e resistências parasitas de cada célula. Esses elementos surgem devido à interações reais dos dispositivos e sua extração é realizada através do leiaute da célula.

A próxima etapa do fluxo, *2.2 - Caracterização da célula*, corresponde a etapa na qual todas as informações da biblioteca serão extraídas. Os dados de entrada para essa etapa são as representações das células criadas, escritas em linguagem de descrição de hardware, como Verilog ou VHDL, além das informações de capacitâncias e resistências parasitas, extraídas na etapa anterior. Utilizando alguma ferramenta de caracterização,

como o *Cadence Liberate* ou o *Synopsis SiliconSmart*, a célula é simulada em diversas condições diferentes e são criadas tabelas de valores para o consumo e atraso para cada célula na biblioteca. Essas tabelas serão usadas posteriormente pelas ferramentas de síntese de circuitos e irão definir os valores finais de potência dinâmica e de atraso total de cada circuito construído com as células padrão.

Para que a ferramenta de caracterização crie essas tabelas, é necessário que o projetista defina *templates* para a caracterização. Os *templates* podem ser de vários tipos, entretanto, nesse trabalho são abordados os *templates* de potência e de atraso. Cada um destes é composto por dois índices distintos, sendo que cada índice possui dois ou mais valores. O primeiro índice representa a variação dos tempos de subida dos pinos de entrada das células, ou seja, o tempo em que o sinal de entrada leva para ir de 10% até 90% do seu valor máximo. O segundo índice representa a variação das capacitâncias conectadas às saídas das células.

A obtenção dos valores que irão compor os índices é realizada através de simulações em ambiente controlado e utilizando uma ou mais células de referência da biblioteca. O primeiro índice deve possuir pelo menos dois valores. Esses valores representam os tempos de subida mínimo e máximo na entrada da célula. O tempo de subida mínimo ocorre quando a célula em teste está sendo alimentada por dispositivos que possuam a maior capacidade de fornecimento de corrente. Por outro lado, o tempo máximo ocorre quando a célula em teste está alimentada por dispositivos com a menor capacidade de fornecer corrente. De maneira similar, o segundo índice também deve possuir pelo menos dois valores. Esses dados representam os valores mínimos e máximos das capacitâncias de saída e dependem da quantidade de dispositivos conectados a célula em teste.

Caso a biblioteca possua tipos diferentes de uma mesma célula, será necessária uma quantidade maior de valores para cada índice. Para encontrar tais valores as simulações devem ser repetidas com as outras células e os valores intermediários incluídos nos índices correspondentes.

Ao final da caracterização, as informações relacionadas ao atraso e ao consumo de todas as células que compõem a biblioteca serão extraídos em um arquivo único que, em conjunto com os leiautes criados, são usados como dados de entrada para outros fluxos de EDA.

A etapa final do fluxo, *2.3 - Verificação das células*, consiste em usar o arquivo de caracterização gerado na etapa anterior além dos leiautes de cada célula e aplicar em um fluxo automatizado de EDA. Nessa etapa, será verificado se as células criadas podem ser aplicadas em um fluxo automático sem que nenhuma regra determinada pela tecnologia criada seja quebrada. Além disso, podem ser realizados nessa etapa, simulações com as células criadas, para que se verifique se as células criadas atendem as restrições colocadas no início do fluxo. Caso ocorra alguma falha nesses testes, será necessário avaliar em qual

momento durante o fluxo em que ocorreu essa falha e refazer aquela etapa. Não havendo nenhum erro, a biblioteca estará finalizada e já poderá ser usada na aplicação desejada.

3.2 Conclusão à metodologia

Ao se analisar os principais métodos existentes na literatura, nota-se a ausência de técnicas que levem em conta, ao mesmo tempo, a escolha da tensão de alimentação e o dimensionamento da células. O fluxo proposto tem como função realizar a complementação das técnicas existentes e servir como uma ferramenta auxiliar para que os projetistas possam fazer as melhores escolhas de acordo com os requisitos do projeto que está trabalhando. Além disso, o fluxo também pode ser aplicado de forma independente, visto que ele cobre todas as etapas necessárias para gerar um biblioteca totalmente funcional e caracterizada.

Com relação às limitações do mesmo, tem-se o fato de que o projetista deve ter total conhecimento da aplicação na qual está trabalhando além de uma boa compreensão dos fenômenos envolvidos, em especial na primeira etapa do fluxo. Esse conhecimento é necessário, pois interpretações erradas podem levar a escolhas não adequadas ao projeto, levando a valores incorretos de atraso e de consumo de energia.

4 Simulações e Resultados

Nesse capítulo serão mostrados os resultados da aplicação do fluxo proposto. Ele será aplicado para dois valores distintos de tensão de alimentação. Ao final do fluxo, a biblioteca gerada será utilizada para a síntese e simulação de circuitos de referência e de um microprocessador de código aberto. Os valores para o consumo de energia, atraso e área serão extraídos e comparados com aqueles gerados por uma biblioteca que utilize o valor padrão para a tensão de alimentação.

O principal objetivo destes testes é demonstrar que o fluxo é capaz de gerar bibliotecas plenamente funcionais e que tenham ganhos relevantes no consumo de energia, se comparadas com bibliotecas tradicionais.

4.1 Aplicação do Fluxo

Para a aplicação do fluxo proposto nesse trabalho, foi escolhida a tecnologia comercial de $130nm$. Essa tecnologia possui uma tensão de alimentação padrão de $1,2V$ e uma tensão de limiar nominal de $410mV$. O comprimento mínimo para os transistores é de $120nm$ e a largura mínima é de $160nm$. As restrições de design utilizadas para esses testes foram setados para que se tenham os menores valores de consumo de energia possível mantendo-se valores de atraso equilibrados.

Para o primeiro passo do fluxo, foram implementados circuitos de teste com a configuração de *fanout-of-4* para as portas do Inversor, NAND2 e NOR2. Através desses circuitos, foi verificado que o menor valor para a tensão de alimentação que consegue gerar dados corretos na saída é $V_{DD} = 0,2V$.

Em seguida, o V_{DD} foi variado, partindo do mínimo de $0,2V$ até o valor padrão da tecnologia de $1,2V$ e foram extraídos o máximo atraso e consumo de energia por transição, para cada valor de V_{DD} . Ambos os parâmetros puderam ser extraídos de diferentes combinações de sinais de entrada nos circuitos de teste. A Figura 14 mostra os resultados encontrados para os circuitos do Inversor(a) e para o NAND2(b).

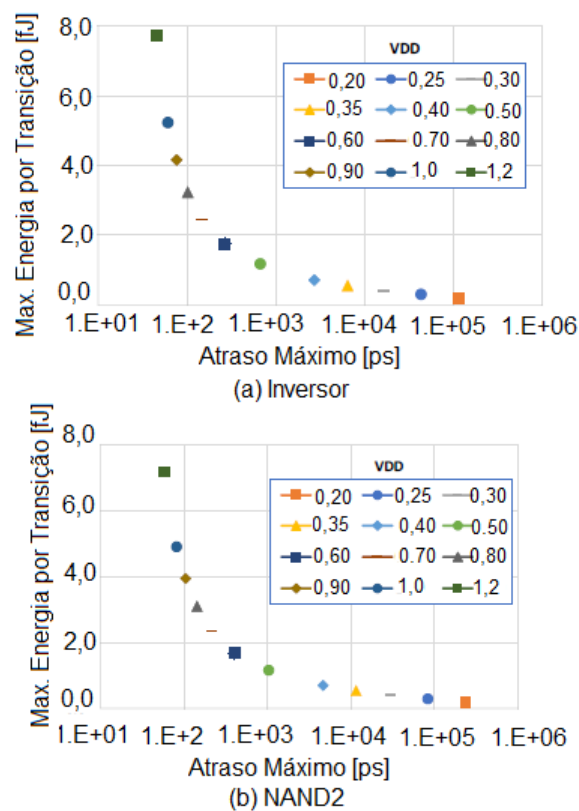


Figura 14 – Comparação do atraso máximo pela energia consumida para os circuitos de Inversor e NAND2

Como pode ser notado pelo gráfico, o consumo de energia pode ser reduzido ao

custo da performance. Tanto na análise do inversor como da NAND2, nota-se que o atraso varia de duas até três ordens de grandeza enquanto que a energia varia em um fator de 20.

De posse desses resultados e seguindo as restrições de design propostas inicialmente, foram escolhidas as tensões de alimentação de $0,25V$ e $0,35V$ para as análises subsequentes. A escolha desses valores de V_{DD} próximos tem como objetivo analisar como será a variação do consumo e do atraso com pequenas variações da tensão de alimentação, utilizando tensões abaixo do limiar e mantendo uma performance equilibrada. Valores menores que $0,25V$ poderiam trazer atrasos muito elevados, enquanto que valores maiores que $0,35V$ já estariam muito próximos ao limiar da tecnologia, fugindo do escopo desse trabalho.

A próxima etapa do fluxo é a definição dos parâmetros de dimensionamento. Na análise dos dados gerados na definição do comprimento do canal dos transistores, foi indicado que para se obter a melhor relação entre a energia e o atraso e atender os requisitos iniciais propostos, os valores para o nMOS e o pMOS deveriam ser diferentes. Dessa forma, o valor encontrado através do fluxo para o nMOS foi de $360nm$ e para o pMOS foi de $240nm$.

Seguindo a análise do dimensionamento, a próxima etapa é a definição dos valores mínimos para a largura dos transistores. Conforme o fluxo, os valores foram setados para os mínimos da tecnologia iguais a $160nm$. Depois das iterações do fluxo através das simulações com os circuitos de teste, definiu-se que a largura base do nMOS seria igual a $W_{nmos} = 200nm$ e do pMOS seria igual a $W_{pmos} = 350nm$.

Os valores encontrados são relacionados à tensão de $0,25V$. Entretanto, uma nova rodada do fluxo foi realizada tendo em vista a tensão de $0,35V$. Nessa análise, concluiu-se que os valores encontrados são válidos para ambas as tensões. Isso quer dizer que não há a necessidade de se redimensionar as células, mas apenas realizar uma caracterização específica para cada valor de V_{DD} .

Esses dados encontrados foram usados como entrada para a segunda etapa do fluxo, que se refere ao design da biblioteca. Assim, foram escolhidas sete células sequenciais de forma que as principais funções lógicas sejam contempladas, além de 3 células combinacionais que implementam tipos diferentes de flip-flops. Também foi adicionada à biblioteca uma célula correspondente à memória SRAM do modelo apropriado para a operação na região sub-limiar, utilizando 10 transistores, como a mostrada na seção 2.5 do capítulo 2.

Inicialmente, deve ser realizada a geração do esquemático e a definição dos tamanhos de cada célula. Essa tarefa foi realizada utilizando o software Cadence Virtuoso. O dimensionamento das células foi realizado de forma manual, tendo como valores base aqueles encontrados na etapa anterior. Para isso, o modelo RC foi utilizado e os tamanhos dos transistores das diferentes células foi escolhido de forma que o valor da resistência efetiva total de cada uma seja igual, ou o mais próximo possível, aos valores de um inversor.

A tabela 1 mostra os valores para a largura de todas as células que compõem a biblioteca. Como dito anteriormente, o valor para o comprimento do canal para todos os transistores é igual a $360nm$ para o nMOS e $240nm$ o pMOS. Os valores entre parêntesis representam as dimensões dos estágios finais de inversão de algumas células.

Tipo de Célula	$W_{nmos}[nm]$	$W_{pmos}[nm]$
INV	200	350
NAND2	350	350
NOR2	200	700
AND2	200 (200)	350 (350)
OR2	200 (200)	700 (350)
XOR2	200	350
AOI22	350	700
Flip-Flop	200 (350)	350 (700)
SRAM	200	350

Tabela 1 – Tabela de dimensionamento da biblioteca

Em seguida, os leiautes foram gerados através do software Cadence Virtuoso. Nessa etapa foram utilizadas as regras de projeto específico da tecnologia de $130nm$. Os testes de DRC comprovou que essas regras foram observadas e o LVS garantiu que os leiautes correspondiam aos esquemáticos. Por fim, foram extraídas as resistências e capacitâncias parasitas, com o auxílio da ferramenta Cadence QRC.

Os arquivos gerados nessa etapa foram usados como base para a caracterização da biblioteca. Essa caracterização foi realizada duas vezes, para contemplar as tensões de alimentação de $0,25V$ e $0,35V$. Para a caracterização foi utilizada a ferramenta Cadence Liberate. Além disso, foram utilizadas condições normais de temperatura, processo e tensão na caracterização.

Para a geração dos *templates* de potência e de atraso, uma série de simulações foram realizadas, utilizando a ferramenta Cadence Virtuoso e uma porta lógica de inversor como célula de referência. Conforme explicado na Seção 3.1.2, o primeiro índice é composto, pelo menos, com os valores mínimos e máximos do tempo de subida da entrada da célula de referência. Para se obter o valor mínimo, a célula foi simulada sendo alimentada por uma configuração de FO4, de forma que ela possa drenar uma maior corrente em sua entrada e assim tenha um menor valor de tempo de subida. De maneira análoga, o valor máximo foi obtido alimentando a célula de referência com uma configuração de FO50, para que a célula de referência drene uma menor corrente. Uma simulação utilizando uma configuração de FO10 também foi realizada, de forma a se obter um valor intermediário para os valores do primeiro índice. Para encontrar os valores que compõem o segundo índice, que é relacionado à carga conectada à célula de referência, o mesmo procedimento foi realizado. Conectando uma configuração de FO4 à saída da célula de referência obtêm-se o valor para o valor

mínimo de carga, enquanto que realizando essa conexão a uma configuração de FO50, obtêm-se o valor para a maior carga. Uma simulação conectando uma configuração de FO10 foi realizada para encontrar um valor intermediário para o índice.

Como resultado dessa etapa, foi gerado um arquivo para cada tensão, que contém todas as informações relacionadas aos consumos dinâmico e vazamento e do atraso para cada uma das células.

Na última etapa do fluxo, os dados gerados pela etapa anterior e as células que compõem a biblioteca foram utilizados para realizar simulações com circuitos de referência e um microprocessador de código aberto. Essa etapa tem como função avaliar a funcionalidade da biblioteca e verificar se os requisitos iniciais propostos foram alcançados.

4.2 Resultados das Simulações

Essa seção tem como função apresentar o desempenho da biblioteca nas simulações realizadas, bem como mostrar o resultado final da aplicação da mesma em um fluxo de EDA.

Vários circuitos foram utilizados para testar o funcionamento e o desempenho das bibliotecas geradas para os dois valores de V_{DD} escolhidos. Além disso, também foi feita a implementação desses circuitos utilizando uma biblioteca comercial na mesma tecnologia e utilizando a tensão de alimentação padrão, com o objetivo de se realizar uma comparação dos resultados obtidos. A biblioteca comercial foi reduzida de modo a conter exatamente as mesmas células que as bibliotecas propostas.

Os circuitos de referência (*benchmarks*) utilizados foram tomados da plataforma ITC'99 (CORNO; RECORDA; SQUILLERO, 2005), que possui um conjunto de circuitos sequenciais, da plataforma EPFL (AMARU; GAILLARDON; MICHELI, 2015), que possui circuitos combinacionais representativos e do núcleo AMBER23 (AMBER, 2011), que é um microprocessador de código aberto com uma arquitetura de 32 bits ARM e um conjunto de instruções ARMv2.

Foram realizados quatro tipos diferentes de síntese para a avaliação do desempenho das bibliotecas e foram coletados os resultados para a área total do circuito, o atraso máximo do caminho crítico e a potências dinâmica, a potência devida ao vazamento e potência total para todos os circuitos.

Na síntese 1, cada circuito foi sintetizado individualmente para cada tensão de alimentação. Essa configuração faz com que a ferramenta defina a melhor arquitetura de cada circuito e faça uma versão otimizada para cada tensão de alimentação. Além disso, um *clock* com frequência de $1MHz$ foi predefinido para todos os circuitos. Esse valor de frequência foi escolhido de forma a definir um desempenho idêntico para todos os

circuitos de teste. Essa configuração tem como objetivo minimizar o consumo de energia, mantendo-se o mesmo desempenho.

Os resultados dessa síntese estão apresentados na Tabela 2 e na Tabela 3.

	Área [mm^2]			Atraso [ns]		
	0,25	0,35	1,2	0,25	0,35	1,2
V_{DD}						
b18	404	393	351	21900	10000	332
b20	76	71	65	21200	9998	354
b20_1	76	74	67	23700	9992	361
b21	80	75	69	23600	9997	365
b21_1	83	77	71	24100	1000	366
b22	117	110	99	24300	1000	361
b22_1	121	113	101	23800	9992	367
Multiplier	142	144	117	31004	7385	557
Mem_ctrl	220	215	168	17530	3928	228
Adder	6	6	6	21137	5068	431
Square	114	111	97	32676	8244	431
BAR	17	17	13	1974	599	51
AMBER	254	251	190	49900	10843	710

Tabela 2 – Resultados de área e atraso para a síntese 1

	Dissipação de Potência [μW]								
	Dinâmica			Vazamento			Total		
V_{DD}	0,25	0,35	1,2	0,25	0,35	1,2	0,25	0,35	1,20
b18	23	45	1317	2	3	68	25	48	1385
b20	1	2	53	0,4	0,5	11	1	2	63
b20_1	0,9	2	53	0,4	0,6	11	1	2	63
b21	0,9	2	51	0,4	0,6	11	1	2	62
b21_1	0,9	2	52	0,4	0,6	11	1	2	63
b22	1,3	2	65	0,6	0,8	16	2	3	82
b22_1	1,3	3	67	0,6	0,9	17	2	3	84
Multiplier	2,9	6	181	0,9	1,3	26	4	7	207
Mem_ctrl	4,6	9	198	1	2	29	6	11	226
Adder	0,4	0,6	6	0,1	0,1	0,9	0,5	0,7	7
Square	2,1	4	132	0,6	0,8	19	3	5	151
BAR	0,4	0,6	15	0,1	0,1	2	0,5	0,7	17
AMBER	4	7	193	1	1	28	5	9	221

Tabela 3 – Resultados de dissipação de potência da síntese 1

A Figura 15 mostra um diagrama de atraso máximo do caminho crítico pela potência total, com resultados para três circuitos da síntese 1.

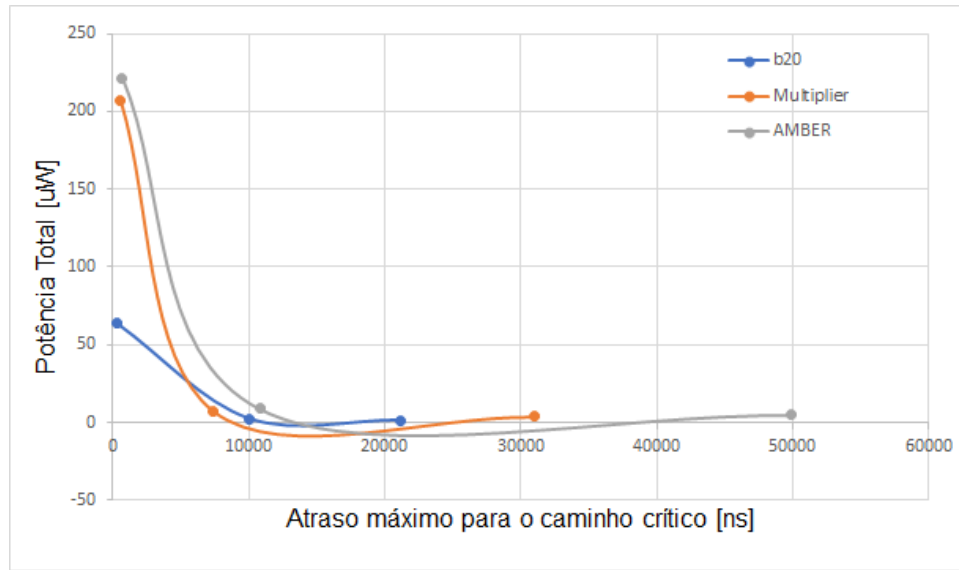


Figura 15 – Atraso do caminho crítico x Potência total para a síntese 1

Analisando os resultados da síntese, mostrados pelas tabelas e pelo diagrama, nota-se que a dissipação de potência teve uma redução média por um fator 44 e 25, para as tensões de 0,25V e 0,35V respectivamente. Por outro lado, o atraso teve um aumento médio por um fator 64 e 16 para as mesmas tensões, respectivamente.

Para a síntese 2, os circuitos também foram sintetizados de forma individual, entretanto para esse caso a frequência de *clock* não foi predeterminada. O objetivo desse tipo de síntese é avaliar o comportamento dos circuitos de testes quando a frequência de trabalho é definida pelo software de síntese. Dessa forma, tem-se uma minimização dos resultados para o atraso, deixando a variação da potência e da área sem restrições. Os resultados desses testes estão mostrados na Tabela 4 e na Tabela 5.

	Área [mm^2]			Atraso [ns]		
	0,25	0,35	1,2	0,25	0,35	1,2
V_{DD}	0,25	0,35	1,2	0,25	0,35	1,2
b18	404	393	351	11800	3824	296
b20	76	71	65	11757	3791	287
b20_1	76	74	67	11617	3737	291
b21	80	75	69	11775	3803	300
b21_1	83	77	71	11617	3737	295
b22	117	110	99	11779	3806	300
b22_1	121	113	101	11616	3740	292
Multiplier	142	144	117	15729	5168	463
Mem_ctrl	220	215	168	8894	2749	189
Adder	6	6	6	10724	3547	358
Square	114	111	97	16578	5769	358
BAR	17	17	13	1002	419	42
AMBER	254	251	190	25316	7587	589

Tabela 4 – Resultados de área e atraso para a síntese 2

	Dissipação de Potência [μW]								
	Dinâmica			Vazamento			Total		
	0,25	0,35	1,2	0,25	0,35	1,2	0.25	0.35	1.20
V_{DD}									
b18	215	449	15263	2	3	66	217	452	15329
b20	34	64	2469	0,3	0,5	11	34	65	2479
b20_1	32	66	2428	0,3	0,5	11	33	66	2439
b21	34	68	2630	0,3	0,5	11	34	68	2641
b21_1	34	70	2609	0,3	0,5	11	35	70	2620
b22	47	98	3078	0,5	0,7	17	48	99	3095
b22_1	48	99	3092	0,5	0,7	17	48	100	3109
Multiplier	75	154	5685	0,7	1,1	26	75	155	5712
Mem_ctrl	117	238	6221	1,1	1,6	29	118	239	6249
Adder	11	14	201	0,1	0,1	1	11	14	202
Square	53	98	4147	0,5	0,7	19	54	99	4166
BAR	11	14	468	0,1	0,1	2	11	14	470
AMBER	96	182	6087	0,9	1,2	28	97	183	6115

Tabela 5 – Resultados de dissipação de potência da síntese 2

Os resultados da síntese 2 são mostrados no diagrama apresentado pela Figura 16, para os circuitos *b20*, *Multiplier* e *AMBER*.

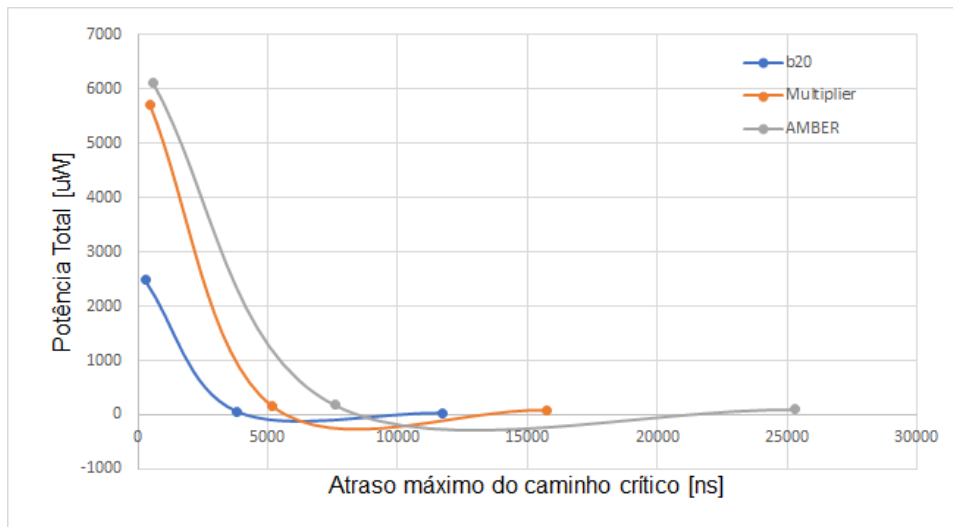


Figura 16 – Atraso do caminho crítico x Potência total para a síntese 2

Nessa síntese a redução na dissipação de potência foi em um fator de cerca de 64 e 33 para as tensões de 0,25V e 0,35V, enquanto que o aumento no atraso ficou com em um fator de 39 e 13 para ambas as tensões, respectivamente.

Ao comparar os diagramas mostrados pelas Figuras 15 e 16 e os valores apresentados pelas tabelas pode-se notar que, ao deixar com que a ferramenta de síntese escolha a melhor frequência para cada circuito, tem-se valores menores de atraso para a síntese 2, se comparado à síntese 1. Entretanto, esse comportamento leva a uma maior dissipação de

potência e, conseqüentemente, um a maior consumo de energia. Essa diferença é esperada, visto que a síntese 2 tem como objetivo a minimização do atraso. Além disso, os valores do eixo Y e o perfil mais acentuado das curvas mostrados pelo diagrama da Figura 15 comprovam a eficácia na minimização da potência dissipada alcançada pela síntese 1.

Na síntese 3, os circuitos foram sintetizados individualmente somente para a tensão padrão de $1,2V$. Após isso, esses mesmos circuitos foram aplicados nas outras tensões e os valores para área, atraso e potência foram aferidos. O objetivo dessa síntese é permitir uma comparação entre circuitos idênticos quando aplicados diferentes a valores de V_{DD} , pois todos possuem as mesmas configurações de portas lógicas. Da mesma forma como na síntese 1, a frequência para todos os circuitos foi setada em $1MHz$. Esses parâmetros foram setados dessa forma para que todos os circuitos tenham um mesmo desempenho e que se tenha uma minimização da potência dissipada.

A Tabela 6 e a Tabela 7 mostram os resultados. Na Figura 17, tem-se o diagrama para os circuitos *b20*, *Multiplier* e *AMBER*.

V_{DD}	Área [mm^2]			Atraso [ns]		
	0,25	0,35	1,2	0,25	0,35	1,2
b18	404	393	351	29350	9689	332
b20	76	71	65	31467	10387	354
b20_1	76	74	67	32042	10584	361
b21	80	75	69	25923	8429	365
b21_1	83	77	71	32500	10736	366
b22	117	110	99	32019	10570	361
b22_1	121	113	101	32613	10773	367
Multiplier	142	144	117	41208	13796	557
Mem_ctrl	220	215	168	23299	7338	228
Adder	6	6	6	28094	9468	431
Square	114	111	97	43430	15401	431
BAR	17	17	13	2624	1119	51
AMBER	254	251	190	66322	20254	710

Tabela 6 – Resultados de área e atraso para a síntese 3

	Dissipação de Potência [μW]								
	Dinâmica			Vazamento			Total		
	0,25	0,35	1,2	0,25	0,35	1,2	0,25	0,35	1,20
V_{DD}									
b18	43	88	1317	8	12	68	51	100	1385
b20	2	4	53	1	2	11	3	6	63
b20_1	2	4	53	1	2	11	3	6	63
b21	2	4	51	1	2	11	3	6	62
b21_1	2	4	52	1	2	11	3	6	63
b22	3	5	65	2	3	16	4	8	82
b22_1	3	5	67	2	3	17	4	8	84
Multiplier	6	13	181	3	4	26	9	17	207
Mem_ctrl	9	20	198	4	7	29	13	26	226
Adder	0,8	1	6	0,4	0,4	0,9	1	2	7
Square	4	8	132	2	3	19	6	11	151
BAR	1	1	15	0	0	2	1	2	17
AMBER	7	15	193	4	5	28	11	20	221

Tabela 7 – Resultados de dissipação de potência da síntese 3

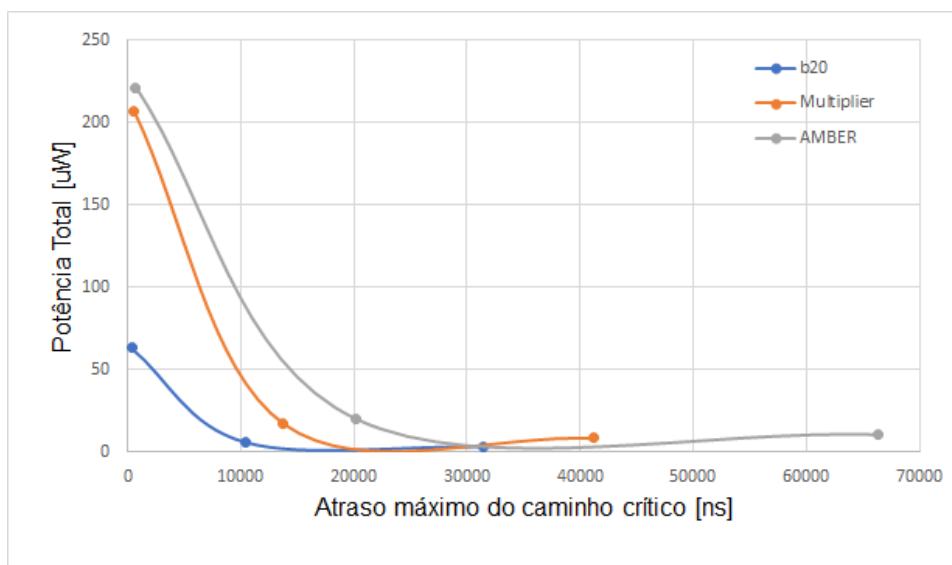


Figura 17 – Atraso do caminho crítico x Potência total para a síntese 3

Assim como nas simulações anteriores, houve uma redução na dissipação de potência nos circuitos sintetizados com as bibliotecas de ultra-baixo consumo, geradas com o auxílio do fluxo proposto. A redução para a tensão de 0,25V foi em um fator de 19 e para a tensão de 0,35V em um fator de 11. O aumento no atraso foi em um fator 84 e 28, respectivamente.

Para a síntese 4 o mesmo procedimento com relação à síntese 3 foi adotado. Da mesma forma como a síntese 2, a frequência do *clock* não foi predeterminada, deixando para a ferramenta de síntese realizar a escolha dos valores adequados. Assim como na síntese 2, o objetivo é a minimização do atraso total.

Os resultados estão mostrados na Tabela 8 e na Tabela 9, enquanto que o diagrama da síntese 4 é mostrado na Figura 18.

	Área [mm^2]			Atraso [ns]		
	0,25	0,35	1,2	0,25	0,35	1,2
V_{DD}						
b18	404	393	351	29346	9689	296
b20	76	71	65	31462	10387	287
b20_1	76	74	67	32042	10584	291
b21	80	75	69	25923	8429	300
b21_1	83	77	71	32500	10736	295
b22	117	110	99	32019	10570	300
b22_1	121	113	101	32613	10773	292
Multiplier	142	144	117	41205	13796	463
Mem_ctrl	220	215	168	23298	7338	189
Adder	6	6	6	28093	9468	358
Square	114	111	97	43427	15401	358
BAR	17	17	13	2624	1119	42
AMBER	254	251	190	66319	20254	589

Tabela 8 – Resultados de área e atraso para a síntese 4

	Dissipação de Potência [μW]								
	Dinâmica			Vazamento			Total		
	0,25	0,35	1,2	0,25	0,35	1,2	0,25	0,35	1,20
V_{DD}									
b18	538	1102	15329	8	12	66	546	1114	15263
b20	86	173	2469	1	2	11	87	178	2479
b20_1	85	173	2428	1	2	11	86	175	2439
b21	95	194	2630	1	2	11	96	196	2641
b21_1	92	189	2609	1	2	11	94	191	2620
b22	104	213	3078	2	3	17	106	213	3095
b22_1	106	218	3092	2	3	17	108	221	3109
Multiplier	188	389	5690	3	4	26	191	393	5716
Mem_ctrl	296	601	6225	4	7	29	300	608	6254
Adder	27	35	201	0,4	0,4	0,9	27	36	202
Square	134	248	4150	2	3	19	136	250	4169
BAR	27	35	469	0,4	0,4	2	27	36	471
AMBER	242	460	6091	4	5	28	246	465	6119

Tabela 9 – Resultados de dissipação de potência da síntese 4

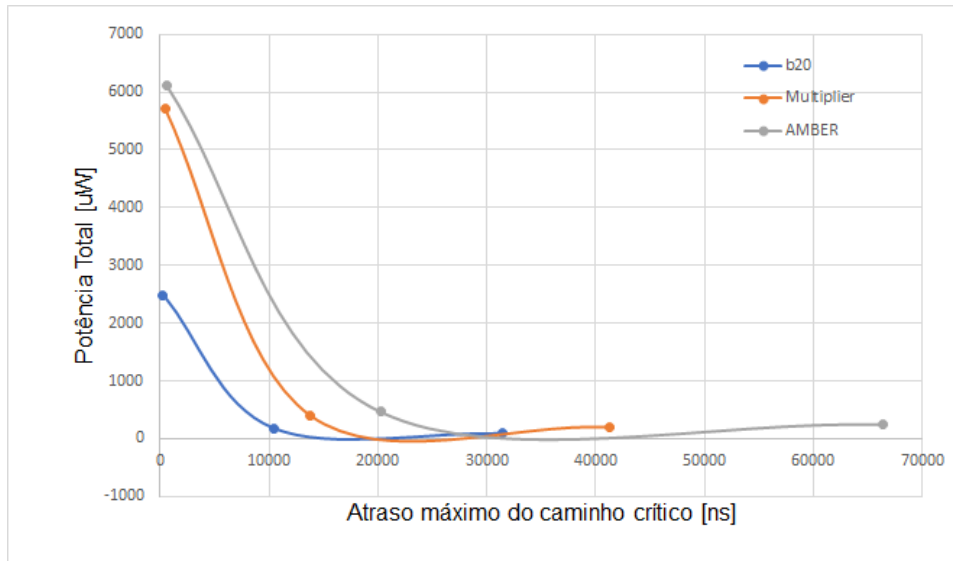


Figura 18 – Atraso do caminho crítico x Potência total para a síntese 4

Analisando os dados mostrados nas tabelas e o diagramas, tem-se que a redução na potência foi em um fator de 25 e 13 para as tensões de 0,25V e 0,35V respectivamente. O aumento no atraso foi similar ao encontrado na síntese 3, mesmo deixando com que a ferramenta faça a escolha da frequência a ser utilizada pelos circuito. Esse comportamento é justificado pelo fato de não haver uma nova síntese dos circuitos e, dessa forma, não há uma reconfiguração das portas lógicas de forma otimizada para a frequência determinada. Além disso, como todos os circuitos possuem as mesmas células, o caminho crítico é idêntico para todas as tensões analisadas, levando à valores de atraso semelhantes.

A Tabela 10 mostra os resultados do consumo de energia médio por instância, considerando todos os circuitos de teste e o microprocessador AMBER, para todas as sínteses.

Consumo Médio de Energia [fJ]			
	0,25	0,35	1,20
Síntese 1	0,015	0,027	0,698
Síntese 2	0,234	0,442	14,7
Síntese 3	0,032	0,062	0,698
Síntese 4	0,588	1,11	14,7

Tabela 10 – Consumo médio de energia por célula para a síntese 3

Pode-se comparar os resultados encontrados com a síntese dos circuitos com os valores encontrados nas simulações realizadas durante o primeiro passo do fluxo, mostrado na seção 4.1, e que são resumidos pela Figura 14. Como visto anteriormente, a simulação utiliza as mesmas portas lógicas independente da tensão que está sendo avaliada. Além disso, a frequência utilizada nessas simulações foi predeterminada antes da realização das

mesmas. Dessa forma, para se utilizar um ambiente similar de comparação, essa deve ser realizada usando os dados da síntese 3.

Como mostrado pelos dados apresentados na Tabela 10, tem-se uma redução no consumo de energia em um fator de aproximadamente 20 para a tensão de 0,25V e em um fator de 11 para a tensão de 0,35V. Ao comparar esses valores com aqueles apresentados pela Figura 14, tem-se diferenças muito próximas, o que mostra que os resultados apresentados pelos testes práticos, são condizentes com aqueles obtidos através de simulações.

Os testes com a memória SRAM foram conduzidos de maneira distinta. Nesse tipo de célula, a avaliação da mesma deve ser em relação às operações de escrita e leitura. Para essa simulação, foi criada uma configuração de 8x8 células de memória, mostrada na Figura 19. Nesse formato, o WWL das células são compartilhados entre todas as memórias das linhas. Além disso, as linhas de bit são compartilhadas com todas as células das colunas.

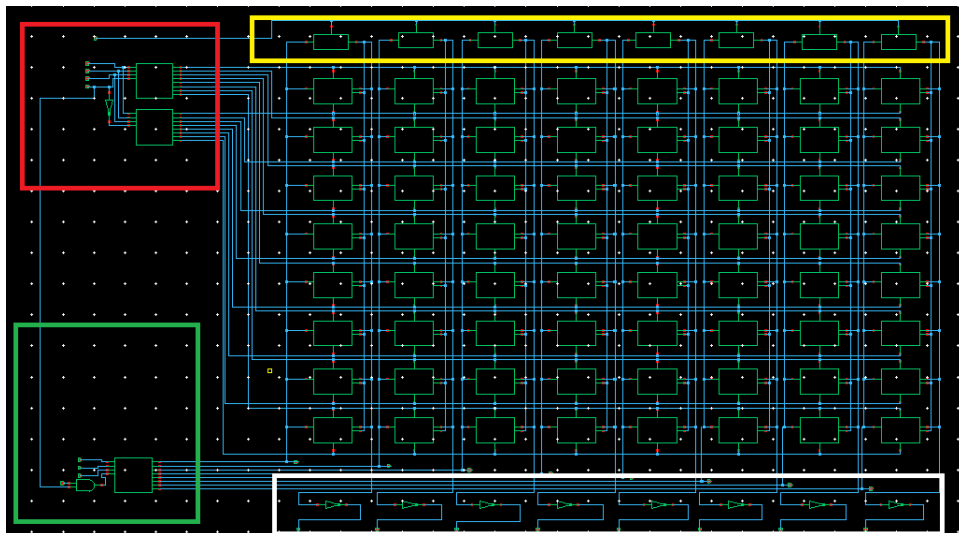


Figura 19 – Estrutura de testes da célula de memória SRAM

Nessa Figura, estão mostradas as memórias e os dispositivos adicionais que garantem o correto funcionamento nas operações de leitura e escrita. Esses circuitos foram construídos utilizando as células combinacionais presentes na biblioteca gerada. No retângulo vermelho estão mostrados os decodificadores que selecionam o endereço da linha que será usada, além de alternar entre as operações de leitura e escrita. Na operação de escrita, o decodificador de dados, mostrado pelo retângulo verde, seleciona em qual coluna ele irá disponibilizar o dado a ser armazenado. Assim, com a combinação dos dois decodificadores, pode-se selecionar qualquer célula para fazer a escrita. Para a operação de escrita, a linha de bit de leitura deve ser levada, inicialmente, para o nível do V_{DD} . Essa operação é realizada pelos circuitos de *pre-charge*, mostrados pelo retângulo amarelo. Em seguida, a operação de leitura deve ser selecionada e um pulso na linha na qual se encontra a célula a ser lida

deve ser aplicado, utilizando o decodificador de endereços. Por fim, o dado a ser lido ficará disponível na saída de coluna correspondente à célula que está sendo lida, mostrada pelo retângulo branco.

Os resultados para o atraso e o consumo de energia para as memórias SRAM são mostradas na Tabela 11.

SRAM	Atraso [ns]			Consumo [pJ]		
	0,25	0,35	1,2	0,25	0,35	1,2
Escrita	12.440	2.426	0.115	4.138	6.795	86.110
Leitura	51.840	23.650	0.223	9.981	13.590	111.900

Tabela 11 – Resultados para a memória SRAM

Ao analisar o estado-da-arte, nota-se que os trabalhos apresentados por [Gao et al. \(2015\)](#) e [Li et al. \(2013\)](#) conseguiram uma redução média de cerca de 97%, enquanto que [Vohrmann et al. \(2015\)](#) conseguiu uma redução de aproximadamente 92% na dissipação de potência. Ao comparar esses valores com aqueles apresentados nesse trabalho, percebe-se que a simples aplicação do fluxo já consegue resultados semelhantes aos apresentados no estudos mais recentes. Além disso, como visto anteriormente, o fluxo pode ser utilizado em conjunto com outras técnicas, o que quer dizer que podem ser obtidas maiores reduções no consumo de energia ou valores de atraso menores.

Para se fazer a verificação das células geradas para a biblioteca, foi aplicado um fluxo de EDA para o processador AMBER23. Após a síntese e o posicionamento das células, foi realizado testes de DRC para garantir que nenhuma regra de design tenha sido violada, sendo que todos os testes passaram sem nenhum erro. Na Figura 20 o resultado do fluxo de EDA é mostrado.

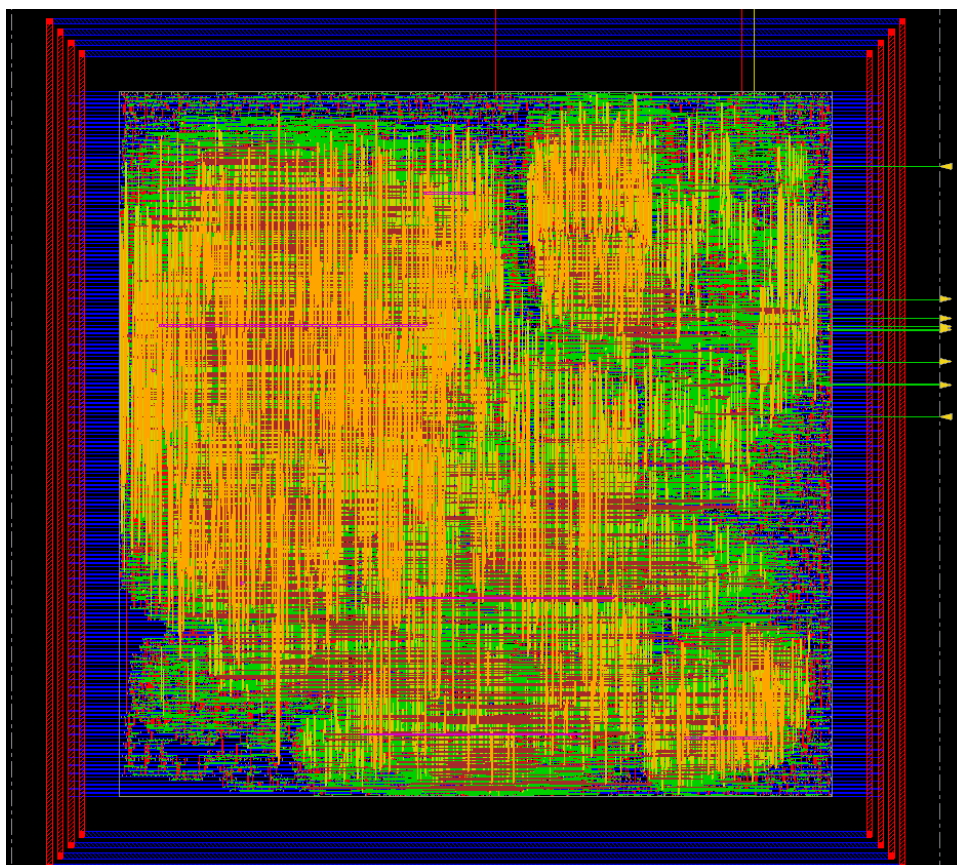


Figura 20 – AMBER23 após fluxo de EDA

5 Conclusão e trabalhos futuros

Diferentes caminhos foram traçados com os recentes avanços no campo da microeletrônica. Dentre eles, destacam-se as pesquisas relacionadas à aplicações no campo de Internet-das-coisas. Nesses sistemas, o consumo de energia é o ator principal, em detrimento da performance geral do sistema. Nesse contexto, surge a necessidade de se desenvolver sistemas que ofereçam essa troca de maneira adequada e otimizada. Em sistemas digitais, é necessária a utilização de bibliotecas de células padrão específicas e que sejam capazes de operar na região sub-limiar.

Inicialmente, foi realizada uma revisão de todos os conceitos e parâmetros que cercam o transistor MOS. Foram discutidas os princípios de funcionamento e as características ideais e não-ideais existentes. Dentro desse contexto, as regiões de operação do transistor foram apresentadas, com ênfase na região sub-limiar, apresentando as vantagens e desvantagens de se trabalhar em cada uma delas. Esse conceito foi ampliado para mostrar como é a lógica CMOS e seus principais componentes. Os parâmetros de potência, consumo de energia e atraso em circuitos integrados foram apresentados e sua importância como métrica para a análise de sistemas microeletrônicos também foi discutida. Por fim, a metodologia de trabalho em bibliotecas de células padrão foi mostrada, além de uma breve discussão sobre memórias CMOS, apresentando um modelo próprio para operação na região sub-limiar.

Tendo esses conceitos em mente, esse trabalho apresentou um fluxo completo para a geração de bibliotecas de células padrão que operem na região sub-limiar. Ele leva em conta as características da tecnologia escolhida, assim como as diferentes técnicas de dimensionamento já existentes na literatura. Além disso, o fluxo também pode ser usado como suporte para os trabalhos de projetistas e pesquisadores que atuam nesse campo.

Assim, duas bibliotecas de células para operação em região sub-limiar foram geradas e caracterizadas com o auxílio do fluxo proposto, para diferentes tensões de alimentação. As células resultantes foram utilizadas para a síntese em diversos circuitos de teste e para um processador AMBER23.

Os resultados encontrados mostraram que a operação na região sub-limiar consegue habilitar um consumo de energia cerca de 96% melhor para a tensão de 0,25V e cerca de 93% melhor com a tensão de 0,35V, quando comparados a uma biblioteca operando com tensão padrão. Entretanto, essa melhora vem ao custo de uma perda de duas a três ordens de grandeza na performance geral do sistema. Além disso, comparações realizadas com o estado-da-arte mostram que a aplicação do fluxo consegue resultados semelhantes ou melhores do que os mostrados por esses trabalhos. É importante notar também que o fluxo

proposto pode ser usado em conjunto com outras técnicas, o que pode gerar resultados ainda mais expressivos.

5.1 Trabalhos Futuros

A aplicação do fluxo proposto conseguiu resultados que se assemelham aos encontrados pelo estado-da-arte. Entretanto, os resultados encontrados ainda podem ser otimizados, principalmente no que se refere ao atraso.

Como dito anteriormente, o dimensionamento utilizado para as células que compõem a biblioteca foi realizado de forma manual, utilizando o modelo de atraso RC. Dessa forma, novas formas de dimensionamento podem ser realizadas, buscando melhores resultados, tendo em vista que o fluxo apresenta apenas os melhores parâmetros iniciais. A aplicação de algumas de técnicas de otimização multiobjetivo para o dimensionamento ([VOHRMANN et al., 2015](#)) ou a exploração adicional do efeito RSCE no dimensionamento ([KIM et al., 2007](#)) podem auxiliar para a melhora dos resultados.

Uma outra proposta é a fabricação de alguns circuitos ou de um processador AMBER com a utilização da biblioteca gerada. A caracterização e testes desses sistemas pode fornecer dados reais do desempenho do fluxo e, através desses testes, novas soluções podem ser apresentadas e inseridas no fluxo proposto.

Referências

- AGARWAL, T. *Difference Between CMOS and NMOS Technology*. [S.l.], 2005. Disponível em: <<https://www.elprocus.com/difference-between-nmos-cmos-technology/>>. Acesso em: 8 out. 2018. Citado 2 vezes nas páginas 7 e 22.
- AMARU, L.; GAILLARDON, P.; MICHELI, G. D. The epfl combinational benchmark suite. *Int. Workshop on Logic Synth*, 2015. Citado na página 57.
- AMBER, C. Amber open source projec. *Technical report*, 2011. Citado na página 57.
- BARRON, M. B. Low level currents in insulated gate field effect transistors. *Solid-State Electronics*, v. 15, p. 293–302, 1972. Citado na página 16.
- BIAGIONI, E.; BRIDGES, K. The application of remote sensor technology to assist the recovery of rare and endangered species. *Special Issue on Distributed Sensor Networks for the International Journal of High Performance Computing Applications*, v. 16, p. 315–324, 2002. Citado na página 16.
- BRITO, A. *Introdução a arquitetura de computadores*. 1. ed. Paraíba: Editora da Universidade Federal da Paraíba, 2014. Citado 2 vezes nas páginas 7 e 39.
- BUTZEN, P.; RIBAS, R. Leakage current in sub-micrometer cmos gates. *Universidade Federal do Rio Grande do Sul*, 2007. Citado 3 vezes nas páginas 26, 35 e 37.
- CHANDRAKASAN, A.; SHENG, S.; BRODERSEN, R. Low-power cmos digital design. *IEEE Journal of Solid State Circuits*, v. 27, p. 473–484, 1992. Citado na página 36.
- CHEN, K. et al. Mosfet carrier mobility model based on gate oxide thickness, threshold and gate voltages. *Solid-State Electronics*, v. 39, p. 1515–1518, 1996. Citado na página 27.
- CHINTALAPUDI, K.; JOHNSON, E.; GOVINDAN, R. Structural damage detection using wireless sensor-actuator networks. *Proceedings of the IEEE International Symposium on Intelligent Control*, p. 322–327, 2005. Citado na página 16.
- COBBOLD, R. *Theory and Application of Field Transistor*. New York: Wiley Interscience, 1970. Citado na página 23.
- COLOMER-FARRARONS, J. et al. A multiharvested self-powered system in a low-voltage lowpower technology. *IEEE Transactions on Industrial Electronics*, v. 58, p. 4250–4263, 2011. Citado na página 15.
- CORNO, F.; RECORDA, M. S.; SQUILLERO, G. Rt-level itc'99 benchmarks and first atpg results. *IEEE Design and Test of Computers*, 2005. Citado na página 57.
- C.SAH. Characteristics of the metal-oxide-semiconductors transistors. *IEEE trans. Electron Devices*, v. 11, p. 324–345, 1964. Citado na página 23.
- C.SAINT; SAINT, J. *IC mask design - Essential Layout Techniques*. 2. ed. New York: McGraw-Hill, 2002. Citado na página 38.

- GAO, F. et al. Development of tfet 0.13 um standard cell library for ultra-low power applications. *IEEE 11th International Conference on ASIC*, 2015. Citado 4 vezes nas páginas 17, 41, 43 e 66.
- HANSON, S. et al. Ultralow-voltage, minimum-energy cmos. *IBM Journal of Research and Development*, p. 469–490, 2006. Citado na página 35.
- HARRIS, D. et al. The fanout-of-4 inverter delay metric. *Stanford University*, 2015. Citado na página 45.
- HU, C. C. *Modern Semiconductor Devices for Integrated Circuits*. [S.l.]: University of California, 2010. Citado 4 vezes nas páginas 7, 20, 21 e 27.
- IEONG, C. et al. Standard cell library design with voltage scaling and transistor sizing for ultra-low-power biomedical applications. *IEEE International Conference of Electron Devices and Solid-state Circuits*, 2013. Citado 3 vezes nas páginas 17, 41 e 43.
- JACOBONI, C. et al. A review of some charge transport properties of silicon. *Solid-State Electronics*, v. 20, p. 77–89, 1977. Citado 2 vezes nas páginas 7 e 28.
- KANITAR, H. Subthreshold circuits: Design, implementation and application. *A thesis in Kate Gleason College of Engineering Rochester Institute of Technology Rochester*, 2009. Citado 3 vezes nas páginas 7, 33 e 34.
- KIM, T. et al. Utilizing reverse short-channel effect for optimal subthreshold circuit design. *IEEE Transactions on VLSI Systems*, v. 15, p. 821–829, 2007. Citado 7 vezes nas páginas 17, 29, 30, 42, 43, 47 e 70.
- KIM, T. et al. A 0.2 v, 480 kb subthreshold sram with 1 k cells per bitline for ultra-low-voltage computing. *IEEE Journal of Solid-State Circuits*, v. 43, p. 518–529, 2008. Citado 4 vezes nas páginas 7, 40, 41 e 42.
- KULKARNI, B. K. Asic back-end design. 2012. Citado 2 vezes nas páginas 7 e 38.
- LI, M. et al. Sub-threshold standard cell library design for ultra-low power biomedical applications. *Annual International Conference of the IEEE EMBS*, 2013. Citado na página 66.
- LU, C. Y.; SUNG, J. M. Reverse short-channel effects on threshold voltage in submicrometer salicide devices. *IEEE Electron Device Lett*, v. 10, p. 446–448, 1989. Citado na página 29.
- MASUHARA, T.; ETOH, J.; NAGATA, M. A precise mosfet model for low- voltage circuits. *IEEE Trans. Electron Devices*, v. 21, p. 363–371, 1974. Citado na página 16.
- NICOLLIAN, E.; BREWS, J. *MOS (Metal Oxide Semiconductor) Technology*. New York: John Wiley and Sons, 1982. Citado na página 20.
- NING, M. Column decoder using ptl for memory. ” *IOSR Journal of Electronics and Communication Engineering*, 2013. Citado na página 42.
- OLIVEIRA, C. R. F. Desenvolvimento de um fluxo de caracterização de células padrão aplicável a bibliotecas de consumo ultra baixo (ulp). *Monografia de TCC apresentada à Universidade Federal de Minas Gerais*, 2017. Citado 2 vezes nas páginas 37 e 38.

- PAUL, B. C. Device optimization for digital subthreshold logic operation. *IEEE Trans. Electron Devices*, v. 52, p. 237–247, 2005. Citado na página 30.
- POPOVIC, Z. Low-power far-field wireless powering for wireless sensors. *Proceedings of the IEEE*, v. 101, p. 1397–1409, 2013. Citado na página 15.
- RABAEY, J.; CHANDRAKASAN, A.; NIKOLIC, B. *Digital Integrated Circuits - A Design Perspective*. 2. ed. New York: Prentice Hall, 2002. Citado 2 vezes nas páginas 7 e 26.
- SCHWIEBERT, L.; GUPTA, S.; WEINMANN, J. Research challenges in wireless networks of biomedical sensors. *Mobile Computing and Networking*, p. 151–165, 2001. Citado na página 16.
- SHOCKLEY, W. A unipolar 'field-effect' transistor. *Proceedings of the IRE*, v. 40, p. 1365–1376, 1952. Citado na página 23.
- S.SEBASTINSURESH et al. Ultra-low power circuit techniques for miniaturized sensor nodes. *Doctoral Thesis in Electronic and Computer Systems*, 2013. Citado na página 40.
- STANGHERLIN, K. H. Energy and speed exploration in digital cmos circuits in the near-threshold regime for very-wide voltage-frequency scaling. *Universidade Federal do Rio Grande do Sul*, 2013. Citado 2 vezes nas páginas 32 e 47.
- SWART, J. W. *Oficina de Microfabricação: Projeto e Construção de CIs MOS*. Campinas: UNICAMP, 2002. Citado na página 21.
- TIMARCHI, S.; ALIOTO, M. Ultra-low voltage standard cell libraries: Design strategies and a case study. *IEEE International Conference on Electronics, Circuits and Systems*, p. 520–523, 2015. Citado na página 41.
- TOH, K.; KO, P.; MEYER, R. An engineering model for short-channel mos devices. *JSSC*, v. 23, p. 950–958, 1988. Citado na página 28.
- TROUTMAN, R. R. Subthreshold slope for insulated gate field-effect transistors. *IEEE Trans. Electron Devices*, v. 22, p. 1049–1051, 1975. Citado na página 16.
- TROUTMAN, R. R. Vlsi limitations from drain-induced barrier lowering. *IEEE Trans. Electron Devices*, v. 26, p. 461–469, 1979. Citado na página 29.
- TROUTMAN, R. R.; CHAKRAVARTI, S. N. Subthreshold characteristics of insulated-gate field-effect transistors. *IEEE Trans. Circuit Theory*, v. 20, p. 659–665, 1973. Citado na página 16.
- VITTOZ, E.; FELLRATH, J. Cmos analog integrated circuits based on weak inversion operation. *IEEE Journal of Solid-State Circuits*, v. 12, p. 224–231, 1977. Citado na página 16.
- VLSI, A. things. *NMOS and PMOS Operating Regions*. [S.l.], 2013. Disponível em: <<https://allthingsvlsi.wordpress.com/2013/04/04/nmos-and-pmos-operating-regions/>>. Acesso em: 23 out. 2017. Citado 2 vezes nas páginas 7 e 25.
- VOHRMANN, M. et al. A 65 nm standard cell library for ultra low-power applications. *2015 European Conference on Circuit Theory and Design (ECCTD)*, 2015. Citado 4 vezes nas páginas 17, 41, 66 e 70.

- WANG, A.; CALHOUN, B.; CHANDRAKASAN, A. *Sub-threshold Design for Ultra Low-Power Systems*. 2. ed. Cambridge, MA: Springer, 2006. Citado na página 16.
- WEINSTEIN, R. Rfid: a technical overview and its application to the enterprise. *IT Professional*, v. 7, p. 27–33, 2005. Citado na página 17.
- WESTE, N. H. E.; HARRIS, D. M. *CMOS VLSI Design - A Circuits and Systems Perspective*. 4. ed. Massachusetts: Addison Wesley, 2011. Citado 10 vezes nas páginas 7, 22, 23, 24, 27, 32, 33, 35, 36 e 37.
- XINFU, L. et al. A study of inverse narrow width effect of 65nm low power cmos technology. *Proc. 9th Int. Conf. Solid-State and Integrated-Circuit Technology*, p. 1138–1141, 2008. Citado na página 42.
- ZHOU, J. et al. A 40 nm dual-width standard cell library for near/sub-threshold operation. *IEEE Transactions On Circuits And Systems*, 2014. Citado 2 vezes nas páginas 42 e 43.