Universidade Estadual de Campinas (UNICAMP)
Curso de Pós-Graduação em Engenharia de Telecomunicações

ANÁLISE DO BALANÇO HARMÔNICO MULTI-NÍVEIS PARA CIRCUITOS DE RF NÃO-LINEARES EM GRANDE-ESCALA VIA OS MÉTODOS DE NEWTON-KRYLOV E DO TENSOR-KRYLOV

Oswaldo Pedreira Paixão

Tese apresentada ao Curso de Pós-Graduação em Engenharia Elétrica e de Computação, do Departamento de Microonda e Óptica da Universidade Estadual de Campinas, para obtenção do grau de Doutor em Engenharia de Telecomunicações, com Área de Concentração em Engenharia de Telecomunicações.

Supervisor: Prof. Dr. Hugo Enrique Hernandez Figueroa

Banca Examinadora:

Prof. Dr. Silvio Ernesto Barbin

Prof. Dr. João Roberto Moreira Neto

Prof. Dr. Leonardo Lorenzo Bravo Roger

Prof. Dr. Aldário Chrestani Bordonalli

Campinas

Junho de 2009

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

Paixão, Oswaldo Pedreira
P167a Análise do balanço harmônico multi-níveis para circuitos de RF não-lineares em grande-escala via os métodos de Newton-Krylov e do tensor-Krylov / Oswaldo Pedreira Paixão. --Campinas, SP: [s.n.], 2009.

Orientador: Hugo Enrique Hernández Figueroa.
Tese de Doutorado - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Espaço de estado. 2. Dispositivos semicondutores. 3. Circuitos elétricos não-lineares. 4. Fourier, Analise de. 5. Projeto auxiliado por computador. I. Figueroa, Hugo Enrique Hernández. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Título em Inglês: Multilevel harmonic balance analysis of large-scale nonlinear RF circuits via Newton-Krylov and tensor-Krylov methods
Palavras-chave em Inglês: Space-times, Semiconductor devices, Nonlinear electric circuits, Fourier analysis, Computer-assisted design
Área de concentração: Engenharia de Telecomunicações
Titulação: Doutor em Engenharia Elétrica
Banca examinadora: Silvio Ernesto Barbin, João Roberto Moreira Neto, Aldário Chrestani Bordonalli, Leonardo Lorenzo Bravo Roger
Data da defesa: 15/06/2009
Programa de Pós Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE DOUTORADO

Candidato: Oswaldo Pedreira Paixão

Data da Defesa: 15 de junho de 2009

Título da Tese: "Análise do Balanço Harmônico Multi-níveis para Circuitos de RF Não-lineares em Grande-escala Via os Métodos de Newton-Krylov e do Tensor-Krylov"

Prof. Dr. Hugo Enrique Hernandez Figueroa (Presidente): _____

Prof. Dr. Silvio Ernesto Barbin: _____

Dr. João Roberto Moreira Neto: _____

Prof. Dr. Leonardo Lorenzo Bravo Roger: _____

Prof. Dr. Aldário Chrestani Bordonalli: _____

Sumário

"Mantenha simples: o mais simples possível, mas não simplório."
- A. Einstein

Este trabalho, tem como objetivo o desenvolvimento de novas técnicas, para análise de regime permanente não-autonoma de circuitos de alta-velocidade não-lineares em grande-escala. Para tal, é proposto um novo método do balanço harmônico (BH) fundamentado em uma eficiente metodologia de decomposição multi-níveis, que subdivide um circuito não-linear em grande-escala em uma estrutura hierarquica de super-redes (SuRs) esparsamente interconectadas. Mais precisamente, em cada nível de hierarquia, o circuito é composto por SuRs intermediárias, SuRs de fundo, e redes de conexão (RCs). As SuRs de fundo são decompostas em um aglomerado de sub-redes não-lineares (SRNs) correspondendo a dispositivos semicondutores, que por sua vez, estão envolvidos por uma sub-rede linear (SRL). A equação de estado e de sonda das SuRs de fundo foram obtidas utilizando uma nova metodologia que combina a formulação de espaço de estado (FEE) para as SRNs com a formulação nodal modificada (FNM) para a SRL. Esta metodologia FEE/FNM produz um sistema quadrado de equações com menor tamanho possível. Para realização das conversões do sinal entre os domínios do tempo e da frequência, foram discutidas e implementadas diferentes transformadas de Fourier discreta (TFDs), para operação em regime multi-tons, incluindo sinais com modulação digital. A equação determinante do BH multi-níveis do circuito assume uma estrutura hierarquica do tipo bloco diagonal com borda , que pode ser eficientemente resolvida utilizando técnicas de processamento paralelo. A matriz jacobiana de cada SuR de fundo é processada utilizando eficientes técnicas de matrizes esparsas, junto com o conceito de espectro de derivada. Para a solução da equação determinante, foram utilizados os métodos de Newton e do tensor para problemas de pequena- e média-escala, e os métodos de Newton inexato e do tensor inexato para problemas em grande-escala. A globalização via pesquisa-em-linha com retrocedimento, foi adotada para nestes solucionadores não-lineares. Entretanto, para o método do tensor e do tensor inexato, também foi adotada a técnica de pesquisa-em-linha curvilinear. Nos métodos inexatos, técnicas de pré-condicionamento foram utilizadas, para aumentar a eficiência e a robustez do solucionador linear iterativo em subespaço de Krylov (GMRES, GMRES-Bt e TGMRES-Bt). Finalmente, a formulação proposta foi validada e a eficiência do método do tensor e do tensor inexato comparada com o método de Newton e de Newton inexato, para diferentes topologias de circuitos utilizando diodos, FETs e HBTs, e operando sob diferentes regimes de excitação multi-tons.

Abstract

"Keep it simple: as simple as possible, but no simpler."
- A. Einstein

This work deals with the development of new techniques for nonautonomous nonlinear steady-state analysis of high-speed large-scale integrated circuits. To this end, it is proposed a novel harmonic balance (HB) method fundamented on a efficient multi-level decomposition methodology, that divides a large-scale circuit into hierarchical structure of sparsely interconnected supernetworks (SuNs). More precisely, the circuit is composed by intermediary SuRs, bottom SuRs and connection networks (CNs). The bottom SuNs are decomposed into a cluster of nonlinear subnetworks (NSNs) corresponding to the opto-electronic semiconductor devices, which in turn, are embedded by a linear subnetwork (LSN). Multi-port elements can be included in the LSN, in order to use measured data or results from electromagnetic analysis of structures with complex geometries. The formulation of the bottom SuN state and probe equations uses an improved table-oriented state-space formulation (SSF), that produces a square system with the lowest possible size, which is equal to the number of nonlinear state-variables (branch voltages and currents) that act as argument of the fuctions representing the semiconductor devices nonlinearities. The SSF is compared with the classical modified nodal formulation (MNF). For dealing with signal time-frequency conversions, discrete Fourier transform (DFT) techniques for different multi-tone regimes are discussed, including complex digitally modulated signals. The multi-level HB determining equation of the circuit assumes a hierarchical block bordered structure that can be efficiently tackled by parallel processing techniques. The HB jacobian matrix is handled using efficient sparse matrix techniques with a proper definition of the derivatives spectra. For the solution of a large-size HB problem, we investigated the applications of inexact tensor method based on Krylov-subspace techniques. Preconditioning are used to improve the robustness of the iterative tensor solver. To determine the circuit DC regime, we employ the tensor method. We adopted the backtracking linesearch technique as a globalisation strategy. However, for the tensor method, in particular, a curvilinear linesearch was also implemented. Finally, the formulation was validated and, the tensor and inexact tensor method efficiency was compared with the Newton and inexact Newton method, respectively, for several different circuits using diodos, FETs and HBTs, and operating under different multi-tone regimes.

Acrogramas

AABR	Antena + Amplificador de Baixo-Ruído
ABR	Amplificador de Baixo Ruído
ACC	Amplificador Classe-C
AGV	Amplificador de Ganho Variável
AP	Amplificador de Potência
APC	Amplificador de Potência Corporativo
BAD	*Buffer* + Acoplador Direcional
BAFC	Balanço de Amostra Fundamentado-em-Convolução
BCF	*Buffer* do Conversor de Frequência
BE	Balanço Espectral
BF	Balanço de Forma-de-onda
BFM	Balanço de Forma-de-onda Modificado
BH	Balanço Harmônico
BHM	Balanço Harmônico Modificado
BM	Buraco Metalizado
BTF	*Buffer* do Triplicador de Frequência
CA	Corrente Alternada
CC	Corrente Contínua
CEE	Circuito Elétrico Equivalente
CIRF	Circuito Integrado de Radio frequência
CF	Conversor (ou Conversão) de Frequência
CFD	Conversão de Frequência de Descida
CFR	Conversão de Frequência Resistivo
CFRB	Conversão de Frequência Resistivo Balanceado
CFS	Conversão de Frequência de Subida
CG	Célula de *Gilbert*
DBF	Dispersão de Baixa Frequência
DD	Decomposição-de-Domínio
DDS	Dispositivo Definido-Simbolicamente
DF	Diferença Finita
DFDT	Diferença Finita no Domínio do Tempo
DIM	Distorção por Intermodulação
EC	Elemento de Contorno
EDPM	Equação Diferencial Parcial Multi-Tempos
EDP	Equação Diferencial Parcial
EF	Elemento Finito
EM	Eletromagnética
EP	Elemento de Processamento
E/S	Entrada/Saída
FARMO	Fonte de Alimentação com Retificação de Meia-Onda
FAROC	Fonte de Alimentação com Retificação de Onda-Completa
FEE	Formulação Espaço-Estado
FEEP	Formulação Espaço-Estado Paramétrica
FI	Frequência Intermediária
FNM	Formulação de Nodal-Modificada

IDT	Integração no Domínio no Tempo
IF	Integração Finita
IM	Intermodulação
LIT	Linear Invariante-no-Tempo
LKC	Lei de *Kirchhoff* para corrente
LKT	Lei de Kirchhoff para tensão
LPTV	Linear Periódico Variante-no-Tempo
MAB	Multiplicador Analógico Balanceado
MAQQ	Multiplicador Analógico de Quatro-Quadrantes
MAQQ-BD	Multiplicador Analógico de Quatro-Quadrantes com Baixo-Deslocamento
MF	Multiplicador de Frequência
MFA	Mapeamento-de-Frequência Artificial
MLT	Matriz Linha de Transmissão
MP	Multi-Porta
NLR	Nó Local de Referência
NQE	Não-Quasi-Estático
OCT	Oscilador Controlado por Tensão
OL	Oscilador Local
RA	Ressoador Ativo
RAE	Rede de Alimentação Externa
RC	Rede de Conexão
RD	Razão de Desempenho
RF	Rádio Frequência
RMP	Rede Multi-Porta
SDF	Série Dupla de Fourier
SRL	Sub-Rede Linear
SRLA	Sub-Rede Linear Ampliada
SRN	Sub-Rede Não-Linear
SRNC	Sub-Rede Não-Linear Compactada
SuR	Super-Rede
QE	Quasi-Estático
TE	Transiente da Envoltória
TF	Triplicador de Frequência
TFD	Transformada de Fourier Discreta
TFDM	Transformada de Fourier Discreta Multi-Dimensional
TFMT	Transformada de Fourier Multi-Tons
TFQP	Transformada de Fourier Quasi-Periódica
TFR	Transformada de Fourier Rápida
TFRM	Transformada de Fourier Rápida Multi-Dimensional
UBB	Unidade Banda-Base
UCF	Unidade de Conversão de Frequência
UCP	Unidade Central de Processamento
UOL	Unidade de Oscilador Local
URF	Unidade de Rádio Frequência

Elementos básicos de circuito:

CAP	Capacitor
CAPN	Capacitor Não-linear
CC	Corrente de Controle

CCN	Corrente de Controle Não-linear
FC	Fonte de Corrente
FCC	Fonte de Corrente Controlada
FCCC	Fonte de Corrente Controlada por Corrente
FCCN	Fonte de Corrente Controlada Não-linear
FCCT	Fonte de Corrente Controlada por Tensão
FCE	Fonte de Corrente Externa
FT	Fonte de Tensão
FTC	Fonte de Tensão Controlada
FTCC	Fonte de Tensão Controlada por Corrente
FTCN	Fonte de Tensão Controlada Não-linear
FTCT	Fonte de Corrente Controlada por Corrente
FTE	Fonte de Tensão Externa
IND	Indutor
INDN	Indutor Não-linear
LT	Linha de Transmissão
LTA	Linha de Transmissão em Circuito Aberto
LTC	Linha de Transmissão em Curto-Circuito
RCT	Resistor Controlado por Tensão
RCC	Resistor Controlado por Corrente
RES	Resistor
SC	Sonda de Corrente
SCE	Sonda de Corrente Externa
ST	Sonda de Tensão
STE	Sonda de Tensão Externa
TC	Tensão de Controle
TCN	Tensão de Controle Não-linear

Acrogramas em inglês:

ACPR	Adjacent Channel Power Ratio
AM	Amplitude Modulation
BARITT	Barrier Injection Transit-Time
BE	Boundary Element
BiCG	Bi-Conjugate Gradient
BJT	Bipolar Junction Transistor
CAD	Computer Aided Design
CCPR	Cochannel Power Ratio
CFD	Computational Fluid Dynamics
CG	Conjugate Gradient
CGS	Conjugate Gradient Squared
CIR	Cochannel Interference Ratio
CN	Connection Network
CPU	Central Processing Unit
CSSB	Convolution-Based Sample Balance
DC	Direct Current
DFB	Distributed Feedback
DFT	Dicrete Fourier Transform
DH	Double Heterostructure
ET	Envelope Transient

FD	Finite Difference
FE	Finite Element
FET	Field-Effect Transistor
FFT	Fast Fourier Transform
FGMBACK	Flexible Generalized Minimal Backward Error
FGMRES	Flexible Generalized Minimal Residual
FI	Finite Integration
GMBACK	Generalized Minimal Backward Error
GMRES	Generalized Minimal Residual
GSM	*Gram-Schmidt* Modificada
HB	Harmonic Balance
HBT	Heterojunction Bipolar Transistor
HEMT	High-Electron Mobility Transistor
IM	Intermodulation
IMD	Intermodulation Distortion
IMPATT	Impact Ionization Avalanche Transit-Time
ISI	Inter-Symbol Interference
JFET	Junction Field-Effect Transistor
LAN	Local Area Network
LDMOS	Laterally Diffused Metal Oxide Semiconductor
LNA	Low-Noise Amplifier
LSN	Linear Sub-Network
MEMS	Microelectromechanical Systems
MESFET	Metal Semiconductor Field Effect Transistor
MGH	Moré/Garbow/Hillstrom
MHB	Modified Harmonic Balance
MIM	Monolithic Integrated Microwave
MMIC	Monolithic Microwave Integrated Circuit
MMID	Millimeter-Wave Identification
MNF	Modified-Nodal Formulation
MOSFET	Metal Oxide Semiconductor Field-Effect Transistor
MPDE	Multi-time Partial Differential Equation
MWB	Modified Waveform Balance
M^3IC	Monolithic Millimeter-wave Integrated Circuit
NRZ	Non-Return-to-Zero
NSN	Nonlinear Sub-Network
TGMRES	Tensor Generalized Minimal Residual
TMHB	Time-Mapped Harmonic Balance
NPR	Noise Power Ratio
OFDM	Orthogonal Frequency-Division Multiplexing
PAN	Personal Area Network
PAPR	Peak-to-Average Power Ratio
PHEMT	Pseudo-morphic HEMT
PIN	p-i-n (diode)
PM	Phase Modulation
QMR	Quasi-Minimal Residual
QPSK	Quadrature Phase-Shift Keying
RFID	Radio Frequency Identification
SB	Spectral Balance
SDD	Symbolically-Defined Device

SINAD	Signal-to-Noise and Distortion
SOR	Successive Over-Relaxation
SSF	State-Space Formulation
SSOR	Symmetric Successive Over-Relaxation
SuN	Super-Network
TLM	Transmission Line Matrix
TRAPATT	Trapped Plasma Avalanche Triggered Transit
VCSEL	Vertical Cavity Surface Emitting Laser
WB	Waveform Balance
WCDMA	Wide-band Code Division Multiple Access

Símbolos

$\hat{\jmath}$	Operador imaginário, $\hat{\jmath} = \sqrt{-1}$.
e	Número natural, $e = 2,71828\ldots$.
$\in, \wp, \approx$	Inclusão de elemento, inclusão de conjunto, união.
$\forall$	Para todo.
$\Re$	Somatória.
$\varnothing$	Elemento não definido.
Z, R, C	Números inteiros, reais, e complexos.
$\boldsymbol{x}^T, \boldsymbol{A}^T$	Transposta do vetor $\boldsymbol{x}$, transposta de $\boldsymbol{A}$.
$\|\Rightarrow\|_a, \|\Rightarrow\|_1, \|\Rightarrow\|_2, \|\Rightarrow\|_p, \|\Rightarrow\|_\infty$	Norma arbitrária, norma-M_1, norma-M_2, norma-M_p, e norma-M_∞ em R^n .
$e^x = \exp(x)$	Função exponencial de $x \in R$.
$\text{sinal}(x)$	Função sinal x , onde $\text{sinal}(x) = 1$, se $x \geq 0$, $\text{sinal}(x) = -1$ se $x < 0$.
$\log_b(x)$	Função logarítimo de $x \in R$ na base $b \in R$.
$\text{máx}(x, y), \text{mín}(x, y)$	Função máximo e mínimo, onde $\text{máx}(x, y) = x$, se $x \geq y$ e $\text{mín}(x, y) = x$, se $x \leq y$.
$\text{típ}(\boldsymbol{x})$	Vetor em R^n com valores típicos de $\boldsymbol{x} \in R^n$.
$\dim(\boldsymbol{x})$	Dimensão do vetor $\boldsymbol{x}$.
$\mathbf{diag}[\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_{n_{SRN}}]$	Matriz bloco diagonal.
t, ω	Tempo, frequência angular.
$\mathfrak{L}$	SRL.
$\mathfrak{L}_a$	SRLA, $\mathfrak{L} \subset \mathfrak{L}_a$.
$\mathfrak{N}_k$	k -ésima SRN.
$\mathfrak{N}_c$	SRNC, $\mathfrak{N}_k \subset \mathfrak{N}_c$ para $k \in [1, n_{SRN}]$.
$\mathfrak{C}_{\upsilon, m}$	m -ésima rede de conexão de nível υ .
$\mathfrak{S}_{\upsilon, k}$	k -ésima super-rede de nível υ .
υ	Índice de nível da hieraquia.
$n_{SuR}^{(\upsilon)}, n_{RC}^{(\upsilon)}$	Número de SuRs e de RCs no nível υ .
$\alpha 1, \alpha 2, \beta 1, \beta 2$	Subescritos utilizados para indicar o tipo de conexão.
$\gamma 1, \gamma 2, \delta 1, \delta 2$	Subescritos utilizados para indicar o tipo de conexão.
n_{TET}, n_{TEC}	Número de terminais alimentados-por-tensão e alimentados-por-corrente.
n_{TE}	Número de terminais externos.
n_{TET}, n_{TEC}	Número de terminais externos alimentados-por tensão e alimentados-por corrente.
$n_{TE_{\mathfrak{N}}}^{(i)}, n_{TE_{1,\mathfrak{L}}}, n_{TE_{2,\mathfrak{L}}}$	Número de terminais externos da i -ésima SRN, número de terminais externos da SRL.
$n_{TE_{\alpha 1}}, n_{TE_{\alpha 2}}, n_{TE_{\beta 1}}, n_{TE_{\beta 2}}$	Número de terminais externos do tipo $\alpha 1$, $\alpha 2$, $\beta 1$, e $\beta 2$ da SRLA e da SRNC.
$n_{TE_{\gamma 1}}, n_{TE_{\gamma 2}}, n_{TE_{\delta 1}}, n_{TE_{\delta 2}}$	Número de terminais externos do tipo $\gamma 1$, $\gamma 2$, $\delta 1$, e $\delta 2$ da SuR.
n_S, n_{SE}, n_{SI}	Número de sondas, número de sondas externas e internas.
n_{VE}, n_{VEL}, n_{VEN}	Número de variáveis de estado, número de variáveis de estado linear e não-linear.
n_{FN}, n_{FNE}, n_{FND}	Número de funções não-lineares, número de funções não-lineares estáticas e dinâmicas.
R, L, C, Z_o, τ	Resistência, capacitância, indutância, impedância característica, tempo de atraso.
$\boldsymbol{x}(t), \boldsymbol{X}(\omega)$	Vetores de equação de estado no domínio do tempo e da frequência.
$\boldsymbol{u}_f(t), \boldsymbol{U}_f(\omega)$	Vetores de função não-linear no domínio do tempo e da frequência.

$\boldsymbol{u}_g(t)$, $\boldsymbol{U}_g(\omega)$	Vetores de fonte independente (gerador) no domínio do tempo e da frequência.
$\boldsymbol{y}_e(t)$, $\boldsymbol{y}_i(t)$, $\boldsymbol{Y}_e(\omega)$, $\boldsymbol{Y}_i(\omega)$	Vetores de sonda externa e de sonda interna no domínio do tempo e da frequência.
$\boldsymbol{u}_{fe}(t)$, $\boldsymbol{u}_{fd}(t)$, $\boldsymbol{U}_{fe}(\omega)$, $\boldsymbol{U}_{fd}(\omega)$	Vetores de função não-linear estática e dinâmica no domínio do tempo e da frequência.
$\boldsymbol{j}_f(t)$, $\boldsymbol{J}_f(\omega)$	Vetores de função não-linear estática no domínio do tempo e da frequência.
$\boldsymbol{e}_f(t)$, $\boldsymbol{E}_f(\omega)$	Vetores de função não-linear dinâmica no domínio do tempo e da frequência.
$\boldsymbol{q}_f(t)$, $\boldsymbol{Q}_f(\omega)$	Vetores de função não-linear estática no domínio do tempo e da frequência.
$\boldsymbol{\phi}_f(t)$, $\boldsymbol{\Phi}_f(\omega)$	Vetores de função não-linear dinâmica no domínio do tempo e da frequência.
$\boldsymbol{V}_@$	Vetor de tensão nodal (FNM).
$\boldsymbol{J}_@$, $\boldsymbol{J}_{@f}$, $\boldsymbol{Q}_{@f}$	Vetores de corrente nodal, de fonte de corrente não-linear nodal e de carga não-linear nodal (FNM).
$\boldsymbol{V}$, $\boldsymbol{I}$	Vetores de tensão e de corrente de ramo na FNM.
$\mathbf{A}_V$, $\mathbf{A}_I$	Matrizes de incidência associadas a LKT e a LKC (FNM).
$\mathbf{K}_L$, $\mathbf{K}_N$, $\mathbf{K}_g$, $\mathbf{K}$, $\mathbf{K}_i$, $\mathbf{K}_e$, $\mathbf{L}_L$, $\mathbf{L}_N$, $\mathbf{L}_f$	Matrizes elementares de permutação utilizadas na FNM.
$\boldsymbol{X}(\omega)$, $\boldsymbol{X}_L(\omega)$, $\boldsymbol{X}_N(\omega)$	Vetores de variáveis de estado, e de variáveis de estado linear e não-linear.
$\boldsymbol{U}_@(\omega)$, $\boldsymbol{U}_{@f}(\omega)$, $\boldsymbol{U}_{@g}(\omega)$, $\boldsymbol{U}_{@e}(\omega)$	Vetores de fonte nodal, de fonte não-linear nodal, de fonte independente nodal e de fonte externa (FNM).
$\boldsymbol{U}_e(\omega)$, $\boldsymbol{U}_g(\omega)$	Vetores de fonte externa e de fonte independente.
$\mathbf{M}_@(\omega)$	Matriz nodal-modificada (FNM).
$\mathbf{H}_@(\omega)$	Matriz híbrida obtida da inversão da matriz nodal-modificada (FNM).
$\boldsymbol{H}_{\nu e}(\omega)$, $\boldsymbol{H}_{\nu g}(\omega)$	Matrizes constitutivas da equação de sonda da SRLA, $\nu = i, e$ (FNM).
$\boldsymbol{V}_t(\omega)$, $\boldsymbol{V}_c(\omega)$, $\boldsymbol{I}_t(\omega)$, $\boldsymbol{I}_c(\omega)$	Vetores de tensão e de corrente associadas a árvore e a co-árvore da SRN (FEE).
$\mathbf{A}_t$, $\mathbf{A}_c$	Matrizes de incidência associadas a árvore e a co-árvore da SRN (FEE).
$\mathbf{D}$	Matriz de corte fundamental da SRN (FEE).
$1a$, $1b$, 2 , $2A$, $3A$, 3	Subescritos utilizados para indicar os grupos de cada váriavel de estado na FEE.
$n_{VE_{1a}}$, $n_{VE_{1b}}$, n_{VE_2} , $n_{VE_{2A}}$, $n_{VE_{3A}}$, n_{VE_3}	Número de variáveis de estado associadas aos grupos 1*a*,1*b*,2,2*A*,3*A*,3 (FEE).
$\boldsymbol{X}_{1a}(\omega)$, $\boldsymbol{X}_{1b}(\omega)$, $\boldsymbol{X}_2(\omega)$, $\boldsymbol{X}_{2A}(\omega)$	Vetores de variável de estado linear associados aos grupos 1*a*,1*b*,2,2*A* (FEE).
$\boldsymbol{X}_{3A}(\omega)$, $\boldsymbol{X}_3(\omega)$	Vetores de variável de estado não-linear associados aos grupos 3*A*,3 (FEE).
$\boldsymbol{\Gamma}_x(\omega)$	Matriz associada aos operadores diferencias, integrais e de diferença aplicado a $\boldsymbol{X}(\omega)$.
$\boldsymbol{\Gamma}_f(\omega)$	Matriz associada ao operador diferencial aplicado a $\boldsymbol{U}_f(\omega)$.
n_{EB}	Número de elementos básicos.
j_x , j_f , j_u , j_y	Índices dos elementos básicos para a construção das equações de estado e de sonda via FEE.
$\underline{\mathbf{A}}_l^{(j_x,j_x)}$, $\underline{\mathbf{A}}_r^{(j_x,j_x)}$, $\underline{\mathbf{B}}_f^{(j_x,j_f)}$, $\underline{\mathbf{B}}^{(j_x,j_u)}$	Matrizes constituivas da equação de estado de um elemento básico.
$\underline{\mathbf{C}}^{(j_y,j_x)}$, $\underline{\mathbf{D}}_f^{(j_y,j_f)}$, $\underline{\mathbf{D}}^{(j_y,j_u)}$	Matrizes constituivas da equação de saída de um elemento básico.
$\underline{\mathbf{N}}^{(j_y,j_u)}$	Matriz consitutiva da equação de sonda.
$\boldsymbol{A}(\omega)$, $\boldsymbol{B}_f(\omega)$, $\boldsymbol{B}_e(\omega)$	Matrizes constituivas da equação de estado da SRN.
$\boldsymbol{M}(\omega)$, $\boldsymbol{N}_f(\omega)$, $\boldsymbol{N}_e(\omega)$	Matrizes constituivas da equação de sonda da SRN.
n_{SRN}	Número de SRNs na SRNC.
e_α , e_β , e_γ , e_δ	Subescritos utilizados para indicar o tipo de conexão externa de um SLRA, SRNC e SuR.
$\boldsymbol{U}_{e_\alpha}^{\mathfrak{S}_a}(\omega)$, $\boldsymbol{U}_{e_\beta}^{\mathfrak{S}_a}(\omega)$, $\boldsymbol{Y}_{e_\alpha}^{\mathfrak{S}_a}(\omega)$, $\boldsymbol{Y}_{e_\beta}^{\mathfrak{S}_a}(\omega)$	Vetores de fonte externa e de sonda externa da SRLA e da SRNC.
$\boldsymbol{U}_{e_\alpha}^{\mathfrak{N}_c}(\omega)$, $\boldsymbol{U}_{e_\beta}^{\mathfrak{N}_c}(\omega)$, $\boldsymbol{Y}_{e_\alpha}^{\mathfrak{N}_c}(\omega)$, $\boldsymbol{Y}_{e_\beta}^{\mathfrak{N}_c}(\omega)$	Vetores de fonte externa e de sonda externa da SRLA e da SRNC.
$\boldsymbol{H}_{\nu\mu}^{\mathfrak{S}_a}(\omega)$	Matrizes constitutivas da equação de sonda da SRLA, $\nu = i, e_\alpha, e_\beta, e_\gamma, e_\delta$ e $\mu = g, e_\alpha, e_\beta, e_\gamma, e_\delta$.

$\boldsymbol{A}^{\mathfrak{R}_c}(\omega)$, $\boldsymbol{B}_f^{\mathfrak{R}_c}(\omega)$, $\boldsymbol{B}_{e_\alpha}^{\mathfrak{R}_c}(\omega)$, $\boldsymbol{B}_{e_\beta}^{\mathfrak{R}_c}(\omega)$	Matrizes constitutivas da equação de estado da SRNC.
$\boldsymbol{M}_\nu^{\mathfrak{R}_c}(\omega)$, $\boldsymbol{N}_{\nu f}^{\mathfrak{R}_c}(\omega)$, $\boldsymbol{N}_{\nu e_\alpha}^{\mathfrak{R}_c}(\omega)$, $\boldsymbol{N}_{\nu e_\alpha}^{\mathfrak{R}_c}(\omega)$	Matrizes constitutivas da equação de sonda da SRNC, $\nu = i, e_\alpha, e_\beta$.
$\boldsymbol{X}^{(\upsilon,k)}(\omega)$, $\boldsymbol{U}_f^{(\upsilon,k)}(\omega)$, $\boldsymbol{U}_g^{(\upsilon,k)}(\omega)$	Vetores de váriavel de estado, de função não-linear, e de fonte independente da SuR, $\mathfrak{S}_{\upsilon,k}$.
$\boldsymbol{U}_{e_\gamma}^{(\upsilon,k)}(\omega)$,$\boldsymbol{U}_{e_\delta}^{(\upsilon,k)}(\omega)$,$\boldsymbol{Y}_{e_\gamma}^{(\upsilon,k)}(\omega)$,$\boldsymbol{Y}_{e_\delta}^{(\upsilon,k)}(\omega)$	Vetores de fonte externa e de sonda externa da SuR, $\mathfrak{S}_{\upsilon,k}$.
$\boldsymbol{X}_\gamma^{(\upsilon,k)}(\omega)$, $\boldsymbol{X}_\delta^{(\upsilon,k)}(\omega)$	Vetores de variável de estado de conexão γ e δ da SuR, $\mathfrak{S}_{\upsilon,k}$.
$\boldsymbol{A}^{\mathfrak{S}}(\omega)$, $\boldsymbol{B}_f^{\mathfrak{S}}(\omega)$, $\boldsymbol{B}_g^{\mathfrak{S}}(\omega)$, $\boldsymbol{B}_{e_\gamma}^{\mathfrak{S}}(\omega)$, $\boldsymbol{B}_{e_\delta}^{\mathfrak{S}}(\omega)$	Matrizes constitutivas da equação de estado da SuR de fundo.
$\boldsymbol{M}_\nu^{\mathfrak{S}}(\omega)$,$\boldsymbol{N}_{\nu f}^{\mathfrak{S}}(\omega)$,$\boldsymbol{N}_{\nu g}^{\mathfrak{S}}(\omega)$,$\boldsymbol{N}_{\nu e_\gamma}^{\mathfrak{S}}(\omega)$,$\boldsymbol{N}_{\nu e_\delta}^{\mathfrak{S}}(\omega)$	Matrizes constitutivas da equação de sonda da SuR de fundo, $\nu = i, e_\gamma, e_\delta$.
$\boldsymbol{A}^{(\upsilon,k)}(\omega)$, $\boldsymbol{B}_f^{(\upsilon,k)}(\omega)$, $\boldsymbol{B}_g^{(\upsilon,k)}(\omega)$, $\boldsymbol{B}_{e_\gamma}^{(\upsilon,k)}(\omega)$, $\boldsymbol{B}_{e_\delta}^{(\upsilon,k)}(\omega)$	Matrizes constitutivas da equação de estado da SuR, $\mathfrak{S}_{\upsilon,k}$.
$\boldsymbol{M}_\nu^{(\upsilon,k)}(\omega)$, $\boldsymbol{N}_{\nu f}^{(\upsilon,k)}(\omega)$, $\boldsymbol{N}_{\nu g}^{(\upsilon,k)}(\omega)$, $\boldsymbol{N}_{\nu e_\gamma}^{(\upsilon,k)}(\omega)$, $\boldsymbol{N}_{\nu e_\delta}^{(\upsilon,k)}(\omega)$	Matrizes constitutivas da equação de sonda da SuR, $\mathfrak{S}_{\upsilon,k}$, $\nu = i, e_\gamma, e_\delta$.
$\boldsymbol{C}^{(\upsilon,k)}(\omega)$, $\boldsymbol{D}_f^{(\upsilon,k)}(\omega)$, $\boldsymbol{D}_g^{(\upsilon,k)}(\omega)$, $\boldsymbol{D}_{e_\gamma}^{(\upsilon,k)}(\omega)$, $\boldsymbol{D}_{e_\delta}^{(\upsilon,k)}(\omega)$, $\boldsymbol{C}_\gamma^{(\upsilon-1,i)}(\omega)$, $\boldsymbol{D}_\delta^{(\upsilon-1,i)}(\omega)$	Matrizes constitutivas de estado associada a RC, $\mathcal{C}_{\upsilon-1,m}$, da SuR, $\mathfrak{S}_{\upsilon-1,i}$.
$\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,k)}$, $\mathbf{B}_{e_\delta\delta}^{(\upsilon,k)}$, $\mathbf{C}_{e_\gamma e_\delta}^{(\upsilon,k)}$, $\mathbf{C}_{e_\delta e_\delta}^{(\upsilon,k)}$	Matrizes de incidência associadas as variáveis de conexão externa e terminações externas da $\mathfrak{S}_{\upsilon,k}$.
$\mathbf{D}_{\delta e_\gamma}^{(\upsilon-1,i)}$, $\mathbf{D}_{\delta e_\delta}^{(\upsilon-1,i)}$	Matrizes de incidência associadas as variáveis de conexão externa e terminações externas da $\mathfrak{S}_{\upsilon-1,i}$.
k , l	Índices de harmônico, de produto de IM ou de número da frequência.
$\boldsymbol{k}$, $\boldsymbol{l}$	Vetores de índices de harmônico de um produto de IM.
NT , NH , NF	Número de tons (frequências fundamentais), de harmônicos, e de frequências.
N_{CF}	Número de componentes (parte real e imaginária) em frequência, $N_{CF} = 2NF - 1$.
M_{IM2} , M_{IM3}	Máxima ordem para os produtos de IM na análise do BH de IM de dois-tons e de três-tons.
M_{OL} , M_{IM2-BL}	Máxima ordem para os produtos de IM na análise do BH de CF de dois-tons e de três-tons.
$\boldsymbol{\Gamma}$, $\boldsymbol{\Gamma}^{-1}$, $\boldsymbol{\Gamma}^{+}$	TFQP direta, inversa e inversa generalizada.
$\kappa_p(\boldsymbol{\Gamma})$, $\beta_p(\boldsymbol{\Gamma})$, $\varepsilon_p(\boldsymbol{\Gamma})$	Número de condição, fator de estabilidade e erro de conversão da TFQP (norma-M_p).
$\mathbf{s}_\infty$	Espectro de frequência infinito contendo todas as harmônicas ou produtos IM.
$\mathbf{s}$	Espectro de frequência truncado.
$\mathbf{s}_d$	Espectro de frequência de derivada.
$\bullet\!\!-\!\!\circ$	Relação da transformada de Fourier.
$\boldsymbol{X}^{(\upsilon,k)}(k)$, $\boldsymbol{X}^{(\upsilon,k),re}(k)$, $\boldsymbol{X}^{(\upsilon,k),im}(k)$	Vetor de variável de estado da $\mathfrak{S}_{\upsilon,k}$, em harmônico, k , e sub-vetores com sua parte real e imaginária.
$\boldsymbol{A}^{(\upsilon,k)}(k)$, $\boldsymbol{A}^{(\upsilon,k),re}(k)$, $\boldsymbol{A}^{(\upsilon,k),im}(k)$	Matriz constitutiva da $\mathfrak{S}_{\upsilon,k}$, em harmônico, k , e sub-matrizes com sua parte real e imaginária.
$\bar{\boldsymbol{X}}$, $\bar{\boldsymbol{U}}_f$, $\bar{\boldsymbol{U}}_g$	Vetores de variável de estado, de função não-linear e de fonte independente do BH.
$\bar{\boldsymbol{F}}(\bar{\boldsymbol{X}})$	Vetor de resíduo não-linear do BH.
$\bar{\boldsymbol{A}}$, $\bar{\boldsymbol{B}}_f$, $\bar{\boldsymbol{B}}_g$	Matrizes da equação de estado do BH.
$\bar{\boldsymbol{M}}$, $\bar{\boldsymbol{N}}_f$, $\bar{\boldsymbol{N}}_s$	Matrizes da equação de sonda do BH.
$\bar{\boldsymbol{G}}_f(\bar{\boldsymbol{X}})$	Matriz jacobiana do BH no ponto $\bar{\boldsymbol{X}}$ e associada ao vetor de função não-linear, $\bar{\boldsymbol{U}}_f$.
$\bar{\boldsymbol{J}}(\bar{\boldsymbol{X}})$	Matriz jacobiana do BH no ponto $\bar{\boldsymbol{X}}$ e associada ao vetor de resíduo não-linear, $\bar{\boldsymbol{F}}$.
i	Índice de cálculo da matriz jacobiana.
k , NI , $NI_{\text{máx}}$	Índice de iteração, número de iterações, e máximo número de iterações do solucionador não-linear.
NCF , NCJ	Número de cálculo da função, número de cálculo da matriz jacobiana.

j , NIP , $NIP_{\text{máx}}$	Índice de iteração, número de iterações, e máximo número de iterações da pesquisa-em-linha.
l , NIL , $NIL_{\text{máx}}$	Índice de iteração, número de iterações, e máximo número de iterações do solucionador linear.
N	Número de variáveis não-lineares.
$\boldsymbol{x}_k$	Vetor de variável não-linear, $\boldsymbol{x}_k \in \mathrm{R}^N$.
$\boldsymbol{x}_{-j,\,k}$	Vetor de variável não-linear, tal que: $\boldsymbol{x}_{-j,\,k} = \boldsymbol{x}_{k-j}$.
$\boldsymbol{F}(\boldsymbol{x}_k)$	Mapeamento não-linear.
$\cos\theta(\boldsymbol{x}, \boldsymbol{y})$	Cosseno do ângulo formado pelos vetores $\boldsymbol{x}$ e $\boldsymbol{y}$.
$f(\boldsymbol{x}_k \vert \boldsymbol{A}_p)$, $N(\boldsymbol{x}_k \vert \boldsymbol{A}_p)$	Função nível e função norma com peso, no ponto $\boldsymbol{x}_k \in \mathrm{R}^N$, com peso $\boldsymbol{A}_p \in \mathrm{R}^{N \times N}$.
$\nabla f(\boldsymbol{x}_k \vert \boldsymbol{A}_p)$, $\nabla N(\boldsymbol{x}_k \vert \boldsymbol{A}_p)$	Gradientes da função nível e da função norma com peso.
$f_{\text{M2}}(\boldsymbol{x}_k)$, $f_{\text{aN}}(\boldsymbol{x}_k)$	Funções nível em norma-M2 e em norma-aN,
$\hat{f}(\lambda_{k,j} \vert \boldsymbol{A}_p)$	Função nível parametrizada, no ponto $\lambda_{k,j} \in \mathrm{R}$, com peso $\boldsymbol{A}_p \in \mathrm{R}^{N \times N}$.
$\hat{f}_{\text{aN}}(\lambda_{k,j})$	Função nível em norma-aN parametrizada.
$\kappa_2(\boldsymbol{J})$	Número de condicão da matriz jacobiana $\boldsymbol{J}$ em norma-M_2.
$\boldsymbol{M}(\boldsymbol{x}_k + \boldsymbol{d})$	Modelo local que define a correção de Newton na k -ésima iteração.
$\boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d})$	Modelo local que define a correção do tensor na k -ésima iteração.
$\boldsymbol{J}(\boldsymbol{x}_k)$	Matriz jacobiana, em $\mathrm{R}^{N \times N}$, no ponto $\boldsymbol{x}_k$ e associada ao mapeamento linear, $\boldsymbol{F}: \mathrm{R}^N \to \mathrm{R}^N$.
$\boldsymbol{d}_{N,\,k}$, $\boldsymbol{d}_{T,\,k}$	Correção de Newton, correção do tensor.
$\boldsymbol{d}_{-1,\,k}$	Vetor auxiliar utilizado na solução inexata do modelo do tensor.
$\lambda_{k,j}$, λ_0 , $\lambda_{\text{mín}}$, λ_*	Fator de amortecimento da pesquisa-em-linha na j -ésima iteração, valor inicial, mínimo e ótimo.
$\boldsymbol{T}(\boldsymbol{x}_k)$	Objeto tensor, em $\mathrm{R}^{N \times N \times N}$, no ponto $\boldsymbol{x}_k$ e associado ao mapeamento linear, $\boldsymbol{F}: \mathrm{R}^N \to \mathrm{R}^N$.
$\boldsymbol{\beta}_k$, β_k	Parâmetros do modelo do tensor.
$\boldsymbol{S}_k$, $\boldsymbol{A}_k$, $\boldsymbol{Z}_k$	Vetores em $\mathrm{R}^{N \times P}$ utilizados na representação do modelo do tensor ($P > 1$).
$\boldsymbol{s}_k$, $\boldsymbol{a}_k$, $\boldsymbol{z}_k$	Vetores em R^N utilizados na representação do modelo do tensor ($P = 1$).
P	Número de pontos passados utilizados na construnção do modelo do tensor
$\boldsymbol{q}_k(\boldsymbol{\beta})$	Vetor de polinômio quadrático envolvido na solução do modelo do tensor ($P > 1$).
$q_k(\beta)$	Polinômio quadrático envolvido na solução do modelo do tensor ($P = 1$).
$q_k(\beta, \lambda_{k,j})$	Polinômio quadrático do modelo do tensor utilizado na pesquisa-em-linha curvilinear.
$\boldsymbol{H}(\boldsymbol{x};\zeta)$	Operador linear que define o caminho de homotopia.
$x_i^{(pred)}(\bar{\zeta}_c)$, $x_i^{(hom)}(\bar{\zeta}_{c,\,-j})$, $L_q^j(\bar{\zeta}_c)$, Q	
ζ_0 , ζ_* , $\bar{\zeta}$, $\bar{\zeta}_c$, $\bar{\zeta}_{c,\,-j}$	Parâmetros utilizados na técnica de continuação (ou homotopia).
η_k , ζ_k , ξ_k , ε_k , ρ_k	Termos forçantes.
$\boldsymbol{r}_k$	Vetor de resíduo associado ao modelo linear.
$?\,\boldsymbol{r}_{T,\,k}\,?$	Vetor de resíduo associado ao modelo do tensor.
τ_k	Parâmetro que define a solução modificada do modelo do tensor.
$\boldsymbol{K}_m(\boldsymbol{A}, \boldsymbol{v})$	Sub-espaço de Krylov cobrindo $\{\boldsymbol{v}, \boldsymbol{A}\boldsymbol{v}, \boldsymbol{A}^2\boldsymbol{v}, \ldots, \boldsymbol{A}^{m-1}\boldsymbol{v}\}$.
$\rho(\boldsymbol{A})$	Raio espectral da matriz $\boldsymbol{A}$.
$\boldsymbol{M}$, $\boldsymbol{M}_1$, $\boldsymbol{M}_2$	Matrizes de pré-condicionamento.

Lista de Figuras

Fig.2.1. (a) Estrutura da decomposição multi-níveis de um circuito. O circuito é decomposto em super-redes (SuRs) hierarquicamente interconectadas através de redes de conexão. (b) Representação em árvore da estrutura hierárquica do circuito. (c) Detalhe da estrutura hierárquica demonstrando a i-ésima SuR intermediária localizada no nível $\upsilon - 1$ de hierarquia, e a k-ésima SuR de fundo localizada no υ-ésimo nível. Em geral, a j- e l-ésima SuR indicadas podem ser do tipo intermediária ou de fundo. 20

Fig.2.2. Estrutura de uma super-rede (SuR) de último nível de hierarquia, localizado no nível υ, e decomposta em partes linear e não-linear. 21

Fig.2.3. (a) Exemplo da decomposição de circuito multi-níveis aplicado ao estágio de recepção de um transceptor de RF. (b) Estrutura em árvore da decomposição multi-níveis do estágio de recepção, ilustrado em (a). 24

Fig.3.1. Estrutura geral de uma sub-rede constituída de elementos básicos estruturais, onde n_{TET} e n_{TEC} representam o número de terminais externos alimentados-por-tensão e alimentados-por-corrrente, respectivamente. Os elementos básicos estruturais para formação da sub-rede não-linear (SRN) e da sub-rede linear (SRL) são destacados. 28

Fig.3.2. (a) Fluxograma da geração das equação topológica. (b) Fluxograma da geração das equações de estado e de sonda. 41

Fig.3.3. (a) Circuito elétrico equivalente (CEE) do FET para regime CC e CA, incluindo a rede de alimentação externa (RAE). (b) CEE do HBT para regime CC e CA, incluindo a RAE. 44

Fig.4.1. Esquema elétrico das fontes e das sondas utilizadas na representação da interconexão entre a *sub-rede linear ampliada* (SRLA) e a *sub-rede não-linear compactada* (SRNC). Também são indicadas as fontes e as sondas utilizadas na conexão da super-rede (SuR) de fundo com um circuito externo. 48

Fig.4.2. Esquema elétrico para conexões externas das super-redes (SuRs) associadas com a estrutura da Fig. 2.1(c). 55

Fig.4.3. Estrutura multi-níveis, tipo bloco diagonal com e sem borda das matrizes constitutivas da equação (a) de estado e (b) de sonda do circuito da Fig. 2.3. 60

Fig.5.1. Topologia do espectro de frequência e grade de frequência para: (a) transformada de Fourier multi-tons (TFMT) e (b) para a transformada de Fourier multi-tons em duas dimensões (TFMT-2D). 66

Fig.5.2. (a) Forma-de-onda de um de um sinal de digital representado por uma sequência pseudo-aleatória de NSímb símbolos com 4 níveis de quantização. (b) Constelação associada à modulação 16 QAM. (c) Passagem de um sinal digital por um filtro de ISI. (d) Forma-de-onda dos sinais $i(t)$ e $q(t)$ obtidos numericamente com formatação de pulso cosseno levantado. (e) Espectro de frequência $I(f)$ e $Q(f)$ dos sinais $i(t)$ e $q(t)$, respectivamente, obtidos via TFMT. 67

Fig.5.3. Teste de precisão das transformadas de Fourier discretas para sinais quasi-periódicos. As frequências fundamentais são $\lambda_1 = 2\pi$ GHz e $\lambda_2 = 2\pi \cdot 0{,}7$ GHz. (a) e (c) $M_{IM2} = 5$. (b) e (d) $M_{IM2} = 10$. O fator de sobre-amostragem, m, é igual a 2. 73

Fig.5.4. As frequências fundamentais são $\lambda_1 = 2\pi$ GHz e $\lambda_2 = \lambda_1 + 2\pi\sqrt{2}$ Hz. (a) $m = 1$, (b) . $m = 2$, (c) $m = 3$, e (d) $m = 4$. As frequências fundamental são $\lambda_1 = 2\pi$ GHz e $\lambda_2 = 2\pi \cdot 0{,}7$ GHz. (e) $m = 1$, (f) $m = 2$, (g) $m = 3$, e (h) $m = 4$. m é o fator de sobre-amostragem. 74

Fig.5.5. Exemplo da aplicação da técnica de mapeamento artificial em frequência para espectro de frequência de dois-tons: (a) grade triangular e (b) grade retangular. 75

Fig.5.6. Estrutura da matriz jacobiana com (a) 1 (um) nível e (b) 2 (dois) níveis de decomposição. (c) Estrutura multi-níveis (bloco diagonal com borda dupla) da matriz jacobiana para sucessivas decomposições incluindo um nível adicional. Assume-se que cada transistor intruduz 3 (três) variáveis de estado não-linear ($\overline{NF} = 2NF - 1$). 77

Fig.5.7. (a) Padrão de esparsidade da matriz jacobiana para análise de único-tom. (b) Padrão de esparsidade da matriz jacobiana para análise de dois-tons (n_{NZB} - número de blocos não-zero). 82

Fig.6.1. Gráficos da razão de desempenho (RD) do método do tensor (novo) vs. método Newton (padrão) para: (a) $J(x_*)$ não-singular e (b) $\boldsymbol{J}(\boldsymbol{x}_*)$ singular com $\text{posto}(\boldsymbol{J}(\boldsymbol{x}_*)) = n - 1$. 109

Fig.6.2. Gráficos da razão de desempenho (RD) das estratégias de pesquisa-em-linha para o método do tensor. (a) Pesquisa-em-linha curvilinear com interpolação quadrática vs. pesquisa-em-linha padrão. (b) Pesquisa-em-linha curvilinear com λ-divindo-pela-metade vs. pesquisa-em-linha padrão. 110

Fig.7.1. Gráficos da razão de desempenho (RD) do método do tensor inexato (novo) vs. Newton inexato (padrão). (a) Processo de solução simplificada. (b) Processo de solução modificada. (c) Processo de solução completa B2. (d) Processo de solução completa B3. 144

Fig.7.2. Gráficos da razão de desempenho (RD) do método do tensor inexato com solução modificada e implementação escalar. (a) Estratégia de pesquisa-em-linha curvilinear com interpolação quadrática (novo) vs. estratégia padrão (padrão). (b) Estratégia de pesquisa-em-linha curvilinear λ-divindo-pela-metade (novo) vs. estratégia padrão (padrão). 145

Fig.7.3. Histórico de convergência para solução do problema 39 utilizando o método de Newton inexato com (a) Escolha 0 e (b) Escolha 5. Histórico de convergência (problema 39) do método do tensor inexato com solução modificada e implementação bloco utilizando (c) Escolha 0 e (d) Escolha 7. 146

Fig.8.1. (a) Esquemático e (b) resultado da fonte de alimentação com retificação de meia-onda (FARMO) e de onda completa (FAROC). (c) Esquemático e (d) resultado do amplificador classe-C (ACC) utilizando BJT. (e) Esquemático e (f) resultado multiplicador de frequência (MF) utilizando BJT. 152

Fig.8.2. (a) Esquemático do amplificador de potência (AP) de microonda utilizando GaAs MESFET. (c) Parâmetros de espalhamento. (b) Trajetórias I/V. (c) Potência de saída versus potência de entrada para regime de único-tom. (e) Potência de saída versus potência de entrada em regime de dois-tons. (f) Formas-de-onda das tensões de entrada e de saída e (g) recrescimento espectral em regime multi-tons com excitaçãode RF modulada digitalmente. 154

Fig.8.3. (a) Microfotografia do amplificador de potência corporativo (APC) fabricado pela Phillips. (b) Esquemático do APC sub-dividido em 2 super-redes (SuRs). (c) Magnitude dos parâmetros de espalhamento medidos e calculados. (d) Ponto de compressão de 1 dB medido e calculado. 157

Fig.8.4. (a) Esquemático e (b)-(c) resultados do conversor de frequência resistivo (CFR) de onda milimétrica utilizando InP pHEMT. (e) Esquemático e (e),(f) resultados do conversor de frequência resistivo balanceado (CFRB) de onda-milimétrica utilizando InP pHEMT. 159

Fig.8.5. (a) Esquemático do circuito para teste do ressoador ativo (RA). (b) Esquemático do circuito elétrico do RA utilizando GaAs MESFETs. 161

Fig.8.6. (a) Esquemático do RA decomposto hierarquicamente em super-redes (SuRs). (b) Esquemático dos circuitos das SuRs utilizadas em (a). (c) Estrutura de dois níveis da matriz jacobiana para o esquema de decomposição (a) (em escala). (d) Impedância de entrada do RA versus frequência para diversos níveis de potência de entrada. (e) Coeficiente reflexão de entrada de pequeno-sinal do RA versus tensão de sintonia, $V_Q = -5{,}5$ V. 163

Fig.8.7. (a) Esquemático do circuito para teste do multiplicador de quatro-quadrantes (MAQQ) e do multiplicador de quatro-quadrantes de baixo-deslocamento (MAQQ-BD) utilizando GaAs MESFETs. (b) Esquemático do MAQQ. (c) Esquemático do MAQQ-BD. (d) Curvas de transferência dos MAQQs operando em regime CC. 164

Fig.8.8. Esquemático (a) do multiplicador analógico de quatro-quadrantes (MAQQ) e (b) do MAQQ-BD decomposto hierarquicamente em super-redes (SuRs). (c) Esquemático dos circuitos das SuRs de fundo utilizadas em (a) e (b). (d) e (e) Estrutura de três-níveis da matriz jacobiana para os esquemas de decomposição (a) e (b), respectivamente (em escala). (f) e (g) Estrutura de dois níveis da matriz jacobiana do MAQQ e do MAQQ-BD, respectivamente (em escala). (a) Formas-de-onda e (b) espectro de frequência da tensão de saída dos MAQQs. 165

Fig.8.9. (a) Microfotografia do multiplicador analógico balanceado (MAB) fabricado pela TRW. (b) Esquemático do MAB. 168

Fig.8.10. (a) Esquemático do multiplicador análogico balanceado (MAB) decomposto hierarquicamente em super-redes (SuRs). (b) Esquemático do circuito das SuRs de fundo utilizadas em (a). (c) Estrutura de dois níveis da matriz jacobiana do MAB (em escala). (d) Estrutura de três níveis da matriz jacobiana para o esquema de decomposição (a) (em escala). 170

Fig.8.11. (a) Esquemático do multiplicador análogico balanceado (MAB) decomposto hierarquicamente em super-redes (SuRs). (b) Esquemático dos circuitos das SuRs de fundo utilizadas em (a). (c) Estrutura de dois níveis da matriz jacobiana do MAB (em escala). (c) Estrutura de três níveis da matriz jacobiana do MAB (em escala). (d) Estrutura de quatro níveis da matriz jacobiana para o esquema de decomposição (a) (em escala). 171

Fig.8.12. (a) Ganho direto versus frequência do MAB operando como amplificador de ganho variável (AGV). (b) Potência de saída versus potência de entrada em regime de único-tom do MAB-AGV. 172

Fig.8.13. Histórico de convergência para solução do BH do MAB-AGV (problema 18) via FNM utilizando o método de método de Newton inexato com (a) Escolha 0 e (b) Escolha 5. Histórico de convergência do problema 18 utilizando o método do tensor curvilinear inexato com (c) Escolha 0 e (d) Escolha 7. Os resultados (e)-(h) se referem à FEE. 177

Fig.8.14. Distribuição dos auto-velores (espectro) da matriz jacobiana do BH para o circuito do APC, calculada na raiz do problema, com potência de entrada, P_{ent} igual (a) 10 e (b) 30 dBm. 178

Fig.8.15. Gráficos da razão de desempenho (RD) da análise do BH multi-níveis. (a) Solução "exata" dos sistemas jacobianos via fatorização/retro-substituição LU. (b) Solução "inexata" via GMRES/pré-condicionador LU. 180

Lista de Tabelas

Tabela 5.1 Topologias do espectro de frequência do sinal para diferentes tipos de análise de equilíbrio de harmônico. 65

Tabela 6.1 Definição de Classe do Problema 90

Tabela 6.2 Resultados para o conjunto de problemas testes de Moré, Garbow e Hillstrom 108

Tabela 7.1 Resultados dos Problemas Testes 141

Tabela 8.1 Função Não-Linear do Modelo do Diodo 153

Tabela 8.2 Parâmetros Elétricos do CEE (top-A) do GaAs MESFET 155

Tabela 8.3 Funções Não-Lineares do Modelo da Phillips do GaAs MESFET 155

Tabela 8.4 Parâmetros das Funções Não-Lineares do Modelo da Phillips do GaAs MESFET 155

Tabela 8.5 Parâmetros Elétricos do CEE (top-B) do InP pHEMT 160

Tabela 8.6 Funções Não-Lineares do Modelo de Lin-Ku do InP pHEMT 160

Tabela 8.7 Parâmetros das Funções Não-Lineares do Modelo de Lin-Ku do InP pHEMT 160

Tabela 8.8 Parâmetros Elétricos do CEE (top-C) do GaAs MESFET 161

Tabela 8.9 Funções Não-Lineares do Modelo HSPICE do GaAs MESFET 161

Tabela 8.10 Parâmetros das Funções Não-Lineares do Modelo HSPICE do GaAs MESFET 161

Tabela 8.11 Parâmetros Elétricos do CEE (top-A) do InP HBT 169

Tabela 8.12 Funções Não-Lineares do Modelo de Gummel-Poon do HBT 169

Tabela 8.13 Parâmetros das Funções Não-Lineares do Modelo de Gummel-Poon do InP HBT 169

Tabela 8.14 Listas dos problemas testes e dos resultados da análise de CC, NC: Nome do Circuito, #T: Número de Transistores, #FN: Número de Funções Não-lineares, #VE: Número de Variáveis de Estado (FEE/FNM) 173

Tabela 8.15 Listas com a Estatística dos Problemas Testes, NC: Nome do Circuito, Dim. TFD: Dimensão da Transformada de Fourier Discreta, #D: Número de Diodos, #T: Número de Transistores, #F: Número de Frequências, #FN: Número de Funções Não-Lineares, #V: Número de Variáveis (FEE/FNM), #VBH: Número de Variáveis na Análise do BH, SNL: Solucionador Não-Linear (“E”=Exato e “I”=Inexato) 174

Tabela 8.16 Listas dos Problemas Testes e dos Resultados da Análise do BH, NC: Nome do Circuito, #T: Número de Transistores, #FN: Número de Funções Não-Lineares, #VP: Número de Variáveis do Problema FEE/FNM 175

Tabela 8.17 Listas com a Estatística dos Problemas Testes, NC: Nome do Circuito, #NH: Número de Níveis da Hierarquia, Dim. TFD: Dimensão da Transformada de Fourier Discreta, #T: Número de Transistores, #F: Número de Frequências, #FN: Número de Funções Não-Lineares, #V: Número de Variáveis, #VBH: Número de Variáveis na Análise do BH, Sol. NL: Solucionador Não-Linear (“E”=Exato e “I”=Inexato) 179

Tabela A.1 Representação de Espaço-de-Estado dos Elementos Concentrados 197

Tabela A.2 Representação de Espaço-de-Estado dos Elementos Distribuídos 197

Tabela A.3 Representação de Espaço-de-Estado das Fontes Lineares Controladas 198
Tabela A.4 Representação de Espaço-de-Estado dos Elementos Não-lineares 198
Tabela A.5 Representação de Espaço-de-Estado dos Elementos de Controle Linear 199
Tabela A.6 Representação de Espaço-de-Estado dos Elementos de Controle Não-linear 199
Tabela A.7 Representação de Espaço-de-Estado das Sondas 199
Tabela B.1 Equações de Estado e de Saída dos Elementos Básicos do CEE do FET 201
Tabela B.2 Equações de Estado e de Saída dos Elementos Básicos do CEE da Parte Intrínseca do FET Representada sob Forma de DDS 202
Tabela B.3 Equações de Estado e de Saída dos Elementos Básicos do CEE do HBT 203
Tabela B.4 Equações de Estado e de Saída dos Elementos Básicos do CEEda Parte Intrínseca do HBT Representada sob Forma de DDS 204

Lista de Algorítimos

Algoritmo BLS: Backtracking Line-Search 92
Algoritmo GN: Global Newton 93
Algoritmo BLS-TS: Backtracking Line-Search - Tensor Standard 101
Algoritmo GT: Global Tensor 104
Algoritmo RGMRES(m): Right-Preconditioned GMRES (with restart) 118
Algoritmo GIN: Global Inexact Newton 121
Algoritmo RGMRES-Bt(m): Right-Preconditioned GMRES - Block (with restart) 128
Algoritmo RTGMRES-Bt(m): Right-Preconditioned TGMRES - Block (with restart) 133
Algoritmo GIT: Global Inexact Tensor 137

Agradecimentos

Eu gostaria de agradecer ao Prof. Dr. Hugo Enrique Hernandez Figueroa e ao Prof. Dr. Rui Fragassi Souza, por tornar possível a apresentação deste trabalho de tese de doutorado.

Eu gostaria de agradecer ao Instituto de Pesquisa Eldorado pelo suporte financeiro a este trabalho, condicionado a minha participação no desenvolvimento do Projeto de um sistema passivo de *radio frequency identification* (RFID), operando em 915 MHz, via a técnica de retro-espalhamento. Em particular, gostaria de agradecer ao Engº. José Eduardo Bertuzzo por ter sido o principal responsável pela execução deste projeto. Também gostaria de agradecer ao Dr. Ricardo Yashioka pelo suporte na compra dos componentes e na assistência durante a caracterização experimental do sistema de RFID. Ainda em referência a este projeto, gostaria de agradecer ao Sr. Wladinei Ricardo Camillo Menegassi pelo suporte em diversas questões referente à minha participação.

Eu gostaria de agradecer ao Prof. Dr. Jacobus Williborus Swart pela oportunidade de atuar como pesquisador visitante no Centro de Componentes Semicondutores (CCS), Universidade Estadual de Campinas (UNICAMP).

Eu gostaria de agradecer a todos no Departamento de Microonda e Óptica (DMO), Faculdade de Engenharia Elétrica e de Computação (FEEC), Universidade Estadual de Campinas (UNICAMP), pelo suporte em diversos aspectos referentes a apresentação deste trabalho.

Eu gostaria de agradecer ao meu grande amigo e colega Engº. Leandro Tiago Manera, pela irmandade, amizade, e pelas inúmeras vêzes em que pude contar com o seu apoio.

Eu gostaria de agradecer ao meu amigo irmão Sr. Lucas Souto Avena (“irmanito”), por ser uma constante fonte de energia positiva na minha vida.

Finalmente, eu gostaria de agradecer a minha querida mãe, minha irmã, meu irmão (bacharel em Engenharia Elétrica, UFBA, e agora doutor em Medicina, USP) e Maria, por todo amor, carinho, atenção e compreensão dedicados ao meu desenvolvimento espiritual e profissional.

Conteúdo

Sumário **i**
Abstract **ii**
Acrogramas **iii**
Símbolos **viii**
Lista de Figuras **xii**
Lista de Tabelas **xv**
Lista de Algoritmos **xvii**
Agradecimentos **xviii**
1. Introdução **1**
1.1. Análise de Regime Permanente para Circuitos Não-Lineares Forçados 4
1.1.1. Método do Balanço Harmônico : Estado-da-Arte 4
1.1.2. Extensões da Análise do Balanço Harmônico 9
1.1.3. Outros Métodos 10
1.2. Contribuições deste Trabalho 12
1.3. Organização da Tese 14
2. Decomposição multi-níveis do Circuito **18**
2.1. Introdução 18
2.2. Decomposição do Circuito em Super-Redes 19
2.3. Decomposição por Partes da Super-Rede 21
2.4. Exemplo Ilustrativo 23
2.5. Conclusão 25
3. Formulação das Equações das Sub-Redes Linear e Não-Linear **26**
3.1. Introdução 26
3.2. Considerações Iniciais 27
3.3. Estrutura das Sub-Redes Linear e Não-Linear 27
3.4. Formulação Nodal-Modificada 30
3.4.1. Formulação da Sub-Rede Não-Linear 32
3.4.2. Formulação da Sub-Rede Linear 34
3.5. Formulação de Espaço-Estado da Sub-Rede Não-Linear 35
3.5.1. Equação Topológica 35
3.5.2. Equação de Estado e de Sonda 37
3.6. Dispositivos Semicondutores 43
3.7. Conclusão 45
4. Formulação Multi-Níveis das Equações do Circuito **46**
4.1. Introdução 46
4.2. Formulação das Equações da Sub-Rede Linear Ampliada 47
4.3. Formulação das Equações da Sub-Rede Não-Linear Compactada 47
4.4. Formulação das Equações da Super-Rede de Fundo 50
4.5. Formulação Multi-Níveis das Equações do Circuito 54
4.6. Eliminação de Níveis da SuR Intermediária 59
4.7. Conclusão 60
5. Análise de Equilíbrio de Harmônico **62**
5.1. Introdução 62
5.2. Espectro de Frequência 63
5.3. Excitação Digital e Multi-Senos 65
5.4. Transformada de Fourier Discreta 68

5.5. Mapeamento-de-Frequência Artificial 73
5.6. Equação Determinante 74
5.7. Matriz Jacobiana 76
5.8. Conclusão 83
6. Método do Tensor 85
6.1. Introdução 85
6.2. Considerações Preliminares 86
6.3. Função Nível 86
6.4. Método de Newton 87
6.4.1. Globalização via Pesquisa-em-Linha 89
6.4.2. Implementação Modificada 93
6.5. Método do Tensor 95
6.5.1. Constução do Modelo Tensor 96
6.5.2. Solução do Modelo Tensor 98
6.5.3. Globalização via Pesquisa-em-Linha 101
6.5.4.1. Estratégia Padrão 101
6.5.5.2. Estratégia Curvilinear 102
6.5.6. Implementação Modificada 104
6.6. Técnica de Continuação 105
6.7 Testes Preliminares 107
6.8 Conclusão 109
7. Método do Tensor-Krylov 112
7.1. Introdução 113
7.2. Método de Newton Inexato 114
7.2.1. Solução Iterativa do Modelo Linear 114
7.2.1.1. Pré-Condicionamento 115
7.2.1.2. Técnicas em Subespaço de Krylov 116
7.2.2. Termo Forçante 119
7.2.4. Globalização via Pesquisa-em-Linha 120
7.2.5. Implementação Modificada 121
7.3. Método do Tensor Inexato 122
7.3.1. Solução Iterativa do Modelo do Tensor 124
7.3.1.1. Pré-Condicionamento 125
7.3.1.2. Técnicas em Subespaço de Krylov 125
7.3.1.2.a. Solução Simplificada e Modificada 126
7.3.1.2.b. Solução Completa 129
7.3.2. Termo Forçante 134
7.3.3. Globalização via Pesquisa-em-Linha 135
7.3.3.1. Estratégia Padrão 135
7.3.3.2. Estratégia Curvilinear 136
7.3.4. Implementação Modificada 136
7.4.Produto Matriz Jacobiana-Vector na Análise do BH 138
7.5. Pré-Condicionadores para Análise do BH 140
7.6. Testes Preliminares 141
7.7. Conclusão 146
8. Validação Numérica 149
8.1. Introdução 149
8.2. Implementação em CAD 150
8.3. Descrição dos Circuitos e Resultados 150
8.3.1. Circuitos Básicos 151
8.3.2. Amplificador de Potência 152

8.3.3. Amplificador de Potência Corporativo 155
8.3.4. Conversor de Frequência Resistivo 158
8.3.5. Ressoador Ativo 158
8.3.6. Multiplicadores Analógicos de Quatro-Quadrantes 162
8.3.7. Multiplicador Analógico Balanceado 166
8.4. Desempenho dos Métodos de Newton e do Tensor 172
8.5. Desempenho da Análise do BH Multi-Níveis 178
8.6. Conclusão 181
9. Conclusão 182
9.1. Considerações Finais 182
9.2. Trabalhos Futuros 184
Referências Bibliográficas 188
Apêndice A. Representação dos Elementos Básicos na Formulação de Espaço-de-Estado 196
Apêndice B. Formulação de Espaço-de-Estado do Circuito Elétrico Equivalente do FET e do HBT 200
Apêndice C. Solução multi-níveis de Sistema Bloco Diagonal com Dupla Borda 205

1. Introdução

MÉTODOS COMPUTACIONAIS para simulação (análise e otimização) de circuitos não-lineares têm sido extensivamente desenvolvidos desde os idos de mil novecentos e sessenta, sendo a análise dinâmica via integração no domínio do tempo (IDT) um dos primeiros métodos que foram desenvolvidos [1],[2]. O programa SPICE [3], amplamente aceito na academia e na indústria, consiste em uma das mais bem suscedidas implementações para solução de problemas de valor inicial. Entretanto, a menos que uma técnica de extrapolação seja utilizada [4], o método de IDT requer a integração do transiente antes de ser capaz de determinar a resposta de regime permanente do circuito. Em circuitos levemente amortecidos, a convergência deste método (e suas variações para problemas duros) é muito lenta, tendo em vista que, neste caso, a integração deve se extender por muitos períodos, o que torna a determinação do regime permanente uma computação de alto custo. A análise de circuitos (ou sistemas) de telecomunicação com portadoras de rádio-frequência (RF) moduladas por sinais digitais na banda-base (BB) também representa uma séria dificuldade. Uma vez que nesta situação, temos um sinal de RF de rápida-variação no tempo (tipicamente maior do que 10^8 ciclos por segundo) e um sinal na BB de lenta-variação no tempo (tipicamente menor do que 10^5 ciclos por segundo). Apesar disto, o método de IDT não possui nenhuma limitação para representação de sinais de entrada (ou excitações) com formas-de-onda complexas. A inclusão de elementos distribuídos não-dispersivos na análise via IDT foi demonstrada em [5], com aplicações em circuitos de alta-frequência. Se os elementos distribuídos forem dispersivos em frequência, pode ser empregada a técnica de convolução descrita em [6].

Para evitar as dificuldades citadas acima, com o método de força bruta de IDT, foi desenvolvido o método das tentativas fundamentado no cálculo direto, no domínio do tempo, da resposta de regime permanente do circuito. O sistema não-linear de equações algébricas diferenciais (EADs) que precisa ser resolvido para este cálculo, equivale a um problema de valor de contorno de dois-pontos, resultante da imposição de uma condição que descarta a solução transiente. A teoria, a implementação e a aplicação númerica deste método podem ser encontradas em [7],[8]. No método das tentativas, o resultado é obtido utilizando técnica de solução de equações não-lineares, e.g., o método de Newton, onde a dimensão do problema é igual ao número de variáveis de estado presentes no circuito. Um outro método relacionado é o de diferença finita no domínio do tempo (DFDT) discutido em [9], que utiliza aproximações por diferenças finitas para discretizar no tempo, sobre um período, o sistema de EADs governantes do circuito. Isto resulta em um sistema de equações algébricas que são resolvidas simultâneamente para a determinação das variáveis de

estado do circuito em todos os pontos de tempo da grade de discretização. A convergência deste método é inferior ao das tentativas e também demanda uma maior quantidade de memória. Infelizmente, o método das tentativas e de DFDT estão restritos à aplicação em circuitos não-lineares excitados por sinais periódicos.

Uma outra alternativa para a determinação do regime permanente via solução direta, porém no domínio da frequência, é o estabelecido método do balanço harmônico (BH) (tradução nossa do termo em inglês *harmonic balance* (HB)) [10],[11],[12]. Neste método, a dimensão do problema é igual ao produto do número de variáveis de estado (parte real e imaginária) vezes o número de linhas espectrais para representação do sinal, e sua aplicação na análise não-linear de circuitos e/ou sistemas de RF pode facilmente envolver um alto custo computacional. Para minimizar este custo, o número de variáveis de estado por harmônico a serem determinadas pode ser significantemente reduzido, utilizando a técnica de decomposição de circuito *por partes* [13]. A solução do problema tem sido conduzida utilizando métodos de relaxação (e.g., *Jacobi*, *Gauss-Seidel* [14]) [12],[15],[16] e métodos iterativos que utilizam a matriz jacobiana do BH de forma explícita (e.g., Newton, secante, modificação [17]) [11],[18],[19],[20] ou implícita (e.g., Newton inexato [21]) [22],[23],[24],[25],[26]. Apesar dos métodos de relaxação não utilizarem a matriz jacobiana, a sua aplicação restringe-se a solução de problemas onde a potência dos sinais de entrada mantém o circuito operando em um regime *fracamente* não-linear. Para aumentar o limite de manipulação de potência, deve-se empregar métodos que utilizam a matriz jacobiana. Neste contexto, com relação à dimensão do problema, é conveniente introduzir as seguintes classificações: problemas de pequena-escala, quando os sistemas jacobianos podem ser resolvidos utilizando representação densa e fatorização QR; média-escala, quando precisam ser resolvidos utilizando representação esparsa e fatorização LU; e grande-escala, quando só podem ser resolvidos via métodos iterativos lineares, como os que operam em subespaço de Krylov (e.g., *generalized minimal residual* (GMRES) [27]). Nos problemas de grande-escala, técnicas de matriz esparsa são necessárias para formação e solução dos pré-condicionadores utilizados para assegurar a robustez dos métodos iterativos lineares.

Uma dificuldade inerente do método do BH é a sua limitação na representação das formas-de-onda que excitam o circuito, tendo em vista que isto resulta em um aumento na dimensão do problema e na complexidade da conversão do sinal de tempo para frequência e vice-versa. O trucamento em frequência depende do tipo de análise a ser realizada e deve efetivar-se de forma que o sinal possa ser devidamente representado. Apesar de classificado como um método no domínio-da-frequência, o método do BH realiza o cálculo das funções não-lineares no domínio do tempo. Com isto, faz-se necessário a representação no tempo das variáveis de estado, e de

conversões de sinal entre o domínio do tempo e da frequência. Estas conversões são conduzidas pela transformada de Fourier discreta (TFD), e sua computação e implementação representam um dos pontos críticos para uma eficiente condução da análise do BH, especialmente para sinais quasi-periódicos, i.e., quando existe mais de uma frequência fundamental (ou tom) não-comensurada.

Neste trabalho, apresentaremos, pela primeira vez, uma nova metodologia para a formulação e a solução do problema do BH envolvendo circuitos não-lineares forçados (não-autônomos), bem como, uma discussão detalhada na teoria e na implementação das TFDs para sinais periódicos (espectro de frequência de único-tom) e quasi-periódicos (espectro de frequência multi-tons). Para formulação das equações de circuito, exporemos um procedimento tabular utilizando a técnica de espaço-de-estado. Convém ressaltar que esta formulação produz expressões mais simples no cálculo e na implementação numérica da matriz jacobiana, quando comparada com a formulação de espaço de estado paramétrica (FFEP) introduzida em [18]. Vale destacar que, quando comparada com à representação nodal-modificada [28], a representação de espaço-de-estado para dispositivos não-lineares produz uma formulação mais econômica em termos do número de variáveis de estado. Para a resolução de problemas de grande-escala, introduziremos a técnica de decomposição de circuito multi-níveis. Nesta decomposição, o circuito assume uma estrutura hierárquica, descrito por um sistema de equação não-linear bloco diagonal com borda dupla.

Para a solução do problema do BH empregar-se-á uma nova família de métodos, entitulados métodos do tensor. Estes métodos utilizam a matriz jacobiana e estão fundamentados na solução de um modelo local, com informação de segunda ordem, para determinação de cada nova iteração. Os métodos do tensor possuem um custo computacional comparável ao do tradicional método de Newton, porém, apresentam significantes vantagens teóricas sobre ele em problemas onde a matriz jacobiana é mal-condicionada ou singular na raiz. O método do tensor para a solução de problemas em pequena-escala, via fatorização QR, é apresentado em [29] e para média-escala, via fatorização LU, é apresentado em [30]. Para problemas em grande-escala, foram propostos, mais recentemente, métodos do tensor inexato [31],[32]. O desempenho dos métodos do tensor e do tensor inexato versus o método de Newton e de Newton inexato são comparados em [29],[30],[33]. Já comparações do desempenho dos métodos do tensor inexato versus o método de Newton inexato podem ser encontradas em [32],[31],[34],[35]. Estas comparações revelam uma superioridade dos métodos do tensor. Neste trabalho, motivado por estes resultados, verificaremos o desempenho destes novos métodos na análise de corrente contínua (CC) e do BH envolvendo circuitos integrados não-lineares de RF.

Abaixo, apresentaremos uma breve revisão do atual estado-da-arte do método do BH para análise de circuitos não-lineares forçados, seguido de um resumo descrevendo as extensões deste

método para inclusão da variação térmica e de sinais com portadoras de RF moduladas por complexos esquemas digitais. Para oferecer uma visão comparativa do método do BH, exporemos uma discussão relatando suas vantagens e desvantagens em relação a outros métodos competitivos. Para finalizar, decreveremos as principais contribuições deste trabalho, a organização deste manuscrito e uma descrição suscinta do conteúdo de cada capítulo.

1.1. Análise de Regime Permanente para Circuitos Não-Lineares Forçados

Nesta seção, apresentaremos uma breve revisão da evolução da técnica do BH para análise de regime permanente em circuitos e sistemas não-lineares de RF não-autônomos (ou forçados). Em seguida, exporemos uma breve revisão nas extensões do método do BH para lidar com efeitos térmicos e sinais de RF digitalmente modulados. Finalizaremos com uma breve revisão de outras técnicas concorrentes.

1.1.1. Método do Equilíbrio Harmônico: Estado-da-Arte

Um interessante resumo histórico e comparativo, na evolução do método do BH até o final da década de oitenta, pode ser encontada em [36]. Em adição, revisões detalhadas descrevendo o estado-da-arte deste método, até o período citado, foram reportadas em [37],[38],[39]. Complementando estas revisões, a seguir, apresentaremos um resumo dos significantes avanços até a presente data.

Conforme mencionado acima, o método do BH caracteriza-se pela determinação direta da resposta de regime permanente de um circuito, através da resolução de um sistema de equação não-linear (ou equação determinante). Em acréscimo, assume-se que a resposta de regime permanente pode ser representada em série de *Fourier*, onde os coeficientes (de Fourier) a serem determinados são fasores representando as componentes em frequência das variáveis de estado (tensões e correntes) do circuito. Desta forma, para um circuito em grande-escala, excitado por sinais com formas-de-onda complexas, o cálculo destes fasores pode facilmente resultar em um problema de alto custo computacional, em termos de memória exigida e de tempo de processamento. Apesar da análise do BH ser considerada uma técnica no domínio-da-frequência, tendo em vista que, os vetores de resíduo (associado à equação determinante) e de variável de estado são definidos neste domínio, as funções não-lineares (cargas e fluxos), que representam as não-linearidades existentes no circuito, são descritas no domínio-do-tempo, o que requer a utilização da TFD para a conversão do sinal de tempo para frequência e de frequência para tempo. Para dispositivos eletrônicos, as

funções não-lineares correspondem às correntes de condução (I/V) e de deslocamento (Q/V), ambas dependente de tensão.

Um avanço significante na redução da dimensão do problema do BH foi obtido com a introdução da técnica de decomposição em pedaços (ou por partes) do circuito, citada acima. Nesta técnica, o circuito é dividido em uma parte linear e outra parte não-linear. Com isto, o tamanho da equação determinante pode ser reduzido a uma dimensão igual ao produto do número de variáveis de estado não-lineares vezes o número de linhas espectrais. Lembremos que, as variáveis de estado não-lineares, por definição, são aquelas que atuam como argumento das funções não-lineares. Utilizando a decomposição em pedaços, foi proposta, em [40], a formulação das equações do circuito através da combinação da clássica formulação nodal-modificada (FNM) [28] aplicada à parte linear, com a formulação espaço-estado paramétrica (FEEP) aplicada à parte não-linear, associada àos dispositivos eletrônicos não-lineares (diodos, transistores, etc). Em geral, a aplicação desta técnica de decomposição está limitada ao domínio da frequência. Fundamentada no princípio de *Diakoptics* para análise de circuitos em grande-escala no domínio do tempo, foi proposta a técnica de decomposição multi-níveis [41]-[43], possibilitando a subdivisão de um circuito em uma estrutura hierárquica de sub-circuitos esparsamente interconectados em cada nível de hierarquia. Esta técnica produz um sistema de equações com uma estrutura tipo bloco diagonal com dupla borda que pode ser eficientemente resolvido com o emprego de técnicas de processamento paralelo em sistemas de computação distribuída. A estruturação do problema é obtida com a introdução de um conjunto adicional de variáveis e equações com dimensão igual ao número de pontos de interconexão entre os sub-circuitos. A aplicação desta técnica, no contexto da análise do BH, utilizando apenas um nível de decomposição, foi apresentada em [44].

Um dos primeiros métodos utilizados na solução da equação determinante do BH foi o clássico método de Newton, conforme descrito no trabalho [11]. Posteriormente, visando reduzir considerávelmente o esforço computacional (espaço de memória e tempo de processamento) associado à formação e a fatorização da matriz jacobiana, foi sugerida a aplicação da técnica de relaxação para circuitos, operando em um regime fracamente não-linear. Uma discussão sobre a aplicação dos método de relaxação não-linear tipo bloco *Jacobi* e bloco *Jacobi-Newton* pode ser encontrada em [15],[16],[45]. Convém ressaltar que, neste último método de relaxação, as equações são resolvidas iterativamente, para cada frequência, sendo também conhecido como técnica de relaxação harmônica. Quando as não-linearidades se manifestarem de forma significante, métodos utilizando a matriz jacobiana devem ser empregados, sendo os métodos de Newton e quasi-Newton [17],[46],[19] os mais utilizados. O método da continuação (homotopia) com iteração do tipo previsão-correção também pode ser utilizado, como no exemplo da análise de

um amplificador de microonda classe-C apresentada em [47]. A iteração-previsão pode ser conduzida via extrapolação polinomial [47] ou racional via *Padé* [48], enquanto a iteração-correção pode ser obtida via o método do tensor ou de Newton.

Infelizmente para a análise do BH para circuitos contendo um número significante de sub-redes não-lineares (SRNs) e operando em complexo regime multi-tons, o tamanho da matriz jacobiana, em representação densa, e a sua fatorização podem facilmente possuir uma complexidade numérica que excede o limite computacional do hardware utilizado. Neste caso, para mitigar os custos computacionais, técnicas de matrizes esparsa devem ser utilizadas para o controle do padrão não-zero da matriz jacobiana e, com isto, possibilitar o seu armazenamento e a sua fatorização na forma LU. Esta situação de solução direta do sistema jacobiano, na definição introduzida acima, corresponde aos problemas de média-escala. O controle de esparsidade da matriz jacobiana é obtido com a introdução do conceito de espectro de derivadas [39],[49]. No método linear cêntrico do BH proposto em [20], a matrix jacobiana do tipo bloco diagonal contém apenas informação de CC, i.e., o espectro de derivada é composto apenas pela componente de frequência zero. Este método, na verdade, é equivalente ao método das cordas paralelas [17].

Utilizando o processo desenvolvido em [50], para a redução da ordem de um circuito não-linear operando em regime permanente, o tamanho da matriz jacobiana do BH pode ser diminuído de forma significante. Este processo de redução está fundamentado na introdução de um parâmetro de continuação, permitindo que o vetor de variável de estado e o vetor de função não-linear sejam expandidos em série de *Maclaurin* finita em relação a este parâmetro. Os coeficientes desta expansão são calculados via um processo recursivo para o casamento de potência dos termos da série. Posteriormente, uma base ortonormal do subespaço, definida pelos coeficientes da série, é construída via fatorização QR. Esta base representa a transformação linear utilizada para a redução drástica da dimensão do sistema não-linear, i.e., da ordem da matriz jacobiana. A precisão desta técnica é determinada pelo número de termos utilizados na expansão em série. É importante destacar que, a decomposição de ciruito por partes, que, por sua vez, representa um processo de redução de ordem, não foi aplicada aos exemplos apresentados em [50].

Na análise do BH para circuitos "fortemente" não-lineares e em grande-escala, métodos lineares (ou solucionadores lineares) iterativos devem ser utilizados para resolução dos sistemas jacobianos que emergem a cada iteração. Para esta tarefa, métodos de subespaço de Krylov [27] podem ser utilizados, conforme sugerido em [51], sendo o método do GMRES com reinício [52], um dos mais utilizados. Estes métodos requerem a matriz jacobiana apenas para operações do tipo produto matriz-por-vetor que, por sua vez, podem ser conduzidas de forma explícita ou implícita. Em adição, a cada iteração do método de Newton inexato, a solução do modelo local, que define a

correção de Newton, é aproximada, com o nível de aproximação sendo determinado por um termo forçante. A escolha correta da sequência de termos forçantes é importante na minimização do problema de sobre-resolução ("oversolving"), que tem como objetivo reduzir o número de iterações do solucionador linear [22],[24],[26],[53],[54]. Infelizmente, em geral, os métodos iterativos utilizando subespaço de Krylov podem apresentar uma dificuldade de convergência. Sendo assim, para ampliar a sua robustez, uma eficiente técnica de pré-condicionamento deve ser aplicada. Lembremos que a decomposição multi-níveis pode ser vista como um tipo de pré-condicionador. Isto por que a eficiência do solucionador linear iterativo pode ser significantemente melhorada utilizando a decomposição multi-níveis, que produz uma matriz jacobiana estruturada na forma bloco diagonal com dupla borda. No contexto da análise do BH, os pré-condicionadores mais utilizados são do tipo bloco diagonal (ou Jacobi) [26] e bloco adaptativo [160]. Em acréscimo, o uso de técnicas de matrizes esparsas devem ser utilizadas para uma eficiente formação e fatorização destes pré-condicionadores, utilizando a matriz jacobiana.

A dimensão da equação determinante também pode ser reduzida, simplificando o espectro de frequência utilizado na representação do sinal que descreve as variáveis de estado não-lineares, porém, isto limita a precisão e a aplicabilidade do método. A necessidade da conversão do sinal entre os domínios do tempo e da frequência representa um ponto crucial na implementação, na eficiência e na precisão da análise do BH. Para sinais periódicos, a forma mais adequada (precisa) de se conduzir estas conversões é através da TFD. Para regime periódico, onde temos apenas uma única frequência fundamental ou inúmeras frequências fundamentais harmonicamente relacionadas, a transformada de Fourier rápida (TFR) representa a escolha ótima. Infelizmente, para conversão de sinais quasi-períodicos, esta transformada assume uma estrutura multi-dimensional, onde o número de dimensões é igual ao número de frequências fundamentais não-harmonicamente relacionadas. Devido a este fato, a transformada de Fourier rápida multi-dimensional (TFRM) é geralmente implementada até três tons [55]. Devemos lembrar que, as TFRs uni- e multi-dimensional operam no plano de fase, produzindo uma amostragem ótima. Com o intuito de resolver esta dificuldade, foram propostas outros tipos transformadas. Pode-se destacar a transformada de Fourier quasi-periódica (TFQP) [56],[57], utilizando amostragem no tempo com pontos igulamente espaçados (escolha trivial) e com espaçamento aleatório. Para regime de dois-tons, pode ser adotado o esquema de amostragem proposto em [58]. A escolha do esquema de amostragem no tempo possui um direto impacto no condicionamento desta transformada, conforme discutido por outras implementações da TFQP [59],[60].

Em paralelo ao surgimento da TFQP, foi proposto o método do balanço harmônico modificado (BHM) (tradução nossa do termo em inglês *modified harmonic balance* (MHB)) [61], para análise

de regime permanente com excitação de dois-tons. Entretanto, devido às operações de translação em frequência (fundamentado no teorema de amostragem passa-faixa), este método requer muitas modificações na estrutura padrão da análise do BH. Para complicar ainda mais, ele possui restrições com relação à ordem das não-linearidades e à largura do espaçamento entre as frequências fundamentais que excitam o circuito.

Uma outra possibilidade de implementação das TFDs para sinais quasi-periódicos é através da técnica de mapeamento-em-frequência artificial (MFA). A grande virtude do MFA é possibilitar a condução das conversões tempo-frequência via TFR-1D (uni-dimensional), sem a inclusão de componentes extras de frequência. Para regime de dois-tons, foram desenvolvidos mapeamentos para a topologia de espectro de frequência com grade triangular [62] e retangular [9]. A generalização do mapeamento com grade retangular para regime multi-tons foi apresentada em [63]. Convém citar que, a TFRM também utiliza truncamento com grade hiper-retangular. A técnica de MFA pode ser aplicada na análise de sistemas de telecomunicação com portadora(s) de RF modulada(s) por sinal digital na banda-base, porém, sua implementação requer cuidados para evitar possíveis erros de *aliasing* [64]. Obedecendo a uma condição de amostragem, a transformada de Fourier multi-tons (TFMT), introduzida em [65], oferece uma alternativa para implementação da TFD para sinais de RF digitalmente modulados. A TFMT é conduzida via uma TFR-1D e em contra-posição a técnica de MFA, não está sujeita ao efeito de *aliasing*.

Atualmente, sistemas avançados de telecomunicação empregam complexos esquemas de modulação digital (QPSK, QAM, OFDM, etc [66]), que, por sua vez, envolvem sinais pseudo-aleatórios que não podem ser considerados como regime permanente. Devido à limitação de memória e tempo de processamento, na análise do BH, estes sinais estão limitados a uma representação através de uma sequência periódica de símbolos de comprimento limitado. O uso deste tipo de sinais e/ou multi-senos, possibilitam a análise de recrescimento espectral ou de vazamento de potência em canal adjacente e a determinação de importantes figuras de mérito: *adjacente channel power ratio* (ACPR), *co-channel power ratio* (CCPR), e *noise power ratio* (NPR), em amplificadores e conversores de frequência.

Um interessante estudo apresentado em [67] considera o efeito do truncamento em frequência na deteriorização da convergência e na perda de precisão do método do BH, e oferece, para melhoria da robustez e eficiência, uma técnica de pré-filtragem de harmônicos de fácil implementação numérica. Em circuitos com fortes não-linearidades e circuitos periódicamente chaveados, onde o número de harmônicos gerados com amplitude considerável é muito elevado, o uso desta técnica de pré-filtragem pode ser mandatório. Um outra alternativa para lidar com circuitos envolvendo sinais de resposta com transições rápidas é o método do balanço harmônico

mapeado no tempo (BHMT) (tradução nossa do termo em inglês *time-mapped harmonic balance* (TMHB)) [68]. Neste método, um dos aspectos importantes é a seleção da grade, no tempo que define o mapeamento [69].

O método do BH pode ser utilizado na análise de um sistema de rádio-enlace envolvendo múltiplos transmissores e receptores de RF. Para este propósito, pode-se lançar mão da metodologia apresentada em [70], incluindo o conceito de matriz de transferência do enlance. Esta metotologia considera as diferentes situações para decomposição das estruturas radiantes e não-radiantes. Para simulação de estruturas eletromagnéticas, pode ser utilizado o método da segmentação introduzido em [71]. Um modelo (no tempo-frequência) simplificado para emissão (antena do transmissor + canal) e recepção (antena do receptor), que pode ser imediatamente incorporado à formulação do BH, é descrito em [72]. Um procedimento para caracterização de antenas ultra-banda-larga, utilizando este modelo simplificado, também é descrito em [72].

Vale ressaltar que, qualquer dispositivo não-linear representado por equações de espaço de estado do tipo integral-diferencial-diferença-não-linear pode ser descrito na formulação do BH. Para exemplificar, podemos citar a análise do BH envolvendo dispositivo fotônico do tipo laser semicondutor [73] e sistemas de estruturas micro eletro-mecânicas, também conhecidas como *microelectromechanical systems* (MEMS) [74].

1.1.2. Extensões da Análise do Equilíbrio Harmônico

Devido à limitação de espaço, consideraremos, neste trabalho, que os circuitos a serem analisados são invariantes com a temperatura. Para introduzir o efeito da variação de temperatura no cálculo simultâneo da reposta elétrica e térmica de regime permanente do circuito utilizando a análise do BH, pode ser empregada a metodologia proposta em [75]. Esta metodologia foi validada com a análise de um amplificador operando em regime multi-tons com excitação tipo RF-pulsada. O comportamento térmico do circuito é descrito por um circuito elétrico equivalente (CEE), composto de uma fonte de corrente, representando a potência dissipada no dipositivo e uma matriz de impedância térmica, no domínio s da transformada de Laplace, que pode ser obtida analíticamente, via série dupla de Fourier (SDF) [76] ou numericamente via métodos de diferença finita (DF) [77]. O desenvolvimento de um modelo matemático para estes dispositivos, capaz de descrever a variação das suas características elétricas com a temperatura em regime de grande-sinal, é de crucial importância. Excluindo as fontes de CC, se apenas excitações de RF (sinais de alta velocidade em relação às constantes térmicas) estiverem presentes, então, a matriz de impedância térmica é meramente resistiva e sua determinação é significativamente simplificada.

A análise do BH [75] também é simplificada, uma vez que, a realimentação térmica ocorre apenas na frequência de CC.

Para resolver a dificuldade do método do BH em lidar com análise envolvendo sinais com rápida e lenta taxa de variação no tempo, pode ser empregada a técnica de transiente envoltória (TE) (tradução nossa do termo em inglês *envelope transient* (ET)) [78],[79],[80]. Nesta técnica de solução de equação diferencial parcial multi-tempos (EDPM) [81], com escalas de tempo amplamente separadas, combina-se a análise do BH, escala de tempo-rápida, com a técnica de IDT, escala de tempo-lenta. Mais precisamente, assume-se que os coeficientes de Fourier a serem determinados a cada intervalo da IDT são variantes no tempo correspondente à escala lenta. Desta forma, o espectro de frequência dos coeficientes de Fourier possuem um faixa de frequência definida pela largura de banda do sinal de baixa-taxa. Na simulação de sistemas de telecomunicações via análise de TE, os sinais digitais (informação modulada em BB) correspondem à escala de tempo-lento, enquanto os sinais de RF (portadoras) correspondem à de tempo-rápida. Para uma introdução na técnica de TE, podem ser consultadas [79],[80]. Para determinar o regime permanente (solução periódica) na escala de tempo rápida, pode ser aplicado o método das tentativas. Sistemas de telecomunicação operando com múltiplas portadoras de RF, e.g., OFDM, também podem ser simulados via a técnica de TE, conforme descrito em [82].

Uma eficiente metodologia para análise eletro-térmica via método de TE em circuitos de RF não-lineares excitados por sinais digitais de baixa-velocidade, pode ser encontrada em [83].

1.1.3. Outros Métodos

O método da corrente não-linear (CNL) [84], fundamentado na clássica teoria de *Wiener-Volterra* [85],[86], representa uma eficiente técnica para análise no domínio da frequência de circuitos não-lineares operando em um regime "fracamente" não-linear. Nesta análise, com a introducão do conceito de função de transferência não-linear, as respostas (de regime permanente) de um circuito são expressas em termos de série funcional de *Volterra*, sendo a ordem da expansão em série ditada pela "força" das não-linearidades. Em adição, as funções não-lineares, descrevendo as não-linearidades dos dispositivos eletrônicos (correntes de condução e de deslocamento), estão limitadas a uma representação em série de potência (i.e., série de Maclaurin). A aplicação do método da corrente não-linear para caracterização da distorção harmônica, em circuitos amplificadores de microondas pode ser encontrada em [87],[88]. Recentemente, em [89], a análise de Volterra foi utilizada na caracterização de distorção por intermodulação (DIM) em sistemas de RF operando em regime multi-tons. O método da CNL também pode ser empregado na análise de

circuito linear periódico variante-no-tempo (LPTV), utilizando a série de Volterra variante no tempo, onde os coeficientes da série podem ser calculados via o método do BH. A extensão desta técnica para análise de DIM em conversores de frequência é ilustrada em [90], para um sinal modulante (pertubação) de único-tom e de dois-tons (figuras de desempenho: pontos de interceptação de segunda e terceira-ordem), e, em [91], para uma perturbação com complexa representação multi-tons (figuras de desempenho: rescrescimento espectral, ACPR, CCPR, NPR).

Um eficiente método que combina a análise de Wiener-Volterra com alguns dos princípios típicos da técnica do BH foi introduzido em [92], para o cálculo do regime permanente, sob excitação periódica, de circuitos com fortes não-linearidades.

Fundamentado na técnica de relaxação de Jacobi, o método da substituição proposto em [93], pode ser utilizado no cálculo da resposta de regime permanente de circuitos não-lineares forçados alimentados por sinais multi-tons. A solução em cada iteração é obtida resolvendo-se um associado circuito linear invariante-no-tempo (LIT) em todas as componentes de frequência do sinal. A construção deste circuito associado é realizada aplicando análise de sensitividade nos elementos não-lineares, eliminando as fontes independentes, colocando-as em curto-circuito ou circuito aberto e alimentando com fontes associadas ao erro residual. Este método possui dificuldades de convergência para circuitos com "fortes" não-linearidades. Sendo assim, para ampliar a capacidade de manipulação de potência deste método de relaxação, foi proposta em [94] a construção do circuito associado, aplicando análise de sensitividade nos elementos lineares e introduzindo elementos de compensação em série. O circuito associado também pode ser construído utilizando um software de propósito-geral, e.g. SPICE [95]. Nesta construção são realizadas análise de CC e/ ou análise de transiente e análise de CA em sub-circuitos.

O método do balanço forma-de-onda (BF) (tradução nossa do termo em inglês *waveform balance* (WB)) foi proposto em [96] e, ao contrário do método do BH, na equação determinante os vetores de resíduo e de variável de estado não-linear são definidos no domínio do tempo. Neste método, a matriz jacobiana associada à derivada do vetor de resíduo com relação ao vetor de variável de estado não-linear pode ser facilmente calculada. Para o cálculo do vetor de resíduo, o par de transformada de Fourier (direta e inversa) é requisitado. Para a eliminação da necessidade do cálculo da transformada de Fourier direta (conversão de tempo-para-frequência), foi proposto o método do balanço forma-de-onda modificado (BFM) (tradução nossa do termo em inglês *modified waveform balance* (MWB)) [97]. Com isto, um grande número de pontos de amostragem pode ser utilizado para a formação da TFQP inversa resultando em um melhor esquema de ortogonalização. Vale mencionar que, na análise do BFM, a TFQP é necessária para se obter a equação determinante. O método do balanço de amostra fundamentado-em-convolução (BAFC)

(tradução nossa do termo em inglês *convolution-based sample balance* (CBSB)), introduzido em [98], está diretamente relacionado com o método EF evita a operação em um domínio misto de tempo-frequência, durante o processo de solução. A grande dificuldade com os métodos BF, BFM e BAFC deve-se ao fato da natureza física da matriz jacobiana tornar difícil a exploração de técnicas de matrizes esparsas para a solução de problemas de grande-escala, principalmente em regime multi-tons. Os métodos de BF e BAFC foram originalmente proposto para a determinação do regime permanente de único-tom de circuitos de RF não-lineares forçados.

Um forte competidor do método do BH, é o método do balanço espectral (BE) (tradução nossa do termo em inglês *spectral balance* (SB)) [99],[100],[101],[102],[103],[104]. A grande vantagem deste método, é o fato de ele operar integralmente no domínio da frequência, eliminando a necessidade de conversões tempo-frequência do sinal via transformadas de Fourier especializadas. Entretanto, esta vantagem é resultante de uma limitação deste método, que é a necessidade de modelar as funções não-lineares do dispositivo ativo intrínseco no domínio da frequência, uma vez que, estas funções são modeladas tradicionalmente no domínio do tempo, utilizando complexas expressões algébricas. Todavia, elementos dispersivos em frequência podem ser naturalmente representados. Em acréscimo, as funções não-lineares (descrevendo as não-linearidades do circuito) devem ser aproximadas por expansão polinomial (*Chebyshev* [103]) e/ou racional (*Hermite* [104]). Vale ressaltar que, as expansões racionais de Hermite aproximam não apenas a função, mas também sua derivada, possibilitando resultados mais precisos para a DIM, quando comparado com expansões polinomiais. Estas expansões no domínio do tempo envolvem as operações de adição, subtração, multiplicação e divisão, que correspondem, respectivamente, à adição, subtração, convolução e deconvolução no domínio da frequência. Finalmente, em [104], assumindo uma topologia regular para o espectro de frequência de um sinal de RF modulado por um sinal de baixa velocidade na BB e utilizando geração de vetor de indexação e técnicas de matrizes de convolução, consegue-se um eficiente algoritmo de BE para a determinação de figuras de mérito em regime multi-tons.

Uma revisão do estado-da-arte dos métodos para simulação de sistemas de RF pode ser encontrada em [105],[106].

1.2. Contribuições deste Trabalho

Além da avançada implementação númerica da teoria proposta, desenvolvida em uma estrutura objeto-orientada com a linguagem de programação C++ [107], as principais contribuições deste trabalho para análise do BH em circuitos de RF não-lineares forçados se concentram na formulação

da equação determinante e na aplicação dos novos métodos do tensor para a solução desta equação em pequena, média, e grande-escala.

Com relação à formulação da equação determinante do BH, foi introduzida pela primeira vez uma eficiente técnica para a decomposição multi-níveis de circuitos (ou sistemas) em grande-escala. Nesta técnica de *diakoptics* [108],[109],[110], o circuito é sub-dividido em uma estrutura hierárquica com múltiplos níveis, onde cada nível é composto de super-redes (SuRs)[1] esparsamente interconectadas. Nesta estrutura, além da SuR topo, são utilizadas SuRs intermediárias e SuRs de fundo. Por definição, e sem perda de generalidade, as SuRs intermediárias são formadas por um número arbitrário de SuRs crianças (intermediárias ou de fundo) e por uma rede de conexão (RC). Quando nenhum nível de hierarquia é especificado, o circuito é representado apenas por uma SuR topo que assume a mesma estrutura de um SuR de fundo. Vale destacar que, as SuRs de fundo não possuem hierarquia e são formadas por elementos estruturais de circuito e/ou por redes multi-terminais, representando resultados de dispositivos medidos experimentalmente ou obtidos utilizando análise de estruturas eletromagnéticas de composição e geometria complexa. Para a separação da SuR de fundo em sub-rede linear ampliada (SRLA) e sub-rede não-linear compactada (SRNC), aplicar-se-á a técnica de decomposição por partes.

Para a formulação das equações das SuRs de fundo, estamos propondo uma nova metodologia que combina a clássica FNM, aplicada a sub-rede linear (SRL), e a uma nova formulação espaço-estado (FEE) aplicada a SRN. Inspirada em [112], a FEE proposta é válida para qualquer topologia e segue um procedimento tabular *ad-hoc*, o que resulta em uma fácil implementação númerica. A motivação para o desenvolvimento da FEE deve-se ao fato desta produzir, em geral, um menor número de váriaveis ou equações não-lineares, quando comparada a FNM. Em adição, quando comparada com a FNM-SRL/FEEP-SRN [39], a metodologia proposta resulta em expressões análiticas mais simples para os elementos da matriz jacobiana, o que torna mais fácil sua implementação númerica e mais eficiente o seu cálculo.

Apesar de não discutida neste trabalho, a FEE proposta para as SRNs foi desenvolvida incluindo a presença de fontes de ruído arbitrariamente correlatas [113]. Isto possibilita a análise e a otimização unificada de sinal-ruído e define o caráter generalizado da metodologia introduzida, i.e., uma ferramenta de propósito geral.

A formulação multi-níveis produz uma equação determinante do BH envolvendo matrizes com estrutura bloco diagonal sem borda, com borda simples e com borda dupla de multiplos níveis. A

[1] A terminologia super-rede é inspirada na palavra *superblock* introduzida em [111].

construção destas matrizes segue a seguinte organização sequencial: por SuR intermediária, por frequência e por SuR de fundo. Com esta estrutura, a matriz jacobiana associada à equação determinante mantém o mesmo padrão não-zero após o processo de fatorização LU. No contexto da análise do BH multi-níveis, foram implementadas técnicas para o controle de esparsidade da matriz jacobiana, estabelecendo-se uma relação de compromisso com a capacidade de manipulação de potência. Vale destacar que a formulação multi-níveis proposta possibilita um alto ganho de desempenho em termos de mémoria e de tempo de *central processing unit* (CPU), particularmente, em sistemas distribuídos, onde recursos para processamento paralelo estão à disposição.

A outra contribuição deste trabalho refere-se à aplicação de um nova classe de métodos, entitulados métodos do tensor para a análise do BH não-autônoma, envolvendo problemas de pequena e média-escala [30] e os métodos do tensor inexato para análise do BH em grande-escala [31],[114]. Neste trabalho, investigaremos o desempenho dos métodos do tensor com relação ao método de Newton (ou método padrão) e dos métodos do tensor inexato com relação ao método de Newton inexato. Também serão feitas comparações entre os métodos do tensor e de Newton para a análise de CC dos circuitos testes utilizados neste trabalho, além de serem conduzidos testes para uma avaliação das estratégias de pesquisa-em-linha curvilinear e padrão, empregadas pelos métodos do tensor e do tensor inexato.

Embora não tenha sido o nosso principal objetivo, foi desenvolvido neste trabalho uma TFD para análise de DIM em circuitos conversores-em-frequência (CFs). Esta TFD para regime quasi-períodico multi-tons está fundamentada em uma extensão da TFMT.

1.3. Organização da Tese

Este manuscrito está dividido em nove capítulos e três apêndices. Excluindo este capítulo introdutório e o último capítulo com as observações finais e as sugestões de trabalhos futuros, cada capítulo inícia e termina com uma seção de introdução e uma seção de conclusão, respectivamente. Como de costume, a seção de introdução destaca os principais aspectos a serem discutidos no capítulo e a seção de conclusão é reservada para os argumentos finais. As referências bibliográficas são listadas no final deste manuscrito. Abaixo, apresentaremos um sumário dos tópicos discutidos em cada capítulo.

No Capítulo 2, descrever-se-á a técnica de decomposição multi-níveis para a formulação das equações de circuitos não-lineares em grande-escala. Este capítulo está dividido em cinco seções. Na segunda seção, após a introdução, iniciamos nossa discussão com a descrição da técnica de

decomposição de circuito multi-níveis. Em seguida, na terceira seção, apresentamos a descrição da técnica de decomposição de rede subdivisãopor partes para a subdivisão de uma SuR de fundo em uma SRLA e uma SRNC. A decomposição multi-níveis de um circuito receptor de onda-milimétrica [115] é apresentada, na seção quatro, como um exemplo ilustrativo da metodologia proposta.

O Capítulo 3 está dividido em seis seções. A segunda seção, apresenta a descrição dos elementos estruturais (resistores, indutores, etc) utilizados na composição da SRL e da SRN. A terceira seção, subdividida em duas subseções, apresenta uma breve revisão na formulação, no domínio da frequência, da equação de sonda da SRL (primeira subseção) e das equações de estado e de sonda da SRN (segunda subseção), utilizando a clássica FNM. Na seção seguinte, introduzimos a FEE para a formulação das equações de estado e de sonda da SRN. Esta seção foi subdividida em uma subseção que apresenta a equação topológica, e em uma outra que apresenta as equações de estado e de sonda. O Apêndice A apresenta as tabelas com as relações constitutivas dos elementos básicos utilizados na construção destas equações. A quinta seção apresenta a aplicação da FEE na geração das equações do CEE de grande-sinal de dispositivos do tipo FET e HBT. O Apêndice B contém detalhes desta formulação incluindo dispositivo definido-simbolicamente (DDS) descrito pela FNM. Em DDS, apenas a parte intrínseca, descrita por correntes de condução e de deslocamento (modelo de carga), é representada.

Utilizando os resultados dos Capítulos 2 e 3, no Capítulo 4 será introduzida a formulação das equações do circuito. Para tal, dividimos este capítulo em sete seções. A segunda e terceira seções apresentam a formulação das equações de sonda da SLRA (SRL + rede de ampliação) e das equações de estado e de sonda da SRNC (SRNs + rede de permutação), respectivamente. Na seção subsequente, utilizando estas equações, é apresentada a formulação das equações da SuR de fundo (SRLA + SRNC). Com estes resultados, a quinta seção apresenta a formulação multi-níveis das equações de estado e de sonda do circuito. Na sexta seção, última antes da conclusão, apresentamos um procedimento para de redução de nível.

No Capítulo 5 discutiremos, em oito seções, a formulação multi-níveis do problema do BH para circuitos não-lineares não-autônomos. Inicialmente, na segunda seção, introduzimos as topologias de espectro de frequência utilizados nas análises do BH de único, dois, três e multi-tons. A terceira seção é dedicada à descrição de sinais de RF digitalmente modulados na análise do BH. Associadas às topologias de espectro introduzidas, serão discutidas, na seção quatro, a teoria e a implementação das TFDs para conversão do sinal de tempo-para-frequência e vice-versa. Na seção cinco, a técnica de MFA é discutida. Na sexta seção, apresentaremos a formulação da equação determinante do BH no contexto da decomposição multi-níveis, introduzida nos capítulos

anteriores. Na sétima seção, apresentamos o cálculo analítico da matriz jacobiana e a técnica de controle de esparsidade fundamentado no conceito de espectro de derivadas. O Apêndice C descreve a fatorização e a solução via retro-substituição da matriz jacobiana multi-níveis.

Capítulo 6 foi organizado em oito seções e descreve a teoria e a implementação do método do tensor para análise de CC e do BH envolvendo problemas de pequena e média-escala. Na primeira seção, após a introdução, apresentamos considerações iniciais referentes à natureza do processo de solução para análise do BH em circuitos forçados. Na seção seguinte, introduzimos o conceito de função nível utilizado pela estratégia de globalização via pesquisa-em-linha. Para facilitar a discussão do método do tensor, apresentamos na quarta seção, em duas subseções, uma breve revisão na teoria do método de Newton (método padrão), sendo a primeira subseção referente à globalização via peequisa-em-linha, e a segunda concernente a uma implementação modificada. A quinta seção diz respeito ao método do tensor e foi organizada em quatro subseções. A primeira e a segunda subseções apresentam uma discussão na constução e na solução do modelo do tensor, respectivamente. As duas subseções seguintes discutem a globalização, via pesquisa-em-linha, e apresenta uma implementação modificada. A subseção referente à pesquisa-em-linha é subdividida em duas subseções que apresentam a estratégia padrão e a curvilinear. A sexta seção discute o método da continuação. Em seguida, na sétima seção, apresentamos os testes preliminares para validação númerica do método do tensor. Algoritmos descrevendo a implementação dos métodos de Newton e do tensor são fornecidos e discutidos.

O Capítulo 7 foi organizado em sete seções e descreve a teoria e a implementação do método do tensor inexato para análise do BH, envolvendo problemas de grande-escala. Seguindo a mesma filosofia do capítulo anterior, para facilitar a discussão no método do tensor inexato, iniciamos nossa discussão (segunda seção) com uma breve revisão na teoria do método de Newton inexato (método padrão). Esta discussão foi organizada em cinco subseções. Nas primeiras duas subseções discutimos a solução iterativa do modelo do tensor e o pré-condicionamento para este modelo. Já na terceira subseção, apresentamos uma discussão na escolha da sequência de termos forçantes para solução aproximada do modelo do linear que define a correção de Newton, seguida, na quarta subseção de uma discussão na teoria e implementação da estratégia de globalização via pesquisa-em-linha. Para concluir a segunda seção, na quinta subseção apresentamos a implementação modificada do método Newton inexato. Na terceira seção, dividida também em quatro subseções, iniciamos a discussão no método do tensor inexato. A primeira subseção discute a solução iterativa aproximada do modelo do tensor, empregando técnicas de subespaço de Krylov e pré-condicionamento. Mais precisamente, para resolução do modelo do tensor, discutimos a teoria e a implementação dos processos de solução simplicada, modificada e completa. Em seguida, na

segunda subseção debatemos a escolha da sequência de termos forçantes, onde é sugerida uma nova escolha utilizando a informação do modelo do tensor. Na terceira subseção apresentamos a teoria e implementação da globalização via pesquisa-em-linha utilizando a estratégia padrão e a estratégia curvilinear. Para encerrar a discussão do método do tensor inexato, na quarta subseção apresentamos uma implementação modificada deste método. Já na quarta seção, discutimos o produto matriz jacobiana-vetor na análise do BH utilizando a formulação proposta. Na quinta seção, a implementação de pré-condicionadores para análise do BH. Na sexta seção, apresentamos os testes preliminares utilizados para validação númerica da implementação modificada. algoritmos descrevendo a implementação do método de Newton e do tensor inexato são fornecidos e discutidos detalhadamente.

Finalmente, no Capítulo 8, organizado em cinco seções, são apresentados os exemplos de circuitos não-lineares forçados e os resultados númericos que avaliam o desempenho e validam a teoria proposta neste trabalho. Na segunda seção, apresentamos uma breve discussão na implementação do sistema de *computer-aided design* (CAD) utilizado para obtenção dos resultados. Na terceira seção, apresentamos a descrição dos circuitos testes utilizados na validação da teoria proposta. Para tal, sub-dividimos esta seção em oito subseções. A primeira subseção apresenta exemplos com circuitos básicos, utilizando diodos ou *bipolar-junction transistors* (BJTs). Estes circuitos possuem descrição e resultados reportados na literatura [7],[8],[13]. Devido ao tamanho destes circuitos, em termos de número de dispositivos não-lineares, estes circuitos permitem realizar validações númericas envolvendo sinais com formas-de-onda complexas. A segunda subseção envolve circuitos de RF utilizando dispositivos do tipo *metal semiconductor field effect transistor* (MESFET) e *pseudomorphic high electron mobility* (PHEMT). Até esta seção os circuitos possuem apenas um nível de hierarquia, i.e., compostos apenas de uma SuR de topo do tipo de fundo. A terceira e última subseção apresenta circuitos com representação hierarquica de múltiplos níveis. Estes circuitos estão fundamentados em dispositivos tipo MESFET e HBT. Para finalizar, utilizando os circuitos testes, a quarta seção, apresenta os resultados da análise do BH conduzida com os métodos do tensor e de Newton e com os métodos do tensor inexato e de Newton inexato.

2. Decomposição Multi-Níveis do Circuito

2.1. Introdução

ANTES DE APRESENTARMOS a formulação das equações, introduziremos uma eficiente técnica para a decomposição multi-níveis de circuitos não-lineares em grande-escala, particularmente circuitos integrados (CIs) de alta-velocidade (i.e., operando na faixa de RF, de microonda e de onda milimétrica). A teoria de *diakoptics*, fundamentada no princípio de *dividir-para-conquistar*, foi introduzida em [108], para solução de sistemas físicos em grande-escala. Nesta teoria, o sistema que se deseja analisar é decomposto em subsistemas de menor escala, e a solução é obtida através da composição das soluções individuais de cada um dos subsistemas. Até a presente data, uma grande quantidade de artigos explorando o conceito de diakoptics tem emergido em diferentes campos de pesquisa, destacando-se, e.g., análise de circuitos lineares [116],[117],[118],[119], análise de circuitos não-lineares [120],[121],[41]-[43], análise e otimização de interconexões em circuitos de alta-velocidade [122], diagnóstico de falha em circuitos [123],[124], análise de distribuição de campo em antenas [125],[126], análise eletromagnética de estrutura planares [127],[128], e análise de estruturas eletromagnéticas descritas por CEE [129],[130].

Como extensão do método de Kron, foi introduzida por Bowden [110] uma técnica de decomposição multi-níveis para sistemas hierárquicos. A decomposição multi-níveis possibilita a utilização de tecnologia de multi-processamento e computação distribuída. Lembremos que um sistema multi-processador é um computador, ou um conjunto de computadores, consistindo de vários elementos de processamento (EPs), onde cada EP consiste de uma unidade central de processamento (UCP), uma memória, e um subsistema de entrada/saída (E/S). A computação distribuída é uma concepção mais geral de multi-processamento, na qual os EPs estão ligados através de algum tipo de *local area network* (LAN). A aplicação/implementação de uma rede "transputers" para a solução do problema decomposto por diakoptics, via métodos não-iterativos, é discutida em [109].

Fundamentada na teoria proposta em [110], na Seção 2.2. veremos que o método de decomposição multi-níveis proposto, possibilita que um circuito não-linear em grande-escala possa ser hierarquicamente sub-dividido em sub-circuitos (ou subsistemas) entitulados de SuRs. Estas SuRs são dos seguintes tipos: intermédiaria e de fundo. A SuR intermediária é compostas de SuRs de níveis superiores e de uma rede de conexão, enquanto, a SuR de fundo é composta por

elementos estruturais (lineares e não-lineares) de circuito. Para este tipo SuR será aplicada a técnica de decomposição de rede *por partes* introduzida em [13]. Com esta técnica, discutida na Seção 2.3, uma SuR de fundo pode ser subdividida em parte linear e não-linear, representadas aqui por uma SRLA e uma SRNC, respectivamente. Com o objetivo de ilustrar a metodologia proposta, a Seção 2.4 apresenta um exemplo prático da decomposição multi-níveis aplicada a um circuito transceptor de onda-milimétrica. Também são discutidos procedimentos para decomposição de um sistema completo de comunicação sem fio. As observações finais são reservadas para a Seção 2.5.

2.2. Decomposição do Circuito em Super-Redes

A estrutura da decomposição multi-níveis de circuito, proposta neste trabalho, é ilustrada na Fig. 2.1(a). Como podemos observar, o circuito foi decomposto em uma estrutura hierárquica de múltiplos níveis compostas de SuRs e redes de conexão. Para maior eficiência, assume-se que em cada nível da hierarquia estas SuRs estão esparsamente interconectadas. A representação em árvore da estrutura hierárquica do circuito, pode ser visualizada na Fig. 2.1(b), onde o Nível 0 (zero) se refere ao topo da hierarquia. Na notação adotada, a *a*-ésima SuR localizada no *b*-ésimo nível da hierarquia é indicada por $\mathfrak{S}_{a,\,b}$; sendo assim, a SuR de topo é indicada por $\mathfrak{S}_{0,\,1}$ ou simplesmente $\mathfrak{S}_0$. Similarmente, a *a*-ésima rede de conexão localizada no *b*-ésimo nível da hierarquia é indicada por $\mathcal{C}_{a,\,b}$, sendo assim, a rede de conexão no nível 0 (zero) é indicada por $\mathcal{C}_{0,\,1}$ ou simplesmente $\mathcal{C}_0$.

Conforme citado na introdução, as SuRs podem ser de dois tipos: intermediária ou de fundo. Convém mencionar, que a SuR de topo será do tipo intermediária se o circuito tiver mais de um nível, ou caso contrário será do tipo de fundo. Este último caso, refere-se à situação convencional, sem decomposição multi-níveis, i.e., só existe o nível 0 e o circuito é descrito apenas por uma SuR de topo. Na situação seguinte, o circuito é descrito por uma SuR de topo composta de diversas SuRs de fundo interconectadas via uma rede de conexão.

As SuRs intermediárias se caracterizam por terem hierarquia, e são compostas exclusivamente de SuRs de nível superior (intermediária ou de fundo) e de uma rede de conexão. Sua introdução possibilita a geração de uma estrutura hierárquica multi-níveis. A RC da SuR intermediária (parente) representa as interconexões entre as suas SuRs de nível superior (crianças), e é formada por nós- e/ou ramos-de-decomposição, ver Capítulo 4. Na Fig. 2.1(c), foi representada a SuR intermediária, $\mathfrak{S}_{\upsilon-1,\,i}$, com $1 \le i \le n_{SuR}^{(\upsilon-1)}$, onde $n_{SuR}^{(\upsilon-1)}$ é igual ao número total de SuRs no nível $\upsilon-1$. Esta SuR intermediária é composta de $l-j+1$ SuRs (intermediárias ou de fundo) de nível

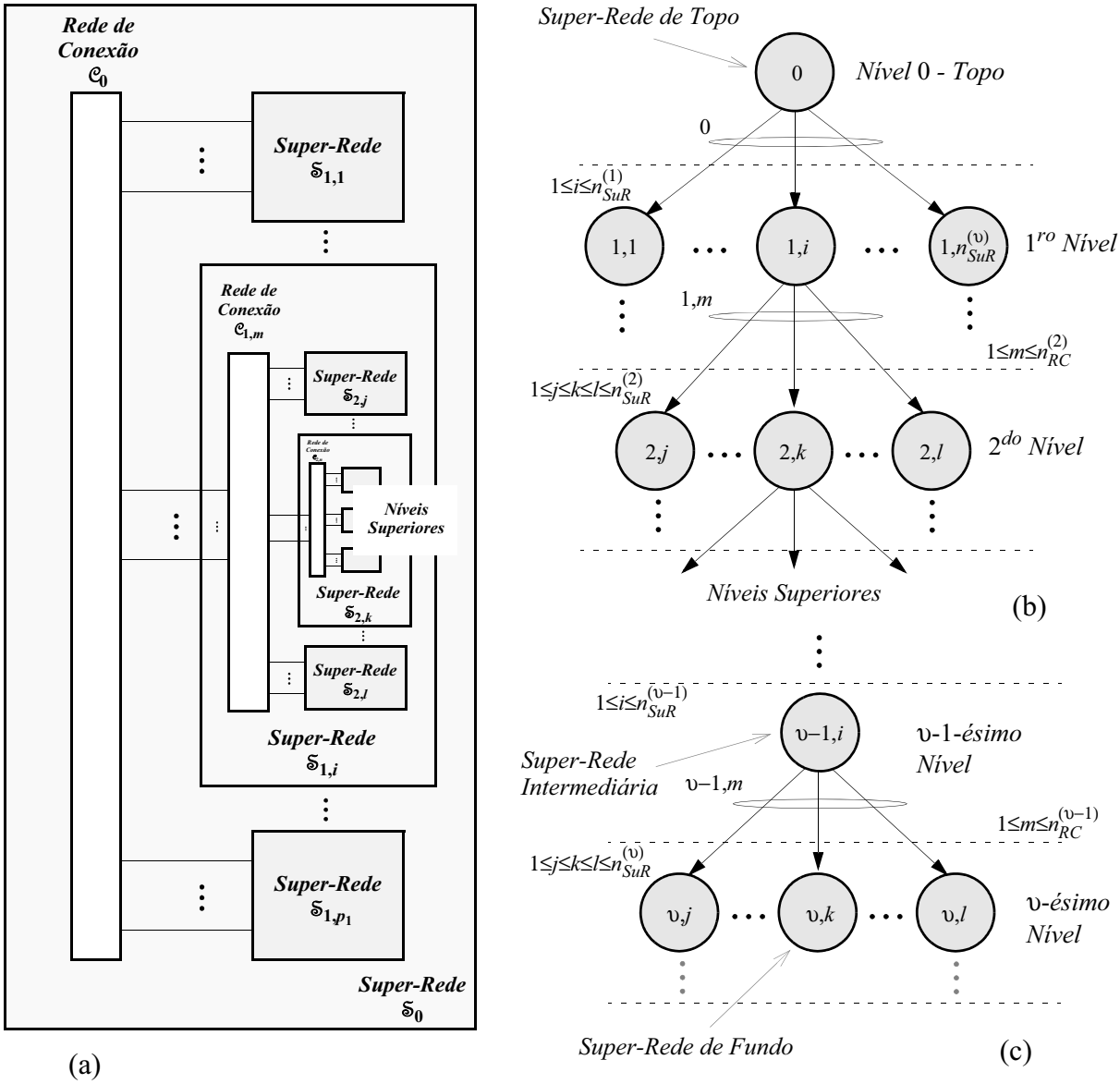


Fig. 2.1 (a) Estrutura da decomposição multi-níveis de um circuito. O circuito é decomposto em super-redes (SuRs) hierarquicamente interconectadas através de redes de conexão. (b) Representação em árvore da estrutura hierárquica do circuito. (c) Detalhe da estrutura hierárquica demonstrando a i-ésima SuR intermediária localizada no nível $\upsilon - 1$ de hierarquia, e a k-ésima SuR de fundo localizada no υ-ésimo nível. Em geral, a j- e l-ésima SuR indicadas podem ser do tipo intermediária ou de fundo.

superior com $1 \le j < l \le n_{SuR}^{(\upsilon)}$, onde $n_{SuR}^{(\upsilon)}$ é igual ao número total de SuRs no nível υ. A RC, $\mathcal{C}_{\upsilon-1,\,m}$, realiza as conexões das SuRs $\mathfrak{S}_{\upsilon,\,j}, \ldots, \mathfrak{S}_{\upsilon,\,l}$, e temos que: $1 \le m \le n_{RC}^{(\upsilon-1)}$, onde $n_{RC}^{(\upsilon-1)}$ é igual ao número total de RCs no nível $\upsilon - 1$.

Sem hierarquia, as SuRs de fundo estão localizadas nos extremos de cada ramo da estrutura hierárquica, conforme ilustrado na Fig. 2.1(c), ver $\mathfrak{S}_{\upsilon,\,k}$ onde $j < k < l$. Estas SuRs são compostas de elementos estruturais lineares (resistor, capacitor, indutor, etc) e não-lineares (resistor não-linear, capacitor não-linear, etc). Como elementos estruturais lineares multi-portas, eles podem conter

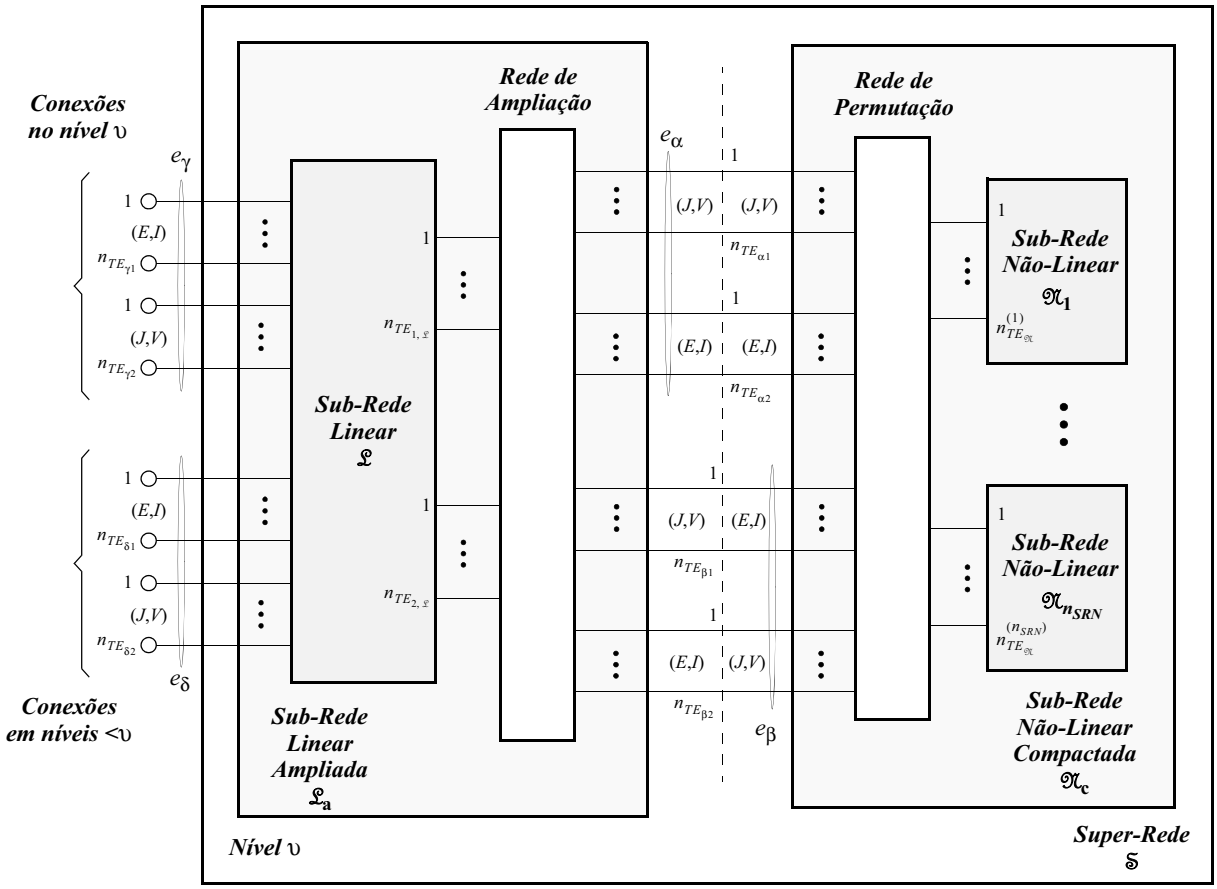


Fig. 2.2 Estrutura de uma super-rede (SuR) de último nível de hierarquia, localizado no nível υ, e decomposta em partes linear e não-linear.

estruturas eletromagnéticas e dispositivos medidos, ambos representados por matrizes híbridas (multi-terminais multi-portas, onde uma porta é formada por um terminal e um nó de referência). Sem hierarquia, e compostas de elementos estruturais lineares (resistor, indutor, etc) e não-lineares (resistor não-linear, capacitor não-linear, etc), as SuRs de fundo estão localizadas nos extremos de cada ramo da estrutura hierárquica, conforme ilustrado na Fig. 2.1(c). Para estes tipos de SuRs, será aplicada a técnica de decomposição de circuito por partes [13]. Lembremos que, nesta decomposição, a SuR de fundo é subdividida em uma SRNA (parte linear) e em uma SRNC (parte não-linear).

2.3. Decomposição por partes das Super-Redes de Fundo

Para descrevermos a técnica de decomposição por partes da SuR de fundo, vamos considerar a representação descrita na Fig. 2.2. De imediato, podemos observar que a SuR de fundo foi decomposta em uma SRLA, $\mathfrak{L}_a$, e em uma SRNC, $\mathfrak{N}_c$, ambas compostas de elementos estruturais

invariantes no tempo. Neste trabalho, as super e sub-redes assumem uma representação multi-terminais, e serão considerados todos os tipos de conexão entre a SRLA e a SRNC. Seguindo o mecanismo sugerido em [131], redes de compensação e de degeneração podem ser incluídas para remover possíveis ambiguidades na interconexão entre a SRLA e a SRNC. A implementação de nó local de referência (NLR) [132], para a análise de circuitos espacialmente distribuídos, está prevista na metodologia discutida.

A parte linear da SuR de fundo, correspondente a SRLA, representa a estrutura de circuito que envolve (completa ou parte intrínseca) os dispositivos semicondutores eletrônicos e fotônicos (e.g., diodos (Schottky, PIN, IMPATT, BARITT, TRAPATT, Gunn, etc), transistores (JFET, MOSFET, LDMOS, MESFET, HEMT, PHEMT, BJT, HBT, etc), diodos lasers (DH, DFB, VCSEL, etc), foto-diodos, etc.), e é composta de uma SRL, $\mathcal{L}$, e uma rede de ampliação. Por outro lado, a parte não-linear, correspondente à SRNC, é composta de n_{SRN} sub-redes não-lineares (SRNs), $\mathcal{N}_i$, $i = 1, \ldots, n_{SRN}$, representando os disposivos semicondutores (completa ou parte intrínseca), e de uma rede de permutação para a correta conexão destes dispositivos com a SRLA. Ressaltamos que, a rede de permutação (sem degeneração-compensação) da SRNC não contém elementos estruturais de circuito e sua função é permutar as terminações externas das SRNs de forma a compatibilizar com a sequência de conexões externas da SRLA. Já a rede de ampliação (sem compensação-degeneração) da SRLA, também não possui elementos estruturais, e sua função é prover diferentes vias para as conexões com as SRNs.

Conforme citado acima, a SRL é constituída de elementos estruturais de circuito concentrados e distribuídos (e.g., resistores lineares e não-lineares, indutores lineares e não-lineares, capacitores lineares e não-lineares, linhas de transmissão, guias de onda, MEMS, etc). Para ser geral, assume-se neste trabalho que os elementos estruturais possam ser descritos por relações constitutivas ou por uma matriz de parâmetros híbridos descrevendo as relações de transferência entre todos os terminais do elemento estrutural (ou dispositivo). Sendo assim, dispositivos caracterizados experimentalmente ou via simulação eletromagnética (EM) podem ser facilmente incluídos na composição da SRL sob forma de uma rede multi-terminais (i.e., multi-portas, onde cada porta possui um nó de referência local ou global).

Como podemos observar na Fig. 2.2, os terminais externos da SuR de fundo, e_γ e e_δ, são providos pela SRL. Os terminais externos, e_γ, são utilizados para conexão com outras SuRs do mesmo nível de hierarquia, enquanto os terminais externos, e_δ, são utilizados para conexão com SuRs de níveis inferiores. Os subescritos "$\gamma 1$" e "$\delta 1$" indicam terminais externos com representação fonte de tensão, *E*, e sonda de corrente, *I*, e os subescritos "$\gamma 2$" e "$\delta 2$" indicam com representação fonte de corrente, *J*, e sonda de tensão, *V*. O número de terminais externos da SuR

de fundo é igual a $n_{TE_{\gamma 1}} + n_{TE_{\gamma 2}} + n_{TE_{\delta 1}} + n_{TE_{\delta 2}}$.

Também podemos observar na Fig. 2.2, que os terminais externos e_{α} e e_{β} representam as interconexões entre a SRLA e a SRNC. Os subescritos "$\alpha 1$" e "$\alpha 2$" indicam mesma representação fonte/sonda para SRLA e SRNC. Os subescritos "$\beta 1$" e "$\beta 2$" indicam a situação contrária. Como podemos observar estes subescritos indicam as condições de contorno a serem utilizadas na formulação das equações do circuito, e são classificados da seguinte forma: ($\alpha 1$) SRLA - fonte de corrente, *J*, e sonda de tensão, *V*, e SRNC - fonte de corrente, *J*, e sonda de tensão, *V*; ($\alpha 2$) SRLA - fonte de tensão, *E*, e sonda de corrente, *I*, e SRNC - fonte de tensão, *E*, e sonda de corrente, *I*; ($\beta 1$) SRLA - fonte de corrente, *J*, e sonda de tensão, *V*, e SRNC - fonte de corrente, *E*, e sonda de tensão, *I*; e ($\beta 2$) SRLA - fonte de tensão, *E*, e sonda de corrente, *I*, e SRNC - fonte de corrente, *J*, e sonda de tensão, *V*. Ainda com relação a Fig. 2.2, podemos observar que a SRL possui $n_{TE_{1,\mathfrak{L}}} + n_{TE_{2,\mathfrak{L}}}$ terminais externos que representam os pontos de conexão com as SRNs, ou com a SRNC via a rede de ampliação. Sendo assim, podemos escrever a seguinte relação, $n_{TE_{\alpha 1}} + n_{TE_{\alpha 2}} + n_{TE_{\beta 1}} + n_{TE_{\beta 2}} = n^{(1)}_{TE_{\mathfrak{N}}} + \ldots + n^{(n_{SRN})}_{TE_{\mathfrak{N}}}$.

2.4. Exemplo Ilustrativo

Para ilustrar a técnica de decomposição multi-níveis, introduzida acima, vamos considerar o exemplo da Fig. 2.3, que consiste do estágio de recepção de um real transceptor de onda-milimétrica [133]. Para realização deste transceptor, que opera em 60 GHz (banda-V), foram empregados transistores bipolares em tecnologia de silício-germânio (SiGe) de 0,12μm, e os seguintes CIs foram fabricados: amplificador de baixo-ruído (ABR), conversor de descida (bloco triplicador de frequência, acoplador direcional, conversor em frequência e ABR2), oscilador controlado por tensão (OCT) e amplificador de potência (AP). Em adição, este transceptor foi idealizado para implementação de uma *personal area network* (PAN) com comunicação sem fio de alta taxa-de-transmissão, e o circuito do receptor possui um total de 61 transistores.

Como podemos observar, na Fig. 2.3(a), o circuito foi decomposto em uma estrutura hierárquica de 3 (três) níveis. No primeiro nível, temos a SuR de topo, $\mathfrak{s}_0$, representando o circuito do receptor. Esta SuR é composta dos seguintes elementos: RC, $\mathfrak{C}_0$, formada pelos nós de conexão $n_{0,1}$-$n_{0,8}$, SuR intermediária, $\mathfrak{s}_{1,1}$, com 1 nível de hierarquia, SuR intermediária, $\mathfrak{s}_{1,2}$, com 2 níveis de hierarquia, SuR intermediária $\mathfrak{s}_{1,3}$ com 2 níveis de hierarquia e SuR intermediária, $\mathfrak{s}_{1,4}$, referente à unidade banda-base (UBB), que não será definida aqui. Como podemos observar, na Fig. 2.3(b), o primeiro nível é formado apenas por SuRs intermediárias.

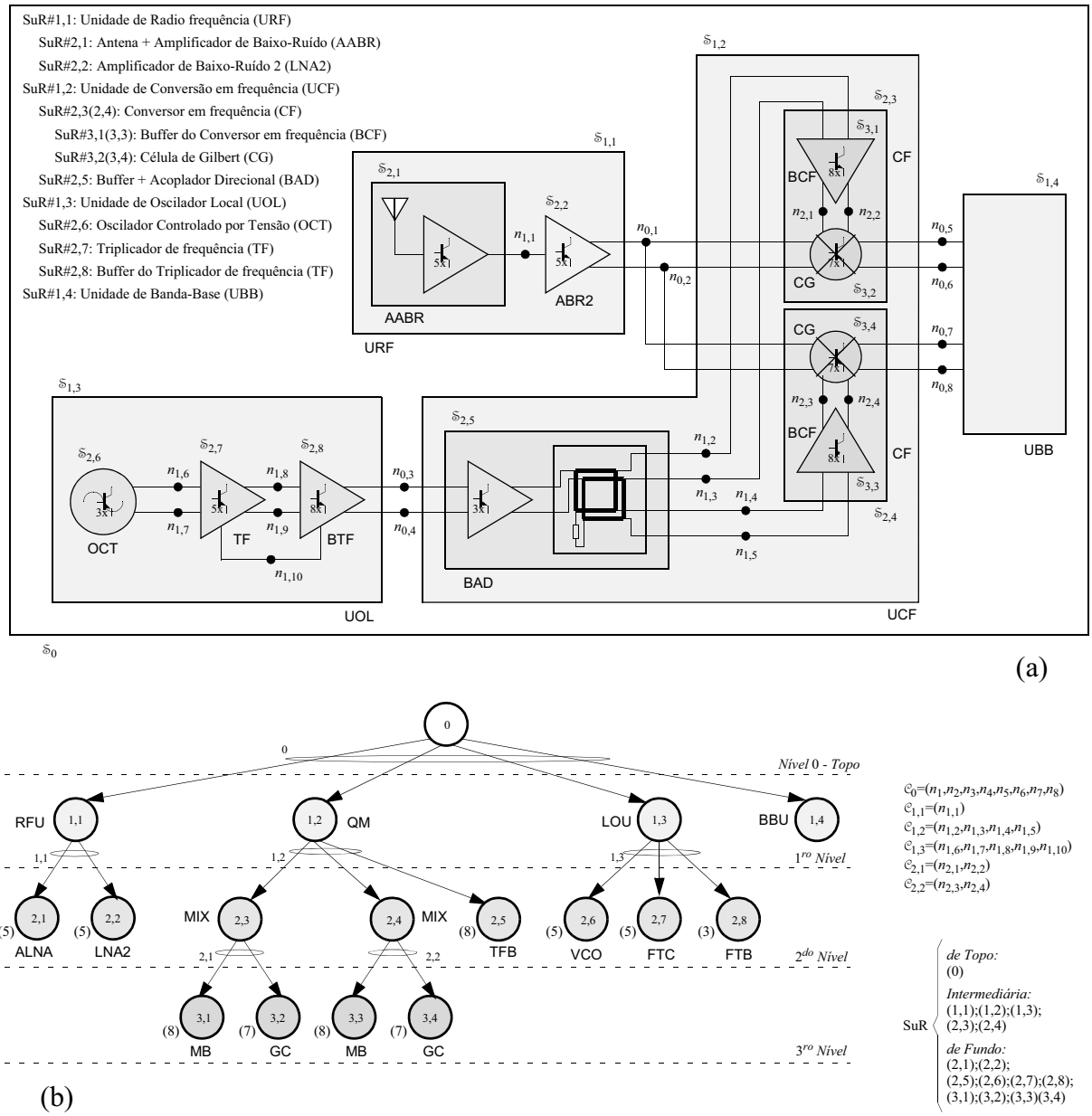


Fig. 2.3 (a) Exemplo da decomposição de circuito multi-níveis aplicado ao estágio de recepção de um transceptor de RF. (b) Estrutura em árvore da decomposição multi-níveis do estágio de recepção, ilustrado em (a).

No segundo nível, temos a SuR intermediária, $\mathfrak{S}_{1,1}$, referente à unidade de rádio frequência (URF), composta dos seguintes elementos: RC, $\mathcal{C}_{1,1}$, formada pelos nós de conexão $n_{1,1}$, SuR de fundo, $\mathfrak{S}_{2,1}$, (com 5 transistores) referente à antena + amplificador de baixo-ruído (AABR) e SuR de fundo, $\mathfrak{S}_{2,2}$, (com 5 transistores) referente ao ABR2. A SuR intermediária, $\mathfrak{S}_{1,2}$, referente à unidade de conversão em frequência (UCF) é composta dos seguintes elementos: RC, $\mathcal{C}_{1,2}$, formada pelos nós de conexão $n_{1,2}$-$n_{1,5}$, SuR intermediária, $\mathfrak{S}_{2,3}$, com 1 nível de hirarquia, SuR intermediária, $\mathfrak{S}_{2,4}$, com 1 nível de hierarquia e SuR de fundo, $\mathfrak{S}_{2,5}$, (com 3 transistores) referente à *buffer* + acoplador direcional (BAD). A SuR intermediária, $\mathfrak{S}_{1,3}$, referente à unidade de oscilador

local (UOL) é composta dos seguintes elementos: RC, $\mathcal{C}_{1,3}$, formada pelos nós de conexão, $n_{1,6}$-$n_{1,9}$, SuR de fundo $\mathfrak{S}_{2,6}$ (com 4 transistores), referente ao OCT, SuR de fundo, $\mathfrak{S}_{2,7}$ (com 5 transistores), referente à triplicador de frequência (TF) e SuR de fundo, $\mathfrak{S}_{2,8}$ (com 8 transistores), referente à *buffer* do triplicador de frequência (BTF).

Finalmente, no terceiro e último nível da hierarquia, temos a SuR intermediária, $\mathfrak{S}_{2,3}$, que é composta dos seguintes elementos: RC $\mathcal{C}_{2,3}$ formada pelos nós de conexão, $n_{2,1}$-$n_{2,2}$, SuR de fundo, $\mathfrak{S}_{3,1}$ (com 8 transistores), referente ao *buffer* do conversor em frequência (BCF) e SuR de fundo, $\mathfrak{S}_{3,2}$ (com 7 transistores), referente à célula de *Gilbert* (CG). A SuR intermediária, $\mathfrak{S}_{2,4}$, é composta dos seguintes elementos: RC, $\mathcal{C}_{2,4}$, formada pelos nós de conexão $n_{2,3}$-$n_{2,4}$, SuR de fundo, $\mathfrak{S}_{3,3}$, referente à BCF, e SuR de fundo, $\mathfrak{S}_{3,4}$, referente à CG.

2.5. Conclusão

Uma eficiente técnica para a decomposição multi-níveis de circuitos (ou sistemas) em grande-escala, foi apresentada. Conforme demonstrado acima, com a aplicação desta técnica de decomposição, o circuito assume uma estrutura hierárquica de multiplos níveis, onde cada nível é formado por um conjunto de SuRs esparsamente interconectadas. Dois tipos de SuR foram introduzidas, a saber: intermediária e de fundo. A SuR intermediária é formada por SuRs (intermediárias e/ou de fundo) de nível superior e por uma rede de conexão. Enquanto a SuR de fundo, contento os elementos estruturais do circuito, é composta de uma SRLA (parte linear) e de uma SRNC (parte não-linear) resultante da aplicação da técnica de decomposição de rede por partes. A rede de conexão da SuR intermediária (parente) representa as interconexões entre as SuRs de nível superior (crianças).

Para ilustração da técnica proposta, a decomposição multi-níveis foi aplicada a um exemplo de circuito transceptor de onda-milimétrica.

3. Formulação das Equações das Sub-Redes Linear e Não-Linear

3.1. Introdução

NESTE CAPÍTULO, DICUTIREMOS A ESTRUTURA E APRESENTAREMOS a formulação, no domínio da frequência, das equações governantes da SRL e da SRN introduzidas anteriormente. Para tal, serão discutidas a clássica FNM [28],[134] e uma nova FEE [112]. Ambas as formulações seguem um procedimento tabular, porém para esta última a sua construção é mais elaborada. A FNM introduzida em [28], representa uma eficiente técnica para geração das equações governantes de um circuito no domínio do tempo ou no domínio da frequência. Em geral, esta formulação produz um sistema de equações bem condicionado, com dominância diagonal. A FNM pode ser empregada na derivação das equações da SRN e da SRL, sendo as variáveis de estado tensões de nó e correntes de ramo. Por envolver tensões de nó, em geral, são necessárias duas variáveis de estado para representar uma tensão de controle. Na FEE as variáveis de estado são tensões de ramo e correntes de ramo, o que torna essa formulação mais eficiente na descrição de uma SRN. Porém, quando consideramos a descrição de uma SRL com elementos densamente interconectados a situação é reversa, i.e., a FNM frequentemente resulta em uma formulação mais compacta, mais fácil de ser gerada e mais bem condicionda, quando comparada com a FEE. Com relação à FEEP [18] para a SRN, a FEE possui uma forma mais simples e não resulta em um sistema de equação não-linear retangular.

Na próxima seção, iniciamos nossa discussão com a introdução da estrutura de circuito da SRL e da SRN. Esta estrutura será formada por um conjunto de elementos básicos, incluindo rede multi-portas para inclusão de estruturas EMs ou resultados experimentais. Na Seção 3.4 será apresentada uma breve revisão da FNM, e sua aplicação na determinação da equação de sonda da SRL (Seção 3.4.1), e das equações de estado e de sonda da SRN (Seção 3.4.2). Em seguida, na Seção 3.5, apresentaremos a FEE para a descrição da SRN, iniciando com a determinação da equação topológica (Seção 3.5.1), e posteriormente com a determinação das equações de estado e de sonda (Seção 3.5.2). Em programas comerciais, sob forma de dispositivo definido-simbolicamente (DDS), a formulação da SRN está restrita apenas à descrição da parte intrínseca dos dispositivos semicondutores (diodos, transistores, lasers, etc). Na formulação proposta, a parte extrínseca pode ser facilmente eliminada, reduzindo a complexidade da formulação da equação do circuito, quando existirem vários dispositivos eletrônicos não-lineares iguais no mesmo circuito. As equações finais de estado e de sonda obtidas via um processo de redução, envolvem apenas variáveis de estado

não-lineares.

Para finalizar, na Seção 3.6 serão apresentados exemplos ilustrando a aplicação da teoria proposta na formulação das equações de transistores de alta-velocidade do tipo FET e HBT. As conclusões são reservadas para a Seção 3.7.

3.2. Considerações Iniciais

Neste trabalho, vamos considerar que os dispositivos eletrônicos, parte intrínseca (não-linear) constituída da região ativa descrita por correntes de deslocamento e de condução [135], e parte extrínseca (linear) representando a estrutura parasita de acesso a região ativa, são descritos por meio de um CEE [136]. Uma outra possibilidade, é o modelo físico fundamentado na representação dinâmica da região ativa através das equações físicas (equação de continuidade-elétron ou buraco, equacão de *Poisson*, equação do *momentum*, equação de conservação de energia) que descrevem o funcionamento eletrônico desta região. Estas equações são geradas após um processo de discretização espacial em escala microspocópica em relação às metalizações (parte extrínseca) das vias de acesso ao dispositivos. Procedimentos quasi-estático (QE) e não-quasi-estático (NQE), no domínio da frequência, para a incorporação em um simulador de circuito de dispositivos descritos por um modelo físico ou por um CEE, são discutidos em [137]. A incorporação de um modelo físico descrito por relações implícitas na análise de EH é descrito em [138]. Para evitar a complexidade da simulação física de dispositivo em 3D e 2D, mantendo um razoável nível de precisão numérica, o problema é comumente aproximado por uma estrutura quasi-2D (ver referência [139]). Vale ressaltar que, a solução das equações do modelo físic, pode ser utilizada na geração de um CEE, conforme demonstrado em [139].

Em adição, o processo de solução discutido neste trabalho, se concentra na situação em que o simulador de circuito utiliza os resultados de simulador de estruturas EM via o método da segmentação [71]. A situação inversa também é possível, via o método de compressão [140],[71], porém, isso resulta em problemas com um significante maior número de equações .

3.3. Estrutura das Sub-Redes Linear e Não-Linear

Sem perda de generalidade, e levando em conta as considerações acima, neste trabalho iremos assumir que a SRL e a SRN possuem a estrutura descrita na Fig. 3.1. Como podemos observar, os terminais externos destas sub-redes são do tipo alimentado-por-tensão e alimentado-por-corrente.

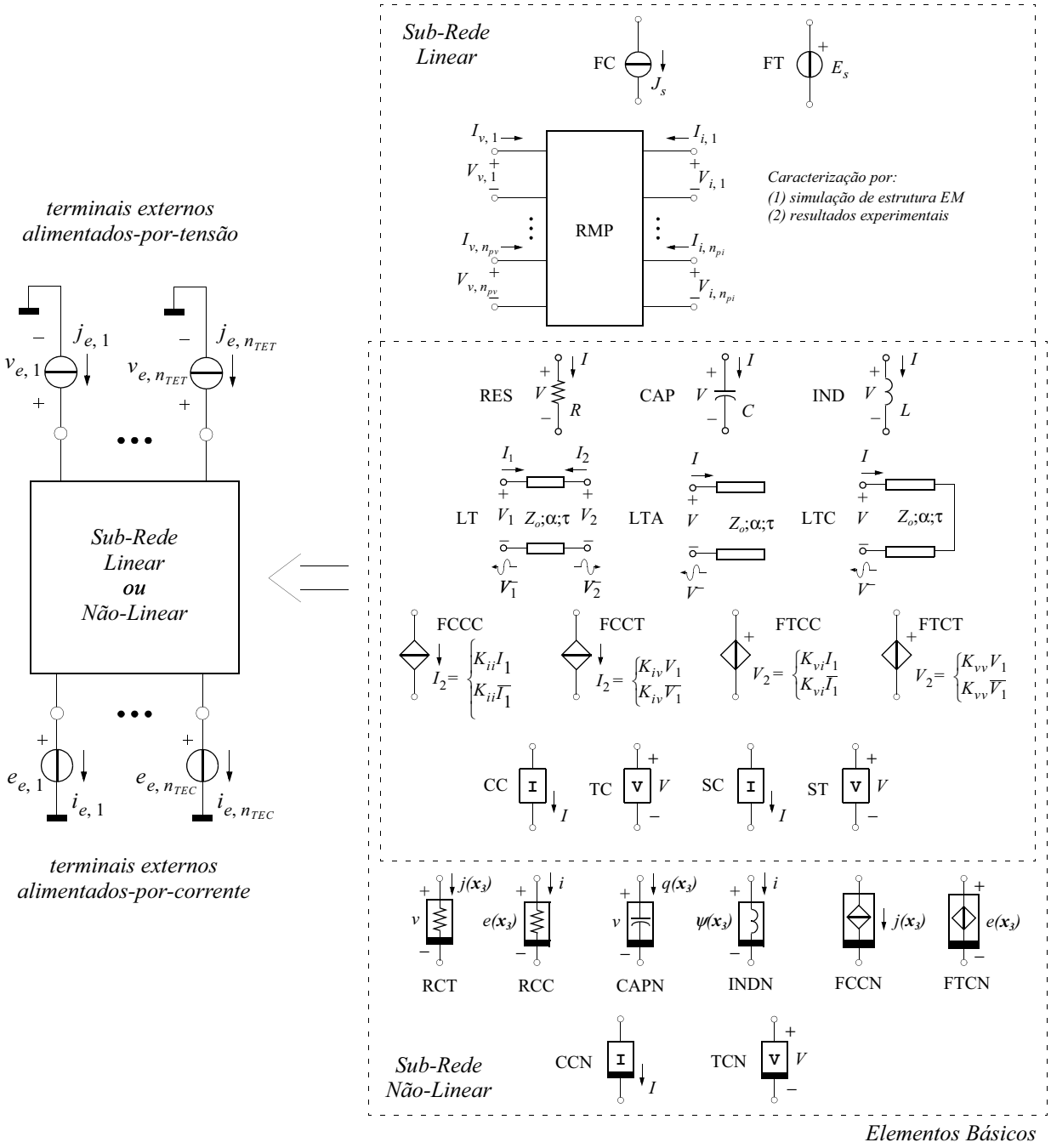


Fig. 3.1 Estrutura geral de uma sub-rede constituída de elementos básicos estruturais, onde n_{TET} e n_{TEC} representam o número de terminais externos alimentados-por-tensão e alimentados-por-corrrente, respectivamente. Os elementos básicos estruturais para formação da sub-rede não-linear (SRN) e da sub-rede linear (SRL) são destacados.

Este primeiro tipo de terminais externos são compostos de uma fonte de tensão externa (FTE) e uma sonda de corrente externa (SCE), enquanto os terminais externos alimentado-por-corrente são compostos de uma fonte de corrente externa (FCE) e sonda de tensão externa (STE). Os elementos básicos para a formação da estrutura interna da SRL e da SRN, são: fonte de corrente (FC), fonte de tensão (FT), resistor (RES), capacitor (CAP), indutor (IND), linha de transmissão (LT), linha de transmissão em circuito aberto (LTA), linha de transmissão em curto-circuito (LTC), fonte de

tensão controlada (FTC), fonte de corrente controlada (FCC), tensão de controle (TC), corrente de controle (CC), sonda de tensão (ST), sonda de corrente (SC), resistor controlado por tensão (RCT), resistor controlado por corrente (RCC), capacitor não-linear (CAPN), indutor não-linear (INDN), fonte de tensão controlada não-linear (FTCN), fonte de corrente controlada não-linear (FCCN), tensão de controle não-linear (TCN), corrente de controle não-linear (CCN), e rede multi-porta (RMP). Vale ressaltar que, a fonte de corrente controlada por corrente (FTCT) é formada por uma FT e uma TC, a fonte de tensão controlada por corrente (FTCC) é formada por uma FT e uma CC, a fonte de corrente controlada por tensão (FCCT) é formada por uma FC e uma TC, e a fonte de corrente controlada por corrente (FCCC) é formada por uma FC e uma CC.

A introdução da RMP como elemento básico, possibilita incluir, na simulação de circuito, estruturas EMs caracterizadas via resultados experimentais ou numéricos (simulação em CAD). Nesta última, os métodos mais utilizados são: elemento finito (EF) (*finite element* (FE)) [141], diferença finita (DF) (*finite difference* (FD)) [142], integração finita (IF) (f*inite integration* (FI)) [143], matriz linha de transmissão (MLT) (*transmission line matrix* (TLM)) [144] ou elemento de contorno (EC) (*boundary element* (BE)). Em alta-frequência, a caracterização experimental é conduzida com base nos parâmetros de espalhamento (formulação em termos de ondas de potência), que podem ser facilmente convertidos para os parâmetros híbridos utilizados internamente pelos simuladores de circuito com formulação em termos de tensões e correntes. Em adição, a RMP também possibilita a inclusão do ruído de substrato em circuitos integrados de alta-velocidade.

Antes de apresentarmos as formulações citadas acima, vamos introduzir o vetor de espaço de estado nos domínios do tempo e frequência $\boldsymbol{x}(t) \bullet\!\!-\!\!\circ\ \boldsymbol{X}(\omega)$ e o vetor de função não-linear $\boldsymbol{u}_f(\boldsymbol{x}(t)) = \left[\boldsymbol{u}_{fe}(\boldsymbol{x}(t))\ \boldsymbol{u}_{fd}(\boldsymbol{x}(t))\right]^T \bullet\!\!-\!\!\circ\ \boldsymbol{U}_f(\omega)$, composto por funções estáticas, $\boldsymbol{u}_{fe}(t) \in \mathbb{R}^{n_{FNE}}$, e dinâmicas, $\boldsymbol{u}_{fd}(t) \in \mathbb{R}^{n_{FND}}$, onde n_{FNE} é o número de funções nã-lineares estáticas e n_{FND} é o número de funções não-lineares dinâmicas. Lembremos que $\boldsymbol{X}(\omega) \in \mathbb{C}^{n_{VE}}$ e $\boldsymbol{U}_f(\omega) \in \mathbb{C}^{n_{FN}}$, onde n_{VE} é o número de variáveis de estado e $n_{FN} = n_{FNE} + n_{FND}$ é o número de funções não-lineares. O vetor de função estática é composto de fontes de corrente controladas não-linearmente, $\boldsymbol{j}_f(t) \bullet\!\!-\!\!\circ\ \boldsymbol{J}_f(\omega)$, e de fontes de tensão controlada não-linearmente, $\boldsymbol{e}_f(t) \bullet\!\!-\!\!\circ\ \boldsymbol{E}_f(\omega)$, tal que: $\boldsymbol{j}_f(t), \boldsymbol{e}_f(t) \in \boldsymbol{u}_{fe}(t)$. Além disso, o vetor de função dinâmica é composto de fontes de carga controlada não-linearmente, $\boldsymbol{q}_f(t) \bullet\!\!-\!\!\circ\ \boldsymbol{Q}_f(\omega)$, e de fontes de fluxo controlado não-linearmente, $\boldsymbol{\phi}_f(t) \bullet\!\!-\!\!\circ\ \boldsymbol{\Phi}_f(\omega)$, tal que: $\boldsymbol{q}_f(t), \boldsymbol{\phi}_f(t) \in \boldsymbol{u}_{fd}(t)$. Em resumo, as funções estáticas estão associadas aos resistores não-lineares controlados por tensão e controlados por corrente, e às fontes de corrente e de tensão de controle não-linear. As funções de carga e de fluxo que descrevem os capacitores e os indutores não-lineares, respectivamente, são denominadas de funções dinâmicas.

As sondas estão organizadas em dois grupos, são eles: sondas externas, referenciadas pelo subescrito "*e*", e sondas internas referenciadas pelo subescrito "*i*". As sondas externas encompassam as sondas de tensão e de corrente associadas aos *terminais externos alimentados-por-corrente* e aos *terminais externos alimentados-por-tensão*, respectivamente, ver Fig. 3.1. As sondas internas permitem o acesso às tensões e às correntes dentro da SRN. Sendo assim, podemos escrever o vetor de sonda da seguinte forma

$$\boldsymbol{Y}(\omega) = \left[\boldsymbol{Y}_e(\omega)\ \boldsymbol{Y}_i(\omega)\right]^T \in \mathbb{C}^{n_S}, \tag{3.1}$$

onde $\boldsymbol{Y}_e(\omega) \in \mathbb{C}^{n_{TE}}$ e $\boldsymbol{Y}_i(\omega) \in \mathbb{C}^{n_{SI}}$ são sub-vetores associados às sondas externas e internas, respectivamente. Lembremos que $\boldsymbol{Y}(\omega) \in \mathbb{C}^{n_S}$, onde $n_S = n_{SE} + n_{SI}$ é o número de sondas, n_{SE} é o número de terminais externos, e n_{SI} é o número de sondas internas. Vale ressaltar que $n_{SE} = n_{TE}$, onde n_{TE} número de terminais externos.

3.4. Formulação Nodal-Modificada

Inicialmente vamos discutir a derivação das equações da SRL e da SRN utilizando a estabelecida FNM, introduzida em [28]. Esta formulação envolve a solução de um sistema linear de equações diagonalmente dominante, e sua descrição detalhada pode ser encontrada em [134]. Sendo assim, nesta seção, iremos nos restringir a uma suscinta descrição desta formulação, incluindo rede multi-terminais.

Fundamentado na Fig. 3.1, iremos assumir que a topologia de uma sub-rede consiste dos seguintes ramos:

• Ramos de corrente de controle de fonte controlada não-linear,
representados pela matriz de incidência $\mathbf{A}_{cix}$ e pelas relações definidas por $\boldsymbol{V}_{cix} = \mathbf{0}$.
• Ramos de tensão de controle de fonte controlada não-linear,
representados pela matriz de incidência $\mathbf{A}_{cvx}$ e pelas relações definida por $\boldsymbol{I}_{cvx} = \mathbf{0}$.
• Ramos de fonte de corrente controlada não-linear e resistor controlado por tensão,
representados pela matriz de incidência $\mathbf{A}_{fj}$ e pelas relações definida por $\boldsymbol{I}_{fj} = \boldsymbol{J}_f(\omega)$.
• Ramos de fonte de tensão controlada não-linear e resistor controlado por corrente,
representados pela matriz de incidência $\mathbf{A}_{fe}$ e pelas relações definida por $\boldsymbol{V}_{fe} = \boldsymbol{E}_f(\omega)$.
• Ramos de capacitor não-linear,
representados pela matriz de incidência $\mathbf{A}_{fq}$ e pelas relações definida por $\boldsymbol{I}_{fq} = \hat{\jmath}\omega\boldsymbol{Q}_f(\omega)$.
• Ramos de indutor não-linear,
representados pela matriz de incidência $\mathbf{A}_{f\phi}$ e pelas relações definida por $\boldsymbol{V}_{f\phi} = \hat{\jmath}\omega\boldsymbol{\Phi}_f(\omega)$.
• Ramos de admitância que inclui todos elementos representados por uma admitância,
representados pela matriz de incidência $\mathbf{A}_y$ e pelas relações constitutivas $\boldsymbol{I}_y = \boldsymbol{K}_y(\omega)\boldsymbol{V}_y$.
• Ramos de impedância que inclui todos elementos representados por uma impedância,
representados pela matriz de incidência $\mathbf{A}_z$ e pelas relações constitutivas $\boldsymbol{V}_z = \boldsymbol{K}_z(\omega)\boldsymbol{I}_z$.
• Ramos de correntes de controle de fontes de tensão controlada por corrente,
representados pela matriz de incidência $\mathbf{A}_{civ}$ e pelas relações definidas por $\boldsymbol{V}_{civ} = \mathbf{0}$.

• Ramos de correntes de controle de fontes de corrente controlada por corrente,
representados pela matriz de incidência $\mathbf{A}_{cii}$ e pelas relações definidas por $\boldsymbol{V}_{cii} = \mathbf{0}$.
• Ramos de tensões de controle de fonte de tensão controlada por tensão,
representados pela matriz de incidência $\mathbf{A}_{cvv}$ e pelas relações definida por $\boldsymbol{I}_{cvv} = \mathbf{0}$.
• Ramos de tensões de controle de fonte de corrente controlada por tensão,
representados pela matriz de incidência $\mathbf{A}_{cvi}$ e pelas relações definida por $\boldsymbol{I}_{cvi} = \mathbf{0}$.
• Ramos de fontes de tensão controlada por corrente,
representados pela matriz de incidência $\mathbf{A}_{vi}$ e pelas relações definida por $\boldsymbol{V}_{vi} = \boldsymbol{K}_{vi}(\omega)\boldsymbol{I}_{vi}$.
• Ramos de fontes de corrente controlada por corrente,
representados pela matriz de incidência $\mathbf{A}_{ii}$ e pelas relações definida por $\boldsymbol{I}_{ii} = \boldsymbol{K}_{ii}(\omega)\boldsymbol{I}_{cii}$.
• Ramos de fontes de corrente controlada por tensão,
representados pela matriz de incidência $\mathbf{A}_{iv}$ e pelas relações definida por $\boldsymbol{I}_{iv} = \boldsymbol{K}_{iv}(\omega)\boldsymbol{V}_{iv}$.
• Ramos de fontes de tensão controlada por tensão,
representados pela matriz de incidência $\mathbf{A}_{vv}$ e pelas relações definida por $\boldsymbol{V}_{vv} = \boldsymbol{K}_{vv}(\omega)\boldsymbol{V}_{cvv}$.
• Ramos de currente da rede multi-portas,
representados pela matriz de incidência $\mathbf{A}_{ih}$ e por relações constitutivas definidas por $\boldsymbol{I}_{ih}(\omega) = \boldsymbol{H}_{iv}(\omega)\boldsymbol{V}_{ih}(\omega) + \boldsymbol{H}_{ii}(\omega)\boldsymbol{I}_{vh}(\omega)$.
• Ramos de tensão da rede multi-portas,
representados pela matriz de incidência $\mathbf{A}_{vh}$ e por relações constitutivas definidas por $\boldsymbol{V}_{vh}(\omega) = \boldsymbol{H}_{vv}(\omega)\boldsymbol{V}_{ih}(\omega) + \boldsymbol{H}_{vi}(\omega)\boldsymbol{I}_{vh}(\omega)$.
• Ramos de fontes de corrente independente,
representados pelas matrizes de incidência $\mathbf{A}_{j}$ e pelas relações definidas por $\boldsymbol{I}_{j} = \boldsymbol{J}$.
• Ramos de fontes de tensão independente,
representados pela matrizes de incidência $\mathbf{A}_{e}$ e pelas relações definidas por $\boldsymbol{V}_{e} = \boldsymbol{E}$.
• Ramos de fontes de corrente externa,
representados pelas matrizes de incidência $\mathbf{A}_{j_e}$ e pelas relações definidas por $\boldsymbol{I}_{j_e} = \boldsymbol{J}_{e}$.
• Ramos de fontes de tensão externa,
representados pela matrizes de incidência $\mathbf{A}_{e_e}$ e pelas relações definidas por $\boldsymbol{V}_{e_e} = \boldsymbol{E}_{e}$.
• Ramos de ???,
representados pela matrizes de incidência $\mathbf{A}_{i_e}$ e pelas relações definidas por $\boldsymbol{V}_{i_e} = \mathbf{0}$.

Convém ressaltar, que os seis primeiros ramos da lista acima são definidos apenas para a SRN, enquanto os quatro últimos ramos são definidos apenas para a SRL.

Para simplificar a notação, iremos suprimir o argumento referente à frequência angular, ω, nos vetores de tensão, de corrente, e de fontes de tensão e de corrente das equações (3.2.a)-(3.2.d) apresentadas abaixo. Como $\boldsymbol{V}_{@}$, $\boldsymbol{V}$ e $\boldsymbol{I}$ encompassam todas as tensões de nó, tensões de ramo, e correntes de ramo da sub-rede, respectivamente. Então, a lei de *Kirchhoff* para corrente (LKC) e lei de Kirchhoff para tensão (LKT) podem ser escritas, respectivamente, como se segue

$$\mathbf{A}_I\boldsymbol{I}(\omega) = \mathbf{0} \tag{3.1.a}$$

$$\boldsymbol{V}(\omega) - \mathbf{A}_V^T\boldsymbol{V}_{@}(\omega) = \mathbf{0}, \tag{3.1.b}$$

onde

$$\boldsymbol{I} = \left[\boldsymbol{I}_y\ \boldsymbol{I}_z\ \boldsymbol{I}_{civ}\ \boldsymbol{I}_{cii}\ \boldsymbol{I}_{vi}\ \boldsymbol{I}_{ii}\ \boldsymbol{I}_{iv}\ \boldsymbol{I}_{vv}\ \boldsymbol{J}\ \boldsymbol{I}_e\ \boldsymbol{J}_f\ \boldsymbol{I}_{fe}\ \boldsymbol{I}_{f\phi}\ \boldsymbol{I}_{ih}\ \boldsymbol{I}_{vh}\right]^T, \tag{3.2.a}$$

$$\boldsymbol{V} = \left[\boldsymbol{V}_y\ \boldsymbol{V}_z\ \boldsymbol{V}_{cvv}\ \boldsymbol{V}_{cvi}\ \boldsymbol{V}_{iv}\ \boldsymbol{V}_{vi}\ \boldsymbol{V}_{vv}\ \boldsymbol{E}\ \boldsymbol{E}_f\ \boldsymbol{E}_{f\phi}\ \boldsymbol{V}_{ih}\ \boldsymbol{V}_{vh}\right]^T, \tag{3.2.b}$$

$$\mathbf{A}_I = \left[\mathbf{A}_y\ \mathbf{A}_z\ \mathbf{A}_{civ}\ \mathbf{A}_{cii}\ \mathbf{A}_{vi}\ \mathbf{A}_{ii}\ \mathbf{A}_{iv}\ \mathbf{A}_{vv}\ \mathbf{A}_j\ \mathbf{A}_e\ \mathbf{A}_{fj}\ \mathbf{A}_{fe}\ \mathbf{A}_{fq}\ \mathbf{A}_{f\phi}\ \mathbf{A}_{ih}\ \mathbf{A}_{vh}\ \mathbf{A}_{j_e}\ \mathbf{A}_{e_e}\right], \tag{3.2.c}$$

$$\mathbf{A}_V = \left[\mathbf{A}_y\ \mathbf{A}_z\ \mathbf{A}_{cvv}\ \mathbf{A}_{cvi}\ \mathbf{A}_{iv}\ \mathbf{A}_{vi}\ \mathbf{A}_{vv}\ \mathbf{A}_e\ \mathbf{A}_{fe}\ \mathbf{A}_{f\phi}\ \mathbf{A}_{ih}\ \mathbf{A}_{vh}\ \mathbf{A}_{i_e}\right]. \tag{3.2.d}$$

As fontes de corrente independentes, $\boldsymbol{J}(\omega)$, as fontes de corrente controladas de controle não-linear, $\boldsymbol{J}_f(\omega)$, e os capacitores não-lineares, $\boldsymbol{Q}_f(\omega)$, podem ser expressos em termos de corrente nodais utilizando as seguintes expressões

$$\boldsymbol{J}_{@}(\omega) = -\mathbf{A}_j\boldsymbol{J}(\omega)\text{,Je,Ee} \tag{3.3.a}$$

$$\boldsymbol{J}_{@f}(\omega) = -\mathbf{A}_{fj}\boldsymbol{J}_f(\omega), \tag{3.3.b}$$

$$\boldsymbol{Q}_{@f}(\omega) = -\mathbf{A}_{fq}\boldsymbol{Q}_f(\omega). \tag{3.3.c}$$

Finalmente, utilizando a expressão da matriz de admitância nodal dada por

$$\boldsymbol{K}_{@y}(\omega) = \mathbf{A}_y\boldsymbol{K}_y(\omega)\mathbf{A}_y^T + \mathbf{A}_{iv}\boldsymbol{K}_{iv}(\omega)\mathbf{A}_{iv}^T + \mathbf{A}_{ih}\boldsymbol{H}_{iv}(\omega)\mathbf{A}_{ih}^T, \tag{3.4}$$

e as equações (3.1.a)-(3.3.a), a equação nodal modificada pode ser escrita da seguinte forma

$$\boldsymbol{U}_{@}(\omega) = \mathbf{M}_{@}(\omega)\boldsymbol{X}(\omega), \tag{3.5}$$

onde $\boldsymbol{U}_{@}(\omega) = \boldsymbol{U}_{@\mathrm{g}}(\omega) + \boldsymbol{U}_{@f}(\omega) + \boldsymbol{U}_{@e}(\omega)$ e $\boldsymbol{V}_{@}(\omega), \boldsymbol{I}(\omega) \subset \boldsymbol{X}(\omega)$. Em adição, temos que , $\boldsymbol{J}_{@}(\omega), \boldsymbol{E}(\omega) \subset \boldsymbol{U}_{@g}(\omega)$, $\boldsymbol{J}_e(\omega), \boldsymbol{E}_e(\omega) \subset \boldsymbol{U}_{@e}(\omega)$ e $\boldsymbol{J}_{@f}(\omega), \boldsymbol{Q}_{@f}(\omega), \boldsymbol{E}_f(\omega), \boldsymbol{\Phi}_f(\omega) \subset \boldsymbol{U}_{@f}(\omega)$. A matriz nodal-modificada, $\mathbf{M}_{@}(\omega)$, pode ser facilmente construída por inspeção utilizando as regras de construção fornecidas em [134]. As barras sobre matriz e vetores na equação (3.5), indicam que o sistema não está ordenado para uma eficiente solução númerica. Em seguida, utilizando a expressão acima, apresentaremos a formulação das equações da SRN (equação de estado e de sonda) e da SRL (equação de sonda). A natureza multi-terminais da formulação apresentada aqui, possibilita a imediata implementação de nós locais de referência (NLRs) para análise de circuitos distribuídos espacialmente [132],[145].

3.4.1. Formulação da Sub-Rede Não-Linear

Considerando que em uma SRN não existem fontes (de tensão e de corrente) independentes, $\boldsymbol{U}_{@g}(\omega) = \mathbf{0}$, e introduzimos, respectivamente, as matrizes elementares de permutação, $\mathbf{L}_L$ e $\mathbf{L}_N$, para separação da parte linear e não-linear da equação nodal-modificada. Então, podemos re-escrever (3.5) da seguinte forma

$$\mathbf{L}_L^T\boldsymbol{U}_{@e}(\omega) = \mathbf{L}_L^T[\mathbf{M}_{@}(\omega)\boldsymbol{X}(\omega) + \boldsymbol{U}_{@f}(\omega)], \tag{3.6.a}$$

$$\mathbf{L}_N^T\boldsymbol{U}_{@e}(\omega) = \mathbf{L}_N^T[\mathbf{M}_{@}(\omega)\boldsymbol{X}(\omega) + \boldsymbol{U}_{@f}(\omega)]. \tag{3.6.b}$$

Para separação das variáveis de estado lineares e não-lineares, introduziremos as matrizes elementares, $\mathbf{K}_L$ e $\mathbf{K}_N$, respectivamente. Utilizando estas matrizes podemos escrever

$$\boldsymbol{X}(\omega) = \mathbf{K}_L \boldsymbol{X}_L(\omega) + \mathbf{K}_N \boldsymbol{X}_N(\omega), \tag{3.7}$$

onde $\boldsymbol{X}_L(\omega) \in \mathbb{R}^{n_{VEL}}$ e $\boldsymbol{X}_N(\omega) \in \mathbb{R}^{n_{VEN}}$ são os vetores de variável de estado linear e de variável de estado não-linear, respectivamente. Em adição, $n_{VE} = n_{VEL} + n_{VEN}$, onde n_{VEL} e n_{VEN} correspondem ao número de variáveis de estado linear e ao número de variáveis de estado não-linear, respectivamente.

A relação entre o vetor de função não-linear e o vetor $\boldsymbol{U}_{@f}(\omega)$, que contém os elementos não-lineares de nó e de ramo sob forma de fontes de corrente, fontes de tensão, cargas e fluxos, pode ser facilmente estabelecida com a introdução da matriz elementar de permutação, $\mathbf{L}_f$. Sendo assim, temos que:

$$\boldsymbol{U}_{@f}(\omega) = \mathbf{L}_f \boldsymbol{U}_f(\omega). \tag{3.8}$$

Introduzindo, $\boldsymbol{U}(\omega) = \mathbf{L}_L^T \boldsymbol{U}_{@}(\omega)$, Substituindo as relações (3.7) e (3.8) nas equações (3.6.a) e (3.6.b), obtemos o seguinte resultado

$$\boldsymbol{U}(\omega) = \boldsymbol{M}_{LL}(\omega)\boldsymbol{X}_L(\omega) + \boldsymbol{M}_{LN}(\omega)\boldsymbol{X}_N(\omega) + \mathbf{N}_L \boldsymbol{U}_f(\omega), \tag{3.9.a}$$

$$\mathbf{0} = \boldsymbol{M}_{NL}(\omega)\boldsymbol{X}_L(\omega) + \boldsymbol{M}_{NN}(\omega)\boldsymbol{X}_N(\omega) + \mathbf{N}_N \boldsymbol{U}_f(\omega), \tag{3.9.b}$$

onde

$$\boldsymbol{M}_{LL}(\omega) = \mathbf{L}_L^T \mathbf{M}_{@}(\omega)\mathbf{K}_L, \tag{3.10.a}$$

$$\boldsymbol{M}_{LN}(\omega) = \mathbf{L}_L^T \mathbf{M}_{@}(\omega)\mathbf{K}_N, \tag{3.10.b}$$

$$\mathbf{N}_L = \mathbf{L}_L^T \mathbf{L}_f, \tag{3.10.c}$$

$$\boldsymbol{M}_{NL}(\omega) = \mathbf{L}_N^T \mathbf{M}_{@}(\omega)\mathbf{K}_L, \tag{3.10.d}$$

$$\boldsymbol{M}_{NN}(\omega) = \mathbf{L}_N^T \mathbf{M}_{@}(\omega)\mathbf{K}_N, \tag{3.10.e}$$

$$\mathbf{N}_N = \mathbf{L}_N^T \mathbf{L}_f. \tag{3.10.f}$$

Antes de obtermos a equação de estado e de sonda, vamos introduzir $\boldsymbol{U}(\omega) = \mathbf{K}_e \boldsymbol{U}_e(\omega)$ e $\boldsymbol{Y}(\omega) = \mathbf{K}^T \boldsymbol{X}(\omega)$, onde $\boldsymbol{U}_e(\omega)$ é o vetor de fonte externa e $\boldsymbol{Y}(\omega)$ é o vetor de sonda externa e interna; e $\mathbf{K}_e$ e $\mathbf{K}$ são matrizes elementares (para seleção de tensões de nós e correntes de ramos referentes às sondas). Com estas introduções, isolando as variáveis de estado lineares, $\boldsymbol{X}_L(\omega)$, em (3.9.a), e substituindo o resultado em (3.9.b), obtemos a equação de estado da SRN, dada por

$$\mathbf{0} = \boldsymbol{A}_N(\omega)\boldsymbol{X}_N(\omega) + \boldsymbol{B}_f(\omega)\boldsymbol{U}_f(\omega) + \boldsymbol{B}_e(\omega)\boldsymbol{U}_e(\omega), \tag{3.11}$$

onde

$$\boldsymbol{A}_N(\omega) = \boldsymbol{M}_{NN}(\omega) - \boldsymbol{M}_{NL}(\omega)\boldsymbol{M}_{LL}(\omega)^{-1}\boldsymbol{M}_{LN}(\omega), \tag{3.12.a}$$

$$\boldsymbol{B}_f(\omega) = \mathbf{N}_N - \boldsymbol{M}_{NL}(\omega)\boldsymbol{M}_{LL}(\omega)^{-1}\mathbf{N}_L, \tag{3.12.b}$$

$$\boldsymbol{B}_e(\omega) = -\mathbf{K}_e . \tag{3.12.c}$$

Utilizando (3.9.a), a equação de sonda externa e interna pode ser escrita da seguinte forma

$$\boldsymbol{Y}(\omega) = \boldsymbol{M}_N(\omega)\boldsymbol{X}_N(\omega) + \boldsymbol{N}_f(\omega)\boldsymbol{U}_f(\omega) + \boldsymbol{N}_e(\omega)\boldsymbol{U}_e(\omega) , \tag{3.13}$$

onde

$$\boldsymbol{M}_N(\omega) = -\mathbf{K}^T\boldsymbol{M}_{LL}(\omega)^{-1}\boldsymbol{M}_{LN}(\omega) , \tag{3.14.a}$$

$$\boldsymbol{N}_f(\omega) = -\mathbf{K}^T\boldsymbol{M}_{LL}(\omega)^{-1}\mathbf{N}_L , \tag{3.14.b}$$

$$\boldsymbol{N}_e(\omega) = -\mathbf{K}^T\mathbf{K}_e . \tag{3.14.c}$$

Como podemos facilmente observar, as equações de estado (3.11) e de sonda (3.13) envolvem apenas variáveis de estado não-lineares.

3.4.2. Formulação da Sub-Rede Linear

Para a formulação da SRL, lembramos que, por definição, a equação (3.9.b) não existe e a equação (3.9.a) envolve apenas os vetores $\boldsymbol{U}_{@}(\omega)$ e $\boldsymbol{X}_L(\omega)$. Em adição, temos que: $\boldsymbol{Y}(\omega) = \mathbf{K}^T\boldsymbol{X}_L(\omega)$. Com estas considerações, a formulação da equação de sonda da SRL relacionando $\boldsymbol{Y}(\omega)$ e $\boldsymbol{U}_e(\omega)$ é imediato. No processo de formulação, estamos apenas interesados em poucas linhas da inversa da matriz nodal-modificada, $\boldsymbol{M}_{LL}(\omega)$. Sabe-se bem [28],[134] que a melhor forma de calcular numericamente estas linhas é via decomposição LU do sistema adjunto de (3.5). Para uma SRL em grande-escala uma fatorização LU densa envolverá $O((\dim(\mathbf{M}_{@}(\omega)))^3)$ operações de ponto flutuante. Fortunadamente, em geral, a matriz nodal modificada, $\mathbf{M}_{@}(\omega)$, é extremamente esparsa e numericamente bem condicionada. Desta forma, para retomar esta forte dominância diagonal, devemos processar trocas de linhas e de colunas da matriz nodal modificada, como descrito em [28]. Após este estágio, obtemos uma matriz fortemente diagonal que pode ser decomposta utilizando técnicas de matriz esparsa [146],[147],[148]. Abaixo, iremos assumir que $\mathbf{H}_{@}(\omega)$ é a matriz híbrida contendo as desejadas linhas da inversa da matriz nodal modificada, $\overline{\mathbf{M}}_{@}(\omega)$. Em adição, vamos introduzir as matrizes elementares $\mathbf{K}_e$, $\mathbf{K}_i$ e $\mathbf{K}_g$ para selecionar as linhas de $\mathbf{H}_{@}(\omega)$ associadas com as sondas externas e internas, e as fontes independentes de excitação (geradores), respectivamente.

Antes de determinar as equações da SRL, utilizando as matrizes elementares e a matriz híbrida $\mathbf{H}_{@}(\omega)$, é conviniente introduzirmos os seguintes vetores: $\boldsymbol{Y}_e(\omega)$ e $\boldsymbol{Y}_i(\omega)$ que encompassam todas as sondas externas, internas e de conexão, respectivamente. Lembrando que $\mathbf{K}_L = \mathbf{1}_{n_{VEN}}$ e $\mathbf{K}_N = \varnothing$, podemos representar estes vetores, como se segue

$$\boldsymbol{Y}_{\mathrm{v}}(\omega) = \mathbf{K}_{\mathrm{v}}^T\boldsymbol{X}_L(\omega) \tag{3.15}$$

onde $\nu = e, i$. Similarmente, podemos representar os vetores, $\boldsymbol{U}_{@g}(\omega)$ e $\boldsymbol{U}_{@e}(\omega)$, em termos dos vetores de fonte independente, $\boldsymbol{U}_g(\omega)$, e de fonte externa, $\boldsymbol{U}_e(\omega)$ utilizando a transposta das matrizes elementares, $\mathbf{K}_s$ e $\mathbf{K}_e$, respectivamente. Sendo assim, temos que

$$\boldsymbol{U}_{@\nu}(\omega) = \mathbf{K}_\nu \boldsymbol{U}_\nu(\omega) \tag{3.16}$$

onde $\nu = e, g$. Agora, utilizando as relações (3.15) e (3.16), e aplicando o princípio da superposição, obtemos as equações híbridas reduzidas, *viz*.

$$\boldsymbol{Y}_\nu(\omega) = \boldsymbol{H}_{\nu e}(\omega)\boldsymbol{U}_e(\omega) + \boldsymbol{H}_{\nu g}(\omega)\boldsymbol{U}_g(\omega), \tag{3.17.a}$$

onde

$$\boldsymbol{H}_{\nu e}(\omega) = \mathbf{K}_\nu^T \mathbf{H}_{@}(\omega)\mathbf{K}_e, \tag{3.17.b}$$

$$\boldsymbol{H}_{\nu g}(\omega) = \mathbf{K}_\nu^T \mathbf{H}_{@}(\omega)\mathbf{K}_g, \tag{3.17.c}$$

e $\nu = e, i$. Vale ressaltar que, a equação de sonda externa serve para descrever a conexão da SRL com um circuito externo, enquanto as sondas internas representam correntes e tensões dentro da SRL.

3.5. Formulação de Espaço-Estado da Sub-Rede Não-Linear

A formulação das equações para circuitos do tipo concentrado/distribuído/não-linear (C/D/N) introduzida em [112], será utilizada na formulação da equação de estado e de sonda da SRN. Lembremos que, nesta formulação, as variáveis de estado são as tensões de ramo e as correntes de ramo. O procedimento para a FEE pode ser basicamente divido em duas etapas. A primeira etapa consiste na formulação da equação topológica, e a segunda na formulação da equação de estado e de sonda. Para o conjunto de elementos básicos, descritos na Fig. 3.1, é possível simplificar a derivação apresentada em [112], sem perda de generalidade.

3.5.1. Equação Topológica

Na Fig. 3.1, podemos observar a estrutura geral de uma SRN composta da interconexão de n_{EB} elementos básicos e de n_{TE} $(= n_{TEC} + n_{TET})$ fontes externas. As fontes externas de tensão e de corrente são introduzidas para possibilitar um meio de conexão da SRN com um circuito externo. Devido à formulação multi-terminais, estas fontes possuem um terminal conectado a um nó de referência (e.g., terra ou nó local de referência).

O conjunto de elementos básicos utilizados na representação geral de uma SRN é indicado na Fig. 3.1, e são dados por: 1) FTE, 2) FTC, 3) CC, 4) FTCN, 5) CAPN, 6) SC e SCE, 7) CAP, 8)

CCN, 9) RCC, 10) RES, 11) LT, 12) LTA, 13) LTC, 14) RCT, 15) TCN, 16) IND, 17) ST e STE, 18) INDN, 19) FCCN, 20) TC, 21) FCC e 23) FCE. Para a lista acima, a árvore e co-árvore que descrevem o grafo da SRN, são as definidas pelos elementos básicos 1-7 e 16-22, respectivamente. Os outros elementos, 8-15, podem ser colocados em ambas, árvore ou co-árvore. Indutores acoplados e transformador ideal foram classificados em [112] como elementos concentrados básicos, porém eles podem ser implementados utilizando fontes controladas lineares e indutores. Sendo assim, estes elementos não foram incluídos na lista de elementos básicos fornecida acima. Em adição, de acordo com a teoria descrita em [149], linhas de transmissão acopladas podem ser implementadas utilizando linhas de transmissão desacopladas e fontes controladas lineares.

A topologia descrevendo as interconexões de uma SRN pode ser representeda através de grafos desconectados. Uma vez que a orientação dos ramos para cada grafo esteja estabelecida, formando um *digrafo* [2], nós podemos escrever a matrix de incidência da SRN como

$$\mathbf{A} = \left[\mathbf{A}_t \quad \mathbf{A}_c\right] \quad . \tag{3.18}$$

Na expressão acima, os subescritos "t" e "c" referem-se à *árvore* (ou *floresta*) e *co-árvore* (ou *co-floresta*) do digrafo, respectivamente. O procedimento para escolha apropriada da árvore será discutido abaixo. Utilizando as partições da matriz de incidência associadas à árvore e à co-árvore, podemos escrever a matriz corte fundamental, $\mathbf{D}$, através da seguinte relação

$$\mathbf{D} = -\mathbf{A}_t^{-1}\mathbf{A}_c \, . \tag{3.19}$$

Com a matriz corte fundamental, podemos expressar mais compactamente as leis de tensão e de corrente de Kirchhoff (LTK e LCK), da seguinte forma

$$\begin{bmatrix} \boldsymbol{I}_t(\omega) \\ \boldsymbol{V}_c(\omega) \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{D} \\ -\mathbf{D}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_t(\omega) \\ \boldsymbol{I}_c(\omega) \end{bmatrix}, \tag{3.20}$$

onde

$$\boldsymbol{I}_t(\omega) = \left[\boldsymbol{I}_{et}(\omega) \; \boldsymbol{I}_{pc}(\omega)\right]^T, \tag{3.21.a}$$

$$\boldsymbol{V}_t(\omega) = \left[\boldsymbol{V}_{et}(\omega) \; \boldsymbol{V}_{pc}(\omega)\right]^T, \tag{3.21.b}$$

$$\boldsymbol{I}_c(\omega) = \left[\boldsymbol{I}_{pt}(\omega) \; \boldsymbol{I}_{ec}(\omega)\right]^T, \tag{3.21.c}$$

$$\boldsymbol{V}_c(\omega) = \left[\boldsymbol{V}_{pt}(\omega) \; \boldsymbol{V}_{ec}(\omega)\right]^T. \tag{3.21.d}$$

Os subescritos "*f*" e "*c*" referem-se aos ramos na floresta e na co-floresta, respectivamente, enquanto os subescritos "*et*", "*ec*", "*pt*", e "*pc*", referem-se àos terminais externos alimentados-por-tensão e alimentados-por-corrente, e as portas alimentadas-por-tensão e alimentadas-por-

corrente, respectivamente.

Seguindo as convenções acima, a matriz (3.19) pode ser organizada da seguinte forma

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_{et,pt} & \mathbf{D}_{et,ec} \\ \mathbf{D}_{pc,pt} & \mathbf{D}_{pc,ec} \end{bmatrix}. \tag{3.22}$$

Neste ponto da formulação, é conviniente introduzirmos os seguintes vetores,

$$\boldsymbol{U}(\omega) = \left[\boldsymbol{I}_{pc}(\omega)\ \boldsymbol{V}_{pt}(\omega)\right]^T \in \mathrm{R}^{2n_{BMP}}, \tag{3.23.a}$$

$$\boldsymbol{Z}(\omega) = \left[\boldsymbol{V}_{pc}(\omega)\ \boldsymbol{I}_{pt}(\omega)\right]^T \in \mathrm{R}^{2n_{BMP}}, \tag{3.23.b}$$

$$\boldsymbol{U}_e(\omega) = \left[\boldsymbol{V}_{ev}(\omega)\ \boldsymbol{I}_{ec}(\omega)\right]^T \in \mathrm{R}^{n_{ET}}, \tag{3.23.c}$$

onde $\boldsymbol{U}$, $\boldsymbol{Z}$, e $\boldsymbol{U}_e$ correspondem aos vetores de entrada, de saída, e das fontes externas, respectivamente [112] e n_{EB} e n_{TE} referem-se ao número de elementos básicos e terminações externas, respectivamente.

Finalmente, utilizando as sub-matrizes em (3.22) e os vetores (3.23.a)-(3.23.c) na equação (3.20), podemos escrever a equação topológica da SRN como

$$\boldsymbol{U}(\omega) = \mathbf{F}\boldsymbol{Z}(\omega) + \mathbf{E}_e\boldsymbol{U}_e(\omega), \tag{3.24}$$

onde

$$\mathbf{F} = \begin{bmatrix} \mathbf{0} & \mathbf{D}_{pc,pt} \\ -\mathbf{D}_{pc,pt}^T & \mathbf{0} \end{bmatrix}, \tag{3.25.a}$$

$$\mathbf{E}_e = \begin{bmatrix} \mathbf{0} & \mathbf{D}_{pc,ec} \\ -\mathbf{D}_{et,pt}^T & \mathbf{0} \end{bmatrix}. \tag{3.25.b}$$

3.5.2. Equação de Estado e de Sonda

Obtida a formulação das equações topológicas da SRN, podemos focalizar a atenção para a derivação da equação de estado e de sonda, seguindo o procedimento tabular apresentado em [112]. Para iniciar, cada elemento básico que introduz pelo menos uma variável-de-estado é classificado de acordo com a natureza das suas relações constitutivas. Sendo assim, foi adotada a seguinte classificação:

- Grupo 1*d*: tensões nos capacitores na floresta e correntes nos indutores na cofloresta;
- Grupo 1*i*: tensões nos capacitores na cofloresta e correntes nos indutores na floresta;
- Grupo 2: ondas de tensão refletidas nos elementos distribuídos;
- Grupo 2*A*: tensões e correntes de controle linear com atraso;
- Grupo 3*A*: tensões e correntes de controle não-linear com atraso; e

• Grupo 3: tensões e correntes de controle não-linear.

Os grupos 1*d*, 1*i*, (2,2*A*) e (3*A*,3) estão associados com equações algébricas do tipo diferencial, integral, diferença e não-linear, respectivamente. A primeira parte das equações não-lineares, associada ào group 3*A*, pode ser classificada como equações diferença não-linear. A classificação acima, resulta em um vetor de variável-de-estado, no domínio do tempo, estruturado da seguinte forma

$$\boldsymbol{x}(t) = \left[\boldsymbol{x}_{1d}(t)\ \boldsymbol{x}_{1i}(t)\ \boldsymbol{x}_2(t)\ \boldsymbol{x}_{2A}(t)\ \boldsymbol{x}_{3A}(t)\ \boldsymbol{x}_3(t)\right]^T \in \mathbb{R}^{n_{VE}}. \tag{3.26}$$

onde podemos observar a inclusão de variáveis-de-estado com atraso (grupos 2*A* e 3*A*). Estas variáveis atuam como variáveis de controle das fontes controladas linearmente e não-linearmente. Neste ponto é conveniente introduzir o seguinte vetor

$$\hat{\boldsymbol{x}}(t) = \left[d\boldsymbol{x}_{1d}(t)/dt \int_{-\infty}^{t} \boldsymbol{x}_{1i}(\tau)d\tau\ \boldsymbol{x}_2(t-\tau_2)\ \boldsymbol{x}_{2A}(t-\tau_{2A})\ \boldsymbol{x}_{3A}(t-\tau_{3A})\ \boldsymbol{x}_3(t)\right]^T \in \mathbb{R}^{n_{VE}}, \tag{3.27}$$

onde $\hat{\tau}_2$, τ_{2A} e τ_{3A} se referem aos tempos de atraso, de tal forma que

$$\boldsymbol{x}_\mu(t-\tau_\mu) = \left[x_{\mu,1}(t-\tau_{\mu,1})\ x_{\mu,2}(t-\tau_{\mu,2})\ \cdots\ x_{\mu,n}(t-\tau_{\mu,n_{X\mu}})\right]^T, \tag{3.28}$$

para $\mu = 2, 2A, 3A$ e $\hat{\tau}_\mu = \tau_2, \tau_{2A}, \tau_{3A}$. Para os elementos distribuídos temos que: $\tau_{2,2i-1} = \tau_{2,2i-1} - \hat{\jmath}(\alpha l)_{2,2i-1}/\omega$ e $\hat{\tau}_{2,2i} = \hat{\tau}_{2,2i-1}$ para $i = 1, n_{LT}$, onde n_{LT} é o número de linhas de transmissão (LTs), e $\tau_{2,i} = 2\tau_{2,i} - \hat{\jmath}2(\alpha l)_{2,i}/\omega$ para $i = 2n_{LT}+1, n_{VE_2}$. Vale ressaltar que, $n_{VE_2} = 2n_{LT} + n_{LTA} + n_{LTC}$, onde n_{LTA} é o número de linhas de transmissão em circuito aberto (LTAs), e n_{LTC} é o número de linhas de transmissão em curto-circuito (LTCs). A parte complexa dos tempos de atraso será diferente de zero se os elementos distríbuidos forem de natureza dissipativa. No domínio da frequência, em associação a (3.27), podemos definir a seguinte matriz

$$\Gamma_x(\omega) = \left[\hat{\jmath}\omega\mathbf{1}_{n_{VE_{1a}}}\ \frac{1}{\hat{\jmath}\omega}\mathbf{1}_{n_{VE_{1b}}}\ e^{-\hat{\jmath}\omega\hat{\boldsymbol{\tau}}_2}\ e^{-\hat{\jmath}\omega\boldsymbol{\tau}_{2A}}\ e^{-\hat{\jmath}\omega\boldsymbol{\tau}_{3A}}\ \mathbf{0}_{n_{VE_3}}\right], \tag{3.29}$$

associada àos operadores diferencial, integral e diferença. Sendo assim, temos que:

$$\hat{\boldsymbol{X}}(\omega) = \Gamma_x(\omega)\boldsymbol{X}(\omega), \tag{3.30}$$

onde

$$\boldsymbol{X}(\omega) = \left[\boldsymbol{X}_{1d}(\omega)\ \boldsymbol{X}_{1i}(\omega)\ \boldsymbol{X}_2(\omega)\ \boldsymbol{X}_{2A}(\omega)\ \boldsymbol{X}_{3A}(\omega)\ \boldsymbol{X}_3(\omega)\right]^T. \tag{3.31}$$

Conforme citado anteriormente, o vetor de função não-linear é composto de funções não-lineares estáticas, $\boldsymbol{u}_{f1}(\boldsymbol{x}(t))$, e dinâmincas, $\boldsymbol{u}_{f2}(\boldsymbol{x}(t))$. Sendo assim, é conveniente introduzir o vetor $\hat{\boldsymbol{u}}_f(t) = \left[\boldsymbol{u}_{f1}(t)\ d\boldsymbol{u}_{f2}(t)/dt\right]^T$. Para facilitar vamos introduzir a seguinte matriz

$$\Gamma_f(\omega) = \left[\mathbf{0}_{n_{F1}}\ \hat{\jmath}\omega\mathbf{1}_{n_{F2}}\right] \tag{3.32}$$

e temos que:

$$\hat{\boldsymbol{U}}_f(\omega) = \Gamma_f(\omega)\boldsymbol{U}_f(\omega)\,. \tag{3.33}$$

Introduzindo os índices $j_x \in [1, n_{VE}]$, $j_{u_f} \in [1, n_{FN}]$, $j \in [1, n_{EB}]$ (n_{EB} número de elementos básicos) e $j_y \in [1, n_S]$, que se referem à posição da variável de estado no vetor $\boldsymbol{X}(\omega)$, da função não-linear (estática e dinâmica) em $\boldsymbol{U}_f(\omega)$, da variável de entrada em $\boldsymbol{U}(\omega)$, e sonda (externa e interna) em $\boldsymbol{Y}(\omega)$, respectivamente. Então, com as preparações acima, os elementos básicos de uma SRN, ilustrados na Fig. 3.1, podem ser descritos por:

$$\underline{\mathbf{A}}_l^{(j_x,j_x)}\hat{\boldsymbol{X}}^{(j_x)}(\omega) = \underline{\mathbf{A}}_r^{(j_x,j_x)}\boldsymbol{X}^{(j_x)}(\omega) + \underline{\mathbf{B}}_f^{(j_x,j_{u_f})}\hat{\boldsymbol{U}}_f^{(j_{u_f})}(\omega) + \underline{\mathbf{B}}^{(j_x,j)}\boldsymbol{U}^{(j)}(\omega)\,, \tag{3.34.a}$$

$$\boldsymbol{Z}^{(j)}(\omega) = \underline{\mathbf{C}}^{(j,j)}\boldsymbol{X}^{(j)}(\omega) + \underline{\mathbf{D}}_f^{(j,j_{u_f})}\hat{\boldsymbol{U}}_f^{(j_{u_f})}(\omega) + \underline{\mathbf{D}}^{(j,j)}\boldsymbol{U}^{(j)}(\omega)\,, \tag{3.34.b}$$

$$\boldsymbol{Y}^{(j_y,j)}(\omega) = \underline{\mathbf{N}}^{(j_y,j)}\boldsymbol{U}^{(j)}(\omega)\,, \tag{3.34.c}$$

onde

$\hat{\boldsymbol{X}}^{(j_x)}(\omega) \in \hat{\boldsymbol{X}}(\omega)$ e $\boldsymbol{X}^{(j_x)}(\omega) \in \boldsymbol{X}(\omega)$ são os sub-vetores de variáveis-de-estado,
$\hat{\boldsymbol{U}}_f^{(j_f)}(\omega) \in \hat{\boldsymbol{U}}_f(\omega)$ é o sub-vetor de função não-linear,
$\boldsymbol{U}^{(j)}(\omega) \in \boldsymbol{U}(\omega)$ e $\boldsymbol{Z}^{(j)}(\omega) \in \boldsymbol{Z}(\omega)$ são os sub-vetores de entrada e de saída,
$\boldsymbol{Y}^{(j)}(\omega) \in \boldsymbol{Y}(\omega)$ é o sub-vetor de sonda,
$\underline{\mathbf{A}}_l^{(j_x)}$, $\underline{\mathbf{A}}_r^{(j_x)}$, $\underline{\mathbf{B}}_f^{(j_x)}$ e $\underline{\mathbf{B}}^{(j_x)}$ são as sub-matrizes reais da equação de estado,
$\underline{\mathbf{C}}^{(j)}$, $\underline{\mathbf{D}}_f^{(j)}$ e $\underline{\mathbf{D}}^{(j)}$ são as sub-matrizes reais das equação de saída, e
$\underline{\mathbf{N}}^{(j)}$ é a sub-matriz real da equação de sonda.

Para um melhor entendimento das equações (3.34.a)-(3.34.c), ver as tabelas constitutivas descritas no Apêndice A.

Neste trabalho, o papel das matrizes elementares $\mathbf{P}_1$ e $\mathbf{P}_2$, definidas em [112], é interpretado pelos índices j_x e j. Em adição, estes índices controlam a posição de cada entrada associada à um elemento básico nas matrizes de estado, de saída e de sonda, definidas acima. Utilizando as representações dos elementos básicos, descritas nas tabelas do Apêndice A, após a inclusão das equações de estado, de saída e de sonda de todos os elementos básicos que integram a SRN, obtemos

$$\underline{\mathbf{A}}_l\hat{\boldsymbol{X}}(\omega) = \underline{\mathbf{A}}_r\boldsymbol{X}(\omega) + \underline{\mathbf{B}}_f\hat{\boldsymbol{U}}_f(\omega) + \underline{\mathbf{B}}\boldsymbol{U}(\omega)\,, \tag{3.35.a}$$

$$\boldsymbol{Z}(\omega) = \underline{\mathbf{C}}\boldsymbol{X}(\omega) + \underline{\mathbf{D}}_f\hat{\boldsymbol{U}}_f(\omega) + \underline{\mathbf{D}}\boldsymbol{U}(\omega)\,, \tag{3.35.b}$$

$$\boldsymbol{Y}(\omega) = \underline{\mathbf{N}}\boldsymbol{U}(\omega)\,, \tag{3.35.c}$$

Lembrando que os vetores $\boldsymbol{X}$, $\boldsymbol{U}_f$, $\boldsymbol{U}$, $\boldsymbol{Z}$ e $\boldsymbol{Y}$ encompassam todos os elementos correspondente aos vetores $\boldsymbol{x}^{(j_x)}$, $\boldsymbol{u}_f^{(j_f)}$, $\boldsymbol{u}^{(j)}$, $\boldsymbol{z}^{(j)}$ e $\boldsymbol{y}^{(j)}$, e as matrizes $\underline{\mathbf{A}}_l$, $\underline{\mathbf{A}}_r$, $\underline{\mathbf{B}}_f$, $\underline{\mathbf{B}}$, $\underline{\mathbf{C}}$, $\underline{\mathbf{D}}_f$, e $\underline{\mathbf{D}}$ encompassam todos os elementos correspondente às matrizes $\underline{\mathbf{A}}_l^{(j_x)}$, $\underline{\mathbf{A}}_r^{(j_x)}$, $\underline{\mathbf{B}}_f^{(j_x)}$, $\underline{\mathbf{B}}^{(j_x)}$, $\underline{\mathbf{C}}^{(j)}$, $\underline{\mathbf{D}}_f^{(j)}$, e $\underline{\mathbf{D}}^{(j)}$.

Substituindo a equação de saída (3.35.b) em (3.24) obtém-se a equação híbrida

$$\mathbf{L}\boldsymbol{U}(\omega) = \mathbf{K}\boldsymbol{X}(\omega) + \mathbf{L}_f\hat{\boldsymbol{U}}_f(\omega) + \mathbf{L}_e\boldsymbol{U}_e(\omega) \ , \tag{3.36}$$

onde

$$\mathbf{L} = \mathbf{1} - \mathbf{F}\underline{\mathbf{D}} \, , \tag{3.37.a}$$

$$\mathbf{K} = \mathbf{F}\underline{\mathbf{C}} \, , \tag{3.37.b}$$

$$\mathbf{L}_f = \mathbf{F}\underline{\mathbf{D}}_f , \tag{3.37.c}$$

$$\mathbf{L}_e = \mathbf{E}_e \, . \tag{3.37.d}$$

Na metodologia acima, a única condição para existência das equações de espaço-de-estado e de sonda, é que a matriz $\mathbf{K}_0$ seja inversível. De acordo com [112], a matriz $\mathbf{K}_0$ será singular, se a SRN:

(i) tiver corte ("cutset") consistindo de apenas indutores e/ou fontes de corrente, ou

(ii) tiver malha fechada ("loop") consistindo apenas de capacitores e/ou fontes de tensão, ou

(iii) na presença de fontes controladas em alguns dispositivos especiais, condição que não ocorre em SRNs.

Finalmente, utilizando a equação híbrida, podemos determinar a equação de estado e de sonda da SRN eliminando o vetor de entrada, $\boldsymbol{U}(\omega)$, nas equações (3.35.a) e (3.35.b), respectivamente. Sendo assim, substituindo (3.36) em (3.35.a), obtemos a equação de estado

$$\mathbf{A}_l\hat{\boldsymbol{X}}(\omega) = \mathbf{A}_r\boldsymbol{X}(\omega) + \mathbf{B}_f\hat{\boldsymbol{U}}_f(\omega) + \mathbf{B}_e\boldsymbol{U}_e(\omega) \, , \tag{3.38}$$

onde

$$\mathbf{A}_l = \underline{\mathbf{A}}_l , \tag{3.39.a}$$

$$\mathbf{A}_r = \underline{\mathbf{A}}_r + \underline{\mathbf{B}}\mathbf{L}^{-1}\mathbf{K} \, , \tag{3.39.b}$$

$$\mathbf{B}_f = \underline{\mathbf{B}}_f + \underline{\mathbf{B}}\mathbf{L}^{-1}\mathbf{L}_f , \tag{3.39.c}$$

$$\mathbf{B}_e = \underline{\mathbf{B}}\mathbf{L}^{-1}\mathbf{L}_e ; \tag{3.39.d}$$

e substituindo (3.36) em (3.35.b), obtemos a equação de sonda

$$\boldsymbol{Y}(\omega) = \mathbf{M}\boldsymbol{X}(\omega) + \mathbf{N}_f\hat{\boldsymbol{U}}_f(\omega) + \mathbf{N}_e\boldsymbol{U}_e(\omega) \, , \tag{3.40}$$

onde

$$\mathbf{M} = \underline{\mathbf{N}}\mathbf{L}^{-1}\mathbf{K} \, , \tag{3.41.a}$$

$$\mathbf{N}_f = \underline{\mathbf{N}}\mathbf{L}^{-1}\mathbf{L}_f , \tag{3.41.b}$$

$$\mathbf{N}_e = \underline{\mathbf{N}}\mathbf{L}^{-1}\mathbf{L}_e \, . \tag{3.41.c}$$

O processo de geração da equação topológica e das equações de estado (3.38) e de sonda (3.40), são ilustradas nos fluxogramas da Fig. 3.2(a) e (b), respectivamente.

Sabemos, de considerações prévias, que as variáveis de estado de uma SRN podem ser organizadas em três diferentes grupos, correspondendo às equações do tipo diferencial, diferença

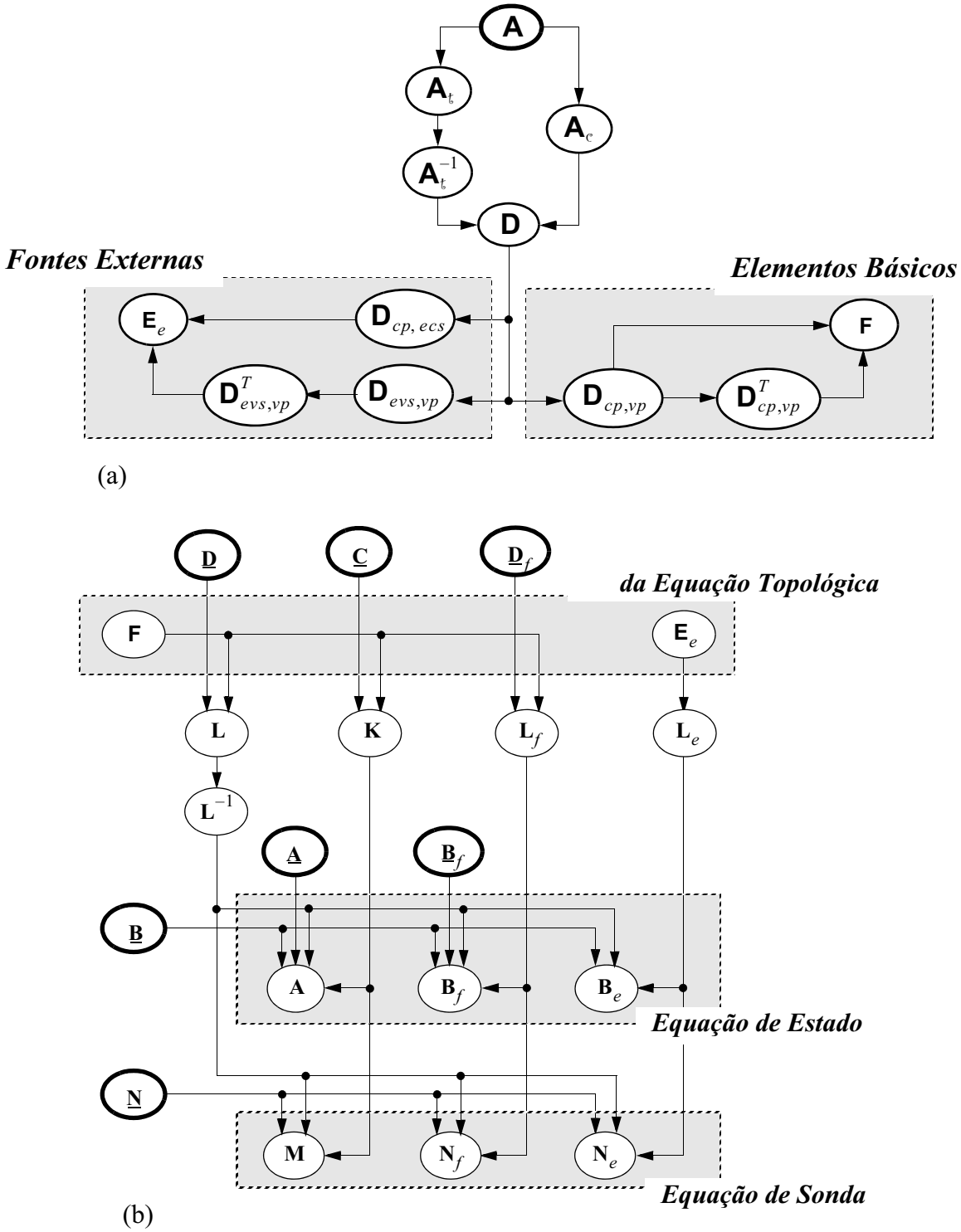


Fig. 3.2 (a) Fluxograma da geração das equação topológica. (b) Fluxograma da geração das equações de estado e de sonda.

e não-linear. Estes grupos formam um sistema misto de equações, descrito em (3.38). Assumindo a organização acima, podemos definir o *vetor de variável de estado linear*, $\boldsymbol{X}_L(\omega) \in \mathbb{C}^{n_{VEL}}$, que encompassa todas as variáveis de estado dos grupos 1 e 2. O grupo restante (3A, 3) está associado com as equações não-lineares em (3.38), e, consequentemente, as variáveis de estado deste grupo

definem o *vetor de variável de estado não-linear*, $\boldsymbol{X}_N(\omega) \in \mathbb{C}^{n_{VEN}}$. Com estas observações temos que

$$\boldsymbol{X}(\omega) = \left[\underbrace{\left[\boldsymbol{X}_1(\omega)\ \boldsymbol{X}_2(\omega)\right]^T}_{\boldsymbol{X}_L(\omega)} \ \underbrace{\left[\boldsymbol{X}_{3A}(\omega)\ \boldsymbol{X}_3(\omega)\right]^T}_{\boldsymbol{X}_N(\omega)} \right]^T \in \mathbb{R}^{n_{VE}} \tag{3.42}$$

onde $n_{VE} = n_{VEL} + n_{VEN}$. As componentes do vetor de função não-linear, nas equações (3.38) e (3.40), são funções não-lineares cujos argumentos são as variáveis de estado dos grupos 3*A* e 3. Isto significa que o vetor de função não-linear, $\boldsymbol{u}_f(\boldsymbol{x}_N(t)) \in \mathbb{R}^{n_{FN}}$, é caracterizado pelo mapeamento $\boldsymbol{u}_f: \mathbb{R}^{n_{VEN}} \rightarrow \mathbb{R}^{n_{FN}}$. Introduzindo as equações (3.30) e (3.33), obtemos

$$\boldsymbol{A}(\omega) = \mathbf{A}_r - \mathbf{A}_l \boldsymbol{\Gamma}_x(\omega)\ , \tag{3.43.a}$$

$$\boldsymbol{B}_f(\omega) = \mathbf{B}_f \boldsymbol{\Gamma}_f(\omega)\ . \tag{3.43.b}$$

Utilizando a partição definida em (3.42), para as variáveis de estado lineares e não-lineares, podemos escrever (3.38) da seguinte forma

$$\mathbf{0} = \mathbf{A}_{LL}(\omega)\boldsymbol{X}_L(\omega) + \mathbf{A}_{LN}\boldsymbol{X}_N(\omega) + \mathbf{B}_{fL}(\omega)\boldsymbol{U}_f(\omega) + \mathbf{B}_{eL}\boldsymbol{U}_e(\omega) \tag{3.44.a}$$

$$\mathbf{0} = \mathbf{A}_{NL}\boldsymbol{X}_L(\omega) + \mathbf{A}_{NN}(\omega)\boldsymbol{X}_N(\omega) + \mathbf{B}_{fN}(\omega)\boldsymbol{U}_f(\omega) + \mathbf{B}_{eN}\boldsymbol{U}_e(\omega) \tag{3.44.b}$$

Para eliminarmos $\boldsymbol{X}_L(\omega)$ no sistema acima é preciso inverter a matriz característica, $\mathbf{A}_{LL}(\omega)$. Após esta inversão, podemos por em evidência $\boldsymbol{X}_L(\omega)$ em (3.44.a), e substituir este resultado em (3.44.b), para produzir a equação de estado da SRN, dada por:

$$\mathbf{0} = \boldsymbol{A}(\omega)\boldsymbol{X}_N(\omega) + \boldsymbol{B}_f(\omega)\boldsymbol{U}_f(\omega) + \boldsymbol{B}_e(\omega)\boldsymbol{U}_e(\omega)\ , \tag{3.45}$$

onde

$$\boldsymbol{A}(\omega) = \mathbf{A}_{NN}(\omega) - \mathbf{A}_{NL}\boldsymbol{A}_{LL}(\omega)^{-1}\mathbf{A}_{LN}\ , \tag{3.46.a}$$

$$\boldsymbol{B}_f(\omega) = \mathbf{B}_{fN} - \mathbf{A}_{NL}\boldsymbol{A}_{LL}(\omega)^{-1}\mathbf{B}_{fL}(\omega)\ , \tag{3.46.b}$$

$$\boldsymbol{B}_e(\omega) = \mathbf{B}_{eN} - \mathbf{A}_{NL}\boldsymbol{A}_{LL}(\omega)^{-1}\mathbf{B}_{eL}(\omega)\ . \tag{3.46.c}$$

A inversão da matriz característica, para eliminação das variáveis de estado lineares, é bem condicionada e representa a absorção da parte linear da SRN. Uma importante característica do sistema de equações (3.45), é o fato deste ser quadrado com dimensão igual a n_{VE}.

Agora, em função de (3.42), introduziremos a seguinte partição, $\mathbf{M} = \left[\mathbf{M}_L\ \mathbf{M}_N\right]$, na equação de sonda (3.40). Sendo assim, substituindo (3.44.a), após o isolamento de $\boldsymbol{X}_L$ neste resultado, obtemos a equação de sonda da SRN, dada por:

$$\boldsymbol{Y}(\omega) = \boldsymbol{M}(\omega)\boldsymbol{X}_N(\omega) + \boldsymbol{N}_f(\omega)\boldsymbol{U}_f(\omega) + \boldsymbol{N}_e(\omega)\boldsymbol{U}_e(\omega)\,, \tag{3.47}$$

onde

$$\boldsymbol{M}(\omega) = \mathbf{M}_N - \mathbf{M}_L \boldsymbol{A}_{LL}(\omega)^{-1} \mathbf{A}_{LN} , \tag{3.48.a}$$

$$\boldsymbol{N}_f(\omega) = \mathbf{N}_f - \mathbf{M}_L \boldsymbol{A}_{LL}(\omega)^{-1} \mathbf{B}_{fL} , \tag{3.48.b}$$

$$\boldsymbol{N}_e(\omega) = \mathbf{N}_e - \mathbf{M}_L \boldsymbol{A}_{LL}(\omega)^{-1} \mathbf{B}_{eL} . \tag{3.48.c}$$

Convém ressaltar, que a dimensão da matriz característica, $\mathbf{A}_{LL}(\omega)$, depende da localização do plano de separação da SRLA e SRNC, ver Fig. 2.2. Em adição, para as SRNs onde não há elementos concentrados com memória (indutores e capacitores) ou elementos distribuídos (linhas de transmissão), nenhuma inversão precisa ser conduzida.

3.6. Dispositivos Semicondutores

Os dispositivos semicondutores eletrônicos e opto-eletrônicos (e.g., diodos, transistores, lasers, etc.) são essencialmente de natureza não-linear, e podem ser descritos através de um circuito elétrico equivalente (CEE) utilizando os elementos básicos descritos na Fig. 3.1. Para ilustrar a aplicação da metodologia desenvolvida acima, vamos considerar a formulação do CEE, com topologia entitulada "A", dos dispositivos FET e HBT, ilustrados na Fig. 3.3(a) e (b), respectivamente. Para o FET, é comum uma topologia B, onde os diodos D_{gs} e D_{gd} estão em paralelo com os capacitores não-lineares q_{gs} e q_{gd}, respectivamente [150]. Em adição, também temos a topologia C, onde os resistores intrínsecos R_{gs} e R_{gd} são iguais a zero [151]. Excluindo, uma possível tensão v_{gs} com atraso, para as topologias A, B e C, a FEE utiliza, respectivamente, 4, 3 e 2 variáveis de estado não-lineares. Já na FNM, estas topologias utilizam 5, 5 e 3, variáveis, respectivamente. Para o HBT considerado, a FEE utiliza 3 variáveis de estado não-lineares, enquanto a FNM utiliza 4. Para o FET com topologia B, utilizando a FEE temos uma econômia de 40% em relação a FNM.

Se os capacitores intrínsecos no CEE do FET e do HBT forem funções de duas ou mais tensões de controle, irá surgir o problema da transcapacitância e de conservação de carga. Sem considerar as transcapacitâncias, só é possível manter a compatibilidade entre o modelo incremental (ou de pequeno-sinal) e o de grande-sinal para apenas um ponto de polarização.

No Apêndice B, é apresentada a FEE para os CEEs ilustrados na Fig. 3.3. Também é apresentada a FEE para o dispositivo representado sob forma de DDS. Lembremos que, para minimizar o número de equações, o DDS deve ser utilizado apenas para representação da parte intrínseca (correntes de condução e de deslocamento) do dispositivo em consideração. Desta

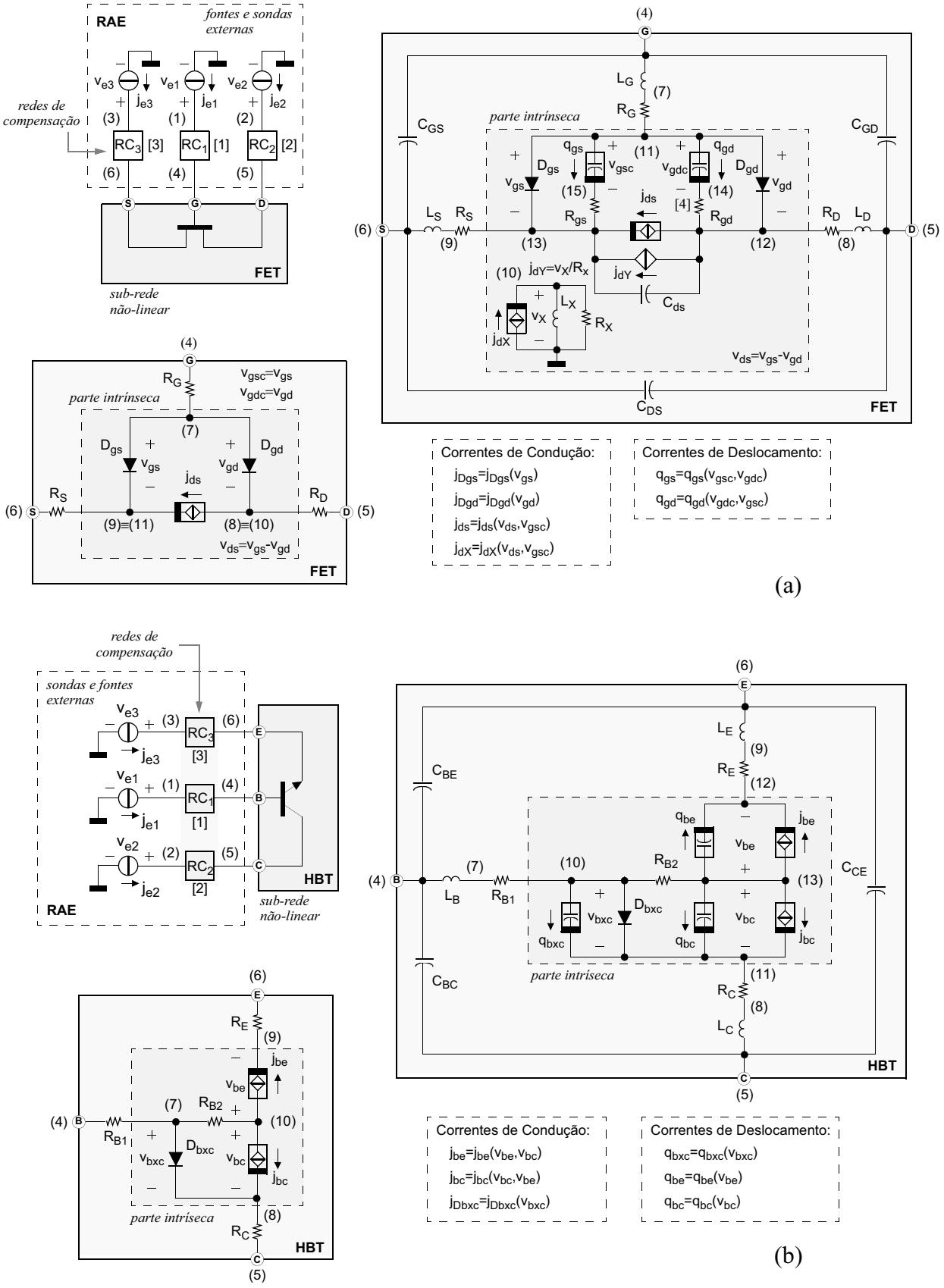


Fig. 3.3 (a) Circuito elétrico equivalente (CEE) do FET para regime CC e CA, incluindo a rede de alimentação externa (RAE). (b) CEE do HBT para regime CC e CA, incluindo a RAE.

forma, todas as variáveis de estado são não-lineares e correspondem às tensões de nós associadas às terminações externas do DDS e às correntes de ramo para as relações auxiliares.

3.7. Conclusão

Neste capítulo, foi apresentada, no domínio da frequência a formulação das equações da SRL e da SRN introduzidas no capítulo anterior. Para formulação da equação de sonda da SRL, foi realizada uma breve revisão da FNM clássica. Também foi apresentada a FNM para a formulação da equação de estado e de sonda da SRN, incluindo uma técnica de redução de ordem para a eliminação das variáveis de estado lineares. Infelizmente, neste caso, são necessárias duas variáveis de estado não-lineares (tensões de nós) para representar uma tensão de controle que atua como argumento de uma função não-linear. Este fato nos motivou ao desenvolvimento da FEE introduzida acima. Nesta formulação a tensão de controle não-linear é representada por uma única variável de estado não-linear (tensão de ramo). Assim como na FNM, a FEE segue um procedimento tabular de fácil implementação numérica. Ao contrário da FEEP, a formulação proposta sempre resultará em um sistema quadrado de equações. Lembremos, também, que não existe nenhuma metologia para geração automática de equações de estado paramétricas.

Para exemplificar a aplicação da nova FEE, foi apresentada a formulação das equações de estado e de sonda dos CEEs utilizados na representação de dispositivos eletrônicos do tipo FET e HBT, lembrando que o dispositivo eletrônico corresponde a uma SRN. Em adição, também foi apresentada a formulação destas equações, com a parte intrínseca (SRN) destes dispositivos representada sob forma de DDS. Neste caso, foi demonstrado que a FEE pode ser configurada de uma forma exatamente equivalente à FNM. Por esta razão, a formulação tradicional, utilizando DDS, pode ser vista como um caso particular da teoria aqui proposta.

4. Formulação Multi-Níveis das Equações do Circuito

4.1. Introdução

UTILIZANDO A TÉCNICA DE DECOMPOSIÇÃO multi-níveis e as formulações das equações da SRL, via FNM, e da SRN, via FEE (ou FNM), apresentadas em detalhe nos capítulos anteriores, iremos introduzir neste capítulo uma nova e eficiente técnica para formulação, no domínio da frequência, das equações de estado e de sonda do circuito. A equação de estado, em geral, de natureza não-linear, representa as relações governantes que descrevem a operação do circuito. Enquanto a equação de sonda, associada às sondas internas, possibilita obter as respostas desejadas, uma vez que a solução da equação de estado for obtida.

Para determinação das equações do circuito, serão discutidas, nas seções 4.2 e 4.3 as formulações da SRLA e da SRNC, que por sua vez, estão fundamentadas nos resultados do capítulo anterior. Utilizando estas formulações, as equações das SuRs de fundo podem ser obtidas empregando o esquema (fonte/sonda externa) de decomposição por partes discutido no Capítulo 2. Conforme citado anteriormente, a equação de estado (ou governante) de uma SuR de fundo, consiste de um sistema quadrado com dimensão igual à soma do número de variáveis de estado não-lineares presentes nas SRNs mais o número de variáveis associadas aos terminais externos de conexão com outras SuRs. A formulação da SuR de fundo será apresentada na Seção 4.4, e comparada com outras formulações [28],[145],[18].

Finalmente, utilizando as equações de estado e de sonda da SuR de fundo, na Seção 4.5 será apresentada a formulação multi-níveis das equações do circuito. Esta formulação produz um sistema quadrado de equações de múltiplos níveis, onde cada nível assume uma estrutura quadrada do tipo bloco diagonal com dupla borda [41]-[43]. Vale ressaltar que, este tipo de estrutura é semelhante às estruturas resultantes da aplicação da técnica de decomposição-de-domínio (DD), via FD ou FE, comumente empregada na solução de problemas de dinâmica computacional dos fluídos [27]. Em geral, a formulação da SuR de fundo envolve a inversão de uma matriz com dimensão igual ao número de terminais externos utilizados na interconexão SRLA/SRNC. Sendo assim, para reduzir o custo desta inversão, no caso de um número elevado de terminais externos, será proposta, na Seção 4.6, um eficiente procedimento que permite o nivelamento (ou eliminação dos níveis de hierarquia) de uma SuR intermediária. Para ilustrar a teoria proposta, vamos considerar, na Seção 4.7, a formulação do exemplo de decomposição multi-níveis introduzido no Capítulo 2. As observações finais são reservadas para a Seção 4.8.

4.2. Formulação das Equações da Sub-Rede Linear Ampliada

Conforme discutido no Capítulo 2, ver Fig. 2.1, a SRLA é composta de uma SRL e uma rede de ampliação (podendo incluir possíveis elementos resistivos de degeneração). Para enterdermos a formulação da equação de sonda que descreve a SRLA, vamos nos referir à Fig. 4.1. Nesta figura, podemos observar, em detalhe, o esquema elétrico empregado na separação da SRLA e da SRNC, utilizando fontes e sondas externas. Lembremos que os terminais alimentados por corrente são representados por uma fonte de corrente e uma sonda de tensão (equivalente de Norton), e os terminais alimentados por tensão são representados por uma fonte de tensão e uma sonda de corrente (equivalente de Thévenin). Para generalização da formulação proposta, todos os tipos possíveis de conexões entre a SRLA e a SRNC foram considerados [152]. Utilizando os resultados do Capítulo 3, e as definições de terminais externos e_α, e_β, e_γ e e_δ, introduzidas no Capítulo 2, podemos escrever a equação da SRLA (via FNM), como se segue

$$\boldsymbol{Y}_{\nu}^{\mathcal{L}_a}(\omega) = \boldsymbol{H}_{\nu e_\alpha}^{\mathcal{L}_a}(\omega)\boldsymbol{U}_{e_\alpha}^{\mathcal{L}_a}(\omega) + \boldsymbol{H}_{\nu e_\beta}^{\mathcal{L}_a}(\omega)\boldsymbol{U}_{e_\beta}^{\mathcal{L}_a}(\omega) + \boldsymbol{H}_{\nu e_\gamma}^{\mathcal{L}_a}(\omega)\boldsymbol{U}_{e_\gamma}^{\mathcal{L}_a}(\omega) + \boldsymbol{H}_{\nu e_\delta}^{\mathcal{L}_a}(\omega)\boldsymbol{U}_{e_\delta}^{\mathcal{L}_a}(\omega) + \boldsymbol{H}_{\nu g}^{\mathcal{L}_a}(\omega)\boldsymbol{U}_{g}^{\mathcal{L}_a}(\omega), \quad (4.1)$$

onde

$$\boldsymbol{H}_{\nu\mu}^{\mathcal{L}_a}(\omega) = \mathbf{K}_{\nu}^{T}\mathbf{H}_{@}(\omega)\mathbf{K}_{\mu}, \quad (4.2)$$

$\nu = i, e_\alpha, e_\beta, e_\gamma, e_\delta$ e $\mu = g, e_\alpha, e_\beta, e_\gamma, e_\delta$. Lembrando que o superescrito "$\mathcal{L}_a$" se refere a SRLA, e serve para diferenciar da SRNC descrita abaixo. Convém relembrar que, os terminais externos, e_α e e_β, estão associados à conexão com a SRNC, e os terminais externos, e_γ e e_δ, estão associados à conexão com outras SuRs de mesmo nível e de níveis inferiores, respectivamente. As conexões externas, "e_α", são conexões nas quais a representação fonte-sonda da SRL/SRN é do tipo corrente-tensão/corrente-tensão ou tensão-corrente/tensão-corrente. As conexões externas, "e_β", correspondem ao caso complementar, i.e., corrente-tensão/tensão-corrente ou tensão-corrente/ corrente-tensão.

4.3. Formulação das Equações da Sub-Rede Não-Linear Compactada

No Capítulo 3, foi apresentada uma eficiente metodologia para a formulação das equações de uma SRN utilizando FEE ou FNM. Estas formulações podem ser aplicadas aos dispositivos eletrônicos de microonda e de onda-milimétrica, e dispositivos fotônicos, e.g.: diodos, transistores (FETs e HBTs), lasers, etc. Conforme já discutido, em geral, a FEE resulta em um sistema não-linear de menor dimensão quando comparado com a FNM. Seguindo a decomposição introduzida

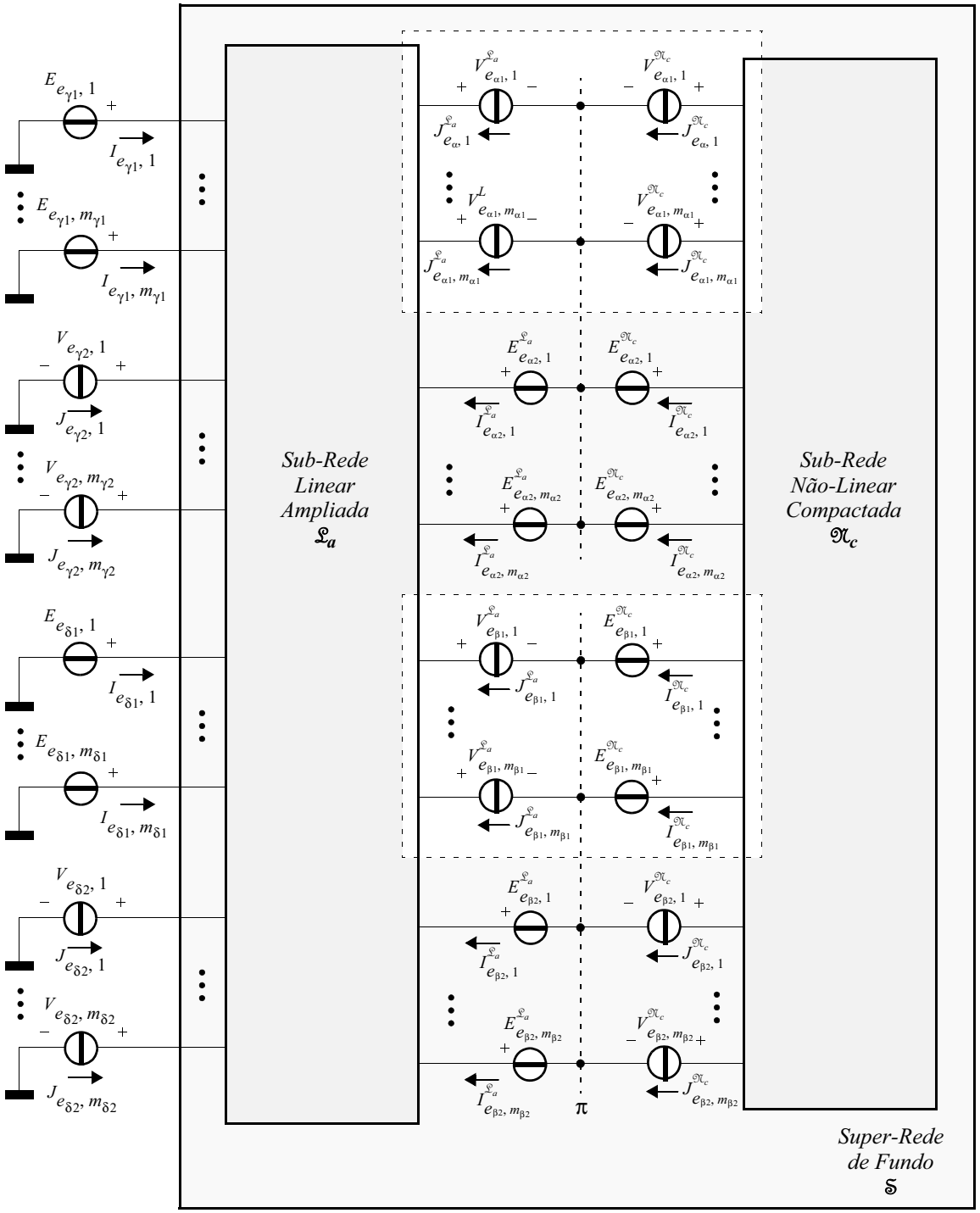


Fig. 4.1 Esquema elétrico das fontes e das sondas utilizadas na representação da interconexão entre a *sub-rede linear ampliada* (SRLA) e a *sub-rede não-linear compactada* (SRNC). Também são indicadas as fontes e as sondas utilizadas na conexão da super-rede (SuR) de fundo com um circuito externo.

no Capítulo 2, ver Fig. 2.1, a SRNC é formada por SRNs e uma rede de permutação (podendo incluir possíveis elementos resistivos de compensação). Utilizando os resultados do capítulo anterior, podemos formular eficientemente as equações de estado e de sonda da SRNC, ilustrada na Fig. 4.1. Sendo assim, assumindo para os superescritos que $(j) \equiv (\mathfrak{N}_j)$, onde $\mathfrak{N}_j$ é a j-ésima SRN,

com $j \in [1, n_{SRN}]$ onde n_{SRN} é o número de SRNs. Então, as equações de estado e de sonda, podem ser escritas da seguinte forma

$$\mathbf{0} = \boldsymbol{A}_N^{(j)}(\omega)\boldsymbol{X}_N^{(j)}(\omega) + \boldsymbol{B}_f^{(j)}(\omega)\boldsymbol{U}_f^{(j)}(\omega) + \boldsymbol{B}_e^{(j)}(\omega)\boldsymbol{U}_e^{(j)}(\omega) \tag{4.3.a}$$

e

$$\boldsymbol{Y}_\nu^{(j)}(\omega) = \boldsymbol{M}_{\nu N}^{(j)}(\omega)\boldsymbol{X}_N^{(j)}(\omega) + \boldsymbol{N}_{\nu f}^{(j)}(\omega)\boldsymbol{U}_f^{(j)}(\omega) + \mathbf{N}_{\nu e}^{(j)}\boldsymbol{U}_e^{(j)}(\omega)\,, \tag{4.3.b}$$

respectivamente, onde $\nu = i, e$. Vale ressaltar que, as equações acima envolvem apenas variáveis de estado não-lineares, Lembremos que as variáveis de estado lineares foram eliminadas antes de compactarmos as SRNs.

Com as variáveis de estado lineares eliminadas do contexto, podemos agrupar as equações de estado e de sonda de todas as SRNs, para formar as equações da SRNC. Para tal, convém introduzir o vetor de variável de estado não-linear que envolve todas as variáveis de estado não-lineares das SRNs ilustradas na Fig. 2.1. Escrevendo este vetor, temos que:

$$\boldsymbol{X}^{\mathfrak{N}_c}(\omega) = \left[\boldsymbol{X}^{(1)}(\omega)\ \boldsymbol{X}^{(2)}(\omega)\ \ldots\ \boldsymbol{X}^{(n_{SRN})}(\omega)\right]^T \in \mathbb{C}^{n_{VEN}^{\mathfrak{N}_c}} \tag{4.4}$$

onde $n_{VEN}^{\mathfrak{N}_c} = n_{VEN}^{(1)} + n_{VEN}^{(2)} + \ldots + n_{VEN}^{(n_{SRN})}$ é o número de variáveis de estado não-lineares na SRNC. Os vetores $\boldsymbol{U}_f^{\mathfrak{N}_c}(\omega) \in \mathbb{C}^{n_{FN}^{\mathfrak{N}c}}$ e $\boldsymbol{U}_e^{\mathfrak{N}_c}(\omega) \in \mathbb{C}^{n_{TE}^{\mathfrak{N}c}}$ assumem a mesma estrutura definida em (4.4), onde $n_{FN}^{\mathfrak{N}_c} = n_{FN}^{(1)} + n_{FN}^{(2)} + \ldots + n_{FN}^{(n_{SRN})}$ e $n_{TE}^{\mathfrak{N}_c} = n_{TE}^{(1)} + n_{TE}^{(2)} + \ldots + n_{TE}^{(n_{SRN})}$ correspondem ao número de funções não-lineares na SRNC e ao número de terminais externos da SRNC, respectivamente. O superescrito "N" é utilizado para referir a SRNC.

Em geral, não existe acoplamento entre as SRNs, o que resulta, para SNRC, em um sistema de equações do tipo bloco diagonal. Desta forma, em virtude de (4.4), convém introduzir

$$\boldsymbol{A}^{\mathfrak{N}_c}(\omega) = \mathbf{diag}[\boldsymbol{A}_N^{(1)}(\omega), \boldsymbol{A}_N^{(2)}(\omega), \ldots, \boldsymbol{A}_N^{(n_{SRN})}(\omega)]\,. \tag{4.5}$$

Adotando a mesma estrutura acima para outras matrizes constitutivas em (4.3.a) e (4.3.b), podemos agrupar todas as equações de estado e de sonda das SRNs. Sendo assim, após incluir a rede de permutação (ver Fig. 2.2), obtemos as equações de estado e de sonda (interna e externa) da SRNC, respectivamente, dadas por:

$$\mathbf{0} = \boldsymbol{A}^{\mathfrak{N}_c}(\omega)\boldsymbol{X}(\omega) + \boldsymbol{B}_f^{\mathfrak{N}_c}(\omega)\boldsymbol{U}_f(\omega) + \boldsymbol{B}_{e_\alpha}^{\mathfrak{N}_c}(\omega)\boldsymbol{U}_{e_\alpha}^{\mathfrak{N}_c}(\omega) + \boldsymbol{B}_{e_\beta}^{\mathfrak{N}_c}(\omega)\boldsymbol{U}_{e_\beta}^{\mathfrak{N}_c}(\omega)\,, \tag{4.6}$$

$$\boldsymbol{Y}_\nu^{\mathfrak{N}_c}(\omega) = \boldsymbol{M}_\nu^{\mathfrak{N}_c}(\omega)\boldsymbol{X}(\omega) + \boldsymbol{N}_{\nu f}^{\mathfrak{N}_c}(\omega)\boldsymbol{U}_f(\omega) + \boldsymbol{N}_{\nu e_\alpha}^{\mathfrak{N}_c}(\omega)\boldsymbol{U}_{e_\alpha}^{\mathfrak{N}_c}(\omega) + \boldsymbol{N}_{\nu e_\beta}^{\mathfrak{N}_c}(\omega)\boldsymbol{U}_{e_\beta}^{\mathfrak{N}_c}(\omega)\,, \tag{4.7}$$

e $\nu = i, e_\alpha, e_\beta$. Como podemos observar, as equações (4.6) e (4.7) envolvem apenas variáveis de estado não-lineares das SRNs.

4.4. Formulação das Equações da Super-Rede de Fundo

Para iniciarmos a formulação das equações das SuRs de fundo, vamos considerar a decomposição por partes ilustrada na Fig. 4.1. Nesta figura, podemos observar que todos os tipos de combinações fonte-sonda externa são permitidas na representação das interconexões entre a SRLA e a SRNC. As conexões externas, $e_{\alpha 1}$, representam as conexões tradicionalmente utilizadas na análise de EH [45],[37],[39], i.e, minimiza-se o erro entre as correntes que entram na SRLA e SRNC, tendo a tensão como variável independente. Considerando a FNM para SRLA, sem inclusão de corrente nos ramos (variáveis extras), é necessário a inversão de uma matriz cuja dimensão é igual ao número de terminais externos. Para evitar esta inversão, em [145] foi proposto o uso exclusivo de conexões externas, $e_{\alpha 2}$, onde se equilibra a tensão, com a corrente sendo a variável independente do problema. Ressaltamos que, os pontos definidos pelo plano de conexão, π, representam pontos de terra virtual.

As fontes e sondas externas, respectivamente, são representadas pelos seguintes vetores

$$\boldsymbol{Y}_{e_{\alpha,\beta}}^{\mathcal{L}_a,\,\mathcal{N}_c}(\omega) = \left[\boldsymbol{V}_{e_{\alpha,\beta}}^{\mathcal{L}_a,\,\mathcal{N}_c}(\omega)\ \boldsymbol{I}_{e_{\alpha,\beta}}^{\mathcal{L}_a,\,\mathcal{N}_c}(\omega)\right]^T, \tag{4.8.a}$$

$$\boldsymbol{U}_{e_{\alpha,\beta}}^{\mathcal{L}_a,\,\mathcal{N}_c}(\omega) = \left[\boldsymbol{J}_{e_{\alpha,\beta}}^{\mathcal{L}_a,\,\mathcal{N}_c}(\omega)\ \boldsymbol{E}_{e_{\alpha,\beta}}^{\mathcal{L}_a,\,\mathcal{N}_c}(\omega)\right]^T, \tag{4.8.b}$$

onde $\nu = e_\alpha, e_\beta$. As condições de contorno para as conexões externas do tipo, e_α, são descritas pelas seguintes equações

$$\boldsymbol{Y}_{e_\alpha}^{\mathcal{N}_c}(\omega) = \boldsymbol{Y}_{e_\alpha}^{\mathcal{L}_a}(\omega), \tag{4.9.a}$$

$$\boldsymbol{U}_{e_\alpha}^{\mathcal{N}_c}(\omega) = \boldsymbol{U}_{e_\alpha}^{\mathcal{L}_a}(\omega). \tag{4.9.b}$$

As condições de contorno para as conexões externas, e_β, são dadas por

$$\boldsymbol{Y}_{e_\beta}^{\mathcal{N}_c}(\omega) = \boldsymbol{U}_{e_\beta}^{\mathcal{L}_a}(\omega), \tag{4.10.a}$$

$$\boldsymbol{U}_{e_\beta}^{\mathcal{N}_c}(\omega) = \boldsymbol{Y}_{e_\beta}^{\mathcal{L}_a}(\omega). \tag{4.10.b}$$

Como podemos observar na Fig. 4.1, os vetores de fonte externa $\boldsymbol{U}_e^{\mathcal{N}_c}$ e $\boldsymbol{U}_e^{\mathcal{L}_a}$, descritos por fontes de tensão equivalentes de Thévenin e por fontes de corrente equivalentes de Norton, estão associados aos terminais externos da SRNC e da SRLA, respectivamente.

Aplicando a FNM para SRLA e a FEE (ou FNM) para SRNC, podemos determinar a equação de estado e de sonda de uma SuR de fundo, $\mathfrak{s}$. Para este propósito, vamos utilizar as expressões (4.1), (4.7), e as condições de contorno (4.9.a)-(4.9.b) e (4.10.a)-(4.10.b). Também vamos adotar $\boldsymbol{X}^{\mathfrak{S}}(\omega) = \boldsymbol{X}^{\mathcal{N}_c}(\omega)$, $\boldsymbol{U}_f^{\mathfrak{S}}(\omega) = \boldsymbol{U}_f^{\mathcal{N}_c}(\omega)$, $\boldsymbol{U}_g^{\mathfrak{S}}(\omega) = \boldsymbol{U}_g^{\mathcal{L}_a}(\omega)$, $\boldsymbol{U}_{e_\gamma}^{\mathfrak{S}}(\omega) = \boldsymbol{U}_{e_\gamma}^{\mathcal{L}_a}(\omega)$ e $\boldsymbol{U}_{e_\delta}^{\mathfrak{S}}(\omega) = \boldsymbol{U}_{e_\delta}^{\mathcal{L}_a}(\omega)$. Assim

sendo, resolvendo estas equações para o vetor $\boldsymbol{U}_e^{\mathfrak{R}_c}$, e substituindo o resultado em (4.6), obtemos a equação de estado da SuR de fundo dada por

$$\mathbf{0} = \boldsymbol{A}^{\mathfrak{S}}(\omega)\boldsymbol{X}^{\mathfrak{S}}(\omega) + \boldsymbol{B}_f^{\mathfrak{S}}(\omega)\boldsymbol{U}_f^{\mathfrak{S}}(\omega) + \boldsymbol{B}_g^{\mathfrak{S}}(\omega)\boldsymbol{U}_g^{\mathfrak{S}}(\omega) + \boldsymbol{B}_{e_\gamma}^{\mathfrak{S}}(\omega)\boldsymbol{U}_{e_\gamma}^{\mathfrak{S}}(\omega) + \boldsymbol{B}_{e_\delta}^{\mathfrak{S}}(\omega)\boldsymbol{U}_{e_\delta}^{\mathfrak{S}}(\omega), \tag{4.11}$$

onde

$$\boldsymbol{A}^{\mathfrak{S}}(\omega) = \boldsymbol{A}^{\mathfrak{R}_c}(\omega) + \boldsymbol{\Psi}(\omega)\boldsymbol{\Phi}(\omega), \tag{4.12.a}$$

$$\boldsymbol{B}_f^{\mathfrak{S}}(\omega) = \boldsymbol{B}_f^{\mathfrak{R}_c}(\omega) + \boldsymbol{\Psi}(\omega)\boldsymbol{\Phi}_f(\omega), \tag{4.12.b}$$

$$\boldsymbol{B}_\mu^{\mathfrak{S}}(\omega) = \boldsymbol{\Psi}(\omega)\boldsymbol{\Phi}_\mu(\omega), \tag{4.12.c}$$

$$\boldsymbol{\Psi}(\omega) = \boldsymbol{B}_e^{\mathfrak{R}_c}(\omega)\boldsymbol{\Gamma}(\omega)^{-1}, \tag{4.12.d}$$

e $\mu = g, e_\alpha, e_\beta$. Das equações acima, podemos observar que (4.12.d) envolve a inversão da matrix $\boldsymbol{\Gamma}(\omega)$, dada por

$$\boldsymbol{\Gamma}(\omega) = \begin{bmatrix} \boldsymbol{H}_{e_\alpha e_\alpha}^{\mathfrak{L}_a}(\omega) - \boldsymbol{N}_{e_\alpha e_\alpha}^{\mathfrak{R}_c}(\omega) + \boldsymbol{H}_{e_\alpha e_\beta}^{\mathfrak{L}_a}(\omega)\boldsymbol{N}_{e_\beta e_\alpha}^{\mathfrak{R}_c}(\omega) & -\boldsymbol{N}_{e_\alpha e_\beta}^{\mathfrak{R}_c}(\omega) + \boldsymbol{H}_{e_\alpha e_\beta}^{\mathfrak{L}_a}(\omega)\boldsymbol{N}_{e_\beta e_\beta}^{\mathfrak{R}_c}(\omega) \\ -\boldsymbol{H}_{e_\beta e_\alpha}^{\mathfrak{L}_a}(\omega) + \boldsymbol{H}_{e_\beta e_\beta}^{\mathfrak{L}_a}(\omega)\boldsymbol{N}_{e_\beta e_\alpha}^{\mathfrak{R}_c}(\omega) & \mathbf{1} - \boldsymbol{H}_{e_\beta e_\beta}^{\mathfrak{L}_a}(\omega)\boldsymbol{N}_{e_\beta e_\beta}^{\mathfrak{R}_c}(\omega) \end{bmatrix}. \tag{4.13}$$

Esta inversão é densa e possui complexidade proporcional ao cubo do número de terminais externos ($n_{TE_{\alpha 1}} + n_{TE_{\beta 1}} + n_{TE_{\alpha 2}} + n_{TE_{\beta 2}}$) das SRNs. As matrizes $\boldsymbol{\Phi}$, $\boldsymbol{\Phi}_f$, $\boldsymbol{\Phi}_s$, $\boldsymbol{\Phi}_{e_\gamma}$ e $\boldsymbol{\Phi}_{e_\delta}$, introduzidas acima, serão utilizadas posteriormente nas equações de sonda, e são dadas por

$$\boldsymbol{\Phi}(\omega) = \begin{bmatrix} \boldsymbol{M}_{e_\alpha}^{\mathfrak{R}_c}(\omega) - \boldsymbol{H}_{e_\alpha e_\beta}^{\mathfrak{L}_a}(\omega)\boldsymbol{M}_{e_\beta}^{\mathfrak{R}_c}(\omega) \\ \boldsymbol{H}_{e_\beta e_\beta}^{\mathfrak{L}_a}(\omega)\boldsymbol{M}_{e_\beta}^{\mathfrak{R}_c}(\omega) \end{bmatrix}, \tag{4.14.a}$$

$$\boldsymbol{\Phi}_f(\omega) = \begin{bmatrix} \boldsymbol{N}_{e_\alpha f}^{\mathfrak{R}_c}(\omega) - \boldsymbol{H}_{e_\alpha e_\beta}^{\mathfrak{L}_a}(\omega)\boldsymbol{N}_{e_\beta f}^{\mathfrak{R}_c}(\omega) \\ \boldsymbol{H}_{e_\beta e_\beta}^{\mathfrak{L}_a}(\omega)\boldsymbol{N}_{e_\beta f}^{\mathfrak{R}_c}(\omega) \end{bmatrix}, \tag{4.14.b}$$

$$\boldsymbol{\Phi}_\mu(\omega) = \begin{bmatrix} \boldsymbol{H}_{e_\alpha \mu}^{\mathfrak{L}_a}(\omega) \\ \boldsymbol{H}_{e_\beta \mu}^{\mathfrak{L}_a}(\omega) \end{bmatrix}, \tag{4.14.c}$$

onde $\mu = g, e_\alpha, e_\beta$. A equação de estado (4.11) se constitui em um sistema quadrado com dimensão igual ao número de variáveis de estado não-lineares.

Para finalizar a formulação das equações da SuR de fundo precisamos determinar as equações de sonda interna e externa. Iniciando com a equação de sonda interna, vamos utilizar as expressões (4.1) e (4.7) e as condições de contorno (4.9.a)-(4.9.b) e (4.10.a)-(4.10.b). Assim sendo, resolvendo estas equações para o vetor $\boldsymbol{U}_e^{\mathfrak{R}_c}$ e substituindo o resultado em (4.7), e observando que

$$\boldsymbol{Y}_i^{\mathfrak{S}}(\omega) = \left[\boldsymbol{Y}_i^{\mathfrak{R}_c}(\omega)\ \boldsymbol{Y}_i^{\mathfrak{L}_a}(\omega)\right]^T, \tag{4.15}$$

obtemos a equação de sonda interna da SuR, dada por

$$\boldsymbol{Y}_i^{\mathcal{S}}(\omega) = \boldsymbol{M}_i^{\mathcal{S}}(\omega)\boldsymbol{X}^{\mathcal{S}}(\omega) + \boldsymbol{N}_{if}^{\mathcal{S}}(\omega)\boldsymbol{U}_f^{\mathcal{S}}(\omega) + \boldsymbol{N}_{ig}^{\mathcal{S}}(\omega)\boldsymbol{U}_g^{\mathcal{S}}(\omega) + \boldsymbol{N}_{ie_\gamma}^{\mathcal{S}}(\omega)\boldsymbol{U}_{e_\gamma}^{\mathcal{S}}(\omega) + \boldsymbol{N}_{ie_\delta}^{\mathcal{S}}(\omega)\boldsymbol{U}_{e_\delta}^{\mathcal{S}}(\omega) \tag{4.16}$$

onde

$$\boldsymbol{M}_i^{\mathcal{S}}(\omega) = \begin{bmatrix} \boldsymbol{M}_i^{\mathfrak{R}_c}(\omega) \\ \boldsymbol{H}_{ie_\beta}^{\mathcal{L}_a}(\omega)\boldsymbol{M}_{e_\beta}^{\mathfrak{R}_c}(\omega) \end{bmatrix} + \boldsymbol{\Psi}_i(\omega)\boldsymbol{\Phi}(\omega), \tag{4.17.a}$$

$$\boldsymbol{N}_{if}^{\mathcal{S}}(\omega) = \begin{bmatrix} \boldsymbol{N}_{if}^{\mathfrak{R}_c}(\omega) \\ \boldsymbol{H}_{ie_\beta}^{\mathcal{L}_a}(\omega)\boldsymbol{N}_{e_\beta f}^{\mathfrak{R}_c}(\omega) \end{bmatrix} + \boldsymbol{\Psi}_i(\omega)\boldsymbol{\Phi}_f(\omega), \tag{4.17.b}$$

$$\boldsymbol{N}_{i\mu}^{\mathcal{S}}(\omega) = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{H}_{i\mu}^{\mathcal{L}_a}(\omega) \end{bmatrix} + \boldsymbol{\Psi}_i(\omega)\boldsymbol{\Phi}_\mu(\omega), \tag{4.17.c}$$

$$\boldsymbol{\Psi}_i(\omega) = \begin{bmatrix} \boldsymbol{N}_{ie_\alpha}^{\mathfrak{R}_c}(\omega) & \boldsymbol{N}_{ie_\beta}^{\mathfrak{R}_c}(\omega) \\ \boldsymbol{H}_{ie_\alpha}^{\mathcal{L}_a}(\omega) + \boldsymbol{H}_{ie_\beta}^{\mathcal{L}_a}(\omega)\boldsymbol{N}_{e_\beta e_\alpha}^{\mathfrak{R}_c}(\omega) & \boldsymbol{H}_{ie_\beta}^{\mathcal{L}_a}(\omega)\boldsymbol{N}_{e_\beta e_\beta}^{\mathfrak{R}_c}(\omega) \end{bmatrix} \boldsymbol{\Gamma}(\omega)^{-1}, \tag{4.17.d}$$

e $\mu = g, e_\gamma, e_\delta$. Seguindo os mesmos passos para a dedução de (4.16), demonstra-se facilmente que a equação de sonda externa da SuR de fundo (referente àos terminais externos e_γ e e_δ), é dada por

$$\boldsymbol{Y}_\nu^{\mathcal{S}}(\omega) = \boldsymbol{M}_\nu^{\mathcal{S}}(\omega)\boldsymbol{X}^{\mathcal{S}}(\omega) + \boldsymbol{N}_{\nu f}^{\mathcal{S}}(\omega)\boldsymbol{U}_f^{\mathcal{S}}(\omega) + \boldsymbol{N}_{\nu g}^{\mathcal{S}}(\omega)\boldsymbol{U}_g^{\mathcal{S}}(\omega) + \boldsymbol{N}_{\nu e_\gamma}^{\mathcal{S}}(\omega)\boldsymbol{U}_{e_\gamma}^{\mathcal{S}}(\omega) + \boldsymbol{N}_{\nu e_\delta}^{\mathcal{S}}(\omega)\boldsymbol{U}_{e_\delta}^{\mathcal{S}}(\omega) \tag{4.18}$$

onde

$$\boldsymbol{M}_\nu^{\mathcal{S}}(\omega) = \boldsymbol{H}_{\nu e_\beta}^{\mathcal{L}_a}(\omega)\boldsymbol{M}_{e_\beta}^{\mathfrak{R}_c}(\omega) + \boldsymbol{\Psi}_\nu(\omega)\boldsymbol{\Phi}(\omega), \tag{4.19.a}$$

$$\boldsymbol{N}_{\nu f}^{\mathcal{S}}(\omega) = \boldsymbol{H}_{\nu e_\beta}^{\mathcal{L}_a}(\omega)\boldsymbol{N}_{e_\beta f}^{\mathfrak{R}_c}(\omega) + \boldsymbol{\Psi}_\nu(\omega)\boldsymbol{\Phi}_f(\omega), \tag{4.19.b}$$

$$\boldsymbol{N}_{\nu\mu}^{\mathcal{S}}(\omega) = \boldsymbol{H}_{\nu\mu}^{\mathcal{L}_a}(\omega) + \boldsymbol{\Psi}_\nu(\omega)\boldsymbol{\Phi}_\mu(\omega), \tag{4.19.c}$$

$$\boldsymbol{\Psi}_\nu(\omega) = \begin{bmatrix} \boldsymbol{H}_{\nu e_\alpha}^{\mathcal{L}_a}(\omega) + \boldsymbol{H}_{\nu e_\beta}^{\mathcal{L}_a}(\omega)\boldsymbol{N}_{e_\beta e_\alpha}^{\mathfrak{R}_c}(\omega) & \boldsymbol{H}_{\nu e_\beta}^{\mathcal{L}_a}(\omega)\boldsymbol{N}_{e_\beta e_\beta}^{\mathfrak{R}_c}(\omega) \end{bmatrix} \boldsymbol{\Gamma}(\omega)^{-1}, \tag{4.19.d}$$

$\nu = e_\gamma, e_\delta$ e $\mu = g, e_\gamma, e_\delta$. Em resumo, as equações de estado e de sonda interna e externa de uma SuR de fundo são dadas por (4.11), (4.16) e (4.18), respectivamente.

Confirmando o caráter geral e a eficiência da formulação acima, para a determinação das equações de uma SuR de fundo contendo apenas SRNs descrevendo a parte intrínseca dos dispositivos eletrônicos representadas por DDSs (ver Apêndice B), nenhuma matriz precisa ser invertida, tendo em vista que $\boldsymbol{\Gamma}(\omega) = \boldsymbol{1}$. Em contraposição, neste caso, toda a parte extrínseca do dispositivo eletrônico vai para a SRLA, o que aumenta a complexidade desta rede, e consequentemente o custo computacional para calcular a sua equação de sonda (4.1). Esta situação é proporcionalmente mais crítica, em favor da representação completa do dispositivo (parte extrínseca+intríseca), quando se emprega uma grande quantidade de dispositivos iguais, fato

comum em muitos CIs. A maioria dos programas comerciais, e.g. ADS^{TM} [2], utilizam DDSs para representação da parte intrínseca dos dispositivos eletrônicos e opto-eletrônicos, que é um caso particular da formulação proposta.

Agora, vamos comparar os desenvolvimentos acima com a formulação da SuR de fundo utilizando a FEEP proposta em [18]. Nesta formulação, a i-ésima SRN é representada por

$$\boldsymbol{y}_{e_\iota}^{\mathfrak{N}_i}(t) = \boldsymbol{\phi}_\iota^{\mathfrak{N}_i}\left\{\boldsymbol{x}^{\mathfrak{N}_i}(t), \frac{d\boldsymbol{x}^{\mathfrak{N}_i}(t)}{dt}, \frac{d^2\boldsymbol{x}^{\mathfrak{N}_i}(t)}{dt^2}, \ldots, \frac{d^m\boldsymbol{x}^{\mathfrak{N}_i}(t)}{dt^m}, \boldsymbol{x}^{\mathfrak{N}_i}(t-\tau)\right\} \tag{4.20.a}$$

$$\boldsymbol{u}_{e_\iota}^{\mathfrak{N}_i}(t) = \boldsymbol{\psi}_\iota^{\mathfrak{N}_i}\left\{\boldsymbol{x}^{\mathfrak{N}_i}(t), \frac{d\boldsymbol{x}^{\mathfrak{N}_i}(t)}{dt}, \frac{d^2\boldsymbol{x}^{\mathfrak{N}_i}(t)}{dt^2}, \ldots, \frac{d^m\boldsymbol{x}^{\mathfrak{N}_i}(t)}{dt^m}, \boldsymbol{x}^{\mathfrak{N}_i}(t-\tau)\right\} \tag{4.20.b}$$

onde $\iota = \alpha, \beta$, $\boldsymbol{\phi}_\iota^{\mathfrak{N}_i}: \mathrm{R}^{n_{TE_\iota}^{(i)}} \to \mathrm{R}^{n_{TE_\iota}^{(i)}}$ e $\boldsymbol{\psi}_\iota^{\mathfrak{N}_i}: \mathrm{R}^{n_{TE_\iota}^{(i)}} \to \mathrm{R}^{n_{TE_\iota}^{(i)}}$ são vetores de função não-linear no domínio do tempo. Seguindo a notação adotada, $\boldsymbol{u}_{e_\iota}^{\mathfrak{N}_i}$ e $\boldsymbol{y}_{e_\iota}^{\mathfrak{N}_i}$ correspondem aos vetores de fonte e de sonda externa, respectivamente. Infelizmente, aparentemente não existe nenhuma metodologia para se gerar automaticamente as relações paramétricas expressas em (4.20.a) e (4.20.b). Utilizando estas relações, no domínio da frequência, as equações (4.9.a), (4.9.b), (4.10.a) e (4.10.b) assumem a seguinte forma

$$\boldsymbol{Y}_{e_\alpha}^{\mathfrak{L}_a}(\omega) = \boldsymbol{\Phi}_\alpha^{\mathfrak{N}_c}(\omega), \tag{4.21.a}$$

$$\boldsymbol{U}_{e_\alpha}^{\mathfrak{L}_a}(\omega) = \boldsymbol{\Psi}_\alpha^{\mathfrak{N}_c}(\omega), \tag{4.21.b}$$

$$\boldsymbol{Y}_{e_\beta}^{\mathfrak{L}_a}(\omega) = \boldsymbol{\Psi}_\beta^{\mathfrak{N}_c}(\omega), \tag{4.21.c}$$

$$\boldsymbol{U}_{e}^{\mathfrak{L}_a}(\omega) = \boldsymbol{\Phi}_\beta^{\mathfrak{N}_c}(\omega), \tag{4.21.d}$$

respectivamente. Se substituirmos as relações (4.21.a)-(4.21.d) na equação híbrida (4.1), obtemos a equação de estado paramétrica que descreve a SuR de fundo, dada por

$$\mathbf{0} = \boldsymbol{\Phi}_\alpha^{\mathfrak{N}_c}(\omega) - \boldsymbol{H}_{e_\alpha e_\alpha}^{\mathfrak{L}_a}(\omega)\boldsymbol{\Psi}_\alpha^{\mathfrak{N}_c}(\omega) - \boldsymbol{H}_{e_\alpha e_\beta}^{\mathfrak{L}_a}(\omega)\boldsymbol{\Phi}_\beta^{\mathfrak{N}_c}(\omega) - \boldsymbol{H}_{e_\alpha g}^{\mathfrak{L}_a}(\omega)\boldsymbol{U}_g^{\mathfrak{L}_a}(\omega) - \boldsymbol{H}_{e_\alpha e_\gamma}^{\mathfrak{L}_a}(\omega)\boldsymbol{U}_{e_\gamma}^{\mathfrak{L}_a}(\omega) - \boldsymbol{H}_{e_\alpha e_\delta}^{\mathfrak{L}_a}(\omega)\boldsymbol{U}_{e_\delta}^{\mathfrak{L}_a}(\omega), \tag{4.22.a}$$

$$\mathbf{0} = \boldsymbol{\Psi}_\beta^{\mathfrak{N}_c}(\omega) - \boldsymbol{H}_{e_\beta e_\alpha}^{\mathfrak{L}_a}(\omega)\boldsymbol{\Phi}_\alpha^{\mathfrak{N}_c}(\omega) - \boldsymbol{H}_{e_\beta e_\beta}^{\mathfrak{L}_a}(\omega)\boldsymbol{\Phi}_\beta^{\mathfrak{N}_c}(\omega) - \boldsymbol{H}_{e_\beta g}^{\mathfrak{L}_a}(\omega)\boldsymbol{U}_g^{\mathfrak{L}_a}(\omega) - \boldsymbol{H}_{e_\beta e_\gamma}^{\mathfrak{L}_a}(\omega)\boldsymbol{U}_{e_\gamma}^{\mathfrak{L}_a}(\omega) - \boldsymbol{H}_{e_\beta e_\delta}^{\mathfrak{L}_a}(\omega)\boldsymbol{U}_{e_\delta}^{\mathfrak{L}_a}(\omega), \tag{4.22.b}$$

Em geral, o sistema não-linear acima é *retangular* ($n_{NSV} \neq n_{TE}$). Sendo assim, para quadrá-lo deve-se prover um conjunto adicional de equações resultante da aplicação da LKT e/ou da LKC na SRN. Apesar de não apresentada aqui, devido à limitação de espaço, a FEE proposta neste trabalho, quando comparada com a FEEP, também oferece uma maior flexibilidade na representação do ruído.

[2] ADS^{TM} = Advanced Design System

4.5. Formulação Multi-Níveis das Equações do Circuito

Para desenvolvermos a formulação multi-níveis das equações do circuito, vamos considerar inicialmente a estrutura descrita na Fig. 2.1(c), e assumir que a hierarquia do circuito é formada de υ níveis, com $\upsilon > 1$. O processo de formulação descrito abaixo, é recursivo, iniciando do último nível superior até o nível 0 (zero) que corresponde ao topo da hierarquia. Seguindo a notação e convenção adotada no Capítulo 2, e assumindo que $j, l \in [1, n_{SuR}^{(\upsilon-1)}]$ são os índices da primeira e da última SuR de nível υ, contidas na SuR $\mathfrak{S}_{\upsilon-1,i}$. Também vamos assumir $(\upsilon, k) \equiv (\mathfrak{S}_{\upsilon,k})$ para que os superescritos que indicam a SuR. Desta forma, $l-j+1$ é o número de SuRs no nível υ, contidas na $\mathfrak{S}_{\upsilon-1,i}$. Os vetores de variável de estado, de função não-linear, de fonte independente, e de sonda (interna e externa) da $\mathfrak{S}_{\upsilon-1,i}$, são dados por:

$$\boldsymbol{X}^{(\upsilon-1,i)}(\omega) = \left[\boldsymbol{X}^{(\upsilon,j)}(\omega) \ \ldots \ \boldsymbol{X}^{(\upsilon,l)}(\omega) \ \boldsymbol{X}_{\gamma}^{(\upsilon-1,i)}(\omega)\right]^T, \tag{4.23.a}$$

$$\boldsymbol{U}_{\mu}^{(\upsilon-1,i)}(\omega) = \left[\boldsymbol{U}_{\mu}^{(\upsilon,j)}(\omega) \ \ldots \ \boldsymbol{U}_{\mu}^{(\upsilon,l)}(\omega)\right]^T, \tag{4.23.b}$$

$$\boldsymbol{Y}_{\nu}^{(\upsilon-1,i)}(\omega) = \left[\boldsymbol{Y}_{\nu}^{(\upsilon,j)}(\omega) \ \ldots \ \boldsymbol{Y}_{\nu}^{(\upsilon,l)}(\omega)\right]^T, \tag{4.23.c}$$

onde $\mu = f, s$ e $\nu = i, e_\gamma, e_\delta$, respectivamente.

Para $j \le k \le l$, as equações de estado e de sonda (interna e externa) da k-ésima SuR de fundo no υ-ésimo nível, $\mathfrak{S}_{\upsilon,k}$, são dadas por:

$$\begin{aligned}\mathbf{0} = {} & \boldsymbol{A}^{(\upsilon,k)}(\omega)\boldsymbol{X}^{(\upsilon,k)}(\omega) + \boldsymbol{B}_f^{(\upsilon,k)}(\omega)\boldsymbol{U}_f^{(\upsilon,k)}(\omega) + \boldsymbol{B}_g^{(\upsilon,k)}(\omega)\boldsymbol{U}_g^{(\upsilon,k)}(\omega) \\ & + \boldsymbol{B}_{e_\gamma}^{(\upsilon,k)}(\omega)\boldsymbol{U}_{e_\gamma}^{(\upsilon,k)}(\omega) + \boldsymbol{B}_{e_\delta}^{(\upsilon,k)}(\omega)\boldsymbol{U}_{e_\delta}^{(\upsilon,k)}(\omega)\end{aligned}, \tag{4.24}$$

$$\begin{aligned}\boldsymbol{Y}_{\nu}^{(\upsilon,k)}(\omega) = {} & \boldsymbol{M}_{\nu}^{(\upsilon,k)}(\omega)\boldsymbol{X}^{(\upsilon,k)}(\omega) + \boldsymbol{N}_{\nu f}^{(\upsilon,k)}(\omega)\boldsymbol{U}_f^{(\upsilon,k)}(\omega) + \boldsymbol{N}_{\nu g}^{(\upsilon,k)}(\omega)\boldsymbol{U}_g^{(\upsilon,k)}(\omega) \\ & + \boldsymbol{N}_{\nu e_\gamma}^{(\upsilon,k)}(\omega)\boldsymbol{U}_{e_\gamma}^{(\upsilon,k)}(\omega) + \boldsymbol{N}_{\nu e_\delta}^{(\upsilon,k)}(\omega)\boldsymbol{U}_{e_\delta}^{(\upsilon,k)}(\omega)\end{aligned}, \tag{4.25}$$

respectivamente, onde $\nu = i, e_\gamma, e_\delta$. Introduzindo, n_{TE_ν} como o número de sondas externas do tipo, e_ν, e os vetores de sonda externa

$$\boldsymbol{Y}_{e_\gamma}^{(\upsilon-1,i)}(\omega) = \left[\ldots \ I_{e_\gamma, r1}^{(\upsilon-1,i)}(\omega) \ \ldots \ V_{e_\gamma, s1}^{(\upsilon-1,i)}(\omega) \ \ldots\right]^T \in \mathbb{C}^{n_{TE_\gamma}^{(\upsilon-1,i)}}, \tag{4.26.a}$$

$$\boldsymbol{Y}_{e_\delta}^{(\upsilon-1,i)}(\omega) = \left[\ldots \ I_{e_\delta, r2}^{(\upsilon-1,i)}(\omega) \ \ldots \ V_{e_\delta, s2}^{(\upsilon-1,i)}(\omega) \ \ldots\right]^T \in \mathbb{C}^{n_{TE_\delta}^{(\upsilon-1,i)}}, \tag{4.26.b}$$

$$\boldsymbol{Y}_{e_\gamma}^{(\upsilon,t)}(\omega) = \left[\ldots \ I_{e_\gamma, p1}^{(\upsilon,t)}(\omega) \ \ldots \ V_{e_\gamma, q1}^{(\upsilon,t)}(\omega) \ \ldots\right]^T \in \mathbb{C}^{n_{TE_\gamma}^{(\upsilon,i)}}, \tag{4.26.c}$$

$$\boldsymbol{Y}_{e_\delta}^{(\upsilon,k)}(\omega) = \left[\ldots \ I_{e_{\delta 1}, p_2}^{(\upsilon,k)}(\omega) \ \ldots \ I_{e_\delta, p_3}^{(\upsilon,k)}(\omega) \ \ldots \ V_{e_\delta, q_2}^{(\upsilon,k)}(\omega) \ \ldots \ V_{e_\delta, q_3}^{(\upsilon,k)}(\omega) \ \ldots\right]^T \in \mathbb{C}^{n_{TE_\delta}^{(\upsilon,k)}}, \tag{4.26.d}$$

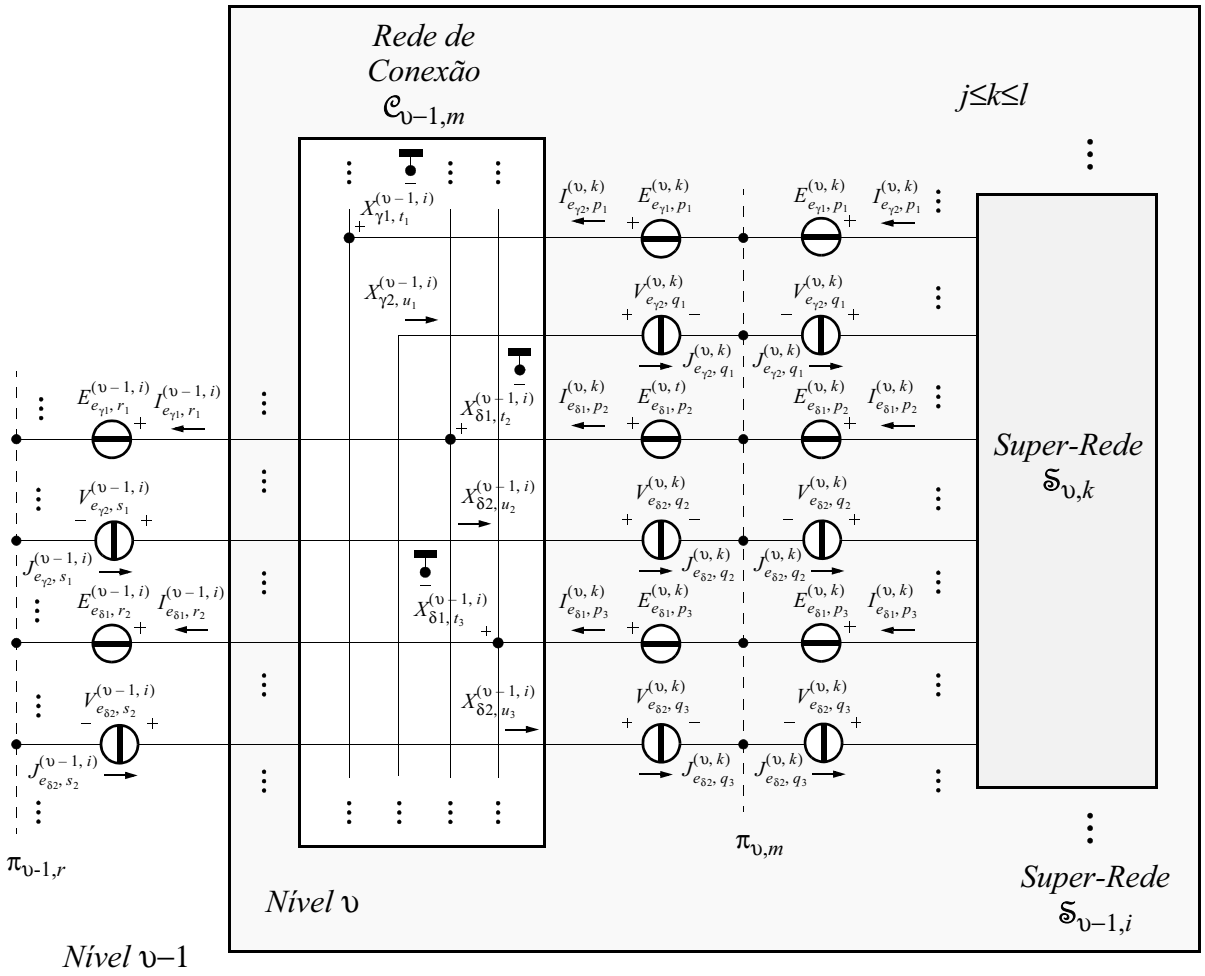


Fig. 4.2 Esquema elétrico para conexões externas das super-redes (SuRs) associadas com a estrutura da Fig. 2.1(c).

para $k = j, \ldots, l$, então, de acordo com o esquema elétrico da Fig. 4.2, podemos escrever as seguintes relações,

$$\boldsymbol{Y}_{e_\gamma}^{(\upsilon-1,\,i)}(\omega) \;=\; \sum_{k=j}^{l} \mathbf{C}_{e_\gamma e_\delta}^{(\upsilon,\,k)} \boldsymbol{Y}_{e_\delta}^{(\upsilon,\,k)}(\omega)\,, \tag{4.27}$$

$$\boldsymbol{Y}_{e_\delta}^{(\upsilon-1,\,i)}(\omega) \;=\; \sum_{k=j}^{l} \mathbf{C}_{e_\delta e_\delta}^{(\upsilon,\,k)} \boldsymbol{Y}_{e_\delta}^{(\upsilon,\,k)}(\omega)\,, \tag{4.28}$$

$$\mathbf{0} \;=\; \sum_{k=j}^{l} \mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)} \boldsymbol{Y}_{e_\gamma}^{(\upsilon,\,k)}(\omega)\,, \tag{4.29}$$

onde $\mathbf{C}_{e_\gamma e_\delta}^{(\upsilon,\,k)}$, $\mathbf{C}_{e_\delta e_\delta}^{(\upsilon,\,k)}$ e $\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)}$, são matrizes de incidência referentes à aplicação da LKT e da LKC na rede de conexão $\mathcal{C}_{\upsilon-1,\,m}$. Introduzindo os vetores de fonte externa

$$\boldsymbol{U}_{e_\gamma}^{(\upsilon-1,\,i)}(\omega)=\left[\ldots\ E_{e_\gamma,\,r_1}^{(\upsilon-1,\,i)}(\omega)\ \ldots\ J_{e_\gamma,\,s_1}^{(\upsilon-1,\,i)}(\omega)\ \ldots\right]^T\in\mathbb{C}^{n_{TE_\gamma}^{(\upsilon-1,\,i)}},\tag{4.30.a}$$

$$\boldsymbol{U}_{e_\delta}^{(\upsilon-1,\,i)}(\omega)=\left[\ldots\ E_{e_\gamma,\,r_2}^{(\upsilon-1,\,i)}(\omega)\ \ldots\ J_{e_\delta,\,s_2}^{(\upsilon-1,\,i)}(\omega)\ \ldots\right]^T\in\mathbb{C}^{n_{TE_\delta}^{(\upsilon-1,\,i)}},\tag{4.30.b}$$

$$\boldsymbol{U}_{e_\gamma}^{(\upsilon,\,k)}(\omega)=\left[\ldots\ E_{e_\gamma,\,p_1}^{(\upsilon,\,k)}(\omega)\ \ldots\ J_{e_\gamma,\,q_1}^{(\upsilon,\,k)}(\omega)\ \ldots\right]^T\in\mathbb{C}^{n_{TE_\gamma}^{(\upsilon,\,k)}},\tag{4.30.c}$$

$$\boldsymbol{U}_{e_\delta}^{(\upsilon,\,k)}(\omega)=\left[\ldots\ E_{e_\delta,\,p_2}^{(\upsilon,\,k)}(\omega)\ \ldots\ E_{e_\delta,\,p_3}^{(\upsilon,\,k)}(\omega)\ \ldots\ J_{e_\delta,\,q_2}^{(\upsilon,\,k)}(\omega)\ \ldots\ J_{e_\delta,\,q_3}^{(\upsilon,\,k)}(\omega)\ \ldots\right]^T\in\mathbb{C}^{n_{TE_\delta}^{(\upsilon,\,k)}},\tag{4.30.d}$$

então, podemos escrever as seguintes relações:

$$\boldsymbol{U}_{e_\gamma}^{(\upsilon,\,k)}(\omega)\ =\ (\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)})^T\boldsymbol{X}_\gamma^{(\upsilon-1,\,i)}(\omega),\tag{4.31}$$

$$\boldsymbol{U}_{e_\delta}^{(\upsilon,\,k)}(\omega)\ =\ \mathbf{B}_{e_\delta\delta}^{(\upsilon,\,k)}\boldsymbol{X}_\delta^{(\upsilon-1,\,i)}(\omega),\tag{4.32}$$

$$\boldsymbol{X}_\delta^{(\upsilon-1,\,i)}(\omega)\ =\ \mathbf{D}_{\delta e_\gamma}^{(\upsilon-1,\,i)}\boldsymbol{U}_{e_\gamma}^{(\upsilon-1,\,i)}(\omega)+\mathbf{D}_{\delta e_\delta}^{(\upsilon-1,\,i)}\boldsymbol{U}_{e_\delta}^{(\upsilon-1,\,i)}(\omega).\tag{4.33}$$

Substituindo (4.25), (4.31) e (4.32) na equação de conexão interna (4.29), obtemos o seguinte resultado

$$\begin{aligned}\mathbf{0}\ =\ \sum_{k=j}^{l}(&\boldsymbol{C}^{(\upsilon,\,k)}(\omega)\boldsymbol{X}^{(\upsilon,\,k)}(\omega)+\boldsymbol{D}_f^{(\upsilon,\,k)}(\omega)\boldsymbol{U}_f^{(\upsilon,\,k)}(\omega)+\boldsymbol{D}_s^{(\upsilon,\,k)}(\omega)\boldsymbol{U}_s^{(\upsilon,\,k)}(\omega))\\&+\boldsymbol{C}_\gamma^{(\upsilon-1,\,i)}(\omega)\boldsymbol{X}_\gamma^{(\upsilon-1,\,i)}(\omega)+\boldsymbol{D}_\delta^{(\upsilon-1,\,i)}(\omega)\boldsymbol{X}_\delta^{(\upsilon-1,\,i)}(\omega)\end{aligned}\tag{4.34}$$

onde

$$\boldsymbol{C}^{(\upsilon,\,k)}(\omega)\ =\ \mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)}\boldsymbol{M}_{e_\gamma}^{(\upsilon,\,k)}(\omega),\tag{4.35.a}$$

$$\boldsymbol{D}_\mu^{(\upsilon,\,k)}(\omega)\ =\ \mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)}\boldsymbol{N}_{e_\gamma\mu}^{(\upsilon,\,k)}(\omega),\tag{4.35.b}$$

$$\boldsymbol{C}_\gamma^{(\upsilon-1,\,i)}(\omega)\ =\ \sum_{k=j}^{l}\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)}\boldsymbol{N}_{e_\gamma e_\gamma}^{(\upsilon,\,k)}(\omega)(\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)})^T,\tag{4.35.c}$$

$$\boldsymbol{D}_\delta^{(\upsilon-1,\,i)}(\omega)\ =\ \sum_{k=j}^{l}\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)}\boldsymbol{N}_{e_\gamma e_\delta}^{(\upsilon,\,k)}(\omega)\mathbf{B}_{e_\delta\delta}^{(\upsilon,\,k)},\tag{4.35.d}$$

com $\mu = f, s$. A equação (4.34) representa as interconexões de todas as SuRs de fundo $j, \ldots, l$ de nível υ, contidas em $\mathfrak{S}_{\upsilon-1,\,i}$.

Finalmente, utilizando os vetores (4.23.a)-(4.23.c), (4.26.a)-(4.26.d) e (4.30.a)-(4.30.d), e substituindo (4.31) em (4.24) e (4.34), obtemos a equação de estado da SuR intermediária, $\mathfrak{S}_{\upsilon-1,\,i}$. Esta equação de estado assume a mesma forma de (4.24), após as seguintes substituições: $(\upsilon, k)\rightarrow(\upsilon-1, i)$. As matrizes constitutivas da equação de estado são dadas por:

$$\boldsymbol{A}^{(\upsilon-1,\,i)}(\omega) = \begin{bmatrix} \boldsymbol{A}^{(\upsilon,j)}(\omega) & & \boldsymbol{A}_{\gamma}^{(\upsilon,j)}(\omega) \\ & \cdots & \cdots \\ & \boldsymbol{A}^{(\upsilon,l)}(\omega) & \boldsymbol{A}_{\gamma}^{(\upsilon,l)}(\omega) \\ \boldsymbol{C}^{(\upsilon,j)}(\omega) \;\ldots\; \boldsymbol{C}^{(\upsilon,l)}(\omega) & & \boldsymbol{C}_{\gamma}^{(\upsilon-1,\,i)}(\omega) \end{bmatrix}, \tag{4.36.a}$$

$$\boldsymbol{B}_{\mu}^{(\upsilon-1,\,i)}(\omega) = \begin{bmatrix} \boldsymbol{B}_{\mu}^{(\upsilon,j)}(\omega) & & \\ & \cdots & \\ & & \boldsymbol{B}_{\mu}^{(\upsilon,l)}(\omega) \\ \boldsymbol{D}_{\mu}^{(\upsilon,j)}(\omega) \;\ldots & & \boldsymbol{D}_{\mu}^{(\upsilon,l)}(\omega) \end{bmatrix}, \tag{4.36.b}$$

$$\boldsymbol{B}_{\nu}^{(\upsilon-1,\,i)}(\omega) = \begin{bmatrix} \boldsymbol{B}_{e_\delta}^{(\upsilon,j)}(\omega)\mathbf{B}_{e_\delta\delta}^{(\upsilon,j)} \\ \cdots \\ \boldsymbol{B}_{e_\delta}^{(\upsilon,l)}(\omega)\mathbf{B}_{e_\delta\delta}^{(\upsilon,l)} \\ \boldsymbol{D}_{\delta}^{(\upsilon-1,\,i)}(\omega) \end{bmatrix} \mathbf{D}_{\delta\nu}^{(\upsilon-1,\,i)}, \tag{4.36.c}$$

$$\boldsymbol{A}_{\gamma}^{(\upsilon,k)}(\omega) = \boldsymbol{B}_{e_\gamma}^{(\upsilon,k)}(\omega)(\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,k)})^{T}, \tag{4.36.d}$$

onde $\mu = f, g$ e $\nu = e_\gamma, e_\delta$. Podemos observar que as matrizes (4.36.a) e (4.36.b) assumem uma estrutura bloco diagonal com borda dupla e bloco diagonal com borda simples, respectivamente.

Para obtermos as equações de sonda (interna e externa) da SuR intermediária, $\mathfrak{s}_{\upsilon-1,\,i}$, devemos substituir (4.31) em (4.25). Estas equações assumem a mesma forma de (4.25) após as seguintes substituições $(\upsilon, k) \rightarrow (\upsilon - 1, i)$. Para as sondas internas, temos que

$$\boldsymbol{M}_{i}^{(\upsilon-1,\,i)}(\omega) = \begin{bmatrix} \boldsymbol{M}_{i}^{(\upsilon,j)}(\omega) & & \boldsymbol{M}_{i\gamma}^{(\upsilon,j)}(\omega) \\ & \cdots & \cdots \\ & \boldsymbol{M}_{i}^{(\upsilon,l)}(\omega) & \boldsymbol{M}_{i\gamma}^{(\upsilon,l)}(\omega) \end{bmatrix}, \tag{4.37.a}$$

$$\boldsymbol{N}_{i\mu}^{(\upsilon-1,\,i)}(\omega) = \begin{bmatrix} \boldsymbol{N}_{i\mu}^{(\upsilon,j)}(\omega) & & \\ & \cdots & \\ & & \boldsymbol{N}_{i\mu}^{(\upsilon,l)}(\omega) \end{bmatrix}, \tag{4.37.b}$$

$$\boldsymbol{N}_{i\nu}^{(\upsilon-1,\,i)}(\omega) = \begin{bmatrix} \boldsymbol{N}_{ie_\delta}^{(\upsilon,l)}(\omega)\mathbf{B}_{e_\delta\delta}^{(\upsilon,j)} \\ \cdots \\ \boldsymbol{N}_{ie_\delta}^{(\upsilon,l)}(\omega)\mathbf{B}_{e_\delta\delta}^{(\upsilon,l)} \end{bmatrix} \mathbf{D}_{\delta\nu}^{(\upsilon-1,\,i)}, \tag{4.37.c}$$

$$\boldsymbol{M}_{i\gamma}^{(\upsilon,k)}(\omega) = \boldsymbol{N}_{ie_\gamma}^{(\upsilon,k)}(\omega)(\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,k)})^{T} \tag{4.37.d}$$

onde $\mu = f, g$ e $\nu = e_\gamma, e_\delta$. Para as sondas externas, temos que

$$\boldsymbol{M}_{\nu}^{(\upsilon-1,\,i)}(\omega) \;=\; \left[\mathbf{C}_{\nu e_\delta}^{(\upsilon,j)}\boldsymbol{M}_{e_\delta}^{(\upsilon,j)}(\omega)\;\ldots\;\mathbf{C}_{\nu e_\delta}^{(\upsilon,\,l)}\boldsymbol{M}_{e_\delta}^{(\upsilon,\,l)}(\omega)\;\boldsymbol{M}_{\nu\gamma}^{(\upsilon-1,\,i)}(\omega)\right], \tag{4.38.a}$$

$$\boldsymbol{M}_{\nu\gamma}^{(\upsilon-1,\,i)}(\omega) \;=\; \sum_{k=j}^{l}\mathbf{C}_{\nu e_\delta}^{(\upsilon,\,k)}\boldsymbol{N}_{e_\delta e_\gamma}^{(\upsilon,\,k)}(\omega)(\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)})^{T}, \tag{4.39.a}$$

$$\boldsymbol{N}_{\nu\mu}^{(\upsilon-1,\,i)}(\omega) \;=\; \left[\mathbf{C}_{\nu e_\delta}^{(\upsilon,j)}\boldsymbol{N}_{e_\delta\mu}^{(\upsilon,j)}(\omega)\;\ldots\;\mathbf{C}_{\nu e_\delta}^{(\upsilon,\,l)}\boldsymbol{N}_{e_\delta\mu}^{(\upsilon,\,l)}(\omega)\right], \tag{4.39.b}$$

$$\boldsymbol{N}_{\nu\nu'}^{(\upsilon-1,\,i)}(\omega) \;=\; \boldsymbol{N}_{\nu\delta}^{(\upsilon-1,\,i)}(\omega)\mathbf{D}_{\delta\nu'}^{(\upsilon-1,\,i)}, \tag{4.39.c}$$

$$\boldsymbol{N}_{\nu\delta}^{(\upsilon-1,\,i)}(\omega) \;=\; \sum_{k=j}^{l}\mathbf{C}_{\nu e_\delta}^{(\upsilon,\,k)}\boldsymbol{N}_{e_\delta e_\delta}^{(\upsilon,\,k)}(\omega)\mathbf{B}_{e_\delta\delta}^{(\upsilon,\,k)}, \tag{4.39.d}$$

onde $\nu = e_\gamma, e_\delta$, $\nu' = e_\gamma, e_\delta$ e $\mu = f, g$. Podemos observar que as matrizes (4.37.a)/(4.38.a) e (4.37.b)/(4.39.b) assumem uma estrutura bloco diagonal com borda simples e bloco diagonal (sem bordas), respectivamente

As matrizes (4.36.c), (4.37.c) e (4.39.c) contém as informações da $\mathfrak{s}_{\upsilon-1,\,i}$ que serão transmitidas para o nível inferior subsequente da estrutura hierárquica. Se esta for a SuR de topo, então as matrizes em questão não são definidas. Caso contrário, o processo acima é repetido recursivamente para cada nível até chegar à SuR de topo.

Excluindo a formulação das equações das SuRs de fundo, a formulação multi-níveis das equações de circuito em grande-escala, apresentada acima, envolve apenas operações de adição, de multiplicação e de manipulação com bloco de matrizes densas e esparsas.

4.6. Eliminação de Níveis da SuR Intermediária

Conforme discutido anteriormente, o aumento no número de SRNs produz uma maior complexidade numérica na formulação das equações de uma SuR de fundo. Isto deve-se ao fato, da dimensão da matriz Γ, em (4.13), ser diretamente proporcional ao número de terminais externos das SRNs. Com a decomposição multi-níveis, a complexidade da formulação (inversão de Γ) pode ser facilmente controlada. Para este fim, vamos introduzir um procedimento para eliminação dos níveis de hierarquia de determinadas SuRs intermediárias, i.e., transformá-las em SuRs de fundo. Isto possibilita maior flexibilidade no controle da estrutura hierárquica do circuito. Para o nivelamento, ou eliminação de níveis, de uma determinada SuR intermediária, vamos introduzir o seguinte procedimento recursivo.

Sem perda de generalidade, vamos considerar que a SuR intermediária (parente), $\mathfrak{s}_{\upsilon,\,i}$, é composta de $l-j+1$ SuRs de fundo (crianças) no nível $\upsilon-1$. Assume-se, que SuR intermediária

(parente), a ser nivelada, só possui SuRs crianças do tipo SuR de fundo (naturais ou niveladas). O nosso objetivo é transformar a SuR intermediária, $\mathfrak{S}_{\upsilon-1,\,i}$, em uma SuR de fundo. Para tal, o vetor de variável de estado associado a $\mathfrak{S}_{\upsilon-1,\,i}$ deve ser organizado da seguinte forma

$$\boldsymbol{X}^{(\upsilon-1,\,i)}(\omega) \;=\; \left[\boldsymbol{X}^{(\upsilon,\,j)}(\omega)\;\ldots\;\boldsymbol{X}^{(\upsilon,\,l)}(\omega)\right]^{T}. \tag{4.40}$$

A equação de estado das SuRs de fundo (crianças), $\mathfrak{S}_{\upsilon,\,k}$, com $k \in [j,\,l]$ é dada por (4.24). Sendo assim, isolando o vetor $\boldsymbol{X}_{\gamma}^{(\upsilon-1,\,i)}(\omega)$ em (4.34), e substituindo o resultado na equação (4.24) via (4.31)-(4.33), obtemos a equação de estado da SuR, $\mathbb{S}_{\upsilon-1,\,i}$, dada por:

$$\boldsymbol{\Lambda}^{(\upsilon-1,\,i)}(\omega) \;=\; \begin{bmatrix} \boldsymbol{\Lambda}^{(\upsilon,\,j)}(\omega)+\boldsymbol{\Xi}_{x}^{(\upsilon,\,j)}(\omega)\boldsymbol{\Pi}^{(\upsilon,\,j)}(\omega) & \ldots & \boldsymbol{\Xi}_{x}^{(\upsilon,\,j)}(\omega)\boldsymbol{\Pi}^{(\upsilon,\,l)}(\omega) \\ \ldots & \ldots & \ldots \\ \boldsymbol{\Xi}_{x}^{(\upsilon,\,l)}(\omega)\boldsymbol{\Pi}^{(\upsilon,\,j)}(\omega) & \ldots & \boldsymbol{\Lambda}^{(\upsilon,\,l)}(\omega)+\boldsymbol{\Xi}_{x}^{(\upsilon,\,l)}(\omega)\boldsymbol{\Pi}^{(\upsilon,\,l)}(\omega) \end{bmatrix}, \tag{4.41}$$

$$\boldsymbol{B}_{\nu}^{(\upsilon-1,\,i)}(\omega) \;=\; \begin{bmatrix} (\boldsymbol{B}_{e_\delta}^{(\upsilon,\,j)}(\omega)\mathbf{B}_{\delta e_\delta}^{(\upsilon,\,j)}+\boldsymbol{\Xi}_{x}^{(\upsilon,\,j)}(\omega)\boldsymbol{C}_{\delta}^{(\upsilon-1,\,i)}(\omega))\mathbf{D}_{e_\delta\nu}^{(\upsilon,\,j)} \\ \ldots \\ (\boldsymbol{B}_{e_\delta}^{(\upsilon,\,l)}(\omega)\mathbf{B}_{\delta e_\delta}^{(\upsilon,\,l)}+\boldsymbol{\Xi}_{x}^{(\upsilon,\,l)}(\omega)\boldsymbol{C}_{\delta}^{(\upsilon-1,\,i)}(\omega))\mathbf{D}_{e_\delta\nu}^{(\upsilon,\,l)} \end{bmatrix}, \tag{4.42}$$

$$\Xi_{x}^{(\upsilon,\,k)}(\omega) \;=\; -\boldsymbol{B}_{e_\gamma}^{(\upsilon,\,k)}(\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)})^{T}(\omega)\boldsymbol{C}_{\gamma}^{(\upsilon-1,\,i)}(\omega)^{-1}, \tag{4.43}$$

onde $\nu = e_\gamma, e_\delta$, e assumindo as seguintes equivalências: $(\boldsymbol{\Lambda}, \boldsymbol{\Pi}) = (\boldsymbol{A}, \boldsymbol{C}), (\boldsymbol{B}_f, \boldsymbol{D}_f), (\boldsymbol{B}_g, \boldsymbol{D}_g)$.

A equação de sonda (interna e externa), $\mathfrak{S}_{\upsilon,\,k}$, com $k \in [j,\,l]$ é dada por (4.25). Seguindo os mesmos passos acima, obtemos a equação de sonda (interna e externa) da SuR, $\mathbb{S}_{\upsilon-1,\,i}$, dada por:

$$\Lambda^{(\upsilon-1,\,i)}(\omega) \;=\; \begin{bmatrix} \Lambda^{(\upsilon,\,j)}(\omega)+\Xi_{y_\mu}^{(\upsilon,\,j)}(\omega)\Pi^{(\upsilon,\,j)}(\omega) & \ldots & \Xi_{y_\mu}^{(\upsilon,\,j)}(\omega)\Pi^{(\upsilon,\,l)}(\omega) \\ \ldots & \ldots & \ldots \\ \Xi_{y_\mu}^{(\upsilon,\,l)}(\omega)\Pi^{(\upsilon,\,j)}(\omega) & \ldots & \Lambda^{(\upsilon,\,l)}(\omega)+\Xi_{y_\mu}^{(\upsilon,\,l)}(\omega)\Pi^{(\upsilon,\,l)}(\omega) \end{bmatrix}, \tag{4.44}$$

$$\boldsymbol{N}_{\mu\nu}^{(\upsilon-1,\,i)}(\omega) \;=\; \begin{bmatrix} (\boldsymbol{N}_{\mu e_\delta}^{(\upsilon,\,j)}(\omega)\mathbf{B}_{\delta e_\delta}^{(\upsilon,\,j)}+\boldsymbol{\Xi}_{y_\mu}^{(\upsilon,\,j)}(\omega)\boldsymbol{C}_{\delta}^{(\upsilon-1,\,i)}(\omega))\mathbf{D}_{e_\delta\nu}^{(\upsilon,\,j)} \\ \ldots \\ (\boldsymbol{N}_{\mu e_\delta}^{(\upsilon,\,l)}(\omega)\mathbf{B}_{\delta e_\delta}^{(\upsilon,\,l)}+\boldsymbol{\Xi}_{y_\mu}^{(\upsilon,\,l)}(\omega)\boldsymbol{C}_{\delta}^{(\upsilon-1,\,i)}(\omega))\mathbf{D}_{e_\delta\nu}^{(\upsilon,\,l)} \end{bmatrix}, \tag{4.45}$$

$$\Xi_{y_\mu}^{(\upsilon,\,k)}(\omega) \;=\; -\boldsymbol{N}_{\mu e_\gamma}^{(\upsilon,\,k)}(\mathbf{A}_{\gamma e_\gamma}^{(\upsilon,\,k)})^{T}(\omega)\boldsymbol{C}_{\gamma}^{(\upsilon-1,\,i)}(\omega)^{-1}, \tag{4.46}$$

onde $\mu = i, e_\gamma, e_\delta$, $\nu = e_\gamma, e_\delta$ e $(\boldsymbol{\Lambda}, \boldsymbol{\Pi}) = (\boldsymbol{M}_\mu, \boldsymbol{C}), (\boldsymbol{N}_{\mu f}, \boldsymbol{D}_f), (\boldsymbol{N}_{\mu s}, \boldsymbol{D}_g)$.

4.7. Exemplo Ilustrativo

Referindo-se ao exemplo da Fig. 2.3, vamos assumir que cada transistor bipolar intruduz 3 (três)

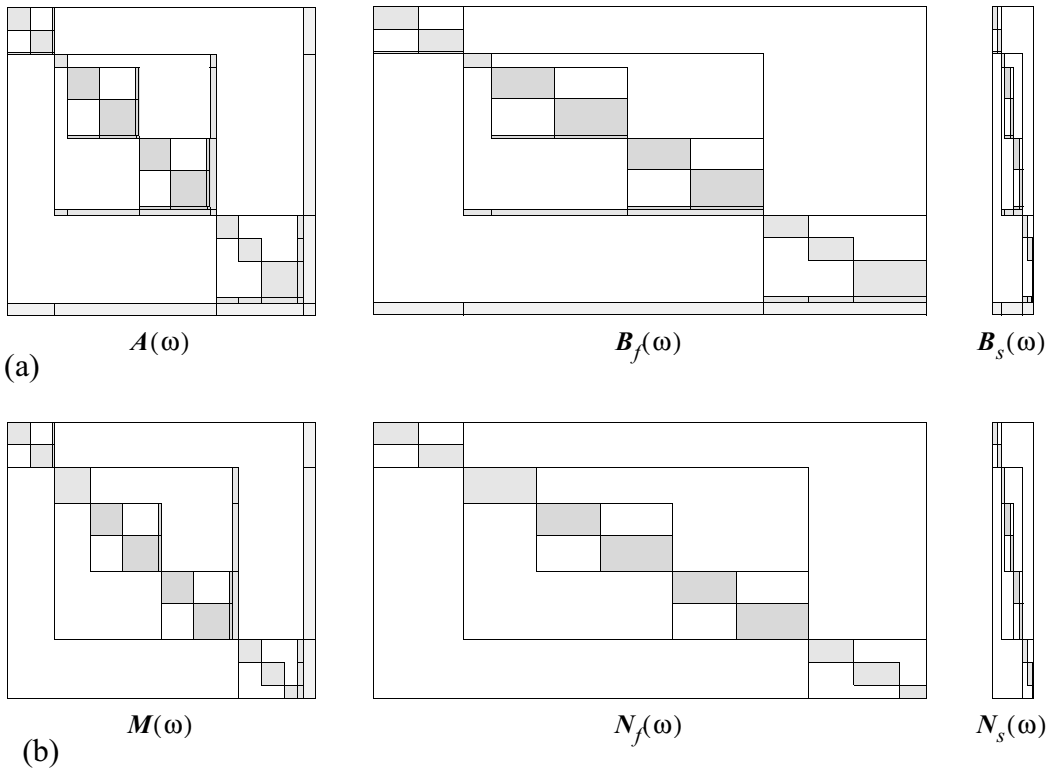


Fig. 4.3 Estrutura multi-níveis, tipo bloco diagonal com e sem borda das matrizes constitutivas da equação (a) de estado e (b) de sonda do circuito da Fig. 2.3.

variáveis de estado não-lineares e 6 (seis) funções não-lineares (ver Fig. 3.3(b)). Desta forma, com uma decomposição em 3 (três) níveis, para o circuito do receptor temos um total de 183 variáveis de estado não-lineares dos transistores, 22 variáveis de estado das redes de conexão, 366 funções não-lineares dos transistores, 27 fontes independentes. Em relação às fontes independentes, temos 26 de polarização CC e 1 (uma) de RF, representando o sinal captado pela antena de recepção. O número de sondas é arbitrário.

Nas Figs. 4.3(a) e (b), podemos observar a estrutura do padrão de blocos não-zero das matrizes que constituem a equação de estado e de sonda, respectivamente. A matriz $\boldsymbol{A}(\omega)$ assume uma estrutura do tipo bloco diagonal com borda dupla, as matrizes $\boldsymbol{B}_f(\omega)$ e $\boldsymbol{B}_s(\omega)$ são do tipo bloco diagonal com borda simples de orientação horizontal, a matriz $\boldsymbol{M}(\omega)$ é do tipo bloco diagonal com borda simples de orientação vertical, e as matriz $\boldsymbol{N}_f(\omega)$ e $\boldsymbol{N}_s(\omega)$ são do tipo bloco diagonal. Conforme será visto mais adiante, a principal vantagem destas estruturas é o fato de preservarem o padrão de blocos não-zero da matriz jacobiana, após uma fatorização LU.

4.8. Conclusão

Apresentamos, neste capítulo, uma eficiente metodologia para a formulação das equações de

estado e de sonda, no domínio da frequência, de um circuito em grande-escala. Esta formulação está fundamentada na técnica de decomposição multi-níveis do circuito e nas formulações da SRL e da SRN, introduzidas nos capítulos 2 e 3, respectivamente. Lembremos que, para uma maior eficiência, assume-se que em cada nível de hierarquia as SuRs intermediárias e de fundo estão esparsamente interconectadas.

Podemos resumir o processo de formulação multi-níveis das equações de circuito, apresentado acima, da seguinte forma: (i) formulação da equação de sonda da SRLA (equações híbridas) obtidas via fatorização LU da matriz nodal-modificada (procedimento padrão), (ii) formulação das equações de estado (com menor dimensão possível) e de sonda (interna e externa) da SRNC via FEE, (iii) formulação das equações de estado e de sonda das SuRs de fundo, envolvendo uma inversão de matriz com dimensão igual ao número de terminais externos da SRNC, e (iv) formulação hierárquica das equações de estado e de sonda do circuito, combinando as equações de estado e de sonda das SuRs de fundo via um eficiente processo recursivo. A etapa (ii) envolve uma inversão de matriz para eliminação da parte linear de cada SRN. Vale ressaltar que, esta inversão possue um custo muito baixo, tendo em vista a sua dimensão. Além disso, um circuito em grande-escala muitas vezes é composto por vários dispositivos (SRNs), se não todos, idênticos. Excluindo a situação de representação de SRNs exclusivamente por DDS, a etapa (iii) também envolve uma inversão de matriz com dimensão igual ao número de terminais externos da SRNC. Considerando as etapas de (i) a (iii), devido à grande flexibilidade da metodologia proposta, podemos dizer que, a formulação das SuRs de fundo apresentada constitui uma extensão das formulações existentes, e resulta sempre em um sistema com o menor número possível de variáveis de estado não-lineares.

Conforme demonstrado, a equação de estado e de sonda do circuito assume uma estrutura multi-níveis tipo bloco diagonal com borda dupla e bloco diagonal com borda horizontal. Esta estrutura possibilita o uso de técnicas de processamento paralelo para resolução do problema. Também foi apresentada uma técnica de nivelamento de SuRs intermediárias, que possibilita o controle da estrutura hierárquica do circuito, além de uma maior eficiência na formulação da etapa (iii).

Finalmente, as operações com as matrizes envolvidas na formulação acima, podem ser eficientemente conduzidas com o auxílio de técnicas de matrizes esparsas [148],[146],[147],[27].

5. Análise do Balanço Harmônico

5.1. Introdução

FOI DESENVOLVIDA NO CAPÍTULO ANTERIOR, uma nova e eficiente metodologia para formulação multi-níveis das equações de circuitos em grande-escala, i.e., contendo um grande número de aglomerados de SRNs (dispositivos eletrônicos e opto-eletrônicos) hierarquicamente interconectados. Nesta formulação, cada aglomerado corresponde a uma SuR de fundo, que, por sua vez, podem ser interconectadas para formar SuRs intermediárias e, com isso, níveis de hierarquia. Neste capítulo, iremos utilizar estas equações para desenvolver uma nova metodologia para análise do BH multi-níveis.

Iniciaremos nossa discussão com a descrição, na Seção 5.2, das principais topologias de espectro de frequência para representação do sinal, utilizadas na determinação do regime permanente em circuitos não-lineares. Estas topologias de espectro dependem do tipo de excitação imposta ao circuito, sendo as mais comumente utilizadas: compressão e distorção com único-tom, dois-tons, três-tons e multi-tons, de compressão, distorção por intermodulação e conversão em frequência, recrescimento espectral. Na Seção 5.3 é apresentada uma breve discussão sobre a representação de sinais de RF digitalmente modulados, utilizando multisenos. Em seguida, na Seção 5.4, apresentaremos uma discussão na teoria e na implementação das transformadas de Fourier empregadas para a conversão do sinal de tempo-para-frequência e de frequência-para-tempo em regime de único- e multi-tons. A condução eficiente destas conversões consiste em um dos aspectos críticos da análise do BH, especialmente quando envolve mais de uma frequência fundamental não-harmonicamente relacionada. Para evitar transformadas de Fourier multi-dimensionais, caso multi-tons, a Seção 5.5 apresenta a técnica de MFA.

Utilizando as equações de estado e de sonda do circuito, obtidas no capítulo anterior, a Seção 5.6 descreve a formulação multi-níveis da equação de estado (ou determinante) e da equação de sonda para análise do BH. Nesta formulação, as equações são construídas seguindo uma orientação-por-SuR (estrutura multi-níveis, tipo bloco diagonal com borda) e depois uma orientação-por-frequência (estrutura bloco diagonal). Para finalizar, na Seção 5.7, será apresentada uma discussão sobre a determinação da matriz jacobiana associada com a equação determinante do BH multi-níveis. Conforme será demonstrado, a matriz jacobiana assume uma estrutura tipo bloco diagonal com dupla borda [41], e com isto pode ser eficientemente processada em sistemas distribuídos de computadores [27],[109]. Para a determinação dos blocos que compõem a matriz

jacobiana é necessário determinar as matrizes jacobianas das SuRs de fundo associadas com o vetor de função não-linear. Serão apresentadas expressões analíticas para o cálculo dos elementos destas matrizes jacobianas, para os regimes de único-tom e multi-tons. Para uma eficiente formação e fatorização da matriz jacobiana do BH, utilizando técnicas de matrizes esparsas, será introduzido um controle de esparsidade fundamentado no conceito de espectro de frequência de derivada. As observações finais são reservadas para a Seção 5.8.

5.2. Espectro de Frequência

Antes de apresentarmos a formulação da equação determinante do BH, é conviniente discutirmos os tipos mais comuns de topologia de espectro de frequência utilizados nas análises do BH com excitação de único-tom, dois-tons, três-tons e multi-tons. Assumimos que o regime transiente foi extinto, e que o circuito não-linear sob análise é estável e alimentado por "NT" fontes senoidais independentes, com frequências angulares $\omega_{f_1}, \omega_{f_2}, \ldots, \omega_{f_{NT}}$. Então, em regime de estado permanente, introduzindo a série de Fourier, as formas-de-onda das variáveis de estado não-lineares (tensões e correntes) do circuito podem ser expressas, de forma geral, como

$$x_p(t) = \mathbf{Re}\left[\sum_{\omega_{\boldsymbol{k}} \in \mathbb{S}_\infty} X_p(\omega_{\boldsymbol{k}}) \mathbf{e}^{\hat{\mathbf{j}}\omega_{\boldsymbol{k}} t} \right] \qquad p = 1, 2, \ldots, n_{NSV}, \tag{5.1}$$

onde

$$\omega_{\boldsymbol{k}} = \sum_{i=1}^{NT} k_i \omega_{f_i}, \; \forall k_i \in \mathbb{Z} \tag{5.2}$$

são as frequências de intermodulação e $\mathbb{S}_\infty$ é o *espectro de frequência do sinal* gerado pela lei de intermodulação (5.2). Acima foi introduzida a seguinte notação: $\omega_{\boldsymbol{k}} = \omega_{k_1, k_2, \ldots, k_{NT}}$, onde $\boldsymbol{k} = k_1, k_2, \ldots, k_{NT}$. Por razões práticas, este espectro deve ser truncado, porém sem sacrificar a precisão da solução do problema. Desta forma, o espectro truncado é dado por $\mathbb{S} = \{\omega_0, \omega_1, \ldots, \omega_{NF-1}\} \subset \mathbb{S}_\infty$, onde NF é o número de frequências presente em $\mathbb{S}$. Vale ressaltar que, ω_0 é igual a zero, correspondendo à componente de CC. Com um número finito de linhas espectrais, o problema de determinação dos fasores complexos, $X_p(\omega_{\boldsymbol{k}})$, em (5.1), torna-se numericamente praticável. O esquema de truncamento em frequência depende da análise a ser desempenhada conforme será discutido a seguir.

Para o caso da análise com uma excitação de único-tom, a expressão (5.2) assume a seguinte forma

$$\omega_k = k\omega_f, \; \forall k \in \mathbb{Z}, \tag{5.3}$$

Nesta situação monocromática, o espectro de frequência do sinal é dado por:

$$\mathbf{S} = \begin{cases} \omega_k = k\omega_f \\ onde \\ \quad |k| \le NH \end{cases} \tag{5.4}$$

onde NH é o número de harmônicos ($NF = NH + 1$) utilizados. Quando apenas uma única frequência angular fundamental, ω_f, está presente na análise do BH, todas as frequências (harmônicos) produzidas pelas não-linearidades do circuito (estável) estarão harmonicamente relacionadas com a fundamental, e toda forma-de-onda será uma função periódica no tempo, com período igual a $T = 2\pi/\omega_f$.

Se mais de uma frequência fundamental incomensurável, i.e., não-harmonicamente relacionada, estiver presente no circuito, o espectro de frequência do sinal assume uma forma mais complexa, quando comparado com (5.4). Neste caso, as frequências positivas que precisam ser consideradas na análise, são dadas por (5.2), i.e., $\omega_j = \left|k_{1,j}\omega_{f,1} + k_{2,j}\omega_{f,2} + \ldots + k_{NT,j}\omega_{f,NT}\right|$, onde $j \in [0, NF-1]$. Para geração destas frequências (produtos de intermodução) distintas, o primeiro $k_{i,j}$ não-zero deve ser maior que zero. Os valores máximos de $|k_{i,j}|$ irão depender da força das não-linearidades do circuito. No método do BH, dependendo do mecanismo de operação do circuito (amplificador, oscilador, multiplicador de frequência, conversor de frequência, divisor de frequência, etc) e do tipo de teste (distorção harmônica, distorção por intermodulação, conversão de frequência, recrescimento espectral, etc) a ser simulado, um determinado esquema de truncamento em frequência deve ser adotado para minimizar o custo de memória e o tempo de processamento. A Tabela 5.1 descreve as topologias do espectro de frequência comumente utilizadas na análise do BH, para circuitos não-autônomos operando em regime de dois- e três-tons. Nas análises de intermodulação (IM) de dois- e três-tons a grade de frequência assume a forma diamante ou pirâmide, repectivamente. Em adição, o número de frequências envolvidas nas análises de distorção por IM de dois-tons e de três-tons são iguais a $M_{IM2}(M_{IM2}-1)+1$ e $(2/3)M_{IM3}(M_{IM3}(M_{IM3}-3/2)+7/2)-1$, respectivamente. O valor do parâmetro de máxima ordem, M_{IM2} e M_{IM3}, dependerá da força das não-linearidades presentes no circuito e da amplitude das fontes de alimentação. A análise do BH de dois-tons para um circuito excitado por um sinal de grande amplitude, referente à portadora de OL (tom 1) representado por M_{LO}, e por um sinal de amplitude incremental, referente à modulação (tom 2) representado por $M_{SB} = 1$, produz resultados equivalentes à análise de conversão de frequência [153]. Nesta análise, o número de frequências (ou linhas espectrais) envolvidas, é igual a $3M_{OL}+2$. Lembramos que, na análise de conversão em frequência, o circuito é linearizado no entorno do regime permanente imposto pelo

Tabela 5.1
TOPOLOGIAS DO ESPECTRO DE FREQUÊNCIA DO SINAL PARA DIFERENTES TIPOS DE ANÁLISE DE EQUILÍBRIO DE HARMÔNICO

Análise de Intermodulação	
dois-tons	três-tons
$\mathbf{S}_{IM2} = \begin{cases} \lvert k_1\rvert + \lvert k_2\rvert \le M_{IM2} \\ \text{onde:} \\ \quad \lvert k_1\rvert \le M_{IM2} = NH_0 \\ \quad \lvert k_2\rvert \le M_{IM2} = NH_1 \end{cases}$	$\mathbf{S}_{IM3} = \begin{cases} \lvert k_1\rvert + \lvert k_2\rvert + \lvert k_3\rvert \le M_{IM3} \\ \text{onde:} \\ \quad \lvert k_1\rvert \le M_{IM3} = NH_0 \\ \quad \lvert k_2\rvert \le M_{IM3} = NH_1 \\ \quad \lvert k_3\rvert \le M_{IM3} = NH_2 \end{cases}$
Análise de Conversão em frequência	Análise de Conversão em frequência com Intermodulação
dois-tons	três-tons
$\mathbf{S}_{CF} = \begin{cases} \lvert k_1 + k_2\rvert \le M_{OL} \\ \text{onde:} \\ \quad \lvert k_1\rvert \le NH_{OL} = NH_0 \\ \quad \lvert k_2\rvert \le NH_{BL} = NH_1 \end{cases}$	$\mathbf{S}_{CF-IM} = \begin{cases} \lvert k_1 + k_2 + k_3\rvert \le M_{OL} \\ \lvert k_2\rvert + \lvert k_3\rvert \le M_{IM2-BL} \\ \text{onde:} \\ \quad \lvert k_1\rvert \le N_{OL} = NH_0 \\ \quad \lvert k_2\rvert \le M_{IM2-BL} = NH_1 \\ \quad \lvert k_3\rvert \le M_{IM2-BL} = NH_2 \end{cases}$

sinal de OL, e o sinal modulante é considerado como uma perturbação incremental (pequeno sinal).

Na Fig. 5.1, podemos observar as topologias do espectro de frequência para determinação da característica de recrescimento espectral nas respostas multi-tons de amplificadores e multiplicadores em frequência, ver Fig. 5.1(a), e conversores em frequência, ver Fig. 5.1(b). Conforme será discutido abaixo, estes espectros de frequência são particularmente úteis na representação de sinais de RF digitalmente modulados. Em adição, discutiremos a conexão destas topologias com eficientes implementações numéricas da transformada de Fourier discreta (TFD) para sinais de natureza multi-tons.

5.3. Excitação Digital e Multi-Senos

Uma excitação representando um sinal digital pode ser implementada numericamente utilizando uma sequência pseudo-aleatória de símbolos. Por sua vez, esta sequência de símbolos pode ser representada por multi-senos [154],[155], que consistem de uma coleção de ondas

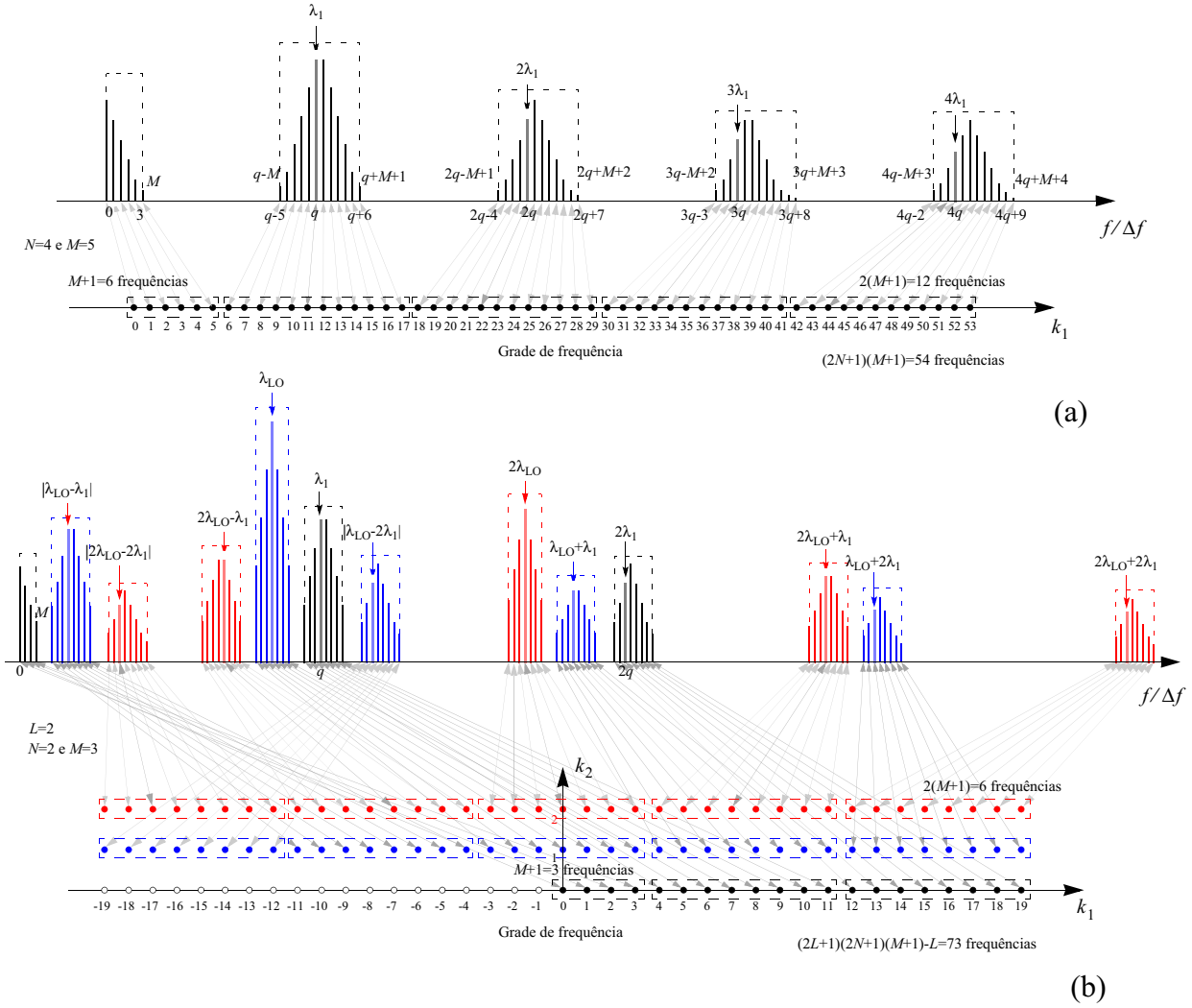


Fig. 5.1 Topologia do espectro de frequência e grade de frequência para: (a) transformada de Fourier multi-tons (TFMT) e (b) para a transformada de Fourier multi-tons em duas dimensões (TFMT-2D).

senoidais geradas simultaneamente. Tipicamente, estas ondas senoidais possuem uma frequência constante e com um espaçamento em frequência Δf. Em [154], foi apresentado um estudo comparativo entre diferentes tipos de exitações multi-senos para determinação da ACPR. Neste estudo, foram considerados multi-senos com magnitude e fase constantes, magnitude e fase aleatórias, magnitude constante e fase de *Schroeder*, e sinal digital representado por uma sequência pseudo-aleatória de 32 símbolos com modulação digital QPSK e formatação de pulso via cosseno levantado. Conforme discutido em [154], diferentes tipos de excitação multi-senos produzirão diferentes resultados de ACPR, principalmente entre excitações com razão de potência entre pico e média (tradução nossa para *peak-to-avarage power ratio)* PAPR bem distintas. As formas-de-onda e a associada densidade espectral de potência, do inglês *power spectral density* (PSD), para sinais GSM e WCDMA 2000, podem ser visualizadas em [106].

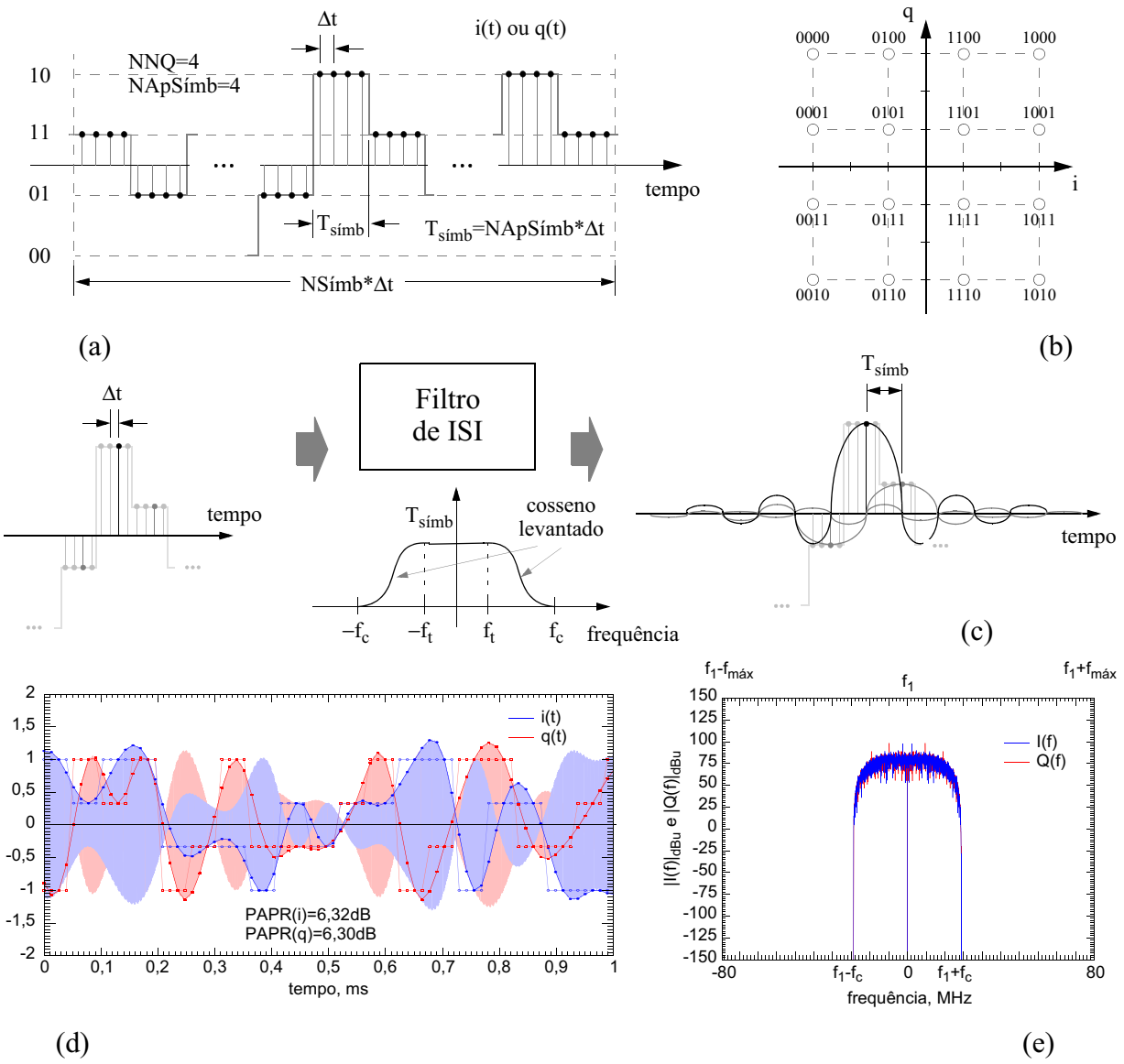


Fig. 5.2 (a) Forma-de-onda de um de um sinal de digital representado por uma sequência pseudo-aleatória de NSímb símbolos com 4 níveis de quantização. (b) Constelação associada à modulação 16 QAM. (c) Passagem de um sinal digital por um filtro de ISI. (d) Forma-de-onda dos sinais $i(t)$ e $q(t)$ obtidos numericamente com formatação de pulso cosseno levantado. (e) Espectro de frequência $I(f)$ e $Q(f)$ dos sinais $i(t)$ e $q(t)$, respectivamente, obtidos via TFMT.

Para ilustrar a implementação de uma excitação digital, vamos considerar a Fig. 5.2(a), onde podemos observar uma sequência pseudo-aleatória de símbolos com quatro níveis de quantização (i.e., $NNQ = 4$) de amplitude. Este tipo de sequência pode ser utilizada para representar as componentes em-fase (i) e em quadratura (q), do sistema descrito pela constelação 16 QAM da Fig. 5.2(b). Mais precisamente, para possibilitar a representação na análise do BH, vamos considerar que sequência de símbolos consiste de uma série periódica de $NSímb$ símbolos. Cada símbolo possue $NApSímb$ amostas no tempo, sendo Δt o intervalo entre cada amostragem. Convém ressaltar

que, o aumento em $NApSímb$ produz um aumento em $f_{máx} = 1/\Delta t$, possibilitando um maior espaço para a determinação do recrescimento espectral. A taxa de transmissão de símbolos é dada por $f_{símb} = 1/T_{símb}$, onde $T_{símb} = NApSímb \cdot \Delta t$.

O problema de interferência entre símbolos, do inglês *inter-symbol interference* (ISI), é particularmente importante em sistemas de comunicação sem fio, devido à faixa estreita de frequência permitida para cada canal. Para mitigar o efeito de ISI, pode ser utilizada a formatação de pulso no transmissor e a equalização no receptor. Para formatação de pulso uma das técnicas mais utilizadas consiste na utilização de um filtro de ISI com resposta em frequência do tipo pulso cosseno levantado, conforme demonstrado na Fig. 5.2(c). A função que define o pulso cosseno levantado no domínio da frequência é dada por [66]:

$$H(f) = \begin{cases} T_{símb}, \ 0 < |f| < f_t = (1-\alpha) f_{símb}/2 \\ (1/2) T_{símb}(1 + \cos(\pi T_{símb}(|f| - f_t)/\alpha)), \ f_t < |f| < f_c = (1+\alpha) f_{símb}/2 \\ 0, \ f_c < |f| \end{cases}. \tag{5.5}$$

onde $0 < \alpha < 1$ é fator de decaimento e f_t e f_c são as frequências de transição e de corte, respectivamente. Na prática, são utilizados filtros cosseno levantado decompostos em dois filtros cuja resposta em frequência é definida pela raiz quadrada de (5.5). Neste caso, temos um filtro no transmissor e o outro no receptor.

As formas-de-onda obtidas numericamente das componentes em fase, $i(t)$, e em quadratura, $q(t)$, de um sinal digital com modulação 16 QAM, utilizando filtro de ISI cosseno levantado com Δt e $T_{símb}$ iguais a 12,5 μs e 50 μs, respectivamente, podem ser visualizadas na Fig. 5.2(d). Estes sinais são do tipo sem retorno-a-zero (NRZ - *non-return-to-zero*) e possuem uma PAPR no entorno de 6 dB. Os espectros de frequência destes sinais podem ser observados na Fig. 5.2(e), onde f_c, $f_{símb}$ e $f_{máx}$ são iguais a 13,5 kHz para $\alpha = 0,35$, 20 ksímbs/s (ou seja, 20 kHz) e 80 kHz, respectivamente. A excitação digital 16 QAM, descrita acima, será utilizada no Capítulo 8, para validação numérica do recrescimento espectral em amplificadores e conversores em frequência.

No ambiente de *software* de simulação desenvolvido, outros tipos de modulação também foram implementadas, e.g., π/4-QPSK, etc [66]. A seguir será apresentado uma TFD que possibilita uma eficiente implementação numérica das conversões de tempo-frequência para sinais de RF digitalmente modulados.

5.4. Transformada de Fourier Discreta

A condução eficiente das conversões do sinal de tempo-para-frequência e de frequência-para-

tempo é um ponto crucial na análise do BH. A razão disto deve-se à necessidade do cálculo dos fasores complexos das funções não-lineares e suas derivadas no domínio do tempo. Desta forma, o método do BH pode ser visto como um método de domínio misto. Lembramos que, devido ao truncamento em frequência e amostragem do tempo, estas conversões estão sujeitas aos erros de "aliasing" [64].

Para um circuito operando em regime de único-tom, as conversões do sinal de tempo-para-frequência e de frequência-para-tempo, podem ser eficientemente conduzidas via [64]

$$X(k) = \sum_{r=0}^{NS-1} x(r) W^{kr} \tag{5.6.a}$$

e

$$x(r) = \sum_{k=-NH}^{NH} X(k) W^{-kr}, \tag{5.6.b}$$

respectivamente, onde $W = \mathbf{e}^{-\hat{\mathbf{j}}2\pi/NS}$. Lembremos que estas transformadas operam no plano de fase. Para que (5.6.b) represente uma forma-de-onda real, devemos ter $X(k) = X(-k)^*$. As somatórias acima podem ser eficientemente calculadas via transformada de Fourier rápida (TFR) e sua inversa, respectivamente. A complexidade da TFR é igual a $O(NS \cdot \log(NS))$, que é bem menor do que a complexidade ($O(NS^2)$), requerida pela computação direta da somatória (5.6.a). Para evitar "aliasing" na última componente espectral produzida pelo algoritmo TFR, devemos selecionar $NS = 2NH + 1$ (NS ímpar).

Em um circuito operando em regime multi-tons, a forma mais precisa de conduzir as conversões dos sinais de tempo-para-frequência e de frequência-para-tempo é via transformada de Fourier discreta multi-dimensional (TFDM) e sua inversa, repectivamente [55],[64],[156]. Sua superior precisão, resulta do fato desta transformada operar no plano de fase e, com isto, produzir uma amostragem ótima. A TFDM e sua inversa são generalizações das transformadas uni-dimensionais (5.6.a) e (5.6.b), dadas por:

$$X(k_1, k_2, \ldots, k_{NT}) = \sum_{r_{NT}=0}^{NS_{NT}-1} \left\{ \ldots \left[\sum_{r_2=0}^{NS_2-1} \left(\sum_{r_1=0}^{NS_1-1} x(r_1, r_2, \ldots, r_{NT}) W_1^{k_1 r_1} \right) W_2^{k_2 r_2} \right] \ldots \right\} W_{NT}^{k_{NT} r_{NT}} \tag{5.7.a}$$

e

$$x(r_1, r_2, \ldots, r_{NT}) = \sum_{k_{NT}=-NH_{NT}}^{NH_{NT}} \left\{ \ldots \left[\sum_{k_2=-NH_2}^{NH_2} \left(\sum_{k_1=-NH_1}^{NH_1} X(k_1, k_2, \ldots, k_{NT}) W_1^{-k_1 r_1} \right) W_2^{-k_2 r_2} \right] \ldots \right\} W_{NT}^{-k_{NT} r_{NT}} \tag{5.7.b}$$

respectivamente, onde $W_j = \mathbf{e}^{-\hat{\mathbf{j}}2\pi/NS_j}$ para $j = 1, \ldots, NT$. Analogamente, à transformada uni-dimensional (5.6.b), para que (5.7.b) represente uma forma-de-onda real devemos ter $X(k_1, k_2, \ldots, k_{NT}) = X(-k_1, -k_2, \ldots, -k_{NT})^*$. As somas acima podem ser eficientemente calculadas

utilizando a transformada de Fourier rápida multi-dimensional (TFRM). Esta transformada possue uma complexidade igual a $O(NS_1 \cdot NS_2 \cdot \ldots \cdot NS_{NT} \cdot \log(NS_1 + NS_2 + \ldots + NS_{NT}))$. Para evitar "aliasing" na última componente espectral produzida pelo algoritmo TFRM, devemos selecionar $NS_j = 2NH_j + 1$, para $j = 1, \ldots, NT$. O número de frequências envolvidas na TFRM (grade quadrada) é igual a $(1/2)(NF_1 \cdot NF_2 \cdot \ldots \cdot NF_{NT} + 1)$, onde $NF_j = 2NH_j + 1$. Em geral, o número de frequências envolvidas nas topologias de espectro multi-tons, sumarizadas na Tabela 5.1, é inferior ao número de frequências presentes na TFRM. Neste caso, para aplicação da transformada de Fourier em uma variável de estado, os fasores complexos correspondente as frequências ausentes na análise do BH são considerados igual a zero. Esta operação corresponde a uma interpolação no domínio do tempo, da forma-de-onda associada à variável de estado [156]. Convém ressaltar que, no estágio de pós-processamento, os resultados de forma-de-onda não podem ser obtidos da TFRM inversa; isto se deve ao fato da decomposição artificial da variável tempo em múltiplas (duas ou três) dimensões. Não obstante isso, as formas-de-onda podem ser facilmente determinadas via série complexa de Fourier [64].

Conforme discutido acima, a grade de frequência utilizada na TFRM é retangular, independente da análise do BH a ser conduzida, e pode incluir frequências não definidas nas topologias do espectro de frequência definidas na Tabela 5.1. Para eliminar esta dificuldade, foi proposta a *transformada de Fourier quasi-periódica* (TFQP) [56],[57] para regime multi-tons. Na TFQP, a grade de frequência é definida pelas linhas espectrais consideradas na análise, e sua implementação númerica é bem mais simples. A conversão de frequência-para-tempo, i.e., a inversa da TFQP, é naturalmente definida por

$$\boldsymbol{\Gamma}_I = \begin{bmatrix} 1 & 2\cos(\omega_1 t_0) & -2\sin(\omega_1 t_0) & 2\cos(\omega_2 t_0) & \ldots & -2\sin(\omega_{NF-1} t_0) \\ 1 & 2\cos(\omega_1 t_1) & -2\sin(\omega_1 t_1) & 2\cos(\omega_2 t_1) & \ldots & -2\sin(\omega_{NF-1} t_1) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 2\cos(\omega_1 t_{NS-1}) & -2\sin(\omega_1 t_{NS-1}) & 2\cos(\omega_2 t_{NS-1}) & \ldots & -2\sin(\omega_{NF-1} t_{NS-1}) \end{bmatrix} . \tag{5.8}$$

O resultado acima decorre diretamente da série de Fourier [64]. Utilizando (5.8), a conversão de frequência-para-tempo, pode ser escrita de forma mais compacta, viz.

$$\boldsymbol{x} = \boldsymbol{\Gamma}\bar{\boldsymbol{X}}, \tag{5.9}$$

onde $\boldsymbol{x}$ e $\bar{\boldsymbol{X}}$ são os vetores de forma-de-onda e de espectro de frequência, respectivamente. Infelizmente, a TFQP, em geral, não é bem condicionada. Para $NS = 2NF - 1$ (caso não-sobre-amostrado), o critério mais simples de amostragem é a utilização de amostras no tempo igualmente espaçadas. Porém, isto resulta em $\boldsymbol{\Gamma}$ mal condicionada numericamente e, como consequência, uma

imprecisa TFQP, que corresponde à inversa de $\boldsymbol{\Gamma}$. Um dos remédios para amenizar este problema de condicionamento consiste na utilização de sobre-amostragem, com as amostras no tempo geradas aleatoriamente [56],[57]. Sendo assim, a conversão de tempo-para-frequência é dada por

$$\bar{\boldsymbol{X}} = \boldsymbol{\Gamma}^{+}\boldsymbol{x} \tag{5.10}$$

onde

$$\boldsymbol{\Gamma}^{+} = \begin{cases} \boldsymbol{\Gamma}^{-1}, & NS = 2NF-1 \\ (\boldsymbol{\Gamma}^{T}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^{T}, & NS > 2NF-1 \end{cases}. \tag{5.11}$$

Na equação (5.11) para o caso sobre-amostrado, podemos observar que a transformada é calculada utilizando a fórmula generalizada de Penrose [157]. A aplicação da TFQP e da sua inversa tem uma complexidade $O(NS\cdot(2NF-1))$. Infelizmente, ao contrário da TFD, que opera no plano de fase, a estabilidade e precisão da TFQP depedende do esquema de amostragem do tempo adotado. Sendo assim, se a amostragem do tempo não for ótima, a transformada (5.11), $\boldsymbol{\Gamma}^{+}$, pode se tornar numericamente mal condicionada. O número de condição de $\boldsymbol{\Gamma}$ é especificado como [157]

$$\kappa_p(\boldsymbol{\Gamma}) = \|\boldsymbol{\Gamma}\|_p \|\boldsymbol{\Gamma}^{+}\|_p \,, \tag{5.12}$$

tendo como objetivo determinar a sensitividade de (5.10) para pequenas mudanças em $\boldsymbol{\Gamma}$ ou em $\boldsymbol{x}$. Se $\bar{\boldsymbol{X}}+\delta\bar{\boldsymbol{X}}$ satisfaz (5.10) com $\boldsymbol{x}+\delta\boldsymbol{x}$, temos que [157]

$$\frac{\|\delta\bar{\boldsymbol{X}}\|_p}{\|\bar{\boldsymbol{X}}+\delta\bar{\boldsymbol{X}}\|_p} \le \kappa_p(\boldsymbol{\Gamma})\frac{\|\delta\boldsymbol{\Gamma}^{+}\|_p}{\|\boldsymbol{\Gamma}^{+}\|_p}. \tag{5.13.a}$$

Similarmente, se $\bar{\boldsymbol{X}}+\delta\bar{\boldsymbol{X}}$ satisfaz (5.10), com $\boldsymbol{\Gamma}+\delta\boldsymbol{\Gamma}$, temos que [157]

$$\frac{\|\delta\bar{\boldsymbol{X}}\|_p}{\|\bar{\boldsymbol{X}}\|_p} \le \kappa_p(\boldsymbol{\Gamma})\frac{\|\delta\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p}. \tag{5.13.b}$$

Na prática, as normas mais comuns são: l_1-norm ($p=1$), l_2-norm ($p=2$) e l_∞-norm ($p=\infty$) [157]. Seguindo [158], podemos definir o fator de estabilidade como se segue

$$\beta(\boldsymbol{\Gamma}) = \begin{cases} \log_{10}(\kappa(\boldsymbol{\Gamma})), & NS = 2NF-1 \\ (\log_{10}(\kappa(\boldsymbol{\Gamma})/(2NF-1))), & NS > 2NF-1 \end{cases}. \tag{5.14}$$

Quando $\beta(\boldsymbol{\Gamma})$ se aproxima de zero, o desempenho da TFQP se aproxima ao limite ótimo imposto pela TFDM. Uma outra figura de mérito que estima a precisão da conversão (5.11) é definida como

$$\varepsilon(\boldsymbol{\Gamma}) = \log_{10}\left\|\mathbf{1}-\boldsymbol{\Gamma}^{+}\boldsymbol{\Gamma}\right\|_\infty. \tag{5.15}$$

Além das TFQPs discutidas acima, podemos destacar os seguintes tipos: com amostragem ortogonal-próxima [59], ortogonalizada via mínimos-quadrados [159], quasi-ortogonal [158], quase TFD-equivalente [58] e a ortogonal [60]. Os resultados apresentados em [159], são

questionados em [158], devido ao fato deles violarem o limite ótimo imposto pela TFDM. O método quasi-ortogonal proposto em [158], possue difíceis parâmetros de sintonização, como é observado em [58]. A transformada proposta em [58], está limitada a operação com sinais de dois-tons. Finalmente, a TFQP ortogonal, proposta em [60], explora o erro de truncamento numérico para poder satisfazer a ortogonalidade da transformada.

Para verificar a precisão das TFQPs, discutidas acima, dois testes simples foram conduzidos. Para este próposito, vamos considerar o simples e hipotético sistema não-linear dado por [159]

$$y(t) = \sum_{i=0}^{\mu} a_i \cdot (x(t))^i \text{ com } a_i = 10^{-i}, \tag{5.16}$$

onde o sinal de entrada é composto por duas frequências fundamentais, e dado por

$$x(t) = \cos(\lambda_1 t) + \cos(\lambda_2 t) \ . \tag{5.17}$$

O primeiro experimento foi conduzido transformando a forma-de-onda (5.16) para o domínio da frequência. Os resultados obtidos com a TFRM e as TFQPs uniformemente amostrada, aleatoriamente amostrada e ortogonal-próxima, são sumarizadas nas Figs. 5.3 (a) e (b). Para a TFQPs aleatoriamente amostrada e ortogonal-próxima, o fator de sobre-amostragem é igual a 2. Para uma ordem de intermodulação, M_{IM}, igual 5, todas as transformadas desempenham bem. Entretanto, para $M_{IM} = 10$ a TFQP igualmente amostrada e a ortogonal-próxima não produzem resultados aceitáveis, indicando a pobre estratégia de amostragem destas transformadas. Estes resultados confirmam os testes de condicionamento numérico ilustrados na Fig. 5.4.. O segundo teste foi conduzido transformando o sinal no domínio da frequência, definido em (5.17), para o domínio do tempo e, então, transformando de volta para o domínio da frequência. Os resultados deste teste são ilustrados nas Figs. 5.3(c) e (d) e, como esperado, conduz às mesmas conclusões do primeiro teste. Em resumo, podemos obsevar o comportamento errático da TFQP com amostragem uniforme e ortogonal-próxima.

Conforme demonstrado em [65], obedecendo uma condição de amostragem não restritiva, utilizando a topologia do espectro de frequência da Fig. 5.1(a), as conversões de tempo-frequência dos sinais podem ser conduzidas em uma dimensão, via TFR. Em [65], a transformada que realiza estas conversões foi entitulada transformada de Fourier multi-tons (TFMT). Este tipo de transformada, pode ser extendido para a análise de conversão em frequência em regime multi-tons, se considerarmos a topologia do espectro de frequência da Fig. 5.1(b). Neste caso, entitulamos esta transformada de *transformada de Fourier multi-tons em duas dimensões* (TFMT-2D), pois utiliza a TFR em duas-dimensões para a realização das conversões dos sinais. A grade de frequência para

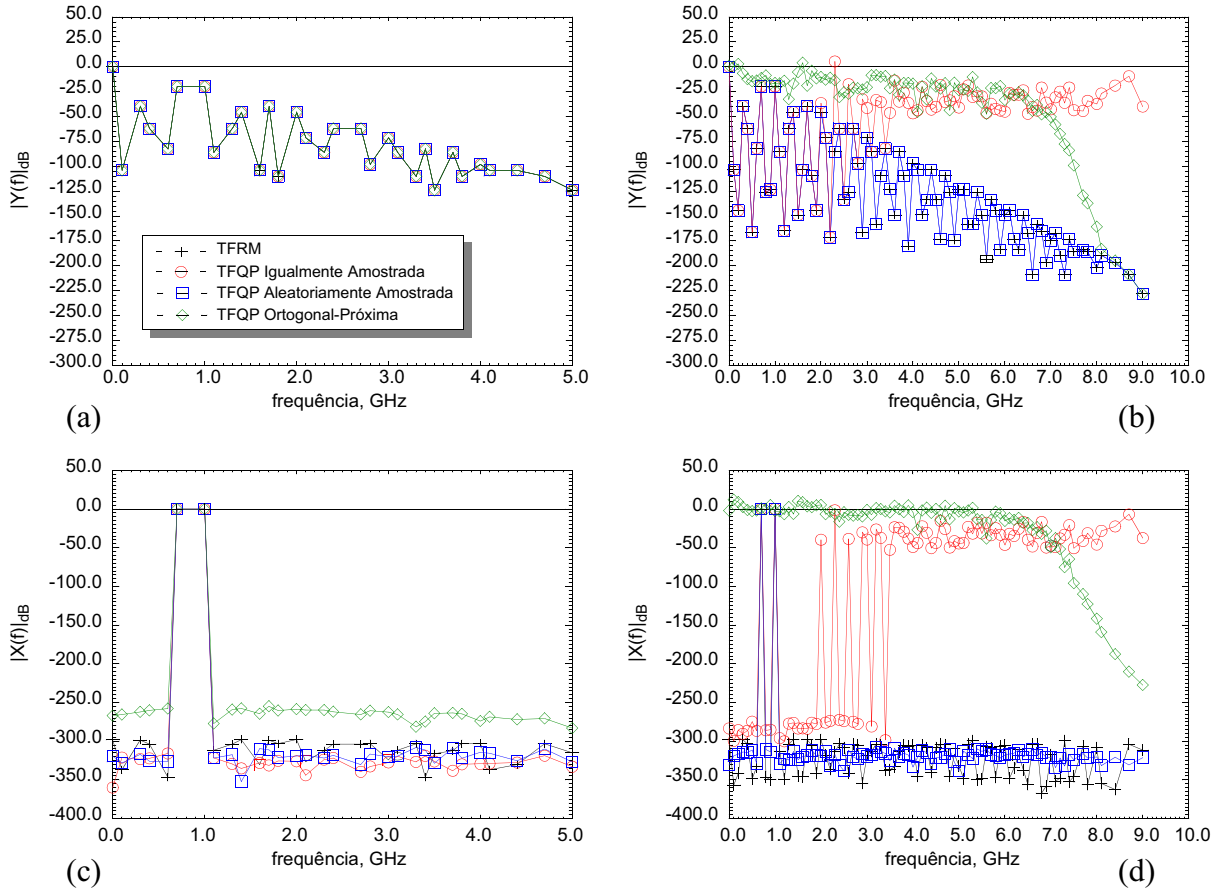


Fig. 5.3 Teste de precisão das transformadas de Fourier discretas para sinais quasi-periódicos. As frequências fundamentais são $\lambda_1 = 2\pi$ GHz e $\lambda_2 = 2\pi \cdot 0{,}7$ GHz. (a) e (c) $M_{IM2} = 5$. (b) e (d) $M_{IM2} = 10$. O fator de sobre-amostragem, m, é igual a 2.

a TFMT-2D é indicada na Fig. 5.1(b). Com estas transformadas, a análise do BH pode ser extendida para sinais de RF digitalmente modulados.

5.5. Mapeamento-de-Frequência Artificial

Assim como a TFMT descrita acima, a técnica de mapeamento-de-frequência articial (MFA), proposta em [62], possibilita que as conversões de tempo-frequência dos sinais em um circuito operando em regime multi-tons, possam ser conduzidas via TFR uni-dimensional. A escolha do mapeamento adequado resulta no conhecido problema de *diophatine* [9]. Nas Fig. 5.5(a) e (b), podemos observar a aplicação da técnica de MFA para um espectro de frequência de dois-tons com grade no formato de diamante e retangular, respectivamente. Para um espectro de frequência definido por uma grade truncada no formato retangular, o mapeamento pode ser generalizado para o caso multi-dimensional [63]. Neste caso, a transformada de Fourier via MAF é equivalente à

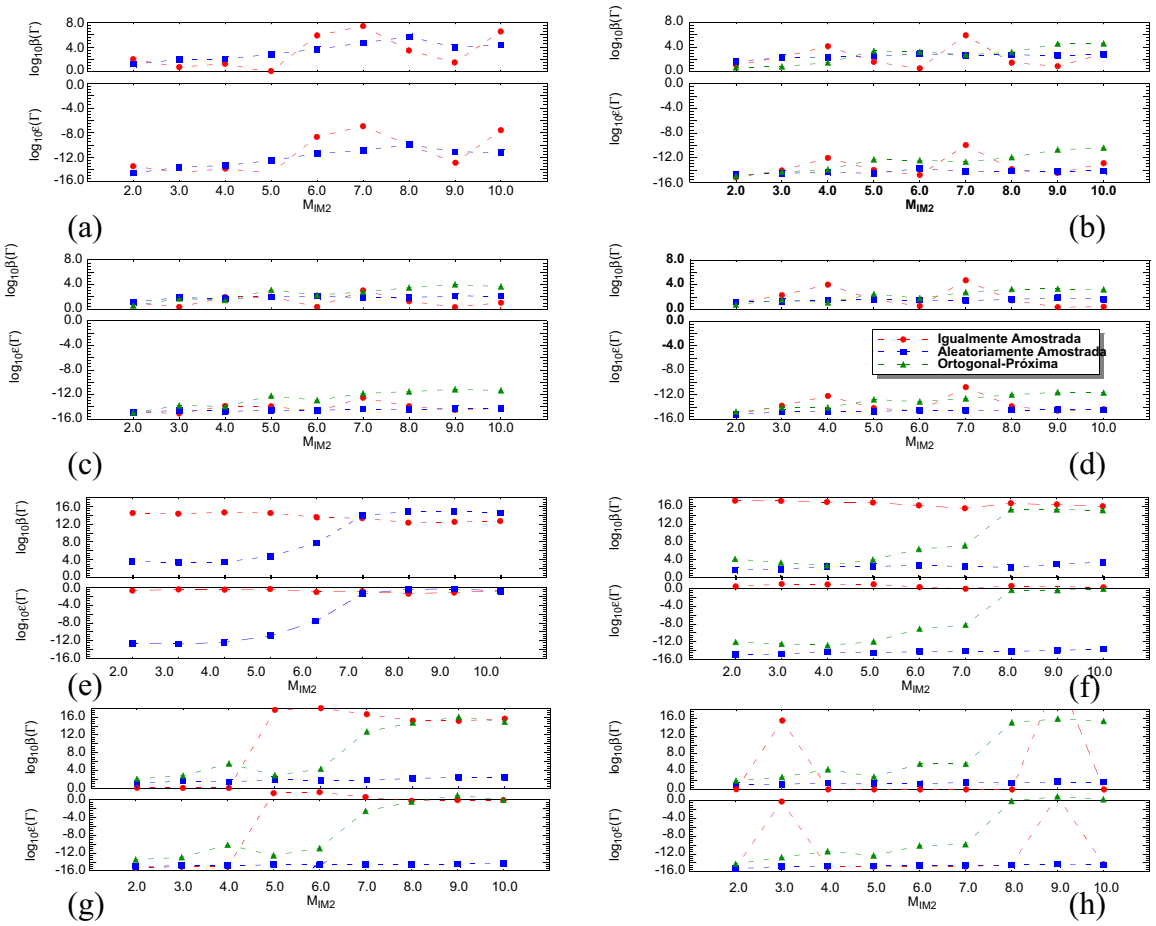


Fig. 5.4 As frequências fundamentais são $\lambda_1 = 2\pi$ GHz e $\lambda_2 = \lambda_1 + 2\pi\sqrt{2}$ Hz. (a) $m = 1$, (b) $m = 2$, (c) $m = 3$, e (d) $m = 4$. As frequências fundamental são $\lambda_1 = 2\pi$ GHz e $\lambda_2 = 2\pi \cdot 0{,}7$ GHz. (e) $m = 1$, (f) $m = 2$, (g) $m = 3$, e (h) $m = 4$. m é o fator de sobre-amostragem.

TFRM. Convém ressaltar que, o ordenamento incorreto das frequências fundamentais para geração dos produtos de IM, ver (5.2), pode conduzir ao problema de *aliasing*. Recentemente, a técnica de MAF foi sugerida para simulação de sistemas digitais com multi-portadoras de RF [82].

5.6. Equação Determinante

Utilizando os resultados dos capítulos anteriores, vamos, nesta seção, obter a equação de estado (ou determinante) para análise do BH. Iniciando com a organização por frequência, e seguindo a clássica notação introduzida em [45], o vetor de variável de estado não-linear da k-ésima SuR de fundo no nível υ da hierarquia, $\mathfrak{s}_{\upsilon,k}$, pode ser organizado por frequência da seguinte forma

$$\bar{\boldsymbol{X}}^{(\upsilon,k)} = \left[\boldsymbol{X}^{(\upsilon,k)}(0)\ \boldsymbol{X}^{(\upsilon,k),re}(1)\ \boldsymbol{X}^{(\upsilon,k),im}(1)\ \boldsymbol{X}^{(\upsilon,k),re}(2)\ \ldots\ \boldsymbol{X}^{(\upsilon,k),im}(NF-1)\right]^T, \tag{5.18}$$

$$\boldsymbol{X}^{(\upsilon,k)}(0) = \left[X_1^{(\upsilon,k)}(0)\ X_2^{(\upsilon,k)}(0)\ \ldots\ X_{n_{VEN}^{(\upsilon,t)}}^{(\upsilon,k)}(0)\right]^T,$$

$$\boldsymbol{X}^{(\upsilon,k),\alpha}(k) = \left[X_1^{(\upsilon,k),\alpha}(k)\ X_2^{(\upsilon,k),\alpha}(k)\ \ldots\ X_{n_{VEN}^{(\upsilon,t)}}^{(\upsilon,k),\alpha}(k)\right]^T \qquad k = 1, 2, \ldots, NF-1 \qquad \alpha = re, im,$$

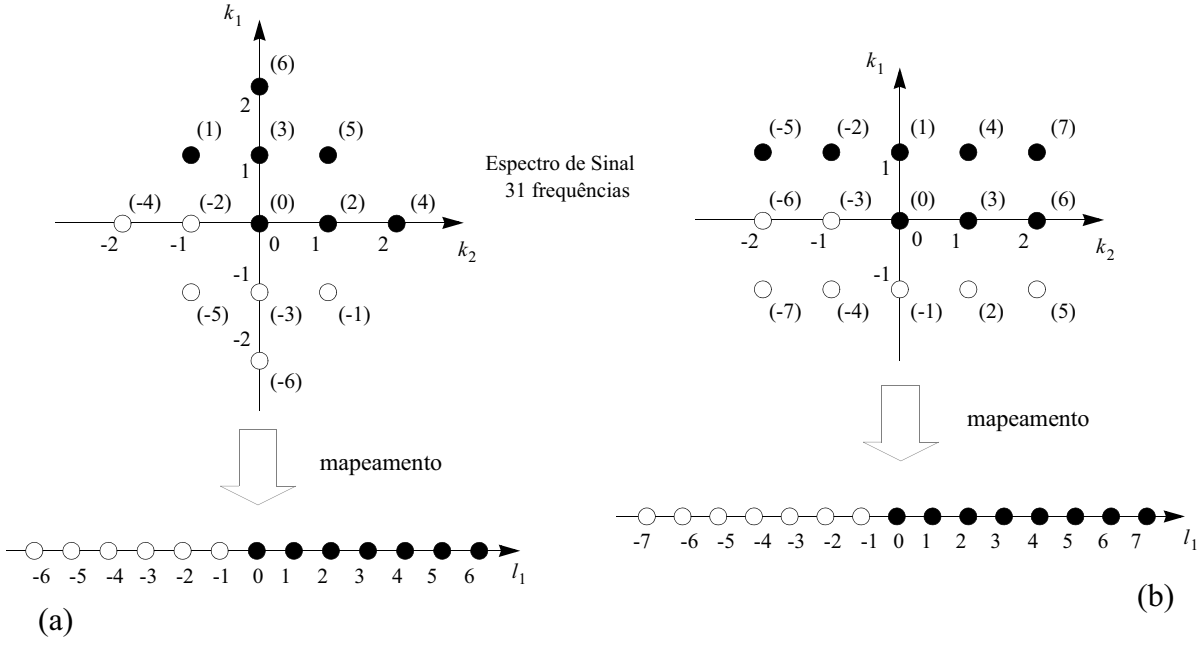


Fig. 5.5 Exemplo da aplicação da técnica de mapeamento artificial em frequência para espectro de frequência de dois-tons: (a) grade triangular e (b) grade retangular.

onde $n_{VEN}^{(\upsilon,\,k)}$ é o número de variáveis de estado não-lineares. Os vetores $\overline{\boldsymbol{U}}_f^{(\upsilon,\,k)}$ e $\overline{\boldsymbol{U}}_s^{(\upsilon,\,k)}$ estão associados com as funções não-lineares e as fontes independentes, respectivamente, e são organizados da mesma forma descrita em (5.18). Seguindo esta notação, a matriz constitutiva $\overline{\boldsymbol{A}}^{(\upsilon,\,t)}$ assume a forma

$$\overline{\boldsymbol{A}}^{(\upsilon,\,t)} = \mathbf{diag}\left[\boldsymbol{A}^{(\upsilon,\,t)}(0),\begin{bmatrix}\boldsymbol{A}^{(\upsilon,\,t),\,re}(1) & -\boldsymbol{A}^{(\upsilon,\,t),\,im}(1)\\ \boldsymbol{A}^{(\upsilon,\,t),\,im}(1) & \boldsymbol{A}^{(\upsilon,\,t),\,re}(1)\end{bmatrix},\begin{bmatrix}\boldsymbol{A}^{(\upsilon,\,t),\,re}(2) & -\boldsymbol{A}^{(\upsilon,\,t),\,im}(2)\\ \boldsymbol{A}^{(\upsilon,\,t),\,im}(2) & \boldsymbol{A}^{(\upsilon,\,t),\,re}(2)\end{bmatrix},\ldots,\begin{bmatrix}\boldsymbol{A}^{(\upsilon,\,t),\,re}(NF-1) & -\boldsymbol{A}^{(\upsilon,\,t),\,im}(NF-1)\\ \boldsymbol{A}^{(\upsilon,\,t),\,im}(NF-1) & \boldsymbol{A}^{(\upsilon,\,t),\,re}(NF-1)\end{bmatrix}\right] \quad (5.19)$$

As matrizes constitutivas $\overline{\boldsymbol{B}}_f^{(\upsilon,\,k)}$ e $\overline{\boldsymbol{B}}_s^{(\upsilon,\,k)}$ são organizadas da mesma forma descrita em (5.19).

No contexto da análise do BH, seguindo a formulação multi-níveis apresentada no Capítulo 4, podemos escrever o vetor de variável de estado associado à SuR intermediária, $\mathfrak{s}_{\upsilon-1,\,i}$, ver (4.23.a), como se segue

$$\overline{\boldsymbol{X}}^{(\upsilon-1,\,i)} = \left[\overline{\boldsymbol{X}}^{(\upsilon,\,j)} \ \ldots\ \overline{\boldsymbol{X}}^{(\upsilon,\,l)}\ \overline{\boldsymbol{X}}_{\gamma}^{(\upsilon)}\right]^T. \quad (5.20)$$

Podemos seguir o processo de construção recursivo, descrito no capítulo anterior, e formulado para um ponto em frequência. Então, podemos compactar as equações constitutivas das SuRs de fundo e as equações das redes de conexão, segundo a organização por frequência descrita acima. A equação determinante do BH do circuito, com decomposição hierárquica em múltiplos níveis, pode ser compactamente escrita como

$$\overline{\boldsymbol{F}}(\overline{\boldsymbol{X}}) = \overline{\boldsymbol{A}}\overline{\boldsymbol{X}} + \overline{\boldsymbol{B}}_f\overline{\boldsymbol{U}}_f(\overline{\boldsymbol{X}}) + \overline{\boldsymbol{W}}, \quad (5.21)$$

onde

$$\overline{\boldsymbol{W}} = \overline{\boldsymbol{B}}_g \overline{\boldsymbol{U}}_g . \tag{5.22}$$

Na equação determinante multi-níveis (5.21), a matriz $\overline{\boldsymbol{A}}$ assume uma estrutura multi-níveis tipo bloco diagonal com dupla borda, enquanto as matrizes $\overline{\boldsymbol{B}}_f$ e $\overline{\boldsymbol{B}}_g$ assumem uma estrutura multi-níveis bloco diagonal com borda horizontal. As matrizes são organizadas por SuR e por frequência. Após a determinação do vetor de variável de estado, podemos calcular as tensões e as correntes associadas com as sondas internas das SuRs de fundo, utilizando a seguinte expressão

$$\overline{\boldsymbol{Y}} = \overline{\boldsymbol{M}}\overline{\boldsymbol{X}} + \overline{\boldsymbol{N}}_f \overline{\boldsymbol{U}}_f(\overline{\boldsymbol{X}}) + \overline{\boldsymbol{N}}_g \overline{\boldsymbol{U}}_g . \tag{5.23}$$

Para a equação de sonda (5.23), a matriz $\overline{\boldsymbol{M}}$ assume uma estrutura multi-níveis bloco diagonal com borda vertical, enquanto as matrizes $\overline{\boldsymbol{N}}_f$ e $\overline{\boldsymbol{N}}_g$ assumem uma estrutura multi-níveis bloco diagonal. As matrizes são organizadas por SuR e por frequência. Uma eficiente representação multi-níveis, para implementação numérica de matrizes bloco diagonal com bordas, foi desenvolvida. Em conjunto, também foram implementadas numericamente técnicas de matrizes densas e esparsas para operar neste tipo de matrizes.

5.7. Matriz Jacobiana

Uma aspecto crucial na análise do BH é a geração, o armazenamento (demanda de memória), e a fatorização (tempo de processamento) da matriz jacobiana associada com a equação determinante. Os sistemas jacobianos que emergem a cada iteração do solucionador não-linear utilizado na solução da equação determinante do BH, são resolvidos via fatorização LU. Para os solucionadores de equação não-linear, com solução iterativa dos sistemas jacobianos, e.g. métodos de Newton inexato e do tensor inexato, a serem discutidos em capítulo subsequênte, exige-se um eficiente pré-condicionador fundamentado na matriz jacobiana. O mais utilizado é o pré-condicionador do tipo bloco diagonal que, por sua vez, é resolvido via fatorização LU [26],[160]. Vale ressaltar que, a técnica de decomposição hierárquica do circuito, introduzida acima, permite reduzir significantemente os custos de formação e de fatorização da matriz jacobiana, que assume uma estrutura multi-níveis tipo bloco diagonal com dupla borda. Em adição, a matriz jacobiana associada ào vetor de função não-linear do circuito assume uma estrutura bloco diagonal, onde cada bloco corresponde à matriz jacobiana associada ào vetor de função não-linear de uma SuR de fundo. Dependendo do número de variáveis de estado não-linear e do número de linhas espectrais envolvidas no problema do BH, a dimensão de cada bloco da matriz jacobiana pode ser excessivamente alta para ser armazenada e processada utilizando técnicas de matriz densa. Nesta

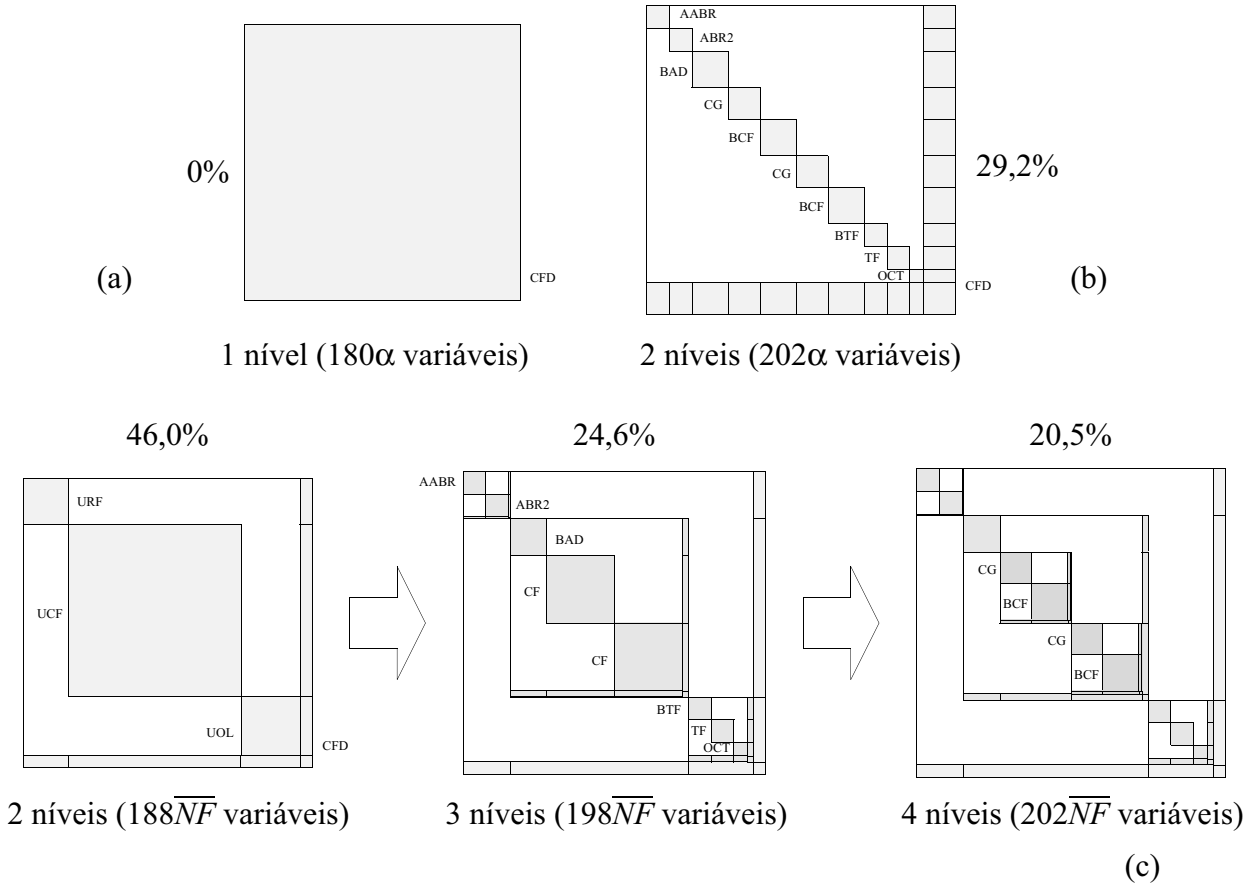


Fig. 5.6 Estrutura da matriz jacobiana com (a) 1 (um) nível e (b) 2 (dois) níveis de decomposição. (c) Estrutura multi-níveis (bloco diagonal com borda dupla) da matriz jacobiana para sucessivas decomposições incluindo um nível adicional. Assume-se que cada transistor intruduz 3 (três) variáveis de estado não-linear ($\overline{NF} = 2NF - 1$).

condição, o uso de técnicas de matrizes esparsas é mandatório [146],[148].

Para ilustrar as observações acima, a estrutura da matriz jacobiana associada ào exemplo de circuito descrito na Fig. 2.3, do Capítulo 2, para diferentes esquemas de decomposição multi-níveis, pode ser visualizada na Fig. 5.6. Mais precisamente, nas Figs. 5.6(a) e (b), podemos observar a estrutura de 1 nível e com decomposição de 2 níveis, respectivamente. Enqunato na Fig. 5.6(c) é apresentada a estrutura da matriz jacobiana do circuito para esquemas de decomposição com 2, 3 e 4 níveis de hierarquia. Vale ressaltar que a estrututura de 2 níveis da Fig. 5.6(b) pode ser obtida apartir das estruturas apresentadas na Fig. 5.6(c), com a aplicação da técnica de nivelamento, introduzida no capítulo anterior. Após o nivelamento, podemos observar que o complemento de Shur, da decomposição da Fig. 5.6(b), possue uma dimensão comparável ao maior bloco diagonal, sendo esta dimensão igual à soma das dimensões dos complementos de Shur da estrutura de 4 níveis da Fig. 5.6(c). É importante destacar, que a fatorização LU de uma matriz na forma multi-níveis tipo bloco diagonal com borda dupla não gera elementos não-zero extras, ver Apêndice C.

A matriz jacobiana correspondente à derivada do vetor de resíduo, definido pela equação determinante multi-níveis (5.21), com relação ao vetor de variável de estado, é dada por

$$\bar{\boldsymbol{J}}(\bar{\boldsymbol{X}}) = \frac{\partial \bar{\boldsymbol{F}}(\bar{\boldsymbol{X}})}{\partial \bar{\boldsymbol{X}}} = \bar{\boldsymbol{A}} + \bar{\boldsymbol{B}}_f \bar{\boldsymbol{G}}_f(\bar{\boldsymbol{X}}), \tag{5.24}$$

onde

$$\bar{\boldsymbol{G}}_f(\bar{\boldsymbol{X}}) = \frac{\partial \overline{\boldsymbol{U}}_f(\bar{\boldsymbol{X}})}{\partial \bar{\boldsymbol{X}}}. \tag{5.25}$$

A matriz jacobiana, $\bar{\boldsymbol{J}}(\bar{\boldsymbol{X}})$, assume uma estrutura multi-níveis tipo bloco diagonal com dupla borda. Os blocos na diagonal correspondem às matrizes jacobianas das SuRs de fundo e os blocos da borda horizontal correspondem às matrizes jacobianas associadas às redes de conexão. Os blocos na borda vertical é igual aos blocos na borda vertical de $\bar{\boldsymbol{A}}$, uma vez que o vetor de funcão não-linear não depende das variáveis de estado das redes de conexão. Seguindo a formulação do capítulo anterior, a matriz jacobiana da SuR intermediária, $\mathcal{S}_{\upsilon-1,\,i}$, é dada por

$$\bar{\boldsymbol{J}}^{(\upsilon-1,\,i)}(\bar{\boldsymbol{X}}^{(\upsilon-1,\,i)}) = \begin{bmatrix} \bar{\boldsymbol{J}}^{(\upsilon,\,j)}(\bar{\boldsymbol{X}}^{(\upsilon,\,j)}) & & \bar{\boldsymbol{A}}_\gamma^{(\upsilon,\,j)} \\ & \cdots & \cdots \\ & \bar{\boldsymbol{J}}^{(\upsilon,\,l)}(\bar{\boldsymbol{X}}^{(\upsilon,\,l)}) & \bar{\boldsymbol{A}}_\gamma^{(\upsilon,\,l)} \\ \bar{\boldsymbol{J}}_\gamma^{(\upsilon,\,j)}(\bar{\boldsymbol{X}}^{(\upsilon,\,j)}) & \cdots\ \bar{\boldsymbol{J}}_\gamma^{(\upsilon,\,l)}(\bar{\boldsymbol{X}}^{(\upsilon,\,l)}) & \bar{\boldsymbol{C}}_\gamma^{(\upsilon-1,\,i)} \end{bmatrix}, \tag{5.26}$$

onde

$$\bar{\boldsymbol{J}}^{(\upsilon,\,t)}(\bar{\boldsymbol{X}}^{(\upsilon,\,t)}) = \frac{\partial \bar{\boldsymbol{F}}^{(\upsilon,\,t)}(\bar{\boldsymbol{X}})}{\partial \bar{\boldsymbol{X}}^{(\upsilon,\,t)}} = \bar{\boldsymbol{A}}^{(\upsilon,\,t)} + \bar{\boldsymbol{B}}_f^{(\upsilon,\,t)} \bar{\boldsymbol{G}}_f^{(\upsilon,\,t)}(\bar{\boldsymbol{X}}^{(\upsilon,\,t)}), \tag{5.27.a}$$

$$\bar{\boldsymbol{J}}_\gamma^{(\upsilon,\,t)}(\bar{\boldsymbol{X}}^{(\upsilon,\,t)}) = \frac{\partial \bar{\boldsymbol{F}}_\gamma^{(\upsilon,\,t)}(\bar{\boldsymbol{X}})}{\partial \bar{\boldsymbol{X}}^{(\upsilon,\,t)}} = \bar{\boldsymbol{C}}^{(\upsilon,\,t)} + \bar{\boldsymbol{D}}_f^{(\upsilon,\,t)} \bar{\boldsymbol{G}}_f^{(\upsilon,\,t)}(\bar{\boldsymbol{X}}^{(\upsilon,\,t)}). \tag{5.27.b}$$

Lembremos que o vetor de resíduo, $\bar{\boldsymbol{F}}(\bar{\boldsymbol{X}})$, possue a mesma organização do vetor de variável de estado, $\bar{\boldsymbol{X}}$, ver e (5.20).

A matriz jacobiana, $\bar{\boldsymbol{G}}_f(\bar{\boldsymbol{X}})$, definida em (5.25), assume uma estrutura multi-níveis tipo bloco diagonal com borda vertical. Os blocos na diagonal correspondentes às matrizes jacobianas das SuRs de fundo $j, \ldots, l$. Em adição, a borda vertical é igual a zero, uma vez que o vetor de função não-linear não depende das variáveis de estado das redes de conexão.

$$\bar{\boldsymbol{G}}_f^{(\upsilon-1,\,i)}(\bar{\boldsymbol{X}}^{(\upsilon-1,\,i)}) = \begin{bmatrix} \bar{\boldsymbol{G}}_f^{(\upsilon,\,j)}(\bar{\boldsymbol{X}}^{(\upsilon,\,j)}) & & \mathbf{0} \\ & \cdots & \cdots \\ & \bar{\boldsymbol{G}}_f^{(\upsilon,\,l)}(\bar{\boldsymbol{X}}^{(\upsilon,\,l)}) & \mathbf{0} \end{bmatrix}, \tag{5.28}$$

onde

$$\bar{\boldsymbol{G}}^{(\upsilon,\,t)}(\bar{\boldsymbol{X}}^{(\upsilon,\,t)}) = \frac{\partial \bar{\boldsymbol{F}}_\gamma^{(\upsilon,\,t)}(\bar{\boldsymbol{X}})}{\partial \bar{\boldsymbol{X}}^{(\upsilon,\,t)}}. \tag{5.29}$$

A seguir, discutiremos o cálculo da matriz jacobiana, $\overline{\boldsymbol{G}}^{(\upsilon,\,t)}(\overline{\boldsymbol{X}}^{(\upsilon,\,t)})$, associada ao vetor de função não-linear de uma SuR de fundo, $\mathcal{S}_{\upsilon,\,t}$. Para simplificar notação, nas expressões abaixo iremos suprimir o superescrito $(\upsilon,\,t)$.

Sendo assim, as derivadas das componentes espectrais do vetor de função não-linear, em relação às componentes espectrais do vetor de variável não-linear, são dadas por:

$$\frac{\partial U_{f,\,p}(\omega_0)}{\partial X_q(\omega_0)} = \sum_{n\,=\,0}^{NS-1} \frac{\partial u_{f,\,p}(t_n)}{\partial x_q(t_n)} \tag{5.30.a}$$

$$\frac{\partial U_{f,\,p}(\omega_k)}{\partial X_q(\omega_0)} = \sum_{n\,=\,0}^{NS-1} \frac{\partial u_{f,\,p}(t_n)}{\partial x_q(t_n)} \gamma_{k,\,n} \tag{5.30.b}$$

$$\frac{\partial U_{f,\,p}(\omega_0)}{\partial X^{re}_q(\omega_l)} = \sum_{n\,=\,0}^{NS-1} \frac{\partial u_{f,\,p}(t_n)}{\partial x_q(t_n)} (2\Gamma^{re}_{n,\,l}) \qquad \frac{\partial U_{f,\,p}(\omega_0)}{\partial X^{im}_q(\omega_l)} = \sum_{n\,=\,0}^{NS-1} \frac{\partial u_{f,\,p}(t_n)}{\partial x_q(t_n)} (-2\Gamma^{im}_{n,\,l}) \tag{5.30.c}$$

$$\frac{\partial U_{f,\,p}(\omega_k)}{\partial X^{re}_q(\omega_l)} = \sum_{n\,=\,0}^{NS-1} \frac{\partial u_{f,\,p}(t_n)}{\partial x_q(t_n)} (2\Gamma^{re}_{n,\,l}) \gamma_{k,\,n} \qquad \frac{\partial U_{f,\,p}(\omega_k)}{\partial X^{im}_q(\omega_l)} = \sum_{n\,=\,0}^{NS-1} \frac{\partial u_{f,\,p}(t_n)}{\partial x_q(t_n)} (2\Gamma^{im}_{n,\,l}) \gamma_{k,\,n} \tag{5.30.d}$$

onde $p \in [1,\, n_{FN}]$ e $q \in [1,\, n_{VEN}]$. Os coeficientes complexos

$$\gamma_{k,\,n} = e^{\hat{\mathbf{j}}\omega_k t_n} \tag{5.31.a}$$

e

$$\Gamma_{n,\,k} = \Gamma^{re}_{n,\,k} + \hat{\mathbf{j}}\Gamma^{im}_{n,\,k} = (1/NS) e^{-\hat{\mathbf{j}}\omega_k t_n}\,, \tag{5.31.b}$$

correspondem aos elementos da matriz que definem a TFD, e sua inversa, respectivamente. Da equação (5.31.b), podemos verificar facilmente que

$$2\Gamma^{re}_{n,\,k} = \Gamma_{n,\,k} + \Gamma_{n,\,-k}\,, \tag{5.32.a}$$

e

$$2\Gamma^{im}_{n,\,k} = (-\hat{\mathbf{j}})(\Gamma_{n,\,k} - \Gamma_{n,\,-k})\,. \tag{5.32.b}$$

Neste ponto, é conveniente definir: $g_{pq}(t) = du_{f,\,p}(\boldsymbol{x}(t))/dx_q(t)$. Agora, substituindo as relações (5.32.a) e (5.32.b) nas expressões (5.30.a)-(5.30.d), e observando que

$$\gamma_{k,\,n}\Gamma_{n,\,\pm l} = (1/NS) e^{\hat{\mathbf{j}}(\omega_k \pm \omega_l)t_n} \equiv \gamma_{k\,\pm\,l,\,n}\,, \tag{5.33}$$

podemos escrever as expressões correspondentes às derivadas do harmônico da função não-linear com relação ao harmônico da variável de estado. Estas expressões são dadas por:

$$\frac{\partial U_{f,\,p}(\omega_0)}{\partial X_q(\omega_0)} = G_{pq}(\omega_0) \tag{5.34.a}$$

$$\frac{\partial U^{re}_{f,p}(\omega_k)}{\partial X_q(\omega_0)} = G^{re}_{pq}(\omega_k) \qquad \frac{\partial U^{im}_{f,p}(\omega_k)}{\partial X_q(\omega_0)} = G^{im}_{pq}(\omega_k) \tag{5.34.b}$$

$$\frac{\partial U_{f,p}(\omega_0)}{\partial X^{re}_q(\omega_l)} = 2G^{re}_{pq}(-\omega_l) \qquad \frac{\partial U_{f,p}(\omega_0)}{\partial X^{im}_q(\omega_l)} = 2G^{im}_{pq}(-\omega_l) \tag{5.34.c}$$

$$\frac{\partial U^{re}_{f,p}(\omega_k)}{\partial X^{re}_q(\omega_l)} = G^{re}_{pq}(\omega_k+\omega_l)+G^{re}_{pq}(\omega_k-\omega_l) \qquad \frac{\partial U^{re}_{f,p}(\omega_k)}{\partial X^{im}_q(\omega_l)} = G^{im}_{pq}(\omega_k+\omega_l)-G^{im}_{pq}(\omega_k-\omega_l) \tag{5.34.d}$$

$$\frac{\partial U^{im}_{f,p}(\omega_k)}{\partial X^{re}_q(\omega_l)} = G^{im}_{pq}(\omega_k+\omega_l)+G^{im}_{pq}(\omega_k-\omega_l) \qquad \frac{\partial U^{im}_{f,p}(\omega_k)}{\partial X^{im}_q(\omega_l)} = -G^{re}_{pq}(\omega_k+\omega_l)+G^{re}_{pq}(\omega_k-\omega_l) \tag{5.34.e}$$

onde $k, l \in [0, NF]$ e $|\omega_k \pm \omega_l| \in \mathbf{S}_d(NF_d) \supset \mathbf{S}(NF)$. $\mathbf{S}_d(NF_d)$ e NF_d correspondem ao espectro de frequência de derivadas e ao número de frequências em $\mathbf{S}_d$. Os coeficientes G_{pq}, G^{re}_{pq}, e G^{im}_{pq} correspondem aos coeficientes de Fourier da função no tempo, g_{pq}, definida previamente. A notação de matriz de bloco, $\boldsymbol{G}(0) = [G_{pq}(0)]$ e $\boldsymbol{G}^{re,\,im}(k) = [G^{re,\,im}_{pq}(k)]$, foi utilizada.

As expressões acima, podem ser generalizadas para regime multi-tons empregando a TFDM, ver somatórias (5.7.a) e (5.7.b). Sendo assim, introduzindo o vetor de índice $\boldsymbol{k} = (k_1, k_2, \ldots, k_{NT})$, para simplificar a notação, podemos escrever

$$\frac{\partial U_{f,p}(\mathbf{0})}{\partial X_q(\mathbf{0})} = G_{pq}(\mathbf{0}) \tag{5.35.a}$$

$$\frac{\partial U^{re}_{f,p}(\boldsymbol{k})}{\partial X_q(\mathbf{0})} = G^{re}_{pq}(\boldsymbol{k}) \qquad \frac{\partial U^{im}_{f,p}(\boldsymbol{k})}{\partial X_q(\mathbf{0})} = G^{im}_{pq}(\boldsymbol{k}) \tag{5.35.b}$$

$$\frac{\partial U_{f,p}(\mathbf{0})}{\partial X^{re}_q(\boldsymbol{l})} = 2G^{re}_{pq}(\boldsymbol{l}) \qquad \frac{\partial U_{f,p}(\mathbf{0})}{\partial X^{im}_q(\boldsymbol{l})} = -2G^{im}_{pq}(\boldsymbol{l}) \tag{5.35.c}$$

$$\frac{\partial U^{re}_{f,p}(\boldsymbol{k})}{\partial X^{re}_q(\boldsymbol{l})} = G^{re}_{pq}(\boldsymbol{k}+\boldsymbol{l})+G^{re}_{pq}(\boldsymbol{k}-\boldsymbol{l}) \qquad \frac{\partial U^{re}_{f,p}(\boldsymbol{k})}{\partial X^{im}_q(\boldsymbol{l})} = G^{im}_{pq}(\boldsymbol{k}+\boldsymbol{l})-G^{im}_{pq}(\boldsymbol{k}-\boldsymbol{l}) \tag{5.35.d}$$

$$\frac{\partial U^{im}_{f,p}(\boldsymbol{k})}{\partial X^{re}_q(\boldsymbol{l})} = G^{im}_{pq}(\boldsymbol{k}+\boldsymbol{l})+G^{im}_{pq}(\boldsymbol{k}-\boldsymbol{l}) \qquad \frac{\partial U^{im}_{f,p}(\boldsymbol{k})}{\partial X^{im}_q(\boldsymbol{l})} = -G^{re}_{pq}(\boldsymbol{k}+\boldsymbol{l})+G^{re}_{pq}(\boldsymbol{k}-\boldsymbol{l}) \tag{5.35.e}$$

onde $\boldsymbol{k}, \boldsymbol{l} \in [-\boldsymbol{NH}, \boldsymbol{NH}]$ ($\boldsymbol{NH} = (NH_1, NH_2, \ldots, NH_{NT})$) e $|\omega_{\boldsymbol{k}\pm\boldsymbol{l}}| \in \mathbf{S}_d(NF_d) \supset \mathbf{S}(NF)$. Convém ressaltar, que as expressões acima assumem uma forma mais simples do que as expressões obtidas com a equação paramétrica, descrita em [18]. Como ilustração, a matriz jacobiana para um regime de único-tom com NH harmônicos é mostrada na próxima página em (5.36).

Utilizando técnicas de matrizes esparsas, o controle de esparsidade da matriz jacobiana, para solução do problema EH no caso de único-tom, pode ser conduzido com

$$\mathbf{S}_d = |k_1| \le M_d . \tag{5.36}$$

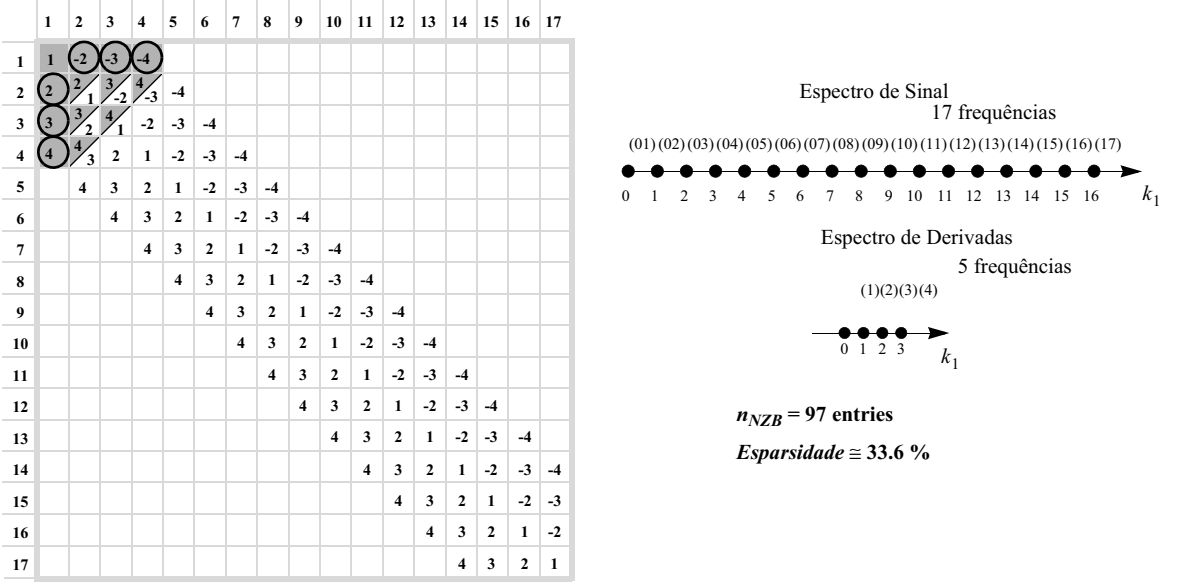
Para entendermos o significado físico da expressão acima, vamos considerar que a amplitude (nível de potência) das fontes de excitação de CA é pequena o suficiente para não afetar as não-linearidades das SuRs de fundo. Nesta condição, apenas a componente de CC ($NF_d = 1$) do espectro de frequência de derivadas, $\mathbf{S}_d$, é necessária para a solução precisa da equação determinante do BH. Com o aumento do nível de potência das fontes de excitação, se torna necessário o aumento da largura-de-banda da matriz jacobiana. Na Fig. 5.7(a), podemos observar o padrão de esparsidade da matriz jacobiana, para o caso de único-tom gerado por (5.36), com $M_d = 4$ e $NH_1 = 16$. Neste caso, a fatorização LU não produzirá nenhum elemento não-zero adicional, i.e., o padrão de esparsidade é preservado [146].

Em regime multi-tons, para a exploração de técnicas de matriz esparsa, a estrutura não-zero da matriz jacobiana associada ào vetor de função não-linear, pode ser definida pelo seguinte espectro de frequência de derivada

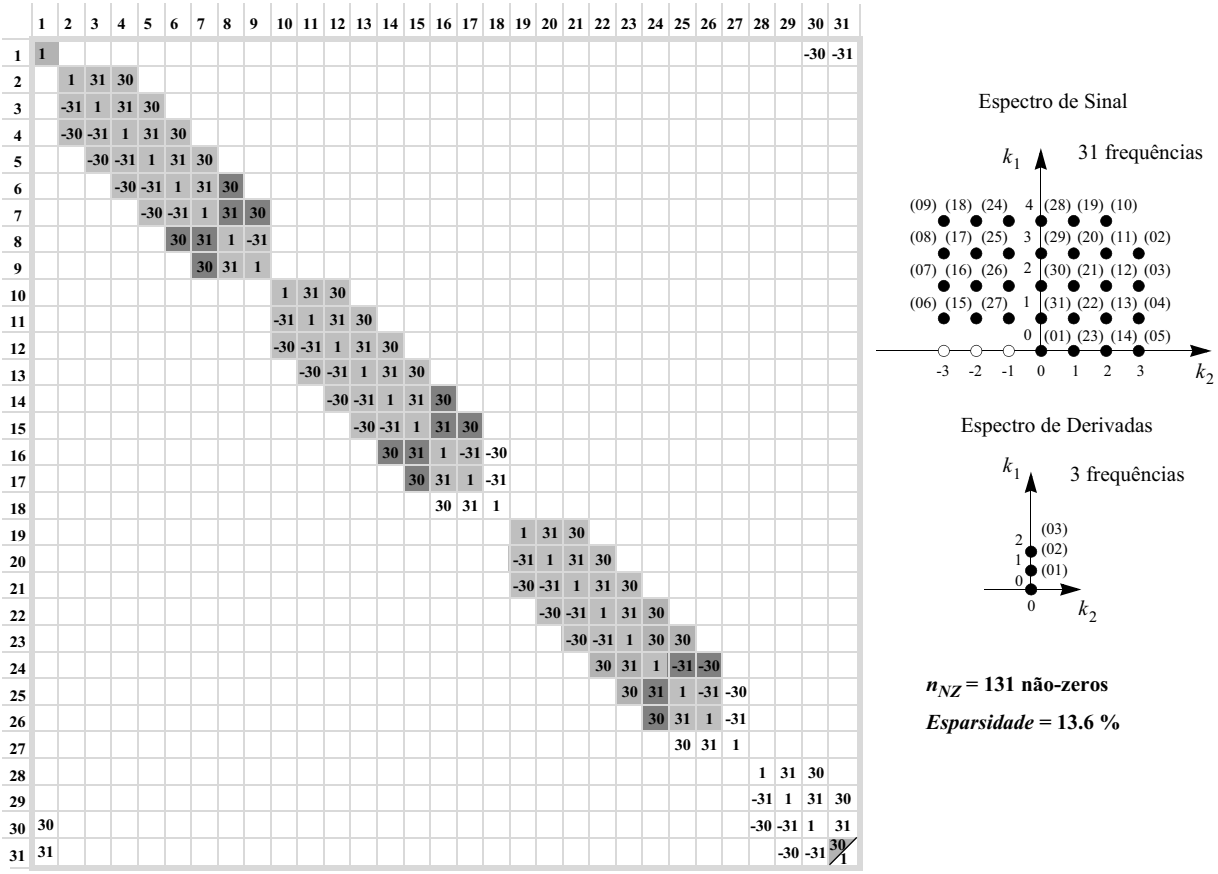
$$\mathbf{S}_d = \begin{cases} \sum_{i=1}^{N_d} |k_i| \le M_d \\ \text{where:} \\ \quad |k_i| \le NH_i,\ 1 \le i \le NT \\ \quad k_i = 0,\ N_d + 1 \le i \le NT \end{cases}, \tag{5.37}$$

onde N_d e M_d são números inteiros, com $N_d < NT$ [18]. Como $\omega_{\boldsymbol{k}\pm\boldsymbol{l}} = |\omega_{\boldsymbol{k}} \pm \omega_{\boldsymbol{l}}|$, então, em vista de (5.35.a)-(5.35.e) e da terceira relação de (5.37), os produtos de IM gerados por $\omega_{\boldsymbol{k}}$ e $\omega_{\boldsymbol{l}}$ serão considerados acoplados apenas quando $|k_i| = |l_i|$, para $k_i = 0,\ N_d + 1 \le i \le NT$. Neste caso, temos que $\omega_{\boldsymbol{k}\pm\boldsymbol{l}} \in \mathbf{S}_d(NF_d)$. Vale notar que o parâmetro M_d controla a largura-de-banda da matriz jacobiana. Conforme destacado em [18], em algumas situações a técnica de *frequency-windowing* [49] produzirá um desacoplamento da matriz jacobiana estruturalmente semelhante à estrutura tipo bloco diagonal, obtida com (5.37). Para ilustrar a aplicação do espectro de derivada (5.37), na Fig. 5.7(b) podemos observar o padrão de esparsidade da matriz jacobiana obtida com $N_d = 1$ e $M_d = 2$. Neste caso, a fatorização LU não produzirá uma entrada não-zero. A topologia do espectro de sinal corresponde a um conversor de frequência de dois-tons ($NT = 2$), ver Tabela 5.1, com $M_{LO} = 6$, $N_{LO} = NH_0 = 4$ e $N_{SB} = NH_1 = 3$.

Dependendo do tipo de análise o espectro de derivadas pode ser diferente de uma SuR para outra. Para exemplificar, com o arquétipo dado no Capítulo 2, podemos considerar a análise de um típico sistema de recepção de RF composto dos seguintes módulos: amplificador de RF de baixo-



(a)



(b)

zero

n não-zero: $\omega_k + \omega_l = \omega_n$

m não-zero: $|\omega_k - \omega_l| = \omega_m$

m/n não-zero: $|\omega_k - \omega_l| = \omega_m$ e $\omega_k + \omega_l = \omega_n$

Fig. 5.7 (a) Padrão de esparsidade da matriz jacobiana para análise de único-tom. (b) Padrão de esparsidade da matriz jacobiana para análise de dois-tons (n_{NZB} - número de blocos não-zero).

ruído, conversor em frequência de descida, oscilador local (OL) e amplificador de frequência intermediária (FI). Assumindo que cada um destes módulos consiste de uma SuR, vamos assumir que estamos interessados na determinação da faixa dinâmica do sistema de recepção. Neste caso, o amplificador de RF de sinais pode ser co. A rede de conexão pode ser considerada com o mesmo espectro de derivadas definido para o CF.

5.8. Conclusão

Iniciamos este capítulo com a apresentação das diferentes topologias do espectro de frequência, comumente utilizadas na análise do BH em regime de dois, três e multi-tons. A análise do BH em regime multi-tons inclui uma versão particularmente útil na determinação da característica de recrescimento espectral em amplificadores, multiplicadores e conversores em frequência, e consequêntemente, na determinação de figuras de mérito para distorção por IM , e.g., ACPR.

Em seguida, foram descritas diferentes técnicas para implementação das TFDs utilizadas nas conversões tempo-frequência do sinal, e requeridas durante o processo de análise do BH. Para a simulação em regime de único-tom, a implementação via TFR (versão radix, i.e., não limitada por uma base de 2) é a opção mais eficiente. Para regime de dois e três-tons a TFR-2D e a TFR-3D., respectivamente, devem ser utilizadas para obter a máxima precisão. Diferentes versões TFQPs foram discutidas e implementadas para comparação com as TFRMs. Para análise de sistemas com multi-portadoras de RF, foi introduzida a técnica de MAF, que permite a condução de conversões via TFR uni-dimensional. Para operações envolvendo sinais com modulação digital, foi discutida e implementada a TFMT para análise de amplificadores, e a TFMT-2D para análise de conversores em frequência. Vale ressaltar que, este tipo de transformada pode ser facilmente extendida para sistemas com multi-portadoras utilizando MAF. tipicamente, para sistemas de única- e multi-portadoras de RF moduladas digitalmente, a análise é conduzida via método do BH-TE. A representação numérica de um sinal digital, utilizando multi-senos, foi discutida e implementada.

Finalmente, utilizando o desenvolvimento dos capítulos anteriores, apresentamos a formulação da equação determinante do BH, no contexto da decomposição multi-níveis de circuito. As matrizes envolvidas nesta formulação, são todas formadas de múltiplos níveis, onde cada nível assume uma forma tipo bloco diagonal com e sem bordas laterais. Para resolução desta equação determinante, foi apresentado o cálculo analítico da matriz jacobiana do BH, que assume uma estrutura hierárquica multi-níveis tipo bloco diagonal com borda dupla. O conceito de espectro de frequência de derivada foi discutido e implementado, oferecendo um controle de esparsidade da

matriz jacobiana. O impacto da formulação multi-níveis na fatorização da matriz jacobiana do BH, para a solução direta ou iterativa via pré-condicionadores dos sistemas jacobianos, será discutido nos próximos capítulos.

6. Métodos de Newton e do Tensor

6.1. Introdução

NO CAPÍTULO ANTERIOR, foi introduzida uma eficiente metodologia para a formulação multi-níveis do problema do BH associado a análise (e otimização) de circuitos de RF em grande-escala. Agora, discutiremos a teoria e a implementação de métodos iterativos, para a eficiente solução deste problema, com o circuito operando em regime *fortemente* não-linear. Consideraremos que as SuRs de fundo produzem subsistemas do BH de pequena- e média-escala. Tradicionalmente, estes métodos fundamentam-se em aproximações sucessivas da raiz, obtida da solução de um modelo linear local (sistema jacobiano), resultante da linearização da equação determinante do BH entorno da iteração corrente. O clássico método de Newton (método padrão) com ou sem iterações de corda e de *Shamanskii*, bem como, os métodos da secante (quasi-Newton) [17],[46] são os mais utilizados na análise do BH, conforme descrito em [37],[38],[45],[18],[19]. Neste capítulo, investigaremos a aplicação (robustez e eficiência) de uma nova classe de métodos, entitulados *métodos do tensor*. O desempenho do método do tensor será avaliado, discutido e implementado em relação ao método padrão. Estes métodos apresentam como principal característica o fato de fundamentarem-se em um modelo local, de extensão do modelo linear que é utilizado na iteração de Newton com a inclusão de um termo quadrático contendo informação de segunda-ordem.

Iniciaremos a nossa discussão, seções 6.2 e 6.3, descrevendo a natureza do problema do BH para análise de circuitos forçados, definindo notação e apresentando o conceito de função nível. Em seguida, na Seção 6.4, apresentaremos um resumo na teoria e na implementação do método de Newton. Neste resumo, que também facilitará a discussão do método do tensor, as subseções 6.4.1 e 6.4.2, respectivamente, discutem a estratégia de globalização via pesquisa-em-linha, e a .implementação numérica adotada para o método padrão. Na Seção 6.5, organizada em quatro subseções, apresentaremos a teoria e a implementação do método do tensor [29],[30]. Para iniciar, as subseções 6.5.1 e 6.5.2 expõem a teoria básica adotada na construção e na solução do modelo do tensor. A Subseção 6.5.3 discute duas diferentes estratégias de globalização via pesquisa-em-linha., específicas para os métodos do tensor, a saber: estratégia padrão (ver Subseção 6.5.3.a) [33] e estratégia curvilinear (ver Subseção 6.5.3.b) [161]. A última Subseção 6.5.4, delineia a implementação do método do tensor adotada neste trabalho. Em adição, são apresentados algoritmos detalhados descrevendo a implementação numérica destes métodos.

Ainda neste capítulo, com o objetivo de ampliar a região de convergência dos solucionadores

não-lineares, na Seção 6.6 é apresentada a técnica de continuação (ou homotopia) [47] implementada neste trabalho. Para validação dos algoritmos implementados e verificação preliminar no desempenho do método de Newton e do tensor, na Seção 6.7 são apresentados os resultados de uma série de testes numéricos realizados. As conclusões finais são reservadas para a Seção 6.8.

6.2. Considerações Preliminares

Conforme discutido anteriormente, a determinação da reposta de regime permanente em circuitos não-lineares forçados, via análise do BH, requer a solução de um sistema não-linear de equações algébricas. Sendo assim, este tipo de análise é equivalente a um problema de determinação-de-raiz, que pode ser postulado da seguine forma

$$\text{Dado } \bar{\boldsymbol{F}}\colon \mathrm{R}^N \to \mathrm{R}^N, \text{ encontrar } \bar{\boldsymbol{X}}^* \in \mathrm{R}^N \text{ tal que } \bar{\boldsymbol{F}}(\bar{\boldsymbol{X}}^*) = \mathbf{0} \in \mathrm{R}^N. \tag{6.1}$$

onde N é a dimensão do problema. Lembramos que, na análise do BH, a dimensão é igual ao produto do número de variáveis de estado vezes o número de linhas espectrais. Alternativamente, introduzindo a função objetivo uni-dimensional, U, podemos postular a mesma análise como um problema de otimização sem restrições, onde

$$\text{Dado } U\colon \mathrm{R}^N \to \mathrm{R}, \text{ encontrar } \bar{\boldsymbol{X}}^* \in \mathrm{R}^N \text{ no qual } U(\bar{\boldsymbol{X}}) \text{ é minimizada.} \tag{6.2}$$

A função objetivo é usualmente fundamentada na norma de *mínimo-quadrado* (M2), ou simplesmente norma-M2, do mapeamento não-linear, $\bar{\boldsymbol{F}}$. De acordo com o debate apresentado em [162], a análise do BH deve ser conduzida utilizando (6.1). Isto deve-se à superior velocidade de convergência local dos métodos de determinação-de-raiz (e.g., método de Newton e do tensor) em relação aos métodos de otimização. No caso de sistemas autônomos e sincronizados, a análise do BH pode ser iniciada como um problema de otimização (6.2), para determinar uma iteração inicial dentro da região de convergência local para solução de (6.1) [163]. Seguindo a clássica notação adotada em análise funcional, neste capítulo iremos representar os vetores $\bar{\boldsymbol{X}}$ e $\bar{\boldsymbol{F}}(\bar{\boldsymbol{X}})$ pelos vetores $\boldsymbol{x}$ e $\boldsymbol{F}(\boldsymbol{x})$, respectivamente.

6.3. Função Nível

A *função nível* (ou *mérito*) é um tipo especial de função, que pode ser utilizada na condução do teste de aceitação (ou monotonicidade) do fator de amortecimento em estratégias de globalização

do tipo pesquisa-em-linha. Esta função, no ponto $\boldsymbol{x}_k$, é definida pela seguinte expressão [166]

$$f(\boldsymbol{x}_k|\boldsymbol{A}_p) \stackrel{\text{def}}{=} \frac{1}{2}\|\boldsymbol{A}_p\boldsymbol{F}(\boldsymbol{x}_k)\|_2^2 = \frac{1}{2}\boldsymbol{F}(\boldsymbol{x}_k)^T\overline{\boldsymbol{A}}_p\boldsymbol{F}(\boldsymbol{x}_k) \tag{6.3}$$

onde $\boldsymbol{A}_p$ é uma matriz (N,N) arbitrária constante não-singular e $\overline{\boldsymbol{A}_p} = \boldsymbol{A}_p{}^T\boldsymbol{A}_p$. O subescrito "$p$" indica que a matriz, $\boldsymbol{A}_p$, é calculada em um determinado ponto estacionário, $\boldsymbol{x}_p$. Mais recentemente, foi introduzida, em [164], a *função norma vetor com peso* definida por

$$N(\boldsymbol{x}_k|\boldsymbol{A}_p) \stackrel{\text{def}}{=} (\boldsymbol{F}(\boldsymbol{x}_k)^T\overline{\boldsymbol{A}}_p\boldsymbol{F}(\boldsymbol{x}_k))^{1/2}. \tag{6.4}$$

Como podemos observar, as funções definidas em (6.3) e (6.4) estão diretamente relacionadas, e os gradientes destas funções possuem a mesma direção, e são dados por

$$\nabla f(\boldsymbol{x}_k|\boldsymbol{A}_p) = \boldsymbol{J}(\boldsymbol{x}_k)^T\overline{\boldsymbol{A}}_p\boldsymbol{F}(\boldsymbol{x}_k) \tag{6.5}$$

e

$$\nabla N(\boldsymbol{x}_k|\boldsymbol{A}_p) = \frac{1}{N(\boldsymbol{x}_k|\boldsymbol{A}_p)}\boldsymbol{J}(\boldsymbol{x}_k)^T\overline{\boldsymbol{A}}_p\boldsymbol{F}(\boldsymbol{x}_k), \tag{6.6}$$

respectivamente, onde $\boldsymbol{J}(\boldsymbol{x}_k) \in \mathbb{R}^{N\times N}$ é a matriz jacobiana, em $\boldsymbol{x}_k$, associada ào mapeamento $\boldsymbol{F}$. Apesar deste fato, a definição de função nível (6.3) não é citada em [164].

6.4. Método de Newton

O método de Newton é fundamentalmente um processo iterativo de aproximações lineares suscessivas, para determinação da raiz de um sistema de equação não-linear. Mais precisamente, fundamenta cada iteração em um vetor de correção, $\boldsymbol{d}$, resultante da solução do modelo linear dado por

$$\boldsymbol{M}(\boldsymbol{x}_k + \boldsymbol{d}) = \boldsymbol{F}(\boldsymbol{x}_k) + \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d}, \tag{6.7}$$

onde $\boldsymbol{x}_k \in \mathbb{R}^N$, $\boldsymbol{F}(\boldsymbol{x}_k) \in \mathbb{R}^N$, e $\boldsymbol{J}(\boldsymbol{x}_k) \in \mathbb{R}^{N\times N}$ correspondem ao vetor de iteração, vetor de resíduo e matrix jacobiana, respectivamente. Observa-se, de imediato, que a raiz do modelo local descrito acima, corresponde à correção de Newton, e pode ser obtida através da solução do seguinte sistema

$$\boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d}_{N,k} = -\boldsymbol{F}(\boldsymbol{x}_k). \tag{6.8}$$

Obviamente, a raiz existe apenas se $\boldsymbol{J}_k$ for uma matriz não-singular. Após o cálculo do vetor $\boldsymbol{d}_{N,k}$, a nova iteração pode ser facilmente calculada como $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_{N,k}$. Sendo assim, o método de Newton pode ser postulado como se segue:

Dado: $\boldsymbol{x}_0 \in \mathbb{R}^N$ (6.9.a)

Para $k \leftarrow 0$ passo 1 até "convergência" faça: (6.9.b)

Encontrar: $\boldsymbol{d}_{N,\,k} \in \mathrm{R}^N$ tal que $\boldsymbol{M}_N(\boldsymbol{x}_k + \boldsymbol{d}_{N,\,k}) = 0$ (6.9.c)

$\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \boldsymbol{d}_{N,\,k}$ (6.9.d)

Assumindo que a convergência é atingida na *k*-ésima iteração, então podemos dizer que o processo iterativo acima é equivalente à solução de "k" sistemas lineares (ou sistemas jacobianos) descritos por (6.7).

A convergência do processo iterativo (6.9.a)-(6.9.d) pode ser caracterizada por importantes resultados teóricos (análise funcional), e.g.: teorema de *Newton-Kantorovich* [17, p. 421], teorema de *Newton-Mysovskii* [17, p. 412] e o teorema apresentado por *Dennis-Schnabel* em [46, p. 95]. Versões refinadas dos dois primeiros teoremas e invariantes sob transformação afim foram apresentadas em [165]. As principais suposições nestes teoremas, são: (i) a matriz jacobiana em um domínio, D_0, onde a solução é procurada, deve ser não-singular, e (ii) o mapeamento não-linear $\boldsymbol{F}: D \subset \mathrm{R}^N \rightarrow \mathrm{R}^N$ deve ser *Fréchet* diferenciável em um conjunto convexo, D, tal que: $D_0 \subset D$. As outras suposições são:

$$\|\boldsymbol{J}(\boldsymbol{y}) - \boldsymbol{J}(\boldsymbol{x})\| \leq \gamma \cdot \|\boldsymbol{y} - \boldsymbol{x}\| \qquad \boldsymbol{x}, \boldsymbol{y} \in D_0 \ , \tag{6.10.a}$$

$$\left\|\boldsymbol{J}(\boldsymbol{x})^{-1}\right\| \leq \beta \qquad \boldsymbol{x} \in D_0 \text{ , e} \tag{6.10.b}$$

$$\left\|\boldsymbol{J}(\boldsymbol{x}_0)^{-1}\boldsymbol{F}(\boldsymbol{x}_0)\right\| \leq \eta \ . \tag{6.10.c}$$

A suposição (6.10.a) corresponde à condição de continuidade *Lipschitz* (caso particular da condição de continuidade de *Hölder*) em D_0. No contexto deste trabalho, esta suposição requer que as funções representando as características não-lineares dos dispositivos ativos sejam Lipschitz contínua, i.e., possuam no mínimo derivadas contínuas de primeira-ordem (de ordem superior, no caso de Hölder contínua) e de suave variação em D. As demais suposições (6.10.b) e (6.10.c), definem um limite superior para o número de condicionamento da matriz jacobiana inversa, e para a norma do vetor de correção inicial, respectivamente. (Problema de deficiência de posto.)

As propriedades de convergência do método de Newton descritas nos teoremas clássicos citados acima são de natureza local, i.e., válidas apenas quando a iteração está suficientemente próxima da raiz. Sendo assim, para extensão do domínio de convergência, uma estratégia de globalização deve ser empregada. No contexto do método de Newton, as estratégias comumente utilizadas são pesquisa-em-linha e região-de-confiança. Esta última é mais adequada aos métodos de otimização. Neste trabalho, vamos considerar a estratégia de pesquisa-em-linha.

A determinação da correção de Newton (6.8), quando a matriz jacobiana for esparsa, pode ser eficientemente conduzida utilizando a fatorização LU e técnicas de matrizes esparsas [146],[148].

Para matriz jacobiana com representação densa, limitada à solução de sistemas de pequena-escala, a fatorização LU e o processo de retro-substituições possuem complexidades $O(N^3/3)$ e $O(N^2)$, respectivamente.

Com o propósito de ampliar o conhecimento teórico, e para comparação de desempenho, além do método de Newton, durante a execução deste trabalho, foram analisados e implementados os seguintes métodos: "Global Affine Invariant Newton" (GAIN) [166],[167],[168] e "Global Approximate Newton" (GAN) [169]. No método GAIN, o processo iterativo consiste basicamente de uma *malha a-priori*, e uma *malha a-posteriori*. Na malha a-priori a matriz jacobiana é sempre calculada e fatorada, e a correção de Newton é utilizada. Já na malha a-posteriori, a matriz jacobiana é mantida constante, e uma correção de Newton simplificada é utilizada. Neste método, o critério de chaveamento entre as malhas, está associado com a função nível natural (ou escalamento natural), e o fator de amortecimento é determinado em função de estimativas, chamadas de quantidades (ou parâmetros) de Kantorovich. No método GAN, o fator de amortecimento é determinado de forma simples, através da norma do resíduo e de um parâmetro adicional que controla a velocidade de convergência. Nos problemas testes considerados, estes métodos demonstraram um desempenho comparável, porém inferior ao método de Newton globalizado com pesquisa-em-linha.

Métodos fundamentados no modelo linear (6.7), no qual a matriz jacobiana, ou a sua inversa, pode ser aproximada em determinadas iterações, são chamados de quasi-Newton. O método da secante de Broyden [170] é um dos mais utilizados. Neste método, a aproximação secante da matriz jacobiana, ou da sua inversa, produz a *boa*, ou a *má,* fórmula de atualização de Broyden, respectivamente [46]. Este tipo de método é mais adequado para solução de sistemas não-lineares de pequena-escala, onde a matriz jacobiana é do tipo densa. A fórmula de atualização de Broyden pode ser eficientemente calculada, utilizando a matriz jacobiana decomposta via fatorização QR [46],[171].

6.4.1. Globalização via Pesquisa-em-Linha

A pesquisa-em-linha via retrocedimento é uma estratégia de globalização, de implementação *ad hoc*, comumente empregada no método de Newton, para melhorar sua robustez na busca da raiz quando a estimativa reside fora da região de convergência local. Assume-se que $\boldsymbol{x}_k$ e $\boldsymbol{d}$ correspondem aos vetores de iteração e de correção (de Newton ou do tensor), respectivamente. Então, a estratégia em questão tenta encontrar uma nova iteração $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \lambda_* \boldsymbol{d}$, com $\lambda_* \in \mathbb{R}:(0, 1]$, que satisfaça um apropriado critério de redução da função nível, i.e., teste de

monotonicidade. Numa j-ésima iteração da pesquisa-em-linha, o valor ótimo do *fator de amortecimento* (ou relaxação), $\lambda_{k,j}$, pode ser obtido analiticamente através da solução de um polinômio de baixa-ordem, tipicamente de ordem quadrática, que interpola a função nível parametrizada $\hat{f}(\lambda_{k,j}|\boldsymbol{A}_k) = f(\boldsymbol{x}_k + \lambda_{k,j}\boldsymbol{d}_k|\boldsymbol{A}_k)$. Notar que, $\hat{f}:\mathrm{R} \to \mathrm{R}$ é uma função de restrição uni-dimensional. Uma estratégia mais simples é a *dividindo-pela-metade*, onde o fator de amortecimento é simplesmente divido pela metade cada vez que o critério de aceitação é violado.

Utilizando a definição de função nível (6.3), e do seu gradiente (6.5), o teste de monotonicidade para aceitação do fator de amortecimento, $\lambda_{k,j}$, pode ser implementado via

$$f(\boldsymbol{x}_k + \lambda_{k,j}\boldsymbol{d}_k|\boldsymbol{A}_k) \le f(\boldsymbol{x}_k|\boldsymbol{A}_k) + \alpha\lambda_{k,j}\mathrm{tax\_inc}_k\,, \tag{6.11}$$

onde $\alpha \in \mathrm{R}$ é um parâmetro tipicamente igual a 10^{-4}, correspondendo a um suave critério de aceitação, e $\mathrm{tax\_inc}_k \in \mathrm{R}$ é a *taxa de inclinação* (ou derivada direcional) de f no ponto $\boldsymbol{x}_k$, dada por

$$\mathrm{tax\_inc}_k = \nabla f(\boldsymbol{x}_k|\boldsymbol{A}_k)^T\boldsymbol{d}_{N,k} = -\boldsymbol{F}(\boldsymbol{x}_k)^T\bar{\boldsymbol{A}}_k\boldsymbol{F}(\boldsymbol{x}_k)\,. \tag{6.12}$$

O teste monotônico (6.11) corresponde à bem conhecida condição de *Armijo* [46],[17] e favorece a característica de rápida convergência local (q-quadrática) dos métodos de Newton e do tensor (a ser discutido), particularmente quando a iteração corrente não está próxima da raiz procurada. Na prática, para assegurar esta característica, o fator de amortecimento deve ser restrito a um valor mínimo, $\lambda_{\min} \in \mathrm{R}:(0,1]$. Adicionalmente, o valor inicial do fator de amortecimento, $\lambda_0 \in \mathrm{R}:(0,1]$, deve ser selecionado de forma a evitar que a correção conduza a uma possível divergência em problemas com acentuadas não-linearidades. Fundamentada em experências númericas em uma grande variedade de problemas, a Tabela 6.I apresenta um simples critério para seleção do fator de amortecimento inicial e mínimo [168]. Como podemos observar, a seleção depende da classe do problema, definida em termos da força das não-linearidades presentes no mapeamento não-linear $\boldsymbol{F}(\boldsymbol{x})$.

Tabela 6.1
DEFINIÇÃO DE CLASSE DO PROBLEMA

Classe do Problema	Fator de amortecimento	
	inicial (λ_0)	mínimo ($\lambda_{\min}$)
linear	1	—
suavemente não-linear	1	0.0001
altamente não-linear	0.01	0.0001
extremamente não-linear	0.0001	0.0001

O teste de aceitação (6.3) pode produzir diferentes resultados a depender da definição da matriz de peso $\boldsymbol{A}_k$. Um importante resultado é o da função nível associada com a norma de *mínimos-quadrados* (M2), ou simplesmente norma-M2, onde temos que

$$\boldsymbol{A}_k = \mathbf{1} \Rightarrow \begin{cases} f_{\mathrm{M2}}(\boldsymbol{x}) \;=\; f(\boldsymbol{x}|\mathbf{1}) \;=\; \dfrac{1}{2}\boldsymbol{F}(\boldsymbol{x})^T\boldsymbol{F}(\boldsymbol{x}) \\ \nabla f_{\mathrm{M2}}(\boldsymbol{x}) \;=\; \nabla f(\boldsymbol{x}|\mathbf{1}) \;=\; \boldsymbol{J}(\boldsymbol{x})^T\boldsymbol{F}(\boldsymbol{x}) \end{cases} . \tag{6.13}$$

Neste caso, utilizando (6.12) e (6.13) deduzimos que a taxa de inclinação em norma-M2 é dada por

$$\mathrm{tax\_inc}_{\mathrm{M2},\,k} \;=\; \nabla f_{\mathrm{M2}}(\boldsymbol{x}_k)^T\boldsymbol{d}_{N,\,k} \;=\; -\boldsymbol{F}(\boldsymbol{x}_k)^T\boldsymbol{F}(\boldsymbol{x}_k)\,. \tag{6.14}$$

Deste resultado, podemos observar que o teste (6.11) irá monitorar a redução da norma-M2 do resíduo (ou mapeamento não-linear), tal que: $\left\|\boldsymbol{F}(\boldsymbol{x}_k + \lambda_{k,\,j}\boldsymbol{d}_{N,\,k})\right\|_2 < (1 - \alpha\lambda_{k,\,j})\left\|\boldsymbol{F}(\boldsymbol{x}_k)\right\|_2$. O ângulo entre a correção de Newton e do *passo de máxima-descida* em, $\boldsymbol{x}_k$, pode ser calculado através da seguinte expressão

$$\cos\theta(-\nabla f_{\mathrm{M2}}(\boldsymbol{x}_k),\, \boldsymbol{d}_{N,\,k}) \;=\; \frac{\boldsymbol{F}(\boldsymbol{x}_k)^T\boldsymbol{F}(\boldsymbol{x}_k)}{\left\|\boldsymbol{J}^T(\boldsymbol{x}_k)\boldsymbol{F}(\boldsymbol{x}_k)\right\|_2\left\|\boldsymbol{J}^{-1}(\boldsymbol{x}_k)\boldsymbol{F}(\boldsymbol{x}_k)\right\|_2} \geq \frac{1}{\kappa_2(\boldsymbol{J}(\boldsymbol{x}_k))}\,, \tag{6.15}$$

onde $\kappa_2(\boldsymbol{J}) = \|\boldsymbol{J}\|_2\|\boldsymbol{J}^{-1}\|_2$ é o número de condicionamento de $\boldsymbol{J}$ em norma-M2 [17]. Conforme discutido em [164], o ângulo definido em (6.15), pode variar de 0 a 90 graus, tendo em vista que $\boldsymbol{J}^{-T}(\boldsymbol{x}_k)\boldsymbol{J}^{-1}(\boldsymbol{x}_k)$ e $\boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{J}^T(\boldsymbol{x}_k)$ são matrizes simétricas do tipo positiva definida compartilhando os mesmos auto-vetores e com recíprocos auto-valores. As direções são coincidentes apenas quando $\boldsymbol{F}(\boldsymbol{x}_k)$ é paralelo a um, e apenas um, dos auto-vetores de $\boldsymbol{J}^{-T}(\boldsymbol{x}_k)\boldsymbol{J}^{-1}(\boldsymbol{x}_k)$, uma rara situação mas de possível ocorrência. em um outro extremo, as direções tendem à ortogonalidade, quando $\boldsymbol{F}(\boldsymbol{x}_k)$ contém uma ampla distribuição de auto-vetores, e $\boldsymbol{J}^{-T}(\boldsymbol{x}_k)\boldsymbol{J}^{-1}(\boldsymbol{x}_k)$ possui uma grande divergência em auto-valores, correspondendo a um elevado número de condicionamento, ver (6.15). Infelizmente, esta última condição co-relaciona bem com o aumento dos erros de arredondamento em operações de ponto flutuante nos procedimentos de fatorização de matriz e retro-substituição. Isto resulta na perda de precisão na determinação da correção de Newton, podendo produzir um ângulo acima de 90 graus entre as direções em consideração. Neste extremo, não existirá nenhum valor para o fator de amortecimento que satisfaça o teste de monotonicidade (6.77).

Para remediar a situação descrita acima, foi proposta em [166], a *função nível natural* que é a função nível associada com correção de Newton, i.e., com a norma de *atualização-de-Newton* (aN), ou norma-aN [164]. Neste caso, temos que

$$\boldsymbol{A}_k = \boldsymbol{J}(\boldsymbol{x}_k)^{-1} \Rightarrow \begin{cases} f_{\mathrm{aN}}(\boldsymbol{x}, \boldsymbol{x}_k) \;=\; f(\boldsymbol{x} | \boldsymbol{J}(\boldsymbol{x}_k)^{-1}) \;=\; \frac{1}{2}\boldsymbol{F}(\boldsymbol{x})^T \boldsymbol{J}(\boldsymbol{x}_k)^{-T} \boldsymbol{J}(\boldsymbol{x}_k)^{-1} \boldsymbol{F}(\boldsymbol{x}) \\ \nabla f_{\mathrm{aN}}(\boldsymbol{x}, \boldsymbol{x}_k) \;=\; \nabla f(\boldsymbol{x} | \boldsymbol{J}(\boldsymbol{x}_k)^{-1}) \;=\; \boldsymbol{J}(\boldsymbol{x})^T \boldsymbol{J}(\boldsymbol{x}_k)^{-T} \boldsymbol{J}(\boldsymbol{x}_k)^{-1} \boldsymbol{F}(\boldsymbol{x}) \end{cases} . \tag{6.16}$$

O ponto estacionário, $\boldsymbol{x}_p$, definido anteriormente, é dado por $\boldsymbol{x}_p = \boldsymbol{x}_k$. As expressões em (6.16) definem a condição de escalamento natural para a correção de Newton. Introduzindo, $\boldsymbol{x} = \boldsymbol{x}_k + \lambda_{k,j} \boldsymbol{d}_{N,k}$, e utilizando (6.16), a função nível natural parametrizada em $\lambda_{k,j}$ é dada por

$$\hat{f}_{\mathrm{aN}}(\lambda_{k,j}) = f_{\mathrm{aN}}(\boldsymbol{x}_k + \lambda_{k,j}\boldsymbol{d}_{N,k}, \boldsymbol{x}_k) = \frac{1}{2}\left\| \boldsymbol{J}(\boldsymbol{x}_k)^{-1} \boldsymbol{F}(\boldsymbol{x}_k + \lambda_{k,j}\boldsymbol{d}_{N,k}) \right\|_2^2 = \frac{1}{2}\left\| \widehat{\boldsymbol{d}}_{N,k,j} \right\|_2^2 . \tag{6.17}$$

Como podemos observar esta função tem como peso a matriz jacobiana inversa, $\boldsymbol{J}(\boldsymbol{x}_k)^{-1}$, e o seu cálculo implica na resolução do seguinte sistema jacobiano

$$\boldsymbol{F}(\boldsymbol{x}_k + \lambda_{k,j}\boldsymbol{d}_{N,k}) + \boldsymbol{J}(\boldsymbol{x}_k)\widehat{\boldsymbol{d}}_{N,k,j} = \boldsymbol{0} . \tag{6.18}$$

Assumindo que a matriz jacobiana, $\boldsymbol{J}(\boldsymbol{x}_k)$, se encontra fatorizada na forma LU, então a solução do sistema acima requer apenas retro-substituições com complexidade $O(N^2)$.

Para a norma-aN, a taxa de inclinação utilizada no teste de monotonicidade (6.11) é dada por

$$\mathrm{tax\_inc}_{\mathrm{aN},k} = \nabla f_{\mathrm{aN}}(\boldsymbol{x}_k, \boldsymbol{x}_k)^T \boldsymbol{d}_{N,k} = -\boldsymbol{d}_{N,k}^T \boldsymbol{d}_{N,k} . \tag{6.19}$$

Isto significa que este teste irá monitorar a redução da norma-M2 de $\widehat{\boldsymbol{d}}_{N,k,j}$ em relação à norma-M2 da correção de Newton, $\left\| \widehat{\boldsymbol{d}}_{N,k,j} \right\|_2 < (1 - \alpha\lambda_{k,j}) \left\| \boldsymbol{d}_{N,k} \right\|_2$.

Com as considerações acima, o algoritmo de pesquisa-em-linha, entitulado Algoritmo BLS, pode ser, então, apresentado como se segue:

Algoritmo BLS
("Backtracking Line-Search")

(L-1) Dado: $\boldsymbol{x}_k \in \mathbf{R}^N$, $\boldsymbol{d}_{N,k} \in \mathbf{R}^N$, $\alpha \in \mathbf{R}$, $\lambda_0, \lambda_{\mathrm{mín}} \in \mathbf{R}$ e $NIP_{\mathrm{máx}} \geq 0$;
(L-2) Calcular: $\nabla f(\boldsymbol{x}_k | \boldsymbol{A}_k) = \boldsymbol{J}(\boldsymbol{x}_k)^T \boldsymbol{F}(\boldsymbol{x}_k)$ (gradiente);
(L-3) Calcular: $\mathrm{tax\_inc}_k = \nabla f(\boldsymbol{x}_k | \boldsymbol{A}_k)^T \boldsymbol{d}_{N,k}$ (taxa de inclinação);
(L-4) **PARA** $j \leftarrow 0$ **PASSO** 1 **ATÉ** $NIP_{\mathrm{máx}}$ **FAÇA**:
(L-5) **SE** $(f(\boldsymbol{x}_k + \lambda_{k,j}\boldsymbol{d}_{N,k} | \boldsymbol{A}_k) \leq f(\boldsymbol{x}_k | \boldsymbol{A}_k) + \alpha\lambda_{k,j}\mathrm{tax\_inc}_k)$ **ENTÃO** $\lambda_k \leftarrow \lambda_{k,j}$ **RETORNE**.
(L-6) **CASO CONTRÁRIO**
(L-7) Para dividindo-pela-metade: $\lambda_{\mathrm{temp}} \leftarrow \lambda_{k,j}/2$, ou
para interpolação: $\lambda_{\mathrm{temp}} \leftarrow \mathrm{quad}(\lambda_{k,j}, f(\boldsymbol{x}_k + \lambda_{k,j}\boldsymbol{d}), f(\boldsymbol{x}_k), \mathrm{tax\_inc})$;
(L-8) $\lambda_{k,j} \leftarrow \max\{\lambda_{\mathrm{temp}}, \lambda_{k,j}/10\}$;
(L-9) $\lambda_{k,j} \leftarrow \max\{\lambda_{k,j}, \lambda_{\mathrm{mín}}\}$.

A condição de convergência, na linha (L-5) do algoritmo acima, é ativada quando o teste de monotonicidade (6.11) é satisfeito. Por outro lado, para ativar a condição de divergência devemos ter $j > NIP_{\text{máx}}$, onde $NIP_{\text{máx}}$ é o máximo número de iterações na pesquisa-em-linha. Esta condição pode ser utilizada para sinalizar divergência no método de Newton. Na linha (L-7) são indicadas as opções para determinação do fator de amortecimento, nomeando, dividindo-pela-metade ou interpolação quadrática [46]. Para evitar que o fator de amortecimento se torne muito pequeno, foram introduzidas as proteções nas linhas (L-8) e (L-9) (ver Tabela 6.1 para seleção de $\lambda_{\text{mín}}$). Para o cálculo do fator de amortecimento utilizando a opção dividindo-pela-metade, a linha (L-8) não possui nenhum efeito.

Apesar de não implementado neste trabalho, conforme introduzido em [172], existe a possibilidade de generalização de (6.11) no seguinte teste não-monotônico,

$$f(\boldsymbol{x}_k + \lambda_{j,k}\boldsymbol{d}_{N,k}) \leq \max_{0 \leq p \leq m(k)} [f(\boldsymbol{x}_{k-p})] + \alpha \cdot \lambda_{k,j} \cdot \text{tax\_inc}_k \tag{6.20}$$

onde

$$m(0) = 0,\ 0 \leq m(k) \leq \text{mín}[m(k-1)+1, M] \text{ para } k > 0, \tag{6.21}$$

e para $M \in \mathbb{Z}_+$. Testes numericos revelam que, quando comparado com a sua versão monotônica ($M = 0$), o teste (6.20) para $M = 5$ ou $M = 10$ é geralmente competitivo, e na maioria dos casos mais eficiente. A aplicação desta técnica na análise de CC e do BH em circuitos eletrônicos pode ser encontrada em [173]. A implementação do teste não-monotônico é relativamente simples.

6.4.2. Implementação Modificada

Uma versão modificada do processo iterativo (6.9.a)-(6.9.d) (método clássico de Newton), equipado com a globalização de pesquisa-em-linha, via retrocedimento, foi implementado incluindo uma *malha interna* dentro da *malha externa*. Na malha interna, cada nova correção, denominada de *correção de Newton simplificada*, é calculada sem atualização da matriz jacobiana. O algoritmo GN, descrito abaixo, foi desenvolvido com base nesta modificação e na estrutura proposta em [19].

Algoritmo GN
("Global Newton")

(L-1) Dado: $\boldsymbol{x}_0 \in \mathbb{R}^N$, $\Xi \in [0, 1)$, $NI_{\text{máx}}, NIJ_{\text{máx}} > 0$ e $NIM_{\text{máx}} \geq 0$;
(L-2) $k \leftarrow 0$;
(L-3) **PARA** $i \leftarrow 0$ **PASSO** 1 **ATÉ** $NIJ_{\text{máx}}$ **FAÇA**: (malha externa)
(L-4) Calcular e fatorar: $\boldsymbol{J}_i = \boldsymbol{J}(\boldsymbol{x}_k) \in \mathbb{R}^{N \times N}$; (matriz jacobiana)

(L-5) Determinar: $\boldsymbol{d}_{N,k}$ tal que $\boldsymbol{M}_N(\boldsymbol{x}_k + \boldsymbol{d}_{N,k}) = \boldsymbol{0}$;

(L-6) **PARA** $j \leftarrow 0$ **PASSO** 1 **ATÉ** $NIM_{\text{máx}}$ **FAÇA**: (malha interna)

(L-7) Determinar: $\lambda_k \in \mathbb{R}$ via Algoritmo BLS;

(L-8) $\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_{N,k}$;

(L-9) **SE** ("CONVERGÊNCIA" **OU** "DIVERGÊNCIA") **ENTÃO** $\boldsymbol{x}_* \leftarrow \boldsymbol{x}_{k+1}$; **RETORNE**.

(L-10) **SE** ($NIM_{\text{máx}} > 0$) **ENTÃO**

(L-11) **SE** ($\Xi = 0$ **OU** $f(\boldsymbol{x}_{k+1})/f(\boldsymbol{x}_k) < \Xi$) **ENTÃO**

(L-12) Determinar: $\bar{\boldsymbol{d}}_{N,k}$ tal que $\bar{\boldsymbol{M}}_N(\boldsymbol{x}_{k+1} + \bar{\boldsymbol{d}}_{N,k}) = \boldsymbol{0}$ †;

(L-13) $\boldsymbol{d}_{N,k} \leftarrow \bar{\boldsymbol{d}}_{N,k}$; (utilizar correção simplificada)

(L-14) **CASO CONTRÁRIO**

(L-15) **SE** ($f(\boldsymbol{x}_{k+1})/f(\boldsymbol{x}_k) > 1$) **ENTÃO FIM PARA**;

(L-16) $k \leftarrow k + 1$.

---

† $\bar{\boldsymbol{M}}_N(\boldsymbol{x}_{k+1} + \bar{\boldsymbol{d}}_N) = \boldsymbol{F}(\boldsymbol{x}_{k+1}) + \boldsymbol{J}_i \bar{\boldsymbol{d}}_N$.

Para o teste de convergência na linha (L-9), no algoritmo acima, foi adotado o seguinte critério, $\|\boldsymbol{F}(\boldsymbol{x}_k)\| < \varepsilon_F$, onde ε_F é uma tolerância de parada. Um outro critério, que também pode ser utilizado é $\|\boldsymbol{F}(\boldsymbol{x}_k)\| < \varepsilon_{F0}\|\boldsymbol{F}(\boldsymbol{x}_0)\|$ [34]. Já a divergência é detectada nas seguintes situações: a primeira, quando é atingido o número máximo de iterações, $k = NI_{max}$; e a sgunda, quando a norma do vetor de correção é menor do que uma determinada tolerância de parada, ε_x [34].

Se fizermos $NIM_{\text{máx}} = 0$, no algoritmo acima, obtemos o método de Newton convencional, i.e., sem correções simplificadas. No outro extremo, gerando apenas correções simplificadas, $\Xi = 0$ e $NIM_{\text{máx}} = NI_{\text{máx}} > 0$, obtemos o método de *corda-paralela*, com a matriz jacobiana calculada no ponto inicial e mantida constante em todas as iterações [17]. Para $0 < NIM_{\text{máx}} << NI_{\text{máx}}$ obtemos o método de *Newton-Shamanskii* [17], sem e com chaveamento para fora da malha interna (L-6)-(L-15), $\Xi = 0$ e $0 < \Xi < 1$, respectivamente. O parâmetro Ξ determina a mínima razão de redução da função nível para se manter na malha interna de correções simplificadas. Em muitas aplicações, este parâmetro pode ser tipicamente selecionado em torno de $0{,}5 - 0{,}95$. Como podemos observar, a matriz jacobiana é mantida constante durante as iterações simplificadas. Na linha (L-7), o emprego da globalização, via pesquisa-em-linha para amortecimento do passo (ou correção), pode ser eliminado fazendo $NPL_{\text{máx}} = 0$ no algoritmo BLS. Uma outra possibilidade, para formulação da correção de Newton simplificada, está fundamentada na fórmula de atualização de posto-dois de *Davidon-Fletcher-Powell* (DFP), desenvolvida para utilização com a matrix jacobiana fatorizada na forma LU [17]. Porém, esta técnica requer uma área extra de mémoria para armazenagem de vetores de trabalho, e a aplicação de uma sequência de operações aritiméticas com estes vetores, que cresce linearmente a cada iteração simplificada.

No contexto da análise do BH, se assumirmos que matriz jacobiana é calculada no ponto de iteração inicial, utilizando um espectro de derivada contendo apenas a componente de frequência de CC, e mantida constante durante todas as iterações. Então, o algoritmo GN (operando como método das cordas) produzirá iterações equivalentes ao método *cêntrico-linear* apresentado em [20]. A análise e a otimização da distorção por intermodulação (DIM) em circuitos não-lineares via BH, tipicamente resulta em um problema de grande-escala, onde este método pode ser a única alternativa viável.

6.5. Método do Tensor

No trabalho apresentado em [29], foi proposto um novo método para solução de sistemas de equações não-lineares, que fundamenta cada correção, $\boldsymbol{d}$, na solução de um modelo com informação de segunda-ordem, representando uma aproximação de $\boldsymbol{F}$ na vizinhança de $\boldsymbol{x}_k$. Este novo modelo local, pode ser visto como uma extensão do modelo linear (6.7), onde foi adicionado um novo termo quadrático, dado por

$$\boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}) \;=\; \boldsymbol{F}(\boldsymbol{x}_k) + \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d} + \frac{1}{2}\boldsymbol{T}(\boldsymbol{x}_k)\boldsymbol{d}^2\,, \tag{6.22}$$

onde $\boldsymbol{T}(\boldsymbol{x}_k) \in \mathbb{R}^{N\times N\times N}$ é um objeto tri-dimensional frequentemente referido como tensor, e com sua composição descrita em [29]. Por esta razão, (6.22) é descrito como modelo do tensor, e os métodos fundamentados nele de *métodos do tensor*. Idealmente, seria interessante utilizar o tensor de derivadas de segunda-ordem, em (6.22), porém sua geração, armazenagem e solução teriam um custo inaceitável por iteração. Não obstante, estas dificuldades podem ser superadas escolhendo $\boldsymbol{T}_k$ com uma forma de baixo posto, gerada através de um processo interpolativo envolvendo o mapeamento não-linear (ou a matriz jacobiana), a ser discutido abaixo.

Em geral, o processo iterativo básico descrevendo o método do tensor, pode ser descrito da seguinte forma

Dado: $\boldsymbol{x}_0 \in \mathbb{R}^N$ (6.23.a)

Para $k \leftarrow 0$ passo 1 até "convergência" faça: (6.23.b)

Encontrar: $\boldsymbol{d}_{N,\,k} \in \mathbb{R}^N$ tal que $\boldsymbol{M}(\boldsymbol{x}_k + \boldsymbol{d}_{N,\,k}) \;=\; 0$ (6.23.c)

Encontrar: $\boldsymbol{d}_{T,\,k} \in \mathbb{R}^N$ que minimiza $\left\|\boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}_{T,\,k})\right\|_2$ (6.23.d)

Escolher: $\boldsymbol{d}_k$ entre $\boldsymbol{d}_{N,\,k}$ e $\boldsymbol{d}_{T,\,k}$ (6.23.e)

$\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \boldsymbol{d}_k$ (6.23.f)

Para que o processo iterativo acima seja eficiente, quando comparado com (6.9.a)-(6.9.d), a

construção e a solução do modelo do tensor, em (6.23.d), deve exigir o mínimo de recursos de memória e tempo de processamento. Convém ressaltar que a determinação da correção do tensor pode resultar em um problema de minimização, pois o modelo do tensor pode não ter raiz. Outro aspecto importante, é a implementação de (6.23.e) utilizando um eficiente critério para seleção da correção entre $\boldsymbol{d}_N$ e $\boldsymbol{d}_T$.

A análise de convergência local desenvolvida em [174], sobre leve suposições, estabelece vantagens teóricas dos métodos do tensor sobre o método de Newton, nos problemas em que a matriz jacobiana na raiz possui uma pequena deficiência de posto. A análise também demonstra que, quando a matriz jacobiana é não-singular na raiz, o método do tensor converge pelo menos quadraticamente. Em adição, como resultado das diversas experiências numéricas reportadas em [29] pode-se afirmar que o método do tensor, a ser discutido abaixo, é muito frequentemente mais eficiente que o método de Newton, e nunca significantemente menos eficiente, principalmente em problemas cuja a matriz jacobiana é singular ou mal condicionada na raiz. Estes resultados motivaram a investigação do seu desempenho com relação ao método de Newton quando aplicado na análise de regime CC e do BH (em pequena escala).

6.5.1. Constução do Modelo Tensor

Conforme discutido acima, o termo quadrático no modelo do tensor (6.22) não contém informação de derivadas de segunda-ordem; ao invés disso, o modelo é construído impondo a condição de interpolação do mapeamento não-linear $\boldsymbol{F}(\boldsymbol{x})$, ou de sua associada matriz jacobiana $\boldsymbol{J}(\boldsymbol{x})$, em pontos passados, não necessariamene consecutivos. Se utilizarmos a condição de interpolação do mapeamento não-linear, assumimos que o modelo do tensor satisfaz

$$\boldsymbol{F}(\boldsymbol{x}_{-j,\,k}) \;=\; \boldsymbol{F}(\boldsymbol{x}_k) + \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{s}_j + \frac{1}{2}\boldsymbol{T}(\boldsymbol{x}_k)\boldsymbol{s}_j\boldsymbol{s}_j \qquad j \;=\; 1,\, \ldots,\, P \tag{6.24}$$

onde

$$\boldsymbol{s}_{j,\,k} \;=\; \boldsymbol{x}_{-j,\,k} - \boldsymbol{x}_k \qquad j \;=\; 1,\, \ldots,\, P\,. \tag{6.25}$$

Seguindo as argumentações apresentadas em [63], a história de pontos passados não deve exceder $P \le \sqrt{N}$. Se introduzirmos, os vetores

$$\boldsymbol{z}_{j,\,k} \;=\; 2(\boldsymbol{F}(\boldsymbol{x}_{-j,\,k}) - \boldsymbol{F}(\boldsymbol{x}_k) - \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{s}_{j,\,k}) \qquad j \;=\; 1,\, \ldots,\, P \tag{6.26}$$

então o sistema (6.24) pode ser escrito mais compactamente como

$$\boldsymbol{T}(\boldsymbol{x}_k)\boldsymbol{s}_{j,\,k}\boldsymbol{s}_{j,\,k} \;=\; \boldsymbol{z}_{j,\,k} \qquad j \;=\; 1,\, \ldots,\, P\,. \tag{6.27}$$

Adotando-se a prática padrão, e bem suscedida, utilizada em métodos da secante para solução de equações não-lineares e problemas de otimização (ver [46]), o objeto tensor $\boldsymbol{T}(\boldsymbol{x}_k)$ pode ser gerado através da solução do seguinte problema [29]

$$\begin{aligned} &\text{minimizar}_{\boldsymbol{T}(\boldsymbol{x}_k) \in \mathbb{R}^{N \times N \times N}} \left\| \boldsymbol{T}(\boldsymbol{x}_k) \right\|_F \\ &\text{tal que } \boldsymbol{T}(\boldsymbol{x}_k)\boldsymbol{s}_{j,k}\boldsymbol{s}_{j,k} = \boldsymbol{z}_{j,k} \qquad j = 1, \ldots, P \end{aligned} \tag{6.28}$$

onde $|\Rightarrow\|_F$ é a norma de Frobenius [157]. Neste ponto, é conviniente definirmos a matriz $\boldsymbol{Z}_k \in \mathbb{R}^{N \times P}$ cujas as linhas são formadas pelos vetores $\boldsymbol{z}_{j,k}$'s, e a matriz $\boldsymbol{M}_k \in \mathbb{R}^{P \times P}$ na qual $[\boldsymbol{M}_k]_{i,j} = (\boldsymbol{s}_{i,k}^T \boldsymbol{s}_{j,k})^2$, $1 \le i, j \le P$. Conforme demonstrado em [29], a solução do problema (6.28), consiste na somatória de P tensores de posto-um, cujas faces horizontais são simétricas e dadas por

$$\boldsymbol{T}(\boldsymbol{x}_k)[l, m, n] = \sum_{j=1}^{P} \boldsymbol{a}_{j,k}[l]\boldsymbol{s}_{j,k}[m]\boldsymbol{s}_{j,k}[n] \qquad 1 \le l, m, n \le N, \tag{6.29}$$

onde $\boldsymbol{a}_{j,k}$ é a j-ésima coluna de $\boldsymbol{A}_k \in \mathbb{R}^{N \times P}$ com $\boldsymbol{A}_k = \boldsymbol{Z}_k \boldsymbol{M}_k^{-1}$ e $\boldsymbol{S}_k$ é a matriz cujas colunas são formadas pelos vetores $\boldsymbol{s}_j$'s. Então, notando que $\boldsymbol{d}^2 \equiv \boldsymbol{d}\boldsymbol{d}$, e utilizando a definição do produto vetor com objeto tensor, podemos finalmente escrever o modelo tensor como

$$\boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}) = \boldsymbol{F}(\boldsymbol{x}_k) + \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d} + \frac{1}{2}\boldsymbol{A}_k(\boldsymbol{S}_k^T \boldsymbol{d})^2. \tag{6.30}$$

Na expressão acima, foi introduzida a notação $(\boldsymbol{v})^2$, onde $\boldsymbol{v} \in \mathbb{R}^M$ denota um vetor $\boldsymbol{w} \in \mathbb{R}^M$ tal que $w[i] = v[i]^2$, $i = 1, \ldots, M$.

Da expressão (6.30) podemos facilmente observar que utilizando interpolação com um ponto passado ($P = 1$), o termo quadrático se reduz para $\boldsymbol{a}_k(\boldsymbol{s}_k^T \boldsymbol{d})^2$, e o seu cálculo envolve um produto interno (ou escalar) e um produto escalar-vetor. O modelo tensor (6.30) se reduz à seguinte forma

$$\boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}) = \boldsymbol{F}(\boldsymbol{x}_k) + \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d} + \frac{1}{2}\boldsymbol{a}_k(\boldsymbol{s}_k^T \boldsymbol{d})^2, \tag{6.31}$$

onde

$$\boldsymbol{s}_k = \boldsymbol{x}_{k-1} - \boldsymbol{x}_k, \tag{6.32}$$

$$\boldsymbol{a}_k = \frac{2}{m_k}(\boldsymbol{F}(\boldsymbol{x}_{k-1}) - \boldsymbol{F}(\boldsymbol{x}_k) - \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{s}_k), \tag{6.33}$$

$$m_k = (\boldsymbol{s}_k^T \boldsymbol{s}_k)^2. \tag{6.34}$$

Conforme discutiremos a seguir, a expansão de segunda-ordem de posto-um (6.31), é a mais simples representação do modelo do tensor, na qual a solução, i.e., a correção do tensor, pode ser obtida analiticamente.

6.5.2. Solução do Modelo Tensor

Em geral, conforme discutido em [29], o modelo do tensor (6.30) pode não possuir uma raiz. Desta forma, sua resolução deve ser considerada como um problema de minimização sem restrição que pode ser enunciado como se segue

$$\min_{\boldsymbol{d} \in \mathrm{R}^N} \left\| \boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}) \right\|_2 . \tag{6.35}$$

Na situação em que o modelo tensor consiste de um sistema denso de equações de pequena-escala, a solução pode ser eficientemente calculada utilizando o método da fatorização QR, ver [29]. O custo aproximado para determinação da solução é de $2n^3/3 + n^2 p + O(n^2)$ operações aritiméticas. Este procedimento possui a virtude de ser numericamente estável mesmo quando a matriz jacobiana é singular ou mal-condicionada. Para a solução eficiente do modelo do tensor constituído de um sistema esparso de equações de pequena- a média-escala, pode ser adotado o procedimento proposto em [30]. Para ser breve, destacaremos apenas os passos básicos deste procedimento adequado para uma matriz jacobiana esparsa não-singular.

Para iniciarmos o processo de solução esparsa de (6.35), convém introduzir a seguinte matriz ortogonal

$$\boldsymbol{Q}^T = \left[\boldsymbol{U}\ \boldsymbol{Z} \right] \in \mathrm{R}^{N \times N} \tag{6.36}$$

onde

- $\boldsymbol{U} \in \mathrm{R}^{N \times P}$ é igual a $\boldsymbol{J}(\boldsymbol{x}_k)^{-T} \boldsymbol{S}_k \boldsymbol{W}_k^{-1/2}$ onde $\boldsymbol{W}_k = \boldsymbol{S}_k^T (\boldsymbol{J}(\boldsymbol{x}_k)^T \boldsymbol{J}(\boldsymbol{x}_k))^{-1} \boldsymbol{S}_k = \boldsymbol{L}\boldsymbol{L}^T$. Esta relação resulta de uma escolha judiciosa a ser discutida abaixo.
- $\boldsymbol{Z} \in \mathrm{R}^{N \times (N-P)}$ é uma base ortonormal para o subespaço ortogonal medido pelas colunas da matriz $\boldsymbol{J}_k^{-T} \boldsymbol{S}_k$.

Da teoria básica de algebra linear sabemos que, se aplicarmos a transformação ortogonal (6.36) em (6.35), o problema de minimização permanece inalterado, e pode ser re-escrito da seguinte forma

$$\min_{\boldsymbol{d} \in \mathbf{R}^N} \left(\left\| \boldsymbol{U}^T \boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}) \right\|_2 + \left\| \boldsymbol{Z}^T \boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}) \right\|_2 \right) . \tag{6.37}$$

Da definição da matriz $\boldsymbol{Z}$, da ortogonalidade de $\boldsymbol{Q}$ e do conceito de matriz de projeção, podemos escrever

$$\boldsymbol{Z}^T \boldsymbol{J}_k^{-T} \boldsymbol{S}_k = \mathbf{0} \tag{6.38}$$

e

$$ZZ^T = \mathbf{1} - UU^T \tag{6.39}$$

respectivamente. Para encontrarmos a solução de (6.37) devemos fazer $\boldsymbol{Z}^T\boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}) = \boldsymbol{0}$ e assumir uma solução na forma $\boldsymbol{d} = \boldsymbol{d}_1 + \xi\boldsymbol{d}_2$. Sendo assim, com a introdução do vetor $\beta = \boldsymbol{S}_k^T\boldsymbol{d} \in \mathrm{R}^P$, e das relações (6.38) e (6.39), pode-se demonstrar que (6.37) se reduz em

$$\min_{\beta \in \mathrm{R}^P} \left\| \boldsymbol{L}^{-1}\boldsymbol{q}_k(\beta) \right\|_2, \tag{6.40}$$

e a correção do tensor que minimiza (6.40) é dada por

$$\boldsymbol{d}_{T,k} = \boldsymbol{J}_k^{-1}\boldsymbol{J}_k^{-T}\boldsymbol{S}_k\boldsymbol{W}_k^{-1}\boldsymbol{q}(\beta) - \boldsymbol{J}_k^{-1}(\boldsymbol{F}_k + \frac{1}{2}\boldsymbol{A}_k\beta^2), \tag{6.41}$$

onde

$$\boldsymbol{q}_k(\beta) = \boldsymbol{S}_k^T\boldsymbol{J}(\boldsymbol{x}_k)^{-1}\boldsymbol{F}_k + \beta + \frac{1}{2}\boldsymbol{S}_k^T\boldsymbol{J}(\boldsymbol{x}_k)^{-1}\boldsymbol{A}_k\beta^2. \tag{6.42}$$

Importante lembrar que $\beta^2 \in \mathrm{R}^P$ denota o vetor cuja *i*-ésima componente é dada por β_i^2. Como podemos observar, a determinação da correção do tensor requer a solução de um problema P-dimensional de minimização sem restrição. O procedimento discutido acima é válido apenas quando a matriz jacobiana possui posto completo, e pode ser visto como um caso particular do procedimento desenvolvido para sistemas de mínimos-quadrados não-lineares [30].

Se considerarmos os métodos do tensor utilizando interpolação com apenas um ponto passado, $P = 1$, (6.40) se reduz a um problema uni-dimensional e, consequentemente pode ser resolvido analiticamente. Neste caso, os objetos do tipo matriz e vetor, relacionados ao termo quadrático do tensor se reduzem a objetos do tipo vetor e escalar, respectivamente, tal que

$$q_k(\beta) = c_{0,k} + \beta + c_{2,k}\beta^2 \tag{6.43}$$

onde

$$c_{0,k} = -\boldsymbol{s}_k^T\boldsymbol{d}_{N,k}, \tag{6.44.a}$$

$$c_{2,k} = -\frac{1}{2}\boldsymbol{s}_k^T\boldsymbol{u}_k. \tag{6.44.b}$$

No cálculo dos coeficientes acima, $\boldsymbol{d}_{N,k}$ denota a correção de Newton, já definida em (6.8), enquanto o vetor $\boldsymbol{u}_k$ é obtido como solução do seguinte sistema jacobiano

$$\boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{u}_k = -\boldsymbol{a}_k. \tag{6.45}$$

Nesta condição, $P = 1$ e, utilizando (6.43) e (6.45), demonstra-se facilmente que a expressão (6.41), referente à correção do tensor, assume a seguinte forma

$$\boldsymbol{d}_{T,k} = \boldsymbol{d}_{N,k} + \frac{1}{2}\beta_k^2\boldsymbol{u}_k + \nu_k\boldsymbol{z}_k \tag{6.46}$$

onde

$$\beta_k = \mathrm{argmin}_{\beta \in \mathrm{R}} |q_k(\beta)|, \tag{6.47}$$

$$\nu_k = w_k^{-1} q_k(\beta_k), \tag{6.48}$$

$$w_k = \boldsymbol{y}_k^T \boldsymbol{y}_k, \tag{6.49}$$

$$\boldsymbol{J}(\boldsymbol{x}_k)^T \boldsymbol{y}_k = \boldsymbol{s}_k, \tag{6.50.a}$$

$$\boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{z}_k = \boldsymbol{y}_k. \tag{6.50.b}$$

A computação de (6.50.a) e (6.50.b) não é requerida quando o modelo do tensor, i.e., $q_k(\beta)$, possuir pelo menos uma raiz real. Neste caso, $\nu_k = 0$, o cálculo da correção do tensor necessita apenas do cálculo de (6.45), via retro-substituições LU (complexidade $O(N^2)$). Caso contrário, precisamos determinar também (6.50.a) e (6.50.b), via retro-substituições LU. Se o polinômio $q_k(\beta)$ possuir raízes reais e distintas, será selecionada a raiz de menor valor absoluto. Esta escolha conservativa mantém a correção do tensor mais próxima da correção de Newton.

Quando a iteração está próxima da solução, o modelo do tensor, em aproximadamente todos os casos, possuirá uma raiz real. Sendo assim, os casos nos quais o modelo do tensor possui apenas raízes complexas, usualmente ocorrem quando o solucionador não-linear está longe da solução e executando grandes passos (ou correções). Nesta situação, o termo do tensor de segunda-ordem, $\boldsymbol{a}_k$, que é formado por uma aproximação secante, e depende da norma da correção (ou tamanho do passo), é muito provavelmente uma aproximação pobre. Como resultado, o modelo do tensor local pode desviar consideravelmente do mapeamento não-linear, $\boldsymbol{F}(\boldsymbol{x})$, em torno de $\boldsymbol{x}_k$ e, consequentemente, $M_T(\boldsymbol{x}_k + \boldsymbol{d})$ pode não ter mais uma raiz real.

A intuição nos sugere que, quando o modelo local está significantemente errado, devemos recuar e ser menos agressivos na tentativa de modelar o mapeamento não-linear ao redor de $\boldsymbol{x}_k$. Assim sendo, foi sugerido em [35], atenuar o modelo do tensor local através do escalamento da informação de segunda-ordem, para produzir um comportamento próximo a um modelo linear (o mesmo que define a correção de Newton). Para realizarmos esta função, multiplicamos o termo do tensor $\frac{1}{2}\boldsymbol{a}_k(\boldsymbol{s}_k^T\boldsymbol{d})^2$ por um parâmetro escalar, τ, de tal forma que

$$M_T(\boldsymbol{x}_k + \boldsymbol{d}) = \boldsymbol{F}(\boldsymbol{x}_k) + \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d} + \frac{1}{2}\tau\boldsymbol{a}_k(\boldsymbol{s}_k^T\boldsymbol{d})^2. \tag{6.51}$$

Este parâmetro é inicialmente igual a um, o que implica na solução do modelo do tensor padrão (6.31), mas se uma raiz real não for encontrada, então escolhemos $\tau \in (0, 1]$, de tal forma que este modelo possua uma raiz real. Isto é obtido facilmente com a formação do polinômio quadrático parametrizado

$$q_k(\beta, \tau) = c_{0,k} + \beta + \tau c_{2,k}\beta^2, \tag{6.52}$$

onde os coeficientes deste polinômino são dados por (6.44.a) e (6.44.b). Se o polinômio quadrático acima não tiver uma raiz real para $\tau = 1$, então o valor de τ que admite uma única raiz real é dado

por

$$\tau = \frac{1}{4c_{0,k}c_{2,k}}. \tag{6.53}$$

Neste caso, para que $\delta = 1 - 4c_{0,k}c_{2,k}$ seja menor que zero (raízes complexas), devemos ter $4c_{0,k}c_{2,k} > 1$ o que, consequentemente, resulta em $\tau < 1$. Sob estas condicões, a correção do tensor (6.46) é simplesmente dada por: $\boldsymbol{d}_T = \boldsymbol{d}_N + \frac{1}{2}\tau\beta_*^2\boldsymbol{u}$, onde

$$\beta_* = -2c_{0,k}. \tag{6.54}$$

Pode-se demonstrar que apesar da exclusão do último termo em (6.46), a correção do tensor ainda retém as mesmas propriedades de convergência local do método do tensor, com resolução completa do seu associado modelo local [114]. A economia produzida com o procedimento acima resulta da eliminação da solução dos sistemas (6.50.a) e (6.50.b) via retro-substituição LU com complexidade $O(N^2)$. As vantagens deste procedimento são significantes no desenvolvimento das versões inexatas do método do tensor, conforme será descrito no próximo capítulo.

6.5.3. Globalização via Pesquisa-em-Linha

A seguir, serão discutidas duas diferentes estratégias de pesquisa-em-linha, particularmente desenvolvidas para o método do tensor. Estas estratégias, entituladas de pesquisa-em-linha *padrão* e de pesquisa-em-linha *curvilinear*, foram originalmente propostas em [33] e [161], respectivamente. Além destas estratégias, vale citar o algoritmo de pesquisa-emlinha proposto em [29], e que segue a mesma filosofia da estratégia padrão considerada.

6.5.3.a. Estratégia Padrão

Na estratégia padrão, a direção do tensor possui a preferência, porém se esta direção não satisfizer um apropriado teste de monotonicidade, e não for uma direção suficientemente de descida, uma pesquisa-em-linha extra na direção de Newton será efetuada. Adicionalmente, uma pesquisa-em-linha na direção do tensor é realizada. O algoritmo BLS-TS, descrito abaixo, representa a realização da estratégia padrão neste trabalho, e segue a mesma implementação utilizada no programa TENSOLVE [33].

Algoritmo BLS-TS
("Backtracking Line-Search - Tensor Standard")

(L-1) Dado: $\alpha, \gamma \in (0, 1)$ $(\alpha = \gamma = 10^{-4})$; $\boldsymbol{x}_k, \boldsymbol{d}_{T,k}, \boldsymbol{d}_{N,k} \in \mathbb{R}^N$;

(L-2) Calcular: $\nabla f(\boldsymbol{x}_k) = \boldsymbol{J}(\boldsymbol{x}_k)^T \boldsymbol{F}(\boldsymbol{x}_k)$ (gradiente);

(L-3) Calcular: $\text{tax\_inc}_k = \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_{N,k}$ (taxa de inclinação);

(L-4) **SE** $(f(\boldsymbol{x}_k + \boldsymbol{d}_{T,k}) < f(\boldsymbol{x}_k) + \alpha \text{tax\_inc}_k)$ **ENTÃO**

(L-5) $\boldsymbol{d}_k \leftarrow \boldsymbol{d}_{T,k}$; $\lambda_k \leftarrow 1$;

(L-6) **CASO CONTRÁRIO**

(L-7) Determinar: $\lambda_{N,k} \in (0, 1]$ na direção $\boldsymbol{d}_{N,k} \in \mathbb{R}^N$ via o algoritmo BLS;

(L-8) **SE** $(\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_{T,k} \geq -\gamma \|\nabla f(\boldsymbol{x}_k)\|_2 \|\boldsymbol{d}_{T,k}\|_2)$ **ENTÃO**

(L-9) $\boldsymbol{d}_k \leftarrow \boldsymbol{d}_{N,k}$; $\lambda_k \leftarrow \lambda_{N,k}$;

(L-10) **CASO CONTRÁRIO**

(L-11) Determinar: $\lambda_{T,k} \in (0, 1]$ na direção $\boldsymbol{d}_{T,k} \in \mathbb{R}^N$ via o algoritmo BLS;

(L-12) **SE** $f(\boldsymbol{x}_k + \boldsymbol{d}_{N,k}) < f(\boldsymbol{x}_k + \boldsymbol{d}_{T,k})$ **ENTÃO**

(L-13) $\boldsymbol{d}_k \leftarrow \boldsymbol{d}_{N,k}$; $\lambda_k \leftarrow \lambda_{N,k}$;

(L-14) **CASO CONTRÁRIO**

(L-15) $\boldsymbol{d}_k \leftarrow \boldsymbol{d}_{T,k}$; $\lambda_k \leftarrow \lambda_{T,k}$.

O teste, na linha (L-4) do algoritmo acima, é equivalente ao teste de monotonicidade (6.11) em norma-M2, quando $\lambda_{k,j} = 1$. Se este teste for satisfeito, a correção do tensor é selecionada. Assumindo que o modelo do tensor tenha pelo menos uma raiz real, então, assim como na correção de Newton, o cálculo da correção do tensor necessita apenas da solução, via retro-substituições, de um sistema jacobiano fatorizado na forma LU. Caso a correção do tensor não seja aprovada, outros testes devem são realizados, para selecionar a melhor direção entre as direções de Newton e do tensor. O primeiro destes testes, na linha (L-8), verifica se a correção do tensor está em uma direção de suficiente descida, i.e, se o ângulo entre a correção do tensor e o seu gradiente está abaixo de um certo máximo definido pelo parâmetro γ. Como podemos observar, a sua condução necessita do cálculo do gradiente da função nível em $\boldsymbol{x}_k$. Se este teste falhar, então um segundo e último teste, ver (L-12), selecionará a correção que proporciona uma maior redução da função nível.

6.5.3.b. Estratégia Curvilinear

Uma das principais vantagens da estratégia curvilinear, proposta em [161], é a eliminação de uma eventual pesquisa-em-linha na direção da correção de Newton, necessária na estratégia padrão descrita anteriormente. Para tal, vamos considerar a seguinte modificação no modelo tensor,

$$\boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}) = \lambda_{k,j} \boldsymbol{F}_k + \boldsymbol{J}_k \boldsymbol{d} + \frac{1}{2} \boldsymbol{a}_k (\boldsymbol{s}_k^T \boldsymbol{d})^2, \tag{6.55}$$

onde podemos observar a presença do fator de amortecimento, $\lambda_{k,j}$, variando com j contador de

iterações de pesquisa-em-linha. Seguindo o mesmo processo utilizado no cálculo da correção do tensor (6.46), podemos escrever a correção do tensor curvilinear, associada ào modelo (6.55), como se segue

$$\boldsymbol{d}_T(\lambda_{k,j}) = \lambda_{k,j}\boldsymbol{d}_{N,k} + \frac{1}{2}\beta(\lambda_{k,j})^2\boldsymbol{u}_k + w_k^{-1}q_k(\beta(\lambda_{k,j}), \lambda_{k,j})\boldsymbol{z}_k \ , \tag{6.56}$$

onde

$$q_k(\beta, \lambda) = -\lambda \boldsymbol{s}_k^T\boldsymbol{d}_{N,k} + \beta - \frac{1}{2}\boldsymbol{s}_k^T\boldsymbol{u}_k\beta^2 = \lambda c_{0,k} + \beta + c_{2,k}\beta^2 \tag{6.57}$$

e $\beta(\lambda_{k,j}) = \text{argmin}_{\beta \in \mathbf{R}}\left|q(\beta, \lambda_{k,j})\right|$. De discussões prévias, sabemos que $\beta(\lambda_{k,j})$ poderá ser a raiz real de menor magnitude do polinômio quadrático (6.57), ou o minimizador deste polinômio, se as raízes forem complexas. Em ambas situações, utilizando uma simples linha de prova, demonstra-se em [161], que $\left\|\boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}_T(\lambda_{k,j}))\right\|$ diminui monotonicamente quando $\lambda_{k,j} = 0 \to 1$. Ou seja, segue, em princípio, a propriedade do passo de Newton, onde $\boldsymbol{M}_N(\boldsymbol{x}_k + \lambda_{k,j}\boldsymbol{d}_{N,k}) = (1 - \lambda_{k,j})\boldsymbol{F}_k$ implica em uma redução linear de $\left\|\boldsymbol{M}_N(\boldsymbol{x}_k + \lambda_{k,j}\boldsymbol{d}_{N,k})\right\|$ em relação a $\left\|\boldsymbol{F}_k\right\|$, com o aumento de $\lambda_{k,j}$ no intervalo $(0, 1]$. Como podemos observar de (6.56) e (6.57), para $\lambda_{k,j} = 1$, a correção do tensor curvilinear (6.56) corresponde à correção do tensor (6.46). Também demonstra-se em [161], que esta correção assintoticamente se aproxima da direção da correção de Newton, quando $\lambda_{k,j} \to 0$. Sendo assim, a pesquisa-em-linha curvilinear, em discussão, tem a virtude de emular o comportamento dos métodos de região de confiança, oferencendo a "melhor" correção como primeira tentativa, e gradualmente regressando para uma direção segura quando comprimento do passo se encolhe para zero.

Devido à possibilidade da correção do tensor curvilinear não estar na direção de descida, o teste de monotonicidade na estratégia curvilinear, deve ser dado por

$$f(\boldsymbol{x}_k + \boldsymbol{d}_T(\lambda_{k,j})) \le f(\boldsymbol{x}_k) + \alpha\lambda_{k,j}\text{tax\_inc}_k\,, \tag{6.58}$$

onde a derivada direcional, tax_inc_k, é o produto-interno entre o gradiente, $\nabla f(\boldsymbol{x}_k)$, e a correção de Newton, $\boldsymbol{d}_{N,k}$. O algoritmo BLS-TC ("Backtracking Line-Search - Tensor Curvilinear"), que representa a implementação da estratégia curvilinear neste trabalho, possui fundamentalmente a mesma estrutura do Algoritmo BLS; por isto sua descrição foi omitida. Devido à dependência não-linear da correção do tensor com o fator de amortecimento, a técnica dividindo-pela-metade foi sugerida em substituição ao procedimento de interpolação. Na determinação da raíz real de menor magnitude do polinômio (6.57), devemos considerar o problema do cancelamento da contribuição de $\lambda_{k,j}c_{0,k}$ no cálculo desta raiz, quando $\boldsymbol{d}_T(\lambda_{k,j}) \to \boldsymbol{d}_{N,k}$.

6.5.4. Implementação Modificada

Seguindo a mesma concepção do algoritmo GN, foi desenvolvida uma versão modificada do processo iterativo (6.23.a)-(6.23.f) (método do tensor), globalizado com estratégia de pesquisa-em-linha. Na malha simplificada deste algoritmo, cada nova correção, denominada de *correção do tensor simplificada*, é calculada sem atualização da matriz jacobiana e dos vetores que definem o termo quadrático do tensor. A implementação envolve um objeto tensor de posto-um com interpolação de um ponto passado na sua formação. Com base nesta concepção, foi desenvolvido o algoritmo GT descrito abaixo.

Algoritmo GT
("Global Tensor")

(L-1) Dado: $\boldsymbol{x}_0 \in \mathbb{R}^N$, $\Xi \in (0, 1)$, $NI_{\text{máx}}, NIJ_{\text{máx}} > 0$ e $NIM_{\text{máx}} \geq 0$;
(L-2) $k \leftarrow 0$;
(L-3) **PARA** $i \leftarrow 0$ **PASSO** 1 **ATÉ** $NIJ_{\text{máx}}$ **FAÇA**: (malha externa)
(L-4) Calcular e fatorar: $\boldsymbol{J}_i = \boldsymbol{J}(\boldsymbol{x}_k) \in \mathbb{R}^{N \times N}$; (matriz jacobiana)
(L-5) Determinar: $\boldsymbol{d}_{N,\,k} \in \mathbb{R}^N$ tal que $\boldsymbol{M}_N(\boldsymbol{x}_k + \boldsymbol{d}_{N,\,k}) = \boldsymbol{0}$;
(L-6) **SE** $(k > 0)$ **ENTÃO**
(L-7) Calcular: $\boldsymbol{s}_k = \boldsymbol{x}_k - \boldsymbol{x}_{k-1}$ e $\boldsymbol{a}_k = \frac{2}{m_k}(\boldsymbol{F}(\boldsymbol{x}_{k-1}) - \boldsymbol{F}(\boldsymbol{x}_k) - \boldsymbol{J}_i\boldsymbol{s}_k)$;
(L-8) Determinar: $\boldsymbol{d}_{T,\,k} \in \mathbb{R}^N$ tal que $\boldsymbol{d}_{T,\,k} = \text{argmin}_{\boldsymbol{d} \in \mathbb{R}^N} \left\| \boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}) \right\|$;
(L-9) **PARA** $l \leftarrow 0$ **PASSO** 1 **ATÉ** $NIM_{\text{máx}}$ **FAÇA**: (malha interna)
(L-10) **SE** $(k > 0)$ **ENTÃO**
(L-11) Determinar: $\lambda_k \in \mathbb{R}$ e $\boldsymbol{d}_k$ entre $\boldsymbol{d}_{T,\,k}$ ou $\boldsymbol{d}_{N,\,k}$ via Algoritmo BLS-TS, ou determinar: $\lambda_k \in \mathbb{R}$ e $\boldsymbol{d}_k \leftarrow \boldsymbol{d}_T(\lambda_k)$ via algoritmo BLS-TC;
(L-12) **CASO CONTRÁRIO**
(L-13) Determinar: $\lambda_k \in \mathbb{R}$ e $\boldsymbol{d}_k \leftarrow \boldsymbol{d}_{N,\,k}$ via Algoritmo BLS;
(L-14) $\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k$;
(L-15) **SE** ("CONVERGÊNCIA" **OU** "DIVERGÊNCIA") **ENTÃO** $\boldsymbol{x}_* \leftarrow \boldsymbol{x}_{k+1}$ **RETORNE**;
(L-16) **SE** $(NIM_{\text{máx}} > 0$ **E** $k > 0)$ **ENTÃO**
(L-17) **SE** $(\Xi = 0$ **OU** $f(\boldsymbol{x}_{k+1})/f(\boldsymbol{x}_k) < \Xi)$ **ENTÃO**
(L-18) Determinar: $\bar{\boldsymbol{d}}_{N,\,k} \in \mathbb{R}^N$ tal que $\overline{\boldsymbol{M}}_N(\boldsymbol{x}_{k+1} + \bar{\boldsymbol{d}}_{N,\,k}) = \boldsymbol{0}$ †;
(L-19) Calcular: $\boldsymbol{s}_k = \boldsymbol{x}_k - \boldsymbol{x}_{k-1}$ e $\boldsymbol{a}_k = \frac{2}{m_k}(\boldsymbol{F}(\boldsymbol{x}_{k-1}) - \boldsymbol{F}(\boldsymbol{x}_k) - \boldsymbol{J}_i\boldsymbol{s}_k)$;
(L-20) Determinar: $\bar{\boldsymbol{d}}_{T,\,k} \in \mathbb{R}^N$ tal que $\bar{\boldsymbol{d}}_{T,\,k} = \text{argmin}_{\bar{\boldsymbol{d}} \in \mathbb{R}^N} \left\| \overline{\boldsymbol{M}}_T(\boldsymbol{x}_{k+1} + \bar{\boldsymbol{d}}) \right\|$ ††;
(L-21) $\boldsymbol{d}_{T,\,k} \leftarrow \bar{\boldsymbol{d}}_{T,\,k}$; $\boldsymbol{d}_{N,\,k} \leftarrow \bar{\boldsymbol{d}}_{N,\,k}$; (utilizar correções simplificadas)
(L-22) **CASO CONTRÁRIO**
(L-23) **SE** $(f(\boldsymbol{x}_{k+1})/f(\boldsymbol{x}_k) > 1)$ **ENTÃO FIM PARA;**
(L-24) $k \leftarrow k + 1$.

† $\bar{\boldsymbol{M}}_N(\boldsymbol{x}_{k+1}+\bar{\boldsymbol{d}}_N) = \boldsymbol{F}(\boldsymbol{x}_{k+1})+\boldsymbol{J}_i\bar{\boldsymbol{d}}_N$.

†† $\bar{\boldsymbol{M}}_T(\boldsymbol{x}_{k+1}+\bar{\boldsymbol{d}}_T) = \boldsymbol{F}(\boldsymbol{x}_{k+1})+\boldsymbol{J}_i\bar{\boldsymbol{d}}_T+\frac{1}{2}\boldsymbol{a}_k(\boldsymbol{s}_k^T\bar{\boldsymbol{d}}_T)^2$.

No algoritmo acima, foram empregados os mesmos testes de convergência e de divergência adotados no Algoritmo GN (método de Newton), assim como os mesmos parâmetros de entrada. Podemos observar facilmente que para $NIM_{\text{máx}} = 0$ obtemos o método do tensor convencional, i.e., sem correções simplificadas. Porém, para $\Xi = 0$ e $NIM_{\text{máx}} = NI_{\text{máx}} > 0$, e para $0 < NIM_{\text{máx}} << NI_{\text{máx}}$, obtemos iterações análogas às definidas no método de Newton-corda-paralela e de Newton-Shamanskii, respectivamente. Em geral, o valor do parâmetro Ξ, que controla o uso da malha interna, (L-9)-(L-23), pode ser selecionado na mesma faixa definida no Algoritmo GN. As pesquisas-em-linha, nas linhas (L-11) e (L-13), para amortecimento do passo (ou correção), pode ser desabilitada fazendo $NPL_{\text{máx}} = 0$ no Algoritmo BLS-TS ou BLS-TC, e no Algoritmo BLS, respectivamente.

6.6. Técnica de Continuação

Em geral, os métodos de Newton e do tensor convergem apenas quando a iteração inicial está localizada suficientemente próxima da raiz procurada. Para ampliar o domínio de convergência destes métodos, quando a iteração inicial estiver fora da hiper-esfera que define a região de convergência local, pode ser empregad a técnica de continuação (ou *homotopia*). A aplicação desta técnica é simples, e envolve a solução da equação do operador não-linear, $\boldsymbol{H}:\mathbb{R}^N\times\mathbb{R}^1\to\mathbb{R}^N$, definido por [17]

$$\boldsymbol{H}(\boldsymbol{x};\zeta) = \boldsymbol{0} \qquad \zeta\in[\zeta_0,\zeta_*]\,, \tag{6.59}$$

onde ζ é o parâmetro de homotopia. Para $\zeta = \zeta_0$, assume-se que (6.59) tem uma solução conhecida dada por $\boldsymbol{x}_0$. Com isto, o objetivo é encontrar a solução desconhecida, $\boldsymbol{x}_*$, para $\zeta = \zeta_* \neq \zeta_0$. Se o problema depender do parâmetro, ζ, tal que: $\boldsymbol{H}(\boldsymbol{x};\zeta) = \boldsymbol{F}(\boldsymbol{x};\zeta)$, então nos referimos a esta situação como continuação natural. Por outro lado, denomina-se continuação artificial quando o parâmetro de continuação é artificialmente introduzido [175],[17]. Esta última técnica, possui a desvantagem de produzir soluções não-físicas no caminho de homotopia, para $\zeta_0 < \zeta < \zeta_*$.

Neste trabalho, nas análises de CC e do BH, assumimos que o problema depende naturalmente do parâmetro de homotopia. Em geral, este parâmetro pode estar associado com: a intensidade de fonte independente de CC e de RF, valor de um componente do circuito (resistência, indutância,

capacitância, etc), temperatura, frequência de operação, etc.

Antes de descrevermos o método de continuação implementado, vamos introduzir o parâmetro de continuação, $\bar{\zeta}$, tal que: $\zeta = \zeta_0 + \bar{\zeta}(\zeta_* - \zeta_0)$ ou $\bar{\zeta} = (\zeta - \zeta_0)/(\zeta_* - \zeta_0)$, e os pontos $\boldsymbol{x}_0$ e $\boldsymbol{x}_*$ conectados por um caminho de homotopia, $\boldsymbol{x}^{(hom)}:[0, 1] \to D$. Neste caso, podemos re-escrever (6.59) como se segue

$$\boldsymbol{H}(\boldsymbol{x}^{(hom)}(\bar{\zeta});\bar{\zeta}) = \boldsymbol{0} \qquad \bar{\zeta} \in [0, 1] \; . \tag{6.60}$$

Da expressão acima, é obvio que para $\bar{\zeta} = 0$ e $\bar{\zeta} = 1$ obtemos a solução inicial e final do problema original dadas por $\boldsymbol{x}_0 = \boldsymbol{x}^{(hom)}(0)$ e $\boldsymbol{x}_* = \boldsymbol{x}^{(hom)}(1)$, respectivamente.

Para implementação da técnica de continução, utilizamos o procedimento duplicando-comprimento-do-passo, proposto em [47]. Para explicar o mecanismo básico de operação deste procedimento, vamos assumir que o processo discreto de continuação já progrediu através dos parâmetros $\bar{\zeta}_{c,-q} < \ldots < \bar{\zeta}_{c,-1} < \bar{\zeta}_{c,0}$, selecionados sob o caminho de homotopia, com $\boldsymbol{x}^{(hom)}(\bar{\zeta}_{c,-q})$, …, $\boldsymbol{x}^{(hom)}(\bar{\zeta}_{c,-1})$, $\boldsymbol{x}^{(hom)}(\bar{\zeta}_{c,0})$, e onde $\bar{\zeta}_c$ é a iteração corrente. Neste caso, a *i*-ésima variável independente pode ser estimada pela interpolação polinomial, dada por

$$x_i^{(pred)}(\bar{\zeta}_c) = \sum_{j=0}^{q} x_i^{(hom)}(\bar{\zeta}_{c,-j}) L_q^j(\bar{\zeta}_c) \qquad i = 1, 2, \ldots, N \tag{6.61}$$

onde $L_q^j(\bar{\zeta}_c)$ denota os polinômios Lagrangeanos de ordem-q [175]. Neste trabalho esta predição é conduzida via extrapolação polinomial de até ordem cubica. Se as derivadas de $x_i(\bar{\zeta})$, com relação a $\bar{\zeta}$, forem suficientemente precisas e estiverem disponíveis, então a extrapolação utilizando polinômios de Hermite pode ser preferencial. Idealmente, gostaríamos de controlar o comprimento-do-passo [175], $\Delta\bar{\zeta}^k$, de tal forma que a iteração predita resida na região de convergência local (e o mais próximo possível da raiz) do solucionador não-linear [166]. Uma discussão sob o controle de $\Delta\bar{\zeta}^k$ para os métodos de continuação, incluindo aqueles fundamentados em técnicas de extrapolação, pode ser encontrado em [175]. Em adição, um interessante método de continuação, fundamentado na extrapolação de Padé, foi proposto em [48], para análise do BH. Com a extrapolação, esperamos obter um caminho predito razoalvelmente próximo do caminho de homotopia, onde cada ponto predito é utilizado como iteração inicial na aplicação do método de Newton ou do tensor (corretor), para encontrar a solução correta. Sendo assim, se a solução correta puder ser encontrada, o método de continuação pode ser caracterizado como iterações do tipo *preditor-corretor*. Entretanto, se uma solução do problema não for encontrada , i.e., algum critério de divergência for satisfeito, então o tamanho-de-passo $\Delta\bar{\zeta}^k$ é dividido pela metade e a solução é novamente procurada. Para problemas fortemente não-lineares, o comprimento-do-passo pode

decrescer abaixo de um limite aceitável. Como consequência, o tempo de computação pode aumentar de forma considerável tornando inviável o processo de homotopia. Por outro lado, se uma nova solução for encontrada, então uma extrapolação (preditor) é feita, e após uma série de bem-suscedidos passos preditor-corretor, o comprimento-do-passo é dobrado em uma tentativa de acelarar o processo de homotopia.

6.7. Testes Preliminares

Para validar a implementação e avaliar o desempenho do método do tensor versus o método de Newton, foram realizados uma série de experimentos númericos utilizando o conjunto de problemas testes da coleção de Moré, Garbow e Hillstrom (MGH) [176]. Os resultados apresentados abaixo, foram obtidos das implementações do Algoritmo GN e do Algoritmo GT, introduzidos acima.

Os primeiros resultados, comparando o método do tensor com o método de Newton, foram sumarizados na Tabela 6.2. Nesta tabela, foram introduzidas os seguintes parâmetros: número de iterações, NI, e número de cálculo da função, NCF. Estes resultados foram obtidos empregando pesquisa-em-linha com interpolação quadrática, e função nível definida em norma-M2. No método do tensor, foi utilizada a pesquisa-em-linha com a estratégia padrão. Convém lembrar, que esta estratégia de pesquisa-em-linha padrão, para o método do tensor, requer um maior, NCF, quando comparada com a pesquisa-em-linha para o método de Newton. Nos resultados abaixo, para os parâmetros ε_F, ε_d, $NI_{\text{máx}}$, $NIJ_{\text{máx}}$, $NIM_{\text{máx}}$ e $NIP_{\text{máx}}$ foram utilizados os seguintes valores 10^{-7}, 10^{-9}, 150, 150, 0, e 10, respectivamente. Para a pesquisa-em-linha foram utilizados $\alpha = 10^{-4}$, $\gamma = 10^{-4}$, e $\lambda_{\min} = 10^{-7}$. De imediato, dos resultados da Tabela 6.2, podemos observar que o método de Newton produz 4 (quatro) falhas, enquanto o método do tensor produz apenas 2 (duas).

A seguir, para estabelecermos uma comparação gráfica entre os métodos discutidos acima, vamos utilizar a seguinte figura de mérito, entitulada de razão de desempenho (RD), e definida por: $RD(X) = \log_2(X\text{ - método padrão} / X\text{ - método novo})$, onde $X = NI, NCF$ [161]. Na Fig. 6.1, podemos observar graficamente a RD, tendo o método de Newton como método padrão, e o método do tensor como método novo. Lembramos que ambos os métodos foram globalizados com pesquisa-em-linha, sendo que, no método do tensor, foi utilizada a estratégia padrão. Mais precisamente, as Figs. 6.1(a) e (b) se referem, respectivamente, aos problemas onde $\boldsymbol{J}(\boldsymbol{x}_*)$ é não-sigular, e aos problemas onde $\boldsymbol{J}(\boldsymbol{x}_*)$ é singular com $\text{posto}(\boldsymbol{J}(\boldsymbol{x}_*)) = n - 1$, sendo n igual à dimensão do problema. Vale ressaltar que, os dados da Tabela 6.2 correspondem ao gráfico da Fig. 6.1(a). Definimos o

Table 6.2

RESULTADOS PARA O CONJUNTO DE PROBLEMAS TESTES DE MORÉ, GARBOW E HILLSTROM [176].

	Função	n	x_0	Método padrão: Newton			Método novo: tensor			$\|\Delta_*\|_2$ †
				NI	NCF	$\|F(x_*)\|_2$	NI	NCF	$\|F(x_*)\|_2$	
1	Box 3D	3	1	2	3	3,7-8	3	4	2,0-16	9,9-5
2			10	3	4	1,4-12	3	4	4,8-10	1,4-3
3	Brown almost linear	10	1	150	162	—	12	53	7,8-14	—
4			10	9	11	2,2-8	8	17	3,5-15	7,3-8
5	Broyden banded	30	1	5	6	1,5-8	5	6	3,0-13	8,7-10
6			10	11	12	5,7-12	9	10	5,7-15	3,2-13
7			100	17	18	3,5-15	13	14	1,7-10	1,0-11
8	Broyden tridiagonal	30	1	4	5	1,1-9	4	5	6,4-12	9,4-11
9			10	8	9	4,6-15	5	6	1,3-10	7,6-12
10			100	11	12	6,1-12	5	6	1,1-8	6,5-10
11	Chebyquad	7	1	7	11	6,2-9	6	8	3,4-8	1,3-1
12		9	1	150	1124	—	7	13	2,9-11	—
13		4	10	53	125	1,1-10	50	196	7,5-10	1,0
14	Discrete boundary	10	1	2	3	3,1-8	2	3	6,9-9	2,7-7
15			10	3	4	1,7-9	3	4	5,4-10	1,1-8
16			100	9	10	1,1-14	8	9	5,8-12	1,0-11
17	Discrete integral	30	1	3	4	9,8-15	2	3	6,4-8	5,1-8
18			10	3	4	3,1-8	3	4	8,6-9	1,8-8
19			100	9	10	2,2-13	7	8	2,7-8	2,5-8
20	Helical valley	3	1	9	12	4,1-11	9	18	2,6-16	3,0-12
21			10	12	19	6,9-8	11	28	1,3-13	5,4-9
22			100	17	25	6,1-8	12	33	1,3-9	4,2-9
23	Powell singular	4	1	14	15	4,7-8	3	4	2,3-16	1,5-4
24			10	17	18	7,4-8	3	4	7,6-15	1,9-4
25			100	21	22	2,9-8	3	4	3,0-12	1,2-4
26	Rosenbrock	2	1	14	27	0,0	7	21	0,0	0,0
27			10	3	5	1,3-14	6	15	2,2-8	1,5-9
28	Trigonometric	30	1	13	26	3,2-8	11	42	1,3-9	5,1-2
29			10	150	1120	—	150	2276	—	—
30			100	150	989	—	115	1049	—	—
31	Variable dimension	10	1	14	15	2,6-12	8	9	4,7-12	5,9-15
32			10	17	18	0,0	10	11	2,7-13	2,3-16
33			100	23	24	5,2-10	17	25	2,7-8	2,2-11
34	Wood gradient	4	1	17	24	4,0-13	11	22	7,0-8	2,6-8
35			10	59	134	4,8-14	61	217	1,9-14	8,2-15

† $\Delta_* = x_*^{\text{tensor}} - x_*^{\text{Newton}}$.

problema fácil como aquele no qual a convergência ocorre em menos de dez iterações, i.e., $NI = 10$. Então, podemos observar que, em termos de NI, para o caso não-singular o método do tensor é sempre superior nos problemas difíceis e nos problemas fáceis é inferior apenas duas vezes. Como esperado, para os problemas com matriz jacobiana singular na raiz, o método do tensor possui uma dramática vantagem sobre o método de Newton. Em termos de NCF, podemos observar que a superioridade do método do tensor é menos significante do que a razão de desempenho para NI. Isto deve-se ao maior custo por iteração (pesquisa nas direções de Newton e do tensor) da estratégia pesquisa-em-linha padrão, comparando-se com a curvilinear.

Nos gráficos da Fig. 6.2., são apresentados os testes para avaliar o desempenho das estratégias de pesquisa-em-linha utilizadas no método do tensor. Detalhando, na Fig. 6.2(a), a estratégia de pesquisa-em-linha curvilinear, com retrocedimento via interpolação quadrática, é comparada com a estratégia de pesquisa-em-linha padrão. Já na Fig. 6.2(b), o retrocedimento com redução do

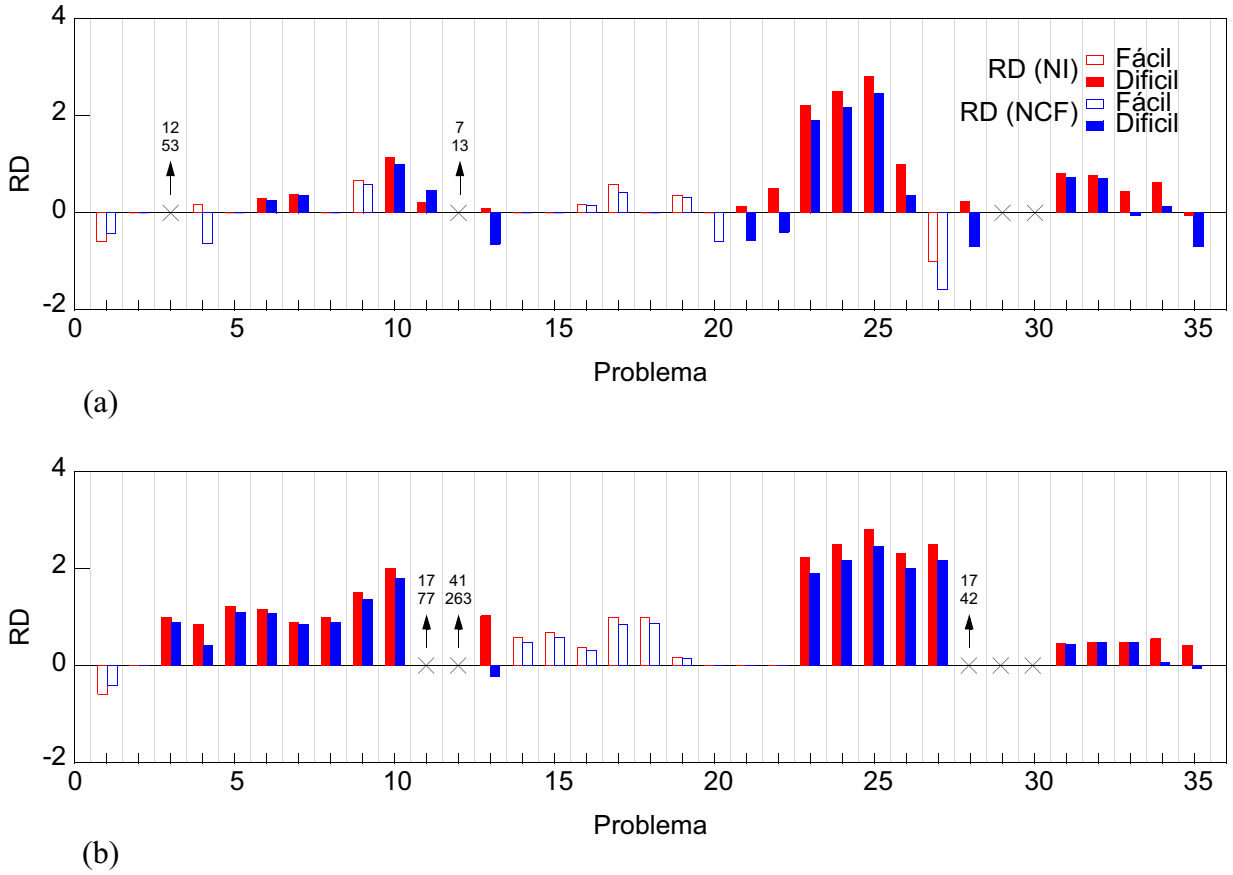


Fig. 6.1 Gráficos da razão de desempenho (RD) do método do tensor (novo) vs. método Newton (padrão) para: (a) $J(x_*)$ não-singular e (b) $\boldsymbol{J}(\boldsymbol{x}_*)$ singular com $\text{posto}(\boldsymbol{J}(\boldsymbol{x}_*)) = n-1$.

parâmetro de amortecimento, via técnica λ-dividindo-pela-metade, é comparada com a estratégia padrão. Os resultados destas figuras demonstram uma superioridade, em termos de NCF e NI, da pesquisa-em-linha curvilinear sobre a padrão, exceto pelos resultados obtidos no problema 4. Conforme esperado, e em contraposição com os resultados da Fig. 6.1, as RDs para pesquisa-em-linha, ilustradas na Fig. 6.2, são significantemente maiores em função de NCF, quando comparadas com as RDs em função de NI.

Os resultados apresentados acima confirmam a superioridade do método do tensor, particularmente quando este utiliza a pesquisa-em-linha curvilinear, em relação ao método de Newton, em termos de tempo de processamento e robustez, com um insignificante custo extra de memória. Em adição, estes resultados estão consistentes com os obtidos em [29] e [161], e validam a implementação dos solucionadores de equação não-linear propostos.

6.8. Conclusão

Visando sua aplicação na análise do BH multi-níveis, foi discutido acima, a teoria e a

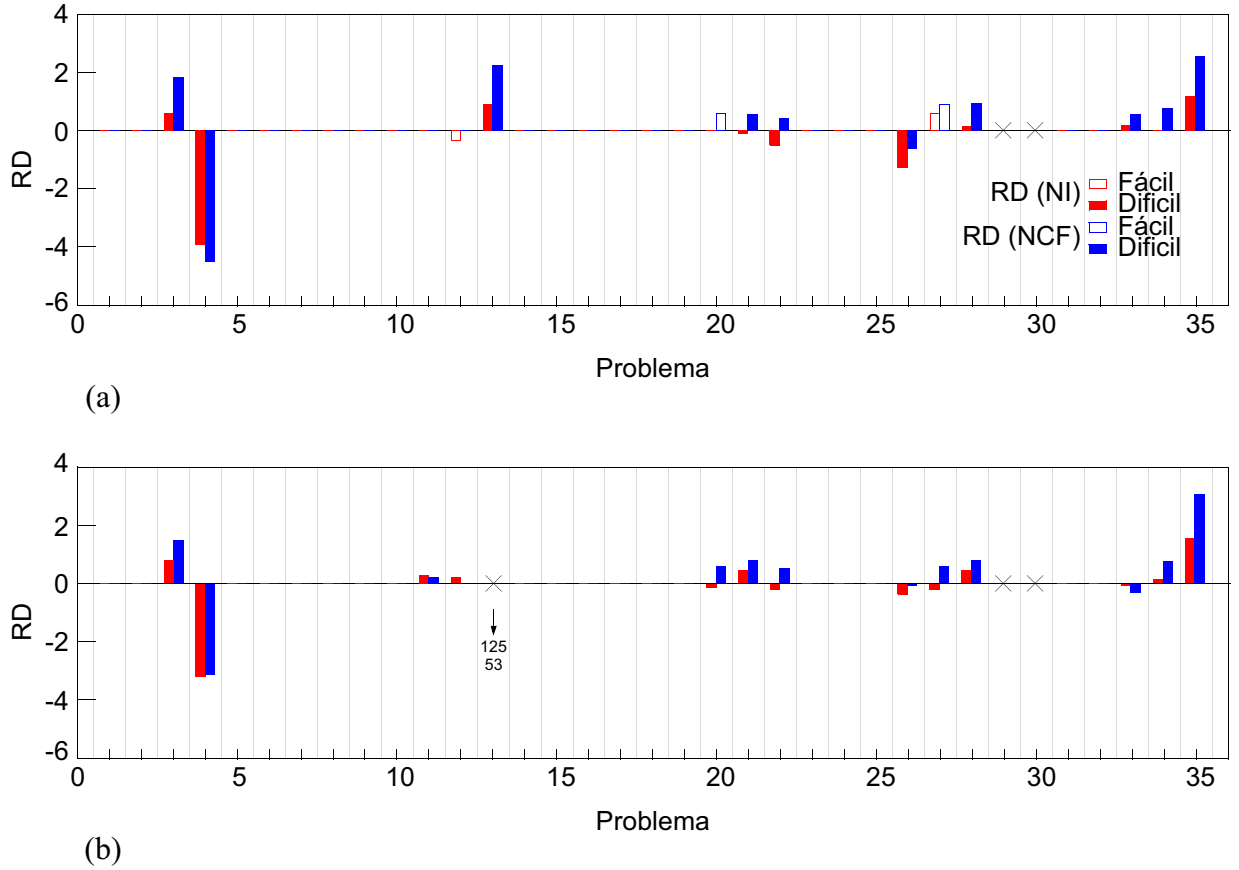


Fig. 6.2 Gráficos da razão de desempenho (RD) das estratégias de pesquisa-em-linha para o método do tensor. (a) Pesquisa-em-linha curvilinear com interpolação quadrática vs. pesquisa-em-linha padrão. (b) Pesquisa-em-linha curvilinear com λ-divindo-pela-metade vs. pesquisa-em-linha padrão.

implementação de solucionadores de equação não-linear com SuRs de fundo em pequena-escala, utilizando matrizes densas, e em média-escala, utilizando a técnica de matrizes esparsas. Estes solucionadores, método de Newton e do tensor, também serão utilizados na análise de CC do circuito, que define a iteração inicial utilizada na solução do regime permanente do circuito. Em adição, nestes métodos de solução direta do sistema jacobiano, a fatorização da matriz jacobiana é o que limita a sua aplicação em problemas com SuRs de fundo em grande-escala.

O conceito de função nível foi discutido e implementado, possibilitando testes de monotonicidade operando com a redução da norma-M2 do mapeamento (resíduo não-linear), ou com a redução da norma-M2 da correção, no método de Newton. A teoria básica do método de Newton (ou método padrão), globalizado com a estratégia de pesquisa-em-linha, foi discutida incluindo sua implementação numérica. O método do tensor discutido e implementado, utiliza um objeto tensor de posto-um com interpolação de um ponto passado, tendo assim um custo por iteração comparável com o método padrão. A correção do tensor está fundamentada na minimização de um modelo quadrático local. Como estratégia de globalização, foram discutidas e

implementadas as estratégias de pesquisa-em-linha padrão e curvilinear. Lembramos que a estratégia curvilinear elimina a necessidade de uma pesquisa-em-linha na direção da correção de Newton, e converge para esta direção para valores críticos do fator de amortecimento. Com isto, esta estratégia irá selecionar uma direção definida entre as correções do tensor e de Newton, emulando o método da região de confiança. Versões não-monotônicas para a pesquisa-em-linha com retrocedimento podem ser incorporadas para ambos os métodos.

Os resultados númericos referentes aos testes preliminares nos problemas da coleção MGH foram apresentados sob forma de tabela e gráfico. Estes resultados confirmam a superioridade do método do tensor sobre o método de Newton, principalmente em problemas onde a matriz jacobiana é mal-condicionada ou singular na raiz. Os algoritmos referentes aos métodos implementados neste trabalho, foram apresentados e discutidos em detalhe. A implementação numérica foi realizada utilizando a linguagem objeto-orientada C++, sob forma de *Standard Template Library* (STL). Para o método de Newton e do tensor, foi incluída uma modificação que permite manter constante a jacobiana em sucessivas iterações (i.e., iterações de cordas paralelas e de Shamanskii). A utilização de escalamento nas funções e/ou variáveis pode ser aplicada em nossa implementação, seguindo a metodologia proposta em [46],[168]. Finalmente, a técnica de continuação (ou homotopia), fundamentada no método duplicando-comprimento-do-passo, foi discutida e implementada.

7. Método de Newton Inexato e do Tensor Inexato

7.1. Introdução

FOI PROPOSTO PREVIAMENTE, a aplicação do método de Newton e do tensor para a análise do BH (um nível ou multi-níveis), quando a dimensão da equação determinante das SuRs de fundo for de pequena- ou média-escala. Com base nestes métodos, discutir-se-á agora a teoria e a implementação do método de Newton inexato [21] e do tensor inexato [31], para eficiente solução do problema do BH, quando a(s) SuR(s) de fundo produzir(em) subsistema(s) de grande-escala. Como os nomes sugerem, estes métodos consistem de uma versão *inexata* dos métodos descritos no capítulo anterior. Mais precisamente, o modelo linear local e o modelo quadrático (ou modelo do tensor) local, que definem a correção de Newton e do tensor, respectivamente, são resolvidos apenas aproximadamente via um apropriado método iterativo (solucionador interno). Vale ressaltar que, a principal diferença entre os métodos de Newton inexato e o do tensor inexato é que, neste último, a resolução aproximada do modelo do tensor está associada, em geral, a um problema de minimização. Por sua vez, isto conduz a um diferente, e mais elaborado, critério de parada para o método iterativo (solucionador linear) utilizado na resolução aproximada do modelo do tensor, quando comparado com o critério de parada para a resolução amproximada do modelo linear utilizado no método de Newton inexato.

A principal vantagem dos métodos inexatos é a eliminação da etapa de fatorização envolvendo a matriz jacobiana. Nestes métodos, a matriz jacobiana é solicitada apenas em operações do tipo produto matriz-por-vetor e em pré-condicionadores. O método de Newton inexato (ou método inexato padrão) tem sido amplamente utilizado na análise do BH, como podemos observar nos trabalhos [54],[22],[110],[26]. Seguindo a mesma organização do capítulo anterior, utilizaremos este método para avaliar o desempenho e facilitar a discussão teórica sobre o método do tensor inexato. Vale destacar, que o suscesso dos métodos inexatos depende diretamente da eficiência do pré-condicionador utilizado pelo solucionador linear interno.

Observando igual estrutura do capítulo anterior, iniciaremos nossa discussão apresentando, na Seção 7.2, um resumo da teoria e da implementação do método de Newton inexato. Esta seção está subdivida em quatro subseções. Na Subseção 7.2.1, discutiremos a solução iterativa do modelo linear local que define a correção de Newton inexata. Para tal, a Subseção 7.2.1.1 apresenta o pré-condicionamento deste modelo linear, e a Subseção 7.2.1.2 técnicas em subespaço de Krylov, em particular o método de GMRES com reinício e pré-condicionamento para a sua solução aproximada. Em seguida, a Subseção 7.2.2 discute a escolha da sequência de termos forçantes que

define o grau de precisão da solução aproximada a cada iteração externa, evitando com isso o problema de sobressolução. Na Subseção 7.2.3, a estratégia de globalização via pesquisa-em-linha é considerada. Concluindo a discussão do método de Newton inexato apresentamos, na Subseção 7.2.4, uma implementação numérica modificada deste método.

Na Seção 7.3, também organizada em quatro subseções, exporemos um resumo da teoria e implementação do método do tensor inexato [29],[30]. Para iniciar, a Subseção 7.3.1 apresenta diferentes processos para a solução iterativa do modelo do tensor, entitulados de processo de solução simplificada [32], modificada [114] e completa [31],[34]. Neste sentido, a Subseção 7.3.1.1 traz o pré-condicionamento do modelo do tensor e a Subseção 7.3.1.2, subdividida em duas partes, discute técnicas em subespaço de Krylov para sua solução. Na primeira parte, Subseção 7.3.1.2.a, discutimos a teoria e implementação dos processos de solução simplificada e modificada via o método GMRES em versão escalar e bloco-2. Enquanto a segunda parte, Subseção 7.3.1.2.b, considera o processo de solução completa, via o método TGMRES-Bt com reinício e pré-condicionamento, e operando com bloco-2 ou bloco-3. Na Subseção 7.3.2, discutimos brevemente a escolha da sequência de termos forçantes para a solução aproximada do modelo do tensor. A teoria das estratégias de globalização via pesquisa-em-linha utilizadas e sua implementação serão discutidas na Subseção 7.3.3. Mais precisamente, na Subseção 7.3.3.1 discutiremos a estratégia padrão, enquanto na Subseção 7.3.3.b, a estratégia curvilinear será apresentada. Finalizando a discussão sobre o método do tensor, a Subseção 7.3.4 apresenta uma implementação numérica modificada deste método.

Assim como no capítulo anterior, os algoritmos apresentados descrevem em detalhe a implementação adotada para os métodos de Newton inexato e do tensor inexato. A Seção 7.4 apresenta uma discussão sobre as diferentes formas de conduzir o produto matriz jacobiana por vetor necessário aos métodos de subespaço de Krylov. A Seção 7.5 discute os diferentes tipos de pré-condicionadores utilizados na análise do BH [26],[160]. Já a Seção 7.6 apresenta um sumário dos testes preliminares realizados para validar a implementação dos métodos inexatos propostos neste capítulo. As conclusões finais são reservadas para a Seção 7.7.

7.2. Método de Newton Inexato

O método de Newton inexato é uma variação do método de Newton descrito no capítulo anterior, onde, a cada iteração, o vetor de correção (ou passo) satisfaz, como raiz, apenas aproximadamente a equação do modelo linear (6.7), porém, produzindo uma redução na associada função de nível (em norma-M2). Este método, pode ser descrito pelo processo iterativo (6.9.a)-

(6.9.d), substituindo-se (6.9.c) por

$$\text{Encontrar: } \boldsymbol{d}_{N,k} \in \mathrm{R}^N \text{ tal que } \left\| \boldsymbol{M}(\boldsymbol{x}_k + \boldsymbol{d}_{N,k}) \right\| \leq \eta_k \left\| \boldsymbol{F}(\boldsymbol{x}_k) \right\| \tag{7.1}$$

onde $\eta_k \in \mathrm{R}:(0,1)$ refere-se ao termo forçante que define o grau de precisão da solução aproximada. A escolha judiciosa da sequência de termos forçantes será discutida a seguir. Para a solução dos sistemas lineares jacobianos associados a (7.1), serão considerados métodos iterativos que envolvem a matriz jacobiana somente em operações do tipo produto matriz por vetor e que exigem uma pequena área de memória para armazenagem de vetores de trabalho. Devido a este fato, o método de Newton inexato é um execelente candidato a problemas em grande-escala, onde a fatorização da matriz jacobiana torna se impraticável. Em geral, a robustez dos métodos iterativos, dependerá da distribuição dos auto-valores (i.e., espectro) da matriz jacobiana em cada iteração. Por sua vez, utilizando a técnica de pré-condicionamento, o espectro da matriz jacobiana pode ser aglomerado dentro de uma elipse que define rápida convergência do método iterativo utilizado.

A característica de convergência local do método de Newton inexato depende diretamente da escolha da sequência de termos forçantes, conforme discutido em [21]. Em [177], foram apresentadas provas de convergência para diferentes escolhas da sequência de termos forçantes. Em adição, podemos citar os trabalhos que apresentam análise de convergência local em problemas de deficiência de posto na raiz [178], em métodos modificados sob transformação afim e quando combinado com métodos de projeção [179]. Para análise de convergência, incluindo estratégias de globalização do tipo de pesquisa-em-linha e de região-de-confiança, temos [180],[181]. Estes resultados teóricos apresentam características de convergência quadrática e superlinear. Vale ressaltar que, a versão inexata do método GAIN descrito no capítulo anterior e que opera com a norma do erro foi desenvolvida em [182], incluido uma teoria para caracterização de convergência. A implementação detalhada deste método pode ser encontrada em [168].

7.2.1. Solução Iterativa do Modelo Linear

Conforme discutido acima, o suscesso no cálculo aproximado da correção de Newton depende da robustez do método iterativo utilizado na solução do modelo linear (6.7). Os tipos mais comuns de métodos iterativos, para esta função, são classificados como: estacionários (e.g., Jacobi, Gauss-Siedel, *successive over-relaxation* (SOR), *symmetric successive over-relaxation* (SSOR), etc) e não-estacionários (e.g., *conjugate gradient* (CG), *bi-conjugate gradient* (BiCG), *conjugate gradient squared* (CGS), *quasi-minimal residual* (QMR), g*eneralised minimal residual* (GMRES), etc). Os métodos não-estacionários, na sua maioria fundamentados na idéia de sequência de vetores

ortogonais, são mais efetivos e sua teoria relativamente mais recente. No contexto destes métodos, é conviniente introduzir o vetor de resíduo definido como

$$\boldsymbol{r}_{N,k,l} \stackrel{\text{def}}{=} -\boldsymbol{F}(\boldsymbol{x}_k) - \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d}_{N,k,l} \tag{7.2}$$

O subescrito l representa o contador de iterações do solucionador linear interno. Lembremos que, $\boldsymbol{F}(\boldsymbol{x}_k)$ e $\boldsymbol{J}(\boldsymbol{x}_k)$ representam, respectivamente, o mapeamento não-linear e sua matriz jacobiana, no ponto estacionário $\boldsymbol{x}_k$, correspondente a *k*-ésima iteração externa referente ao solucionador não-linear. Utilizando a norma de (7.2), podemos definir o seguinte critério de parada para o solucionador linear interno, da seguinte forma [183]:

$$\|\boldsymbol{r}_{N,k,l}\| / \|\boldsymbol{F}(\boldsymbol{x}_k)\| < \eta_k, \tag{7.3}$$

onde η_k é uma tolerância relativa de parada (ou termo forçante) na *k*-ésima iteração do solucionador externo. Este critério de parada está diretamente relacionado com (7.1), onde observarmos que o vetor $\boldsymbol{r}_{N,k,j}$ está diretamente relacionado com o modelo local, $\boldsymbol{M}_N(\boldsymbol{x}_k + \boldsymbol{d}_{N,k,l})$. O vetor de erro associado a correção de Newton inexata é dado por:

$$\boldsymbol{e}_{N,k,l} \stackrel{\text{def}}{=} \boldsymbol{d}_{N,k,l} - \boldsymbol{d}_{N,k}, \tag{7.4}$$

onde $\boldsymbol{d}_{N,k}$ é a correção de Newton (exata) dada por (6.8). O cálculo do vetor de erro possui um alto custo, pois, necessita da solução do sistema linear que se deseja resolver. Entretanto, um método fundamentado na minimização de estimativas precisas de (7.4) foi proposto em [182] com implementação em [168].

7.2.1.1. Pré-Condicionamento

Conforme descrito em [184],[183],[27], a razão de convergência dos métodos iterativos para a solução de sistemas de equações lineares algébricas pode variar dramaticamente dependendo do tipo de problema (espectro de auto-valores da matriz de iteração) no qual são aplicados. Devido a este fato, para a utilização bem suscedida destes métodos, na solução dos sistemas jacobianos que emergem a cada iteração nos métodos de Newton inexato e do tensor inexato, o desenvolvimento de pré-condicionadores é mandatório. A aplicação de um pré-condicionador permite a redução do raio espectral definido como $\rho(\boldsymbol{J}) = \text{máx}|\text{auto-valor}(\boldsymbol{J})|$, onde $\boldsymbol{J}$ é a matriz jacobiana [184] e, consequentemente, um aumento na velocidade de convergência. Em geral, a matriz pré-condicionadora pode ser escrita como $\boldsymbol{M} = \boldsymbol{M}_1\boldsymbol{M}_2$ e a versão pré-condicionada do sistema jacobiano (6.8) escrita como se segue [183]

$$\underbrace{\boldsymbol{M}_1^{-1}\boldsymbol{J}(\boldsymbol{x}')\boldsymbol{M}_2^{-1}}_{\tilde{\boldsymbol{J}}(\boldsymbol{x}')}\underbrace{\boldsymbol{M}_2\boldsymbol{d}}_{\tilde{\boldsymbol{d}}} = -\underbrace{\boldsymbol{M}_1^{-1}\boldsymbol{F}(\boldsymbol{x})}_{\tilde{\boldsymbol{F}}(\boldsymbol{x})} \tag{7.5}$$

onde $\boldsymbol{M}_1$ e $\boldsymbol{M}_2$ são matrizes de pré-condicionamento. Da expressão (7.5), para $\boldsymbol{M}_1 = \boldsymbol{M}_L$ e $\boldsymbol{M}_2 = \mathbf{1}$ obtemos o *pré-condicionamento à esquerda*, para $\boldsymbol{M}_1 = \mathbf{1}$ e $\boldsymbol{M}_2 = \boldsymbol{M}_R$ obtemos o *pré-condicionamento à direita*, e para $\boldsymbol{M}_1 = \boldsymbol{L}$ e $\boldsymbol{M}_2 = \boldsymbol{U}$ ($\boldsymbol{M}_2 = \boldsymbol{L}^T$ em sistemas simétricos) o *pré-condicionamento dividido*. Da definição do vetor de resíduo (7.2), podemos concluir que o pré-condionamento a esquerda afeta diretamente este vetor e, com isto, modifica o teste baseado em sua norma. Já o pré-condicionamento à direita afeta diretamente o vetor de erro (7.4) e não o vetor de resíduo, por esta razão, utilizaremos este tipo de pré-condicionamento. Interessante observar que, no método de Newton proposto em [182],[168], a norma do vetor de erro é utilizada como critério de convergência (de parada), sendo neste caso, o pré-condicionamento a esquerda utilizado na sua implementação.

Por razões práticas, a matriz $\tilde{\boldsymbol{J}}(\boldsymbol{x})$ deve ter uma estrutura, tal que, a solução do sistema pré-condicionado (7.5) exija um pequeno esforço computacional e a distribuição de auto-valores esteja aglomerada entorno do ponto complexo (1:0). Entretanto, estes objetivos são conflitantes, e.g., se escolhermos $\boldsymbol{M}$ igual a matriz identidade, não teremos custo computacional associado com a formação de $\boldsymbol{M}$, porém, sem melhoria no espectro, i.e., sem redução do raio espectral $\rho(\mathbf{1} - \tilde{\boldsymbol{J}}(\boldsymbol{x}))$ da matriz jacobiana. No outro extremo, se escolhermos $\boldsymbol{M}$ igual a inversa de $\boldsymbol{J}(\boldsymbol{x})$, temos um espectro concentrado no ponto complexo (1:0) (i.e., raio espectral igual a zero), porém, com um custo computacional de formação igual ao custo de solução do problema original. Claramente, uma relação de compromisso entre melhoria do espectro da matriz jacobiana e custo computacional deve ser estabelecida.

7.2.1.2. Técnicas em Subespaço de Krylov

Os métodos iterativos mais utilizados para solução de sistemas lineares não-simétricos e indefinidos, em subespaço de Krylov, estão fundamentados no processo de ortogonalização de *Arnoldi* [185], ou no processo de bi-ortogonalização de *Lanczos* [186],[187]. Apesar de exigir um pequeno esforço computacional e um baixo requerimento de memória por iteração, o processo de Lanczos pode produzir instabilidade númerica e uma conseqüente quebra, i.e., uma divisão por zero. Adicionalmente, o produto entre a matriz de coeficiente (i.e., matriz jacobiana) transposta e um vetor faz se necessário. Excluindo a situação de quebra-incurável, para evitar o problema de quebra, a estratégia *olhando-adiante* pode ser utilizada, o que resulta em uma implementação mais

complexa e em aumento do tempo de processamento e de o espaço de memória por iteração. Por sua vez, no processo de Arnoldi, as soluções aproximadas a cada iteração são caracterizadas por uma propriedade de minimização sobre os subespaços de Krylov. Neste processo, em aritimética exata, uma quebra significa que a solução foi obtida e é chamada de *quebra-afortunada* [52]. A dificuldade deste método deve-se ao fato da complexidade computacional e do espaço de memória crescerem linearmente com o número de iteração.

Neste trabalho, empregaremos o método GMRES com reinício para solução de (7.1). Proposto em [52] (ver também [157],[27]), este método é um dos mais bem suscedidos métodos de projeção em subespaço de Krylov e está fundamentado no processo de ortogonalização de Arnoldi. Abaixo apresentaremos uma breve discussão sobre a teoria e a implementação deste método. Para simplificar a notação, suprimiremos o subescrito "k" referente ao contador de números de iterações do solucionador externo e assumiremos $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{x}_k)$ e $\boldsymbol{J} = \boldsymbol{J}(\boldsymbol{x}_k)$.

A primeira etapa do método GMRES consiste na construção de uma base bem-condicionada, que define o subespaço de Krylov, dado por $\mathcal{K}_m(\boldsymbol{J}, \boldsymbol{v}_1) \equiv \text{span}\{\boldsymbol{v}_1, \boldsymbol{J}\boldsymbol{v}_1, \boldsymbol{J}^2\boldsymbol{v}_1, \ldots, \boldsymbol{J}^{m-1}\boldsymbol{v}_1\}$, onde $\boldsymbol{v}_1 = \boldsymbol{r}_{l-1} / \|\boldsymbol{r}_{l-1}\|$. Para tal construção pode ser empregada as seguintes técnicas de ortogonalização: *Gram-Schmidt* padrão, Gram-Schmidt modificada (GSM), GSM com re-ortogonalização ou *Householder*. Sendo as duas últimas técnicas as mais eficientes, porém, computacionalmente mais caras. A ortogonalização de Householder, conforme demonstrado em [188], produz uma melhoria na estabilidade numérica, quando a norma do vetor de resíduo tende ao limite inferior de precisão de máquina. Porém, quando comparada com a técnica GSM, este tipo de ortogonalização resulta em um aumento no número de operações aritiméticas e em perda de paralelismo. O processo de ortogonalização de Arnoldi constroi $\boldsymbol{V}_m = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\} \in \mathbf{R}^{N \times m}$ cujas colunas formam uma base M2-ortonormal em subespaço de Krylov, e a matriz $\overline{\boldsymbol{H}}_m$ de dimensão $(m+1) \times m$ do tipo *Hessenberg* superior. Em adição, esta construção satisfaz a seguinte relação

$$\boldsymbol{J}\boldsymbol{V}_m = \boldsymbol{V}_{m+1}\overline{\boldsymbol{H}}_m . \tag{7.6}$$

Utilizando a matriz definida por $\boldsymbol{V}_m$, assumimos que a solução de (7.1), i.e., minimização de (7.2), impondo (7.4), na l-ésima iteração possui a seguinte forma

$$\boldsymbol{d}_l = \boldsymbol{d}_{l-1} + \boldsymbol{V}_m\boldsymbol{y}, \tag{7.7}$$

onde $\boldsymbol{y} \in \mathbb{R}^m$ com $\boldsymbol{V}_m\boldsymbol{y} \in \mathcal{K}_m(\boldsymbol{J}, \boldsymbol{r}_{l-1})$. Utilizando (7.6) e (7.7) e Lembremos que $\boldsymbol{V}_{m+1}$ é uma base M2-ortonormal, podemos escrever a norma do vetor de resíduo (7.2), como se segue

$$\|\boldsymbol{r}_l\|_2 = \left\|\beta\mathbf{e}_1 - \overline{\boldsymbol{H}}_m\boldsymbol{y}\right\|_2, \tag{7.8}$$

onde $\beta = \|\boldsymbol{r}_{l-1}\|_2$ e $\mathbf{e}_1 \in \mathbb{R}^{m+1}$ é um vetor unitário com um único elemento diferente de zero na

primeira linha. A equação (7.8) forma um sistema de mínimo-quadrados tendo o vetor $\boldsymbol{y}$, que minimiza esta expressão, como solução. Aplicando-se nesta equação a fatorização QR, via rotações de *Givens*, o primeiro e o segundo termo da norma-M2 no lado direito de (7.8), são dados por $\beta \boldsymbol{Q}_m^T \mathbf{e}_1 = \bar{\boldsymbol{g}}_m$ e $\bar{\boldsymbol{R}}_m \boldsymbol{y}$, respectivamente. A matriz M2-ortonormal, $\boldsymbol{Q}_m \in \mathbb{R}^{(m+1)^2}$, contém os coeficientes das rotações de Givens para eliminação da sub-diagonal de $\bar{\boldsymbol{H}}_m = \boldsymbol{Q}_m \bar{\boldsymbol{R}}_m$. Com esta fatorização, o vetor solução, $\boldsymbol{y}_{N^*}$, é eficientemente obtido, resolvendo-se, via retro-substituições, o sistema triangular superior formado por $\bar{\boldsymbol{R}}_m \in \mathbb{R}^{(m+1)\times m}$, após remover a sua última linha igual a zero. Nesta resolução, o vetor de lado da mão-direita é dado por $\bar{\boldsymbol{g}}_{m,\,1:m}$. Consequentemente, o vetor de correção de Newton inexato é dado por $\boldsymbol{d}_{N,\,l} = \boldsymbol{d}_{l-1} + \boldsymbol{V}_m \boldsymbol{y}_{N^*}$ (ver (7.7)). Cada estágio da iteração de Arnoldi, $|g_{m,\,m+1}|$ é igual a norma-M2 do vetor de resíduo, o que oferece uma forma barata para realização do critério de parada (7.3).

Em arimética exata, quando $m = N$, i.e., ortogonalização completa, o GMRES converge em, no máximo, N iterações. Entretanto, por razões práticas, devemos restringir o valor máximo de m ($m<<N$) tal que os requisitos de memória sejam aceitáveis. Lembremos que, para um valor muito baixo de m poderá ocorrer o problema de estagnação [27]. A análise de convergência do método GMRES(m) utiliza polinômios de Chebyshev e pode ser encontrada em [52],[27].

O Algoritmo RGMRES(m), descrito abaixo, descreve a implementação do método do GMRES com reinício, ortogonalização GSM e pré-condicionamento à direita.

Algoritmo RGMRES(m)
("Right-preconditioned Generalised Minimal Residual with restart")

(L-1) Dado: $\boldsymbol{d}_{N,\,0} \in \mathbb{R}^N$, $NIL_{\text{máx}} \in \mathbb{Z}_+$ e $m \in (1, N]$;
(L-2) **PARA** $l \leftarrow 1$ **PASSO** 1 **ATÉ** $NIL_{\text{máx}}$ **FAÇA**:
(L-3) Calcular: $\boldsymbol{r}_{N,\,l-1} = -\boldsymbol{F} - \boldsymbol{J}\boldsymbol{d}_{N,\,l-1}$, $\beta = \|\boldsymbol{r}_{N,\,l-1}\|_2$ e $\boldsymbol{v}_1 = \boldsymbol{r}_{N,\,l-1}/\beta$;
(L-4) **PARA** $j \leftarrow 1$ **PASSO** 1 **ATÉ** m **FAÇA**:
(L-5) Resolver: $\boldsymbol{M}_R \boldsymbol{z}_j = \boldsymbol{v}_j$;
(L-6) Calcular: $\boldsymbol{w} = \boldsymbol{J}\boldsymbol{z}_j$;
(L-7) **PARA** $i \leftarrow 1$ **PASSO** 1 **ATÉ** j **FAÇA**:
(L-8) Calcular: $h_{i,j} = \boldsymbol{w}^T \boldsymbol{v}_i$ e $\boldsymbol{w} = \boldsymbol{w} - h_{i,j}\boldsymbol{v}_i$;
(L-9) Calcular: $h_{j+1,j} = \|\boldsymbol{w}\|_2$;
(L-10) **SE** $h_{j+1,j} \neq 0$ **ENTÃO** $\boldsymbol{v}_{j+1} \leftarrow \boldsymbol{w}/h_{j+1,j}$ **CASO CONTRÁRIO** $j \leftrightarrow m$;
(L-11) Resolver: $\boldsymbol{y}_{N^*} = \underset{\boldsymbol{y}_N \in \mathbf{R}^m}{\text{argmin}} \|\beta \mathbf{e}_1 - \bar{\boldsymbol{H}}_m \boldsymbol{y}_N\|_2$; onde $\bar{\boldsymbol{H}}_m = [h_{i,j}] \in \mathbb{R}^{(m+1)\times m}$;
(L-12) Calcule: $\boldsymbol{d}_{N,\,l} = \boldsymbol{d}_{N,\,l-1} + \boldsymbol{M}_R^{-1} \boldsymbol{V}_m \boldsymbol{y}_{N^*}$;
(L-13) **SE** ("CONVERGÊNCIA") **ENTÃO** $\boldsymbol{d}_N \leftarrow \boldsymbol{d}_{N,\,l}$; **RETORNE**.

No algoritmo da página anterior, a ação dos pré-condicionadores (à direita) ocorrem nas linhas

(L-5) e (L-12). O teste de quebra do processo de ortogonalização é realizado em (L-10). A convergência na linha (L-13) ocorre quando o critério de parada (7.3) for satisfeito e a divergência quando o contador de iterações, l, exceder o limite máximo definido por $L_{\text{máx}}$.

Uma variante flexível do método GMRES entitulada *flexible generalised minimal residual* (FGMRES) foi apresentada em [188], no qual, cada nova iteração, $\boldsymbol{d}_l$, é calculada em um espaço afim $\boldsymbol{d}_{l-1} + \text{span}\{\boldsymbol{Z}_m\}$, formado utilizando as m soluções do sistema pré-condicionado. O GMRES utiliza a norma do vetor de resíduo (6.71) como condição de parada, entretanto, isto pode não refletir aproximações precisas da solução, quando a matriz de iteração, $\boldsymbol{J}(\boldsymbol{x}_k)$, for mal-condicionada. Para mitigar este problema, o método *generalised minimal backward error* (GMBACK) com reinício foi proposto em [190], incluindo a variante flexível entitulada de *flexible generalised minimal backward error* (FGMBACK).

7.2.2. Termo Forçante

O termo forçante, introduzido em (7.1), é o parâmetro que determina o grau de precisão da solução aproximada do modelo linear que define a correção de Newton. Mais precisamente, este parâmetro corresponde ao erro relativo da norma do resíduo linear (7.3) utilizado no critério de parada do solucionador linear iterativo (e.g., método GMRES(m)). Sendo que, a cada iteração deve ser escolhido de forma a evitar o problema de *sobressolução*, principalmente, longe da raiz como nas primeiras iterações. Por exemplo, se escolhermos os termos forçantes de forma a impor um alto grau de precisão na aproximação da correção de Newton, a convergência do solucionador linear iterativo pode ser muito lenta. Em outro extremo, se os termos forçantes forem escolhidos com baixo grau de precisão, a convergência do *solucionador não-linear* (i.e., método de Newton) poderá ser muito lenta ou nunca ocorrer. Desta forma, uma judiciosa relação de compromisso deve ser estabelecida, tendo em mente, que a convergência local do método de Newton inexato é controlada pela sequência de termos forçantes.

Em [177], foram analisadas as diversas escolhas para a definição da sequência de termos forçantes, sendo elas:

Escolha 0: $\eta_k \leftarrow \eta_0 \qquad \eta_0 \in (0, 1)$, (7.9.a)

Escolha 1: $\eta_k \leftarrow 1/2^{k+1}$, (7.9.b)

Escolha 2: $\eta_k \leftarrow \min(1/(k+2), \|\boldsymbol{F}(\boldsymbol{x}_k)\|)$, (7.9.c)

Escolha 3: $\eta_k \leftarrow \gamma \cdot (\|\boldsymbol{F}(\boldsymbol{x}_k)\| / \|\boldsymbol{F}(\boldsymbol{x}_{k-1})\|)^{\alpha} \qquad \gamma \in [0, 1], \alpha \in (1, 2]$, (7.9.d)

Escolha 4: $\eta_k \leftarrow \|\boldsymbol{F}(\boldsymbol{x}_k) - \boldsymbol{M}_N(\boldsymbol{x}_{k-1} + \boldsymbol{d}_{N,k-1})\| / \|\boldsymbol{F}(\boldsymbol{x}_{k-1})\|$, e (7.9.e)

Escolha 5: $\eta_k \leftarrow \left\| \|\boldsymbol{F}(\boldsymbol{x}_k)\| - \|\boldsymbol{M}_N(\boldsymbol{x}_{k-1} + \boldsymbol{d}_{N,\,k-1})\| \right\| / \|\boldsymbol{F}(\boldsymbol{x}_{k-1})\|$. (7.9.f)

A Escolha 0 é trivial e produz uma sequência constante de termos forçantes sem controle no problema de sobressolução. Por exemplo, para uma sequência uniforme de aproximações modestas e próximas, podemos escolher $\eta_0 = 10^{-1}$ e $\eta_0 = 10^{-4}$, respectivamente. A Escolha 1, proposta em [16], permite aproximações com baixo-grau de precisão para k pequeno, evitando o problema de sobressolução no estágio inicial, mas, sem incorporar nenhuma informação sobre $\boldsymbol{F}$. Na Escolha 2, proposta em [32], tenta-se incorporar esta informação, porém, depende da escala (norma) de $\boldsymbol{F}$. Finalmente, as escolhas 4 e 5 refletem diretamente a concordância entre $\boldsymbol{F}$ e seu modelo local $\boldsymbol{M}_N$, calculados na iteração atual ($\boldsymbol{x}_k$) e prévia ($\boldsymbol{x}_{k-1}$), respectivamente. Apesar deste fato, resultados númericos demonstram um pequeno índice de sobressolução quando aplicado a problema práticos [30]. A caracterização teórica da razão de convergência local do método de Newton inexato, para as escolhas citadas acima, pode ser encontrada em [21],[181].

Obviamente, excluindo (7.9.a)-(7.9.c), as demais escolhas requerem a introdução de um mecanismo prático de proteção, para prevenir que o termo forçante possa ficar muito pequeno muito rapidamente. Lembremos que, para $k = 0$ temos $\eta_0 \leftarrow \eta_{\text{máx}}$, onde $\eta_{\text{máx}} < 1$ (está tipicamente entre 0,1 e 0,9), então, foi sugerida em [177] a seguinte proteção:

$$\eta_k \leftarrow \begin{cases} \text{máx}[\eta_k,\, \gamma \eta_{k-1}^{\alpha}], & \gamma \eta_{k-1}^{\alpha} > 0{,}1 \\ \eta_k, & \text{caso contrário} \end{cases}, \tag{7.10}$$

para $k > 0$. Para a Escolha 3, os parâmetros α e γ correspondem aos parâmetros em (7.9.d), enquanto, para as escolhas 4 e 5 devemos utilizar $\alpha = (1 + \sqrt{5})/2$ e $\gamma = 1$. Para definir um limite superior no termo forçante, uma proteção adicional requer que $\eta_k \leftarrow \min(\eta_k, \eta_{\max})$.

7.2.3. Globalização via Pesquisa-em-Linha

O critério de redução da função nível, na estratégia de pesquisa-em-linha, para o método de Newton inexato, é fundamentado na norma-M2 (abaixo o subescrito "M2" foi suprimido) do resíduo ou mapeamento não-linear. Neste caso, esta estratégia de globalização é praticamente a mesma utilizada no método de Newton, ver Algoritmo BLS. A única diferença reside no cálculo da taxa de inclinação, tax_inc, utilizada nos testes monotônico (6.11) e não-monotônico (6.20) para aceitação do fator de amortecimento. Para a determinação deste parâmetro, podemos substituir o vetor de resíduo (7.2) e o gradiente (6.13) em (6.12). Desta forma, obtemos:

$$\text{tax\_inc}_k = \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_{N,\,k} = -\boldsymbol{F}(\boldsymbol{x}_k)^T (\boldsymbol{F}(\boldsymbol{x}_k) + \boldsymbol{r}_{N,\,k}). \tag{7.11}$$

Como podemos observar, para o cálculo da expressão anterior, não é necessário determinar o gradiente através de um produto matriz por vetor, envolvendo a matriz jacobiana transposta, $\boldsymbol{J}(\boldsymbol{x}_k)^T$. No método de Newton inexato, a utilização da função nível associada com a norma-aN apresenta dificuldades, devido a necessidade da solução iterativa de um sistema jabobiano a cada passo da pesquisa-em-linha. Uma discussão detalhada na análise de convergência de métodos inexatos utilizando pesquisa-em-linha e outras estratégias pode ser encontrado em [180],[181]. Da teoria de convergência apresentada em [180], sabemos que para satisfazer simultaneamente a condição de norma residual (7.3) e o teste de monotonicidade (6.11), devemos ter $\alpha < \frac{1}{2}$, onde α é um parâmetro do teste (6.11).

7.2.4. Implementação Modificada

Com base na estrutura do algoritmo GN e na teoria descrita acima, foi desenvolvido o algoritmo GIN, versão inexata, que corresponde a uma implementação modificada do método de Newton inexato, globalizado com a técnica de pesquisa-em-linha. Este algoritmo pode ser descrito da seguinte forma:

Algoritmo GIN
("Global Inexact Newton")

(L-1) Dado: $\boldsymbol{x}_0 \in \mathbb{R}^N$, $\Xi \in [0, 1)$, $NI_{\text{máx}} \geq 0$, $NIJ_{\text{máx}} \geq 0$, $NIM_{\text{máx}} > 0$ e $0 < \eta_{\text{máx}} < 1$;
(L-2) $k \leftarrow 0$; $\eta_0 \leftarrow \eta_{\text{máx}}$;
(L-3) **PARA** $i \leftarrow 0$ **PASSO** 1 **ATÉ** $NI_{\text{máx}}$ **FAÇA** (malha principal):
(L-4) Escolher: η_k (termo forçante) utilizando (7.9.a)-(7.9.f);
(L-5) Aplicar proteção (7.10) em η_k tal que $\eta_k \in (0, \eta_{\text{máx}}]$;
(L-6) Calcular: $\boldsymbol{J}_i = \boldsymbol{J}(\boldsymbol{x}_k) \in \mathbb{R}^{N \times N}$ (matriz jacobiana);
(L-7) **SE** (atual_pc) **ENTÃO** calcule pré-condicionador;
(L-8) Determinar: $\boldsymbol{d}_{N,k} \in \mathbb{R}^N$ tal que $\|\boldsymbol{r}_{N,k}\|_2 \leq \eta_k \|\boldsymbol{F}(\boldsymbol{x}_k)\|_2$ †;
(L-9) **PARA** $l \leftarrow 0$ **PASSO** 1 **ATÉ** $l_{\text{máx}}$ **FAÇA** (malha simplificada):
(L-10) Determinar: $\lambda_k \in \mathbb{R}$ via Algoritmo BLS (com modificação (7.11));
(L-11) $\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_{N,k}$;
(L-12) **SE** ("CONVERGÊNCIA" **OU** "DIVERGÊNCIA") **ENTÃO** $\boldsymbol{x}_* \leftarrow \boldsymbol{x}_{k+1}$; **RETORNE**.
(L-13) **SE** ($l_{\text{máx}} > 0$) **ENTÃO**
(L-14) **SE** ($\Xi = 0$ **OU** $f(\boldsymbol{x}_{k+1})/f(\boldsymbol{x}_k) < \Xi$) **ENTÃO** (executar iteração modificada)
(L-15) Determinar: $\bar{\boldsymbol{d}}_{N,k} \in \mathbb{R}^N$ tal que $\|\bar{\boldsymbol{r}}_{N,k+1}\| \leq \eta_{k+1} \|\boldsymbol{F}(\boldsymbol{x}_{k+1})\|$;
(L-16) $\boldsymbol{d}_{N,k} \leftarrow \bar{\boldsymbol{d}}_{N,k}$ (correção simplificada);
(L-17) **CASO CONTRÁRIO**

(L-18) **SE** ($f(\boldsymbol{x}_{k+1})/f(\boldsymbol{x}_k) > 1$) **ENTÃO FIM PARA**;

(L-19) $k \leftarrow k+1$.

---

† $\bar{\boldsymbol{r}}_{N,k+1} = -\boldsymbol{F}(\boldsymbol{x}_{k+1}) - \boldsymbol{J}_i \boldsymbol{d}_{N,k}$.

No algoritmo acima, foram empregados os mesmos testes de convergência e de divergência adotados no Algoritmo GN. Exceto pelo parâmetro, η_{max}, os parâmetros de entrada são os mesmos descritos no Algoritmo GN. A escolha do termo forçante é efetuada nas linhas (L-4) e (L-5). Como podemos observar, em (L-7), o pré-conditionador é atualizado somente se bandeira atual_pc estiver alta. Na linha (L-8) o solucionador interno é acionado para a determinação da correção de Newton inexata. O Algoritmo RGMRES(m) pode ser utilizado como solucionador interno. Conforme indicado na linha (L-10), o Algoritmo BLS com a modificação (7.11) no cálculo da taxa de inclinação é utilizado para a pesquisa-em-linha com retrocedimento. As condições para entrada e saída da malha simplificada, definida pelas linhas (L-9)-(L-18), são as mesmas adotadas para o Algoritmo GN.

7.3. Método do Tensor Inexato

O método do tensor inexato pode ser considerado como uma variação do método do tensor, discutido no capítulo anterior, no qual a correção, definida como *correção do tensor inexata*, corresponde a solução aproximada do modelo do tensor (6.31). Antes de iniciarmos a discussão na teoria dos métodos iterativos para a solução aproximada de (6.31), é conveniente re-escrever a correção do tensor (6.46) da seguinte forma

$$\boldsymbol{d}_{T,k} = (1-\mu_*)\boldsymbol{d}_{N,k} + \mu_* \boldsymbol{d}_{-1,k} + \mu_* \boldsymbol{s}_k + \nu \boldsymbol{z}\,, \tag{7.12}$$

onde

$$\boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d}_{-1,k} = -\boldsymbol{F}(\boldsymbol{x}_{k-1})\,, \tag{7.13}$$

$$\mu_* = \beta_*^2 / m_k\,, \tag{7.14}$$

e $\boldsymbol{d}_{N,k}$ é a correção de Newton dada por (6.8). Os vetores $\boldsymbol{s}_k$ e $\boldsymbol{z}$ são definidos em (6.32) e (6.50.a)-(6.50.b), respectivamente, e os escalares m_k e ν são definidos em (6.34) e (6.48), respectivamente. O parâmetro β_* é a solução de (6.47), onde o polinômio quadrático $q_k(\beta)$ definido em (6.43), possui coeficientes dados por (6.44.a) e

$$c_{2,k} = -\frac{1}{m_k}\boldsymbol{s}_k^T(\boldsymbol{d}_{N,k} + \boldsymbol{d}_{-1,k} + \boldsymbol{s}_k) = \frac{1}{m_k}(c_{0,k} + \boldsymbol{s}_k^T \boldsymbol{d}_{-1,k} + \sqrt{m_k})\,, \tag{7.15}$$

A expressão (7.15) produz um resultado equivalente a (6.44.b), porém, em uma forma conveniente

para a determinação da correção do tensor inexata.

A dificuldade na aplicação de técnicas em subespaço de Krylov para a solução aproximada do modelo do tensor (6.30) se deve à presença do termo quadrático neste modelo. Não obstante, para o modelo do tensor de posto um (6.31), com uma interpolação de um ponto passado, eficientes metodologias tem sido recentemente propostas. Estas metodologias resultam nos processos de solução simplificada [32], modificada [114] e completa [31],[34] a serem discutidos abaixo.

Em vista de (7.12), o método do tensor inexato (sem amortecimento) pode ser descrito como uma versão inexata do processo iterativo (6.23.a)-(6.23.e). A correção de Newton inexata associada à (6.23.c), e presente em (6.23.d), pode ser obtida via (7.1). De imediato, a correção do tensor inexata, resolução aproximada de (6.23.d), poderia ser obtida utilizando um método iterativo (e.g. GMRES) para a resolução aproximada dos sistemas jacobianos (7.13), (6.50.a) e (6.50.b) associados com

$$\text{Encontrar: } \boldsymbol{d}_{-1,\,k} \in \mathrm{R}^N \text{ tal que } \left\| \boldsymbol{F}(\boldsymbol{x}_{k-1}) + \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d}_{-1,\,k} \right\| \leq \zeta_k \left\| \boldsymbol{F}(\boldsymbol{x}_{k-1}) \right\| \tag{7.16}$$

e

$$\text{Encontrar: } \boldsymbol{y} \in \mathrm{R}^N \text{ tal que } \left\| \boldsymbol{s}_k - \boldsymbol{J}(\boldsymbol{x}_k)^T \boldsymbol{y} \right\| \leq \xi_k \left\| \boldsymbol{s}_k \right\|, \tag{7.17.a}$$

$$\text{Encontrar: } \boldsymbol{z} \in \mathrm{R}^N \text{ tal que } \left\| \boldsymbol{y} - \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{z} \right\| \leq \varepsilon_k \left\| \boldsymbol{y} \right\|, \tag{7.17.b}$$

respectivamente. Porém, além da solução iterativa (7.1) para determinação da correção de Newton inexata, três sistemas adicionais (7.16)-(7.17.b), sendo um transposto (7.17.a), devem ser resolvidos iterativamente para determinação da correção do tensor inexata. Obviamente, este tipo solução possui um custo computacional (tempo de processamento e implementação) consideravelmente mais elevado do que o método padrão. Destarte, torna se mais difícil a escolha da sequência de termos forçantes ζ_k, ξ_k e ε_k. Para contornar estas dificuldades, discutiremos a teoria e a implementação dos processos de solução simplificada e modificada, propostos em [32] e [114], respectivamente. Também discutir-se-á um processo para a solução completa de (6.23.d), proposto em [31],[34], que resolve o seguinte problema de minimização,

$$\text{Encontrar: } \boldsymbol{d}_{T,\,k} \in \mathrm{R}^N \text{ tal que } \left\| \boldsymbol{M}_T(\boldsymbol{x}_k + \boldsymbol{d}_{T,\,k}) \right\| - \left\| \boldsymbol{M}_{TE}(\boldsymbol{x}_k + \boldsymbol{d}_{T,\,k}) \right\| \leq \rho_k \left\| \boldsymbol{F}(\boldsymbol{x}_k) \right\|, \tag{7.18}$$

onde $\rho_k \in \mathrm{R}{:}(0,1)$ refere-se ao termo forçante que define o grau de precisão da solução aproximada. Esta minimização considera o quanto a norma do modelo do tensor se aproxima da norma de um modelo do tensor que assume a correção do tensor como exata. A definição de $\boldsymbol{M}_{TE}(\boldsymbol{x}_k + \boldsymbol{d}_T)$ apresentar-se-á abaixo.

Além da solução iterativa do modelo do tensor para obtenção da correção do tensor inexata, apresentar-se-á uma discussão na escolha da sequência de termos forçantes e na técnica de

globalização via pesquisa-em-linha: estratégia padrão [32] e curvilinear [31],[34]. algoritmos apresentar-se-ão descrevendo detalhadamente a implementação proposta. No melhor do nosso conhecimento, uma análise de convergência local para os métodos do tensor inexato citados acima ainda não foi apresentada na literatura técnica especializada. Além do método do tensor descrito anteriormente, convém citar o método do tensor-GMRES [191]. Os processos de solução simplificada e modificada permitem uma total flexibilidade na escolha do solucionador linear interno. Enquanto o processo de solução completa está limitado ao uso do método TGMRES-Bt.

7.3.1. Solução Iterativa do Modelo do Tensor

Para a versão escalar da solução iterativa do modelo do tensor, nos processos de solução simplificada e modificada, além de (7.2), é conviniente introduzir o vetor de resíduo associado com (7.13), dado por

$$\boldsymbol{r}_{-1,\,k,\,l} \overset{\text{def}}{=} -\boldsymbol{F}(\boldsymbol{x}_{k-1}) - \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d}_{-1,\,k,\,l} \tag{7.19}$$

O subescrito l representa o contador de iterações do solucionador interno. Utilizando a norma de (7.19), pode se definir o seguinte critério de parada para o solucionador linear interno [183]

$$\left\|\boldsymbol{r}_{-1,\,k,\,l}\right\| / \left\|\boldsymbol{F}(\boldsymbol{x}_{k-1})\right\| < \eta_k\,, \tag{7.20}$$

onde η_k é uma tolerância relativa de parada (ou termo forçante) na k-ésima iteração do solucionador externo.

Para facilitar a discussão deste método, introduziremos os seguintes vetores de resíduo

$$\boldsymbol{r}_{i,\,k,\,l} = -\boldsymbol{F}_{i,\,k} - \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d}_{i,\,k,\,l}\,, \tag{7.21}$$

onde $i = 1, \ldots, t$. Convém ressaltar, que estamos trabalhando com $t = 2$, porém, a discussão e a implementação apresentada aqui é válida para a solução simultânea de sistemas lineares com t múltiplos vetores de lado da mão-direita. Em adição, temos que: $\boldsymbol{F}_{1,\,k} = \boldsymbol{F}(\boldsymbol{x}_k)$, $\boldsymbol{F}_{2,\,k} = \boldsymbol{F}(\boldsymbol{x}_{k-1})$, $\boldsymbol{r}_{1,\,k,\,l} = \boldsymbol{r}_{N,\,k,\,l}$, $\boldsymbol{r}_{2,\,k,\,l} = \boldsymbol{r}_{-1,\,k,\,l}$, $\boldsymbol{d}_{1,\,k,\,l} = \boldsymbol{d}_{N,\,k,\,l}$ e $\boldsymbol{d}_{2,\,k,\,l} = \boldsymbol{d}_{-1,\,k,\,l}$. Introduzindo o bloco de vetores: $\boldsymbol{\Xi} = \left[\boldsymbol{\xi}_1 \ldots \boldsymbol{\xi}_i \ldots \boldsymbol{\xi}_t\right] \in \mathrm{R}^{N\times t}$ com $(\boldsymbol{\Xi}, \boldsymbol{\xi}) = (\boldsymbol{R}_{k,\,l}, \boldsymbol{r}_{k,\,l}), (\boldsymbol{B}\boldsymbol{F}_k, \boldsymbol{F}_k), (\boldsymbol{D}_{k,\,l}, \boldsymbol{d}_{k,\,l})$, podemos re-escrever (7.21) de forma mais compacta como

$$\boldsymbol{R}_{k,\,l} = -\boldsymbol{B}\boldsymbol{F}_k - \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{D}_{k,\,l}\,. \tag{7.22}$$

No caso da versão em bloco, podemos definir os seguintes critérios de parada:

$$\left\|\boldsymbol{R}_{k,\,l}\right\| / \left\|\boldsymbol{F}(\boldsymbol{x}_k)\right\| < \eta_k\ , \tag{7.23}$$

e

$$\|\boldsymbol{r}_{i,k,l}\| / \|\boldsymbol{F}(\boldsymbol{x}_k)\| < \eta_{i,k}, \tag{7.24}$$

onde $i = 1, \ldots, t$. Pode-se observar, facilmente, que o critério (7.24) busca a aproximação isolada de cada componente do bloco $\boldsymbol{D}_{k,l}$, enquanto (7.23) considera uma aproximação global.

No contexto do processo de solução completa do modelo do tensor, é conviniente introduzir o vetor de resíduo associado com (6.31), e definido como

$$\boldsymbol{r}_{T,k,l} \stackrel{\text{def}}{=} \boldsymbol{F}(\boldsymbol{x}_k) + \frac{1}{2}\boldsymbol{a}_k(\boldsymbol{s}_k^T \boldsymbol{d}_{T,k,l})^2 + \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d}_{T,k,l}. \tag{7.25}$$

Introduzindo o vetor de resíduo $\boldsymbol{r}_{TE,k,l}$ associado a $\boldsymbol{M}_{TE}(\boldsymbol{x}_k + \boldsymbol{d}_{T,k,l})$, o critério de parada para minimização do vetor de resíduo (7.25) está fundamentado no seguinte teste

$$(\|\boldsymbol{r}_{T,k,l}\| - \|\boldsymbol{r}_{TE,k,l}\|) / \|\boldsymbol{F}(\boldsymbol{x}_k)\| < \eta_k. \tag{7.26}$$

Por sua vez, o teste acima está diretamente associado com (7.18). Se houver interesse, no processo de solução completa, o vetor de erro associado a correção do tensor pode ser calculado via

$$\boldsymbol{e}_{T,k,l} \stackrel{\text{def}}{=} \boldsymbol{d}_{T,k,l} - \boldsymbol{d}_{T,k}, \tag{7.27}$$

onde $\boldsymbol{d}_{T,k}$ é a correção do tensor (exata) dada por (6.46). Podemos facilmente observar que o cálculo do vetor de erro implica na solução (exata) do modelo do tensor.

7.3.1.1. Pré-Condicionamento

Da discussão acima, podemos concluir que o pré-condicionamento de (7.1) e (7.16), nos processos de solução simplificada e modificada, pode ser implementado seguindo a mesma estrutura aplicada em (7.5). Já no processo de solução completa, precisamos realizar o pré-condicionamento de (6.31) para a resolução de (7.18). Utilizando as definições previamente estabelecidas em (7.5), podemos escrever o modelo do tensor pré-condicionado da seguinte forma

$$\underbrace{\boldsymbol{M}_1^{-1}\boldsymbol{J}(\boldsymbol{x})\boldsymbol{M}_2^{-1}}_{\tilde{\boldsymbol{J}}(\boldsymbol{x})} \underbrace{\boldsymbol{M}_2\boldsymbol{d}}_{\tilde{\boldsymbol{d}}} = -\underbrace{\boldsymbol{M}_1^{-1}\boldsymbol{F}(\boldsymbol{x})}_{\tilde{\boldsymbol{F}}(\boldsymbol{x})} - \frac{1}{2}\boldsymbol{M}_1^{-1}\boldsymbol{a}\Big(\underbrace{\boldsymbol{s}^T\boldsymbol{M}_2^{-1}\overbrace{\boldsymbol{M}_2\boldsymbol{d}}^{\tilde{\boldsymbol{d}}}}_{\beta}\Big)^2. \tag{7.28}$$

Conforme veremos a seguir o tipo de pré-condicionamento (à esquerda ou à direita) resultará em diferentes metodologias para o processo de solução completa.

7.3.1.2. Técnicas em Subespaço de Krylov

Abaixo discutiremos a teoria para a aplicação da técnica em subespaço de Krylov nos processos

de solução simplificada, modificada e completa do modelo do tensor. A implementação será apresentar-se-á sob forma de algoritmos.

7.3.1.2.a. Solução Simplificada e Modificada

Para entendermos os processos de solução simplificada e modificada do modelo do tensor, consideremos a expressão da correção do tensor (7.12). Como podemos observar, para determinar a versão inexata desta correção precisamos calcular, com um grau de aproximação desejado, os vetores $\boldsymbol{d}_{N,k}$, $\boldsymbol{d}_{-1,k}$ e $\boldsymbol{z}_k$. Lembremos que $\boldsymbol{d}_{N,k}$ é a correção de Newton inexata definida em (7.1), $\boldsymbol{d}_{-1,k}$ é definido em (7.16) e $\boldsymbol{z}_k$ é definido em (7.17.a)-(7.17.b). No processo de solução simplificada devemos assumir que o vetor $\boldsymbol{z}$ não produz nenhuma contribuição na correção do tensor e $\zeta_k = \eta_k$. Esta suposição só é falsa quando o modelo do tensor não possuir uma raiz. Isto é, quando o polinômio quadrático $q_k(\beta)$ (ver (6.43)) possui raízes complexas e a correção do tensor é obtida via minimização deste polinômio. Esta situação geralmente está limitada as primeiras iterações, quando a correção ainda está longe da raiz. À medida em que melhores estimativas vão sendo obtidas, o modelo do tensor tende a produzir uma melhor interpolação do mapeamento não-linear resultando em um modelo com raiz e, consequentemente com o vetor $\boldsymbol{z}$ igual a zero.

No processo de solução modificada, proposto em [114], o modelo do tensor padrão (6.31) é substituido pela expressão (6.51). Lembremos da discussão no capítulo anterior, esta expressão só introduz uma modificação no modelo padrão quando este não possui uma raiz, i.e., a solução é conduzida via minimização. Neste caso, o parâmetro, $\tau \in \mathrm{R}:(0, 1]$, que define o peso do termo quadrático no modelo do tensor, é selecionado de forma que este possua uma raiz real. Mais precisamente, o fator de peso é selecionado no limite quando as raízes do polinômio quadrático $q(\beta)$ forem iguais. Desta forma, o termo quadrático não é completamente negligenciado como na solução simplificada. O processo de solução modificada apresentado acima requer apenas a solução de (7.1) e (7.16), sendo assim, podemos adotar a mesma metodologia utilizada no processo de solução simplificada.

Para calcularmos os vetores $\boldsymbol{d}_{N,k}$ e $\boldsymbol{d}_{-1,k}$ necessários nos processos de solução simplificada e modificada, a metodologia mais simples consiste em utilizar duas vezes um método iterativo linear (e.g. GMRES(m)) para a determinação individual de cada um dos vetores. Seguindo o procedimento proposto em [32], na solução do sistema (6.8), a iteração inicial é igual a zero ($\boldsymbol{d}_{N,k,0} = \boldsymbol{0}$), porém, na solução do sistema (7.13), conforme sugerido em [32], a iteração inicial é igual a solução aproximada do sistema (6.8) na iteração anterior ($\boldsymbol{d}_{-1,k,0} = \boldsymbol{d}_{N,k-1}$). Esta escolha fundamenta-se na suposição da matriz jacobiana não variar muito entre duas iterações,

$\boldsymbol{J}(\boldsymbol{x}_k) \sim \boldsymbol{J}(\boldsymbol{x}_{k-1})$.

Uma outra possibilidade é a solução simultânea dos problemas (7.1) e (7.16), associados com as correções inexatas $\boldsymbol{d}_{N,k}$ e $\boldsymbol{d}_{-1,k}$, utilizando um método iterativo operando em subespaço de Krylov de bloco 2 [114]. Neste trabalho, foi empregado um método que consiste em uma extensão para operação em bloco do método escalar GMRES, visto anteriormente. Conforme procedimento descrito acima, os vetores de correção iniciais dados por $\boldsymbol{d}_{N,k,0} = \boldsymbol{0}$ e $\boldsymbol{d}_{-1,k,0} = \boldsymbol{d}_{N,k-1}$ são determinados conforme procedimento descrito acima. Seguindo notação introduzida acima, iremos assumiremos que $\boldsymbol{D}_{0,k} \in \mathbb{R}^{N\times t}$ é a aproximação inicial da solucão para o bloco de vetores $\boldsymbol{D}_k$, e $\boldsymbol{R}_{0,k} \in \mathbb{R}^{N\times t}$ é o bloco de vetores de resíduo iniciais associado a $\boldsymbol{D}_{0,k}$ calculado via (7.22).

A seguir, suprimiremos o subescrito k, para simplificar a notação e assumir $\boldsymbol{J} = \boldsymbol{J}(\boldsymbol{x}_k)$. Assim como no método GMRES, a primeira etapa do método GMRES-Bt consiste na construção de uma base bem-condicionada que define o subespaço de Krylov de bloco, dado por: $\mathcal{K}_m(\boldsymbol{J}, \boldsymbol{V}_t) \equiv \text{span}\{\boldsymbol{V}_t, \boldsymbol{J}\boldsymbol{V}_t, \ldots, \boldsymbol{J}^{m-1}\boldsymbol{V}_t\}$, onde $\boldsymbol{V}_t = \left[\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_t\right]^T \in \mathbb{R}^{N\times t}$ e $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_t \in \mathbb{R}^N$ são vetores M2-ortonormais. Também temos que: $\mathcal{K}_m(\boldsymbol{J}, \boldsymbol{V}_t) = \bigcup_{i=1}^{t} \mathcal{K}_m(\boldsymbol{J}, \boldsymbol{v}_i)$. Os vetores iniciais $\boldsymbol{v}_i$'s são gerados apartir de um processo de ortogonalização (e.g., MGS), de tal forma que

$$\boldsymbol{R}_0 = \left[\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_t\right]\boldsymbol{\beta} \in \mathbb{R}^{N\times t}, \tag{7.29}$$

onde $\boldsymbol{\beta} \in \mathbb{R}^{t\times t}$ é uma matriz triangular superior. Agora, utilizando o bloco de vetores iniciais, $\boldsymbol{V}_t$, e aplicando a variante da ortogonalização bloco Arnoldi [27], entitulada bloco *Arnoldi-Ruhe* [192], obtemos o seguinte resultado

$$\boldsymbol{J}\boldsymbol{V}_m = \boldsymbol{V}_{m+t}\overline{\boldsymbol{H}}_m, \tag{7.30}$$

que podemos comparar com (7.5). Na expressão acima a matriz, $\overline{\boldsymbol{H}}_m$, construída pelo processo de ortogonalização, possui uma dimensão $(m+t)\times m$ e estrutura do tipo *banda-Hessenberg* superior com t sub-diagonais. Caso seja detectada uma divisão por zero (ou quebra) em uma determinada etapa do processo de ortogonalização de Arnoldi-Ruhe, a dimensão do bloco inicial é reduzida de um, antes de seguir com a ortogonalização. Se a dimensão do bloco, após redução de um, for igual a zero, então, em aritimética exata, a solução foi atingida e pode ser construída conforme descrita abaixo, utilizando um subespaço de Krylov de bloco com a dimensão dos vetores de bloco menor ou igual a m.

Utilizando a mesma matriz $\boldsymbol{V}_m$, assume-se que a solução simultânea de (7.1) e (7.16), i.e., minimização de (7.2) impondo (7.3), e de (7.19) impondo (7.20), respectivamente, assume a forma descrita em (7.7). Generalizando estes resultados para qualquer dimensão de bloco t e escrevendo sob forma de bloco de vetores, assumiremos uma solução do tipo:

$$\boldsymbol{D}_l = \boldsymbol{D}_{l-1} + \boldsymbol{V}_m \boldsymbol{Y}, \tag{7.31}$$

onde $\boldsymbol{Y} \in \mathbb{R}^{m \times t}$ e $\boldsymbol{V}_m \boldsymbol{Y} \in \mathcal{K}_m(\boldsymbol{J}, \boldsymbol{R}_{l-1})$. Utilizando (7.30) e (7.31) e lembrando que $\boldsymbol{V}_{m+t}$ é uma base ortonormal-M2, podemos escrever a norma-M2 (norma de *Frobenius*) do bloco de vetores de resíduo (7.22), como se segue

$$\left\| \boldsymbol{R}_l \right\|_2 = \left\| \mathbf{E}_t \boldsymbol{\beta} - \overline{\boldsymbol{H}}_m \boldsymbol{Y} \right\|_2 \tag{7.32}$$

na qual podemos observar a analogia com (7.8). Na expressão acima, foi introduzido $\mathbf{E}_t = \left[\mathbf{e}_1 \ \cdots \ \mathbf{e}_i \ \cdots \ \mathbf{e}_t \right]^T \in \mathbb{R}^{(m+t) \times t}$, onde $\mathbf{e}_i$ é um vetor unitário com um único elemento diferente de zero na *i*-ésima linha. A equação (7.32) forma um sistema de mínimos-quadrados tendo o bloco de vetores $\boldsymbol{Y}$ que minimiza a expressão como solução. Semelhante a versão escalar, aplicando-se em (7.32) a fatorização QR, via rotações de *Givens*, podemos eliminar as t sub-diagonais em $\overline{\boldsymbol{H}}_m = \boldsymbol{Q}_m \overline{\boldsymbol{R}}_m$, onde $\boldsymbol{Q}_m \in \mathbf{R}^{(m+t)^2}$ é uma matriz M2-ortonormal que contém os coeficientes das rotações. Com isto, o bloco de vetores solução, $\boldsymbol{Y}_*$, pode ser eficientemente obtido, resolvendo-se, via retro-substituições, t idênticos, sistemas lineares triangulares formados por $\overline{\boldsymbol{R}}_m \in \mathbb{R}^{(m+t) \times m}$, removendo-se as suas últimas t linhas iguais a zero. O bloco de vetores de lado da mão-direita nesta solução é dado por $\boldsymbol{Q}_m^T \mathbf{E}_t \boldsymbol{\beta} = \overline{\boldsymbol{G}}_m = \left[\bar{\boldsymbol{g}}_{1,m} \ \cdots \ \bar{\boldsymbol{g}}_{2,m} \right]$. Com este resultado o bloco de vetores de correção inexata, $\boldsymbol{D}_*$, pode ser determinado via (7.31). Em cada estágio da iteração de Arnoldi, a norma-M2 do vetor de resíduo associado com o *i*-ésimo lado de mão-direita é dada por $\left\| g_{p,\, m,\, m+1:m+p} \right\|_2$. Isto nos oferece uma forma barata para realização do critério de parada (7.3) e (7.20) para determinação de $\boldsymbol{d}_{N,k} = \boldsymbol{d}_{1*}$ e de $\boldsymbol{d}_{-1,k} = \boldsymbol{d}_{2*}$, respectivamente.

Para ortogonalização completa, $m = N$, em arimética exata, o GMRES-Bt convergirá em, no máximo, N iterações. Entretanto, pelas mesmas razões discutidas no método GMRES devemos ter preferencialmente, $m \ll N$, porém com m acima do limite inferior, no qual ocorre o problema de estagnação. A análise de convergência para o método GMRES-Bt(m) é mais difícil do que em sua versão de único-vetor, $t = 1$, que equivale ao método GMRES. Isto possivelemente deve-se à dificuldade de se estabelecer um convincente análogo para a relação com os polinômios de Chebyshev [27].

Abaixo é descrita a versão de bloco do Algoritmo RGMRES(m) entitulado Algoritmo RGMRES-Bt(m).

Algoritmo RGMRES-Bt(m)
("Right Preconditioned Generalised Minimal Residual-Block with restart")

(L-1) Dado: $\boldsymbol{D}_0 \in \mathbb{R}^{N \times t}$, $NIL_{\text{máx}} \in \mathbb{Z}_+$ e $m \in (1, N]$;

(L-2) **PARA** $l \leftarrow 1$ **PASSO** 1 **ATÉ** $NIL_{\text{máx}}$ **FAÇA**:

(L-3) Calcular: $\boldsymbol{R}_{l-1} = -\boldsymbol{BF} - \boldsymbol{J}(\boldsymbol{x})\boldsymbol{D}_{l-1}$;

(L-4) Resolver: $\boldsymbol{R}_{l-1} = \boldsymbol{V}_t\boldsymbol{\beta}$ tal que $\boldsymbol{V}_t \in \mathbb{R}^{N\times t}$ é M2-ortonormal e $\boldsymbol{\beta} \in \mathbb{R}^{t\times t}$ é triangular superior;

(L-5) **PARA** $j' \leftarrow 1$ **PASSO** 1 **ATÉ** m **FAÇA**:

(L-6) $j \leftarrow j' + t - 1$;

(L-7) Resolver: $\boldsymbol{M}_R\boldsymbol{z}_j = \boldsymbol{v}_j$;

(L-8) Calcular: $\boldsymbol{w} = \boldsymbol{J}(\boldsymbol{x})\boldsymbol{z}_j$;

(L-9) **PARA** $i \leftarrow 1$ **PASSO** 1 **ATÉ** j **FAÇA**:

(L-10) Calcular: $h_{i,j} = \boldsymbol{w}^T\boldsymbol{v}_i$ e $\boldsymbol{w} = \boldsymbol{w} - h_{i,j}\boldsymbol{v}_i$;

(L-11) Calcular: $h_{j+1,j} = \|\boldsymbol{w}\|_2$;

(L-12) **SE** $h_{j+1,j} \neq 0$ **ENTÃO** $\boldsymbol{v}_{j+1} \leftarrow \boldsymbol{w}/h_{j+1,j}$ **CASO CONTRÁRIO** $t' \leftarrow t' - 1$;

(L-13) **SE** $t' = 0$ **ENTÃO** $j' \leftrightarrow m$;

(L-14) Resolver: $\boldsymbol{Y}_* = \operatorname{argmin}_{\boldsymbol{Y}\in\mathbb{R}^{m\times t}} \left\|\mathbf{E}_t\boldsymbol{\beta} - \overline{\boldsymbol{H}}_m\boldsymbol{Y}\right\|_2$, onde $\overline{\boldsymbol{H}}_m = [h_{i,j}] \in \mathbb{R}^{(m+t)\times m}$;

(L-15) Calcular: $\boldsymbol{D}_l = \boldsymbol{D}_{l-1} + \boldsymbol{M}_R^{-1}\boldsymbol{V}_m\boldsymbol{Y}_*$;

(L-16) **SE** ("CONVERGÊNCIA") **ENTÃO** $\boldsymbol{D}_* \leftarrow \boldsymbol{D}_l$; **RETORNE**.

Como podemos observar, para $t = 1$, o algoritmo acima representa uma versão escalar coincidente com o Algoritmo RGMRES(m). Em adição, a ação dos pré-condicionadores (a direita) ocorrem nas linhas (L-7) e (L-15). O teste de quebra do processo de ortogonalização é realizado em (L-12). A convergência, na linha (L-16), ocorre quando o critério de parada (7.23) ou (7.24) for satisfeito e a divergência quando o contador de iterações, l, exceder o limite máximo definido por $L_{\text{máx}}$.

Convém ressaltar, que a variante flexível do método GMRES-Bt, com reinício e pré-condicionamento, também é possível.

7.3.1.2.b. Solução Completa

No processo de solução completa, a resolução aproximada do modelo do tensor é obtida iterativamente e consiste no processo de miminização descrito em (7.18). Devido a existência do termo quadrático, o subespaço de Krylov deve ser ampliado de forma a conter as direções $\boldsymbol{a}_k$ e/ou $\boldsymbol{s}_k$, além da direção do vetor de resíduo inicial associado com a parte linear do modelo do tensor. Isto resulta em métodos operando com bloco de subespaço de Krylov inicial com duas ou três colunas. Para simplificar notação, abaixo iremos suprimiremos o subescrito "k" referente ao contador de número de iterações do solucionador externo.

Exceto pela definição do bloco de vetores inciais, $\boldsymbol{R}_0$, a primeira etapa do método TGMRES-Bt para a construção de um bem-condicionado subespaço de Krylov de bloco, $\mathcal{K}_m(\boldsymbol{J}, \boldsymbol{V}_t)$, é idêntica a

descrita para o método GMRES-Bt. De acordo com a teoria a ser descrita abaixo, o bloco de vetores iniciais para o método TGMRES de bloco 2 e 3 são dados por: $\boldsymbol{R}_0 = \left[\boldsymbol{r}_0 \; \boldsymbol{a}\right]$ e $\boldsymbol{R}_0 = \left[\boldsymbol{s} \; \boldsymbol{r}_0 \; \boldsymbol{a}\right]$, respectivamente.

Assumindo uma solução do tipo (7.7), podemos expressar o termo quadrático do modelo do tensor (7.25), como se segue

$$\beta = \boldsymbol{s}^T \boldsymbol{d}_l = \boldsymbol{s}^T \boldsymbol{d}_{l-1} + \boldsymbol{s}^T \boldsymbol{V}_m \boldsymbol{y}. \tag{7.33}$$

Então, a idéia básica do método TGMRES-Bt é transformar o termo quadrático em (6.31) em uma função apenas do último elemento do vetor $\boldsymbol{y}$. Esta transformação produzirá duas diferentes metodologias, dependendo de o vetor $\boldsymbol{s}$ estar ou não contido no espaço definido por $\boldsymbol{V}_m$.

Quando $\boldsymbol{s} \subset \boldsymbol{R}_0$, podemos facilmente observar que $\boldsymbol{s}^T \boldsymbol{V}_m = \|\boldsymbol{s}\|_2 \boldsymbol{e}_p^T$ com $1 \le p \le t$. Nesta situação, podemos fazer a seguinte transformação:

$$\boldsymbol{y} = \boldsymbol{P}_m \hat{\boldsymbol{y}}, \tag{7.34}$$

onde $\boldsymbol{P}_m \in \mathrm{R}^{m \times m}$ é uma matriz de permutação utilizada para mover p-ésima componente, $\|\boldsymbol{s}\|_2$, do vetor $\boldsymbol{s}^T \boldsymbol{V}_m$, para a última componente deste vetor. Esta transformação operará na permutação das colunas da matriz $\overline{\boldsymbol{H}}_m$. Conforme veremos a seguir, esta situação é de menor interesse prático (só é valida para o modelo do tensor sem ou com pré-condicionamento à esquerda).

A outra situação, quando $\boldsymbol{s} \not\subset \boldsymbol{R}_0$, é mais complexa por envolver um processo de ortogonalização. Neste caso, devemos introduzir a seguinte matriz M2-ortonormal, $\boldsymbol{Q}_m^{(1)} \in \mathrm{R}^{m \times m}$ (ver equação (7.35) no fundo da próxima página), tal que

$$\boldsymbol{y} = \boldsymbol{Q}_m^{(1)} \hat{\boldsymbol{y}}, \tag{7.36}$$

onde $\hat{q}_{i,j}^{(1)} = -\left\|\boldsymbol{V}_j^T \boldsymbol{s}\right\|_2 / \boldsymbol{v}_i^T \boldsymbol{s}$ para $i = t, \ldots, m$ e $j = 1, \ldots, m-1$. Os elementos $\hat{q}_{i,j}^{(1)}$ da matriz $\hat{\boldsymbol{Q}}_m^{(1)} \in \mathrm{R}^{m \times m}$ são calculados de forma a assegurar a ortogonalidade desta matriz. Enquanto os elementos da matriz diagonal, $\boldsymbol{D}_m^{(1)} \in \mathrm{R}^{m \times m}$, são calculados de forma a transfomar $\boldsymbol{Q}_m^{(1)}$ em uma

$$\boldsymbol{Q}_m^{(1)} = \underbrace{\begin{bmatrix} \boldsymbol{v}_1^T \boldsymbol{s}_k & \ldots & \boldsymbol{v}_1^T \boldsymbol{s}_k & \boldsymbol{v}_1^T \boldsymbol{s}_k & \ldots & \boldsymbol{v}_1^T \boldsymbol{s}_k & \boldsymbol{v}_1^T \boldsymbol{s}_k \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \hat{q}_{t,1}^{(1)} & \ldots & \boldsymbol{v}_t^T \boldsymbol{s} & \boldsymbol{v}_t^T \boldsymbol{s} & \ldots & \boldsymbol{v}_t^T \boldsymbol{s} & \boldsymbol{v}_t^T \boldsymbol{s} \\ 0 & \ldots & \hat{q}_{t+1,2}^{(1)} & \boldsymbol{v}_{t+1}^T \boldsymbol{s} & \ldots & \boldsymbol{v}_{t+1}^T \boldsymbol{s} & \boldsymbol{v}_{t+1}^T \boldsymbol{s} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & 0 & \ldots & \boldsymbol{v}_{m-1}^T \boldsymbol{s} & \boldsymbol{v}_{m-1}^T \boldsymbol{s} \\ 0 & \ldots & 0 & 0 & \ldots & \hat{q}_{m,m-1}^{(1)} & \boldsymbol{v}_m^T \boldsymbol{s} \end{bmatrix}}_{\hat{\boldsymbol{Q}}_m^{(1)}} \underbrace{\begin{bmatrix} d_1^{(1)} & \ldots & 0 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & d_t^{(1)} & 0 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & d_{t+1}^{(1)} & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & 0 & \ldots & d_{m-1}^{(1)} & 0 \\ 0 & \ldots & 0 & 0 & \ldots & 0 & d_m^{(1)} \end{bmatrix}}_{\boldsymbol{D}_m^{(1)}}, \tag{7.35}$$

matriz M2-ortonormal.

Para simplificar a notação, assumiremos para $\boldsymbol{s} \not\subset \boldsymbol{R}_{l-1}$ que a matriz $\boldsymbol{\Gamma}_m$ e o escalar γ referem-se à $\boldsymbol{Q}_m^{(1)}$ e à $\left\|\boldsymbol{V}_m^T\boldsymbol{s}\right\|_2$, respectivamente. Enquanto que, para $\boldsymbol{s} \subset \boldsymbol{R}_{l-1}$ esta matriz e este escalar referem-se à $\boldsymbol{P}_m$ e à $\|\boldsymbol{s}\|_2$, respectivamente. Substituindo a expressão (7.34)/(7.36) em (7.25) obtemos

$$\bar{\boldsymbol{r}}_T = \bar{\boldsymbol{r}}_{l-1} + \overline{\boldsymbol{H}}_m\boldsymbol{\Gamma}_m\hat{\boldsymbol{y}} + \frac{1}{2}\bar{\boldsymbol{a}}(\boldsymbol{s}^T\boldsymbol{d}_{l-1} + \gamma\hat{\boldsymbol{y}}_m)^2 \tag{7.37}$$

onde

$$\bar{\boldsymbol{r}}_T = \boldsymbol{V}_{m+t}^T\boldsymbol{r}_T, \tag{7.38.a}$$

$$\bar{\boldsymbol{r}}_{l-1} = \boldsymbol{V}_{m+t}^T\boldsymbol{r}_{l-1}, \tag{7.38.b}$$

$$\bar{\boldsymbol{a}} = \boldsymbol{V}_{m+t}^T\boldsymbol{a}. \tag{7.38.c}$$

A matriz $\overline{\boldsymbol{H}}_m\boldsymbol{\Gamma}_m$ é uma matriz banda-Hessenberg superior com $t+1$ sub-diagonais. Da expressão acima, podemos observar que $\beta = \boldsymbol{s}^T\boldsymbol{d}_{l-1} + \gamma\hat{\boldsymbol{y}}_m$. Embora envolva longas operações matemáticas, pode-se demonstrar com facilidade que o produto $\overline{\boldsymbol{H}}_m\boldsymbol{Q}_m^{(1)}$ pode ser eficientemente calculado em apenas $2(m+t-1)+t$ multiplicações e $4(m+t-1)+t$ adições. Para o cálculo de $\overline{\boldsymbol{H}}_m\boldsymbol{P}_m$ o custo são apenas m permutações.

Assim como no GMRES, o processo de eliminação das $t+1$ sub-diagonais da matriz $\overline{\boldsymbol{H}}_m\boldsymbol{\Gamma}_m$ pode ser conduzido de forma eficiente através de uma série de rotações de Givens ou reflexões de Householder. Estas operações podem ser representadas por uma matriz ortogonal $\boldsymbol{Q}_m^{(2)} \in \mathbb{R}^{m\times m}$. Aplicando esta matriz aos vetores de resíduo associado às correções de Newton e do tensor, obtemos:

$$\tilde{\boldsymbol{r}}_{N,\,l} = \tilde{\boldsymbol{r}}_{l-1} + \tilde{\boldsymbol{H}}_m\hat{\boldsymbol{y}}_N, \tag{7.39}$$

e

$$\tilde{\boldsymbol{r}}_{T,\,l} = \tilde{\boldsymbol{r}}_{l-1} + \tilde{\boldsymbol{H}}_m\hat{\boldsymbol{y}}_T + \frac{1}{2}\tilde{\boldsymbol{a}}(\boldsymbol{s}^T\boldsymbol{d}_{l-1} + \gamma\hat{\boldsymbol{y}}_{T,\,m})^2, \tag{7.40}$$

respectivamente, onde

$$\tilde{\boldsymbol{r}}_{T,\,l} = \boldsymbol{Q}_m^{(2)}\bar{\boldsymbol{r}}_{T,\,l}, \tag{7.41.a}$$

$$\tilde{\boldsymbol{r}}_{l-1} = \boldsymbol{Q}_m^{(2)}\bar{\boldsymbol{r}}_{l-1}, \tag{7.42.a}$$

$$\tilde{\boldsymbol{a}} = \boldsymbol{Q}_m^{(2)}\bar{\boldsymbol{a}}, \tag{7.42.b}$$

$$\tilde{\boldsymbol{H}}_m = \boldsymbol{Q}_m^{(2)}\overline{\boldsymbol{H}}_m\boldsymbol{\Gamma}_m. \tag{7.43.a}$$

A determinação da correção do tensor inexata envolve a minimização da norma-M2 do vetor de resíduo (7.40), o que é equivalente a minimização das últimas $t+1$ linhas deste vetor, dadas por:

$$\tilde{\boldsymbol{r}}_{T,\,m:m+t} = \tilde{\boldsymbol{r}}_{l-1,\,m:m+t} + \left[\tilde{h}_{m,\,m}\ \underbrace{0\ \ldots\ 0\ 0}_{t\text{ zeros}}\right]^T\hat{\boldsymbol{y}}_{T,\,m} + \frac{1}{2}\tilde{\boldsymbol{a}}_{m:m+t}(\boldsymbol{s}^T\boldsymbol{d}_{l-1} + 2\boldsymbol{s}^T\boldsymbol{d}_{l-1}\gamma\hat{\boldsymbol{y}}_{T,\,m} + \gamma^2\hat{\boldsymbol{y}}_{T,\,m}^2). \tag{7.44}$$

Como podemos observar, minimizar a expressão acima em norma-M2 é equivalente a minimização de uma equação quártica em $\hat{y}_m$. Por sua vez, este problema é equivalente a determinação da raiz da equação cúbica resultante da diferenciação da equação quártica que se deseja minimizar. Resolvendo a equação cúbica, e escolhendo a raiz com menor magnitude absoluta, obtemos $\hat{y}_{T^*,\,m}$ e consequentemente $\beta_* = \boldsymbol{s}^T\boldsymbol{d}_{l-1} + \gamma\hat{y}_{T^*,\,m}$. Com este tipo de escolha conservadora, obtém-se uma correção do tensor inexata mais próxima da correção de Newton, conforme foi adotado no método do tensor. Utilizando β_* podemos obter as outras componentes do vetor $\hat{\boldsymbol{y}}_{T^*}$. Para tal, vamos considerar as seguintes componentes do vetor de resíduo (7.40)

$$\tilde{\boldsymbol{r}}_{T,\,l,\,1:m-1} = \tilde{\boldsymbol{r}}_{l-1,\,1:m-1} + \tilde{\boldsymbol{H}}_{m,\,1:m-1,\,m}\hat{y}_{T^*,\,m} + \tilde{\boldsymbol{H}}_{m,\,1:m-1,\,1:m-1}\hat{\boldsymbol{y}}_{T,\,1:m-1} + \frac{1}{2}\tilde{\boldsymbol{a}}_{1:m-1}\beta_*^2. \tag{7.45}$$

Fazendo o vetor acima igual a zero, obtemos $\hat{\boldsymbol{y}}_{T^*,\,1:m-1}$, e consequentemente a solução $\hat{\boldsymbol{y}}_{T^*} = \left[\hat{\boldsymbol{y}}_{T^*,\,1:m-1}\ \hat{y}_{T^*,\,m}\right]^T$. Com este resultado, a correção do tensor inexata pode ser determinada através de uma expressão do tipo (7.7), dada por: $\boldsymbol{d}_{T,\,l} = \boldsymbol{d}_{l-1} + \boldsymbol{V}_m\boldsymbol{\Gamma}_m\hat{\boldsymbol{y}}_{T^*}$. Utilizando os produtos resultantes da solução iterativa do modelo do tensor, também podemos calcular facilmente e praticamente sem custo computacional a correção de Newton inexata. Para tal, devemos considerar as seguintes componentes do vetor de resíduo (7.39)

$$\tilde{\boldsymbol{r}}_{N,\,l,\,1:m} = \tilde{\boldsymbol{r}}_{l-1,\,1:m} + \tilde{\boldsymbol{H}}_{m,\,1:m,\,1:m}\hat{\boldsymbol{y}}_N. \tag{7.46}$$

Analogamente a (7.45), fazendo o vetor acima igual a zero obtemos $\hat{\boldsymbol{y}}_{N^*}$. Com este resultado, a correção de Newton inexata pode ser calculada através de uma expressão do tipo (7.7), dada por: $\boldsymbol{d}_{N,\,l} = \boldsymbol{d}_{l-1} + \boldsymbol{V}_m\boldsymbol{\Gamma}_m\hat{\boldsymbol{y}}_{N^*}$. Vale ressaltar que a norma de $\boldsymbol{r}_{N,\,l}$ é dada por $\|\tilde{\boldsymbol{r}}_{N,\,l,\,m+1:m+t}\|$.

Com base nos resultados acima e seguindo a mesma filosofia do método GMRES, o critério de parada (7.26) pode ser elaborado utilizando as linhas do vetor de resíduo (7.40), localizadas abaixo da parte triangular da matriz $\tilde{\boldsymbol{H}}_m$. Sendo assim, a implementação de (7.26), assume a seguinte forma [31],[34]:

$$\|\tilde{\boldsymbol{r}}_{T,\,l,\,m:m+t}\| - \|\tilde{\boldsymbol{r}}_{T,\,l,\,m}\| \le \rho\|\boldsymbol{F}(\boldsymbol{x})\|. \tag{7.47}$$

Sem incluir uma contribuição de $\tilde{\boldsymbol{H}}_m$ no seu cálculo, uma alternativa mais simplificada para a implementação de (7.26), também proposta em [31],[34], é dada por

$$\|\tilde{\boldsymbol{r}}_{T,\,l,\,m+1:m+t}\| \le \rho\|\boldsymbol{F}(\boldsymbol{x})\|. \tag{7.48}$$

A determinação do termo do lado esquerdo da inequação acima, implica apenas na minimização de uma equação quadrática em β^2. Vale ressaltar que, quando o modelo do tensor possuir uma raiz, a solução iterativa descrita acima, corresponderá a uma aproximação da correção do tensor obtida via solução direta (ver capítulo anterior), com precisão definida pela a tolerância, ρ. Na situação

contrária, temos um problema de minimização e a correção do tensor obtida com (7.47) ou (7.48) divergirá da solução direta. Nesta situação, a norma do resíduo, definido em (7.25), será mais elevada para a correção do tensor inexata. Isto ocorre devido ao fato da minimização do vetor $\boldsymbol{y}_T$ ter sido separada em duas sub-minimizações envolvendo a princípio o escalar $y_{T,\,m}$, e depois o sub-vetor $\boldsymbol{y}_{T,\,1:m-1}$. É fácil verificar esta situação quando ocorre uma quebra-afortunada.

Apesar de estarmos interessado na implementação com pré-condicionamento à direita, vale ressaltar que o método TGMRES-B3 requer distintas implementações para o pré-condionamento à direita e à esquerda. Em adição, a implementação sem pré-condicionamento possui a mesma estrutura de quando pré-condicionado à esquerda. Já o método TGMRES-B2, não possui esta particularidade, isto, por que não introduz $\boldsymbol{s}_k$ no bloco de subespaço de Krylov inicial, $\boldsymbol{R}_0$. Em [34], mais detalhadamente em [31] é sugerido também uma versão bloco-2+.

Seguindo a teoria descrita acima, os métodos TGMRES-Bt com reinício e pré-condicionamento à direita, foram implementados sob forma do algoritmo RTGMRES-Bt(m), descrito abaixo. Destacando-se o fato do pre-condicionamento a direita produzir, $\gamma_m = \left\| \boldsymbol{V}_m^T \boldsymbol{M}_R^{-T} \boldsymbol{s} \right\|_2$, o que elimina a possibilidade de uma solução via transformação (7.34), quando $\boldsymbol{s} \subset \boldsymbol{V}_m$.

Algoritmo RTGMRES-Bt(m)

("Right-preconditioned Tensor Generalised Minimal Residual of Block-t with restart")

(L-1) Dado: $\boldsymbol{d}_0 \in \mathbb{R}^N$, $NIL_{\text{máx}} \in \mathbf{Z}_+$ e $m \in (1, N]$;

(L-2) **PARA** $l \leftarrow 1$ **PASSO** 1 **ATÉ** $NIL_{\text{máx}}$ **FAÇA**:

(L-3) Calcular: $\boldsymbol{r}_{l-1} = \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{J}(\boldsymbol{x})\boldsymbol{d}_{l-1}$;

(L-4) Para bloco-2: $t = 2$, $\boldsymbol{R}_{l-1} = \left[\boldsymbol{r}_{l-1} \;\; \boldsymbol{a} \right] \in \mathbb{R}^{N \times 2}$ (algoritmo RTGMRES-B2(m)), ou
para bloco-3: $t = 3$, $\boldsymbol{R}_{l-1} = \left[\boldsymbol{s} \;\; \boldsymbol{r}_{l-1} \;\; \boldsymbol{a} \right] \in \mathbb{R}^{N \times 3}$ (algoritmo RTGMRES-B3(m));

(L-5) Resolver: $\boldsymbol{R}_{l-1} = \boldsymbol{V}_t \boldsymbol{\beta}$ tal que $\boldsymbol{V}_t \in \mathbb{R}^{N \times t}$ é M2-ortonormal e $\boldsymbol{\beta} \in \mathbb{R}^{t \times t}$ é triangular superior;

(L-6) **PARA** $j' \leftarrow 1$ **PASSO** 1 **ATÉ** m **FAÇA**:

(L-7) $j \leftarrow j' + t - 1$;

(L-8) Resolver: $\boldsymbol{M}_R \boldsymbol{z}_j = \boldsymbol{v}_j$;

(L-9) Calcular: $\boldsymbol{w} = \boldsymbol{J}(\boldsymbol{x})\boldsymbol{z}_j$;

(L-10) **PARA** $i \leftarrow 1$ **PASSO** 1 **ATÉ** j **FAÇA**:

(L-11) Calcular: $h_{i,j} \leftarrow \boldsymbol{w}^T \boldsymbol{v}_i$ e $\boldsymbol{w} \leftarrow \boldsymbol{w} - h_{i,j}\boldsymbol{v}_i$;

(L-12) Calcular: $h_{j+1,j} \leftarrow \|\boldsymbol{w}\|_2$ e ;

(L-13) **SE** $h_{j+1,j} \neq 0$ **ENTÃO** $\boldsymbol{v}_{j+1} \leftarrow \boldsymbol{w} / h_{j+1,j}$ **CASO CONTRÁRIO** $t \leftarrow t - 1$;

(L-14) **SE** $t = 1$ **ENTÃO** $j \leftarrow m$;

(L-15) Se $(t = 3 \wedge \boldsymbol{M}_R = \mathbf{1}_N)$ então $\boldsymbol{\Gamma}_m = \boldsymbol{P}_m$ e $\gamma = \|\boldsymbol{s}\|_2$; caso contrário $\boldsymbol{\Gamma}_m = \boldsymbol{Q}_m^{(1)}$ e $\gamma = \left\| \boldsymbol{V}_m^T \boldsymbol{M}_R^{-T} \boldsymbol{s} \right\|_2$;

(L-16) Resolver: $\hat{\boldsymbol{y}}_T = \operatorname{argmin}_{\hat{\boldsymbol{y}} \in \mathbf{R}^m} \left\| \tilde{\boldsymbol{r}}_{l-1} + \overline{\boldsymbol{H}}_m \boldsymbol{\Gamma}_m \hat{\boldsymbol{y}} + \frac{1}{2}\tilde{\boldsymbol{a}}(\boldsymbol{s}^T \boldsymbol{d}_{l-1} + \gamma \hat{y}_m)^2 \right\|_2$,

(L-17) onde $\overline{\boldsymbol{H}}_m\boldsymbol{\Gamma}_m = [h_{i,j}] \in \mathbb{R}^{(m+t)\times m}$;

(L-18) Calcular: $\boldsymbol{d}_{T,\,l} = \boldsymbol{d}_{l-1} + \boldsymbol{M}_R^{-1}\boldsymbol{V}_m\boldsymbol{\Gamma}_m\hat{\boldsymbol{y}}_T$;

(L-19) **SE** ("CONVERGÊNCIA") **ENTÃO**

(L-20) Resolver: $\hat{\boldsymbol{y}}_N = \operatorname{argmin}_{\hat{\boldsymbol{y}} \in \mathbf{R}^m} \left\| \beta \boldsymbol{e}_1 + \overline{\boldsymbol{H}}_m\boldsymbol{\Gamma}_m\hat{\boldsymbol{y}} \right\|_2$;

(L-21) Calcular: $\boldsymbol{d}_{N,\,l} = \boldsymbol{d}_{l-1} + \boldsymbol{M}_R^{-1}\boldsymbol{V}_m\boldsymbol{\Gamma}_m\hat{\boldsymbol{y}}_N$; **RETORNE**;

(L-22) **CASO CONTRÁRIO** $\boldsymbol{d}_l \leftarrow \boldsymbol{d}_{T,\,l}$.

Para $t = 2$ (bloco-2) e $t = 3$ (bloco-3), o algoritmo descrito acima corresponde aos algoritmos RTGMRES-B2(m) e RTGMRES-B3(m), respectivamente. Na linha (L-4) é definido o bloco de vetores iniciais a depender do algoritmo B2 ou B3. O teste de quebra no processo de ortogonalização é conduzido em (L-13), o que resulta na redução do bloco, até satisfazer a condição imposta em (L-14) e que corresponde a quebra-afortunada. O pré-condicionamento à direita é aplicado nas linhas (L-8), (L-15) e (L-21). Convém observar, em (L-15), a diferente implementação no método TGMRES-B3 para a versão sem pré-condicionamento. A covergência na linha (L-19) ocorre, se o critério de convergância (7.26), implementado por (7.47) ou (7.48), for satisfeito. O algoritmo acima fundamenta-se no processo de *Arnoldi-Ruhe* desenvolvido para solução com bloco de vetor [192],[27].

7.3.2. Termo Forçante

No contexto dos métodos do tensor inexato, apenas a Escolha 0 (7.9.a) tem sido utilizada [32],[31],[34],[114]. Os métodos publicados na literatura técnica especializada não abordam, sob o ponto de vista experimental e/ou teórico (análise de convergência), a questão da escolha da sequência dos termos forçantes para se evitar o problema de sobressolução do modelo do tensor, utilizando os processos de solução discutidos acima. Intuitivamente, no processo de solução simplificada ou modificada, a escolha da sequência de termo forçante para solução de (7.1) e (7.16) evitando o problema de sobressolução pode ser a mesma adotada no método de Newton inexato, ver (7.9.a)-(7.9.f). Lembremos que, a mesma sequência de termos forçantes em (7.1) será empregada em (7.16) [32]. Também investigaremos experimentalmente, as seguintes escolhas, respectivamente análogas à Escolha 4 (7.9.e) e à Escolha 5 (7.9.e), e dadas por:

Escolha 6: $\eta_k \leftarrow \left\| \boldsymbol{F}(\boldsymbol{x}_k) - \boldsymbol{M}_T(\boldsymbol{x}_{k-1} + \boldsymbol{d}_{T,\,k-1}) \right\| / \left\| \boldsymbol{F}(\boldsymbol{x}_{k-1}) \right\|$, e (7.49.a)

Escolha 7: $\eta_k \leftarrow \left| \left\| \boldsymbol{F}(\boldsymbol{x}_k) \right\| - \left\| \boldsymbol{M}_T(\boldsymbol{x}_{k-1} + \boldsymbol{d}_{T,\,k-1}) \right\| \right| / \left\| \boldsymbol{F}(\boldsymbol{x}_{k-1}) \right\|$. (7.49.b)

Para aplicação das escolhas acima, foi adotado o seguinte procedimento. Na pesquisa-em-linha

com a estratégia padrão, quando a correção do tensor for selecionada, a escolha 6 ou 7 é utilizada, caso contrário, a escolha 4 ou 5 será utilizada. Na estratégia de pesquisa-em-linha curvilinear, do capítulo anterior, sabemos que quando o fator de amortecimento se aproxima de zero, $\lambda \to 0$, a direção da correção do tensor curvilinear se aproxima da direção da correção de Newton. De fato, para um fator de amortecimento muito pequeno temos que $\mu_* = O(\lambda^2)$, o que resulta em $\boldsymbol{d}_{T,k-1}(\lambda_{k-1}) = \lambda_{k-1}\boldsymbol{d}_{N,k-1} + O(\lambda_{k-1}^2)(-\boldsymbol{d}_{N,k-1} + \boldsymbol{d}_{-1,k-1} + \boldsymbol{s}_{k-1} + \nu z)$, ver equação (7.12). Nesta situação, temos que: $\boldsymbol{M}_T(\boldsymbol{x}_{k-1} + \overbrace{\boldsymbol{d}_{T,k-1}(\lambda_{k-1})}^{\boldsymbol{d}_{T,k-1}}) \to \boldsymbol{M}_N(\boldsymbol{x}_{k-1} + \overbrace{\lambda_{k-1}\boldsymbol{d}_{N,k-1}}^{\boldsymbol{d}_{N,k-1}})$, e com isto, as escolhas 6 e 7, devem produzir aproximadamente o mesmo efeito das escolhas 4 e 5, respectivamente. A adição de proteções, para ampliar a margem de segurança das escolha 6 e 7, conduzir-se-á através de (7.10).

Por envolver um problema de minimização, a escolha da sequência de termo forçante para evitar sobressolução, empregando o processo de solução completa, possui uma maior complexidade, quando comparado com o procedimento acima.

7.3.3. Globalização via Pesquisa-em-Linha

No método do tensor inexato, pelas mesmas razões apresentadas na discussão da versão inexata do método de Newton, consideraremos apenas a pesquisa-em-linha fundamentada na função nível associada com a norma-M2.

7.3.3.1. Estratégia Padrão

A aplicação da estratégia padrão desenvolvida para o método tensor, requer o cálculo da norma-M2 do gradiente da função nível, que por sua vez, envolve o produto matriz-jacabiana-transposta por vetor. Caso este produto não possa ser efetuado, o teste executado na linha (L-8) do Algoritmo BLS-TS, deve ser simplifcado para

$$\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_{T,k} < 0\,. \tag{7.50}$$

Com isto, o parâmetro γ, definido no Algoritmo BLS-TS, não atua no teste de suficiente descida. A inequação (7.50) pode ser calculada utilizando os produtos da solução iterativa do modelo do tensor. De fato, para os processos de solução simplificada e modificada, temos os vetores de resíduo, $\boldsymbol{r}_{N,k}$ e $\boldsymbol{r}_{-1,k}$, disponíveis das aproximções (7.1) e (7.16), respectivamente. Neste caso, o teste (7.50) pode ser eficientemente calculado como se segue,

$$\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_{T,k} = -\boldsymbol{F}(\boldsymbol{x}_k)^T \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{d}_{T,k} = -\boldsymbol{F}(\boldsymbol{x}_k)^T((1-\mu_k)(\boldsymbol{F}(\boldsymbol{x}_k) + \boldsymbol{r}_{N,k}) + \mu_k(\boldsymbol{F}(\boldsymbol{x}_k) + \boldsymbol{r}_{-1,k} - \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{s}_k))\,. \tag{7.51}$$

Lembremos que, para a solução modificada, temos $\mu_k = \mu_k(\tau)$. Para o processo de solução

completa, temos disponível o vetor de resíduo, $\boldsymbol{r}_{T,\,k}$, da solução aproximada de (7.18), significando que (7.50) pode ser calculada atráves da seguinte expressão

$$\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_{T,\,k} \;=\; -\boldsymbol{F}(\boldsymbol{x}_k)^T(\boldsymbol{F}(\boldsymbol{x}_k) + \frac{1}{2}\boldsymbol{a}_k(\boldsymbol{s}_k^T\boldsymbol{d}_{T,\,k})^2 - \boldsymbol{r}_{T,\,k})\,. \tag{7.52}$$

Considerando as colocações acima, o Algoritmo BLS-TS pode ser eficientemente utilizado para condução da pesquisa-em-linha no método do tensor inexato.

7.3.3.b. Estratégia Curvilinear

Para a aplicação da técnica de pesquisa-em-linha curvilinear, utilizando os subprodutos do método TGMRES-Bt, é preciso introduzir o seguinte bloco de vetores iniciais: $\boldsymbol{R}_{k,\,l-1} = \left[\lambda_{k,\,j}\boldsymbol{r}_{k,\,l-1}\;\boldsymbol{a}_k\right]$ para bloco-2 e $\boldsymbol{R}_{k,\,l-1} = \left[\boldsymbol{s}_k\;\lambda_{k,\,j}\boldsymbol{r}_{k,\,l-1}\;\boldsymbol{a}_k\right]$ para bloco-3. Com esta modificação, o vetor de resíduo (7.40) é dado por:

$$\tilde{\boldsymbol{r}}_{T,\,k,\,l} \;=\; \lambda_{k,\,j}\tilde{\boldsymbol{r}}_{k,\,l-1} + \tilde{\boldsymbol{H}}_m\boldsymbol{\Gamma}_m\hat{\boldsymbol{y}} + \frac{1}{2}\tilde{\boldsymbol{a}}_k(\boldsymbol{s}_k^T\boldsymbol{d}_{k,\,l-1} + \gamma\hat{\boldsymbol{y}}_m)^2\,. \tag{7.53}$$

Como podemos observar na expressão acima, o fator de amortecimento, $\lambda_{k,\,j}$, multiplica $\tilde{\boldsymbol{r}}_{k,\,l-1}$ após as transformações envolvendo as matrizes $\boldsymbol{\Gamma}_m$ e $\boldsymbol{Q}_m^{(2)}$, desta forma, a maior parte do trabalho para a obtenção da solução em subespaço de Krylov é poupado. Em adição, pode-se demonstrar, facilmente, que a minimização do vetor de resíduo (7.53) resulta na seguinte correção do tensor curvilinear inexata (com pré-condicionamento):

$$\boldsymbol{d}_T(\lambda_{k,\,j}) \;=\; \lambda_{k,\,j}\boldsymbol{d}_{k,\,l-1} + \boldsymbol{M}_R^{-1}\boldsymbol{V}_m\boldsymbol{\Gamma}_m\hat{\boldsymbol{y}}_T, \tag{7.54}$$

onde $\hat{\boldsymbol{y}}_T$ é o minimizador em norma-M2 de (7.53) e também uma função do fator de amortecimento $\lambda_{k,\,j}$. O custo computacional para o cálculo da correção curvilinear via (7.53) é relativamente baixo (complexidade $O(m^2)$). Após as modificações introduzidas acima o Algoritmo BLS pode ser utilizado para realização da pesquisa-em-linha curvilinear.

7.3.4. Implementação Modificada

Fundamentado na teoria discutida acima, foi desenvolvida uma implementação modificada do método do tensor inexato globalizado com a técnica de pesquisa-em-linha (estratégia padrão ou curvilinear). O Algoritmo GIT, que descreve esta implementação, representa a versão inexata do Algoritmo GT e pode ser descrito como se segue:

Algoritmo GIT

("Global Inexact Tensor")

(L-1) Dado: $\boldsymbol{x}_0 \in \mathbb{R}^N$, $\Xi \in (0, 1)$, $NI_{\text{máx}} \geq 0$, $NIJ_{\text{máx}} \geq 0$, $NIM_{\text{máx}} > 0$ e $0 < \eta_{\text{máx}} < 1 \in \mathbb{R}$;

(L-2) $k \leftarrow 0$; $\eta_0 \leftarrow \eta_{\text{máx}}$;

(L-3) **PARA** $i \leftarrow 0$ **PASSO** 1 **ATÉ** $NJI_{\text{máx}}$ **FAÇA**: (malha principal)

(L-4) Escolher: η_k (termo forçante) utilizando (7.9.a)-(7.9.f),(7.49.a)-(7.49.b);

(L-5) Aplicar proteção: $\eta_k \in (0, \eta_{\text{máx}}]$ utilizando (7.10);

(L-6) Calcular: $\boldsymbol{J}_i = \boldsymbol{J}(\boldsymbol{x}_k) \in \mathbb{R}^{N \times N}$; (matriz jacobiana)

(L-7) **SE** (atual_pc) **ENTÃO** calcule pré-condicionador;

(L-8) Determinar: $\boldsymbol{d}_{N,k} \in \mathbb{R}^N$ tal que $\|\boldsymbol{r}_{N,k}\|_2 \leq \eta_k \|\boldsymbol{F}(\boldsymbol{x}_k)\|_2$;

(L-9) **SE** $(k > 0)$ **ENTÃO**

(L-10) Calcular: $\boldsymbol{s}_k = \boldsymbol{x}_k - \boldsymbol{x}_{k-1}$ e $\boldsymbol{a}_k = \frac{2}{m_k}(\boldsymbol{F}(\boldsymbol{x}_{k-1}) - \boldsymbol{F}(\boldsymbol{x}_k) - \boldsymbol{J}(\boldsymbol{x}_k)\boldsymbol{s}_k)$;

(L-11) Determinar: $\boldsymbol{d}_{T,k} \in \mathbb{R}^N$ tal que $\left| \|\boldsymbol{r}_{T,k}\|_2 - \|\boldsymbol{r}_{TE,k}\|_2 \right| \leq \eta_k \|\boldsymbol{F}(\boldsymbol{x}_k)\|_2$;

(L-12) **PARA** $l \leftarrow 0$ **PASSO** 1 **ATÉ** $NIM_{\text{máx}}$ **FAÇA**: (malha simplificada)

(L-13) **SE** $(k > 0)$ **ENTÃO**

(L-14) Determinar: $\lambda_k \in \mathbb{R}$ via Algoritmo BLS-TS, com $\boldsymbol{d}_k$ entre $\boldsymbol{d}_{T,k}$ ou $\boldsymbol{d}_{N,k}$, ou

(L-15) determinar: $\lambda_k \in \mathbb{R}$ via algoritmo BLS-TC, com $\boldsymbol{d}_k = \boldsymbol{d}_T(\lambda_k)$;

(L-16) **CASO CONTRÁRIO**

(L-17) Determinar: $\lambda_k \in \mathbb{R}$ via Algoritmo BLS, com $\boldsymbol{d}_k \leftarrow \boldsymbol{d}_{N,k}$;

(L-18) $\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k$

(L-19) **SE** ("CONVERGÊNCIA" **OU** "DIVERGÊNCIA") **ENTÃO** $\boldsymbol{x}_* \leftarrow \boldsymbol{x}_{k+1}$ **RETORNE**;

(L-20) **SE** $(l_{max} > 1)$ **ENTÃO**

(L-21) **SE** $(\hat{f}_+ / \hat{f}_c < \Xi)$ **ENTÃO** (realize a iteração modificada)

(L-22) Determinar: $\bar{\boldsymbol{d}}_{N,k} \in \mathbb{R}^N$ tal que $\|\bar{\boldsymbol{r}}_{N,k+1}\| \leq \eta_{k+1} \|\boldsymbol{F}(\boldsymbol{x}_{k+1})\|$ †;

(L-23) Calcular: $\boldsymbol{s}_{k+1} = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$ e $\boldsymbol{a}_{k+1} = \frac{2}{m_{k+1}}(\boldsymbol{F}(\boldsymbol{x}_k) - \boldsymbol{F}(\boldsymbol{x}_{k+1}) - \boldsymbol{J}_i \boldsymbol{s}_{k+1})$;

(L-24) Determinar: $\bar{\boldsymbol{d}}_{T,k} \in \mathbb{R}^N$ tal que $\left| \|\bar{\boldsymbol{r}}_{T,k+1}\|_2 - \|\bar{\boldsymbol{r}}_{TE,k+1}\|_2 \right| \leq \eta_{k+1} \|\boldsymbol{F}(\boldsymbol{x}_{k+1})\|$ ††;

(L-25) $\boldsymbol{d}_{T,k} \leftarrow \bar{\boldsymbol{d}}_{T,k}$; $\boldsymbol{d}_{N,k} \leftarrow \bar{\boldsymbol{d}}_{N,k}$; (correções simplificadas)

(L-26) **CASO CONTRÁRIO**

(L-27) **SE** $(\hat{f}_+ / \hat{f}_c > 1)$ **ENTÃO FIM PARA**;

(L-28) $k \leftarrow k + 1$.

---

† $\bar{\boldsymbol{r}}_{N,k+1} = -\boldsymbol{F}(\boldsymbol{x}_{k+1}) - \boldsymbol{J}_i \bar{\boldsymbol{d}}_{N,k}$.

†† $\bar{\boldsymbol{r}}_{T,k+1} = \boldsymbol{F}(\boldsymbol{x}_{k+1}) + \boldsymbol{J}_i \bar{\boldsymbol{d}}_{T,k} + \frac{1}{2}\boldsymbol{a}_{k+1}(\boldsymbol{s}_{k+1}^T \bar{\boldsymbol{d}}_{T,k})^2$.

No algoritmo acima, foram empregados os mesmos testes de convergência e de divergência adotados no Algoritmo GT. Assim como, no Algoritmo GIN foi introduzida na (L-7) uma condição para a atualização do pré-condicionador. As linhas (L-11) e (L-24) referem se ao processo de

solução completa, porém, para os processos de solução simplificada e modificada basta alterar estas linhas à determinação das correções $\boldsymbol{d}_{T,k}$ e $\bar{\boldsymbol{d}}_{T,k}$, tal que, as expressões (7.3) e (7.20), implementação escalar, sejam satisfeitas, respectivamente. Para implementação em bloco deve-se satisfazer a expressão (7.23) ou (7.24). A globalização, via estratégia de pesquisa-em-linha, foi conduzida após as modificações citadas acima, utilizando o Algoritmo BLS para correção de Newton inicial e os algoritmos BLS-TS e BLS-TC para a correção do tensor. As condições para entrada e saída da malha simplificada (L-12)-(L-27) são as mesmas adotadas no Algoritmo GT.

7.4. Produto Matriz Jacobiana-Vector na Análise do BH

Da teoria discutida acima, podemos observar a importância da implementação númerica do produto matriz jacobiana por vetor na eficiência do solucionador interno. Se matriz jacobiana for densa, a complexidade deste produto, via multiplicação direta, é igual a $O(N^2)$. Na análise do BH, podemos escrever este produto da seguinte forma

$$\bar{\boldsymbol{Y}}(\bar{\boldsymbol{X}}^k) = \bar{\boldsymbol{J}}(\bar{\boldsymbol{X}}^k)\bar{\boldsymbol{Z}}, \tag{7.55}$$

onde $\bar{\boldsymbol{X}}^k$ é o vetor de variável de estado na iteração k e $\bar{\boldsymbol{J}}(\bar{\boldsymbol{X}}^k)$ é a matriz jacobiana (associada com vetor de resíduo não-linear) no ponto definido por este vetor. O vetor $\bar{\boldsymbol{Z}}$ é um vetor arbitrário. Utilizando os resultados do Capítulo 5, podemos re-escrever o produto acima como

$$\bar{\boldsymbol{Y}}(\bar{\boldsymbol{X}}^k) = \bar{\boldsymbol{A}}\bar{\boldsymbol{Z}} + \bar{\boldsymbol{B}}_f\bar{\boldsymbol{Z}}_G(\bar{\boldsymbol{X}}^k), \tag{7.56}$$

onde

$$\bar{\boldsymbol{Z}}_G(\bar{\boldsymbol{X}}^k) = \bar{\boldsymbol{G}}_f(\bar{\boldsymbol{X}}^k)\bar{\boldsymbol{Z}}. \tag{7.57}$$

Conforme já discutido, a expressão (7.56) envolve apenas matrizes multi-níveis. Lembremos que estas matrizes possuem uma estrutura hierárquica, onde cada nível assume uma forma do tipo bloco diagonal com bordas. Em adição, os blocos associados com os pontos de fundo (associado as SuRs de fundo) da hierarquia assumem uma representação bloco diagonal. Sendo assim, estas matrizes são altamente esparsas e a sua computação numérica torna se irrelevante quando comparado com (7.57). Em geral, o produto (7.57) também envolve uma matriz multi-níveis. Porém, neste caso, cada nível da hierarquia assume uma forma do tipo bloco diagonal e os blocos associados as SuRs de fundo, assumem uma representação do tipo bloco diagonal. Lembremos que, a largura de banda, definida em cada un destes blocos, dependerá proporcionalmente do limite de manipulação de potência desejado. Abaixo, discutiremos formas alternativas para o cálculo do produto matriz por vetor (7.57), envolvendo a matriz jacobiana associada ào vetor de função não-

linear.

Utilizando a notação e os resultados do Capítulo 5 e introduzindo a soma multi-dimensional $\sum_{i=(a_1, a_2, \ldots, a_l)}^{(b_1, b_2, \ldots, b_l)} = \sum_{i_1=a_1}^{b_1} \sum_{i_2=a_2}^{b_2} \cdots \sum_{i_l=a_l}^{b_l}$, o vetor $\bar{\boldsymbol{Z}}_G$, em (7.57), pode ser escrito da seguinte forma:

$$Z_{G,p}(\mathbf{0}) = G_{pq}(\mathbf{0})Z_q(\mathbf{0}) + 2 \sum_{\boldsymbol{l}=(1,1,\ldots,1)}^{(NH_1, NH_2, \ldots, NH_{NT})} (G^{re}_{pq}(-\boldsymbol{l})Z^{re}_q(\boldsymbol{l}) + G^{im}_{pq}(-\boldsymbol{l})Z^{im}_q(\boldsymbol{l})), \tag{7.58.a}$$

$$Z^{re}_{G,T,p}(\boldsymbol{k}) = G^{re}_{pq}(\boldsymbol{k})Z_q(\mathbf{0}) + \sum_{\boldsymbol{l}=(1,1,\ldots,1)}^{(NH_1, NH_2, \ldots, NH_{NT})} (G^{re}_{pq}(\boldsymbol{k}-\boldsymbol{l})Z^{re}_q(\boldsymbol{l}) - G^{im}_{pq}(\boldsymbol{k}-\boldsymbol{l})Z^{im}_q(\boldsymbol{l})), \tag{7.58.b}$$

$$Z^{im}_{G,T,p}(\boldsymbol{k}) = G^{im}_{pq}(\boldsymbol{k})Z_q(\mathbf{0}) + \sum_{\boldsymbol{l}=(1,1,\ldots,1)}^{(NH_1, NH_2, \ldots, NH_{NT})} (G^{im}_{pq}(\boldsymbol{k}-\boldsymbol{l})Z^{re}_q(\boldsymbol{l}) + G^{re}_{pq}(\boldsymbol{k}-\boldsymbol{l})Z^{im}_q(\boldsymbol{l})), \tag{7.58.c}$$

$$Z^{re}_{G,H,p}(\boldsymbol{k}) = \sum_{\boldsymbol{l}=(1,1,\ldots,1)}^{(NH_1, NH_2, \ldots, NH_{NT})} (G^{re}_{pq}(\boldsymbol{k}+\boldsymbol{l})Z^{re}_q(\boldsymbol{l}) + G^{im}_{pq}(\boldsymbol{k}+\boldsymbol{l})Z^{im}_q(\boldsymbol{l})), \text{ e} \tag{7.58.d}$$

$$Z^{im}_{G,H,p}(\boldsymbol{k}) = \sum_{\boldsymbol{l}=(1,1,\ldots,1)}^{(NH_1, NH_2, \ldots, NH_{NT})} (G^{im}_{pq}(\boldsymbol{k}+\boldsymbol{l})Z^{re}_q(\boldsymbol{l}) - G^{re}_{pq}(\boldsymbol{k}+\boldsymbol{l})Z^{im}_q(\boldsymbol{l})). \tag{7.58.e}$$

onde os subescritos H e T referem-se as contribuições do tipo *Toeplitz* e *Hankel*, respectivamente. Adicionando estas contribuições, temos que:

$$\boldsymbol{Z}^{re}_{G,p}(\boldsymbol{k}) = \boldsymbol{Z}^{re}_{G,T,p}(\boldsymbol{k}) + \boldsymbol{Z}^{re}_{G,H,p}(\boldsymbol{k}), \tag{7.59.a}$$

$$\boldsymbol{Z}^{im}_{G,p}(\boldsymbol{k}) = \boldsymbol{Z}^{im}_{G,T,p}(\boldsymbol{k}) + \boldsymbol{Z}^{im}_{G,H,p}(\boldsymbol{k}). \tag{7.59.b}$$

Se definirmos $g_{pq}(\boldsymbol{r})$ e $z_q(\boldsymbol{r})$, $\boldsymbol{r} = (0, 0, \ldots, 0), \ldots, (NS_1 - 1, NS_2 - 1, \ldots, NS_{NT} - 1)$ como as formas-de-onda associadas aos espectros de frequência definidos por $\bar{G}_{pq}$ e $\bar{\boldsymbol{Z}}_q$, respectivamente. Então, utilizando o teorema da convolução [64], as somatórias no domínio da frequência (7.59.a)-(7.59.b) são equivalentes à operação de convolução no domínio do tempo e podem ser escritas como se segue:

$$Z_{G,p}(\mathbf{0}) = \frac{1}{N} \sum_{\boldsymbol{r}=(0,0,\ldots,0)}^{(NS_1 - 1, NS_2 - 1, \ldots, NS_{NT} - 1)} g_{pq}(\boldsymbol{r})z_q(\boldsymbol{r}), \tag{7.60.a}$$

$$Z^{re}_{G,p}(\boldsymbol{k}) = \frac{1}{N} \sum_{\boldsymbol{r}=(0,0,\ldots,0)}^{(NS_1 - 1, NS_2 - 1, \ldots, NS_{NT} - 1)} g_{pq}(\boldsymbol{r})z_q(\boldsymbol{r})W^{\boldsymbol{k}\cdot\boldsymbol{r}}_{re}, \tag{7.60.b}$$

$$Z^{im}_{G,p}(\boldsymbol{k}) = \frac{1}{N} \sum_{\boldsymbol{r}=(0,0,\ldots,0)}^{(NS_1 - 1, NS_2 - 1, \ldots, NS_{NT} - 1)} g_{pq}(\boldsymbol{r})z_q(\boldsymbol{r})W^{\boldsymbol{k}\cdot\boldsymbol{r}}_{im}. \tag{7.60.c}$$

As somatórias acima podem ser eficientemente calculadas via a TFRM [157] com complexidade

$O(NH_\Sigma \cdot \log(NH_\Sigma))$, onde $NH_\Sigma = NH_1 + NH_2 + \ldots + NH_{NT}$. A complexidade no cálculo direto do produto matriz densa por vetor é igual a $O(NH_\Sigma^2)$. Se a técnica de matriz esparsa, discutida no Capítulo 5, for utilizada na formação da matriz jacobiana, o custo em termos de operações numéricas do produto direto pode ser menor do que no produto denso via o teorema da convolução.

Por fim, o produto matriz jacobiana por vetor, pode ser aproximado utilizando uma aproximação de diferenças-finitas [16], através da seguinte fórmula (derivada direcional),

$$\bar{\boldsymbol{J}}(\bar{\boldsymbol{X}}^k)\bar{\boldsymbol{Z}} \approx \frac{1}{\sigma}(\bar{\boldsymbol{F}}(\bar{\boldsymbol{X}}^k + \sigma\bar{\boldsymbol{Z}}) - \bar{\boldsymbol{F}}(\bar{\boldsymbol{X}}^k)) \tag{7.61}$$

onde

$$\sigma = \sqrt{\varepsilon} \cdot \max\{|(\bar{\boldsymbol{X}}^k)^T\bar{\boldsymbol{Z}}|, \text{típ}(\bar{\boldsymbol{X}}^k)^T|\bar{\boldsymbol{Z}}|\}\frac{\text{sinal}((\bar{\boldsymbol{X}}^k)^T\bar{\boldsymbol{Z}})}{\|\bar{\boldsymbol{Z}}\|_2}, \tag{7.62}$$

$\text{típ}(\bar{\boldsymbol{X}}^k)$ é um vetor com valores típicos de $\bar{\boldsymbol{X}}^k$, $|\bar{\boldsymbol{Z}}| = \left[|Z_1| \cdots |Z_{\dim(\bar{\boldsymbol{Z}})}|\right]^T$, e ε é o parâmetro de precisão da máquina. Neste trabalho, assumimos $\text{típ}(\bar{\boldsymbol{X}}^k) = \mathbf{1}$ para as componentes em frequência das tensões e correntes do circuito, i.e., norma-M1 do vetor $\bar{\boldsymbol{Z}}$ em (7.62). Em termos de tempo de processamento, o produto (7.61) pode ser considerado a versão mais eficiente, pois, não necessita do cálculo da matriz jacobiana, $\bar{\boldsymbol{J}}(\bar{\boldsymbol{X}}^k)$ e sim do vetor de resíduo $\bar{\boldsymbol{R}}$ no ponto $\bar{\boldsymbol{X}}^k + \sigma\bar{\boldsymbol{Z}}$. Entretanto, sob o ponto de vista da precisão numérica, este produto consiste em uma aproximação $O(\sigma\|\bar{\boldsymbol{Z}}\|_2)$. Convém ressaltar que, os produtos matriz por vetor via convolução (7.58.a)-(7.60.c) e diferenças-finitas (7.61) produzem resultados equivalentes ao produto direto utilizando a matriz jacobiana completa (densa), i.e., sem o controle de esparsidade introduzido no Capítulo 5.

7.5. Pré-Condicionadores para Análise do BH

Conforme menciondo acima, nos métodos de Newton inexato e do tensor inexato o uso da técnica de pré-condicionamento é, em geral, vital para garantir a convergência do solucionador interno (método iterativo linear e iterativo do tensor). No Capítulo 5, foi discutido a aplicação de eficientes técnicas de matriz esparsa para a fatorização da matriz jacobiana em regime de único- e multi-tons. Este procedimento, pode ser utilizado para formar pré-condicionadores suficientemente robustos. Por exemplo, em regime de único-tom, a matriz bloco diagonal (ou bloco Jacobi) com informação apenas da componente de CC no espectro de derivadas, pode servir como pré-condicionador, com considerável capacidade de manipulação de potência [26]. Para uma extensão do limite de manipulacão de potência, pode-se ampliar a banda diagonal da matriz jacobiana. Um procedimento adaptativo para esta ampliação pode ser encontrado em [160]. Utilizando técnicas de aproximação da matriz jacobiana inversa, tal como, iterações de mínimo

resíduo auto-pré-condicionado pode-se refinar o pré-condicionador para solução dos sistemas jacobianos [193].

A decomposição e formulação multi-níveis para circuitos em grande-escala, introduzida nos capítulos anteriores, possibilitam a formação de eficientes pré-condicionadores para análise do BH envolvendo problemas de grande-escala.

7.6. Testes Preliminares

Para uma avaliação preliminar do desempenho do método do tensor inexato, em relação ao método de Newton inexato (método padrão), foram realizados uma série de experimentos númericos utilizando os problemas testes listados na Tabela 7.1. Os problemas 1-6 foram extraídos da coleção de *Moré*/*Garbow*/*Hillstrom* (MGH) [176]. Já os problemas 7-12, 13-18 e 19-24, correspondem a uma modificação dos problemas 1-6, onde a matriz jacobiana possui, na raiz, uma deficiência de posto-um, posto-dois e posto-três, respectivamente. Este tipo de modificação é descrita em [35] e [191] e consiste em quadrar a última equação para deficiência de posto-um (7-12), quadrar as duas últimas equações para deficiência de posto-dois (13-18) e quadrar as três últimas para deficiência de posto-três (19-24). Os demais problemas correspondem a sistemas de equação diferencial parcial (EDP) em duas dimensões, sendo todos do tipo elíptico de valor de contorno [32],[177],[168]. Estes problemas foram resolvidos utilizando os seguintes pré-condicionadores: *Jacobi* (probs. 1-4,7-10,13-16,19-22) [183], tri-diagonal (probs. 5,6,11,12,17,18,23,24,33-36) e os operadores laplaciano (probs. 25-32,37-40) e biharmônico (probs. 37-40) com decomposição *Choleski* [157].

Na Tabela 7.1, a primeira e a segunda coluna referem se ao número e ao nome do problema, respectivamente. Para os problemas de EDPs, a terceira coluna refere se a grade de discretização utilizada. A dimensão do problema (número de variáveis) é fornecida na quarta coluna. A quinta coluna refere se ao ponto inicial, ver [29]. O parâmetro $k_{\dim}$, que corresponde a dimensão do subespaço de Krylov utilizado em cada problema, é listado na sexta coluna. Nas demais colunas, são listados: o número de iterações, NI, o número de cálculo da função, NCF, o número de iterações do solucionador linear, NIL, e a norma-M2 da função não-linear. A pesquisa-em-linha foi conduzida utilizando interpolação quadrática para atualização do fator de amortecimento. Sendo que, no método do tensor inexato foi utilizada a estratégia padrão. Para o solucionador linear iterativo, foi adotado $NIL_{\text{máx}} = 500$ e uma sequência de termo forçante definida pela Escolha 0 com $\eta_0 = 10^{-3}$. Para o solucionador não-linear, foram utilizados os seguintes parâmetros: $\varepsilon_F = 10^{-10}$,

Tabela 7.1
RESULTADOS DOS PROBLEMAS TESTES

	Função	Grade N x N	n	x_0	k_{dim}	Método padrão: Newton inexato				Método novo: tensor inexato pesquisa-em-linha: estratégia padrão			
						NI	*NCF*	*NIL*	$\|F(x_*)\|_2$	*NI*	*NCF*	*NIL*	$\|F(x_*)\|_2$
1	"Broyden Banded"	—	1000	100	5	18	19	18	7,9–11	11	12	20	8,6–11
2		—	1000	100	10	18	19	18	7,9–11	11	12	20	8,6–11
3	"Broyden Triangular"	—	1000	100	5	12	13	14	5,0–11	8	13	20	1,9–13
4		—	1000	100	10	12	13	12	4,5–11	8	13	13	1,9–13
5	"Discrete Boundary"	—	1000	50	10	6	7	18	1,4–11	7	11	41	4,6–11
6		—	1000	50	20	6	7	6	1,4–11	7	11	13	1,6–12
7	"Broyden Banded with Rank-One Deficiency"	—	1000	1	10	21	22	21	2,7–11	8	9	15	3,3–11
8		—	1000	1	15	21	22	21	2,7–11	8	9	15	3,3–11
9	"Broyden Triangular with Rank-One Deficiency"	—	1000	1	10	19	20	63	5,0–11	7	8	32	7,2–13
10		—	1000	1	20	19	20	19	5,0–11	7	8	13	7,9–13
11	"Discrete Boundary with Rank-One Deficiency"	—	1000	50	15	14	15	14	3,5–11	9	13	21	1,3–12
12		—	1000	50	25	14	15	14	3,5–11	9	13	17	3,6–15
13	"Broyden Banded with Rank-Two Deficiency"	—	1000	1	10	21	22	21	3,7–11	9	10	17	1,0–11
14		—	1000	1	20	21	22	21	3,7–11	9	10	17	1,0–11
15	"Broyden Triangular with Rank-Two Deficiency"	—	1000	1	20	19	20	19	5,1–11	8	9	15	3,7–13
16		—	1000	1	25	19	20	19	5,1–11	8	9	15	3,7–13
17	"Discrete Boundary with Rank-Two Deficiency"	—	1000	50	15	12	13	932	—	9	10	865	4,6–11
18		—	1000	50	25	14	15	14	4,2–11	10	14	19	3,6–14
19	"Broyden Banded with Rank-Three Deficiency"	—	1000	1	10	21	22	21	4,4–11	12	13	23	1,1–12
20		—	1000	1	20	21	22	21	4,4–11	12	13	23	1,1–12
21	"Broyden Triangular with Rank-Three Deficiency"	—	1000	1	20	19	20	19	5,1–11	7	8	13	3,2–11
22		—	1000	1	25	19	20	19	5,1–11	7	8	13	3,2–11
23	"Discrete Boundary with Rank-Three Deficiency"	—	1000	50	15	3	4	568	—	5	6	873	—
24		—	1000	50	25	14	15	14	5,0–11	9	13	17	5,0–14
25	PDE κ=275	64	4096	1	15	14	21*	14	8,4–11	11	24	20	1,4–11
26		64	4096	1	20	14	21*	14	8,4–11	11	24	21	5,4–11
27	PDE κ=280	64	4096	1	20	150	1140	150	—	10	26	18	6,9–12
28		64	4096	1	25	150	1140	150	—	10	26	18	6,9–12
29	"Bratu" λ=6,8067408	32	1024	0	5	14	15	15	8,2–12	6	7	11	4,5–11
30		32	1024	0	10	14	15	14	8,2–12	6	7	11	4,5–11
31	"Modified Bratu" λ=κ=10	64	4096	0	5	5	6	18	4,2–13	5	6	23	1,3–13
32		64	4096	0	10	5	6	5	4,6–13	4	5	7	4,6–11
33	"Porous Medium" d=50	64	4096	1	20	9	10	77	1,4–12	8	9	111	1,2–13
34		64	4096	1	35	9	10	39	1,7–12	7	8	51	5,7–11
35	"Porous Medium" d=–50	64	4096	1	20	7	9	32	4,4–11	8	10	50	6,0–12
36		64	4096	1	35	7	9	19	5,3–11	8	10	30	3,3–12
37	"Lid Driven Cavity" Re=250	63	7938	0	10	6	7	30	9,9–11	6	7	49	9,1–11
38		63	7938	0	20	6	7	14	9,7–11	6	7	23	6,8–11
39	"Lid Driven Cavity" Re=500	63	7938	0	15	8	9	41	1,8–11	8	9	72	1,3–11
40		63	7938	0	30	8	9	19	1,8–11	8	9	33	1,3–11

$\varepsilon_x = 10^{-12}$, $NI_{\text{máx}} = NIJ_{\text{máx}} = 150$ e $NIM_{\text{máx}} = 0$. Os parâmetros utilizados na pesquisa-em-linha, foram: $\alpha = 10^{-4}$, $NIP_{\text{máx}} = 10$ e $\lambda_{\min} = 10^{-7}$. Finalmente, no método do tensor inexato foi utilizado o processo de solução modificada com implementação escalar, para resolução aproximada do modelo do tensor. A diferença entre as soluções obtidas com os métodos em consideração foi insignificante, i.e., $\|\Delta_*\|_2 < 10^{-9}$. Antecipadamente, nas condições acima, podemos observar que o método de Newton inexato produz quatro falhas (probs. 17,23,27,28), enquanto o método do tensor inexato produz apenas uma (prob. 23). As falhas nos problemas 17 e 23 são eliminadas com o aumento de k_{dim}. Entretanto, as falhas do método padrão nos problemas 27 e 28 estão associadas à falta de robustez do método de Newton inexato.

Utilizando a definição de razão de desempenho, RD, introduzida no capítulo anterior, podemos comparar graficamente o desempenho do método do tensor inexato em relação ao método de

Newton inexato. Neste sentido, nas Figs. 7.1(a)-(c), são representadas graficamente as RDs com o modelo do tensor sendo (aproximadamente) resolvido via os processos de solução simplificada, modificada, completa de bloco 2 e completa de bloco 3. Mais precisamente, os histogramas apresentados nas Figs. 7.1(a), (b) e (c) correspondem as razões de desempenho em termos do número de iterações, $RD(NI)$, do número de cálculo da função, $RD(NCF)$ e do número de iterações do solucionador linear, $RD(NIL)$. Vale ressaltar que, os resultados da Tabela 7.1, correspondem aos histogramas referente à solução modificada. Definindo um problema fácil como aquele, no qual a convergência ocorre em menos de cinco iterações, i.e., $NI = 5$. Podemos observar, pelos histogramas, que os métodos do tensor inexato possuem um superior desempenho em termos de NI e NCF, principalmente para os problemas difíceis. Também podemos identificar, facilmente, a superioridade nos problemas com deficiência de posto (um, dois e três) da matriz jacobiana na raiz, conforme previsto teoreticamente [174]. Porém, se considerarmos o desempenho em termos de NIL, observarmos que os métodos do tensor inexato possuem um desempenho inferior ao método padrão. Ou seja, os métodos do tensor inexato, quando comparado com o método padrão, produzem um menor NI e NCF abrindo mão do NIL, conforme discutido em [114]. Finalmente, podemos observar que o processo de solução modificada é o mais eficiente entre os processos de solução comparados. As falhas e a perda de desempenho do método do tensor inexato empregando o processo de solução completa (bloco 2 ou 3) estão associadas ao baixo limite imposto na dimensão do subespaço de Krylov utilizado.

Considerando o método do tensor inexato, com resolução do modelo do tensor, via o processo de solução modificada (melhor desempenho nos resultados acima), apresentamos na Fig. 7.1(d) o desempenho da implementação em bloco versus a implementação escalar. Como podemos observar, excluindo as falhas, a implementação em bloco é apenas marginalmente superior a implementação escalar se considerarmos o NIL e apenas marginalmente inferior considerando NI e NCF. Não obstante, considerando o hardware utilizado, a implementação em bloco pode ser muito superior a escalar em termos de tempo de processamento, conforme destacado em [114].

Ainda considerando o método do tensor inexato com solução modificada e implementação escalar, foram realizados testes para determinar o desempenho da estratégia de pesquisa-em-linha curvilinear versus a estratégia de pesquisa-em-linha padrão. Os resultados destes testes com atualização do fator de amortecimento via redução divindindo-pela-metade e s interpolação quadrática são ilustrados no histograma da Fig. 7.2(a) e Fig. 7.2(b), respectivamente. Assim, como na versão exata, ver resultados do Capítulo 6, em ambos os casos, a estratégia curvilinear é apenas marginalmente mais eficiente em termos de NCF. Em adição, a estratégia com redução divindindo-

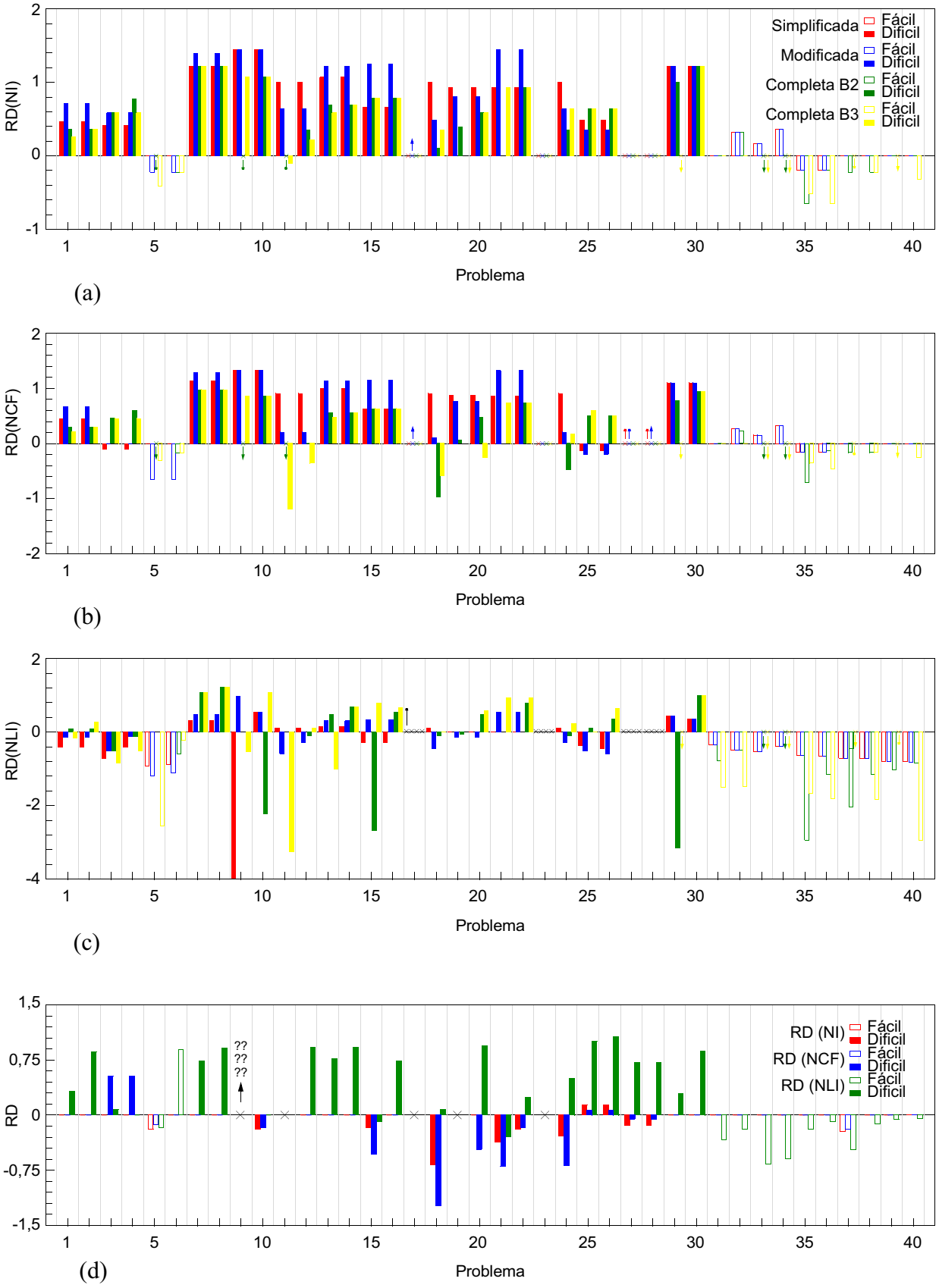


Fig. 7.1 Gráficos da razão de desempenho (RD) do método do tensor inexato (novo) vs. Newton inexato (padrão). (a) Processo de solução simplificada. (b) Processo de solução modificada. (c) Processo de solução completa B2. (d) Processo de solução completa B3.

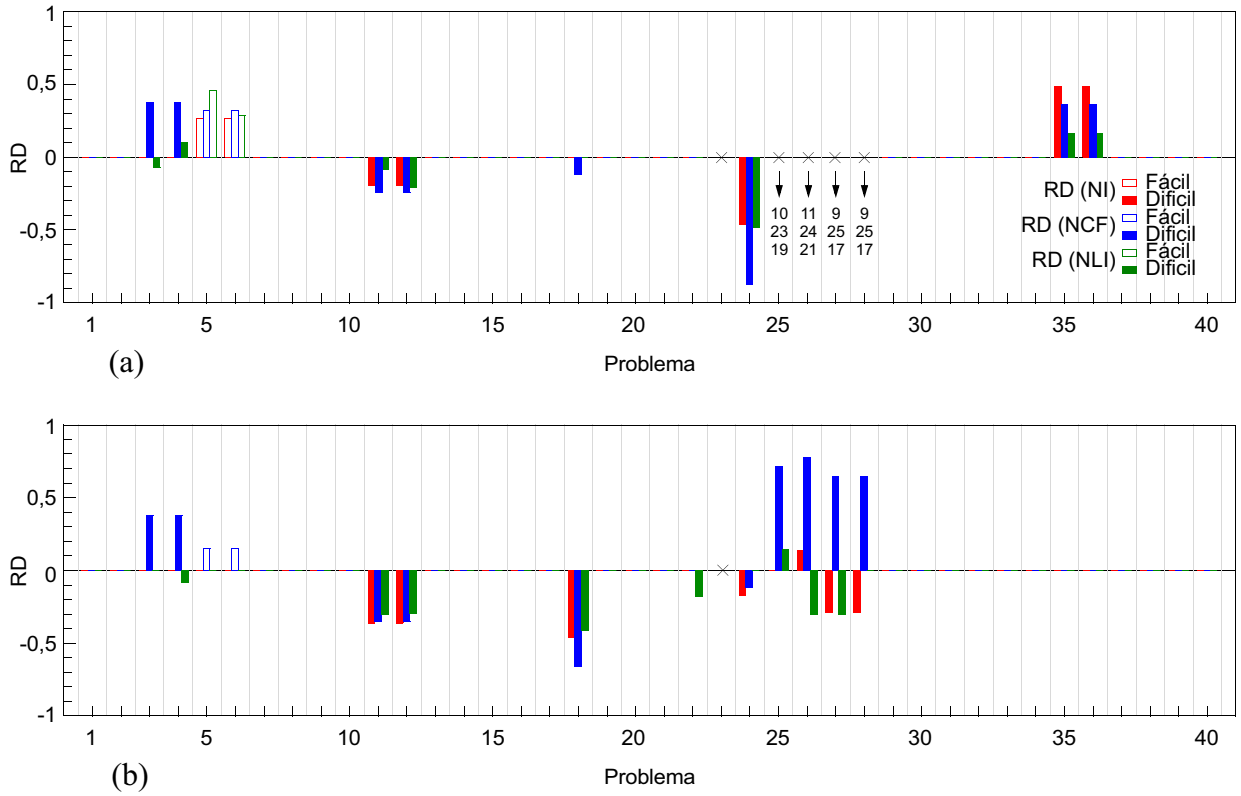


Fig. 7.2 Gráficos da razão de desempenho (RD) do método do tensor inexato com solução modificada e implementação escalar. (a) Estratégia de pesquisa-em-linha curvilinear com interpolação quadrática (novo) vs. estratégia padrão (padrão). (b) Estratégia de pesquisa-em-linha curvilinear λ-divindo-pela-metade (novo) vs. estratégia padrão (padrão).

pela-metade produz quatro falhas adicionais, como podemos observar na Fig. 7.2(a).

Nos resultados acima, o termo forçante foi mantido constante (escolha trivial ou Escolha 0), o que, em geral, resulta no problema de sobressolução. Sendo assim, utilizando o Problema 39 (“Lid Cavity”) da Tabela 7.1, foram realizados testes para avaliar o efeito de outras escolhas para definição da sequência de termo forçante. Os resultados destes testes, sob forma de históricos de convergência, são ilustrados na Fig. 7.3 e foram obtidos com as mesmas condições adotadas para a geração da Tabela 7.1. Ressaltando que, os gráficos referentes ao método do tensor inexato foram obtidos com a implementação em bloco em substituição a implementação escalar. Na Fig. 7.3(a) e Fig. 7.3(b) podemos observar o histórico de convergência do método de Newton inexato utilizando as escolhas 0 e 5, respectivamente. No gráfico da Fig. 7.3(a), podemos observar claramente a característica *dente de serra* causada pelo efeito de sobressolução. Para a Escolha 5, o termo forçante inicial é dado por $\eta_0 = 0,1$. Conforme previsto em [177], o uso da Escolha 5, praticamente elimina o problema de sobressolução que ocorre com a Escolha 0. Para avaliarmos o

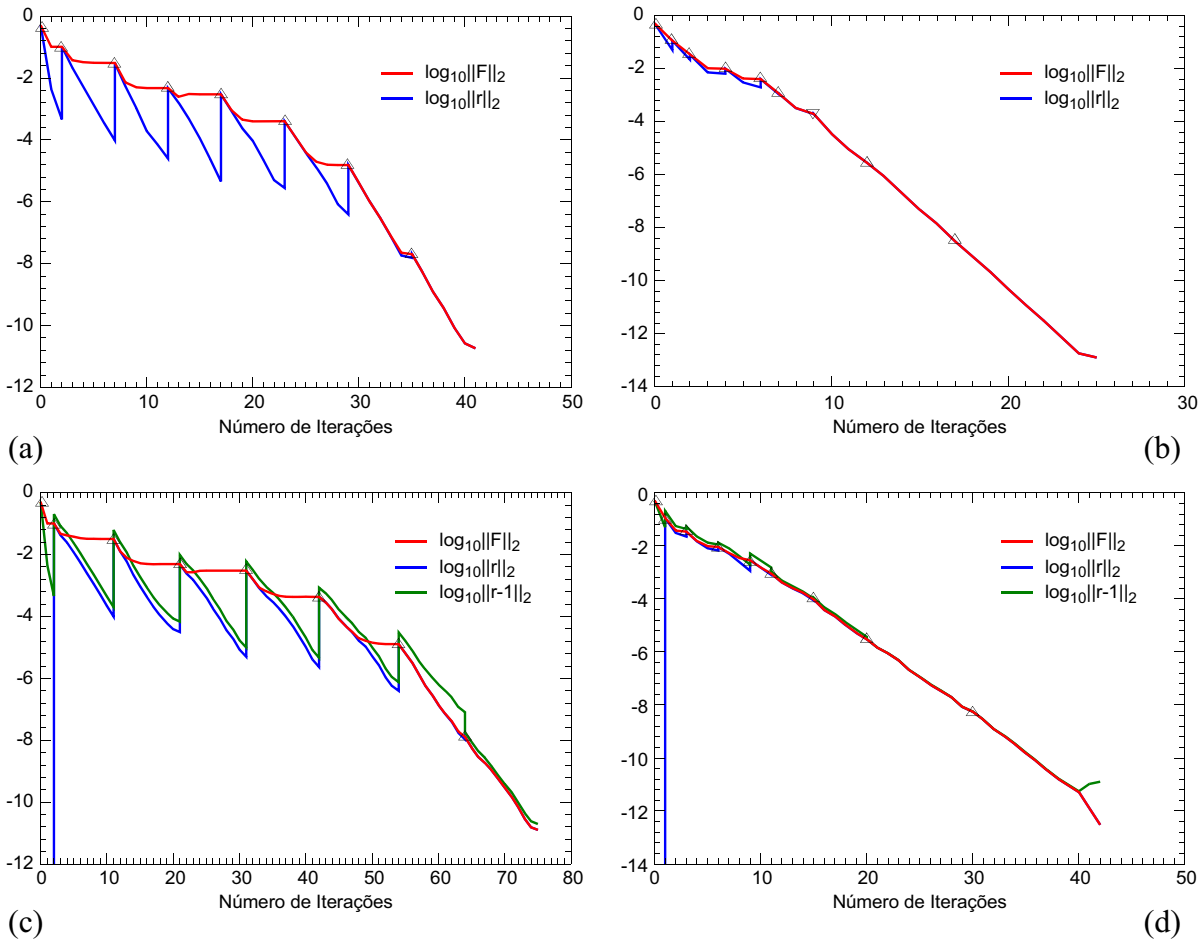


Fig. 7.3 Histórico de convergência para solução do problema 39 utilizando o método de Newton inexato com (a) Escolha 0 e (b) Escolha 5. Histórico de convergência (problema 39) do método do tensor inexato com solução modificada e implementação bloco utilizando (c) Escolha 0 e (d) Escolha 7.

efeito de sobressolução com o método do tensor inexato com solução modificada e implementação em bloco, foi realizado um teste com a Escolha 0, ver Fig. 7.3(c) e um outro com a Escolha 7, ver Fig. 7.3(d). É interessante observar que a Escolha 7, proposta neste trabalho, também produz uma eliminação do problema de sobressolução. Este tipo de resultado não está apresentado em nenhuma das referências listadas, e sendo assim, representa uma contribuição deste trabalho. Nos gráficos da Fig. 7.3, os símbolos " △ " e " ▽ " indicam cada passo de correção inexata do solucionador não-linear (iterações externas). Sendo que, o último símbolo indica também a ação da proteção (7.10).

Finalmente, os resultados descritos acima estão em plena concordância com os resultados apresentados em [31]-[114], e desta forma, validam a implementação numérica proposta acima.

7.7. Conclusão

Para análise do BH em grande-escala, foram discutidos e implementados as versões inexatas

dos métodos de Newton e do tensor. Conforme foi destacado, na discussão teórica anterior, o método do tensor inexato possui uma maior complexidade, na solução iterativa do modelo local, que define a correção para o cálculo da próxima iteração. Para resolução deste modelo, foram considerados os processos de solução simplificada, modificada e completa (terminologias introduzidas neste trabalho). Sendo o processo de solução completa, um problema de minimização (com dimensão igual ao subespaço de Krylov) do modelo do tensor.

No método de Newton inexato, como solucionador interno, foi discutido e implementado o método GMRES com re-inícialização e pré-condicionamento a direita, RGMRES(m). Este método, realiza a solução iterativa e aproximada do modelo linear local que define a correção de Newton inexata. O pré-condicionamento tem o caráter de assegurar robustez e velocidade de convergência. Conforme apresentado acima, os processos de solução simplificada e modificada exigem a resolução iterativa e simultânea de dois sistemas de equação não-linear. Para tal, foram discutidos e implementados a versão escalar RGMRES(m) e a versão bloco-2 RGMRES-B2(m). Para o processo de solução completa do modelo do tensor, como solucionador-interno, foi discutido o método iterativo TGMRES-Bt com reinício e pré-condicionamento. Mais precisamente, foi discutido e implementado uma versão que opera em subespaço de Krylov de bloco-2, RTGMRES-B2(m) e uma outra versão de bloco-3, RTGMRES-B3(m). Esta última versão trabalha com informação completa do modelo do tensor na formação base inicial do subespaço de Krylov.

Para validar a nossa implementação númerica, foram realizados uma série de testes preliminares comparando o desempenho do método de Newton inexato com o método do tensor inexato.

Na avaliação preliminar de desempenho apresentada acima, o método do tensor inexato com processo de solução modificada apresentou uma maior eficiência quando comparado com o método de Newton inexato. Porém, vale ressaltar que, os processos de solução completa demandam uma maior dimensão para o subespaço de Krylov utilizado. Nos testes realizados, esta dimensão foi estabelecida no limite de solução do método de Newton inexato, e sob este aspecto, a comparação pode não ter sido justa. Com relação a pesquisa-em-linha, os resultados que comparam a estratégia curvilinear e a estratégia padrão, aplicado ao processo de solução modificada, demonstram um comportamento semelhante aos resultados envolvendo o método do tensor (versão exata) discutido no capítulo anterior. Diferentes escolhas para a sequência de termos forçantes foram analisadas e testadas. Incluindo resultados que demonstram a eficácia da Escolha 7 (e da Escolha 8), proposta neste trabalho, para eliminação do problema de sobressolução. Uma comparação do desempenho da implementação em bloco versus implementação escalar no processo de solução modificada, também foi apresentada. Em resumo, os testes realizados

demonstram a superioridade do método do tensor inexato, particularmente, na solução de problemas singulares e mal-condicionados entorno da raiz. Em adição, pode se observar pelos resultados que o método do tensor inexato negocia mais iterações internas por menos iterações externas.

Para maior eficiência do solucionador-interno, na análise do BH, foram discutidas eficientes técnicas para a condução do produto entre a matriz jacobiana do BH (associada à função não-linear) e um vetor, sem formação explícita desta matriz. Também foram discutidas diferentes técnicas de pré-condicionamento da matriz jacobiana (associada ào resíduo), incluindo a metodologia proposta de decomposição multi-níveis do circuito.

Os algoritmos referentes aos métodos inexatos implementados, neste trabalho, foram descritos em detalhe, seguindo a mesma estrutura de implementação numérica discutida no capítulo anterior.

8. Validação Numérica

8.1. Introdução

NESTE CAPÍTULO, serão apresentados os exemplos de circuitos forçados utilizados na validação da teoria proposta para formulação e resolução do problema do BH em regime de único-tom, dois-tons, três-tons e multi-tons (distorção harmônica, distorção por intermodulação, conversão em frequência e conversão em frequência com distorção por intermodulação). Por conveniência, estes exemplos foram sub-divididos em três grupos. O primeiro grupo consiste de circuitos básicos utilizando um número pequeno de dispositivos ativos do tipo diodo ou transistor BJT (modelo de Ebers-Moll utilizando 2 diodos). Estes circuitos são: fonte de alimentação com retificação de meia-onda (FARMO) [7],[13],[4]; fonte de alimentação com retificação de onda-completa (FAROC); amplificador classe-C (ACC) [8],[13],[4] e; multiplicador de frequência (MF) [7],[8]. No segundo e terceiro grupo, foram considerados circuitos fundamentados em transistores de microonda e de onda milimétrica do tipo MESFET, pHEMT e HBT, cujos CEEs foram discutidos no Capítulo 3. Para o segundo grupo, foram considerados os seguintes circuitos: amplificador de potência (AP); amplificador de potência corporativo (APC) [196]; conversor de frequência resistivo (CFR) e; conversor de frequência resistivo balanceado (CFRB) [113]. No terceiro e último grupo, foram considerados os circuitos com dezenas de transistores, onde a aplicação da decomposição multi-níveis, introduzida no Capítulo 2, produz uma significante redução na complexidade da análise do BH em termos de memória e tempo de processamento. Os circuitos em consideração são: ressoador ativo (RA) [194]; multiplicador analógico de quatro-quadrantes (MAQQ) [195]; multiplicador analógico de quatro-quadrantes de baixo-deslocamento (MAQQ-BD) [195] e; multiplicador analógico balanceado (MAB) [197].

Na Seção 8.2, discutir-se-á, de forma resumida, a implementação em CAD da teoria proposta neste trabalho. Em seguida, na Seção 8.3, serão descritos os circuitos forçados utilizados como exemplos e os resultados numéricos que validam o processo de formulação e de solução introduzido nos capítulos anteriores. Esta seção está subdividida em 7 subseções, descrevendo cada um dos circuitos citados acima. Nestas subseções, são apresentados esquemáticos, modelo de grande-sinal, tabela de parâmetros e as funções não-lineares dos circuitos. Para os circuitos com estrutura hierárquica, os esquemas de decomposição multi-níveis serão expostos detalhadamente. Em adição, para cada circuito de teste, são ilustrados os resultados obtidos com as SRNs descritas pela FEE e pela FNM e o circuito com e sem hierarquia. Estes resultados, obtidos da análise de CC,

análise MP, análise de CA, análise do BH, e análise de CF-BH [153], demostram a precisão da nossa implementação númerica. A Seção 8.4 apresenta os resultados que comparam o desempenho do método do tensor versus método de Newton em problemas de pequena e média-escala e do método do tensor inexato versus método de Newton inexato, em problemas de grande-escala. A Seção 8.5 apresenta o desempenho da análise do BH utilizando a técnica de decomposição multi-níveis. Finalmente, as conclusões são reservadas para a Seção 8.6.

8.2. Implementação em CAD

Toda a teoria descrita neste trabalho foi implementada numericamente e integrada no ambiente de programa para simulação de circuitos integrados, entitulado CDSys - *Circuit Design System*. A linguagem de programação objeto-orientada C++ [107], foi adotada nesta implementação por oferecer avançados recursos de programação, tais como: classes, polimorfismo, funções virtuais, sobre-carregamento ("*overloading*") de funções e de operadores, classes abstratas, derivação de classe, template, etc. O ambiente do programa está sendo desenvolvido sob forma de *Standard Template Library* (STL) o que garante a sua modularidade e portabilidade. Vale ressaltar que, o sistema CDSYS está apoiado na biblioteca numérica, entitulada NuPack - *Numerical Package*. Esta biblioteca é composta dos seguintes módulos: ILESOLV - *Iterative Linear Equation Solver;* ITESOLV - *Iterative Tensor Equation Solver;* NESOLV - *Nonlinear Equation Solver* (Tensor, Newton, multi-level, homotopy); DLESOLV - *Direct Linear Solver*; EIGENSOLV - *Eigen Solver;* MATRIX (operações com matrizes de ponto e de bloco armazenada sob forma densa e esparsa); VECTOR; DFT - *Discrete Fourier Transform*; DATFIT - *Data Fitting;* etc.

8.3. Descrição dos Circuitos e Resultados

Nesta seção, descreveremos os circuitos e os resultados numéricos que validam a teoria desenvolvida neste trabalho. Inicialmente, apresentamos os circuitos básicos: FARMO, FAROC, ACC e MF. Neste primeiro grupo, os circuitos utilizam diodos e BJTs descritos por um simples CEE e a função não-linear padrão para corrente de condução do diodo. Em seguida, no segundo grupo, são apresentados os circuitos de microonda e de onda-milimétrica: AP, CFR e CFRB. Exceto pelo CFRB, que utiliza dois dispositivos do tipo FET, os demais circuitos utilizam apenas um dispositivo. Vale lembrar, dos capítulos anteriores, que cada dispositivo corresponde a uma SRN. Finalmente, empregando um maior número de dispositivos, apresentaremos os resultados

para o terceiro grupo de circuitos: RA, MAQQ, MAQQ-BD, AAPC e MAB que utilizam a decomposição multi-níveis. Os circuitos do segundo e terceiro grupo, estão fundamentados nos dispositivos de alta-velocidade do tipo MESFET, pHEMT e HBT, cujos CEEs foram descritos no Capítulo 2.

8.3.1. Circuitos Básicos

Os circuitos básicos descritos na Fig. 8.1 tem sido tradicionalmente utilizados na validação numérica de métodos para análise de regime permanente em circuitos não-lineares forçados [7],[8],[13],[4],[57],[198]. São circuitos simples que possuem apenas um nível de hierarquia (nível 0) referente a SuR topo, $\mathbb{S}_0$. O primeiro circuito básico, utilizando 1 diodo, consiste de uma fonte de alimentação com retificação de meia-onda (FARMO) operando em 60 Hz, ver esquemático da Fig. 8.1(a) [7],[13],[4]. Apesar de não ilustrado, também foi considerado o circuito de uma fonte de alimentação com retificação de onda completa (FAROC) utilizando uma ponte de 4 diodos. As formas-de-onda das tensões retificadas geradas pelas fontes de alimentação FARMO e FAROC são ilustradas na Fig. 8.1(b). Estes resultados foram obtidos da análise do BH de único-tom com 32 harmônicos (i.e., 33 linhas espectrais), o que resulta em 65 variáveis na FEE; em 130 variáveis na FNM para a FARMO; em 260 variáveis na FEE e em 520 variáveis na FNM para a FAROC.

Em seguida, consideramos o ACC operando em 100 MHz, cujo esquemático é fornecido na Fig. 8.1(c) [8],[13],[8]. Este circuito, utiliza um dispositivo do tipo BJT como elemento ativo, representado pelo CEE de Ebers-Moll que utiliza 2 diodos e 2 FTCCs. A forma-de-onda da tensão de saída pode ser observada na Fig. 8.1(d). Este resultado foi obtido da análise do BH de único-tom com 16 hramônicos (i.e., 17 linhas espectrais), o que resulta em 33 variáveis na FEE e na FNM

O último circuito básico a ser considerado, representado no esquemático da Fig. 8.1(e) , é o MF utilizando BJT [7],[8]. Este circuito opera com frequência de entrada igual a 21 MHz e fator de multiplicação em frequência igual a 2. O CEE do BJT é o mesmo utilizado no ACC. Os resultados da análise do BH de único-tom, com a forma-de-onda e o espectro de frequência da tensão de saída são apresentados nas Figs. 8.1(f) e (g), respectivamente. Estes resultados foram obtidos da análise do BH de único-tom com 16 harmônicos.

Na Tabela 8.1 é listada a função não-linear padrão que descreve a corrente de condução de um diodo de junção PN ou de barreira Schottky. Para evitar o problema de estouro do valor numérico, em alta corrente, o modelo do diodo é aproximado por uma função quadrática com continuidade de derivada de primeira e de segunda-ordem no ponto de transição. O modelo paramétrico do diodo

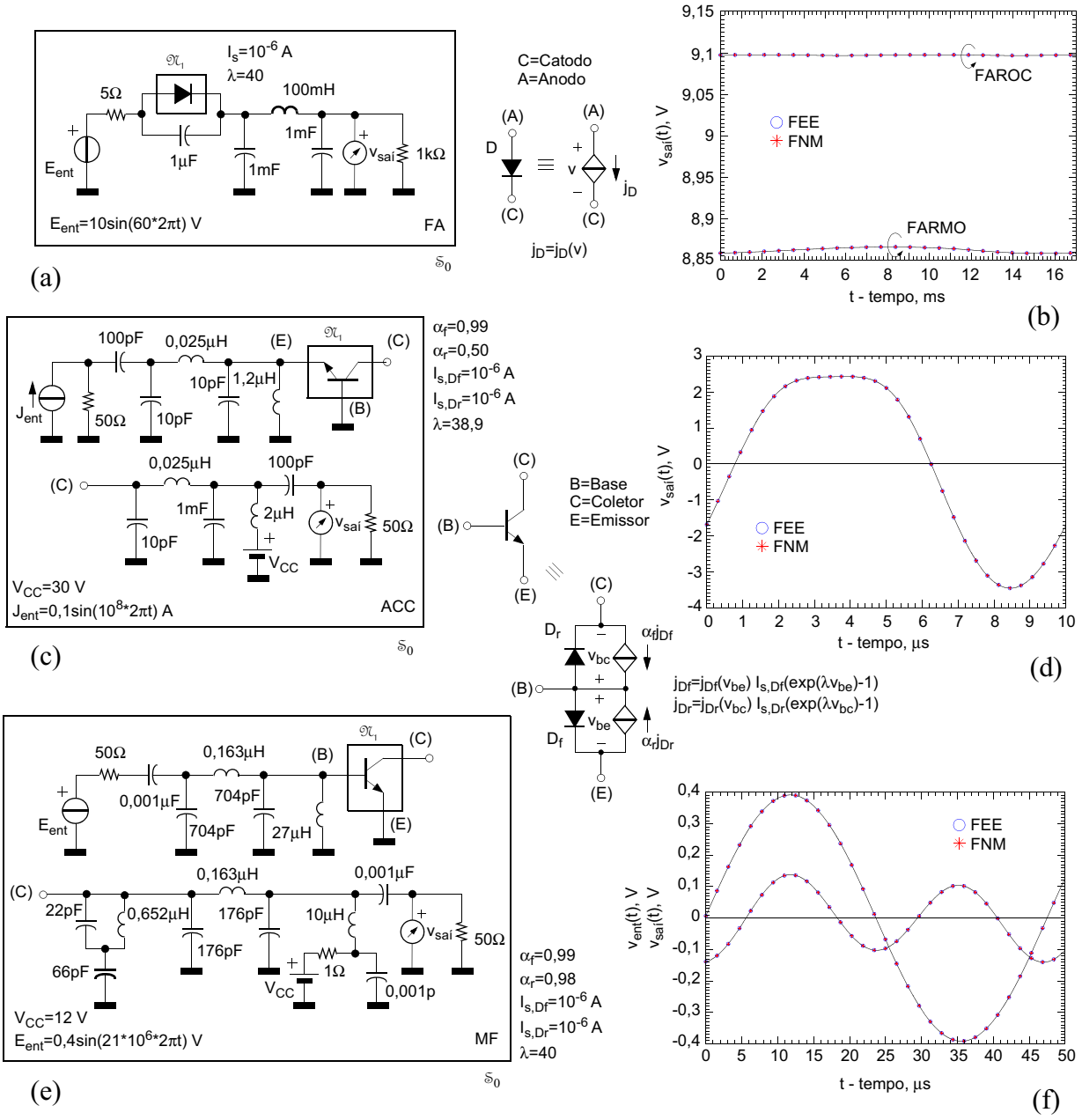


Fig. 8.1 (a) Esquemático e (b) resultado da fonte de alimentação com retificação de meia-onda (FARMO) e de onda completa (FAROC). (c) Esquemático e (d) resultado do amplificador classe-C (ACC) utilizando BJT. (e) Esquemático e (f) resultado multiplicador de frequência (MF) utilizando BJT.

também é comumente utilizado para solucionar este problema [199],[39].

8.3.2. Amplificador de Potência

O circuito do AP, ilustrado no esquemático da Fig. 8.2(a), está fundamentado em um dispositivo transistor do tipo GaAs MESFET de 8 dedos, fabricado pela Phillips Microwave, RU, com comprimento de porta de 0,7μm (processo D07) e largura de porta total de 900 μm. As redes de

Tabela 8.1
FUNÇÃO NÃO-LINEAR DO MODELO DO DIODO

Corrente de Condução do Diodo
$j_{D,1}(v(t)) = I_s(\exp(v(t)/(nV_T)) - 1)$
$j_{D,2}(v(t)) = I_s(\exp(V_{máx}/(nV_T)) - 1) + A_2 v(t)^2 + A_1 v(t) + A_0$
$A_0 =$, $A_1 =$ e $A_2 =$
$j_D(v(t)) = \text{if}(v(t) > V_{máx}, j_{D,1}(v(t)), j_{D,2}(v(t)))$

adaptação de entrada e de saída foram projetadas para um casamento de potência simultâneo, na frequência de 10 GHz (banda-X), resultando em um ganho de potência de aproximadamente 6 dB em regime de pequeno-sinal. O ponto de polarização do MESFET corresponde -1 V para a tensão de porta e 8 V para a tensão de dreno. As redes de terminação TE1 (porta 1 - entrada) e TE2 (porta 2 - saída) são formadas por uma sonda de tensão, uma sonda de corrente e uma excitação com representação de Thevènin ou de Norton. Esta rede de terminação será utilizada em todos os exemplos a seguir.

Para representar os MESFETs, foi utilizado o CEE (top-B) do FET descrito na Fig. 3.3(a) do Capítulo 3. Porém, neste exemplo, estamos considerando o CEE sem a presença do diode D_{gd} e com R_{gs} e q_{gd} linear. Lembrando que, para este modelo temos 4 ou 6 variáveis de estado não-lineares, se usarmos FEE ou FNM, respectivamente. Isto representa aproximadamente uma redução de 33% na dimensão do problema do BH, a favor da FEE. Na Tabela 8.5 estão listados os parâmetros elétricos do CEE do MESFET. As funções não-lineares, que descrevem o comportamento do MESFET em regime de grande-sinal, estão fundamentadas no modelo desenvolvido pela Phillips Microwave [196]. Neste modelo, a corrente dos diodos de porta-fonte e de porta-dreno são representadas pelo funcional descrito na Tabela 8.1. Para a corrente de dreno, a transição entre as regiões de sub-limiar e de alta-corrente, é realizada da seguinte forma: quando a tensão porta-fonte, v_{gsc}, assume um valor menor que $V_c = V_{th}(1 - \Delta)$ (tensão crítica), a função f_1 assume a forma: $f_1(v) = K_1 \exp(K_2 v)$. Os coeficientes K_1 e K_2 são determinados impondo continuidade na função e na sua derivada de primeira-ordem. Este modelo possui apenas continuidade na derivada de primeira ordem e, por este motivo, possui limitações na reprodução dos efeitos de DIM. O efeito de dispersão em baixa-frequência (DBF) da condutância do canal foi considerado. Os parâmetros do CEE do MESFET de oito-portas estão listados na Tabela 8.2. Os parâmetros das funções não-lineares, associadas as correntes de condução e deslocamento deste modelo, se encontram listadas nas Tabelas 8.3 e 8.4.

Na Fig. 8.2(a) podemos observar a trajetória I/V da corrente de dreno versus tensão dreno-fonte

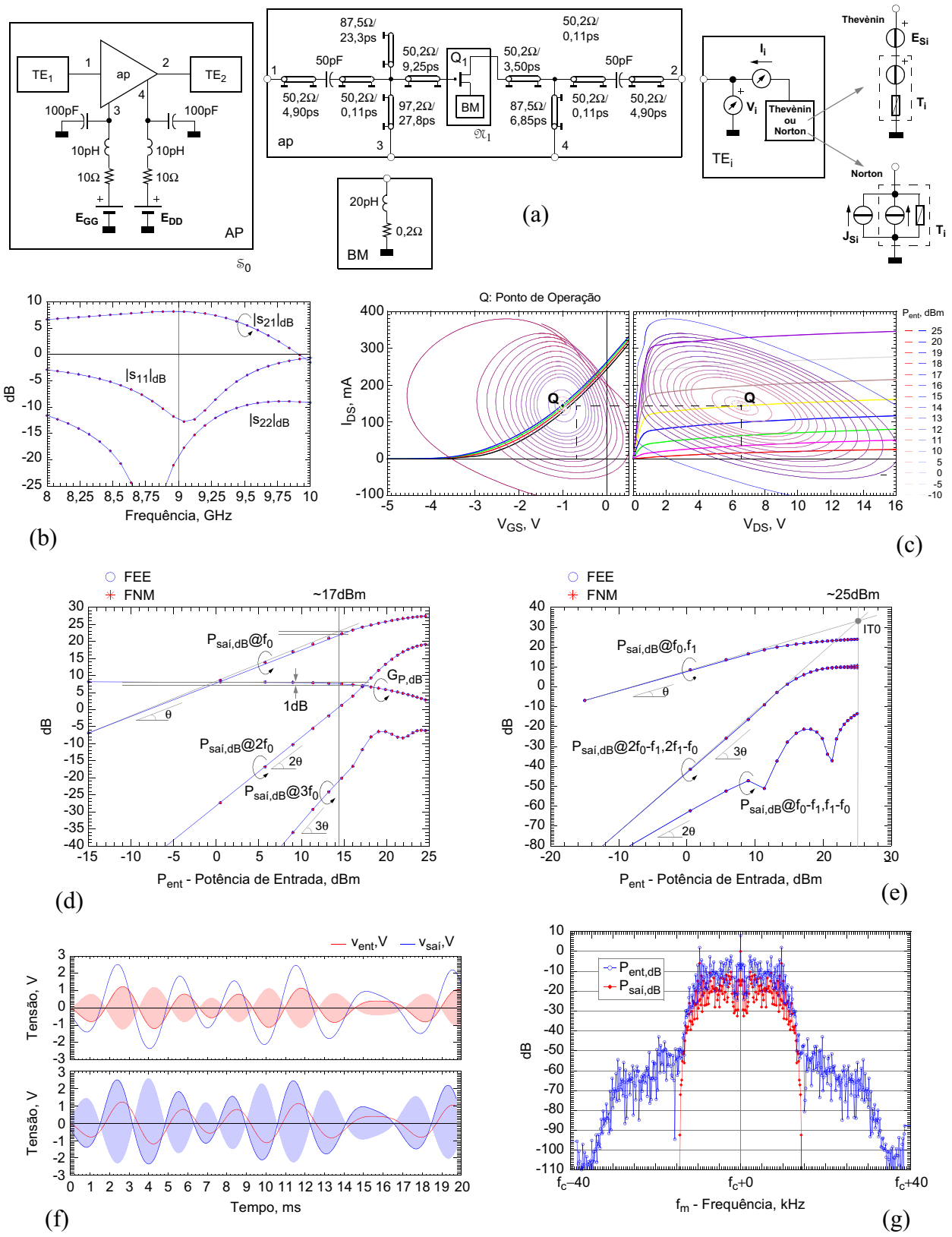


Fig. 8.2 (a) Esquemático do amplificador de potência (AP) de microonda utilizando GaAs MESFET. (c) Parâmetros de espalhamento. (b) Trajetórias I/V. (c) Potência de saída versus potência de entrada para regime de único-tom. (e) Potência de saída versus potência de entrada em regime de dois-tons. (f) Formas-de-onda das tensões de entrada e de saída e (g) recrescimento espectral em regime multi-tons com excitaçãode RF modulada digitalmente.

Tabela 8.2
PARÂMETROS ELÉTRICOS DO CEE (TOP-A) DO GAAS MESFET

Extrínsecos									Intrínsecos		
L_{gpar} (pH)	L_{dpar} (pH)	R_{dpar} (Ω)	L_{spar} pH	R_{spar} (Ω)	R_{gw} (Ω·mm)	R_{dw} (Ω·mm)	R_{sw} (Ω·mm)	C_{dsw} pF/ (mm)	τ (ps)	R_{gsw} (Ω·mm)	C_{gdw} (pF/mm)
20,0	20,0	0,01	12,8	0,01	0,603 (0,64)	0,306 (0,34)	0,306 (0,34)	$1,8\overline{22}$ (1,64)	3,7	1,287 (1,43)	$0,13\overline{77}$ (0,124)

Tabela 8.3
FUNÇÕES NÃO-LINEARES DO MODELO DA PHILLIPS DO GAAS MESFET

Cargas de Porta-Fonte ($x = gs$) e de Porta-Dreno ($x = gd$)

$$q_x(v_x(t)) = C_{x0}V_{bi}(1-(1-(v_x(t)+0{,}5V_{sat})/V_{bi})^m)/m + C_{x1}v_x(t) \qquad V_{Cmax} = F_cV_{bi} - 0{,}5V_{sat}$$
$$e_x(v_x(t)) = v_x(t) - q_x(v_x(t))/C_{x0}$$

Corrente de Dreno-Fonte

$$f_1(v_{gsd}(t)) = (1-v_{gsd}(t)/V_{th})^p H(v_{gsd}(t)-V_{th}) \qquad f_2(v_{ds}(t)) = 1-\exp(-k(v_{ds}(t)/V_{sat})^2)/(1+v_{ds}(t)/V_{sat})$$
$$j_{ds}(v_{gsd}(t), v_{ds}(t)) = I_{dss}f_1(v_{gsd}(t))f_2(v_{ds}(t))$$

Condutância de Saída CC

$$g_{ds}^{(DC)}(v_{gsd}(t), v_{ds}(t)) = \alpha g_{ds0}/(1+a\cdot\exp(V_{th}-v_{gsd}(t))/V_a)\cdot\ln(1+bv_{ds}(t))/(bv_{ds}(t))$$
$$j_{ds}^{(DC)}(v_{gsd}(t), v_{ds}(t)) = g_{ds}^{(DC)}(v_{gsd}(t), v_{ds}(t))v_{ds}(t)$$

Condutância de Saída CA

$$g_{ds}^{(AC)}(v_{gsd}(t), v_{dsb}(t)) = (1-\alpha)g_{ds0}/(1+a\cdot\exp(V_{th}-v_{gsd}(t))/V_a)/(1+bv_{dsb}(t))$$
$$j_{ds}^{(AC)}(v_{gsd}(t), v_{dsb}(t)) = g_{ds}^{(AC)}(v_{gsd}(t), v_{dsb}(t))v_{dsb}(t)$$

Tabela 8.4
PARÂMETROS DAS FUNÇÕES NÃO-LINEARES DO MODELO DA PHILLIPS DO GAAS MESFET

Corrente de Dreno-Fonte (Dispersão de Baixa-frequência)										
I_{dssw} (mA/mm)	V_{sat} (V)	V_{th} (V)	k	p	Δ	g_{d0w} (mS/mm)	a	b	V_a (V)	α
332,4 (???)	0,4	-2,98	0,7	1,75	0,0025	22,725 (???)	0,827	0,3	0,296	0,333
Cargas de Porta-Fonte e de Porta-Dreno						Diodos de Porta-Fonte e de Porta-Dreno				
C_{gs0w} (pF/mm)	C_{gd0w} (pF/mm)	C_{gsBw} (pF/mm)	C_{gdBw} (pF/mm)	V_{bi} (V)	m	I_{sgsw} (pA/mm)	I_{sgdw} (pA/mm)	n_{gs}	n_{gd}	V_T (mV)
1,667 (???)	1,667 (???)	0,146 (???)	0,146 (???)	0,8	0,331	10,0 (???)	10,0 (???)	1,0	1,0	31,2

em regime não-linear. Na Fig. 8.2(b) podemos observar o recrescimento espectral da potência de saída pra um sinal de entrada digital.

8.3.3. Amplificador de Potência Corporativo

A micrografia do circuito integrado de microonda monolítico (CIMM) (tradução nossa do termo

em inglês *monolithic microwave integrated circuit* (MMIC)) do APC fabricado pela Phillips Microwave, RU, em tecnologia de GaAs, pode ser visualizada na Fig. 8.3(a). Este circuito emprega 8 MESFETs com as mesmas dimensões e características elétricas do transistor utilizado no AP descrito anteriormente, ver Fig. 8.2(a). Assim como no AP, o circuito do APC foi projetado para operar em 10 GHz. As redes de adaptação de entrada e de saída, utilizadas para divisão e combinação de potência, respectivamente, formam uma estrutura binária de oito-ramos conforme podemos observar no diagrama esquemático da Fig. 8.3(b). Estas redes de adaptação e de polarização foram implementadas em linhas de transmissão (LTs), físicamente realizadas utilizando estruturas de microfita em substrato de GaAs semi-isolante, cujas as dimensões físicas e os parâmetros elétricos são apresentados na Fig. 8.3(b). Os buracos metalizados (BMs) ("via-holes"), modelados por uma rede RL (resistor-indutor) série, fornecem o aterramento dos capacitores *microwave integrated monolithic* (MIM) do tipo "overlay" utilizados nas redes de adaptação, e dos terminais de fonte dos FETs. A rede externa de polarização de porta (REPP) e de dreno (REPD), compostas de fios de solda e "pads", foram modeladas por uma rede LC passa-baixa. O esquemático completo do AP utilizado é apresentado na Fig. 8.3(b).

As discontinuidades do tipo junção-T e sobre-passo existentes no layout do APC produzem um efeito desprezível na faixa de frequência de operação, sendo assim, foram desprezadas. Porém, o mesmo não se pode dizer sobre as componentes harmônicas de alta-frequência com APC operando em regime de grande-sinal.

Os parâmetros de espalhamento (ou parâmetros-s) calculados e medidos do APC, polarizado com $E_{GG} = -1$ V e $E_{DD} = 8$ V, na faixa de frequência de 2 a 20 GHz, podem ser visualizados na Fig. 8.3(c). Destes resultados observamos que o APC possui uma banda de passagem de 3-dB de 8,5 a 11,0 GHz, um ganho de potência de 8 dB, e uma potência de saída de +36 dBm no ponto de compressão de 1 dB. Os parâmetros-s calculados foram obtidos via análise MP e análise CA de pequeno-sinal, utilizando a FEE e a FNM para as SRNs. Os resultados de grande-sinal da potência de saída componente fundamental e das componentes de segunda e de terceira harmônica, em função da potência de entrada, são mostrados na Fig. 8.3(d). O ponto de compressão de 1 dB é também indicado nesta figura. Observe que as inclinações das curvas estão de acordo com a teoria descrita em [201].

Se considerarmos a análise eletro-térmica [76], cada MESFET de 8 dedos de porta pode ser considerado uma SuR intermediária composta de 8 SuRs de fundo representando um arranjo de 8 células de MESFET de porta-única e de uma RC para a interligação destas SuRs (células) com os terminais externos do dispositivo.

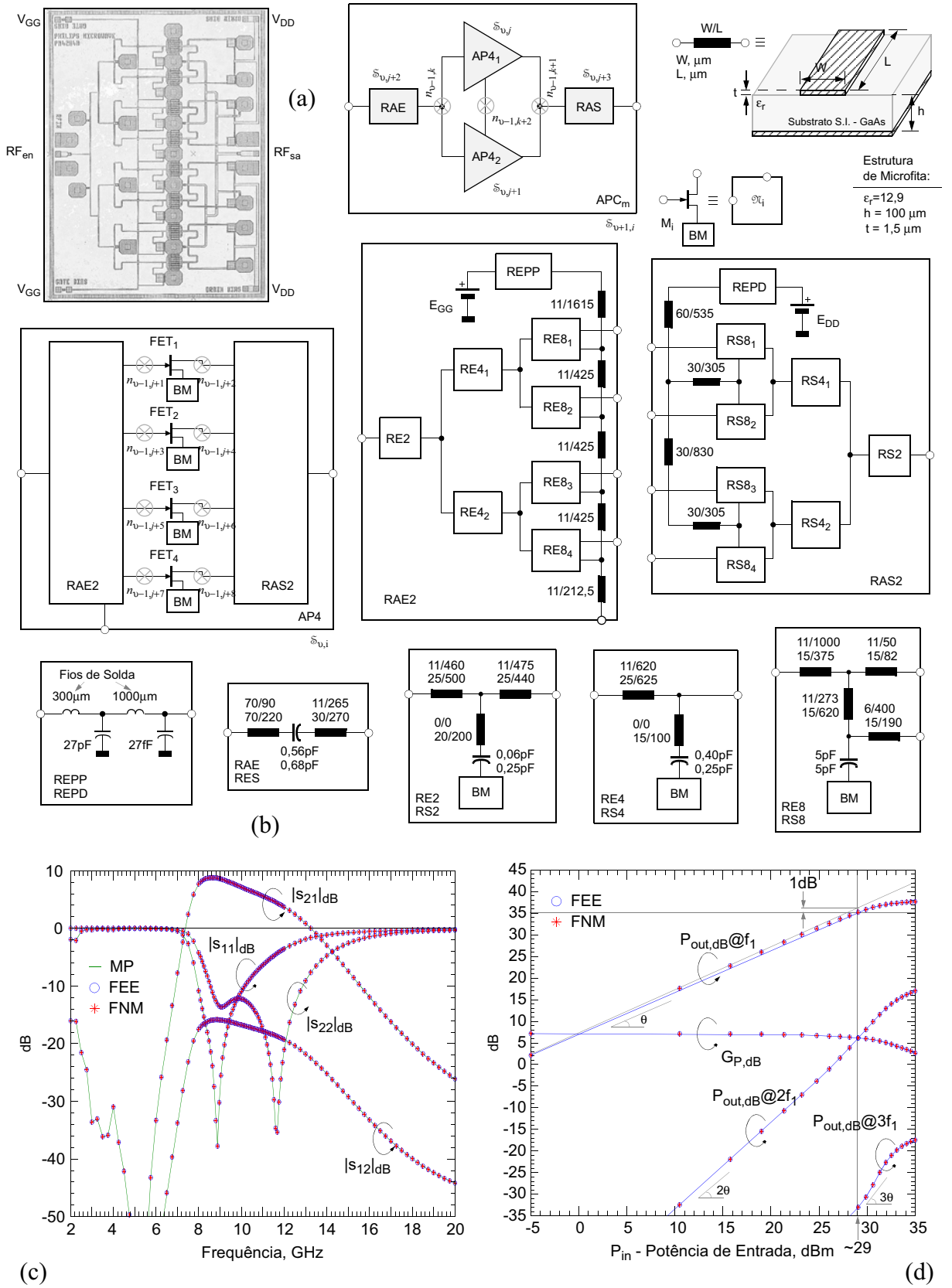


Fig. 8.3 (a) Microfotografia do amplificador de potência corporativo (APC) fabricado pela Phillips. (b) Esquemático do APC sub-dividido em 2 super-redes (SuRs). (c) Magnitude dos parâmetros de espalhamento medidos e calculados. (d) Ponto de compressão de 1 dB medido e calculado.

8.3.4. Conversores de Frequência Resistivos

O esquemático do circuito do CFR, projetado para operação em 94 GHz (banda-W), pode ser visualizado na Fig. 8.4(a). Fig. 8.4 O projeto deste conversor de frequência foi desenvolvido para a tecnologia de guia-de-onda coplanar utilizando as técnicas discutidas em [113]. Em acréscimo, este conversor emprega um transistor do tipo pHEMT (ou MODFET), que atua como resistor variável de alta-velocidade. Os transistores foram polarizados de forma a operar na região linear, i.e., entorno da tensão de dreno-fonte igual a zero.

Os pHEMTs fabricados em tecnologia de InP com um valor nominal de comprimento de porta de 80 nm [113] foram modelados utilizando a metodologia de aproximação quasi-estática [136], conforme descrito em [200]. O CEE (top-A) descrito na Fig. 3.3(a) foi utilizado para representar o comportamento elétrico do pHEMT, com seus parâmetros elétricos fornecidos na Tabela 8.5. As funções não-lineares do modelo de Lin-Ku [200], utilizadas para representar as correntes de condução e de deslocamento no dispositivo intríseco, são listadas na Tabela 8.6 e os seus parâmetros na Tabela 8.7. Ao contrário do AP, no modelo do FET, não foi utilizado a tensão intrínseca de porta-fonte com atraso, $v_{gsc}(t-\tau)$, i.e., assume-se $\tau = 0$.

Em operação heteródina com $f_{\mathrm{OL}}=94\,\mathrm{GHz}$ e $f_{\mathrm{FI}}=500\,\mathrm{MHz}$, podemos observar nas Fig. 8.4(b) e (c) os resultados da análise do BH. Mais precisamente, foi representado graficamente, na Fig. 8.4(b), a perda de conversão, L_c, versus a tensão de polarização de porta, E_{PP}, para diferentes níveis de potência de OL, P_{OL}, e na Fig. 8.4(c), a variação de L_c com P_{OL}, para $E_{\mathrm{PP}}=-0{,}2\,\mathrm{V}$ análise de intermodulação de dois-tons. Estes resultados foram obtidos com uma potência de RF, P_{RF}, igual a -25dBm. A análise foi conduzida utilizando 8 harmônicos para representação da portadora (sinal na frequência de OL) e 1 harmônico para modulação (sinal na frequência de RF), totalizando 26 linhas espectrais (incluindo a componente de CC).

O esquemático do circuito do CFRB, versão com arquitetura de rejeição de frequência imagem do conversor acima, pode ser visualizado na Fig. 8.4(d). Nas Figs. 8.4(e) e (f) são apresentados os resultados referente a análise do BH para ganho de conversão e isolação OL-RF versus potência de OL. Estes resultados foram obtidos utilizando 8 harmônicos para a frequência de OL (portadora/ grande-sinal) e 1 harmônico para a frequência de RF (modulação/pequeno-sinal).

8.3.5. Ressoador Ativo

O diagrama esquemático do RA proposto em [194], utilizando GaAs MESFETs de 1 μm modo

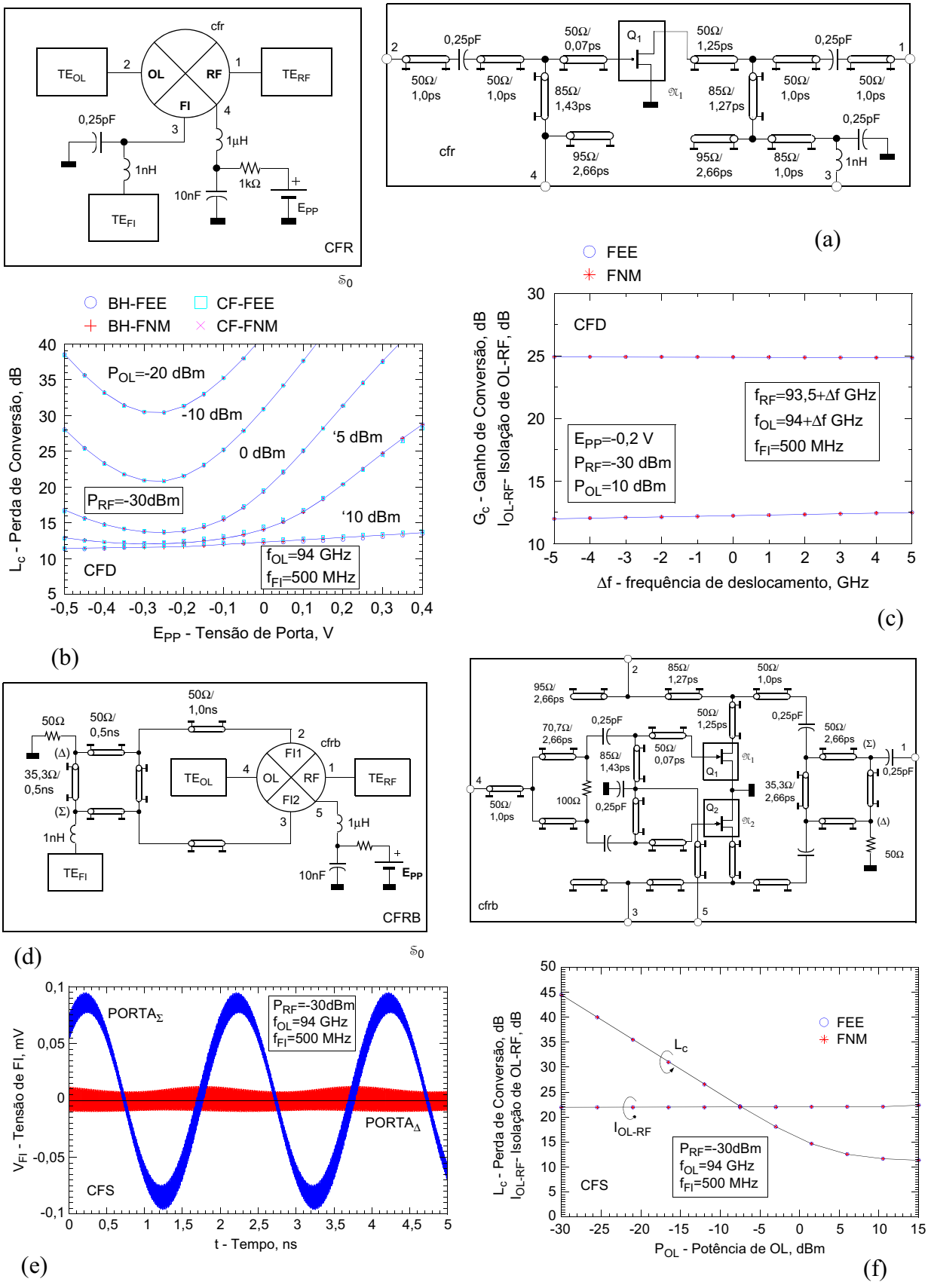


Fig. 8.4 (a) Esquemático e (b)-(c) resultados do conversor de frequência resistivo (CFR) de onda milimétrica utilizando InP pHEMT. (e) Esquemático e (e),(f) resultados do conversor de frequência resistivo balanceado (CFRB) de onda-milimétrica utilizando InP pHEMT.

Tabela 8.5
PARÂMETROS ELÉTRICOS DO CEE (TOP-B) DO INP PHEMT

Parasitas							Intrínsecos		
$L_{g\text{par}}$ (pH)	$L_{d\text{par}}$ (pH)	$L_{s\text{par}}$ (pH)	R_g (Ω)	R_d (Ω)	R_s (Ω)	C_{ds} (fF)	τ (ps)	R_{gs} (Ω)	R_{gd} (Ω)
24	14	1	13	25	20	3	0	14	14

Tabela 8.6
FUNÇÕES NÃO-LINEARES DO MODELO DE LIN-KU DO INP PHEMT

Cargas de Porta-Fonte ($x = gs$) e de Porta-Dreno ($x = gd$)
$q_x(v_x(t)) = C_{x0}\tanh((A_{x1}v_x(t) + A_{x2}) + A_{x3})$
$e_x(v_x(t)) = v_x(t) - q_x(v_x(t))/C_{x0}$
Corrente de Dreno-Fonte
$g_{ds}(v_{gsd}(t)) = G_{ds0}(\tanh(A_{ds1}v_{gsd}(t) + A_{ds2}) + 1)$
$j_{ds}(v_{gsd}(t), v_{ds}(t)) = g_{ds}(v_{gsd}(t))v_{ds}(t)$

Tabela 8.7
PARÂMETROS DAS FUNÇÕES NÃO-LINEARES DO MODELO DE LIN-KU DO INP PHEMT

Corrente de Dreno-Fonte							
G_{ds0} (mS)	A_{ds1} (1/V)	A_{ds2}	I_{sgs} (μA)	I_{sgd} (μA)	n_{gs}	n_{gd}	V_T (mV)
31,2	4,53	0,05	0,63	0,63	4,5	4,5	???
Cargas de Porta-Fonte e de Porta-Dreno					Diodos de Porta-Fonte e de Porta-Dreno		
C_{gs0} (fF)	A_{gs1} (1/V)	A_{gs2}	A_{gs3}	C_{gd0} (fF)	A_{gd1} (1/V)	A_{gd2}	A_{gd3}
10,60	5,45	1,26	2,45	5,24	5,22	0,79	3,68

depleção (com tensão de acionamento de -1.0 V) é descrito na Fig. 8.5(a) e (b). Como podemos observar, o circuito do RA é composto por 2 integradores em configuração inversora e não-inversora ligados em anel. As tensões de polarização de fonte, E_{FF}, e de dreno, E_{DD} são iguais a 6V e -6V, respectivamente. Os transistores M_1, M_2, M_8 e M_{10} compõem o integrador não-inversor, enquanto os demais transistores compõem o integrador não-inversor. O transistor M_{11} é utilizado para sintonia em frequência. Os transistores M_1-M_{10} possuem uma largura de porta igual a 50 μm, enquanto para o transistor M_{11} esta largura é igual a 375 μm.

Para descrever as correntes de condução e de deslocamento do GaAs MESFET foram utilizados as funções não-lineares descritas na Tabela 8.9 e que correspondem ao modelo HSPICE nível 1 [195]. Os parâmetros destas funções não-lineares e os valores dos componentes adotados para CEE

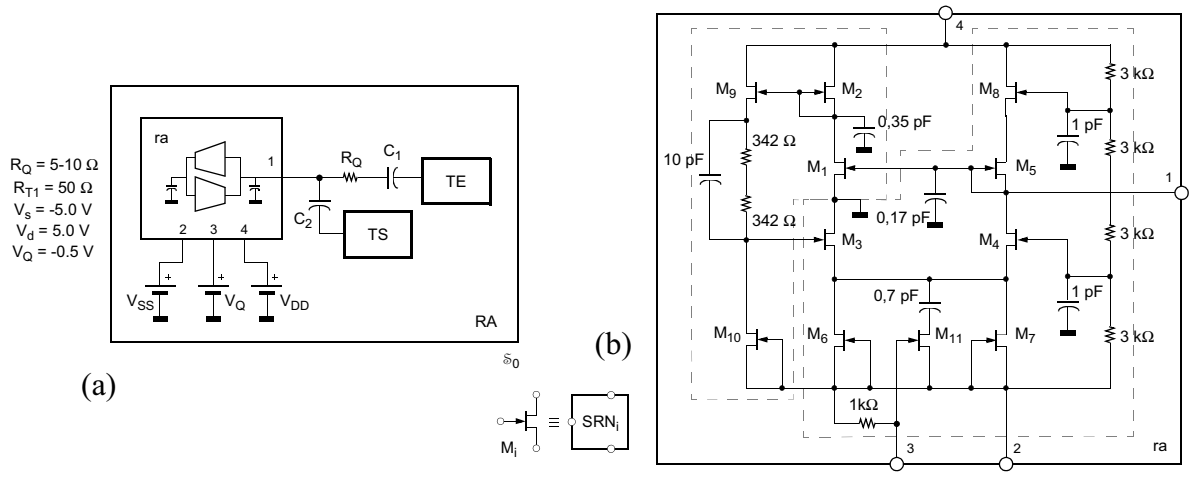


Fig. 8.5 (a) Esquemático do circuito para teste do ressoador ativo (RA). (b) Esquemático do circuito elétrico do RA utilizando GaAs MESFETs.

Tabela 8.8
PARÂMETROS ELÉTRICOS DO CEE (TOP-C) DO GAAS MESFET

Extrínsecos									Intrínsecos		
$L_{g\text{par}}$ (pH)	$L_{d\text{par}}$ (pH)	$R_{d\text{par}}$ (Ω)	$L_{s\text{par}}$ pH	$R_{s\text{par}}$ (Ω)	R_{gw} (Ω·mm)	R_{dw} (Ω·mm)	R_{sw} (Ω·mm)	C_{dsw} (pF/mm)	τ (ps)	R_{gsw} (Ω·mm)	C_{gdw} (pF/mm)
20,0	20,0	0,01	12,8	0,01	0,64	0,34	0,34	1,64	3,7	1,43	0,124

Tabela 8.9
FUNÇÕES NÃO-LINEARES DO MODELO HSPICE DO GAAS MESFET

Cargas de Porta-Fonte ($x = gs$) e de Porta-Dreno ($x = gd$)

$$q_x(v_x(t)) = C_{x0}V_{bi}(1-(1-(v_x(t)+0{,}5V_{sat})/V_{bi})^m)/m + C_{x1}v_x(t) \qquad V_{Cmax} = F_cV_{bi} - 0{,}5V_{sat}$$
$$e_x(v_x(t)) = v_x(t) - q_x(v_x(t))/C_{x0}$$

Corrente de Dreno-Fonte

$$f_1(v_{gsd}(t)) = (1-v_{gsd}(t)/V_{th})^p H(v_{gsd}(t)-V_{th}) \qquad f_2(v_{ds}(t)) = 1-\exp(-k(v_{ds}(t)/V_{sat})^2)/(1+v_{ds}(t)/V_{sat})$$
$$j_{ds}(v_{gsd}(t), v_{ds}(t)) = I_{dss}f_1(v_{gsd}(t))f_2(v_{ds}(t))$$

Tabela 8.10
PARÂMETROS DAS FUNÇÕES NÃO-LINEARES DO MODELO HSPICE DO GAAS MESFET

Corrente de Dreno-Fonte (Dispersão de Baixa-frequência)										
I_{dssw} (mA/mm)	V_{sat} (V)	V_{th} (V)	k	p	Δ	g_{d0w} (mS/mm)	a	b	V_a (V)	α
332,4	0,4	-2,98	0,7	1,75	0,0025	22,725	0,827	0,3	0,296	0,333
Cargas de Porta-Fonte e de Porta-Dreno						Diodos de Porta-Fonte e de Porta-Dreno				
C_{gs0w} (pF/mm)	C_{gd0w} (pF/mm)	C_{gsBw} (pF/mm)	C_{gdBw} (pF/mm)	V_{bi} (V)	m	I_{sgsw} (pA/mm)	I_{sgdw} (pA/mm)	n_{gs}	n_{gd}	V_T (mV)
1,667	1,667	0,146	0,146	0,8	0,331	10,0	10,0	1,0	1,0	31,2

(ver Capítulo 3) do MESFET normalizados para uma largura de 1 μm são descritos na Tabela 8.8 e Tabela 8.10, respectivamente. Esta topologia envolve, na FNM, 4 variáveis não-lineares, e, na

FEE, 3 variáveis não-lineareas.

Para uma primeira demonstração da técnica de decomposição multi-níveis, introduzida no Capítulo 2, vamos considerar o esquema ilustrado na Fig. 8.6(a). Neste esquema, o circuito do RA foi decomposto em uma estrutura hirarquica de dois níveis, cujas as SuRs de fundo $\mathfrak{S}_{1,1}$, $\mathfrak{S}_{1,2}$ e $\mathfrak{S}_{1,3}$, são representadas pelas células CEL_1, CEL_2 e CEL_3, respectivamente, ver Fig. 8.6(b). O número de váriaveis de conexão é igual a 3 associadas aos nós de conexão: $n_{0,1}$, $n_{0,2}$ e $n_{0,3}$ da RC $\mathcal{C}_0$. Neste caso, o número de váriaveis de conexão (via FEE) é igual ao número de váriaveis não-lineares por transistor. Ou seja, para este circuito com 11 transistores, a decomposição multi-níveis introduz uma sobre-carga equivalente a introdução de um transistor. A estrutura do padrão "não-zero" da matriz jacobiana associada ao esquema de decomposição de dois-níveis da Fig. 8.6(a) é representada na Fig. 8.6(c).

Para ilustrar a operação em frequência do RA em regime de pequenos sinais foi realizada uma análise em frequência dos parâmetros de espalhamento de 45 MHz a 8 GHz. O resultado desta análise pode ser visualizado na Fig. 8.5(d) e indica um ressonância entorno da frequência de 1,7 GHz para uma tensão de -5.5 V.

8.3.6. Multiplicadores Analógicos de Quatro-Quadrantes

Os esquemáticos dos circuitos dos MAQQs utilizando GaAs MESFETs (operando no modo-depleção) são ilustrados nas Figs. 8.7(a)-(c). O princípio de operação destes multiplicadores, se fundamenta na característica aproximadamente quadrática da curva de transferência da corrente de dreno versus tensão de porta-fonte. O MAQQ ilustrado na Fig. 8.7(b) utiliza 18 MESFETs e, como podemos observar na Fig. 8.7(d), apresenta um alto deslocamento na sua curva de transferência devido a modulação da corrente de dreno com a tensão de dreno-fonte. Para eliminação deste efeito de deslocamento, pode ser utilizado o MAQQ-BD ilustrado na Fig. 8.7(c) [195]. Na Fig. 8.7(d) são apresentadas as curvas de transcondutância estática dos MAQQs sem e com compensação, onde fica claro a ação de compensação do circuito com baixo-deslocamento. O MAQQ-BD emprega 35 MESFETs correspondendo aproximadamente ao dobro de dispositivos da versão sem compensação.

O CEE e as funções não-lineares que descrevem os MESFETs utilizados nos MAQQs são os mesmos utilizados no circuito do RA, descrito anteriormente. Entretanto, para o MAQQ e o MAQQ-BD foram utilizados dispositivos com largura de porta de 32 μm e 16 μm, respectivamente.

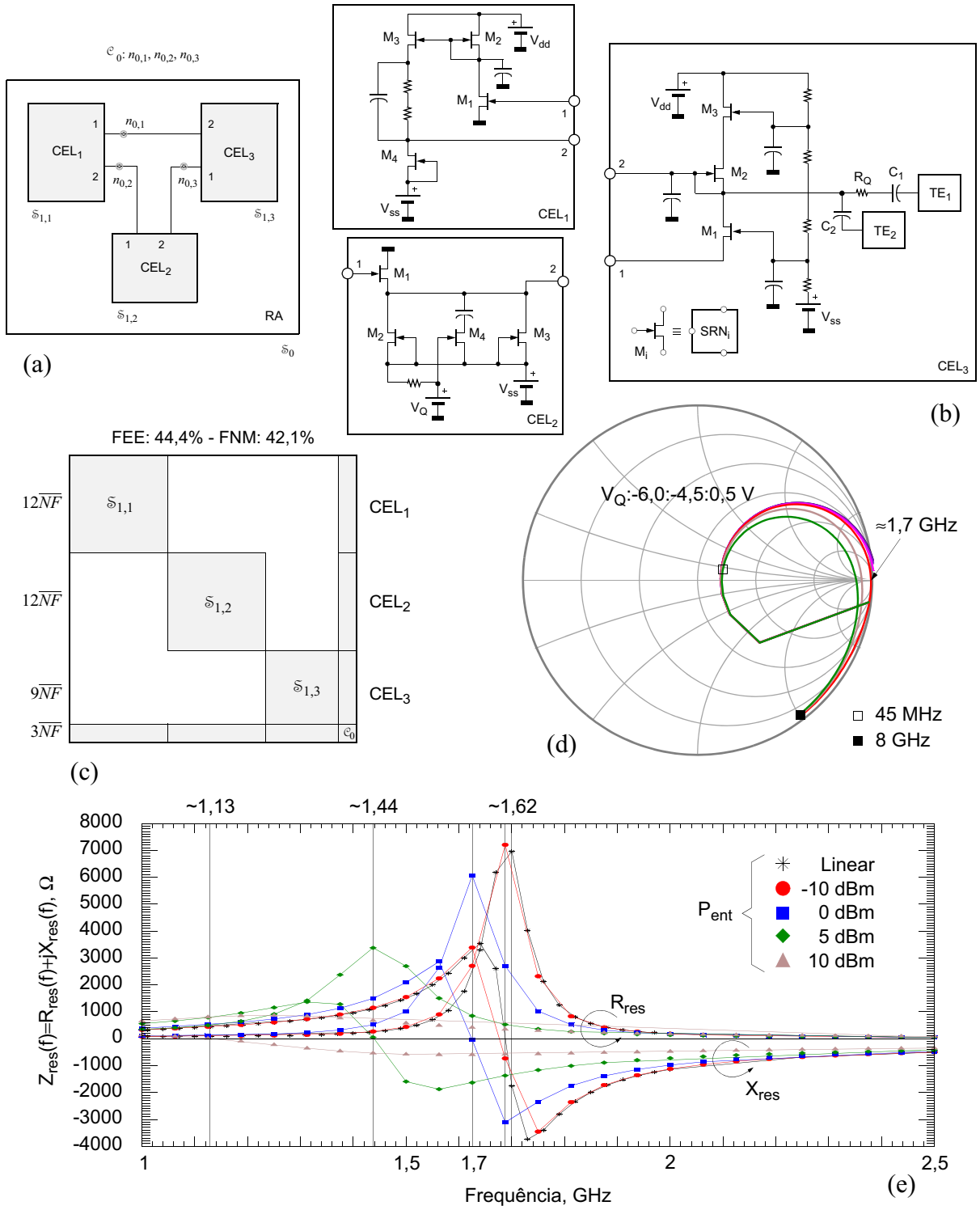


Fig. 8.6 (a) Esquemático do RA decomposto hierarquicamente em super-redes (SuRs). (b) Esquemático dos circuitos das SuRs utilizadas em (a). (c) Estrutura de dois níveis da matriz jacobiana para o esquema de decomposição (a) (em escala). (d) Impedância de entrada do RA versus frequência para diversos níveis de potência de entrada. (e) Coeficiente reflexão de entrada de pequeno-sinal do RA versus tensão de sintonia, $V_Q = -5{,}5$ V.

Os esquemáticos descrevendo a estrutura da decomposição multi-níveis dos MAQQs são ilustrados nas Fig. 8.8(a) e (b). Como podemos observar o MAQQ foi decomposto em uma

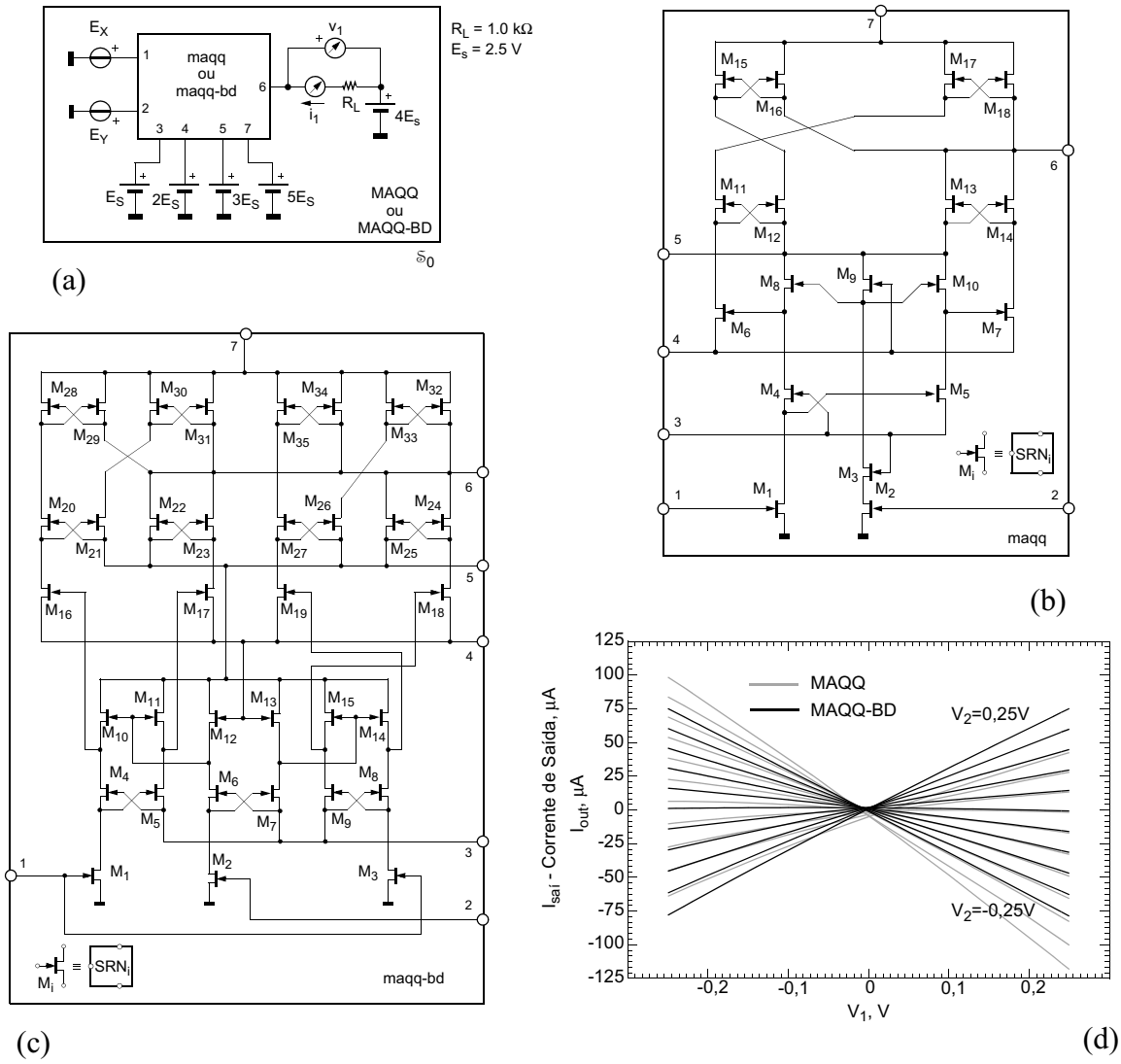


Fig. 8.7 (a) Esquemático do circuito para teste do multiplicador de quatro-quadrantes (MAQQ) e do multiplicador de quatro-quadrantes de baixo-deslocamento (MAQQ-BD) utilizando GaAs MESFETs. (b) Esquemático do MAQQ. (c) Esquemático do MAQQ-BD. (d) Curvas de transferência dos MAQQs operando em regime CC.

estrutura hirarquica de três níveis. No primeiro nível temos 2 SuRs intermediárias ($\mathfrak{S}_{1,1}$,$\mathfrak{S}_{1,2}$) e RC ($\mathcal{C}_0$) com 2 nós de conexão ($n_{0,1}$,$n_{0,2}$). Já no segundo nível, temos a SuR intermediária, $\mathfrak{S}_{1,1}$, que possui 2 SuRs de fundo ($\mathfrak{S}_{2,1}$,$\mathfrak{S}_{2,2}$) e RC ($\mathcal{C}_{1,1}$) com 2 nós de conexão ($n_{1,1}$,$n_{1,2}$), e a SuR intermediária, $\mathfrak{S}_{1,2}$, que possui 3 SuRs de fundo ($\mathfrak{S}_{2,3}-\mathfrak{S}_{3,5}$) e RC ($\mathcal{C}_{1,2}$) com 2 nós de conexão ($n_{1,3}$,$n_{1,4}$). O MAQQ-BD também foi decomposto em uma estrutura de três níveis, com 2 SuRs intermediárias ($\mathfrak{S}_{1,1}$,$\mathfrak{S}_{1,2}$) e RC com 4 nós de conexão ($n_{0,1}-n_{0,4}$) no primeiro nível. No nível seguinte, a SuR intermediária, $\mathfrak{S}_{1,1}$, foi decomposta em 3 SuRs de fundo ($\mathfrak{S}_{2,1}-\mathfrak{S}_{2,3}$) e RC ($\mathcal{C}_{1,1}$) com 2 nós de conexão ($n_{1,1}$,$n_{1,2}$), enquanto a SuR intermediária, $\mathfrak{S}_{1,2}$, foi decomposta em 5 SuRs de fundo ($\mathfrak{S}_{2,4}-\mathfrak{S}_{2,8}$) e RC com 3 nós de conexão, ($n_{1,3}-n_{1,4}$). Neste caso, com a FEE, o número de variáveis introduzido pelas RCs, na decomposição do MAQQ e do MAQQ-BD, produz uma

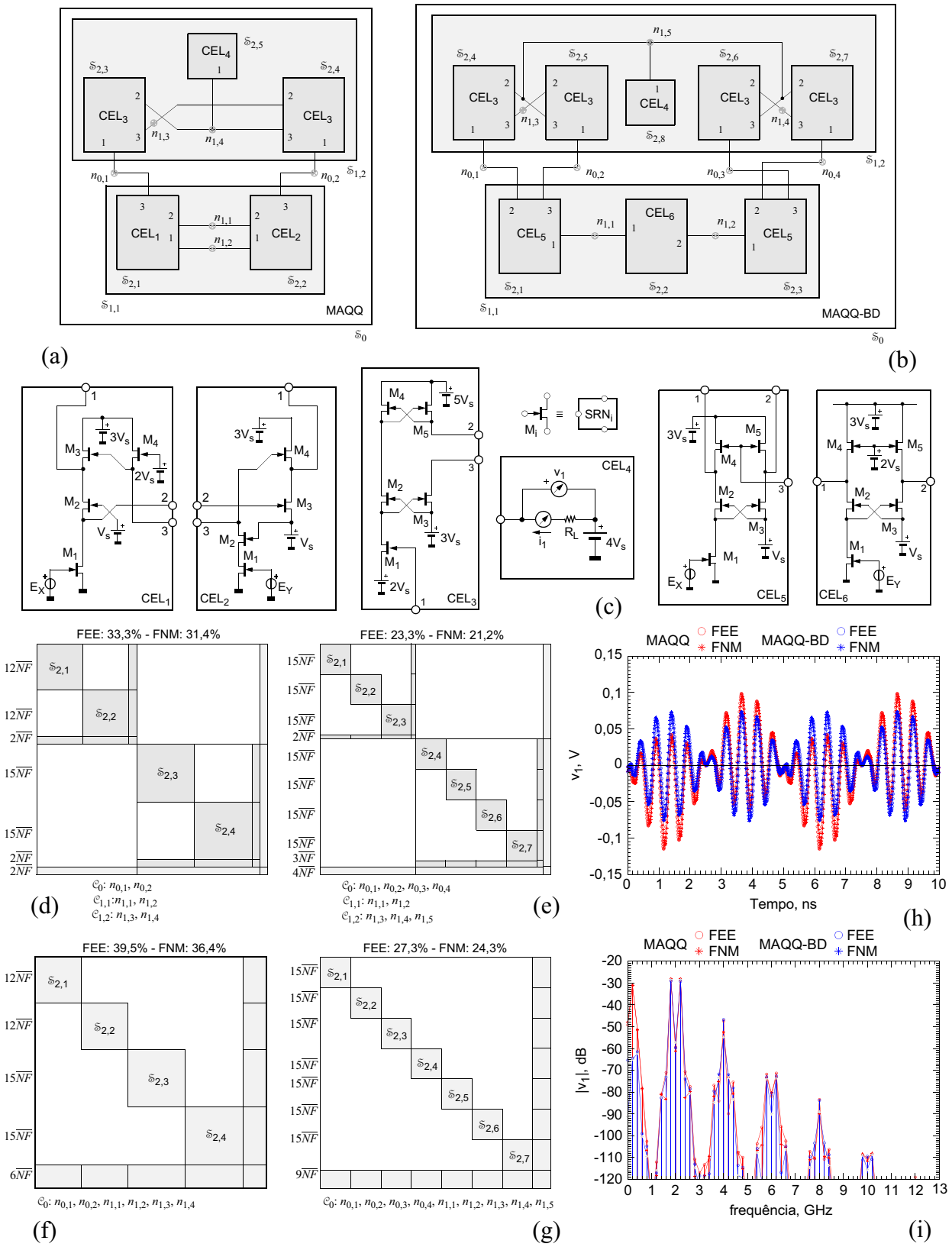


Fig. 8.8 Esquemático (a) do multiplicador analógico de quatro-quadrantes (MAQQ) e (b) do MAQQ-BD decomposto hierarquicamente em super-redes (SuRs). (c) Esquemático dos circuitos das SuRs de fundo utilizadas em (a) e (b). (d) e (e) Estrutura de três-níveis da matriz jacobiana para os esquemas de decomposição (a) e (b), respectivamente (em escala). (f) e (g) Estrutura de dois níveis da matriz jacobiana do MAQQ e do MAQQ-BD, respectivamente (em escala). (a) Formas-de-onda e (b) espectro de frequência da tensão de saída dos MAQQs.

sobre-carga igual ao número de variáveis não-lineares em 2 e 3 transistores num circuito composto por 18 e 35 transistores, respectivamente.

As estruturas do padrão não-zero da matriz jacobiana associadas as decomposições de três-níveis da Fig. 8.8(a) e (b), são representadas nas Fig. 8.8(d) e (e), respectivamente. Se eliminarmos um nível de hierarquia da decomposição multi-níveis da Fig. 8.8(a) e (b), i.e., nivelando as SuRs intermediárias $\mathcal{S}_{1,1}$ e $\mathcal{S}_{1,2}$, obtemos as estruturas de dois-níveis da Fig. 8.8(f) e (g), respectivamente. Observe que neste caso os nós das redes de conexão $\mathcal{C}_{1,1}$ e $\mathcal{C}_{1,2}$ passam a integrar a rede de conexão $\mathcal{C}_0$. Convém ressaltar que a SuR de fundo $\mathcal{S}_{2,5}$ do MAQQ, ver Fig. 8.8(a), e $\mathcal{S}_{2,8}$ do MAQQ-BD, ver Fig. 8.8(b), não introduzem nenhuma variável de estado adicional ao problema, pois representa apenas a terminação de saída, ver Fig. 8.8(c).

Na Fig. 8.8(h) e (i) são ilustrados respectivamente a forma-de-onda e o espectro de frequência da tensão de saída dos MAQQs obtidos via simulação do BH com topologia do espectro de frequência definida para análise de IM de dois-tons. Os circuitos foram excitados pelas fontes de entrada, E_X e E_Y (ver Fig. 8.7(a) e Fig. 8.8(c)), operando com frequência fundamental de 200 MHz e de 2 GHz, e amplitude de 0,25 V.

8.3.7. Multiplicador Analógico Balanceado

O último exemplo a ser considerado é o MAB desenvolvido pela antiga TRW [197], cuja microfotografia do circuito fabricado, utilizando tecnologia de InP-HBT com emissores de 1x10 μm^2, pode ser visualizada na Fig. 8.9(a). O circuito do MAB é composto de um amplificador de RF, um amplificador de OL, um conversor de frequência, um amplificador de FI e redes de adaptação de entrada de RF, de entrada de OL e de saída de FI, conforme ilustrado na Fig. 8.9. Os circuitos dos amplificadores de RF, de OL e de FI, como podemos observar nesta figura, são basicamente compostos de deslocadores de nível de CC de entrada e de saída e um amplificador diferencial. Em particular, o estágio diferencial do amplificador de FI, inclui uma rede de realimentação para obtenção de uma resposta em frequência ultra-banda-larga com sacrifício do ganho, mantendo constante o produto ganho-largura-de-banda. O circuito do conversor de frequência é implementado utilizando a bem-conhecida célula de Gilbert (CG) [202]. Os amplificadores diferenciais de RF (Q_5-Q_6) e de OL (Q_3-Q_4) e o amplificador transcondutivo da CG (Q_1-Q_2) utilizam degeneração de emissor, para ampliação da característica de transferência linear com relação a tensão de entrada.

Além da funções de conversão em frequência de descida e de subida, aplicando-se uma tensão de CC variável na entrada de OL (tensão de controle), o circuito do MAB pode operar como um

amplificador de ganho variável (AGV) controlado por tensão com uma resposta em frequência ultra-banda-larga.

Como podemos observar, na Fig. 8.9, o circuito do MAB utiliza 10, 6 e 6 HBTs de emissor-único no amplificador de RF (ARF), no amplificador de OL (AOL) e na célula de Gilbert (CG), respectivamente. O amplificador de FI (AFI) utiliza 8 HBTs de emissor-único e 2 HBTs de emissor-quádruplo (estágio final de amplificação). Sendo assim, temos um total de 30 HBTs de emissor-único e 2 HBTs de emissor-quádruplo. Se utilizarmos o modelo do HBT distribuído em quatro fatias para representar os HBTs de emissor-quádruplo, então, temos um total de 38 transistores HBTs de emissor-único. Este modelo distribuído do HBT pode ser diretamente utilizado para análise eletro-térmica [76] e, assim como no APC, os transistores de emissor emissor-quádruplo podem ser respresentados por um SuR intermediária, composta de 4 SuRs de fundo correspondendo aos HBTs de emissor-único e, de 1 RC.

O CEE (top-A), utilizado para representar os HBTs de emissor-único e -quádruplo, é ilustrado na Fig. 3.3(b) do Capítulo 3. Os parâmetros elétricos do CEE são listados na Tabela 8.11. Na FEE e FNM, cada HBT introduz 3 e 4 variáveis de estado não-lineares, respectivamente. O modelo de Wei *et al*. [203], empregando 5 diodos, foi utilizado para representar as correntes de condução e de deslocamento do HBT intrínseco. As funções não-lineares que descrevem a operação do HBT em regime de grande-sinal são listadas na Tabela 8.12. Sendo assim, a FEE produzirá uma economia de 25% na dimensão do problema, quando comparado com a FNM. O modelo foi desenvolvido, considerando um HBT de emissor-quádruplo com área de emissor de 1x10 μm^2 e densidade de corrente de aproximadamente 50 kA/cm^2. A tensão de acionamento de base-emissor é entorno de 0,5 V, valor típico para InP HBTs [197]. As frequências f_t e $f_{máx}$ são iguais a 70 GHz e 150 GHz, respectivamente. Estes valores simulados com tensão de coletor-emissor igual a 2,0 V, estão próximos dos valores experimentais fornecidos em [204]. Os parâmetros associados às funções não-lineares utilizadas no modelo de grande-sinal são fornecidos na Tabela 8.13.

O esquemático da decomposição multi-níveis do circuito do MAB é ilustrado na Fig. 8.10(a). Como podemos observar, o circuito foi sub-dividido numa estrutura hierárquica de 3 níveis. Podemos observar na Fig. 8.10(b) que o circuito do ARF é composto dos sub-circuitos ARF1 e ARF2 com 6 e 4 HBTs, respectivamente. Similarmente, o circuito do AFI é composto do AFI1 e AFI2 com 8 e 2 HBTs, respectivamente. No segundo nível (nível 1) da hierarquia, temos a SuR intermediária $\mathfrak{S}_{1,1}$ (AMP-RF), a SuR de fundo $\mathfrak{S}_{1,2}$ (AMP-LO), a SuR de fundo $\mathfrak{S}_{1,3}$ (CG), a SuR intermediária $\mathfrak{S}_{1,4}$ (AMP-FI), e os nós $n_{0,}$ - $n_{0,}$ da rede de conexão $\mathcal{C}_0$. No terceiro nível, temos as SuRs de fundo $\mathfrak{S}_{2,1}$ (AMP-RF1) e $\mathfrak{S}_{2,2}$ (ARF2) que são crianças da SuR intermediária $\mathfrak{S}_{1,1}$;

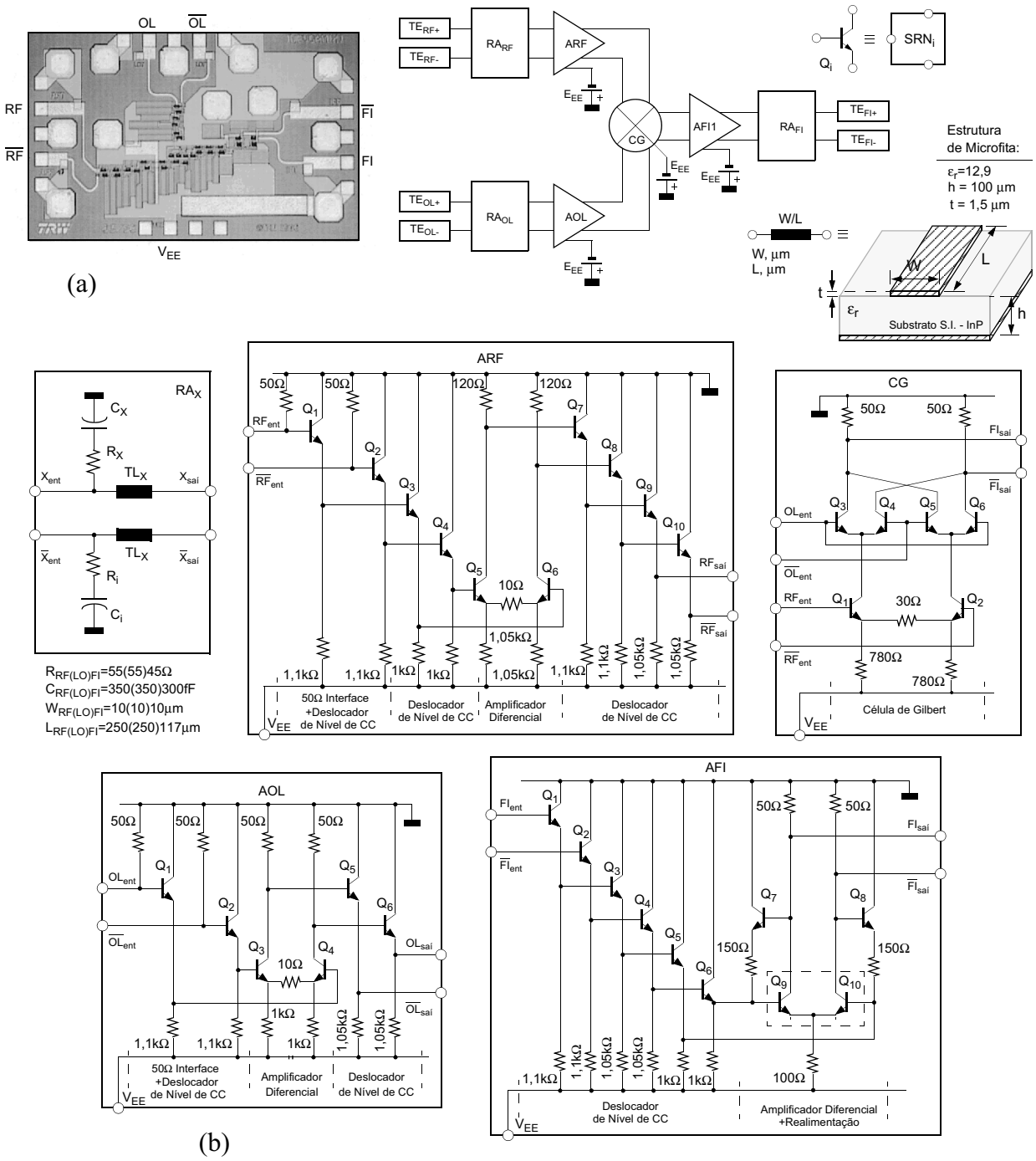


Fig. 8.9 (a) Microfotografia do multiplicador analógico balanceado (MAB) fabricado pela TRW. (b) Esquemático do MAB.

temos as SuRs de fundo $\mathfrak{s}_{2,3}$ (AFI1) e $\mathfrak{s}_{2,2}$ (AMP-FI2) que são crianças da SuR intermediária $\mathfrak{s}_{1,4}$, e os nós da rede de conexão contidas nas SuRs intermediárias $\mathfrak{s}_{1,1}$ e $\mathfrak{s}_{1,4}$, respectivamente. Neste exemplo, temos um total de 114 variáveis de estado não-lineares (FEE). O estruturamento da matriz jacobiana produzido pela decomposição multi-níveis, resulta num custo adional de 10 variáveis de conexão, i.e, acrescenta aproximadamente 10% a dimensão em relação a dimensão do problema original. A estrutura "não-zero" da matriz jacobiana é ilustrada na Fig. 8.10(c).

Tabela 8.11
PARÂMETROS ELÉTRICOS DO CEE (TOP-A) DO INP HBT

Extrinsecos						Intrinsecos			
L_{bpar}^{SE} (pH)	R_{bpar}^{SE} (Ω)	L_{cpar}^{SE} (pH)	R_{cpar}^{SE} (Ω)	L_{epar}^{SE} (pH)	R_{epar}^{SE} (Ω)	R_{b1} (Ω)	R_{b2} (Ω)	R_c (Ω)	R_e (Ω)
10,0	0,05	10,0	0,05	1,0	0,01	6,4	16,0	12,0	3,88
L_{bpar}^{QE} (pH)	R_{bpar}^{QE} (Ω)	L_{cpar}^{QE} (pH)	R_{cpar}^{QE} (Ω)	L_{epar}^{QE} (pH)	R_{epar}^{QE} (Ω)				
2,5	0,0125	2,5	0,0125	0,25	0,0025				

Tabela 8.12
FUNÇÕES NÃO-LINEARES DO MODELO DE GUMMEL-POON DO HBT

Corrente Coletada Direta

$$j_f(v_{be}(t), v_{bc}(t)) = \alpha_f(1 + v_{bc}(t)/V_{A_f})^{-1} j_{D_f}(v_{be}(t))$$

$$j_r(v_{bc}(t), v_{be}(t)) = \alpha_r(1 + v_{be}(t)/V_{A_r})^{-1} j_{D_r}(v_{bc}(t))$$

Cargas de Depleção Base-Emissor ($\mu = be$), Base-Coletor ($\mu = bc$) e Base-Coletor ($\mu = bxc$)

$$q_\mu(v_\mu(t)) = -C_{\mu 0} V_{j\mu}(1 - (1 - v_\mu(t)/V_{j\mu})^{1+m_\mu})/(1+m_\mu) \quad V_{max} = F_c V_{jx}$$

$$e_\mu(v_\mu(t)) = v_\mu(t) - q_\mu(v_\mu(t))/C_{\mu 0}$$

Carga de Depleção Base-Coletor

$$q_{bc}'(v_{bc}(t), v_{be}(t)) = q_{bc}(v_{bc}(t))(1 - j_f(v_{be}(t), v_{bc}(t))/i_0) \quad V_{max} = F_c V_{jx}$$

$$e_{bc}'(v_{bc}(t), v_{be}(t)) = v_{bc}(t) - q_{bc}'(v_{bc}(t), v_{be}(t))/C_{bc0}$$

Carga de Difusão Base-Emissor

$$q_f(v_{be}(t)) = \tau_f \cdot j_{D_f}(v_{be}(t))$$

$$e_f(v_{be}(t)) = v_{be}(t) - q_f(v_{be}(t))/C_{f0}$$

$$C_{f0} = \tau_f R_{D_{f0}}$$

Carga de Difusão Base-Coletor (separar)

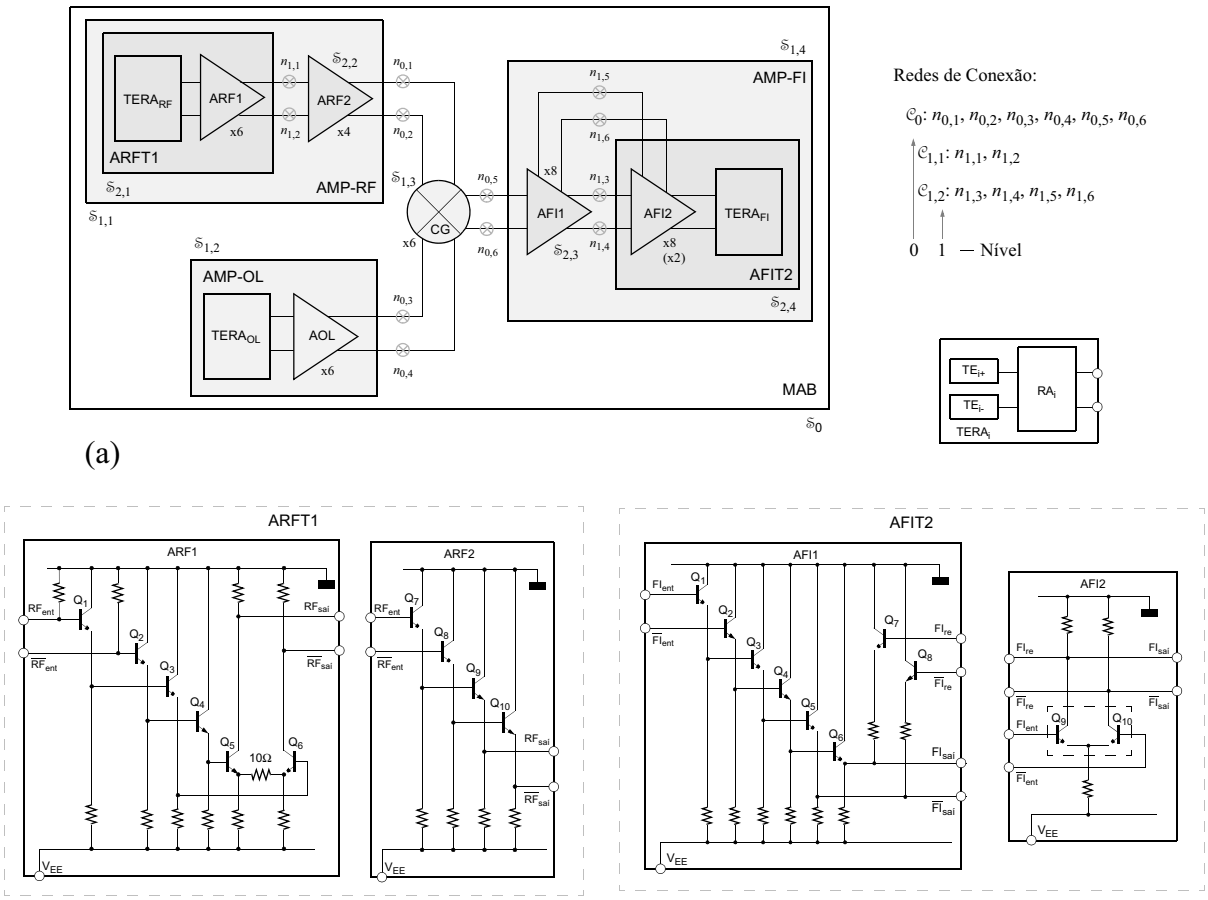
$$q_r(v_{bc}(t), v_{be}(t)) = \tau_r \cdot j_{D_r}(v_{bc}(t)) + \tau_c \cdot j_{D_f}(v_{be}(t))$$

$$e_r(v_{bc}(t), v_{be}(t)) = v_{bc}(t) - q_r(v_{bc}(t), v_{be}(t))/C_{r0}$$

$$C_{r0} = \tau_r R_{D_{r0}}$$

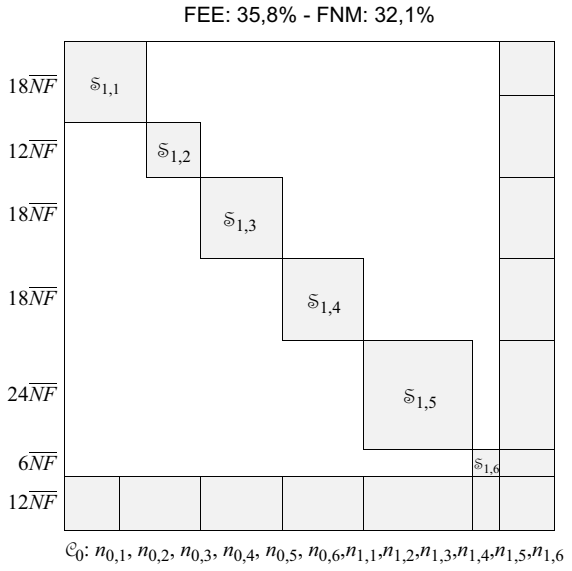
Tabela 8.13
PARÂMETROS DAS FUNÇÕES NÃO-LINEARES DO MODELO DE GUMMEL-POON DO INP HBT

Correntes dos Diodos										
I_{sf} (pA)	n_f	I_{se} (pA)	n_e	I_{sr} (pA)	n_r	I_{sc} (pA)	n_c	I_{sx} (fA)	n_x	V_T (mV)
421,0	1,2	9,15	1,9	872,0	1,6	128,0	2,1	418,0	2,1	26,0
Correntes Coletadas				Cargas de Difusão						
α_f	α_r	V_{A_f} (V)	V_{A_r} (V)	τ_f (ps)	τ_r (ps)	τ_c (ps)				
0,96	0,1	1000	1000	0,4	2,0	1,4				
Cargas de Depleção										
C_{jbe0} (fF)	V_{jbe} (V)	m_{be}	C_{jbc0} (fF)	V_{jbc} (V)	m_{bc}	I_0 (mA)	C_{jbxc0} (fF)	V_{jbxc} (V)	m_{bxc}	F_c
610,0	1,2	0,5	23,8	1,1	0,5	95	15,8	1,1	0,5	0,95

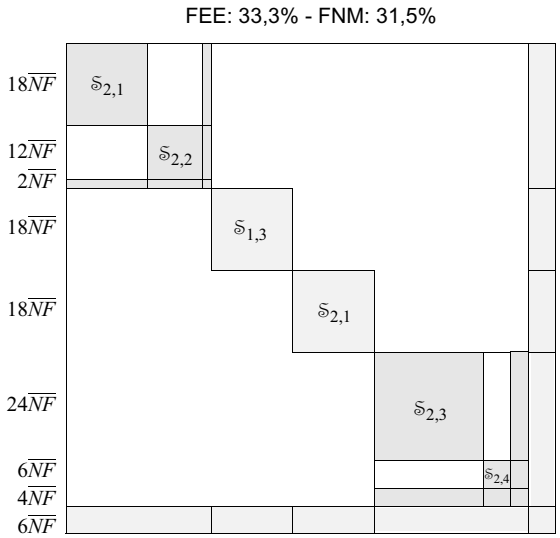


(a)

(b)



(c)



(d)

Fig. 8.10(a) Esquemático do multiplicador análogico balanceado (MAB) decomposto hierarquicamente em super-redes (SuRs). (b) Esquemático do circuito das SuRs de fundo utilizadas em (a). (c) Estrutura de dois níveis da matriz jacobiana do MAB (em escala). (d) Estrutura de três níveis da matriz jacobiana para o esquema de decomposição (a) (em escala).

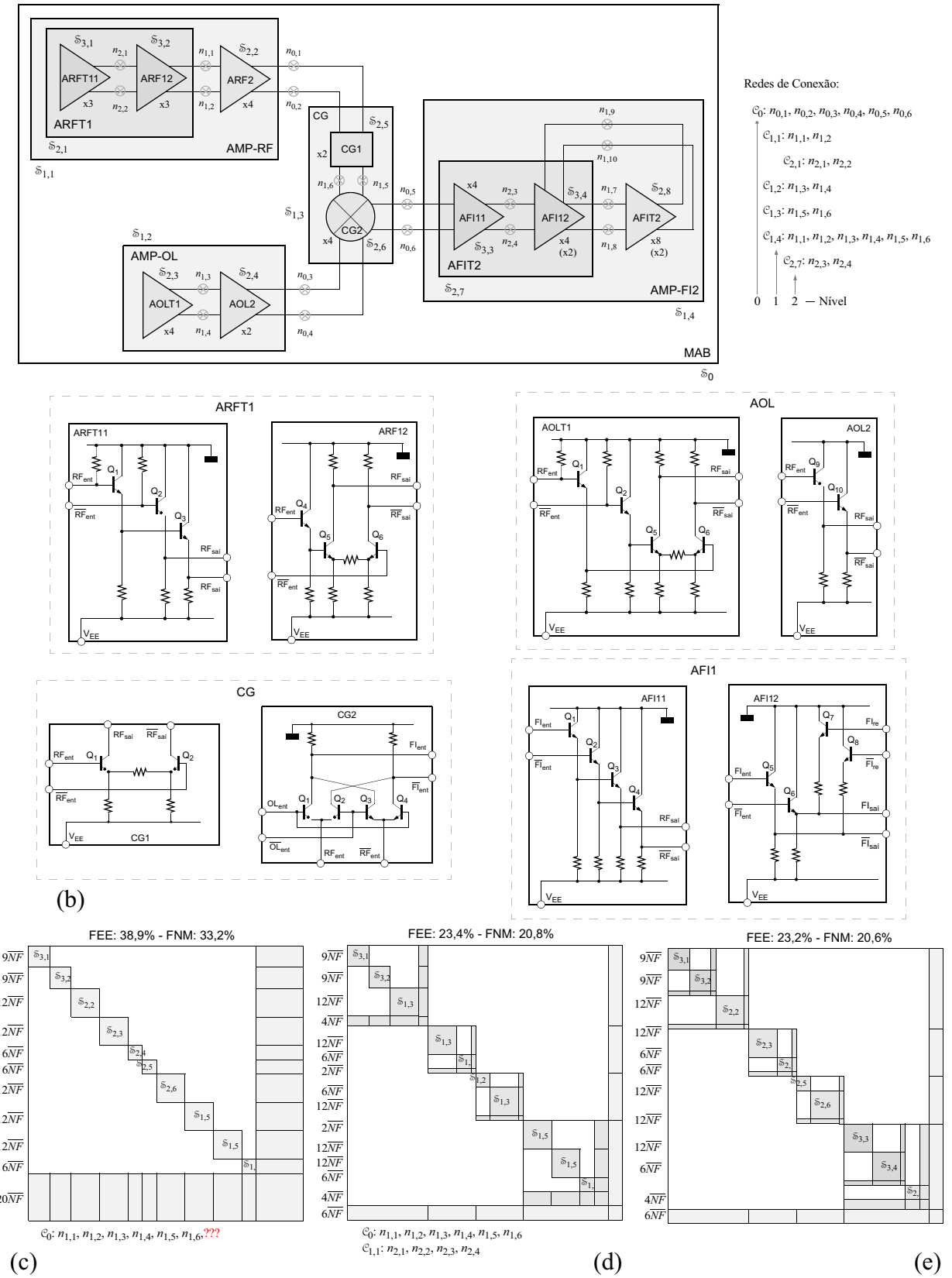


Fig. 8.11(a) Esquemático do multiplicador análogico balanceado (MAB) decomposto hierarquicamente em super-redes (SuRs). (b) Esquemático dos circuitos das SuRs de fundo utilizadas em (a). (c) Estrutura de dois níveis da matriz jacobiana do MAB (em escala). (c) Estrutura de três níveis da matriz jacobiana do MAB (em escala). (d) Estrutura de quatro níveis da matriz jacobiana para o esquema de decomposição (a) (em escala).

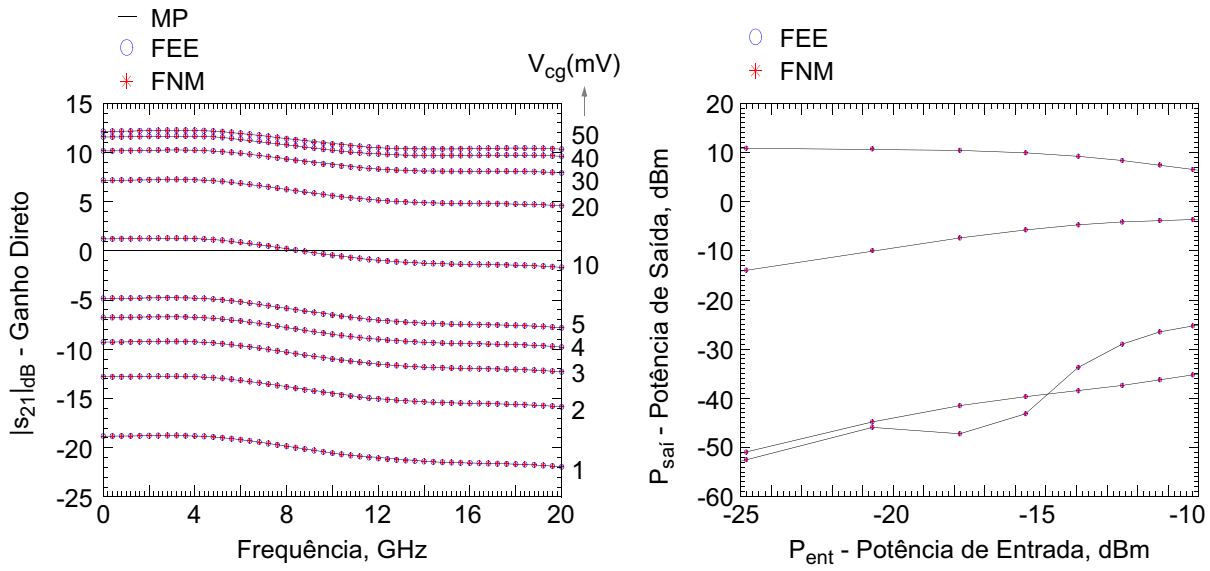


Fig. 8.12(a) Ganho direto versus frequência do MAB operando como amplificador de ganho variável (AGV). (b) Potência de saída versus potência de entrada em regime de único-tom do MAB-AGV.

Os parâmetros de espalhamento do MAB operando como um AGV, em regime de pequeno sinal, podem ser visualizados na Fig. 8.12(a) para diferentes valores da tensão de controle. Como podemos observar, em 10 GHz, para uma variação da tensão de controle 10 à 50 mV obtemos uma variação de 0 à 11 dB no ganho de potência, respectivamente. Na Fig. 8.10(e), utilizando análise do BH de único-tom para determinação de DH com $NH = 8$, podemos observar o efeito de compressão no ganho de potência, para uma tensão de controle igual a 50 mV.

8.4. Desempenho dos Métodos de Newton e do Tensor

Utilizando os circuitos acima, foram conduzidos uma série de testes para avaliar o desempenho do método do tensor quando comparado ao método de Newton (ver Capítulo 6) e do método do tensor inexato quando comparado com o método de Newton inexato (ver Capítulo 7).

Na Tabela 8.14 são listados os resultados de convergência do método do tensor e do método de Newton na determinação do regime de CC dos circuitos testes. Estes resultados foram obtidos adotando-se os seguintes parâmetros para os solucionadores não-lineares, são eles: $\varepsilon_F = 10^{-10}$, $\varepsilon_d = 10^{-12}$, $NI_{\text{máx}} = 150$, $NIP_{\text{máx}} = 10$ e $\lambda_{\text{mín}} = 10^{-7}$. Como podemos observar, para a maioria dos problemas a convergência ocorre em menos de 5 iterações sem necessidade de amortecimento via pesquisa-em-linha. Exceto pelos problemas 7 e 10, envolvendo os circuitos do RA e do MAB-AGV, respectivamente. No problema 7 com FNM, observamos que o método do tensor

Tabela 8.14
LISTAS DOS PROBLEMAS TESTES E DOS RESULTADOS DA ANÁLISE DE CC
NC: NOME DO CIRCUITO, #T: NÚMERO DE TRANSISTORES,
#FN: NÚMERO DE FUNÇÕES NÃO-LINEARES,
#VE: NÚMERO DE VARIÁVEIS DE ESTADO (FEE/FNM)

	NC	Polarização	#T	#FN	#VE FEE/ FNM	Newton FEE/FNM		Tensor PL† Padrão FEE/FNM		Tensor PL† Curvilinear FEE/FNM	
						NI	*NCF*	*NI*	*NCF*	*NI*	*NCF*
1	ACC	V_{CC}=30V	1	2	2 3	2 2	3 3	3 3	4 4	3 3	4 4
2	MF	V_{CC}=12V	1	2	2 3	2 2	3 3	2 2	3 3	2 2	3 3
3	AP	V_{GG}=−1V;V_{DD}=8V	1	6	4 6	4 4	5 5	5 6	6 10	5 5	6 7
4	CFR	V_{GG}=−1.0V	1	6	4 6	2 2	3 3	2 2	3 3	2 2	3 3
5	CFRB	V_{GG}=−1.0V	2	12	8 12	2 2	3 3	2 2	3 3	2 2	3 3
6	APC	V_{GG}=−1V;V_{DD}=8V	8	6	4 6	4 4	5 5	4 4	5 5	4 4	5 5
7	RA	V_{DD}=6V;V_Q=−5,5V	11	66	33 44	17 17	40 41	21 27	100 131	22 12	46 23
8	MAQQ	VCC=12,5V	18	108	54 72	5 5	6 6	4 4	5 5	4 4	5 5
9	MAQQ-BD	VCC=12,5V	35	210	105 140	5 5	6 6	4 4	5 5	4 4	5 5
10	MAB-AGV	V_{EE}=−6V;V_{ctr}=50mV	32	192	96 128	20 20	50 50	32 41	127 254	21 31	52 84
11	MAB-CF	V_{EE}=−6V	32	192	96 128						

†PL = pesquisa-em-linha

metade do NCF, quando comparado como o método de Newton, e também menos NI. Para este mesmo problema com FEE, o desempenho do método do tensor curvilinear é comparável ao do método de Newton. Já no problema 10 com FNM o desempenho do método de Newton, em termos de NCF e NI, é superior ao método do tensor com pesquisa-em-linha curvilinear possui um desempenho comparável ao do método de Newton. Conforme esperado, podemos verificar que o método do tensor empregando a estratégia curvilinear produz uma redução no *NCF* quando comparada a estratégia padrão.

Na Tabela 8.15, é fornecida uma lista dos problemas utilizados para avaliar, na análise do BH, o desempenho do método do tensor versus o método de Newton e do método do tensor inexato versus o método de Newton. Os problemas Nestes problemas os circuitos testes estão operando em diversas situações envolvendo excitação de único-tom, dois-tons e três-tons Para todos os solucionadores foram adotados os mesmos parâmetros utilizados na análise CC, sendo que para os métodos inexatos foi adotada a Escolha 0 para a sequência de termos forçantes com $\eta_0 = 0,001$. O limite no número de iterações do solucionador foi imposto adotando-se $NIL_{\text{máx}} = 500$.

Da Tabela 8.16. podemos observar o baixo desempenho do método do tensor em relação ao

Tabela 8.15

LISTAS COM A ESTATÍSTICA DOS PROBLEMAS TESTES

NC: NOME DO CIRCUITO, DIM. TFD: DIMENSÃO DA TRANSFORMADA DE FOURIER DISCRETA, #D: NÚMERO DE DIODOS, #T: NÚMERO DE TRANSISTORES, #F: NÚMERO DE FREQUÊNCIAS, #FN: NÚMERO DE FUNÇÕES NÃO-LINEARES, #V: NÚMERO DE VARIÁVEIS (FEE/FNM), #VBH: NÚMERO DE VARIÁVEIS NA ANÁLISE DO BH, SNL: SOLUCIONADOR NÃO-LINEAR ("E"=EXATO E "I"=INEXATO)

	NC	Descrição e Parâmetros	DIM. TFD	#D	#T	#F	#FN	#V FEE/ FNM	#VBH (k_{dim})	SNL
1	FARMO	Análise de Distorção Harmônica (DH): NH=32; $E_{ent,máx}$=10V	1	1	0	33	1	1 2	65 130 (5)	E-I
2	FAROC	Análise de DH: NH=32; $E_{ent,máx}$=10V	1	4	0	33	4	4 8	260 520 (10)	E-I
3	ACC	Análise de DH: NH=16; $J_{ent,máx}$=0,1A	1	0	1	17	2	2 3	66 99 (5)	E-I
4	MF	Análise de DH: NH=16; $E_{ent,máx}$=0,4V	1	0	1	17	2	2 3	18 27 (5)	E-I
5	AP	Análise de DH: NH=8; P_{RF}=20dBm; f_{RF}=9GHz	1	0	1	9	6	4 6	85 102 (5)	E-I
6		Análse de Intermodulação (IM) Dois-Tons: NH_{RF1}=6; NH_{RF2}=6; P_{RF1}=P_{RF2}=15dBm; f_{RF1}=9GHz; f_{RF2}=9,01GHz	2			43			425 510 (5)	E-I
7		Análse de IM Três-Tons: NH_{RF1}=4; NH_{RF2}=4; NH_{RF3}=4; P_{RF1}=P_{RF2}=P_{RF3}=10dBm; f_{RF1}=9GHz, f_{RF2}=9,01GHz; f_{RF3}=8,99GHz	3			32			315 378 (5)	E-I
8	CFR	Análise de Conversão em frequência (CF): NH_{RF}=1; NH_{OL}=8; P_{RF}=−10dBm; P_{OL}=10dBm; f_{RF}=94,5GHz; f_{OL}=94GHz	2	0	1	26	6	5 6	255 306 (5)	E-I
9		Análise de CF com IM dois-tons: NH_{RF1}=2; NH_{RF2}=2; NH_{OL}=4; P_{RF1}=P_{RF2}=−10dBm; P_{OL}=10dBm; f_{RF1}=94,49GHz; f_{RF2}=94,51GHz; f_{OL}=94GHz	3			51			505 606 (5)	E-I
10	CFRB	Análise de CF: NH_{RF}=1; NH_{OL}=8; P_{RF}=−10dBm; P_{OL}=10dBm; f_{RF}=94,5GHz; f_{OL}=94GHz	2	1	2	26	12	10 12	510 612 (5)	E-I
11		Análise de CF com IM dois-tons: NH_{RF1}=2; NH_{RF2}=2; NH_{OL}=4; P_{RF1}=P_{RF2}=−10dBm; P_{OL}=10dBm; f_{RF1}=94,49GHz; f_{RF2}=94,51GHz; f_{OL}=94GHz	2			51			1010 1212 (5)	E-I
12	APC	Análise de DH: NH=8; P_{RFin}=30dBm; f_{RF}=10GHz	1	0	8	9	48	40 48	680 816 (5)	E-I
13		Análise de IM dois-tons: NH_{RF1}=3; NH_{RF2}=3; P_{RF1}=P_{RF2}=30dBm; f_{RF1}=10,1GHz; f_{RF2}=10GHz	2			13			1000 1200 (5)	E-I
14		Análse de IM três-tons: NH_{RF1}=2; NH_{RF2}=2; NH_{RF3}=2; P_{RF1}=P_{RF2}=P_{RF3}=25dBm; f_{RF1}=10,1GHz; f_{RF2}=10GHz; f_{RF3}=10,1GHz	3			13			1000 1200 (5)	E-I
15	RA	Análise de DH: NH=8; P_{RF}=15dBm; f_{RF}=1,7 GHz	1	0	11	9	66	33 44	561 748 (5)	E-I
16	MAQQ	Análise de IM dois-tons: NH_X=5; NH_Y=5; V_X=V_Y=1V; f_X=1,0GHz; f_Y=0,2GHz	2	0	18	13	108	54 72	1350 1800 (10)	I
17	MAQQ-BD	Análise de IM dois-tons: NH_X=5; NH_Y=5; V_X=V_Y=1V; f_X=1,0GHz; f_Y=0,2GHz	2	0	35	13	210	105 140	2625 3500 (10)	I
18	MAB-AGV	Análise de DH: NH=16; P_{RF}=10dBm; f_{RF}=8,0GHz	1	0	32	17	192	96 128	1632 2176 (15)	I
19		Análse de IM dois-tons: NH_{RF1}=4; NH_{RF2}=4; P_{RF1}=P_{RF2}=−10dBm; f_{RF1}=8,0 GHz; f_{RF2}=8,0GHz	2			21			3936 5248 (15)	I
20	MAB-CF	Análise de CF (subida): NH_{OL}=8; NH_{RF}=1; P_{RF}=−30dBm; P_{OL}=0dBm; f_{RF}=8,5GHz; f_{OL}=8,0GHz	2			26			4896 6528 (20)	I

Tabela 8.16
LISTAS DOS PROBLEMAS TESTES E DOS RESULTADOS DA ANÁLISE DO BH
NC: NOME DO CIRCUITO, #T: NÚMERO DE TRANSISTORES,
#FN: NÚMERO DE FUNÇÕES NÃO-LINEARES,
#VP: NÚMERO DE VARIÁVEIS DO PROBLEMA FEE/FNM

	NC	#VP FEE FNM	SOL. NL	E=Newton I=Newton Inexact FEE/FNM			E=Tensor I= Tensor inexato PL Padrão FEE/FNM			E=Tensor I=Tensor inexato PL Curvilinear FEE/FNM		
				NI	*NCF*	*NIL*	*NI*	*NCF*	*NIL*	*NI*	*NCF*	*NIL*
1	FARMO	65 130	E	11 11	23 23	—	52 57	233 260	—	34 33	75 73	—
			I	11 11	23 23	67 74	32 31	163 146	223 216	63 65	134 137	308 304
2	FAROC	260 520	E	28 28	65 67	—	94 46	508 206	—	56 52	147 129	—
			I	29 28	66 67	108 92	47 39	212 219	1168 359	46 29	100 69	560 160
3	ACC	18 27	E	5 5	7 7	—	6 6	8 6	—	6 6	8 8	—
			I	6 6	8 8	23 23	6 6	8 8	30 30	6 6	8 8	30 30
4	MF	18 27	E	2 3	3 4	—	2 3	3 4	—	2 3	3 4	—
			I	3 3	4 4	3 3	3 3	4 4	4 4	3 3	4 4	4 4
5	AP	85 102	E	22 21	41 39	—	20 15	74 47	—	20 18	37 32	—
			I	23 21	42 39	61 51	22 20	60 50	89 73	20 18	36 32	73 66
6		425 510	E	20 20	37 36	—	22 16	85 55	—	18 18	33 31	—
			I	21 20	38 36	60 52	21 18	55 46	88 68	18 17	33 31	75 68
7		315 378	E	20 19	37 34	—	19 13	64 37	—	17 13	30 22	—
			I	21 19	38 34	31 27	17 16	41 38	39 37	15 12	26 20	35 29
8	CFR	255 306	E	5 6	6 9	—	5 9	6 21	—	5 8	6 13	—
			I	6 6	7 9	16 14	6 9	7 21	24 43	6 8	7 13	24 32
9		505 606	E	5 6	6 9	—	5 9	6 21	—	5 8	6 13	—
			I	6 6	7 9	16 15	6 9	7 21	25 43	6 8	7 13	25 33
10	CFRB	510 612	E	6 8	7 11	—	5 12	6 29	—	5 9	6 14	—
			I	6 7	7 10	16 18	6 9	7 25	28 50	6 9	7 14	28 40
11		270 324	E	6 8	7 11	—	5 13	6 31	—	5 9	6 14	—
			I	6 7	7 10	16 18	6 9	7 26	26 44	6 9	7 14	26 39
12	APC	680 816	E	11 10	18 16	—	7 7	17 13	—	8 7	11 10	—
			I	12 10	19 16	21 17	9 9	15 13	29 34	8 8	11 12	30 28
13		1000 1200	E	26 26	49 48	—	27 25	113 97	—	22 21	41 39	—
			I	26 26	49 48	115 106	22 22	56 56	146 135	22 21	40 38	135 126
14		1000 1200	E	27 27	51 50	—	26 27	108 109	—	24 23	45 43	—
			I	28 27	52 50	71 63	23 24	61 62	94 91	23 25	43 46	96 98
15	RA	561 748	E	24 24	45 45	—	26 26	109 103	—	22 22	40 40	—
			I	25 24	46 45	171 162	29 26	126 99	289 289	22 23	40 42	225 233
16	MAQQ	1350 1800	I	23 21	40 37	40 39	19 28	50 111	70 97	20 31	36 57	61 82
17	MAQQ-BD	2625 3500	I	22 25	40 39	39 43	16 24	56 90	51 53	21 32	37 57	72 69
18	MAB (AGV)	1632 2176	I	28 29	62 66	116 104	26 27	76 81	161 148	28 27	66 61	178 148
19		3936 5248	I	45 44	117 114	280 203	43 42	150 146	376 292	44 42	114 110	381 292
20	MAB (CF)	4896 6528	I	28 25	65 56	94 89	28 25	87 75	171 158	28 24	64 54	172 151

método de Newton para os circuitos básicos FARMO e FAROC. Estes circuitos operam em regime fortemente não-linear. No caso do circuito do APC é interessante observar a superioridade, em termos de NI e NCF, do método do tensor com pesquisa-em-linha curvilinear quando comparado com os outros métodos. Observamos também, que para quase todos os circuitos o método do tensor com pesquisa-em-linha curvilinear minimiza o número de NCF em relação ao método do tensor com estratégia padrão de pesquisa-em-linha.

Para ilustrar o problema de sobressolução discutido no Capítulo 7, vamos considerar a resolução do Problema 10 listado na Tabela 8.15, e referente ao circuito do MAB-AGV. Mais precisamente, vamos considerar inicialmente a solução utilizando o método de Newton inexato com a sequência de termos forçantes definida pela Escolha 0 e com um termo forçante inicial, η_0, igual a 0,001. O resultado desta solução é apresentado nos gráficos de histórico de convergência da Fig. 8.13(a). Já na Fig. 8.13(b), podemos observar o histórico de convergência utilizando a Escolha 5 e $\eta_0 = 0,1$. Assim como no exemplo do Capítulo 7 (ver Fig. 7.3(a)-(b)), podemos observar que a Escolha 5 minimiza o problema de sobressolução. Em adição, aos testes acima, também foi realizado um teste numérico utilizando o mesmo problema porém solucionado com o método de tensor inexato. O resultado deste teste utilizando a Escolha 0 e $\eta_0 = 0,001$ pode ser observado na Fig. 8.13(c). Para verificar o efeito da Escolha 7 foi realizado um teste com $\eta_0 = 0,1$ cujo resultado é ilustrado na Fig. 8.13(d). Como podemos observar o método do tensor inexato com a Escolha 7 também produz uma minimização do problema de sobressolução. Para análise do BH utilizando a FEE podemos observar que as conclusões são as mesmas, exceto por um pequeno aumento no número maior de iterações. Vale ressaltar que a FEE exige um menor número de variáveis, o que pode resultar num menor tempo de processamento por iteração.

Conforme destacado no Capítulo 7 o uso de pré-condicionadores pode ser vital para o sucesso do solucionador linear iterativo utilizado pelos métodos inexatos de Newton e do tensor. Sendo assm, considerando o circuito do APC (problema #10), operando em regime de único-tom aproximado por 8 harmônicos. Na Fig. 8.14, podemos observar graficamente a ação do pré-condicionador do tipo bloco Jacobi, i.e., formado só com informação de CC, sobre o espectro da matriz jacobiana do BH para uma potência de entrada de 10 e 30 dBm. Estes níveis de potência de entrada estão abaixo do ponto de compressão de 1 dB. Vale ressaltar que a matriz jacobiana foi calculada na solução do problema. O raio espectral, ρ, e o número de condicionamento, κ, da matriz jacobiana são fornecidos nos gráficos da Fig. 8.14. Como podemos observar da Fig. 8.14(a), para uma simulação com potência de entrada 10 dBm, o pré-condicionador na raiz do problema possue um efeito dramático na distribuição dos auto-valores, aglomerando-os em torno do ponto

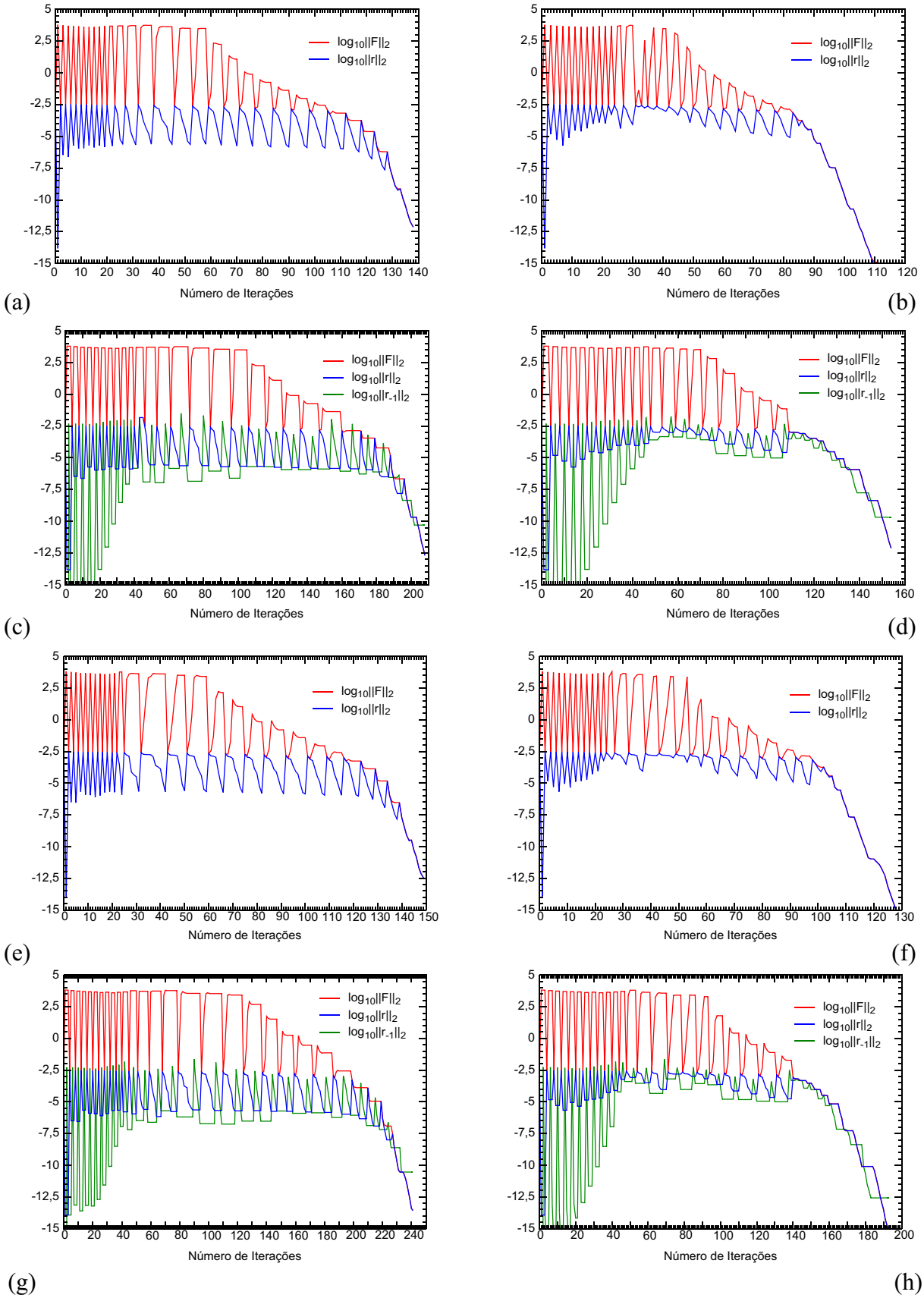


Fig. 8.13Histórico de convergência para solução do BH do MAB-AGV (problema 18) via FNM utilizando o método de método de Newton inexato com (a) Escolha 0 e (b) Escolha 5. Histórico de convergência do problema 18 utilizando o método do tensor curvilinear inexato com (c) Escolha 0 e (d) Escolha 7. Os resultados (e)-(h) se referem à FEE.

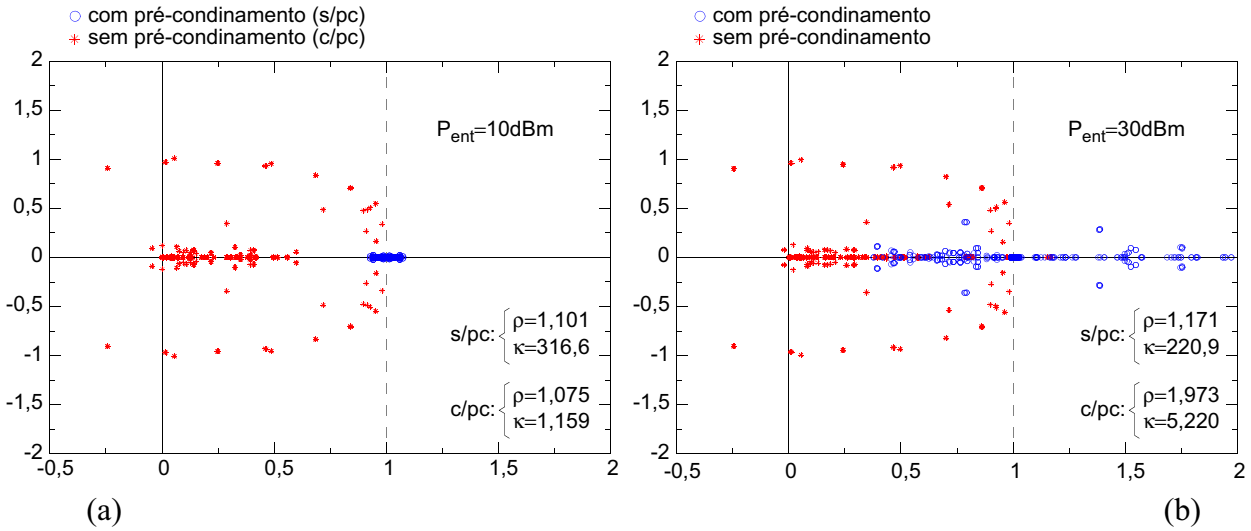


Fig. 8.14Distribuição dos auto-velores (espectro) da matriz jacobiana do BH para o circuito do APC, calculada na raiz do problema, com potência de entrada, P_{ent} igual (a) 10 e (b) 30 dBm.

$(1, 0)$. Se elevarmos a potência de entrada para 30 dBm, podemos observar que o pré-condicionador ainda é bastante eficaz em mover os auto-valores da origem, i.e., ponto $(0, 0)$. São estes auto-valores que causam problemas para os métodos lineares iterativos.

8.5. Desempenho da Análise do BH Multi-Níveis

O desempenho da análise do BH utilizando a técnica de decomposição multi-níveis foi avaliado utilizando os circuitos do RA, MAQQ, MAQQ-BD e MAB, descritos anteriormente. Convém destacar que os resultados apresentados a seguir foram obtidos via solução explícita, conforme discutido em [41]-[43]. Na Tabela 8.17 apresentamos a lista dos problemas utilizados nos testes descritos abaixo.

Para avaliarmos graficamente o desempenho da decomposição multi-níveis, vamos introduzir uma RD definida em termos do tempo de processamento por iteração, t_{PI}. Mais precisamente, como: $RD(t_{PI}) = \log_2(t_{PI}\text{ - sem decomposição} / t_{PI}\text{ - com decomposição multi-níveis})$. Utilizando esta definição, na Fig. 8.15(a), podemos observar a $RD(t_{PI})$ para os circuitos do RA (ver Fig. 8.6(a)-(c)), dos MAQQs (ver Fig. 8.8(a)-(g)) e do MAB operando como AGV (ver Fig. 8.10 e Fig. 8.11). Estes resultados foram obtidos utilizando o método de Newton como solucionador não-linear. No solucionador empregado os sistemas jacobianos são resolvidos utilizando um processo de fatorização e retro-substituição LU com técnicas de matriz esparsa. Para globalização foi adotada a estratégia de pesquisa-em-linha com interpolação quadrática. Em adição, foi empregado um

Tabela 8.17
LISTAS COM A ESTATÍSTICA DOS PROBLEMAS TESTES
NC: NOME DO CIRCUITO, #NH: NÚMERO DE NÍVEIS DA HIERARQUIA,
DIM. TFD: DIMENSÃO DA TRANSFORMADA DE FOURIER DISCRETA,
#T: NÚMERO DE TRANSISTORES, #F: NÚMERO DE FREQUÊNCIAS,
#FN: NÚMERO DE FUNÇÕES NÃO-LINEARES, #V: NÚMERO DE VARIÁVEIS,
#VBH: NÚMERO DE VARIÁVEIS NA ANÁLISE DO BH,
SOL. NL: SOLUCIONADOR NÃO-LINEAR ("E"=EXATO E "I"=INEXATO)

	NC	#NH	Descrição e Parâmetros		DIM. TFD	#T	#F	#FN	#V FEE	#VBH	k_{dim}	SOL. NL
1	RA	2	Análise de DH: P_{in}=15dBm; f=1,7 GHz	NH=4 NH=8	1	11	5 9	66	33	324 612	– 5	E I
2				NH=6 NH=12			7 13			468 900	– 5	E I
3				NH=8 NH=16			9 17			612 1188	– 5	E I
4	MAQQ	2	Análise de IM dois-tons: V_X=V_Y=1V; f_X=1,0GHz; f_Y=0,2GHz	NH_{RF1}=3; NH_{RF2}=3 NH_{RF1}=6; NH_{RF2}=6	2	18	13 13	108	54	1500 5100	– 5	E I
5				NH_{RF1}=4; NH_{RF2}=4 NH_{RF1}=7; NH_{RF2}=7			21 21			2460 6780	– 5	E I
6				NH_{RF1}=5; NH_{RF2}=5 NH_{RF1}=8; NH_{RF2}=8			31 31			3660 8700	– 5	E I
7		3		NH_{RF1}=3; NH_{RF2}=3 NH_{RF1}=6; NH_{RF2}=6			13 13			1500 5100	– 5	E I
8				NH_{RF1}=4; NH_{RF2}=4 NH_{RF1}=7; NH_{RF2}=7			21 21			2460 6780	– 5	E I
9				NH_{RF1}=5; NH_{RF2}=5 NH_{RF1}=8; NH_{RF2}=8			31 31			3660 8700	– 5	E I
10	MAQQ-BD	2	Análise de IM dois-tons: V_X=V_Y=1V; f_X=1,0GHz; f_Y=0,2GHz	NH_{RF1}=3; NH_{RF2}=3 NH_{RF1}=6; NH_{RF2}=6	2	35	13 13	210	105	2850 9690	– 5	E I
11				NH_{RF1}=4; NH_{RF2}=4 NH_{RF1}=7; NH_{RF2}=7			21 21			4674 12882	– 5	E I
12				NH_{RF1}=5; NH_{RF2}=5 NH_{RF1}=8; NH_{RF2}=8			31 31			6954 16530	– 5	E I
13		3		NH_{RF1}=3; NH_{RF2}=3 NH_{RF1}=6; NH_{RF2}=6			13 13			2850 9690	– 5	E I
14				NH_{RF1}=4; NH_{RF2}=4 NH_{RF1}=7; NH_{RF2}=7			21 21			4674 12882	– 5	E I
15				NH_{RF1}=5; NH_{RF2}=5 NH_{RF1}=8; NH_{RF2}=8			31 31			6954 16530	– 5	E I
16	MAB-AGV	2	Análise de DH: P_{RF}=10dBm; f_{RF}=8,0GHz	NH=4 NH=8	1	32	5 9	192	96	972 1836	– 15	E I
17				NH=6 NH=12			7 13			1404 2700	– 15	E I
18				NH=8 NH=16			9 17			1836 3564	– 15	E I
19		3		NH=4 NH=8			5 9			972 1836	– 15	E I
20				NH=6 NH=12			7 13			1404 2700	– 15	E I
21				NH=8 NH=16			9 17			1836 3564	– 15	E I
22		2		NH=4 NH=8			5 9			1044 1972	– 15	E I
23				NH=6 NH=12			7 13			1508 2900	– 15	E I
24				NH=8 NH=16			9 17			1972 3828	– 15	E I
25		3		NH=4 NH=8			5 9			1044 1972	– 15	E I
26				NH=6 NH=12			7 13			1508 2900	– 15	E I
27				NH=8 NH=16			9 17			1972 3828	– 15	E I
28		4		NH=4 NH=8			5 9			1044 1972	– 15	E I
29				NH=6 NH=12			7 13			1508 2900	– 15	E I
30				NH=8 NH=16			9 17			1972 3828	– 15	E I

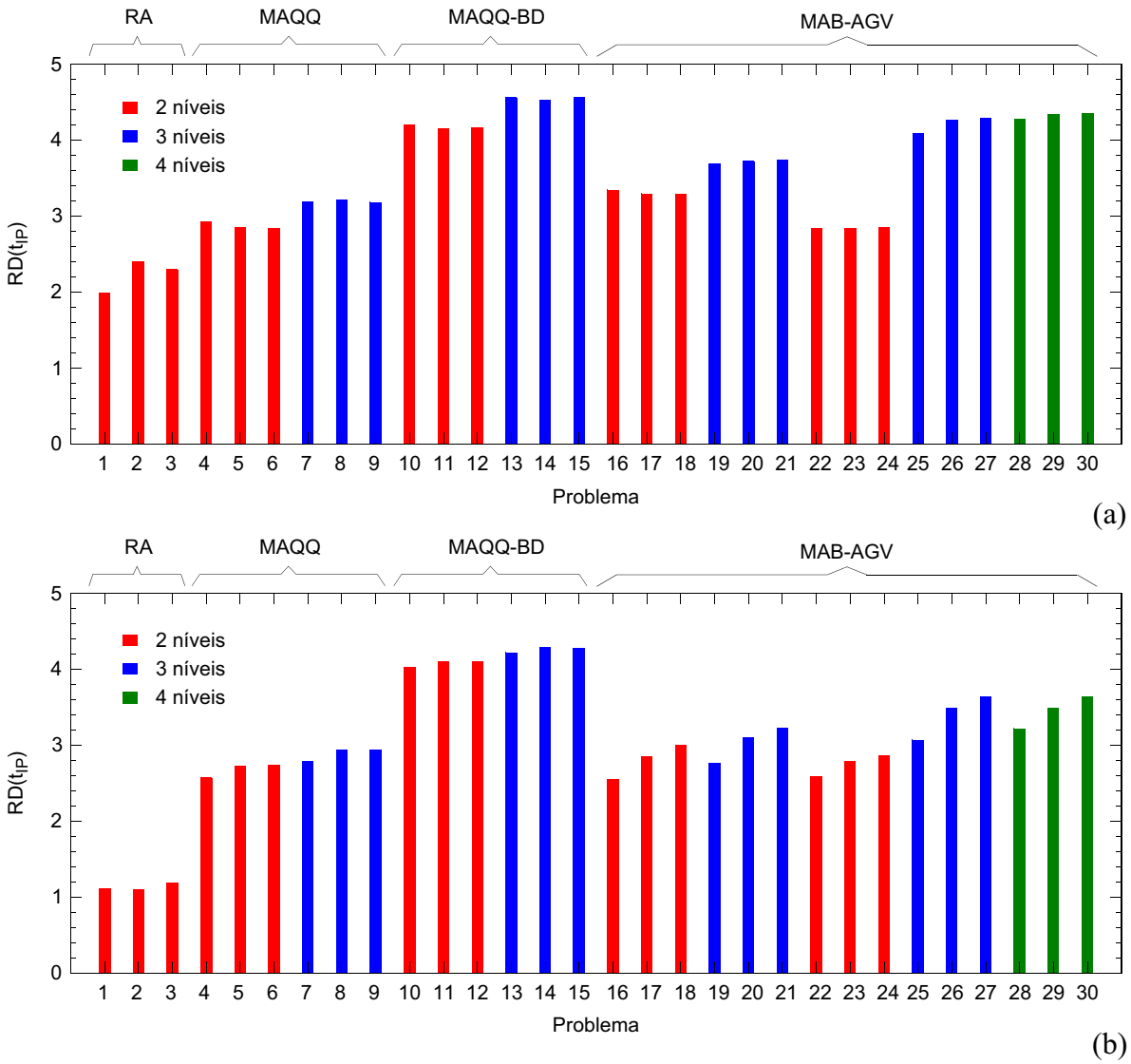


Fig. 8.15Gráficos da razão de desempenho (RD) da análise do BH multi-níveis. (a) Solução "exata" dos sistemas jacobianos via fatorização/retro-substituição LU. (b) Solução "inexata" via GMRES/pré-condicionador LU.

espectro de derivada (ver definição no Capítulo 5), que inclui todas as frequências empregadas nas TFDs para conversão do sinal do domínio do tempo para o domínio da frequência, e vice-versa. Como podemos observar pelos resultados o impacto da decomposição multi-níveis é dramático produzindo um aumento na velocidade de resolução do problema acima de 16(=2^4) vezes. Também podemos observar que as RDs são praticamente insensitivas a variação no número de linhas espectrais utilizadas para representação do sinal.

Complementando a análise de desempenho, na Fig. 8.15(b) podemos observar a $RD(t_{PI})$ para a análise do BH utilizando o método de Newton inexato como solucionador não-linear. Assim como no teste acima, a pesquisa-em-linha via interpolação quadrática foi utilizada como estratégia de globalização. Adicionalmente, como pré-condicionador foi utilizado a matriz jacobiana com

espectro de derivada incluindo apenas a componente de CC. Como podemos observar a decomposição multi-níveis produz impacto apenas um pouco menor do que na solução "exata" via o método de Newton. Observe que o circuito do RA é o que apresenta o menor impacto na redução do t_{PI} via decomposição multi-níveis. Isto deve-se a menor complexidade deste circuito em relação aos outros circuitos considerados. Na solução via solucionadores inexatos a maior contribuição da decomposição multi-níveis está relacionada a fatorização do pré-condicionador.

8.6. Conclusão

Apresentamos neste capítulo, os resultados numéricos referente a análise do BH forçada em regime multi-tons, para uma série de circuitos testes. As diversas comparações entre os resultados obtidos validam a teoria apresentada nos capítulos anteriores.

Nos exemplos de circuito com hierarquia multi-níveis, podemos observar que a escolha dos pontos de decomposição do circuito para a formação de uma estrutura hierárquica de SuRs, pode ser eficientemente realizada de forma intuitiva. Os parâmetros de espalhamento fornecidos acima foram obtidos utilizando análise MP e utilizando análise de CA de pequeno-sinal. Nesta última análise, as equações para cada ponto de frequência, são as mesmas utilizadas na composição da equação determinante do BH. Sendo assim, possibilitam a validação do processo de formulação das equações de circuito para análise do BH. Para o CFR os resultados da análise do BH foram validados pelos resultados obtidos pela análise de CF. Lembrando que neste caso, a análise do BH é conduzida utilizando uma excitação de dois-tons referente ao sinal de OL e de RF. Sendo que o sinal de RF é considerado como um pequeno-sinal representado por apenas 1 harmônico.

Uma série de teste foram realizados para determinar o desempenho do método do tensor versus o método de Newton, e do método do tensor inexato versus o método de Newton inexato, utilizando os circuitos testes introduzido neste capítulo. Estes testes indicam que o método do tensor com pesquisa-em-linha curvilinear e o método de Newton possuem um desempenho comparável. Para avaliar o desempenho da decomposição multi-níveis na análise do BH foram realizados diversos testes, que indicaram uma redução dramática no tempo de processamento, quando comparada com a solução sem decomposição multi-níveis.

9. Conclusão

9.1. Considerações Finais

APRESENTAMOS, NESTE TRABALHO, uma nova e eficiente metodologia para análise do BH de circuitos de RF não-lineares forçados, operando em regime multi-tons. Para circuitos em grande-escala, foi introduzida uma técnica de decomposição multi-níveis que permite a subdivisão do circuito (ou sistema) em uma estrutura hierarquica composta de SuRs e RCs. Incluindo a SuR de tôpo (nível 0), foram definidas SuRs do tipo intermédiaria e de fundo. As SuRs intermédiarias são compostas de SuRs (intermediárias e/ou de fundo) de nível superior e de uma RC. Os elementos estruturais que integram o circuito são todos incluídos nas SuRs de fundo. Estas SuRs sofrem decomposição de circuito por partes, para separação em uma parte linear, representada pela SRLA, e outra parte não-linear, representada pela SRNC.

Para a formulação da SRNC, foi introduzida a FEE, que permite, em geral, uma representação mais eficiente dos dispositivos semicondutores (diodos, transistores, lasers, etc), em termos do números de váriaveis de estado não-lineares, quando comparada com a FNM. Em adição, produz expressões mais simples que a FEEP, e ao contrário desta, sua derivação é automática, e obtida via um eficiente procedimento tabular. Refletindo o cárater generalizado da FEE proposta, caso seja conveniente, esta pode assumir a mesma forma da FNM utilizada na formulação de DDSs. Devido à limitação de espaço, e por estar fora do contexto deste trabalho, não foi apresentada a extensão original da FEE incluindo fontes de ruído arbitrariamente correlatas. Para uma eficiente formulação da SRLA adotamos a FNM.

Vale ressaltar que a formulação das SuRs de fundo pode ser conduzida sem decomposição de circuito por partes. Neste caso, as matrizes constitutivas das equações de estado e de sonda destas SuRs, obitidas via FEE ou FNM, são invariantes em frequência e assumem uma estrutura altamente esparsa. Em SuRs de fundo de grande-escala, a FNM é a mais indicada por ter um superior condicionamento numérico e uma forte dominância diagonal. Este tipo de representação é ideal para análise e otimização no domínio do tempo [134] ou em um domínio misto frequência-tempo [81]. Apesar da maior dimensão do problema, pela simplicidade na formulação das equações do circuito, a análise do BH sem decomposição em pedaços e empregando a FNM é comumente utilizada [45],[22],[51].

Lançando mão da combinação SRNC/FEE e SRLA/FNM para formulação das SuRs de fundo, foi apresentado um procedimento generalizado para a obtenção das equações de um circuito

hierárquico composto de múltiplos níveis. Nesta formulação, as matrizes constitutivas assumem uma estrutura multi-níveis, onde cada nível possui uma estrutura esparsa do tipo bloco diagonal sem e com bordas. Para a realização das conversões tempo-frequência do sinal, apresentamos uma discussão detalhada na teoria e na implementação da TFD para sinais quasi-periódicos resultante de excitação dois-tons, três-tons e multi-tons. A inclusão de sinais com modulação digital utilizando a teoria de multi-senos foi discutida. Na verdade, qualquer forma de excitação multi-senos é possível neste trabalho. Para lidar com este tipo de sinais, foi utilizada a TFMT, incluindo uma generalização que possibilita a análise do BH com múltiplas portadoras de RF moduladas por sinais complexos. Definida a formulação das equações do circuito, a topologia de espectro de frequência e as TFDs, apresentamos como formar a equação determinante do BH no contexto da decomposição e formulação multi-níveis. A matriz jacobiana associada à esta equação determinante possui múltiplos níveis, onde cada nível assume uma estrutura do tipo bloco diagonal com borda dupla. Em adição, as expressões analíticas que descrevem os elementos da matriz jacobiana associada às funções não-lineares das SuRs de fundo para SRNC/SRLA com FEE/FNM, confirmam a maior simplicidade destas expressões, quando comparada com a técnica FEEP/FNM [18]. O conceito de espectro de frequência de derivadas foi utilizado para o controle de esparsidade ou da largura de banda desta matriz jacobiana e dos pré-condicionadores que estão fundamentados nela.

Para a análise de CC e para análise do BH em pequena- e média-escala foram discutidos e implementados os métodos do tensor e de Newton globalizados com a estratégia de pesquisa-em-linha via retrocedimento. Para o método do tensor, foram discutidas e implementadas as estratégias de pesquisa-em-linha padrão e curvilinear [161],[31],[34]. O método do tensor está fundamentado na solução de um modelo local com informação de segunda-ordem, descrita por um objeto tensor de posto-um. A formação e a solução do modelo do tensor possuem um custo comparável à do método de Newton (ou método padrão). Os benefícios esperados do método do tensor são a maior robustez e o menor número de iterações, quando comparado ao método de Newton, que está fundamentado na solução de um modelo linear local a cada iteração. Os testes preliminares, que foram realizados utilizando problemas especiais, confirmam estes benefícios.

Para análise do BH, em grande-escala, foram discutidos e implementados os métodos do tensor inexato e de Newton inexato. Para a solução do modelo do tensor inexato, foram empregados os processos de solução simplificada [32], solução modificada [114] e solução completa [31],[34]. Nos processos de solução simplificada e modificada, o solucionador linear iterativo é o GMRES(m) implementação escalar ou o GMRES-B2(m) implementação em bloco. Caso o GMRES(m) seja utilizado, duas soluções aproximadas são necessárias para determinação da

correção do tensor. O método TGMRES(m) com implementações em bloco B2 e B3 foram utilizados no processo de solução completa. O método de Newton inexato utiliza o GMRES(m) como solucionador linear interno. O pré-condicionamento à direita foi adotado para aumentar a robustez e acelarar a convergência dos solucionadores iterativos para solução do modelo linear (Newton) e do modelo do tensor. Nos métodos do tensor inexato e de Newton inexato (ou padrão), adotamos a globalização via pesquisa-em-linha com retrocedimento. As estratégias de pesquisa-em-linha padrão e curvilinear, para o método do tensor inexato, foram consideradas. Testes preliminares em problemas especiciais foram conduzidos para a validação das implementações propostas e também confirmam uma superioridade do método do tensor inexato frente ao método padrão.

Algoritmos, com uma descrição detalhada de todos os métodos citados acima, foram apresentados. Nos solucionadores não-lineares, eles descrevem uma implementação modificada que permite a re-utilização da matriz jacobiana, produzindo iterações de corda paralela ou iterações do tipo Shamaskii. Para simplificar a avaliação de desempenho, alguns monitores de convergência implementados não foram discutidos e nem utilizados neste trabalho.

Para validação da teoria proposta nesta tese, foram desenvolvidos vários exemplos de circuito de RF. Em particular, a validação do processo de formulação foi conduzida comparando os resultados da análise CA, em regime de pequeno sinal, com os resultados obtidos com a análise multi-porta utilizando parâmetros híbridos. Ressaltando que, a equação determinante do BH consiste da combinação de equações da análise de CA para cada linha espectral considerada na análise. Devido a limitação de espaço, apresentamos apenas uma suscinta discussão da sofisticada implementação númerica desenvolvida durante e antes deste trabalho de doutorado.

9.2. Trabalhos em Adamento e Futuros

Abaixo apresentamos um resumo das atividades que já estão e que serão conduzidas como extensão do trabalho apresentado neste manuscrito. Entretanto, o ideal é que estas atividades sejam desenvolvidas através da formação de um sólido grupo de *pesquisa e desenvolvimento* (P&D). Com a formação deste grupo, podemos ter um significante auxílio para conduzir novas implementações em software e realizar validações numéricas, ficando mais tempo para que possamos desenvolver novas teorias e novos algoritmos. Principalmente, se considerarmos que estamos lidando com o desenvolvimento de um avançado ambiente de software para simulação de circuitos integrado de RF não-lineares e em grande-escala. Em adição, estas atividades podem ser conduzidas, de forma eficiente, como programas de pesquisa de iniciação científica, mestrado, e

doutorado, dependendo da complexidade do problema. Nestes projetos, seria interessante incluir a fabricação e a caracterização, em laboratório de circuitos e sistemas de telecomunicação em tecnologia de MMICs e M^3ICs. Isto possibilita uma validação experimental dos resultados númericos gerados por este trabalho. Vale destacar que, para o projeto preciso de um circuito ou sistema de RF, é de grande importância viabilizar uma estrutura de caracterização experimental e de modelagem numérica dos dispositivos semicondutores utilizados no projeto.

Apartir do final deste trabalho de tese, pretendemos investigar o desempenho dos métodos de Newton e do tensor utilizando o modelo da região de confiança como estratégia de globalização. Também pretendo implementar e avaliar a pesquisa-em-linha não-monotônica citada anteriormente. Com relação a eficiência e robustez dos solucionadores lineares iterativos de subespaço de Krylov, é interessante desenvolver novos tipos de pré-condicionadores. No momento já foi implementado a técnica da *matriz inversa aproximada* [193] e, assim que possível, avaliaremos o seu desempenho na formação de pré-condicionadores para a análise do BH, incluindo a extensão deste tipo de pré-condicionador na formulação multi-níveis. Na metodologia apresentada neste trabalho, já está prevista a possibilidade de definir diferentes topologias de espectro de frequência de derivadas para cada SuR de fundo presernte no circuito. Isto possibilitará um novo tipo de controle de esparsidade multi-níveis. Também, pretendemos desenvolver uma nova técnica de solução multi-níveis para análise do BH, onde as SuRs de fundo podem ter diferentes topologias de espectro de frequência, para a representação do sinal, dependendo da sua operação no circuito.

Realizando poucas modificações, o método do BH apresentado neste trabalho pode facilmente ser utilizado na análise e otimização de circuitos autônomos e sincronizados [163] e também na análise de estabilidade de circuitos. Na análise destes circuitos, seria interessante investigar o desempenho do método do tensor em relação ao método de Newton, principalmente na vizinhança de um ponto de virada (*turning point*) [114]. Para otimização de circuitos de não-lineares (forçados, autônomos e sincronizados), será preciso implementar o cálculo das sensitividades com relação ao parâmetro de otimização. Lembramos que vários métodos de otimização, incluindo o método do tensor [205],[206], já se encontram implementados na biblioteca NUPACK. A solução de dois-níveis para circuitos autônomos deve ser desenvolvida, onde, neste caso, uma aproximação inicial da solução do problema é obtida via otimização e, no segundo nível, a solução final é obtida via um solucionador não-linear que utiliza a solução do primeiro nível como iteração inicial. Obviamente, espera-se que esta solução inicial esteja dentro da região de convergência do solucionador não-linear utilizado.

Apesar de não ser descrita neste trabalho, por uma limitação de espaço, o método da análise de

conversão (AC) [153], foi desenvolvido em paralelo para validação dos resultados do BH em circuitos conversores de frequência. Em um próximo passo, estaremos extentendo o método AC para análise de conversão de frequência e de modulação do ruído. As aplicações deste tipo de análise são a determinação da figura de ruído em CFs e a determinação do ruído de fase em sistemas autônomos e sincronizados.

Com a introdução da metodologia proposta em [70], é possível a aplicação imediata do método do BH na análise e na otimização de sistemas de rádio-enlace, envolvendo múltiplas antenas de transmissão e de recepção. Esta metodologia está fundamentada no conceito da matriz de ligação, que, por sua vez, descreve as estruturas de radiação do transmissor e do receptor e o meio de transmissão utilizado. A matriz de ligação pode ser determinada, considerando campo próximo ou campo distante, via um simulador eletromagnético. Em casos simplificados, a matriz de ligação pode ser derivada analiticamente. Uma aplicação interessante desta metodologia é a caracterização de sistemas de *radio frequency identication* (RFID) [115], *milimeter-wave identification* (MMID) [207] e sistemas de radar em geral. No contexto deste trabalho, caso a análise simultânea transmissor-receptor seja necessária, as antenas de transmissão e de recepção e o meio podem ser representados por uma SuR de fundo linear.

Utilizando a avançada técnica de decomposição de circuito multi-níveis originalmente desenvolvida neste trabalho, já está sendo implementado o método de TE. A validação númerica desta implementação conduzir-se-á utilizando os resultados da análise do BH multi-tons empregando a TFMT. O método de TE pertence a família de métodos para a solução de EDPM [81] e representa um método mais eficiente que o do BH, na determinação do regime permanente de um circuito envolvendo portadoras de RF moduladas digitalmente. Após o desenvolvimento e implementação do método de TE, pretendemos desenvolver uma metodologia de otimização para linearização de amplificadores e de conversores de frequência, excitados por sinais BB gerados por complexos esquemas de modulação digital, conforme demonstrado em [208]. Também está prevista a implementação da análise de TE em circuitos autônomos e sincronizados e a análise de sistemas de RF heterogêneos, i.e., operando com um mistura de sinais periódicos de RF e de relógio digital e aperiódicos do tipo modulação de amplitude (*amplitude modulation* (AM)) e modulação de fase (*phase modulation* (PM)). Um exemplo de aplicação deste tipo de análise é a simulação de um transmissor polar para comunicação sem fio (“wireless”).

Considerando a análise eletro-térmica via BH e TE, será adotada, como base, a metodologia proposta em [75]. Para o cálculo da matriz de impedância térmica e que representa o sistema térmico, será utilizado o método da SDF no espaço-s da transformada de Laplace [76]. Na verdade, utilizando este método, na condição estática, $s = 0$, já foi desenvolvido um programa de

computador que efetua o cálculo da matriz de resistência térmica em dispositivos semicondutores do tipo FET e HBT. No FET os pontos quentes estão associados com os terminais de porta, enquanto no HBT, os pontos quentes estão associados aos terminais de emissor. Este tipo de análise é muito importante no projeto de amplificadores de alta-potência utilizando dispositivos semicondutores com múltiplas portas (FET) ou emissores (HBT).

A formulação multi-níveis, apresentada neste trabalho, abre caminho para uma nova linha de pesquisa, explorando os recursos de processamento paralelo. Visando este tipo de desenvolvimento, a implementação proposta já disponibiliza arquivos de dados, com todas as informações do problema, em uma estrutura que pode ser facilmente utilizada em sistemas de processamento distribuído. Com relação ao solucionador não-linear, podem ser utilizados os métodos desenvolvidos em [41]-[43],[209],[210]. Estes métodos consistem de generalizações do método de Newton para a solução de sistemas não-lineares do tipo bloco diagonal com borda dupla e foram aplicados apenas em problemas com dois níveis de hierarquia. A nossa intenção é desenvolver e avaliar estes métodos em problemas com descrição multi-níveis. Em adição, é preciso avaliar a possibilidade de adaptação do método do tensor para este tipo de solução. Existem também possibilidades de pesquisa para avaliar a potencialidade do método de Newton inexato e do tensor inexato, neste tipo de solução multi-processamentos.

A integração do nosso simulador de circuito, CDSYS - *Circuit Design System*, com o sistema de simulação de estruturas eletromagnéticas tri-dimensionais de geometria complexa, EDSYS - *Electromagnetic Design System,* também esta prevista utilizando a teoria e as implementações apresentadas em [113]. Com esta integração, vamos na direção de um ambiente de software para o modelamento global de sistemas de telecomunicações. Atualmente, o software desenvolvido para análise de campos EDSYS opera no domínio do tempo, e está fundamentado nas técnicas da MLT e de DF [113]. A implementação destes métodos no domínio da frequência pode representar um trabalho interessante de pesquisa. Lembremos que, até a presente data, estes métodos são bem mais explorados no domínio do tempo.

Finalmente, seria interessante desenvolver uma interface gráfica, utilizando a biblioteca VISLIB - *Visual Library* [113], para definição do circuito (ou sistema) a ser simulado via um editor de diagrama esquemático. A biblioteca VISLIB oferece recursos de construção automática para desenvolvimento de aplicações no sistema X Windows e combina as bibliotecas Xt Intrinsics e Motif. Para o sistema EDSYS, uma interface gráfica já foi desenvolvida para edição e simulação de estruturas EM. Os resultados podem ser exibidos utilizando o pacote gráfico AGPACK - *Advanced Grafics Package* [113], implementado utilizando a VISLIB.

Referências

[1] C. W. GEAR, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Inc., Englewood Cliffs, 1971.

[2] L. O. CHUA E P. -M. LIN, *Computer-Aided Analysis of Electronic Circuits: Algorithms and Computational Techniques*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.

[3] L. W. NAGEL E D. O. PEDERSON, *SPICE (Simulation Program with Integrated Circuit Emphasis)*, Electronics Research Laboratory, University of California, Memorandum ERL-M382, Abr. 1973.

[4] S. SKELBOE, "Computation of the periodic steady-state response of nonlinear networks by extrapolation methods," *IEEE Trans. Circuits Syst.*, vol. 27, no. 3, pg. 161-175, Mar. 1980.

[5] M. I. SOBHY, S. S. BEDAIR E M. H. KERIAKOS, "State-space approach for the analysis of networks containing lossy coupled transmission lines in inhomogeneous media," *IEE Proc.*, vol. 129, no. 3, pg. 89-95, Jun. 1982.

[6] T. J. BRAZIL, "Causal-convolution - A new method for the transient analysis of linear systems at microwave frequencies," *IEEE Trans. Microwave Theory Tech.*, vol. 43, no. 2, pg. 315-323, Fev. 1995.

[7] T. J. APRILLE JR. E T. N. TRICK, "Steady-state analysis of nonlinear circuits with periodic inputs," *Proc. IEEE*, vol. 60, no. 1, pg. 108-114, Jan. 1972.

[8] F. R. COLON E T. N. TRICK, "Fast periodic steady-state analysis for large-signal electronic circuits," *IEEE J. Solid-State Circuits*, vol. 8, no. 4, pg. 260-269, Ago. 1973.

[9] K. S. KUNDERT, J. K. WHITE, E A. SANGIOVANNI-VINCENTELLI, *Steady-State Methods for Simulating Analog and Microwave Circuits*, Boston, MA: Kluwer Academic, 1990.

[10] E. M. BAILY, *Steady State Harmonic Analysis of Nonlinear Networks*, Ph.D. Thesis, Stanford University, 1968.

[11] S. EGAMI, "Nonlinear, linear analysis and computer-aided design of resistive mixers," *IEEE Trans. Microwave Theory Tech.*, vol. 22, no. 3, pg. 270-275.

[12] A. R. KERR, "A technique for determining the local oscillator waveforms in a microwave mixer," *IEEE Trans. Microwave Theory Tech.*, vol. 23, no. 10, pg. 828-831, Out. 1975.

[13] M. S. NAKHLA E J. VLACH, "A piecewise harmonic balance technique for determination of periodic response of nonlinear systems," *IEEE Trans. Circuits Syst.*, vol. 23, no. 2, pg. 85-91, Fev. 1976.

[14] J. J. MORÉ, "Nonlinear generalizations of matrix diagonal dominanc with application to Gauss-Seidel iterations," *SIAM J. Numer. Anal.*, vol. 9, no. 2, pg. 357-378, 1972.

[15] R. G. HICKS E P. J. KHAN, "Numerical analysis of nonlinear solid-state device excitation in microwave circuits," *IEEE Trans. Microwave Theory Tech.*, vol. 30, no. 2, pg. 251-259, Mar. 1981.

[16] C. CAMACHO-PEÑALOSA, "Numerical steady-state analysis of nonlinear microwave circuits with periodic excitation," *IEEE Trans. Microwave Theory Tech.*, vol. 31, no. 9, pg. 724-730, Set. 1983.

[17] J. M. ORTEGA E W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in SeveralVariables*, London: Academic Press, Inc., 1970.

[18] V. RIZZOLI *ET AL.*, "State-of-the-art harmonic-balance simulation of forced nonlinear microwave circuits by the piecewise technique," *IEEE Trans. Microwave Theory Tech.*, vol. 40, no. 1, pg. 12-28, Jan. 1992.

[19] D. D'AMORE, P. MAFFEZZONI, E M. PILLAN, "A Newton-Powell modification algorithm for harmonic balance-based circuit analysis," *IEEE Trans Circuits Syst.-I: Fund. Theory Appl.*, vol. 41, no. 2, pg. 177-180, Fev. 1994.

[20] P. LI E L. PILEGGI, "A linear-centric modeling approach to harmonic balance analysis," *Proc. IEEE/ACM Design, Automation and Test Eur. Conf.*, pg. 634-639, 2002.

[21] R. S. DEMBO, S. C. EISENSTAT E T. STEIHAUG, "Inexact Newton methods," *SIAM J. Numer. Anal.*, vol. 19, no. 2, pg. 400-408, Ago. 1982.

[22] P. FELDMANN, R. MELVILLE, E D. LONG, "Efficient frequency domain analysis of large nonlinear analog circuits," *Proc. IEEE Custom Integrated Circuits Conf.*, pg. 241-244, 1995.

[23] H. G. BRACHTENDORF, G. WELSCH, E R. LAUR, "Fast simulation of the steady-state of circuits by the harmonic balance technique," *Proc. ISCAS Conf.*, pg. 1388-1391, 1995.

[24] R. MELVILLE, P. FELDMANN, E J. ROYCHOWDHURY, "Efficient multi-tone distortion analysis of analog integrated circuits," *Proc. IEEE Custom Integrated Circuits Conf.*, pg. 241-244, 1995.

[25] R. TELICHEVESKY, K. S. KUNDERT, E J. WHITE, "Efficient steady-state analysis based on matrix-free Krylov-subspace methods," *Proc. 32rd ACM/IEEE Design Automation Conf.*, pg. 437-444, 1995.

[26] V. RIZZOLI, F. MASTRI, C. CECCHETTI, E F. SGALLARI, "Fast and robust inexact Newton approach to the harmonic-balance analysis of nonlinear microwave circuits," *IEEE Microwave Guided Lett.*, vol. 7, no. 10, pg. 359-361, Out. 1997.

[27] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, 1996.

[28] C. HO, A. E. RUEHLI, E P. A. BRENNAN, "The modified nodal approach to network analysis", *IEEE Trans. Circuits Syst.*, vol. 22, no. 6, pg. 504-509, Jun. 1975.

[29] R. B. SCHNABEL E P. D. FRANK, "Tensor methods for nonlinear equations," *SIAM J. Numer. Anal.*, vol. 21, no.

5, pg. 815-843, Out. 1984.
[30] A. BOUARICHA E R. B. SCHNABEL, *Tensor Methods for Large, Sparse Systems of Nonlinear Equations*, Tech. Report MCS-P473-1094, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, USA, 1994.
[31] B. W. BADER, *A Tensor-Krylov Method for Solving Large-Scale Systems of Nonlinear Equations*, PhD thesis, Department of Computer Science, University of Colorado, Boulder, CO, Ago. 2003.
[32] A. BOUARICHA, *Tensor-Krylov Methods for Large Nonlinear Equations*, Tech. Report MCS-P482-1194, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, USA, 1994.
[33] A. BOUARICHA, "Algorithm 768: TENSOLVE: A software package for solving systems of nonlinear equations and nonlinear least-squares problems using tensor methods," *ACM Trans. Math. Softw.*, vol. 23, no. 2, pg. 174-195, Jun. 1997.
[34] B. W. BADER, "Tensor-Krylov methods for solving large-scale systems of nonlinear equations," Tech. Report SAND2004-1837, Computational Science Department, Sandia National Laboratories, Albuquerque, NW, USA, Maio 2004.
[35] B. W. BADER E R. B. SCHNABEL, "On the performance of tensor methods for solving ill-conditioned problems," Tech. Report SAND2004-1944, Computational Science Department, Sandia National Laboratories, Albuquerque, NW, USA, Set. 2004.
[36] M. B. STEER, "Simulation of nonlinear microwave circuits - an historical perspective and comparisons," *IEEE MTT-S Int. Microwave Symp. Dig.*, pg. 599-602, 1991.
[37] R. J. GILMORE E M. B. STEER, "Nonlinear circuit analysis using the method of harmonic balance - A review of the art. I. Introductory concepts," *Int. J. Microwave Millimeter-Wave CAE*, vol. 1, pg. 22-37, 1991.
[38] R. J. GILMORE E M. B. STEER, "Nonlinear circuit analysis using the method of harmonic balance - A review of the art. II. Advanced concepts," *Int. J. Microwave Millimeter-Wave CAE*, vol. 1, pg. 159-180, 1991.
[39] V. RIZZOLI E A. NERI, "State-of-the-art and present trends in nonlinear CAD techniques," *IEEE Trans. Microwave Theory Tech.*, vol. 36, no. 2, pg. 343-365, Fev. 1988.
[40] V. RIZZOLI, A. LIPPARINI, E E. MARAZZI, "A general-purpose program for nonlinear microwave circuit design," *IEEE Trans. Microwave Theory Tech.*, vol. 31, no. 9, pg. 762-770, Set. 1983.
[41] X. ZHANG, "Parallel computation for the solution of block bordered nonlinear equations and their applications," Ph.D. dissertation, Univ. Colorado, Boulder, Jul. 1989.
[42] X. ZHANG, "Dynamic and static load balancing for solving nonlinear block bordered circuit equations on multiprocessors," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 11, no. 9, pg. 1086-1094, Set. 1992.
[43] X. ZHANG, R. BYRD, E R. B. SCHNABEL, "Parallel methods for solving nonlinear block bordered systems of equations," *SIAM J. Sci. Stat. Comput.*, vol. 13, no. 4, pg. 841-859, Jul. 1992.
[44] V. RIZZOLI, A. LIPPARINI, D. MASOTTI, E F. MASTRI, "Efficient circuit-level analysis of large microwave systems by Krylov-subspace harmonic balance," *IEEE MTT-S Int. Microwave Symp. Dig.*, pg. 25-28, 2001.
[45] K. S. KUNDERT E A. SANGIOVANNI-VINCENTELLI, "Simulation of nonlinear circuits in the frequency domain," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 5, no. 4, pg. 521-535, Out. 1986.
[46] J. E. DENNIS, JR. E R. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, New Jersey: Printice-Hall, Inc., 1983.
[47] F. FILICORI, V. A. MONACO, E C. NALDI, "Simulation and design of microwave class-C amplifiers through harmonic analysis," *IEEE Trans. Microwave Theory Tech.*, vol. 27, no. 12, pg. 1043-1051, Dez. 1979.
[48] M. CELIK, A. ATALAR, E M. A. TAN, "A new method for the steady-state analysis of periodically excited nonlinear circuits," *IEEE Trans. Circuits Syst.-I: Fund. Theory Appl.*, vol. 43, no. 12, pg. 964-972, Dez. 1996.
[49] F. FILICORI, V. A. MONACO, E G. VANINI, "Computationally efficient intermodulation analysis of GaAs MESFET amplifiers and mixers," *Proc. 3rd. International Workshop on GaAs Telecom.*, pg. 117-129, 1991.
[50] E. GAD, R. KHAZAKA, M. S. NAKHLA, E R. GRIFFITH, "A circuit reduction technique for finding the steady-state solution of nonlinear circuits," *IEEE Trans. Microwave Theory Tech.*, vol. 48, no. 12, pg. 2389-2396, Dez. 2000.
[51] H. G. BRACHTENDORF, G. WELSCH, E R. LAUR, "A simulation tool for the analysis and verification of the steady state of circuit designs," *Int J. Circuit Theory Appl.*, pg. 311-323, 1995.
[52] Y. SAAD E M. H. SCHULTZ, "GMRES: A generalised minimum residual algorithm for solving non-symmetric linear systems," *SIAM J. Sci. Statist. Comput.*, vol. 7, no. 3, pg. 856-869, July 1986.
[53] P. FELDMANN E J. ROYCHOWDHURY, "Computation of circuit waveform envelopes using an efficient, matrix-decomposed harmonic balance algorithm," *Proc. ICCAD 96*, pg. 295-300, 1996.
[54] R. TELICHEVESKY, K. S. KUNDERT, I. ELFADEL E J. WHITE, "Fast simulation algorithms for RF circuits," *Proc. IEEE Custom Integrated Circuits Conf.*, pg. 437-444, 1996.
[55] V. RIZZOLI, C. CECCHETTI, E A. LIPPARINI, "A general-purpose program for the analysis of nonlinear microwave circuits under multitone excitation by multidimensional Fourier transform," *Proc. 17th European Microwave Conf.*, pg. 635-640, 1987.

[56] L. O. CHUA, E A. USHIDA, "Algorithms for computing almost periodic steady-state response of nonlinear systems to multiple input frequencies," *IEEE Trans. Circuits Syst.*, vol. 28, no. 10, pg. 953-971, Out. 1981.
[57] A. USHIDA E L. O. CHUA, "Frequency-domain analysis of nonlinear circuits driven by multi-tone signals," *IEEE Trans. Circuits Syst.*, vol. 31, no. 9, pg. 766-779, Set. 1984.
[58] J. KUNISCH E I. WOLFF, "Determination of sampling points for nearly DFT-equivalent almost-periodic Fourier transform," *Proc. 23rd European Microwave Conf.*, pg. 710-713, 1993.
[59] K. S. KUNDERT, G. B. SORKIN E A. SANGIOVANNI-VINCENTELLI, "Applying harmonic balance to almost-periodic circuits," *IEEE Trans. Microwave Theory Tech.*, vol. 36, no. 2, pg. 366-378, Fev. 1988.
[60] P. J. C. RODRIGUES, "An orthogonal almost-periodic fourier transform for use in nonlinear circuit simulation," *IEEE Microwave Guided Wave Lett.*, vol. 4, no. 3, pg. 74-76, Mar. 1994.
[61] R. GILMORE, "Nonlinear circuit design using modified harmonic balance algorithm," *IEEE Trans. Microwave Theory Tech.*, vol. 34, no. 12, pg. 1294-1307, Dez. 1986.
[62] D. HENTE E R. H. JANSEN, "Frequency domain continuation method for the analysis and stabitlity investigation of nonlinear microwave circuits," *Proc. Inst. Elec. Eng.*, vol. 133, pt. H, pg. 351-362, Out. 1986.
[63] P. J. C. RODRIGUES, "A general mapping technique for fourier transform computation in nonlinear circuit analysis," *IEEE Microwave Guided Wave Lett.*, vol. 7, no. 11, pg. 374-376, Nov. 1997.
[64] E. O. BRIGHAM, *The Fast Fourier Transform FFT and Its Applications*, Prentice-Hall Signal Processing Series, Englewood Cliffs, New Jersey, 1988.
[65] V. BORICH, J. EAST, E G. HADDAD, "An efficient Fourier transform algorithm for multitone harmonic balance," *IEEE Trans. Microwave Theory Tech.*, pg. 117-129, 1991.
[66] B. RAZAVI, *RF Microelectronics*, Prentice-Hall, Inc., 1998.
[67] A. BRAMBILLA E D. D'AMORE, "A filter-based technique for the harmonic balance method," *IEEE Trans. Circuit Syst. - I: Fund. Theory Appl.*, vol. 43, no. 2, pg. 92-98, Fev. 1996.
[68] O. J. NASTOV E J. K. WHITE, "Time-mapped harmonic balance," *Proc. IEEE Data Automation Circuits Conf.*, pg. 641-646, Jun. 1999.
[69] O. J. NASTOV E J. K. WHITE, "Grid selection strategies for the time-mapped harmonic balance simulation of circuits with rapid transitions," *Proc. IEEE Custom Integrated Circuits Conf.*, pg. 13-16, Maio 1999.
[70] V. RIZZOLI *ET AL.*, "Prediction of the end-to-end performance of a microwave/RF linkby means of nonlinear/ electromagnetic co-simulation,"*IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 12, pg. 4149-4159, Dez. 2006.
[71] J. KUNISCH E I. WOLFF, "Steady-state analysis of nonlinear forced and autonomous microwave circuits using the compression approach", *Int. J. Microwave Millimeter-Wave Computer-Aided Eng.*, vol. 5, no. 4, pg. 241-255, 1995.
[72] Y. DUROC, T. -P. VUONG, E S. TEDJINI, "A time/frequency model of ultrawideband antennas," *IEEE Trans. Antennas Propag.*, vol. 55, no. 8, pg. 2342-2350, Ago. 2007.
[73] A. SHARAIHA E J. LE BIHAN, "Harmonic balance analysis of IM harmonic distortion in semiconductor lasers," *Microwave Optical Tech. Lett.*, vol. 14, no. 6, pg. 327-330, Abr. 1997.
[74] T. VEIJOLA E T. MATTILA, "Modeling of nonlinear micromechanical resonators and their simulation with the harmonic-balance method," *Int. J. RF and Microwave CAE*, vol. 11, pg. 310-321, 2001.
[75] V. RIZZOLI *ET AL.*, "Simultaneous thermal and electrical analysis of nonlinear microwave circuits," *IEEE Trans. Microwave Theory Tech..*, vol. 40, no. 7, pg. 1446-1455, Jul. 1992.
[76] W. BATTY *ET AL.*, "Electrothermal CAD of power devices and circuits with fully physical time-dependent compact thermal modeling of complex nonlinear 3-D systems," *IEEE Trans. Components Packaging Tech.*, vol. 24, no. 4, pg. 566-590, Dez. 2001.
[77] P. LI, L. T. PILEGGI, M. ASHEGHI, E R. CHANDRA, "IC thermal simulation and modeling via efficient multigrid-based approaches," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 25, no. 9, pg. 1763-1776, Set. 2006.
[78] D. SHARRIT, "New circuit simulation analysis methods for communication system," *IEEE MTT-S Work-Shop WMFA*, pg. 29-41, Jun. 1996.
[79] E. NGOYA E R. LARCHEVÈQUE, "Envelop transient analysis: A new method for the transient and steady-state analysis of microwave communication circuits and systems," *IEEE MTT-S Int. Microwave Symp. Dig.*, pg. 1365-1368, 1996.
[80] V. RIZZOLI, A. NERI, F. MASTRI, E A. LIPPARINI, "A Krylov-subspace technique for the simulation of integrated RF/microwave subsystems driven by digitally modulated carriers," *Int. J. RF and Microwave CAE*, vol. 9, pg. 490-505, 1999.
[81] J. ROYCHOWDHURY, "Analyzing circuits with widely-separated time scales using numerical PDE methods," *IEEE Trans. Circuits. Syst.*, vol. 48, no. 5, pg. 578–594, Maio 2001.
[82] N. B. CARVALHO, J. C. PEDRO, W. JANG E M. B. STEER, "Nonlinear RF circuits and systems simulation when driven by several modulated signals," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 2, pg. 572-579, Fev.

2006.
[83] V. RIZZOLI, F. MASTRI, A. NERI, E A. LIPPARINI, "Analysis of electrothermal transients and digital signal processing in electrically and thermally nonlinear microwave circuits," *IEEE MTT-S Int. Microwave Symp. Dig.*, pg. 367-370, 1997.
[84] J. J. BUSSGANG, L. EHRMAN, E J. W. GRAHAM, "Analysis of nonlinear systems with multiple inputs," *Proc. IEEE*, vol. 62, no. 8, pg. 1088-1119, Ago. 1974.
[85] V. VOLTERRA, *Theory of Functionals and of Integral and Integro-Differential Equations*, New York: Dover, 1959.
[86] N. WIENER, *Nonlinear Problems in Random Theory*, New York: Technology Press, 1958.
[87] N. NØRHOLM, C. IVERSEN, AND T. LARSEN, "GaAs MESFET large-signal modelling for multiport Volterra series analysis," *IEE Proc. -Circuits Devices Syst.*, vol. 144, no. 1, pg. 40-44, Fev. 1997.
[88] R. A. MINASIAN, "Intermodulation distortion analysis of MESFET amplifiers using the Volterra series representation,"*IEEE Trans. Microwave Theory Tech.*, vol. 28, no. 1, pg. 1-8, Jan. 1980.
[89] J. C. PEDRO E N. B. CARVALHO, "On the use of multitone techniques for assessing RF components' intermodulation distortion," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 12, pg. 2393-2402, Dez. 1999.
[90] S. PENG, P. J. MCCLEER, E G. I. HADDAD, "Nonlinear models for the intermodulation analysis of FET mixers," *IEEE Trans. Microwave Theory Tech.*, pg. 1037-1044, Maio 1995.
[91] J. A. GARCIA *ET AL.*, "Time-varying Volterra-series analysis of spectral regrowth and noise power ration in FET mixers," *IEEE Trans. Microwave Theory Tech.*, vol. 49, no. 3, pg. 545-549, Mar. 2001.
[92] E. VAN DEN EIJNDE E J. SCHOUKENS, "Steady-state analysis of a periodically excited nonlinear systems," *IEEE Trans. Circuits Syst.*, vol. 37, no. 2, pg. 232-242, Fev. 1990.
[93] A. USHIDA, L. O. CHUA, E T. SUGAWARA, "A substitution algorithm for solving nonlinear circuits with multi-frequency components," *Int J. Circuit Theory Appl.*, vol. 15, pg. 327-355, 1987.
[94] A. USHIDA E L. O. CHUA, "Steady-state response of non-linear circuits: A frequency-domain relaxation method," *Int J. Circuit Theory Appl.*, vol. 17, pg. 249-269, 1989.
[95] Y. YAMAGAMI, Y. HISHIO, A. USHIDA, M. TAKAHASHI, E K. OGAWA, "Analysis of communication circuits based on multidimensional Fourier transform," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 18, no. 8, pg. 1165-1177, Ago. 1999.
[96] V. D. HWANG, Y. SHIH, H. M. LE E T. ITOH, "Nonlinear modeling and verification of MMIC amplifiers using the waveform-balance method," *IEEE Trans. Microwave Theory Tech.*, vol. 37, no. 12, pg. 2125-2132, Dec. 1989.
[97] J. DREIFUSS, A. MADJAR E A. BAR-LEV, "An improved version of the almost periodic fourier transform algorithm with applications in the large-signal domain," *IEEE Trans. Microwave Theory Tech.,* vol. 39, no. 3*,* pg. 571-575, Mar. 1991.
[98] P. J. C. RODRIGUES, M. J. HOWES E J. R. RICHARDSON, "Efficient computation of the steady-state response of periodic nonlinear microwave circuits using convolution-based sample-balance technique," *IEEE Trans. Microwave Theory Tech.*, vol. 39, no. 4, pg. 732-737, Abr. 1991.
[99] J. H. HAYWOOD E Y. L. CHOW, "Intermodulation distortion analysis using a frequency-domain harmonic balance technique," *IEEE Trans. Microwave Theory Tech.*, vol. 36, no. 8, pg. 1251-1257, Ago. 1988.
[100]D. RESCA *ET AL.*, "Scalable nonlinear FET model based on a distributed parasitic network description," *IEEE Trans. Microwave Theory Tech.*, vol. 56, no. 4, pg. 755-766, Apr. 2008.
[101]C. -R. CHANG, M. B. STEER, E G. W. RHYNE, "Frequency-domain spectral balance using the arithmetic operator method," *IEEE Trans. Microwave Theory Tech.*, vol. 37, no. 11, pg. 1681-1688, Nov. 1989.
[102]C. -R. CHANG E M. B. STEER, "Frequency-domain nonlinear microwave circuit simulation using the arithmetic operator method," *IEEE Trans. Microwave Theory Tech.*, vol. 38, no. 8, pg. 1139-1143, Ago. 1990.
[103]T. NÄRHI, "Frequency domain analysis of strongly nonlinear circuits using a consistent large-signal model," *IEEE Trans. Microwave Theory Tech.*, vol. 44, no. 2, pg. 182-192, Fev. 1996.
[104]N. B. CARVALHO E J. C. PEDRO, "Multitone frequency-domain simulation of nonlinear circuits in large- and small-signal regimes," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 12, pg. 2016-2024, Dez. 1998.
[105]K. S. KUNDERT, "Introduction to RF simulation and its application," *IEEE J. Solid-State Circuits*, vol. 34, no. 9, pg. 1298-1319, Set. 1999.
[106]J. C. PEDRO E N. B. CARVALHO, "An integrated overview of CAD/CAE tools and their use on wireless-communication circuit design," *Int. J. RF Microwave CAE*, vol. 14, pg. 507-524, 2004.
[107]B. STROUSTRUP, *The C++ Programming Language*, Addison-Wesley Publishing Company, 1991.
[108]G. KRON, *Diakoptics: Piecewise Solution of Large-Scale Systems*, London, UK: MacDonald, 1963.
[109]K. BOWDEN, "Kron's method of tearing on a transputer array," *The Computer Journal*, vol. 33, no. 5, pg. 453-459, 1990.
[110]K. BOWDEN, "Hierarchical tearing: An efficient holographic algorithm for system decomposition," *Int. J. General Systems*, vol. 23, no. 1, pg. 23-37, Oct. 1994.

[111]P. SAVIZ E O. WING, "Circuit simulation by hierarchical waveform relaxation", *IEEE Trans. Comput.-Aided Integr. Circuits Syst.*, vol. 12, no. 6, pg. 845-860, Jun. 1993.
[112]M. I. SOBHY, E. A. HOSNY, E M. A. NASSEF, "Multiport approach for the analysis of microwave nonlinear networks," *Int. J. Num. Modelling: Electronic Networks, Devices and Fields*, vol. 6, pg. 67-81, 1993.
[113]O. P. PAIXÃO, in Contributions to "Millimetre-wave integrated circuits at 94 and 140 GHz," Tech. Rep., EPSRC, Research Grant Number ???, England, July 1997.
[114]B. W. BADER, R. PAWLOWSKI E T. G. KOLDA, *Robust Large-Scale Parallel Nonlinear Solvers for Simulations*, Tech. Report SAND2005-6864, Computational Science Department, Sandia National Laboratories, Albuquerque, NW, USA, Nov. 2005.
[115]K. FINKENZELLER, *RFID Handbook*, 2da ed., New York: Wiley, 2003.
[116]L. CHUA E L. -K. CHEN, "Diakoptic and generalized hybrid analysis", *IEEE Trans. Circuits and Syst.*, vol. 23, no. 12, pg. 694-705, Dez. 1976.
[117]A. SANGIOVANNI-VINCENTELLI L. -K. CHEN, E L. CHUA, "A new tearing approach - The node-tearing nodal analysis," *Proc. IEEE Int. Symp. Circuit and Systems*, pg. 143-148, 1977.
[118]*IBID*, "An efficient heuristic cluster algorithm for tearing large-scale networks," *IEEE Trans. Circuits and Systems*, vol. 24, no. 12, pg. 709-717, Dez. 1977.
[119]F. F. WU, "Solution of large-scale networks by tearing", *IEEE Trans. Circuits and Systems*, vol. 23, no. 12, pg. 706-713, Dez. 1976.
[120]N. B. G. RABBAT, A. SANGIOVANNI-VINCENTELLI, E H. Y. HSIEH, "A multilevel Newton algorithm with macromodeling and latency for the analysis of large-scale nonlinear circuits in the time domain", *IEEE Trans. Circuits and Systems*, vol. 26, no. 9, pg. 733-741, Set. 1979.
[121]V. RIZZOLI *ET AL.*, "Computer-aided optimization of nonlinear microwave circuits with the aid of electromagnetic simulation,"*IEEE Trans. Microwave Theory Tech.*, vol. 52, no. 1, pt. 2, pg. 362-377, Jan. 2004.
[122]Y. WEI, Q. -J. ZHANG, E M. NAHKLA, "Multilevel optimization of high speed VLSI interconnect networks by decomposition", *IEEE Trans. Microwave Theory Tech.*, vol. 42, no. 9, pg. 1638-1650, Set. 1994.
[123]A. SALAMA, A. E. STARZYK E J. W. BANDLER, "A unified decomposition apporach for fault location in large analogue circuit," *IEEE Trans. Circuits and Syst.*, vol. 31, no. 7, pg. 609-622, Jul. 1984.
[124]H. -T. SHEU E Y. -H. TANG, "Hierarchical frequency-domain robust component failure detection scheme for large-scale analogue circuits with component tolerances," *IEEE Proc. -Circuits Devices Syst.*, vol. 143, no. 1, pg. 53-60, Fev. 1996.
[125]G. GOUBAU, N. N. PURI, E F. SCHWERING, "Diakoptic theory for multielement antennas", *IEEE Trans. Antennas Propagat.*, vol. 30, pg. 15-26, Jan. 1982.
[126]F. SCHWERING, N. N. PURI, E C. M. BUTLER, "Modified diakoptic theory of antennas," *IEEE Trans. Antennas Propagat.*, vol. 34, pg. 1273-1281, Nov. 1986.
[127]G. E. HOWARD E Y. L. CHOW, "Diakoptic theory for the microstrip structures", *IEEE AP-S Int. Symp. Dig.*, pg. 1079-1082, 1990.
[128]S. OOMS E D. DE ZUTTER, "A new iteration diakoptics-based multilevel moments method for planar circuits," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 3, pg. 280-291, Mar. 1998.
[129]W. GEYL, "Numerical analysis of waveguide discontinuity problems using the network model decomposition method," *IEEE Trans. Microwave Theory Tech.*, vol. 39, no. 10, pg. 1766-1770, Out. 1991.
[130]L. N. MERUGU E V. F. FUSCO, "Concurrent network diakoptics for electromagnetic field problems", *IEEE Trans. Microwave Theory Tech.*, vol. 41, no. 4, pg. 85-91, Abr. 1993.
[131]C. E. CHRISTOFFERSEN, M. OZKAR, M. B. STEER, M. G. CASE, E M. RODWELL, "State-variable based transient analysis using convolution", *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 6, pt. II, pg. 882-889, Jul. 1999.
[132]C. E. CHRISTOFFERSEN E M. B. STEER, "Implementation of the local reference node conceptfor spatially distributed circuits," *Int. J. RF Microwave CAE*, vol. 9, pg. 376–384, 1999.
[133]B. FLOYD *ET AL.*, "SiGe bipolar transceiver circuits operating at 60 GHz," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pg. 156-167, Jan. 2005.
[134]J. VLACH E K. SINGHAL, *Computer Methods for Circuit Analysis and Design*, Van Nostrand Reinhold, New York, New York, 1983.
[135]A. SANTARELLI *ET AL.*, "A nonquasi-static empirical model of electron devices," *EEE Trans. Microwave Theory Tech.*, vol. 54, no. 12, pg. 4021-4031, Dez. 2006.
[136]C. RAUSCHER E H. A. WILLING, "Simulation of nonlinear microwave FET performance using a quasi-static model," *IEEE Trans. Microwave Theory Tech.*, vol. 27, no. 10, pg. 834-840, Out. 1979.
[137]Y. HU E K. MAYARAM, "Comparison of algorithms for frequency domain coupled device and circuit simulation", *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 12, no. 11, pg. 1726-1733, Nov. 1993.
[138]R. SOMMET E E. NGOYA, "Full implementation of an implicit nonlinear model with memory in an harmonic balance software," *IEEE Microwave Guided Wave Lett.*, pg. 153-155, vol. 7, no. 6, Jun. 1997.
[139]R. SINGH E C. M. SNOWEDEN, "Small-signal characterization of microwave and millimeter-wave HEMT's based

on a physical model," IEEE Trans. Microwave Theory Tech., vol. 44, no. 1, pg. 114-121, Jan. 1996.
[140]I. WOLFF, "Finite difference time-domain simulation of electromagnetic fields and microwave circuits," *Int. J. Num. Modelling: Electronic Networks, Devices and Fields*, vol. 5, pg. 163-182, 1992.
[141]J. JIN, *The Finite Element Method in Electromagnetics*, John Wiley & Sons, Inc., 1993.
[142]A. TAFLOVE, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, Artech House, Inc., 1995.
[143]T. WEILAND, "Time-domain electromagnetic field computation with finite-difference methods," *Int. J. Num. Modelling: Electronic Networks, Devices and Fields*, vol. 9, pg. 295-319, 1996.
[144]C. Christopoulos, *The Transmission-Line Modeling Method: TLM*, IEEE Press/Oxford University Press, 1996.
[145]C. E. CHRISTOFFERSEN, M. B. STEER, E M. A. SUMMERS, "Harmonic balance analysis for systems with circuit-field iterations," *IEEE MTT-S Int. Microwave Symp. Dig.*, pg. 1131-1134, Jun. 1998.
[146]I. S. DUFF, A. M. ERISMAN E J. K. REID, *Direct Methods for Sparse Matrices*, Oxford University Press, 1986.
[147]A. GEORGE E J. W. H. LIU, *Computer Solution of Large Sparse Positive Systems*, Prentice-Hall, Inc., 1981.
[148]S. PISSANETSKY, *Sparse Matrix Technology*, Academic Press Inc., 1984.
[149]V. K. TRIPATHI E J. B. RETTIG, "A SPICE model for multiple coupled microstrips and other transmission lines," *IEEE Trans. Microwave Theory Tech.*, vol. 33, no. 12, pg. 1513-1518, Dez. 1985.
[150]I. ANGELOV *ET AL.*, "An empirical table-based FET model," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 12, pg. 2350-2357, Dez. 1999.
[151]C. -J. WEI, Y. A. TKACHENKO, E D. BARTLE, "An accurate large-signal model of GaAs MOSFET which accounts for charge conservation, dispersion and self-heating," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 11, pg. 1650-1644, Nov. 1998.
[152]D. M. HIMMELBLAU, *Decomposition of Large-Scale Problems*, Amsterdam: North-Holland, 1973.
[153]H. JOKINEN E M. VALTONEN, "Small-signal harmonic analysis of non-linear circuits," *Int J. Circuit Theory Appl.*, vol. 23, pg. 325-343, 1995.
[154]K. A. REMLEY, "Multisine excitation for ACPR measurements," *IEEE MTT-S Int. Microwave Symp. Dig.*, pg. 2141-2144, 2003.
[155]J. C. PEDRO E N. B. CARVALHO, "Designing multisine excitations for nonlinear model testing," *IEEE Trans. Microwave Theory Tech.*, vol. 53, no. 1, pg. 45-54, Jan. 2005.
[156]W. H. PRESS, B. P. FLANNERY, S. A. TEUKOWLSKI, E W. T. VETTERLING, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 1988.
[157]G. H. GOLUB E C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1983. (The John Hopkins Press Ltd., London, 3rd ed., 1996.)
[158]E. NGOYA *ET AL.*, "Efficient algorithms for spectra calculations in nonlinear microwave circuits simulators," *IEEE Trans. Circuits Syst.*, vol. 37, no. 11, pg. 1339-1355, Dez. 1989.
[159]Z. XIANG-DONG, H. XING-NAN, E G. BAO-XIN, "Accurate fourier transform method for almost-periodic response simulation of microwave nonlinear circuits," *Electron. Lett.*, vol. 25, no. 6, pg. 404-406, Mar. 1989.
[160]M. M. GOURARY *ET AL.*, "Adaptive preconditioners for the simulation of extremely non-linear circuits using harmonic balance," *IEEE MTT-S Int. Microwave Symp. Dig.*, pg. 779-782, 1999.
[161]B. W. BADER E R. B. SCHNABEL, "Curvilinear linesearch for tensor methods," *SIAM J. Sci. Comput.*, vol. 25, no. 2, pg. 604-622, 2003.
[162]G. W. RHYNE E M. B. STEER, comentários (e réplica) sobre o "Simulation of nonlinear circuits in the frequency domain " de K. S. Kundert e A. Sangiovanni-Vincentelli, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 8, no. 8, pg. 927-929, Ago. 1989.
[163]E. NGOYA, A. SUÁREZ, R. SOMMET E R. QUÉRÉ, "Steady-state analysis of free or forced oscillators by harmonic balance and stability investigation of periodic and quasi-periodic regimes," *Int. J. Microwave Millimetre-Wave CAE*, vol. 5, pg. 210-223, 1995.
[164]H. R. YEAGER E R. W. DUTTON, "Improvement in norm-reducing Newton methods for circuit simulation," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 8, No. 5, pg. 538-546, Maio 1989.
[165]P. DEUFLHARD E G. HEINDL, "Affine invariant convergence theorems for Newton's method and extensions to related methods," *SIAM J. Numer. Anal.*, vol. 16, no. 1, pg. 1-10, Fev. 1979.
[166]P. DEUFLHARD, "A modified Newton method for the solution of ill-conditioned systems of nonlinear equations with application to multiple shooting," *Numer. Math.*, 22, pg. 289-315, 1974.
[167]P. DEUFLHARD, "A relaxation strategy for the modified Newton Method," in Bulirsch/Oettli/Stoer: *Optimazation and Optimal Control*, *Springer Lecture Notes*, pg. 59-73, 1975.
[168]U. NOWAK E L. WEIMANN, "GIANT - A software package for the numerical solution of very large systems of highly nonlinear equations," *Technical Report TR-90-11*, *Konrad-Zuse-Zentrum für Informationstechnik Berlin: Preprint*, Dez. 1990.
[169]R. E. BANK E D. J. ROSE, "Global approximate Newton methods," *Numer. Math.*, 37, pg. 279-295, 1981.
[170]C. G. BROYDEN, "A class of methods for solving nonlinear simultaneous equations," *Math. Comput.*, vol. 19, pg.

577-593, 1965.
[171]M. J. D. POWELL, "A hybrid method for nonlinear equations," in *Numerical Methods for Nonlinear Algebraic Equations*, edited by P. RABINOWITZ, Gordon-Breach, London, 1970.
[172]L. GRIPPO, F. LAMPARIELLO, E S. LUCIDI, "A nonmotone line search technique for Newton's method", *SIAM J. Numer. Anal.*, vol. 23, no. 4, pg. 707-716, Ago. 1986.
[173]M. HONKALA, *Nonmonotone norm-reduction method in numerical circuit analysis*, Circuit Theory Laboratory Report Series, No. CT-46, Espoo 2002, 11 págs. [Em rede]. Disponível: http://www.aplac.hut.fi/publications/ct-46/ct-46.pdf.
[174]D. FENG, P. D. FRANK, E R. B. SCHNABEL, "Local convergence analysis of tensor methods for nonlinear equations," *Math. Programming*, 62, pg. 427-459, Ago. 1993.
[175]P. DEUFLHARD, "A stepsize control continuation methods and its special application to multiple shooting techniques," *Numer. Math.*, vol. 33, pg. 115-146, 1979.
[176]J. J. MORÉ, B. S. GARBOW, E K. E. HILLISTROM, "User guide for MINPACK-1," Argonne National Labs Report ANL-80-74, 1980.
[177]S. C. EISENSTAT E H. F. WALKER, "Choosing the forcing terms in an inexact Newton method," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pg. 16-32, Jan. 1996.
[178]C. T. KELLEY E Z. Q. XUE, "Inexact Newton methods for singular problems," *Optimization Methods Soft.*, vol. 2, pg. 249-267, 1993.
[179]P. N. BROWN, "A local convergence theory for combined inexact-Newton/finite-difference projection method," *SIAM J. Numer. Anal.*, vol. 24, no. 2, pg. 407-434, Abr. 1987.
[180]P. N. BROWN E Y. SAAD, "Convergence theory of nonlinear Newton-Krylov algorithms," *SIAM J. Optim*., vol. 4, no. 2, pg. 297-330, Maio 1994.
[181]S. C. EISENSTAT E H. F. WALKER, "Globally convergent inexact Newton methods," *SIAM J. Optimization*, vol. 4, no. 2, pg. 393-422, Maio 1994.
[182]P. DEUFLHARD, "Global inexact methods for very large scale nonlinear problems," *Comput. Science Engineering*, 3, pg. 366-393, 1991.
[183]R. **BARRET**T *ET AL*., *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM, 1994.
[184]O. AXELSSON, *Iterative Soultion Methods*, Cambridge University Press, 1996.
[185]W. E. ARNOLDI, "The principle of minimized iterations in the solution of the matrix eigenvalue problem," *Quart. Appl. Math*., vol. 9, pg. 17-29, 1951.
[186]C. LANCZOS, "An iterative method for the solution of the eigenvalue problem of linear differential and integral operators," *J. Natl. Bur. Stand.*, vol. 45, no. 4, pg. 255-282, Out. 1950.
[187]C. LANCZOS, "Solution of systems of linear equations by minimized iterations," *J. Res. Natl. Bur. Stand.*, vol. 49, no. 1, pg. 33-53, Jul. 1952.
[188]H. F. WALKER, "Implementation of the GMRES method using Householder transformations," *SIAM J. Sci. Statist. Comput.*, vol. 6, no. 1, pg. 152-163, Jan. 1988.
[189]Y. SAAD, "A flexible inner-outer preconditioned GMRES algorithm," *SIAM J. Sci. Statist. Comput.*, vol. 14, no. 2, pg. 461-469, Mar. 1993.
[190]E. M. KASENALLY, "GMBACK: A generalised minimum backward error algorithm for nonsymmetric linear systems," *SIAM J. Sci. Statist. Comput.*, vol. 16, no. 3, pg. 698-719, Maio 1995.
[191]D. FENG E T. H. PULLIAM, "Tensor-GMRES method for large systems of nonlinear equations," *SIAM J. Optim.*, vol. 7, no. 3, pg. 757-779, Ago. 1997.
[192]A. RUHE, "Implementation aspects of band Lanczos algorithm for computation of eigenvalues of large sparse symmetric matrices," *Math. Comp.*, vol. 33, pg. 680-687, 1979.
[193]E. CHOW E Y. SAAD, "Approximate inverse preconditioners via sparse-sparse iterations," *SIAM J. Sci. Comput.*, vol. 19, no. 3, pg. 995-1023, Maio 1998.
[194]D. G. HAIGH, "GaAs MESFET active resonant circuit for microwave filter applications," *IEEE Trans. Microwave Theory Tech.*, vol. 42, no. 7, pg. 1419-1423, Jul. 1994.
[195]C. TOUMAZOU, D. G. HAIGH, E J. M. FOPMA, "High frequency gallium arsenide four-quadrant analogue multiplier," *Electron. Lett.*, vol. 26, no. 20, pg. 1650-1652, Set. 1990.
[196]C. M. BUCK, "A 2.0 W X-band MMIC power amplifier", Private Communication, Phillips Microwave, UK, Jun. 1995.
[197]K. W. KOBAYASHI, *ET AL*., "A dc-20 GHz InP HBT balanced analog multiplier for high-data-rate direct-digital modulation and fiber-optic receiver applications", *IEEE Trans. Microwave Theory Tech.*, vol. 48, no. 2, pg. 194-202, Fev. 2000.
[198]D. R. FREY E O. NORMAN, "An integral equation approach to the periodic steady-state problem in nonlinear circuits," *IEEE Trans. Circuits Syst.-I: Fund. Theory Appl.*, vol. 39, no. 9, pg. 744-755, Set. 1992.
[199]L. O. CHUA E N. N. WANG, "A new approach to overcome the overflow problem in computer-aided analysis of

nonlinear resistive circuits," *Int. J. Circuit Theory Appl.*, vol. 15, pg. 261-284, Dec. 1975.
[200]E. W. LIN E W. H. KU, "Device considerations and modeling for the design of an InP-Based MODFET milimiter-wave resistive mixer with superior conversion efficiency", *IEEE Trans. Microwave Theory Tech.*, vol. 43, no. 8, pg. 1951-1959, Ago. 1995.
[201]S. A. MAAS, *Nonlinear Microwave Circuits*, Artech House, Inc., 1998.
[202]B. GILBERT, "A precise four-quadrant multiplier with subnanosecond response", *IEEE J. Solid-State Circuits*, vol. 3, no. 4, pg. 365-373, Dez. 1968.
[203]C. -J. WEI, J. C. M. HWANG, W. -J. HO, E J. A. HIGGINS, "Large-signal modeling of self-heating, collector transit-time, and RF-breakdown effects in power HBT's," *IEEE Trans. Microwave Theory Tech.*, vol. 44, no. 12, pg. 2641-2647, Dez. 1996.
[204]K. W. KOBAYASHI, *ET AL.*, "A 44-GHz high IP3 InP-HBT amplifier with practical current reuse biasing", *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 12, pg. 2541-2552, Dez. 1998.
[205]R. B. SCHNABEL E T. -H. CHOW, "Tensor methods for unconstrained optimization using second derivatives," *SIAM J. Optim.*, vol. 1, no. 3, pg. 293-315, Ago. 1991.
[206]A. BOURICHA, "Tensor methods for large,sparse unconstrained optimization," *SIAM J. Optim.*, vol. 7, no. 3, pg. 732-756, Ago. 1997.
[207]P. PURSULA, *ET AL.*, "Millimeter-wave identification—A new short-range radio system for low-power high data-rate applications," *IEEE Trans. Microwave Theory Tech.*, vol. 56, no. 10, pg. 2221-2228, Out. 2008.
[208]V. BORICH, J. EAST, E G. HADDAD,"Gradient optimization of RF amplifiers for digital communications," *Int. J. RF Microwave CAE*, vol. 10, pg. 353-365, 2000.
[209]D. FENG E R. B. SCHNABEL, *Globally Convergent Parallel Algorithms for Solving Block. Bordered Systems of Nonlinear Equations*, Tech. Report CU-CS-633-92, Department of Computer Science, University of Colorado at Boulder, Boulder, CO, USA, Dez. 1992.
[210]M. HONKALA, V. KARANKO, E J. ROOS, "Improving the convergence of combined Newton-Raphson and Gauss-Newton multilevel iteration method", *Proc. Int. Symp. Circuits Systems, ISCAS*, pg. 229-232, 2002.

Apêndice A. Representação dos Elementos Básicos para Análise de Espaço-de-Estado

Neste apêndice, seguindo a teoria apresentada no Capítulo 3, serão apresentadas as tabelas com a equação de estado, de saída (relações constitutivas), e de sonda de cada elemento básico utilizado na construção de uma SRN. Estas tabelas estão organizadas de acordo com os grupos de variável-de-estado (1d,1i,2,2A,3A,3) introduzidos no Capítulo 3. Os índices que controlam a posição de inserção dos sub-vetores e sub-matrizes são mostrados nestas tabelas. As entradas hachuriadas nas tabelas abaixo correspondem á formulação sem equação de estado do tipo integral, i.e., ideal para análise no domínio do tempo.

Seguindo a organização adotada, o vetor de variável-de-estado descrito nas tabelas abaixo é organizado da seguinte forma

$$\boldsymbol{X} = \left[\boldsymbol{X}_{1d}\ \boldsymbol{X}_{1i}\ \boldsymbol{X}_{2}\ \boldsymbol{X}_{2A}\ \boldsymbol{X}_{3A}\ \boldsymbol{X}_{3}\right]^{T}$$

onde

- $\boldsymbol{X}_{1d}$ indutores na co-floresta e capacitores na floresta,
- $\boldsymbol{X}_{1i}$ indutores na floresta e capacitores na co-floresta,
- $\boldsymbol{X}_{2}$ linhas de transmissão, linhas de transmissão (toco) em aberto e linhas de transmissão (toco) em curto, sondas de tensão, e sondas de corrente,
- $\boldsymbol{X}_{2A}$ tensões e correntes de controle com atraso das fontes de tensão e de corrente de controle linear,
- $\boldsymbol{X}_{3A}$ tensões e correntes de controle não-linear com atraso dos elementos não-lineares, e
- $\boldsymbol{X}_{3}$ tensões e correntes de controle não-linear dos elementos não-lineares.

O vetor de função não-linear descrito nas tabelas abaixo é organizado da seguinte forma

$$\boldsymbol{u}_{f}(\boldsymbol{x}_{3}) = \left[\boldsymbol{u}_{f1}(\boldsymbol{x}_{3})\ \boldsymbol{u}_{f2}(\boldsymbol{x}_{3})\right]^{T}$$

onde

- $\boldsymbol{u}_{f1}(\boldsymbol{x}_{3})$ funções não-lineares estáticas (resistores não-lineares, fontes de tensão de controle não-linear e fontes de corrente de controle não-linear), e
- $\boldsymbol{u}_{f2}(\boldsymbol{x}_{3})$ funções não-lineares dinâmincas (indutores não-lineares e capacitores não-lineares).

As funções não-lineares estáticas e dinâmicas associadas aos resistores não-lineares, indutores não-lineares e capacitores não-lineares, em geral, possuem na sua lista de argumentos uma variável de controle local e qualquer número de variáveis de controle remota.

Tabela A.1:
REPRESENTAÇÃO DE ESPAÇO-DE-ESTADO DOS ELEMENTOS CONCENTRADOS

Elemento	Estado	Entrada	Saída	Matrizes da Equação de Estado e de Saída				
				$\underline{\mathbf{A}}_l$	$\underline{\mathbf{A}}_r$	$\underline{\mathbf{B}}$	$\underline{\mathbf{C}}$	$\underline{\mathbf{D}}$
RES	-	$U = V$	$Z = I$	-	-	-	-	G
	-	$U = I$	$Z = V$	-	-	-	-	R
CAP	$X_{1d} = V$	$U = I$	$Z = V$	C	0	1	1	0
	$X_{1i} = I$	$U = V$	$Z = I$	$1/C$	0	1	1	0
IND	$X_{1d} = I$	$U = V$	$Z = I$	L	0	1	1	0
	$X_{1i} = V$	$U = I$	$Z = V$	$1/L$	0	1	1	0

onde
$R = 1/G$ é a resistência,
L é a indutância, e
C é a capacitância.

Tabela A.2:
REPRESENTAÇÃO DE ESPAÇO-DE-ESTADO DOS ELEMENTOS DISTRIBUÍDOS

Elemento	Estado(s)	Entrada(s)	Saída(s)	Matrizes da Equação de Estado e de Saída				
				$\underline{\mathbf{A}}_l$	$\underline{\mathbf{A}}_r$	$\underline{\mathbf{B}}$	$\underline{\mathbf{C}}$	$\underline{\mathbf{D}}$
LT	$\boldsymbol{X}_2 = \begin{bmatrix} V_1^- \\ V_2^- \end{bmatrix}$	$\boldsymbol{U} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$	$\boldsymbol{Z} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} -2Y_o & 0 \\ 0 & -2Y_o \end{bmatrix}$	$\begin{bmatrix} Y_o & 0 \\ 0 & Y_o \end{bmatrix}$
		$\boldsymbol{U} = \begin{bmatrix} V_1 \\ I_2 \end{bmatrix}$	$\boldsymbol{Z} = \begin{bmatrix} I_1 \\ V_2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & Z_o \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} -2Y_o & 0 \\ 0 & 2 \end{bmatrix}$	$\begin{bmatrix} Y_o & 0 \\ 0 & Z_o \end{bmatrix}$
		$\boldsymbol{U} = \begin{bmatrix} I_1 \\ V_2 \end{bmatrix}$	$\boldsymbol{Z} = \begin{bmatrix} V_1 \\ I_2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ Z_o & 0 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 0 & -2Y_o \end{bmatrix}$	$\begin{bmatrix} Z_o & 0 \\ 0 & Y_o \end{bmatrix}$
		$\boldsymbol{U} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}$	$\boldsymbol{Z} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} Z_o & 0 \\ 0 & Z_o \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$	$\begin{bmatrix} Z_o & 0 \\ 0 & Z_o \end{bmatrix}$
LTA	$\boldsymbol{X}_2 = \begin{bmatrix} V^- \end{bmatrix}$	$U = V$	$Z = I$	1	-1	1	$-2Y_o$	Y_o
		$U = I$	$Z = V$	1	1	Z_o	2	Z_o
LTC	$\boldsymbol{X}_2 = \begin{bmatrix} V^- \end{bmatrix}$	$U = V$	$Z = I$	1	1	-1	$-2Y_o$	Y_o
		$U = I$	$Z = V$	1	-1	$-Z_o$	2	Z_o

onde
$Y_o = 1/Z_o$ é a impedância característica,
α é a constante de atenuação, e
τ é o tempo de atraso.

Tabela A.3:
REPRESENTAÇÃO DE ESPAÇO-DE-ESTADO DAS FONTES LINEARES CONTROLADAS

Elemento	Estado	Entrada(s)	Saída(s)	Matrizes da Equação de Estado e de Saída				
				$\underline{\mathbf{A}}_l$	$\underline{\mathbf{A}}_r$	$\underline{\mathbf{B}}$	$\underline{\mathbf{C}}$	$\underline{\mathbf{D}}$
Fonte de Corrente Controlada por Corrente	-	$U = V_2$	$Z = I_2$	-	-	-	-	K_{ii}
	$X_{2A} = U_1$			0	–1	1	K_{ii}	0
Fonte de Corrente Controlada por Tensão	-	$U = V_2$	$Z = I_2$	-	-	-	-	K_{iv}
	$X_{2A} = U_1$			0	–1	1	K_{iv}	0
Fonte de Tensão Controlada por Corrente	-	$\boldsymbol{U} = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}$	$\boldsymbol{Z} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$	-	-	-	-	K_{vi}
	$X_{2A} = U_1$			0	–1	1	K_{vi}	0
Fonte de Tensão Controlada por Tensão	-	$\boldsymbol{U} = \begin{bmatrix} V_1 \\ I_2 \end{bmatrix}$	$\boldsymbol{Z} = \begin{bmatrix} I_1 \\ V_2 \end{bmatrix}$	-	-	-	-	K_{vv}
	$X_{2A} = U_1$			0	–1	1	K_{vv}	0

onde

K_{ii} é o ganho de corrente (adimensional),
K_{iv} é o ganho transcondutivo,
K_{vi} é o ganho transresistivo,
K_{vv} é o ganho de tensão (adimensional),
τ é o tempo de atraso .

Tabela A.4:
REPRESENTAÇÃO DE ESPAÇO-DE-ESTADO DOS ELEMENTOS NÃO-LINEARES

Elemento	Estado e Função Não-linear	Entrada	Saída	Matrizes da Equação de Estado e de Saída						
				$\underline{\mathbf{A}}_l$	$\underline{\mathbf{A}}_r$	$\underline{\mathbf{B}}_f$	$\underline{\mathbf{B}}$	$\underline{\mathbf{C}}$	$\underline{\mathbf{D}}_f$	$\underline{\mathbf{D}}$
Resistor Controlado por Tensão	$X_3 = V$	$U = V$	$Z = I$	0	–1	0	1	0	1	0
	$U_{fs} = J$	$U = I$	$Z = V$	0	0	–1	1	1	0	0
Resistor Controlado por Corrente	$X_3 = I$	$U = I$	$Z = V$	0	–1	0	1	0	1	0
	$U_{fs} = E$	$U = V$	$Z = I$	0	0	–1	1	1	0	0
Capacitor de Controle Não-Linear	$X_3 = V$	$U = V$	$Z = I$	0	–1	0	1	0	1	0
	$U_{fd} = Q$	$U = I$	$Z = V$	0	0	–1	1	1	0	0
Indutor de Controle Não-Linear	$X_3 = I$	$U = I$	$Z = V$	0	–1	0	1	0	1	0
	$U_{fd} = \Psi$	$U = V$	$Z = I$	0	0	–1	1	1	0	0
Fonte de Corrente de Controle Não-Linear	$U_{fs} = J$	$U = V$	$Z = I$	0	-	-	-	0	1	0
Fonte de Tensão de Controle Não-Linear	$U_{fs} = E$	$U = I$	$Z = V$	0	-	-	-	0	1	0

Tabela A.5:
REPRESENTAÇÃO DE ESPAÇO-DE-ESTADO DOS ELEMENTOS DE CONTROLE LINEAR

Elemento	Estado	Entrada	Saída	Matrizes da Equação de Estado e de Saída				
				$\underline{\mathbf{A}}_l$	$\underline{\mathbf{A}}_r$	$\underline{\mathbf{B}}$	$\underline{\mathbf{C}}$	$\underline{\mathbf{D}}$
Corrente de Controle Linear	-	$U = I$	$Z = V$		-	-	0	0
Tensão de Controle Linear	-	$U = V$	$Z = I$		-	-	0	0
Tensão ou Corrente de Controle Não-Linear com Atraso	$\begin{bmatrix} U \\ X_{2A} = U \end{bmatrix}$	-	-		–1	1	0	0

Tabela A.6:
REPRESENTAÇÃO DE ESPAÇO-DE-ESTADO DOS ELEMENTOS DE CONTROLE NÃO-LINEAR

Elemento	Estado(s)	Entrada	Saída	Matriz da Equação de Estado e de Saída				
				$\underline{\mathbf{A}}_l$	$\underline{\mathbf{A}}_r$	$\underline{\mathbf{B}}$	$\underline{\mathbf{C}}$	$\underline{\mathbf{D}}$
Corrente de Controle Não-Linear	$X_3 = I$	$U = I$	$Z = V$		–1	1	0	0
	$X_3 = I$	$U = V$	$Z = I$		0	1	1	0
Tensão de Controle Não-Linear	$X_3 = V$	$U = V$	$Z = I$		–1	1	0	0
	$X_3 = V$	$U = I$	$Z = V$		0	1	1	0
Tensão ou Corrente de Controle Não-Linear com Atraso	$\begin{bmatrix} X_3 \\ X_{3A} = X_3 \end{bmatrix}$	-	-		$\begin{bmatrix} 1 & 0 \end{bmatrix}$	0	-	-

Tabela A.7:
REPRESENTAÇÃO DE ESPAÇO-DE-ESTADO DAS SONDAS

Elemento	Entrada	Saída	Sonda	Matrizes da Equação de Saída e de Sonda		
				$\underline{\mathbf{C}}$	$\underline{\mathbf{D}}$	$\underline{\mathbf{N}}$
Sonda de Tensão	$U = V$	$Z = I$	$Y = I$	-	0	1
Sonda de Corrente	$U = I$	$Z = V$	$Y = V$	-	0	1

Apêndice B. Formulação de Espaço-de-Estado do Circuito Elétrico Equivalente do FET e do HBT

O digrafo correspondente ao CEE (top-A) do FET da Fig. 3.3(a), é ilustrado na Fig. B.1(a). Os vetores de váriavel de estado, de função não-linear, de entrada, de saída, e de sonda externa, são dados por:

$$\boldsymbol{X} = \left[V_{R_{11}}\ V_{R_{12}}\ V_{R_{13}}\ V_{R_{31}}\ V_{R_{32}}\ V_{R_{33}}\ I_{R_{21}}\ I_{R_{22}}\ I_{R_{23}}\ V_{C_{dS}}\ V_{C_{DS}}\ I_{L_G}\ I_{L_D}\ I_{L_X}\ I_{C_{GS}}\ I_{C_{GD}}\ V_{L_S}\ V_{gsc}\ V_{gdc}\ V_{gs}\ V_{gd}\ V_{gscD}\right]^T,$$

$$\boldsymbol{U}_f = \left[J_{D_{gs}}\ J_{D_{gd}}\ J_{ds}\ J_{dX}\ Q_{gs}\ Q_{gd}\right]^T,$$

$$\boldsymbol{U}_e = \left[J_{e1}\ J_{e2}\ J_{e3}\right]^T,\ \boldsymbol{Y}_e = \left[V_{e1}\ V_{e2}\ V_{e3}\right]^T,$$

$$\boldsymbol{U} = \Big[\ I_{R_{11}}\ I_{R_{12}}\ I_{R_{13}}\ I_{R_{31}}\ I_{R_{32}}\ I_{R_{33}}\ I_{Q_{gs}}\ I_{Q_{gd}}\ I_{C_{ds}}\ I_{C_{DS}}\ I_{R_G}\ I_{R_D}\ I_{R_S}\ I_{R_{gd}}\ I_{R_X}\ I_{L_S}$$
$$V_{C_{GS}}\ V_{C_{GD}}\ V_{R_{gs}}\ V_X\ V_{gs}\ V_{gd}\ V_{L_G}\ V_{L_D}\ V_{L_X}\ V_{J_{dY}}\ V_{J_{Dgs}}\ V_{J_{Dgd}}\ V_{J_{ds}}\ V_{J_{dX}}\ V_{R_{21}}\ V_{R_{22}}\ V_{R_{23}}\ V_{e1}\ V_{e2}\ V_{e3}\ \Big]^T,$$

$$\boldsymbol{Y} = \Big[\ V_{R_{11}}\ V_{R_{12}}\ V_{R_{13}}\ V_{R_{31}}\ V_{R_{32}}\ V_{R_{33}}\ V_{Q_{gs}}\ V_{Q_{gd}}\ V_{C_{ds}}\ V_{C_{DS}}\ V_{R_G}\ V_{R_D}\ V_{R_S}\ V_{R_{gd}}\ V_{R_X}\ V_{L_S}$$
$$I_{C_{GS}}\ I_{C_{GD}}\ I_{R_{gs}}\ I_X\ I_{gs}\ I_{gd}\ I_{L_G}\ I_{L_D}\ I_{L_X}\ I_{J_{dY}}\ I_{J_{Dgs}}\ I_{J_{Dgd}}\ I_{J_{ds}}\ I_{J_{dX}}\ I_{R_{21}}\ I_{R_{22}}\ I_{R_{23}}\ I_{e1}\ I_{e2}\ I_{e3}\ \Big]^T.$$

Na Tabela B.1 são descritas as equações de estado, de saída e de sonda dos elementos básicos do CEE do FET da Fig. 3.3(a). A construção das matrizes constitutivas descritas nas tabelas do Apêndice A, pode ser facilmente visualizada.

O digrafo do CEE da parte intrínseca do FET, representada sob forma de DDS, é ilustrado na Fig. B.1(b). Os vetores de váriavel de estado, de função não-linear, de entrada, de saída, e de sonda externa, são dados por:

$$\boldsymbol{X} = \left[V_{R_{11}}\ V_{R_{12}}\ V_{R_{13}}\ V_{R_{14}}\ V_{R_{15}}\ V_g\ V_{sc}\ V_{dc}\ V_s\ V_d\ V_{gscD}\right]^T,$$

$$\boldsymbol{U}_f = \left[J_{D_{gs}}\ J_{D_{gd}}\ J_{ds}\ J_{dX}\ Q_{gs}\ Q_{gd}\right]^T,$$

$$\boldsymbol{U}_e = \left[J_{e1}\ J_{e2}\ J_{e3}\right]^T,\ \boldsymbol{Y}_e = \left[V_{e1}\ V_{e2}\ V_{e3}\right]^T,$$

$$\boldsymbol{U} = \left[I_{R_{11}}\ I_{R_{12}}\ I_{R_{13}}\ I_{R_{14}}\ I_{R_{15}}\ I_{Q_{gs}}\ I_{Q_{gd}}\ I_g\ I_{sc}\ I_{dc}\ I_s\ V_X\ V_{gs}\ V_{gd}\ V_{J_{dY}}\ V_{J_{Dgs}}\ V_{J_{Dgd}}\ V_{J_{ds}}\ V_{J_{dX}}\ V_{e1}\ V_{e2}\ V_{e3}\right]^T,$$

$$\boldsymbol{Y} = \left[V_{R_{11}}\ V_{R_{12}}\ V_{R_{13}}\ V_{R_{14}}\ V_{R_{15}}\ V_{Q_{gs}}\ V_{Q_{gd}}\ V_g\ V_{sc}\ V_{dc}\ V_s\ I_X\ I_{gs}\ I_{gd}\ I_{J_{dY}}\ I_{J_{Dgs}}\ I_{J_{Dgd}}\ I_{J_{ds}}\ I_{J_{dX}}\ I_{e1}\ I_{e2}\ I_{e3}\right]^T,$$

Na Tabela B.2 são descritas as equações de estado, de saída e de sonda dos elementos básicos do CEE da parte intrínseca do FET da Fig. 3.3(b) (ver Apêndice A).

O digrafo correspondente ao CEE (top-A) do HBT da Fig. 3.3(b), é ilustrado na Fig. B.1(a). Os vetores de variável de estado, de função não-linear, de entrada, de saída, e de sonda externa, são

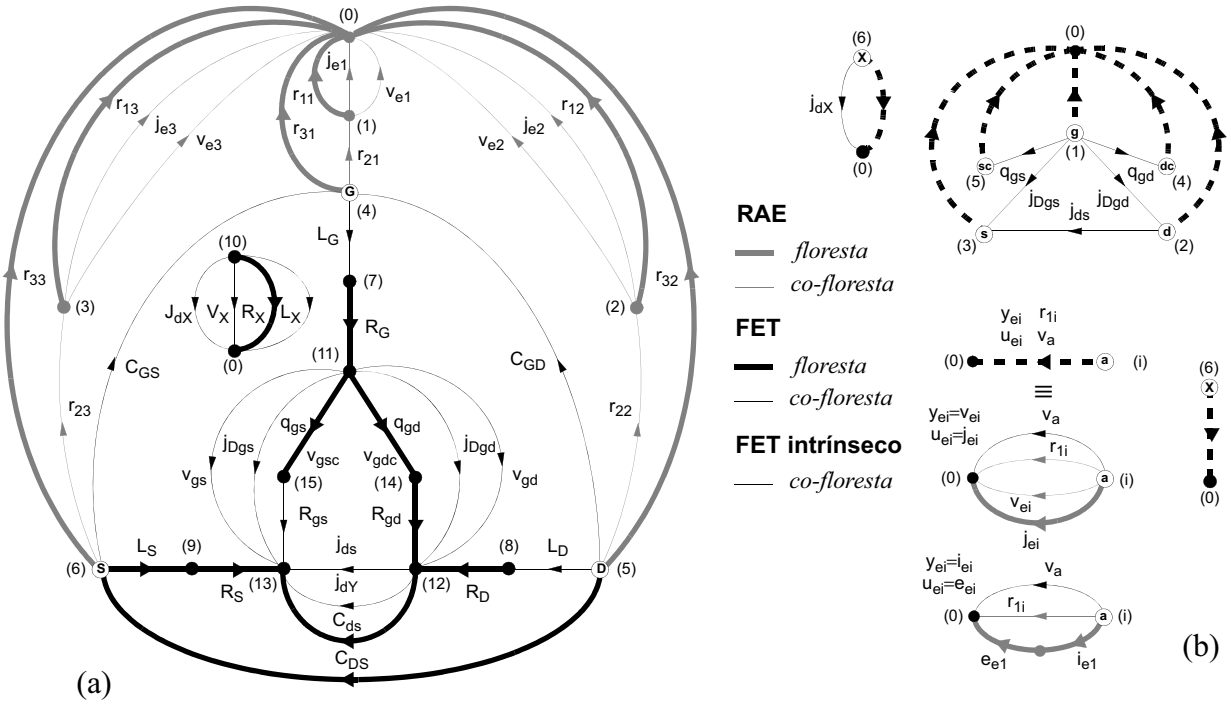


Fig. B.1 (a) Digrafo do circuito elétrico equivalente (CEE) do FET incluindo rede de alimentação externa (RAE). (b) Digrafo do CEE da parte intrínseca do FET sob forma de dispositivo definido-simbolicamente (DDS) incluindo RAE.

Tabela B.1:
EQUAÇÕES DE ESTADO E DE SAÍDA DOS ELEMENTOS BÁSICOS DO CEE DO FET

floresta						cofloresta					
j	j_x	j_f	EB	Equação de Saída	Equação de Estado	j	j_x (j_e)	j_f	EB	Equação de Saída	Equação de Estado (de Sonda)
1	1	–	r_{11}	$z_1 = x_1$	$0 = r_{11}^{-1}x_1 - u_1$	16	15	–	C_{GS}	-	$x_{15}/\hat{\jmath}\omega C_{GS} = u_{16}$
2	2	–	r_{12}	$z_2 = x_2$	$0 = r_{12}^{-1}x_2 - u_2$	17	16	–	C_{GD}	-	$x_{15}/\hat{\jmath}\omega C_{GS} = u_{16}$
3	3	–	r_{13}	$z_3 = x_3$	$0 = r_{13}^{-1}x_3 - u_3$	18	17	–	C_{DS}	-	$x_{15}/\hat{\jmath}\omega C_{GS} = u_{16}$
4	4	–	r_{31}	$z_4 = x_4$	$0 = r_{31}^{-1}x_4 - u_4$	19	-	–	R_{gs}	$z_{19} = R_{gs}^{-1}u_{19}$	
5	5	–	r_{32}	$z_5 = x_5$	$0 = r_{32}^{-1}x_5 - u_5$	20	-	–	X	$z_{20} = 0$	
6	6	–	r_{33}	$z_6 = x_6$	$0 = r_{33}^{-1}x_6 - u_6$	21	20	–	gs	$z_{21} = 0$	
7	18	5	q_{gs}	$z_7 = x_{18}$	$0 = \hat{\jmath}\omega u_{f5} - u_7$	22	21	–	gd	$z_{22} = 0$	
8	19	6	q_{gd}	$z_8 = x_{19}$	$0 = \hat{\jmath}\omega u_{f6} - u_8$	23	11	–	L_G	$z_{23} = x_{11}$	$0 = \hat{\jmath}\omega L_G x_{11} - u_{23}$
9	10	–	C_{ds}	$z_9 = x_{10}$	$\hat{\jmath}\omega C_{ds}x_{10} = u_9$	24	12	–	L_D	$z_{24} = x_{12}$	$0 = \hat{\jmath}\omega L_D x_{12} - u_{24}$
10	–	–	R_G	$z_{10} = R_G u_{10}$	–	25	13	–	L_X	$z_{25} = x_{13}$	$0 = \hat{\jmath}\omega L_X x_{13} - u_{25}$
11	–	–	R_D	$z_{11} = R_D u_{11}$	–	26	–	–	j_{dY}	$z_{26} = R_X^{-1}u_{20}$	–
12	–	–	R_S	$z_{12} = R_S u_{12}$	–	27	–	1	j_{Dgs}	$z_{27} = u_{f1}$	–
13	–	–	R_{gd}	$z_{13} = R_{gd}u_{13}$	–	28	–	2	j_{Dgd}	$z_{28} = u_{f2}$	–
14	–	–	R_X	$z_{14} = R_X u_{14}$	–	29	–	3	j_{ds}	$z_{29} = u_{f3}$	–
15	14	–	L_S	$z_{15} = x_{14}$	$x_{14}/\hat{\jmath}\omega L_S = u_{15}$	30	–	4	j_{dX}	$z_{30} = u_{f4}$	–
						31	7	–	r_{21}	$z_{31} = x_7$	$0 = r_{21}x_7 - u_{31}$
						32	8	–	r_{22}	$z_{32} = x_8$	$0 = r_{22}x_8 - u_{32}$
						33	9	–	r_{23}	$z_{33} = x_9$	$0 = r_{23}x_9 - u_{33}$
						34	(1)	–	e_1	$z_{34} = 0$	$(y_{e1} = u_{34})$
						35	(2)	–	e_2	$z_{35} = 0$	$(y_{e2} = u_{35})$
						36	(3)	–	e_3	$z_{36} = 0$	$(y_{e3} = u_{36})$
						37	–	–	j_{e1}	–	–
						38	–	–	j_{e2}	–	–
						39	–	–	j_{e3}	–	–

Tabela B.2:
EQUAÇÕES DE ESTADO E DE SAÍDA DOS ELEMENTOS BÁSICOS DO CEE
DA PARTE INTRÍNSECA DO FET REPRESENTADA SOB FORMA DE DDS

floresta						cofloresta					
j	j_x	j_f	EB	Equação de Saída	Equação de Estado	j	j_x (j_e)	j_f	EB	Equação de Saída	Equação de Estado (de Sonda)
1	1	–	r_{11}	$z_1 = x_1$	$0 = r_{11}^{-1}x_1 - u_1$	16	15	–	C_{GS}	-	$x_{15}/\hat{\jmath}\omega C_{GS} = u_{16}$
2	2	–	r_{12}	$z_2 = x_2$	$0 = r_{12}^{-1}x_2 - u_2$	17	16	–	C_{GD}	-	$x_{15}/\hat{\jmath}\omega C_{GS} = u_{16}$
3	3	–	r_{13}	$z_3 = x_3$	$0 = r_{13}^{-1}x_3 - u_3$	18	17	–	C_{DS}	-	$x_{15}/\hat{\jmath}\omega C_{GS} = u_{16}$
4	4	–	r_{31}	$z_4 = x_4$	$0 = r_{31}^{-1}x_4 - u_4$	19	-	–	R_{gs}	$z_{19} = R_{gs}^{-1}u_{19}$	
5	5	–	r_{32}	$z_5 = x_5$	$0 = r_{32}^{-1}x_5 - u_5$	20	-	–	X	$z_{20} = 0$	
6	6	–	r_{33}	$z_6 = x_6$	$0 = r_{33}^{-1}x_6 - u_6$	21	20	–	gs	$z_{21} = 0$	
7	18	5	q_{gs}	$z_7 = x_{18}$	$0 = \hat{\jmath}\omega u_{f5} - u_7$	22	21	–	gd	$z_{22} = 0$	
8	19	6	q_{gd}	$z_8 = x_{19}$	$0 = \hat{\jmath}\omega u_{f6} - u_8$	23	11	–	L_G	$z_{23} = x_{11}$	$0 = \hat{\jmath}\omega L_G x_{11} - u_{23}$
9	10	–	C_{ds}	$z_9 = x_{10}$	$\hat{\jmath}\omega C_{ds}x_{10} = u_9$	24	12	–	L_D	$z_{24} = x_{12}$	$0 = \hat{\jmath}\omega L_D x_{12} - u_{24}$
10	–	–	R_G	$z_{10} = R_G u_{10}$	–	25	13	–	L_X	$z_{25} = x_{13}$	$0 = \hat{\jmath}\omega L_X x_{13} - u_{25}$
11	–	–	R_D	$z_{11} = R_D u_{11}$	–	26	–	–	j_{dY}	$z_{26} = R_X^{-1}u_{20}$	–
12	–	–	R_S	$z_{12} = R_S u_{12}$	–	27	–	1	j_{Dgs}	$z_{27} = u_{f1}$	–
13	–	–	R_{gd}	$z_{13} = R_{gd}u_{13}$	–	28	–	2	j_{Dgd}	$z_{28} = u_{f2}$	–
14	–	–	R_X	$z_{14} = R_X u_{14}$	–	29	–	3	j_{ds}	$z_{29} = u_{f3}$	–
15	14	–	L_S	$z_{15} = x_{14}$	$x_{14}/\hat{\jmath}\omega L_S = u_{15}$	30	–	4	j_{dX}	$z_{30} = u_{f4}$	–
						31	7	–	r_{21}	$z_{31} = x_7$	$0 = r_{21}x_7 - u_{31}$
						32	8	–	r_{22}	$z_{32} = x_8$	$0 = r_{22}x_8 - u_{32}$
						33	9	–	r_{23}	$z_{33} = x_9$	$0 = r_{23}x_9 - u_{33}$
						34	(1)	–	e_1	$z_{34} = 0$	$(y_{e1} = u_{34})$
						35	(2)	–	e_2	$z_{35} = 0$	$(y_{e2} = u_{35})$
						36	(3)	–	e_3	$z_{36} = 0$	$(y_{e3} = u_{36})$
						37	–	–	j_{e1}	–	–
						38	–	–	j_{e2}	–	–
						39	–	–	j_{e3}	–	–

dados por:

$$\boldsymbol{X} = \left[V_{R_{11}}\ V_{R_{12}}\ V_{R_{13}}\ V_{R_{31}}\ V_{R_{32}}\ V_{R_{33}}\ I_{R_{21}}\ I_{R_{22}}\ I_{R_{23}}\ V_{C_{CE}}\ I_{L_B}\ I_{L_C}\ I_{C_{BE}}\ I_{C_{BC}}\ V_{L_E}\ V_{be}\ V_{bc}\ V_{bxc}\right]^T$$

$$\boldsymbol{U}_f = \left[J_{be}\ J_{bc}\ J_{D_{bxc}}\ Q_{be}\ Q_{bc}\ Q_{D_{bxc}}\right]^T$$

$$\boldsymbol{U}_e = \left[J_{e1}\ J_{e2}\ J_{e3}\right]^T \quad , \quad \boldsymbol{Y}_e = \left[V_{e1}\ V_{e2}\ V_{e3}\right]^T \quad ,$$

$$\boldsymbol{U} = \left[\ I_{R_{11}}\ I_{R_{12}}\ I_{R_{13}}\ I_{R_{31}}\ I_{R_{32}}\ I_{R_{33}}\ I_{Q_{be}}\ I_{Q_{bc}}\ I_{Q_{bxc}}\ I_{C_{CE}}\ I_{R_{B1}}\ I_{R_C}\ I_{R_E}\ I_{L_E}\right.$$
$$\left. V_{C_{BC}}\ V_{C_{BE}}\ V_{R_{B2}}\ V_{L_B}\ V_{L_C}\ V_{J_{be}}\ V_{J_{bc}}\ V_{J_{Dbxc}}\ V_{R_{21}}\ V_{R_{22}}\ V_{R_{23}}\ \right]^T \quad ,$$

$$\boldsymbol{Y} = \left[\ V_{R_{11}}\ V_{R_{12}}\ V_{R_{13}}\ V_{R_{31}}\ V_{R_{32}}\ V_{R_{33}}\ V_{Q_{be}}\ V_{Q_{bc}}\ V_{Q_{bxc}}\ V_{C_{CE}}\ V_{R_{B1}}\ V_{R_C}\ V_{R_E}\ V_{L_E}\right.$$
$$\left. I_{C_{BC}}\ I_{C_{BE}}\ I_{R_{B2}}\ I_{L_B}\ I_{L_C}\ I_{J_{be}}\ I_{J_{bc}}\ I_{J_{Dbxc}}\ I_{R_{21}}\ I_{R_{22}}\ I_{R_{23}}\ \right]^T \quad .$$

Na Tabela B.3 são descritas as equações de estado e de sonda dos elementos básicos do CEE do HBT da Fig. 3.3(b) (ver Apêndice A).

O digrafo do CEE da parte intrínseca do HBT, representada sob forma de DDS, é ilustrado na

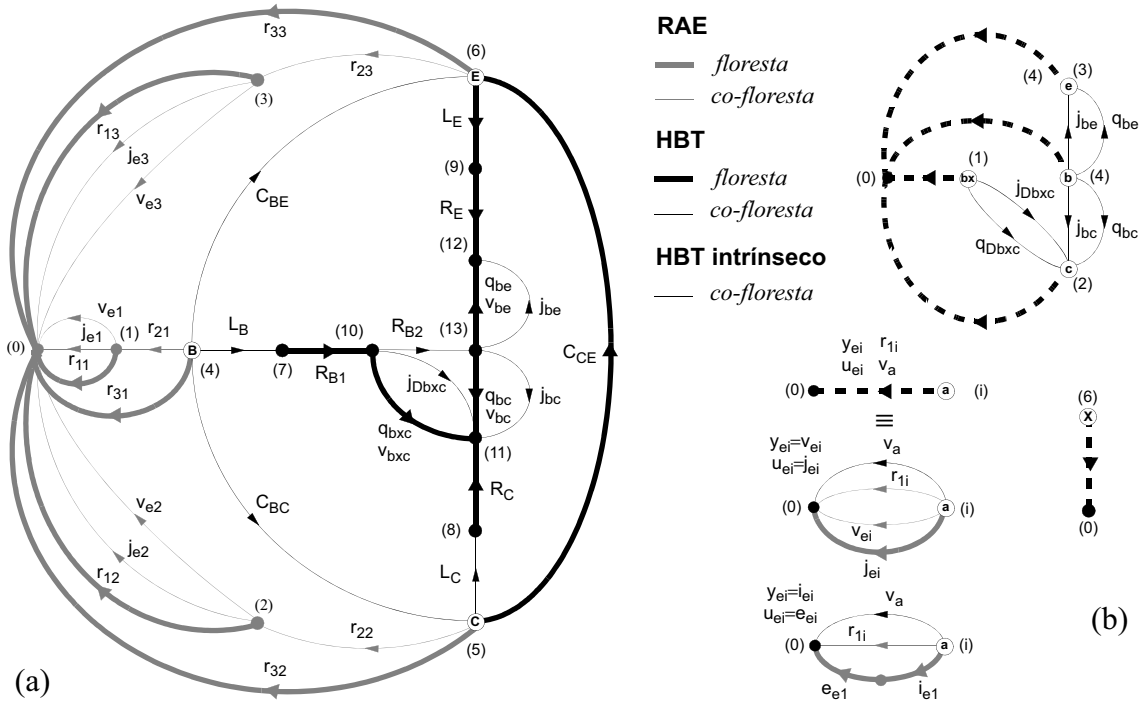


Fig. B.1 (a) Digrafo do circuito elétrico equivalente (CEE) do HBT incluindo rede de alimentação externa (RAE). (b) Digrafo do CEE da parte intrínseca do HBT sob forma de dispositivo definido-simbolicamente (DDS) incluindo RAE.

Tabela B.3:
EQUAÇÕES DE ESTADO E DE SAÍDA DOS ELEMENTOS BÁSICOS DO CEE DO HBT

floresta						cofloresta					
j	j_x	j_f	EB	Equação de Saída	Equação de Estado	j	j_x (j_e)	j_f	EB	Equação de Saída	Equação de Estado (de Sonda)
1	1	–	r_{11}	$z_1 = x_1$	$0 = r_{11}^{-1}x_1 - u_1$	14	13	–	C_{BE}	-	$x_{13}/\hat{\jmath}\omega C_{BE} = u_{14}$
2	2	–	r_{12}	$z_2 = x_2$	$0 = r_{12}^{-1}x_2 - u_2$	15	14	–	C_{BC}	-	$x_{14}/\hat{\jmath}\omega C_{BC} = u_{15}$
3	3	–	r_{13}	$z_3 = x_3$	$0 = r_{13}^{-1}x_3 - u_3$	16	15	–	C_{CE}	-	$x_{15}/\hat{\jmath}\omega C_{CE} = u_{16}$
4	4	–	r_{31}	$z_4 = x_4$	$0 = r_{31}^{-1}x_4 - u_4$	17	-	–	R_{B2}	$z_{17} = R_{B2}^{-1}u_{17}$	
5	5	–	r_{32}	$z_5 = x_5$	$0 = r_{32}^{-1}x_5 - u_5$	18	10	–	L_B	$z_{23} = x_{11}$	$0 = \hat{\jmath}\omega L_B x_{10} - u_{18}$
6	6	–	r_{33}	$z_6 = x_6$	$0 = r_{33}^{-1}x_6 - u_6$	19	11	–	L_C	$z_{24} = x_{12}$	$0 = \hat{\jmath}\omega L_C x_{11} - u_{19}$
7	16	4	q_{be}	$z_7 = x_{18}$	$0 = \hat{\jmath}\omega u_{f4} - u_7$	20	–	1	j_{Dgs}	$z_{20} = u_{f1}$	–
8	17	5	q_{bc}	$z_8 = x_{19}$	$0 = \hat{\jmath}\omega u_{f5} - u_8$	21	–	2	j_{Dgd}	$z_{21} = u_{f2}$	–
9	18	6	q_{bxc}	$z_9 = x_{10}$	$0 = \hat{\jmath}\omega u_{f6} - u_9$	22	–	3	j_{ds}	$z_{22} = u_{f3}$	–
10	–	–	R_{B1}	$z_{10} = R_{B1}u_{10}$	–	23	7	–	r_{21}	$z_{23} = x_7$	$0 = r_{23}x_7 - u_{23}$
11	–	–	R_C	$z_{11} = R_C u_{11}$	–	24	8	–	r_{22}	$z_{24} = x_8$	$0 = r_{24}x_8 - u_{24}$
12	–	–	R_E	$z_{12} = R_E u_{12}$	–	25	9	–	r_{23}	$z_{25} = x_9$	$0 = r_{25}x_9 - u_{25}$
13	12	–	R_{gd}	$z_{13} = x_{12}$	$x_{12}/\hat{\jmath}\omega L_E = u_{13}$	26	(1)	–	e_1	$z_{26} = 0$	$(y_{e1} = u_{26})$
						27	(2)	–	e_2	$z_{27} = 0$	$(y_{e2} = u_{27})$
						28	(3)	–	e_3	$z_{28} = 0$	$(y_{e3} = u_{28})$
						29	–	–	j_{e1}	–	–
						30	–	–	j_{e2}	–	–
						31	–	–	j_{e3}	–	–

Tabela B.4:
EQUAÇÕES DE ESTADO E DE SAÍDA DOS ELEMENTOS BÁSICOS DO CEE DA PARTE INTRÍNSECA DO HBT REPRESENTADA SOB FORMA DE DDS

floresta						cofloresta					
j	j_x	j_f	EB	Equação de Saída	Equação de Estado	j	j_x (j_e)	j_f	EB	Equação de Saída	Equação de Estado (de Sonda)
1	1	–	r_{11}	$z_1 = x_1$	$0 = r_{11}^{-1}x_1 - u_1$	14	13	–	C_{BE}	-	$x_{13}/\hat{\jmath}\omega C_{BE} = u_{14}$
2	2	–	r_{12}	$z_2 = x_2$	$0 = r_{12}^{-1}x_2 - u_2$	15	14	–	C_{BC}	-	$x_{14}/\hat{\jmath}\omega C_{BC} = u_{15}$
3	3	–	r_{13}	$z_3 = x_3$	$0 = r_{13}^{-1}x_3 - u_3$	16	15	–	C_{CE}	-	$x_{15}/\hat{\jmath}\omega C_{CE} = u_{16}$
4	4	–	r_{31}	$z_4 = x_4$	$0 = r_{31}^{-1}x_4 - u_4$	17	-	–	R_{B2}	$z_{17} = R_{B2}^{-1}u_{17}$	
5	5	–	r_{32}	$z_5 = x_5$	$0 = r_{32}^{-1}x_5 - u_5$	18	10	–	L_B	$z_{23} = x_{11}$	$0 = \hat{\jmath}\omega L_B x_{10} - u_{18}$
6	6	–	r_{33}	$z_6 = x_6$	$0 = r_{33}^{-1}x_6 - u_6$	19	11	–	L_C	$z_{24} = x_{12}$	$0 = \hat{\jmath}\omega L_C x_{11} - u_{19}$
7	16	4	q_{be}	$z_7 = x_{18}$	$0 = \hat{\jmath}\omega u_{f4} - u_7$	20	–	1	j_{Dgs}	$z_{20} = u_{f1}$	–
8	17	5	q_{bc}	$z_8 = x_{19}$	$0 = \hat{\jmath}\omega u_{f5} - u_8$	21	–	2	j_{Dgd}	$z_{21} = u_{f2}$	–
9	18	6	q_{bxc}	$z_9 = x_{10}$	$0 = \hat{\jmath}\omega u_{f6} - u_9$	22	–	3	j_{ds}	$z_{22} = u_{f3}$	–
10	–	–	R_{B1}	$z_{10} = R_{B1}u_{10}$	–	23	7	–	r_{21}	$z_{23} = x_7$	$0 = r_{23}x_7 - u_{23}$
11	–	–	R_C	$z_{11} = R_C u_{11}$	–	24	8	–	r_{22}	$z_{24} = x_8$	$0 = r_{24}x_8 - u_{24}$
12	–	–	R_E	$z_{12} = R_E u_{12}$	–	25	9	–	r_{23}	$z_{25} = x_9$	$0 = r_{25}x_9 - u_{25}$
13	12	–	R_{gd}	$z_{13} = x_{12}$	$x_{12}/\hat{\jmath}\omega L_E = u_{13}$	26	(1)	–	e_1	$z_{26} = 0$	$(y_{e1} = u_{26})$
						27	(2)	–	e_2	$z_{27} = 0$	$(y_{e2} = u_{27})$
						28	(3)	–	e_3	$z_{28} = 0$	$(y_{e3} = u_{28})$
						29	–	–	j_{e1}	–	–
						30	–	–	j_{e2}	–	–
						31	–	–	j_{e3}	–	–

Fig. B.1(b). Os vetores de váriavel de estado, de função não-linear, de entrada, de saída, e de sonda externa, são dados por:

$$\boldsymbol{X} = \left[V_{R_{11}}\ V_{R_{12}}\ V_{R_{13}}\ V_{R_{31}}\ V_{R_{32}}\ V_{R_{33}}\ I_{R_{21}}\ I_{R_{22}}\ I_{R_{23}}\ V_{C_{CE}}\ I_{L_B}\ I_{L_C}\ I_{C_{BE}}\ I_{C_{BC}}\ V_{L_E}\ V_{be}\ V_{bc}\ V_{bxc}\right]^T \quad ,$$

$$\boldsymbol{U}_f = \left[J_{be}\ J_{bc}\ J_{D_{bxc}}\ Q_{be}\ Q_{bc}\ Q_{D_{bxc}}\right]^T \quad ,$$

$$\boldsymbol{U}_e = \left[J_{e1}\ J_{e2}\ J_{e3}\right]^T \quad , \qquad \boldsymbol{Y}_e = \left[V_{e1}\ V_{e2}\ V_{e3}\right]^T \quad ,$$

$$\boldsymbol{U} = \left[\ I_{R_{11}}\ I_{R_{12}}\ I_{R_{13}}\ I_{R_{31}}\ I_{R_{32}}\ I_{R_{33}}\ I_{Q_{be}}\ I_{Q_{bc}}\ I_{Q_{bxc}}\ I_{C_{CE}}\ I_{R_{B1}}\ I_{R_C}\ I_{R_E}\ I_{L_E}\right.$$
$$\left. V_{C_{BC}}\ V_{C_{BE}}\ V_{R_{B2}}\ V_{L_B}\ V_{L_C}\ V_{J_{be}}\ V_{J_{bc}}\ V_{J_{Dbxc}}\ V_{R_{21}}\ V_{R_{22}}\ V_{R_{23}}\ \right]^T \quad ,$$

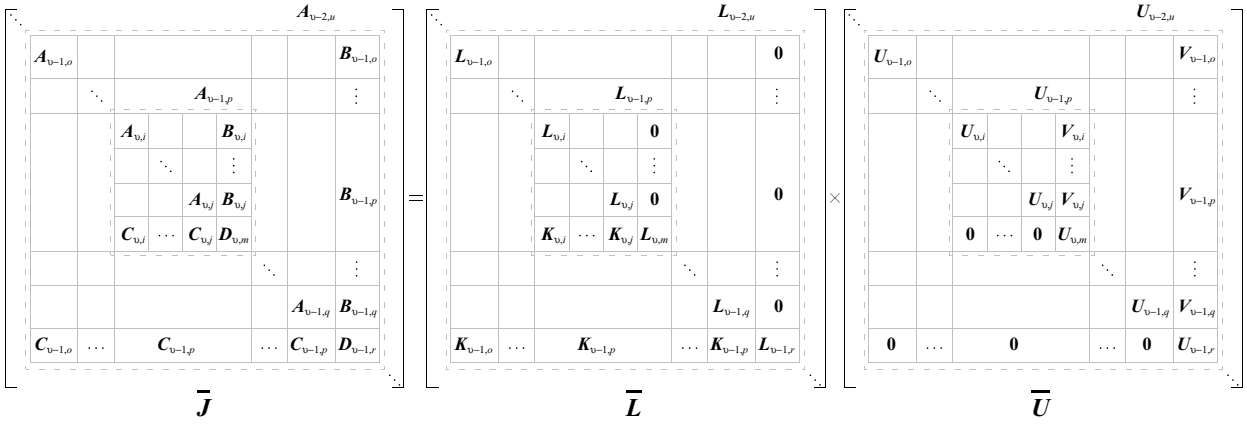
$$\boldsymbol{Y} = \left[\ V_{R_{11}}\ V_{R_{12}}\ V_{R_{13}}\ V_{R_{31}}\ V_{R_{32}}\ V_{R_{33}}\ V_{Q_{be}}\ V_{Q_{bc}}\ V_{Q_{bxc}}\ V_{C_{CE}}\ V_{R_{B1}}\ V_{R_C}\ V_{R_E}\ V_{L_E}\right.$$
$$\left. I_{C_{BC}}\ I_{C_{BE}}\ I_{R_{B2}}\ I_{L_B}\ I_{L_C}\ I_{J_{be}}\ I_{J_{bc}}\ I_{J_{Dbxc}}\ I_{R_{21}}\ I_{R_{22}}\ I_{R_{23}}\ \right]^T \quad .$$

$$\boldsymbol{Y} = \left[\ V_{R_{11}}\ V_{R_{12}}\ V_{R_{13}}\ V_{R_{14}}\ V_{R_{15}}\ V_g\ V_{sc}\ V_{dc}\ V_s\ V_d\ V_{R_X}\right.$$
$$\left. I_X\ I_{gs}\ I_{gd}\ I_{J_{dY}}\ I_{J_{Dgs}}\ I_{J_{Dgd}}\ I_{J_{ds}}\ I_{J_{dX}}\ I_{e1}\ I_{e2}\ I_{e3}\ \right]^T$$

Na Tabela B.4 são descritas as equações de estado, de saída e de sonda dos elementos básicos do CEE da parte intríseca do HBT da Fig. 3.3(b) (ver Apêndice A).

Apêndice C. Solução multi-níveis de Sistema Bloco Diagonal com Dupla Borda

A solução multi-níveis de um sistema jacobiano de estrutura bloco diagonal com borda dupla, pode ser eficientemente calculada utilizando a fatorização LU. Para ilustrarmos o processo de fatorização, vamos considerar a seguinte representação



associada à FIG. Da representação acima, podemos obter os termos LU associado a matriz $\boldsymbol{A}_{\upsilon-1,\,p}$, no nível υ, através da seguinte sequência de operações:

- 1) Fatorização: $\boldsymbol{A}_{\upsilon,\,t} \Rightarrow \boldsymbol{L}_{\upsilon,\,t}\boldsymbol{U}_{\upsilon,\,t}$;
- 2) Retro-substituição: $\boldsymbol{L}_{\upsilon,\,t}\boldsymbol{V}_{\upsilon,\,t} = \boldsymbol{B}_{\upsilon,\,t} \Rightarrow \boldsymbol{V}_{\upsilon,\,t}$;
- 3) Retro-substituição: $\boldsymbol{K}_{\upsilon,\,t}\boldsymbol{U}_{\upsilon,\,t} = \boldsymbol{C}_{\upsilon,\,t} \Rightarrow \boldsymbol{U}_{\upsilon,\,t}$;

onde $t = i, \ldots, j$;

- Fatorização: $\boldsymbol{D}_{\upsilon,\,m} - \sum_{t=j}^{k} \boldsymbol{K}_{\upsilon,\,t}\boldsymbol{V}_{\upsilon,\,t} \Rightarrow \boldsymbol{L}_{\upsilon,\,m}\boldsymbol{U}_{\upsilon,\,m}$.

Outros blocos no nível υ podem ser processados da mesma forma. No nível $\upsilon - 1$ a determinação dos fatores LU.

- 4) Fatorar: $\boldsymbol{A}_{\upsilon-1,\,t}$, para determinar $\boldsymbol{L}_{\upsilon-1,\,t}$ e $\boldsymbol{U}_{\upsilon-1,\,t}$;
- 5) Retro-resolver: $\boldsymbol{L}_{\upsilon-1,\,t}\boldsymbol{V}_{\upsilon-1,\,t} = \boldsymbol{B}_{\upsilon-1,\,t}$, para determinar $\boldsymbol{V}_{\upsilon-1,\,t}$;
- 6) Retro-resolver $\boldsymbol{K}_{\upsilon-1,\,t}\boldsymbol{U}_{\upsilon-1,\,t} = \boldsymbol{C}_{\upsilon-1,\,t}$, para determinar $\boldsymbol{U}_{\upsilon-1,\,t}$;

onde $t = p, \ldots, q$;

- Fatorização: $\boldsymbol{D}_{\upsilon-1,\,r} - \sum_{t=p}^{q} \boldsymbol{K}_{\upsilon-1,\,t}\boldsymbol{V}_{\upsilon-1,\,t} \Rightarrow \boldsymbol{L}_{\upsilon-1,\,r}\boldsymbol{U}_{\upsilon-1,\,r}$;

determinação do passo 4. A determinação dos passos 5-6 requerem a solução a ser discutida abaixo.

A solução do sistema pode ser obtida via retro-substituição, utilizando os fatores L e U

$$\boldsymbol{J}\,\boldsymbol{x} = \boldsymbol{b}$$

$$\boldsymbol{L} \times \underbrace{\begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & U_{\upsilon,m} & \cdots & \boldsymbol{0} & U_{\upsilon-1,r} \end{bmatrix}}_{\boldsymbol{z}} = \underbrace{\begin{bmatrix} U_{\upsilon-1,r} & \boldsymbol{0} & \cdots & U_{\upsilon,m} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}}_{\boldsymbol{b}}$$

$$\boldsymbol{U} \times \underbrace{\begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & U_{\upsilon,m} & \cdots & \boldsymbol{0} & U_{\upsilon-1,r} \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} U_{\upsilon-1,r} & \boldsymbol{0} & \cdots & U_{\upsilon,m} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}}_{\boldsymbol{z}}$$

Da representação acima, podemos calcular a solução do sistema jacobiano multi-níveis. ??? Esta solução é obtida através da seguinte sequência de operações