

MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Escola de Engenharia
Programa de Pós-Graduação em Engenharia Metalúrgica e dos Materiais-
PPGEMM

Síntese semi-automática de amplificadores operacionais

Marcelo Thomas

Dissertação para obtenção do título de Mestre
em Engenharia

Porto Alegre
1995

MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Escola de Engenharia
Programa de Pós-Graduação em Engenharia Metalúrgica e dos Materiais-
PPGEMM

Síntese semi-automática de amplificadores operacionais

Marcelo Thomas

Dissertação apresentada ao programa de Pós-Graduação
em Engenharia Metalúrgica e dos Materiais-PPGEMM, como parte dos
requisitos para a obtenção do título de Mestre em Engenharia.
Área de concentração: Instrumentação Eletro-eletrônica - Microeletrônica

Porto Alegre
1995

ESCOLA DE ENGENHARIA
BIBLIOTECA

Esta dissertação foi julgada adequada para a obtenção do Título de Mestre em Engenharia e aprovada em sua forma final pelo Orientador e pela Banca Examinadora do Curso de Pós-Graduação.

Orientador: Prof. Dr. Altamiro Amadeu Suzim

Banca Examinadora:

Prof. Hércules Pereira Neves

Prof. Marcelo Lubaszewski

Prof. Sérgio Bampi

DEDICATÓRIA

Dedico este trabalho a minha noiva Andréa Simões Sebben, por tantos motivos que não posso escrevê-los todos aqui. Também aos meus pais Flavio e Ines Thomas, por terem confiado incondicionalmente em mim, desde os idos tempos da graduação. Aos colegas de universidade, àqueles que enxergavam companheiros dentro das salas de aula e não concorrentes. A quase todos os professores, pela sua compreensão, bom-humor e presença de espírito. E finalmente aos amigos e familiares, por acreditarem que eu estava realmente trabalhando.

AGRADECIMENTOS

Ao final desta etapa, sinto-me compelido a agradecer a todas as pessoas que acreditam que o aprimoramento da técnica como benefício social não é somente uma frase de impacto, àquelas que crêem que a evolução tecnológica é consequência de um trabalho sério, compenetrado, recompensador e sobretudo árduo. Ainda, gostaria de prestar as devidos agradecimentos e homenagens às pessoas que, por acreditarem na pesquisa acadêmica como instrumento de desenvolvimento, valorizam e incentivam esta carreira que tem contribuído tanto para a construção de uma sociedade melhor. Pensando nisso, não posso deixar de mencionar o Prof. Altamiro Suzim, Luigi Carro, Renato Brito e Sérgio Bampi e os colegas Alexandre Junqueira, Pedro Gonçalves, Cristiano Pinto e Fábio Speggorin.

SUMÁRIO

LISTA DE FIGURAS	VIII
LISTA DE TABELAS	X
LISTA DE SÍMBOLOS	XI
LISTA DE ANEXOS	XIII
LISTA DE GRÁFICOS	XIV
RESUMO	XVI
ABSTRACT	XVII
1 Introdução	18
2 Revisão bibliográfica	22
3 Circuitos analógicos	27
3.1 Inversor	28
3.2 Amplificador Operacional	33
3.3 Compensação	40
3.4 Buffers	43
3.5 Miller: análise matemática	44
3.6 Implementações alternativas do capacitor de compensação	50
4 Modelamento	59
4.1 Modelo Quadrático	60
4.2 Modelo Vladimirescu	63
4.3 Modelo EKV	68
4.4 O modelo ótimo	73
5 Implementação do Modelo Vladimirescu	75
5.1 Objetivos Gerais	75
5.2 Descrição do algoritmo	77
5.3 Utilização e principais características	78
5.4 Análise da qualidade da implementação	80
6 Síntese de amplificadores operacionais	85
6.1 Generalidades	85
6.2 Aplicação para topologia Miller	89
6.3 Versão usada no <i>testchip</i>	100
6.4 Resultados da síntese	105
7 Aprimoramento do método	116
7.1 Canal N x Canal P	116
7.2 Capacitor de compensação	117
7.3 Resultado das rotinas aperfeiçoadas	121
7.4 Alterando parâmetros tecnológicos	123
7.5 Método alternativo: gm/Id	125
8 <i>Lay-out</i> de circuitos analógicos	134
8.1 Descrição de transistores	134

8.2 Modificação do projeto original	137
8.3 Métodos de verificação	139
9 Conclusão	151
REFERÊNCIAS BIBLIOGRÁFICAS	154
Anexos	157
Gráficos	161

LISTA DE FIGURAS

3.1- Função de transferência e esquemático de um inversor simples	28
3.2- Representação dos terminais do transistor e seu equivalente de pequenos sinais	29
3.3- Circuito de pequenos sinais do circuito inversor da fig 3.1	30
3.4- Inversor com fonte de corrente para melhorar a performance	31
3.5- Inversor em configuração <i>PUSH-PULL</i>	32
3.6- Deslocador de nível (<i>Level Shifter</i>)	33
3.7- Estágio diferencial de um amplificador operacional genérico	35
3.8- Amplificador Operacional Miller, com compensação interna	36
3.9- Aplicação da técnica <i>Cascode</i> a um inversor simples	36
3.10- Comparação qualitativa da performance de inversores com um ou dois estágios <i>Cascode</i> em relação a um desprovido do mesmo	37
3.11- Amplificador Operacional <i>Folded Cascode</i>	38
3.12- Relação $I_d \times V_{ds}$ para dois valores de V_{gs} , evidenciando o processo de rearranjo de tensões	40
3.13- Gráficos de magnitude e ângulo do ganho de operacionais, evidenciando problemas vinculados à estabilidade	41
3.14- Três implementações de um <i>buffer</i> , com características	43
3.15- Representação com modelo de pequenos sinais do Miller	45
3.16- Representação do transistor sendo usado como elemento de capacitância	51
3.17- Primeiro circuito de teste usado na determinação da característica Capacitância por Área do transistor MOS	53
3.18- Circuito de teste usado na determinação da característica Capacitância por Tensão DC no transistor MOS	54
3.19- Integrador simples usando uma lógica adicional para comutar dois capacitores com características lineares em semi-planos diferentes de polarização	57
4.1- Característica $I_d \times V_{gs}$ e $I_d \times V_{ds}$ de um transistor MOS genérico	62
4.2- Gráfico de velocidade de portadores por campo elétrico aplicado, evidenciando a degradação de mobilidade	65
4.3- Representação de um transistor com V_{ds} igual a tensão de <i>pich-off</i> e outro com V_{ds} superior a tensão de <i>pich-off</i> , retratando a questão da modulação do comprimento de canal	66
5.1- Análise do transistor levando em conta as resistências dos terminais de fonte e dreno	81

5.2- Configuração do transistor usado no circuito de teste para verificar a precisão do modelo implementado, em comparação com as simulações	82
6.1- Obtenção dos parâmetros de pequenos sinais do transistor	90
6.2- Representação completa do Miller, incluindo circuito de polarização	90
7.1- Amplificador Miller com entrada PMOS, para reduzir a área necessária e a potência consumida	116
7.2- Repetição do circuito inversor, com ênfase em sua descrição baseada no parâmetro g_m/I_d	126
7.3- Gráfico de g_m/I_d em relação à corrente normalizada $I=I_d/\beta$	126
8.1- Representação em corte de um circuito integrado com um transistor NMOS e PMOS	134
8.2- Representação a nível de <i>lay-out</i> de um transistor PMOS, mostrando suas dimensões e alguns efeitos parasitas	136
8.3- Distribuição de ondas de calor dentro do integrado, enfatizando a importância da equidistância dos elementos das isoterms	139
8.4- Representação a nível de <i>lay-out</i> de um transistor com a dimensão W excessivamente grande, em configuração empilhada	140
8.5- Representação a nível de <i>lay-out</i> de um transistor com a dimensão W excessivamente grande, em configuração salsicha	141
8.6- Circuito usado em testes de determinação de erro no particionamento de transistores	142
8.7- Técnica através da qual simuladores como SPICE contam o número de retângulos de dreno e fonte	143
8.8 - Exemplo do <i>lay-out</i> de um dos amplificadores Miller	145
8.9- Típico erro falso acusado por DRC que não analisa os erros globalmente	146
8.10- Partição do par diferencial em geometria centróide, para amenizar os efeitos de imperfeições na construção e tornar o bloco mais robusto e resistente a falhas	150

LISTA DE TABELAS

1 Relação entre a razão f_2/F_u e a margem de fase do sistema	49
2 Tamanhos originais obtidos pelo programa de síntese	110
3 Tamanhos corrigidos para adequação ao <i>snap</i> e redução de <i>off-set</i>	110
4 Especificações mais importantes dos quatro circuitos gerados	110
5 Características variadas para obtenção das quatro versões	111
6 Cálculo da máxima taxa de variação da saída em função de um degrau de diferente amplitude na entrada	114

LISTA DE SÍMBOLOS

A_v - Ganho de tensão
 $A_v(s)$ - Ganho de tensão em frequência
 BW - Bandwidth - Largura da banda passante
 C_l - Capacitâncias parasitas que influem no cálculo de C_c
 C_c, C_m - Capacitor de compensação
 C_{gb} - Capacitância porta-substrato
 C_{bd} - Capacitância substrato-dreno
 C_{bs} - Capacitância substrato-fonte
 C_{gd} - Capacitância porta-dreno
 C_{gs} - Capacitância porta-fonte
 C_{ds} - Capacitância dreno-fonte
 CMRR - *Common Mode Rejection Ratio* - Relação de rejeição ao Modo comum
 C_{ox} - Capacitância do óxido
 E_x - Campo elétrico longitudinal
 f_u - Frequência de ganho unitário
 g_d, g_s - Condutância de dreno, condutância de fonte
 g_{ds} - Condutância entre dreno e fonte
 g_m - Transcondutância de porta
 g_{mb} - Transcondutância de bulk
 g_{mg}, g_{ms}, g_{md} - Transcondutâncias de porta, fonte e dreno, no modelo EKV
 g_{oxy} - Condutância de M_x somada à condutância de M_y
 I_f, I_r - *Forward Current, Reverse Current* - Corrente direta e Corrente inversa
 K_p - Constante de transcondutância (equivalente a $\mu * C_{ox}$)
 L - Comprimento do transistor
 L_d - Difusão lateral
 L_{eff} - Comprimento eficaz do transistor
 GBW - Gain Bandwidth - Produto ganho-banda
 n_i - Concentração intrínseca
 N_{sub} - Concentração de portadores no substrato
 P_d - Frequência do pólo dominante
 PM - Phase Margin - Margem de fase
 P_{nd} - Frequência do pólo não-dominante
 PSRR - *Power Supply Rejection Ratio* - Relação de rejeição à fonte de alimentação
 quads - Função MATLAB que implementa o modelo Vladimirescu

Q_{inv} - Carga elétrica acumulada na zona de inversão
R_l, g_l - Resistência de carga, condutância de carga (R_l = 1/g_l)
SR - *Slew Rate*
t_r - *Rising time* - Tempo de subida
U_t - Tensão termodinâmica
V_{an}, V_{ap} - Tensão de Early para dispositivos NMOS e PMOS de tamanho mínimo
V_{ch} - Tensão elétrica ao longo do canal
V_{dd} - Alimentação positiva
V_{ds} - Tensão dreno-fonte
V_{dsat} - Tensão dreno-fonte de saturação
V_e - Tensão de Early
V_{fb} - *Flat Band Voltage* - Tensão de bandas planas
V_{gs} - Tensão porta-fonte
V_{il} - Máxima tensão de entrada ainda equivalente a nível baixo
V_{ih} - Mínima tensão de entrada ainda equivalente a nível alto
V_{ss} - Alimentação negativa ou terra
V_t - *Threshold voltage* - Tensão limiar
W - Largura do transistor
WL - Fator de forma do transistor, equivalente a W/L
Z - Frequência de zero
Z₁, Z₂ - Impedâncias relacionadas aos elementos parasitas
ε_{ox} - Permissividade do óxido
γ - Coeficiente de efeito de corpo
λ - Coeficiente de modulação do comprimento do canal

RELAÇÃO DE ANEXOS

- 1 Resolução das equações de malhas do inversor *CASCADE*
- 2 Codificação MATLAB do algoritmo que implementa o modelo de transistor usado neste trabalho

RELAÇÃO DE GRÁFICOS

- 2 Estudos relativos à capacitância do transistor MOS
 - 2.1 Gráfico de capacitância por área do transistor ($W \cdot L$)
 - 2.2 Gráfico de capacitância por tempo, em um circuito de teste de determinação do perfil capacitivo
 - 2.3 Extração dos dados do gráfico anterior para montagem de um gráfico de capacitância por tensão DC aplicada aos terminais
 - 2.4 Capacitância da associação de um transistor PMOS em paralelo com NMOS (variação máxima = 53.1%)
 - 2.5 Capacitância da associação de um transistor PMOS em paralelo com NMOS invertido (variação máxima = 32.8%)
 - 2.6 Capacitância da associação de um transistor NMOS em paralelo com NMOS invertido (variação máxima = 35.5%)
 - 2.7 Capacitância da associação de um transistor PMOS em paralelo com PMOS invertido (variação máxima = 32.7%)
 - 2.8 Gráfico comparativo entre a resposta AC de um amplificador com capacitor simples e outro usando um capacitor MOS, com mesmo valor nominal de capacitância
 - 2.9 Gráfico comparativo entre a resposta temporal de um amplificador usando um capacitor simples e outro usando um capacitor MOS
- 4 Gráficos de teste usados na determinação do erro máximo entre as respostas do modelo implementado e do simulador
 - 4.1 Comparação entre Corrente de dreno e Tensão Porta-fonte com dados oriundos do SPICE (linha cheia) e MATLAB (pontilhada)
 - 4.2 Erro percentual entre a curva $I_d \times V_{gs}$
 - 4.3 Comparação entre Corrente de dreno e Tensão Dreno-fonte
 - 4.4 Erro percentual entre a curva $I_d \times V_{ds}$
 - 4.5 Comparação entre Transcondutância e Tensão Porta-fonte
 - 4.6 Erro percentual entre a curva $g_m \times V_{gs}$
 - 4.7 Comparação entre Condutância de saída e Tensão Dreno-fonte
 - 4.8 Erro percentual entre a curva $g_{ds} \times V_{ds}$
- 5 Análise térmica da performance do amplificador operacional em termos de *off-set* referido à saída

- 6.1 Gráfico ilustrando a possibilidade de diminuir a área total de um circuito através do aumento da relação de transcondutâncias k
- 6.2 Redução do tamanho dos transistores por efeito do aumento da razão k
- 6.3 Redução do valor do capacitor de compensação por efeito do aumento da razão k
- 6.4 Aumento da corrente total por efeito do aumento da razão k
- 7 Dados extraídos de uma planilha eletrônica, contendo 28 projetos diferentes de amplificadores operacionais na topologia Miller
- 8.1 Desvio percentual da corrente de dreno de um transistor quando este é particionado para *lay-out*
- 8.2 Desvio percentual da corrente de dreno de um transistor levando em conta unicamente o efeito de canal estreito

RESUMO

Os continuados progressos tecnológicos na área de fabricação de dispositivos semicondutores aliados à maturação no projeto de circuitos mistos convergiram na produção de *System on Silicon* [1]. Essa técnica permite a integração dos vários blocos constituintes de um sistema genérico em um circuito integrado, onde operam em conjunto e harmonia circuitos digitais e analógicos. Este trabalho se propõe a relatar de forma didática e abrangente as experiências acumuladas ao longo do processo de produção de circuitos analógicos CMOS como o amplificador operacional, para aplicações em Instrumentação. Para tanto, três versões de modelos de transistor MOS são propostos e analisados e uma implementação de acordo com as necessidades de projeto é realizada. Para elevar o grau de automação na produção de blocos analógicos, rotinas iterativas de cálculo de tamanhos e polarizações de transistores foram implementadas, de modo a aumentar a qualidade dos circuitos projetados e simultaneamente reduzir sensivelmente o tempo gasto na fase de síntese. Os programas destinam-se ao cálculo do tamanho e das polarizações de todos os transistores de um dada topologia, mas permitem que, sem mudanças estruturais significativas, se consiga sintetizar também amplificadores em outras configurações. Um estudo minucioso sobre técnicas alternativas de implementação de capacitores como elementos de compensação é demonstrado, em conjunto com circuitos de auxílio para determinação do perfil dessa capacitância e suas não-idealidades. As fases seguintes do projeto também fazem parte do escopo deste trabalho, ou seja, *lay-out*, verificação das regras impostas pela tecnologia usada (ES2 - 1.2 μ m) e extração. Todas essas etapas foram executadas com ferramentas de software desenvolvidas dentro do Grupo de Microeletrônica (GME). Dessa maneira, todas as fases da produção de um sistema analógico são especificadas, dando subsídios suficientes para uma análise crítica do processo de desenvolvimento de blocos analógicos usando tecnologia digital. Finalmente, exemplos numéricos dos vários amplificadores operacionais projetados são dados, validando, através de uma análise comparativa, as rotinas de síntese e a qualidade do *lay-out*. Os circuitos tiveram seus *lay-outs* levantados e conferidos e foram enviados para fabricação do primeiro circuito analógico integrado do Laboratório de Instrumentação Eletro-eletrônica.

ABSTRACT

The continuous technological developments in the field of semiconductor devices manufacture, together with the mixed circuits design have converged to Systems on Silicon production. These techniques allow the complete integration of the numerous blocks of a generic system into a unique analog and/or digital integrated circuit. The aim of this work is to report, in an instructive and comprehensive way the experiences acquired in the production of CMOS integrated analog circuits, like the operational amplifier, for Instrumentation applications. For that, three versions of the model of the transistor (Quadratic, Vladimirescu and EKV) are proposed and analyzed and one implementation is done according to the design needs. To increase the degree of automation in the analog block production and design, iterative routines to calculate size and bias are implemented, in order to improve the quality of the projects and at the same time reduce the time spent in the synthesis phase. The program's main purpose is to determine the size and bias of all transistors in a given topology, though it's also possible to design an amplifier in different configurations, with minimum software modifications. A detailed study was carried out concerning alternative ways of implementing the compensation capacitor, together with the circuits used for capacitance-voltage curve determination. The following steps in an analog project are also subject of interest in this work, i.e., *lay-out*, rules checking and extraction. All of these steps were carried out using tools developed in the Microeletronic Group (GME). In this way, all the aspects in an analog system design are specified, allowing critical analysis of analog blocks design using digital technology. Finally, numeric examples of four operational amplifiers are given, demonstrating the efficiency of the routines and the *lay-out* quality. The lay-out of all the circuits was checked and sent to fabrication, in the first experience of analog integrated circuits of the Eletric-eletronic Instrumentation Lab.

1 Introdução

Nos últimos anos, os sistemas digitais conheceram um progresso qualitativamente muito expressivo, culminando nas linguagens de descrição de alto nível, que, em combinação com softwares apropriados, geram *lay-out* de circuitos com mais de 100.000 portas [2]. Infelizmente, os sistemas analógicos ainda não têm ferramentas e recursos similares sendo que os mais simples blocos analógicos ainda são manualmente projetados e encontra-se muito pouca automação neste setor, tanto na fase de síntese quanto na fase de *lay-out*. A natureza eminentemente analógica da maioria dos fenômenos que nos rodeiam garantem que sempre haja necessidade da utilização de circuitos que implementam funções analógicas.

Mesmo assim, cresce o interesse pela integração dessas duas classes de circuitos, criando uma terceira: os circuitos mistos [3]. Integrar sistemas inteiros, compostos de funções analógicas e digitais, em um circuito integrado tornou-se uma saída interessante que proporciona várias vantagens. Especialmente em aplicações de uso repetido, isto é, quando são necessárias várias réplicas do mesmo sistema, a transformação de sistemas discretos em equivalentes integrados é ainda mais recompensadora economicamente. Um caso típico dentro da Instrumentação são os condicionadores de sinal que estão presentes ao lado de cada transdutor em um sistema de controle distribuído, por exemplo. As funções desempenhadas por esses condicionadores são amplificação de sinal, filtragem, conversão A/D, codificação, operações aritméticas variadas, etc. Geralmente, essas funções são implementadas através de blocos discretos, necessitando assim uma placa de circuito impresso para interconexão destes blocos. Se fosse possível projetar um circuito integrado que contivesse todas as funções implementadas diretamente, a redução de área e de potência compensaria o custo de produção de circuitos integrados em larga escala. A atual tecnologia permite a construção desse tipo de sistema misto, que integra funções analógicas e digitais com o mínimo possível de elementos externos.

Além da instrumentação, outros setores podem ser beneficiados com essa técnica. Qualquer aplicação que requeira uma quantidade expressiva (superior a 1000 unidades/mês, no atual contexto tecnológico) de dispositivos integrados e que tenha um mínimo de complexidade para justificar o processo de integração é candidata potencial à transformação em equivalente integrado. Por exemplo, um sistema telefônico que incluía

um cancelador de eco e de ruído, compressão de dados, codificação e modulação é passível de integração, desde que tenha uma demanda que justifique o investimento inicial.

Para que realmente exista a viabilidade econômica no processo de integração de sistemas, é necessário que o custo unitário de cada circuito integrado seja suficientemente baixo. Por isso, a fabricação de circuitos mistos, isto é, circuitos eletrônicos integrados compostos de parte analógica e digital sob o mesmo encapsulamento, a preços competitivos deve ser feita, preferencialmente, com tecnologia digital, por duas razões: primeiro porque o custo do milímetro quadrado em uma tecnologia digital é inferior ao custo da mesma área em tecnologia analógica e em segundo por que a tecnologia digital sempre está um passo à frente da analógica, seja no comprimento mínimo dos transistores ou na disponibilidade de camadas. Naturalmente, uma implementação mais densa de circuitos analógicos poderia resolver este problema, mas infelizmente, os projetos analógicos não são como os digitais na característica de construção, isto é, pode-se fazer um grande circuito digital unicamente através da conexão de blocos básicos que desempenham funções determinadas, sem um aumento considerável na complexidade do conjunto. Em analógicos, por outro lado, a inclusão de novos blocos dificilmente dá-se de forma isolada, repercutindo no funcionamento do circuitos adjacentes, obrigando assim o projetista a levar em conta a inclusão do bloco inserido no desempenho total do conjunto. Por isso, as tecnologias digitais são preferidas em relação às analógicas na fabricação de sistemas mistos de baixo custo unitário. O processo de fabricação de circuitos analógicos é utilizado quando os requisitos da aplicação demandam um desempenho superior, não atingido pelo processo digital.

Transformar um sistema discreto em integrado é um processo que demanda muito esforço, pois implica em rever todas as partes integrantes do sistema e refazê-lo adequando-se às necessidades e idiosincrasias da realidade dos circuitos integrados. Para que esse processo tenha êxito, devem existir vantagens em número suficiente para o capital investido nesse processo retorne e ainda dê lucros. Entre essas vantagens, podemos destacar:

* Redução dos custos de produção - A produção de dispositivos semicondutores em larga escala permite um decréscimo sensível no custo unitário de cada CI (Circuito Integrado). Esse custo, quando comparado ao necessário para implementação discreta do sistema, torna-se ainda menos expressivo, fazendo da redução dos custos uma das mais atrativas características dessa técnica.

* Redução da potência consumida - Especialmente em sistemas onde os transdutores e os condicionadores são alimentados através do próprio barramento do sistema de controle, o consumo individual de cada nó é um dado muito relevante, pois limita o número máximo de dispositivos conectados ao sistema ao mesmo tempo.

* Redução da probabilidade de erros - Uma vez que as operações são executadas dentro do integrado, a quantidade de conexões com o mundo externo cai consideravelmente, isto quando não se limita à alimentação e saída de dados, e com isso a probabilidade de erro cai proporcionalmente, uma vez que estas operações internamente executadas tem uma confiabilidade maior. Pode-se dizer então que a probabilidade de falhas de um sistema integrado é inferior àquela do equivalente discreto.

* Facilidade de uso - Quando um sistema discreto apresenta alguma falha, um profissional capacitado faz uma análise de blocos para determinar a origem do problema. Quando essa é descoberta, uma substituição das partes defeituosas é feita e o sistema geralmente volta ao funcionamento normal. Sob a ótica do usuário final que utiliza circuitos integrados, buscar a origem do problema é uma tarefa sem sentido e como o custo unitário do circuito é baixo, a sua reposição tem um custo agregado inferior ao do processo anterior. Naturalmente esta idéia não é válida quando se está desenvolvendo só circuitos.

* Portabilidade - Em algumas aplicações específicas a diminuição do tamanho dos componentes é de vital importância, não só pela economia de espaço físico que proporciona, mas também pela redução proporcional dos elementos parasitas que degradam a performance do sistema.

O objetivo primeiro deste trabalho é dar uma visão geral de circuitos analógicos integrados, enfatizando a possibilidade de produzir-se circuitos mistos dedicados a uma aplicação em especial. Baseado em um conjunto de especificações, o projetista deve ser capaz de produzir um circuito eficiente em um tempo suficientemente pequeno, e para isso, técnicas de síntese semi-automática são propostas. A ênfase maior do trabalho é o projeto do amplificador operacional, célula básica dos circuitos analógicos, e para isso, as rotinas foram desenvolvidas para uma topologia em especial: o amplificador Miller. Usando estes recursos, juntamente com o simulador SPICE, quatro versões de operacionais foram projetadas e enviadas para fabricação. Além disso, novas técnicas e propostas ainda fazem parte do trabalho, retratando o estado atual do desenvolvimento de circuito analógicos integrados.

Este trabalho foi desenvolvido a partir de um acordo firmado entre a Universidade Federal do Rio Grande do Sul e a Universidade Católica de Louvain-la-Neuve, na Bélgica. Através deste acordo, a universidade belga se propôs a colaborar com conhecimentos e tecnologia, proporcionando assim que a amplitude do trabalho desenvolvido fosse ainda maior. Em especial os professores Paul Jaspers, Altamiro Suzim e Jean-Pierre Colinge foram os maiores responsáveis pelo sucesso da interação entre as universidades. Através de um curso de 60 horas ministrado pelo Professor Paul Jaspers na UFRGS e de um estágio de um mês na UCL, foi possível alavancar de maneira substancial e sólida os avanços obtidos a partir deste trabalho.

2 REVISÃO BIBLIOGRÁFICA

Em 1960, a tecnologia de fabricação de circuitos integrados já estava desenvolvida e funcional, permitindo construir um circuito várias ordens de grandeza menor que o correspondente discreto. Durante muito tempo, essa tecnologia foi dirigida exclusivamente para aplicações digitais, entre outras razões devido a sua previsibilidade e consequente controlabilidade facilitada desse tipo de circuito. Posteriormente, surgiu a necessidade de implementar funções analógicas em circuitos integrados, o que criou, também, a necessidade de desenvolvimento de ferramentas de auxílio a este tipo de projeto, como por exemplo, modelos de transistores que permitissem uma descrição mais precisa da característica I-V e de pequenos sinais. A tecnologia disponível nesse tempo contava somente com um tipo de dispositivo, preferencialmente NMOS, mas a ausência de complementaridade tornava os projetos sensivelmente mais complexos. O material usado para a porta do transistor era o alumínio e a espessura do óxido estava em torno de $15\mu\text{m}$. Com essa tecnologia, em inversão forte, o ganho intrínseco dificilmente superava 100 vezes, fazendo dos amplificadores de dois estágios circuitos com ganho inferior a 10.000 vezes (80dB). A inclusão de mais estágios de ganho esbarrava em problemas de estabilidade e em dificuldades de implementação de capacitores de compensação, de forma que estudos deveriam ser conduzidos para melhorar este desempenho.

Mesmo com estas dificuldades, em 1977, já haviam sido publicados trabalhos consistentes no sentido de modelar com satisfatória precisão o comportamento do transistor em regime de inversão fraca [4]. No trabalho citado, os autores propõem um modelo simples para descrever o comportamento do transistor em inversão fraca, com resultados experimentalmente confirmados. Naturalmente, o objetivo deste estudo era polarizar os transistores em zonas de baixo consumo mas alto potencial de amplificação. Os contínuos desenvolvimentos tecnológicos tornavam cada vez mais eficientes os circuitos analógicos, não só pela redução da espessura do óxido mas pela disponibilização de um maior número de camadas, maior precisão na fabricação dos dispositivos (advinda da maior controlabilidade dos processos envolvidos na fabricação), menor probabilidade de erro e desalinhamento, etc [1]. Esses progressos aconteceram e acontecem ainda hoje; prova disso é que a tecnologia usada ao longo deste trabalho foi a ES2 - $1.2\mu\text{m}$ e quando este estava terminado, já estava disponível no mercado a ES2 - $1.0\mu\text{m}$. Embora se soubesse que esses progressos só tendem a aumentar, outras técnicas

de melhoria de desempenho foram sendo investigadas, como por exemplo, a utilização de topologias alternativas, no caso específico dos amplificadores operacionais. A idéia de colocar os estágios em cascata foi aos poucos sendo substituída pela técnica *Cascode*, que prevê sensível aumento de ganho sem consequências negativas para a estabilidade do circuito [5].

Para viabilizar projetos analógicos completos, ainda seria necessário o desenvolvimento de ferramentas que tornassem os sistemas mais simples de serem projetados. A modelagem analógica do transistor foi o primeiro passo, embora vários exemplos possam ser citados: por exemplo, em 1980, Hodges [6] propôs um estudo detalhado sobre a construção dos blocos básicos de qualquer sistema de conversão de sinal, como resistores e capacitores de alta precisão relativa, chaves, comparadores, etc. Mais tarde, em 1982, [7] propôs um modelo analítico para o comportamento do transistor, com a característica de ter uma derivada da corrente de dreno contínua desde a inversão fraca até a inversão forte, com boa aproximação na inversão moderada. O modelo foi usado em simulação e provou ser de boa eficiência. Posteriormente, em 1984, [5] mostrou a viabilidade de projetar amplificadores operacionais voltados a obter determinadas características de especificação do usuário, refletindo assim o fato de já se dominar com mais eficiência a síntese destes circuitos. Para este fim, o referido trabalho faz um extensivo uso de técnica *Cascode*, com o objetivo de, entre outras coisas, aumentar a rejeição do circuito clássico à variação de tensão de alimentação.

A publicação de tutoriais, como [13], disseminava não somente o equacionamento de circuitos completos, mas também indicava maneiras de iniciar um projeto manual, além da implicação de fatores como off-set, ruído, CMRR, PSRR, etc, no desempenho do conjunto. Posteriormente, [8] apresentou uma técnica que permitia a integração de circuitos a capacitor chaveado, usando tecnologia digital com porta de silício, implementando estes capacitores através da polarização de transistores na região de acumulação. Uma das mais importante características deste trabalho é a criação de certas técnicas a serem aplicadas a uma grande quantidade de circuitos, criando a possibilidade da extensão do trabalho para versões automáticas. A associação série e paralela proposta em [9] é uma importante contribuição no sentido de melhorar o desempenho de um transistor individualmente, através de sua combinação com outros transistores. Com isso, o alto ganho, que até então só era possível à custo de um transistor comprido, que, em contrapartida, oferecia uma frequência de transição pequena, pode ser agora obtido através da associação de dois transistores individualmente curtos, com alta frequência de corte, e ganho igualmente alto. Este tipo de melhoria permite que usando uma mesma topologia, obtenha-se resultados melhores,

aumentando o ganho intrínseco da célula básica do circuito. Uma característica em comum de todos os exemplos citados acima é o fato de serem eminentemente manuais; ou seja, as técnicas de síntese dos amplificadores, a utilização de *Cascode*, o projeto de chaves analógicas, resistores e capacitores, era feito de maneira manual, com auxílio de equações, tabelas, ábacos, ou qualquer outra fonte de dados. A metodologia empregada fazia uso da força bruta do cálculo aliada à criatividade do projetista, para convergir a uma resposta final, sendo a experiência uma qualidade de muito valor, que por vezes definia o sucesso ou o fracasso do projeto.

A medida que a descrição dos circuitos passava a ser mais detalhada e os modelos utilizados forneciam erros menores, as técnicas de síntese, embora ainda completamente manuais, começavam a incluir parâmetros mais detalhadas, que refletem um aspecto bastante definido do funcionamento do circuito. Exemplo disso é o PSRR (*Power Supply Rejection Ratio*), CMRR (*Common Mode Rejection Ratio*), SR (*Slew Rate*), tr (*Time response*), etc. Esses parâmetros foram introduzidos na etapa de análise permitindo assim que os circuitos projetados tivessem estas características. Embora tendo aumentado o nível de complexidade da síntese, essa inclusão permitiu uma previsibilidade maior relativa ao desempenho do bloco analógico.

Apesar dos avanços percebidos na área de projeto de circuitos analógicos, estes são pouco expressivos quando em comparação aos acontecimentos no universo digital. A proliferação de ferramentas automáticas de síntese, verificação, extração e *lay-out* para circuitos digitais permitiu que um projeto qualquer pudesse ser implementado totalmente através de uma simples descrição funcional. Esse grau de automação torna o projeto mais simples, rápido e eficiente, pois os algoritmos que formam esses programas geralmente convergem para respostas próximas daquelas ditas ótimas. Situação similar não sucedeu no campo analógico, onde o nível de automação, ainda hoje, é relativamente baixo, restringindo-se a alguns casos de aplicações específicas. A complexa síntese de circuitos analógicos era feita através da técnica de tentativa e erro, com auxílio de um simulador elétrico. Esse processo, além de dispendioso, fornecia respostas pouco otimizadas. A verificação de regras também tornou-se um inconveniente, no momento em que se passou a utilizar extratores feitos para circuitos digitais em circuitos analógicos. O problema advém do fato de que os extratores usados em circuitos digitais não dão muita importância à natureza resistiva ou capacitiva das ligações, fator este da maior relevância em circuitos analógicos. Dessa maneira, por ter um outro tipo de prioridade, é possível que estes extratores não desempenhem o seu papel corretamente quando ao analisar um circuito analógico. Também o processo de *lay-out* torna-se mais complexo, pois as ligações influenciam de maneira decisiva no funcionamento do

circuito, de forma que a construção automática do *lay-out* pode criar elementos parasitas em número muito grande. Por essas e outras razões, o grau de automação presente nas etapas de projeto de um circuito analógico ainda não é comparável àquela observada nos digitais.

Uma contribuição muito importante foi o trabalho de [10], que trazia não só o modelamento contínuo do transistor nas três regiões de operação, mas também um método de cálculo automático que gerava um amplificador operacional completo, a partir das especificações do usuário. A linha mestra do trabalho era a grandeza gm/Id (relação de transcondutância pela corrente de dreno) e a partir deste valor, um algoritmo fazia um ciclo de cálculos sequenciais e mostrava como resultado o tamanho, a polarização e a compensação de todos os transistores do amplificador Miller com compensação resistiva. A maior diferença entre o referido trabalho e este é que no último, o espaço de projeto está restrito à inversão forte (por razões que serão posteriormente explicadas) e por isso o método de síntese é inteiramente diferente.

Talvez ainda mais relevante no contexto deste trabalho seja citar [11], que propõe a utilização de amplificadores operacionais em inversão forte em filtros sigma-delta. Por utilizar um método de síntese manual, guiado por premissas simples, os resultados, a nível de tamanho de transistores e potência dissipada, se mostram passíveis de otimização, representando desta forma um exemplo concreto da necessidade de utilização de rotinas de síntese nestes casos. No referido trabalho, o autor propõe um método de síntese relativamente clássico, conforme a literatura em geral. Nesse método, a desconsideração de várias grandezas importantes torna o complexo projeto do amplificador Miller em uma sequência de cálculos que podem inclusive ser manualmente realizados. O inconveniente é que o projetista obriga-se a fazer ajustes iterativos com o auxílio de um simulador e eventualmente distancia-se muito da solução original provida pelo método.

As mais recentes técnicas de síntese automática de amplificadores operacionais e blocos analógicos em geral permitem a construção desses circuitos de acordo com as necessidades impostas pelas especificações do usuário, sem limitação de topologia fixa. Para isso, utiliza-se os conceitos de arquitetura e hierarquia. Arquitetura representa um bloco funcional capaz de implementar uma ou mais funções. Como exemplo mais comum, cita-se o espelho de corrente: as opções disponíveis são espelho simples, espelho *Cascode*, espelho Widlar, espelho Wilson etc. Para um circuito em especial, o programa tenta as arquiteturas mencionadas até que o desempenho do sistema seja adequada. Hierarquia consiste em decompor um circuito complexo como o amplificador

operacional ao nível de transistor e fazer com que estes elementos básicos ofereçam a especificação desejada. Com esses conceitos e outros, programas de síntese bastante poderosos podem ser construídos. É fácil perceber que o projeto de circuitos analógicos teve um salto expressivo, considerando-se a técnica manual em um extremo e as técnicas multi-topologia em outro. Na verdade, esse salto teve algumas escalas, mas sua magnitude é grande devido à importante dificuldade inicial de construir algoritmos capazes de comparar desempenhos de blocos e caracterizar corretamente o seu funcionamento. Para que o esforço destinado ao desenvolvimento de um *software* como este seja recompensado, a demanda de circuitos analógicos deve ser bastante expressiva, do contrário, técnicas de topologia única são empregadas, o que reduz sensivelmente a complexidade. O passo seguinte, que está sendo investigado, é o projeto automático conjunto de circuitos mistos, de forma a poder caracterizar o funcionamento de um circuito qualquer e obter um projeto detalhado e completo capaz de reproduzir as especificações solicitadas na caracterização.

Assim, pode-se constatar que o progresso no campo de desenvolvimento de circuitos analógicos ainda está em andamento e longe de uma solução final. Um dos fatores que mantém a questão em aberto é a ausência de um modelo único que responda às necessidades do projetista, o que dificulta sensivelmente o trabalho de síntese e simulação. Além disso, e talvez ainda mais importante, é que não faz muito tempo que se começou a dar a verdadeira importância para circuitos analógicos e mistos integrados, o que retardou bastante o desenvolvimento de tecnologias e técnicas referentes aos processos analógicos. Historicamente, o desenvolvimento de circuitos analógicos integrados começou simultaneamente nos Estados Unidos e Europa, mas por questões políticas e filosóficas, os primeiros decidiram investir de maneira intensa nas técnicas digitais. Os países da Europa, ao contrário, sempre mantiveram suas pesquisas nessa área, o que justifica o seu estado atual de liderança no setor.

3 Circuitos Analógicos

Circuitos analógicos integrados CMOS são compostos, em 99% dos casos, de transistores e capacitores. Os resistores têm papel fundamental na maioria dos blocos analógicos discretos, mas dificuldades tecnológicas tornam seu uso pouco interessante em aplicações integradas, não só por sua natural baixa eficiência (relação entre resistência e área necessária para implementação [$\Omega/\mu\text{m}^2$]), mas também pela pequena controlabilidade do valor absoluto destes resistores. Processos alternativos de fabricação como *Laser Trimming* e Implantação Iônica permitem um ajuste fino no valor dos resistores mas obviamente representam custo adicional e não são disponíveis em todas as tecnologias. Assim, a controlabilidade do valor absoluto da resistência destes elementos é pequena, oferecendo uma tolerância muito grande, embora essa questão possa ser resolvida através da utilização de circuitos que utilizem a relação de resistências (que pode ser mais eficientemente controlada), ao invés de seu valor absoluto. Dessa forma a implementação de resistores é geralmente feita usando técnicas de capacitor chaveado, cuja teoria já é bem conhecida e está muito bem fundamentada [12]. Para valores de resistor inferiores a 300 k Ω aproximadamente, um transistor polarizado na região linear pode ser usado para simular o comportamento de resistor, embora seja necessário assegurar que o transistor não sairá da região linear, pois a impedância do dispositivo aumenta muito quando este entra na região de saturação. Ainda, a polarização do transistor deverá ser mantida constante para que a resistência permaneça constante.

Dessa forma, para aplicações do tipo *Countinuous Time*, onde a introdução de capacitor chaveado não é conveniente, a topologia dos blocos analógicos deve ser composta somente de transistores e capacitores. O Amplificador Operacional é o circuito chave nas aplicações analógicas [13], sendo indiscriminadamente usado em praticamente todas as funções. Suas aplicações mais comuns vão além de simples amplificação de sinal usando elementos de realimentação, sendo usados também em filtros, *buffers*, comparadores, conversores A/D, somadores, etc. Logo, projetar blocos analógicos quaisquer significa projetar amplificadores operacionais com especificações de acordo com as necessidades. Para tanto, algumas das possíveis topologias serão abordadas neste capítulo.

3.1. Inversores

Um dos circuitos básicos de amplificação é o inversor [14,15]. Consagrado no uso digital como implementação mais simples de inversão lógica de um sinal, o inversor é usado aqui com objetivos ligeiramente diferentes. Um gráfico típico do sinal de saída em relação ao de entrada para um inversor genérico juntamente com o esquemático da implementação de um inversor é dado na figura 3.1:

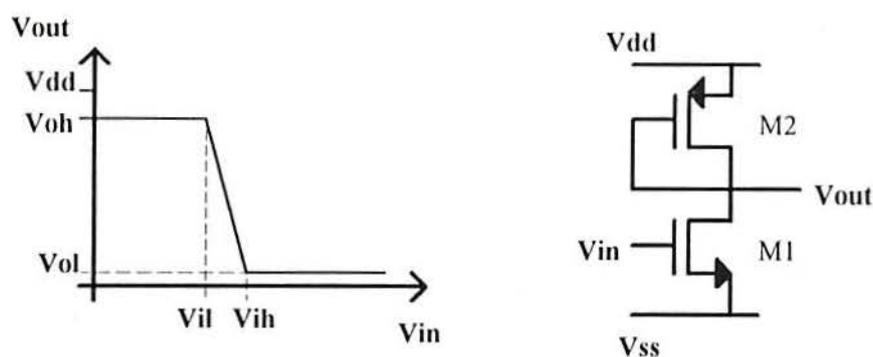


Figura 3.1

A função digital do inversor pode ser facilmente compreendida através do gráfico: se uma tensão de entrada (V_{in}) qualquer menor que V_{il} (Tensão de entrada de nível baixo) a saída vai para nível alto; se, ao contrário, a tensão de entrada for maior que V_{ih} (Tensão de entrada de nível alto), a saída está forçosamente em nível baixo. No circuito este funcionamento pode ser compreendido da seguinte forma: M2 é um transistor carga ativa, ou seja, representa uma carga resistiva; dependendo da tensão V_{in} aplicada, a corrente drenada por M1 (igual a de M2) fará V_{out} assumir valores próximos a V_{dd} e V_{ss} . Infelizmente, este tipo de configuração não permite que V_{out} assumira os valores limites das fontes de alimentação, pelos seguintes motivos:

* Quando V_{in} é insuficiente para gerar corrente no circuito, V_{out} é igual a V_{dd} descontada o módulo da tensão V_t que deve existir entre porta e fonte de M2. Com isso a tensão máxima disponível para V_{out} no caso de entrada em nível baixo é aproximadamente $V_{dd} - |V_{tp}|$.

* V_{out} tende assintoticamente a V_{ss} quando V_{in} gera uma tensão de porta muito grande e a corrente tende a subir muito. Mesmo assim, em ambas extremidades da curva anterior, existe um encurvamento causado pela continuidade das características do dispositivo, impedindo assim uma mudança abrupta de declividade, conforme foi esquematicamente mostrado. Assim, quando a corrente é muito grande, a queda de

tensão na carga ativa é aproximadamente $V_{dd} + |V_{ss}|$, restando muito pouca tensão disponível para M1, de modo que V_{out} tende a assumir o valor da tensão dreno-fonte do transistor 1.

O exemplo dado pode ser construído na forma em que se apresenta ou aplicando o sinal em um PMOS e usando um NMOS como carga. Embora apresentando falhas, as limitações deste circuito podem ser contornadas através do uso de topologias melhoradas, mas o estudo desse bloco básico é de vital importância para a boa compreensão dos circuitos amplificadores. Como foi dito, a região de interesse para o funcionamento digital compreende o intervalo $[V_{ss}, V_{il}] \cup [V_{ih}, V_{dd}]$, sendo que a região formada pelo intervalo (V_{il}, V_{ih}) produz respostas numa região de incerteza, e portanto de pouco interesse para aplicações digitais. Entretanto, para aplicações analógicas, essa região permite que pequenas variações na tensão de entrada propiciem grandes variações na tensão de saída, configurando assim uma amplificação de tensão. Essa propriedade é amplamente explorada no sentido de construir circuitos cuja atuação se restrinja a esse intervalo de comportamento praticamente linear. O ganho de tensão de um circuito amplificador qualquer pode ser calculado como sendo a relação entre a variação de V_{out} em relação à variação de V_{in} , ou, levando o conceito aos limites infinitesimais, $A_v = \partial V_{out} / \partial V_{in}$. Naturalmente, essa interpretação só é útil dentro do intervalo de operação linear, ou seja, quando a saída é proporcional à entrada, pois fora deste o valor do ganho cai praticamente a zero.

A substituição do transistor por equivalentes de pequenos sinais é uma importante ferramenta no auxílio da compreensão e do projeto de circuitos amplificadores [14,15]. Idealmente, um transistor na região de saturação é um dispositivo gerador de corrente (I_d) controlado por tensão (V_{gs}). Na prática, a corrente principal também sofre influência da tensão em outros pontos, assim como de capacitâncias e resistências parasitas, conforme a figura 3.2

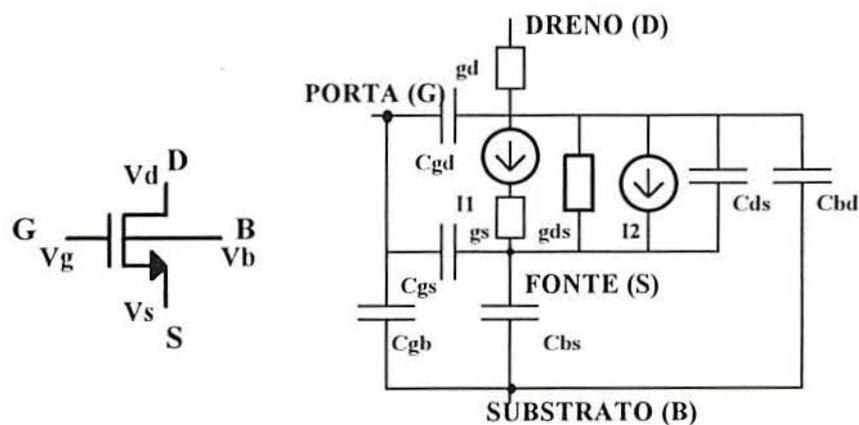


Figura 3.2

Neste modelo de pequenos sinais, a corrente total de dreno é dada por $I_1 + I_2$, onde $I_1 = g_m \cdot (V_g - V_s)$ e $I_2 = g_{mb} \cdot (V_b - V_s)$, g_m representa a transcondutância e g_{mb} a transcondutância de corpo. Entre dois quaisquer terminais do transistor existe sempre uma capacitância, que é desprezada na análise DC, mas desempenha papel vital na análise dinâmica do circuito. Entre elas, C_{gs} é a de maior valor numérico, sendo proporcional à área da porta, embora C_{dg} represente na maioria das vezes uma capacitância que degrada severamente o desempenho do circuito em frequências mais elevadas.

O circuito da figura 3.1 pode ter seu desempenho em DC analisado através do modelo de pequenos sinais simplificado, conforme a Figura 3.3.

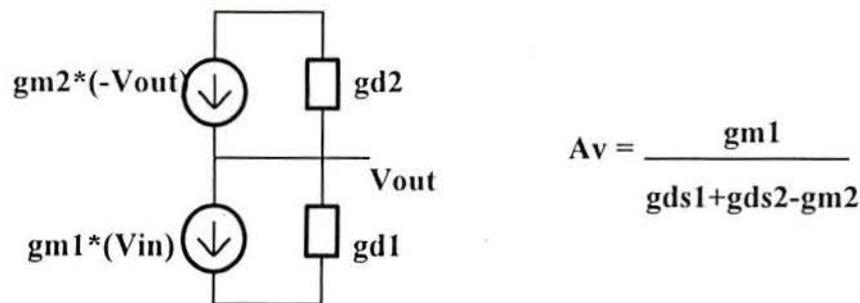


Figura 3.3

Com este tipo de análise pode-se identificar os parâmetros que podem otimizar determinadas especificações, como ganho DC, por exemplo. Deve-se ressaltar também o fato de que essa análise é independente do modelo usado. Qualitativamente, um dos modelos empregados, o modelo quadrático, nos informa que a transcondutância de um transistor pode ser incrementada aumentando sua relação de aspecto (W/L) e GVO ou diminuindo a corrente. Para frequências médias este modelo já não fornece respostas precisas, devendo então ser incluídos os efeitos parasitas. Além disso, a redução da condutância de saída de ambos transistores é interessante pelo incremento de ganho e isso geralmente é feito com o aumento do comprimento do dispositivo

Ainda sobre o inversor da figura 3.1, um grande problema que ele representa é o fato de conectar a entrada com a saída através de C_{dg} de M_1 [14]. Devido ao efeito Miller, o valor da capacitância fica multiplicado por $|A_v| + 1$, degradando seriamente a resposta em frequências mais altas. Para evitar este efeito, a configuração Cascode, que será objeto de estudo mais tarde, é extremamente conveniente.

Uma das maneiras de melhorar o desempenho do inversor é usar uma fonte de corrente em vez de uma carga ativa (no lugar de M2). Como teoricamente a impedância interna de uma fonte de corrente é infinita (e efetivamente se consegue valores bem mais altos do que com carga ativa), existe um substancial aumento do valor do ganho. A tensão V_g (Figura 3.4) determina a máxima excursão positiva do sinal, uma vez que para o bom funcionamento do inversor, deve-se garantir que ambos estejam em saturação: $V_{out} < V_g + V_{tp}$. Igualmente, existe um limite inferior para a tensão de saída: $V_{out} = V_{in} - V_t$.

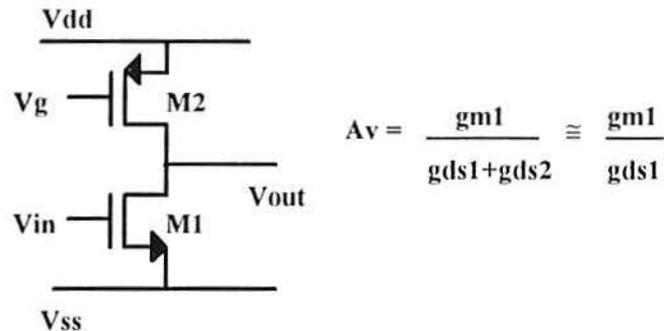


Figura 3.4

Neste circuito, como a corrente em M2 é determinada quase que totalmente por V_g e muito pouco por V_{out} (através do efeito de modulação de canal), a corrente no transistor M1 terá pequena amplitude no funcionamento dinâmico do inversor. Se por um lado este procedimento conduz a um aumento de ganho, por outro introduz a necessidade de projetar também um circuito de polarização como um divisor de tensão formado por um par de transistores ligados como carga ativa, para ser ligado à porta de M2. Como a impedância de M1 é muito alta, ele é propício para receber sinais fracos e amplificá-los algumas ordens de grandeza. Outra maneira de executar este procedimento é através de reflexões de corrente, na qual uma corrente de referência é copiada e amplificada usando transistores com diferentes valores de fator de forma. Entretanto, colocar vários estágios série não é um bom procedimento, pois os elementos parasitas (em especial as capacitâncias) são responsáveis por rotações na fase do sinal, podendo tornar os circuitos potencialmente instáveis. Dessa forma, é costume não colocar mais de três estágios de ganho em série em um circuito, a menos que este esteja munido de um circuito que garanta a estabilidade.

Uma topologia alternativa que combina algumas das características dos blocos anteriores é o *PUSH-PULL*, que apresenta, em determinadas condições, desempenho

superior aos inversores anteriormente apresentados. Seu esquema consta na Figura 3.5 e como se pode perceber prescinde de circuito de polarização.

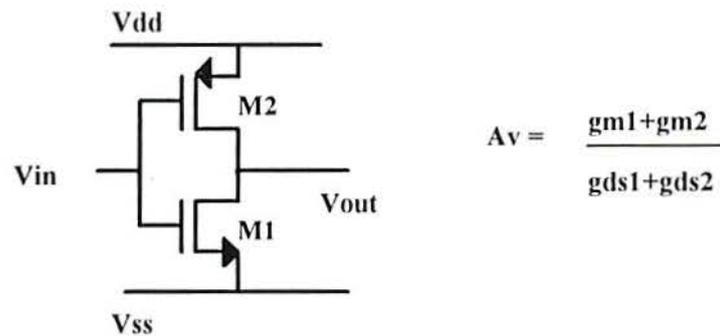


Figura 3.5

A característica de amplificação deste circuito pode ser evidenciada quando V_{in} está variando. Se V_{in} está aumentando, GVO de M1 aumenta e com isso a corrente I_d comum aos dois transistores tende a aumentar, embora a GVO de M2 se reduza na mesma proporção, tendendo a reduzir também I_d . Na verdade, existe um valor de V_{in} intermediário (aproximadamente no meio de V_{dd} e V_{ss}) que causa uma corrente máxima; se V_{in} aumenta ou diminui, a corrente cai, de uma forma aproximadamente gaussiana, enquanto que o ganho de tensão se mantém relativamente constante. O problema intrínseco a este circuito é que, no ponto de corrente máxima, nenhum transistor (M1 ou M2) está na região de máximo ganho, sendo interessante portanto anexar um *Level Shifter* (Deslocador de Nível) na porta de M1 ou M2. Este deslocador de nível pode ser da forma do circuito descrito esquematicamente na figura 3.6.

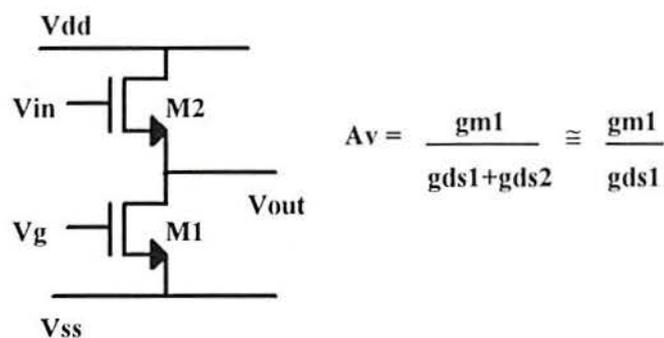


Figura 3.6

Embora o esquema seja semelhante, a presença de dois transistores do mesmo tipo causa um efeito diferente: como M1 opera como simples fonte de corrente, quando V_{in} varia, V_{out} deve acompanhar essa mudança, para que GVO de M2 se mantenha fixa

e a corrente efetivamente se mantenha constante. Dessa forma, o valor de $V_{out} - V_{in}$ é função da corrente de polarização fornecida por M1. Pode-se projetar o bloco deslocador para funcionar com ganho unitário, executando unicamente a função de deslocamento.

3.2 Amplificador Operacional

Embora os circuitos mostrados até aqui permitam um ganho substancial, sofrem de um grave problema que tem repercussões importantes, especialmente para aplicações em Instrumentação: a ausência de entrada para sinais diferenciais [14,15,16]. Como a maioria dos transdutores são montados em configurações que geram sinais de tensão diferenciais, é bastante interessante ter um estágio que permita a entrada de dois sinais, como o da figura 3.7.

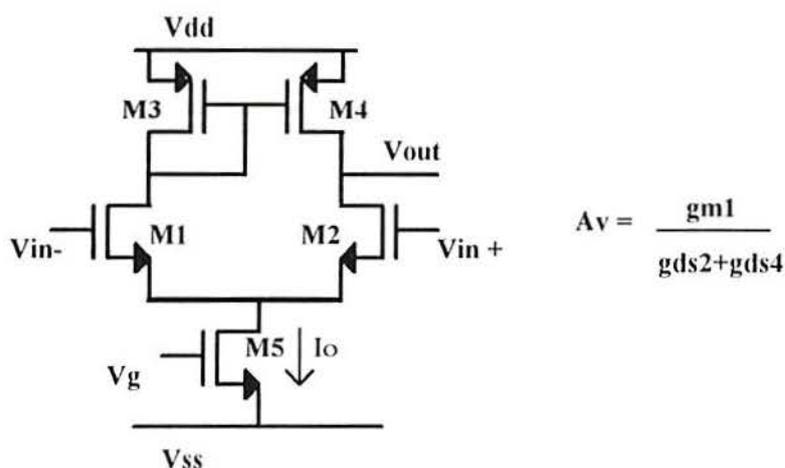


Figura 3.7

A função desse estágio é, inicialmente, servir de entrada para sinais diferenciais, posteriormente fornecer o maior ganho possível e ao mesmo tempo, fazer a conversão do sinal de entrada em um sinal do tipo unipolar. A entrada NMOS foi usada para tirar vantagem do fato de que a mobilidade apresentada por este tipo de dispositivo é maior que a do seu complementar, propiciando assim um ganho maior, embora nada impedisse de usar uma entrada PMOS. A análise deste circuito é simples e usa os princípios da análise dos inversores já mencionados. M5 tem uma configuração de fonte de corrente, não tendo portanto outra função além de polarizar o circuito. A corrente I_o gerada por M5 se distribui equitativamente por M1 e M2 na situação de repouso, isto é, quando $V_{in+} = V_{in-} = 0$. M3 e M4 formam um espelho de corrente, isto é, supondo que o fator de forma de M3 e M4 seja igual, a corrente que passa por M3 gera uma tensão apropriada em sua porta, que está diretamente conectada à porta de M4. Assim a corrente é

"copiada" para M4, posto que a tensão no dreno de M4 não tem influência muito relevante no valor da corrente de dreno. Para entradas em 0V, a saída (Vout) assume um valor geralmente não nulo, que será o ponto em torno do qual haverá excursão de sinal. Se V_{in+} aumenta, a corrente em M2 aumenta (ΔI) de forma aproximadamente quadrática (e, como a corrente gerada por M5 é constante, há uma redução na corrente em M1). Essa corrente diminuída é refletida pelo espelho de corrente para M4, de modo que a lei de Kirchhoff de correntes deve ser observada no nó de saída: uma corrente $I_{o/2} - \Delta I$ em M4 e $I_{o/2} + \Delta I$ em M2 obriga a entrada de corrente ($2\Delta I$) no circuito através do ponto de saída. Como geralmente o referido ponto está ligado a uma porta de outro transistor (alta impedância), não é possível a circulação de corrente DC e por isso, deve existir um rearranjo de distribuição de tensões para garantir que a soma total de correntes seja nula. No caso de ausência de carga ou ligação a uma porta no próximo estágio, pelo fato de que não existe a possibilidade de entrada ou saída de corrente DC pelo nó de Vout, essa tensão aumenta, de modo a simultaneamente aumentar Vds de M2 e diminuir Vds de M4. Para pequenas variações da tensão de entrada, o efeito de modulação do comprimento do canal permite essa acomodação de tensões para que as correntes se equivalham em M2 e M4. Uma vez que quando V_{in+} aumenta Vout varia no mesmo sentido, V_{in+} é dita a entrada não inversora. Todo o raciocínio pode ser empregado para V_{in-} , que conduzirá à conclusão que essa é a entrada inversora. Para perfeito funcionamento, as dimensões de M1 devem ser iguais a de M2, valendo o mesmo para M3 e M4. O tamanho de M5 deve ser definido em função da corrente de polarização, embora não se possa esquecer o fato de que a rejeição a modo comum é função da semelhança de M5 com uma fonte de corrente, de modo que, à medida do possível, é interessante aumentar o comprimento de M5 para aumentar sua impedância. Independentemente da alimentação do circuito, o ponto quiescente de Vout deve coincidir com o ponto quiescente do próximo estágio, seja ele de ganho ou simplesmente um *buffer*, de modo que haja encadeamento entre os blocos e garanta-se o bom funcionamento do conjunto.

Embora o circuito da figura 3.8 tenha um ganho comparável ao inversor simples, seu uso isolado é mais indicado para simples conversões *Dual-single Ended*, pois o sinal de entrada é diferencial e o de saída é referenciado a um potencial fixo. Para o caso de amplificadores operacionais, costuma-se usar mais de um estágio, sendo que o segundo pode ser o inversor. Com isso, os circuitos são uma simples justaposição de dois estágios *a priori* independentes, formando a conhecida topologia Miller, conforme a figura 3.8.

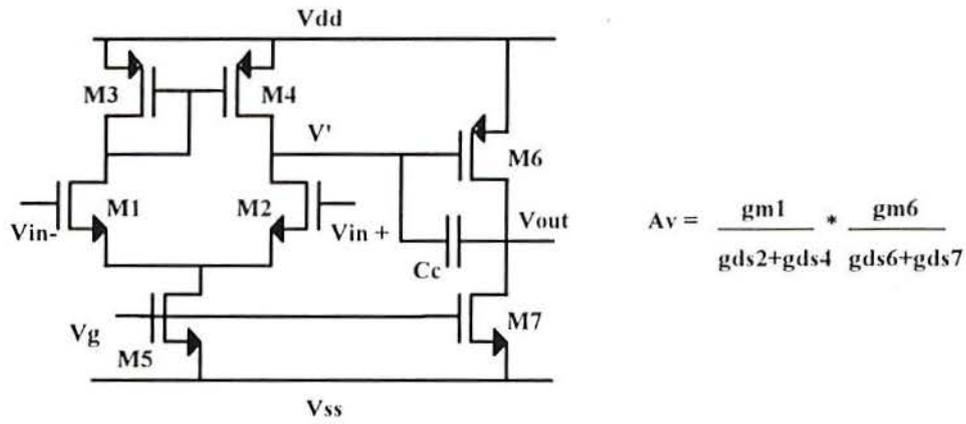


Figura 3.8

O segundo estágio foi anexado ao primeiro com objetivo de dar ganho e, para que as variações na tensão V' estejam dentro do intervalo linear de $M6$, este deve ser um dispositivo complementar ao de entrada, ou seja, entrada NMOS, estágio inversor PMOS. Conforme a equação da figura 3.8, em DC não existe interação entre as correntes do primeiro e segundo estágio, de modo que o ganho total é dado pelo produto dos ganhos individuais. Naturalmente, $M5$ deve polarizar o primeiro estágio de tal maneira que V' coincida com a tensão quiescente do segundo estágio. A inclusão do capacitor tem razões de estabilidade e está intimamente ligado aos outros parâmetros do circuito. A necessidade de sua existência, assim como seu cálculo, serão alvos de estudo em capítulos seguintes. O funcionamento do circuito pode ser dividido em duas partes:

- * Transformação do sinal diferencial em sinal unipolar (através do primeiro estágio)
- * Amplificação de tensão, promovida pelos dois estágios individualmente.

Embora de complexidade aparentemente inexpressiva, o projeto de um amplificador operacional como o da figura 3.8 não é uma tarefa fácil, sendo executada geralmente de forma manual, posto que ferramentas de síntese analógica não são muito comuns e eficientes. A topologia do amplificador Miller é bastante conhecida e seu uso é muito difundido, especialmente para aplicações que não exijam resposta em frequências muito altas. Como este amplificador já foi analisado por muitos autores, um capítulo especial é dedicado à análise do Miller: sua resposta em frequência, ganho DC, banda passante, margem de fase, etc. Embora existam técnicas mais funcionais de melhorar o desempenho deste operacional, sua natureza simples o torna didaticamente interessante para discussão e construção.

Uma das técnicas mais difundidas no aumento de ganho e melhoria da resposta em frequência é o *Gain Boosting* proporcionado pelo estágios *Cascode* [15]. Essa

metodologia apresenta-se eficiente na maioria das aplicações, permitindo um aumento da impedância do transistor, em conjunto com uma redução do efeito Miller (capacitância que liga entrada à saída). A figura 3.9 mostra o equacionamento de um estágio genérico por efeito da inclusão de um transistor *Cascode* (M2).

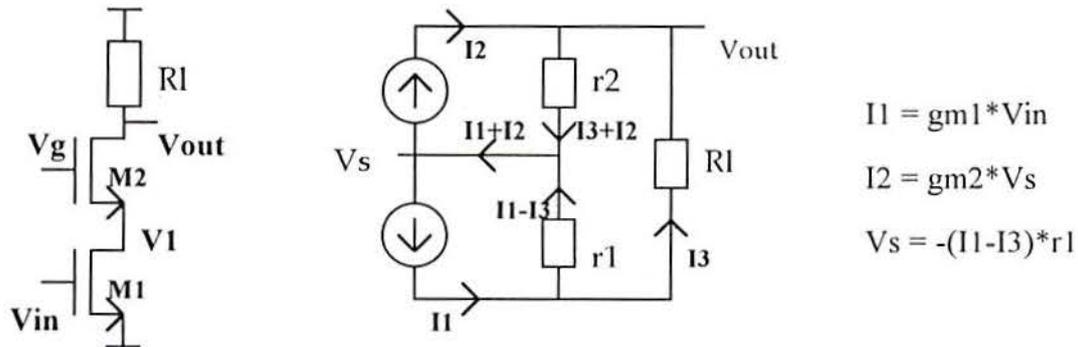


Figura 3.9

A resolução da equação de malha que relaciona entrada e saída consta no Anexo 1, resultando numa equação de ganho como segue:

$$Av = -RL gm1 \cdot \frac{gm2 \cdot r2 \cdot r1 + r1}{gm2 \cdot r2 \cdot r1 + r1 + r2 + RL} \quad Av2 = gm2 \cdot r2 = \frac{gm2}{gds2}$$

$$Av = -RL gm1 \cdot \frac{Av2 \cdot r1 + r1}{Av2 \cdot r1 + r1 + r2 + RL}$$

Equação 1

Ela mostra, entre outras coisas, que uma impedância $rds1$ pode ser aumentada até duas ordens de grandeza pela inclusão de um estágio Cascode. Então, essa impedância será $rds1 * Ai + rds1$ (Ai = ganho intrínseco), ou seja, $rds1 * gm2 * rds2 + rds1$, que é aproximadamente igual a $rds1 * rds2 * gm2$. Como no exemplo essa impedância que foi aumentada está em paralelo com a carga RL , muitas vezes ela pode ser desprezada, por ser muito maior que a carga. É o caso da equação 1, onde a relação que multiplica $gm1$ pode ser negligenciada sem perda de precisão no valor do ganho. A substituição do ganho intrínseco na equação permite visualizar mais facilmente a propriedade de amplificação de impedância.

Dessa maneira, percebe-se que a inclusão de um dispositivo pode aumentar enormemente a impedância de um transistor, causando expressivo aumento de ganho. A

figura 3.10 mostra dois exemplos de aplicações destas técnicas a serem comparados com um circuito comum.

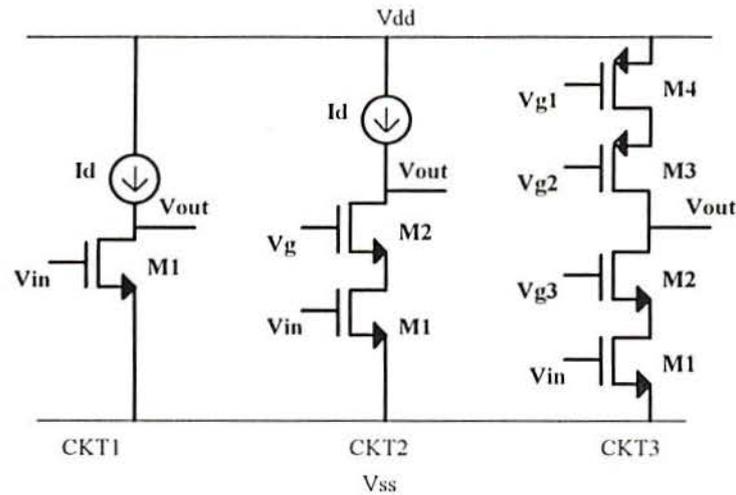


Figura 3.10

O primeiro circuito da figura 3.10 (CKT1) é o inversor padrão com fonte de corrente, cujo ganho total não é mais que o próprio ganho intrínseco do transistor MOS (g_m/g_d), pois a impedância da fonte de corrente foi desprezada, restando somente a impedância de M1. Em CKT2 o transistor M2 foi introduzido em série com M1 de modo a permitir a passagem da mesma corrente (através da estabilização da tensão da fonte de M2 em um nível adequado), porém aumentando substancialmente a impedância a partir no nó de saída em direção a V_{ss} . Com isso, ganhos bem superiores podem ser obtidos. Além disso, a capacitância parasita que liga porta a dreno (C_{gd}), que estava presente em CKT1, ainda pode ser evidenciada em CKT2, embora sua ação degradadora não seja mais observável, pelo efeito de separação da entrada e da saída pela inserção de M2. CKT3 é uma implementação real (pois não tem fonte ideal de corrente) para uma aplicação específica onde a excursão de saída não é grande. M4 representa a fonte de corrente real, enquanto que M3 e M2 são transistores *Cascode* e M1 é o transistor que recebe o sinal de entrada. Tanto a carga como o próprio transistor amplificador têm suas impedâncias aumentadas sensivelmente por efeito *Cascode*, garantindo um ganho bastante superior a CKT1. Entretanto, a colocação em série de vários dispositivos tem uma séria desvantagem: a redução da excursão útil do sinal de saída, posto que todos os transistores devem estar polarizados na região de saturação e deste modo devem ter tensões dreno-fonte suficientemente altas, consumindo parte da tensão disponível pela fonte de alimentação.

Uma forma alternativa de usar a técnica de *Cascode* em amplificadores operacionais é ilustrada na figura 3.11, usando a propriedade de alta e baixa impedância do transistor no terminal de dreno e fonte, respectivamente [5]. O primeiro estágio tem funcionamento muito similar ao estágio diferencial do Miller, sendo que o segundo estágio usa os princípios de inversores e reflexão de correntes.

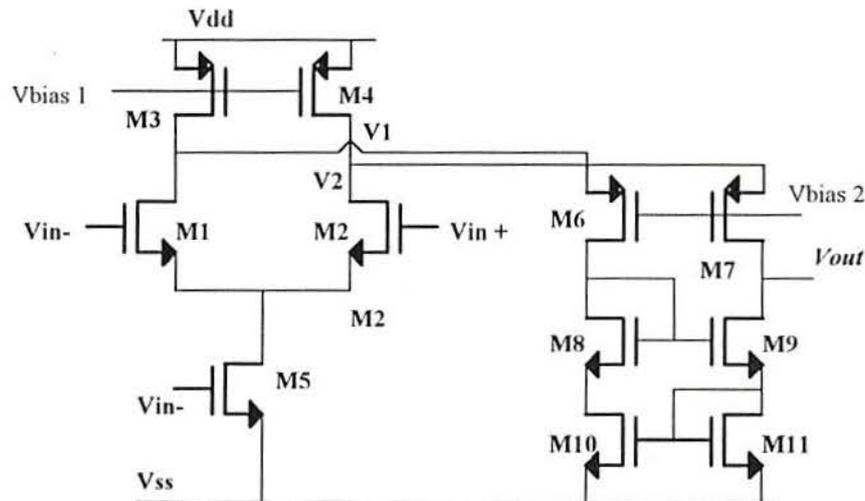


Figura 3.11

Para o bom entendimento do funcionamento deste circuito, é necessária a compreensão da diferença da funcionalidade dos terminais de dreno e fonte do transistor. Por exemplo, supondo-se que M5 gere uma corrente I_o na condição de equilíbrio do circuito, essa corrente não variará de maneira significativa mesmo que V_{in+} , V_{in-} e a tensão no dreno de M5 assumam valores diversos. Esse fato se deve à alta impedância observada no terminal de dreno de M5, impedindo assim variações de corrente. Qualitativamente essa impedância pode ser considerada alta devido ao fato de que grandes variações de tensão de dreno de M5 são necessárias para provocar pequenas variações de corrente. Esse comportamento está em pleno acordo com a descrição teórica do transistor, que diz que sua corrente de dreno é controlada pela tensão de porta-fonte e tem pequena influência frente a variações de V_{ds} .

Exatamente o oposto ocorre com M6 e M7, pois as variações de corrente causadas pelos sinais V_{in+} e V_{in-} são totalmente absorvidas por elementos de baixa impedância, como a fonte de um transistor. A caracterização teórica do transistor dada no parágrafo anterior permite também compreender o funcionamento de baixa impedância, da seguinte forma: se uma fonte externa causa uma variação de corrente, a tensão de fonte do transistor pode se alterar, de maneira a criar um GVO adequado à

passagem da corrente imposta pela fonte externa. Uma vez que grandes variações de corrente são possíveis mediante pequena variação de GVO do transistor, o dispositivo pode ser interpretado como sendo uma baixa impedância.

À luz destes comentários, torna-se mais simples a compreensão do circuito da figura 3.11.

* Conforme exemplos anteriores, M5 funciona como fonte de corrente para polarizar o estágio diferencial.

* M3 e M4 não desempenham mais o papel de espelhos de corrente; agora operam como fontes independentes de corrente que suprem corrente suficiente não só para M1 e M2 mas também para o estágio seguinte.

* M1 e M2 são responsáveis por dividir a corrente disponível proveniente de M5. Se V_{i+} aumenta, a corrente de M2 aumenta e a de M1 diminui. Com isso, a corrente em M7 tende a diminuir (posto que resta menos corrente disponível para o segundo estágio), enquanto a de M6 aumenta.

* M8-M11 são uma simples implementação do espelho de corrente Wilson melhorado, que copia a corrente em M8 para M9. Logo, no nó de saída deverá existir um elemento capaz de fornecer a corrente solicitada pela variação de tensão da entrada.

* Novamente, como a saída costuma ser conectada a um nó de alta impedância, ao invés de variações de corrente, o circuito perceberá uma acomodação nas tensões para permitir a observância da lei de Kirchhoff das correntes. No exemplo dado, onde V_{i+} aumenta, a corrente em M7 tenderia a diminuir, mas como em M6 a corrente está aumentando e essa corrente será refletida para ficar em série com M7, a diminuição de corrente em M7 é impossível. A figura 3.12 mostra a dinâmica das variações de tensão para estabilização do circuito.

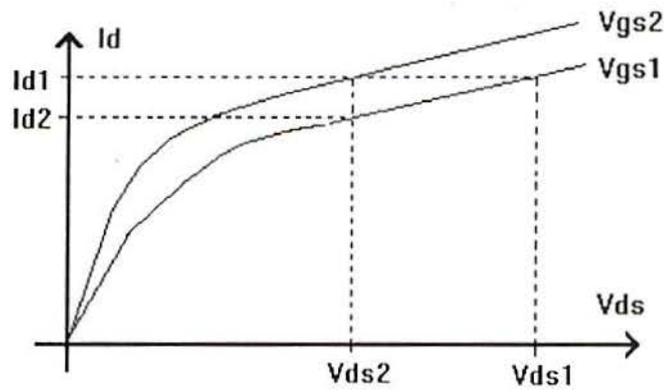


Figura 3.12

Na situação de repouso ($V_{in} = 0$), M7 apresenta V_{ds1} e V_{gs1} como suas tensões terminais. Quando V_{in} aumenta, a corrente sobre M7 tende a cair e, para evitar isso, (por efeito do espelho de corrente) a tensão no fonte de M7 aumenta, para garantir o GVO necessário para aquela corrente. Dessa forma, o transistor muda sua característica para a curva de V_{gs2} , onde, para que a corrente seja a mesma, V_{ds} deve ser V_{ds2} . A diminuição de V_{ds} é feita com o aumento de V_{out} , que produz o efeito contrário em M9. Outros arranjos podem ser feitos, combinando características bem determinadas de cada topologia, como *Cascode*. O circuito apresentado na figura 3.11, por exemplo, tem duas vantagens básicas que o tornam melhor que o Miller: ausência de capacitor de compensação e melhor resposta em frequência.

3.3 Compensação

Um problema inerente a amplificadores de dois pólos realimentados são as rotações de fase impostas por pólos e zeros [5,14,15,16,17]. A figura 3.13 mostra o comportamento do módulo e da fase do ganho de um amplificador genérico, dependendo da posição relativa entre os pólos e frequência de ganho unitário (F_u).

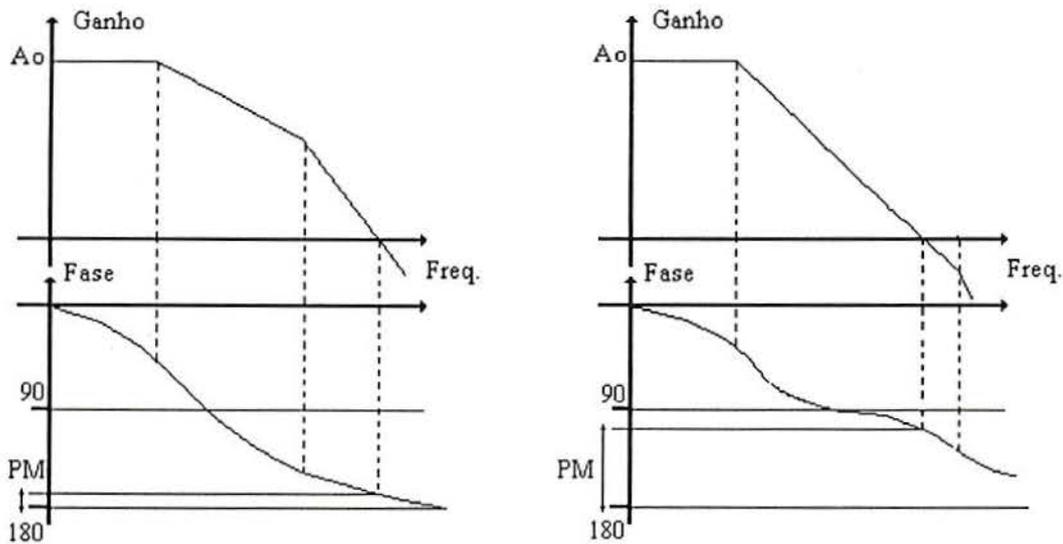


Figura 3.13

No primeiro gráfico pode-se ver que um pólo causa, além de uma diminuição na declividade de 20dB por década no gráfico de módulo de ganho, uma rotação de -90° no gráfico de fase. Na verdade, essa aproximação assintótica é o modelo de comportamento de sistemas, usando a técnica do gráfico de Bode. Na exata frequência do pólo, o sinal já teve aproximadamente -45° de rotação, sendo que os outros -45° serão obtidos em frequências superiores. Como o segundo pólo está relativamente próximo, existe a interação entre as duas frequências, de modo que a fase do sinal tende a se aproximar de -180° para frequências maiores que a do segundo pólo. Na frequência de ganho unitário, a diferença entre a fase do sinal e -180° é definida como sendo Margem de Fase (*Phase Margin* ou PM). No segundo gráfico, a colocação do segundo pólo em frequências acima da de ganho unitário retarda a segunda rotação de fase e permite que a margem de fase do sinal na frequência F_u seja maior. Esse valor deve ser mantido, por questões de estabilidade e segurança, num valor superior a 60° , pois a realimentação de sinal pode trazer instabilidade ao funcionamento do dispositivo.

Assim, garantir a adequada posição do segundo pólo é um quesito importante na manutenção da estabilidade do amplificador. Além do pólo, um outro fator que pode degradar a resposta em frequência do circuito é o zero. Devido ao fato do zero ser positivo, sua contribuição de fase não é positiva, como se espera comumente, trazendo consequências negativas ao funcionamento do circuito. Se o zero for colocado numa frequência próxima à F_u , a sua presença vai interagir com a fase do sinal antes da sua própria frequência, diminuindo assim a margem de fase. Logo, deve-se colocar o zero numa frequência suficientemente mais elevada, de modo que não atue sobre a fase do sinal.

Para projetar circuitos que obedeçam às premissas expostas, é necessária a introdução de um capacitor de compensação conectando primeiro e segundo estágio. Embora tendo algumas inconveniências, esse método, chamado de *Pole Splitting* (Separação de pólos), permite um aumento na frequência do pólo não-dominante, juntamente com uma redução da frequência do pólo dominante e com isso a garantia da estabilidade do circuito. Em frequências mais baixas, o papel desempenhado pelo capacitor é praticamente imperceptível, pois sua alta impedância praticamente não cria interação entre os estágios. Em frequências muito altas, por outro lado, a redução da impedância do capacitor faz com que o transistor de amplificação do segundo estágio fique em *diode-connection*, ou seja, com o dreno ligado à porta, prejudicando severamente o ganho do estágio. É claro que, em frequências tão altas, raramente se tem uma aplicação que ainda requeira ganhos significativos, mas a introdução do capacitor degrada de uma maneira suave e perceptível a performance do amplificador com o aumento da frequência.

Esse método de compensação é dito puramente capacitivo e atua somente nos pólos do amplificador. A introdução de uma resistência em série com o capacitor permite controlar de forma mais satisfatória a posição do zero, permitindo assim colocá-lo em frequências bem superiores e reduzindo sua influência no circuito. Na verdade, à medida que a resistência série aumenta, o zero se desloca em direção a frequências maiores, tendendo ao infinito quando a resistência é igual ao inverso do valor da transcondutância do transistor M6. Aumentando mais ainda a resistência, o zero aparece no semi-plano esquerdo. Uma outra maneira de interpretar a importância do posicionamento do zero na estabilidade do sistema é através da análise do diagrama de *root locus*, no qual se percebe que, para ganhos muito elevados e colocação do zero em frequências (positivas) reduzidas, os pólos (de parte real negativa) podem tender a encontrar o zero no semi-plano positivo e dar origem a raízes com parte real positiva (sistemas instáveis). Para evitar isso, a solução é a mesma, ou seja, assegurar que a frequência do zero seja suficientemente alta de modo que qualquer rota dos pólos necessite de um ganho maior que o máximo disponível no circuito para chegar ao semi-plano direito. Um inconveniente dessa técnica é a implementação do resistor, feita através de um transistor polarizado adequadamente na região linear. Entretanto, é importante garantir que a polarização mantenha o transistor na região linear independente das variações percebidas no nó de saída, pois se o transistor entrar na região de saturação, o valor dessa resistência sobe algumas ordens de grandeza.

3.4. Buffers

De forma geral, os amplificadores operacionais são classificados em dois tipos, quanto à sua aplicação: internos e externos. Amplificadores internos são aqueles cujo sinal de saída é empregado em outro bloco dentro do próprio integrado, enquanto que os externos possuem um bloco de potência que torna o sinal amplificado mais propício para o uso em circuitos exteriores ao chip. No primeiro caso, as cargas ligadas na saída destes amplificadores são puramente capacitivas e da ordem de 0.1 a 5 pF, dependendo do tamanho e quantidade de transistores de entrada dos estágios seguintes. Logo, na fase de projeto, essa carga pode ser levada em conta, de modo que o operacional poderá alimentá-la diretamente. No caso de amplificadores externos, é necessário inserir um estágio conhecido como *buffer*, cuja função é dar um alto ganho de corrente com ganho de tensão unitário. Essa necessidade advém do fato de que as cargas capacitivas representadas pelas entradas dos circuitos externos representam séria degradação de desempenho para o circuito, de modo que se for possível isolar os estágios de amplificação dos estágios de alimentação de carga, este procedimento trará benefícios à funcionalidade e à vida útil do dispositivo.

Um *buffer* pode ser implementado através de um amplificador operacional, configurado com realimentação negativa de ganho unitário, mas essa técnica peca por ter uma alta impedância de saída, representada pelos drenos dos transistores do estágio de saída. Para resolver este problema, topologias com saída em fonte são empregadas, mas estas, entretanto, apresentam pequena excursão de sinal. Na figura 3.14 três implementações distintas de *buffers* são propostas.

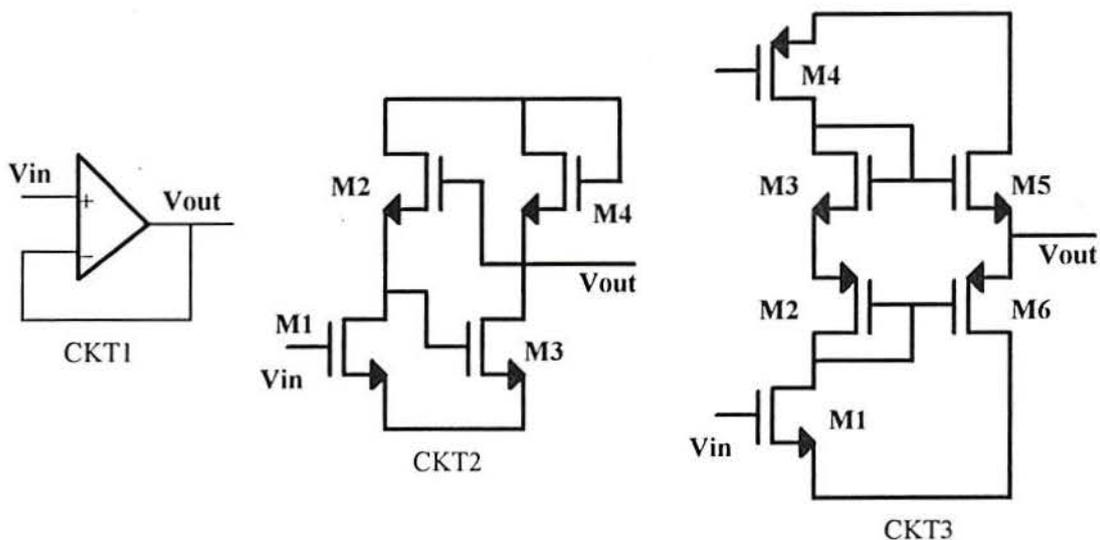


Figura 3.14

A amplificação de corrente pode ser feita através de um espelho de corrente, onde o fator de forma dos transistores que o compõe não são iguais ($W/L = 1$ e $W/L = 10$, por exemplo). CKT1 representa a implementação mais simples de um estágio de saída, pois proporciona um ganho de tensão unitário e significativa melhora na capacidade de fornecimento de corrente. Essa configuração permite vislumbrar mais facilmente um problema intrínseco ao próprio *buffer*: a rotação de fase do sinal de saída em relação à entrada. Teoricamente, seu funcionamento seria o de simplesmente copiar a tensão da entrada para saída, sem atrasos e sem distorção. Entretanto, pelo fato de ser composto de transistores reais, rotações de fase significativas podem ser observadas, especialmente quando se chega próxima à frequência de corte do circuito.

Também na figura 3.14, CKT2 [15] representa um *buffer* com saída em dreno-fonte, cujo funcionamento se baseia em princípios de realimentação. Se a tensão V_{out} começar a cair (por efeito de impedância de carga baixa), a corrente em M2 tenderá a diminuir, assim como a tensão na porta de M3. Com isso, M3 conduzirá menos e mais corrente estará disponível para as solicitações de carga. Uma análise simples permite verificar que a impedância de saída é o resultado da associação da condutância de M4 vista a partir do fonte e a de M3 visto pelo terminal de dreno. Geralmente, esse valor é suficientemente baixo para as especificações de carga, mas caso não o seja, uma outra alternativa é dada em CKT3 [14]. Aqui a condutância de saída é dada pela associação das condutâncias de dois transistores vistas a partir do fonte. A razão dos fatores de forma de M3 e M5 e M2/M6 dá o ganho de corrente, enquanto que o ganho de tensão pode ser unitário para uma larga faixa de frequências. O único inconveniente da saída totalmente em fonte é a consequente redução da excursão do sinal de saída, por causa da elevada sensibilidade do terminal de fonte em relação à corrente do transistor.

Naturalmente, a análise qualitativa dos circuitos que foi feita até agora é insuficiente para a execução da etapa de projeto. Para tanto, é necessária uma análise mais aprofundada das equações matemáticas que relacionam parâmetros tecnológicos com as especificações do circuito. Embora não seja a melhor topologia em todos os casos, o amplificador operacional Miller foi escolhido para ser analisado a fundo neste trabalho, para servir de base para as rotinas de síntese.

3.5 Miller: Análise matemática

O amplificador operacional Miller de dois estágios (figura 3.8) é um dos circuitos mais conhecidos entre as inúmeras possíveis combinações de amplificadores. Seu uso é bastante difundido e por isso seu equacionamento matemático, isto é, as equações que

relacionam parâmetros de performance com geometria e polarização, já foi analisado por vários autores [15,16,18]. Como todo operacional, o amplificador Miller preenche somente alguns dos muitos requisitos de desempenho solicitados a este circuito. Entre estas solicitações, destaca-se:

- * Alto ganho em malha aberta
- * Boa resposta em frequência
- * Equacionamento pouco complexo
- * Reduzido consumo de potência
- * Capacidade de corrente de saída
- * Alta rejeição a modo comum
- * Alta impedância de entrada
- * Baixa impedância de saída e
- * Reduzido off-set.

Através do uso do equivalente de pequenos sinais do transistor é possível extrair uma quantidade razoável de informações sobre este circuito, conforme a figura 3.15:

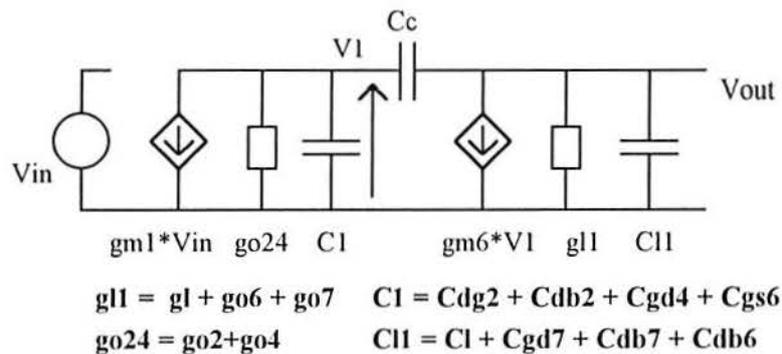


Figura 3.15

A resolução das duas equações de malha do circuito conduzem a uma equação que relaciona diretamente a entrada à saída, onde $I1 = gm1 \cdot Vin$, $I2 = gm6 \cdot V1$ e $I3$ é corrente que passa por $go24$ e $C1$.

$$\frac{I3}{g_{o24} + s \cdot C1} - \frac{I1 - I3}{s \cdot Cc} - \frac{I1 - I3 + I2}{g_{l1} + s \cdot C11} = 0$$

$$\frac{I3}{g_{o24} + s \cdot C1} - \frac{gm1 \cdot Vin - I3}{s \cdot Cc} - \frac{gm1 \cdot Vin - I3 + I2}{g_{l1} + s \cdot C11} = 0$$

Equação 2

A análise DC (desprezando os capacitores), permite uma forma compacta para o ganho de tensão V_{out}/V_{in} , uma vez que g_l é geralmente negligenciada, sem excessiva perda de precisão:

$$A_{v0} = \frac{g_{m1}}{g_{o2} + g_{o4}} \cdot \frac{g_{m6}}{g_{o6} + g_{o7}}$$

Equação 3

Essa equação indica que quanto menores os níveis de condutância (maiores as impedâncias), maior será o ganho A_{v0} e esse é um dos motivos pelos quais os circuitos analógicos não têm os comprimentos mínimos permitidos pela tecnologia, pois a relação de dimensões L/W é aproximadamente proporcional à impedância. Outro importante motivo para isso é que as pequenas variações no tamanho efetivo dos transistores podem representar percentuais expressivos se as dimensões forem reduzidas, degradando sensivelmente a performance do circuito. Igualmente importante é o papel desempenhado pela transcondutância g_m , que relaciona a variação de corrente de dreno devido à variação na tensão porta-fonte. Também este valor pode ser aumentado através do aumento da relação do fator de forma do transistor, originando assim dispositivos de tamanho superiores aos comumente vistos em aplicações digitais. Tamanhos de $500\mu\text{m}$ por $10\mu\text{m}$ e ainda maiores são comuns e de maneira alguma devem ser encarados como absurdos em aplicações como amplificadores operacionais.

A presença de dois pontos de alta impedância (saída do primeiro e segundo estágio) leva o sistema a ter um comportamento de segunda ordem, pois dois pólos são gerados nesses pontos. Levando em conta todos os efeitos parasitas mais importantes em uma análise AC, a equação resultante é bem mais complexa, onde Z_1 e Z_2 são impedâncias relacionadas aos elementos parasitas:

$$A_v(s) = \frac{-Z_2 \cdot g_{m1} \cdot Z_1 \cdot (g_{m6} + s \cdot C_c)}{-Z_1 \cdot s \cdot C_c - 1 - Z_2 \cdot s \cdot C_c + Z_2 \cdot g_{m6} \cdot Z_1 \cdot s \cdot C_c}$$

$$Z_1 = \frac{1}{g_{o2} + g_{o4}} + \frac{1}{s \cdot C_{ds1}}$$

$$Z_2 = \frac{1}{g_{o6} + g_{o7}} + \frac{1}{s \cdot (C_{ds2} + C_l)}$$

Equação 4

$A_v(s)$ representa a função ganho de tensão em relação à frequência e é, na verdade, de segundo grau, pois Z_1 e Z_2 são fatores que contém o termo s . A substituição das impedâncias Z_1 e Z_2 na equação principal produz:

$$A_v = \frac{(s \cdot C_{ds2} + s \cdot C_l + g_{o67})}{g_{o67} [s^2 \cdot (C_{ds2} + C_l)]} \cdot g_{m1} \cdot \frac{(s \cdot C_{ds1} + g_{o24})}{(g_{o24} \cdot C_{ds1})} \cdot \frac{(g_{m6} + s \cdot C_c)}{\left[\frac{(s \cdot C_{ds1} - g_{o24})}{(g_{o24} \cdot C_{ds1})} \cdot C_c - 1 - \frac{(s \cdot C_{ds2} + s \cdot C_l + g_{o67})}{(g_{o67} \cdot (C_{ds2} + C_l))} \cdot C_c + \frac{(s \cdot C_{ds2} + s \cdot C_l + g_{o67})}{(g_{o67} \cdot (s \cdot (C_{ds2} + C_l)))} \cdot g_{m6} \cdot \frac{(s \cdot C_{ds1} + g_{o24})}{(g_{o24} \cdot C_{ds1})} \cdot C_c \right]}$$

Equação 5

A complexidade dessa equação impede a extração de alguns parâmetros importantes, como a frequência dos pólos, do zero e a frequência de ganho unitário. Com o objetivo de obter uma equação mais compacta e mais manuseável algebricamente, embora ainda de segundo grau, para descrever corretamente o comportamento do circuito, mantém-se o equacionamento e simplifica-se algumas capacitâncias parasitas, de modo a obter a seguinte função de transferência:

$$A_v(s) = g_{m1} \cdot \frac{(g_{m6} + s \cdot C_c)}{s^2 \cdot C_c \cdot C_l + (C_c \cdot g_l - C_c \cdot g_{m6} + g_{o24} C_l + C_c \cdot g_{o24}) \cdot s + g_{o24} g_l}$$

Equação 6

A partir dessa equação simplificada, pode-se relacionar algumas especificações do operacional com parâmetros de pequenos sinais e capacitâncias. Pelo fato de alguns termos terem sido desprezados, a análise em frequência não é perfeita e com isso os resultados obtidos não devem ser encarados como absolutamente precisos. Pelo numerador da equação, é possível determinar facilmente a frequência do zero do sistema. Esse valor, juntamente com o valor dos dois pólos do sistema, é dado abaixo:

$$Z = \frac{g_{m6}}{C_c} \quad P_d = 0$$

$$P_{nd} = \frac{(g_{o24} C_l + C_c \cdot g_{o24} - C_c \cdot g_{m6})}{C_c \cdot C_l} = \frac{g_m}{C_l}$$

Equação 7

Estas simplificações forma obtidas através da supressão de termos por comparação. Ou seja: quando $C_c \cdot g_l$ é subtraído de $C_c \cdot g_{m6}$, o primeiro termo é desprezado, por contribuir muito pouco significativamente para o resultado final. Na

verdade, o pólo dominante de um amplificador operacional não é zero, mas pode ser colocado em frequências suficientemente baixas para os quesitos de estabilidade sejam satisfeitos. Já o pólo não-dominante, que nesse caso está pelo menos quatro décadas acima, é relativamente bem descrito pela equação anterior e é extremamente útil na síntese do próprio operacional.

A banda passante do operacional é definida como sendo a frequência na qual a curva do ganho cai aproximadamente 3dB. A determinação da banda passante também pode ser feita de forma analítica e, valendo-se do fato de que o produto ganho-banda passante é aproximadamente constante, a F_u também pode ser derivada [16].

$$BW = \frac{gm1}{Av2 \cdot Cc} \quad Fu = GBW \cdot \sin(PM) = \frac{gm1}{Cc} \cdot \sin(PM)$$

Equação 8

De posse de todos estes parâmetros, é possível projetar dispositivos cujas características de pequenos sinais resultem em um desempenho igual ou superior a certas especificações de projeto. Embora ainda havendo pólos em frequências ainda maiores, estes não desempenham papel importante no intervalo de frequências geralmente usado (entre DC e F_u).

A restrição mais importante na etapa de síntese de operacionais diz respeito à sua estabilidade. Para tanto, as condutâncias e transcondutâncias devem ser tais que o amplificador é dito intrinsecamente estável, ou seja, sob qualquer condição de polarização ou tensão de entrada, o circuito é sempre estável. A resposta em frequência esquemática na figura 3.13 mostra a necessidade de respeitar certa ordem na colocação dos pólos e dos zeros. Para que o sistema se assemelhe ao máximo com um sistema de primeira ordem (um pólo em F_d), a curva de resposta deve ser constante para frequências baixas, iniciando uma queda com 20dB por década, cruzando o eixo de ganho unitário, encontrando mais um pólo (40dB/década) e finalmente um zero (Total: 20dB/dec).

A inserção do capacitor de compensação (conforme 3.4) possibilita reduzir a frequência do pólo dominante ao mesmo tempo que aumenta a frequência do pólo não-dominante, impedindo assim que rotações de fase de 180° aconteçam dentro do intervalo útil de frequências. Mesmo assim, apenas reposicionar os pólos não é suficiente para garantir a estabilidade. A análise das equações extraídas a partir do modelo de pequenos sinais relaciona grandezas como transcondutâncias e condutâncias com especificações

como ganho e produto ganho-banda. Com isso, para assegurar que o sistema será estável, essas equações devem ser respeitadas. Na literatura sobre este assunto, duas heurísticas são comuns:

1) Projetar o transistor de amplificação do segundo estágio com uma transcondutância 10 vezes maior que o equivalente do primeiro estágio, ou seja: $g_{m6} = 10 \cdot g_{m1}$. Analisando as equações da frequência de ganho unitário e do zero, conclui-se que, se essa relação for satisfeita, o zero estará uma década acima dessa frequência e, com isso, sua influência na degradação da margem de fase será pequena.

2) Relacionar o capacitor de compensação com o capacitor de carga, de modo a posicionar o pólo não-dominante numa distância segura da F_u . Na verdade, essa premissa está baseada na observação do comportamento da resposta em fase de um sistema de um pólo fixo e outro variável. Supondo um pólo em baixas frequências (f_1), uma frequência de ganho unitário (F_u) e um segundo pólo em uma frequência f_2 não fixa ($f_2 \geq F_u$), pode-se verificar a importância da relação f_2/F_u sobre a margem de fase, de acordo com a tabela 1:

f_2/F_u	PM [°]
1	45
1.5	52
2	65
3	72
4	76

Tabela 1

Para sistemas reais, diferentes deste exemplo simples e idealizado, a relação entre a frequência do pólo não-dominante e a de ganho unitário é fixa em 2.2 vezes, para gerar uma margem de fase de 60°. Usando as equações que definem a F_u e P_{nd} , em conjunto com a definição anterior, pode-se extrair a seguinte equação, simplificando a equação que descreve a frequência do pólo não-dominante:

$$P_{ND} = 2.2 \cdot F_u \quad \rightarrow \quad \frac{g_{m6}}{C_l} = 2.2 \frac{g_{m1}}{C_c}$$

$$\frac{C_c}{C_l} = 0.22$$

Equação 9

Uma vez que se conheça a carga nominal que deve ser ligada ao amplificador operacional (que geralmente já inclui uma margem de segurança), esse cálculo permite que se tenha o valor do capacitor de compensação. A simplicidade desse cálculo compensa sua intrínseca falta de precisão, pois a negligência de algumas capacitâncias importantes torna-se cada vez mais perceptível à medida que a frequência e a geometria dos dispositivos aumentam. De maneira geral, quanto maior o capacitor de compensação, melhor a resposta em frequência do operacional, mas seu valor não pode ser arbitrariamente aumentado por razões construtivas, uma vez que as atuais técnicas de implementação oferecem uma eficiência pequena, exigindo assim grandes áreas. Uma das características que fazem o projeto e o funcionamento do Miller um pouco mais complexo é que a posição exata do segundo pólo é função, entre outras coisas, da carga a qual o amplificador é conectado. Com isso, um circuito calculado e projetado para funcionar com uma margem de fase de 70° para 5pF pode se tornar instável simplesmente pelo fato de aumentar a carga para 20pF, por exemplo.

Outro parâmetro importante do desempenho do Miller é o *Slew Rate*, ou seja, a máxima taxa de variação temporal de saída em função de um estímulo de entrada. Essa propriedade está diretamente ligada à capacidade do circuito de fornecer corrente e de não distorcer as características do sinal de entrada quando este é transportado para a saída. O seu valor, geralmente dado em V/ μ s, é calculado da seguinte forma:

$$SR = \frac{I_b}{C_c}$$

Equação 10

O número de parâmetros de especificação de um amplificador operacional é muito grande, sendo portanto difícil concatenar todas estas solicitações em uma metodologia de síntese única. Logo, alguns parâmetros são escolhidos como críticos e uma técnica específica de projeto deve ser desenvolvida. Neste trabalho, as especificações usadas até agora serão empregadas no sentido de produzir um operacional que atenda a essas necessidades, sendo que os outros parâmetros, supostamente de menor importância, serão consequência das escolhas feitas.

3.6 Implementação alternativa do capacitor de compensação

A utilização de amplificadores operacionais que necessitam de elementos compensadores internos é dificultada por dois motivos básicos:

1) Segundo a literatura convencional, num circuito amplificador genérico, estatisticamente entre 40 e 60% da área é consumida para construção do capacitor de compensação.

2) A construção desse capacitor em tecnologia digital é sempre através da sobreposição de duas camadas: poli-silício e metal. Infelizmente, a tecnologia não assegura com precisão a distância entre estas duas camadas. O valor do capacitor é uma função quadrática inversa dessa distância, de modo que o valor nominal deste capacitor tem uma incerteza muito grande. Logo, a falta de controlabilidade da espessura do óxido entre as camadas (que funciona como dielétrico) impede que os valores de capacitância efetivamente observados coincidam com os originalmente projetados.

Esses dois fatos tornam a busca de novas formas de implementação uma técnica muito interessante que objetiva não só reduzir a área dos circuitos mas também construir capacitores com valores mais precisos [15,19,20]. Tecnologias analógicas permitem a construção de capacitores com duas camadas de poli-silício e com isso aumenta-se a eficiência, ou seja o valor de capacitância por unidade de área, entre 7 e 15 vezes. Naturalmente, devido ao maior número de etapas no processo de fabricação, essas tecnologias têm um custo mais elevado.

A idéia proposta aqui é utilizar um transistor MOS numa configuração específica, de modo a explorar a capacitância porta-fonte, que conta com a possibilidade de redução de área. Dessa forma, o valor total desse "capacitor" é função das dimensões W e L do transistor. A figura 3.16 mostra o tipo de ligação de um transistor NMOS numa configuração de capacitor e sua estrutura típica.

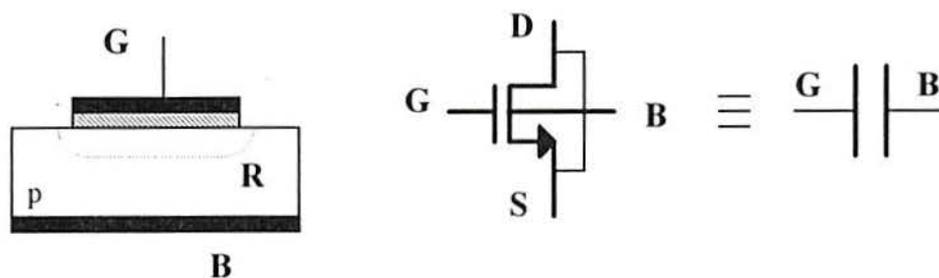


Figura 3.16

A teoria [15] prevê que a capacitância de uma estrutura como a da figura 3.16 pode ser expressa da seguinte forma:

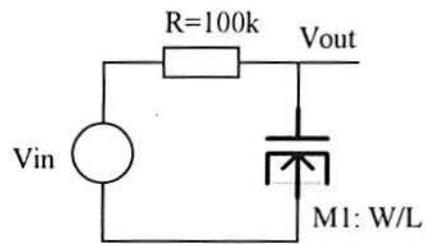
$$C = \frac{\epsilon_{OX}}{L}$$

Equação 11

onde ϵ_{ox} é a permissividade do óxido, a Área é dada pelas dimensões $W \cdot L$ do transistor e L é a espessura do óxido que separa o terminal de porta do corpo. Supondo uma pequena tensão de porta V_g positiva, as lacunas do substrato tipo P serão repelidas eletrostaticamente da região da porta, deixando a região de depleção com carga negativa. Dessa maneira, as cargas estão separadas por um dielétrico e um capacitor é formado. Para pequenos valores de V_g ($V_g \ll 1$), o valor dessa capacitância permanece constante (acumulação). Porém, à medida que V_g aumenta, a quantidade de cargas induzidas na região de depleção aumenta e com isso o valor efetivo de L aumenta, diminuindo a capacitância (depleção). Aumentando mais ainda V_g , o campo elétrico criado no interior do dispositivo tem magnitude suficiente para atrair elétrons termicamente gerados para a região de depleção (inversão). Esses elétrons, que se movem em direção à superfície, promovem uma carga exatamente sobre o canal, de maneira que o valor efetivo de L passa novamente a ser a espessura do óxido e com isso a capacitância volta a ter seu valor nominal. O análogo ocorre se V_g tem valores negativos: a porta forma uma placa com carga negativa e a superfície do semiconductor apresenta carga positiva, separados pelo óxido de porta. O terminais de dreno, fonte e substrato estão eletricamente conectados, de modo que o transistor é empregado como sendo um dispositivo de dois terminais [15].

Para realmente poder usar o transistor como capacitor deve-se conhecer seu comportamento de uma maneira razoavelmente precisa. As duas mais importantes características a serem extraídas são: relação área - capacitância e perfil de capacitância em relação a tensão. Estabelecer uma relação entre a área do transistor e a capacitância por ele gerada permite construir capacitores dos mais variados valores, sem, para isso, repetir uma bateria de teste para determinar seu valor. Por outro lado, como se sabe que o comportamento dessa capacitância é variável com a tensão aplicada, essa dependência deve ser analisada, para determinar se sua presença interfere negativamente de alguma maneira no funcionamento do circuito.

Para poder obter dados numéricos dessas características, a simulação (SPICE) foi o recurso empregado, em conjunto com a criação de circuitos de teste, que permitiram que se extraísse informações muito úteis, conforme a figura 3.17.



$$C = \frac{1}{2 \cdot \pi \cdot R \cdot f}$$

Figura 3.17

O funcionamento do circuito é o seguinte:

- 1) Para baixas frequências, a impedância do transistor usado como capacitor é muito maior que a do resistor, de modo que a tensão da fonte tende a ficar completamente no transistor.
- 2) Em frequências mais elevadas, a impedância do "capacitor" tende a cair de modo que o divisor resistivo aumenta a tensão sobre o resistor.
- 3) Em um diagrama de Bode, a tensão da saída pode ser identificada por um pólo em $1/(R \cdot C)$. Como o valor de R é conhecido, o valor da capacitância pode ser extraído pela equação da figura 3.17.

A exata determinação da frequência na qual ocorre o pólo não é de todo precisa, especialmente se se utiliza o gráfico de módulo, onde o pólo é identificado por uma queda de 3dB. Por outro lado, se o gráfico de fase for usado, o pólo pode ser determinado pela frequência na qual existe uma rotação de fase de 45°. Para uma polarização fixa, vários tamanhos W/L foram testados e os valores de capacitâncias foram tabulados em curvas no Gráfico 2. O Gráfico 2.1 mostra a relação entre a área (W*L) do transistor e o valor de capacitância entre seus terminais. Convém observar que foi feito também um teste de variação de geometria do transistor, isto é, um dispositivo com fator de forma de 20µm x 20µm mostra a mesma capacitância que outro de 40µm x 10µm, pois têm a mesma área de porta. Na verdade, poderia se esperar alguma diferença, devido ao efeito de borda presente em ambos dispositivos, mas atuando de maneira diferente e mais acentuada em função do aumento do perímetro.

A análise da curva do Gráfico 2.1 mostra uma relação linear entre área capacitância, como seria de se esperar. Usando uma regressão linear, os coeficientes

dessa reta foram extraídos, permitindo que se escreva a seguinte equação, (desprezando o termo independente):

$$\text{Área} = \frac{C}{0.0013884}$$

Equação 12

Embora esse mesmo circuito pudesse ser usado na extração do perfil de capacitância em relação à tensão, um outro foi escolhido, de forma que o processo não fosse manual. O circuito usado é simplesmente uma fonte de corrente em paralelo com um transistor desempenhando o papel de capacitor. Conforme a teoria, a tensão sobre este capacitor deve subir linearmente e a corrente deve permanecer fixa. O esquemático do circuito, assim como a relação entre tensão e corrente de um capacitor ideal, é mostrado na figura 3.18.

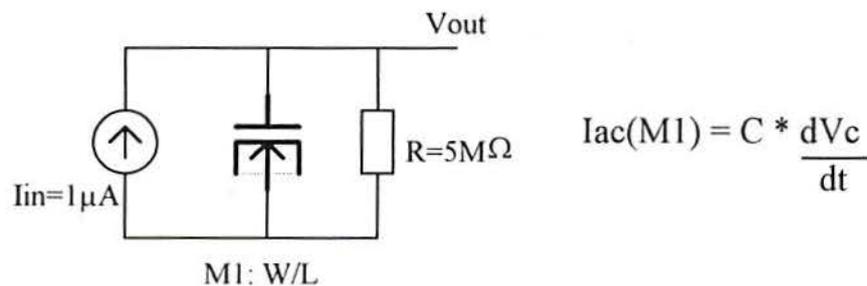


Figura 3.18

O resistor foi colocado por razões de simulação, isto é, o simulador não é capaz de avaliar este circuito sem fixar um valor de tensão em Vout. Naturalmente, a inclusão do resistor modifica o funcionamento do circuito, cuja tensão não subirá mais linearmente e sim através de uma exponencial. Entretanto, a equação que relaciona tensão e corrente em um capacitor pode ser usada, independente da forma de onda desses parâmetros. O funcionamento do circuito é o seguinte:

- 1) A fonte de corrente inicialmente tem um valor negativo ($-1\mu\text{A}$), de forma que toda essa corrente flui pelo resistor, logo este tem uma queda de tensão de -5V .
- 2) Num instante determinado, acontece um salto e a corrente vai de $-1\mu\text{A}$ para $+1\mu\text{A}$. Essa descontinuidade de corrente fará com que o capacitor tenda a manter a tensão de seus terminais (ou seja, manter a corrente em R), logo, por breves instantes, a corrente que circula pelo capacitor é $2\mu\text{A}$.

3) À medida que o tempo passa, o capacitor vai se carregando exponencialmente, até que a tensão na carga chegue a 5V novamente, momento no qual, por não existir variação de tensão, a corrente do capacitor cai a zero.

Os valores de resistência e corrente usados foram tais que a tensão sobre o capacitor excursionasse dentro do limite de interesse para este teste: $\pm 2.5V$. A grande virtude deste método é que em um só teste pode-se levantar a característica de capacitância do transistor para várias tensões, simplesmente invertendo a equação da figura 3.18 ($C = I_d(M1) / d(V_c)/dt$). Naturalmente, o gráfico resultante desse procedimento não é $C \times V$ e sim $C \times t$ (tempo), pois a análise usada, segundo o Gráfico 2.2, é temporal. Embora o SPICE permita que se substitua o eixo horizontal por qualquer outra forma de onda do circuito, devido a problemas operacionais com este simulador, esses parâmetros foram extraídos para o MATLAB, através de um arquivo ASCII e da opção *PRINT*.

As curvas contidas no gráfico 2.3 permitem visualizar o comportamento da capacitância em relação a variações de tensão aplicada para transistores NMOS e PMOS. A primeira característica a ser mencionada é o fato de que a região de não-idealidade compreende aproximadamente o intervalo $[0 ; V_t]$, sendo negativo para o PMOS e positivo para o NMOS. Percentualmente, a variação máxima no valor da capacitância para um transistor NMOS simples é 70.9%, tornando seu uso restrito a algumas aplicações que não exijam valores precisos. Para o PMOS, essa variação é menor: 65.5%, mas ainda é muito pronunciada para uso indiscriminado. A melhor forma de reduzir essa variação percentual entre valor máximo e mínimo é através da associação paralela de dois transistores. Para tanto, vários testes foram feitos e a curva de capacitância extraída está mostrada para várias ligações:

- | | |
|--|---------------------|
| 1) PMOS em paralelo com NMOS - | 53.1% (Gráfico 2.4) |
| 2) PMOS em paralelo com NMOS invertido*- | 32.8% (Gráfico 2.5) |
| 3) NMOS em paralelo com NMOS invertido*- | 35.5% (Gráfico 2.6) |
| 4) PMOS em paralelo com PMOS invertido*- | 32.7% (Gráfico 2.7) |

* Invertido significa rotar o transistor 180°, intercambiando os terminais de porta e substrato.

Estas simulações foram obtidas através de um circuito simples, onde os terminais de corpo e porta forma ligados às tensões mais convenientes. Num circuito real, construído com substrato P, por exemplo, os transistores NMOS não podem ter sua

tensão de porta ligado a um terminal flutuante, pois essa tensão é comum a vários outros dispositivos. Ao contrário, o PMOS será construído sob um poço N e a tensão de corpo poderá ser mantida em qualquer potencial que se desejar, fixo ou não. Isto acarreta um problema de ordem prática para construir capacitores mais lineares em circuitos reais, pois o corpo não pode ser ligado à qualquer tensão.

A colocação de uma fonte de tensão em série com uma das portas permite deslocar horizontalmente a característica de capacitância de um transistor, simulando o efeito de alteração da tensão de limiar. Dessa forma, seria possível combinar as características de um NMOS e um PMOS, de modo a mostrarem perfis semelhantes no mesmo intervalo de tensão. Infelizmente, esse recurso não serve para reduzir a variação de capacitância, uma vez que a sobreposição destas características através da associação paralelo produz um efeito aditivo e não subtrativo, tornando ainda pior o comportamento de capacitância.

Para aplicações como integradores, um capacitor cujo valor nominal varia 32% gera uma distorção proporcional no sinal integrado, de maneira que seu uso é totalmente desaconselhável. Entretanto, em aplicações como compensadores (amplificador Miller), onde a função do capacitor é simplesmente ser maior que determinado valor, esse tipo de configuração permite uma redução considerável de área. Naturalmente, como o capacitor C_c influencia a resposta em frequência do operacional, a resposta do circuito será variável e testes devem ser feitos para demonstrar que o desempenho do circuito não será degradada pela presença deste elemento não-linear.

Os expressivos benefícios na redução de área causados pela substituição do capacitor por um transistor causariam uma sensível economia em blocos que utilizem extensivamente estes capacitores. Mas como já foi dito, sua característica não linear dificulta o seu uso direto em aplicações analógicas. Entretanto usando o esquema proposto na figura 3.19, toma-se partido do fato de que a característica de capacitância do transistor PMOS é absolutamente constante para polarizações reversas, ou seja, de 0V até uma tensão máxima, a capacitância é sempre a mesma. O mesmo não acontece com o NMOS, cuja não-idealidade localiza-se exatamente no cruzamento por zero.

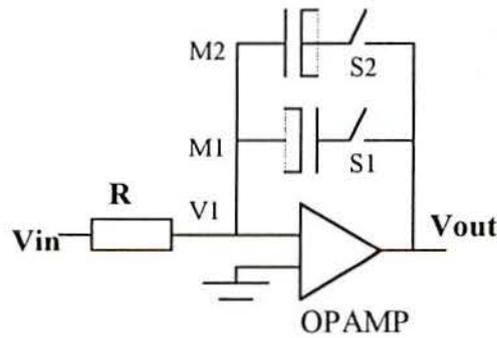


Figura 3.19

O integrador da figura 3.19 funciona da seguinte maneira: quando a tensão V_{out} é maior que V_1 , S2 fecha e S1 abre, de maneira que o integrador seja composto de R e M2. Se, ao contrário, $V_{out} < V_1$, S1 fecha e S2 abre, para conectar M1. Dessa forma, em qualquer instante de tempo, o capacitor conectado sempre está polarizado na região de capacitância constante. As chaves S1 e S2 são comandadas por um comparador, cujas entradas estão ligadas a V_{out} e V_1 . Uma vez que os capacitores serão comutados no cruzamento por zero, a integração não sofrerá com isso, uma vez que neste momento a carga interna do capacitor é zero ($Q=C \cdot V$). Teoricamente o circuito funciona como um integrador comum com capacitor linear, uma vez que as chaves sejam ideais e o intervalo de comutação do comparador muito estreito.

Essa implementação só teria razão de ser se a utilização de capacitores MOS permitisse uma economia de área que compensasse o custo extra de engenharia para manipulá-los. Entretanto, quando se coloca chaves reais (transistores), a resistência série torna-se significativa e atua no funcionamento do integrador, prejudicando seu desempenho. De mesma maneira, quando o intervalo de comutação do comparador é aproximado de um valor real, os cruzamentos por zero do sinal integrado apresentam visíveis descontinuidades. Além disso, a implementação de um comparador eficiente, por mais simples que seja, em conjunto com os capacitores e as chaves, dificilmente ocuparia menos área que um capacitor linear poli-metal, para aplicações comuns.

Em vista destes problemas, a utilização de capacitores MOS em circuitos como o integrador mostrou-se pouco eficiente. Entretanto, para aplicação no Amplificador Miller, essa técnica é bastante promissora. Para evidenciar a sua eficiência, o Gráfico 2.8 traz um curva comparativo entre a resposta AC de um amplificador operacional usando um capacitor simples e outro usando um capacitor MOS, com mesmo valor nominal de capacitância. Similarmente, a curva do Gráfico 2.9 mostra a resposta temporal de dois amplificadores, ambos com realimentação para ter ganho unitário e excitação senoidal na

entrada, mas um com capacitor linear e outro com capacitor MOS. Duas características do amplificador Miller fazem seu uso propício para a aplicação de capacitores MOS:

1) Em frequências relativamente baixas, onde geralmente ganhos elevados são necessários, a importância do capacitor não é grande, de modo que variações mesmo que significativas do capacitor de compensação não serão percebidas.

2) Em frequências maiores, próximas a frequência de ganho unitário, a tensão no capacitor tem excursão reduzida, uma vez que o ganho é baixo. A característica inversora do segundo estágio encarrega-se de polarizar adequadamente o capacitor, juntamente com o fato de que nestas frequências, como a impedância deste é baixa, o transistor de amplificação está praticamente em *diode connection*. Nessas condições, a queda de tensão entre porta e fonte é necessariamente maior que V_t , de modo que o transistor estará com polarização tal que sua capacitância será constante.

Por estes motivos, teoricamente o funcionamento do Miller não deve ser degradado, em nenhum intervalo de frequência, pela substituição do capacitor pelo transistor. Entretanto, simulações foram feitas, para que fosse possível certificar-se de que estas premissas são verdadeiras. Para tanto, foi utilizado um projeto de operacional comum e sua resposta no tempo foi comparada à outro circuito igual, com a substituição do capacitor por um transistor PMOS. A resposta AC de ambos circuitos, assim como a DC é exatamente igual, sendo que a transiente tem ligeiras diferenças.

A análise da resposta do amplificador com capacitor e MOS mostra que seu emprego cumpre as exigências de não mudar a resposta do circuito, ao mesmo tempo que permite que a implementação do capacitor seja feita de maneira mais precisa. Em termos numéricos, um capacitor de 2.2pF implementado na tecnologia ES2- 1.2 μm ocupa uma área de 44.000 μm^2 quando implementado com as camadas polisilício e metal, enquanto que se for utilizado um transistor MOS, 1600 μm^2 serão necessários, ou seja, somente 3.6 % da área.

4 Modelamento

Projetar circuitos analógicos exige um conhecimento bem mais profundo do comportamento do transistor do que o necessário em circuitos digitais, uma vez que nestes, o reduzido número de estados possíveis facilita o bom funcionamento do circuito e a obtenção de determinado desempenho. Por isso, antes de efetivamente fazer um esforço em direção ao projeto de blocos analógicos é necessário conhecer o perfil tensão-corrente do transistor MOS a partir de seus terminais, bem como seu comportamento de pequenos sinais.

Por "modelamento" entenda-se a capacidade de prever o comportamento de um dispositivo qualquer frente a determinadas condições de operação. Assim, modelar precisamente um dispositivo (como o transistor) é etapa imprescindível em qualquer processo de síntese de circuitos com base neste elemento. Em circuitos integrados CMOS, a complementaridade dos dispositivos empregados permite que se use somente transistores nos circuitos, de maneira que um modelo preciso de transistor é ferramenta suficiente para projetos das mais variadas naturezas.

Uma das tendências mais difundidas nos dias de hoje sugere uma redução severa nos níveis de corrente e potência empregada nos circuitos integrados, levando os transistores a operarem em regiões de alto ganho, chamada de inversão fraca. Aqui a corrente de dreno de um transistor é composta quase que exclusivamente de corrente de difusão. Por outro lado, para valores mais elevados de corrente e potência, a região de operação é dita inversão forte, o valor nominal do ganho intrínseco de um estágio inversor é consideravelmente menor e a corrente é essencialmente devida à deriva. No meio desse intervalo existe uma região chamada inversão moderada, na qual a corrente de dreno é uma composição dos dois tipos acima referidos. Infelizmente, ainda não existe um modelo consagrado pelo uso que modele com precisão o comportamento de corrente dos transistores independentemente da região de operação. De maneira geral, a corrente de dreno depende das tensões terminais do transistor, através de relações complexas, que podem ser simplificadas nos casos limites.

Como regra geral, pode-se dizer que a precisão de um modelo é proporcional à sua complexidade, de maneira que modelos mais simplificados e computacionalmente pouco intensivos fornecem resultados pouco confiáveis e com tolerâncias muito altas.

Por outro lado, modelos complexos, com embasamento físico ou com formulação empírica, pelo fato de levarem em conta efeitos de segunda ordem, retratam de maneira mais fidedigna o comportamento dos dispositivos. Neste trabalho, as equações que relacionam tensão e corrente em um transistor não foram derivadas, por motivo de simplicidade, mas a sua forma final é estudada de modo a permitir a compreensão dos fenômenos internos ao transistor. Além disso, somente os modelos relacionados à inversão forte serão vistos, posto que o modelamento e projeto de blocos analógicos em inversão fraca e moderada é uma tarefa bastante mais complexa e portanto deixada para o futuro. A decisão de fazer uma abordagem enfatizando somente a inversão forte deve-se ao fato de que o recurso disponível na época de execução deste trabalho, a nível de simulação, era única e exclusivamente o simulador SPICE 2.G, que sabidamente não modela bem as regiões de inversão fraca e moderada. Para que se tivesse um parâmetro de comparação capaz de auxiliar o desenvolvimento da implementação do modelo, este simulador foi usado como sendo representativo da realidade e, uma vez que o seu modelamento é limitado quanto às regiões de operação, essa característica foi mantida no modelo usado neste trabalho. Embora sabendo que o projeto de amplificadores em inversão fraca e moderada provou-se bem mais eficiente, tanto em ganho quanto em potência, essa possibilidade foi deixada para um trabalho futuro.

4.1 Modelo quadrático

A formulação mais simples de transistor em inversão forte é o Modelo Quadrático, cuja equação fundamental é dada na Equação 1 [14,16].

$$I_d = K_p \cdot \frac{W}{L - 2 \cdot L_d} \cdot \left[(V_{gs} - V_t) - \frac{V_{ds}}{2} \right] \cdot V_{ds} \cdot (1 + \lambda \cdot V_{ds}) \quad V_{ds} < V_{gs} - V_t$$

$$I_d = K_p \cdot \frac{W}{L - 2 \cdot L_d} \cdot (V_{gs} - V_t)^2 \cdot (1 + \lambda \cdot V_{ds}) \quad V_{ds} \geq V_{gs} - V_t$$

Equação 13.1 e 13.2

Este equacionamento propõe equações distintas para modelar as regiões linear e de saturação, da seguinte forma:

* Se a tensão dreno-fonte é inferior à tensão de saturação (dada por $V_{gs} - V_t$), a equação 13.1 é usada.

* Se, ao contrário, $V_{ds} > V_{gs} - V_t$, a segunda equação é usada e posteriores incrementos de V_{ds} produzirão pequenos aumentos na corrente de dreno, ponderados pelo parâmetro λ . Este fator proporciona uma impedância de saída não infinita ao transistor na região de saturação.

K_p representa a constante de transcondutância ($K_p = \mu \cdot C_{ox}$), e L_d representa a difusão lateral sob a porta. Uma característica a ser observada é que o modelo proposto é contínuo, isto é, as equações que o compõe são individualmente contínuas e o ponto de transição entre a primeira e a segunda gera valores iguais nas duas equações. Entretanto, existe uma descontinuidade na primeira derivada da corrente em relação a tensão V_{gs} , que pode ser comprovada traçando a curva de transcondutância em função de V_{gs} . Este problema causa repercussões importantes pois o ganho dos circuitos é diretamente proporcional à transcondutância, logo seu valor absoluto deve ser dado com o mínimo de erro. O parâmetro V_t representa a tensão limiar, cujo valor é expresso pela Equação 14.

$$V_t = V_{fb} + \phi_b + \gamma \cdot \sqrt{\phi_b + V_{sb}}$$

$$\phi_b = U_t \cdot \ln\left(\frac{N_{sub}}{n_i}\right)$$

Equação 14

Conforme mostra a equação, V_t não é constante, mas mostra uma dependência com a tensão de corpo, através do parâmetro γ (Coeficiente de efeito de Corpo). O modelo apresentado prima pela simplicidade e, embora não sendo muito preciso, pode ser usado como uma primeira aproximação em circuitos mais complexos. Apesar disso, esse modelo é usado pelo simulador SPICE no nível 1 e provoca erros da ordem de até 40% quando comparado ao modelo implementado no nível 2 do SPICE. A simplicidade do modelo quadrático permite que este seja usado em sua forma analítica completa, permitindo assim análise de pequenos sinais de forma também analítica. Embora não podendo ser usado indiscriminadamente para análises quantitativas, o modelo quadrático se presta muito para visualizações qualitativas do funcionamento dos dispositivos. Por exemplo, a transcondutância de um transistor é dada pela derivada da corrente de dreno em relação à tensão V_{gs} . Como a equação de corrente desse modelo é simples, ela permite uma derivação direta e a obtenção de uma expressão analítica fechada para o cálculo da transcondutância, o que possibilita análises qualitativas eficientes na fase de projeto.

A figura 4.1 permite a visualização qualitativa do comportamento da corrente de dreno variando dois parâmetros individualmente: V_{gs} e V_{ds} , para esse modelo simplificado. Analisando inicialmente V_{gs} para V_{ds} fixo, constata-se que a relação é primeiramente quadrática e posteriormente linear, separadas pelo ponto onde V_{gs} é suficientemente alto para levar o dispositivo para a região linear. A simplicidade do modelo permite verificar que quando V_{gs} causa uma tensão de saturação igual a V_{ds} , existe uma mudança de equação, mas que produz, para o mesmo ponto, o mesmo valor de corrente de dreno, garantindo assim a continuidade da função. Por outro lado, ao analisar o segundo gráfico da figura 4.1, vemos o perfil de corrente para variações da tensão V_{ds} , mantendo V_{gs} fixo. O ponto onde I_d perde sua característica linear é o mesmo do gráfico anterior, sendo representado agora pela tensão V_{ds} igual a tensão de saturação causada pelo V_{gs} aplicado. Uma análise de pequenos sinais indica que a impedância do transistor na região linear é baixa (tipicamente $10k\Omega$). Entretanto, quando o dispositivo entra na região de saturação, grandes variações de V_{ds} são necessárias para produzir uma mudança significativa na corrente de dreno, implicando assim em uma alta impedância. A impedância de saída da transistor é modelada pelo parâmetro λ , supostamente constante ao longo de qualquer faixa de operação.

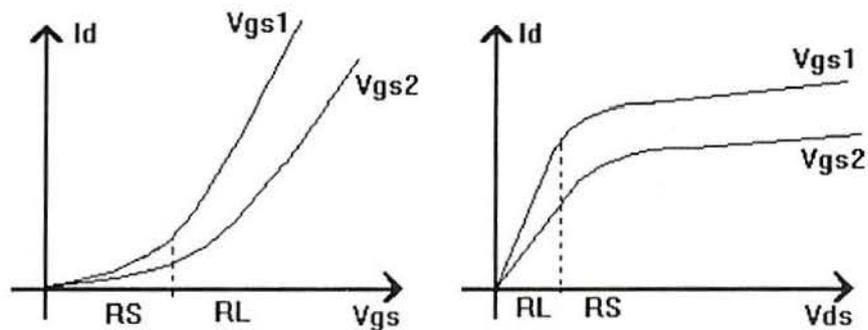


Figura 4.1

O gráfico da figura 4.1 permite a visualização de duas regiões de operação distintas: RS (Região de saturação, onde $V_{ds} \geq V_{gd} - V_t$) e RL (Região Linear, onde $V_{ds} < V_{gs} - V_t$). A partir destas curvas é possível também extrair os parâmetros de pequenos sinais mais importantes do transistor, ou seja: a condutância e a transcondutância. Analiticamente, as expressões que relacionam estes parâmetros com os dados tecnológicos e de polarização podem ser deduzidas diretamente das equações que fornecem a corrente de dreno, de acordo com as seguintes expressões:

$$g_m = \frac{d}{d V_{gs}} I_d = \sqrt{K_p \cdot \frac{W}{L} \cdot (V_{gs} - V_t)}$$

$$g_d = \frac{d}{d V_{ds}} I_d = \frac{1}{\lambda \cdot I_d}$$

Equação 15

As equações anteriores são válidas somente na região de saturação, mas suas formas na região linear podem ser igualmente obtidas derivando a equação de corrente nessa região. Novamente é interessante ressaltar que o mais importante nessas equações é a possibilidade da análise qualitativa da relação entre os parâmetros de pequenos sinais e os dados físicos do transistor, e não do valor absoluto resultante da aplicação destas equações, posto que estes geralmente apresentam um grave problema de falta de precisão.

4.2 Modelo Vladimirescu

Em 1980, Andrei Vladimirescu propôs um modelo eficiente para descrever a característica tensão-corrente de um transistor MOS na região de inversão forte [21]. Infelizmente a complexidade do modelo impede cálculos à mão e seu uso só pode ser feito com auxílio computacional. Embora esse modelo preveja somente uma equação para a corrente de dreno, a descontinuidade na primeira derivada se mantém, pela seguinte razão:

- * Em qualquer ponto de operação, uma função $G(V_{gs}, V_{ds}, V_{bs}, W, L_{eff})$ calcula a tensão dreno-fonte de saturação: V_{dssat} . Se a tensão dreno-fonte for menor que este valor, o transistor está na região linear; ou na saturação, em caso contrário.
- * Se estiver na região linear, uma função $F(V_{gs}, V_{ds}, V_{bs}, W, L_{eff})$ calcula a corrente de dreno para as tensões terminais do transistor.
- * Se estiver na região de saturação, o modelo propõe que se substitua V_{ds} por V_{dssat} na função $F(V_{gs}, V_{ds}, V_{bs}, W, L_{eff})$ para qualquer tensão dreno-fonte que o transistor estiver submetido. Com isso, posteriores aumentos de V_{ds} aumentarão a corrente de dreno somente através da redução do comprimento efetivo do canal (L_{eff}).

Dessa forma, mesmo que se garanta que não existirá uma descontinuidade absoluta na curva de corrente, o fato de utilizar a equação de forma diferente causará uma descontinuidade na primeira derivada da corrente em relação a V_{gs} e V_{ds} . Além dessa característica, o modelo proposto engloba uma grande quantidade de efeitos de segunda ordem, modelando seu comportamento para dimensões de até $2\mu\text{m}$. A seguir, vários pontos importantes da metodologia de cálculo serão abordados, bem como os fenômenos por trás dos efeitos principais.

Uma tensão limiar específica é calculada para cada polarização e tamanho de transistor, de modo a computar os efeitos de geometria e polarização. O parâmetro Delta é responsável pelo aumento do valor nominal de V_t (e conseqüente diminuição da corrente) quando a dimensão W diminui. Esse efeito é conhecido como Efeito de Canal Estreito e pode ser explicado da seguinte maneira: em operação normal a região depleta não está limitada à zona diretamente em baixo da porta, pois algumas linhas de campo que saem da região de porta terminam em átomos ionizados nas áreas implantadas. Se o valor de W é suficientemente grande (maior que $10\mu\text{m}$), essa zona extra pode ser negligenciada, pois sua representação percentual é pequena. Em caso contrário, essa região pode se tornar relevante e desempenhar um papel importante. Nesse caso, para que a porta de um transistor estreito forme um canal com a mesma intensidade de um largo, V_{gs} deverá ser maior no primeiro, uma vez que parte da região depleta estará fora da área do canal. Esse modelo prevê este funcionamento adequando o valor de V_t de acordo com a largura do canal, de maneira linear.

Os efeitos de canal curto tem um equacionamento algo mais complexo, mas os fenômenos envolvidos podem ser explicados da seguinte forma: existe um encurvamento na região de depleção nas zonas próximas às extremidades do dreno e fonte. Este encurvamento é chamado de efeito de canto (*Edge Effect*). Quando o valor de L é grande, este fenômeno pode ser negligenciado, mas assume proporções mais importantes quando a dimensão L se reduz. Experimentalmente, mostra-se que a tensão V_{gs} necessária para produzir uma dada corrente de dreno é menor do que aquela fornecida pela teoria de canal longo, devido a efeitos como rebaixamento da barreira de potencial. Esse efeito sugere que quando as extremidades de dreno e fonte estão muito próximas a profundidade do canal tende a aumentar, causando um potencial superficial mais significativo. Pelo fato de que isto torna o canal mais atrativo para os portadores, a corrente que passa pelo canal será maior do que a esperada. Esse fenômeno é equacionado no modelo através de uma redução de V_t quando L é muito pequeno.

Um efeito muito importante que não pode ser deixado de lado, sob pena de incorrer em graves erros, é a saturação de velocidade. Esse fenômeno ocorre para tensões V_{gs} mais altas e é representado por uma não-proporcionalidade entre a velocidade dos portadores V_d e o campo elétrico longitudinal E_x (no sentido do fluxo de portadores). Para pequenos valores de campo elétrico, a velocidade dos portadores é proporcional a este campo e a derivada dessa curva representa a mobilidade dos portadores, conforme a representação da Figura 4.2.

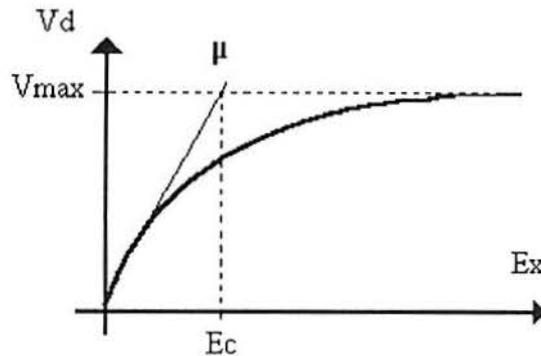


Figura 4.2

Dessa maneira, a mobilidade nominal imposta pela tecnologia pode ser usada para valores de campo elétrico inferiores a E_c (Campo Crítico). Após este valor uma formulação empírica é usada para descrever o comportamento de redução de mobilidade. Percebe-se também que, para valores muito grandes do campo elétrico longitudinal (E_x), a velocidade dos portadores atinge um valor máximo, no qual a mobilidade tende a zero. Esse efeito se faz sentir na característica tensão-corrente do transistor através de uma redução na corrente de dreno em relação à esperada para tensões V_{gs} muito grandes e uma conseqüente descaracterização do regime quadrático. Nesses valores de tensão, o campo elétrico induzido no interior do transistor não é proporcional à velocidade dos portadores, que por sua vez é proporcional à corrente. Logo a corrente que efetivamente passa pelo transistor é bem inferior àquela prevista pela modelo sem a inclusão da redução da mobilidade. Embora os fenômenos envolvidos na saturação de velocidade sejam conhecidos, a correção da mobilidade é feita através de uma equação empírica, que usa três parâmetros: U_{crit} , U_{tra} e U_{exp} . U_{crit} é o valor máximo do campo elétrico longitudinal para o qual ainda não existe necessidade de corrigir a mobilidade, U_{tra} é um parâmetro de relaciona a redução de mobilidade com o campo elétrico transversal do interior do dispositivo e U_{exp} é um valor empírico que permite o cálculo aproximado da verdadeira mobilidade. Para valores elevados de tensão de porta, a corrente de dreno não responde de maneira aproximadamente quadrática para variações em V_{gs} , conforme a teoria. Ao contrário, a relação passa a ser quase linear e existe uma tendência de

degradar mais ainda a mobilidade quando a dimensão L começa a se reduzir. Embora tenha sido dito que os portadores não podem ter velocidades superiores a V_{max} (valor assintótico da velocidade), por breves instantes de tempo e pequenas distâncias, os portadores podem experimentar uma velocidade muito grande, conhecida como *Velocity Overshoot*. Os fenômenos por trás deste efeito tem uma natureza complexa, reportando-se à física quântica e ainda não foram modelados satisfatoriamente por equações empíricas.

Um outro efeito de importância fundamental e que é modelado pelo equacionamento proposto por Vladimirescu é a Modulação do comprimento do canal [15].

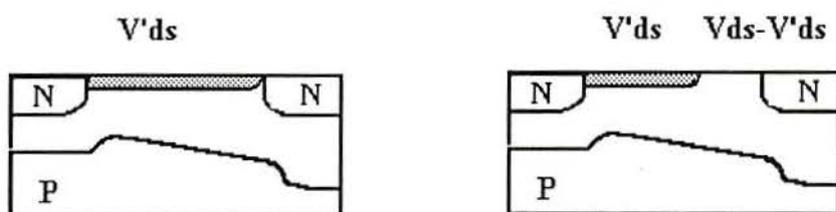


Figura 4.3

Ambos desenhos da figura 4.3 representam um transistor NMOS com um canal (hachurado) ligando duas extremidades (dreno e fonte) construídas de material N, acima da região de depleção, dentro de um substrato P. Na primeira, a tensão V_{ds} aplicada nos terminais do transistor é exatamente igual à tensão de *pinch-off* (V'_{ds}), ou seja, a tensão necessária para fazer o canal se estender exatamente até o terminal de fonte. Para valores de V_{ds} inferiores a V'_{ds} , o transistor está na região linear e sua impedância é substancialmente inferior àquela observada em regime de saturação. Se a tensão V_{ds} assume valores maiores que V'_{ds} , o canal não pode suportar esse excesso de tensão, criando uma região de baixa impedância entre o canal e o dreno. O resultado desse fenômeno é uma redução do comprimento efetivo do canal por efeito de aumento de V_{ds} . Este mecanismo está presente no modelo da seguinte forma: quando V_{ds} é maior que a tensão de saturação, a equação que calcula a corrente de dreno terá como parâmetro o valor dessa tensão de saturação e não o valor real de V_{ds} . Para evitar que valores continuamente crescentes de V_{ds} produzam sempre o mesmo valor de corrente, a redução do comprimento efetivo é implementada, permitindo assim aumentos no valor da corrente. Esse efeito é conhecido como Modulação de canal, pois o comprimento efetivo do canal depende das condições de polarização do transistor, sendo portanto modulado por este. Esse foi um dos primeiros "Efeitos de canal curto" a ser estudado,

mas seu efeito não se aplica só para transistores de geometria reduzida, sendo perfeitamente observável para dispositivos de tamanhos maiores.

Um dos parâmetros mais importantes no cálculo da corrente de dreno na região de saturação do transistor é a tensão de saturação. Embora tenha um equacionamento bastante complexo, um aspecto de seu cálculo será discutido aqui, devido a uma simplificação que pode levar a erros significativos. Para o cálculo deste valor, duas equações devem ser simultaneamente resolvidas, pois uma delas requer o valor de L_{eff} (o comprimento efetivo do canal), que só pode ser calculado de posse do valor da tensão de saturação. Dada a complexidade das expressões individuais, a resolução de um sistemas de equações como esse seria um processo por demais trabalhoso e demorado para ser implementado num modelo. Por isso, uma simplificação é feita no sentido de admitir que em uma das equações $L_{eff} = L - 2 \cdot L_d$. Essa aproximação causa erros mais significativos quando os dispositivos usados tem dimensões reduzidas, porque a influência percentual de tal simplificação aumenta o erro. Segundo a própria literatura deste modelo, que é usado no nível 2 do SPICE, a resolução simultânea das equações seria um processo dispendioso, cujo acréscimo de precisão não compensaria o custo de demora. Embora exista um método iterativo para resolução deste sistema (Método Ferrarri) onde se pode achar uma solução com no máximo 3 iterações, o SPICE impede, por razões filosóficas, que existam laços iterativos dentro do modelo dos dispositivos. Dessa maneira, o modelo sugere que se conviva com os erros e não resolva-se o sistema de equações.

A maior vantagem da utilização desse modelo é o seu compromisso entre complexidade e precisão. Embora seu caráter eminentemente numérico impeça qualquer tipo de tratamento analítico, mesmo que qualitativo, uma boa descrição do comportamento DC, isto é, característica I-V, é fornecida. Infelizmente isto não é suficiente, especialmente quando se trata de circuitos analógicos. A análise de pequenos sinais é a ferramenta mais importante que se dispõe para projeto de blocos analógicos, de maneira que uma boa descrição dessa característica seria muito interessante para um modelo usado nesse tipo de projeto. Infelizmente, não é o caso desse modelo, que proporciona uma descrição pobre da característica de pequenos sinais do transistor. Naturalmente, o modelo não é o único responsável por essa má representação. As *foundries* costumam fornecer os parâmetros tecnológicos para serem usadas nos simuladores comercialmente disponíveis. Esses parâmetros possuem métodos de extração definidos, que, quando em conjunto, fazem com que o modelo usado dê o mínimo possível de erro em comparação a um circuito fabricado naquela tecnologia. Entretanto, a extração deste parâmetros visa primeiramente obter uma boa concordância

com a característica I-V do dispositivo, isto é, obter um conjunto de parâmetros que possam conjuntamente reproduzir a curva de tensão-corrente do transistor. A análise de pequenos sinais, quando tem algum papel, é secundário, relegando a segundo plano o modelamento das características dinâmicas do transistor. Essa negligência, em conjunto com a utilização de modelos com descontinuidades na derivada primeira da corrente, proporciona a criação de modelos com sérias limitações de desempenho, mesmo para dispositivos de geometria grande.

Embora tendo suas limitações, esse modelamento foi usado para implementação no auxílio da síntese semi-automática de amplificadores operacionais. Essa escolha foi feita basicamente por dois motivos:

* Diminuir os erros percentuais abaixo de 2 ou 3% na previsão de corrente de um modelo significa multiplicar os esforços de cálculo, inviabilizando-se portanto, uma vez que as incertezas no processo de fabricação estatisticamente gerarão erros com essa ordem de grandeza. Assim, concentrar esforços no sentido de reduzir os erros a níveis inferiores a 1% não é uma tática inteligente. Esse fato permite a utilização deste modelo que, embora tendo erros na característica tensão-corrente, estas discrepâncias não chegam a ser críticas frente ao processo de fabricação.

* Implementar um modelo de transistor não é uma tarefa simples, especialmente quando o modelo utilizado possui uma grande complexidade. Para tanto, é necessário ter um parâmetro de comparação para determinar se a implementação foi feita de uma forma matematicamente correta. Como esse modelo já foi consagrado por sua eficiência em inversão forte, ele está presente em alguns simuladores elétricos comercialmente disponíveis, como o SPICE. Dessa forma, comparações diretas entre a resposta produzida pelo SPICE e pela implementação do modelo podem validar essa implementação.

4.3 Modelo EKV

Lançado ao público recentemente [22], o modelo EKV (iniciais de três pesquisadores: Christian C. Enz, François Krummenacher e Eric A. Vittoz) propõe integrar as três regiões de operação em um único equacionamento, que, além de simples e preciso, não possui descontinuidades. A origem deste modelo reporta-se à análise das correntes de difusão e deriva em inversão fraca, moderada e forte. Através do uso de

uma função interpoladora permite-se a supressão das descontinuidades ao longo das três regiões de operação. Assintoticamente, essa função se assemelha a uma exponencial em inversão fraca e a uma quadrática em inversão forte, com um comportamento intermediário na região de inversão moderada. A simplicidade do equacionamento de correntes desse modelo torna seu uso propício para cálculos iterativos, mas a sua ausência nos simuladores elétricos mais conhecidos torna seu uso difícil. Alguns dos parâmetros tecnológicos desse modelo se assemelham aos do SPICE nível 2, sendo que outros requerem técnicas especiais de extração. Esse equacionamento usa o princípio da simetria funcional do transistor MOS em relação aos terminais de dreno e fonte; por isso, todas as tensões do dispositivo são referenciadas à tensão do substrato local.

Um dos parâmetros mais importantes desse modelo é a tensão de *pinch-off*, na qual a carga de inversão se torna zero. Naturalmente, essa tensão é função direta da excitação da porta, através da equação 16 [23]:

$$V_P = V_G - V_{to} - \gamma \cdot \left[\sqrt{V_G - V_{to} + \left(\Psi_0 + \frac{\gamma}{2} \right)^2} - \left(\Psi_0 + \frac{\gamma}{2} \right) \right]$$

$$\gamma = \frac{\sqrt{2 \cdot q \cdot \epsilon_s \cdot N_{sub}}}{C_{ox}}$$

Equação 16

onde V_{to} é definida como a tensão de porta na qual a densidade de cargas móveis na região de inversão é zero [19], γ é o conhecido efeito de corpo e Ψ_0 é o potencial superficial. A derivada da tensão de *pinch-off* pela tensão de porta resulta no fator de declividade n , que pode também ser expresso em função de V_G .

$$n = \frac{d}{dV_G} V_P = 1 + \frac{\gamma}{2 \cdot \sqrt{\Psi_0 + V_P}} = 1 - \frac{\gamma}{2 \cdot \sqrt{V_G - V_{to} + \left(\frac{\gamma}{2} + \sqrt{\Psi_0} \right)}}$$

$$V_P = \frac{V_G - V_{to}}{n(V_G)}$$

Equação 17

Para as tecnologias usualmente empregadas, V_P pode ser representado por uma relação linear de V_G , através do parâmetro n , dado na equação. As relações colocadas até aqui prevêem dopagem uniforme do substrato, embora esteja se tornando cada vez

mais comuns tecnologias que utilizam implantação iônica para correção do valor de V_t . Esses casos merecem equacionamento diferenciado, para levar em conta a não-uniformidade do substrato e, por sua adicional complexidade, não serão abordados aqui. Analisando gráficos de carga de inversão por V_p , constata-se que o equacionamento sugerido prevê um acúmulo maior do que aquele previsto pela teoria, mesmo se for feita uma simplificação da equação, tomando simplesmente a primeira ordem da expansão de Taylor. Para ajustar essa resposta de maneira conveniente, a seguinte técnica é empregada: ao invés de admitir parâmetros como V_{to} , Ψ e γ como dados relacionados a quantidades físicas; estas constantes são passíveis de ajuste para melhorar a aproximação. Dessa forma, estes parâmetros de ajuste permitirão uma concordância maior com a teoria e assim uma melhor descrição do desempenho do dispositivo.

A expressão da corrente de dreno é obtida diretamente a partir da sua definição, que inclui correntes de deriva e difusão. Separando os limites de integração de modo a ter uma corrente direta e outra reversa a partir desse equacionamento, é possível que se tenha uma descrição genérica do comportamento do transistor em qualquer região de operação, posto que não foi feita nenhuma restrição quanto a isso. Beta representa o produto da mobilidade, capacitância do óxido e relação de aspecto e V_{ch} representa a tensão ao longo do canal.

$$I_d = \beta \cdot \int_{V_s}^{\infty} \frac{-Q_{inv}(V_{ch})}{C_{ox}} dV_{ch} - \beta \cdot \int_{V_d}^{\infty} \frac{-Q_{inv}(V_{ch})}{C_{ox}} dV_{ch} = I_f - I_r$$

Equação 18

A corrente direta (I_f) depende somente da diferença $V_p - V_s$ (onde V_s é a tensão no terminal fonte) e a corrente reversa depende de $V_p - V_d$. Integrando essa expressão, obtém-se as expressões de corrente direta e inversa nas regiões de inversão forte e fraca, como segue:

$I_f = \frac{n \cdot \beta}{2} \cdot (V_p - V_s)^2$	$V_s < V_p$	$I_r = \frac{n \cdot \beta}{2} \cdot (V_p - V_d)^2$	$V_d < V_p$	Inversão fort
$I_f = 0$	$V_s \geq V_p$	$I_r = 0$	$V_d \geq V_p$	
$I_f = 2 \cdot n \cdot \beta \cdot U_t^2 \cdot e^{\frac{V_p - V_s}{U_t}}$		$I_r = 2 \cdot n \cdot \beta \cdot U_t^2 \cdot e^{\frac{V_p - V_d}{U_t}}$		Inversão frac

Equação 20

A maioria dos modelos descreve o comportamento do transistor em diferentes regiões de operação usando equações diferentes. O compromisso entre complexidade e precisão que um modelamento nesses moldes proporciona é razoável, mas o equacionamento falha seriamente ao ser descontínuo em uma análise que envolva a primeira derivada da corrente. Para evitar isso, o EKV propõe a interpolação de uma função entre as duas regiões de comportamento conhecido (inversão fraca e forte), normalizando o valor de todas as tensões em relação à tensão termodinâmica U_t e todas as correntes em relação à corrente específica I_s ($I_s = 2 \cdot n \cdot \beta \cdot U_t^2$). O coeficiente de inversão i_f é dado pela normalização da corrente direta: I_f / I_s . Uma proposta simples e eficiente de interpolação foi proposta por H. Oguey e S. Cserveny [27], que propõe uma equação logarítmica para prever o comportamento do transistor na região de inversão moderada.

$$\begin{aligned}
 F(v) &= \left(\frac{v}{2}\right)^2 & V \gg 0 & & F(v) &= \ln\left(1 + e^{\frac{v}{2}}\right)^2 \\
 F(v) &= e^v & V \ll 0 & & & &
 \end{aligned}$$

Equação 21

Aplicando essa função na definição de correntes e invertendo para ter a relação de tensões, tem-se:

$$\begin{aligned}
 i_f &= F(v_p - v_s) = \ln\left(1 + e^{\frac{v_p - v_s}{2}}\right)^2 & v_p - v_s &= 2 \cdot \ln\left(e^{\sqrt{i_f}} - 1\right) \\
 i_r &= F(v_p - v_d) = \ln\left(1 + e^{\frac{v_p - v_d}{2}}\right)^2 & v_p - v_d &= 2 \cdot \ln\left(e^{\sqrt{i_r}} - 1\right)
 \end{aligned}$$

Equação 22

Embora sendo uma boa aproximação, uma interpolação mais precisa pode ser obtida integrando a característica de transcondutância do transistor. Por fugir do tema principal deste trabalho, essa metodologia será deixada de lado.

Uma nova definição de parâmetros de pequenos sinais é feita para este modelo, dando origem a três transcondutâncias, referidas a variações de fonte-dreno, porta-dreno e porta-fonte. Na equação 23 tem-se a definição de cada uma dessas variáveis, assim como suas relações com os parâmetros de pequenos convencionais.

$$\begin{array}{ccc}
g_{mg} = \frac{d}{dV_g} I_d & g_{ms} = \frac{d}{dV_s} I_d & g_{md} = \frac{d}{dV_d} I_d \\
g_m = g_{mg} & g_{mb} = g_{ms} - g_{mg} - g_{md} & g_{ds} = g_{md}
\end{array}$$

Equação 23

Utilizando estas definições nas equações de corrente mencionadas anteriormente, chega-se a expressões para esses valores em regime de inversão forte e fraca. Similarmente, uma função será usada para fazer a interpolação na região de inversão moderada. De modo a fazer uma interpolação independente do tamanho do transistor, as transcondutâncias são normalizadas para seus máximos valores, obtidos na região de inversão fraca. Esses parâmetros podem ser escritos em função da corrente direta normalizada, usando uma função de interpolação $G(i)$.

$$\begin{array}{ll}
G_{if} = \frac{g_{mg} \cdot n \cdot U_t}{I_f} & G(i) = 1 \quad i \ll 1 \text{ (inversão fraca)} \\
G_{ir} = \frac{g_{md} \cdot n \cdot U_t}{I_r} & G(i) = \sqrt{i} \quad i \gg 1 \text{ (inversão forte)}
\end{array}$$

$$G_{if} = \left[\frac{1}{if} \cdot \left(\frac{d}{dvs} if \right) \right] = \frac{1}{if} \cdot \frac{d}{dv} F(v) = \frac{1 - e^{-\sqrt{if}}}{\sqrt{if}}$$

Equação 24

A função transcondutância deve parecer uma constante em inversão fraca e cair linearmente (em um gráfico logarítmico) em inversão forte. Usando a função de corrente dada, uma expressão é obtida na equação 25, que prevê valores maiores em inversão fraca e menores em inversão forte, quando comparado aos dados numéricos. Esse erro é da ordem de 6% e pode ser sensivelmente reduzido usando a seguinte equação:

$$G_{if} = \frac{1}{\sqrt{if + \frac{1}{2} \cdot \sqrt{if} + 1}}$$

Equação 25

O erro dos parâmetros de pequenos sinais correspondente ao emprego dessa equação é menor que 3%, sendo assim extremamente conveniente para seu uso em simulação, por sua simplicidade e precisão. Pode-se obter, a partir desse modelamento de pequenos sinais, uma descrição da característica I-V do transistor através de uma

integração. Infelizmente, a função resultante dessa operação é complexa demais para ser usada com objetivos de simulação, devendo ser deixada de lado.

A idéia de usar curvas interpoladoras para descrever de forma contínua o comportamento do transistor em todas as regiões de inversão provou ser muito eficiente. Unindo a isso a correção de alguns parâmetros por efeitos de segunda ordem pode-se construir um modelo completo capaz de ser implementado em um simulador elétrico. Com isso, a restrição da maioria dos simuladores analógicos (a má descrição da inversão moderada) tende a acabar, permitindo que se projete eficientemente circuitos em qualquer região de operação. Esse modelo (derivado do *Charge Sheet*) privilegia a descrição precisa da característica I-V e de pequenos sinais do transistor, sem se restringir a uma região de operação em especial e com a supressão das descontinuidades da derivada primeira da corrente de dreno.

4.4 O Modelo Ótimo do transistor MOS.

Vários modelos de transistor foram criados e publicados nos últimos anos, com o objetivo de descrever de forma acurada certo tipo de fenômeno ou geometria. Por exemplo, a crescente necessidade de miniaturização tem impelido os projetistas de circuitos analógicos a usar geometrias cada vez mais reduzidas. Logo as topologias empregadas devem tornar-se cada vez mais eficientes no quesito ocupação de área. Por outro lado, para projetar circuitos com essas características, modelos apropriados devem ser desenvolvidos, com objetivo de enfatizar os efeitos secundários causados pela geometria pequena. Nesses casos, esses efeitos ditos "secundários" começam a ter papel significativo na composição total da corrente.

O EKV é um exemplo de modelo que surgiu frente à necessidade de projetar circuitos analógicos cujos elementos não funcionassem somente dentro do regime de inversão forte. Conclui-se portanto que necessidades de projeto patrocinam os esforços orientados no sentido de criação de novos modelos. Esses modelos devem ter algumas características que os diferenciarão de um simples conjunto de equações interconectadas por algumas grandezas físicas. A facilidade de implementação, complexidade das equações, precisão, facilidade na extração de parâmetros, etc, são requisitos que um modelo deve atender para merecer a designação de "um bom modelo". De maneira geral, esses modelos devem observar alguns pressupostos, entre os quais pode-se destacar [17,22]:

* Descrever de forma precisa a característica I-V, parâmetros de pequenos sinais e elementos parasitas.

* Ser independente da geometria usada

* Possuir uma metodologia simples de extração de parâmetros

* Funcionar de maneira simplificada para aplicações não críticas e testes, com um conjunto menor de parâmetros

* Modelar o funcionamento do dispositivo em qualquer região de operação

* Permitir modelamento do ruído interno do dispositivo

* Implementar análise térmica

* Ser computacionalmente simples

Infelizmente, a despeito das várias tentativas que já foram feitas e das muitas que ainda o serão, não existe um modelo de aceitação difundida que englobe a maioria das necessidades expostas. Até que modelos com estas características estejam disponíveis no mercado, projetos de circuitos de alto desempenho com dimensões reduzidas operando em qualquer região de inversão terão importantes dificuldades de simulação e conseqüentemente de projeto.

5. Implementação do modelo Vladimirescu

5.1 Objetivos Gerais

Para a maioria das aplicações, os recursos de análise fornecidos pelos simuladores é suficiente para resolver os problemas mais comuns de forma rápida e eficiente. Um exemplo disso seria o de polarizar adequadamente um transistor operando como fonte de corrente. Mantendo fixas as suas dimensões, qual deve ser o valor de V_{gs} para que este gere uma corrente I_d ? A resposta para esse problema pode ser dada por qualquer simulador que permita uma análise DC, pois variando a tensão de porta do transistor, cria-se variações na corrente de dreno e deve existir um valor de V_{gs} que cause a corrente desejada. Entretanto, o problema contrário não permite soluções tão rápidas, isto é, supondo um transistor com polarização fixa, qual deve ser seu fator de forma para que conduza a mesma corrente I_d ? Obviamente, o mesmo simulador usado no problema anterior também seria capaz de oferecer uma resposta, mas à custa de algum trabalho manual do projetista, que, à luz de algumas equações, tentará variar as dimensões do dispositivo até que este conduza à corrente desejada.

É óbvio que o tempo envolvido na resolução do segundo problema é maior, uma vez que o simulador não permite uma análise na qual o tamanho de um transistor assuma mais de um valor. A restrição imposta nesse exemplo reflete bem a precariedade de recursos automáticos de projetos dentro do ambiente de simulação. Devido ao fato de que somente as simulações mais clássicas estão presentes na maioria dos simuladores (DC, AC, temporal, Monte Carlo, etc...), o método de tentativa e erro torna-se a única ferramenta de que dispõe o projetista quando se vê frente a um problema específico como o que foi sugerido.

A partir dessa deficiência, surge a necessidade de criar softwares que permitam, dentro de certos limites, resolver problemas simples, mas não muito comuns (ponto nos quais os simuladores são pobres). A utilização de rotinas abertas (com acesso ao código fonte) permite a resolução desses problemas específicos de maneira automática, o que representa um ganho de tempo considerável, sem mencionar o fato de que a resposta final geralmente é mais otimizada que aquela gerada pela método manual.

Naturalmente, a qualidade das respostas geradas por essas rotinas está diretamente ligada ao modelo de transistor usado. Para ter uma base de comparação disponível, o modelo empregado nessas rotinas é o mesmo do SPICE nível 2 (baseado no modelo proposto por Andrei Vladimirescu). As rotinas de "solução de problemas" estão estruturalmente separadas da rotina que implementa o modelo pois, dessa forma, mudanças na implementação do modelo ou mesmo troca de modelo são possíveis sem alterações na sintaxe ou no algoritmo da rotina principal. Embora conheça-se as deficiências e idiosincrasias do SPICE, o modelo usado neste trabalho foi o deste simulador, pois dessa forma poder-se-ia confirmar e validar as técnicas de síntese de uma maneira muito rápida. Se, ao contrário, se tivesse utilizado um modelo baseado em dados experimentais, por exemplo, a validação das referidas técnicas só poderia ser feita através de medições no circuito construído; um processo por demais demorado. Para promover uma maior agilização nessa etapa, tentou-se reproduzir o modelo usado no SPICE, de modo que as rotinas de síntese e o simulador usado para conferência apresentassem resultados semelhantes. Por outro lado, nada impede que, quando as rotinas já tiverem provado ser eficientes, que se altere o modelo para outro baseado em dados experimentais, ou mesmo outro tipo de modelamento. A ausência do simulador trará alguns problemas de insegurança e desconforto, mas o processo se dará com precisão ainda maior.

A plataforma de trabalho usada nessa implementação foi o MATLAB, cujo desempenho a nível de velocidade é somente razoável, mas dispõe de uma grande quantidade de funções matemáticas que facilitam enormemente a resolução das equações que compõem o modelo. Entretanto, uma vez que a rotina que implementa o modelo não necessita mais de alterações ou correções de nenhuma natureza, a utilização de outra plataforma de trabalho mais rápida (como C) está sob estudo. O Anexo 2 traz a codificação MATLAB da implementação do modelo e sua característica mais importante pode ser observada na primeira linha de código: a despeito de toda a complexidade das equações envolvidas, a rotina retorna um e somente um valor: **a corrente de dreno**. Embora existam outros fatores, inclusive mais importantes que este, as rotinas foram estruturadas dessa maneira porque verificou-se a possibilidade de executar um grande número de funções usando somente a informação de corrente. Estas funções serão posteriormente analisadas em detalhe, pois elas formam o núcleo das rotinas de síntese de amplificadores operacionais.

5.2 Descrição do algoritmo

O funcionamento dessa rotina é sequencial, isto é, prescinde de técnicas iterativas para o cálculo da corrente. Os passos podem ser estruturados da seguinte forma:

- 1) Recebimento dos dados de entrada (tensões terminais, tamanhos e tipo)
- 2) Carregamento de parâmetros tecnológicos para o tipo de transistor solicitado.
- 3) Desconto da difusão lateral no comprimento do transistor
- 4) Cálculo do potencial interno corrigido para efeitos de canal estreito
- 5) Cálculo do coeficiente de efeito de corpo corrigido devido à realimentação estática dreno-porta.
- 6) Cálculo da tensão limiar específica para geometria e polarização
- 7) Redução de mobilidade se o campo elétrico é superior a UCRIT
- 8) Cálculo da tensão dreno-fonte de saturação em duas formas possíveis:
 - 8.1) Se V_{max} (Máxima velocidade dos portadores) não é fornecido, o programa calcula V_{dsat} através da equação:

$$V_{dsat} = \frac{V_{gs} - V_{bin}}{v} + \frac{1}{2} \cdot \frac{\gamma_s^2}{v} \left[1 + 4 \cdot \left(\frac{v}{\gamma_s} \right)^2 \cdot \left(\frac{V_{gs} - V_{bin}}{v} + 2 \cdot \phi_f - V_{bs} \right)^{\frac{1}{2}} \right]$$

Equação 26

8.2) Se V_{max} é fornecido (que é o caso), o cálculo da tensão de saturação se baseia no fato de que a corrente de dreno satura porque os portadores de carga atingem suas velocidades máximas antes da tensão de *pitch-off*, para dispositivos de canal curto. Dessa forma, duas equações muito complexas devem ser resolvidas simultaneamente e a simplificação mencionada no Capítulo 6 é feita ($L_{eff} = L$), de forma que a obtenção de V_{dsat} implica da solução de uma equação de quarto grau. A menor raiz positiva encontrada na resolução desse polinômio (através da função ROOTS.M) é a solução procurada. Se não houver raízes reais, o programa deve calcular V_{dsat} através da

equação anterior e mostrar uma mensagem de erro. Essa precaução não foi tomada nessa implementação, pois ao longo de todo processo de programação e utilização das rotinas esse fato nunca foi observado.

9) Cálculo do comprimento efetivo do transistor, causado pelo encurtamento do canal para tensões dreno-fonte superiores à de saturação.

10) Cálculo da corrente de dreno, de duas formas:

10.1) Se $V_{ds} < V_{dsat}$, o transistor está na região linear e a corrente é obtida inserindo-se o valor real de V_{ds} na equação principal.

10.2) Se $V_{ds} \geq V_{dsat}$, o transistor está saturado e o valor dessa corrente é calculado usando V_{dsat} no lugar de V_{ds} , para qualquer $V_{ds} \geq V_{dsat}$. Os incrementos de corrente oriundos de aumentos de V_{ds} são implementados através da redução do comprimento efetivo do canal.

Embora esse mesmo modelo seja usado também em inversão moderada pelo SPICE (através da multiplicação por uma exponencial), essa implementação garante uma corrente essencialmente zero para quaisquer dispositivos com V_{gs} abaixo de V_{th} (onde V_{th} é a tensão limiar específica para um transistor). Além disso, evita-se utilizar o transistor com tensão de porta muito pequena, pois a influência do comportamento de inversão moderada pode se sobrepor ao do inversão forte, causando erros. Dessa forma, somente as características de inversão forte são cobertas por essa implementação, deixando de lado os outros regimes de funcionamento .

5.3 Utilização e principais características

Para que a rotina possa efetivamente calcular a corrente de dreno do transistor, vários dados devem ser informados, conforme mostra a sintaxe do chamamento da função:

```
function I = quads(Vgs, Vds, Vbs, W, L, tipo)
```

onde os valores são explicitados usando o sistema MKS (Volts e metros). Tipo é um *flag* com dois valores possíveis: 0 para NMOS e 1 para PMOS. Como a rotina retorna somente um valor, um trecho de programa que faz uma chamada à rotina QUADS (que

contém a implementação do modelo) deve usar a operação de atribuição a uma variável escalar (ou elemento de um vetor ou matriz), da seguinte forma:

```
...  
I_dreno = quads(1.3 , 2.0 , 0 , 100e-6 , 20e-6 , 0);  
...
```

Obviamente, simplesmente determinar o valor da corrente não é suficiente para fazer um projeto, de forma que essa propriedade deverá ser usada em conjunto com algumas técnicas de programação para tornar seu uso funcional. Por exemplo, para resolver o problema da determinação do tamanho do transistor que gera uma corrente I_d para uma polarização fixa, a rotina QUADS.M é extensamente usada, conforme o segmento de programa abaixo:

```
I1 = 20µA; tol = 1e-4; Id1 = 0; WL1 = 1.0;  
  
while abs( (Id1-I1)/I1 ) > tol  
  Id1 = quads(Vgs1,Vds1,Vbs1,WL1*L1,L1,1);  
  WL1 = I1*WL1 / Id1;  
end
```

Supondo que V_{gs1} , V_{ds1} e V_{bs1} são as tensões terminais de um transistor PMOS, $WL1$ é o seu fator de forma (logo $WL1 \cdot L1 = W1$), $L1$ é seu comprimento (fornecido pelo usuário) e $I1$ é a corrente que deve circular por ele. O programa funciona da seguinte maneira:

- 1) Uma vez definidas todas as variáveis (onde tol é a tolerância permitida), a relação de aspecto deste transistor é arbitrariamente igualada a 1.0 ($W1 = L1$), embora pudesse ser qualquer valor maior que zero.
- 2) O programa entra em um laço do qual só sairá quando o erro entre $Id1$ (corrente que realmente passa no transistor) e $I1$ (corrente que deveria passar) for menor que 0.01%.
- 3) Usando o valor inicial de $WL1 = 1.0$, o programa chama QUADS.M para calcular a corrente com este tamanho.
- 4) A menos de uma causalidade, a corrente resultante da relação $WL1$ não será igual à corrente esperada, de modo que uma correção deve ser feita em $WL1$. Essa correção pode ser feita de muitas formas, como por exemplo: $WL1 = WL1 + \delta$, onde δ é um passo pequeno, mas não nulo. Duas consequências negativas advém do uso dessa

técnica: primeiramente a reduzida precisão oferecida pelo incremento delta e depois o tempo consumido para reduzir o erro. Por isso, uma técnica alternativa é usada, que corrige WLI através de uma multiplicação, que se baseia na aproximada proporcionalidade que existe entre corrente e a relação de aspecto. Logo, um tamanho errado gera uma corrente errada, mas a relação dessas grandezas é aproximadamente a relação tamanho certo-corrente certa, de modo que a geometria certa pode ser calculada usando o tamanho errado, a corrente certa e a errada, de acordo com a equação do programa.

5) Com esse novo valor de fator de forma, uma nova corrente será calculada, que se aproximará velozmente do valor de corrente solicitada. Para os dados do problema (que são típicos), não mais que 4 iterações são necessárias para obtenção de um tamanho ótimo para conduzir a corrente desejada.

Dessa forma, com apenas 5 linhas de programação e menos de 1 segundo, é possível resolver um problema que com um simulador convencional levaria alguns minutos. Obviamente, a proposta aqui não é construir um simulador, mas sim criar meios para a solução automática de problemas que são relativamente simples e exigem considerável tempo de processamento. A viabilidade da construção dessa ferramenta foi provada na etapa de síntese de operacionais, na qual vários problemas sem solução automática apareciam e faziam da síntese um processo longo e pouco eficiente. Estes fatos são confirmações dos benefícios que se obtém ao se trabalhar com um código fonte próprio, passível de mudanças e confirmações.

5.4 Análise da qualidade da implementação

Antes de ser efetivamente usado, o programa QUADS passou por uma bateria de testes, nos quais pode-se evidenciar suas virtudes e defeitos. A determinação da eficiência deste modelo ao descrever o comportamento do transistor foi feita através de comparações com os resultados fornecidos pelo SPICE, embora comparações com a realidade teriam sido mais úteis. Infelizmente, como esta é a primeira versão do programa de síntese de caráter automático, era necessária uma base sólida de comparação onde se pudesse verificar o funcionamento correto do circuito como um todo. Como este é o mesmo modelo que o SPICE usa no nível 2, seria de se esperar que os resultados fossem exatamente iguais. Entretanto, duas características forçam a existência de certa diferença:

1) A simplificação feita para o cálculo de V_{dsat} ($L_{eff} = L - 2 \cdot L_d$) é explicitamente sugerida no próprio trabalho onde o modelo é apresentado. Entretanto, é possível que versões mais modernas do SPICE já tenham implementado uma nova forma de resolução rápida das equações, que permita a supressão da simplificação. Infelizmente, não se pode assegurar este fato com certeza, posto que informações mais específicas sobre funcionamento interno do simulador são difíceis de obter.

2) A segunda razão é o efeito de resistências série nos terminais de dreno e fonte, de acordo com a figura 5.1. Mesmo para um circuito simples como esse, o procedimento interno dos simuladores (que têm facilidade com cálculo iterativo e atribuição de nós) é incluir resistências como R_s e R_d , criando assim nós intermediários que consomem parte da tensão V_{ds} disponível, causando assim um erro nem sempre negligenciável.

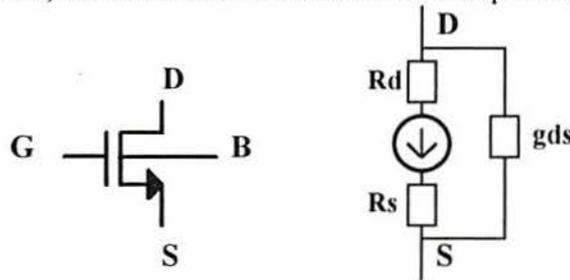


Figura 5.1

Esse problema poderia ser resolvido da seguinte forma:

- 1 - Calcula-se a corrente ignorando-se as resistências séries
- 2 - Calcula-se um novo V_{ds} , resultante da diferença entre o verdadeiro V_{ds} e a queda de tensão $I_d \cdot R_s$ e $I_d \cdot R_d$ no transistor.
- 3 - Usando-se este novo V_{ds} , calcula-se novamente a corrente.
- 4 - Repetindo (2) e (3) várias vezes, o valor final da corrente incluirá o efeito de resistência série.

Embora teoricamente certo, esse método é pouco prático, pois torna um simples cálculo de corrente em um processo iterativo relativamente longo, indo na direção contrária à sua principal função: economizar tempo. Mesmo assim, análises comparativas entre SPICE e QUADS foram feitas para um dispositivo de dimensões grandes ($50\mu\text{m} \times 10\mu\text{m}$), revelando uma concordância suficientemente boa para as aplicações desejadas, tanto na característica I-V quanto na de pequenos sinais.

Para fazer uma comparação entre um conjunto de dados gerado pelo SPICE e pela rotina QUADS da maneira mais simples e rápida possível, um programa especial para verificação de erro foi criado, sendo que os dados do SPICE foram obtidos a partir

de um arquivo ASCII contendo as informações relevantes, através da opção *PRINT* (sintaxe SPICE). No Gráfico 4, todos os testes usados na determinação do erro estão presentes, bem como a rotina usada para tal tarefa. Os gráficos extraídos são os seguintes:

- * Corrente de dreno x Tensão porta-fonte - Gráfico 4.1
- * Erro percentual na corrente de dreno - Gráfico 4.2
- * Corrente de dreno x Tensão dreno-fonte- Gráfico 4.3
- * Erro percentual na corrente de dreno - Gráfico 4.4
- * Transcondutância x Tensão porta-fonte - Gráfico 4.5
- * Erro percentual na transcondutância - Gráfico 4.6
- * Condutância x Tensão dreno-fonte - Gráfico 4.7
- * Erro percentual na condutância de saída - Gráfico 4.8

Por convenção, as curvas fornecidas pelo SPICE foram grafadas em linha cheia (c) e as do QUADS em linha tracejada (t). O circuito de teste, assim como seus parâmetros, é mostrado na figura 5.2.

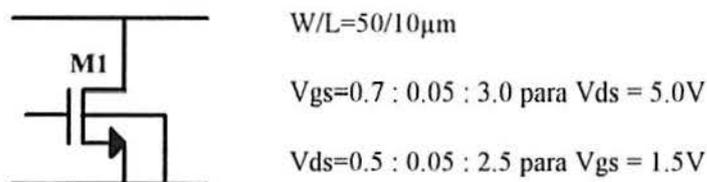


Figura 5.2

O primeiro teste consiste em levantar uma curva de corrente de dreno por tensão porta-fonte ($I_d \times V_{gs}$), mantendo V_{ds} fixo. Essa curva típica é composta de uma região de comportamento aproximadamente quadrático, seguida de outro, de caráter linear. Para permitir uma análise mais detalhada, o gráfico seguinte representa as diferenças percentuais entre estas duas curvas. V_{gs} foi variado de 0.7 até 3.0V, em passos de 0.05V, com $V_{ds} = 5V$. A curva de erro percentual baseia-se na equação clássica de erro, admitindo que o SPICE produz a medida correta:

$$E\% = \frac{I_{spice} - I_{quads}}{I_{spice}} \cdot 100$$

Equação 27

A curva de corrente de dreno por tensão de dreno-fonte para um V_{gs} fixo também foi extraída no SPICE e calculada pelo QUADS. V_{ds} foi variado de 0.5 até 2.5V, em passos de 0.05V. Para a análise dos parâmetros de pequenos sinais (condutância de saída e transcondutância), a característica de corrente foi numericamente derivada em relação a uma das tensões: V_{gs} ou V_{ds} . Para a transcondutância ($g_m = \partial I_d / \partial V_{gs}$), a relação se tornou: $g_m = \Delta I_d / \Delta V_{gs}$, da seguinte forma:

```
for j = 1:maxI(1)-1
gms(j) = (Is1(j+1) - Is1(j)) / (Vgs(j+1) - Vgs(j));
gmm(j) = (Im1(j+1) - Im1(j)) / (Vgs(j+1) - Vgs(j));
end
gms(maxI(1)) = gms(maxI(1)-1);
gmm(maxI(1)) = gmm(maxI(1)-1);
```

onde: $\max I(1)$ é o número de elementos a serem derivados,
 g_{ms} é a transcondutância SPICE,
 g_{mm} é a transcondutância QUADS,
 V_{gs} é o vetor de tensões porta-fonte,
 I_{s1} é a corrente SPICE para o primeiro teste e
 I_{m1} é a corrente QUADS para o mesmo teste.

Um ponto qualquer é derivado levando em consideração a tendência de variação do ponto imediatamente seguinte, de modo que essa operação só pode ser feita até o penúltimo ponto ($I_{s1}(\max I(1)-1)$). Para evitar problemas, o último ponto ($I_{s1}(\max I(1))$) é uma simples cópia do anterior. A análise da condutância de saída é feita de maneira semelhante, substituindo, porém, V_{gs} por V_{ds} .

Embora os dados usados tenham 10 dígitos de precisão em ponto flutuante (opção do SPICE), a análise do erro entre os parâmetros de pequenos sinais fornecidos pelo SPICE e por QUADS revela valores um pouco mais expressivos, causados basicamente por dois efeitos. Primeiro, como já foi dito, o esforço na implementação de um modelo é feito no sentido de reduzir ao máximo as discordâncias da característica I-V, deixando os parâmetros de pequenos sinais como consequência do sucesso da primeira operação. Além disso, como se sabe, a derivada é uma operação matemática com certo grau de instabilidade numérica, de modo que pequenos desvios nos valores absolutos da característica I-V podem se tornar erros expressivos na análise da derivada dessa curva.

De qualquer maneira, os gráficos mostram a eficiência da implementação do modelo e assim possibilitam que este seja usado em rotinas mais avançadas de síntese de blocos analógicos. Embora na síntese de um amplificador operacional os parâmetros de pequenos sinais tenham um papel muito mais relevante que as tensões e correntes DC, o erro máximo fornecido pelo modelo é inferior a 3%, geralmente da ordem de grandeza do erro introduzido por desvios de fabricação. Logo, conclui-se que o modelo empregado cumpre as exigências de confiabilidade e pode ser usado nas rotinas de síntese.

6 Síntese de amplificadores operacionais

6.1 Generalidades

O grau de automação existente na diversas etapas de projetos analógicos é muito menor que nos projetos digitais. Vários fatores causam essa situação, dentre eles o reduzido número de estados de saída dos circuitos digitais, sua modularidade e independência funcional, o grande número de blocos, permitindo que quaisquer funções sejam facilmente implementadas, etc. Dessa forma, um circuito digital complexo (pelo menos 20.000 transistores) nada mais é de que a interconexão de muitos blocos que desempenham funções específicas, cujas relações entre si são apenas sinais de tensão ou corrente em limites bem estabelecidos. Essa característica torna esse tipo de projeto extremamente adequado para automação, não só na etapa de projeto, mas também na de *lay-out*. Através de uma linguagem específica, a descrição funcional do circuito é feita e um *software*, usando uma biblioteca de funções, realiza a montagem do circuito a nível de transistor, com cálculo de tamanhos e polarização. Posteriormente, outra rotina se encarrega de fazer o *lay-out* minimizado desse projeto e verificar as regras de projeto. O projeto final é então concluído, prescindindo da atuação do projetista em qualquer das fases, a menos a de descrição do circuito.

No campo analógico, a maior complexidade dos projetos, aliada à sua enorme diversidade proporcionada por inúmeras combinações topológicas, torna difícil a criação de rotinas com funcionamento simplesmente mecânico. Essa dificuldade não é válida somente na fase de projeto, mas também na construção do *lay-out*, cujas ferramentas agora é que estão se tornando disponíveis no mercado. O estabelecimento das longas linhas de conexão, capacitâncias e resistências parasitas, que não afetam significativamente o desempenho de circuitos digitais, são de vital importância em projetos analógicos, pois podem significar a diferença entre um circuito funcionar e não funcionar. Por isso, até os passos mais determinísticos como a execução de *lay-out* necessitam especial cuidado quando o projeto tem cunho analógico. Por essas razões, o desenvolvimento de circuitos analógicos tem sofrido certo grau de defasagem em comparação a outros segmentos da eletrônica. Tanto isso é verdade que a construção de circuitos analógicos muito grandes (mais de 1.000 transistores) é absolutamente incomum, pois demandaria um tempo de projeto muito grande, inviabilizando assim o produto final.

Comumente, o projeto de um circuito analógico como o amplificador operacional se processa da seguinte maneira:

1) A escolha da topologia a ser empregada é o primeiro passo na direção de projetar um amplificador operacional. Essa escolha pode ser baseada em vários fatos, como as especificações que se deseja, a área aproximada que a topologia demanda, a simplicidade ou complexidade necessária para empregá-la, etc.

2) Usando como ponto de partida um equacionamento bastante simplificado, como o quadrático, por exemplo, o projetista faz um cálculo aproximado de tamanhos e polarizações dos transistores para a sua aplicação. Outra maneira comum de se começar um projeto analógico é inspirar-se em um projeto já feito que conste na literatura.

3) Segue-se então a parte de simulação, na qual o projetista transfere para um programa a topologia que ele empregou, assim como os dados que ele encontrou usando o seu equacionamento. Na maioria dos casos, essa etapa revela muitas surpresas, entre as quais, a incapacidade dos modelos simples de descreverem com precisão o comportamento do transistor e a ineficiência das técnicas simples de síntese. A diferença entre o desempenho esperado e o realmente obtido pode exceder 50%, especialmente para geometrias e correntes pequenas.

4) Para dirimir ao máximo estas diferenças, o projetista deve alterar alguns parâmetros do circuito, de modo a que o desempenho do circuito simulado aproxime-se do projetado. Essa alteração manual geralmente consome muito tempo, pois deve ser sempre seguida de uma outra simulação, que a validará. Em sistemas com alguma complexidade, onde os dados de um elemento devem se relacionar com outro, mudanças em dois pontos do circuito ao mesmo tempo às vezes se fazem necessárias. Essas mudanças, além de serem muito demoradas, geralmente fornecem respostas longe do que poderia se chamar de ótimas.

5) Depois de simular o circuito muitas vezes, alterando tamanhos e polarizações, o projetista consegue reproduzir, mesmo que parcialmente, o funcionamento do circuito desejado. Feito isso, a etapa de *lay-out* deve ser iniciada, na qual os dados que foram alterados com tanto trabalho devem ser adequados às imposições da tecnologia empregada. Determinar a influência dos elementos parasitas do *lay-out* no desempenho do circuito só é possível através de uma nova simulação, cujos resultados determinarão

se é necessário re-projetar o circuito, fazer um novo *lay-out* ou se o circuito desempenha suas funções satisfatoriamente.

Embora não seja uma técnica muito elegante, a totalidade dos circuitos analógicos produzidos dentro do Grupo de Microeletrônica da UFRGS foi feita dessa forma, prejudicando não só a qualidade dos circuitos, mas também a produtividade dos projetistas. Pensando nestas restrições e usando o modelo do transistor implementado em rotinas do MATLAB, foram criadas rotinas específicas de projeto de amplificadores operacionais, capazes de reduzir o tempo usualmente destinado a projetos e aumentar a qualidade das especificações que estes oferecem. Atualmente, a tecnologia de ponta na síntese de circuitos analógicos usa conceitos como arquitetura e hierarquia [24,25] para construir o melhor circuito capaz de responder a dadas solicitações. Inicialmente, um conjunto de especificações é dada pelo usuário e o programa de síntese se encarregará de montar um circuito, a partir de certas células básicas em um nível imediatamente acima do transistor. No caso em que o emprego de uma dada célula não permita que essa topologia forneça as especificações desejadas, essa será substituída por outra, até que se encontre a melhor solução. Essa técnica de topologia variável necessita de algoritmos de comparação e biblioteca de funções que tornam o *software* um sistema muito complexo. Mesmo assim, o desempenho do programa produz blocos de alta eficiência, com grande capacidade de adaptação e mudança.

Naturalmente, para as aplicações deste projeto, um sistema como esse seria muito bem-vindo, mas não compensaria de maneira alguma o esforço de sua criação e implementação. Por isso, programas muito mais simples, mas de funcionalidade marcante, foram implementados, no sentido de fazer o projeto de circuitos analógicos de forma mais automática. Para tanto, a utilização de uma topologia única dentro de um programa torna mais simples a criação do mesmo, embora este possa ser expandido para outras topologias sem mudanças estruturais significativas. A iteração é a técnica mais amplamente empregada nestas rotinas, pois elas permitem, de maneira relativamente fácil, que as soluções apareçam como resultados de cálculos repetidos, garantindo assim que um transistor ou bloco funcional esteja operando de forma adequada. A partir de determinados valores iniciais, o programa pode fazer alterações sistemáticas na busca de certo funcionamento. Técnicas de aceleração de convergência foram incluídas nos laços do programa, à medida do possível, para fazer com que este seja ainda mais rápido e proporcione um ganho de tempo ainda mais expressivo.

A qualidade e precisão dos projetos obtidos com essa ferramenta é função direta do modelo de transistor usado no programa. Para que novos modelos possam ser

usados, incluindo aqueles em inversão fraca e moderada, o programa foi estruturado de tal maneira que existe uma modularidade nas funções desempenhadas, do modo que a parte de modelamento está concentrada numa rotina externa, podendo ser facilmente alterada ou substituída. Uma outra limitação que as técnicas empregadas nas rotinas proporcionam é que os parâmetros de pequenos sinais são obtidos através de relação de diferenças e não diferenciação analítica. Este fator pode ser um gerador de erro, pois o intervalo usado para essa diferença (por exemplo $V_{gs1} - V_{gs0}$) às vezes não é tão pequeno quanto deveria, especialmente nos casos de declividade mais acentuada.

Infelizmente, em um mesmo projeto, não é possível incluir todas as especificações possíveis, sendo que algumas devem ser eleitas como críticas e usadas no equacionamento, sendo que as outras devem ser meras consequências das primeiras. Isso é devido ao fato de que o número de variáveis geralmente é superior ao número de equações disponíveis para serem resolvidas, de modo que frequentemente alguns valores devem ser fornecidos pelo usuário. Para ser mais preciso e severo na qualificação, as rotinas desenvolvidas não devem ser chamadas de automáticas, e sim de semi-automáticas, pelo fato de que necessitam de certo grau de interação com o projetista para funcionarem a contento. Através da alteração de alguns valores iniciais ou tamanhos pré-fixados, o projetista tem condição de conduzir o desenvolvimento do projeto para certa direção, de acordo com seu interesse. Por exemplo, as rotinas não trabalham com as duas dimensões: W e L . Isto se explica da seguinte maneira: quando é necessário diminuir a corrente em um dispositivo, duas alternativas são possíveis: diminuir W ou aumentar L , de modo que o fator de forma diminua. Entretanto, alterações na dimensão L causam variação direta na condutância do transistor, de modo que este fator tem uma sensibilidade bastante crítica quanto ao funcionamento do circuito (tanto ganho quanto banda passante). Para evitar instabilidades na convergência de uma subrotina que tem alteração de tamanho, essa é sempre feita na dimensão W , sendo que L é fornecida sempre pelo usuário neste sistema. O fornecimento da dimensão L de todos os transistores do circuito é uma tarefa que cabe somente ao usuário, sendo que estes valores tem uma considerável influência no comportamento, especialmente de pequenos sinais, do circuito. A escolha ótima destes valores não é uma tarefa simples, mas pode ser enormemente auxiliada pelo emprego das rotinas, que podem adiantar, com certo grau de precisão, o comportamento do circuito com aqueles valores. Logo, repetir o programa algumas vezes já dará ao usuário visão necessária sobre a importância e a influência deste parâmetro no desempenho geral do circuito.

Dependendo das especificações que se quer privilegiar, várias rotinas podem ser implementadas e diferentes valores iniciais devem ser fornecidos, de acordo com as

tensões de alimentação. Esses valores, por sua vez, também são configuráveis no sistema. Embora a grande maioria dos parâmetros do circuito sejam calculados pelo programa, alguns devem ser fornecidos pelo usuário, em qualquer configuração de sistema que se use. Esses valores se referem a constantes físicas e amplitude de alimentação, que também é variável, permitindo assim que se projete amplificadores para trabalhar em regime de $\pm 5V$ ou $\pm 1.5V$.

6.2 Aplicação para topologia Miller

Utilizar um modelo de transistor para projetar circuitos analógicos não significa somente polarizar adequadamente um transistor para que este conduza a corrente desejada. Sobretudo, a síntese de um amplificador operacional faz uso de parâmetros de pequenos sinais, como condutâncias e transcondutâncias. Para a eficiente utilização destes fatores, rotinas de adaptação devem ser criadas, no sentido de compatibilizar a natureza do modelo (intrinsecamente modelando a característica I-V) com as necessidades de projeto (pequenos sinais). A determinação da condutância de um transistor pode ser feita da seguinte forma:

$$Ic1 = quads(Vgs, Vds, 0, WL * L, L, I);$$

$$Ic2 = quads(Vgs, Vds + v, 0, WL * L, L, I);$$

$$gd6 = (Ic2 - Ic1) / v;$$

Para os dados completos de um transistor (tensões terminais, tamanho e tipo), a corrente de dreno é calculada. Posteriormente, aumenta-se a tensão Vds de um valor pequeno (v, no exemplo) e calcula-se novamente a corrente, através da rotina QUADS. Graficamente, o que acontece pode ser visualizado na figura 6.1, onde o comportamento de corrente do transistor em relação a tensão Vds apresenta pequenos aumentos na região de saturação, quando Vds aumenta. Esse aumento representa a própria condutância de saída do transistor, referenciada à variação de Vds, conforme o programa. Finalmente, a condutância pode ser calculada relacionando a diferença entre essas duas correntes calculadas e a variação de Vds entre as mesmas. Na região de saturação, esse valor costuma variar de 0.01 a $1\mu S$.

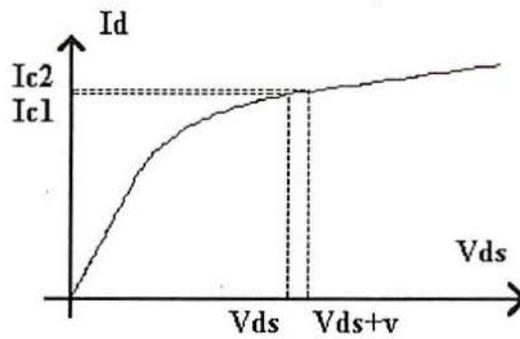


Figura 6.1

Analogamente, a transcondutância pode ser calculada da seguinte forma:

$$Ic1 = quads(Vgs, Vds, 0, WL * L, L, 1);$$

$$Ic2 = quads(Vgs+v, Vds, 0, WL * L, L, 1);$$

$$gp = (Ic2 - Ic1) / v;$$

O procedimento aqui é o mesmo, com a diferença que a tensão que sofre incremento é V_{gs} , de forma que o aumento de corrente é bem mais expressivo, conforme também pode ser visualizado em uma análise visual do gráfico da figura 6.1 ($I_d \times V_{gs}$). Como o incremento de corrente é maior que no caso anterior, os valores típicos de transcondutância na região de saturação variam entre 0.01 e 1mS.

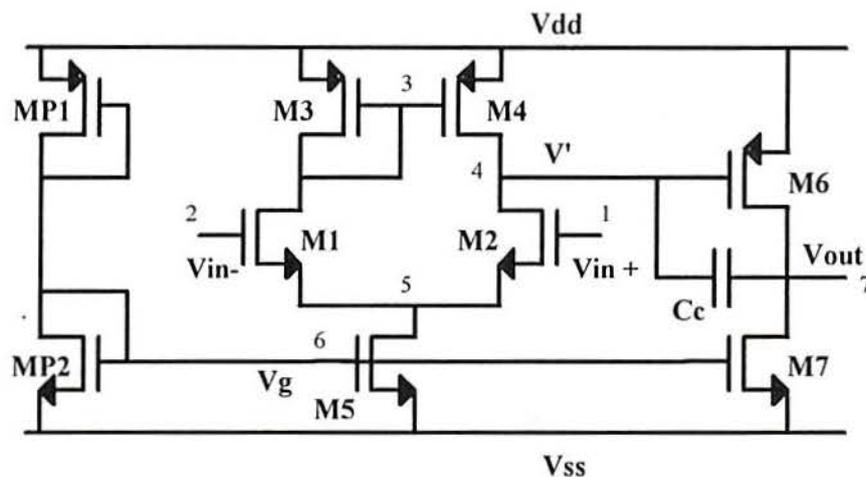


Figura 6.2

Para facilitar o processo de análise, o circuito Miller foi repetido na figura 6.2, incluindo-se também o número dos nós do circuito. A primeira implementação de uma rotina que objetiva a síntese semi-automática de um amplificador operacional na topologia Miller é funcionalmente descrita neste parágrafo. O programa PRG1.M, que faz uso da rotina que implementa o modelo do transistor (QUADS.M), destina-se ao

uso da tecnologia ES2 - 1.2 μ m. Os dados de entrada são as tensões de todos os pontos do circuito:

```
prg1(V3,V4,V5,V6,V7,I1,I2);
```

Através de cálculos relativamente simples, o programa permite que se calcule os tamanhos de todos os transistores do circuito, sofrendo porém de uma carência fundamental: a falta de inter-relação entre os parâmetros dos dispositivos. Uma vez que todas as tensões e correntes sobre o circuito são conhecidas, o cálculo do tamanho do transistor x pode ser feito através da seguinte rotina:

```
WLx = 1.0; Idx = 0; Ix = 10 $\mu$ A;
```

```
while abs( (Idx - Ix)/Ix ) > 1e-4  
  Idx = quads(Vgsx,Vdsx,Vbsx,WLx*Lx,Lx,tipox);  
  WLx = Ix * WLx / Idx;  
end
```

Embora não seja eficiente no cálculo do comportamento global do amplificador operacional, este método permite facilmente polarizar todos os transistores do circuito, com tensões e correntes definidas. Dessa maneira, um circuito mais simples, que não necessite de inter-relacionamento entre os parâmetros, pode ser projetado de maneira muito rápida através dessa técnica.

A despeito de suas qualidades, essa maneira de projetar circuitos prevê o que o projetista saiba, *a priori*, as tensões e as correntes em todos os nós do circuito. Esse conhecimento nem sempre é possível, ainda mais em circuitos mais complexos. Por isso, uma nova classe de rotinas foi implementada, que não necessita de tantas informações e, sobretudo, que é capaz de relacionar grandezas de acordo com as premissas de estabilidade mencionadas no Capítulo 5. A sintaxe deste novo tipo de rotina (PRG2.M) é:

```
prg2(Av2,Ib, Ic,Vout);
```

onde Av_2 é o ganho do segundo estágio, I_b é a corrente do primeiro estágio;
 I_c é a corrente de segundo estágio e V_{out} é a tensão DC de saída

1) O cálculo começa naturalmente pelo segundo estágio. Sabendo o valor dessa corrente, é simples determinar a relação de aspecto de M7, uma vez que sua polarização é fixa e conhecida.

2) O cálculo de M6 deve satisfazer duas exigências simultâneas: quando a corrente fornecida por M7 for igual à corrente quiescente, a tensão na saída (dreno de M6) deve ser igual à tensão V_{out} , solicitada pelo usuário. Em segundo lugar, esse estágio deve ter ganho igual a A_{v2} . Para calcular um transistor que apresente estas características, um laço iterativo deve ser utilizado, com o parâmetro V_{gs} e WL variando simultaneamente (WL representa a relação de aspecto, conforme já mencionado). Segue um segmento do programa que implementa essa função:

```

I6 = Ic;
gp = 1; gm6 = 0; Vgs6 = 2.0;

while abs( (I6 - Id7)/I6 ) > 1e-4
Id6 = quads(Vgs7,V7-Vss,0,WL7*L7,L7,0);
WL7 = I6 * WL7 / Id7;
end
Ic1 = quads(Vgs7,V7-Vss      ,0,WL7*L7,L7,0);
Ic2 = quads(Vgs7,V7-Vss+0.05,0,WL7*L7,L7,0);
gd7 = (Ic2 - Ic1) / 0.05;

while abs( (gm6 - gp)/ gm6 ) > 1e-4
Id6 = 0;
WL6 = 1.0;

while abs( (Id6 - I6)/I6 ) > 1e-4
Id6 = quads(Vgs6,Vdd-V7,0,WL6*L6,L6,1);
WL6 = I6 * WL6 / Id6;
end

Ic1 = quads(Vgs6,Vdd-V7      ,0,WL6*L6,L6,1);
Ic2 = quads(Vgs6,Vdd-V7+0.05,0,WL6*L6,L6,1);
gd6 = (Ic2 - Ic1) / 0.05;

Ia1 = quads(Vgs6,      Vdd-V7,0,WL6*L6, L6,1);
Ia2 = quads(Vgs6+0.05,Vdd-V7,0,WL6*L6, L6,1);

```

```

gp = (Ia2 - Ia1) / 0.05;
gm6 = Av2*(gd6+gd7);

delta6 = (gm6 - gp)/gm6 * 0.05;
Vgs6 = Vgs6 + delta6;
end
end

```

O programa funciona da seguinte maneira:

1) Inicialmente g_p (variável intermediária) é arbitrariamente igualada a 1, g_{m6} (transcondutância de M6) a zero e V_{gs6} (um valor inicial para a tensão porta-fonte de M6) a 2.0V. A corrente do segundo estágio, assim como a do primeiro, é fixada na corrente solicitada pelo usuário.

2) Como a corrente e a polarização são conhecidas (embora não mostradas no exemplo), M7 tem suas dimensões calculadas através do primeiro laço. Uma vez conhecida sua geometria, o cálculo da condutância de saída (g_{d7}) usa as variáveis intermediárias I_{c1} e I_{c2} .

2) Em seguida, inicia-se um laço que será iterativamente repetido, até que a diferença percentual entre g_p (transcondutância real de M6) e g_{m6} (transcondutância requerida para M6) seja suficientemente pequena (inferior a 0.01%).

3) Usando a informação de V_{gs6} disponível (valor inicial na primeira iteração), calcula-se o correto tamanho de M6 para que conduza a corrente desejada. Depois disto, avalia-se a condutância de saída de M6 (g_{d6}), assim como sua transcondutância (g_p). Para que o estágio tenha o ganho desejado (Av_2), a transcondutância de M6 deve ser tal que satisfaça a equação de ganho: $g_{m6} = Av_2 * (g_{d6} + g_{d7})$. Para isso, é calculado um desvio percentual da transcondutância real (g_p) e da desejada (g_{m6}). Este valor é ponderado através de um valor adequado (escolhido para permitir a convergência mais rápida) denominado *delta*.

4) Faz-se então uma correção no valor de V_{gs6} , para que a transcondutância mostrada por M6 permita que o ganho se aproxime mais de Av_2 . Esse cálculo é repetido até que as variações em V_{gs6} causarem ganhos muito próximos aos desejados. O fator de correção é indiscriminadamente somado a V_{gs6} , de modo que erros negativos diminuam a tensão na porta de M6, aumentando em caso contrário.

5) Uma vez conhecido o valor de V_{gs6} que causa o ganho desejado, este valor será usado para calcular os transistores restantes do circuito, começando pelo espelho de corrente. A tensão no ponto 4 (V_4) pode ser calculada da seguinte forma: $V_4 = V_{dd} - V_{gs6}$. Admitindo que, para condições DC $V_4 = V_3$, o cálculo de M_3 e M_4 se dá da seguinte forma:

$$I_3 = I_b/2;$$

$$WL_3 = 1.0;$$

$$I_{d3} = 0;$$

while abs((I_{d3}-I₃)/I₃) > 1e-4

*I_{d3} = quads(V_{dd}-V₃,V_{dd}-V₃,0,WL₃*L₃,L₃,1,z₃);*

*WL₃ = I₃*WL₃ / I_{d3};*

end

$$WL_4 = WL_3;$$

Os transistores M_3 e M_4 são, por definição, exatamente iguais, de modo que o cálculo de um é válido para o outro. Observe-se entretanto que a corrente circulante por cada transistor é metade da corrente disponível na fonte de corrente M_5 , conforme o funcionamento teórico do circuito. Nessa etapa do processo, M_5 ainda não pode ser calculado, uma vez que sua tensão de dreno (V_5) ainda não é conhecida. Entretanto, M_1 e M_2 podem ser calculados, usando com fator de ajuste exatamente a tensão V_5 . Supondo $V_{in+} = V_{in-} = 0$, $V_{gs1} = V_{in-} - V_5 = -V_5$. De igual modo, $V_{ds1} = V_3 - V_5$. De maneira análoga ao cálculo de M_6 , para determinar com precisão M_1 e M_2 duas exigências devem ser simultaneamente satisfeitas:

- a) A corrente em M_1 deve ser igual a de M_2 igual à metade de M_5 .
- b) A transcondutância de M_1 e M_2 deve ser pelo menos 10 vezes menor que a de M_6 , por razões de estabilidade (já discutido no Capítulo 5).

Uma vez que uma nova restrição entrou no equacionamento (a relação de transcondutâncias), o ganho do primeiro estágio não pode ser um parâmetro de entrada, sendo assim uma consequência do resto do equacionamento. Como a variação de mais de um parâmetro ao mesmo tempo exige um controle mais acurado no programa,

costuma-se variá-los individualmente. Segue a rotina que implementa a procura do melhor tamanho de M1 e M2:

```
I1 = Ib/2; Vgs1 = 1;
```

```
Ic3 = quads(Vdd - V3, Vdd - V3, Vbs1, WL3 * L3, L3, 1);
```

```
Ic4 = quads(Vdd - V3, Vdd - V3 + 0.05, Vbs1, WL3 * L3, L3, 1);
```

```
gd4 = (Ic4 - Ic3) / 0.05;
```

```
while abs( gm6/10 - gm1) / gm6/10 > 1e-4
```

```
  Vds1 = V3 + Vgs1;
```

```
  Vbs1 = Vdd + Vgs1;
```

```
  WL1 = 1.0;
```

```
  Id1 = 0;
```

```
    while abs( (I1 - Id1) / I1 ) > 1e-4
```

```
      Id1 = quad3(Vgs1, Vds1, Vbs1, WL1 * L1, L1, 0);
```

```
      WL1 = I1 * WL1 / Id1;
```

```
    end
```

```
Ic1 = quad3(Vgs1, Vds1, Vbs1, WL1 * L1, L1, 0);
```

```
Ic2 = quad3(Vgs1, Vds1 + 0.05, Vbs1, WL1 * L1, L1, 0);
```

```
gd2 = (Ic2 - Ic1) / 0.05;
```

```
Ic1 = quad3(Vgs1, Vds1, Vbs1, WL1 * L1, L1, 0);
```

```
Ic2 = quad3(Vgs1 + 0.05, Vds1, Vbs1, WL1 * L1, L1, 0);
```

```
gp = (Ic2 - Ic1) / 0.05;
```

```
delta1 = (gm6/10 - gm1) / gm6/10 * 0.05;
```

```
Vgs1 = Vgs1 + delta1;
```

```
end
```

O funcionamento do programa baseia-se em variar a tensão do ponto 5, através de variações em Vgs1, até que a transcondutância de M1 seja pelo menos 10 vezes menor que a de M6, da seguinte forma:

1) Como a corrente em M3 e seu tamanho já são conhecidos, a cálculo de sua condutância de saída pode ser feito facilmente. Este valor será útil quando se quiser calcular o ganho do primeiro estágio.

2) A seguir, entra-se em um laço, do qual só se pode sair se a transcondutância do primeiro estágio for 10 vezes menor que a do segundo.

3) A tensão V_{ds1} pode ser escrita sob a forma: $V_{ds1} = V_3 + V_{gs1}$, pois $V_{gs1} = -V_5$ quando $V_{in} = 0$. M1 tem seu tamanho calculado para uma corrente igual a $I_b/2$ e a partir destes dados calcula-se sua condutância de saída (g_{ds2}). Deve-se atentar, entretanto, que V_{bs1} não é zero e também não é igual a um valor fixo: V_{bs1} deve ser re-calculado a cada nova iteração, pois seu valor depende de V_5 , ou seja, $-V_{gs1}$.

4) A transcondutância de M1 (g_p) pode então ser calculada e caso não esteja dentro dos limites esperados, V_{gs1} será corrigido para minimizar este erro. *Delta* pode ser tanto positivo quanto negativo, permitindo assim que V_5 assuma qualquer tensão dentro de um intervalo especificado. Naturalmente, V_5 (e V_3) determina não só a excursão mas também o ponto quiescente da entrada, não podendo exceder $-V_{tn}$, caso onde a entrada só poderia excursionar positivamente.

Os segmentos de programa até agora detalhados, em conjunto com a rotina QUADS de modelamento, permite que um amplificador operacional seja projetado, de forma automática, embora pouco eficiente. O maior problema referente a essa técnica de projeto onde as tensões podem assumir praticamente quaisquer valores no circuito é que facilmente uma destas tensões sai do intervalo possível de funcionamento, comprometendo assim o resto do circuito. Simplesmente limitar os valores das tensões em alguns pontos não resolve o problema, uma vez que às vezes, para conseguir certa transcondutância ou condutância de saída as polarizações de porta necessárias devem ser altas. Este problema se torna mais importante e visível quando a tensão de alimentação é $\pm 2.5V$ e a excursão das tensões do circuito é consideravelmente pequena.

Para evitar esse tipo de problema, um outro de rotina foi proposto (PRG3.M), usando uma nova especificação: o *Slew Rate*. Conforme a equação 10, o *Slew Rate* (SR) relaciona a corrente do primeiro estágio com o capacitor de compensação. Uma vez que o último está fixado em razão da carga e para que o pólo não-dominante diste 2.2 vezes da frequência de ganho unitário, a corrente I_b pode ser determinada através de SR. Para que o cálculo seja possível, os parâmetros de entrada devem ser:

prg3(V3, SR, Av1);

A tensão V3 deve ser fornecida pois sem ela o cálculo de M1 e M2 tornar-se-ia muito complexo e computacionalmente dispendioso, além de poder incorrer no problema de encontrar soluções em valores de V3 fora da faixa útil. A sequência de cálculos é a seguinte:

1) Através do valor de SR fornecido, calcula-se a corrente que M5 deverá fornecer, para uma polarização fixa. De posse deste valor, segue-se o cálculo do tamanho de M5.

2) Como V3 é conhecido, a tensão Vds de M1 é também conhecida, de modo que se pode calcular seu tamanho para que tenha um ganho igual a Av1. Naturalmente, este processo se dá através de análise da transcondutância de M1, conforme a rotina:

```

Ib = SR / Cc;
Id3 = 0; WL3 = 1;
while abs( (Ib/2-Id3)/ib/2 ) > 1e-4
Id1 = quads(Vdd-V3,Vdd-V3,WL3*L3,L3,1);
WL3 = Ib/2*WL3 / Id3;
end

while abs( (gm1 - gp) / gm1 ) > 1e-4
Id1 = 0;
WL1 = 1.0;

while abs( (Ib/2-Id1)/Ib/2 ) > 1e-4
Id1 = quads(Vgs1,Vds1,WL1*L1,L1,0);
WL1 = Ib/2*WL1 / Id1;
end

Ic1 = quads(Vgs1,Vds1 ,WL1*L1,L1,0);
Ic2 = quads(Vgs1,Vds1+0.05,WL1*L1,L1,0);
gd2 = (Ic2 - Ic1) / 0.05;

gm1 = Av1 * (gd2 + gd4);
Ia1 = quad3(Vgs1 ,Vds1, WL1*L1, L1,0,z1);
Ia2 = quad3(Vgs1+0.05,Vds1, WL1*L1, L1,0,z1);
gp = (Ia2 - Ia1) / 0.05;

delta1 = abs( (gp - gm1)/gm1 ) * 0.05;

```

```

Vgs1 = Vgs1 + delta;
end

```

O objetivo principal dessa rotina é calcular uma polarização adequada para o par diferencial, de modo que este apresente o ganho desejado. Para isso, varia-se V_{gs1} , ou seja, $-V_5$. Inicialmente, o cálculo de I_b é feito, levando em conta o SR solicitado. Depois, M_3 e M_4 podem ser calculados, uma vez que V_3 é fornecido pelo usuário. O valor de V_3 tem duas implicações importantes: determina o fator de forma de M_3 e M_4 , tornando estes valores muito altos se a escolha não for bem feita; e estabelece o *overdrive* de M_6 , que determina a sua transcondutância. O laço principal da rotina inicia e é repetido até que a diferença percentual entre g_p (transcondutância real) e g_m (transcondutância desejada) seja inferior a 0.01%. Para o V_{gs1} dado, o tamanho de M_1 é calculado, assumindo que M_1 é idêntico a M_2 . Uma vez conhecendo a geometria de M_1 , sua condutância de saída (g_{d2}) pode ser calculada. Para que o estágio tenha o ganho desejado, a transcondutância de M_1 deve ser tal que satisfaça a equação de ganho. A transcondutância do transistor M_1 é então calculada e se seu valor não estiver dentro dos limites de erro aceitável, a tensão V_{gs1} será corrigida até que este desvio seja minimizado. Posto que M_3 tem seu tamanho determinado pela corrente I_b e por V_3 , o cálculo de sua condutância de saída não precisa entrar dentro do laço, o que permite uma economia de tempo na execução da rotina. Por outro lado, como o tamanho de M_1 está sendo constantemente alterado, cada nova iteração deve contar também com o cálculo de g_{ds2} e g_m .

3) Uma vez estabelecida a geometria de M_1 , o conhecimento da tensão V_5 permite o cálculo do tamanho do transistor M_5 , que deve gerar a corrente I_b .

4) No segundo estágio, os transistores M_6 e M_7 serão calculados de maneira a respeitar as restrições de estabilidade, usando para isso a variação do valor absoluto da corrente quiescente do segundo estágio. De acordo com o modelo quadrático, a equação que relaciona transcondutância com corrente e tamanho de um transistor é dada por:

$$g_m = \sqrt{K_P \cdot I_b \cdot \frac{W}{L}}$$

Equação 27

Para que a transcondutância de M_6 (g_{m6}) seja dez vezes maior que a de M_1 , a corrente do segundo estágio deve ser 10 vezes maior que a do primeiro e a relação de

aspecto de M6 deve ser 10 vezes maior que a de M1. Naturalmente, outras combinações podem levar a resultados similares. Neste caso, tal situação não é verdadeira, mas como as polarizações são fixas, nessa etapa do programa, espera-se que tanto a corrente de M6 quanto seu tamanho sejam bem maiores que os correspondentes do primeiro estágio. Existe porém a necessidade de fixar o limite superior para essas duas grandezas, para evitar que se projete circuitos com excesso de dissipação de potência ou tamanhos exageradamente grandes. Para isso, a análise dos dados fornecidos pelo programa é extremamente útil para poder separar os amplificadores potencialmente eficazes daqueles que são resultados puramente matemáticos.

Determinar o melhor valor de corrente para o segundo estágio é a tarefa principal do segmento de programa que segue, no qual faz-se um cálculo baseado no respeito da equação que relaciona as transcondutâncias de M1 e M6. Pelo fato de que se deve observar essa relação, o estágio não pode ter o valor de seu ganho DC programado pelo usuário, sendo este uma mera consequência do valor de transcondutância encontrado.

```

Ic = Ib;      % Valor inicial de corrente para segundo estágio

while abs( (10*gm1 - gp) / (10*gm1) ) > 1e-4
Ia1 = quads(Vdd-V3      ,Vdd-Vout, WL7*L7, L7,1);
Ia2 = quads(Vdd-V3+0.05,Vdd-Vout, WL7*L7, L7,1);
gp = (Ia2 - Ia1) / 0.05;

delta1 = (gp - gm1)/gm1 * 0.1;
Ic = (1+delta) * Ic;
end

Id7 = 0;
WL7 = 1.0;

while abs( (Id7-Ic)/Id7 ) > 1e-4
Id7 = quads(Vgs7, Vds7,WL7*L7,L7,0);
WL7 = Ic*WL7 / Id7;
end

Ic1 = quads(V6-Vss,Vout-Vss      ,WL1*L1,L1,0);
Ic2 = quads(V6-Vss,Vout-Vss+0.05,WL1*L1,L1,0);
gd7 = (Ic2 - Ic1) / 0.05;

```

$$Ic1 = quads(Vdd-V3, Vdd-Vout, WL6*L6, L6, 1);$$

$$Ic2 = quads(Vdd-V3, Vdd-Vout+0.05, WL6*L6, L6, 1);$$

$$gd6 = (Ic2 - Ic1) / 0.05;$$

$$Av2 = gp / (gd6+gd7);$$

Usando um valor inicial de corrente do segundo estágio igual ao do primeiro, o laço do programa que encontra o corrente é iterativamente executado até que o erro entre dez vezes a transcondutância do primeiro estágio e a real transcondutância do segundo seja inferior a 0.01%. De maneira similar às rotinas anteriores, a transcondutância é calculada e seu desvio em relação à transcondutância desejada é usado no cálculo de um fator de correção delta, que modificará o valor da corrente quiescente. Através de várias iterações, a corrente I_c será aumentada ou diminuída até que se encontre um valor de transcondutância satisfatório. Nesse ponto, a rotina deixa o laço de iteração e parte para o cálculo do tamanho e da condutância de saída de M7, para que o valor total do ganho possa ser definido.

Com isso, todos transistores da topologia tiveram seu tamanho calculado. É conveniente salientar que estes cálculos são feitos para condições DC, enquanto que a relação entre as transcondutâncias e a relação entre capacitor de carga e capacitor de compensação garantem o funcionamento do amplificador operacional como um sistema simples de primeira ordem.

6.3 Versão usada no *testchip*

A simplicidade das rotinas implementadas, aliada à grande variedade de combinações e técnicas de projeto proporcionado pelo atendimento de determinadas especificações conduz à criação de um grande número de algoritmos, que podem ser alterados de acordo com aplicações específicas. Estendendo ao máximo essa característica de modularidade, chega-se a possibilidade de adaptação das rotinas até agora empregadas no sentido de permitir que estas sejam empregadas na síntese de amplificador operacional em topologias alternativas, como Cascode. Uma vez que o estudo que culminou nessa dissertação de Mestrado também tem como objetivo a consolidação dos conhecimentos concernentes a circuitos analógicos integrados, a topologia Miller foi a única empregada neste trabalho a nível de síntese, em primeiro

lugar para que o seu funcionamento seja detalhado ao máximo e em segundo para facilitar a criação dessa classe de rotina. Este novo tipo de rotina que será proposto nessa seção foi usado como ferramenta de projeto em um circuito integrado de teste, executado dentro do Laboratório de Instrumentação. Todos os amplificadores operacionais que constam neste *testchip* foram projetados usando essa ferramenta, que, embora apresentando algumas falhas, mostrou-se eficiente na tarefa de projeto. Depois de projetados os amplificadores, foram feitos os respectivos *lay-out's* e os circuitos foram enviados para fabricação. Infelizmente, não houve retorno a tempo de permitir que resultados experimentais fossem tabulados neste trabalho.

A análise comparativa das rotinas apresentadas até agora e dos resultados por estas fornecidos elegeu como mais eficiente e fidedigna a quarta classe de programas de síntese semi-automática: PRG4.M. Essa nova técnica engloba mais especificações e dessa maneira pode conduzir de maneira mais acurada o processo de síntese. Como as outras rotinas, essa também deve ser executada algumas vezes até que o resultado final seja satisfatório. A sintaxe da chamada dessa função mostra a quantidade reduzida de dados de entrada:

prg4(SR,GBW,V3,V7);

O *Slew Rate* serve para determinar a corrente quiescente do primeiro estágio; GBW (*Gain x Band Width*) é o produto ganho banda passante e é empregado para o cálculo de gm_1 , conforme equação 8. Finalmente, V7 dá a tensão quiescente da saída, geralmente igual a zero ou um ponto fixo determinado pelo próximo estágio, como um buffer ou um estágio subsequente qualquer. Deve-se observar, no entanto, que o valor do produto GBW não deve ser confundido com a frequência de ganho unitário. Na verdade, $F_u = GBW * \sin(PM)$; mas se a margem de fase se aproximar de 90° , estes dois valores podem ser confundidos.

O funcionamento dessa classe de rotinas é basicamente o seguinte:

1) A transcondutância do primeiro estágio gm_1 é calculada da seguinte forma: $gm_1 = BW * C_c * 2 * \pi$. A tensão V3 permite que se calcule a transcondutância de M6 variando para isso a corrente do segundo estágio, de acordo com o segmento de programa:

$$gm_6 = 25 * gm_1;$$

$$g_p = 0;$$

$$V_{ds6} = V_{dd} - V_7;$$

$$V_{gs6} = V_{dd} - V_3;$$

```

while abs( (gm6 -gp)/gm6 ) > tol
  Id6 = 0;
  WL6 = 1.0;
  while abs( (I6-Id6)/I6 ) > tol
    Id6 = quads(Vgs6,Vds6,0,WL6*L6,L6,1);
    WL6 = I6*WL6 / Id6;
  end

  Ic1 = quads(Vgs6 ,Vds6,0,WL6*L6,L6,1);
  Ic2 = quads(Vgs6+0.05,Vds6,0,WL6*L6,L6,1);
  gp = (Ic2 - Ic1) / 0.05;
  delta = (gm6-gp)/gp;
  I6 = I6 + I6* delta;
end

```

Análises usando recursos de simulação provaram que os elevados tamanhos dos transistores causam capacitâncias parasitas muito grandes, influenciando sensivelmente a resposta do circuito. Dessa forma, uma das maneiras de garantir a estabilidade do circuito e uma margem de fase superior a 60° , é manter a relação entre a transcondutância do segundo para o primeiro estágio em 25 vezes. Inicia-se então um laço que busca o valor de corrente que cause a relação desejada entre as transcondutâncias em questão. Dessa forma, o conhecimento da corrente de M6 permite que se calcule a geometria de M7.

2) O cálculo de M3 e M4 é trivial, posto que V_3 é fornecido pelo usuário e como a corrente do primeiro estágio é fixada por SR. A transcondutância de M1 será tal que o produto ganho-banda se iguale ao solicitado, variando para isso a polarização V_{gs1} , conforme o programa:

```

I1 = SR*Cc / 2;
gp = I; gm1 = 0

while abs( (gm1-gp)/gm1 ) > tol
  Vds1 = V3 + Vgs1;

  Id1 = 0;

```

```

WL1 = 1.0;
while abs( (Id1-I1)/I1 ) > tol
    Id1 = quads(Vgs1,Vds1,0,WL1*L1,L1,0);
    WL1 = I1*WL1 / Id1;
end

Ic1 = quads(Vgs1 ,Vds1,0,WL1*L1,L1,0);
Ic2 = quads(Vgs1+0.05,Vds1,0,WL1*L1,L1,0);
gp = (Ic2 - Ic1) / 0.05;

delta1 = abs( (gp - gm1)/gm1 ) * .1;
Vgs1 = Vgs1 + delta1;

if Vgs1 <= Vtn
    Vgs1 = Vtn + 0.05;
end

end

Ic1 = quads(Vgs1,Vds1 ,0,WL1*L1,L1,0);
Ic2 = quads(Vgs1,Vds1+0.05,0,WL1*L1,L1,0);
gd2 = (Ic2 - Ic1) / 0.05

Avp = gm1/(gd2+gd4)

```

A corrente DC que deve circular em M1 é dada pela metade da corrente de M5, conforme já mencionado. Através de uma variação em V_{gs1} , o programa deve achar, para uma corrente fixa, a polarização que cause uma transcondutância g_p tal que o produto ganho banda solicitado seja respeitado.

3) Para cada novo valor de V_{gs1} , um novo tamanho $WL1$ deve ser calculado, que proporcionará uma nova transcondutância. O fator de correção δ_1 é ponderado através de uma constante: 0.1. Este valor foi escolhido por representar um bom compromisso entre velocidade de convergência e probabilidade de divergência. A tensão V_{gs1} , embora sendo corrigida pelo desvio de transcondutância, pode assumir valores abaixo do mínimo: tensão limiar, na qual a corrente é essencialmente zero no modelamento empregado. Para evitar que grandes desvios desloquem V_{gs1} para valores muito pequenos, uma cláusula é imposta no final do laço: se V_{gs1} for inferior à tensão

mínima, este será desprezado e um novo valor será calculado, com base na equação dada. Isto é feito porque uma vez que V_{gs1} for igualado a um valor inferior a V_{tn} , o programa é abortado, pois correntes nulas aparecem nos denominadores de algumas equações. Dessa maneira, esse mecanismo serve como um ajuste grosso e digital da convergência da iteração.

4) Como a equação que deve ser observada é a da relação de transcondutâncias, o ganho do estágio não pode ser programado pelo usuário, sendo dado pela fórmula abaixo:

$$A_v = \frac{g_{m1}}{g_{d2} + g_{d4}}$$

Ao contrário dos outros programas, neste nem o ganho do primeiro nem do segundo estágio podem ser especificados pelo usuário, cabendo a este introduzir outros parâmetros. Duas razões obrigam este funcionamento: primeiro porque cada estágio tem uma limitação concernente à quantidade de especificações que pode atender simultaneamente. Em segundo, porque os ganhos obtidos nessa e em outras rotinas geralmente são acima da faixa de 80dB, valor na maioria das vezes suficiente e aceitável. Logo, determinar o exato valor do ganho não é tão importante quanto especificar o valor da banda passante, por exemplo, desde que o ganho se mantenha em níveis satisfatórios.

O valor de V_3 desempenha um papel crucial neste circuito, sendo que o intervalo de valores funcionais é pequeno. A sensibilidade dos tamanhos de M_6 e M_3 em relação a V_3 é muito grande, sendo que a escolha ótima só pode ser feita através da repetição iterativa do programa, numa análise crítica na qual o usuário pode avaliar a sua relação no desempenho do circuito. Aumentar muito o valor de SR e GBW conduz a respostas impossíveis de implementação, seja pelo fator de forma muito elevado ou por excesso de corrente no segundo estágio. Na verdade, existe um compromisso que deve ser satisfeito entre o estado da arte das especificações e o custo de fabricação e até mesmo de potência consumida pelo circuito. Para isso, convencionou-se limitar a corrente do segundo estágio em $200\mu A$, valor já razoavelmente alto, mas necessário para permitir altos valores de transcondutância no segundo estágio. De igual maneira, a dimensão máxima dos transistores foi estipulada em $1000\mu m$ (com certa tolerância, para reduzir os custos de fabricação ao mesmo tempo em que se diminui a probabilidade de efeitos indesejados causado pelo particionamento na fase *lay-out*).

6.4 RESULTADOS DA SÍNTESE

O fato de ter negligenciado algumas capacitâncias importantes na fase de equacionamento, juntamente com o fato de que a extração dos parâmetros de pequenos sinais não é perfeita, introduz uma quantidade de erro no sistema, o que explica as discrepâncias entre as especificações solicitadas e as realmente encontradas na simulação. Rever o equacionamento no sentido de incluir capacitâncias parasitas é uma das maneiras de aproximar os dados da realidade e aumentar a precisão do sintetizador. Infelizmente, o custo dessa inclusão refletir-se-á na complexidade das rotinas, no momento em que as especificações mais importantes não serão função de um mas de vários parâmetros do circuito.

A utilização da classe de rotinas PRG4.M permitiu que quatro versões do amplificador operacional Miller fossem projetadas, com especificações ligeiramente diferentes, com o objetivo de sintetizar vários circuitos cujo desempenho pudesse ser comparado e analisado. Pelos dados disponíveis na literatura, não se pode requerer muito de um operacional com uma topologia simples como o Miller. De forma geral, as especificações mais genéricas são:

Ganho: 60 a 85dB

Produto Ganho Banda: 1 e 2MHz

Capacitor de Carga: 2 a 10pF

Capacitor de compensação: 2 a 10pF

Margem de fase: 60 a 80°

Slew Rate: 1 a 20V/μs

Deve-se observar, no entanto, que as melhores especificações são obtidas quando a tecnologia utilizada é analógica, de maneira que o capacitor de compensação pode ser construído com valores nominais maiores, entre duas camadas de poli-silício. Na verdade, os parâmetros fornecidos pelas *foundries* são garantidos para operação digital, sendo que circuitos analógicos não estão cobertos por qualquer tipo de garantia de funcionamento. Mesmo assim, dada a economia resultante do emprego de circuitos mistos, a tecnologia digital é largamente empregada na construção de blocos analógicos.

A essa altura já se pode perceber que a qualidade dos resultados fornecidos pelo programa é função dos dados de entrada que o usuário fornece. Uma das qualidades do sistema é exatamente a de permitir uma análise rápida entre a alteração de um parâmetro e sua repercussão no circuito. Este é o caso dos comprimentos L dos dispositivos do

circuito. O programa foi usado várias vezes até que se encontrasse a melhor relação de comprimentos, pois cada um destes desempenha uma tarefa específica. Por exemplo, os transistores envolvidos na parte de ganho (todos, à exceção de M5) devem ter suas condutâncias de saída minimizadas, de maneira a contribuir mais efetivamente para o ganho. Em contrapartida, o aumento da dimensão L que separa dreno e fonte aumenta o tempo médio que os portadores demoram para atravessar o canal, contribuindo negativamente para a resposta em frequência. Dessa forma, a dimensão L deve ser uma relação de compromisso entre ganho e banda, ou outro parâmetro do circuito. O transistor M5, que desempenha unicamente o papel de fonte de corrente, deve ter sua condutância minimizada, para diminuir o ganho de modo comum. M6 e M7 poderiam ter comprimentos maiores, mas como a dimensão W já é muito grande nestes dispositivos, um aumento em L comprometeria o tamanho do circuito. A determinação aproximada do melhor comprimento para cada transistor pode ser feita usando as próprias rotinas até agora mencionadas e através de variações em L pode-se determinar a sua influência nos principais aspectos do circuito. O extensivo uso deste programa permitiu que se chegasse aos seguintes valores:

$$L1 = L2 = 20\mu\text{m}$$

$$L3 = L4 = 10\mu\text{m}$$

$$L5 = 20\mu\text{m}$$

$$L6 = L7 = 10\mu\text{m}$$

Os transistores do par diferencial não têm $L = 10\mu\text{m}$ e sim $20\mu\text{m}$ devido ao fato de que capacitor de compensação age sobre a resposta em frequência muito mais agressivamente do que o aumento no comprimento, permitindo assim que se aumente L e conseqüentemente o ganho. Outro aspecto importante diz respeito à polarização das fontes de corrente, até agora não mencionadas. Admite-se para os dispositivos de polarização uma tensão $V_{gs} - V_t$ de 0.5V, em média, de acordo com as técnicas de síntese propostas em [16]. Nesta mesma linha de pensamento, esta tensão em dispositivos ligados à amplificação deveria estar em torno de 0.2V, mas como exatamente este parâmetro é usado para ajustar a transcondutância, esta heurística não é usada. Naturalmente não se usa uma fonte de tensão de valor V_g no circuito. O procedimento mais simples é implementar um divisor resistivo através de dois transistores em *diode-connection*, cujos tamanhos podem ser facilmente calculados de acordo com o seguinte procedimento:

* Estipular uma corrente de polarização que circulará pelos transistores para criar a tensão desejada. Esta corrente não tem restrição alguma quanto ao seu valor, uma vez

que características como resistência de entrada fazem do transistor CMOS um dispositivo com impedância praticamente infinita. Dessa forma, o circuito polarizado não carrega o circuito polarizador. Naturalmente, quanto mais alta a esta corrente, maior a potência dissipada, de modo que seu valor não deve ser grande. Por outro lado, sabe-se que para correntes pequenas, abaixo de 1 ou 2 μ A, os efeitos de troca de regime de inversão começam a aparecer, juntamente com os erros relativos ao modelamento impreciso neste região de transição. Logo, a determinação dessa corrente é também um compromisso entre precisão e potência, sendo que o valor 10 μ A foi empregado como ótimo para este circuito.

* Para que exista uma tensão V_g no ponto 6 (Figura 6.1), a queda de tensão drenofonte de MP1 (V_{dsp1}) deve ser igual a $V_{dd}-V_g$, enquanto que MP2 fica com o resto da tensão disponível: $V_{ss} + V_g$. Na configuração que os transistores MP1 e MP2 estão, $V_{gs} = V_{ds}$ e $V_{bs} = 0$, de maneira que só é necessário descobrir o tamanho de ambos, o que é possível através da seguinte rotina:

```

Ip1 = 0; Ip2 = 0; Ip = 10 $\mu$ A;
Vgsp1 = Vdd+Vg; Vdsp1 = Vgsp1;
Vgsp2 = Vss- Vg; Vdsp2 = Vgsp2;
Lp1 = 10 $\mu$ m; Lp2 = Lp1;

```

```

while abs( (Ip-Ip1)/Ip ) > 1e-4
Ip1 = quads(Vgsp1,Vdsp1,0,WLp1*Lp1,Lp1,1);
WLp1 = Ip*WLp1 / Ip1;
end

```

```

while abs( (Ip-Ip2)/Ip ) > 1e-4
Ip2 = quads(Vgsp2,Vdsp2,0,WLp2*Lp2,Lp2,0);
WLp2 = Ip*WLp2 / Ip2;
end

```

O cálculo é bastante simples e rápido, dispensando, portanto, maiores comentários. Entretanto, como será posteriormente evidenciado, MP2 tem uma relação de aspecto maior que MP1, contrariamente ao que se poderia pensar. A explicação é a seguinte: tomando como exemplo a alimentação de ± 2.5 V e uma tensão $V_6 = -1.3$ V, então $V_{gsp1} = 3.8$ V e sua corrente, 10 μ A. Analogamente, a tensão $V_{gsp2} = 1.2$ V e a corrente conduzida é essencialmente a mesma. Para permitir que transistores com diferentes *overdrives* conduzam a mesma corrente, tamanhos diferentes devem ser

empregados. Como MP1 tem um overdrive maior que MP2 e as correntes são iguais, MP2 tem um tamanho superior a MP1. Ao contrário do que poder-se-ia pensar (o maior transistor fica com a maior tensão V_{ds} , por analogia a resistores), o menor transistor fica com a maior parte da tensão disponível entre as fontes de alimentação. Este método de polarização é bastante simples e seu uso é intuitivo. Entretanto, sofre de alguns problemas sérios, como por exemplo a sensibilidade a variações das tensões de alimentação. Como a tensão V_6 é proporcional às alimentações, 10% de variação em V_{dd} ou V_{ss} incorrem em uma variação de V_g também de 10%. Naturalmente, uma variação dessa magnitude despolarizaria tanto o primeiro quanto o segundo estágio, criando assim a necessidade de utilizar circuitos de polarização com sensibilidade à variação de tensão de alimentação inferiores a 100%, como é o caso. Técnicas que empregam transistores em configuração de porta comum podem atenuar até 200 vezes as variações das tensões de alimentação, enquanto que usando tensão de referência do tipo *Band Gap* as variações tendem a ser ainda menores. Foram usados os seguintes dados na execução de PRG4.M.

$$SR = 10V/\mu s$$

$$GBW = 2MHz$$

$$gm6 / gm1 = 25$$

$$V3 = 1.0$$

$$V_{out} = 0$$

$$V_{dd} = -V_{ss} = 2.5V$$

e os dados resultados foram:

$$W1 = 20.58\mu m - L1 = 20\mu m$$

$$W2 = 20.58\mu m - L2 = 20\mu m$$

$$W3 = 59.94\mu m - L3 = 10\mu m$$

$$W4 = 59.94\mu m - L4 = 10\mu m$$

$$W5 = 81.05\mu m - L5 = 20\mu m$$

$$W6 = 872.08\mu m - L6 = 10\mu m$$

$$W7 = 281.40\mu m - L7 = 10\mu m$$

A implementação destes dados diretamente no simulador revela um problema emergente: o *off-set*. Devido à própria natureza amplificadora do circuito, aliada aos pequenos erros na característica I-V, a tensão quiescente da saída se desloca do seu ponto original. Entretanto, alterações manuais simples podem corrigir este off-set e levar

a tensão V_7 mais próxima à solicitada. Para implementar esta alteração, M6 é a primeira escolhida, devido ao alto valor de sua relação de aspecto e a pouca importância percentual que uma pequena variação possa provocar. Infelizmente, a alteração na relação de aspecto de M6 não varia de forma significativa o valor do *off-set*, necessitando dessa forma, modificações mais expressivas. Para evitar se mudar a característica AC do circuito por efeito de alteração de g_{m6} , o problema foi analisado de forma alternativa: a tensão na saída tem um valor de $-1.33V$ e isto representa um V_{ds} $1.33V$ maior que o solicitado. Para diminuir este V_{ds} extra, pode-se diminuir a polarização V_{gs} de M6, através de alterações no tamanho de M3 e M4, que não desempenham outro papel, a não ser de espelhar corrente e polarizar M6. Diminuir V_{gs6} significa aumentar a tensão de porta, para que o *overdrive* caia. Aumentar a tensão de porta de M6 significa diminuir V_{ds3} , mantendo a mesma corrente. Assim, reduzindo sucessivamente o tamanho de M3 e M4, pode-se polarizar adequadamente M6 e reduzir assim o *off-set*. Dependendo da necessidade, esta alteração para redução de *off-set* pode ser feita em mais de um dispositivo, de modo a não sobrecarregar nenhum em especial, evitando assim que suas características DC e AC sejam alteradas significativamente. Uma outra alternativa, que geralmente não compromete o funcionamento do circuito é alterar o tamanho de MP2, como um ajuste fino para a corrente do primeiro e segundo estágio. Deve-se salientar, entretanto, que este alto valor de *off-set* é causado por pequenas discrepâncias na tensão V_4 , que polariza o segundo estágio. Dessa forma, a diferença entre a tensão V_4 esperada e a real é amplificada através do segundo estágio, unindo aqui a sua própria contribuição negativa, culminando na tensão de saída polarizada inadequadamente.

O último aspecto a ser salientado é que não se tem liberdade para usar qualquer tamanho de transistor que se queira, por limitações construtivas. É fácil perceber que não se pode construir um transistor com dimensão de $10.2157\mu m$, pois o processo de fabricação não provê uma precisão suficiente para construir com segurança um dispositivo com esta geometria. Longe disso, o limite imposto pela ES2 - $1.2\mu m$ é $0.15\mu m$ (*snap*), ou seja, todos os dispositivos do circuito devem ter suas dimensões múltiplas inteiras de $0.15\mu m$. Esta restrição obriga que se refaça alguns cálculos, pois um comprimento que era $10\mu m$ passa agora a ser de $10.05\mu m$, o que não seria relevante, caso a dimensão W não fosse da mesma ordem de grandeza. Parece óbvio que este tipo de limitação deve ser levado em conta na fase de síntese automática, isto é, quando for solicitado ao projetista especificar os valores dos comprimentos, este já deverá fazê-lo respeitando as regras tecnológicas, evitando ao máximo posteriores manipulações e alterações manuais. De igual maneira, pode-se normalizar a saída de dados para as dimensões W de forma que estas também saiam formatadas como múltiplos inteiros de

0.15 μ m, embora esta técnica ainda não tenha sido implementada. Reunindo tudo que foi mencionado nos últimos parágrafos e inserindo também o tamanho dos transistores de polarização (MP1 e MP2), pode-se construir duas tabelas: uma com os tamanhos originais fornecidos pelo programa e outra com os tamanhos corrigidos para acerto de off-set e adequação às regras tecnológicas, nas quatro versões de operacional:

Versão	W1/L1	W3/L3	W5/L5	W6/L6	W7/L7	WP1/LP1	WP2/LP2
OP1	20.58/20	58.50/10	81.05/20	872.08/10	281.40/10	2.40/10	17.40/10
OP2	30.49/20	41.49/10	80.98/20	945.55/10	440.18/10	2.40/10	17.70/10
OP3	20.59/20	41.49/10	81.06/20	756.45/10	352.16/10	2.40/10	17.40/10
OP4	30.48/20	59.95/10	80.97/20	1090.1/10	351.73/10	2.40/10	17.40/10

Tabela 2

Versão	W1/L1	W3/L3	W5/L5	W6/L6	W7/L7	WP1/LP1	WP2/L2
OP1	20.55/	58.50/	81.00/	868.45/	281.40/	2.40/	17.40/
	19.95	10.05	19.95	10.05	10.05	10.05	10.05
OP2	30.45/	40.80/	81.00/	943.35/	440.10/	2.40/	17.70/
	19.95	10.05	19.95	10.05	10.05	10.05	10.05
OP3	20.70/	41.10/	81.15/	756.75/	352.05/	2.40/	18.60/
	19.95	10.05	19.95	10.05	10.05	10.05	10.05
OP4	30.45/	58.45/	81.00/	1090.5/	351.75/	2.40/	17.40/
	19.95	10.05	19.95	10.05	10.05	10.05	10.05

Tabela 3

A tabela 4 mostra os parâmetros mais importantes dos quatro amplificadores operacionais gerados a partir do programa PRG4.M, com os transistores otimizados para melhorar o desempenho.

Versão	Ganho	Fu	GBW*	PM	Cl	Off-set	Área	Potência	Slew-rate
	[dB]	[MHz]	[MHz]	[°]	[pF]	[μ V]	[μ m ²]	[mW]	[V/ μ s]
OP1	87.35	1.55	1.78	60.45	3	-731.52	15.366.72	0.95	7.78
OP2	88.34	1.80	2.07	60.0	3	185.877	17.756.66	1.35	7.52
OP3	86.97	1.51	1.74	60.1	4	-214.42	14.625.47	1.08	7.67
OP4	890.04	1.83	1.59	60.42	2	959.64	18,699.35	1.14	7.58

GBW* - Produto ganho banda real, obtido através de: $GBW = F_u * \sin(PM)$

Tabela 4

Conforme se pode verificar, as especificações não diferem muito entre si, mesmo porque não foi possível melhorar significativamente uma característica em especial, pois a topologia usada cria certo grau de encadeamento entre seus parâmetros. Para que mesmo assim fosse possível criar versões diferentes de amplificadores operacionais dentro da mesma topologia, os dados de entrada foram alterados levemente, de acordo com a tabela 5.

Versões	V3 [V]	Rg* []	SR [V/μs]	GBW [MHz]	Ganho' [dB]
OP1	1.0	25	10	2.0	90.5
OP2	0.9	25	10	2.5	91.1
OP3	0.9	25	10	2.0	89.4
OP4	1.0	25	10	2.5	92.0

Tabela 5

Rg* - Relação entre a transcondutância do segundo e primeiro estágio

Ganho' - Calculado pelo programa a partir dos outros parâmetros.

Através de simples simulações, é possível verificar que os amplificadores obtidos a partir das rotinas não têm as especificações esperadas. Este fato é atribuído ao equacionamento simplificado que negligencia elementos parasitas de significado expressivo, especialmente em altas frequências. Mesmo assim, a funcionalidade e operacionalidade do sistema de síntese é mantida, pois um erro percentual aproximadamente fixo nas especificações permite um projeto direcionado ao alcance de certas especificações com certa precisão. As capacitâncias intrínsecas do transistor, como Cdg (dreno-porta) e Cgs (porta-fonte) têm relação com a geometria do transistor e como a topologia empregada faz uso de dispositivos grandes, este aspecto toma proporções cada vez mais significativas. De forma clara, percebe-se que o erro existe, mas pelo fato de estar dentro de limites aceitáveis e ter um comportamento conhecido e explicável, o funcionamento das rotinas não sofre grandes perdas por efeitos destas não idealidades.

Conforme já foi dito, a tensão V3 tem um papel fundamental no funcionamento do amplificador Miller e seu intervalo de valores possíveis é pequeno. Assim, V3 foi

usado como um dos parâmetros de variação para obter versões distintas deste circuito. Alguns fatores que são afetados pelo valor quiescente de V3:

* A máxima entrada de modo comum é aquela que ainda mantém todos os transistores do circuito na região de saturação. Dessa forma, V3 deve ser o maior possível, de modo que:

$$V3 - V5 > V_{in} - V5 - V_t$$

$$V3 > V_{in} - V_{tn}$$

$$V_{in} < V3 - V_{tn}$$

Equação 28

Dessa forma, o fator limitante para a excursão positiva do sinal de entrada é V3 e, como geralmente se quer a maior entrada de modo comum possível, V3 deve ser colocado na tensão mais alta possível.

* A tensão V3, estando ligada diretamente à porta de M6, fixa a tensão Vgs6 e conseqüentemente o seu *overdrive*. Conforme já foi cansativamente mencionado, gm6 deve ter valores altos e, para que isso aconteça, altos valores de tamanhos e/ou Vgs6 são necessários, de modo que V3 não pode ser tão alto quanto se previa.

* Pelo fato de que a corrente do segundo estágio é alta, se Vgs6 for mantido em um valor baixo, será necessário construir um transistor com dimensões muito grandes para permitir a passagem de uma corrente elevada; logo V3 deve baixar.

* Se V3 é muito alto, ou seja, muito próximo de Vdd, Vgs3 fica muito reduzido e M3 também precisaria de uma geometria muito grande para conduzir a metade da corrente de M5, aumentando as capacitâncias parasitas relacionadas a um ponto de alta impedância. Logo a economia de área também deve ser relevada na escolha do valor de V3.

Respeitar todas estas características da melhor forma significa encontrar um intervalo de valores para V3 onde existe um compromisso ótimo entre as especificações do amplificador. Por isso, testes foram feitos usando vários valores para V3 e aqueles que mostraram um desempenho melhor estavam no intervalo [0.85, 1.1]V. Na execução de PRG4.M dois valores de V3 foram usados: 0.9 e 1.0V. Embora o produto ganho-banda que se usa como dado de entrada não coincida com a que se observa no

funcionamento do operacional, este parâmetro também foi objeto de variação. Conforme pode se observar na tabela 4, os amplificadores que foram projetados com uma especificação de produto ganho-banda igual a 2.5MHz tem uma frequência de ganho unitário inferior, provando assim que a influência de parasitas traduz-se em erro, mas de comportamento previsível e em concordância com o tamanho dos dispositivos empregados.

Combinando estes dois tipos de alteração (V3 e GBW), quatro versões foram projetadas, com tamanhos e características peculiares, conforme a tabela 5. O ganho DC previsto pelas rotinas tem um erro aproximadamente constante em torno de 3dB em comparação ao obtido no simulador SPICE. Este desvio é considerado plenamente aceitável, pois seu valor nominal é função direta da qualidade com a qual se modela a característica de pequenos sinais do transistor. Obviamente, pelo fato de que a análise DC despreza as capacitâncias parasitas envolvidas no circuito, era realmente de se esperar que o erro não fosse nulo. Esta discrepância entre valor previsto e realmente encontrado é um indício muito relevante na verificação de eficiência dessa técnica de síntese, pois a magnitude deste desvio permite que se deposite mais ou menos confiança no desempenho das rotinas.

Os valores do produto GBW, PM e Cl (capacitor de carga) foram extraídos da seguinte maneira. É bem frequente na literatura [14], usar um capacitor de compensação com valor igual ao capacitor de carga, naturalmente para tecnologias analógicas. As premissas de afastamento entre F_u e pólo não-dominante e F_u e zero não foram capazes de posicionar estes valores de modo a obter uma operação seguramente estável para a carga programada. Para permitir o uso sem o perigo de instabilidade (margem de fase igual ou superior a 60°), a carga capacitiva foi sistematicamente reduzida até que a margem de fase estivesse próxima a 60° . Dessa forma, os dados que constam na tabela 5 (GBW, PM e Cl) foram coletados para uma capacitância de carga que permitisse margem de fase de 60° . Novamente, a negligência de algumas capacitâncias parasitas convergiu para erros significativos em frequências mais elevadas, obrigando assim a reduzir o valor da capacitância de carga. Este procedimento não teve maiores repercussões, posto que amplificadores internos geralmente são conectados a estágios cuja capacitância geralmente não ultrapassa 1 ou 2pF.

A área mencionada na tabela 5 refere-se somente ao espaço físico que ocupariam os transistores do circuito, isto é, ela foi obtida simplesmente multiplicando W por L de todos os dispositivos. O capacitor não está presente neste cálculo, assim como a área necessária às interconexões dos transistores, de modo que o valor dessa área só tem

sentido quando comparado às outras áreas dos outros operacionais. Na etapa de execução de *lay-out*, percebeu-se que a área efetivamente necessária para construir um amplificador operacional é significativamente maior que a área calculada na tabela.

A potência do circuito é calculada em condições de ausência de sinal, isto é, refere-se somente à potência estática. Mas como os sinais de entrada geralmente são de pequeno valor de pico, a potência efetivamente dissipada quando o amplificador está em funcionamento é muito similar à potência estática. Finalmente, o *Slew Rate* mede a máxima variação da saída frente a uma variação na entrada, para um valor de tensão de entrada que não faça a corrente do estágio diferencial variar. Em outras palavras, quando as duas tensões diferenciais são iguais, a corrente de M5 é equitativamente distribuída para M1 e M2. Qualquer valor pequeno de desbalanço de tensão produz um *off-set* de corrente proporcional, que carrega o capacitor de compensação. Naturalmente este efeito se mantém para certos valores de tensão de entrada, pois se esta continua aumentando a curva tende a estabilizar-se num ponto onde toda a corrente de M1 passa por um dos transistores e o outro está cortado. Nesta situação, onde toda a corrente da fonte M5 está disponível para carregar C_c , mede-se o valor do *Slew Rate*. Como se vê, a máxima taxa de variação da saída em função de um estímulo na entrada é função do valor absoluto da entrada, e quando este valor é alto suficiente para bloquear um dos transistores, esta taxa recebe o nome de *Slew Rate*. A tabela 6 mostra os resultados das simulações de teste da velocidade de resposta do amplificador, para vários valores de degraus aplicados à entrada. Quando $V_{in} = 3V$, encontra-se a máxima taxa de variação na saída e aumentando ainda mais a excitação, não se observa posteriores incrementos nesse valor. Para valores inferiores, entretanto, a taxa de variação da saída em relação à entrada é inferior ao *Slew Rate* nominal.

V_{in}	s. <i>feedback</i>	Realim. unitária	Realim.ganho 100
3V	8.08 V/ μ s	7.64 V/ μ s	8.15 V/ μ s
3mV	16.75 V/ms	6.01 V/ μ s	31.26 V/ms
3μV	14.28 V/s	3.05 V/ μ s	15.72 V/s

Tabela 6

Um problema que pode ser evidenciado, mas que não deve ser relacionado ao sistema de síntese, é a sensibilidade do circuito frente a variações de temperatura. Para demonstrar mais claramente a relevância deste fato, o Gráfico 5 traz a análise térmica do

off-set de um dos amplificadores, na qual se pode comprovar a importância da temperatura no funcionamento destes blocos. A explicação para essas variações pode ser compreendida lembrando que os fenômenos que regem a geração e transporte de portadores dentro do dispositivo semiconductor são fortes funções da temperatura. Para evitar este tipo de relação indesejada, métodos como *Band Gap Reference* são usados. Este método consiste em criar circuitos cuja relação entrada-saída têm seus coeficientes de temperatura cancelados, restando unicamente o termo E_g (*Band Gap*), que representa a distância entre a banda de valência e a de condução do silício, sendo este valor usado como referência na produção de fontes de tensão dentro do circuito integrado. Embora tendo também uma certa variação com a temperatura, através da utilização destes métodos, é possível ter uma variação da ordem de miliVolt dentro da faixa de -25 a $+85^\circ$ C. Para evitar fugas da topologia sugerida e simplificar o processo de síntese, as técnicas de fontes de referência foram deixadas de lado, tornando o integrado particularmente susceptível a variação de temperatura.

Além de susceptibilidade a variações de temperatura, o circuito também apresenta sensibilidade quanto a variações das tensões de alimentação. Embora a performance do circuito se mantenha suficientemente estável com estas variações, sérios problemas advêm do tipo de circuito de polarização que está sendo usado. Usando a técnica de divisor resistivo (implementado através de transistores), a tensão V_6 é 100% dependente da tensão de alimentação e, dessa forma, quaisquer variações na fonte, ainda que de pequeno valor, podem polarizar ambos estágios fora da zona ideal, degradando o desempenho geral do operacional. Para evitar isso, técnicas de reflexão de corrente são mais eficientes, pois garantem que transistores próximos sejam construídos de maneira bem semelhante. Mesmo assim, o problema de gerar a primeira corrente ainda existe, sendo então resolvido utilizando estruturas em configuração *Cascode* e reduzindo as possíveis variações da fonte.

redução de tamanho de M6, um pequeno aumento na largura de M1, que tinha uma relação de aspecto perigosamente próxima à unidade. Costuma-se evitar o projeto de transistores com esta relação, porque nesses casos, quaisquer variações do tamanho efetivo das dimensões do dispositivo podem incorrer em erros percentuais altos, enquanto que se uma dimensão for bem superior à outra, as variações do processo tecnológico de fabricação não terão repercussões tão negativas. Dessa forma, na tentativa de otimizar uma topologia que não oferece muitos recursos de ganho e banda passante, essa inversão se mostrou eficaz e simples de implementar, sendo portanto adotada de imediato.

Naturalmente, quanto maior a geometria do transistor, melhor é o seu desempenho na análise de ruído e, com esta alteração, as reduções nas geometrias podem fazer o circuito mais sensível à esta característica. Por outro lado, a redução das dimensões diminui também as capacitâncias parasitas, sempre proporcionais à área do dispositivo e, com isso, vislumbra-se também uma melhor resposta em frequência. Na prática, esta alteração é implementada nas rotinas simplesmente através da alteração do parâmetro tipo (de 0 - NMOS para 1 - PMOS e vice-versa). Além disso, no cálculo de M1, quando a tensão V_{gs} é calculada, existe uma limitação de segurança para este valor, que não pode ser igual ou inferior à tensão limiar daquele tipo de dispositivo. Como o canal era NMOS, este valor era $V_{tn}(0,7)$, mas agora, como o circuito foi invertido, este valor passa a ser $V_{tp}(1,1)$. Para a tecnologia ES2, a relação entre as mobilidades dos dispositivos NMOS e PMOS vale 2.43, de modo que a redução no fator de forma pode ser da ordem de até 7 vezes.

7.2 Capacitor de compensação

A utilização da equação $gm1 = GBW * C_c * 2*\pi$ pode ser usada sem erros significativos, desde que se possa garantir que o comportamento do amplificador operacional será igual a um sistema de primeira ordem, isto é, com um pólo em frequências relativamente baixas e um comportamento de módulo com declividade estável até frequências superiores à F_u . Se o pólo e o zero estiverem suficientemente afastados, suas influências não serão perceptíveis e a equação pode ser usada sem problema. O mesmo não ocorre quando se calcula o valor do capacitor de compensação através da relação: $C_c = 0.22 * C_l$. Aqui, admite-se que a frequência do pólo não dominante é unicamente dependente da capacitância de carga, desprezando as capacitâncias parasitas proporcionadas por M1 e M6 (M7, M3 e M4 geralmente não

entram no cálculo porque seu tamanho é comparativamente menor). Esta suposição não é verdadeira e Fndp deve ser calculada da seguinte forma:

$$F_{ndp} = \frac{gm_6}{2 \cdot \pi \cdot \left[CL \cdot \left(1 + \frac{C1}{CL} \right) + C1 \right]}$$

$$C1 = \frac{2}{3} \cdot Cox \cdot W6 \cdot L6 + W1 \cdot xp \cdot Cjp$$

Equação 29

onde C1 representa a capacitância parasita do primeiro e do segundo estágio. Logo, o cálculo do capacitor de compensação (Cm) que faz com que a relação de transcondutâncias seja k e a frequência do pólo não dominante seja y vezes maior que a frequência de ganho unitário, pode ser derivado e calculado da seguinte forma:

$$C_c = \frac{1}{(2k)} \cdot \left(y \cdot CL - y \cdot C1 - \sqrt{y \cdot \sqrt{y \cdot CL^2 + 2y \cdot C1 \cdot CL + y \cdot C1^2 + 4k \cdot C1 \cdot CL}} \right)$$

$$C_c = \frac{y}{2k} \cdot \left[(CL + C1) + \sqrt{(CL + C1)^2 + 4 \cdot C1 \cdot CL \cdot \frac{k}{y}} \right]$$

Equação 30

Na verdade, negligenciando o efeito das capacitâncias parasitas e fazendo $gm_6 = 10 \cdot gm_1$, a equação 30 resulta na forma simplificada que era usada até então: $C_c = 0.22 \cdot C1$. Esta maneira de equacionar o problema das capacitâncias parasitas foi sugerida pelo Prof. Paul Jaspers, em conversações individuais em uma de suas visitas à esta Universidade, em agosto de 1995. A equação de Cm (capacitor de compensação) permite que se calcule o valor necessário de compensação num circuito onde o tamanho dos transistores é levado em conta, juntamente com o afastamento desejado entre o zero e o produto ganho banda e o pólo não-dominante e o mesmo produto. A resolução analítica dessa equação permite que cria dois novos parâmetros passíveis de alteração: k e y. O último tem influência quase que somente na margem de fase, não sendo, portanto, alterado. Por outro lado, k é um fator cujas alterações repercutem em todo o circuito, variando o tamanho de todos os transistores e o valor do capacitor de compensação. Plotando um gráfico de Cm por k, observa-se que quanto maior a relação entre as transcondutâncias, menor é o valor de Cm, através de uma característica exponencial. Por outro lado, aumentar k significa aumentar o tamanho de M6, a corrente do segundo estágio e conseqüentemente a potência consumida. Para tecnologias digitais, como a que

está sendo empregada, aumentar k significa economia de área pois a redução do valor do capacitor de compensação dá uma economia muito superior ao aumento das dimensões de M6 e M7. Dessa forma, pode-se levantar uma curva, como a do Gráfico 6.1, relacionando um incremento na razão k com um decréscimo da área total do circuito, resultante da diminuição do capacitor e do aumento de M6 e M7. Mais especificamente, pode-se constatar, através das curvas dos Gráficos 6.2 e 6.3, que a diminuição do tamanho dos transistores acontece simultaneamente a uma diminuição do valor do capacitor de compensação quando se aumenta a razão k , o que prova que o incremento deste parâmetro pode trazer uma economia grande de área com uma pequeno aumento de potência. Os adicionais graus de liberdade que se obtém a partir deste novo equacionamento permitem que vários amplificadores sejam projetados, facilitando que determinada característica seja favorecida. Espera-se que com este equacionamento mais preciso das capacitâncias do circuito, a previsão de resposta em frequência seja mais acurada que utilizando o método anterior ($C_c=0.22*C_l$). De fato, simulações de circuitos que foram projetados usando esta nova técnica mostraram um desempenho visivelmente superior, tanto no quesito área ocupada como produto ganho-banda.

Entretanto, de acordo com a equação 29 (definição de C_l), comprova-se que o cálculo do capacitor de compensação só pode ser feito depois que se tiver o valor de C_l , que é função do tamanho dos transistores mais importantes do circuito. Este fato causa uma impossibilidade de adequar diretamente as rotinas existentes a este novo método. Lembrando, as rotinas do tipo PRG4 calculam em primeiro lugar o capacitor de compensação e depois, usando o valor do *Slew Rate*, calculam a corrente do primeiro estágio, para depois determinar a relação de aspecto do par diferencial. Mas como agora o cálculo de C_m é função do tamanho dos transistores do par diferencial (que ainda não se conhece), uma nova técnica deve ser empregada. A solução para esta questão faz uso dos conhecimentos obtidos na área de rotinas iterativas e propõe que o valor da capacitor de compensação seja obtido através de uma série de cálculos iterativos com o mesmo formato das rotinas PRG4. Em outras palavras, esta nova classe de rotinas, dita PRG5, opera da seguinte forma:

* Inicialmente calcula-se um valor de capacitor de compensação através da equação simplificada $C_c=0.22*C_l$. Este valor, embora portador de um erro significativo, servirá como valor inicial de um processo iterativo que vem a seguir.

* A partir deste valor inicial de C_c , calcula-se a corrente e a transcondutância do primeiro estágio, a transcondutância do segundo estágio e assim por diante, de acordo com PRG4.

* Depois de calcular a relação de aspecto do par diferencial, já se tem informações suficientes para calcular um novo valor de $C1$ e Cc que, muito provavelmente, não é o correto, posto que a corrente do primeiro estágio foi calculada com o valor anterior de Cc .

* Novamente, repete-se o procedimento de cálculo ao estilo PRG4, com o novo valor de Cc ditando a corrente do primeiro estágio, até que o valor de Cc calculado no final da rotina seja igual ao usado para calcular a corrente no começo da rotina.

Ao final dessa convergência, tipicamente 7 ou 8 iterações, todos os valores calculados pelo programa podem ser usados e simulados. Para acelerar esse processo iterativo, tomou-se o cuidado de zerar as variáveis fora do laço de iteração, de modo que quando $Cc = 1.5\text{pF}$, por exemplo, o programa calcula $W/L1 = 12$ e quando Cc é corrigido para 1.4pF , o valor 12 é usado como aproximação inicial do fator de forma de $M1$, economizando-se portanto um tempo não desprezível em relação à técnica de atribuir indistintamente o valor unitário à $W/L1$.

Outro aspecto importante que sofreu alteração foi quanto às especificações de entrada: GBW e SR . Até então, estas variáveis eram consideradas independentes, podendo-se incorrer na possibilidade de inserir especificações contraditórias, ou que levassem o circuito a uma operação indevida. No capítulo seguinte será dada uma base teórica que facilitará a compreensão da relação GBW / SR , assim como as razões pelas quais este valor deve ser aumentado.

Naturalmente, a rotina PRG5 toma mais tempo do que qualquer uma das outras mencionadas até agora, mas seus resultados são expressivamente melhores. Além disso, o fato de ter mais variáveis passíveis de alteração permite maior flexibilidade no projeto, podendo-se dar prioridade a uma especificação em especial, como produto ganho-banda ou área ocupada. Este programa foi executado várias vezes, dando origem a uma quantidade grande de projetos com características bastante distintas, como banda passante, área ocupada, potência., etc. Estes projetos foram tabelados e os melhores foram escolhidos para que se permitisse uma comparação com os dados obtidos segundo a metodologia proposta por PRG4.

Uma outra característica que foi adicionada ao software é a geração automática de *netlist* formato SPICE. Ao invés de executar o programa e anotar os resultados de tamanhos e polarizações para posteriormente aplicá-los em um arquivo de dados de um

simulador, este processo foi agregado ao sistema, permitindo que todo o circuito seja transposto para um arquivo cujo nome o usuário pode definir. Este aspecto tem por finalidade tornar ainda menor o tempo necessário para o projeto do circuito, uma vez que a conferência do desempenho do bloco em questão também faz parte do processo e, dessa forma, é passível de otimização. Foi considerada também a hipótese de gerar um arquivo em formato RS, com uma parte ou a totalidade do *lay-out*, de maneira a reduzir o trabalho do desenhista de *lay-out*. Esta opção foi deixada para um trabalho futuro, pois é conveniente fazer primeiramente uma pesquisa no sentido de determinar se já não existe uma ferramenta que execute este trabalho, de modo que o programa de síntese devesse simplesmente gerar um arquivo compatível com esse eventual utilitário de *lay-out* automático.

7.3 RESULTADOS DAS ROTINAS APERFEIÇOADAS

Aplicando os conhecimentos desenvolvidos até agora, em conjunto com as alterações propostas em 9.1 e 9.2, a rotina PRG5 foi usada para projetar um conjunto de amplificadores operacionais cujas características mostram uma discrepância menor em comparação às especificações iniciais do que as propostas pelo método anterior. A descrição mais precisa das capacitâncias envolvidas é fator preponderante no aumento da precisão na previsão da resposta em frequência do circuito. Porém, é conveniente salientar que as duas únicas mudanças que foram feitas referem-se à inserção do tipo de dispositivo (NMOS para PMOS e vice-versa) e ao cálculo mais preciso do capacitor de compensação. Mesmo assim, várias especificações foram melhoradas, permitindo que um circuito com desempenho idêntico aos fabricados seja construído com menos área, menos potência e capaz de alimentar uma carga maior.

Um dos aspectos mais importantes que se pode observar nesta nova família de amplificadores é sua capacidade de alimentar cargas relativamente grandes (10pF). Conforme já foi dito, a carga capacitiva de um operacional Miller define a posição do pólo não-dominante e, dessa forma, a estabilidade do sistema. Quanto maior a carga, menor o valor do pólo e conseqüentemente menor a margem de fase. Nos operacionais obtidos com PRG4, a carga não podia passar de 5pF, sob pena de diminuir a margem de fase abaixo de 60°. Aqui, ao contrário, percebe-se que absolutamente todos os operacionais têm margem de fase superior a 60° com cargas capacitivas de 10pF, o que representa um caso de uma carga grande, em se tratando de um amplificador interno. Atribui-se esse fato ao cálculo melhorado do capacitor de compensação, que leva em conta as capacitâncias parasitas dos transistores de maior tamanho do circuito.

A inversão de entrada NMOS para PMOS permitiu não somente reduzir o tamanho de M6, mas também reduzir a potência do circuito, posto que a corrente do segundo estágio foi reduzida. Em PRG4 e PRG5, a corrente do segundo estágio é aumentada até que se obtenha a transcondutância desejada. Como agora M6 é um NMOS, que tem uma transcondutância intrinsecamente maior, esta corrente assume valores menores, tipicamente da ordem de $100\mu\text{A}$.

O parâmetro k , que relaciona as transcondutâncias do primeiro e segundo estágio, possibilita um ajuste interessante no circuito. Segundo a teoria, o seu valor mínimo é 10, de modo que o zero seja colocado uma década acima de F_u . Entretanto, usando $k = 10$, o valor do capacitor de compensação que é encontrado é grande demais para tecnologia digital (entre 3 e 5 pF). Portanto, k foi sendo gradativamente aumentado, verificando-se então que o capacitor de compensação assumia valores cada vez menores, como pode ser observado no Gráfico 6.3. Simultaneamente à redução do valor do capacitor de compensação, verifica-se um pequeno aumento na corrente total do circuito, posto que transcondutâncias maiores devem ser encontradas. Através do Gráfico 6.4, verifica-se que este acréscimo de corrente, embora existente, não é muito relevante.

Naturalmente, forçar que a transcondutância do segundo estágio seja 100 vezes maior que a do primeiro não é nada econômico, nem sob o ponto de vista de economia de área nem de potência; mas usar valores da ordem de 15 e 30 vezes representa um compromisso interessante entre o aumento de potência e tamanho de M6 e a redução do valor do capacitor de compensação, muito interessante em tecnologia digital.

A redução do valor dos comprimentos dos transistores usados também desempenha um papel importante na redução do valor do capacitor de compensação, embora determine também uma redução no ganho do amplificador, por efeito do aumento da condutância de saída dos dispositivos. A tensão V_6 , que polariza as fontes de corrente do primeiro e segundo estágio também foi alvo de mudanças, permitindo que se efetue uma redução no tamanho dos transistores M5 e M7, da seguinte forma: se I_1 é da mesma ordem de grandeza que I_2 , a tensão V_6 , que é comum à porta de M5 e M7 vai gerar relações de aspectos também da mesma ordem de grandeza. Porém de $I_2 \gg I_1$, $W/L_7 \gg W/L_5$, o que ocasiona um dispêndio desnecessário de área. Para equilibrar os tamanhos de M5 e M7 e evitar discrepâncias entre os dois transistores, V_6 pode ser usado como parâmetro de ajuste, sendo aumentado para diminuir o tamanho de M5 e M7. Contudo, se a corrente do segundo estágio for realmente muito maior que a do

primeiro, pode eventualmente acontecer que M7 tenha um tamanho considerável e M5 seja um transistor quadrado ou até com $L > W$. Embora tenha uma atuação perceptível no tamanho dos transistores de polarização, V6 não tem nenhuma influência no desempenho do circuito, sendo absolutamente independente das características de pequenos sinais. Outro aspecto interessante é que a dimensão L foi diminuída até valores próximos do mínimo ($1.2\mu\text{m}$) e este fato permitiu que se construísse um operacional em uma área mínima (somente contando os transistores) de $430\mu\text{m}^2$, com 73 dB de ganho e mais de 2.2MHz de produto ganho-banda. Neste caso em especial, cria-se uma situação interessante, pois somente $430\mu\text{m}^2$ são necessários para construir todo o amplificador, mas no mínimo $20.000\mu\text{m}^2$ são necessárias para fabricar o capacitor de 1.0pF, representando um custo muito grande, em tecnologia digital. Para entender como foi obtida a diminuição da dimensão L, deve-se lembrar como os circuitos estavam sendo calculados em PRG3 e PRG4.M. Usando estas técnicas, o usuário deveria entrar com duas variáveis: GBW e SR. Embora não aparente, estes parâmetros não são independentes e são função direta da razão g_m/I_d (proporcional ao ganho do amplificador). Os valores que estavam sendo usados anteriormente causavam um g_m/I_d muito baixo, que estava sendo compensado por altos valores de impedância de saída, ou seja, altos valores de L. Em PRG5, ao contrário, GBW e SR assumem valores mais interessantes, ou seja, valores que conduzem a uma relação g_m/I_d mais alta, mas não alta a ponto de entrar em inversão moderada. Como g_m/I_d aumentou, para manter o ganho na ordem de 80dB o comprimento dos transistores de amplificação não necessitava mais ser tão alto e este foi natural e paulatinamente reduzido, o que permitiu o projeto de circuitos consideravelmente mais econômicos no quesito área. Por último, a diminuição do *Slew Rate* foi muito útil na redução da corrente do primeiro estágio e com isso a redução da potência.

Vários projetos obtidos através dessa metodologia constam do Gráfico 7, no qual uma tabela com os parâmetros mais importantes foi montada, permitindo assim que o usuário escolha um operacional que se encaixe melhor à sua aplicação. Da mesma maneira, o projeto de novos circuitos pode ser dirigido de maneira conveniente, para que se encontre as especificações desejadas.

7.4 ALTERANDO PARÂMETROS TECNOLÓGICOS

Analisando a filosofia das técnicas de síntese apresentadas até agora, vemos que sempre houve a preocupação em fazer o modelo usado (Vladimirescu) o mais parecido

possível com o SPICE, embora saiba-se *a priori* que este não corresponde necessariamente à realidade. Mesmo assim, por motivos já citados, a implementação do modelo SPICE nível 2 foi feita no sentido de aproximar ao máximo as respostas fornecidas por estes dois métodos. A preocupação que surge a partir dessa metodologia é a seguinte: uma vez que se consegue obter um modelo de transistor com código fonte acessível, cuja concordância com o SPICE é suficientemente boa, é possível garantir que os circuitos construídos a partir dessa técnica funcionarão exatamente da maneira prevista? A resposta, infelizmente, é não, pois as discrepâncias existentes entre os resultados fornecidos pelo SPICE e os observados através de medições nem sempre podem ser desprezadas. Dessa forma, existe uma lacuna que deve ser preenchida entre o SPICE e a realidade.

A causa destas discrepâncias está intimamente ligada à formulação do modelo utilizado, assim como os métodos de extração de seus parâmetros. Numa comparação entre os dados fornecidos pelo SPICE e aqueles obtidos a partir de experimentos, percebe-se um erro que pode chegar a 40%, em alguns casos. Antes de concluir que os parâmetros fornecidos pela *foundry* não são de boa qualidade, deve-se lembrar que a ES2 - 1.2 μ m é uma tecnologia digital, isto é, o fabricante garante o seu funcionamento com segurança somente para aplicações digitais. O fato de usar transistores de forma analógica traz inconvenientes como falta de precisão no simulador, uma vez que os parâmetros foram extraídos para aplicações de outra natureza. Com isso, as rotinas de síntese, usando os parâmetros fornecidos pela *foundry*, efetuam uma cópia quase idêntica do comportamento do SPICE que, por sua vez, não é igual à realidade, pelos motivos mencionados. O erro causado por esta filosofia de síntese provavelmente não poderá ser negligenciado, mas só poderá ser quantificado quando os circuitos que foram enviados para fabricação retornarem.

Naturalmente, existe uma alternativa para solucionar ou pelo menos minimizar o problema. Admitindo-se uma situação ideal, na qual empregando um conjunto de parâmetros de tecnologia os dados fornecidos pelo SPICE sejam idênticos aos experimentais, pode-se garantir que os dados fornecidos pela rotina QUADS.M também o serão, com erro máximo de 3%, usando naturalmente os mesmos parâmetros. Neste caso, então, os amplificadores projetados com os programas sugeridos neste trabalho teriam desempenhos finais muito próximos àquelas especificações que serviram de ponto de partida na síntese. É claro que esta situação não existe, pois o SPICE erra consideravelmente na caracterização do transistor, especialmente em operações analógicas. Porém, utilizando rotinas de otimização de parâmetros e dados experimentais, pode-se fazer ajustes nesses parâmetros de forma a fazer com que os

resultados do SPICE se aproximem muito da realidade, reduzindo consideravelmente os erros. Estas rotinas manipulam adequadamente todos os parâmetros tecnológicos, de modo a encontrar um conjunto que propicie o menor erro quadrático em comparação a uma curva dita de referência (obtida a partir de medições). O método, implementado através de um rotina chamada SIMPLEX.M é relativamente demorado (20 minutos em um PC 486 DX2 66MHz), mas conduz a resultados relativamente precisos.

O objetivo final deste trabalho seria de obter um conjunto novo de parâmetros para uma tecnologia, no caso ES2 - 1.2 μ m, cuja aplicação no simulador SPICE e na rotina QUADS.M permita que se obtenha respostas muito mais próximas àquelas medidas experimentalmente. Com isso, seria possível aumentar consideravelmente a confiabilidade do sistema de síntese, pois sabe-se, *a priori*, que o modelo utilizado é muito semelhante à realidade. Na verdade, a utilização do simulador SPICE em conjunto com os parâmetros otimizados pela rotina SIMPLEX.M não foram capazes fazer as curvas experimentais iguais às curvas obtidas via simulação. Acredita-se que este fato deva-se à impossibilidade da rotina SIMPLEX de produzir um conjunto de parâmetros válidos em um intervalo grande operações. Sem esta garantia, o que se pode obter são dados que não podem ser diretamente aplicados, pois gerariam erros consideráveis.

7.5 MÉTODO ALTERNATIVO: gm/Id

Uma das características negativas das técnicas de síntese apresentadas até agora é que o sistema converge para uma solução matematicamente possível, mas não necessariamente ótima, em termos de ganho. Uma das maneiras de corrigir este erro é empregar a técnica de síntese apresentada nesta seção [26] que, embora não tenha sido ainda implementada em sua totalidade, permite ao usuário obter amplificadores nesta mesma topologia, com características muito interessantes. Para compreender esta abordagem, é necessário retroceder um pouco e voltar a um dos blocos básicos, vistos no Capítulo 5: o inversor. Este circuito está esquematizado na figura 7.2, onde a polarização é feita por uma fonte de corrente ideal.

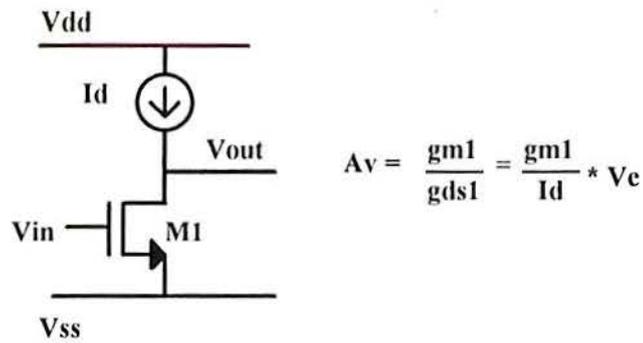


Figura 7.2

Naturalmente, aplicando a teoria da análise por pequenos sinais, como a impedância da fonte de corrente é infinita, o ganho do estágio é a relação da transcondutância pela condutância de saída de M1. Reescrevendo de outra maneira, g_{ds1} é igual a I_d / V_e , onde V_e é a tensão de Early. Este valor é obtido através do prolongamento da característica $I_d \times V_{ds}$ do transistor, na região de saturação. Este parâmetro é razoavelmente constante, para qualquer valor fixo de V_{gs} que se utilize. A compatibilidade dimensional da equação é mantida, uma vez que uma condutância foi substituída por uma relação de corrente por tensão. Dessa forma, considerando que o parâmetro V_e é fixo e dependente da tecnologia empregada, para um dado comprimento L , o ganho do circuito é diretamente proporcional à razão g_m/I_d . Dessa forma, maximizar o ganho do estágio significa maximizar g_m/I_d .

O comportamento de g_m/I_d é conhecido e está esquematizado na figura 7.3.

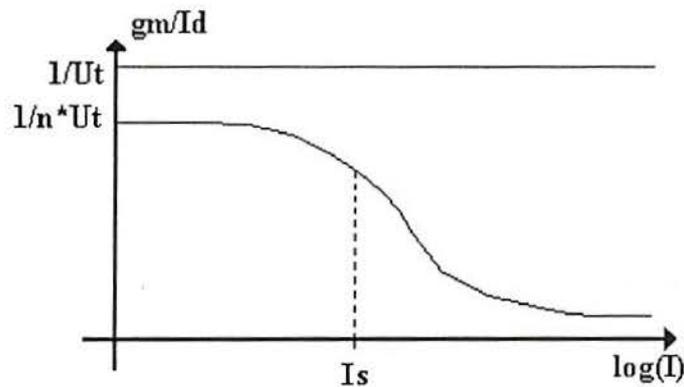


Figura 7.3

I_s representa a corrente de saturação, que divide a região de inversão forte da região de inversão fraca, sendo que a região intermediária é denominada inversão moderada. I representa a corrente I_d/β , onde $\beta = K_p * W/L$, levando em conta dessa forma a geometria do transistor. Todos os amplificadores projetados até agora operam

em inversão forte, de modo que o valor de gm/Id é relativamente baixo e podem ser localizados no gráfico na parte direita de I_s . Quando se deseja obter altos valores de ganho usando gm/Id de valor pequeno, a única alternativa é aumentar o valor da tensão de Early, aumentando o comprimento dos transistores, conforme foi feito. Quando se utiliza dispositivos bipolares, o valor máximo de gm/Id é dado pelo inverso da tensão térmica ($U_t = 25.9\text{mV}$), mas quando se utiliza dispositivos CMOS, este valor máximo é somente $1/n \cdot U_t$, onde n é um parâmetro tecnológico, de valor entre 1.3 e 1.5. Como o eixo horizontal é logarítmico, quanto mais à esquerda, menor o nível de corrente no dreno do dispositivo e assim, menor o seu consumo de potência. Naturalmente, poder-se-ia fazer a seguinte pergunta: porque não usar o valor máximo de gm/Id ? Em primeiro lugar porque este valor ocorre para correntes de dreno realmente pequenas, da ordem de centésimos e décimos de nano-ampères. Em segundo porque quanto mais se desloca à esquerda no gráfico, pior é a resposta em frequência do amplificador, embora o ganho seja muito alto. Dessa maneira, a região ótima de operação depende das especificações que o amplificador deve ter, especialmente ganho e área ocupada. Em inversão fraca, valores comuns de relação de aspecto podem chegar a 800 e até 1000, de forma que circuitos realmente grandes são projetados, muito embora o consumo de potência tenda a ser reduzir sensivelmente.

A curva real que relaciona gm/Id com I foi aproximada através de uma equação proposta por [27], sob a forma:

$$gm_{Id} = \frac{1}{n \cdot U_t} \cdot \frac{1 - e^{-\sqrt{\frac{I}{I_s}}}}{\sqrt{\frac{I}{I_s}}}$$

Equação 31

Esta equação não é analiticamente inversível, ou seja, não se pode encontrar uma forma fechada para descrever a corrente I em função de gm/Id , necessitando-se dessa maneira recorrer a métodos numéricos. Dessa forma, para um determinado valor de gm/Id , pode-se calcular I , que é diretamente proporcional ao tamanho. Esta metodologia traz inúmeros benefícios em especial devido à sua simplicidade, pois prescinde da utilização de um modelo completo que descreva o comportamento do transistor, sendo somente necessária uma equação que relaciona o parâmetro gm/Id e I_d/β . Abaixo segue um exemplo de uma rotina capaz de determinar o valor de I através de uma inversão numérica da equação dada:

```
function Y = conver(gmId)
```

```
n = 1.3; limi = 0.000001; lims = 1; Ut = 0.02585215;
```

```
Is=2*n*Ut*Ut; cnt = 1; F1 = 1;
```

```
if gmId > 1/(n*Ut)
```

```
    sprintf('O valor passado a funcao (%f) eh superior a %f e assim a funcao nao
```

```
    converge',gmId,1/(n*Ut))
```

```
    pause
```

```
    return
```

```
end
```

```
while abs(F1) > 1e-5
```

```
    I = (limi + lims) / 2;
```

```
    F1 = 1/(n*Ut) * (1-exp(-sqrt(I/Is))) / sqrt(I/Is) - gmId;
```

```
    if F1 > 0;
```

```
        limi = I;
```

```
    end
```

```
    if F1 < 0;
```

```
        lims = I;
```

```
    end
```

```
    I = (lims + limi) / 2;
```

```
    cnt = cnt + 1;
```

```
end
```

```
Y = I;
```

O funcionamento é muito simples:

* Primeiro, o programa recebe o parâmetro gmId e o avalia para determinar se ele é menor que o valor máximo que a função pode assumir. Em caso negativo, o sistema indica a impossibilidade de achar a raiz e é automaticamente abortado.

* A função $F1$ é a igual à equação 31, com a diferença de passar $gmId$ para o outro lado da igualdade, de forma que quando for encontrado o valor de I que faz $F1 = 0$, este a é a raiz da equação.

* O valor exato da raiz é encontrado dentro de um intervalo inferior ($limi$) e superior ($lims$). Quando o resultado da função é positivo, significa que o limite inferior deve ser aumentado e quando este é negativo, o limite superior é que deve ser diminuído.

* Ao final de cada iteração, o valor da variável I é o ponto médio do intervalo em questão, sendo que este intervalo fica menor a cada iteração. Em geral, menos de 15 iterações são necessárias para encontrar a solução. Na verdade, outros métodos de inversão podem proporcionar convergências mais rápidas, mas como esta não é o ponto crítico da rotina, este aspecto foi deixado de lado.

Dessa forma, um valor qualquer de gm/Id representa um e somente um valor de I , que pode ser encontrado usando a rotina `CONVER.M`.

É óbvio que, quando se usa esta metodologia para projetar amplificadores operacionais, as mesmas premissas de separação de zero e frequência de corte e frequência do pólo não-dominante e frequência de corte devem ser observadas, sob pena de incorrerem problemas de estabilidade. Na verdade, pelo fato de se utilizar este método mais direto e analítico, a inserção destas restrições é bem mais simples, facilitando a construção de um programa de síntese.

A maior qualidade deste método é permitir ao usuário, que deve ter certo conhecimento de projeto de circuitos analógicos em inversão fraca e moderada, decidir em qual região de operação cada transistor do circuito vai estar. Com isso, é possível melhorar ainda mais as especificações fornecidas pela topologia Miller, como se pode comprovar com o uso de um simulador elétrico que contenha um modelo válido em todas as regiões de operação. No exemplo que será dado, o usuário deve informar o valor de gm/Id de $M1$ e $M6$, sendo o ganho também uma consequência dos valores escolhidos. Naturalmente, pode ser que os valores fornecidos pelo usuário não sejam os melhores, de modo que o sistema de síntese deveria ser inteligente o suficiente para se deslocar horizontalmente no gráfico da figura 7.3, à busca da melhor solução. Esta última parte ainda não foi implementada, pois o algoritmo precisa de certa quantidade de inteligência no controle das variáveis, sendo esse um trabalho futuro.

O programa abaixo propõe a síntese do amplificador Miller para um dado valor de produto ganho banda. Neste caso, cabe ao usuário definir a região na qual ele quer trabalhar, isto é, os transistores que compõem o circuito, embora estejam dentro do mesmo encapsulamento, não necessitam operar sob o mesmo regime de inversão. Dessa forma, o par diferencial pode estar operando sob inversão fraca, para dar alto ganho e o espelho e as fontes de corrente sob inversão forte, posto que a probabilidade de erros e desalinhamentos é menor em inversão forte. Segue a implementação dessa técnica.

```
function Z = miller(gmld1, gmld2, gmld4, w)
wt = w * 2 * pi; Cm = 3*10^(-12); C2 = 10*10^(-12); Cox= 0.8*10^(-3);
Cj = 0.15*10^(-3);
mun = 0.10625; mup = 3.794*10^(-2); xp = 6 *10^(-6);
L1 = 6*10^(-6); L2 = 3*10^(-6); L = 3*10^(-6);

I1= conver(gmld1); I2 = conver(gmld2); I4 = conver(gmld4);
VAN = 10; VAP = 15;
V1n = (L1/L) * VAN; V1p = 2*(L1/L) * VAP; VA1 = (V1n * V1p) / (V1n + V1p);
V2n = (L2 / L) * VAN; V2p = (L2/L) * VAP; VA2 = (V2n * V2p) / (V2n + V2p);
A1 = 20 * log10(gmld1 .* VA1); A2 = 20 * log10(gmld2 * VA2);
A = A1 + A2;

while num2 < 10
num2 = num2 + 1;
gm1 = Cm * wt;
Id1 = gm1 ./ gmld1;
beta1 = Id1 ./ I1;
wl1 = beta1 ./ (mup .* Cox);
w1 = wl1 .* L1;

gm2 = 10 * gm1;
Id2 = gm2 ./ gmld2;
beta2 = Id2 ./ I2;
wl2 = beta2 ./ (mun * Cox);
w2 = wl2 .* L2;
C1 = (Cox*w2*L2)^2/3 + w1*xp*Cj;
Cm = (C1+C2 + sqrt((C1+C2).^2 + 2*C1*C2/0.11)) .* 0.11;
end
```

$$w3 = Id1 ./ (mun * Cox * I2);$$

$$w3 = w3 .* L1;$$

$$w4 = Id2 ./ (mup * Cox * I4);$$

$$w4 = w4 .* L2;$$

$$w5 = 2 * Id1 ./ (mup * Cox * I4);$$

$$w5 = w5 .* L1;$$

O programa também tem caráter iterativo, uma vez que, para calcular o valor exato do capacitor de compensação, é necessário saber o tamanho dos transistores do par diferencial (M1) e do segundo estágio (M6). Os passos são:

* Passar ao programa o valor de gm/Id do par diferencial ($gmId1$), do transistor M6 ($gmId2$) e de Mx ($gmId4$), além do valor do produto ganho banda passante (w).

* C_m é o valor inicial do capacitor de compensação e $C2$ refere-se à carga do amplificador. Cox , Cj , mun , mup e xp são parâmetros tecnológicos. Neste programa também o cálculo dos tamanhos baseia-se no cálculo da relação de aspecto dos transistores, posto que o comprimento é fornecido pelo usuário.

* Usando a rotina *CONVER.M*, encontra-se o valor de I referente a cada um dos valores $gmId$.

* V_{an} e V_{ap} representam a tensão de Early para os dispositivos NMOS e PMOS, mas este valor refere-se ao comprimento mínimo da tecnologia, devendo portanto ser corrigido quando o comprimento usado não for este. V_{A1} e V_{A2} representam a associação paralela entre V_{1n} e V_{1p} e V_{2n} e V_{2p} , indicando que as condutâncias dos dois transistores de um inversor são igualmente importantes na determinação do ganho.

* Conforme foi dito, o ganho pode ser calculado simplesmente através do parâmetro gm/Id e a tensão de Early corrigida para o tamanho desejado. Dessa forma, pode-se calcular diretamente o ganho do primeiro e segundo estágios em dB.

* O laço de iteração é repetido 10 vezes, mas poderia eventualmente ser-lo menos, desde que a variação de C_m fosse suficientemente pequena.

Inicialmente, a transcondutância de M1 é calculada com base no produto ganho-banda e no capacitor de compensação. Uma vez que se conhece o valor de gm_1 e de gm/Id_1 , pode-se calcular facilmente o efetivo valor da corrente de dreno do par diferencial. A partir daí, como já se sabe o valor de I_1 e Id_1 , o parâmetro β_1 pode ser calculado e em decorrência deste, a relação de aspecto wl_1 . Como o valor do comprimento é fornecido pelo usuário, encontra-se o valor de W_1 .

* Usando a relação mínima de 10 para as transcondutâncias do primeiro e segundo estágio, obtém-se o valor de gm_2 (transcondutância de M6) e analogamente pode-se determinar sua corrente de dreno. Da mesma forma calcula-se β_2 e W_2 .

* C_1 representa a soma das capacitâncias parasitas criadas pelo efeito de M1 e M6. Usando a mesma equação usada em PRG5, para $k = 10$ e $y = 2.2$, o capacitor de compensação C_m é calculado.

* Os mesmos passos são repetidos até que a variação de C_m possa ser negligenciada. Então, calcula-se o tamanhos dos transistores passivos, da seguinte forma:

a) Q3 (a e b) são calculados usando a corrente Id_1 (corrente efetiva de dreno) e I_2 , obtida a partir da inversão de gm/Id_2 . Dessa forma, obtém-se um par de transistores operando na mesma zona de inversão que Q2. Admite-se também que a largura de Q3 será idêntica a de Q1.

b) Q4 é calculado de maneira mais direta, uma vez que a sua corrente é conhecida e o valor de gm/Id para este transistor também o é. Assim, obtém-se diretamente o valor do tamanho, admitindo, porém, uma largura igual a de Q2, para maior simetria.

c) Q5, por sua vez, é calculado com os seguintes dados: corrente igual a duas vezes o valor de Id_1 e $gm/Id_5 = gm/Id_4 = gm/Id_2$. Isto significa que geralmente Q1, Q2 e Q3 operam em inversão fraca ou moderada e Q4 e Q5 em inversão forte.

Esta técnica provou ser eficiente no projeto de amplificadores operacionais, permitindo que se defina, *a priori*, a região de inversão qual cada transistor vai atuar. Em

contrapartida, o inconveniente do método reside no fato de que o usuário **deve** saber o significado dos termos inversão fraca, transcondutância, gm/Id , etc., exigindo dessa forma um conhecimento prévio que um projetista novato geralmente não tem. Para resolver esta questão propõe-se a seguinte metodologia:

- 1) Como condição inicial, todos os transistores têm um valor de gm/Id que conduza a inversão moderada.
- 2) A rotina de cálculo sugerida é implementada na íntegra e os dados de W e L são anotados.
- 3) A partir de uma determinada técnica de controle de variáveis, o valor de gm/Id de $Q1$ e $Q2$ seriam sistematicamente alterados, usando as especificações do circuito como parâmetros. Assim, para obter maior ganho, $gm/Id1$ deve aumentar, mas quando isso acontece, a resposta em frequência fica prejudicada. Logo, deve existir certa inteligência no programa de modo que seja encontrado um compromisso ótimo entre as especificações de maior interesse.

Esta filosofia de projeto não foi implementada neste trabalho, nem tampouco foi encontrada em nenhuma literatura do ramo. Embora aparentemente conduza a uma resposta ótima, especialmente se as rotinas de controle multi-variáveis utilizarem algum critério de minimização, o algoritmo acima peca pela complexidade de implementação e por isso a discussão pode ficar somente à nível qualitativo.

8 LAY-OUT DE CIRCUITOS ANALÓGICOS

8.1 DESCRIÇÃO DOS CIRCUITOS

A etapa subsequente no projeto de circuitos integrados de qualquer natureza, analógico ou digital, é o *lay-out*. Aqui, um desenho hierárquico da estrutura é feito, retratando da maneira mais fiel possível a composição dos transistores, assim como sua interconexão no circuito. Com o uso de um *software* dedicado, é possível montar um esquema composto de várias peças, representando as camadas existentes no processo de fabricação de dispositivos semicondutores. Na execução deste projeto, o programa empregado com esta finalidade foi o SELA [28], um aplicativo para PC desenvolvido por um grupo de pesquisadores dentro do Grupo de Microeletrônica da UFRGS. Na etapa de ligação e colocação, isto é, quando os projetos de amplificadores operacionais já estão prontos e é necessário conectá-los adequadamente às fontes de tensão e aos *pads* de entrada e saída, o *software* usado foi o EDILEX, para uso em *workstation*. Ambas ferramentas têm uma característica que as diferenciam dos aplicativos de *lay-out* digital: embora possuindo interface gráfica que auxilia o projetista a transpor para uma tela o seu circuito, estes programas não tem nenhuma característica automática, sendo que absolutamente todos os passos devem ser manualmente executados pelo projetista. A ausência de criação e ligação automática de transistores, presente em ferramentas para projetos digitais, torna o processo de *lay-out* demorado, passível de erro e geralmente não otimizado da melhor maneira possível. Os arquivos gerados por estes programas tem formatos padronizados, para que possam ser compreendidos por outros sistemas, sendo as linguagens mais comuns GDSII e CIF, sendo RS uma adaptação usada nos laboratórios da UFRGS e, portanto, de pequena expressão.

O *lay-out* é a ferramenta através da qual a *foundry* reconhece um projeto e o constrói, usando um processo automático de sobreposição de camadas, conforme pode ser detalhado na literatura. Na tecnologia ES2 - 1.2 μ m, o substrato disponível é do tipo P, sendo que transistores NMOS são construídos diretamente sobre este e os PMOS são construídos através de um poço tipo N, de acordo com a figura 8.1.

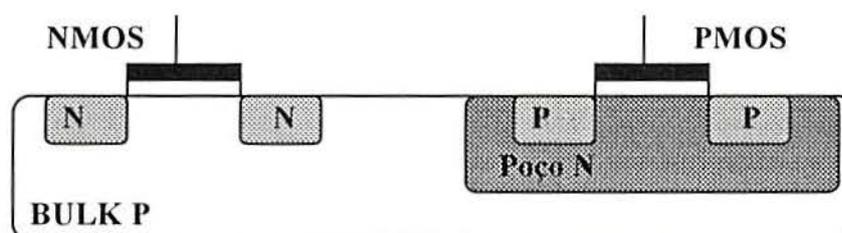


Figura 8.1

No *lay-out*, o substrato tipo P não precisa ser incluído, pois este já está subentendido, embora o poço N, construído através de um implante, deve ser especificado para diferenciar os tipos de transistor. As regiões usadas para construir as regiões de dreno e fonte são denominadas implantes, que podem ser de dois tipos, de acordo com o tipo de transistor que se quer construir. A região de porta é formada através de uma camada de poli-silício, que é separada da região do substrato por uma camada fina de óxido de silício. O procedimento para execução do *lay-out* se baseia nas técnicas de fabricação do próprio dispositivo, de modo que a descrição do PMOS pode ser feita da seguinte forma, conforme esquema da figura 8.2:

* Como o transistor é PMOS, necessita de um substrato tipo N, que será construído sob a forma de um poço. Este transistor terá a característica de poder ter uma tensão de substrato individual, o que não acontece com os transistores NMOS, cujos substratos estão todos ligados no mesmo potencial. A área do poço deve compreender todo o transistor, incluindo certas distâncias mínimas, conforme as regras de tecnologia.

* Marca-se então a área ativa, na qual será implantado o material que comporá os terminais de dreno e fonte. Esta área ativa deve ter dimensões superiores a W e L do próprio transistor, de acordo com as regras de construção impostas pela tecnologia.

* Após isso, a região de porta é determinada por uma camada de poli-silício, ou simplesmente *poly*. As dimensões do *poly* da porta devem ser as seguintes: comprimento igual ao L do transistor e largura superior à dimensão W.

* A alimentação dos terminais de dreno e fonte é feita através de contatos na extremidade da zona implantada. Usa-se o primeiro nível de metal (Metal 1) para a conexão destes terminais com os outros transistores do circuito.

* Configura-se um transistor quando existe uma porta sobre uma área ativa com contatos nas extremidades. Estes contatos são, na verdade, "furos" que permitem que a camada de metal que vem sobre o implante de conecte eletricamente a este. Os contatos são enfileirados na direção da dimensão W para criar uma linha de equipotencialidade e permitir um fluxo homogêneo de portadores de um terminal a outro, caso contrário, a resistência vista em um dos

lados seria maior e haveria uma tendência do fluxo de portadores se concentrar na parte de menor resistência, comprometendo o funcionamento do dispositivo.

* A área ativa não é uma camada, mas sim um artifício usado na fabricação. Conforme já foi mencionado, o transistor PMOS é composto de um substrato (poço N), uma porta (*poly*) e duas zonas de implante P (dreno e fonte). Para poder fazer o implante N dos terminais dreno e fonte, a área ativa e o *poly* (já construído) serão usados como máscara, de maneira que somente a área dos terminais dreno e fonte será atingida, posto que a porta funciona como um escudo, protegendo o implante na região imediatamente abaixo desta. Ainda assim, percebe-se uma disseminação indesejada de portadores para baixo da região de porta, causado pelo conhecido fenômeno de difusão lateral.

* Metal 1 é a forma de conectar eletricamente os terminais de dreno e fonte que não sejam contíguos, sendo que Metal 2 é uma forma alternativa, geralmente empregada quando há necessidade de cruzar duas linhas. Neste caso, o procedimento é usar uma via, que permite a conexão elétrica entre estas duas camadas. Dessa forma, pode-se usar Metal 1 até certo ponto, no qual se "sobe" com uma via até a camada de Metal 2, continua-se a linha com este metal e quando for adequado, se "desce" uma camada, através de outra via, de volta para Metal 1. Com isso, implementa-se um *jumper* em circuitos integrados.

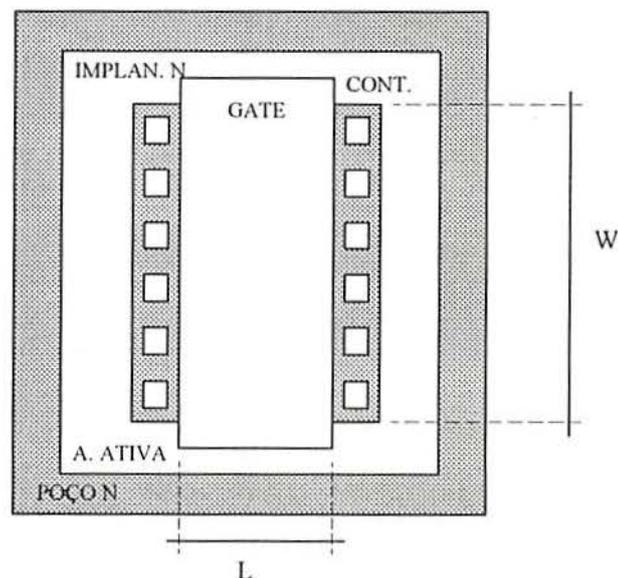


Figura 8.2

A alimentação do substrato é outro ponto que merece especial atenção. Uma vez que a resistividade do material usado para o substrato é relativamente alta, é necessário que *body ties* ou simplesmente *ties* sejam espalhados pela área do circuito, de forma a manter todos os pontos do substrato no mesmo potencial. Isto é válido tanto para o substrato P quanto para o poço N. A tecnologia recomenda uma distância não superior a $100\mu\text{m}$ entre cada *tie*, para que o substrato seja devidamente alimentado. As correntes envolvidas no circuito devem ser relevadas durante a fase de *lay-out*, pois elas definem se a largura das linhas de metal.

8.2 Modificações do projeto original

Embora o objetivo do *lay-out* seja fazer uma réplica do circuito numa linguagem mais propícia à fabricação, uma quantidade significativa de elementos parasitas são introduzidos no circuito, fazendo com que este tenha um comportamento levemente diferente do esperado. Em circuitos digitais, estas mudanças geralmente são desprezadas, posto que a influência no circuito é quase imperceptível, à exceção de alguns nanossegundos a mais na resposta. Se a má alimentação do substrato ou as capacitâncias parasitas adicionadas por efeito da partição de transistores introduzirem desvios no funcionamento do bloco digital, estes nem serão analisados, a menos que interfiram significativamente na performance do circuito. Em parte, isto se deve ao fato de que os circuitos digitais só podem oferecer dois tipos de resposta, separados por uma grande região proibida, de maneira que a falha introduzida pelo *lay-out* deve ser realmente expressiva para poder modificar o comportamento do circuito. O fato de que os desvios causados geralmente não são cumulativos, isto é, uma tensão abaixo do esperado ainda faz com que a tensão na saída do bloco seja V_{dd} ou essencialmente 0 V .

Estas variações introduzidas na fase de *lay-out* podem ter uma influência bem mais significativa em circuitos analógicos, onde uma tensão V_{gs} com 2% de erro pode gerar uma corrente com 4% de erro e sucessivamente pode-se culminar em um problema de inviabilidade de funcionamento. A introdução de resistências e capacitâncias parasitas extras é o maior responsável por esta degradação, posto que cada ligação, cada conexão entre camadas diferentes e na mesma camada é implementada a partir de um material não ideal. Por exemplo, se dois transistores tem suas portas conectadas e, conforme já foi dito, a porta é constituída de *poly*, este mesmo material poderia ser usado para efetuar a conexão entre os dispositivos. O inconveniente deste procedimento é que o *poly* tem uma alta resistividade, o que insere uma resistência parasita de valor significativo entre as duas portas. Na análise DC esta resistência extra não tem maior relevância, posto que

está em série com outra de valor bem maior, a própria resistência de porta. Entretanto, na análise AC, a inclusão dessa resistência adicional, em combinação com a capacitância intrínseca da porta do transistor, pode modificar certas características do funcionamento do dispositivo, especialmente em altas frequências. Esse problema poderia ser minimizado usando uma linha de metal com vias nas extremidades, para formar a conexão com a porta. Dessa forma a resistência adicional é minimizada e o transistor terá um comportamento mais próximo do esperado.

Quando um projeto prevê a construção de transistores de tamanho grande, é costume usar técnicas de partição para construí-lo de forma mais conveniente. Existem três razões básicas que validam a afirmação anterior [30]:

- * Existe uma sensível economia de área no circuito se os transistores de maior tamanho forem partidos, permitindo assim que a forma geométrica do conjunto seja mais homogênea.

- * Um transistor muito longo ou largo pode sofrer influência de variação de dopagem do substrato e ter seu desempenho degradado por este fator. Embora a tecnologia garanta a uniformidade da dopagem do substrato, este pode apresentar um gradiente em uma direção definida. Se um transistor com uma ou duas dimensões muito grandes for construído sobre um substrato cuja concentração de impurezas estiver variando, o seu desempenho pode sofrer sérias consequências. Dessa forma, se existe a possibilidade de partir o transistor de modo a reduzir suas dimensões máximas, este método deve ser empregado.

- * Um transistor único que apresenta-se disposto numa área muito grande pode sofrer de diferenças de temperatura, ocasionada por dissipadores de potência localizados. Esses gradientes térmicos também podem trazer problemas ao funcionamento do transistor, pela mudança de suas características elétricas e físicas. Uma sugestão interessante e útil é sempre colocar os dispositivos inteiros dentro de uma mesma isoterma, ou seja, a região delimitada por duas frentes de onda consecutivas produzidas por uma fonte de calor, conforme a figura 8.3.

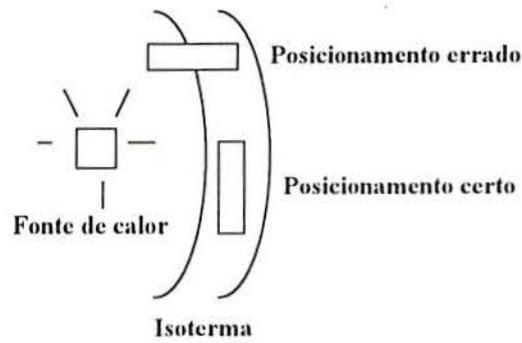


Figura 8.3

Existem várias maneiras de se particionar um transistor, dependendo em especial de qual das dimensões tem um tamanho exageradamente grande para ser construído "ao natural". No caso de circuito analógicos, geralmente a dimensão W é muito maior que L , de modo que dois métodos são largamente empregados:

1) *Stacked Configuration* (Configuração Empilhada). Esquemático na figura 8.4, este método de partição permite quebrar a dimensão W e colocar os pedaços lado a lado, desfrutando de uma fonte ou dreno comum. A porta, em *poly*, está curto-circuitada na extremidade superior e poderia estar também na inferior e sua disposição geométrica permite que vários transistores sejam construídos. Todos os terminais de dreno devem ser curto-circuitados por uma linha de Metal 1, valendo o mesmo para os terminais de fonte. Como são camadas diferentes, separadas por uma camada de óxido, não existe problema em cruzar linhas de *poly* e metal. A dimensão L continua a mesma, ou seja, não sofreu alteração e W_t (largura total) é dado da seguinte forma: $W_t = n \cdot W$, onde n é o número de transistores, determinado pelo número de portas (5, na figura 8.4) e W é a largura resultante. W é medido entre os dois contatos mais afastados no terminal de fonte ou dreno e L é a distância que separa estes terminais, ou mais precisamente, a largura do próprio *poly*. Deve-se observar também que existe uma considerável redução no número de contatos, posto que um terminal de fonte, por exemplo, serve a dois transistores, valendo o mesmo para o dreno. Como a linha de Metal 1 cria uma equipotencialidade nestes terminais, existem portadores suficientes para migrarem através do canal abaixo da porta para os dois lados, sem prejuízo à operação normal. Como o material de porta sai da zona de implante, à cada re-entrada está associado um efeito de borda, isto é, a mesma não-linearidade que se percebe no campo elétrico das extremidades entre duas placas paralelas. Este efeito, embora de pequena magnitude, é proporcional a n , ou seja, ao número de partições. Esta configuração permite facilmente a inclusão de *ties* entre os vários transistores, uma vez que seu comportamento é independente.

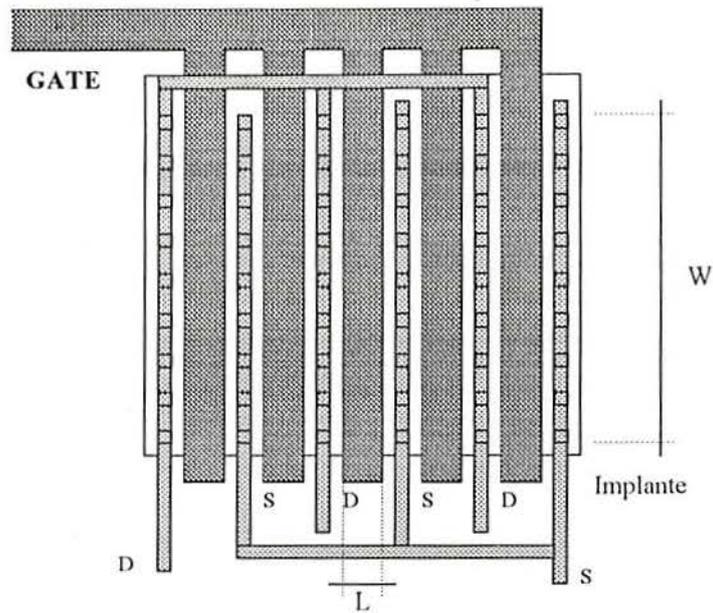


Figura 8.4

2) Configuração Interdigitada. A figura 8.5 mostra um exemplo similar ao anterior, usando, porém, uma outra técnica de partição. Aqui, todo transistor está contido dentro da área de implante, de modo que não existe tanto efeito de borda quanto havia na configuração anterior. Por outro lado, nas partes onde o *poly* de porta é paralelo aos contatos, o funcionamento do transistor é normal, mas nas curvas do *poly* existem regiões de comportamento desconhecido, tornando o cálculo da largura e comprimento efetivo mais difícil. A inclusão de *ties* nesta topologia de partição oferece mais problemas que na anterior, pois as modificações estruturais no interior do dispositivo partido podem acarretar significativas mudanças na geometria resultante.

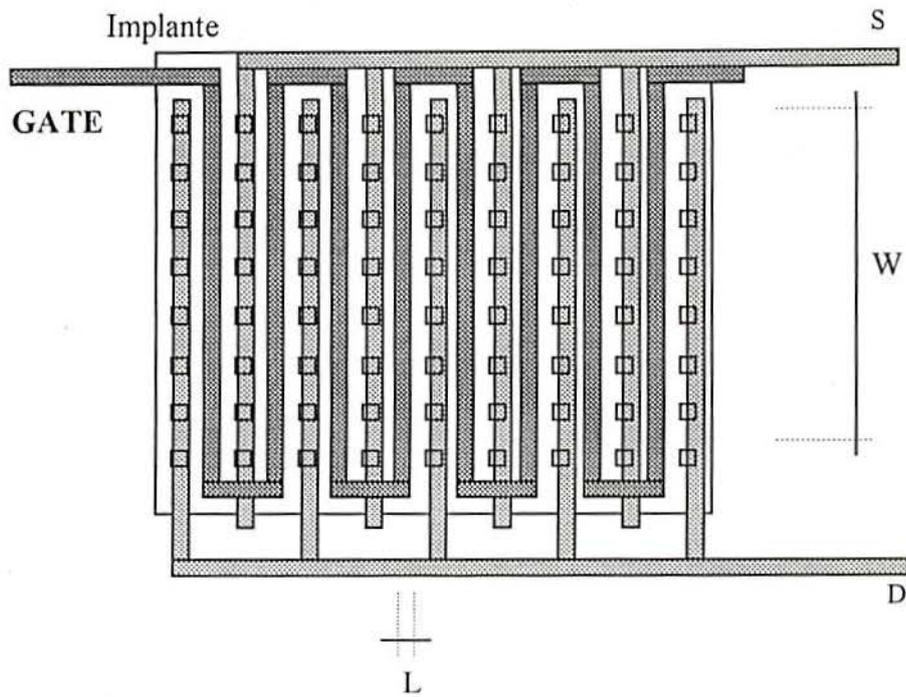


Figura 8.5

No *test chip* que foi executado, ambas técnicas de partição foram usadas, numa mesma topologia, possibilitando assim que se evidencie eventuais problemas decorrentes de cada uma das técnicas de partição. Independentemente da metodologia de partição empregada, qualquer alteração na geometria de um transistor acarreta um certo tipo de erro. Por exemplo, a corrente que um transistor de $100\mu\text{m} \times 10\mu\text{m}$ conduz não é igual àquela conduzida por dois transistores de $50\mu\text{m} \times 10\mu\text{m}$, para uma mesma polarização. Este fato se deve a pequenas variações na tensão limiar, causadas por alterações na geometria do dispositivo. Efeitos de canal curto e mais frequentemente de canal estreito são os responsáveis por estas variações, que podem afetar de maneira significativa a corrente conduzida. Para determinar a quantidade de erro introduzido por efeito de partição, a curva do Gráfico 8.1 mostra o desvio percentual de corrente de dreno quando o transistor é partido. O parâmetro DELTA do modelamento prevê um aumento de V_t por efeito da redução de W , oriundo do particionamento. Com isso, seria de se esperar que quando um transistor é particionado n vezes, como a dimensão W será dividida por n , a tensão V_t aumenta e a corrente diminui. Calcula-se o erro da seguinte forma:

$$e\% = \frac{I_0 - n \cdot I_p}{I_0} \cdot 10$$

Equação 32

onde I_o é a corrente de dreno do transistor original
 n é o número de partições e
 I_p é a corrente individual de um transistor partido.

O circuito usado na determinação deste erro está esquematizado na figura 8.6:

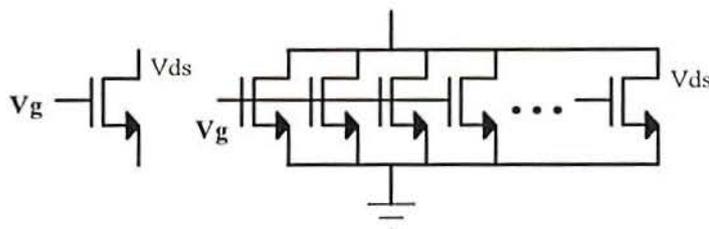


Figura 8.6

Dessa forma, o erro deve ser positivo e proporcional à redução de W , uma vez que o aumento de V_t é função linear da dimensão W . Entretanto, o que se pode comprovar através da observação da curva de erro do Gráfico 9.1, é que existe um erro negativo, isto é, o somatório das correntes dos n transistores em paralelo é maior que a corrente original de um transistor. Analisando as resistências de dreno e fonte da topologia do transistor inteiro, é possível verificar que estas são equivalentes à associação paralela das resistências do transistor partido, de modo que, teoricamente, não deveria haver diferenças. A explicação para as discrepâncias observadas no gráfico reporta-se à interpretação que o simulador SPICE faz das resistências parasitas, contadas em quadrados, ponderadas pelo parâmetro RSH (*Drain and Fonte Sheet Resistance*). Cada transistor é caracterizado por sua geometria, sua área e perímetro, além do número de quadrados de resistência da fonte e dreno. Este número de quadrados (NRD e NRS) é contado da seguinte maneira [31]: para um transistor de dimensão $W \times L$, a dimensão W forma um dos lados do quadrado, de forma que se a distância entre o final do contato e o início do *poly* fosse igual a W , o transistor seria caracterizado por $NRD = 1$. Conforme a figura 8.7, pode-se verificar que esta distância é muito menor que W , de modo que o valor de $NRD \ll 1$.

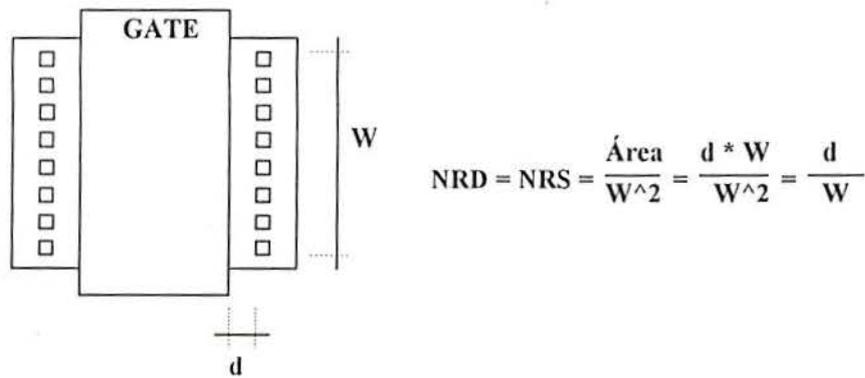


Figura 8.7

Infelizmente, o valor *default* de NRD e NRS no SPICE é unitário e não zero, como seria de se esperar. A consequência disto é que um transistor de $W \times L$ tem uma resistência série igual a $NRD \times RSH$, onde NRD é igual a 1 e RSH tem um valor típico entre 55 e 75Ω. Quando o transistor é partido, a dimensão W é reduzida e com isso a resistência série de cada transistor também o é, mas devido ao valor default de $NRD = 1$, a resistência de cada transistor continua sendo $1 \times RSH$. Com isso, a tensão V_{ds} efetiva, isto é, aquela descontada a queda de tensão de valor $I_d \times R_d$ na resistência parasita, é menor, sobrando mais tensão para o transistor, aumentando assim sua corrente. Para corrigir este efeito, basta informar ao simulador o valor correto de NRD e NRS, dado pela seguinte equação:

$$NRD = NRS = \frac{d}{W}$$

Equação 33

A redução do valor do número de quadrados fará com a corrente diminua, permitindo que se visualize o real problema concernente à partição: o efeito de canal estreito. O modelamento simplificado implementado através do SPICE prevê um aumento da tensão limiar proporcional à diminuição da largura do transistor. Este modelamento não é correto, posto que os efeitos de canal estreito só começam a efetivamente se fazer notar para transistores de largura inferior a aproximadamente 5μm, não interferindo em geometrias maiores. Logo, os erros observados no gráfico são resultados de um modelamento simplificado e com certeza não se farão presentes na implementação física de circuitos que usam técnicas de particionamento. Prova disso é o Gráfico 8.2, no qual figura uma curva de desvio percentual de corrente no qual os transistores componentes do circuito têm a informação de número de quadrados de dreno e fonte. Com isso, o único efeito responsável pelo desvio entre a corrente original

e o somatório das correntes dos transistores partidos é o efeito de canal estreito, linearmente modelado pela formulação SPICE.

Ao invés disso, um problema intrínseco à tecnologia e que deve ser relevado é a incerteza nas dimensões de qualquer dispositivo. Estes valores costumam ser da mesma ordem de grandeza do passo mínimo utilizado no *lay-out*, ou seja, $0.15\mu\text{m}$ (vale dizer que não se pode fazer *lay-outs* com dimensões não múltiplas deste valor, pois o processo de fabricação não oferece precisão suficiente para tanto). Dessa forma, para operação segura, é interessante projetar circuitos cujo funcionamento é garantido mesmo que os transistores que o compõe tenham as máximas variações estatisticamente previstas pela *foundry*. Atender a este quesito significa basicamente usar transistores com altos valores de fator de forma, comprimento e largura, de modo que possíveis variações em quaisquer das dimensões não sejam percentualmente significativas para degradar a performance do circuito. Assim evidencia-se que a nível de simulação, os erros vinculados ao processo de particionamento de transistores não são significativos, podendo dessa forma ser negligenciados.

Dentro de um mesmo circuito integrado, é possível construir uma grande quantidade de amplificadores operacionais. Cada ponto do circuito que tem alguma relação com o mundo externo deve ser conectado ao circuito através de *pads*. Os *pads* de entrada são responsáveis por proteger as partes internas do integrado contra possíveis descargas eletrostáticas que podem danificar seriamente o circuito. Eletricamente, sua função é permitir uma conexão entre a fonte externa de sinal e a entrada do circuito, com o mínimo de interferência, distorção, alteração de magnitude e rotação de fase. Os *pads* de saída, por outro lado, destinam-se a fornecer uma grande quantidade de corrente às cargas externas, tarefa impossível para os operacionais do circuito integrado. Estes circuitos têm ganho de tensão unitário, mínima rotação de fase, mesmo em altas frequências a pequeno tempo de resposta. As alimentações V_{dd} e V_{ss} também são ligadas aos amplificadores operacionais através de *pads*, permitindo operação segura frente a qualquer comportamento espúrio das fontes. Todos os componentes dos circuitos devem estar dentro de um anel, que é um tipo de proteção, sugerido pela própria tecnologia, na qual linhas de alimentação e *ground* permitem isolamento e proteção aos circuitos nele inseridos. A composição deste anel é dada pela própria *foundry*, de modo que os fenômenos que determinam a sua forma não são conhecidos e seu uso limita-se a simples cópias do modelo original. Pelo fato de que este anel é composto de módulos, é possível construí-lo de duas formas: retangular ou quadrado, dependendo do número de pinos do integrado e da quantidade de circuitos a serem construídos.

Como exemplo, a figura 8.8 mostra o *lay-out* de um dos operacionais. Nele, pode-se visualizar facilmente uma grande área composta de pequenos quadrados, representando um capacitor de 2.2pF, através da união de uma camada de poly e outra de metal. Todos os transistores estão na vertical e o par de entrada (M1,M2) se localiza da parte esquerda, entre o espelho de corrente (M3,M4 - esquerda em cima) e a fonte de corrente do primeiro estágio (M5 - esquerda em baixo). Na parte central o transistor de amplificação do segundo estágio (M6) é encontrado sobre a fonte de corrente deste estágio (M7).

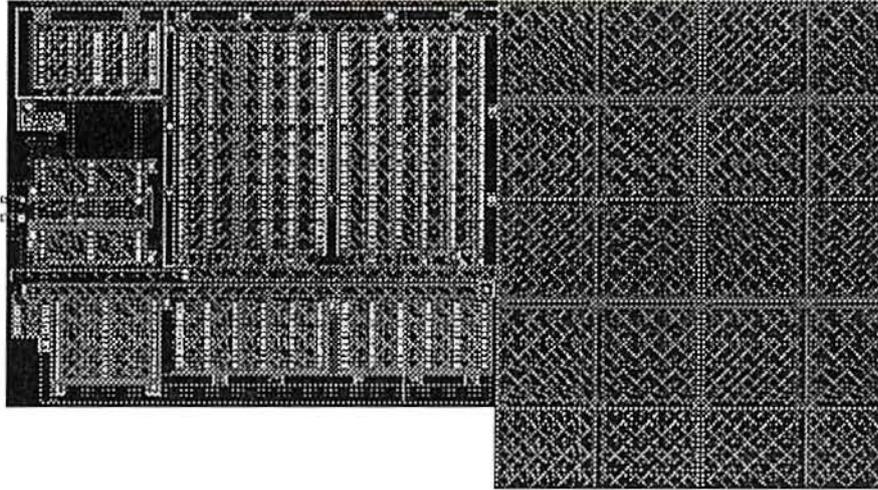


Figura 8.8

No *testchip*, vários circuitos de teste foram enviados, para que sua performance pudesse ser avaliada. As configurações são as seguintes:

Operacional normal:	OP1b	OP2b	OP3b
Operacional com compensação MOS:	OP1b	OP2b	
Integrador com capacitor linear:			OP4b
Integrador com capacitor MOS:			OP4b

Através destes testes poderão ser realmente validadas as rotinas de síntese, pois as respostas DC, AC e transiente de todos os circuitos serão analisadas, no intuito de determinar se a performance simulada se equivale à real. Ainda, poderá se verificar o desempenho do capacitor de compensação implementado através da porta de um MOS. A comparação será direta, posto que o mesmo circuito foi construído duas vezes, sendo que a única diferença é a construção do elemento capacitivo.

8.3. MÉTODOS DE VERIFICAÇÃO

A análise visual de um *lay-out* pode se tornar uma tarefa muito complexa, especialmente quando o número de transistores é grande. Para assegurar que a formação dos transistores é adequada, existem *softwares*, conhecidos como DRC (*Design Rule Checkers*), capazes de interpretar corretamente a linguagem de descrição de *lay-out*. Estes programas permitem que seja feita uma conferência automática das regras de projeto impostas pela tecnologia. Todas as distâncias mínimas entre entes diferentes são conferidas, evitando assim que se envie para fabricação um *lay-out* com erros. O grande número de regras tecnológicas faz do *lay-out* manual um processo passível de erro, tornando imprescindível, portanto, o uso de uma ferramenta automática de conferência. A análise que os DRC's utilizam para determinar e encontrar erros é baseada nos mais variados algoritmos, diferindo entre si pela velocidade de sua análise e quantidade de "erros falsos". A ferramenta usada neste trabalho chama-se DARC3 [32] e, apesar de sua funcionalidade, frequentemente apresenta uma certa quantidade de erros na *lay-out* que na verdade não existem (são os referidos erros falsos). Um exemplo simples de erro falso está ilustrado na figura 8.9. Conforme as regras da tecnologia, um contato deve distar no mínimo $1.5\mu\text{m}$ da borda da zona implantada, de modo que se um DRC encontrar um contato que desobedeça esta regra, irá acusar um erro.

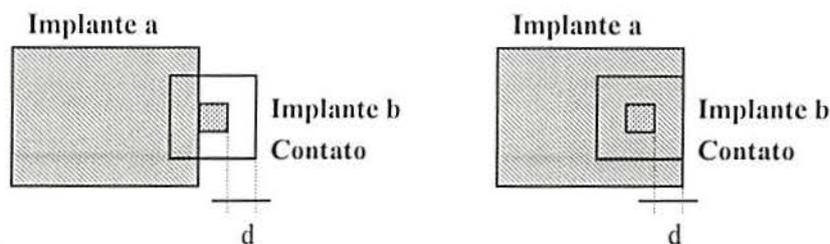


Figura 8.9

Se o contato for analisado como estando dentro do Implante b, as regras não são infringidas, posto que a distância d é observada em todas as dimensões. Porém, se o DRC considerar que o contato pertence ao Implante a (no primeiro desenho da figura 8.9), a regra da distância mínima entre contato e implante é desrespeitada, acusando um erro. Como regra geral, o DRC utilizado analisa sempre o maior retângulo de uma certa camada, em comparação às outras, de forma que no exemplo dado, o contato seria analisado com respeito ao Implanta a, acusando um erro. Naturalmente, este erro é

considerado falso, obrigando o projetista a reposicionar o Implante b e o contato de forma a não permitir que esta análise errônea produza um erro inexistente, de acordo com a segunda versão da figura 8.9. O programa DARC3 utiliza um arquivo de tecnologia, no qual estão definidas todas as regras de projeto e que serve, portanto, como fonte de consulta na procura de erros. Um aspecto que é também considerado erro de *lay-out* é o desrespeito ao passo mínimo, neste caso $0.15\mu\text{m}$. Todo e qualquer retângulo que compuser o *lay-out* deve ter dimensões múltiplas inteiras do passo, valendo o mesmo para as distâncias entre retângulos quaisquer, inclusive entre-camadas. Esta exigência tecnológica tem suas causas na limitação física da precisão na construção de uma camada, não só de seu valor absoluto, mas também do seu posicionamento na célula. Conforme já foi salientado, uma das imperfeições na construção do transistor é a migração de impurezas das zonas implantadas para baixo da porta, fenômeno conhecido como difusão lateral. Este efeito tem um comportamento aproximadamente constante, sendo modelado através do parâmetro LD no simulador SPICE. Este fenômeno de natureza determinística deve ser somado à outro estatisticamente previsto pela tecnologia, de modo que o comprimento real de um transistor é dado por:

$$L_r = L_m - 2 \cdot L_{dl} + \Delta$$

Equação 34

onde L_r é o comprimento real,

L_m é o comprimento da máscara (original do projeto)

L_{dl} é a difusão lateral e

ΔL é a variação acarretada pela incerteza no processo de fabricação.

Embora ΔL não seja nulo, seu valor médio é zero, de forma que pode se admitir, na fase de projeto, que $L_r = L_m - 2 \cdot L_{dl}$. O mesmo não ocorre na dimensão W , posto que não existe difusão lateral aumentando ou diminuindo a largura do transistor, uma vez que o material de porta é mais largo que a zona implantada exatamente para impedir este efeito indesejado. Dessa forma, a largura real do transistor é dada por:

$$W_r = W_m + \Delta W$$

Equação 35

onde ΔW tem o mesmo comportamento estatístico que ΔL .

Uma outra fonte de erro de difícil detecção visual são as interconexões entre os transistores do circuito, implementadas através de linhas de Metal 1 e 2. Uma ferramenta automática largamente utilizada com este propósito são os extratores. Neste trabalho, o software utilizado foi o EX7, um produto dos laboratórios da UFRGS, que, além de verificar que o transistor está bem formado, permite conferir as ligações do circuito, evitando assim que um pequeno curto-circuito ou circuito aberto torne o circuito inoperante. A análise do *lay-out* permite que se extraia todos os transistores do circuito, bem como suas conexões com os outros dispositivos. No caso de transistores partidos, *alguns* deles são unidos, sendo que esse processo pelo qual essa união não é generalizada não é conhecido. Além das dimensões W e L do transistor, as áreas e perímetros do dispositivo são levantadas, tornando mais acurada a resposta em altas frequências. Este programa também oferece a possibilidade de extrair os elementos parasitas intrínsecos ao *lay-out*, mas pelo fato de que o tamanho do circuito excede o máximo permitido pelo programa, esta característica não foi usada. Um problema intrínseco a esta ferramenta, que foi feita preferencialmente para circuitos digitais, é que, quando da extração de transistores partidos, o parâmetro NRD e NRS não é mencionado, mantendo portanto seu valor *default* unitário e ocasionando os erros já evidenciados. Uma característica facilitadora do extrator utilizado é que este apresenta a possibilidade de atribuir números aos diferentes pontos do *lay-out*, permitindo assim que se visualize mais rapidamente a existência das conexões.

Utilizando os DRC's e os extratores é possível garantir a qualidade do *lay-out*, evitando assim que um dispendioso e demorado processo como o de construção de um circuito integrado, seja feito inutilmente, ou seja, com erros. Mas simplesmente garantir que o *lay-out* não está errado nem sempre é suficiente, devendo-se recorrer a certas normas, algumas das quais já mencionadas neste trabalho, que objetivam a obtenção de uma melhor performance do circuito. A precisão não ideal das técnicas de fabricação é uma característica com a qual o projetista deve se acostumar, pois mudá-las está fora de seu alcance. Entretanto, algumas medidas podem ser tomadas de maneira que esta lacuna não assuma proporções suficientemente expressivas para degradar o desempenho do bloco. O primeiro aspecto diz respeito a transistores casados, isto é, cujas relações de aspecto devem ter um valor determinado, geralmente unitário. O espelho de corrente formado por M3 e M4 (vide figura 6.1), embora não sendo o mais importante, é o forte candidato à aplicação destas técnicas de *lay-out*. A seguir são dadas sugestões que podem melhorar a qualidade de circuitos que usam transistores iguais, embora os princípios valham igualmente para transistores proporcionais.

* Dispositivos casados devem ter a mesma estrutura, isto é, se partidos, o número de partições deve ser igual; as camadas de implante, área ativa, *poly* e poço devem ter as mesmas dimensões. Isto vale para transistores, onde W/L, W e L devem ter o mesmo valor e para capacitores e eventuais resistores, que devem apresentar a mesma forma geométrica.

* Sempre que possível, manter os dispositivos casados na mesma temperatura, o que significa posicioná-los equidistantes das principais zonas de dissipação de potência.

* Reduzir ao máximo a distância entre os transistores, evitando assim possíveis variações na concentração de impurezas do substrato local.

* Dispor os transistores invariavelmente na mesma orientação, embora isto possa trazer um aumento de área ocupada, em comparação aos circuitos que desobedecem este princípio. Ensaio experimentais já demonstraram que a orientação dos transistores afeta sensivelmente o seu desempenho, sendo que, para dispositivos casados, esta técnica tem valor dobrado.

* Além de seguir a mesma orientação, é aconselhável, sempre que possível, que os transistores sejam dispostos sobre linhas imaginárias, pois isso garante operações semelhantes. Deve-se isso ao fato que variações causadas pelos diferentes processos, aliadas ao stress que o circuito é submetido na montagem final, geralmente se processam em direções perpendiculares ou paralelas ao próprio circuito, de modo que se a orientação sugerida for mantida, estes efeitos tendem a ter pouco influência.

* Posicionar dispositivos semelhantes à volta dos transistores casados, de forma que se situem igualmente, também permite que o comportamento do bloco seja melhor. Se não houver dispositivos para serem colocados, pode-se recorrer à técnica de *dummy devices*, ou seja, posicionar adequadamente dispositivos sem função nenhuma, mas cuja presença permite maior simetria para os transistores casados.

* A não utilização de tamanhos mínimos aumenta a probabilidade de que os transistores sejam efetivamente construídos com pequenas diferenças percentuais de tamanho.

Um outro exemplo é o par diferencial, que se encontra no primeiro estágio do amplificador Miller, representado pelos transistores M1 e M2. O desempenho destes transistores é de vital importância para o resto do circuito, de maneira que algumas precauções são tomadas para que seu funcionamento esteja dentro dos limites satisfatórios [30]. Inicialmente, deve-se dizer que M1 e M2 são implementados na verdade através de quatro transistores, conforme a figura 8.10.

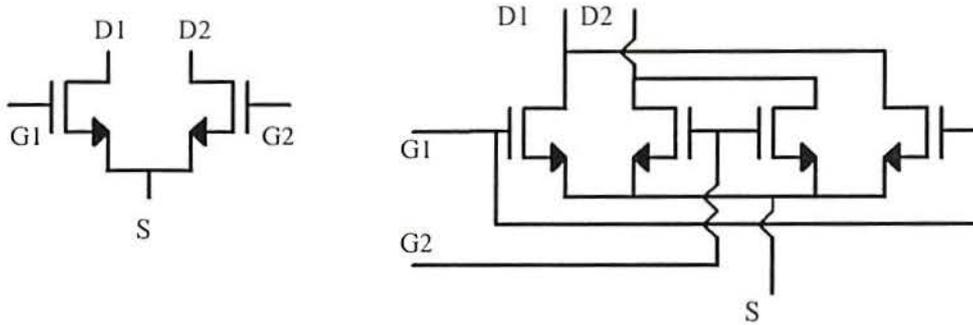


Figura 8.10

Conforme pode-se observar no *lay-out* do operacional, o par diferencial (primeiro esquema da figura 8.10), identificado por seus terminais G1, G2, D1, D2 e S está funcionalmente repetido no segundo esquema, com os correspondentes terminais. Para que o comportamento se equivalha nas duas implementações, a dimensão W dos transistores do par diferencial de projeto deve se transformar em $W/2$ na implementação real de *lay-out*. A disposição em cruz desses quatro transistores permite que oscilações no perfil de impurezas do substrato sejam compensadas, pois ao mesmo tempo que um transistor é prejudicado, o seu correspondente é favorecido, em qualquer direção de variação que existir. Além disso, esta disposição também torna o circuito imune a gradientes térmicos, que geralmente causam desvios de corrente que ocasionam *off-set*.

9 Conclusão

O objetivo deste trabalho consistia em se lançar na pesquisa de um segmento até hoje pouco explorado pela comunidade científica nacional: a síntese automática de circuitos analógicos. A razão disso é o crescente interesse da indústria por soluções melhores, mais rápidas, mais eficientes e sobretudo mais baratas. Empregando técnicas como *System on Silicon*, tornou-se possível integrar um projeto genérico qualquer em um único circuito integrado, de custo reduzido e eficiência superior. Porém quando estes sistemas possuem partes analógicas, como é a maioria dos casos, torna-se imprescindível saber projetar esses blocos de forma integrada. Dessa forma, este trabalho vem ao encontro à uma carência que já há muito se fazia notar: a de um sistema capaz de projetar um circuito analógico, como um amplificador operacional, em menos tempo, com mais confiabilidade e economia, tanto de área quanto de potência.

Para executar este trabalho, várias etapas foram vencidas, entre as quais pode-se citar como sendo as mais importantes:

- * Estudo da modelagem do dispositivo transistor. Através de uma abordagem matemática, procurou-se reunir um conjunto de equações capazes de prever o comportamento elétrico do transistor sob quaisquer condições. Somente a implementação dessas equações em um programa apropriado permitiria que as etapas seguintes fossem realizadas com êxito.

- * Análise minuciosa dos fatores que influenciam as principais características do circuito em questão. Naturalmente este estudo foi realizado para servir de base às próximas etapas do projeto, pois para construir um operacional com determinadas características fornecidas pelo usuário, o programa deve saber quais são as condições de contorno para a resolução do problema, ou, dito de outra forma, deve saber quais são os elementos responsáveis por aquela característica e como alterá-los de modo a obter a especificação solicitada.

- * Criação de algoritmos de síntese automática. Sem dúvida, a parte mais complexa do projeto só foi superada através da criação de ferramentas auxiliares que viabilizaram o processo de síntese. Várias versões desses algoritmos foram criadas, privilegiando uma ou outra característica, com mais

ou menos necessidade de interação com o usuário; de maneira que ao final obteve-se o melhor compromisso entre precisão, complexidade e velocidade dos dados fornecidos por este programa. A iteração é a principal das ferramentas, presente em quase a totalidade dos algoritmos de síntese. Devido ao fato de que geralmente o número de incógnitas é bem superior ao número de equações disponíveis, freqüentemente é necessário arbitrar valores iniciais a algumas variáveis e somente através de processos iterativos torna-se possível encontrar o valor real dessa variável. A derivação numérica foi outro recurso amplamente utilizado, capaz de descrever, com suficiente precisão, as características de pequenos sinais do transistor que são fatores de alta relevância em um projeto de amplificar operacional.

* Projeto do circuito de polarização, simulação e conferência. Para dar por encerrado o processo de síntese, o SPICE 2G foi utilizado como simulador padrão para verificação final dos circuitos antes que os *lay-out's* fossem feitos. O circuito de polarização, embora sendo muito simples, foi projetado utilizando as ferramentas desenvolvidas ao longo do trabalho, provando que estas são suficientemente versáteis para serem reaproveitadas em outros lugares.

* *Lay-out* manual, DRC e extrator. Ao final da etapa de síntese, iniciaram-se os trabalhos para realização dos desenhos de *lay-out*, que, infelizmente, não contaram com nenhuma ferramenta automática e foram, portanto, manualmente desenhados. Tanto o DRC como o extrator usado neste trabalho são *softwares* desenvolvidos em trabalhos anteriores, dentro dos laboratórios da UFRGS. Uma vez concluído o trabalho de *lay-out* de cada um dos amplificadores, a ligação do circuito completo foi feito, usando para isso ferramentas mais avançadas, como EDILEX, em uma estação de trabalho. Naturalmente para que tudo não fique no campo puramente teórico, um projeto de um CI de teste foi enviado para fabricação, contendo várias configurações de amplificadores propícias para teste e comparação com a performance simulada.

Após este breve relato sobre as etapas que compuseram este trabalho, seria interessante analisar suas características funcionais e observar as perspectivas para trabalhos futuros. Primeiramente, o que mais chama a atenção nesta ferramenta é a velocidade com que se pode projetar um circuito como o amplificador operacional Miller. O fato ainda de que o usuário tem certo grau de liberdade para fixar alguns parâmetros e "dirigir" o projeto para uma direção segundo seu interesse, torna o sistema versátil no momento em que pode responder a várias solicitações de especificações

diferentes. Alterando parâmetros, valores iniciais e especificações de performance, a gama de projetos possíveis é muito grande e, com certeza quase absoluta, uma destas configurações há de se adequar às necessidades do usuário. Ainda, a possibilidade de aplicar as ferramentas de *software* desenvolvidas neste trabalho para síntese de operacionais em outras configurações abre as portas para um sem-fim de possibilidades na área de síntese de amplificadores multi-topologia.

Por outro lado, seria um erro imaginar que este trabalho esteja concluído; longe disso, verificou-se que existe uma quantidade de alterações e melhoramentos a fazer que podem aumentar o grau de automatismo e ainda reduzir o tempo de projeto. A principal modificação deixada para o futuro é a substituição do modelo Vladimirescu pelo EKV, da Escola Politécnica de Lausanne, Suíça. Dessa forma, a convergência final será obtida mais rapidamente devido à natureza intrinsecamente contínua da curva que descreve a característica elétrica do transistor. Ainda, a boa resposta do modelo em regime de inversão fraca e moderada permitirá que se faça projetos nestas condições de inversão, fato totalmente proibido até agora, devido a limitações do modelo empregado.

Uma outra opção que foi incluída no programa foi a de gerar automaticamente um arquivo com a descrição do circuito, tamanhos e polarização, em formato SPICE. Cogitou-se a hipótese de iniciar os trabalhos de *lay-out* automático, gerando um arquivo RS a partir dos tamanhos calculados no programa, mas a idéia, por ser trabalhosa demais, foi deixada de lado, mesmo porque os algoritmos de *lay-out* automático de circuitos analógicos ainda não estão suficientemente funcionais e são por demais complexos. Uma nova versão do programa foi criada, permitindo entrada de dados de forma gráfica padrão Windows, facilitando o trabalho do projetista e tornando o *software* mais apresentável.

Concluindo, verifica-se que a criação de rotinas de síntese semi-automática foi exitosa no momento em que possibilitou o projeto de amplificadores operacionais de forma muito mais rápida e confiável que os métodos anteriormente empregados. Ainda, pelo fato de possuir versatilidade e modularidade, muitas combinações podem ser feitas, originando uma grande quantidade de rotinas e conseqüentemente de novos projetos. Os algoritmos se baseiam na necessidade do usuário e procuram atender às especificações mais importantes para este. Procurou-se fazer com que os dados sejam mostrados ao usuário no menor tempo possível, facilitando assim o processo de escolha da melhor configuração.

. Embora trabalhos similares a este já tenham sido publicados, eles referem-se a um contexto distinto e surgem como propostas de solução para problemas diferentes em realidades diferentes. Uma das características mais interessantes deste trabalho é o fato de que, uma vez reconhecidas as carências e dificuldades de projeto que enfrenta-se nos nossos laboratórios, foi possível transpô-las para o sistema de síntese, tornando-o particularmente hábil para sanar essas dificuldades. A conclusão final do trabalho é que, na tentativa de se atualizar com as novas tendências mundiais na área de pesquisa de circuitos analógicos, foi obtido um programa de bom desempenho, que foi concebido de acordo com realidade local e que responde a uma necessidade bem específica da nossa situação atual. Dessa forma, lançar-se em um terreno desconhecido, no qual a pesquisa de base foi absolutamente imprescindível, permitiu um sensível avanço tecnológico, que refletir-se-á não somente nos circuitos que serão doravante produzidos, mas também nos novos trabalhos que virão a complementar e aperfeiçoar este.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] - Gray, Paul R. - Basic MOS Operational Amplifier Design - An overview, University of Berkeley, Califórnia, 1980
- [2] - Anais da IV EBMicro - Escola Brasileira de Microeletrônica - Microeletronica para telecomunicações - Tutorial sobre VHDL - Janeiro de 1995 - Recife - Brasil
- [3] - Duchene, Philippe e Declerq, Michel - A highly flexible Sea-of-gates Structure for Digital and Analog Applications - IEEE Journal of Solid State Circuits Vol 24, No 3 - Junho 1989
- [4] - Vittoz, Eric e Fellrath, Jean - CMOS Analog Integrated Circuits Based on Weak Inversion Operation - IEEE Journal of Solid State Circuits, Vol SC-12, No 3 - Junho 1977
- [5] - Ribner, David B. e Copeland, Miles A. - Design Techniques for Cascoded CMOS Op Amps with Improved PSRR and Common-Mode Input Range - IEEE Journal of Solid State Circuits, vol SC-19, no 6, Dezembro - 1984
- [6] - Hodges, David A. - Analog Switches and Passive Elements in MOSLSI - Department of Electrical Engineering and Computer Science, University of Califórnia - Berkeley - Março de 1980
- [7] - Antognetti, P. e Caviglia, D. D. - CAD Model for Threshold and Subthreshold Conduction in MOSFET's - Manuscrito recebido e revisado em Janeiro de 1982 - Istituto di Elettrotecnica, Università de Genova - Itália
- [8] - Montoro, Carlos Galup e Bermudez, José Carlos M. - Switched Capacitor Circuits Fully Compatible with Digital Si-Gate Single Poly Technology - Midwest, Sant Louis ??
- [9] - Montoro, Carlos Galup, Schneider, Márcio C. e Loss, J. B. - Series-Parallel Association of FET's for High Gain and High Frequency Applications - IEEE Journal of Solid-State Circuits, Vol 29, No 9 - Setembro 1994
- [10] - Gouveia Filho, Oscar da Costa - Programa de auxílio no projeto de amplificadores operacionais CMOS - Dissertação de mestrado da Universidade Federal de Santa Catarina - Florianópolis - Junho 1992.
- [11] - Choi, Jung Hyun - Uma análise comparativa entre os conversores A/D Sigma-Delta direcionada ao projeto de uma arquitetura VLSI CMOS - Porto Alegre - CPGCC - UFRGS - 1994
- [12] - Unbenhauen, Rolf e Cichocki, Andrzej - MOS Switched Capacitor and Continuous-Time Integrated Circuits and Systems - Berlim, Alemanha - 1989

- [13] - Gray, Paul R. e Meyer G. Robert - MOS Operational Amplifier Design - A Tutorial overview, IEEE Journal of Solid State Circuits, vol SC-17, no 6, Dezembro - 1982
- [14] - Allen, Philip E., Holdberg, Douglas L - CMOS Analog Circuit Design. Califórnia, 1985.
- [15] - Gregorian, Roubik e Temes, Gabor C. - Analog MOS Integrated Circuits for Signal Processing, Wiley series on filters - 1986
- [16] - Laker, R. Kenneth e Sansen, Willy M.C. - Design of Analog Integrated Circuits and Systems, McGraw Hill - 1994
- [17] - Tsvividis, Yannis P. - An Integrated NMOS Operational Amplifier with Internal Compensation - IEEE Journal of Solid State Circuits, vol SC-1, Dezembro - 1976
- [18] - Advanced Analog Digital Design - Eurochip Course - Universidade Católica de Louvain-la-Neuve, 1994
- [19] - Wu, Wen-Chung, Helms, Ward J. e Byrket, Bruce E. - Digital Compatible High Operational Amplifier with Rail-to-Rail Input and Output Ranges, IEEE Journal of Solid State Circuits, vol 29, no 1, Janeiro - 1994
- [20] - Tsvividis, Yannis, MOS Modelling for Analog circuit CAD: Problems and Prospects - IEEE Journal of Solid State Circuits, Vol 29, No 3 - Março 1994.
- [21] - Vladimirescu, Andrei e Liu, Sally - The simulation of MOS integrated circuits using SPICE2 - Universidade da Califórnia, Berkeley - Memorandum UCB/ERL M80/7
- [22] - Enz, Christian, Proceedings of IV EBMicro, IV Brazilian Summer School on Microelectronics, Recife - 1995.
- [23] - Machado, Gerson - On the usage of EKV MOST model for circuit design and simulation, Proceedings of IV EBMicro, IV Brazilian Summer School on Microelectronics, Recife - 1995.
- [24] - Toumazou, Christofer e Makris, Costas - Analog IC Design Automation: Part I - Automated Circuit Generation: New Concepts and Methods, IEEE Transactions on Computer-Aided Design of Integrated Circuits, vol 14, no 2, Fevereiro - 1995
- [25] - Makris, Costas e Toumazou, Christofer - Analog IC Design Automation: Part II - Automated Circuit Correction by Qualitative Reasoning, IEEE Transactions on Computer-Aided Design of Integrated Circuits, vol 14, no 2, Fevereiro - 1995
- [26] Jaspers, Paul - Basic Building Blocks (large and small signal) - Eurochip Course - Universidade Católica de Louvain-la-Neuve, 1994
- [27] - Vittoz, Eric e Oguey, Cservensky - Analog Circuits modelling, Eurochip Course - Universidade Católica de Louvain-la-Neuve, 1994
- [28] - Casacurta, Alexandre; Suzim, Altamiro Amadeu; Barone, Dante Augusto Couto; Carro, Luigi; Stemmer, Marcos Augusto. - SELA: sistema de edição de *lay-out's*. Seminário Interno da Microeletrônica - Capão da Canoa, CPGCC - 1991.

- [30] - Cohn, John M.; Garrod, David J. - Analog Device-Level Lay-out Automation - Boston, Kluwer Academic Publishers - 1994.
- [31] - SPICE: A guide to Circuit Simulation & Analysis using PSPICE - Paul W. Tuinenga - Microsim Corporation. Ed. Prentice Hall - Segunda edição - New Jersey - 1992.
- [32] - Cleto, Laute Davi; Dossa, Marcos Kordyas - The DARC3 design rule Checkers: new features and modifications - Seminário Interno da Microeletrônica - Capão da Canoa, CPGCC - 1991.

ANEXO 1 - Resolução das equações do CASCODE

Nesse anexo está contido a dedução das equações que pertinem visualizar a utilidade de circuitos com transistores Cascode, segundo o equacionamento simplificado da página 33.

$I_1 = g_{m1} \cdot V_{in}$	Definições
$I_2 = g_{m2} \cdot V_1$	
$V_1 = (I_1 - I_3) \cdot r_{ds1}$	
$(I_1 - I_3) \cdot r_{ds1} - (I_3 - I_2) \cdot r_{ds2} - R_1 I_3 = 0$	Equação principal
$I_3 = \frac{(r_{ds1} \cdot I_1 + r_{ds2} \cdot I_2)}{((r_{ds1} + r_{ds2}) + R_1)}$	Isolando para I3
$V_1 = \left[I_1 - \frac{(r_{ds1} \cdot I_1 + r_{ds2} \cdot I_2)}{(r_{ds1} + r_{ds2} + R_1)} \right] \cdot r_{ds1}$	Aplicando na definição de V1
$I_2 = g_{m2} \cdot \left[I_1 - \frac{(r_{ds1} \cdot I_1 + r_{ds2} \cdot I_2)}{(r_{ds1} + r_{ds2} + R_1)} \right] \cdot r_{ds1}$	Aplicando na definição de I2
$I_2 = \frac{\left[-g_{m2} \cdot r_{ds1} \cdot I_1 + g_{m2} \cdot \frac{r_{ds1}^2}{(r_{ds1} + r_{ds2} + R_1)} \cdot I_1 \right]}{\left[1 + g_{m2} \cdot \frac{r_{ds1}}{(r_{ds1} + r_{ds2} + R_1)} \cdot r_{ds2} \right]}$	Resolvendo para I2
$I_2 = \frac{\left[g_{m2} \cdot r_{ds1} \cdot g_{m1} \cdot V_{in} - g_{m2} \cdot \frac{r_{ds1}^2}{(r_{ds1} + r_{ds2} + R_1)} \cdot g_{m1} \cdot V_{in} \right]}{\left[1 + g_{m2} \cdot \frac{r_{ds1}}{(r_{ds1} + r_{ds2} + R_1)} \cdot r_{ds2} \right]}$	
$V_{out} = R_1 I_3$	Definição de Vout
$V_{out} = R_1 \cdot \frac{(r_{ds1} \cdot I_1 + r_{ds2} \cdot I_2)}{(r_{ds1} + r_{ds2} + R_1)}$	Substituindo I3
$V_{out} = R_1 \cdot \frac{(r_{ds1} \cdot g_{m1} \cdot V_{in} + r_{ds2} \cdot I_2)}{(r_{ds1} + r_{ds2} + R_1)}$	Substituindo I1, I2

$$V_{out} = R_L \left[\frac{r_{ds1} \cdot g_{m1} + r_{ds2} \cdot \left[\frac{g_{m2} \cdot r_{ds1} \cdot g_{m1} - g_{m2} \cdot \frac{r_{ds1}^2}{(r_{ds1} + r_{ds2} + R_L)} \cdot g_{m1}}{1 + g_{m2} \cdot \frac{r_{ds1}}{(r_{ds1} + r_{ds2} + R_L)} \cdot r_{ds2}} \right]}{(r_{ds1} + r_{ds2} + R_L)} \right] \cdot V_i$$

$$A_v = R_L \left[\frac{r_{ds1} \cdot g_{m1} + r_{ds2} \cdot \left[\frac{g_{m2} \cdot r_{ds1} \cdot g_{m1} - g_{m2} \cdot \frac{r_{ds1}^2}{(r_{ds1} + r_{ds2} + R_L)} \cdot g_{m1}}{1 + g_{m2} \cdot \frac{r_{ds1}}{(r_{ds1} + r_{ds2} + R_L)} \cdot r_{ds2}} \right]}{(r_{ds1} + r_{ds2} + R_L)} \right]$$

$$A_v = (1 + r_{ds2} \cdot g_{m2}) \cdot g_{m1} \cdot r_{ds1} \cdot \frac{R_L}{(r_{ds1} + r_{ds2} + R_L + g_{m2} \cdot r_{ds1} \cdot r_{ds2})} \quad \text{Substituindo } r_{ds} \cdot g_m \text{ por } A_v$$

$$A_v = R_L \cdot g_{m1} \cdot \frac{A_{v2} \cdot r_{ds1} + r_{ds1}}{A_{v2} \cdot r_{ds2} + r_{ds1} + r_{ds2} + R_L} \quad \text{Ganho de tensão mostrando a amplificação de resistência causada pelo estágio } \textit{Cascode}$$

ANEXO 2 - Codificação MATLAB da rotina que implementa o modelo de transistor proposto por Vladimirescu

Nesse anexo está listado o programa desenvolvido sob a plataforma MATLAB para implementar o comportamento elétrico estático e dinâmico do transistor. A partir de parâmetros tecnológicos, tamanho e polarização do transistor, o modelo abaixo fornece o valor da corrente de dreno, usando para isso o equacionamento fornecido em [21], cobrindo somente a região de inversão forte.

```
function I = quads1(Vgs,Vds,Vbs,W,L,tip0)
% Calcula a corrente total de dreno com base no modelo do Vladimirescu
% Usa 6 parametros:
% Vgs, Vds, Vbs, W, L, tipo (0,1)

if tipo == 0
    Neff = 4;
    Pb = 0.68;
    Vfb = -0.588536478036;
    Vt = 0.7;
    Ld = 0.125e-6;
    Uo = 510e-4;
    Ucrit = 24.3e5;
    Uexp = 0.22;
    Utra = 0;
    Nsub = 2e22;
    Xj = 0.4e-6;
    Vmax = 54e3;
else
    Neff = 0.88;
    Pb = 0.78;
    Vfb = -0.253743435;
    Vt = 1.1;
    Ld = 0.1e-6;
    Uo = 210e-4;
    Ucrit = 51e5;
    Uexp = 0.33;
    Utra = 0;
    Nsub = 5e22;
    Xj = 0.5e-6;
    Vmax = 47e3;
end

Delta = 0.4;
L = L - 2*Ld;
Fi = Ut * log(Nsub/Ni);
Vbi = Vfb + 2*Fi;
Vbin = Vbi + Delta*pi*Esi*(2*Fi-Vbs) / (4*Cox*W);
Xd = sqrt(2*Esi/(q*Nsub));
```

```

Ws = Xd * sqrt(2*Fi-Vbs);
Wd = Xd * sqrt(2*Fi - Vbs + Vds);
alfaS = Xj/(2*L) * ( sqrt(1+ 2*Ws/Xj) - 1);
alfaD = Xj/(2*L) * ( sqrt(1+ 2*Wd/Xj) - 1);
Gamma = sqrt(2*Esi*q*Nsub) / Cox;
Gamma = 0.65;
Gammass = Gamma*(1-alfaS-alfaD);
Vfb = Vt - 2*Fi - Gamma*sqrt(2*Fi);
Vth = Vbin + Gammass*sqrt(2*Fi-Vbs) + Esi*Delta*pi*(2*Fi-Vbs) / (4*Cox*W);

```

```

if Vgs >= Vth
    if Ucrit *Esi/Cox < Vgs-Vth-Ultra*Vds
        Mus = Uo * ( Ucrit*Esi / (Cox * (Vgs - Vth - Ultra*Vds)) )^Uexp;
    else
        Mus = Uo;
    end

```

```

N = 1 + pi*Esi/(4*Cox*W);
V1 = (Vgs-Vbin)/N + 2*Fi-Vbs;
V2 = 2*Fi-Vbs;
V = Vmax*L/Mus;
a = 4*Gammass/3;
b = -2* (V1+V);
c = -2*Gammass* V;
d = -(V2^2) -4*Gammass/3 * V2^1.5 + 2*V1*V2 + 2*V*V1;
R = [1 a b c d];
X = roots(R);
sat = X.^2 - 2*Fi+Vbs;
Vdssat = min(sat);

```

```

Xdl = sqrt(2*Esi/(q*Nsub*Neff));
L = L - Xdl*sqrt( (Xdl*Vmax/(2*Mus))^2 + Vds-Vdssat) + Xdl^2*Vmax/(2*Mus);
Beta = W*Mus*Cox/L;

```

```

    if Vds >= Vdssat
        Ids = Beta * ( (Vgs-Vbin-N*Vdssat/2)*Vdssat - 2*Gammass/3*( (2*Fi+Vdssat-
Vbs)^1.5 - (2*Fi-Vbs)^1.5 ) );
    else
        Ids = Beta * ( (Vgs-Vbin-N*Vds/2)*Vds - 2*Gammass/3*( (2*Fi+Vds-Vbs)^1.5 -
(2*Fi-Vbs)^1.5 ) );
    end
else
    Ids = 0;
end

```

```

I = Ids;

```

GRAFICO 2.1 - Cap x Area. Ponto: real Cheia:interpolacao

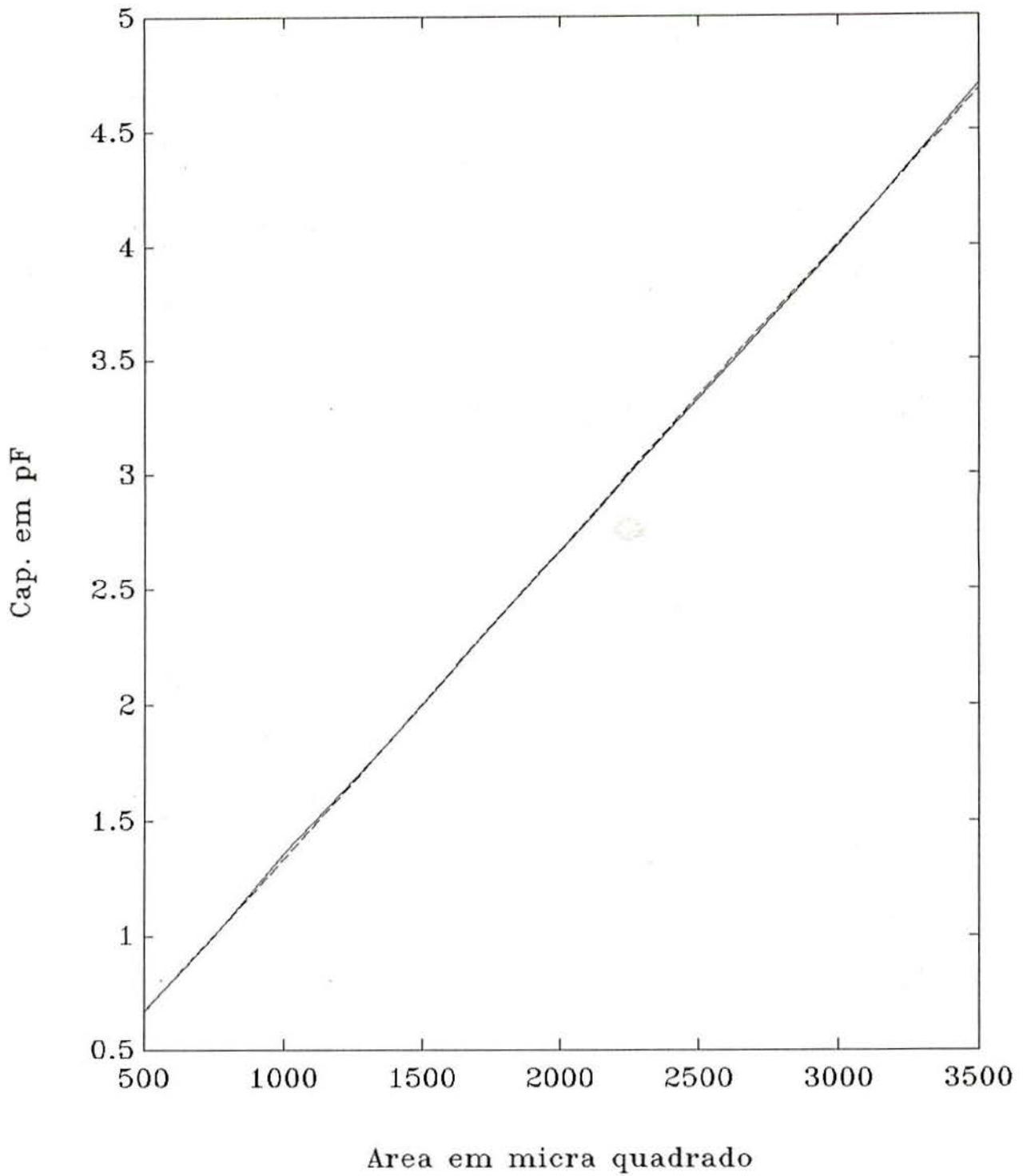


GRAFICO 2.2 - Capacitancia versus tempo

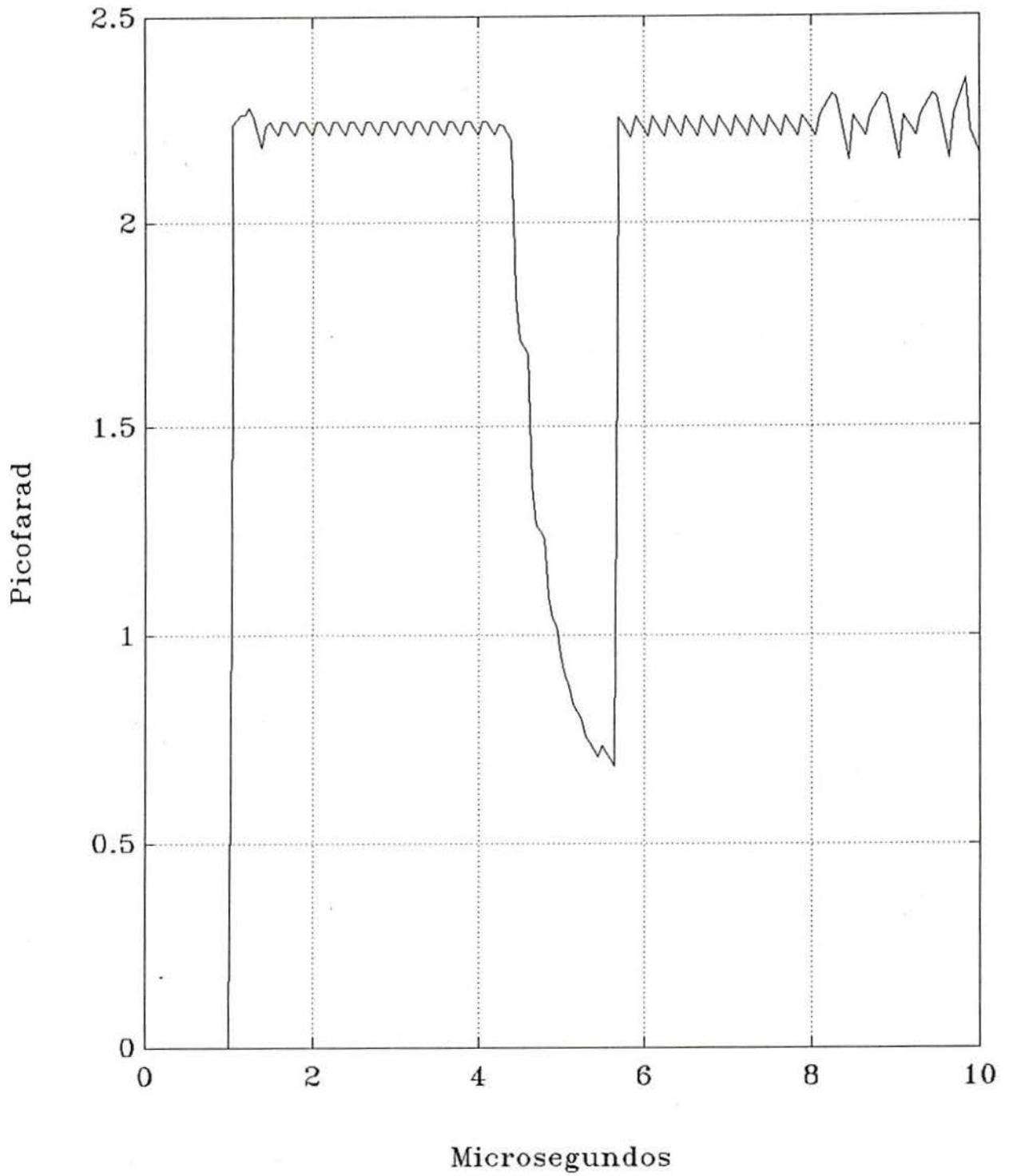


GRAFICO 2.3 - Capacitancia versus tensao

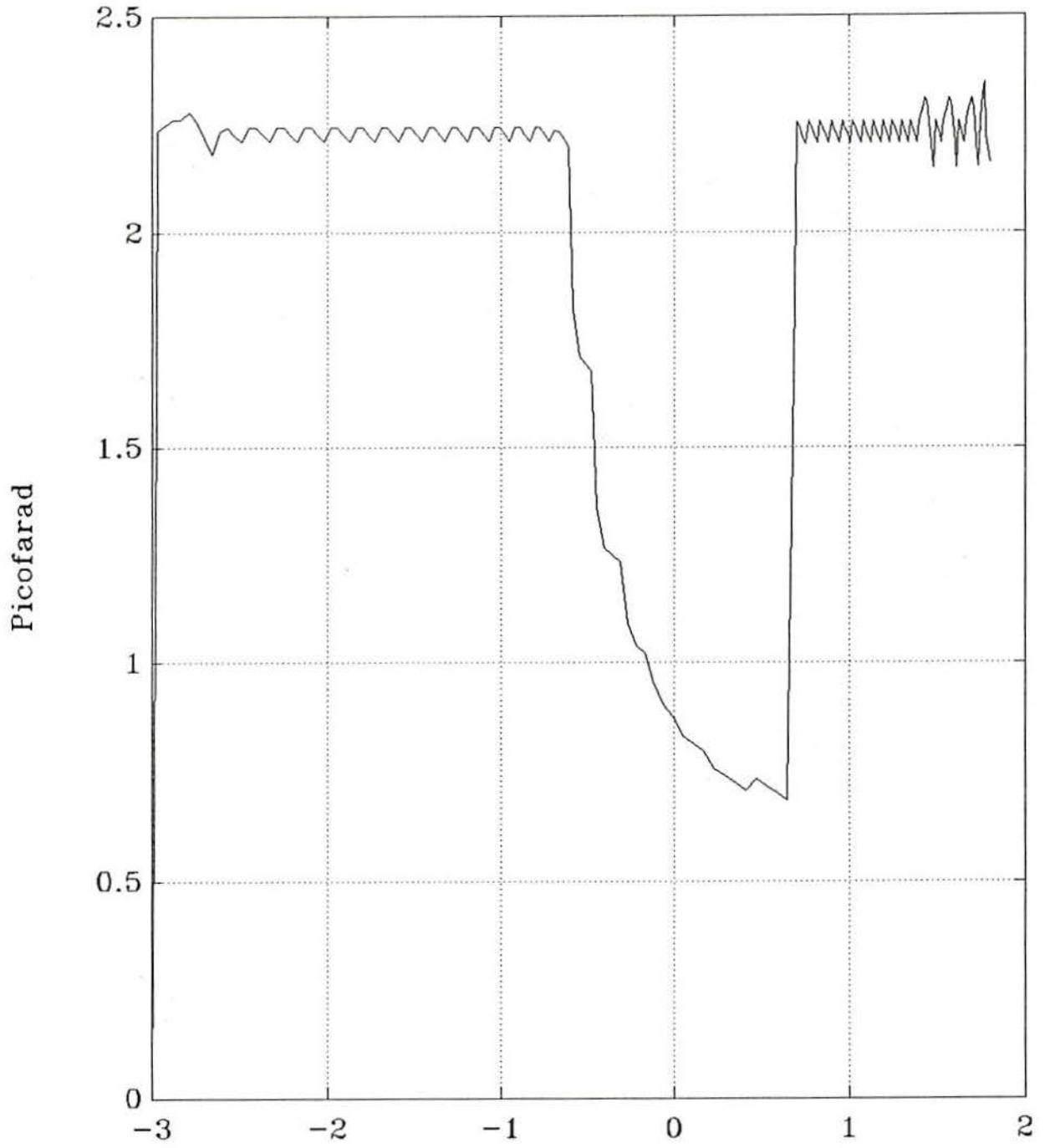


Grafico 2.4 - Associacao PMOS // NMOS

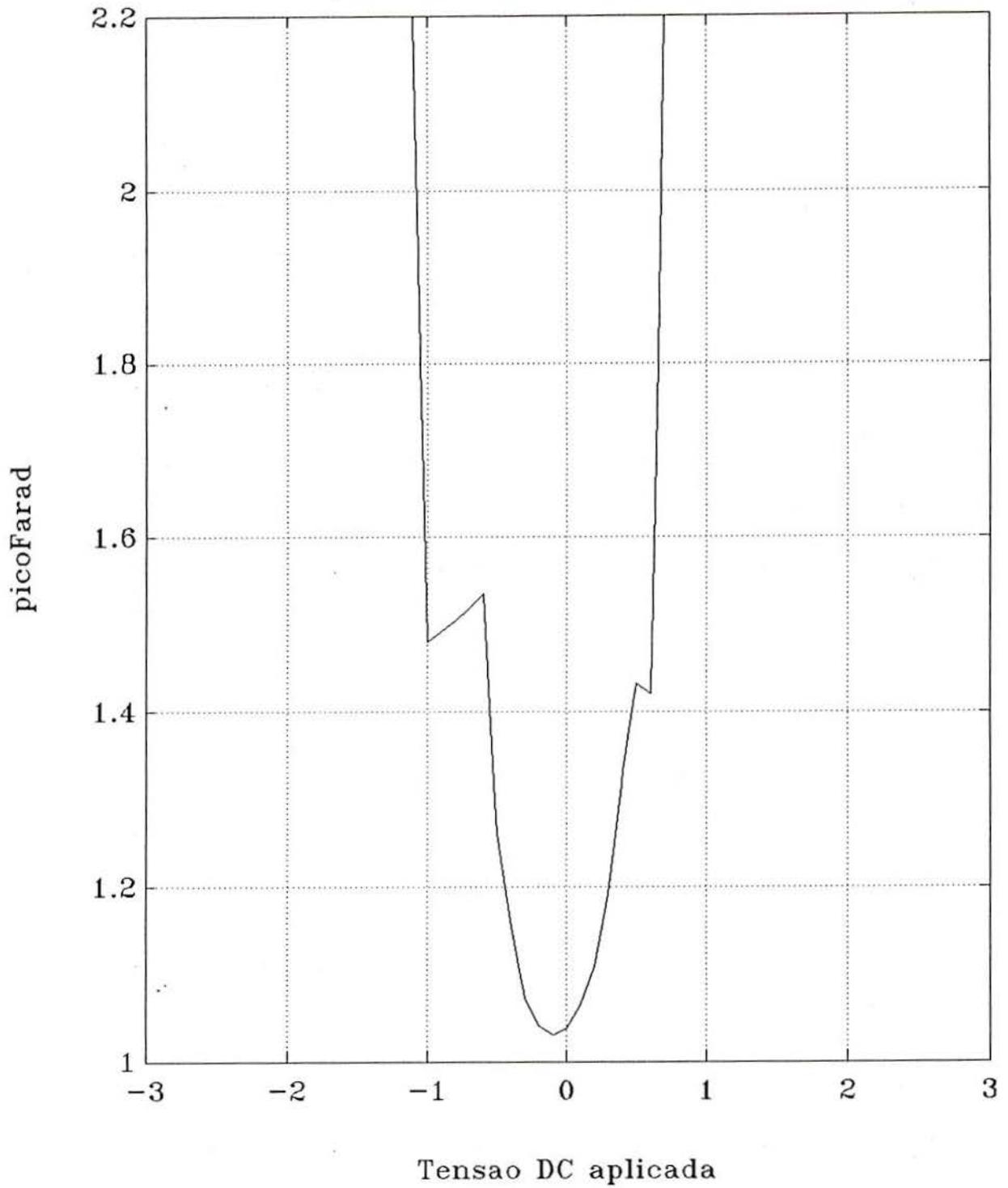


Grafico 2.5 - Associacao PMOS // NMOS invertido

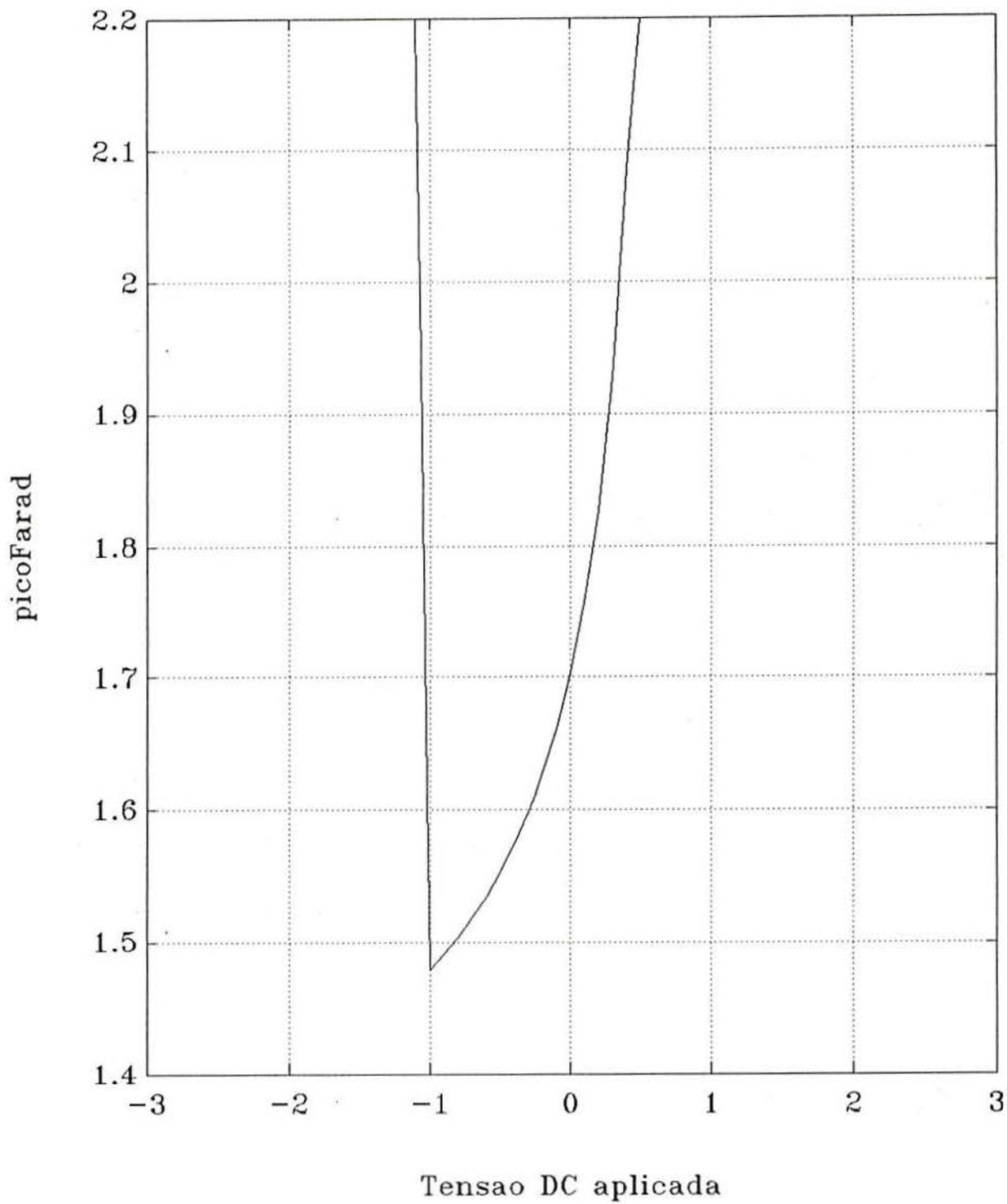


Grafico 2.6 - Associacao NMOS // NMOS invertido

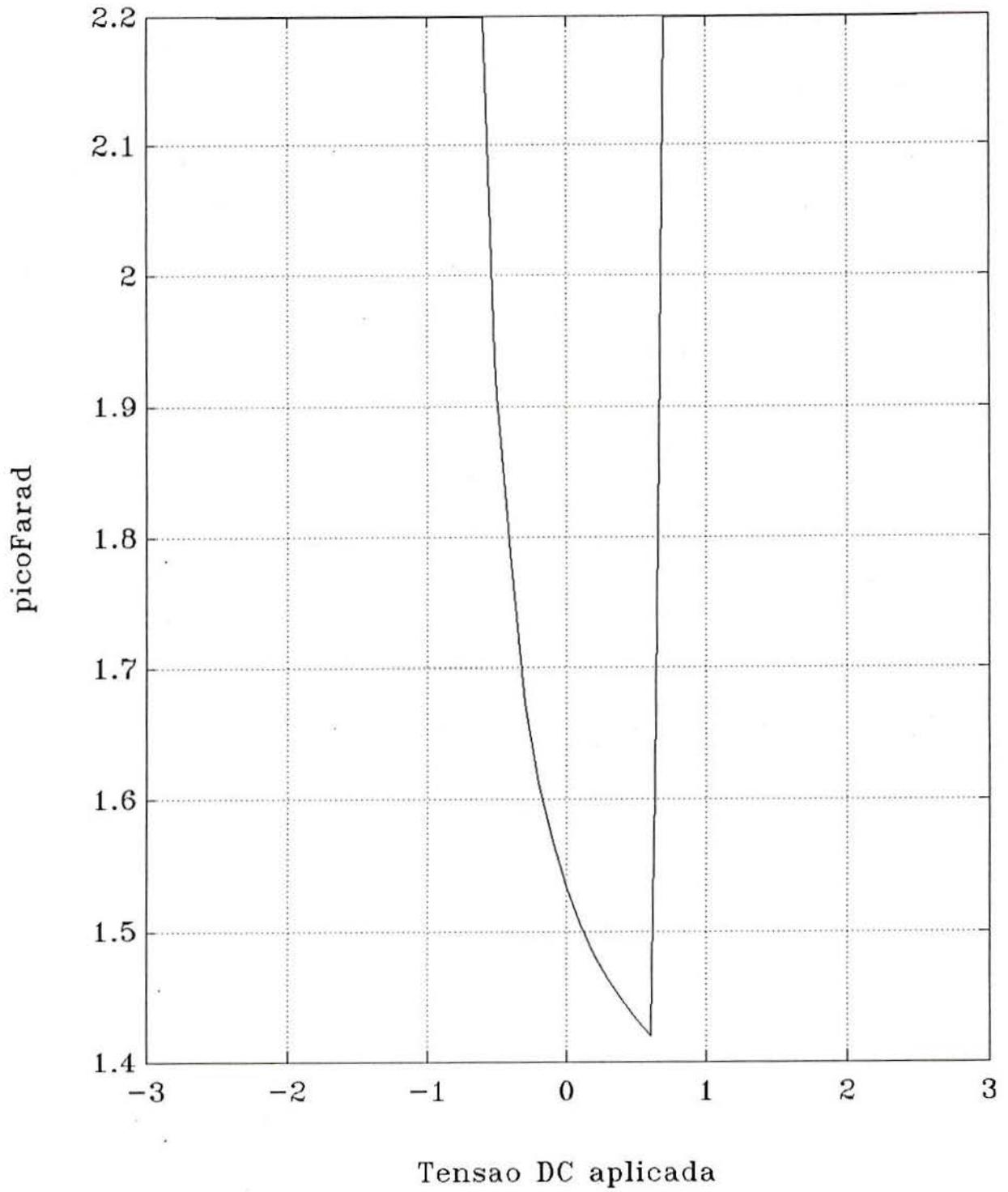


Grafico 2.7 - Associacao PMOS // PMOS invertido

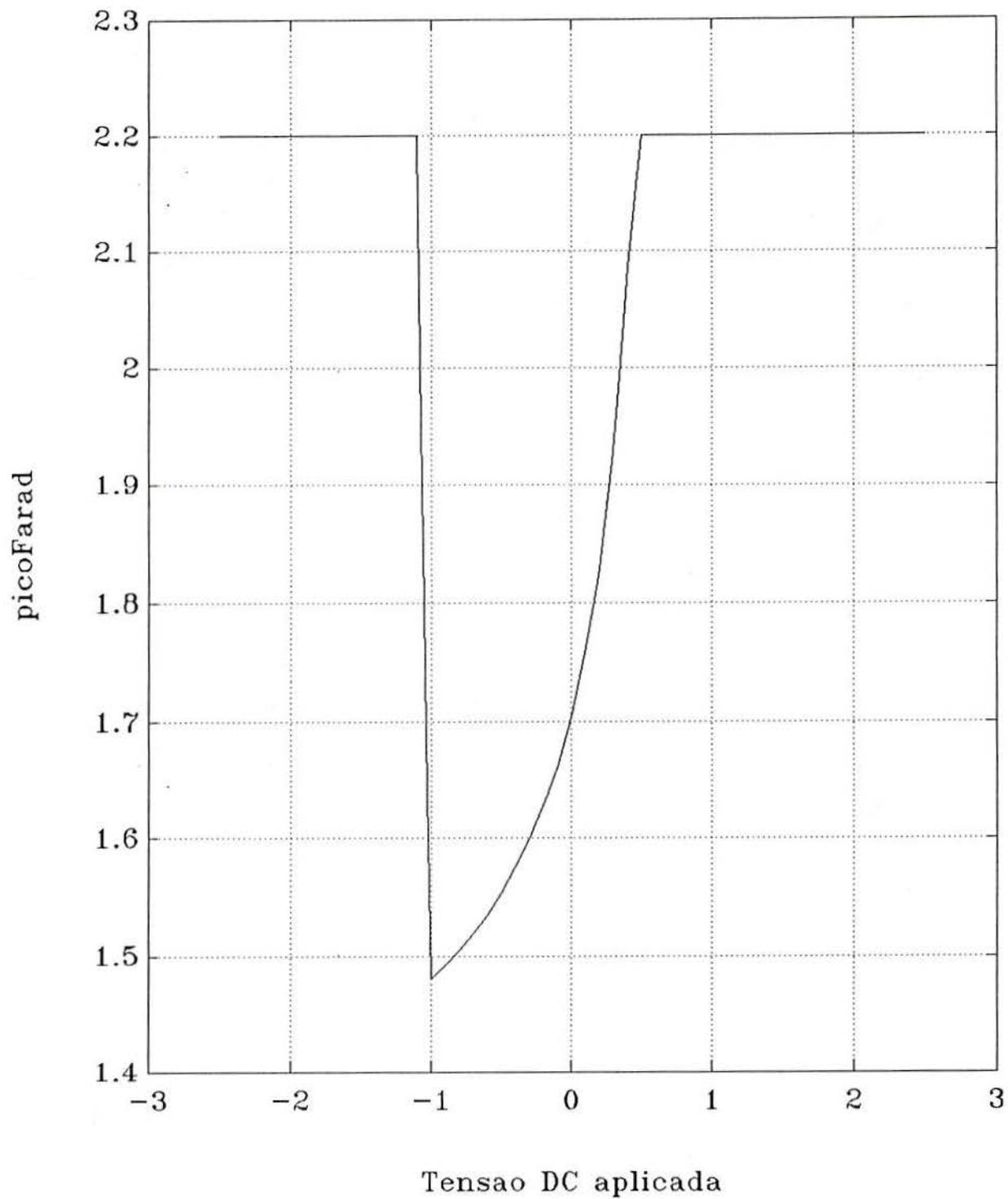


Grafico 2.8: Resposta AC dos dois amplificadores

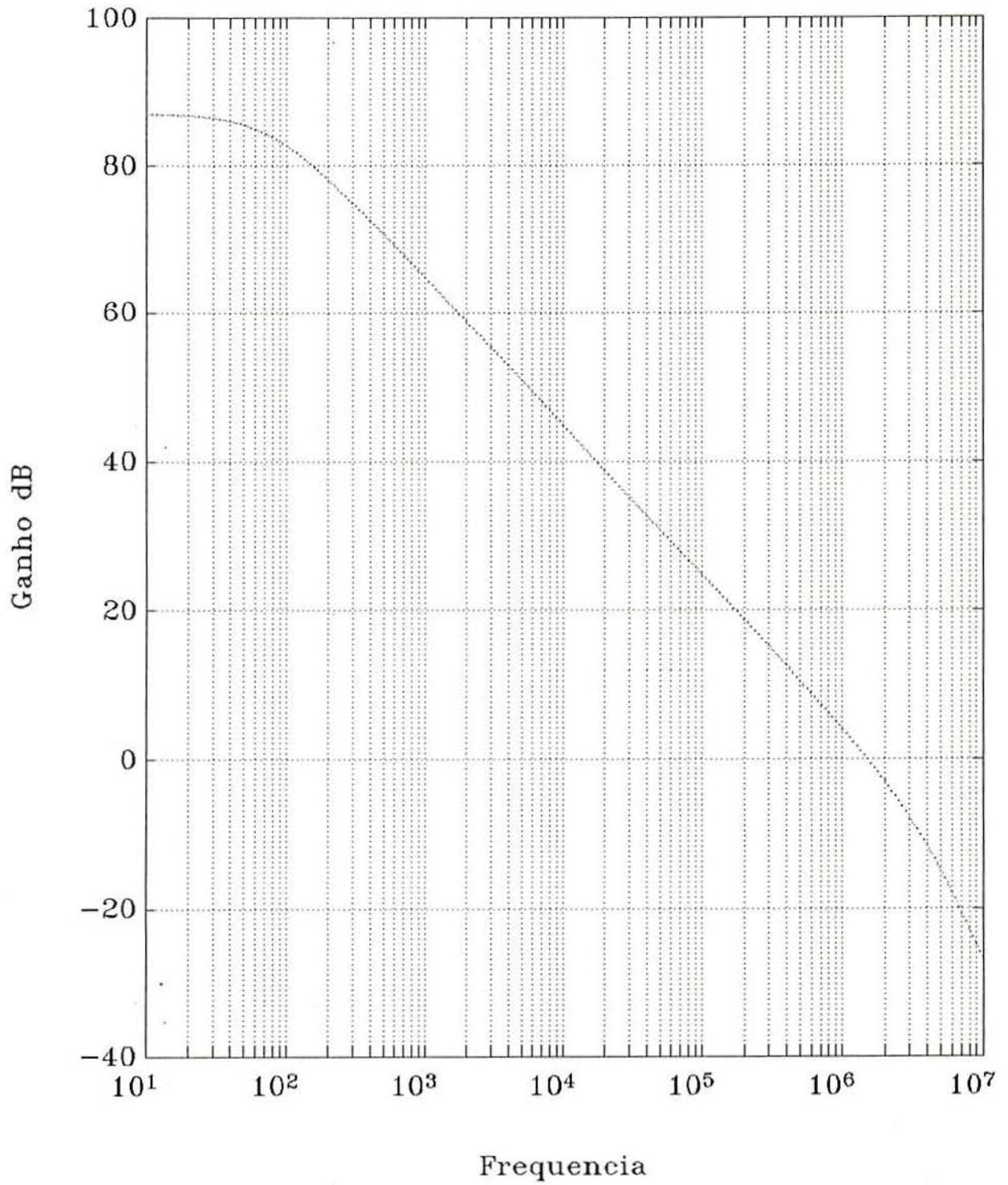
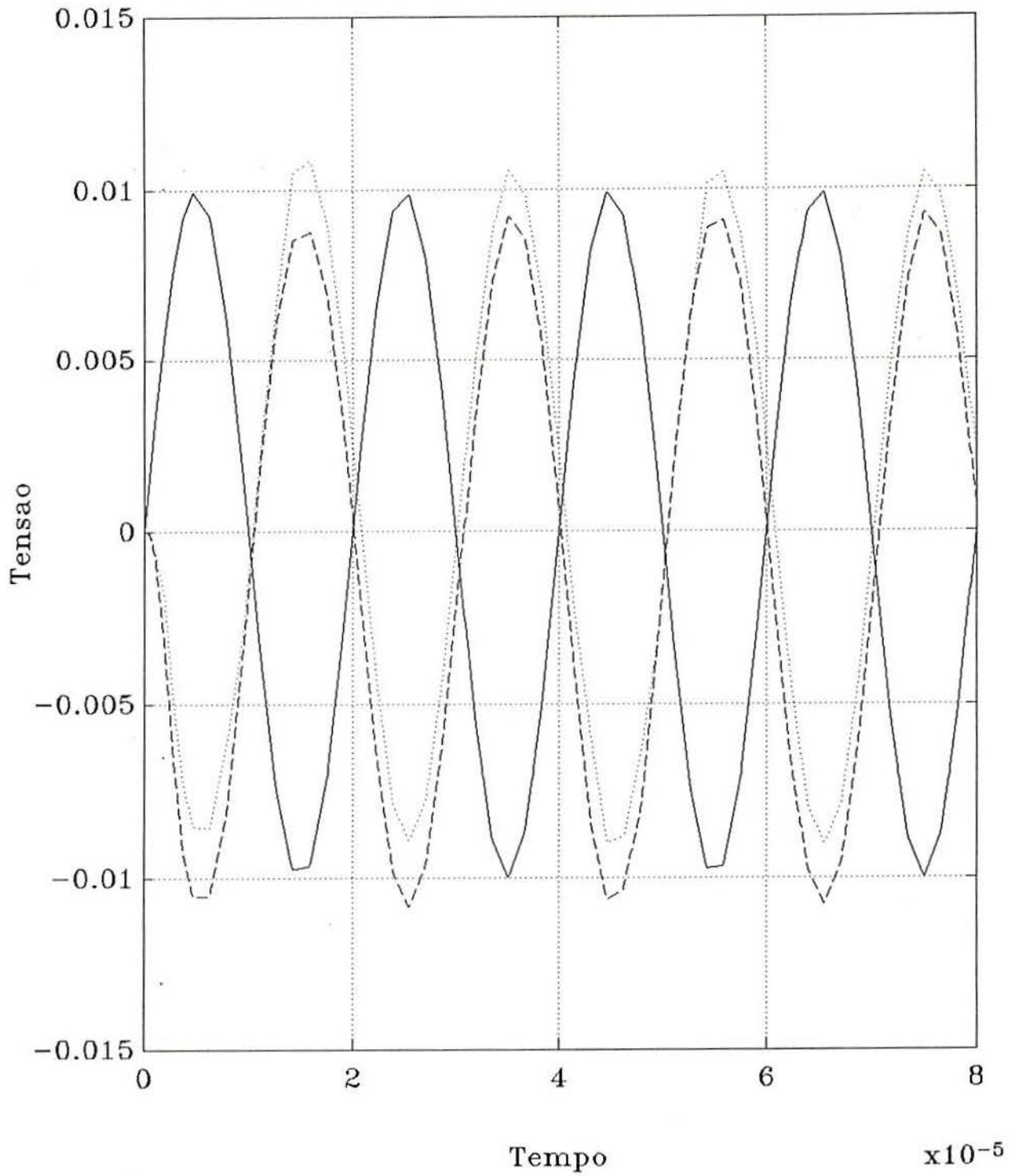


Grafico 2.9: Cheia=IN - Pontilhada=OUT1 - Tracejada=OUT2



$\times 10^{-4}$ GRAFICO 4.1 - Id: SPICE(c) e QUADS(p)

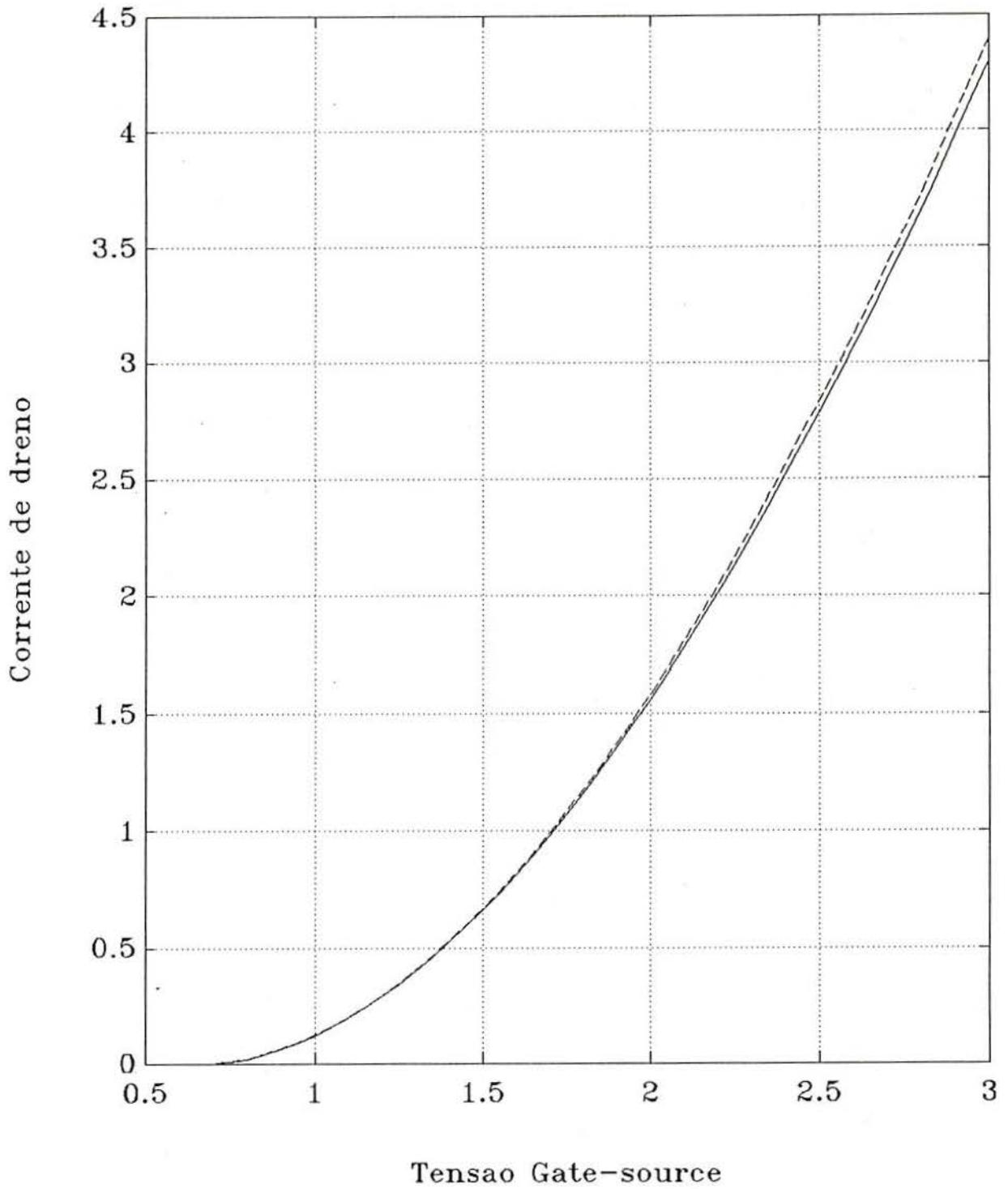
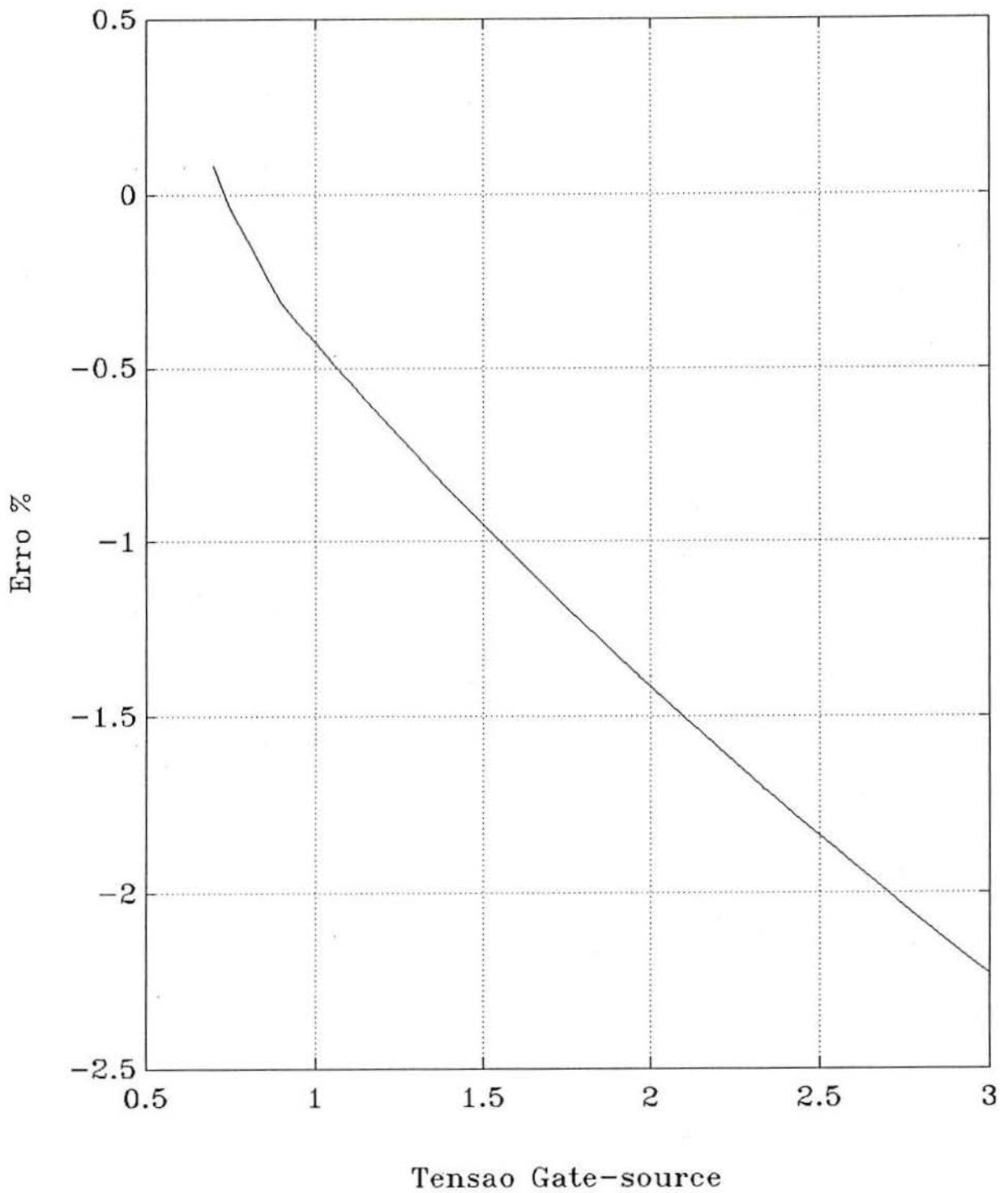


GRAFICO 4.2 - Erro percentual de corrente



$\times 10^{-5}$ GRAFICO 4.3 - Id: SPICE(c) e QUADS(p)

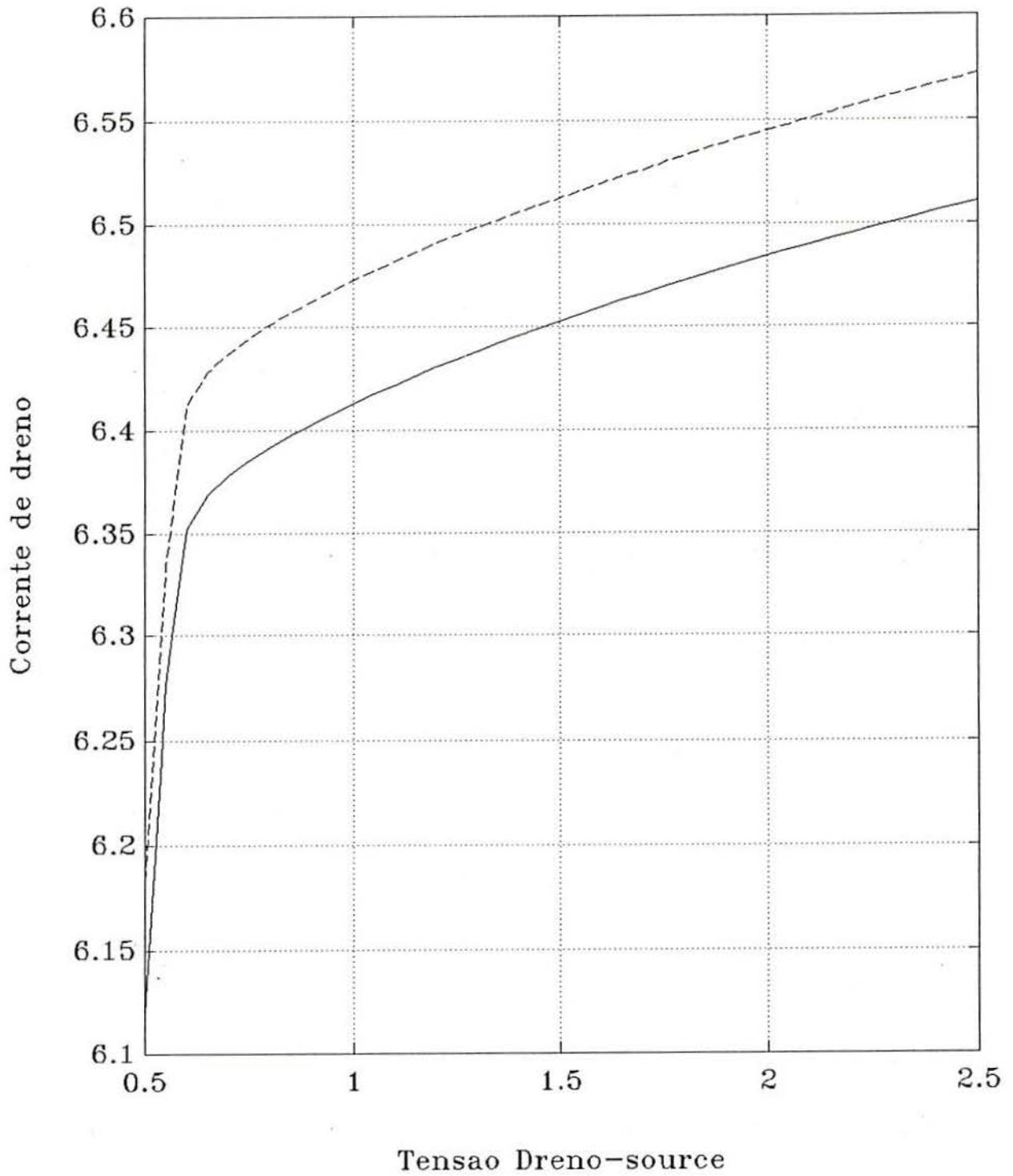
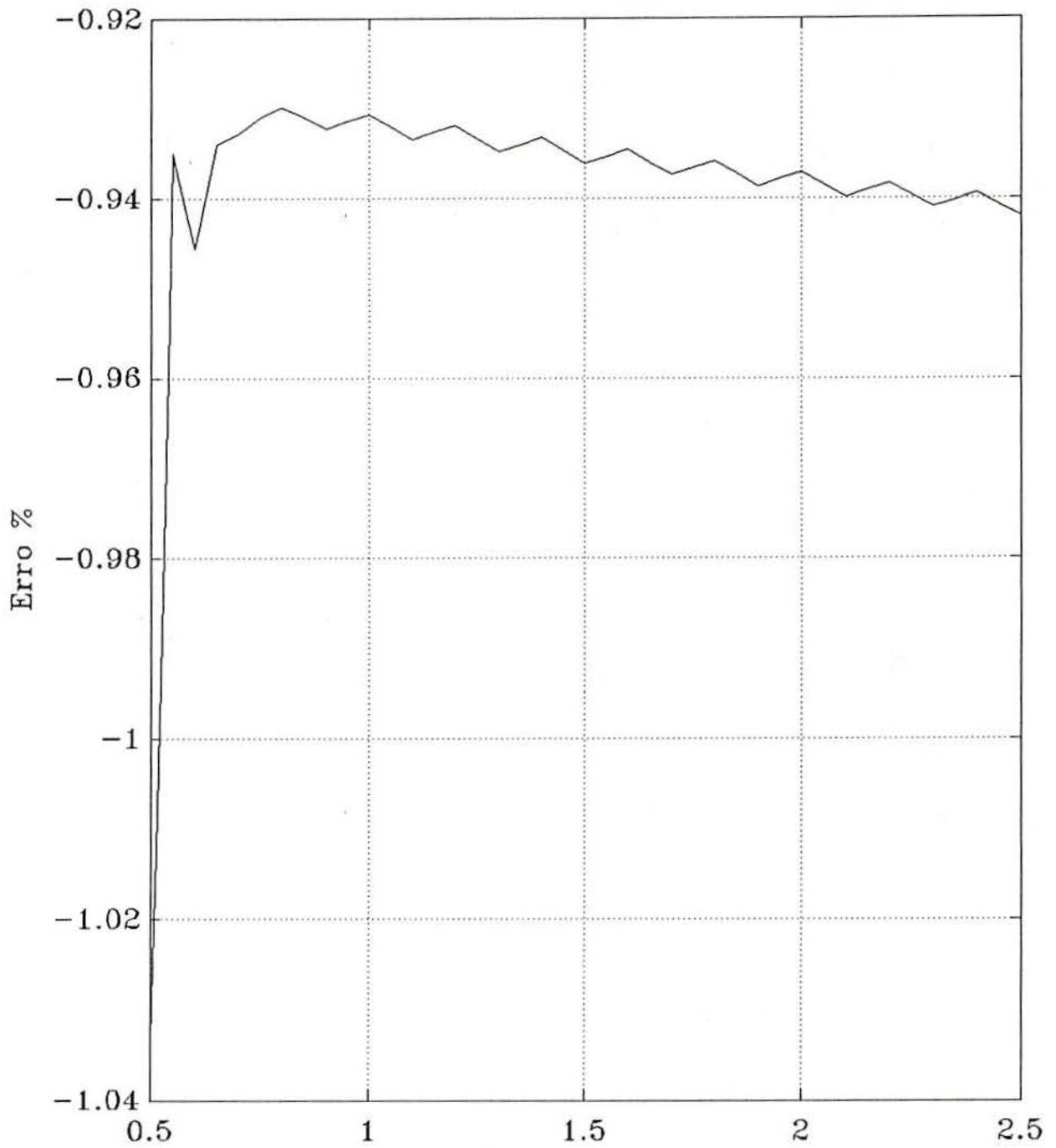


GRAFICO 4.4 - Erro percentual de corrente



Tensao Dreno-source

$\times 10^{-4}$ GRAFICO 4.5 - gm: SPICE(c) e QUADS(p)

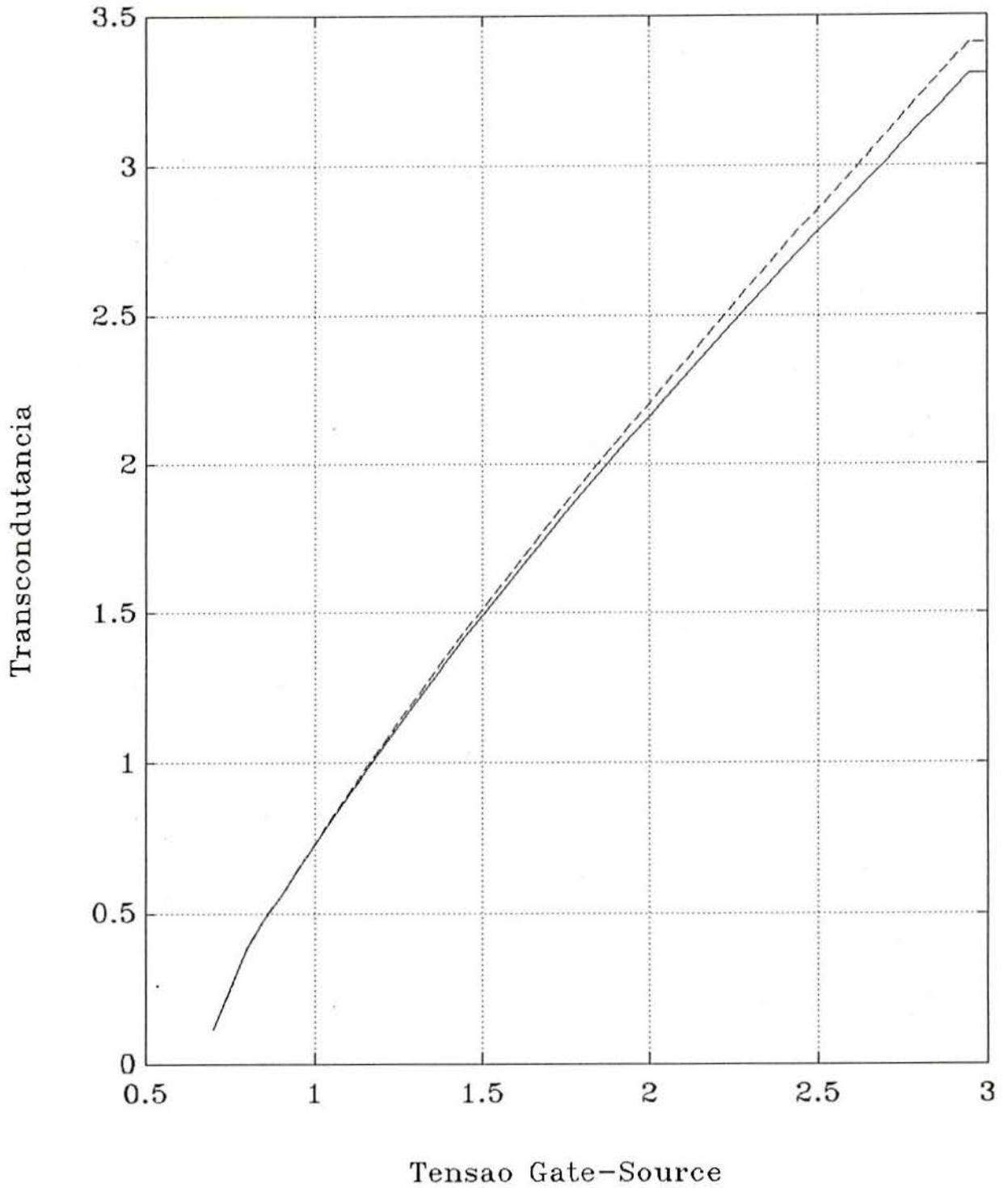
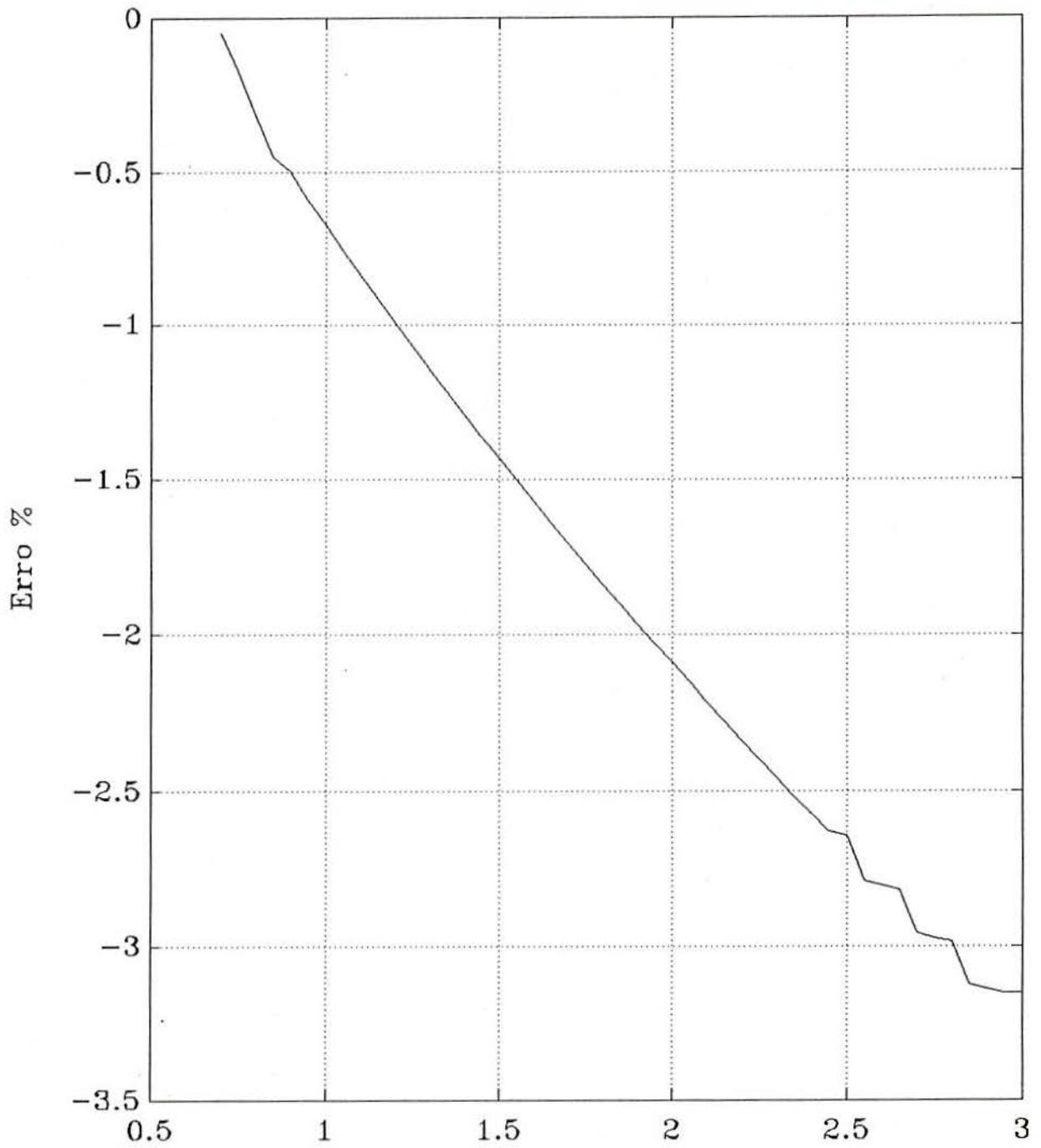


GRAFICO 4.6 - Erro percentual de transcondutancia



Tensao Dreno-source

$\times 10^{-5}$ GRAFICO 4.7 - go: SPICE(c) e QUADS(p)

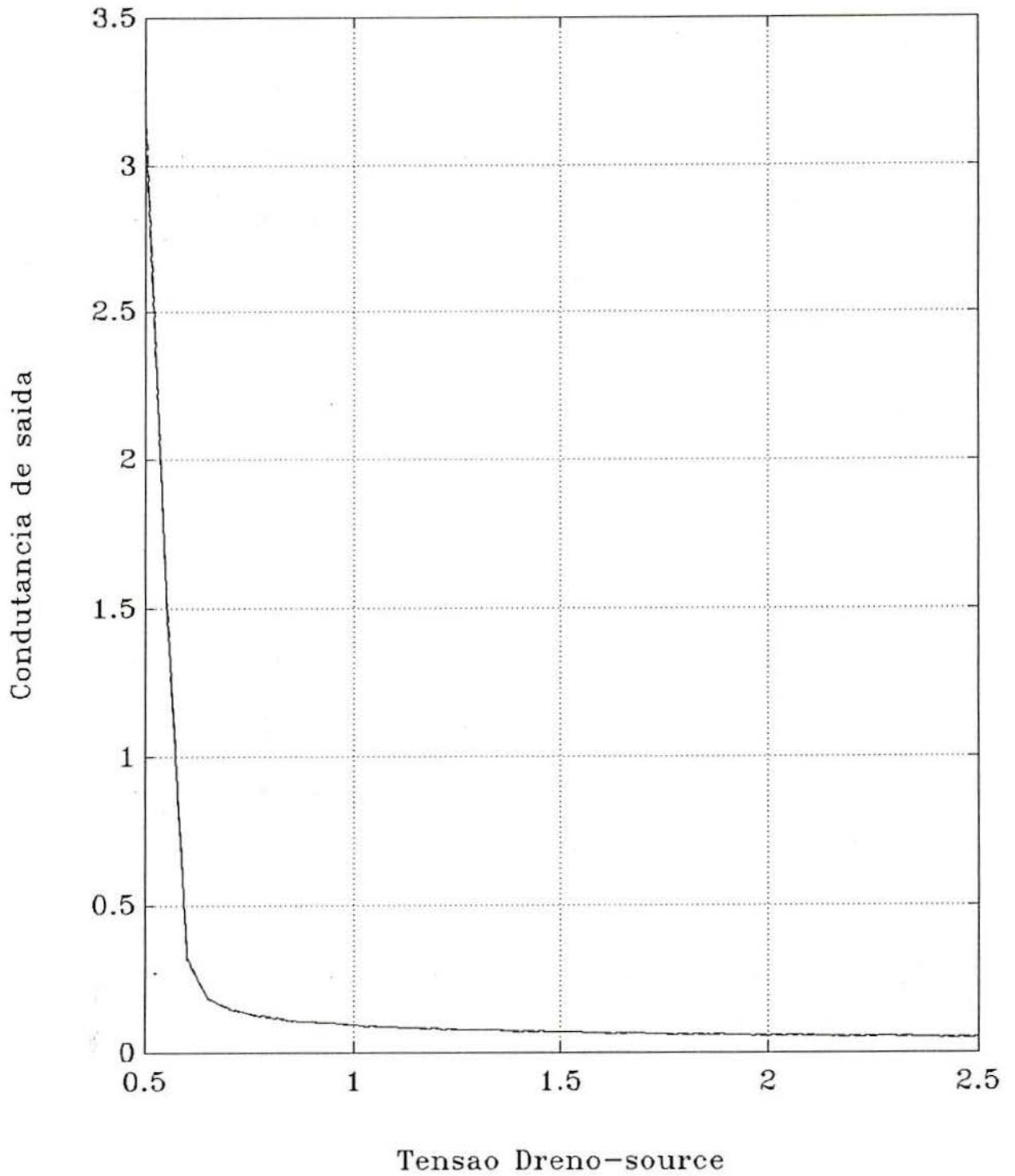


GRAFICO 4.8 - Erro percentual de condutancia

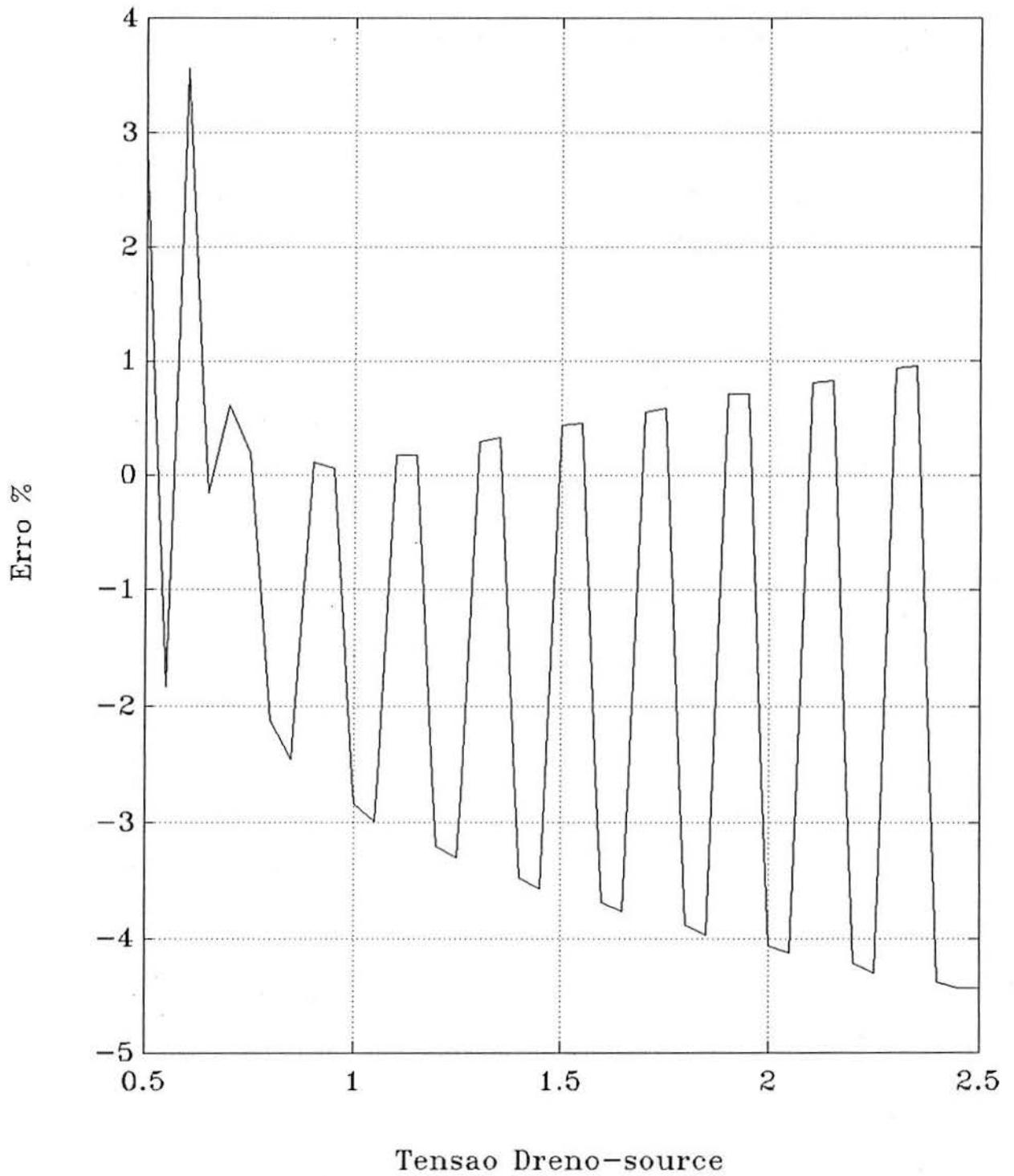
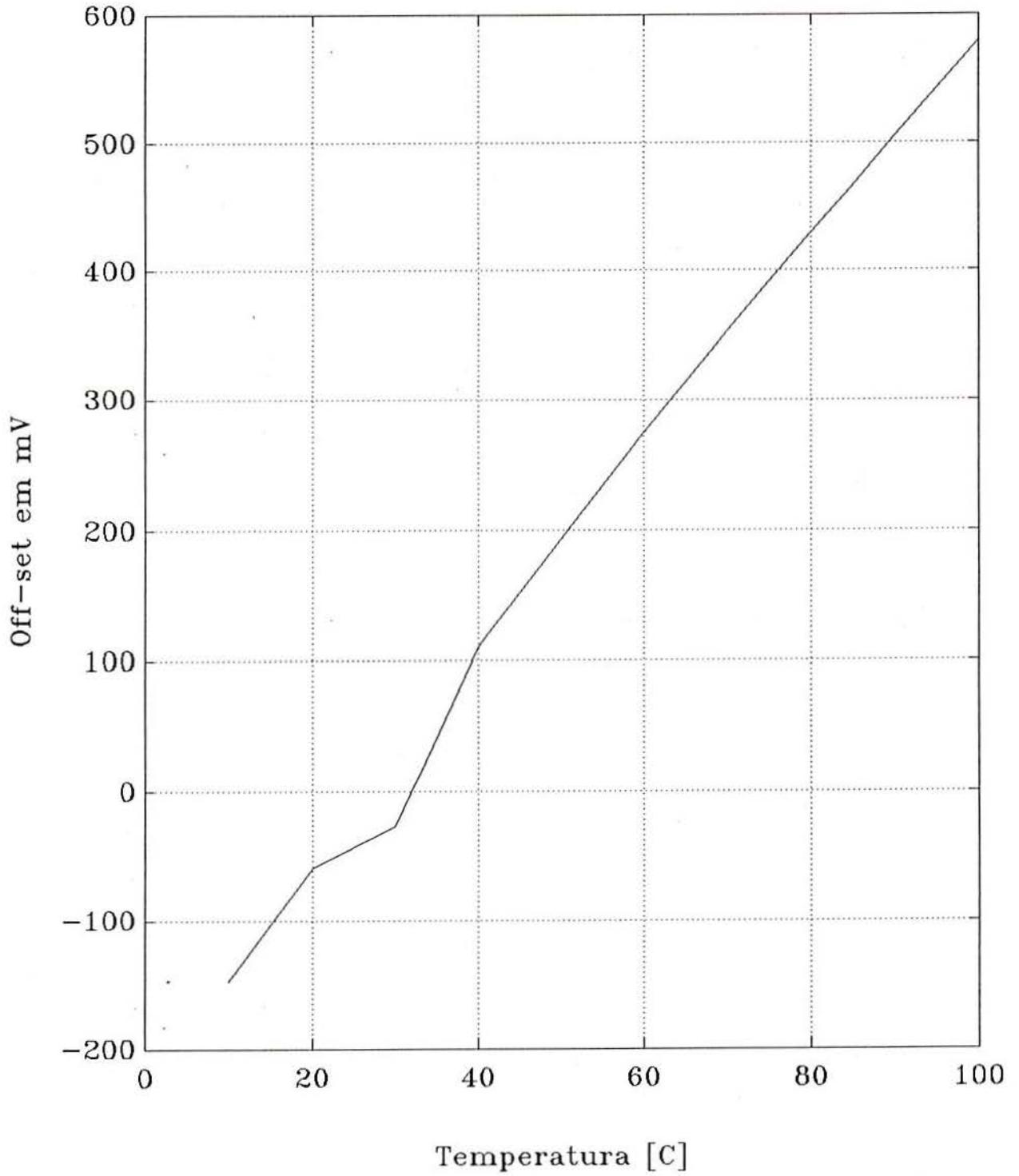


GRAFICO 5 - Analise termica de OFF-SET



$\times 10^4$ GRAFICO 6.1 - Area total versus k

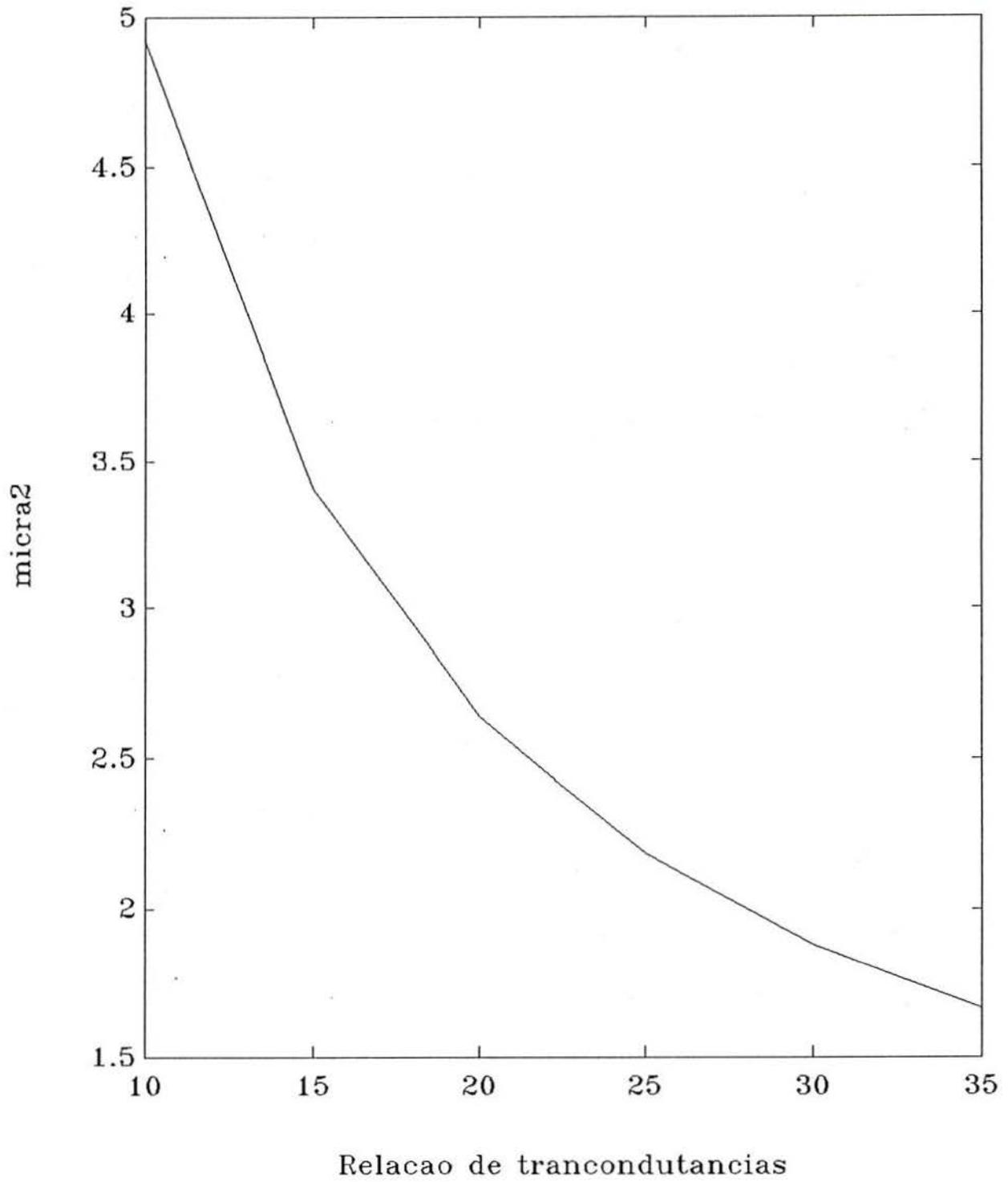


GRAFICO 6.2 - Area de transistor versus k

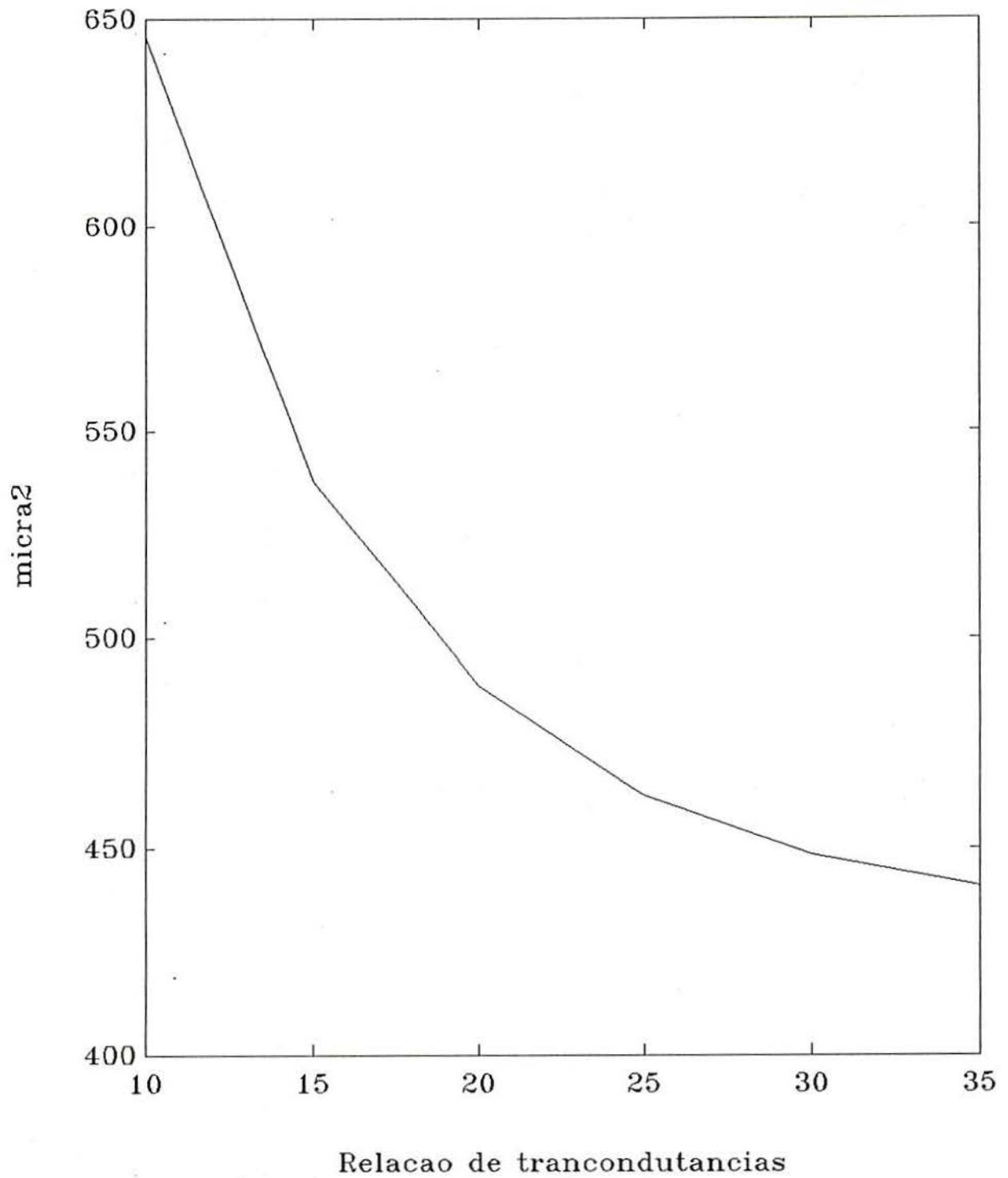


GRAFICO 6.3 - Capacitancia de compensacao versus k

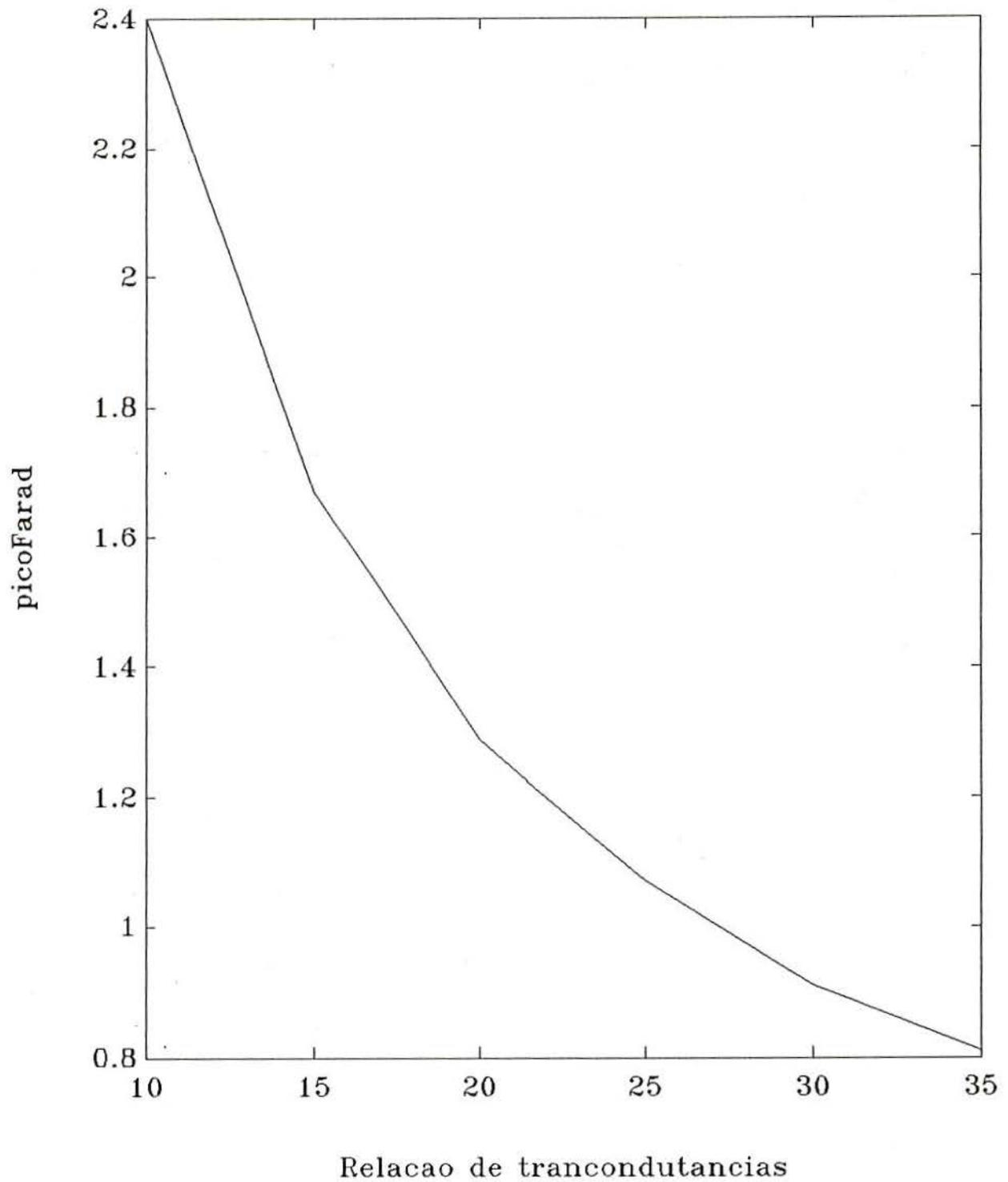


GRAFICO 6.4 - Corrente do prim e seg estagios e total

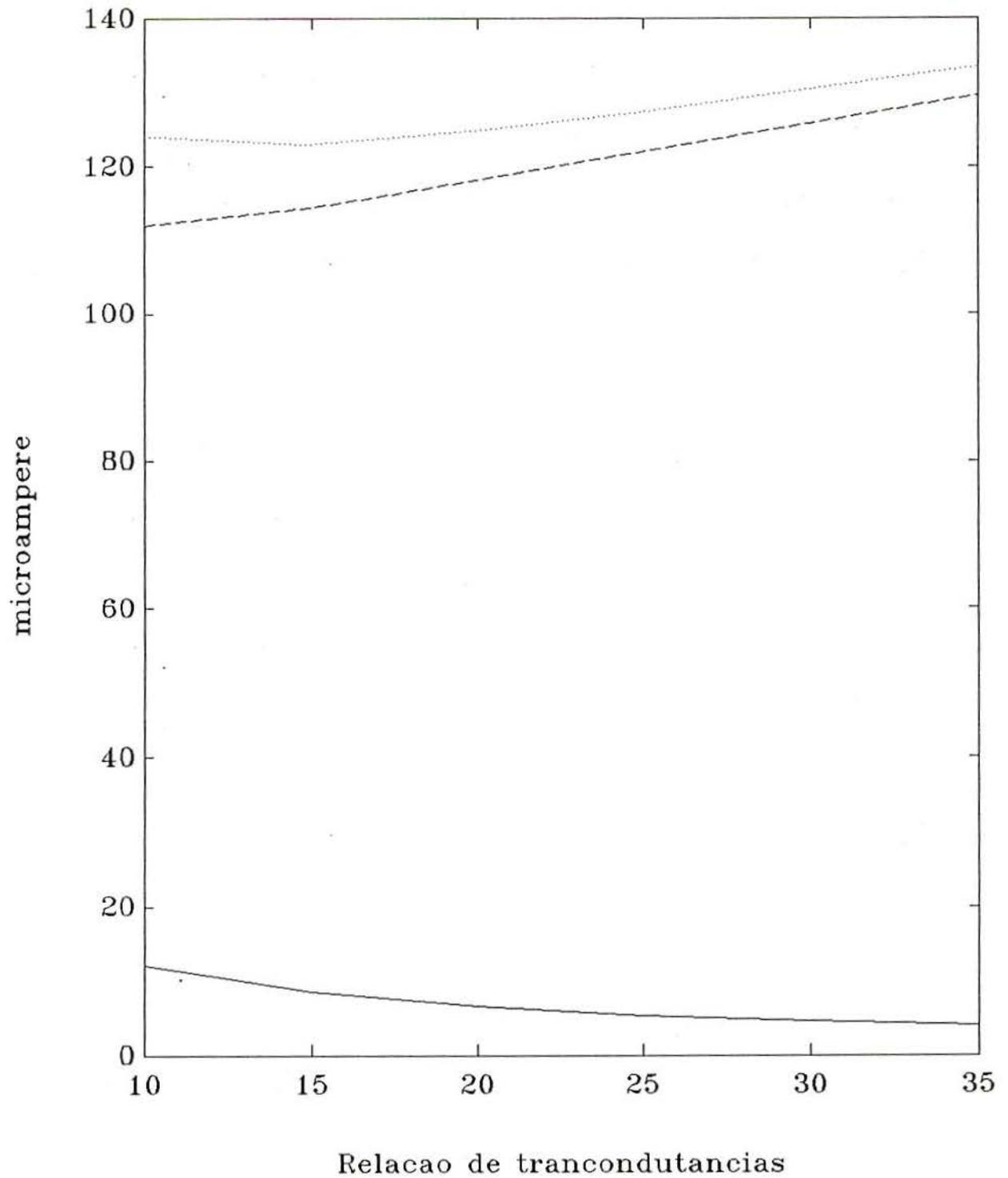


GRAFICO 8.2 - Erro % de corrente devido a particao

