



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
UNIDADE ACADÊMICA DE ENGENHARIA ELÉTRICA

Trabalho de Conclusão de Curso

Elaboração de Módulos para o Ensino de Microeletrônica

Aluno

Antonio Agripino da Costa Filho

Orientador

Prof. D.Sc. Raimundo Carlos Silvério Freire

Campina Grande, Fevereiro de 2010

ANTONIO AGRIPINO DA COSTA FILHO

Elaboração de Módulos para o Ensino de Microeletrônica

Trabalho de conclusão de curso apresentado à
Universidade Federal de Campina Grande como
requisito parcial para a obtenção do título de
Engenheiro Eletricista.

Orientador: Prof. D.Sc. Raimundo Carlos Silvério Freire

Campina Grande, Fevereiro de 2010

C837e Costa Filho, Antonio Agripino da.
Elaboração de módulos para o ensino de
Microeletrônica. / Antonio Agripino da Costa Filho. -
Campina Grande - PB: [s.n], 2010.

111f.

Orientador: Professor Dr. Raimundo Carlos Silvério
Freire.

Trabalho de Conclusão de Curso - Monografia; (Curso
de Bacharelado em Engenharia Elétrica) - Universidade
Federal de Campina Grande; Centro de Engenharia Elétrica
e Informática.

1. Microeletrônica. 2. Ensino de Microeletrônica. 3.
História da Microeletrônica. 4. Layout - Microeletrônica.
5. Fotolitografia. 6. Transistor MOS. 7. Semicondutores.
8. Junção PN. 9. Elaboração e material didático -
Microeletrônica. 10. Módulos de ensino - Microeletrônica
I. Freire, Raimundo Carlos Silvério. II. Título.

CDU:621.3.049.77(043.1)

Elaboração da Ficha Catalográfica:

Johnny Rodrigues Barbosa
Bibliotecário-Documentalista
CRB-15/626

Dedicatória

A todos meus amigos e familiares que me apoiaram no decorrer do curso para que eu pudesse superar várias barreiras.

Sumário

1	Introdução	4
2	A História da Microeletrônica	5
3	A Microeletrônica no Brasil	9
4	Formulação dos Módulos.....	11
5	Conclusões	12
6	Bibliografia	13
7	Anexos	14

1 Introdução

A microeletrônica é um ramo da eletrônica voltado à integração de circuitos eletrônicos, promovendo uma miniaturização dos componentes. A área engloba tanto os processos físico-químicos de fabricação dos circuitos integrados como o projeto do circuito em si.

Os componentes utilizados na microeletrônica são construídos na escala de micrômetros ou mesmo nanômetros. A redução no tamanho dos componentes utilizados vem, ao longo da história, seguindo a Lei de Moore.

A microeletrônica vem assumindo crescente importância no mundo atual, estando presente na informática, nas telecomunicações, nos controles de processos industriais, nos bens de consumo, enfim, numa vasta gama de aplicações de uso residencial ou industrial.

Tendo em vista isso, e o crescimento que a microeletrônica vem tendo no Brasil, propôs-se a elaboração de módulos que tratam da física básica dos semicondutores até a parte do leiaute, para que o aluno interessado em ter seu primeiro contato com esse mundo tenha um material no qual veja os conceitos importantes para tal.

2 A História da Microeletrônica

Em 1958, Jack Kilby construiu o primeiro circuito de flip-flop integrado com dois transistores na *Texas Instruments*. Em 2003, o microprocessador *Intel Pentium 4*, Figura 1(b), contendo 55 milhões de transistores é construído. Isso corresponde a uma taxa de crescimento anual de 53% em 45 anos. Nenhuma outra tecnologia na história sofreu uma taxa de crescimento tão elevada por muito tempo (1).

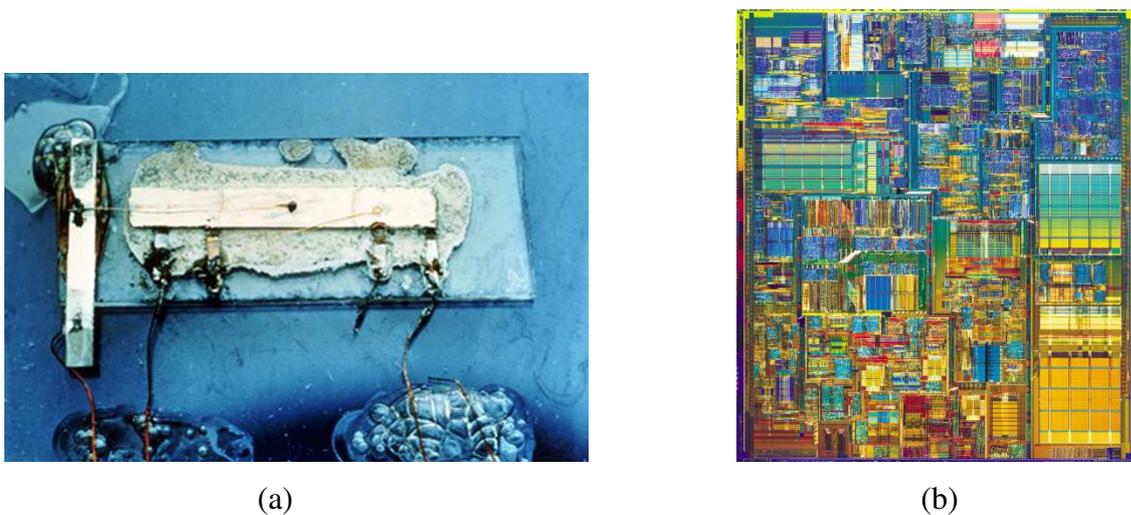


Figura 1 – (a) Primeiro circuito integrado (b) Circuito integrado do microprocessador *Intel Pentium 4*

Esse incrível crescimento advém da miniaturização constante dos transistores e aperfeiçoamentos nos processos de fabricação. Os transistores, além de estarem cada vez menores, têm-se tornados rápidos, dissipam menos potência e são baratos para fabricar. Essas ações em comum têm revolucionado não somente a eletrônica, mas também a sociedade como um todo.

A Figura 2 mostra as vendas anuais do mercado mundial de semicondutores (1).

Fatos anteriores contribuíram para a evolução da microeletrônica, mas o marco histórico nessa evolução foi o descobrimento do efeito de transistor bipolar em 1947 na *Bell Labs* por John Bardeen e Walter Brattain. O transistor, Figura 3, era construído por uma base de germânio tipo N e duas junções de contato tipo P na superfície, sendo um de emissor e outro de coletor, feitos um próximo ao outro.

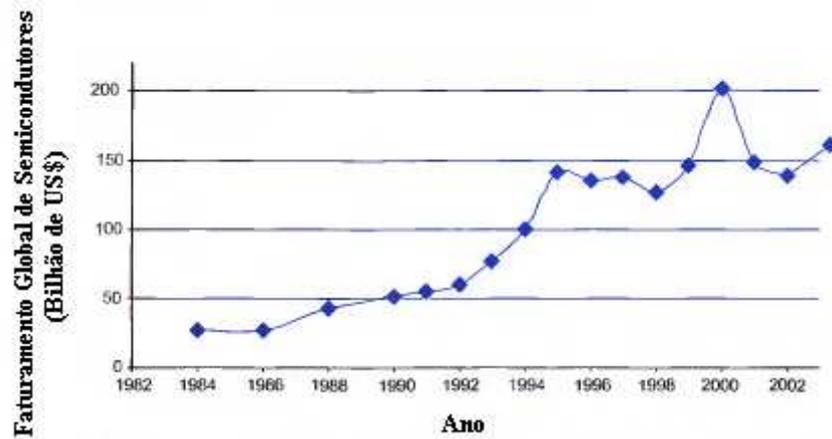


Figura 2 – Vendas anuais do mercado mundial de semicondutores

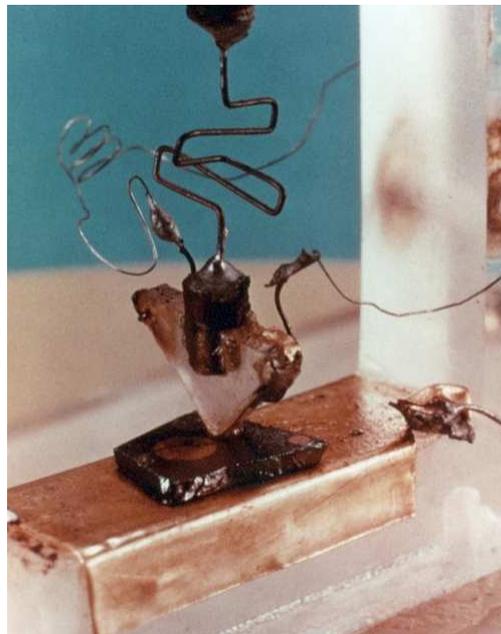


Figura 3 – Primeiro transistor bipolar de contato

Inicialmente, essa descoberta foi classificada como um segredo militar, mas a *Bell Labs* anunciou publicamente a descoberta no ano seguinte.

Onze anos depois, Jack Kilby na *Texas Instruments* realizou a possibilidade, até então, de se múltiplos transistores poderiam ser construídos em uma única peça de silício. A Figura 1(a) mostra seu primeiro protótipo de um circuito integrado, construído de um único bloco de silício contendo um transistor, um capacitor e um resistor (2).

Os primeiros circuitos integrados usavam o transistor bipolar de junção. Esses transistores podem ser vistos como chaves controladas eletricamente com um terminal de controle e dois outros terminais que são ligados ou não dependendo da tensão

aplicada ao terminal de controle. Eles requerem uma pequena corrente no terminal (base) de controle para chavear correntes muito maiores entre os outros dois terminais (emissor e coletor).

Os transistores de efeito de campo metal/óxido semiconductor (MOSFET) surgiram com a atrativa vantagem de que eles drenam uma corrente de controle quase nula enquanto inativo. A ideia original do transistor MOSFET é datada ao cientista germânico Julius Lilienfeld em 1925; uma estrutura proxicamente parecida foi proposta em 1935 por Oskar Heil, mas problemas de materiais frustraram tentativas posteriores de se fazer tais dispositivos funcionarem. Foi graça ao estudo e desenvolvimento dos processos de oxidação do silício que permitiram o desenvolvimento do transistor MOSFET. Em 1960, D. Kahng e M. Atalla, da *Bell Labs*, demonstraram o transistor MOSFET. Porém, problemas com a falta de controle de contaminação de impurezas fizeram com que tais dispositivos apresentassem uma estabilidade pobre, causando um atraso de mais de 10 anos para seu uso em grande escala.

Em 1963, Frank Wanlass, da *Fairchild*, propôs a combinação de transistores MOSFET tipo N e P num mesmo substrato, conhecido como tecnologia CMOS. Ele construiu portas lógicas que consumiam somente alguns nanowatts de potência, uma magnitude de seis vezes menos que os transistores bipolares de junção.

Gordon Moore, em 1965, observou que se esboçasse o número de transistores que podem ser fabricados mais economicamente em um circuito integrado dava uma linha reta em uma escala semilogarítmica. Na época, ele notou que a quantidade de transistores dobrava a cada dezoito meses. Essa observação é conhecida como *Lei de Moore* e tornou-se uma profecia que sempre se cumpre. Em 1975, Moore revisou a sua previsão para cada dois anos. A Figura 4 mostra que o número de transistores nos microprocessadores da *Intel* tem dobrado a cada vinte e seis meses desde a invenção do microprocessador 4004.

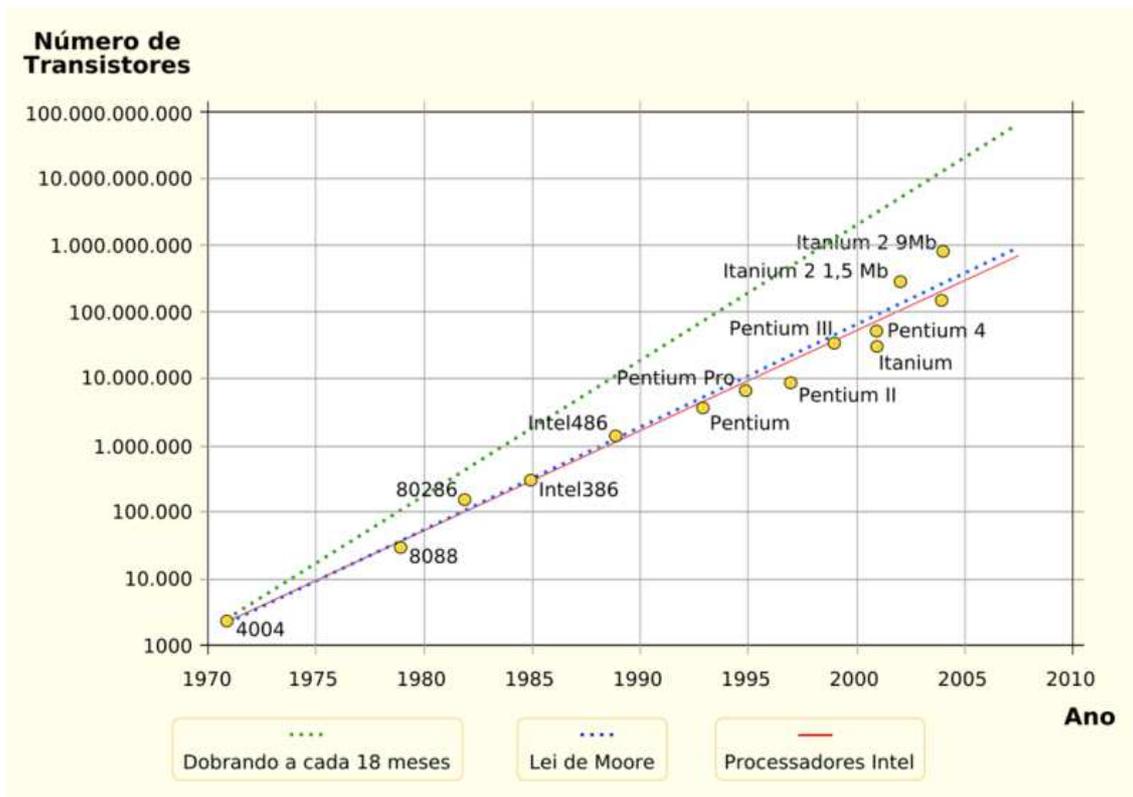


Figura 4 – Observação da Lei de Moore nos processadores da Intel

3 A Microeletrônica no Brasil

O Brasil foi um dos primeiros países em desenvolvimento a utilizar e produzir equipamentos eletrônicos. Nos anos 80 havia uma significativa produção local de computadores e periféricos, além de uma indústria de componentes microeletrônicos. Entretanto, diversos fatores de ordem econômica, política e tecnológica, tanto no âmbito interno quanto internacional, levaram a uma perda relativa de capacitação industrial local, principalmente em semicondutores. Das 23 empresas produtoras de semicondutores existentes no Brasil ao final dos anos 80, restaram apenas quatro (Itaucom, Aegis, Semikron e Heliodinâmica) atuando em etapas mais simples do processo produtivo. A consequência econômica imediata da perda de capacidade industrial é o crescente peso negativo da microeletrônica na balança comercial do país, em uma época de difícil acesso a créditos externos (3).

Atualmente, o Ministério da Ciência e Tecnologia tem investido na microeletrônica através de três programas: CI-Brasil, PADIS e PDP.

O PADIS (Programa de Apoio ao Desenvolvimento Tecnológico da Indústria de Semicondutores) é determinado pela Lei nº 11.484 de 31 de maio de 2007 (4). O PADIS abrange incentivos restritos às empresas que atuem, exclusivamente, na área dos dispositivos semicondutores e *displays* do tipo LCD, PDP (*plasma display panel*), LED, OLED e TFEL (*Thin Film Electroluminescent*) (exceto cinescópios) visando investimentos para atividades de:

- Concepção, desenvolvimento e projeto;
- Difusão ou processamento físico-químico;
- Encapsulamento e testes.

O PDP (Política de Desenvolvimento Produtivo) tem como objetivo ampliar a produção local e exportações de componentes microeletrônicos, tendo como meta para alcançar tal objetivo o de construir duas empresas de circuitos integrados envolvendo a etapa de *front-end* e elevar o número de *Design Houses* do programa CI-Brasil para fortalecer sua atuação.

Por fim, o programa CI-Brasil é baseado em um conjunto de três ações principais que visam:

- Incentivar a atividade econômica na área de projeto de CIs;
- Expandir a formação de projetistas de circuitos integrados;
- Promover a criação de uma indústria nacional de semicondutores.

Uma das ações do CI-Brasil é o Programa Nacional de Formação de Projetistas de Circuitos Integrados, que tem por objetivo ampliar a formação no País de projetistas de circuitos integrados de nível internacional, e por meta a graduação de 1.000 projetistas em três anos (5). Essa ação é implementada através de uma rede de quatro Centros de Treinamento (CT) distribuídos no país. Atualmente esta rede conta com os seguintes centros em operação:

- Centro de Treinamento 1 (CT1)
 - NSCAD Microeletrônica
Centro de Empreendimentos em Informática (CEI)
Instituto de Informática – UFRGS
Porto Alegre, Rio Grande do Sul
- Centro de Treinamento 2 (CT2)
 - Centro de Tecnologia da Informação Renato Archer (CTI)
Campinas, São Paulo

4 Formulação dos Módulos

A disciplina “Estrutura e Concepção de Circuitos Integrados” é uma disciplina de conteúdo profissional específico da ênfase de eletrônica do curso de engenharia elétrica da UFCG. Possui uma carga horária de 60 horas e apresenta como pré-requisitos as seguintes disciplinas: Eletrônica e Laboratório de Eletrônica.

A ementa da disciplina é a seguinte: Introdução à Microeletrônica. Elementos de física de semicondutores, de tecnologia e modelagem de transistores. Ferramentas matemáticas. Componentes fundamentais. Amplificação. Técnicas de capacitores comutados. Técnica de modo corrente. Introdução à filtragem. Introdução à conversão analógico-digital e digital-analógico. Ruído nos circuitos integrados. Simulação. Teste de circuitos integrados. Introdução ao CAD.

Baseado nos temas acima sublinhados e no primeiro contato que o aluno terá com o projeto de leiaute de um circuito integrado, três módulos foram desenvolvidos para apresentar os conceitos necessários ao entendimento dos semicondutores, dos transistores MOS (há outros transistores que são usados no desenvolvimento de circuitos integrados, mas esses são os mais largamente utilizados), da fabricação de um CI e das técnicas usadas para criar um leiaute de um circuito impresso.

Os módulos são os seguintes:

- Semicondutores e a Junção PN;
- Transistor MOS;
- Projeto de Leiaute.

O primeiro módulo trata do conceito de semicondutores e o estudo do movimento das cargas neles. Trata-se também da junção PN, que serve como diodo, e da formulação do seu comportamento, já que em circuitos integrados fabricados por processo MOS existem diodos inerentes ao processo.

No segundo módulo são abordados os princípios de funcionamento e as curvas características do transistor MOS. Também é abordado o modelo BSIM, que descreve os parâmetros utilizados para a caracterização elétrica de dispositivos.

No terceiro e último módulo, o processo de fabricação, os conceitos das regras de leiaute e sua relação com os microcircuitos são enfatizados.

5 Conclusões

Neste trabalho foram apresentados módulos visando o auxílio no ensino da microeletrônica. Os módulos foram desenvolvidos para fornecer aos alunos que cursem a disciplina de “Estrutura e Concepção de Circuitos Integrados” um material que os auxiliem a compreender a física básica dos semicondutores e dos transistores MOS e a fabricação de um CI e das técnicas usadas para criar um leiaute de um circuito impresso.

Muitas outras informações podem ser acrescentadas aos módulos, tais como sobre as técnicas de leiaute para componentes passivos (resistores e capacitores) nos circuitos integrados, sobre as simulações dos circuitos, maiores detalhamentos da física dos semicondutores, etc. Melhorias na didática do texto também podem ajudar no desenvolvimento de módulos mais esclarecedores.

6 Bibliografia

1. **Weste, Neil e Harris, David.** *CMOS VLSI Design: A Circuits and Systems Perspective*. Boston : Addison Wesley, 2004. 0-321-26977-2.
2. **Swart, Jacobus W.** *Evolução de Microeletrônica a Micro-Sistemas*. s.l. : UNICAMP.
3. **Ministério da Ciência e Tecnologia.** *Programa Nacional de Microeletrônica: Contribuições para a formulação de um plano estruturado de ações*. Brasília : s.n., 2002.
4. **Miguel, Henrique de Oliveira.** Política para o Desenvolvimento Tecnológico e Industrial para o Setor de Semicondutores. *Instituto de Pesquisa e Desenvolvimento Tecnológico do Complexo Eletroeletrônico e Tecnologia da Informação*. [Online] [Citado em: 22 de Fevereiro de 2010.] <http://www.ipdeletron.abinee.org.br/mct.pdf>.
5. **Ministério da Ciência e Tecnologia.** *Programa CI-Brasil*. [Online] [Citado em: 22 de Fevereiro de 2010.] <http://www.ci-brasil.gov.br/>.
6. **Sedra, Adel S. e Smith, Kenneth C.** *Microeletrônica*. s.l. : Makron Books, 2000.
7. **Razavi, Behzad.** *Design of Analog CMOS Integrated Circuits*. s.l. : McGraw-Hill Book, 2001.
8. **Sicard, Etienne e Bendhia, Sonia Delmas.** *Basics of CMOS Cell Design*. s.l. : McGraw-Hill Book, 2007.
9. **Razavi, Behzad.** *Fundamentals of Microelectronics*. s.l. : Wiley Publishing, 2006.
10. **Hastings, Alan e Hastings, Roy Alan.** *The Art Of Analog Layout*. s.l. : Prentice Hall, 2000.
11. **Clein, Dan.** *CMOS IC Layout: Concepts, Methodologies And Tools*. s.l. : Newnes, 1999.
12. **Johns, David e Martin, Ken.** *Analog Integrated Circuit Design*. s.l. : Wiley, 1996.

7 Anexos

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
UNIDADE ACADÊMICA DE ENGENHARIA ELÉTRICA
ESTRUTURA E CONCEPÇÃO DE CIRCUITOS INTEGRADOS

MÓDULO 1

Semicondutores e a Junção PN

Campina Grande, Fevereiro de 2010

Sumário

1	Introdução	4
2	Semicondutores.....	3
2.1	Semicondutor Intrínseco.....	4
2.2	Semicondutor Extrínseco.....	9
2.3	Difusão e Deriva.....	11
3	Junção PN	13
3.1	Junção PN em Equilíbrio.....	13
3.2	Junção PN Sob Polarização Reversa	16
3.3	Junção PN Sob Polarização Direta	19
3.4	Característica I/V	21
3.5	Ruptura da Junção PN	23
3.5.1	Efeito Zener.....	23
3.5.2	Efeito Avalanche	24

1 Introdução

Os semicondutores têm sido bastante pesquisados há mais de décadas e ocupam um lugar de destaque no conjunto dos materiais. Isso se deve ao alto grau que se tem alcançado no conhecimento de suas propriedades básicas, assim como no de suas aplicações. Pode-se dizer que hoje em dia os semicondutores são peças básicas em toda a eletrônica, a qual nos últimos anos tem mostrado um crescimento espetacular. Exemplo disso ocorre na microeletrônica, onde os circuitos são baseados nas estruturas dos semicondutores.

A indústria da microeletrônica continua a estabelecer grande demanda dos circuitos, empurrando para projetos ambiciosos que puxam os dispositivos semicondutores para seus limites. Assim, um bom entendimento da operação interna dos dispositivos é necessário.

Este módulo é iniciado com a apresentação do conceito de semicondutores e o estudo do movimento das cargas neles. Em seguida, tratar-se-á da junção PN, que serve como diodo, e da formulação do seu comportamento, já que em circuitos integrados fabricados por processo MOS existem diodos inerentes ao processo.

2 Semicondutores

Lembre-se de que cada átomo consiste de um núcleo de prótons e nêutrons rodeado por uma nuvem de elétrons. Os elétrons ocupam uma série de camadas, sendo a camada mais externa chamada *camada de valência* e os elétrons que a ocupam chamados *elétrons de valência*.

Átomos com a camada de valência completa, ou seja, com oito elétrons na camada de valência (Teoria do Octeto), possuem uma configuração estável. Aqueles com a camada de valência incompleta irão adquirir ou compartilhar elétrons até que a preencham. Dependendo da estratégia para preencher a camada de valência, um dos três tipos de ligação ocorrerá:

- Ligação metálica: ocorre entre os metais e a ligação é explicada pelo modelo da “nuvem eletrônica”;
- Ligação iônica: ocorre entre os metais e os não metais, onde há a formação de íons; e
- Ligação covalente: ocorre entre átomos não metálicos, onde há um compartilhamento de pares de elétrons da camada de valência.

Do ponto de vista elétrico, os materiais se dividem em três categorias, conforme a sua condutividade, em:

- Condutores;
- Isolantes; e,
- Semicondutores.

Os semicondutores formam um grupo de materiais que apresentam uma condutividade elétrica intermediária entre os condutores e os isolantes. Como será visto mais adiante, o semicondutor no estado puro e a baixas temperaturas apresenta uma condutividade relativamente baixa, assemelhando-se a de um material isolante. Porém, a condutividade desse material é uma função crescente com a temperatura, de forma que

sob temperatura ambiente ele apresenta uma condutividade apreciável, embora sempre seja bem menor que a de um metal.

O termo semicondutor também é utilizado para se falar dos dispositivos eletrônicos baseados nos materiais semicondutores, como o diodo e o transistor bipolar por exemplo.

Dentre os materiais semicondutores, o mais utilizado pelas indústrias de semicondutores é o silício (Si), sendo encontrado na areia, no quartzo e argila, por exemplo. Outro material usado, porém bem menos se comparado ao silício, é o germânio (Ge).

A existência de grandes quantidades de silício na crosta terrestre não é o principal motivo para seu uso como material de base na indústria de semicondutores, mas sim suas vantagens sobre os demais materiais semicondutores.

A utilização do silício começou após o uso, nos anos quarenta, do germânio. Experimentou-se numa câmara com vácuo o crescimento de uma camada de óxido de silício (SiO_2) em substratos de silício apenas com o aumento de temperatura (temperaturas entre 700 e 1200 °C) e a introdução de oxigênio. Verificou-se que o processo era muito simples e podia-se controlar a espessura do filme de SiO_2 com o tempo que demorava a experiência. Além disso, o SiO_2 é um isolante elétrico, o que permite a sua deposição entre duas camadas de metal para isolamento elétrico.

Outro motivo é que o silício não perde suas propriedades de semicondutor em temperaturas mais altas, se comparado ao germânio.

2.1 Semicondutor Intrínseco

Semicondutor intrínseco é a termo usado para designar o semicondutor que se encontra no estado puro e cristalizado.

Observando a tabela periódica – Figura 5, observa-se que o átomo de silício tem quatro elétrons de valência na sua terceira camada (a camada de valência). No estado cristalino, os átomos de silício ocupam posições tetraédricas, em uma rede similar ao do diamante, compartilhando cada um dos seus quatro elétrons de valência com átomos vizinhos por meio de ligações covalentes – Figura 6.

IIIA		IVA		VA	
5	2 3	6	2 4	7	2 5
B		C		N	
Boro		Carbono		Nitrogênio	
10.811		12.0107		14.00674	
...					...
13	2 8 3	14	2 8 4	15	2 8 5
Al		Si		P	
Alumínio		Silício		Fósforo	
26.981538		28.0855		30.973761	
31	2 8 18 3	32	2 8 18 4	33	2 8 18 5
Ga		Ge		As	
Gálio		Germânio		Arsênio	
69.723		72.64		74.92160	
		⋮			

Figura 5 – Seção da Tabela Periódica

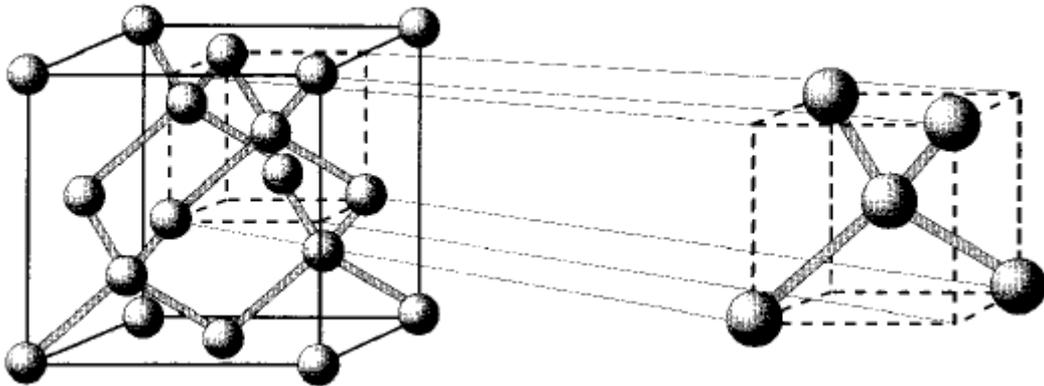


Figura 6 – Representação de uma estrutura de silício e suas ligações

A Figura 7 é uma representação bidimensional da estrutura e das ligações químicas do semicondutor silício, também válido para o germânio, onde se observa mais claramente o compartilhamento dos elétrons de valência entre os átomos.

Em temperaturas próximas do zero absoluto, os elétrons de valência são confinados as suas respectivas ligações covalentes, evitando o movimento livre dos elétrons pela estrutura. Em outras palavras, o cristal de silício comporta-se como um isolante para $T \rightarrow 0$ K. Porém, em temperaturas maiores (temperatura ambiente), experimentos realizados detectam uma pequena, mas mensurável, quantidade de elétrons livres no cristal de silício puro. A presença desses elétrons livres implica que eles receberam

energia a partir das vibrações dos átomos, devido à excitação térmica, em quantidade suficiente para romperem a ligação covalente.

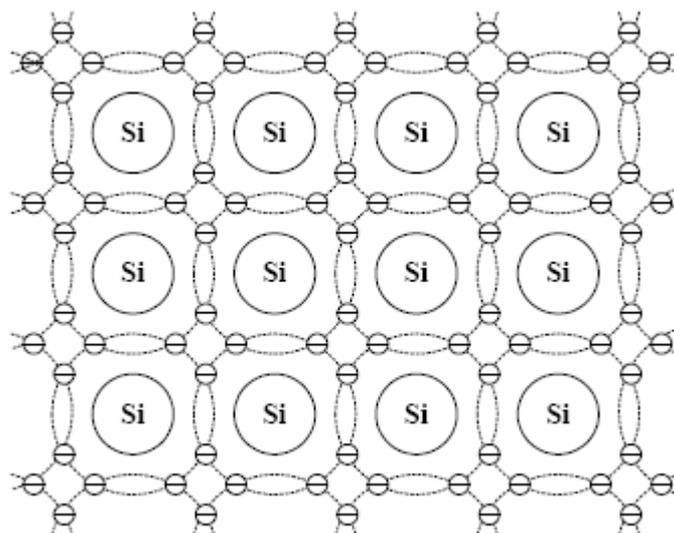


Figura 7 – Representação bidimensional de uma estrutura de silício

Esse fenômeno é explicado pela teoria das bandas de energia. Conforme ilustrado na Figura 8, um material possui duas bandas de energia: a banda de valência e a de condução. Existe uma energia associada a cada uma dessas duas bandas, determinadas de energia de valência (E_v) e energia de condução (E_c). Neste contexto, num material semiconductor as propriedades de condução situam-se entre a dos materiais condutores e a dos isolantes. De fato, a diferença entre um semiconductor e um isolante é muito pequena, arbitrando-se que um material é semiconductor se a sua energia da banda proibida (*bandgap energy*) que separa as duas bandas, $E_{bp} = E_c - E_v$, é aproximadamente inferior a $3,0 \text{ eV}^1$ – Figura 8(b). A baixas temperaturas, a banda de valência de um semiconductor está completamente ocupada por elétrons, comportando-se como isolante. A energia da banda proibida de um isolante é muito elevada, sendo muito difícil, senão impossível, a transferência de elétrons da banda de valência para a de condução. No caso dos materiais condutores, não existe uma separação clara entre as bandas de valência e de condução. De fato, essas se interceptam – Figura 8(c), explicando desta forma a abundância de elétrons livres ou que facilmente tornam-se livres. Na Tabela 1

¹ A unidade eV (elétron-Volt) representa a energia necessária para mover um elétron por uma diferença de potencial elétrico de 1 V. Equivale a $1,6 \times 10^{-19} \text{ J}$.

encontram-se listadas algumas energias da banda proibida para os materiais isolantes e semicondutores mais utilizados.

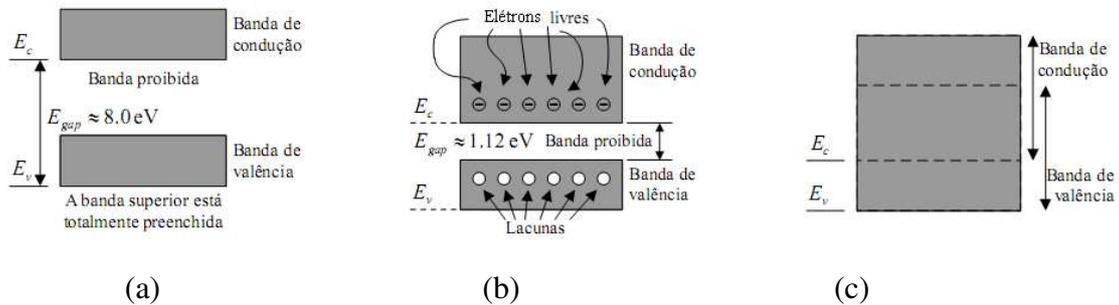


Figura 8 – Estrutura de bandas de energia (a) Para um material isolante (dióxido de silício), (b) Para um material semiconductor (silício) e (c) Para um material condutor

Tabela 1 – Energia da banda proibida para alguns materiais isolantes e semicondutores a 300 K

Material	E_{bp} (eV)	Material	E_{bp} (eV)
Germânio	0,67	Nitreto de gálio	3,3
Silício	1,12	Nitreto de silício	5,1
Arseneto de gálio	1,43	Dióxido de silício	8,0

Quando um elétron deixa a estrutura, uma lacuna ocorre. Um dos átomos que anteriormente possuía uma camada de valência completa agora carece de um elétron de valência. Essa situação é descrita na Figura 9. O átomo ionizado pode recuperar a camada de valência completa se ele tomar posse de um elétron de um átomo vizinho. Isso é aceitável desde que esse átomo vizinho ainda compartilhe elétrons com três átomos adjacentes. A lacuna não é eliminada, ela meramente desloca para o átomo adjacente. Como a lacuna é passada de átomo para átomo, ela move-se através da estrutura.

Suponha que um campo elétrico seja estabelecido através do cristal. Os elétrons livres movem-se em direção ao polo positivo do cristal. Já as lacunas comportam-se como se elas fossem partículas carregadas positivamente e movem-se em direção ao polo negativo do cristal – Figura 10.

O movimento de uma lacuna em direção ao pólo negativo do cristal é explicado assumindo-se que lacunas são carregadas positivamente. Similarmente, sua taxa de movimento pelo cristal é medida por uma quantidade chamada *mobilidade*². Lacunas têm baixa mobilidade se comparada aos elétrons. A baixa mobilidade das lacunas torna-

² Mais detalhe no tópico 2.3

as portadoras de carga menos eficientes. O comportamento de um dispositivo, então, depende de se sua operação envolver lacunas ou elétrons.

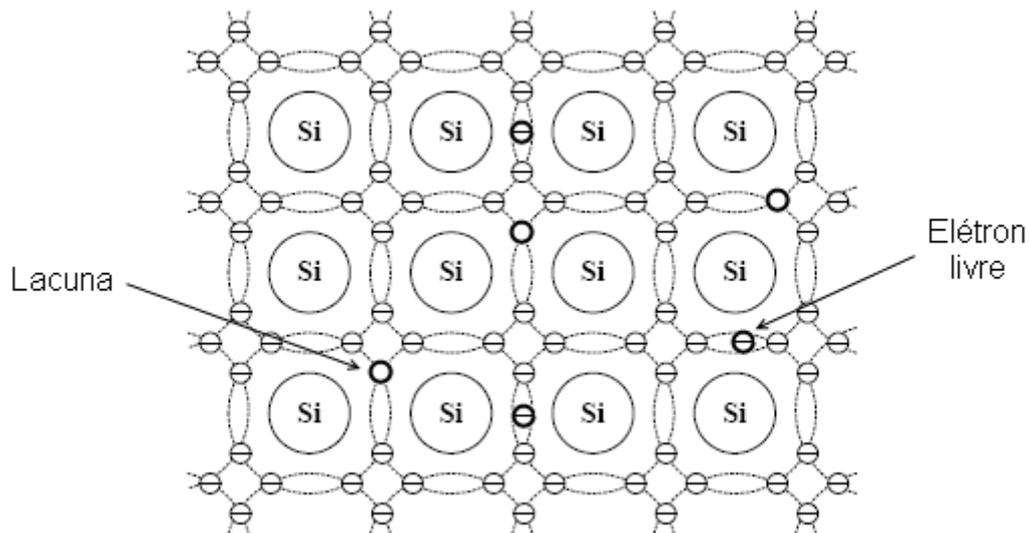


Figura 9 – Elétrons livres liberados pela energia térmica

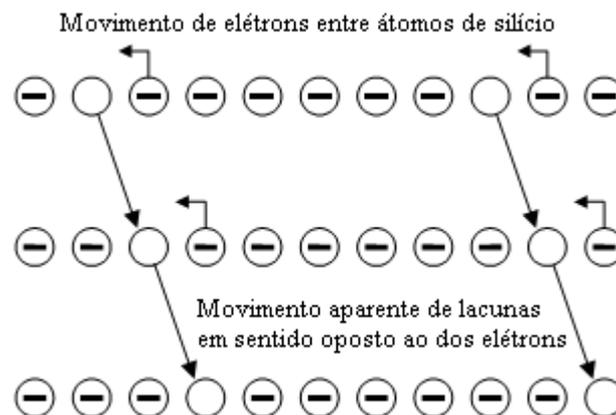


Figura 10 – Movimento aparente das lacunas

Lacunas e elétrons são chamados portadores por causa de sua função em transportar carga elétrica. Os portadores são sempre gerados em pares desde que a remoção de um elétron de valência da estrutura forma, simultaneamente, uma lacuna. As gerações de pares elétron-lacuna podem ocorrer a toda hora que alguma energia é absorvida pela estrutura. Vibrações térmicas geram portadores, assim como luz, radiação nuclear, fricção mecânica, entre outros processos.

Assim como portadores são gerados em pares, eles também se recombinam em pares. O mecanismo exato da recombinação dos portadores depende da natureza do semicondutor.

Disso, percebe-se que em um semicondutor intrínseco, em equilíbrio térmico, a concentração de elétrons livres, n , deve ser igual à concentração de lacunas, p . Assim:

$$n = p = (n_i)^2, \quad (1)$$

onde n_i representa a concentração de elétrons livres ou lacunas em um semicondutor intrínseco a uma dada temperatura.

E mais, o valor de n e p deve ser constante com o tempo se a temperatura do material for constante.

2.2 Semicondutor Extrínseco

A condutividade dos semicondutores depende de sua pureza. Os semicondutores intrínsecos têm pouca condutividade, pois contêm poucos portadores gerados por meio do calor. O processo de dopagem, que é a adição de impurezas no semicondutor, aumenta enormemente o número de portadores disponíveis. Esses semicondutores dopados, ou extrínsecos, podem aproximar-se da condutividade de um metal.

As impurezas podem ser elementos do grupo IIIA, que possuem a camada de valência preenchida com três elétrons, ou elementos do grupo VA, que possuem a camada de valência preenchida com cinco elétrons (observe a Figura 5). Um semicondutor levemente dopado pode conter somente um pouco de parte por bilhão de dopante. Mesmo um semicondutor dopado fortemente contém somente um pouco de centenas de partes por milhão devido à solubilidade limitada dos dopantes no silício.

O silício dopado com fósforo é um exemplo. O átomo de fósforo, elemento do grupo VA, compartilha quatro dos cinco elétrons de valência. Desde que oito elétrons completam inteiramente a camada de valência, nenhum lugar fica para o nono elétron, que é expelido do átomo de fósforo e vagueia através do cristal – Figura 11(a). Cada átomo de fósforo adicionado ao cristal de silício gera, então, um elétron livre. Elementos que doam elétrons ao semicondutor dessa maneira são chamados *doadores*. O arsênio (As), o antimônio (Sb) e, como comentado, o fósforo (P) são os elementos do

grupo VA mais usados no processamento de semicondutor como doadores para o silício.

Um semicondutor dopado com um grande número de doadores tem um predomínio de elétrons como portadores, aumentando enormemente sua condutividade. Devido isso, esses elétrons são considerados portadores majoritários.

Um poucas lacunas geradas pelo calor ainda existem, porém seu número diminui na presença de elétrons extras, já que a probabilidade de uma lacuna achar um elétron e recombinar aumenta. Essas lacunas são consideradas portadores minoritários.

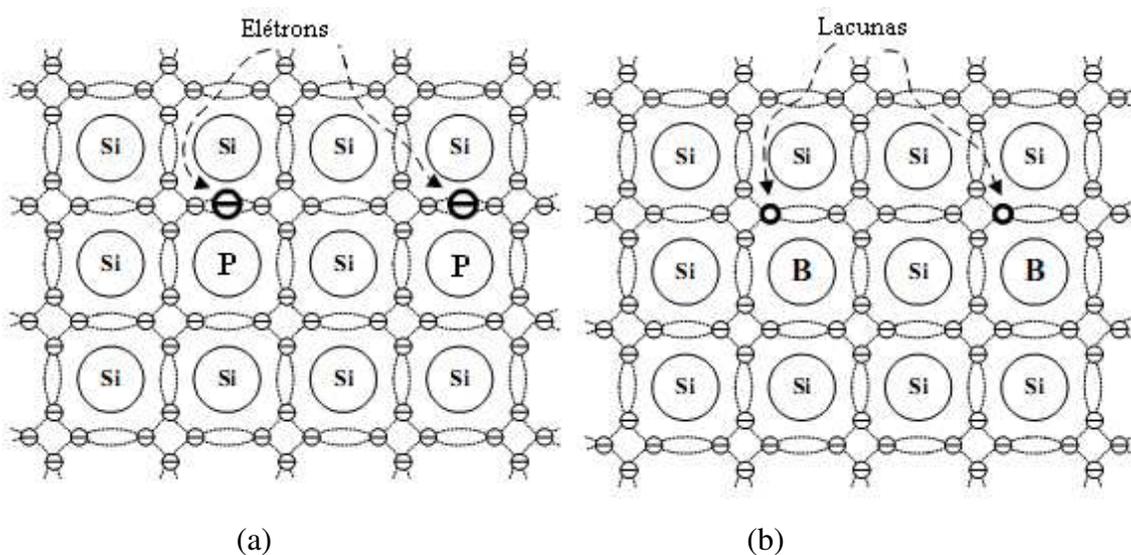


Figura 11 – Cristal de silício (a) Dopado por um doador e (b) Dopado por um aceitador

Um semicondutor dopado com doadores é dito ser do tipo N (semicondutor carregado negativamente). Um silício do tipo N fortemente dopado é, algumas vezes, sinalizado como N^+ , enquanto o levemente dopado é sinalizado como N^- .

Atenção, não confunda os símbolos + e - como representação da polaridade, pois eles denotam o número relativo de doadores presentes no semicondutor.

Outro tipo de semicondutor extrínseco é formado dopando o silício com boro, por exemplo. O boro, um elemento do grupo IIIA, tenta compartilhar seus elétrons de valência com seus quatro átomos vizinhos, mas por causa dele só conter três elétrons a quarta ligação não é completada, formando uma lacuna – Figura 11(b). Cada átomo de boro adicionado ao silício contribui para o aparecimento de uma lacuna.

Os elementos do grupo IIIA mais usados como dopante são o alumínio (Al), o gálio (Ga), o índio (In) e, como comentado, o boro (B). Esses elementos aceitam elétrons dos átomos adjacentes, por isso são chamados *aceitadores*.

Um semicondutor dopado com aceitadores é dito ser do tipo P (semicondutor carregado positivamente). O silício tipo P dopado fortemente é sinalizado como P⁺, enquanto que levemente dopado é sinalizado como P⁻. As lacunas são os portadores majoritários e os elétrons os portadores minoritários.

A tabela abaixo sumariza algumas das terminologias usadas para descrever os semicondutores extrínsecos.

Tabela 2 – Terminologias para o semicondutor extrínseco

Semicondutor	Tipo dopante	Dopante típico	Portador majoritário	Portador minoritário
Tipo N	Doador	P, As e Sb	Elétron	Lacunas
Tipo P	Aceitador	B, Al, Ga e In	Lacunas	Elétrons

Comentado na seção 2.1, a concentração de elétrons e lacunas em um semicondutor intrínseco é igual. Mas, o que pensar dessa concentração em um material dopado? Pode ser provado que mesmo nesse caso:

$$n_e p_e = (n_i)^2 \quad (2)$$

onde n_e e p_e denotam, respectivamente, a concentração de elétrons e lacunas em um semicondutor extrínseco. Lembre-se que n_i representa a concentração em um semicondutor intrínseco e é, então, independente do nível de dopagem.

2.3 Difusão e Deriva

A movimentação de portadores por meio de um cristal de silício resulta de dois processos separados: difusão e deriva. A difusão é uma movimentação aleatória dos portadores que ocorre a todo tempo e lugar, enquanto a deriva é um movimento unidirecional de portadores sobre a influência de um campo elétrico. Ambos os processos contribuem para a condução em semicondutores.

Difusão aproximadamente lembra a movimentação Browniana. Isso é, portadores individuais movem-se por meio do semicondutor até eles colidirem com os átomos da estrutura. O processo de colisão dispersa os portadores por ângulos imprevisíveis. Após

algumas poucas colisões, a movimentação dos portadores torna-se completamente aleatória. Os portadores vagueiam “sem destino”.

A difusão de portadores por meio de um semiconductor é análoga a difusão de moléculas de corante em uma água parada. Quando uma gota de corante cai na água, as moléculas do corante ocupam, inicialmente, um pequeno volume de líquido. As moléculas gradualmente espalham-se das regiões de alta concentração para regiões de baixa concentração. Ao final, o corante torna-se distribuído uniformemente por toda a solução de água. Similarmente, a difusão de portadores por gradientes de concentração gera uma corrente de difusão. A menos que algum mecanismo adicione mais portador, a difusão redistribui-os, no fim das contas, uniformemente pelo silício.

A movimentação de um portador sob a influência de um campo elétrico é chamado deriva. Embora o portador colida com a estrutura e, dessa forma, mova-se em um caminho aleatório, ele gradualmente dirige-se em uma direção específica. A aceleração devido ao campo elétrico e a colisão com o cristal neutralizam-se, conduzindo a uma velocidade constante para os portadores. Espera-se que a velocidade, v , seja proporcional a força do campo elétrico, E , ou seja:

$$v = \mu E \quad (3)$$

onde μ é chamado de mobilidade e é expresso em $\text{cm}^2/(\text{V}\cdot\text{s})$. No silício, por exemplo, a mobilidade dos elétrons é de $\mu_n = 1350 \text{ cm}^2/(\text{V}\cdot\text{s})$, enquanto que a mobilidade das lacunas é de $\mu_p = 480 \text{ cm}^2/(\text{V}\cdot\text{s})$.

3 Junção PN

Semicondutores dopados uniformemente têm poucas aplicações. Quase todos os dispositivos de estado sólido contêm uma combinação de múltiplas regiões do tipo N e P. A interface entre uma região tipo P e tipo N é chamada junção PN, constituindo um diodo.

Normalmente, os dispositivos semicondutores são formados por um mesmo material, no qual se introduz impurezas de um e outro tipo em duas regiões adjacentes do semiconductor. A Figura 12 mostra um esquema da seção transversal de uma lâmina (*wafers*) de silício no qual foi preparada uma junção PN para formar um diodo. Para conseguir essa junção, parte-se de uma lâmina do material semiconductor, tipo N por exemplo, e em uma região da lâmina adiciona-se uma dopagem com impurezas do tipo oposto, neste caso tipo P, em concentração suficiente para inverter o tipo de portadores nessa região. Finalmente, sobre a região dopada e na parte inferior da lâmina, deposita-se uma camada condutora com objetivo de fazer os contatos elétricos com o exterior.

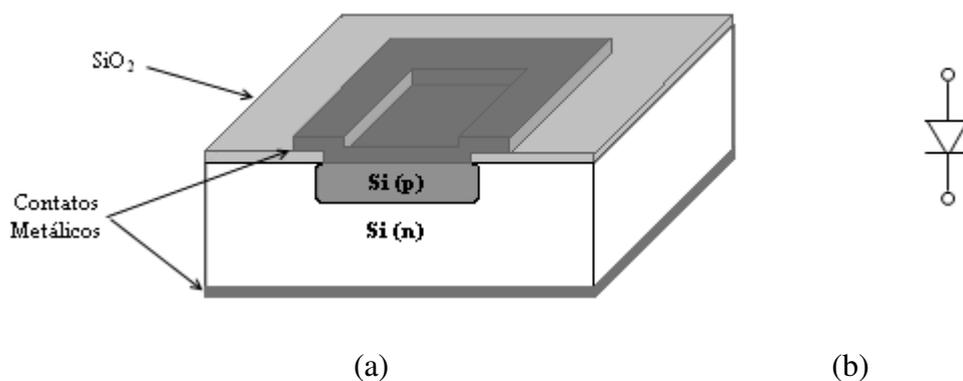


Figura 12 – (a) Esquema de um diodo semicondutor (b) Símbolo do diodo

3.1 Junção PN em Equilíbrio

Inicialmente, é estudada a junção PN com os terminais em aberto, ou seja, nenhuma tensão é aplicada sob o dispositivo. Assim, diz-se que a junção está em equilíbrio. Embora, aparentemente, de nenhum valor prático, essa condição fornece uma percepção

que se revela útil no entendimento do funcionamento no estado não equilibrado também.

Examinando a junção PN, reconhece-se que de um lado há um grande excesso de lacunas e do outro um grande excesso de elétrons. Ao longo da junção, o gradiente de concentração acentuado, tanto para os elétrons como para as lacunas, leva a duas grandes correntes de difusão: os elétrons fluindo do lado N para o lado P, enquanto as lacunas fluindo na direção oposta. Desde que se irá tratar com ambas as concentrações de elétrons e lacunas de cada lado da junção, as notações mostradas na Figura 13 serão usadas.

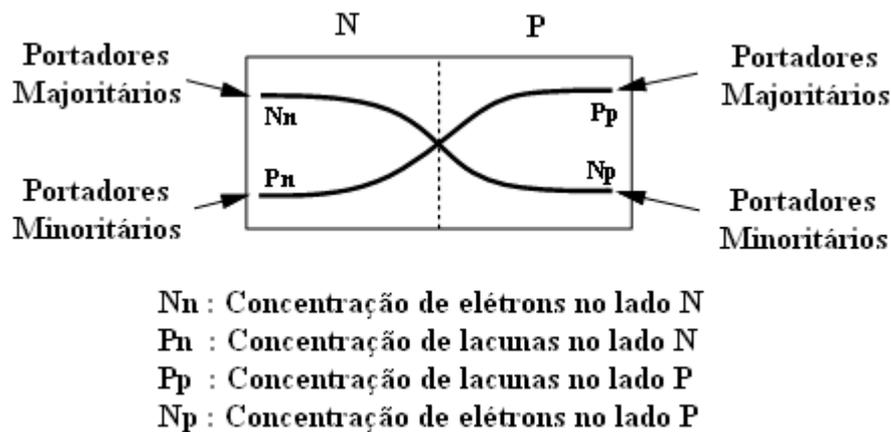


Figura 13 – Notações usadas

As correntes de difusão transportam uma grande quantidade de cargas de cada lado para o outro, mas elas eventualmente decaem para zero. Isso porque, se os terminais estão abertos (condição de equilíbrio), o dispositivo não pode conduzir uma corrente indefinidamente.

Mas o que cessa a corrente de difusão? Pode-se postular que as correntes parem após suficientes portadores livres moverem-se através da junção de modo a igualar as concentrações em ambos os lados. Porém, outro efeito domina a situação e cessa as correntes de difusão bem antes de esse ponto ser atingido.

Para entender esse efeito, admite-se que para cada elétron que parte do lado N, um íon positivo é deixado pra trás, isto é, a junção desenvolve-se com o tempo como visto na Figura 14. Nessa ilustração, a junção é repentinamente formada no instante $t = 0$. Com o passar do tempo ($t = t_i$), as correntes de difusão continuam a revelar mais íons.

Consequentemente, a vizinhança imediata da junção é esgotada de portadores livres e, então, chamada de região de depleção ($t = \infty$).

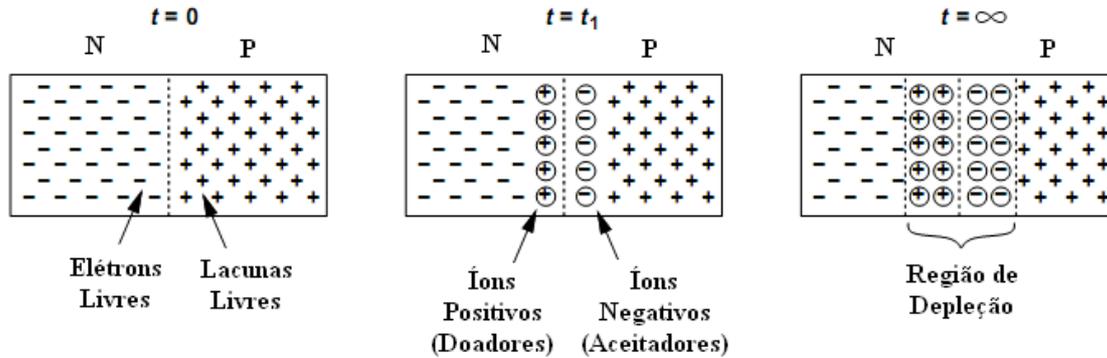


Figura 14 – Evolução da concentração de cargas em uma junção PN

Da física básica, uma partícula carregando uma carga líquida gera um campo elétrico ao seu redor. Assim, com a formação da região de depleção, um campo elétrico surge como mostrado na Figura 15. Curiosamente, o campo tende a forçar um fluxo de carga positiva da esquerda para a direita, enquanto o gradiente de concentração exige que o fluxo de lacunas seja da direita para a esquerda (e elétrons da esquerda para a direita). Então, suspeita-se que a junção atinge o equilíbrio uma vez que o campo elétrico é forte suficiente para cessar completamente as correntes de difusão. Alternativamente, pode-se dizer que, em equilíbrio, as correntes de deriva resultantes do campo elétrico cancelam as correntes de difusão devido aos gradientes.

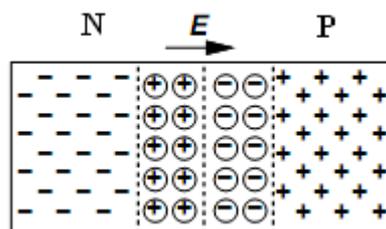


Figura 15 – Campo elétrico numa junção PN

Da observação sobre as correntes de deriva e difusão, talvez a pessoa seja tentada a escrever:

$$\left| I_{der,p} + I_{der,n} \right| = \left| I_{dif,p} + I_{dif,n} \right| \quad (4)$$

Essa condição, porém, permite um fenômeno falso: se o número dos elétrons fluindo do lado N para o lado P é igual àquele das lacunas indo do lado P para o lado N, então cada lado dessa equação é nulo, enquanto os elétrons continuam a acumular-se no lado P e as lacunas no lado N. Assim, impõe-se a seguinte condição de equilíbrio em **cada** portador:

$$|I_{der,p}| = |I_{dif,p}| \quad (5)$$

$$|I_{der,n}| = |I_{dif,n}| \quad (6)$$

A existência de um campo elétrico na região de depleção sugere que a junção pode exibir um potencial interno, V_0 . Tal valor pode ser determinado através da expressão:

$$V_0 = \left(\frac{kT}{q} \right) \ln \left(\frac{N_n P_p}{n_i^2} \right), \quad (7)$$

onde k é a constante de Boltzmann, T é a temperatura absoluta e q é a carga elementar do elétron.

Neste ponto, uma importante questão pode ser levantada. A junção não transporta nenhuma corrente líquida (pois seus terminais permanecem abertos), mas ela sustenta uma tensão. Como isso é possível? Observe que o potencial interno desenvolveu-se para opor-se ao fluxo das correntes de difusão (e, por isso, a região de depleção é algumas vezes chamada de “barreira de potencial”).

3.2 Junção PN Sob Polarização Reversa

Tendo analisado a junção PN em equilíbrio, pode-se estudar seu comportamento sobre condição mais interessante e útil. Começa-se aplicando uma tensão externa no dispositivo como mostra a Figura 16, onde a fonte de tensão deixa o lado N mais positivo que o lado P. Diz-se que a junção está sobre polarização reversa para enfatizar a conexão da polaridade positiva ao terminal do lado N.

Agora, reavaliam-se os resultados obtidos em equilíbrio para o caso da polarização reversa. Primeiro, determinar se a tensão externa aumenta o campo elétrico interno ou

se opõe a ela. Desde que em equilíbrio, \vec{E} é direcionado do lado N para o lado P, V_R aumenta o campo. Mas, um campo elétrico maior pode ser sustentado somente se uma grande quantidade de carga fixa é fornecida, requerendo que mais íons aceitadores e doadores sejam expostos e, assim, a região de depleção seja ampliada.

O que acontece com as correntes de difusão e de deriva? Desde que a tensão externa tenha fortalecido o campo elétrico, a barreira aumenta até mais que aquela em equilíbrio, dessa forma proibindo o fluxo de corrente. Em outras palavras, a junção transmite uma corrente insignificante sob polarização reversa.

Com nenhuma condução de corrente, uma junção PN polarizada reversamente não parece ser útil. Porém, uma importante observação mostrará o contrário. Note que, na Figura 16, à medida que V_R aumenta, mais carga positiva aparece no lado N e mais carga negativa no lado P.

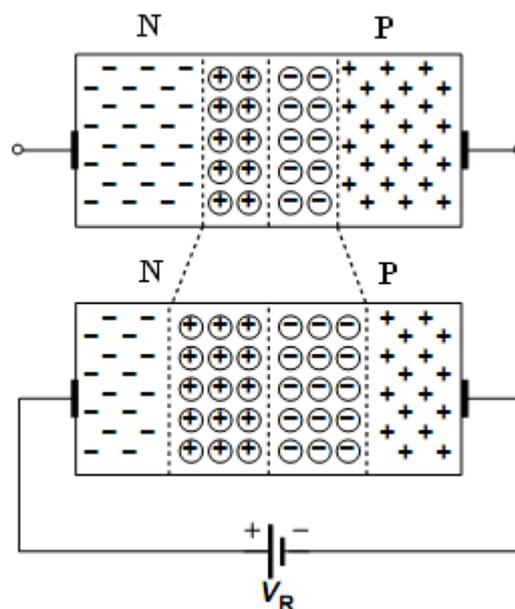


Figura 16 – Junção PN sob polarização reversa

Assim, o dispositivo opera como um capacitor – Figura 17(a). Veja as seções condutoras N e P como as duas placas de um capacitor e assumamos que a carga na região de depleção equivalentemente reside em cada placa.

A junção PN polarizada reversamente exibe uma propriedade única que se torna útil em projeto de circuitos. Reobservando a Figura 16, reconhece-se que, com o aumento de V_R , o mesmo acontece com a largura da região de depleção. Isso é, o diagrama

conceitual da Figura 17(a) pode ser desenhado como na Figura 17(b) para aumentar os valores de V_R , mostrando que a capacitância da estrutura diminui com as duas placas afastando-se uma da outra. A junção, assim, mostra uma capacitância dependente da tensão e esse valor é chamado de *capacitância de junção*.

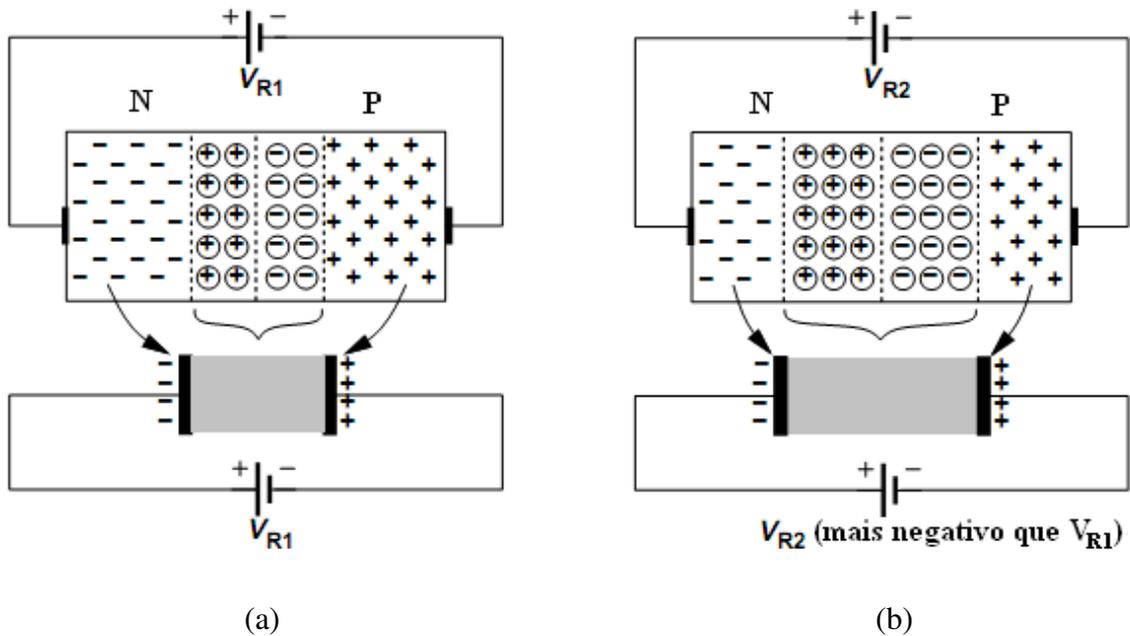


Figura 17 – Redução da capacitância de junção com a polarização reversa

Pode ser provado que a capacitância de junção por unidade de área é igual a:

$$C_j = \frac{C_{j0}}{\sqrt{1 - \frac{V_R}{V_0}}} \quad (8)$$

C_{j0} denota a capacitância correspondente a polarização nula ($V_R = 0$). O valor de C_{j0} é dado por:

$$C_{j0} = \sqrt{\frac{\epsilon_{si} q}{2} \frac{P_p N_n}{P_p + N_n} \frac{1}{V_0}} \quad (9)$$

onde ϵ_{si} representa a constante dielétrica do silício.

A variação da capacitância com a tensão aplicada torna o dispositivo um capacitor não linear, já que ele não satisfaz $Q = CV$.

Uma importante aplicação dos diodos polarizados reversamente é em câmeras digitais. Se a luz de energia suficiente é aplicada à junção PN, elétrons são arrancados de suas ligações covalentes e, desta maneira, pares elétrons-lacunas são criados. Com uma polarização reversa, os elétrons são atraídos para o terminal positivo da bateria e as lacunas para o terminal negativo. Como resultado, uma corrente flui por meio do diodo que é proporcional à intensidade luminosa. Pode-se dizer que a junção PN opera como um fotodiodo.

3.3 Junção PN Sob Polarização Direta

Nesta seção, o objetivo é mostrar como que a junção PN transporta uma corrente se o lado P é ligado a uma tensão positiva – Figura 18. Essa condição é dita ser polarizada diretamente.

Do que foi visto do dispositivo em equilíbrio e em polarização reversa, nota-se que a barreira de potencial desenvolvida na região de depleção determina o desejo do dispositivo conduzir. Na polarização direta, a tensão externa, V_F , tende a criar um campo direto do lado P em direção ao lado N (opondo-se ao campo interno que foi desenvolvido para cessar as correntes de difusão). Suponha que V_F , de fato, diminui a barreira de potencial pelo enfraquecimento do campo, permitindo assim maiores correntes de difusão.

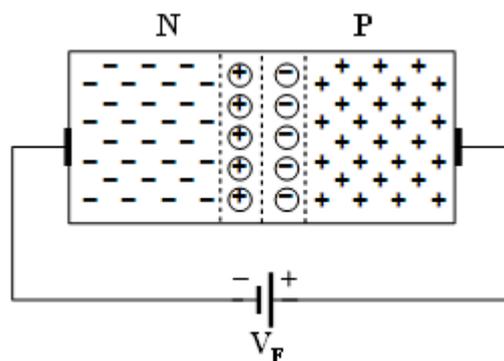


Figura 18 – Junção PN sob polarização direta

Para derivar as características $I-V$ na polarização direta, começar-se-á reformulando-se a equação (7) como a seguir:

$$P_{n,e} = \frac{P_{p,e}}{\exp\left(\frac{V_0}{V_T}\right)} \quad (10)$$

onde o índice e indica a condição de equilíbrio e $V_T = kT/q$ é chamado de *tensão térmica*.

Na polarização direta, a barreira de potencial é diminuída por uma quantidade igual à tensão aplicada. Logo:

$$P_{n,f} = \frac{P_{p,f}}{\exp\left(\frac{V_0 - V_F}{V_T}\right)} \quad (11)$$

Desde que o denominador exponencial decai consideravelmente, suponha-se que $P_{n,f}$ seja muito maior que $P_{n,e}$. Em outras palavras, a concentração de portadores minoritários no lado P decai rapidamente com a tensão de polarização direta, enquanto a concentração de portadores majoritários permanece relativamente constante. Essa afirmação aplica-se ao lado N também.

Como a junção vai do equilíbrio a polarização direta, N_p e P_n aumenta dramaticamente, conduzindo a uma variação proporcional nas correntes de difusão. Pode-se expressar a variação na concentração de lacunas no lado N como:

$$\Delta P_n = P_{n,f} - P_{n,e} \quad (12)$$

$$\Delta P_n = \frac{P_{p,f}}{\exp\left(\frac{V_0 - V_F}{V_T}\right)} - \frac{P_{p,e}}{\exp\left(\frac{V_0}{V_T}\right)} \quad (13)$$

$$\Delta P_n \approx \frac{P_p}{\exp\left(\frac{V_0}{V_T}\right)} \left[\exp\left(\frac{V_F}{V_T}\right) - 1 \right] \quad (14)$$

Igualmente, para a concentração de elétrons no lado P:

$$\Delta N_p \approx \frac{N_n}{\exp\left(\frac{V_0}{V_T}\right)} \left[\exp\left(\frac{V_F}{V_T}\right) - 1 \right] \quad (15)$$

O aumento na concentração de portadores minoritários sugere que as correntes de difusão devem aumentar por uma quantidade proporcional acima do seu valor de equilíbrio, isso é:

$$I_{tot} = \frac{P_p}{\exp\left(\frac{V_0}{V_T}\right)} \left[\exp\left(\frac{V_F}{V_T}\right) - 1 \right] - \frac{N_n}{\exp\left(\frac{V_0}{V_T}\right)} \left[\exp\left(\frac{V_F}{V_T}\right) - 1 \right] \quad (16)$$

De fato, pode-se provar que:

$$I_{tot} = I_s \left[\exp\left(\frac{V_F}{V_T}\right) - 1 \right] \quad (17)$$

Onde I_s é chamada de *corrente de saturação reversa*, a qual é função de dopagens, coeficiente de difusão, tempo de vida das cargas minoritárias e dimensões do diodo.

3.4 Característica $I-V$

Resumindo o que foi visto até o momento: na polarização direta, a tensão externa opõe-se ao potencial interno, aumentando as correntes de difusão substancialmente. Na polarização reversa, por outro lado, a tensão aplicada aumenta o campo, impedindo um fluxo de corrente. Além disso, escreveu-se a equação de junção como:

$$I_D = I_s \left[\exp\left(\frac{V_D}{V_T}\right) - 1 \right], \quad (18)$$

onde I_D e V_D denotam a corrente e tensão de diodo, respectivamente. À medida que V_D torna-se positivo e supera V_T , o termo exponencial cresce rapidamente e

$I_D \approx I_S \exp(V_D/V_T)$. Além disso, assume-se que $\exp(V_D/V_T) \gg 1$ na região de polarização direta. A equação (18) também se mantém na polarização reversa, isso é, para V_D negativo. Se $V_D < 0$ e $|V_D|$ supera V_T , então $\exp(V_D/V_T) \ll 1$ e:

$$I_D \approx -I_S \quad (19)$$

A Figura 19 esboça a característica $I-V$ geral da junção, revelando o porquê de I_S ser chamado de corrente de saturação reversa. Vê-se que a corrente sob polarização reversa como uma corrente de fuga (leakage). I_S , e assim a corrente de junção, é proporcional à área transversal do dispositivo. Por exemplo, dois dispositivos idênticos colocados em paralelo – Figura 20 – comporta-se como uma junção única com duas vezes o valor de I_S .

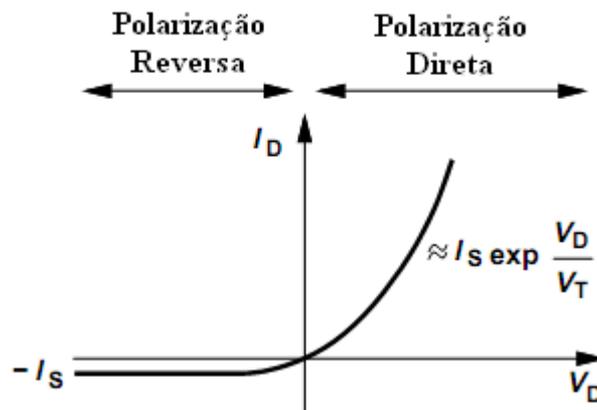


Figura 19 – Característica I/V de uma junção PN

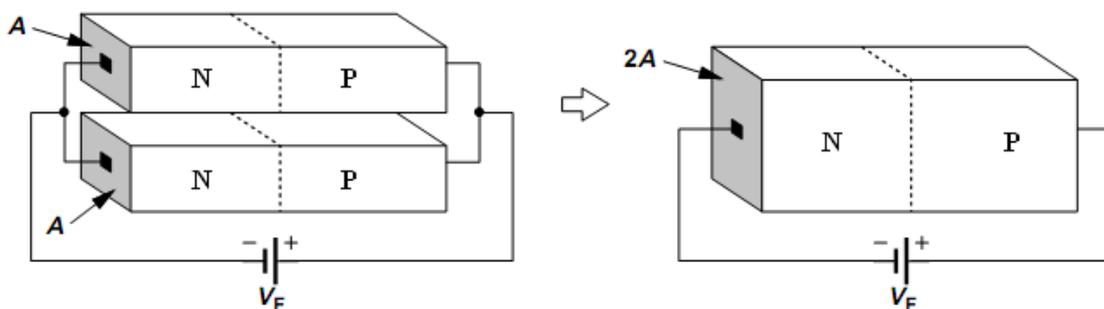


Figura 20 – Equivalência de dispositivos paralelos

3.5 Ruptura da Junção PN

Da Figura 19, a junção PN transporta somente uma corrente relativamente constante e pequena na região de polarização reversa. Porém, à medida que a tensão reversa em cima do dispositivo aumenta, eventualmente ocorrerá uma ruptura (*breakdown*) e uma corrente imprevista e enorme é observada. A Figura 21 esboça a característica $I-V$ do dispositivo mostrando esse efeito.

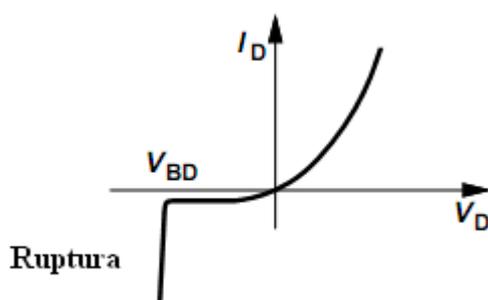


Figura 21 – Característica da Ruptura

A ruptura resultante de uma tensão alta (e, assim, um campo elétrico alto) pode ocorrer em qualquer material.

O fenômeno da ruptura numa junção PN ocorrer por um de dois possíveis mecanismos: *efeito Zener* e *efeito de avalanche*.

3.5.1 Efeito Zener

A região de depleção em uma junção PN contém átomos que tenham perdido um elétron ou uma lacuna e, assim, não fornece portadores conectados vagamente. Porém, um alto campo elétrico nessa região pode transmitir energia suficiente para que os restantes dos elétrons covalentes rompessem de suas ligações – Figura 22(a). Uma vez livre, os elétrons são acelerados pelo campo e arrastados para o lado N da junção. Esse efeito ocorrer em um campo elétrico de cerca de 10^6 V/cm (1 V/ μ m).

Para criar tais campos elétricos altos com tensões razoáveis, uma região de depleção estreita é requerida, o que da equação (9) traduz-se em altos níveis de

dopagem em ambos os lados da junção. Chamado de *efeito Zener*, esse tipo de ruptura aparece para tensões de polarização reversa da ordem de 3-8 V.

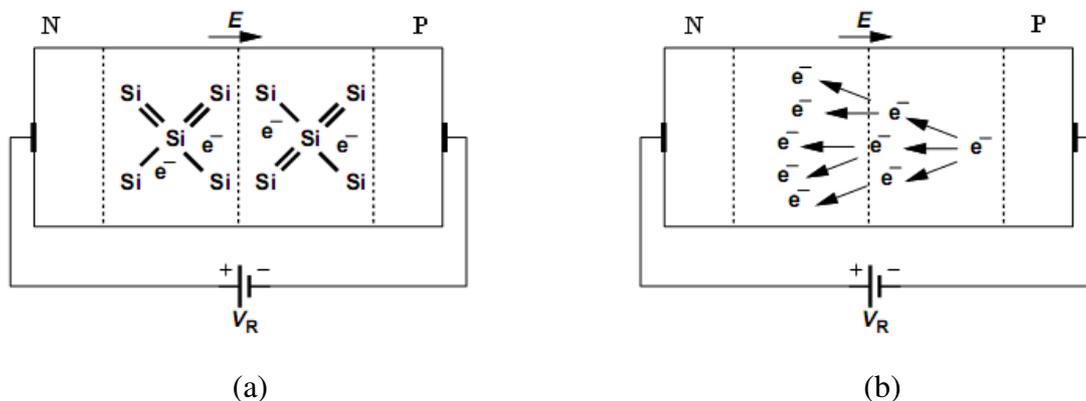


Figura 22 – (a) rompimento de elétrons devido o alto campo elétrico (b) efeito avalanche

3.5.2 Efeito Avalanche

Junções com níveis de dopagem moderado ou baixo ($< 10^{15} \text{ cm}^3$) geralmente não apresentam efeito Zener. Mas, à medida que a tensão de polarização reversa em tais dispositivos aumenta, um efeito de avalanche toma lugar. Embora a corrente de fuga seja muito pequena, cada portador entra a região de depleção experimentando um campo elétrico muito alto e então uma grande aceleração, ganhado energia bastante para liberar os elétrons de suas ligações covalentes. Chamada de *ionização de impacto*, esse fenômeno pode conduzir a avalanche: cada elétron liberado pelo impacto pode acelerar-se tanto no campo elétrico como a colidir com outro átomo com energia suficiente, com isso liberando mais um elétron. Agora, esses dois elétrons pode novamente adquirir energia e provocar mais colisões de ionização, aumentando rapidamente o número de portadores livres.

Um interessante contraste entre os fenômenos de Zener e de avalanche é que eles exibem opostos coeficientes de temperatura: a tensões de ruptura (V_{BD}) tem um coeficiente de temperatura negativo para o efeito Zener e positivo para o efeito avalanche.

Os efeitos Zener e avalanche não danifica o diodo se a corrente resultante permanecer abaixo de certo limite dado pelos níveis de dopagem e a geometria da

junção. Ambas as V_{BD} e a máxima corrente reversa permissível são especificados pelo fabricante do diodo.

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
UNIDADE ACADÊMICA DE ENGENHARIA ELÉTRICA
ESTRUTURA E CONCEPÇÃO DE CIRCUITOS INTEGRADOS

MÓDULO 2

Transistor MOS

Campina Grande, Fevereiro de 2010

Sumário

1	Introdução	2
2	Estrutura do Transistor MOS	3
3	Operação do Transistor MOS	6
3.1	O Transistor MOS como um Resistor Variável	7
3.2	Estrangulamento (<i>Pinch-off</i>) do Canal	10
3.3	Derivação das Características <i>I/V</i>	12
3.3.1	Densidade de Carga do Canal	12
3.3.2	Corrente de Dreno	13
3.3.3	Regiões de Triodo e Saturação.....	17
3.4	Modulação do Comprimento de Canal.....	19
3.5	Transcondutância do Transistor MOS.....	20
3.6	Efeitos de Segunda Ordem	21
3.6.1	Efeito de Corpo	21
3.6.2	Condução de <i>Subthreshold</i>	22
3.7	Transistor PMOS	23
4	Modelos de Dispositivo	24
4.1	Modelo Nível 1	24
4.2	Modelos Níveis 2 e 3	25
4.3	Modelos BSIM	25

1 Introdução

O transistor de efeito de campo do tipo metal/óxido semicondutor (MOSFET – *Metal/Oxide Semiconductor Field Effect Transistor*, ou simplesmente MOS) constitui, atualmente, um dos dispositivos eletrônicos mais importantes, senão o mais importante. O advento da tecnologia MOS, que se baseia fundamentalmente na utilização desse tipo de transistor, permitiu a gigantesca miniaturização dos circuitos observada atualmente, ampliando em muito a capacidade de processamento de informações.

Neste módulo serão abordados seus princípios de funcionamento e curvas características. Também é abordado o modelo BSIM, que descreve os parâmetros utilizados para a caracterização elétrica.

2 Estrutura do Transistor MOS

Para chegar à estrutura do transistor MOS, começa-se com uma simples geometria consistindo de uma placa condutora (metal, por exemplo), um isolador (dielétrico) e uma peça dopada de silício. Ilustrado na Figura 23(a), tal estrutura opera como um capacitor, já que o silício tipo P é, de certa forma, condutivo, espelhando qualquer carga depositada na placa superior.

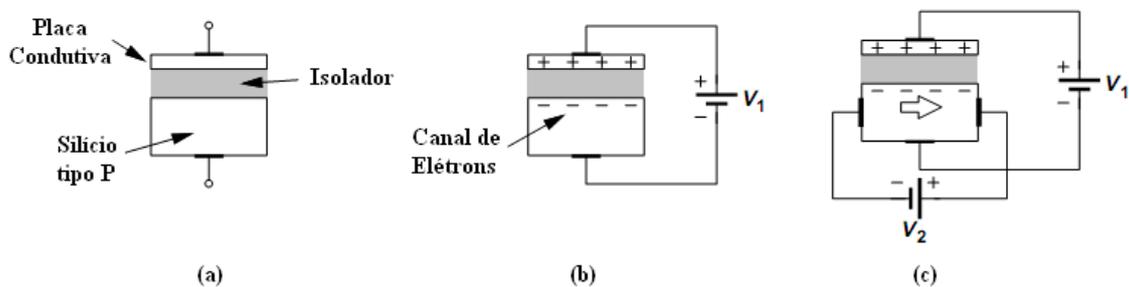


Figura 23 – (a) Dispositivo semiconductor hipotético (b) Operação como um capacitor (c) Fluxo de corrente como resultado de uma diferença de potencial

O que acontece se uma diferença de potencial é aplicada como visto na Figura 23(b)? Como uma carga positiva é depositada na placa superior, ela atrai carga negativa, ou seja, elétrons, do pedaço de silício (mesmo dopado de qualquer maneira com aceitadores, o silício tipo P contém um número pequeno de elétrons). Observe, então, que um canal de elétrons livres pode ser gerado na interface entre o isolador e o pedaço de silício, potencialmente servindo como um bom caminho condutivo se a densidade de elétrons é suficientemente alta. O ponto chave aqui é que a densidade de elétrons no canal varia com V_1 , evidentemente conforme a equação do capacitor $Q = CV$, onde C denota a capacitância entre as duas placas.

A dependência da densidade de elétrons por V_1 conduz a uma propriedade interessante: se, como observado na Figura 23(c), permite-se uma corrente fluindo da esquerda para direita através do material de silício, V_1 pode controlar a corrente pelo ajuste da resistividade do canal (lembre-se que a corrente prefere tomar o caminho de menor resistência, assim fluindo primeiramente através do canal, ao invés do corpo todo de silício).

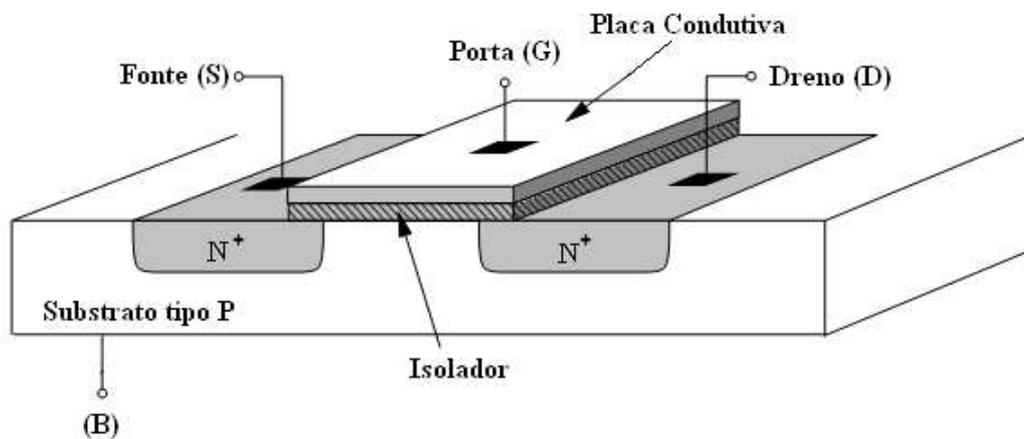
A equação $Q = CV$ sugere que para alcançar um forte controle de Q por V , o valor de C deve ser maximizado, por exemplo, pela redução da espessura da camada de dielétrico que separa as duas placas³. A capacidade da tecnologia de fabricação de silício para produzir camadas de dielétrico extremamente fino, mas uniforme, tem se mostrado essencial para o rápido avanço nos dispositivos microeletrônicos.

A discussão anterior leva a estrutura do MOSFET observada na Figura 24(a). Chamado de porta ($G - Gate$), a placa condutiva superior reside sobre uma camada fina de dielétrico (isolador), a qual está depositada sobre o substrato de silício tipo P ($B - Body$). Para permitir um fluxo de corrente através do material de silício, dois contatos são ligados ao substrato através de duas regiões fortemente dopadas do tipo N, devido à conexão direta de metal ao substrato não gerar um bom contato ôhmico⁴. Esses dois terminais são chamados de fonte ($S - Source$) e dreno ($D - Drain$) para indicar que o primeiro pode fornecer portadores de carga e o segundo pode absorvê-los. A Figura 24(a) revela que o dispositivo é simétrico com relação à fonte e o dreno, isso é, dependendo das tensões aplicadas ao dispositivo, um ou outro desses terminais pode drenar os portadores de carga do outro. Como explicado na seção seguinte, com uma fonte/dreno tipo N e um substrato tipo P, esse transistor opera com elétrons ao invés de lacunas e, assim, é dito ser um dispositivo MOS tipo N (NMOS). Quando uma fonte/dreno tipo P e um substrato tipo N constituem o transistor, o dispositivo é dito ser MOS tipo P (PMOS).

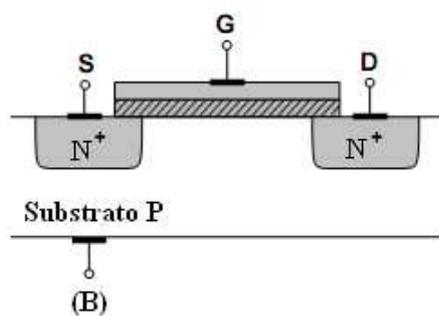
A Figura 25 mostra alguns dos símbolos que são comumente usados para os transistores MOS. Lembre-se que os transistores MOS são dispositivos de quatro terminais, com o substrato sendo o quarto terminal. Assim, usam-se os símbolos observados na Figura 25(a), para o NMOS, e Figura 25(c), para o PMOS, já que esses símbolos podem ser usados para mostrar a conexão do substrato explicitamente. Porém, em dispositivos NMOS o substrato é normalmente conectado à tensão mais negativa no circuito, enquanto que para dispositivos PMOS o substrato é normalmente conectado à tensão mais positiva no circuito. Nesse caso, usam-se os símbolos observados na Figura 25(b) e Figura 25(d), respectivamente.

³ A capacitância entre duas placas é dada por $\epsilon A/t$, onde ϵ é a constante dielétrica, A é a área de cada placa e t é a espessura do dielétrico.

⁴ O termo “contato ôhmico” enfatiza o fluxo de corrente bidirecional.

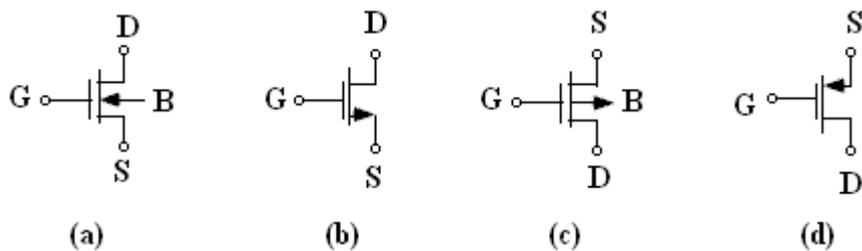


(a)



(b)

Figura 24 – (a) Estrutura de um MOS (b) Vista de frente (c) Símbolo do circuito



(a)

(b)

(c)

(d)

Figura 25 – Símbolos comumente usados (a) NMOS (b) NMOS com o substrato ligado ao menor nível de tensão (c) PMOS (d) PMOS com o substrato ligado ao maior nível de tensão

3 Operação do Transistor MOS

Desde que o transistor apresenta três terminais (ignorando, por enquanto, o fato do substrato formar o quarto terminal), pode-se abranger várias combinações de tensões e correntes terminais. Felizmente, com a corrente de porta sendo nula (para baixas frequências), a única corrente de interesse é a que flui entre a fonte e o dreno. É estudada a dependência dessa corrente sob a tensão de porta (por exemplo, para uma tensão de dreno constante) e sob a tensão de dreno (por exemplo, para uma tensão de porta constante). Esses conceitos ficarão claros a seguir.

Primeiro, considere o arranjo observado na Figura 26(a), onde a fonte e o dreno são aterrados e a tensão é variada. Lembre-se da Figura 23(b) que, enquanto V_G decai, a carga positiva sobre a porta deve ser espelhada por uma carga negativa no substrato. Foi dito na Seção 2 que os elétrons são atraídos para a superfície, na realidade, outro fenômeno precede a formação do canal. Enquanto V_G aumenta a partir de zero, a carga positiva na porta repele as lacunas no substrato, com isso expondo íons negativos e gerando uma região de depleção – Figura 26(b)⁵. Note que o dispositivo ainda age como um capacitor, mas nenhum canal de cargas móveis é gerado ainda. Assim, nenhum corrente pode fluir da fonte para o dreno. Diz-se que o transistor MOS está desligado.

Pode as junções fonte-substrato e dreno-substrato transportar corrente nesse modo? Para evitar esse efeito, o substrato é ligado à tensão mais negativa no circuito, geralmente o terra (0 V), garantindo que esses diodos não estão polarizados diretamente. Isso é válido para o NMOS, já que para o PMOS o substrato é ligado à tensão mais positiva no circuito para evitar tal efeito. Assim, por simplicidade, não é mostrado a conexão do substrato nos diagramas.

O que ocorre enquanto V_G aumenta? Para espelhar a carga na porta, mais íons negativos são expostos e a região de depleção sob o óxido torna-se profunda. Isso significa que o transistor nunca ligará? Felizmente, se V_G torna-se suficientemente positivo, elétrons livres são atraídos para a interface óxido-silício, formando um canal

⁵ Note que essa região de depleção contém somente uma polaridade de cargas imóveis, enquanto que a região de depleção de uma junção PN consiste de duas áreas de íons negativos e positivos nos dois lados da junção.

condutivo – Figura 26(c). Diz-se que o transistor MOS está ligado. O potencial de porta no qual o canal inicia a aparecer é chamada de tensão de *threshold*, V_{TH} , e recai na faixa de 300 mV a 500 mV. Note que os elétrons são rapidamente fornecidos pelas regiões de fonte/dreno N^+ e não necessita ser fornecida pelo substrato.

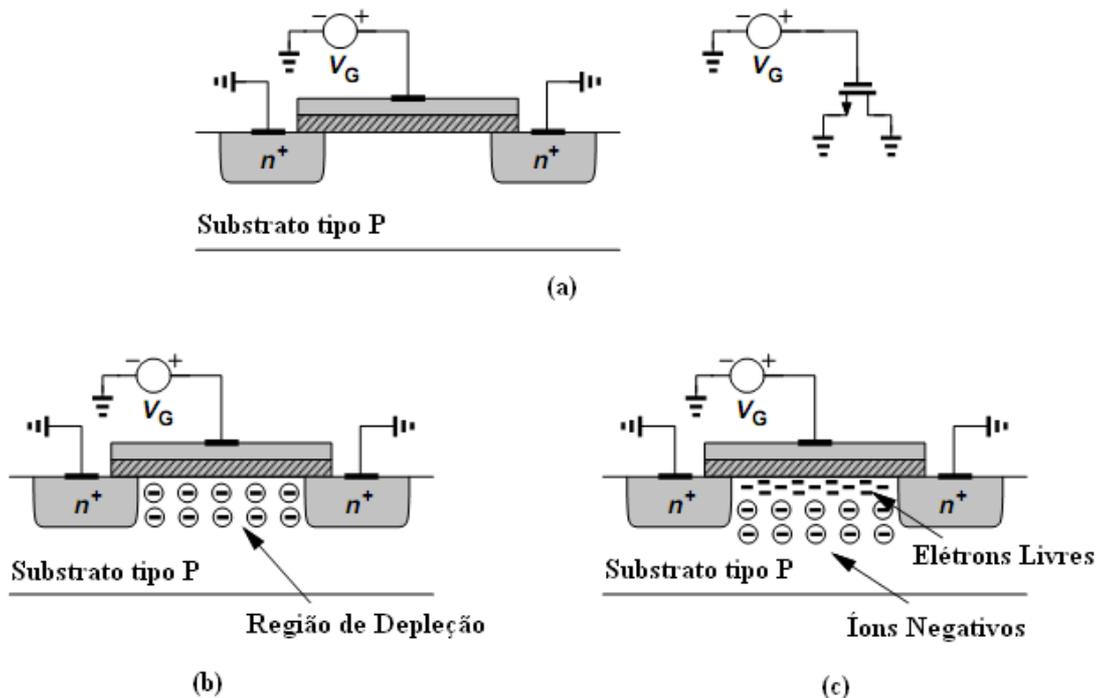


Figura 26 – (a) MOS com tensão de porta (b) Formação da região de depleção (c) Formação do canal

3.1 O Transistor MOS como um Resistor Variável

O canal condutivo entre a fonte e o dreno pode ser vista como um resistor. Além disso, desde que a densidade de elétrons no canal aumenta com V_G tornando-se mais positivo, o valor desse resistor muda com a tensão de porta. Conceitualmente ilustrado na Figura 27, tal resistor dependente da tensão mostra-se extremamente útil em circuitos analógicos e digitais.

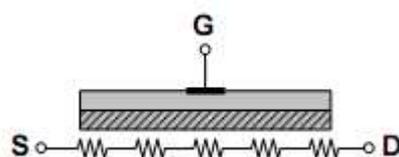


Figura 27 – Visão do MOS como um resistor dependente da tensão

No circuito observado na Figura 26(c), nenhuma corrente flui entre a fonte e dreno, visto que os dois terminais estão no mesmo potencial. Agora, eleva-se a tensão de dreno como observado na Figura 28(a) e examina-se a corrente de dreno (que é igual a corrente de fonte). Se $V_G < V_{TH}$, nenhum canal existe, o dispositivo está desligado e $I_D = 0$ sem levar em conta o valor de V_D . Por outro lado, se $V_G > V_{TH}$, então $I_D > 0$ – Figura 28(b). De fato, o caminho fonte-dreno pode agir como um simples resistor, rendendo a característica I_D - V_D observada na Figura 28(c). A inclinação da característica é igual a $1/R_{on}$, onde R_{on} denota a “resistência ligada” do transistor.

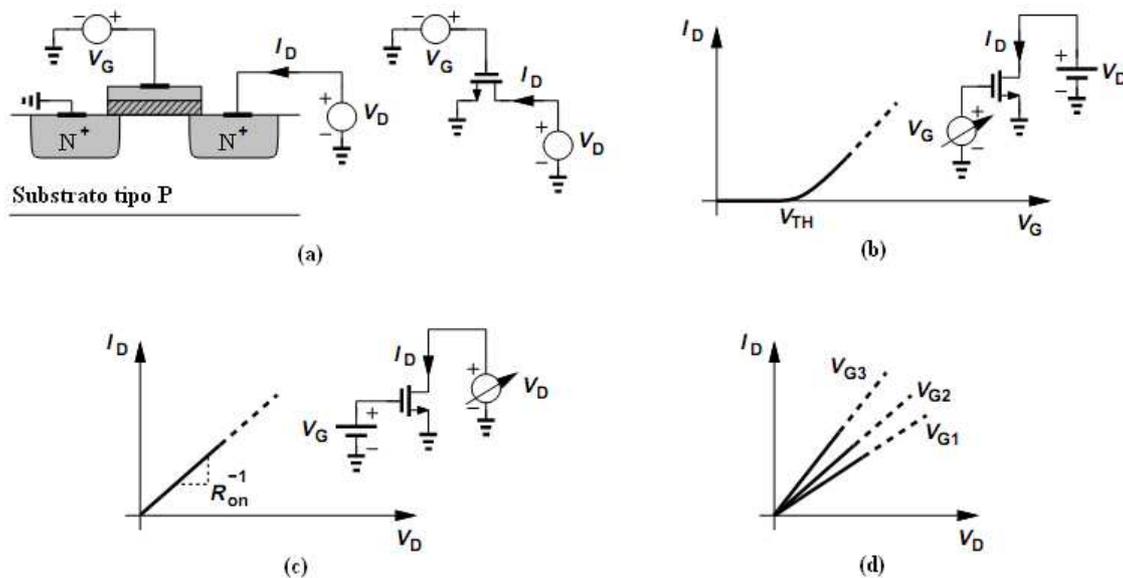


Figura 28 – (a) MOS com tensões de porta e de dreno (b) Característica I_D - V_G (c) Característica I_D - V_D (d) Características I_D - V_D para várias tensões de porta

O breve tratamento das características I/V do transistor MOS até aqui aponta para duas visões diferentes de operação: na Figura 28(b), V_G é variado enquanto V_D permanece constante, enquanto que na Figura 28(c), V_D é variado enquanto V_G permanece constante. Cada visão produz observações importantes da operação do transistor MOS.

Como a característica da Figura 28(b) muda se V_G aumenta? A alta densidade de elétrons no canal diminui a resistência entre fonte e dreno, rendendo uma grande inclinação. Observado na Figura 28(d), as características resultantes reforçam a noção de resistência dependente da tensão.

Lembre-se do Módulo 1 que as cargas que fluem nos semicondutores ocorrem por difusão ou deriva. O que achar do mecanismo de transporte no transistor MOS? Desde que a fonte de tensão está presa ao dreno criando um campo elétrico ao longo do canal, a corrente resulta da deriva de carga.

As características I_D-V_G e I_D-V_D observada na Figura 28(b) e Figura 28(c), respectivamente, representa uma função central no entendimento dos dispositivos MOS.

Enquanto ambos o comprimento e a espessura de óxido afetam o desempenho do transistor MOS, somente o primeiro está sob controle do projetista de circuito, isso é, ele pode ser especificado no leiaute do transistor. O segundo, por outro lado, é definido durante a fabricação e permanece constante para todos os transistores de uma dada geração da tecnologia. Outro parâmetro do transistor MOS controlado pelo projetista é a largura do transistor – Figura 29(a).

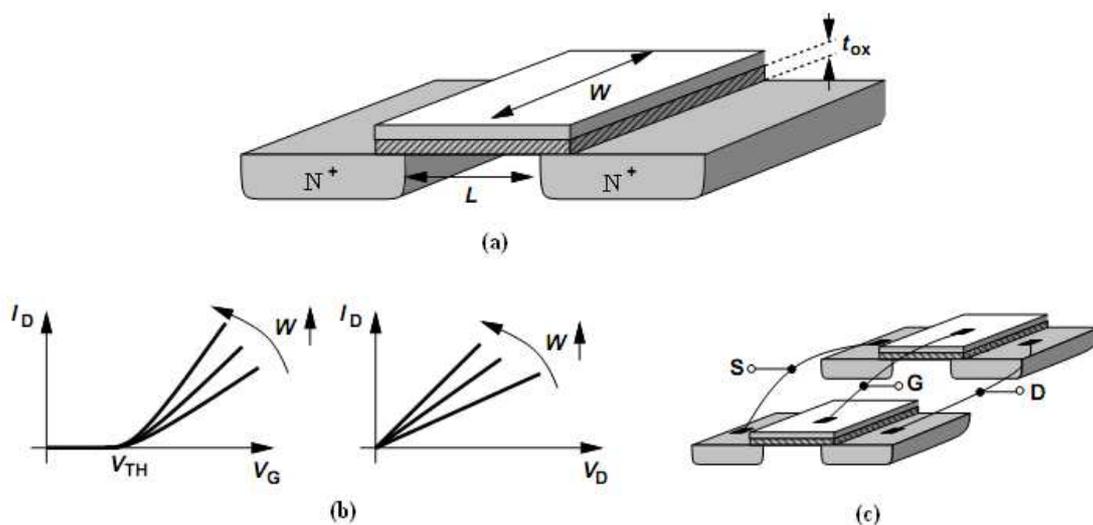


Figura 29 – (a) Dimensões de um transistor MOS (b) Características I_D para diferentes valores de W (c) Equivalente para os dispositivos em paralelo

Como a largura da porta impacta nas características I-V? Enquanto W aumenta, por conseguinte a largura do canal, a resistência entre a fonte e o dreno diminui, dando origem as curvas observadas na Figura 29(b). De outra perspectiva, um dispositivo largo pode ser visto como dois transistores estreitos em paralelo, produzindo uma corrente de dreno alta – Figura 29(c). Pode-se supor, então, que W deve ser maximizado, mas deve-se notar também que a capacitância de porta total aumenta com W , limitando possivelmente a velocidade do circuito. Assim, a largura de cada dispositivo no circuito deve ser escolhida cuidadosamente.

3.2 Estrangulamento (*Pinch-off*) do Canal

Do que foi visto até o momento, lembre que o dispositivo age como um resistor dependente da tensão se a tensão de porta exceder V_{TH} . Na realidade, porém, o transistor opera como uma fonte de corrente se a tensão de dreno é suficientemente positiva. Para entender esse efeito, duas observações são feitas: (1) para formar um canal, a diferença de potencial entre a porta e a interface óxido-silício deve exceder V_{TH} ; (2) se a tensão de dreno permanecer mais alta que a tensão de fonte, então a tensão em cada ponto ao longo do canal, com relação ao terra, aumenta a medida que desloca-se da fonte em direção ao dreno. Ilustrado na Figura 30(a), esse efeito decorre da queda de tensão gradual ao longo da resistência de canal. Desde que a tensão de porta é constante e o potencial na interface óxido-silício aumenta da fonte para o dreno, a diferença de potencial entre a porta e a interface óxido-silício diminui ao longo do eixo x – Figura 30(b). A densidade de elétrons no canal segue a mesma tendência, diminuindo para um mínimo de $x = L$.

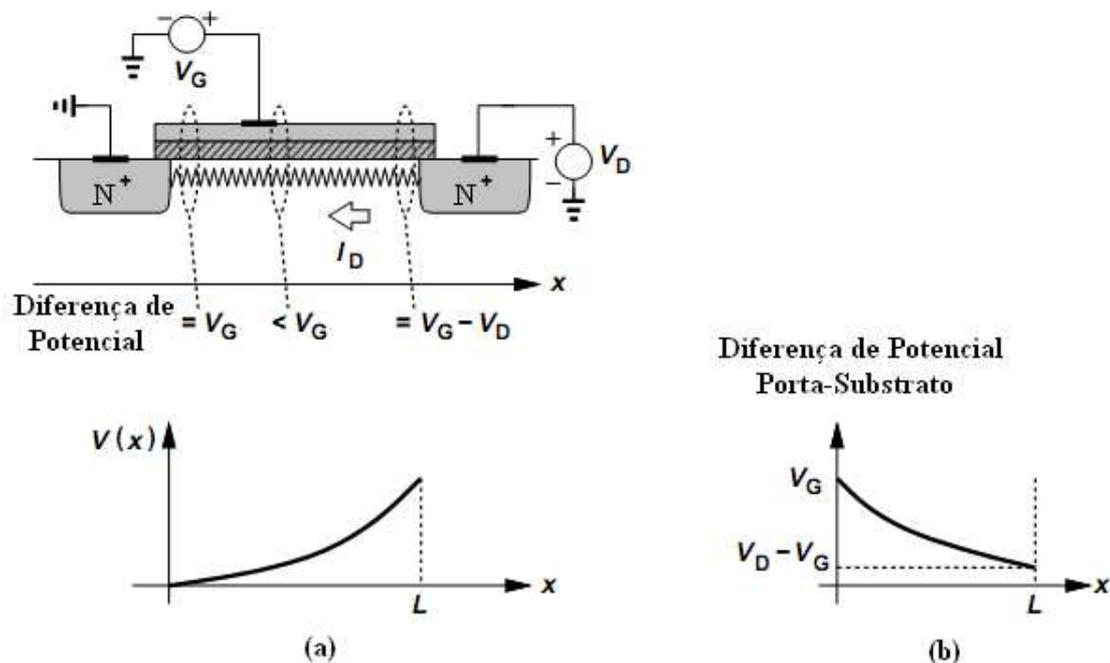


Figura 30 – (a) Variação de potencial do canal (b) Diferença de tensão porta-substrato ao longo do canal

Dessas observações, concluí-se que se a tensão de dreno é alta o suficiente para produzir $V_G - V_D \leq V_{TH}$, então o canal cessa para existir perto do dreno. Diz-se que a

diferença de potencial porta-substrato não é suficiente em $x = L$ para atrair elétrons e o canal é estrangulado (*pinched-off*).

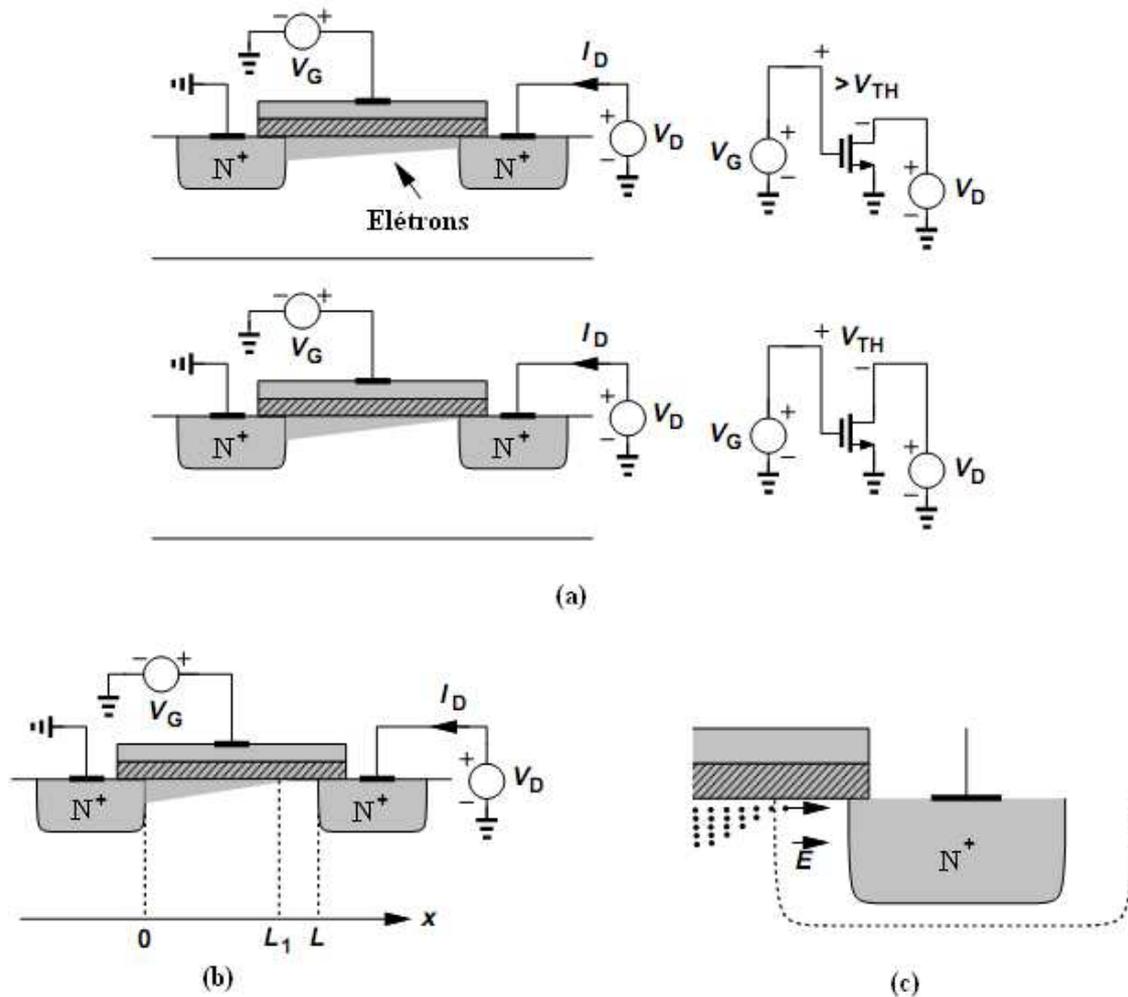


Figura 31 – (a) Pinch-off (b) Variação do comprimento com a tensão de dreno (c) Efeito detalhado perto do dreno

O que acontece se V_D aumentar até maior que $V_G - V_{TH}$? Desde que $V(x)$ parta de 0 em $x = 0$ para $V_D > V_G - V_{TH}$ em $x = L$ – Figura 31(b). Dessa forma, o dispositivo não contém canal entre L_1 e L . Isso significa que o transistor não pode conduzir corrente? Não, o dispositivo ainda conduz: como ilustrado na Figura 31(c), uma vez os elétrons alcançaram o final do canal, eles experimentam um alto campo elétrico na região de depleção ao redor da junção de dreno e são rapidamente arrastados para o terminal de dreno. Apesar de tudo, como observado na seção seguinte, a tensão de dreno não afeta

longamente a corrente significativamente e o transistor MOS age como uma fonte de corrente constante. Note que as junções de fonte-substrato e dreno-substrato não transportam corrente.

3.3 Derivação das Características I/V

Com o estudo anterior, pode-se agora formular o comportamento do transistor MOS em termos de suas tensões terminais.

3.3.1 Densidade de Carga do Canal

A densidade de carga refere-se a uma expressão para a carga do canal por unidade de comprimento. De $Q = CV$, note que se C é a capacitância de porta por unidade de comprimento e V a diferença de tensão entre a porta e o canal, então Q é a densidade de carga desejada. Denotando a capacitância de porta por unidade de área por C_{ox} (expresso em F/m^2 ou $fF/\mu m^2$), escreve-se $C = WC_{ox}$ para explicar a largura do transistor – Figura 32. Além disso, tem-se $V = V_{GS} - V_{TH}$ devido nenhuma carga móvel existir para $V_{GS} < V_{TH}$. Segue que:

$$Q = WC_{ox}(V_{GS} - V_{TH}) \quad (20)$$

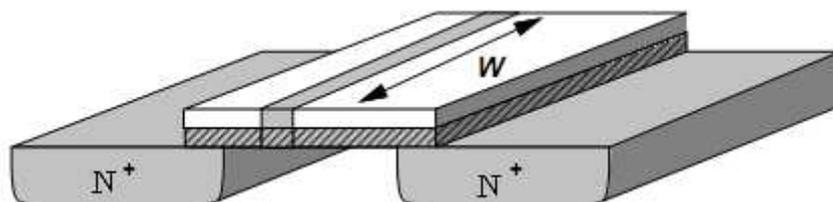


Figura 32 – ilustração da capacitância por unidade de comprimento

Note que Q é expresso em Coulomb/metro. Agora, lembre-se da Figura 30(a) que a tensão de canal varia ao longo do comprimento do transistor e a densidade de carga decai à medida que se desloca da fonte para o dreno. Assim, a Eq. (20) é válida somente

perto do terminal de fonte, onde o potencial do canal permanece próximo de zero. Como observado na Figura 33, denota-se o potencial do canal em x por $V(x)$ e escreve-se:

$$Q(x) = WC_{ox} [V_{GS} - V(x) - V_{TH}] \quad (21)$$

Note que $V(x)$ vai de zero a V_D se o canal não está estrangulado (*pinched-off*).

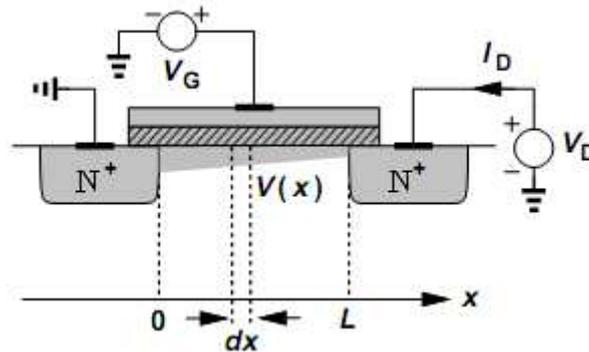


Figura 33 – Ilustração do dispositivo para cálculo da corrente de dreno

3.3.2 Corrente de Dreno

Qual é a relação entre a densidade de carga móvel e a corrente? Considere uma barra de semiconductor tendo uma densidade de carga uniforme (por unidade de comprimento) igual a Q e transportando uma corrente I – Figura 34. Observe que: (1) I é dada pela carga total que passa através o corte transversal da barra em um segundo e (2) se os portadores movem-se com uma velocidade de v m/s, então a carga confinada em v metros ao longo da barra passa através do corte transversal em um segundo. Desde que a carga confinada em v metros é igual a $Q v$, tem-se:

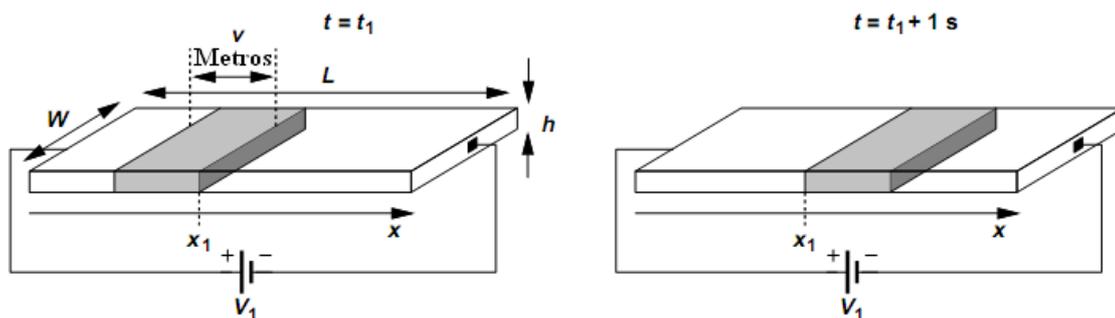


Figura 34 – Relação entre velocidade de carga e corrente

$$I = Qv \quad (22)$$

Do Módulo 1,

$$v = -\mu_n E \quad (23)$$

$$v = +\mu_n \frac{dV}{dx} \quad (24)$$

Onde dV/dx denota a derivada da tensão em um dado ponto. Combinando (21), (22) e (24), obtém-se:

$$I_D = WC_{ox} [V_{GS} - V(x) - V_{TH}] \mu_n \frac{dV(x)}{dx} \quad (25)$$

Curiosamente, I_D deve permanecer constante ao longo do canal, $V(x)$ e dV/dx deve variar tal que o produto de $[V_{GS} - V(x) - V_{TH}]$ e dV/dx seja independente de x .

Enquanto é possível solucionar a equação diferencial acima para obter $V(x)$ em termos de I_D , necessita-se achar uma expressão para I_D em termos das tensões terminais. Para tal finalidade, escreve-se:

$$\int_{x=0}^{x=L} I_D dx = \int_{V(x)=0}^{V(x)=V_{DS}} \mu_n WC_{ox} [V_{GS} - V(x) - V_{TH}] dV \quad (26)$$

Isso é:

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} [2(V_{GS} - V_{TH}) V_{DS} - V_{DS}^2] \quad (27)$$

Analisando a Eq. (27) de diferentes perspectivas, primeiro note a dependência linear de I_D sob μ_n , C_{ox} e W/L é para ser esperado: uma alta mobilidade gerando uma grande corrente para uma dada tensão de dreno-fonte; uma alta capacitância de óxido de porta conduzindo a uma grande densidade de elétron no canal para uma dada tensão porta-fonte; e, um grande W/L (chamado de “razão de aspecto” do dispositivo) é equivalente a colocar mais transistores em paralelo – Figura 29(c). Segundo, para uma

V_{GS} constante, I_D varia parabolicamente com V_{DS} – Figura 35, alcançando um máximo de:

$$I_{D,\max} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2, \quad (28)$$

em $V_{DS} = V_{GS} - V_{TH}$. É comum escrever W/L como a razão de dois valores, por exemplo, $5 \mu\text{m}/0,18 \mu\text{m}$ (ao invés de 27,8) para enfatizar a escolha de W e L . Enquanto somente a razão aparece em várias equações MOS, os valores individuais de W e L também tornam-se críticos em vários casos. Por exemplo, se ambos W e L são dobrados, a razão permanece a mesma, mas as capacitâncias de porta aumentam.

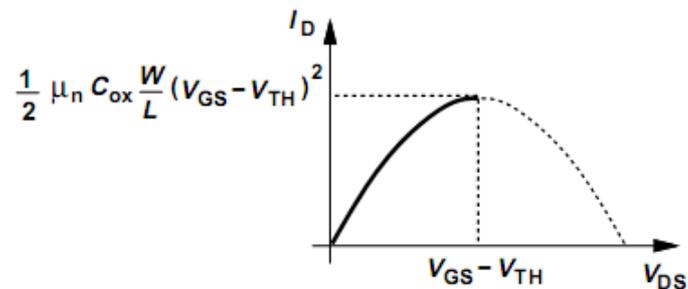


Figura 35 – Característica parabólica de I_D - V_{DS}

Se esboçado as curvas características de I_D - V_{DS} para diferentes valores de V_{GS} , observa-se que as características exibem máximas que seguem um formato parabólico entre elas, já que $I_{D,\max} \propto (V_{GS} - V_{TH})^2$ – Figura 36.

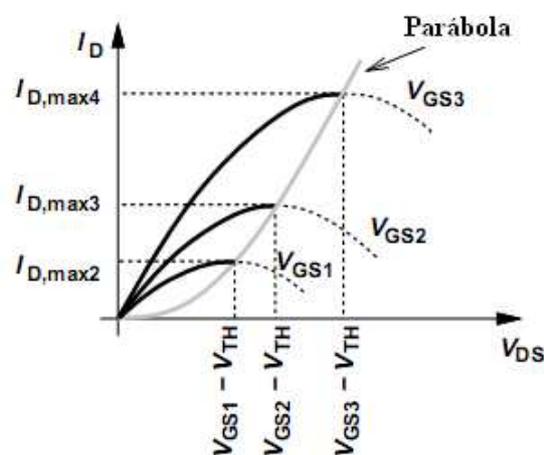


Figura 36 – Características do transistor MOS para diferentes tensões de porta-fonte

A relação não linear entre I_D e V_{DS} revela que o transistor não pode, geralmente, ser modelado como um simples resistor linear. Porém, se $V_{DS} \ll 2(V_{GS} - V_{TH})$, a Eq. (27) reduz para:

$$I_D \approx \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH}) V_{DS}, \quad (29)$$

exibindo um comportamento I_D - V_{DS} linear para um dado V_{GS} . De fato, a resistência equivalente é dada por V_{DS}/I_D :

$$R_{on} = \frac{1}{\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})} \quad (30)$$

De outra perspectiva, para um pequeno V_{DS} (próximo a origem), as parábolas na Figura 36 podem ser aproximadas por linhas retas tendo diferentes inclinações – Figura 37.

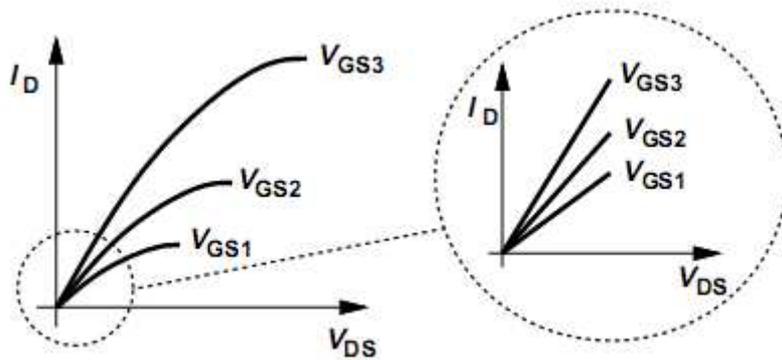


Figura 37 – Características detalhadas para um pequeno V_{DS}

Como visto no início Seção 3, a Eq. (30) sugere que a resistência pode ser controlada pela tensão porta-fonte. Em particular, para $V_{GS} = V_{TH}$, $R_{on} = \infty$, isso é, o dispositivo pode operar como uma chave eletrônica.

Na maioria dos casos, é desejável atingir uma baixa resistência para as chaves MOS. O projetista de circuito deve, assim, maximizar W/L e V_{GS} .

3.3.3 Regiões de Triodo e Saturação

A Eq. (27) expressa a corrente de dreno em termos das tensões terminais do dispositivo, implicando que a corrente começa a cair para $V_{DS} > V_{GS} - V_{TH}$. Diz-se que o dispositivo opera na região de triodo (também conhecida como região linear) se $V_{DS} < V_{GS} - V_{TH}$. Também se usa o termo região de triodo profunda para $V_{DS} \ll 2(V_{GS} - V_{TH})$, onde o transistor opera como um resistor.

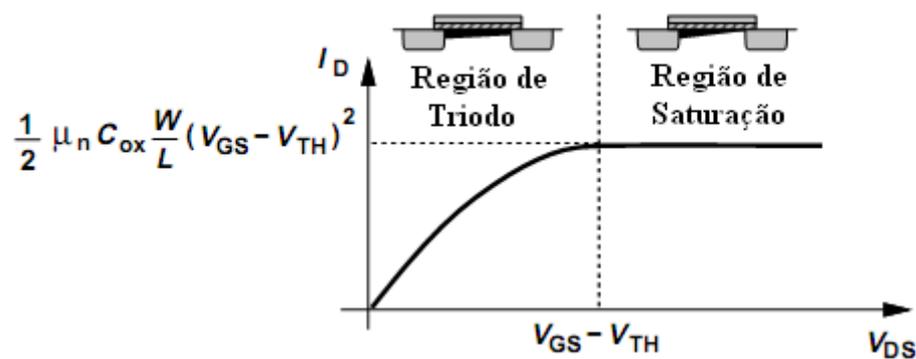


Figura 38 – Característica geral do transistor MOS

Na realidade, a corrente de dreno alcança a saturação, isso é, torna-se constante para $V_{DS} > V_{GS} - V_{TH}$ – Figura 38. Para entender o porquê, lembre-se da Figura 31 que o canal experimenta um estrangulamento se $V_{DS} = V_{GS} - V_{TH}$. Assim, o aumento em V_{DS} simplesmente desloca o ponto de estrangulamento ligeiramente em direção ao dreno. Também, relembre que as Eqs. (25) e (26) são válidas somente onde cargas do canal existem. Segue que a integração em (26) deve incluir somente o canal, isso é, de $x = 0$ a $x = L_1$ na Figura 31(b) e ser modificado para:

$$\int_{x=0}^{x=L_1} I_D dx = \int_{V(x)=0}^{V(x)=V_{GS} - V_{TH}} \mu_n C_{ox} W [V_{GS} - V(x) - V_{TH}] V dx \quad (31)$$

Note que os limites superiores correspondem ao ponto de estrangulamento do canal. Em particular, a integral à direita é avaliada até $V_{GS} - V_{TH}$, ao invés de V_{DS} . Consequentemente:

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L_1} (V_{GS} - V_{TH})^2, \quad (32)$$

como resultado independente de V_{DS} e idêntico a $I_{D,max}$ em (28) se assumido que $L_1 \approx L$. Dita de excesso de tensão, a quantidade $V_{GS} - V_{TH}$ dita uma regra chave em circuitos MOS. Os transistores MOS são, algumas vezes, chamados de dispositivos de ordem quadrática para enfatizar a relação entre I_D e o excesso de tensão. Pela razão de redução, daqui em diante denota-se L_1 com L .

Uma ilustração conceitual pode ser usada para determinar a região de operação – Figura 39. Note que a diferença de potencial porta-dreno satisfaz essa finalidade e não se necessita calcular as tensões de porta-fonte e de porta-dreno separadamente.

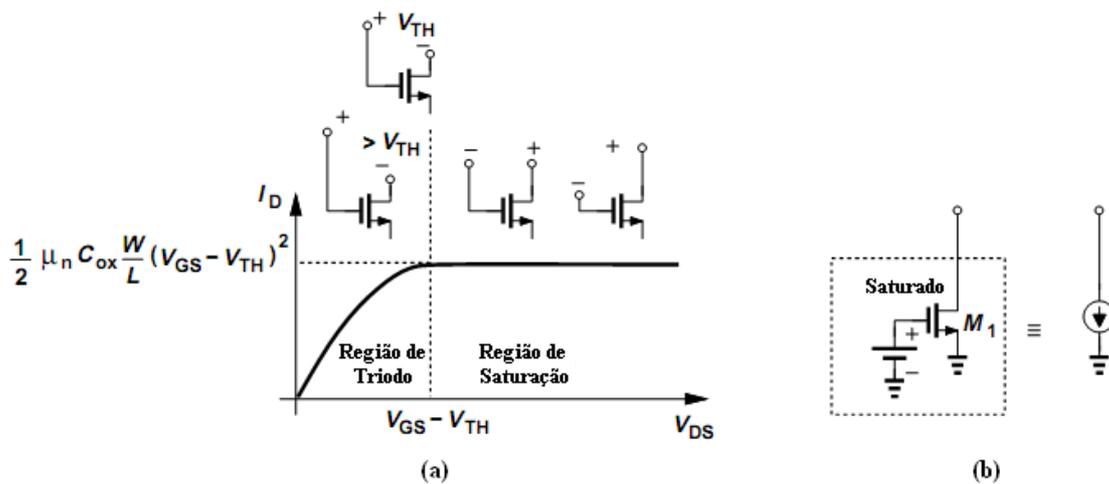


Figura 39 – Ilustração das regiões de triodo e de saturação baseadas nas tensões de porta e dreno

Exibindo uma corrente plana na região de saturação, um transistor MOS pode operar como uma fonte de corrente tendo um valor dado pela Eq. (32). Além disso, a dependência de ordem quadrática de I_D sob $V_{GS} - V_{TH}$ sugere que o dispositivo pode agir como uma fonte de corrente controlada por tensão.

3.4 Modulação do Comprimento de Canal

No estudo do efeito de estrangulamento (*pinch-off*), observou-se que o ponto no qual o canal desaparece de fato move em direção à fonte a medida que a tensão de dreno aumenta. Em outras palavras, o valor de L_1 na Figura 31(b) varia com V_{DS} por alguma quantidade. Chamada de modulação do comprimento de canal e ilustrada na Figura 40, esse fenômeno gera uma grande corrente de dreno à medida que V_{DS} aumenta, devido $I_D \propto 1/L_1$ na Eq. (32). A modulação do comprimento de canal resulta em uma impedância de saída finita dada pelo inverso da inclinação $I_D - V_{DS}$ na Figura 40.

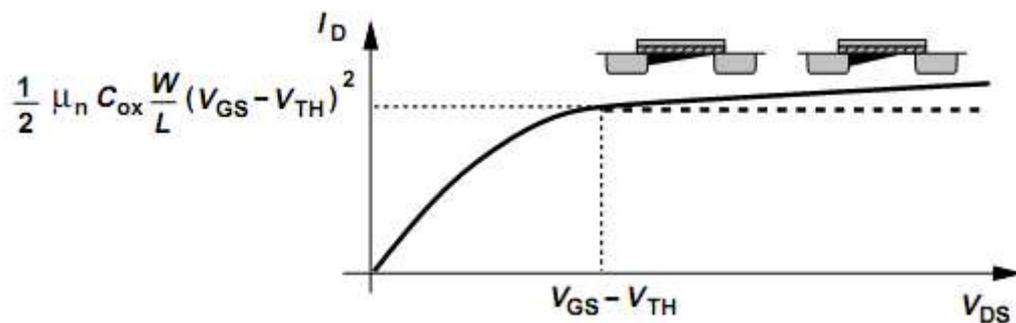


Figura 40 – Variação de I_D na região de saturação

Para calcular-se a modulação do comprimento de canal, assume-se que L é constante, mas multiplica-se o lado direito da Eq. (32) por um termo de correção:

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2 (1 + \lambda V_{DS}) \quad (33)$$

onde λ é chamada de coeficiente de modulação da largura de canal. Enquanto somente uma aproximação, essa dependência linear de I_D sob V_{DS} ainda fornece uma grande quantidade de observações nas implicações de projeto do circuito de modulação do comprimento de canal.

A quantidade de modulação encontra-se sob controle do projetista. Isso porque λ é inversamente proporcional a L : para um longo canal, a mudança relativa em L (e, assim, em I_D) para uma dada mudança em V_{DS} é menor – Figura 41.

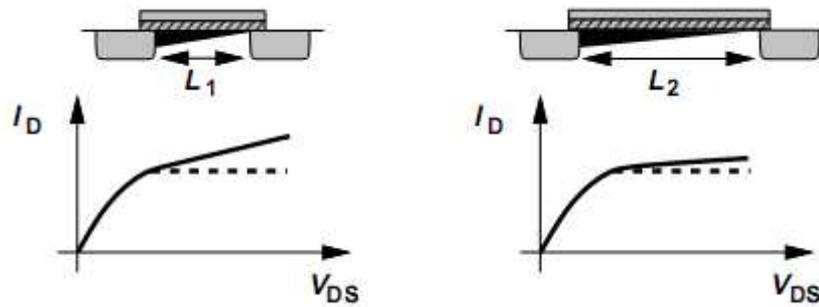


Figura 41 – Modulação do comprimento de canal

3.5 Transcondutância do Transistor MOS

Como uma fonte de corrente controlada por tensão, um transistor MOS pode ser caracterizado por sua transcondutância:

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \quad (34)$$

Essa quantidade serve como medida da “força” do dispositivo: um alto valor corresponde a uma grande mudança na corrente de dreno para uma dada mudança em V_{GS} . Usando a Eq. (32) para a região de saturação, tem-se:

$$g_m = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH}) \quad (35)$$

Concluindo que: (1) g_m é linearmente proporcional a W/L para um dado $V_{GS} - V_{TH}$; e (2), g_m é linearmente proporcional a $V_{GS} - V_{TH}$ para um dado W/L . Também substituindo a $V_{GS} - V_{TH}$ da Eq. (32), obtém-se:

$$g_m = \sqrt{2\mu_n C_{ox} \frac{W}{L} I_D} \quad (36)$$

Isso é: (1) g_m é proporcional a $\sqrt{W/L}$ para um dado I_D ; e (2), g_m é proporcional a $\sqrt{I_D}$ para um dado W/L . Além do mais, dividindo-se a Eq. (35) pela Eq. (32), dá:

$$g_m = \frac{2I_D}{V_{GS} - V_{TH}} \quad (37)$$

Sendo que: (1) g_m é linearmente proporcional a I_D para um dado $V_{GS} - V_{TH}$; e (2), g_m é inversamente proporcional a $V_{GS} - V_{TH}$ para um dado I_D .

A Tabela 3 sumariza essas dependências, que se mostram críticas no entendimento na tendência de desempenho dos dispositivos MOS. Dentre essas três expressões para g_m , a Eq. (36) é a mais frequentemente usada, devido I_D pode ser predeterminada pelos requerimentos de dissipação de potência.

Tabela 3 – Várias dependências de g_m

W/L Constante $V_{GS} - V_{TH}$ Variável	W/L Variável $V_{GS} - V_{TH}$ Constante	W/L Variável $V_{GS} - V_{TH}$ Constante
$g_m \propto \sqrt{I_D}$	$g_m \propto I_D$	$g_m \propto \sqrt{W/L}$
$g_m \propto \frac{1}{V_{GS} - V_{TH}}$	$g_m \propto \frac{1}{W/L}$	$g_m \propto \frac{1}{V_{GS} - V_{TH}}$

3.6 Efeitos de Segunda Ordem

3.6.1 Efeito de Corpo

No decorrer do estudo visto até o momento, assumiu-se que tanto a fonte quanto o substrato estavam ligados ao terra (0 V). Porém, essa condição não necessita permanecer em todos os circuitos. Por exemplo, se o terminal de fonte eleva-se a uma tensão positiva enquanto o substrato permanece em zero, então a junção fonte-substrato permanece polarizada reversamente e o dispositivo ainda opera propriamente.

A Figura 42 ilustra esse caso. O terminal de fonte é ligado a um potencial V_S com relação ao terra, enquanto que o substrato é aterrado através de um contato P⁺⁶. A linha

⁶ A parte de P⁺ é necessária para alcançar um contato ôhmico com baixa resistência.

tracejada adicionada ao símbolo do transistor indica o terminal de substrato. Denota-se a diferença de tensão entre a fonte e o substrato por V_{SB} .

Um interessante fenômeno ocorre à medida que a diferença de potencial fonte-substrato parti do zero: a tensão de *threshold* do dispositivo muda. Em particular, conforme a fonte torna-se mais positiva com relação ao substrato, V_{TH} aumenta. Chamado de efeito de corpo, esse fenômeno é formulado como:

$$V_{TH} = V_{TH0} + \gamma \left(\sqrt{|2\phi_F + V_{SB}|} - \sqrt{|2\phi_F|} \right) \quad (38)$$

Onde V_{TH0} denota a tensão de *threshold* com $V_{SB} = 0$ e γ e ϕ_F são parâmetros dependentes da tecnologia.

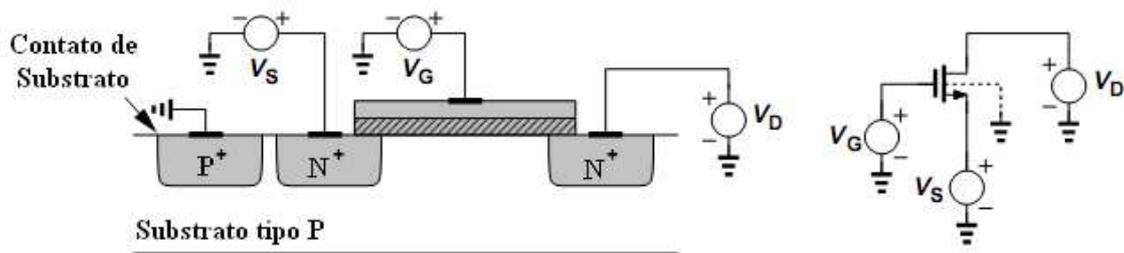


Figura 42 – Efeito de corpo

3.6.2 Condução de *Subthreshold*

A derivação da característica $I-V$ do transistor MOS assumiu que o transistor liga abruptamente conforme V_{GS} atinja V_{TH} . Na realidade, a formação do canal é um efeito gradual e o dispositivo conduz uma pequena corrente para $V_{GS} < V_{TH}$. Chamada de condução de *subthreshold*, esse efeito torna-se um caso importante nos dispositivos MOS modernos e é estudado em textos avançados.

3.7 Transistor PMOS

Ilustrado na Figura 43(a), mudando-se as polaridades dos dopantes do substrato e das áreas de fonte/dreno, resulta em um dispositivo PMOS. O canal agora consiste de lacunas e é formado se a tensão de porta está abaixo do potencial de fonte por uma tensão de *threshold*. Isso é, para ligar o dispositivo, $V_{GS} < V_{TH}$, onde V_{TH} é negativo. O transistor opera na região de triodo se a tensão de dreno é próxima à tensão de fonte, que se aproxima da saturação à medida que V_D decai para $V_{GS} - V_{TH} = V_{GS} - |V_{TH}|$. A Figura 43(b) conceitualmente ilustra as tensões de porta-dreno requeridas para cada região de operação.

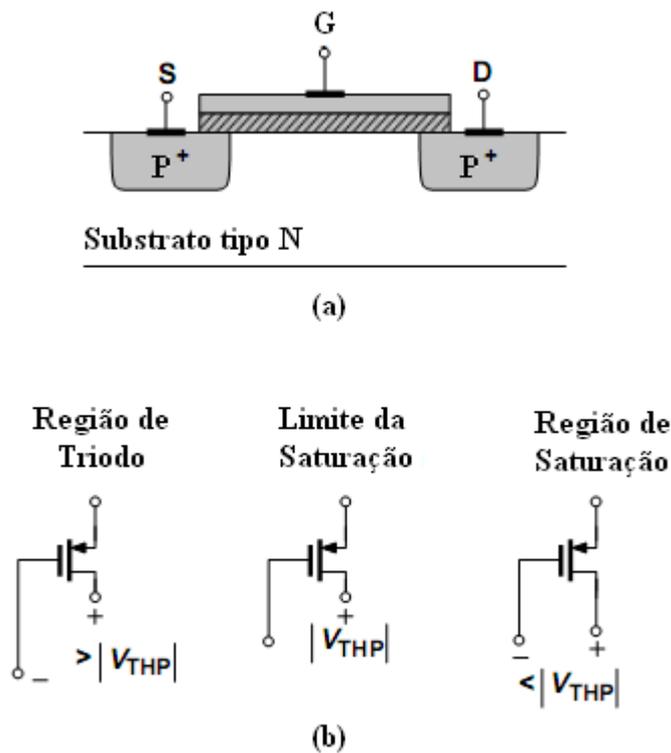


Figura 43 – (a) Estrutura de um dispositivo PMOS (b) Ilustração das regiões de triodo e de saturação baseadas nas tensões de porta e dreno.

4 Modelos de Dispositivo

O SPICE fornece uma larga variedade de modelos de transistores MOS com várias trocas entre complexidade e precisão. Os modelos Nível 1 e Nível 3 foram historicamente importantes, mas eles não são largamente adequados aos modelos de transistores modernos muito pequenos. Os modelos BSIM são mais precisos e são, atualmente, os mais usados. Algumas empresas usam seus próprios modelos. Essa seção descreve brevemente as principais características de cada desses modelos.

4.1 Modelo Nível 1

O SPICE nível 1, ou modelo *Shichman-Hodges*, aumenta com a modulação do comprimento de canal e o efeito de corpo. O modelo de corrente básico é:

$$I_D = \begin{cases} 0 & V_{GS} < V_{TH} \\ \frac{KP}{L_{eff}} \frac{W_{eff}}{L_{eff}} (1 + LAMBDA V_{DS}) \left(V_{GS} - V_{TH} - \frac{V_{DS}}{2} \right) V_{DS} & V_{DS} > V_{GS} - V_{TH} \\ \frac{KP}{2} \frac{W_{eff}}{L_{eff}} (1 + LAMBDA V_{DS}) (V_{GS} - V_{TH})^2 & V_{DS} < V_{GS} - V_{TH} \end{cases} \quad (39)$$

Os parâmetros do modelo SPICE são dados todos em maiúsculos.

A tensão de *threshold* é modulada pela tensão de fonte-substrato, V_{SB} , através do efeito de corpo. Para V_{SB} não negativo, a tensão de *threshold* é:

$$V_{TH} = V_{TO} + GAMMA \left(\sqrt{PHI + V_{SB}} - \sqrt{PHI} \right) \quad (40)$$

A capacitância de porta é calculada da espessura de óxido TOX.

Os modelos nível 1 são úteis para ensinar, porque eles são facilmente correlacionados com a análise feita a mão, mas são muito simplistas para projetos modernos. A

“encaixotados” (*binned*) com diferentes modelos abrangendo diferentes faixas de comprimento e larguras especificados pelos parâmetros LMIN, LMAX, WMIN e WMAX.

À medida que os modelos BSIM vão se tornando mais complicados, é impraticável derivar equações explícitas para o atraso de propagação, *threshold* de chaveamento, margens de ruído, etc. Porém, não é difícil achar essas propriedades através da simulação do circuito.

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
UNIDADE ACADÊMICA DE ENGENHARIA ELÉTRICA
ESTRUTURA E CONCEPÇÃO DE CIRCUITOS INTEGRADOS

MÓDULO 3

Projeto de Leiaute

Campina Grande, Fevereiro de 2010

Sumário

1	Introdução	3
2	Tecnologia de Processo CMOS	4
2.1	Formação da Pastilha.....	4
2.2	Fotolitografia	5
2.3	Formação do Poço e do Canal	7
2.4	Dióxido de Silício (SiO ₂).....	9
2.5	Isolamento	10
2.6	Óxido de Porta.....	12
2.7	Formação da Porta e Fonte/Dreno	12
2.8	Contatos e Metalização.....	15
2.9	Passivação.....	17
3	Confiabilidade.....	18
3.1	Autoaquecimento.....	18
3.2	Portadores Quentes	18
3.3	<i>Latchup</i>	19
4	Regras de Projeto de Leiaute	21
4.1	Largura Mínima.....	22
4.2	Espaçamento Mínimo	22
4.3	Delimitação Mínima.....	23
4.4	Extensão Mínima.....	24
4.5	Efeito Antena.....	25
4.6	Regras de Leiaute Escalonável MOSIS.....	26
5	Técnicas de Leiaute Analógico.....	28
5.1	Transistores Seccionados.....	28
5.2	Simetria.....	29
5.3	Distribuição de Referência	33
6	Anexo A – Regras de Leiaute Escalonável MOSIS (Rev 8.00)	35
6.1	Well	35
6.2	Ativa	35

6.3	Poli.....	36
6.4	Rigorouso	37
6.5	Contato Simples para Poli	38
6.6	Contato Alternativo para Poli.....	38
6.7	Contato Simples para Ativa.....	39
6.8	Contato Alternativo para Ativa.....	39
6.9	Metal 1.....	40
6.10	Via 1	41
6.11	Metal 2.....	42
6.12	Via 2 (Opção Metal Triplo).....	42
6.13	Metal 3 (Opção Metal Triplo)	43

1 Introdução

Leiaute é a parte do projeto da fabricação de circuito integrado (CI), no qual a geometria dos elementos de circuito e as ligações são definidas. Pode-se dizer mais: é o processo de criar uma representação física fiel de um desenho de engenharia que está de acordo com as limitações impostas pelo processo de fabricação, o fluxo de projeto e os requerimentos de desempenho que se mostraram possíveis por meio de simulação.

Um bom entendimento do processo de fabricação do dispositivo prova ser essencial no projeto e leiaute de CI's. Em muitos casos, se o projetista entende o processo físico, ele irá compreender as razões para o entendimento dos fundamentos das regras de leiaute e usar esse conhecimento para criar um projeto otimizado. O entendimento dos passos da fabricação também é importante quando depurando algumas falhas difíceis.

Além do processo de fabricação, os conceitos das regras de leiaute e sua relação com os microcircuitos serão enfatizados.

2 Tecnologia de Processo CMOS

Os passos básicos envolvidos no processo de um CI CMOS (*Complementary MOS*) serão apresentados. As principais tecnologias CMOS são:

- Processo *N-well*;
- Processo *P-well*;
- Processo *Twin-well*; e,
- Processo *Triple-well*.

Em um processo *P-well*, os transistores NMOS são construídos em um poço tipo P (*P-well*) e os transistores PMOS são construídos em um substrato tipo N. O processo *P-well* foi usado para melhorar o desempenho do transistor PMOS.

Técnicas melhoradas permitiram que bons transistores PMOS fossem construídos em um poço tipo N (*N-well*) e excelentes transistores NMOS fossem construídos em um substrato tipo P num processo *N-well*.

Um processo *twin-well* permite a otimização de cada tipo de transistor. Um terceiro poço pode ser adicionado para criar um processo *triple-well*. O processo *triple-well* surgiu para fornecer bom isolamento entre blocos analógicos e digitais em CI's de sinal mixado (*mixed-signal*). É também usado para isolar memória dinâmica de alta densidade do circuito lógico.

Nesta seção, apresenta-se um processo CMOS na geração 130 ou 90 nm que é representativo dos processos comerciais correntes.

2.1 Formação da Pastilha

A matéria-prima usada nas fábricas de semicondutores modernas é uma pastilha (*wafer*) de silício, que atualmente varia de aproximadamente 75 mm a 300 mm de diâmetro e menos de 1 mm de espessura. As pastilhas são cortadas a partir de barras de

silício monocristalino, que foram extraídos de um cadinho⁸ de silício fundido puro. Isso é conhecido como método de *Czochralski* e é correntemente o método mais comum para produzir material monocristalino. Quantidades controladas de impurezas são adicionadas à fusão para conceder o cristal com as propriedades elétricas requeridas.

Um cristal semente (*seed crystal*) é mergulhado na fusão para iniciar o crescimento cristalino. A barra de silício assume a mesma orientação cristalina da semente. Um radiador de grafite aquecido pela indução de rádio frequência rodeia o cadinho contendo a fusão, mantendo a temperatura poucos graus acima do ponto de fusão do silício (1425° C). O ambiente é, tipicamente, de hélio ou de argônio para prevenir a oxidação do silício.

A semente é gradualmente retirada verticalmente da fusão, enquanto sendo girada. O silício fundido liga-se à semente e recristaliza-se enquanto ele é retirado. As taxas de retirada e rotação da semente determinam o diâmetro da barra.

2.2 Fotolitografia

As regiões de dopantes, polissilício, metal e contatos são definidos usando máscaras. Por exemplo, em locais protegidos pela máscara, a implantação de íon não deve ocorrer ou a camada de dielétrico ou metal pode ser deixada intacta. Em áreas onde a máscara está ausente, a implantação pode ocorrer ou o dielétrico ou metal pode ser gravado. A modelagem é conseguida por um processo chamado de fotolitografia ou litografia. O método primário para definir as áreas de interesse, isso é, onde se deseja que o material esteja presente ou ausente em uma pastilha, é pelo uso de fotoresistes ou resistes. A pastilha é coberta com o resiste e exposta à iluminação seletiva através da fotomáscara. Após a modelagem inicial do resiste, outras camadas, tais como o de silício policristalino, dióxido de silício ou nitreto de silício, podem ser usadas como máscaras físicas no CI.

Uma fotomáscara é construída com cromo cobrindo um vidro de quartzo. Uma fonte de luz ultravioleta é usada para expor o resiste. A Figura 45 ilustra o processo de fotolitografia. A fotomáscara tem cromo onde a luz deve ser bloqueada. A luz

⁸ Vaso de material resistente ao fogo, usado para fundir minérios e minerais ou para realizar certas operações químicas ou físico-químicas que exigem altas temperaturas.

ultravioleta (UV) incide sobre a máscara e atravessa as aberturas da máscara para expor o resiste revestido sobre a pastilha. Um solvente revelador é então usado para dissolver a solução de resiste não exposta ao UV, deixando ilhas de resistes expostas ao UV. Isso é designado de resiste negativo. Um resiste positivo é inicialmente insolúvel e quando exposto ao UV torna-se solúvel ao solvente. Os resistes positivos proporcionam melhor resolução que os negativos, mas são menos sensíveis a luz. À medida que os tamanhos característicos tornam-se menores, as camadas de resiste têm sido feitas finas. Isso está tornando-os menos robustos e mais sujeitos a falhas. Sucessivamente, isso pode impactar toda a produção de um processo e o custo para produzir o CI.

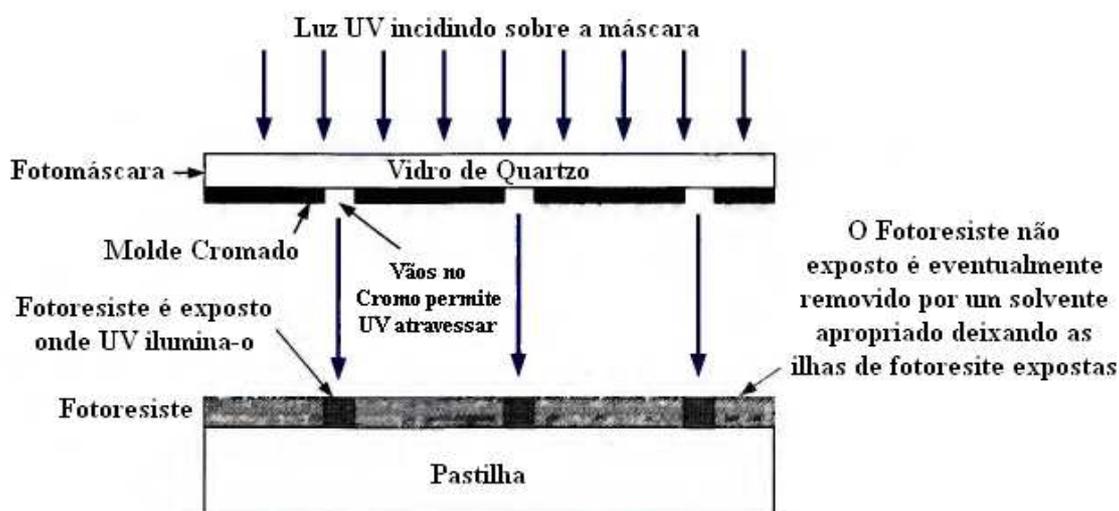


Figura 45 – Fotomáscara com um resiste negativo

O comprimento de onda da fonte de luz influencia na característica do tamanho mínimo que pode ser “impresso”. Na década de 80, lâmpadas de mercúrio com comprimentos de onda de 436 nm ou 365 nm eram usadas. Na geração do processo de 250 nm, laser de *excimer*⁹ com 248 nm foi adotado e usado até diminuir para 180 nm. Correntemente, laser de fluoreto de argônio de 193 nm é usado para camadas críticas, diminuindo para um tamanho de até 90 nm. As camadas críticas são aquelas que definem o comportamento do dispositivo. Exemplo: as máscaras da porta (polissilício), da fonte e do dreno (difusão), do primeiro metal e de contato.

⁹ Uma molécula excitada instável que é formada pela combinação de duas moléculas menores e dissocia rapidamente com a emissão de radiação.

Comprimentos de onda semelhantes a ou maiores que o tamanho característico causam distorção nos moldes expostos sobre o resiste. Técnicas de aprimoramento da resolução pré-compensam essa distorção à medida que o molde desejado é obtido. Essas técnicas envolvem modificar a amplitude, fase ou direção da luz recebida. As terminações de uma linha em um leiaute recebem menos luz que o centro, causando uma exposição não uniforme. A correção de proximidade ótica faz pequenas mudanças para os moldes sobre as máscaras compensar essas distorções locais. Múltiplas linhas que estão em paralelo em uma máscara comportam-se como rede de difusão. Máscaras de deslocamento de fase variam a espessura da máscara para mudar a fase tal que a luz das linhas adjacentes esteja fora de fase e cancele-se onde nenhuma luz é desejada. Iluminação fora de eixo pode também fornecer contraste para certos tipos de moldes repetitivos e densos. Usando essas técnicas, a resolução pode ser estendida a um oitavo do comprimento de onda da luz.

2.3 Formação do Poço e do Canal

Variadas proporções de impurezas doadoras e aceitadoras podem ser alcançadas usando técnicas de epitaxia, deposição ou implantação.

A epitaxia envolve o crescimento de um filme monocristalino sobre a superfície do silício (que já é um monocristalino), sujeitando-se a superfície da pastilha de silício a uma temperatura elevada e uma fonte de material dopante. Pode ser usada para produzir uma camada de silício com menos defeitos que a superfície da pastilha original e também pode ajudar a prevenir *latchup*¹⁰.

A deposição envolve colocar material dopante sobre a superfície de silício e então levá-lo para o interior do substrato usando uma etapa de difusão térmica. Isso pode ser usada para construir junções profundas. Uma etapa chamada deposição de vapor químico (CVD – *Chemical Vapor Deposition*) pode ser usada para a deposição. Como seu nome sugere, a CVD ocorre quando gases aquecidos reagem na vizinhança da pastilha e produz um produto que é depositado sobre a superfície de silício. Também é usada para depositar, posteriormente, finos filmes de material no processo CMOS.

¹⁰ Ver Tópico 3.3

A implantação iônica envolve submeter o substrato de silício a átomos doadores ou aceitadores altamente energizados. Quando esses átomos chocam sobre a superfície de silício, eles movem-se por debaixo da superfície, formando regiões com variadas concentrações de dopagem. Sob temperatura elevada ($> 800^\circ \text{C}$), a difusão ocorre entre regiões de silício que tem diferentes densidades de impurezas, com impurezas tendendo a difundir-se de áreas de alta concentração para áreas de baixa concentração. Assim, é importante manter as etapas restantes do processo na temperatura mais baixa possível, uma vez que as áreas dopadas tenham sido colocadas no lugar. Porém, uma etapa de recozimento a alta temperatura é muitas vezes realizado após a implantação iônica para redistribuir os dopantes mais uniformemente. A implantação iônica é o método padrão de implante de poço e fonte/dreno usado atualmente.

O primeiro passo na maioria dos processos CMOS é definir as regiões de poço. Em um processo *triple-well*, um *N-well* profundo é primeiro induzido no substrato tipo P, usualmente usando implantação iônica de alta energia (da ordem de mega elétron-volt – MeV) em contraste a operação de difusão térmica. Isso evita o ciclo termal, isso é, as pastilhas não tem que ser elevadas, significativamente, na temperatura, o que melhora o rendimento e a confiabilidade. Uma implantação de 2 a 3 MeV pode produzir um *N-well* de profundidade de 2,5 a 3,5 μm . Tal poço tem uma concentração máxima de dopante sob a superfície e por essa razão é chamado de um poço retrógrado. Isso pode elevar o desempenho do dispositivo pela determinação das características melhoradas do *latchup*. Um resiste grosso (3,5 a 5,5 μm) tem sido usado para bloquear a implantação de alta energia onde nenhum poço deve ser formado. Resistes grossos e implantes profundos necessariamente conduzem a dimensões características razoavelmente grosseiras para os poços, comparado ao tamanho característico mínimo. Regiões *P-well* e *N-well* rasas são então implantadas. Após os poços terem sido formados, os níveis de dopagem podem ser ajustados (chamado de implante de limiar – *threshold implant*) para configurar as tensões de *threshold* desejadas para ambos os transistores NMOS e PMOS. Para uma dada porta e substrato, a tensão de limiar (V_{TH}) depende somente do nível de dopagem no substrato (N_A), da espessura de óxido (t_{ox}) e da carga de superfície (Q_{fc}). O implante pode afetar ambos N_A e Q_{fc} e, portanto, V_{TH} . A Figura 46 mostra uma estrutura típica *triple-well*. Como discutido, o transistor NMOS está situado no P-well localizado no N-well profundo. O transistor PMOS está localizado no N-well raso (normal).

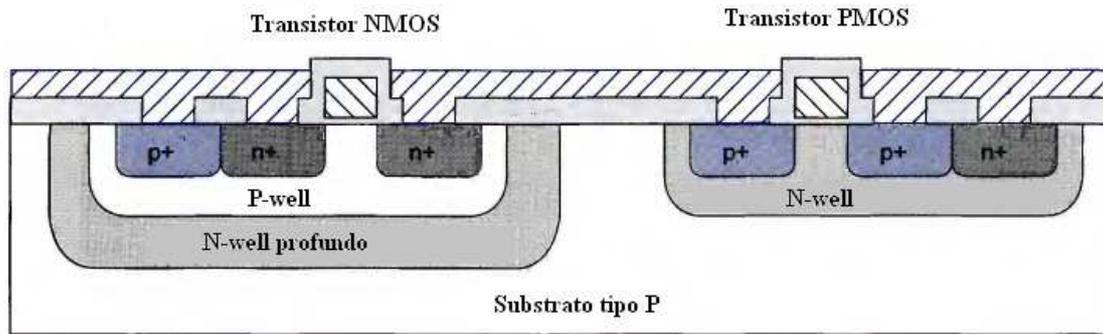


Figura 46 – Estrutura do poço no processo *triple-well*

2.4 Dióxido de Silício (SiO_2)

Várias das estruturas e técnicas de fabricação usadas para criar circuitos integrados de silício conta com a propriedades do SiO_2 . Então, a produção confiável de SiO_2 é extremamente importante. Várias espessuras de SiO_2 podem ser requeridas, dependendo do processo em particular. Óxidos finos são requeridos para portas de transistores. Óxidos grossos podem ser exigidos para dispositivos de alta tensão, enquanto até mesmo camadas de óxido mais grosso poder ser exigidos para garantir que os transistores não sejam formados acidentalmente no silício debaixo das linhas de polissilício.

A oxidação do silício é conseguida pelo aquecimento das pastilhas de silício em uma atmosfera oxigenada. Algumas técnicas comuns são:

- Oxidação úmida: quando a atmosfera oxigenada contém vapor de água;
- Oxidação seca: quando a atmosfera oxigenada é formada somente por oxigênio; e,
- Deposição de camada atômica: um processo no qual uma camada química fina (material A) é ligada a uma superfície e então uma substância química (material B) é introduzida para produzir a camada fina requerida (isso também pode ser usado para outras camadas, como a dielétrica e a de metal). Esse processo é então repetido, formando camada por camada.

O processo de oxidação normalmente consome silício. Desde que SiO_2 tem aproximadamente duas vezes o volume do silício, a camada de SiO_2 cresce quase que

igualmente em ambas as direções verticais. Assim, após o processo, o SiO₂ projeta acima e abaixo da superfície de silício não oxidada.

2.5 Isolamento

Embora não seja imediatamente óbvio, os dispositivos, em um processo CMOS, necessitam ser isolados um dos outros tal que eles não tenham interações inesperadas. A fonte e o dreno dos transistores formam junções PN polarizadas reversamente com o substrato ou poço, isolando-os de seus vizinhos. Em seguida, a formação de quaisquer canais parasitas MOS devem ser prevenidos. Isso é comumente conseguido usando um óxido de porta (*gate oxide*) fino para transistores e um óxido de campo (*field oxide*) muito mais grosso em outro lugar. O óxido grosso aumenta a tensão de limiar a um valor acima da tensão de alimentação e, assim, previne um canal de se formar no substrato, a menos que haja uma condição de sobretensão (atualmente, esses dispositivos de campo são usados para proteção de entradas e saídas – E/S). Além disso, para usar o óxido grosso, o substrato nas áreas onde os transistores não são requeridos pode ser implantado mais adiante com dopantes para criar uma difusão de parada de canal (*channel stop*). O implante aumenta a concentração de impurezas no substrato, que por sua vez aumenta a tensão de limiar e previne a inversão de um canal indesejado.

Historicamente, a oxidação local de silício (LOCOS – *Local Oxidation of Silicon*) foi usada para produzir espessuras variantes de óxido. Um problema comum dos processos baseados em LOCOS era a transição entre óxido grosso e fino, que prolongava algumas distâncias lateralmente devido à maneira como o óxido crescia. Isso limitou a densidade de agrupamento dos transistores. A etapa de isolamento, que é usada para conseguir o isolamento entre dispositivos em processos a e abaixo de 180 nm, é para formar fossos separados de SiO₂ que circunda as áreas ativas. Tipicamente, fossos em um processo de 90 nm podem ter 140 nm de largura e 400 nm de profundidade. Isso é chamado de isolamento de fossa rasa (STI – *shallow trench isolation*). A STI – Figura 47 – inicia com um óxido de ligação e uma camada de nitreto de silício, que age como as camadas de mascaramento. As aberturas no óxido de ligação são então usadas para corroer a região de poço ou substrato. Um óxido de revestimento é então cultivado para cobrir o silício exposto. As fossas são preenchidas com SiO₂ usando deposição de vapor químico que não consome o silício de base. O óxido de base

e o nitreto são removidos e um polidor mecânico químico (CMP – *Chemical Mechanical Polishing*) é usado para planar a estrutura. O CMP, como seu nome sugere, combina uma ação de polimento mecânico no qual a pastilha em rotação é contatada por uma cabeça de polimento estacionária, enquanto uma mistura abrasiva é aplicada. CMP é usado para conseguir superfícies planas, que são de importância central em processos modernos com várias camadas.

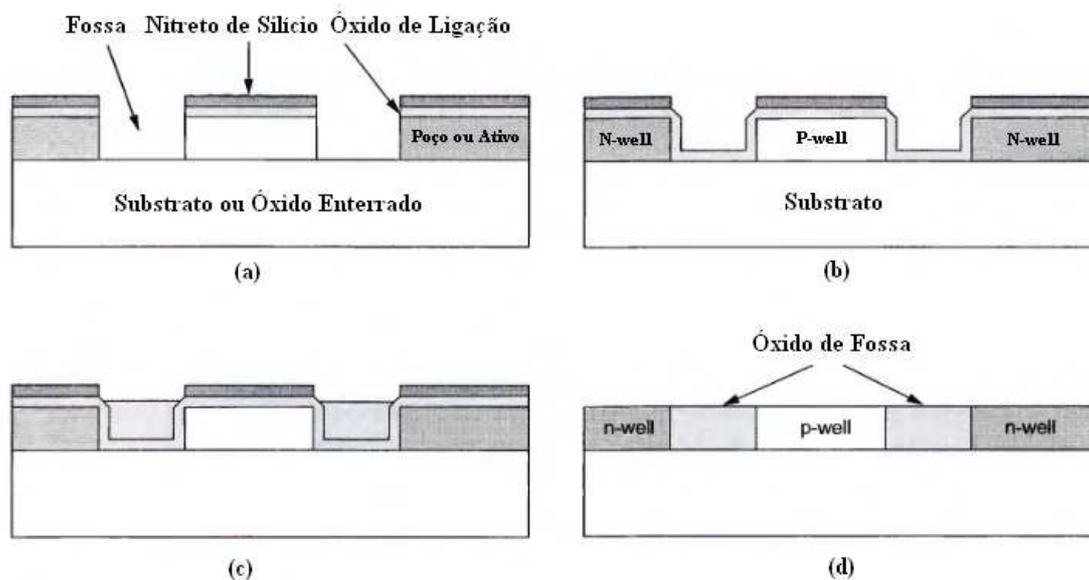


Figura 47 – Isolamento de fossas rasas (a) Fossa entalhada (b) Oxidação de revestimento (c) Fossa preenchida com dielétrico (d) CMP para planar

Da perspectiva do projetista, a presença de um *N-well* profundo e/ou isolamentos escavados torna fácil isolar as porções sensíveis ao ruído (circuitos analógicos ou memória) em um CI de seções digitais. O isolamento também permite que transistores NMOS e PMOS sejam colocados próximos, devido o isolamento fornecer uma alta tensão de ruptura de fonte/dreno. A tensão de ruptura deve exceder a tensão de alimentação, assim as junções não rompem durante operação normal. Fossas profundas aumentam a tensão de ruptura.

Poços são definidos por máscaras separadas. No caso de um processo *twin-well*, somente uma máscara necessita ser definida, porque o outro poço é por definição seu complemento. Nos processos *triple-well* tem que definir no mínimo duas máscaras, uma para o poço profundo e outra para ou o *N-well* ou o *P-well*.

2.6 Óxido de Porta

O próximo passo no processo é para formar o óxido de porta para os transistores. Esse é mais comumente na forma de dióxido de silício (SiO_2).

No caso das regiões de fonte/dreno definida por STI, o óxido de porta é cultivado no topo da estrutura planar que ocorre no estágio mostrado na Figura 47(d). Isso é mostrado na Figura 48. A estrutura de óxido é chamada de pilha de porta (*gate stack*). Esse termo é devido o corrente processo raramente usar um óxido de porta puramente de SiO_2 , preferindo-se produzir uma pilha que consiste de um pouco de camadas atômicas, cada uma de 3 a 4 Å de espessura, de SiO_2 revestido com poucas camadas de um óxido oxinitroso (*oxynitrided oxide*). A presença do nitrogênio aumenta a constante dielétrica, que diminui a espessura efetiva do óxido (EOT – *Effective Oxide Thickness*). Isso significa que para uma dada espessura de óxido, ele atua como um óxido fino, podendo-se usar um óxido mais grosso para melhorar a robustez do processo.

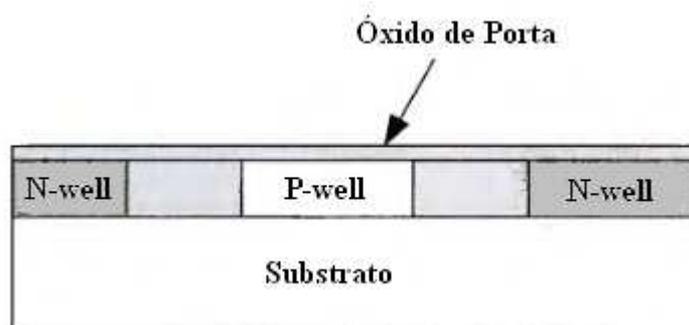


Figura 48 – Formação do óxido de porta

2.7 Formação da Porta e Fonte/Dreno

Quando o silício é depositado no SiO_2 ou outra superfície sem orientação cristalina, forma-se o silício policristalino, comumente chamado polissilício ou simplesmente poli. Um processo de recozimento é usado para controlar o tamanho do cristal e fornecer a qualidade do polissilício. Polissilício não dopado tem uma alta resistividade. A resistência pode ser reduzida implantando-o com dopantes e/ou combinando-o com um metal. A porta de polissilício serve como uma máscara para permitir alinhamento preciso da fonte e dreno com a porta. Esse processo é chamado de porta de polissilício

autoalinhada (*self-aligned polysilicon gate*). Alumínio não pode ser usado devido ele derreter durante formação da fonte e do dreno.

Os passos para definir a porta e fonte/dreno em uma porta de polissilício autoalinhada são as seguintes:

- O óxido de porta cresce em todos os lugares em que os transistores são exigidos – Figura 49(a). Nos outros lugares haverá óxido grosso;
- Polissilício depositado sobre o CI – Figura 49(b);
- Polissilício moldado (ambas as portas e interconexões) – Figura 49(c);
- Óxido de porta exposta a corrosão, ou seja, a área do óxido de porta que não foi coberta pelo polissilício. Neste ponto, o CI tem janelas inativas para o poço ou substrato em todos os lugares onde uma difusão fonte/dreno é requerida – Figura 49(d); e,
- Implante regiões de fonte/dreno PMOS e NMOS – Figura 49(e).

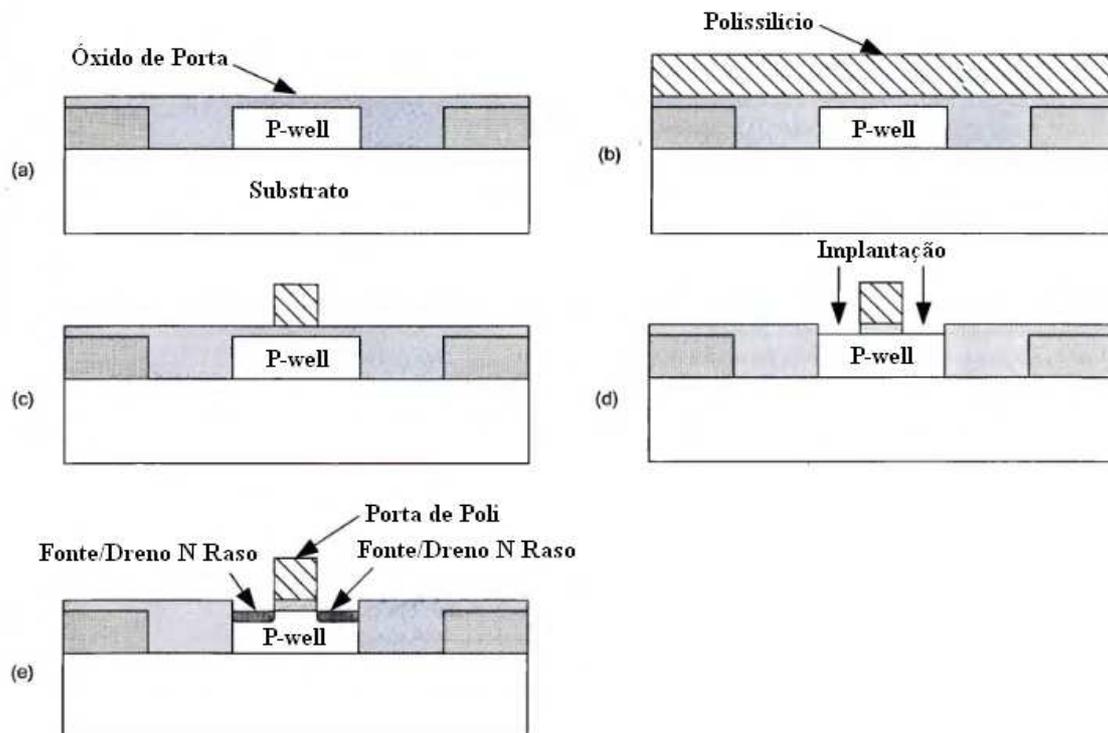


Figura 49 – Definição da fonte/dreno raso e da porta

O implante de fonte/dreno é relativamente lento, tipicamente na ordem de 10^{18} a 10^{20} átomos de impureza por cm^3 . Tal como uma estrutura de dreno dopado levemente

(LDD – *lightly doped drain*) reduz o campo elétrico na junção de dreno, o qual aumenta a imunidade do dispositivo a danos de elétron quente. Os implantes de dreno dopado levemente são rasos, assim eles exibem baixa capacitância, mas alta resistência. Isso reduz o desempenho do dispositivo um pouco, devido à resistência em série com o transistor. Conseqüentemente, implantes de fonte/dreno dopados mais fortemente e profundos são necessários para fornecer dispositivos que combinam supressão de elétron quente com baixa resistência de fonte/dreno. Um separador de nitreto de silício (Si_3N_4) ao longo da borda da porta serve como uma máscara para definir a localização dessa difusão profunda, como mostra a Figura 50(a).

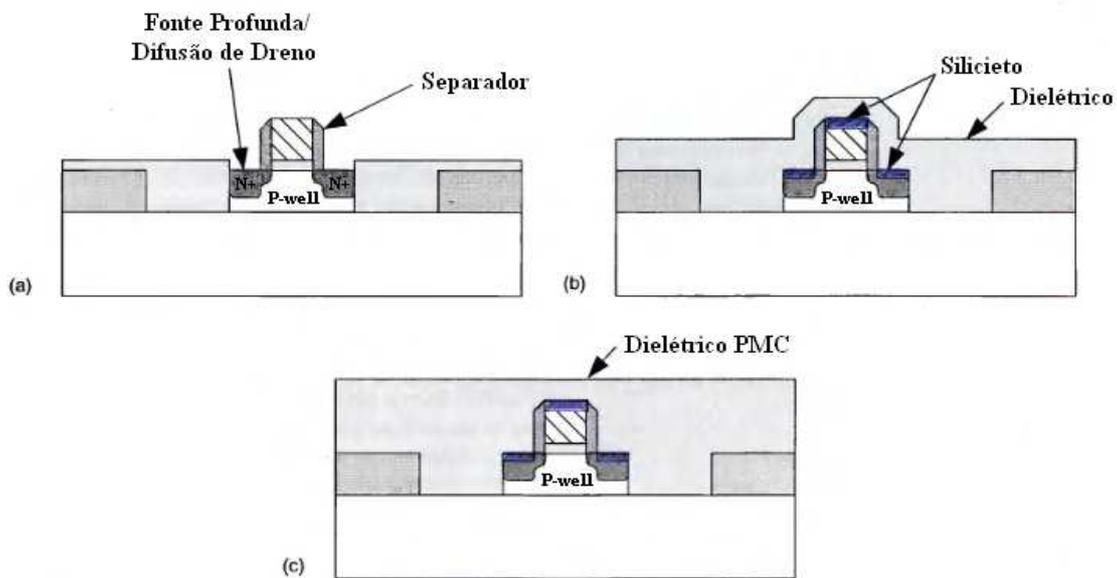


Figura 50 – Estrutura do dreno dopado levemente

Como mencionado, a porta de polissilício e a difusão fonte/dreno tem alta resistência devido à resistividade do silício e suas dimensões extremamente pequenas. Processos modernos formam uma camada de superfície de um metal refratário sobre o silício para reduzir a resistência. Um metal refratário é um com um alto ponto de fusão que não será danificado durante subsequente processamento. Tântalo (Ta), molibdênio (Mo), titânio (Ti) ou cobalto (Co) são comumente usados. O metal é depositado sobre o silício, especificadamente sobre a porta de polissilício e/ou as regiões de dreno/fonte. Uma camada de siliceto (*silicide*) é formada quando as duas substâncias reagem a elevadas temperaturas. Em um processo de polycido (*polycide*), somente o polissilício recebe a camada de siliceto. Em um processo de siliceto, ambas as portas de

polissilício e as regiões de dreno/fonte recebem a camada. Esse processo diminui a resistência da interconexão de polissilício e/ou a fonte e dreno.

A Figura 50(b) mostra a estrutura resultante com a porta e as regiões de fonte/dreno com a camada de siliceto. Além disso, o SiO_2 ou um dielétrico alternativo tem sido usado para cobrir todas as áreas anteriores para os próximos passos de processo.

A Figura 50(c) mostra a estrutura quando o CMP foi utilizado. Obter um acabamento muito plano permite que as camadas sejam empilhadas sem ocorrer os problemas de metal tem para atravessar rápidas transições em superfícies altas, como mostra a Figura 50(b), o que pode conduzir a quebras e um excesso de regras de projeto relacionadas às bordas de metal.

Polissilício por cima da difusão normalmente forma uma porta de transistor, assim uma curta ligação de metal é necessária para conectar um nó de saída de difusão para uma entrada de polissilício. Alguns processos adicionam uma região de contato ao processo de modo que a camada de polissilício pode conectar diretamente à difusão. Tais ligações de polissilício são chamadas interconexões locais (*local interconnect*).

2.8 Contatos e Metalização

Cortes de contato são feitos para fonte, dreno e porta de acordo com a máscara de contato. São buracos entalhados no dielétrico ao final do passo de fonte/dreno. O alumínio (Al) é comumente usado para as ligações, mas o tungstênio (W) pode ser usado como um “plugue” para preencher os cortes do contato. Em alguns processos, o tungstênio pode também ser usado como uma camada de interconexão local.

Metalização é o processo de construir ligações para conectar os dispositivos. Como mencionado, uma metalização convencional usa alumínio. O alumínio pode ser depositado ou por evaporação ou por pulverização. A evaporação é realizada pela passagem de uma alta corrente elétrica por meio de um fio de alumínio grosso em uma sala a vácuo. Alguns dos átomos de alumínio são vaporizados e depositados sobre a pastilha. Uma forma melhorada de evaporação que tolera menos da contaminação foca em um feixe de elétron em um recipiente de alumínio para evaporar o metal. A pulverização é realizada pela geração de um gás plasma pela ionização de um gás inerte usando radio frequência ou campo elétrico DC. Os íons são focados sobre um alumínio e o plasma emite átomos de metal, que são então depositados sobre a pastilha.

Gravação a seco ou úmida pode ser usada para remover metal indesejado. Uma solução de piranha (*piranha solution* ou *piranha etch*) é uma mistura 3:1 para 5:1 de ácido sulfúrico e peróxido de hidrogênio que é usado para limpar pastilhas de contaminação orgânica e metálica ou resiste após moldura de um metal. Corrosão por plasma é um processo de gravação a seco com flúor ou cloro em gás usado para os passos de metalização. O plasma carrega os íons do gás corrosivo, que são atraídos para a superfície do silício propriamente carregada. Vários perfis de gravuras distintos podem ser conseguidos usando corrosão por plasma. O resultado da moldagem do contanto e da metalização é mostrado na Figura 51.

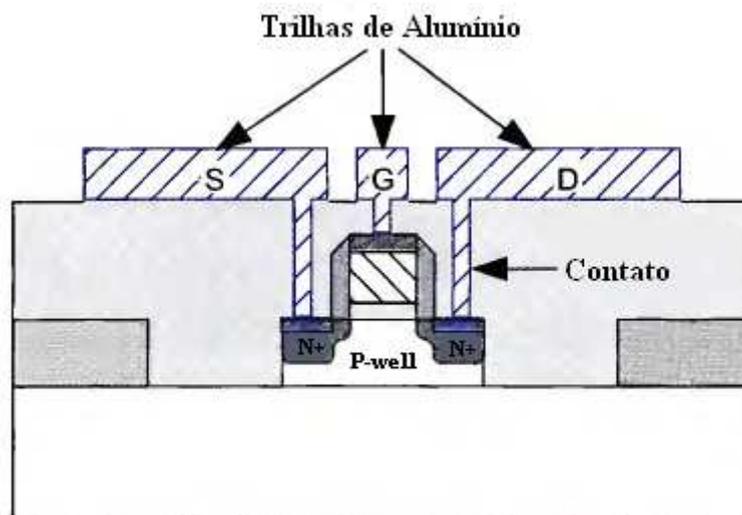


Figura 51 – Metalização de alumínio

Subsequentes vias intermetal e metalização são então aplicadas. Alguns processos tentam usar um esquema de metalização não uniforme de no mínimo nível 2 a $n-1$, onde n é o maior nível de metal. O maior nível é normalmente uma camada grossa para uso em distribuição de alimentação e como tal tem restrições de largura e espaçamento relaxados. Um corte transversal da metalização é mostrado na Figura 52.

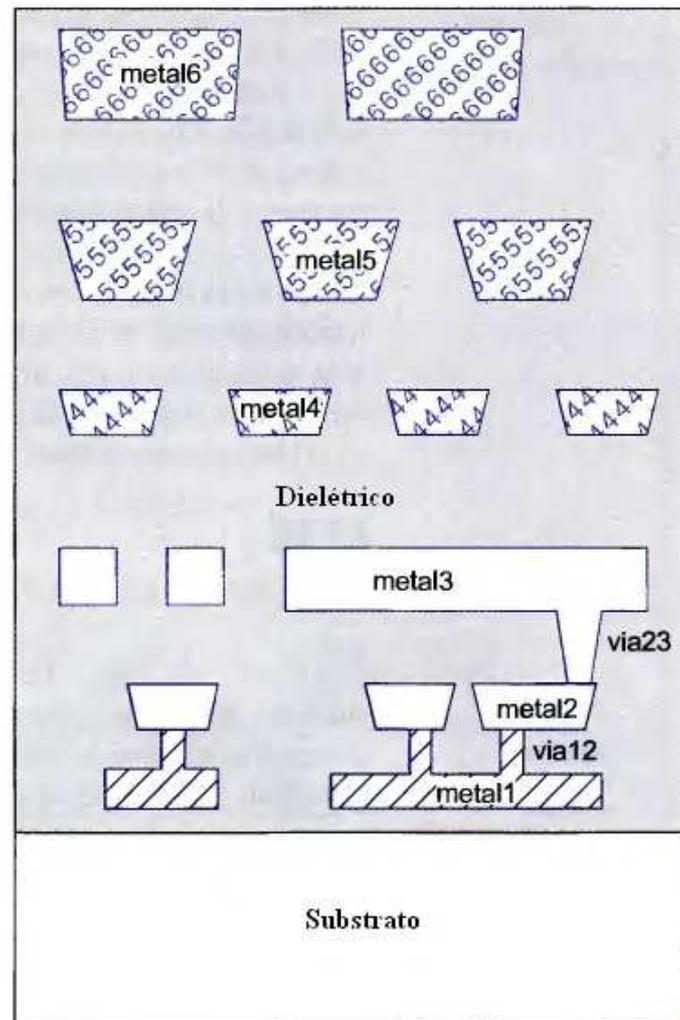


Figura 52 – Corte transversal de uma metalização típica

2.9 Passivação

A etapa final do processo é adicionar uma camada de vidro protetora chamada passivação, que previne a entrada de contaminantes. Aberturas na camada de passivação permite a conexão de pernas de E/S e pontos de prova se necessários.

3 Confiabilidade

Projetar um CI CMOS confiável envolve entender e indicar os modos de falha em potencial. Esta seção indica alguns poucos problemas de confiabilidade que causa aos CI's falhar permanentemente.

3.1 Autoaquecimento

A densidade de corrente das linhas é limitada pelo autoaquecimento. Correntes elevadas dissipam potência na linha, o qual eleva sua temperatura ambiente. As ligações aquecidas exibem grande resistência e atraso. O autoaquecimento é dependente do valor RMS da densidade de corrente.

Este problema limita a densidade de corrente RMS em linhas de sinal bidirecionais. Porém, não se podem negligenciar as correntes unidirecionais que fluem através das linhas conectando transistores NMOS e PMOS.

3.2 Portadores Quentes

À medida que os transistores chaveiam, alguns portadores com alta energia (“quentes”) podem ser injetados no óxido de porta e torna-se preso lá. O óxido danificado muda as características $I-V$ do dispositivo, reduzindo a corrente no transistor NMOS, enquanto aumenta no transistor PMOS. O dano é maximizado quando a corrente de substrato I_{sub} é grande, o qual ocorre tipicamente quando transistores NMOS estão em saturação enquanto a entrada eleva. Assim, o problema é péssimo para inversores e portas NOR com entradas que elevam rapidamente, saídas excessivamente carregadas e para altas tensões de alimentação.

Portadores quentes causa um desgaste de circuito, como os transistores NMOS que se tornam lentos. Eles podem também causar falhas de amplificadores sensíveis e outros circuitos casados se os componentes casados desgastarem-se diferentemente.

Um mecanismo de envelhecimento é a instabilidade de temperatura de polarização negativa (NBTI – *Negative Bias Temperature Instability*), que conduz a diminuição da

corrente em transistores PMOS à medida que os transistores experimentam altas temperaturas. NBTI resulta de buracos presos no óxido juntamente com a criação de estados de interface. Como os portadores quentes, leva o circuito a falhas de aumento no atraso e casamento pobre.

3.3 *Latchup*

Pouco depois da adoção dos processos CMOS, foi percebida uma curiosa tendência de CI's CMOS desenvolver caminhos de baixa resistência entre V_{DD} e GND , causando uma catastrófica fusão. O fenômeno, conhecido como *latchup*, ocorre quando transistores bipolares parasitas formados pelo substrato, poço e difusão ativam. Com o avanço do processo e procedimentos de leiaute adequados, os problemas de *latchup* podem ser facilmente evitados.

A causa do efeito de *latchup* pode ser entendida examinando-se um corte transversal de um inversor CMOS, mostrado na Figura 53, no qual está sobreposto um circuito equivalente. Para os transistores NMOS e PMOS esperados, o esquema descreve um circuito composto de um transistor NPN, um transistor PNP e dois resistores conectados entre a fonte e o terra. Os transistores transversalmente casados formam um retificador controlado biestável (SCR). Geralmente, ambos os transistores bipolares estão desligados. O *latchup* pode ser acionado quando correntes transientes fluem pelo substrato durante a ligação do CI ou quando tensões externas além da faixa de operação normal são aplicadas. Se houver substancial fluxo de corrente no substrato, V_{sub} irá elevar, ligando o transistor NPN. Isso leva corrente pelo resistor de poço, derrubando a tensão do poço e ligando o transistor PNP. A corrente do transistor PNP tende a elevar V_{sub} , iniciando uma realimentação positiva com uma elevada corrente fluindo entre V_{DD} e GND , que persiste até a fonte de alimentação ser desligada ou as linhas de alimentação derreterem.

Felizmente, a prevenção de *latchup* é completamente fácil pela minimização de R_{sub} e R_{well} . Alguns processos usam uma camada epitaxial fina de silício dopado levemente sobre o topo de um substrato dopado fortemente que oferece uma baixa resistências de substrato. O mais importante, o projetista deve pôr as ligações de substrato e poço próximo a cada transistor. Se isso não for útil, você pode tentar seguir os seguintes passos:

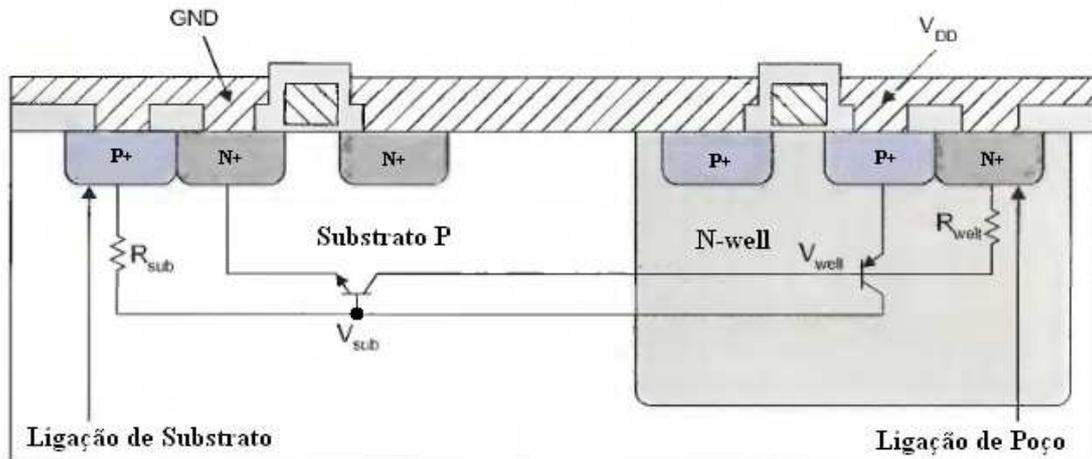


Figura 53 – Origem do latchup no processo CMOS

- Cada poço deve ter no mínimo uma ligação;
- Todas as ligações de substrato e poço devem ser conectadas diretamente a fonte apropriada por metal;
- Uma ligação deve ser colocada a cada 5 a 10 transistores ou a cada 25 a 100 μm ; e,
- Transistores NMOS devem ser agrupados juntamente próximos ao GND , enquanto os transistores PMOS devem ser agrupados juntamente próximos ao V_{DD} , evitando estruturas convolutas que entrelaçam transistores NMOS e PMOS em padrões tipo tabuleiro de damas.

4 Regras de Projeto de Leiaute

Enquanto a largura e comprimento de cada transistor são determinados pelo projeto do circuito, a maioria das outras dimensões no leiaute são ditadas pelas regras de leiaute, isso é, um conjunto de regras que garantem a fabricação de transistores e interconexões apropriadas, apesar das várias tolerâncias em cada etapa do processo.

O principal objetivo das regras de leiaute é de construir circuitos funcionais com segurança em áreas tão pequenas quanto possível. Em geral, representam um compromisso entre desempenho e produção. Elas especificam ao projetista certas limitações geométricas sobre as ilustrações do leiaute de modo que os moldes sobre a pastilha produzida preservará a topologia e geometria dos projetos. É importante observar que as regras de leiaute não representam o quão severo é a fronteira entre a fabricação correta e incorreta. De preferência, elas representam uma tolerância que garante uma grande probabilidade de uma fabricação correta e, subsequentemente, operação correta. Qualquer desvio significativo das regras de leiaute prejudicará seriamente o êxito de um projeto.

As regras de leiaute industrial são usualmente especificadas em microns. Isso torna a migração de um processo para um processo mais avançado difícil, porque nem todas as escalas de regras são do mesmo modo. Mead e Conway popularizaram as regras de leiaute baseada sobre um único parâmetro, o lambda (λ), que caracteriza a resolução do processo. O λ geralmente é metade do mínimo comprimento de canal do transistor extraído.

As regras baseadas em λ são necessariamente conservadoras, porque elas arredondam para cima as dimensões para um inteiro múltiplo de λ . Porém, elas tornam o escalonamento do leiaute trivial. O mesmo leiaute pode ser movido para um novo processo simplesmente especificando um novo valor de λ .

O serviço MOSIS (*Metal Oxide Semiconductor Implementation Service*) é um serviço de criação de protótipo de baixo custo que coleta projetos de clientes como universidades, comércios e governos e agrega-os sobre um conjunto de máscara para repartir custos e gerar volume suficiente de produção para empresas de fabricação interessadas. A MOSIS tem desenvolvido um conjunto de regras de leiaute escalonáveis baseadas em λ que abrange uma larga faixa de processos. As regras tem se tornado um

pouco mais conservadoras para processos submícron mais avançados, mas são fundamentalmente muito similares àquele proposto por Mead e Conway. As regras descrevem a largura mínima para evitar quebras em uma linha, o espaçamento mínimo para evitar curtos entre linhas e a delimitação mínima para assegurar que duas camadas se sobreponham completamente.

4.1 Largura Mínima

As larguras (e comprimentos) das geometrias definidas sobre uma máscara deve exceder um valor mínimo impostos pela litografia e recursos da produção da tecnologia. Por exemplo, se um retângulo de polissilício é excessivamente estreito, então, devido às tolerâncias de fabricação, ele pode simplesmente quebrar ou pelo menos sujeitar-se de uma grande resistência. Em geral, quanto mais grossa uma camada é, maior sua largura mínima permitida, indicando que com a escala de tecnologia, a espessura deve ser diminuída proporcionalmente. A Figura 54 demonstra exemplos de larguras mínimas em uma tecnologia de $0,25\ \mu\text{m}$. Note que a espessura das camadas não está sob o controle do projetista.

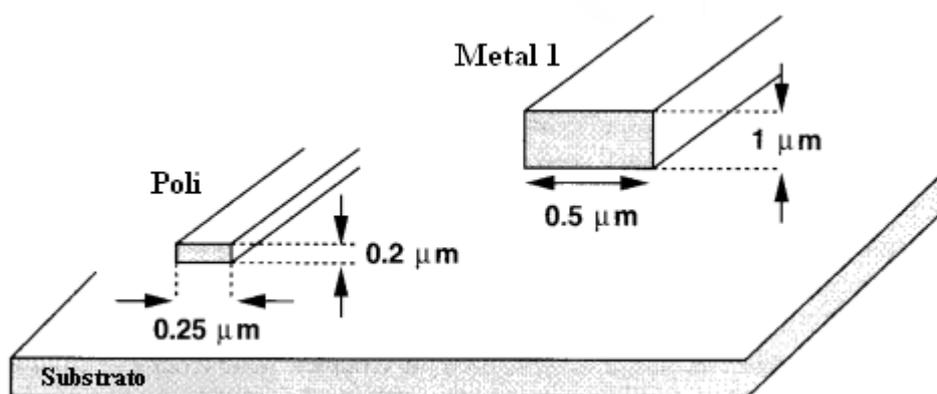


Figura 54 – Larguras e espessuras de linhas de poli e metal 1

4.2 Espaçamento Mínimo

As geometrias construídas sobre a mesma máscara ou, em alguns casos, diferentes máscaras, podem ser separadas por um espaçamento mínimo. Por exemplo, como visto

na Figura 55(a), se duas linhas de polissilício são colocadas muito próximas uma da outra, elas podem causar curto-circuito. Como outro exemplo, considere o caso visto na Figura 55(b), onde uma linha de polissilício passa perto da área de fonte/dreno de um transistor. Um espaçamento mínimo é requerido aqui para garantir que o implante em volta do transistor não seja sobreposto com uma linha de polissilício.

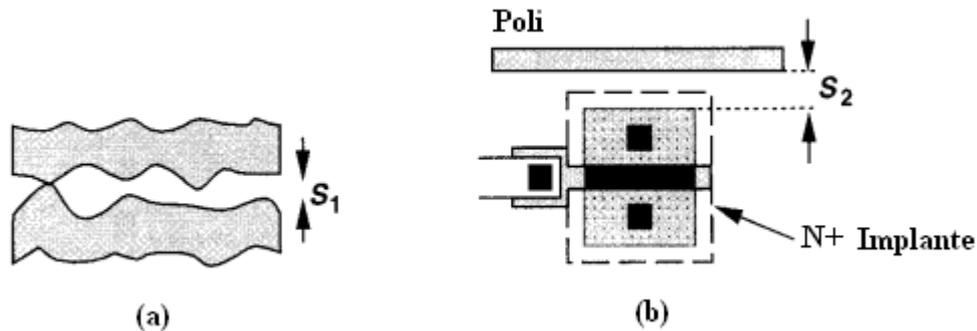


Figura 55 – (a) Curto-circuito entre duas linhas de polissilício muito próximas (b) Espaçamento mínimo entre uma linha de poli e uma região ativa

4.3 Delimitação Mínima

Em um leiaute, um N-well e o implante P+ devem cerca o transistor com margem suficiente para garantir que o dispositivo está contido por essas geometrias em desfeita as tolerâncias. Isso é um exemplo de regra de delimitação mínima. A Figura 56 representa outro exemplo, onde um contato de poli conecta uma linha de poli a uma linha de metal 1. Para ter certeza que o contato permaneça dentro dos quadrados de poli e metal 1, ambas as geometrias devem incluir o contato com margem o suficiente.

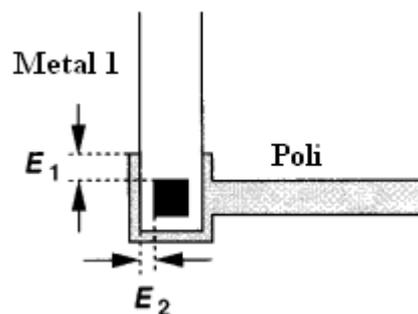


Figura 56 – Regra de delimitação para poli e metal circundando um contato

4.4 Extensão Mínima

Algumas geometrias podem ser estendidas além da borda por um valor mínimo. Por exemplo, como visto na Figura 57, o polissilício da porta tem uma extensão mínima além da área ativa para garantir uma ação adequada do transistor na borda.

Para as dimensões mínimas especificadas nas quatro categorias anteriores, algumas dimensões máximas permitidas podem também ser forçadas. Por exemplo, para longas linhas de metal, a largura mínima é tipicamente maior que aquela para linhas curtas para evitar problemas de “decolagem”.

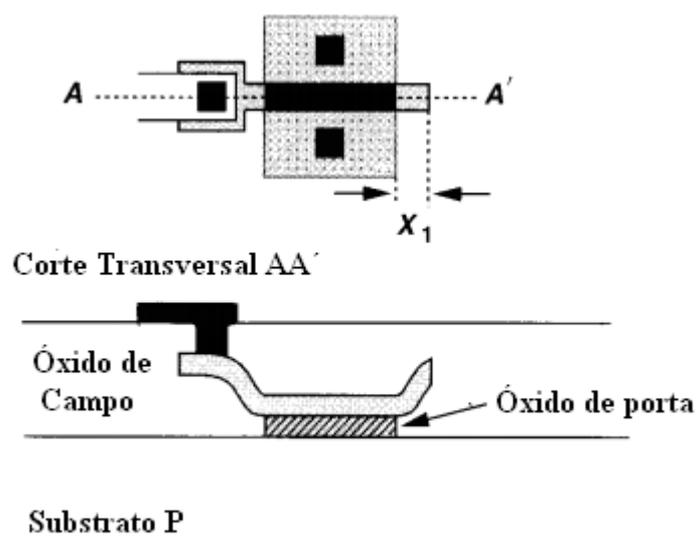


Figura 57 – Extensão de uma linha de poli além da área de porta

A Figura 58 resume um pequeno subconjunto de regras de layout que dominam o layout de um par diferencial NMOS com fonte de corrente de carga PMOS. Modernas tecnologias CMOS envolvem, tipicamente, mais de 150 regras de layout.

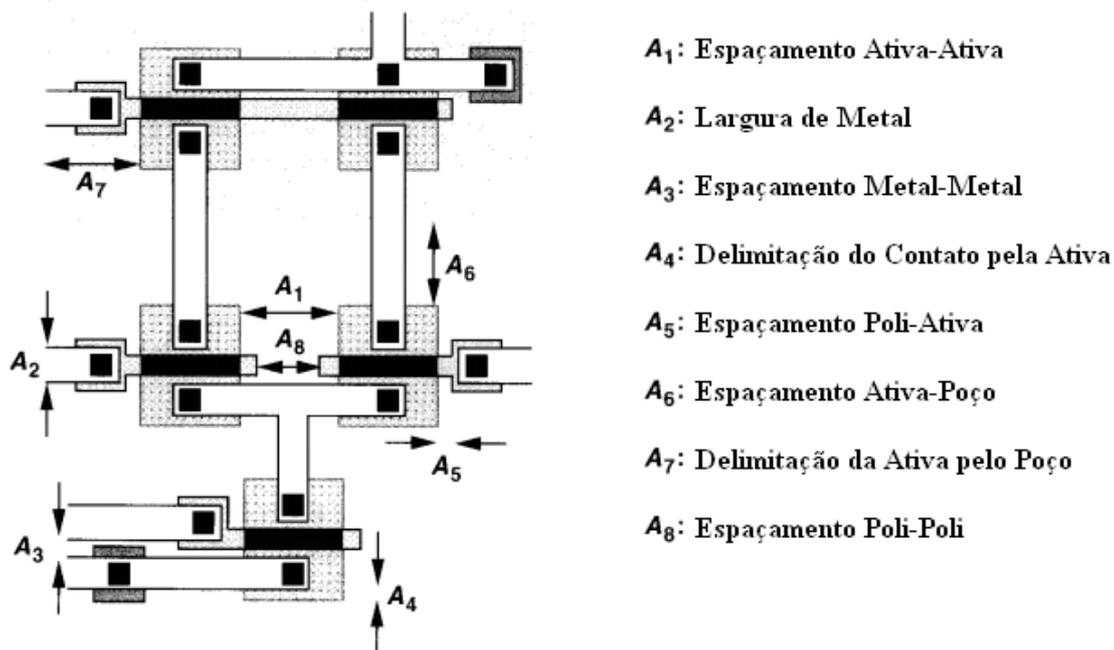


Figura 58 – Leiaute de um par diferencial com fonte de corrente de carga PMOS

4.5 Efeito Antena

Suponha que a porta de um pequeno MOSFET é ligada a um metal 1 interconectado, tendo uma grande área – Figura 59(a). Durante o entalhamento do metal 1, a área de metal age como uma antena coletando íons e elevando em potencial. É então possível que a tensão de porta do dispositivo MOS aumente tanto que o óxido de porta quebre durante a fabricação.

O efeito antena (*Antenna Effect*) pode ocorrer para qualquer peça larga de material condutivo ligado a porta, incluindo o próprio poli. Por essa razão, tecnologias CMOS submícron limitam a área total de tais geometrias, com isso minimizando a probabilidade do óxido de porta danificar.

Se as áreas largas são inevitáveis, então uma descontinuidade pode ser criada como ilustrado na Figura 59(b), de forma que quando o metal 1 é sendo entalhado, a área larga não é conectada à porta.

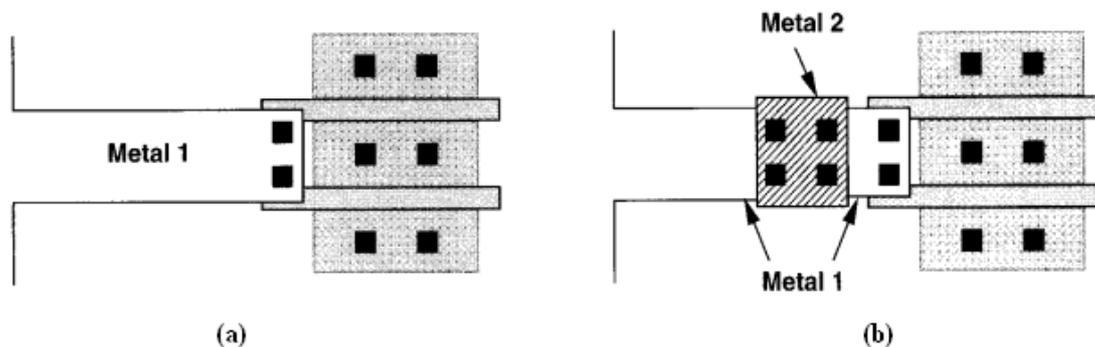


Figura 59 – (a) Leiaute suscetível ao efeito antena (b) Descontinuidade na camada de metal 1 evita o efeito antena

4.6 Regras de Leiaute Escalonável MOSIS

MOSIS atualmente tem três conjuntos de regras: SCMOS, SUBM e DEEP. As regras SUBM são um tanto mais conservadora que as regras SCMOS. As regras DEEP são até mais conservadoras. As regras mais conservadoras permitem que você use um valor levemente menor de λ , enquanto ainda satisfazendo todas as regras de leiaute para um processo. A Tabela 4 lista alguns dos processos da fábrica MOSIS tem oferecido e o valor associado de λ para os diferentes conjuntos de regras. Por exemplo, o processo AMI 0,5 μm pode usar as regras SCMOS com $\lambda = 0,35 \mu\text{m}$ ou as regras SUBM com $\lambda = 0,35 \mu\text{m}$. As regras SUBM são uma boa escolha para projetos de classe porque elas são um tanto fáceis para usar do que as regras DEEP (sem a regra da metade de λ), enquanto ainda sendo compatível com a maioria dos processos. Alguns processos oferecem uma segunda camada de polissilício para transistores de porta flutuante e capacitores poli-isolador-poli usado em circuitos analógicos.

Para as regras de leiaute onde o mínimo comprimento de canal extraído excede o tamanho característico, a MOSIS aplica uma distorção de polissilício (*polysilicon bias*) para contrair as portas por uma quantidade uniforme antes das máscaras serem feitas. Por exemplo, nas regras SUBM para o processo AMI 0,5 μm com $\lambda = 0,30 \mu\text{m}$, uma distorção de $-0,1 \mu\text{m}$ é aplicada a todos os polissilícios. Assim, uma porta de transistor de 2λ é $0,50 \mu\text{m}$, ao invés de $0,60 \mu\text{m}$, e uma porta de 4λ é igual a $1,1 \mu\text{m}$, ao invés de $1,2 \mu\text{m}$. Quando simulando circuitos, esteja certo de estar usando os comprimentos de canal distorcidos para modelar mais precisamente o comportamento do transistor.

Tabela 4 – Opções de regras MOSIS

Fornecedor	Tamanho Característico	Camadas de Interconexão	Vias Sobrepostas	SCMOS	SUBM	DEEP
Orbit	2,0 μm	2 metais	Não	$\lambda = 1,0 \mu\text{m}$		
AMI	1,5 μm	2 metais, 2 polis	Não	$\lambda = 0,80 \mu\text{m}$	$\lambda = 0,80 \mu\text{m}$	
AMI	0,5 μm	3 metais, 1-2 poli	Sim	$\lambda = 0,35 \mu\text{m}$	$\lambda = 0,30 \mu\text{m}$	
TSMC	0,35 μm	4 metais, 1-2 poli	Sim		$\lambda = 0,20 \mu\text{m}$	
TSMC	0,25 μm	5 metais	Sim		$\lambda = 0,15 \mu\text{m}$	$\lambda = 0,12 \mu\text{m}$
TSMC	0,18 μm	6 metais	Sim		$\lambda = 0,10 \mu\text{m}$	$\lambda = 0,09 \mu\text{m}$

No Anexo 6, obtido no site da MOSIS¹¹, mostra-se algumas regras básicas de importância ao aluno que está começando a trabalhar com leiaute. Os leiautes consistem de um conjunto de retângulos sobre várias camadas tais como polissilício ou metal. Largura é a largura mínima de um retângulo sobre uma camada em particular. Espaçamento é o espaçamento mínimo entre dois retângulos sobre a mesma ou diferentes camadas. Sobreposição especifica quanto um retângulo deve circular outro retângulo sobre outra camada. Dimensões são todas especificadas em λ , exceto para cortes de passivação que não dimensiona bem devido eles deverem entra em contato com pontas de prova.

¹¹ <http://www.mosis.com/Technical/Designrules/scmos/scmos-main.html>

5 Técnicas de Leiaute Analógico

O grande conjunto de regras impostas pelos processos CMOS apontam para maximizar a produção de CI's digitais enquanto permitem um projeto de circuito moderadamente agressivo. Sistemas analógicos, por outro lado, demandam muito mais precauções com o leiaute para minimizar efeitos tais como descasamento, ruído, linha cruzada, etc. Assim, quando alguém projeta circuitos analógicos, várias questões importantes do leiaute devem ser consideradas para construir um circuito de alta qualidade.

5.1 Transistores Seccionados

Transistores largos são usualmente dobrados para reduzir ambas as áreas de junção fonte/dreno e a resistência de porta. Uma simples estrutura dobrada, Figura 60(a), pode ser inadequada para vários dispositivos largos, necessitando o uso de múltiplas secções – Figura 60(b). Como uma regra prática, a largura de cada secção é escolhida de modo que a resistência das secções seja menor que a transcondutância inversa associada com a secção. Em aplicações de baixo ruído, a resistência de porta deve ser 1/5 a 1/10 de $1/g_m$.

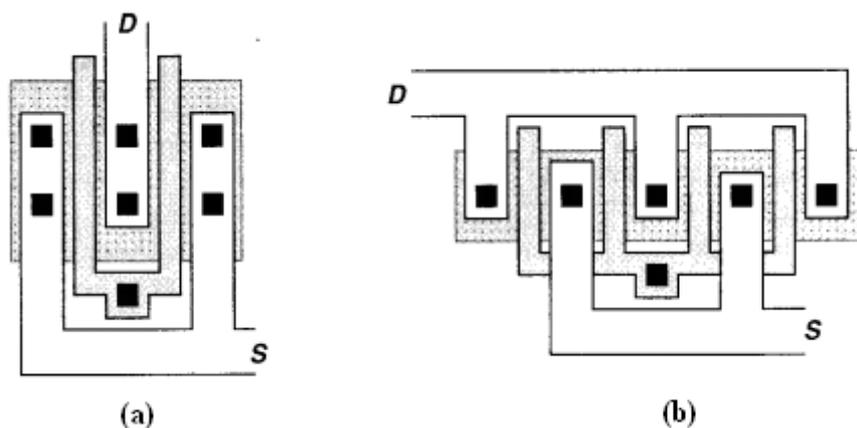


Figura 60 – (a) Dobradiça simples de um MOSFET (b) Uso de secções

Enquanto a resistência de porta pode ser reduzida para decompor o transistor em mais secções, a capacitância associada com os parâmetros das áreas de fonte/dreno aumenta. Na prática, esse requerimento pode conflitar com aquele para minimizar o ruído de resistência de porta, requisitando um compromisso entre os dois ou conectando a porta sobre ambas as extremidades para reduzir a resistência.

O leiaute de um circuito em cascata pode ser simplificado se o dispositivo M_1 e o dispositivo em cascata M_2 tiverem larguras iguais. Como visto na Figura 61(a), o dreno de M_1 e a fonte de M_2 podem compartilhar a mesma junção. Mais importante, desde que essa junção não é conectada a qualquer outro nó, ela não necessita conter um contato e pode, assim, ser muito pequeno – Figura 61(b). Consequentemente, a capacitância no dreno de M_1 é reduzida substancialmente, aumentando o desempenho em alta frequência. Para transistores largos, cada transistor pode usar dois ou mais secções – Figura 61(c).

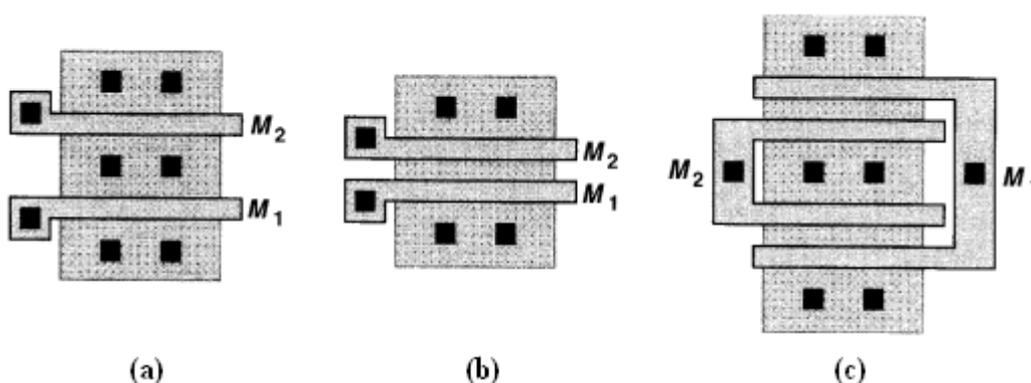


Figura 61 – Leiaute de dispositivos em cascata que tem a mesma largura

5.2 Simetria

Assimetria em circuitos diferenciais introduz um deslocamento de referência de entrada, limitando o nível de sinal mínimo que pode ser detectado. Enquanto algum casamento entre dispositivos seja inevitável, uma atenção inadequada à simetria no leiaute pode resultar em grandes deslocamentos. A simetria também anula o efeito de ruído de modo comum e a não linearidade de ordem par. É importante notar que a simetria deve ser aplicada para ambos os dispositivos de interesse e seu ambiente em volta.

Considere o par diferencial da Figura 62(a) como ponto de partida. Se, como visto na Figura 62(b), os dois transistores são dispostos com orientações diferentes, o casamento sofre enormemente devido várias etapas no processo de litografia e da pastilha comportarem-se diferentemente ao longo de diferentes eixos. Assim, uma das configurações na Figura 62 (c) e (d) fornece uma melhor solução. A escolha entre essas duas é determinada por um sutil efeito chamado de “sombreamento de porta” (*gate shadowing*). Ilustrado na Figura 63, o sombreamento é causado pelo polissilício de porta durante a implantação do fonte/dreno, devido o implante (ou a pastilha) ser inclinado por cerca de 7° para evitar a canalização. Como resultado, uma faixa estreita na região de fonte e dreno recebe menos implantação, criando uma pequena assimetria entre os lados de difusão de dreno e fonte após as áreas implantadas serem temperadas.

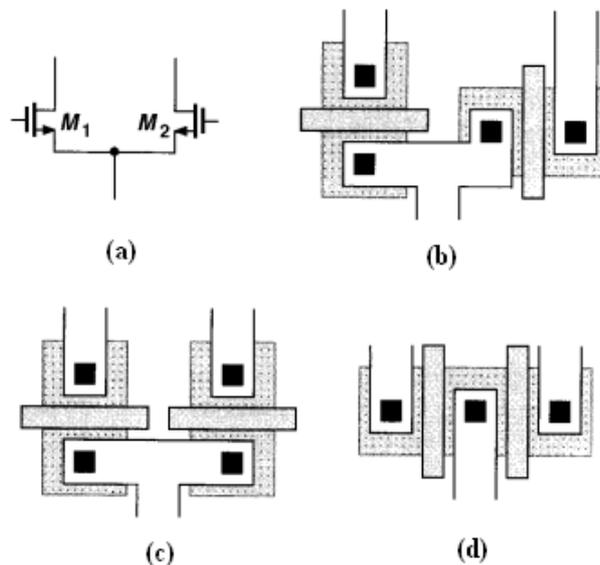


Figura 62 – (a) Par diferencial (b) Leiaute de M_1 e M_2 com diferentes orientações (c) Leiaute com dispositivos alinhados pela porta (d) Leiaute com dispositivos com as portas em paralelo

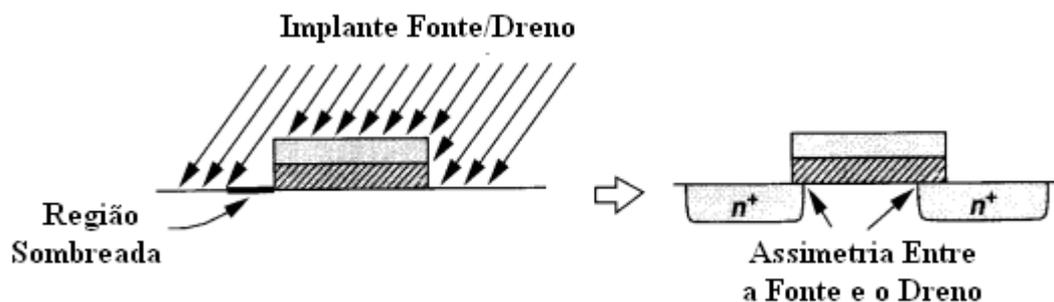


Figura 63 – Sombreamento devido à inclinação de implante

Agora considere as estruturas das Figura 62(c) e (d) na presença do sombreamento de porta – Figura 64. Na Figura 64(a), se o terminal sombreado é diferenciado como o dreno (ou a fonte), então os dois dispositivos não mantêm assimetria resultando do sombreamento. Na Figura 64(b), por outro lado, os transistores não são idênticos mesmo se os terminais sombreados são distinguidos, devido à região de fonte de M_1 “enxergar” M_2 a sua direita, enquanto que a região de fonte de M_2 “enxerga” somente o óxido de campo. Similarmente, os drenos de M_1 e M_2 verem diferentes estruturas as suas esquerdas. Em outras palavras, o ambiente ao redor de M_1 não é idêntico a aquele de M_2 . Por essa razão, a topologia da Figura 64(a) ser preferível.

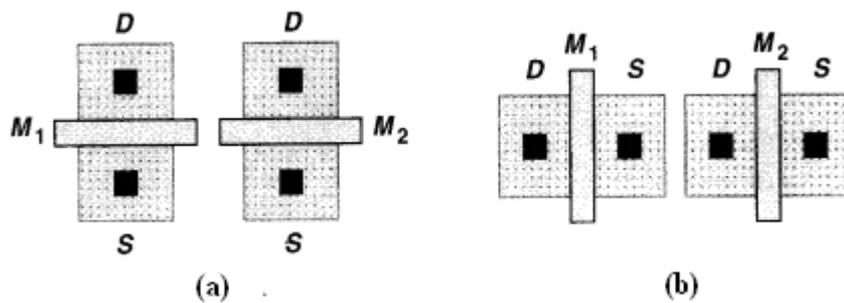


Figura 64 – Efeito de sombreamento sobre transistores (a) Alinhados pela porta (b) Portas paralelas

A assimetria inerente a estrutura da Figura 64(b) pode ser melhoradas adicionando-se transistores “mortos” (*dummy*) nos dois lados, de forma que M_1 e M_2 vejam aproximadamente o mesmo ambiente – Figura 65. Porém, em circuitos mais complexos, tais meios não podem ser aplicados facilmente.

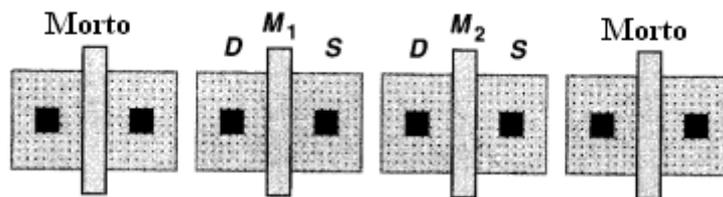


Figura 65 – Adição de dispositivos “mortos” para melhorar a simetria

Deve-se enfatizar a importância de manter o mesmo ambiente dos dois lados dos eixos de simetria. Por exemplo, na estrutura da Figura 66(a), uma linha de metal desvinculada sobre somente um transistor reduz, de fato, a simetria, aumentando o descasamento entre M_1 e M_2 . Em tais casos, ou uma réplica deve ser produzida sobre o

outro lado (mesmo que a réplica permaneça flutuante) – Figura 66(b) – ou, preferencialmente, a fonte de assimetria deve ser removida.

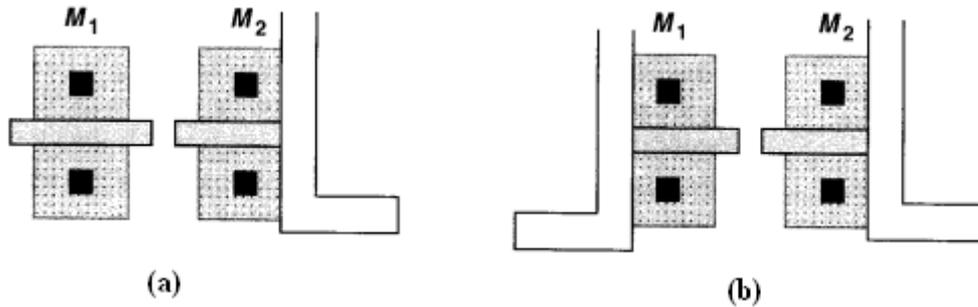


Figura 66 – (a) Assimetria resultante de uma linha de metal passando sobre M_2 (b) Removendo a assimetria pela replicação da linha sobre M_1

A simetria torna-se mais difícil de ser estabelecida para transistores grandes. No par diferencial da Figura 67, por exemplo, os dois transistores tem uma largura grande de forma a alcançar uma pequena tensão de deslocamento de entrada, mas gradientes ao longo do eixo x dão origem à apreciáveis descasamentos. Para reduzir o erro, uma configuração de centro comum (*common-centroid*) pode ser usada tal que o efeito de primeira ordem dos gradientes ao longo de ambos os eixos é cancelado. Ilustrado na Figura 68, a ideia é decompor cada transistor em duas metades que são colocadas frente a frente e conectadas em paralelo (as linhas interconectadas mostradas nessa figura são somente conceituais). Porém, o modo de interconectar nesse leiaute é muito difícil, frequentemente conduzindo a assimetrias sistemáticas do tipo mostrada na Figura 66(a) ou nas capacitâncias das linhas de terra e entre as linhas.

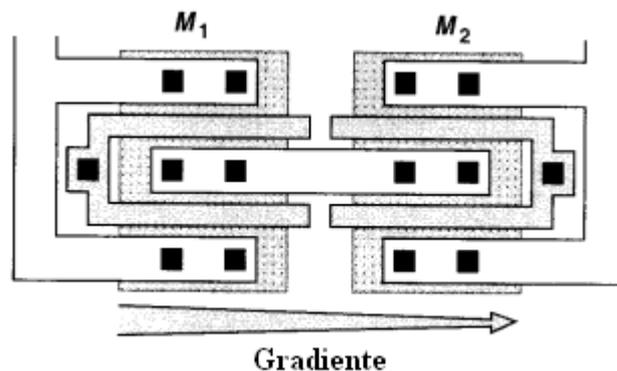


Figura 67 – Efeito de gradiente em um par diferencial

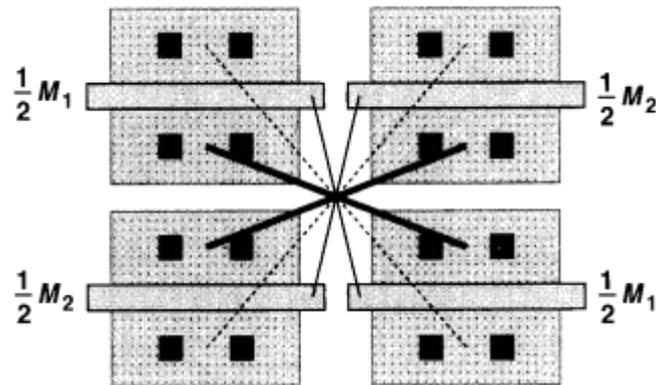


Figura 68 – Leiaute de centro comum

5.3 Distribuição de Referência

Em sistemas analógicos, as tensões e correntes de polarização de vários blocos são derivados de um ou mais geradores de referência. A distribuição de tais referências através de um CI grande exige um número de questões importantes. Considere o exemplo visto na Figura 69, onde I_{REF} é gerado por uma referência e M_1-M_n serve como fontes de corrente de polarização de blocos que são localizados longe de M_{REF} e de cada outro. Se o casamento entre $I_{D1}-I_{Dn}$ e I_{REF} é importante, então a queda de tensão ao longo da linha de terra deve ser levada em conta. De fato, para um número grande de circuitos conectados a mesma linha de terra, o casamento sistemático entre as fontes de corrente e I_{REF} pode ser inaceitável.

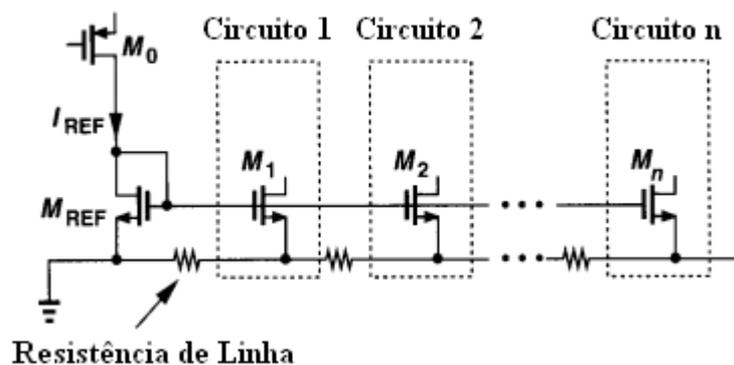


Figura 69 – Distribuição de uma tensão de referência por fontes de corrente polarizando

Para prevenir a dificuldade anterior, a referência pode ser distribuída no domínio da corrente ao invés do domínio da tensão. Ilustrada na Figura 70, a ideia é direcionar a corrente de referência à vizinhança dos blocos e realizar a operação da fonte de corrente

localmente. Pondo a resistência de conexão em série com as fontes de corrente, essa técnica diminui erros sistemáticos se os blocos aparecerem em grupos densos em regiões diferentes no CI. Porém, casamentos entre I_{REF1} e I_{REF2} e entre M_{REF1} e M_{REF2} introduz erro. Em sistemas grandes, pode ser vantajoso empregar vários circuitos de referências locais para aliviar problemas de roteamento.

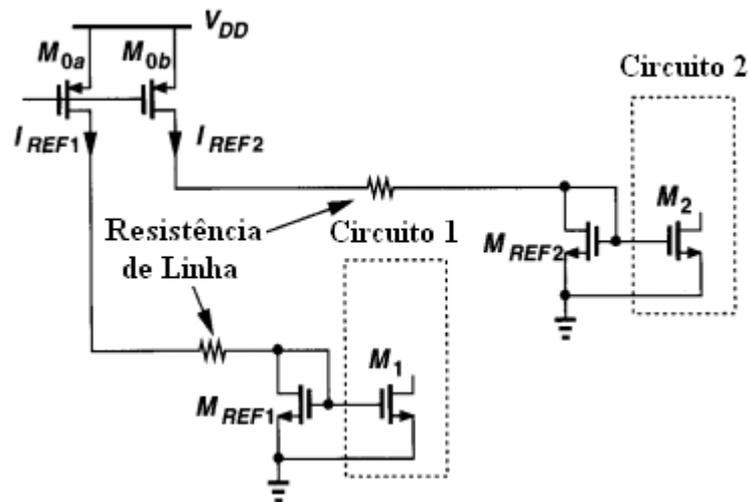
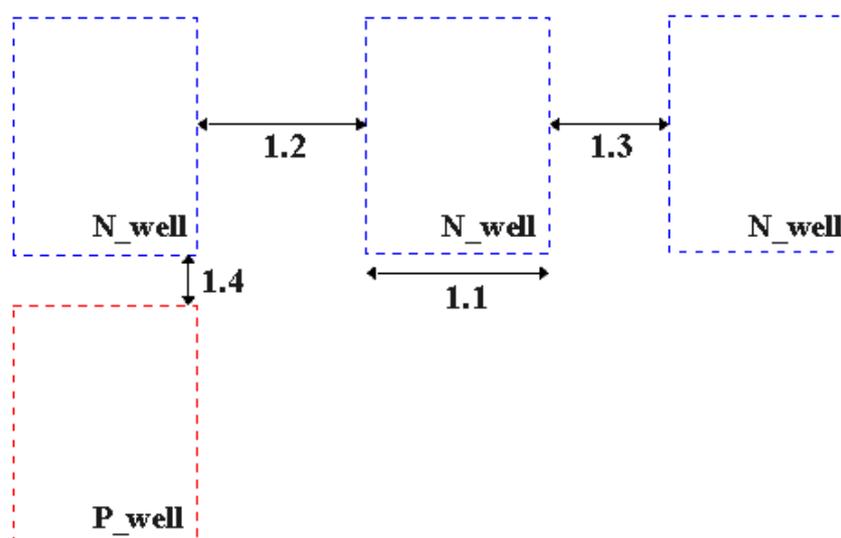


Figura 70 – Distribuição de corrente para reduzir o efeito de resistência de conexão

6 Anexo A – Regras de Leiaute Escalonável Mosis (Rev 8.00)

6.1 Well

Rule	Description	Lambda		
		SCMOS	SUBM	DEEP
1.1	Minimum width	10	12	12
1.2	Minimum spacing between wells at different potential	9	18	18
1.3	Minimum spacing between wells at same potential	6	6	6
1.4	Minimum spacing between wells of different type (if both are drawn)	0	0	0

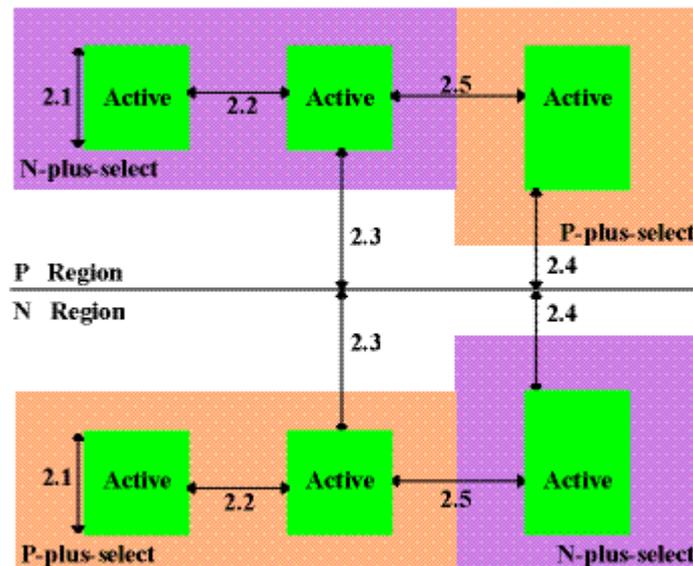


6.2 Ativa

Rule	Description	Lambda		
		SCMOS	SUBM	DEEP
2.1	Minimum width	3 *	3 *	3
2.2	Minimum spacing	3	3	3
2.3	Source/drain active to well edge	5	6	6
2.4	Substrate/well contact active to well edge	3	3	3
2.5	Minimum spacing between non-abutting active of different implant. Abutting active ("split-active") is illustrated under <u>Select Layout Rules</u> .	4	4	4

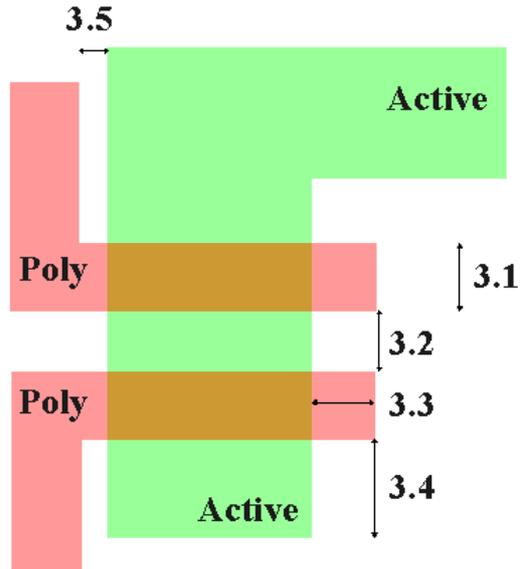
* Note: For analog and critical digital designs, MOSIS recommends the following minimum MOS channel widths (active under poly) for ON Semiconductor designs. Narrower devices, down to design rule minimum, will be functional, but their electrical characteristics will not scale, and their performance is not predictable from MOSIS SPICE parameters.

Process	Design Technology	Design Lambda (micrometers)	Minimum Width (lambda)
AMI_C5F/N	SCN3M, SCN3ME	0.35	9
AMI_C5F/N	SCN3M_SUBM, SCN3ME_SUBM	0.30	10



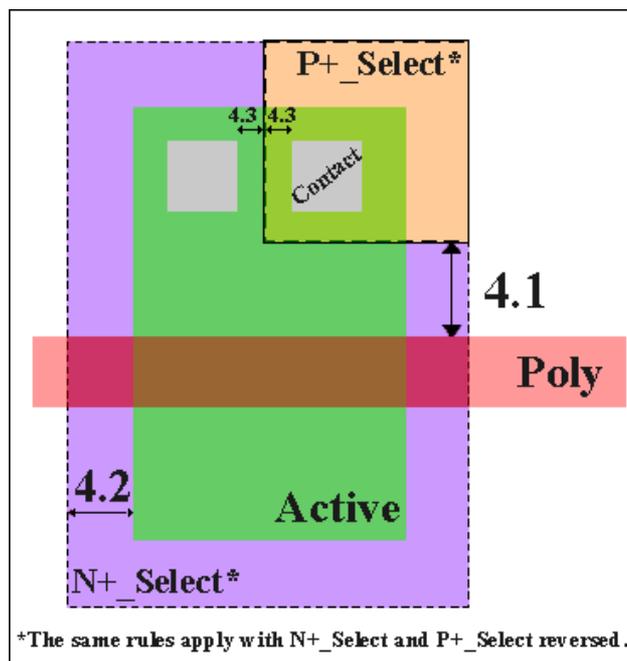
6.3 Poli

Rule	Description	Lambda		
		SCMOS	SUBM	DEEP
3.1	Minimum width	2	2	2
3.2	Minimum spacing over field	2	3	3
3.2.a	Minimum spacing over active	2	3	4
3.3	Minimum gate extension of active	2	2	2.5
3.4	Minimum active extension of poly	3	3	4
3.5	Minimum field poly to active	1	1	1



6.4 Rigoroso

Rule	Description	Lambda		
		SCMOS	SUBM	DEEP
4.1	Minimum select spacing to channel of transistor to ensure adequate source/drain width	3	3	3
4.2	Minimum select overlap of active	2	2	2
4.3	Minimum select overlap of contact	1	1	1.5
4.4	Minimum select width and spacing (Note: P-select and N-select may be coincident, but must <i>not</i> overlap) (not illustrated)	2	2	4

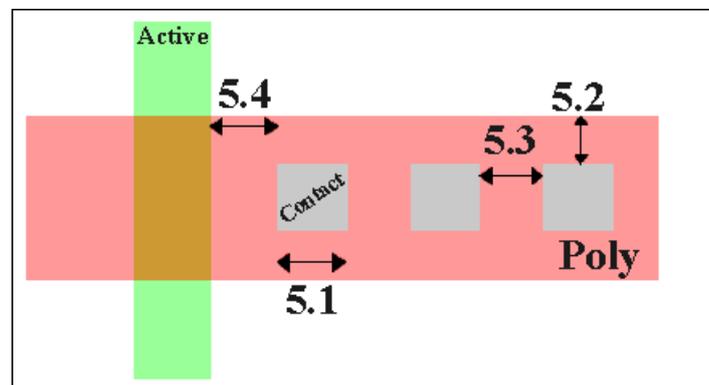


6.5 Contato Simples para Poli

On 0.50 micron process (and all finer feature size processes), it is required that all features on the insulator layers (CONTACT, VIA, VIA2) must be of the single standard size; there are no exceptions for pads (or logos, or anything else); large openings must be replaced by an array of standard sized openings. Contacts must be drawn orthogonal to the grid of the layout. Non-Manhattan contacts are not allowed.

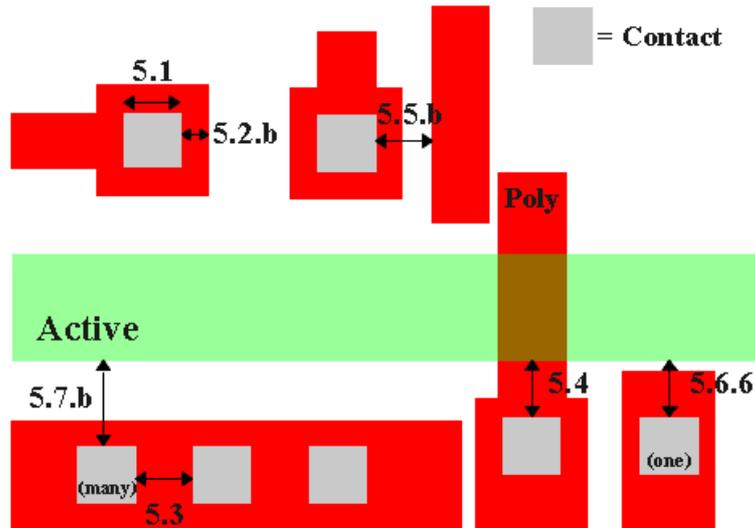
If your design cannot tolerate 1.5 lambda contact overlap in 5.2, use the alternative rules which reduce the overlap but increase the spacing to surrounding features. Rules 5.1, 5.3, and 5.4, still apply and are unchanged.

Rule	Description	Lambda		
		SCMOS	SUBM	DEEP
5.1	Exact contact size	2x2	2x2	2x2
5.2	Minimum poly overlap	1.5	1.5	1.5
5.3	Minimum contact spacing	2	3	4
5.4	Minimum spacing to gate of transistor	2	2	2



6.6 Contato Alternativo para Poli

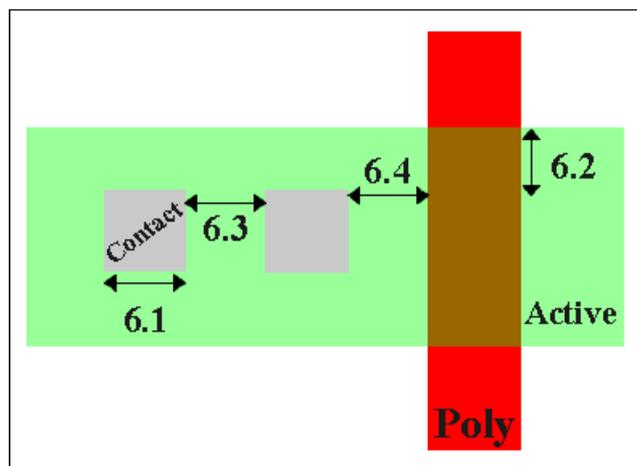
Rule	Description	Lambda		
		SCMOS	SUBM	DEEP
5.2.b	Minimum poly overlap	1	1	1
5.5.b	Minimum spacing to other poly	4	5	5
5.6.b	Minimum spacing to active (one contact)	2	2	2
5.7.b	Minimum spacing to active (many contacts)	3	3	3



6.7 Contato Simples para Ativa

If your design cannot handle the 1.5 lambda contact overlap in 6.2, use the alternative rules which reduce the overlap but increase the spacing to surrounding features. Rules 6.1, 6.3, and 6.4, still apply and are unchanged. Contacts must be drawn orthogonal to the grid of the layout. Non-Manhattan contacts are not allowed.

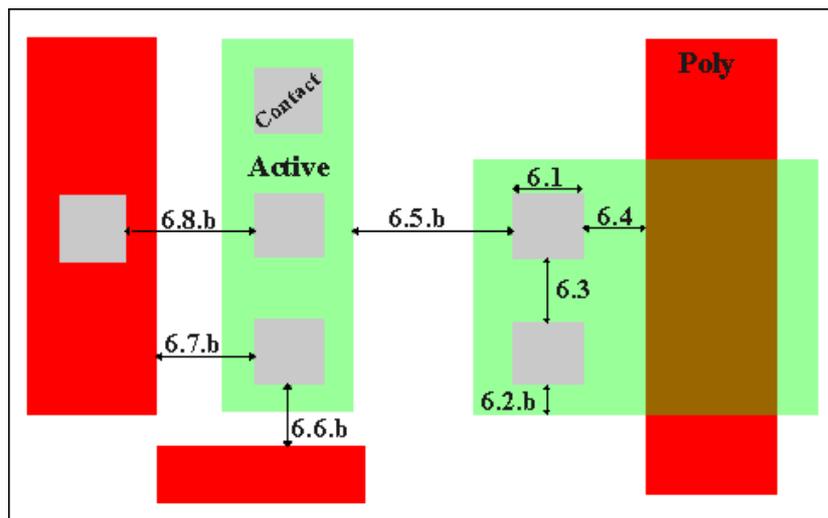
Rule	Description	Lambda		
		SCMOS	SUBM	DEEP
6.1	Exact contact size	2x2	2x2	2x2
6.2	Minimum active overlap	1.5	1.5	1.5
6.3	Minimum contact spacing	2	3	4
6.4	Minimum spacing to gate of transistor	2	2	2



6.8 Contato A

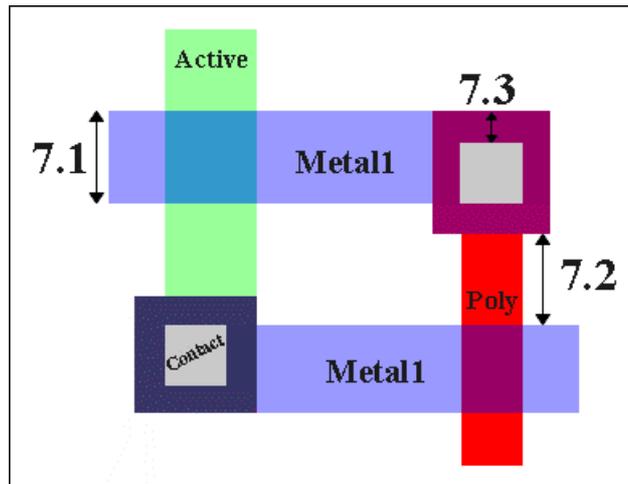
Rule	Description	Lambda
------	-------------	--------

		SCMOS	SUBM	DEEP
6.2.b	Minimum active overlap	1	1	1
6.5.b	Minimum spacing to diffusion active	5	5	5
6.6.b	Minimum spacing to field poly (one contact)	2	2	2
6.7.b	Minimum spacing to field poly (many contacts)	3	3	3
6.8.b	Minimum spacing to poly contact	4	4	4



6.9 Metal 1

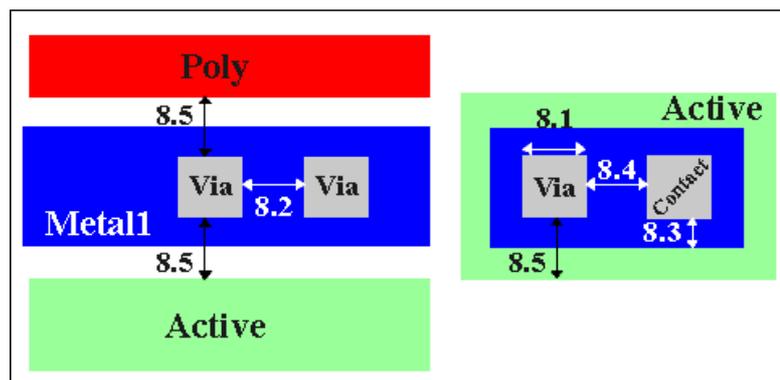
Rule	Description	Lambda		
		SCMOS	SUBM	DEEP
7.1	Minimum width	3	3	3
7.2	Minimum spacing	2	3	3
7.3	Minimum overlap of any contact	1	1	1
7.4	Minimum spacing when either metal line is wider than 10 lambda	4	6	6



6.10 Via 1

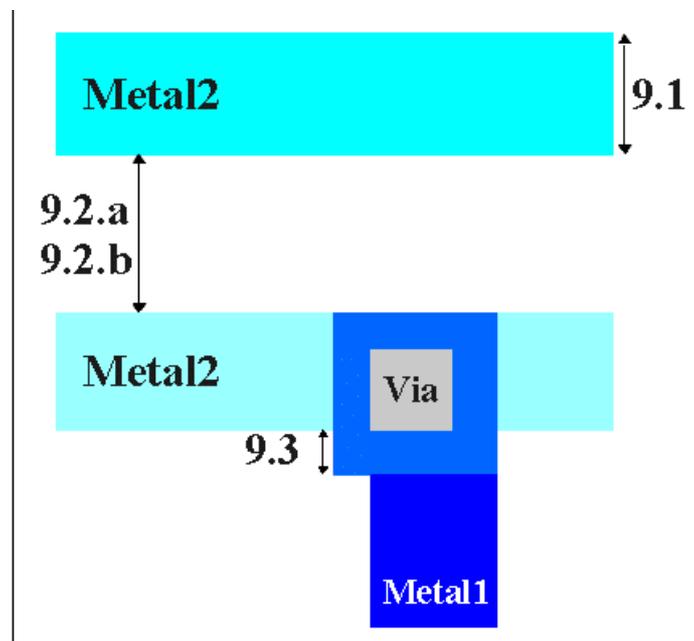
Vias must be drawn orthogonal to the grid of the layout. Non-Manhattan vias are not allowed.

Rule	Description	Lambda					
		2 Metal Process			3+ Metal Process		
		SCMOS	SUBM	DEEP	SCMOS	SUBM	DEEP
8.1	Exact size	2 x 2	n/a	n/a	2 x 2	2 x 2	3 x 3
8.2	Minimum via1 spacing	3	n/a	n/a	3	3	3
8.3	Minimum overlap by metal1	1	n/a	n/a	1	1	1
8.4	Minimum spacing to contact for technology codes mapped to processes that do not allow stacked vias (SCNA, SCNE, SCN3M, SCN3MLC)	2	n/a	n/a	2	2	n/a
8.5	Minimum spacing to poly or active edge for technology codes mapped to processes that do not allow stacked vias (NOTE: list is not same as for 8.4)	2	n/a	n/a	2	2	n/a



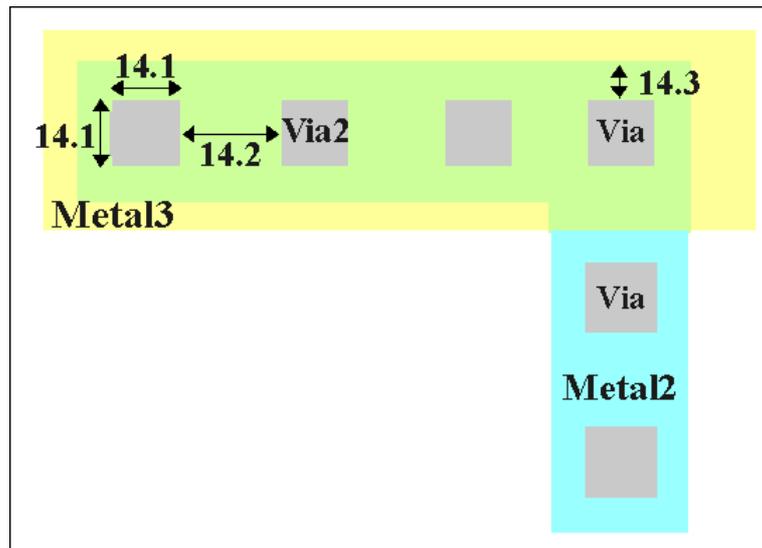
6.11 Metal 2

Rule	Description	Lambda					
		2 Metal Process			3+ Metal Process		
		SCMOS	SUBM	DEEP	SCMOS	SUBM	DEEP
9.1	Minimum width	3	n/a	n/a	3	3	3
9.2	Minimum spacing	3	n/a	n/a	3	3	4
9.3	Minimum overlap of via1	1	n/a	n/a	1	1	1
9.4	Minimum spacing when either metal line is wider than 10 lambda	6	n/a	n/a	6	6	8



6.12 Via 2 (Opção Metal Triplo)

Rule	Description	Lambda					
		3 Metal Process			4+ Metal Process		
		SCMOS	SUBM	DEEP	SCMOS	SUBM	DEEP
14.1	Exact size	2x2	2x2	n/a	2x2	2x2	3x3
14.2	Minimum spacing	3	3	n/a	3	3	3
14.3	Minimum overlap by metal2	1	1	n/a	1	1	1
14.4	Via2 may be placed over via1						
14.5	Via2 may be placed over contact						



6.13 Metal 3 (Opção Metal Triplo)

Rule	Description	Lambda					
		3 Metal Process			4+ Metal Process		
		SCMOS	SUBM	DEEP	SCMOS	SUBM	DEEP
15.1	Minimum width	6	5	n/a	3	3	3
15.2	Minimum spacing to metal3	4	3	n/a	3	3	4
15.3	Minimum overlap of via2	2	2	n/a	1	1	1
15.4	Minimum spacing when either metal line is wider than 10 lambda	8	6	n/a	6	6	8

