Projeto Brasil 6G
Fase III

Atividade 2.2 - Concepção da Rede de Acesso para Rede 6G - Parte 1

31/07/2025

Histórico de Atualizações:

Versão	Data	Autor(es)	Notas
1	14/7/2024	Abraão C. Albuquerque Bianca S. de C. da Silva Daniely Gomes Silva Eduardo Nunes Velloso Fabbryccio Akkazzha C. M. Cardoso Francisco Raimundo A. Parente Igor M. Guerreiro Isaac Andrés Balbuca Guachizaca Joanna Manjarres Jonathan N. Gois José Cândido Silveira Santos Filho Lucas dos Santos Costa Luiz Fernando Gontijo Maikon Renan Pereira da Silva Mariana Baracat de Mello Matheus D. Carneiro Paulo Cardieri Plínio Santini Dester Richard Demo Souza Samuel S. Silva Victor F. Monteiro Victoria D. P. Souto Yuri C. B. Silva	Elaboração de conteúdo
2	31/07/2025	Daniely Gomes Silva Filipe Henrique Cardoso José Cândido Silveira Santos Filho Luciano Leonel Mendes Paulo Cardieri Richard Demo Souza	Revisão de texto

Lista de Figuras

1 Análise de desempenho da PAPR (1a) e da BER (1b). Avaliação da complexidade para valores de $\mathcal{P}_{\max} = 5$ dB (1c) e $\mathcal{P}_{\max} = 6$ dB (1d). 7
2 Taxa de soma alcançável versus o número de elementos M da BD-RIS, para NOMA e OMA. 11
3 Taxa de soma alcançável versus o número de antenas N da BS, para NOMA e OMA. 12
4 Diagrama em blocos do transceptor FTN-GFDMA no *downlink*, integrado ao sistema A-RoF, considerando o terminal do usuário u. As etapas de adição e remoção do CP foram omitidas para simplificação. 17
5 Desempenho de BER dos sistemas FTN-GFDM utilizando BPSK e filtro RRC com fator de roll-off $\alpha = 0.5$. 18
6 PAPR dos sistemas FTN-GFDM e GFDM, com CCDF obtida empiricamente a partir de 10^5 quadros, cada um contendo 40 blocos de FTN-GFDM, e fator de superamostragem de 4. 19
7 Curvas ROC das estatísticas de teste T_1 e T_2 e respectivas AUCs. 24
8 Curvas ROC ED no NCSS; com SNR média de $\bar{\gamma} = -12$ dB, na ausência de NU (linha pontilhada, com $\rho = 0$) e sob NU ($\rho \neq \{0{,}1\}$), com $\rho = \{1{,}06; 1{,}09; 1{,}12; 1{,}15\}$ na combinação entre o intervalo e a fonte tradicional de NU (linhas tracejadas) e $\rho = \{0{,}06; 0{,}09; 0{,}12; 0{,}15\}$ na combinação entre a nova fonte e o novo intervalo de NU (linhas sólidas). 47
9 Curvas ROC (teóricas e simuladas) da estatística de teste ED no NCSS e CSS (com $M = 5$ SUs) via SD com os cômputos tradicional (32) (linha pontilhada) e novo (38) (linha sólida) da estatística de teste, e via HD sob as regras OR ($k = 1$), MAJ ($k = \lfloor M/2 + 1 \rfloor = 3$), e AND ($k = 5$), com SNR média $\{\bar{\gamma}_1; \bar{\gamma}_2; \ldots; \bar{\gamma}_M\} = \{-13; -14; -15; -16; -17\}$ dB, combinação da nova fonte com o novo intervalo de NU e fatores de incerteza de ruído $\{\rho_1; \rho_2; \ldots; \rho_M\} = \{0{,}02; 0{,}015; 0{,}012; 0{,}01; 0{,}0075\}$. 49
10 1-TSA com $n = 100$, $A_1 = 0$, $p_1 = 0.01$, correspondente ao *slotted* Aloha clássico. *Throughput* médio: 0.3704, *Age of Information* (AoI) médio: 270.63, desvio padrão: 26.33. 54
11 1-TSA com $n = 100$, $A_1 = 217$, $p_1 = 0.0443$. *Throughput* médio: 0.3379, AoI médio: 147.85, desvio padrão: 9.42. 55
12 2-TSA com $n = 100$, $A_1 = 217$, $p_1 = 0.0443$, $A_2 = 300$, $p_2 = 0.03$. *Throughput* médio: 0.3594, AoI médio: 148.48, desvio padrão: 9.84. 55
13 Drift do TSA com $n = 100$, $A = 217$, e vários valores de p. 56
14 Exemplo de realização de $\Delta(t)$. As marcas t_i e t'_i representam os instantes de geração e de recepção de i-ésima amostra; os pontos A e B indicam os valores de AoI da amostra 1 quando recebida e quando substituída pela amostra 2, respectivamente. 60
15 Sistema de rastreamento remoto de processos empregando a estratégia de comunicação orientada a tarefa. 63
16 Diagrama do sistema de comunicação semântica baseado em [1]. 67
17 Resultados de BLEU para o canal Rayleigh. 68
18 Resultados de BERTScore para os canais AWGN, Rayleigh e Rice. 69
19 Resultados preliminares de treinamento para o algoritmo MRC. 70

20 Resultados preliminares de treinamento para o algoritmo MIC. 70
21 Modelo de compressão para uma fonte semântica. 72
22 A arquitetura DtS-IoT é composta por um *gateway* embarcado em um satélite LEO, dispositivos IoT distribuídos sobre a área de interesse, uma estação terrestre e um servidor de rede. 76
23 Número médio de bytes recebidos com sucesso por volta, em função do número de dispositivos IoT, para os esquemas propostos FTP e CTP, assim como para o ALOHA e a estratégia *Max Angle*. 78
24 Eficiência energética em função do número de dispositivos IoT para os esquemas propostos FTP (em vermelho, marcador x) e CTP (em verde claro, marcador quadrado), bem como para as abordagens ALOHA (em marrom, marcador triangular) e Max Angle (em roxo, triângulo invertido). 78
25 Descrição da estrutura do pacote LR-FHSS quando a concatenação é aplicada com $c = 2$. O padrão hachurado indica que as mensagens foram codificadas em um único *payload*. 82
26 Estrutura do pacote LR-FHSS quando a retransmissão embutida é aplicada com $p = 2$. O padrão hachurado indica que as mensagens foram codificadas em um único *payload*. 82
27 Descrição da estrutura do pacote LR-FHSS quando a replicação é aplicada com $n = 2$. 83
28 Probabilidade média de sucesso, considerando o melhor valor de parâmetro para cada técnica. Os resultados para combinações das três técnicas também são apresentados. Por exemplo, $(c = 2, p = 2)$ significa que duas mensagens são concatenadas em um único pacote e que as mensagens concatenadas são retransmitidas utilizando a técnica de retransmissão embutida em um pacote subsequente. 84
29 CF-mMIMO em um ambiente com múltiplas CPUs. Esta figura ilustra o agrupamento virtual formado pela seleção de APs por cada UE e o serviço compartilhado fornecido por diferentes APs controlados por diferentes CPUs. 87
30 Função de distribuição acumulada empírica (CDF) da eficiência espectral para diferentes métodos de DCC no ambiente simulado. Entre as configurações testadas, o melhor desempenho foi alcançado com $\alpha = 0{,}7$. 91
31 Representação do sinal transmitido num enlace com linha de visada direta (LoS). 96
32 Comparação da magnitude da matriz de canal H (LoS vs NLoS) 98
33 Mapas de fase da matriz de canal $\mathbf{H}$ para os cenários LoS e NLoS 99
34 Diagrama de blocos do *dataset* Raymobtime tomado de [2]. 101
35 Acurácia da rede WiSARD em função do tamanho de memória, para diferentes tipos de entrada e cenários LOS/NLOS. 103
36 Acurácia do modelo WiSARD com inversão dos conjuntos de treinamento e teste (cross-domain), considerando separação entre cenários LOS, NLOS e ALL. . . . 104
37 Comparação de desempenho dos modelos de ML com e sem inversão dos conjuntos de treinamento e teste *(cross-domain)* usando dados de coordenadas. . . . 105
38 comparação de desempenho dos modelos de ML com e sem inversão dos conjuntos de treinamento e teste *(cross-domain)* usando dados de LiDAR. 106
39 Comparação de desempenho dos modelos de ML com e sem inversão dos conjuntos de treinamento e teste (cross-domain) usando dados de LiDAR + coordenadas. 107
40 Sistema *cell-free* com TDD dinâmico. 111
41 CDF do atraso de pacote no UL. 116

42 CDF do atraso de pacote no DL. 117
43 Arquitetura D-MIMO com D-TDD, na qual dois *clusters* disjuntos operando em modos de transmissão distintos sofrem interferência cruzada (CLI). 120
44 CDFs da SE para diferentes valores do parâmetro α do FPC. 125
45 CDFs da EE para diferentes valores do parâmetro α do FPC. 126
46 CDF da SE dos UEs em UL e DL variando ρ_{UL}. 127
47 CDF da EE dos UEs em UL e DL variando ρ_{UL}. 128
48 Modelo de um sistema OFDMA-MIMO. 141
49 Modelo de transmissor para NOMA-OFDMA-MIMO. 142
50 Modelo de receptor para NOMA-OFDMA-MIMO. 143
51 Grid de recursos OFDM para alocação dos *streams* MIMO e NOMA 146
52 Comparação de desempenho do NOMA para diferentes valores de α e β. 147
53 Comparação entre MIMO e NOMA para taxa de codificação 1/2 e modelo de canal CDL-A. 148
54 Comparação entre MIMO e NOMA para taxa de codificação 1/2 e modelo de canal CDL-C. 149
55 Comparação entre MIMO e NOMA para taxa de codificação 1/2 e modelo de canal CDL-E. 150

Lista de Tabelas

1 Parâmetros dos sistemas FTN-GFDM e GFDM. Modulação BPSK, filtro RRC, $\mathcal{P} = 4$, e $\mathcal{S} = 5$. 17
2 Modelos de canal para avaliação dos sistemas FTN-GFDM e GFDM. 18
3 SNRW da ED no i-ésimo SU, γ_{w_i}, no NCSS para as quatro combinações envolvendo os modelos tradicional e novo de fonte e intervalo de NU. 32
4 SNRW da ED no i-ésimo SU, γ_{w_i}, no CSS via HD sob a regra k-em-M para as quatro combinações envolvendo os modelos tradicional e novo de fonte e intervalo de NU. 37
5 ANRW da ED em cada SU, γ_{w}, no CSS via SD e cômputo tradicional de T em (32) para as quatro combinações envolvendo as fontes e intervalos tradicional e novo de NU. 41
6 SNRW da ED referente ao i-ésimo SU, γ_{w}, no CSS via SD com o novo cômputo da estatística de teste em (38) para as quatro combinações envolvendo as fontes e intervalos tradicional e novo de NU. 46
7 SNRWs, γ_{w}, em dB, e AUCs para a ED no NCSS, com SNR média de $\bar{\gamma} = -12$ dB e para as quatro combinações entre as fontes e intervalos tradicional e novo de NU, com $\rho = \{1{,}06; 1{,}09; 1{,}12; 1{,}15\}$ para o intervalo tradicional e $\rho = \{0{,}06; 0{,}09; 0{,}12; 0{,}15\}$ para o novo intervalo de NU. 48
8 SNRWs, γ_{w}, em dB, e AUCs para a ED no NCSS, CSS via SD com $M = 5$ SU para os cômputos tradicional (32) e novo (38) da estatística de teste e SNRWs e AUCs no CSS via HD sob as regras OR ($k = 1$), MAJ ($k = \lfloor M/2 + 1 \rfloor = 3$) e AND ($k = M = 5$), com fatores de NU $\{\rho_1; \ldots; \rho_M\} = \{0{,}02; 0{,}015; 0{,}012; 0{,}01; 0{,}0075\}$ e $\{\rho_1; \ldots; \rho_M\} = \{1{,}02; 1{,}015; 1{,}012; 1{,}01; 1{,}0075\}$ conforme as quatro combinações entre as fontes e intervalos tradicional e novo de NU, e SNR média $\{\bar{\gamma}_1; \ldots; \bar{\gamma}_M\} = \{-13; -14; -15; -16; -17\}$ dB. 50
9 Estatísticas dos *dataset* s008 e s009. 101
10 Resumo comparativo da acurácia dos modelos nas configurações a (s008→s009) e b (s009→s008), considerando diferentes tipos de entrada e top-k. 109
11 Parâmetros de simulação das estratégias de agrupamento de usuários. 116
12 Parâmetros de simulação das estratégias de alocação de potência 125
13 Modelos de canal CDL do 3GPP TR 38.901. 144

Acrônimos

1G Primeira Geração de Rede Móvel Celular
3GPP *3rd Generation Partnership Project*
5G Quinta Geração de Rede Móvel Celular
6G Sexta Geração de Rede Móvel Celular
AND *AND Decision Fusion Rule*
AoI *Age of Information*
AP *Access Point*
AUC *Area Under the Curve*
AWGN *Additive White Gaussian Noise*
BD-RIS *Beyond-Diagonal RIS*
BER *Bit Error Rate*
BS *Base Station*
CDF *Cumulative Distribution Function*
CFD *Cyclostationary Feature Detection*
CLI *Cross Link Interference*
CP *Cyclic Prefix*
CPU *Central Processing Unit*
CR *Cognitive Radio*
CSI *Channel State Information*
CSS *Cooperative Spectrum Sensing*
CTP *Controlled Transmit Power*
DCC *Dynamic Cooperation Clustering*
DL *Downlink*
D-MIMO *Distributed Multi-input Multi-output*
DPD *Digital Pre-Distortion*
DSA *Dynamic Spectrum Access*
D-TDD *Dynamic TDD*
DtS *Direct-to-Satellite*
ED *Energy Detection*
EE *Energy Efficiency*
FPC *Fractional Power Control*
FTP *Fixed Transmit Power*
GFDM *Generalized Frequency Division Multiplexing*
GFDMA *Generalized Frequency Division Multiple Access*

GPON *Gigabit Passive Optical Network*

HD *Hard-Decision*

HPA *High Power Amplifier*

IAI *Inter-AP Interference*

ICI *Intercarrier Interference*

IFFT *Inverse Fast Fourier Transform*

IoE *Internet of Everything*

IoT *Internet of Things*

ISI *Intersymbol Interference*

IUI *Inter-User Interference*

KPI *Key Performance Indicator*

LEO *Low Earth Orbit*

LoRa *Long Range*

LoRaWAN *Long Range Wide Area Network*

LOS *Line of Sight*

LPWAN *Low Power Wide Area Network*

LR-FHSS *Long Range - Frequency Hopping Spread Spectrum*

LSF *Large Scale Fading*

MAJ *Majority Decision Fusion Rule*

MCSA *Memoryless Continuous Search Algorithm*

MF *Matched Filter*

MFD *Matched Filter Detection*

MIMO *Multiple-Input Multiple-Output*

MP *Max Power*

MR *Maximum Ratio*

MRC *Maximal Ratio Combining*

MRT *Maximal Ratio Transmission*

MTC *Machine Type Communications*

M-TSA *Multiple Age Threshold Slotted ALOHA*

MZM *Mach-Zehnder Modulator*

NCSS *Non-Cooperative Spectrum Sensing*

NLOS *Non-Line of Sight*

NOMA *Non-Orthogonal Multiple Access*

NTN *Non-Terrestrial Networks*

NU *Noise Uncertainty*

OFDM *Orthogonal Frequency Division Multiplexing*

OMA *Orthogonal Multiple Access*

OOBE *Out-of-Band Emissions*

OR *OR Decision Fusion Rule*

PA *Power Amplifier*

PAPR *Peak-to-Average Power Ratio*

PD *Photo Detector*

PDF *Probability Density Function*

PMF *Probability Mass Function*

PSO *Particle Swarm Optimization*

PTS *Partial Transmit Sequence*

PU *Primary User*

QAM *Quadrature Amplitude Modulation*

QoS *Quality of Service*

RF Radiofrequência

RIS *Reconfigurable Intelligent Surfaces*

ROC *Receiver Operating Characteristic*

RRM *Radio Resource Management*

SD *Soft-Decision*

SE *Spectral Efficiency*

SIC *Successive Interference Cancellation*

SINR *Signal to Interference plus Noise Ratio*

SNR *Signal-to-Noise Ratio*

SNRW *Signal-to-Noise Ratio Wall*

SS *Spectrum Sensing*

SSF *Small Scale Fading*

S-TDD *Static TDD*

SU *Secundary User*

TDD *Time Division Duplex*

TSA *Age-Threshold Slotted ALOHA*

TV *Television*

TVWS *TV White Space*

UE *User Equipment*

UHF *Ultra High Frequency*

UIT União Internacional de Telecomunicações

UL *Uplink*

ULA *Uniform Linear Array*

VHF *Very High Frequency*

VR *Virtual Reality*

xG Próxima Geração de Rede Móvel Celular

Índice

1 Introdução **1**

2 Técnica de Redução de PAPR para Sistema de Comunicações Móveis com Base em Aprendizagem de Padrões **5**
2.1 Introdução . . . 5
2.2 Solução Proposta . . . 5
2.3 Resultados . . . 6
2.4 Conclusão . . . 7

3 Sistema de Comunicações Sem Fio NOMA Assistido por *Beyond-Diagonal* RIS **8**
3.1 Introdução . . . 8
3.2 Modelo do Sistema . . . 9
3.3 Solução Proposta . . . 10
3.4 Resultados Obtidos . . . 11
3.5 Conclusões . . . 12

4 Integração do FTN-GFDMA com A-RoF: Desafios e Benefícios para Comunicações em Áreas Remotas **14**
4.1 Introdução . . . 14
4.2 Solução Proposta . . . 14
4.3 Resultados Obtidos . . . 17
4.4 Conclusão . . . 20

5 Sensoriamento Espectral para Redes 6G **21**
5.1 Fundamentos sobre sensoriamento espectral . . . 22
5.1.1 Comparação entre o Modelo Tradicional e o Novo Modelo Proposto para Incerteza de Ruído . . . 24
5.2 Detecção de energia . . . 26
5.2.1 Sensoriamento não cooperativo . . . 26
5.2.2 Sensoriamento cooperativo . . . 26
5.2.2.1 Fusão de decisão abrupta . . . 26
5.2.2.2 Fusão de decisão suave . . . 27
5.3 Modelamento da incerteza de ruído . . . 28
5.4 Detecção de energia sob incerteza de ruído . . . 29
5.4.1 Sensoriamento não cooperativo . . . 29
5.4.1.1 Fonte tradicional combinada com intervalo tradicional/novo de NU . . . 29
5.4.1.2 Nova fonte combinada com intervalo tradicional/novo de NU . . . 30
5.4.2 Sensoriamento cooperativo . . . 31
5.4.2.1 Fusão de decisão abrupta . . . 31
5.4.2.2 Fusão de decisão suave . . . 37
5.5 Resultados Numéricos . . . 46
5.6 Conclusões . . . 50

6 Análise do Slotted ALOHA com Múltiplos *Age Thresholds* 52
6.1 Introdução . . . 52
6.2 Modelo do Sistema . . . 53
6.3 Resultados . . . 53
6.4 Conclusão . . . 56

7 Métricas para Comunicação Orientada a Tarefa 58
7.1 Introdução . . . 58
7.2 Comunicação orientada a tarefas . . . 58
7.3 A Métrica Idade da Informação e suas variantes . . . 59
7.4 Valor da Informação . . . 61
7.5 O problema do rastreamento remoto . . . 63
7.6 Conclusões . . . 64

8 Sistemas de Comunicação com Abordagens Semânticas 65
8.1 Introdução . . . 65
8.2 Propostas de Sistemas de Comunicação Baseadas em Semântica . . . 66
8.3 Conclusão . . . 70

9 Modelos de Fontes Semânticas para Relações Taxa-distorção Não-separáveis 72
9.1 Introdução . . . 72
9.2 Formulação do Problema . . . 73
9.3 Conclusão . . . 74

10 Estratégias de Acesso Não-Ortogonal para IoT Direto ao Satélite 75
10.1 Introdução . . . 75
10.2 Desenvolvimento . . . 76
10.3 Conclusão . . . 79

11 Técnicas de Transmissão Alternativas para LR-FHSS 80
11.1 Introdução . . . 80
11.2 Desenvolvimento . . . 81
11.3 Conclusão . . . 84

12 Escalabilidade e Alocação de Recursos em Cell-free Massive MIMO usando múltiplas CPUs 85
12.1 Introdução . . . 85
12.2 Modelo do Sistema . . . 87
12.3 Método Proposto . . . 89
12.4 Resultados Parciais . . . 90
12.5 Conclusão . . . 92

13 Algoritmos de AI/ML para Seleção de Feixe na Transmissão em mmWaves 93
13.1 Introdução . . . 93
13.2 Fundamentação Teórica . . . 93
13.3 Método de ML proposto . . . 100
13.4 Análise do *Dataset* . . . 100
13.5 Metodologia proposta . . . 102

13.6 Resultados parciais . . . 103
13.7 Conclusão e próximos passos . . . 108

14 Agrupamento de Usuários em Redes *Cell-Free* com TDD Dinâmico 110
14.1 Introdução . . . 110
14.2 Modelo do sistema . . . 111
14.3 Configuração de quadro . . . 113
14.4 Algoritmos de agrupamento . . . 114
14.5 Resultados de simulação . . . 115
14.6 Conclusão . . . 118

15 Análise da Eficiência Energética na Alocação de Potência em Sistemas D-MIMO com TDD Dinâmico 119
15.1 Introdução . . . 119
15.2 Modelo do Sistema . . . 121
15.3 Algoritmo de Controle de Potência . . . 123
15.4 Resultados Numéricos . . . 124
15.5 Conclusão . . . 129

16 Métodos Baseados em NOMA 130
16.1 Introdução . . . 130
16.2 Métodos baseados no PD-NOMA . . . 130
16.3 Métodos baseados no CD-NOMA . . . 131
16.4 Estratégias de integração sistêmica do NOMA . . . 132
16.5 NOMA como uma tecnologia habilitadora para a IoT e o Agronegócio . . . 134
16.6 Conclusão e Direcionamentos . . . 137

17 Análise de Eficiência Espectral entre MIMO e NOMA no Domínio de Potência para Redes 6G OFDM 139
17.1 Introdução . . . 139
17.2 Modelo Sistêmico . . . 141
17.3 Resultados de Simulação . . . 144
17.4 Conclusão . . . 149

18 Conclusões Gerais 151

1 Introdução

Esse relatório apresenta resultados de estudos exploratórios sobre tecnologias inovadoras e habilitadoras das redes 6G, conduzidos durante o primeiro ano da Atividade 2.2, do Projeto Brasil 6G - Fase III. Os estudos concentraram-se em aspectos da rede de acesso, abrangendo as camadas física e de controle de acesso.

O relatório é composto por dezesseis seções (além das seções de Introdução e Conclusões Gerais), cada uma dedicada aos temas investigados pelas equipes das instituições participantes da Atividade 2.2: Instituto Nacional de Telecomunicações, Universidade Estadual de Campinas, Universidade Federal de Santa Catarina, Universidade Federal do Rio de Janeiro, Universidade Federal do Ceará e Fundação CPqD. A seguir, apresenta-se uma breve descrição do conteúdo de cada seção do relatório.

A **Seção 2** apresenta uma técnica de redução da *Peak-to-Average Power Ratio* (PAPR) em sistemas *Orthogonal Frequency Division Multiplexing* (OFDM), com foco na operação em espectros como os canais *TV White Space* (TVWS), onde a potência de transmissão é limitada. O método proposto, denominado *Memoryless Continuous Search Algorithm* (MCSA), utiliza uma abordagem de força bruta para selecionar aleatoriamente conjuntos de subportadoras piloto que produzam sinais com PAPR abaixo de um limiar desejado, sem exigir alterações no receptor. A seção inclui análises comparativas com o método *Partial Transmit Sequence* (PTS), evidenciando que o MCSA alcança reduções de PAPR mais expressivas, com leve degradação da *Bit Error Rate* (BER) e custo computacional controlável via parâmetro de iteração. Os resultados demonstram que o MCSA é uma alternativa eficaz para melhorar a linearidade do sistema e mitigar distorções não lineares, sendo promissor para comunicações móveis baseadas em OFDM com restrições de potência.

A futura Sexta Geração de Rede Móvel Celular (6G) exigirá links mais rápidos e confiáveis. Para isso, a combinação de *Reconfigurable Intelligent Surfaces* (RIS) com *Non-Orthogonal Multiple Access* (NOMA) surge como uma possibilidade interessante, conforme é discutido na **Seção 3**. Os autores mostram que, ao trocar a RIS tradicional por uma *Beyond-Diagonal RIS* (BD-RIS), ganha-se liberdade para ajustar fase e amplitude do sinal refletido, enquanto NOMA garante que dois usuários dividam o mesmo recurso de tempo-frequência. É então formulado um problema de maximização da soma de taxas, otimizando simultaneamente *beamforming* na estação-base, coeficientes da BD-RIS e alocação de potência, tudo resolvido via uma heurística de baixa complexidade baseada em *Particle Swarm Optimization* (PSO). Simulações indicam que o esquema proposto, empregando BD-RIS e NOMA, supera esquemas ortogonais em todas as arquiteturas, obtendo ganho de *throughput*. Também é mostrado que a configuração *group-connected* da BD-RIS, com grupos de quatro elementos, praticamente replica o desempenho totalmente conectado, reduzindo custo de controle e hardware e permitindo cortar até 20% das antenas na estação-base sem perda de taxa nos cenários investigados.

A **Seção 4** propõe a integração da técnica *Faster-than-Nyquist* (FTN)-*Generalized Frequency Division Multiple Access* (GFDMA) com sistemas *Analog Radio over Fiber* (A-RoF) como solução para ampliar a conectividade em áreas remotas, onde a infraestrutura tradicional é inviável. A proposta visa aumentar a eficiência espectral e a capacidade de transmissão, aproveitando enlaces ópticos de longo alcance e espectro não licenciado. O estudo inclui modelagem detalhada dos efeitos não lineares e dispersivos do canal óptico, bem como desafios como o alto PAPR e a complexidade do receptor. Resultados indicam ganhos expressivos em eficiência espectral, tornando a solução promissora para aplicações em redes 6G voltadas à inclusão digital.

A escassez espectral decorre da política tradicional de alocação fixa de faixas pelos órgãos

reguladores, tornando-se obsoleta frente ao aumento exponencial da demanda por conectividade. Paradoxalmente, grande parte do espectro permanece subutilizada, motivando o uso de políticas de *Dynamic Spectrum Access* (DSA) e o *Spectrum Sensing* (SS) via rádio cognitivo como soluções promissoras. A **Seção 5** detalha como o SS permite que usuários secundários detectem oportunidades de uso de canais ociosos sem interferir em usuários primários, abordando conceitos como *Signal-to-Noise Ratio Wall* (SNRW), incerteza de ruído *Noise Uncertainty* (NU) e estratégias cooperativas e não cooperativas para maximizar a eficiência espectral em redes 6G.

Em comunicações com restrições de latência, a idade da informação que chega no receptor, conhecida na literatura como AoI, é crucial. Por exemplo, uma informação correta mas significativamente atrasada pode não ter mais valor em alguns sistemas ciberfísicos avançados que devem operar na 6G. Neste sentido, a **Seção 6** propõe o método Multiple-Threshold Slotted ALOHA (M-TSA), com o intuito de diminuir o valor médio do AoI em uma rede operando sobre um protocolo ALOHA, típico de *Internet of Things* (IoT). Neste esquema, cada nó transmite com uma probabilidade que depende da faixa de AoI em que este se encontra. Simulações com 100 nós mostram que o esquema com dois limiares (2-TSA), portanto três faixas de AoI, mantém o AoI médio consideravelmente menor que o ALOHA clássico e lida melhor com congestionamento que o TSA de um único limiar, sem perda significativa de throughput. Os resultados positivos com dois limiares motivam a investigação da abordagem multilimiar na continuação deste projeto.

A **Seção 7** apresenta uma revisão das principais métricas associadas à comunicação orientada a tarefa, abordando tanto o nível semântico quanto o nível de efetividade da comunicação, conforme proposto por Shannon e Weaver. São discutidas métricas como Idade da Informação (AoI), Idade da Informação Incorreta (AoII), Urgência da Informação (UoI) e Custo do Erro de Atuação, além do conceito de Valor da Informação (VoI), em suas formas observada e esperada. A seção destaca as limitações das métricas convencionais e argumenta em favor de abordagens que priorizem a transmissão de informações mais relevantes para a realização de tarefas específicas. A aplicação prática considerada é o rastreamento remoto de processor.

A **Seção 8** explora sistemas de comunicação baseados em semântica, que visam transmitir o significado essencial das mensagens, em vez de sua representação bit a bit. São apresentadas arquiteturas baseadas em redes neurais, especialmente Transformers e modelos de linguagem pré-treinados, para codificação e decodificação semântica. São discutidas também estratégias de modulação adaptadas ao espaço semântico, como esquemas QAM não uniformes. Resultados preliminares mostram que os modelos propostos apresentam desempenho promissor em diferentes canais de comunicação, especialmente sob condições adversas. A pesquisa visa demonstrar a robustez e a aplicabilidade desses sistemas para futuras redes 6G.

A **Seção 9** apresenta uma formulação teórica para o estudo de relações taxa-distorção não-separáveis em fontes semânticas, onde o objetivo é preservar simultaneamente características sintáticas e semânticas da informação. O modelo considerado combina aspectos de transmissão indireta e codificação de múltiplas fontes, refletindo a necessidade de compressão orientada à tarefa em aplicações como diagnóstico médico ou interpretação de imagens com significado semântico. A análise é conduzida a partir de uma generalização do problema clássico de compressão com perdas, com restrições acopladas sobre diferentes métricas de distorção. São discutidas as condições de otimalidade para a taxa de compressão mínima sob tais restrições, incluindo formulações lagrangeanas e soluções iterativas para o caso discreto. Essa abordagem estabelece fundamentos teóricos relevantes para o projeto de sistemas de comunicação semântica sob métricas compostas.

A **Seção 10** apresenta estratégias de NOMA no domínio da potência aplicadas ao enlace de subida em redes IoT direto ao satélite *Direct-to-Satellite* (DtS)-IoT, com foco na superação das limitações de escalabilidade dos protocolos ALOHA em cenários com visibilidade intermitente e alta densidade de dispositivos. Considerando a tecnologia LoRaWAN e o conhecimento prévio da órbita de satélites *Low Earth Orbit* (LEO), são propostas duas abordagens: *Fixed Transmit Power* (FTP), baseada em instantes discretos de transmissão com potência fixa, e *Controlled Transmit Power* (CTP), que ajusta dinamicamente a potência ao longo da janela de visibilidade. Resultados demonstram que as estratégias propostas ampliam o *goodput* e a eficiência energética em comparação com soluções tradicionais, especialmente em redes densas, consolidando o uso de NOMA como alternativa viável para a conectividade massiva em cenários *Non-Terrestrial Networks* (NTN).

A **Seção 11** propõe e avalia três técnicas de transmissão para redes LR-FHSS voltadas a aplicações DtS-IoT: concatenação de mensagens, retransmissão embutida e replicação. As técnicas visam melhorar a escalabilidade e a confiabilidade da comunicação em cenários de alta densidade, sem demandar alterações na infraestrutura da rede. Simulações com o simulador LR-FHSS-Sim indicam que a concatenação apresenta melhor desempenho em redes densas, enquanto a retransmissão embutida equilibra bem redundância e carga da rede. Combinações entre as técnicas resultam em ganhos adicionais, sugerindo abordagens híbridas como promissoras para ambientes de IoT massivo.

A **Seção 12** investiga a escalabilidade de redes cell-free Massive MIMO em cenários com múltiplas CPUs distribuídas, visando reduzir a complexidade de processamento centralizado e melhorar a alocação de recursos. O modelo considera diversos pontos de acesso (APs) conectados a diferentes unidades de processamento central (CPUs), que cooperam para atender usuários distribuídos em uma área geográfica. A proposta é baseada em um esquema de agrupamento dinâmico que vincula cada usuário a uma única CPU, considerando métricas de força de canal e balanceamento de carga. A alocação de recursos é formulada como um problema de maximização da soma de taxas com restrições de potência, e a solução é obtida por uma heurística que combina seleção de usuários e programação convexa. Os resultados parciais indicam que o uso de múltiplas CPUs melhora significativamente a eficiência espectral e o balanceamento da rede em comparação ao modelo centralizado, com impacto positivo na escalabilidade da arquitetura cell-free em redes 6G.

A **Seção 13** aborda o uso de algoritmos de inteligência artificial (AI) e aprendizado de máquina (ML) para a seleção eficiente de feixes em sistemas de comunicação mmWave, uma etapa crítica para reduzir a latência de inicialização de enlace e maximizar a taxa de transmissão. A proposta fundamenta-se na previsão do melhor feixe com base em informações contextuais, como posição, orientação e características do ambiente. O estudo avalia diferentes modelos de ML supervisionado e redes neurais profundas (DNNs), utilizando dados reais coletados com sensores em cenários indoor. A análise considera dois conjuntos de dados e avalia o desempenho em diferentes configurações de entrada e valores de top-k accuracy. Os resultados mostram que modelos baseados em DNN apresentam alta acurácia na predição do feixe ideal, especialmente quando se usa a trajetória do usuário como entrada. A proposta visa acelerar o processo de beam alignment com baixa complexidade, contribuindo para aplicações móveis de alta velocidade nas futuras redes 6G.

A **Seção 14** investiga estratégias de agrupamento centradas no usuário para redes cell-free com *Dynamic TDD* (D-TDD), visando mitigar interferências e melhorar a qualidade de serviço em cenários com tráfego heterogêneo. O estudo propõe algoritmos de reagrupamento de *Access Point*s (APs) com base em critérios de desvanecimento de larga escala e *Signal-to-Noise*

Ratio (SNR), permitindo que múltiplos usuários na mesma direção de tráfego compartilhem APs. Simulações em ambiente *Multiple-Input Multiple-Output* (MIMO) cell-free com formação de feixe e tráfego assimétrico mostram que essas estratégias reduzem significativamente o atraso de pacotes, principalmente no enlace reverso, quando comparadas ao agrupamento com clusters disjuntos. Os resultados demonstram o potencial do reagrupamento adaptativo como solução escalável para redes densas com D-TDD.

Dentro das possíveis inovações propostas para a 6G temos a arquitetura *Distributed Multi-input Multi-output* (D-MIMO), onde dezenas de pontos de acesso cooperam para servir cada *User Equipment* (UE). Para maximizar o desempenho individual de acada UE, na literatura foi proposto o uso de *Time Division Duplex* (TDD) dinâmico. Porém, esta estratégia introduz forte interferência cruzada entre enlaces *Uplink* (UL) e *Downlink* (DL), tornando o controle de potência peça-chave para manter consumo baixo e qualidade alta. Nesta linha, a **Seção 15** modela detalhadamente o sistema, incluindo sinal desejado, interferências e métricas de *Spectral Efficiency* (SE) e *Energy Efficiency* (EE) em um cenário com 30 UEs e 120 APs. Dois esquemas são comparados: transmissão na potência máxima, aqui chamado de *Max Power* (MP), e o método clássico *Fractional Power Control* (FPC). Simulações mostram que o FPC reduz ligeiramente a SE, mas entrega ganhos consistentes de EE sobre o MP, especialmente quando a fração de usuários em uplink é alta. Além disso, o FPC mantém desempenho estável enquanto a proporção UL/DL varia, provando ser alternativa prática para redes sustentáveis sob interferência cruzada sem recorrer a métodos mais complexos.

A **Seção 16** expõe os fundamentos, avanços e desafios da tecnologia NOMA como solução promissora para comunicações massivas na Quinta Geração de Rede Móvel Celular (5G) e 6G, com foco em aplicações de IoT. São analisados métodos baseados em domínio de potência (PD-NOMA) e domínio de código (CD-NOMA), destacando técnicas de alocação de potência, estratégias de cancelamento de interferência (SIC) e códigos customizados para separação de usuários. A seção também discute limitações práticas da tecnologia, como complexidade computacional e latência, propondo estratégias de integração com MIMO, HARQ adaptativo e aprendizado de máquina para mitigar esses entraves. Por fim, são apresentados casos de uso em agricultura inteligente e retroespalhamento de baixo consumo, consolidando o NOMA como tecnologia habilitadora para cenários de alta densidade e dispositivos com restrições energéticas.

A **Seção 17** analisa comparativamente a eficiência espectral das técnicas MIMO (Multiple-Input Multiple-Output) e NOMA (Non-Orthogonal Multiple Access) aplicadas a sistemas OFDM para redes 6G. Por meio de simulações em canais com desvanecimento Rayleigh, o estudo avalia o desempenho dessas técnicas em diferentes regimes de SNR e configurações de antenas e alocação de potência. Os resultados mostram que, em baixas SNRs, o NOMA é mais eficiente, especialmente quando há grande diferença de potência entre os usuários, permitindo multiplexação no domínio de potência com ganhos notáveis. Em altas SNRs, o MIMO se destaca devido à sua capacidade de explorar diversidade espacial, resultando em maior capacidade espectral. A seção também destaca que a combinação de MIMO e NOMA pode ser vantajosa, aproveitando os pontos fortes de ambas as técnicas. Por fim, conclui-se que a escolha entre essas abordagens deve considerar as condições operacionais, os requisitos de equidade e a complexidade de implementação.

2 Técnica de Redução de PAPR para Sistema de Comunicações Móveis com Base em Aprendizagem de Padrões

Bianca S. de C. da Silva, Mariana B. de Mello, Luciano L. Mendes
Inatel

2.1 Introdução

O avanço dos sistemas de comunicação móvel aumentou a demanda por larguras de banda mais amplas, incentivando a utilização de espectros subutilizados, como o TVWS[3]. No Brasil, as regulamentações permitem o uso de canais de TV ociosos nas faixas *Very High Frequency* (VHF) e *Ultra High Frequency* (UHF), desde que não haja interferência nos usuários primários [4]. Entretanto, a potência de transmissão no TVWS é limitada a 1W, o que pode restringir a cobertura e comprometer a eficiência da rede — especialmente em sistemas baseados em OFDM, que apresentam elevados níveis de PAPR[5].

Define-se a PAPR como a razão entre a potência de pico e a potência média de um determinado sinal, deste modo, como o sinal OFDM é formado pela soma de K sinais independentes, resultantes da modulação de cada subportadora por um símbolo M-QAM aleatório, sua envoltória no domínio do tempo tende a se comportar como uma variável aleatória com distribuição Gaussiana, especialmente para um número elevado de subportadoras. Portanto, existe a possibilidade de que essas diversas componentes se somem em fase, resultando em picos instantâneos de amplitude e, consequentemente, em uma elevada PAPR.

Picos de potência no sinal podem levar o amplificador de potência do transmissor à saturação. Nessa condição, o amplificador limita o sinal de saída, fazendo com que sua resposta deixe de ser linear e introduza distorções não lineares no sinal transmitido. Essas distorções resultam em *Intercarrier Interference* (ICI) e interferência em canais adjacentes, devido ao aumento da *Out-of-Band Emissions* (OOBE).

Dessa forma, a redução da PAPR em sinais OFDM torna-se um tema de pesquisa de grande relevância, pois contribui para mitigar o crescimento da OOBE. Com esse objetivo, um algoritmo é proposto e detalhado no próximo tópico.

2.2 Solução Proposta

Nesta seção, é apresentado um algoritmo proposto para a redução de PAPR em sistemas OFDM.

O MCSA é uma abordagem de força bruta que explora aleatoriamente um grande número de combinações possíveis sem empregar estratégias de otimização inteligente. No contexto da redução da PAPR, o método proposto utiliza essa abordagem para selecionar aleatoriamente um conjunto de subportadoras piloto capaz de gerar um sinal com PAPR abaixo de um valor alvo. A PAPR de cada sinal OFDM, gerado com diferentes configurações de subportadoras piloto, é avaliado; caso ultrapasse o limite pré-estabelecido, um novo conjunto de pilotos é sorteado. Esse processo se repete até que a PAPR desejado seja atingido. Em [6] pode ser encontrado o fluxograma do algoritmo MCSA para a redução de PAPR. O método pode ser facilmente escalado para conjuntos maiores de subportadoras piloto.

Para cada vetor de dados $\mathbf{s}$, são definidos N_p subportadoras como pilotos, o que implica que a carga útil do sinal OFDM é composta por $N - N_p$ símbolos de dados. Inicialmente, os N_p pilotos são configurados com os valores $\pm\sqrt{E}$, onde E representa a energia média do vetor

s. A PAPR do sinal OFDM resultante é então comparado a um valor de referência $\mathcal{P}_{\max}$. Se a PAPR estiver abaixo desse limite, o sinal é transmitido. Caso contrário, um novo conjunto de pilotos é gerado aleatoriamente e incorporado ao vetor de dados. O processo de avaliação se repete até que se obtenha uma PAPR abaixo do valor alvo ou até que seja atingido um número máximo de tentativas N_t. Nesse último caso, o sinal com a menor PAPR encontrado é transmitido. No receptor, não são necessárias modificações, pois as subportadoras piloto não interferem nos procedimentos de detecção convencionais.

Vale destacar que, quando o valor alvo de PAPR é relativamente alto, o MCSA tende a encontrar rapidamente uma configuração satisfatória, exigindo poucas iterações. Por outro lado, se o alvo estiver vários decibéis abaixo da PAPR típico de um sinal OFDM convencional, o número de tentativas necessárias pode aumentar significativamente, elevando a complexidade computacional do sistema.

2.3 Resultados

Esta seção apresenta a avaliação do desempenho dos algoritmos propostos por meio de simulações, considerando as métricas de PAPR, BER e complexidade computacional em termos do número de *Inverse Fast Fourier Transform*s (IFFTs) necessárias para gerar um símbolo OFDM.

As simulações foram conduzidas considerando um sistema OFDM com modulação 16-*Quadrature Amplitude Modulation* (QAM) e 64 subportadoras. O MCSA é avaliado para dois limiares de PAPR, $\mathcal{P}$max = 6 dB e 5 dB, utilizando $\mathbf{N}_{\mathrm{t}} = 30$ e $\mathbf{N}_{\mathrm{t}} = 100$, respectivamente. Para este caso, duas subportadoras piloto foram alocadas nas posições $k = 1$ e $k = 62$. No total, são transmitidos 1.000.000 de símbolos OFDM. Para fins de comparação, utiliza-se o esquema PTS, amplamente empregado na literatura devido à sua complexidade computacional constante, igual a L, e à sua implementação relativamente simples.

A partir da Figura 1a, observa-se que o MCSA apresenta desempenho equivalente na redução de PAPR para os alvos considerados ($\mathcal{P}_{\max} = 5$ dB e 6 dB), superando o desempenho do PTS, que alcança uma redução de aproximadamente 1 dB. Assim, o algoritmo proposto demonstra maior eficácia na mitigação da PAPR.

Para analisar o impacto das técnicas de redução de PAPR sobre o desempenho em termos de BER, considerou-se um canal *Additive White Gaussian Noise* (AWGN), cujos resultados são apresentados na Figura 1b. Nota-se uma leve degradação no desempenho de BER para o MCSA, atribuída à inserção de subportadoras piloto, o que resulta em uma pequena perda na eficiência espectral. No entanto, essa penalização é marginal quando comparada ao benefício obtido na redução do PAPR.

A avaliação da complexidade computacional, em termos do número de IFFTs requeridos, é ilustrada nas Figuras 1c e 1d, referentes aos casos em que $\mathcal{P}_{\max} = 5$ dB e $\mathcal{P}_{\max} = 6$ dB, respectivamente. Utilizou-se a *Probability Mass Function* (PMF) do número de IFFTs computados para caracterizar a complexidade geral do esquema. Por sua vez, o número de IFFTs exigido pelo PTS é constante e igual a $L = 4$.

As Figuras 1d e 1c apresentam a PMF do número de IFFTs para os casos em que $\mathcal{P}_{\max} = 6$dB e $\mathcal{P}_{\max} = 5$dB, respectivamente. A Figura1d mostra que, embora o método MCSA apresenta maior complexidade em comparação ao PTS, essa comparação deve ser feita com cautela, uma vez que o algoritmo proposto atinge valores de PAPR significativamente menores. De forma semelhante, a Figura1c evidencia que, para $\mathcal{P}$max = 5 dB, o número de IFFTs necessários para alcançar a PAPR desejado aumenta de forma significativa em ambos os métodos, reforçando

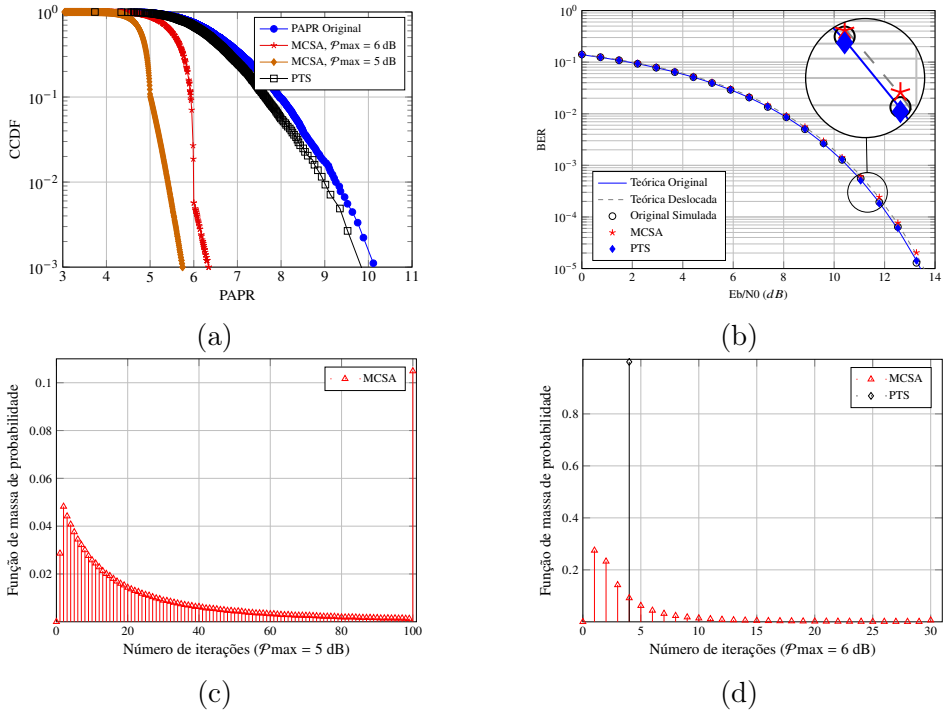


Figura 1: Análise de desempenho da PAPR (1a) e da BER (1b). Avaliação da complexidade para valores de $\mathcal{P}_{\max} = 5$ dB (1c) e $\mathcal{P}_{\max} = 6$ dB (1d).

a existência de um *trade-off* entre a complexidade computacional e a eficiência na redução da PAPR.

2.4 Conclusão

Com base nos resultados apresentados, conclui-se que o algoritmo MCSA é capaz de atingir reduções de PAPR superiores ao método tradicional PTS, ainda que à custa de maior complexidade computacional. No entanto, essa complexidade pode ser gerenciada por meio do parâmetro de controle $\mathbf{N}_{\mathrm{t}}$, permitindo um equilíbrio entre desempenho e custo computacional. Por fim, a leve degradação de desempenho em BER é compensada pelos ganhos substanciais na redução de PAPR, tornando o algoritmo proposto promissor para aplicações em sistemas OFDM.

3 Sistema de Comunicações Sem Fio NOMA Assistido por *Beyond-Diagonal* RIS

Victoria D. P. Souto e Richard Demo Souza
Inatel e UFSC

3.1 Introdução

Aplicações futuras como realidade estendida, robótica conectada, automação em larga escala, saúde inteligente, comunicação biomédica, interações cérebro-computador sem fio e a *Internet of Everything* (IoE) exigirão capacidades ainda maiores do que as obtidas com as Redes 5G, com taxas de dados mais altas, menor latência e confiabilidade ultra-alta [7]. Para alcançar esses requisitos é necessário o uso de novas tecnologias emergentes, como as RIS, capazes de melhorar a qualidade do sinal para usuários ocultos, contribuir para a supressão de interferências e aumentar a capacidade da rede [7]. No entanto, nem todas as RIS compartilham a mesma arquitetura, e seu desempenho varia dependendo de como seus componentes são conectados e dos modos de operação considerados. As RIS são classificadas em três categorias de acordo com a topologia do circuito entre seus elementos [8]: *Single-connected*, *group-connected* e *fully-connected*. A RIS convencional, também conhecida como *single-connected RIS*, apresenta uma arquitetura baseada em uma matriz de deslocamento de fase diagonal, onde cada elemento é conectado individualmente a uma carga de terra. Embora essa arquitetura simplifique a complexidade computacional, ela também limita sua flexibilidade, pois só pode controlar a fase dos sinais incidentes e refleti-los em um único plano, afetando o desempenho e a cobertura do sistema [8]. Para explorar plenamente as capacidades da RIS e superar as limitações da RIS convencional, os autores de [9] propuseram a BD-RIS, que introduz as arquiteturas totalmente conectada e conectada em grupo, cujo modelo matemático não se limita a uma matriz diagonal. Essas superfícies podem operar em diferentes modos (reflexivo, transmissivo e híbrido), proporcionando maior flexibilidade para manipular a fase e a amplitude das ondas incidentes, resultando em maior desempenho e cobertura do sistema [8].

Além disso, a capacidade das redes 6G de suportar um número massivo de dispositivos por metro cúbico, juntamente com sua latência ultrabaixa, alta confiabilidade, taxa de transferência e justiça, exige o desenvolvimento de técnicas avançadas de múltiplo acesso para usar os recursos do espectro de forma mais eficiente. Tecnologias como o NOMA permitem que múltiplos usuários compartilhem o mesmo canal por meio de multiplexação no domínio da potência, alcançando melhores resultados em termos de latência, mobilidade, justiça, eficiência espectral e conectividade massiva, em comparação com as técnicas tradicionais *Orthogonal Multiple Access* (OMA) [10]. Isso é alcançado por meio do uso de codificação externa e *Successive Interference Cancellation* (SIC) [10]. Portanto, pode-se verificar, que o estudo de sistemas de comunicação sem fio assistidos por BD-RIS e NOMA é essencial para alcançar os rigorosos requisitos da 6G.

Portanto, neste trabalho, investigamos um sistema de comunicação sem fio NOMA assistido por BD-RIS. Especificamente, o principal objetivo deste trabalho é maximizar a taxa de soma total alcançável, otimizando conjuntamente o *beamforming* na *Base Station* (BS), a fase e a amplitude nos elementos da BD-RIS e a potência alocada para cada usuário. Para resolver esse problema de otimização, diferente de [11], propomos uma solução de baixa complexidade

[0]O conteúdo desta seção é um resumo do artigo aceito para publicação no PIERS 2025: Duarte, D. L. F., Souto, V. D. P., and Souza, R. D., "*NOMA Wireless Communications System Assisted by Beyond-Diagonal RIS*", PIERS 2025, Abu Dhabi.

baseada na técnica PSO, que não requer estimação explícita do *Channel State Information* (CSI). Além disso, o desempenho da solução proposta é investigado considerando diferentes arquiteturas de BD-RIS (*single-connected*, *group-connected* e *fully-connected*), e os esquemas de transmissão OMA e NOMA são avaliados. As principais contribuições são:

- Propomos uma solução de baixa complexidade baseada em PSO para maximizar a taxa de soma alcançável de um sistema sem fio NOMA assistido por BD-RIS.
- Demonstramos que o sistema NOMA assistido por BD-RIS supera os sistemas OMA assistidos por BD-RIS.
- Demonstramos que as novas arquiteturas de BD-RIS (*group-connected* e *fully-connected*) superam a RIS tradicional (*BD-RIS single-connected*) em sistemas NOMA e OMA para diferentes cenários.

3.2 Modelo do Sistema

Neste trabalho foi considerado um sistema de comunicação sem fio de *downlink* composto por uma BS equipada com uma *Uniform Linear Array* (ULA) com N antenas transmissoras, uma BD-RIS com M elementos refletores e dois usuários, cada um com uma única antena. Além disso, consideramos a técnica NOMA, então, o sinal recebido pelo k-ésimo usuário é dado por

$$y_{\mathrm{k}}^{\mathrm{NOMA}} = C_k \times \left(\sum_{j \in \{1,2\}} \sqrt{\rho_j} \mathbf{w}_j s_j \right) + n_{\mathrm{k}}, \tag{1}$$

onde $k, j \in \{1, 2\}$, ρ_j denota a potência de transmissão atribuída ao j-ésimo usuário sob um esquema de transmissão NOMA, de modo que $\sum_j \rho_j = P_{\mathrm{T}}$ onde P_{T} denota a potência total de transmissão. Além disso, $\mathbf{w}_j \in \mathbb{C}^{N \times 1}$ é o vetor de *beamforming* da BS para o j-ésimo usuário, s_j denota o sinal transmitido para o j-ésimo usuário e é modelado como uma variável aleatória independente com média zero e variância unitária. Além disso, o ruído branco gaussiano no k-ésimo usuário é representado por n_{k} com variância σ^2. Para finalizar, $C_k = \mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \mathbf{\Phi}_k \mathbf{G} + \mathbf{h}_{\mathrm{d},k}$, onde $\mathbf{G} \in \mathbb{C}^{M \times N}$ representa a matriz de canal entre a BS e o BD-RIS, $\mathbf{h}_{\mathrm{r},k}^{\mathrm{H}} \in \mathbb{C}^{1 \times M}$ denota o vetor de canal entre o BD-RIS e o k-ésimo usuário, $\mathbf{h}_{\mathrm{d},k} \in \mathbb{C}^{1 \times N}$ representa o vetor de canal direto entre a BS e o k-ésimo usuário, e $\mathbf{\Phi}_k \in \mathbb{C}^{M \times M}$ denota a matriz de coeficientes de reflexão do BD-RIS.

Com base no ganho do canal, definido como $G_k = |C_k \mathbf{w}_k|^2$, é possível calcular a taxa alcançável de cada usuário. Então, consideramos um sistema NOMA, onde os usuários compartilham os mesmos recursos de tempo e frequência de forma não ortogonal, diferenciando-se no domínio da potência. O usuário com melhores condições de canal deve decodificar e cancelar o sinal dos usuários com canais mais fracos [12]. Portanto, para uma representação geral, assume-se que $G_1 > G_2$, então a taxa alcançável para os usuários é dada por

$$R_1^{\mathrm{NOMA}} = \log_2 \left(1 + \frac{\rho_1 G_1}{\rho_2 G_2 + \sigma^2} \right). \tag{2}$$

$$R_2^{\mathrm{NOMA}} = \log_2 \left(1 + \frac{\rho_2 G_2}{\sigma^2} \right). \tag{3}$$

O principal objetivo é maximizar a taxa de soma alcançável $\left(R_{\mathrm{T}} = \sum_{k\in\{1,2\}} R_k^{\mathrm{NOMA}}\right)$ otimizando conjuntamente o *beamforming* na BS, os coeficientes de reflexão na BD-RIS e a alocação de potência para cada dispositivo, considerando uma restrição mínima de taxa. Esse problema de otimização pode ser expresso da seguinte forma

$$\begin{aligned} \underset{\mathbf{w}_k,\boldsymbol{\Phi}_k,\rho_k}{\text{Maximizar}} \quad & \sum_k R_k \\ \text{Sujeito a} \quad & R_k \geq \gamma_k \ \forall k \in \{1{,}2\}, \\ & \|\mathbf{w}_k\|^2 = 1 \ \forall k \ , \\ & \sum_k \rho_k = P_{\mathrm{T}} \ \forall k \ , \\ & \boldsymbol{\Phi}_g = \boldsymbol{\Phi}_g^T, \\ & \boldsymbol{\Phi}_g \boldsymbol{\Phi}_g^H = I_{M_{\mathrm{G}}} \ \forall g \in \{1,\ldots,N_{\mathrm{G}}\}, \end{aligned} \tag{4}$$

em que γ_k denota a restrição mínima de taxa alcançável do k-ésimo usuário, $||\mathbf{w}_k||^2 = 1$ denota a restrição de *beamforming* analógico, e $\sum_k \rho_k = P_T$ é a restrição de alocação de potência para o sistema NOMA. Além disso, a matriz de espalhamento da BD-RIS $\boldsymbol{\Phi}_g$, é representada como uma matriz complexa, simétrica e ortogonal que corresponde a uma arquitetura conectada em grupo. Essa configuração permite que N_g varie, considerando tanto arquiteturas clássicas quanto totalmente conectadas. É possível verificar que (4) é um problema não convexo e não pode ser resolvido por métodos de otimização padrão. Por esse motivo, e também motivados pelas vantagens oferecidas pela técnica PSO, propomos uma solução baseada nessa abordagem.

3.3 Solução Proposta

Motivados por sua flexibilidade e baixa complexidade computacional do PSO, uma nova abordagem baseada em PSO é proposta para resolver (4), cujos passos são detalhados abaixo.

1. Inicialize L partículas com valores aleatórios correspondentes ao vetor de *beamforming* no BS, a matriz de coeficientes de reflexão da BD-RIS e os coeficientes de alocação de potência $\left([\mathbf{w}_k]_i, [\boldsymbol{\Phi}_k]_i, [\rho_k]_i \ \forall k \in \{1,2\}\right)$.

2. Para cada partícula, se R_k^{NOMA} exceder o limite mínimo, o fitness é calculado com base em (4).

3. Atualize p_{best} e g_{best} com base no fitness calculado na etapa anterior.

4. Atualize a velocidade e a posição de cada partícula com base em

$$\mathbf{v}_i^{q+1} = w\mathbf{v}_i^q + a_1 r_1 (p_{\mathrm{best}_i}^q - \mathbf{x}_i^q) + a_2 r_2 (g_{\mathrm{best}}^q - \mathbf{x}_i^q), \tag{5}$$

$$\mathbf{x}_i^{q+1} = \mathbf{x}_i^q + \mathbf{v}_i^{q+1}, \tag{6}$$

onde $q \in \{1,\ldots,N_{it}\}$ denota a iteração atual, N_{it} é o número total de iterações, $i \in \{1,\ldots,L\}$ representa o índice da partícula, L denota o número total de partículas, a_1 e a_2 são os coeficientes de aceleração cognitiva e social, r_1 e r_2 são dois valores aleatórios uniformemente distribuídos entre 0 e 1, $\mathbf{v}_i^q$ e $\mathbf{x}_i^q$ representam a velocidade e a posição da partícula i na iteração q, respectivamente.

5. Verifique se a condição de parada foi atingida. Se sim, selecione e retorne a melhor partícula: $(\mathbf{w}_{\text{best}}, \mathbf{\Phi}_{\text{best}}, \boldsymbol{\rho}_{\text{best}})$. Caso contrário, retorne ao Passo 2.

Finalmente, é importante destacar que, em nosso método proposto, os pares de vetores de *beamforming* na BS e na BD-RIS a serem testados são definidos na BS, enquanto o *beamforming* da BD-RIS é enviado pela BS para o controlador. Em seguida, para cada par de beamforming, o feedback da *Signal to Interference plus Noise Ratio* (SINR) nos usuários é recebido na BS/BD-RIS, e o par de *beamforming* atual é avaliado com base em $\sum_{k\in\{1,\,2\}} R_k$. Esse processo é realizado para todos os pares. Observe que os canais BS/BD-RIS/Usuários não são explicitamente estimados.

3.4 Resultados Obtidos

A avaliação da abordagem proposta considera um cenário onde a BS está localizada em $(x_{\text{BS}}, y_{\text{BS}}) = (0, 0)$ m, a BD-RIS está localizado em $(x_{\text{RIS}}, y_{\text{RIS}}) = (70, 0)$ m, a posição horizontal dos dispositivos (x_{UE1} e x_{UE2}) é uniformemente distribuída e sua posição vertical é fixa em 2 m, ou seja, $y_{\text{UE1}} = y_{\text{UE2}} = 2$ m. A Figura 2 ilustra a taxa de soma alcançável versus o número de elementos na BD-RIS (M) para sistemas NOMA e OMA considerando $N = 2$, $(x_{\text{UE1}}, y_{\text{UE1}}) = (35, 2)$ m e $(x_{\text{UE1}}, y_{\text{UE1}}) = (45, 2)$ m. A partir dos resultados, pode-se observar que, mesmo quando os dispositivos estão próximos uns dos outros e longe da BS e da BD-RIS, o sistema NOMA supera o sistema OMA em todas as arquiteturas de RIS. Especificamente, a abordagem NOMA aumenta a taxa de soma alcançável em até 6% em comparação com os sistemas OMA. Além disso, pode-se verificar que é possível alcançar um desempenho próximo ao da BD-RIS totalmente conectada considerando um tamanho de grupo de 4 ($M_{\text{G}} = 4$), o que reduz consideravelmente a complexidade de *hardware* e o custo computacional para otimizar os coeficientes dos elementos da BD-RIS.

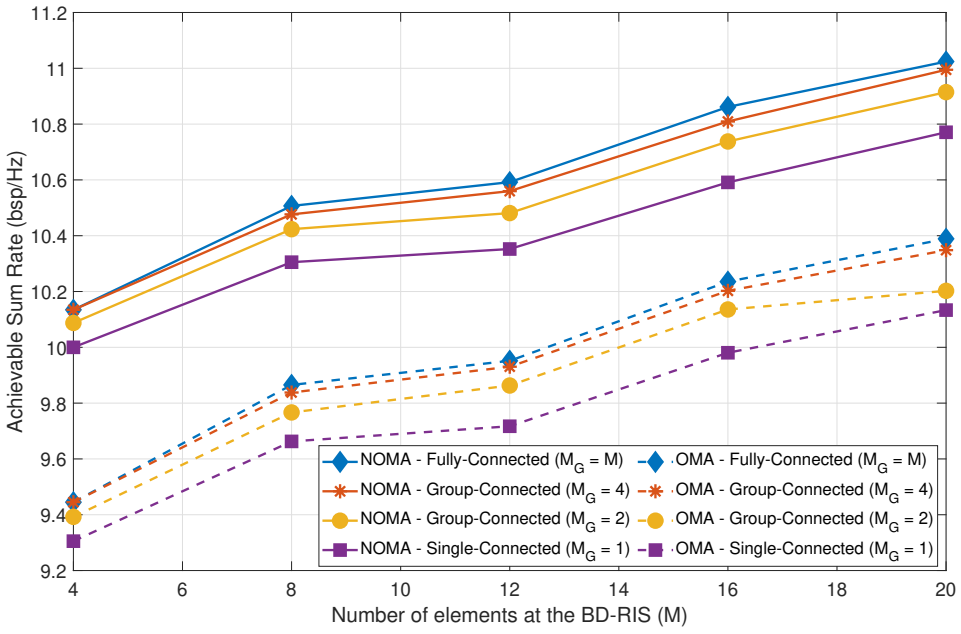


Figura 2: Taxa de soma alcançável versus o número de elementos M da BD-RIS, para NOMA e OMA.

Além disso, para avaliar a influência do número de antenas na BS (N), a Figura 3 mostra a taxa de soma alcançável versus N para $M = 20$, $(x_{\mathrm{UE1}}, y_{\mathrm{UE1}}) = (35, 2)$ m e $(x_{\mathrm{UE1}}, y_{\mathrm{UE1}}) = (45, 2)$ m. Os resultados são consistentes com os da Figura 2, indicando que, mesmo quando os dispositivos estão próximos uns dos outros e longe tanto da BS quanto da BD-RIS, o sistema NOMA supera o sistema OMA em todas as arquiteturas de BD-RIS. Além disso, os resultados demonstram que, à medida que N aumenta, o desempenho da arquitetura *single-connected* se aproxima do desempenho da arquitetura totalmente conectada para ambos os sistemas NOMA e OMA. Esses achados destacam a importância da BD-RIS em futuros sistemas de comunicação sem fio, pois sua implantação permite uma redução de até 20% no número de antenas na BS, mantendo a mesma taxa de soma alcançável. Além disso, para N grande, pode-se observar que um desempenho próximo ao da arquitetura totalmente conectada pode ser alcançado mesmo com um pequeno M_{G}, já que o design do *beamforming* na BS tem um impacto mais significativo no desempenho geral do sistema.

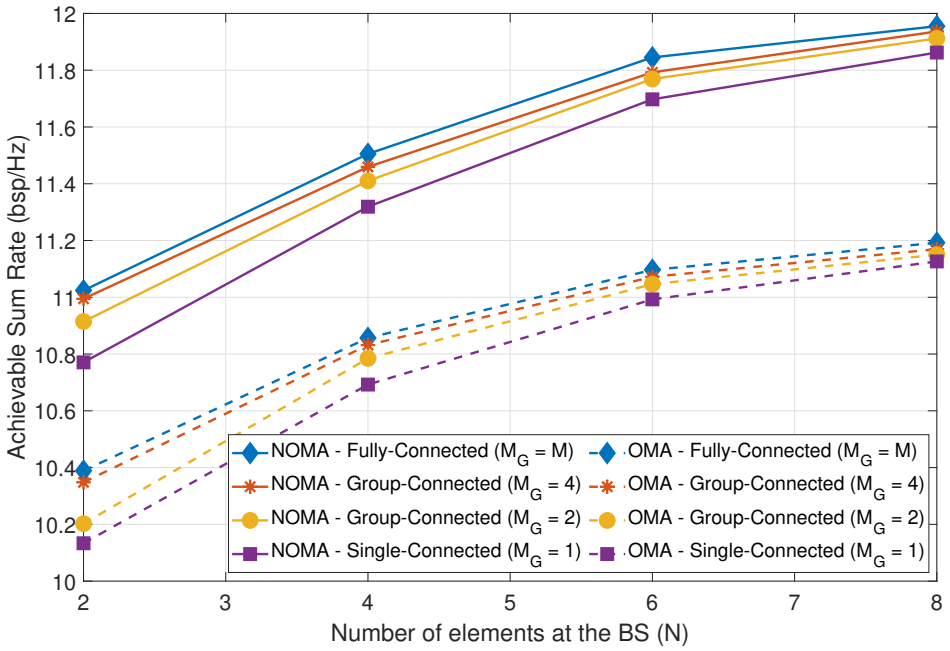


Figura 3: Taxa de soma alcançável versus o número de antenas N da BS, para NOMA e OMA.

3.5 Conclusões

Neste trabalho, analisamos o desempenho de um sistema sem fio NOMA de *downlink* assistido por BD-RIS. Especificamente, propomos uma nova solução baseada em PSO para maximizar a taxa de soma alcançável, otimizando conjuntamente o *beamforming* na BS, a matriz de coeficientes de reflexão na BD-RIS e a alocação de potência para os dispositivos, considerando arquiteturas clássicas, totalmente conectada e conectada em grupo. Foram realizadas extensivas simulações computacionais para comparar a taxa de soma alcançável do sistema NOMA assistido por BD-RIS proposto com a do sistema OMA assistido por BD-RIS. Os resultados demonstraram que o sistema NOMA supera consistentemente o sistema OMA em todas as arquiteturas de BD-RIS. Além disso, os achados destacam que o desempenho do sistema pode ser significativamente melhorado mesmo com uma arquitetura de BD-RIS simples e econômica, destacando sua potencial aplicabilidade nas futuras redes sem fio. Como direção para pesquisas

futuras, planejamos investigar o desempenho de sistemas sem fio assistidos por BD-RIS considerando deslocamentos de fase discretos nos elementos da BD-RIS.

4 Integração do FTN-GFDMA com A-RoF: Desafios e Benefícios para Comunicações em Áreas Remotas

Mariana B. de Mello, Luiz A. M. Pereira e Luciano L. Mendes
Inatel

4.1 Introdução

A evolução das redes móveis rumo à era 6G requer a integração de tecnologias avançadas, como comunicação, computação móvel, sensoriamento e *Artificial Intelligence* (AI), para viabilizar aplicações emergentes, tais como IoT, *Virtual Reality* (VR), transporte inteligente e comunicações ubíquas. No entanto, aproximadamente 30% da população mundial ainda carece de acesso a serviços básicos de comunicação, especialmente em áreas remotas e rurais, onde as infraestruturas legadas se mostram insuficientes. Para assegurar uma cobertura global e mitigar a exclusão digital, torna-se imprescindível o desenvolvimento de soluções tecnológicas que sejam simultaneamente eficientes e economicamente viáveis.

Nesse cenário, o principal desafio para a oferta de conectividade em regiões remotas reside no elevado custo associado à implantação e manutenção de infraestruturas complexas, especialmente em locais com baixa densidade populacional e limitado potencial econômico. Além disso, as frequências mais altas previstas para o 6G reduzem significativamente a área de cobertura de cada célula, inviabilizando em regiões remotas a aplicação de soluções tradicionais, como MIMO massivo e mmWave.

Diante desse contexto, este trabalho propõe e analisa a integração da técnica FTN-GFDMA com sistemas A-RoF como uma solução promissora para comunicações em áreas remotas, no âmbito de *Enhanced Remote Areas Communication* (eRAC). A proposta visa aprimorar a eficiência espectral e a capacidade de sistemas multiusuário, valendo-se da arquitetura *Centralized Radio Access Network* (C-RAN), de enlaces ópticos e do uso de espectro não licenciado (TVWS), configurando uma alternativa de baixo custo para ampliar a cobertura e melhorar a qualidade de serviço nessas regiões. O estudo inclui a modelagem do sistema em banda base, levando em conta as não-linearidades e os efeitos ópticos, e contribuindo, assim, para o avanço da integração óptico-*wireless* no contexto das redes 6G.

4.2 Solução Proposta

A proposta de integração do FTN-GFDMA com sistemas A-RoF para aplicações em eRAC é viabilizada por meio de uma modelagem precisa do sistema em banda base, contemplando as especificidades da alocação de recursos, a geração do sinal, bem como os efeitos não-lineares e lineares introduzidos ao longo do enlace óptico e do canal sem fio.

No esquema FTN-GFDMA, os recursos no domínio tempo-frequência são alocados por meio de uma matriz de símbolos de dados $\mathbf{S} \in \mathbb{C}^{K \times M}$, onde K representa o número de subportadoras e M o número de subsímbolos transmitidos por subportadora. Cada símbolo $s_{k,m}$ é transmitido na subportadora $k \in \{0, \ldots, K-1\}$ e no subsímbolo $m \in \{0, \ldots, M-1\}$, utilizando uma versão do filtro protótipo deslocado no tempo e na frequência, representado por $\mathbf{g}_{k,m}$. O número total de amostras do filtro é dado por $\mathcal{N} = \mathcal{P}S$, sendo $\mathcal{P}$ o número de períodos e S o número de amostras por período.

A ν-ésima amostra do filtro deslocado é expressa por

$$g_{k,m}[\nu] = g\left[\langle \nu - m\tau S \rangle_{\mathcal{N}}\right] \exp\left(j2\pi k \frac{\phi}{S}\nu\right), \tag{7}$$

em que $\tau = \Delta_K/S$ é o fator de sobreposição temporal e $\phi = \Delta_M/\mathcal{P}$ é o fator de sobreposição espectral. As distâncias entre subsímbolos e subportadoras são representadas por Δ_K e Δ_M, respectivamente. Valores de Δ_K e Δ_M menores que S e $\mathcal{P}$ promovem sobreposição temporal e espectral, resultando, respectivamente, em *Intersymbol Interference* (ISI) e ICI, mas aumentando a taxa de transmissão e eficiência espectral do sistema.

Os números de subsímbolos e subportadoras no sistema FTN-GFDMA podem ser calculados por

$$M = \left\lfloor \frac{\mathcal{N}}{\Delta_K} \right\rfloor = \frac{\mathcal{P}}{\tau}, \quad K = \left\lfloor \frac{\mathcal{N}}{\Delta_M} \right\rfloor = \frac{S}{\phi}. \tag{8}$$

A matriz de transmissão $\mathbf{A}$, composta por $N = K \times M$ colunas, organiza as versões deslocadas do filtro protótipo da seguinte maneira:

$$\mathbf{A} = \left[\begin{array}{cccccccc} \mathbf{g}_{0,0} & \mathbf{g}_{1,0} & \cdots & \mathbf{g}_{K-1,0} & \mathbf{g}_{0,1} & \cdots & \mathbf{g}_{K-1,M-1} \end{array} \right]. \tag{9}$$

Em um cenário multiusuário composto por U usuários, cada usuário $u \in \{1, \ldots, U\}$ recebe um subconjunto específico de recursos, representado pelos conjuntos $\mathcal{K}_{\text{on}}^{(u)}$ de subportadoras e $\mathcal{M}_{\text{on}}^{(u)}$ de subsímbolos. Assim, a matriz de símbolos de cada usuário é dada por

$$\mathbf{S}_{k,m}^{(u)} = \begin{cases} s_{k,m} & \text{se } k \in \mathcal{K}_{\text{on}}^{(u)} \text{ e } m \in \mathcal{M}_{\text{on}}^{(u)} \\ 0 & \text{caso contrário.} \end{cases} \tag{10}$$

O sinal transmitido pela BS resulta da superposição das matrizes de símbolos de todos os usuários, ou seja,

$$\mathbf{S} = \sum_{u=1}^{U} \mathbf{S}^{(u)}, \tag{11}$$

cuja vetorização é dada por $\mathbf{s} = \text{vec}(\mathbf{S})$, resultando no seguinte sinal:

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \tag{12}$$

Assim, amostra ν-ésima do sinal transmitido é expressa por

$$x[\nu] = \sqrt{\frac{\mathcal{N}}{N}} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} s_{k,m} g\left[\langle \nu - m\tau S \rangle_{\mathcal{N}}\right] \exp\left(j2\pi k \frac{\phi}{S}\nu\right), \tag{13}$$

com a normalização $\sqrt{\mathcal{N}/N}$ assegurando a manutenção da potência de transmissão.

Antes da transmissão, adiciona-se um *Cyclic Prefix* (CP) de comprimento $\mathcal{N}_{\text{CP}}$, ampliando a robustez contra interferências inter-bloco. A eficiência espectral alcançada é dada por

$$\eta = \frac{R_c M K \log_2(J)}{\tilde{\mathcal{N}}}, \tag{14}$$

onde R_c é a taxa de codificação *Forward Error Correction* (FEC), J é a ordem da modulação, e $\tilde{\mathcal{N}} = \mathcal{N} + \mathcal{N}_{\text{CP}}$ é o comprimento total do bloco transmitido.

Na arquitetura A-RoF, o sinal gerado é transportado por um enlace óptico de longo alcance entre a *Central Office* (CO) e o site remoto. Para compensar as perdas no enlace, é necessário aumentar as potências óptica e de Radiofrequência (RF), induzindo, entretanto, efeitos não-lineares significativos na resposta do *Mach-Zehnder Modulator* (MZM). Este comportamento não-linear é modelado como

$$b[\nu] = \sum_{\iota=0}^{I-1} \zeta_\iota |x[\nu]|^\iota x[\nu], \tag{15}$$

onde ζ_ι são os coeficientes polinomiais e I é a ordem não-linear considerada.

Adicionalmente, o enlace óptico pode ser afetado pela *Chromatic Dispersion* (CD), a qual altera a resposta em frequência da fibra óptica, descrita pela seguinte função de transferência:

$$\Psi(e^{j\omega T}) = e^{-j\mathcal{A}(\omega T)^2}, \quad \mathcal{A} = \frac{D\lambda^2 L}{4\pi c T^2}, \tag{16}$$

onde D é o parâmetro de dispersão da fibra, λ o comprimento de onda óptico, L o comprimento da fibra, c a velocidade da luz, e T o período de amostragem. A resposta no domínio temporal pode ser representada por um filtro *Finite Impulse Response* (FIR) com N_t coeficientes:

$$\psi[\mu] = \sqrt{\frac{1}{j4\mathcal{A}\pi}} e^{j\frac{\mu^2}{4\mathcal{A}}}, \quad -\left\lfloor \frac{N_t}{2} \right\rfloor \leq \mu \leq \left\lfloor \frac{N_t}{2} \right\rfloor \tag{17}$$

O sinal na saída do *Photo Detector* (PD) resulta da convolução entre o sinal modulado e a resposta do canal óptico, sendo expresso por

$$z[\nu] = b[\nu] * \psi[\mu]. \tag{18}$$

Para evitar interferência inter-bloco, o comprimento do CP deve ser superior à soma das durações das respostas impulsivas do canal óptico (N_t) e do canal sem fio ($\mathcal{L}$):

$$N_{\text{CP}} = N_t + \mathcal{L} - 1 \tag{19}$$

Além disso, o *Power Amplifier* (PA), localizado no site remoto, introduz não-linearidades adicionais, modeladas como

$$\bar{x}[\nu] = \sum_{q=0}^{Q-1} \sum_{\gamma=0}^{\Gamma-1} \xi_{q,\gamma} |z[\nu - q]|^\gamma z[\nu - q], \tag{20}$$

onde $\xi_{q,\gamma}$ são os coeficientes do modelo, Q é a profundidade de memória, e Γ a ordem não-linear.

Na recepção, o terminal do usuário u observa o seguinte sinal

$$\mathbf{y}^{(u)} = \mathbf{H}^{(u)} \bar{\mathbf{x}} + \mathbf{w}^{(u)}, \tag{21}$$

onde $\mathbf{H}^{(u)}$ é a matriz circulante representando o canal sem fio, e $\mathbf{w}^{(u)}$ corresponde ao ruído AWGN. Após a remoção do CP, o sinal é processado por um *Matched Filter* (MF), resultando em:

$$\mathbf{r}^{(u)} = \mathbf{A}^{\text{H}} \left(\mathbf{H}^{(u)}\right)^{-1} \mathbf{H}^{(u)} \bar{\mathbf{x}} + \mathbf{A}^{\text{H}} \left(\mathbf{H}^{(u)}\right)^{-1} \mathbf{w}^{(u)} \tag{22}$$

Por fim, cada terminal de usuário emprega um detector não-linear para mitigar a *Inter-User Interference* (IUI) e realiza posteriormente o demapeamento dos recursos alocados para a recuperação correta de seus dados.

Este fluxo completo, desde a geração do sinal até a sua recepção e detecção no terminal do usuário, encontra-se ilustrado na Figura 4, que apresenta o diagrama em blocos do transceptor FTN-GFDMA *downlink*, integrado ao sistema A-RoF, para o caso do terminal do usuário u.

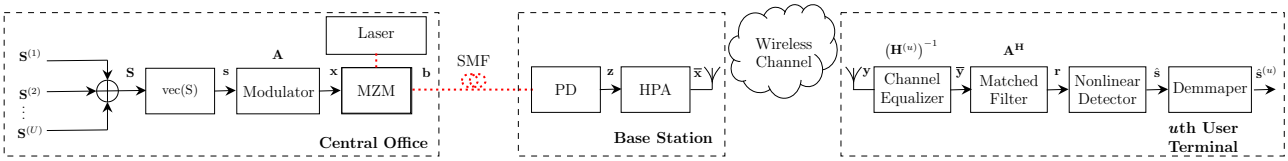


Figura 4: Diagrama em blocos do transceptor FTN-GFDMA no *downlink*, integrado ao sistema A-RoF, considerando o terminal do usuário u. As etapas de adição e remoção do CP foram omitidas para simplificação.

4.3 Resultados Obtidos

A principal vantagem da integração proposta reside na possibilidade de aumentar a eficiência espectral e a capacidade de transmissão utilizando a infraestrutura óptica já existente. Em ambientes rurais, nos quais as redes de fibra óptica podem ser limitadas e compartilhadas com outros serviços, como redes *Gigabit Passive Optical Network* (GPON), a adoção do FTN-GFDMA permite maximizar o aproveitamento da largura de banda disponível, evitando a necessidade de investimentos expressivos na ampliação da infraestrutura.

Além disso, o FTN-GFDMA apresenta uma redução significativa da OOBE em comparação com técnicas tradicionais baseadas em OFDM, como resultado da aplicação de filtragem circular e da possibilidade de sobreposição temporal e espectral dos pulsos. Essa característica é particularmente relevante para o uso oportunístico dos canais TVWS, pois permite mitigar a interferência com comunicações adjacentes, facilitando a coexistência entre usuários primários e secundários, e promovendo maior flexibilidade na fragmentação espectral.

Outro ponto positivo está na capacidade do FTN-GFDMA de ajustar diversos parâmetros do sistema, como o número de subportadoras, subsímbolos e fatores de sobreposição no tempo e na frequência. Essa flexibilidade torna possível a otimização do comportamento espectral e o ajuste fino das características do sistema de acordo com os requisitos específicos de cada aplicação. Com isso, é possível não apenas acomodar um maior número de usuários sem comprometer a *Quality of Service* (QoS), mas também aumentar a taxa de dados por usuário quando necessário, adaptando-se às demandas heterogêneas de tráfego.

O ganho em eficiência espectral promovido pelo FTN-GFDMA pode ser visualizado na Figura 5, que apresenta o desempenho em termos de BER não codificado para diferentes valores de E_b/N_0, considerando as configurações especificadas na Tabela 1 e os modelos de canal da Tabela 2. Nota-se que, mesmo com a sobreposição temporal e espectral, o sistema alcança ganhos expressivos de eficiência espectral, com 25% para o FTN-GFDM-I e 50% para o FTN-GFDM-II, sem comprometer significativamente a BER.

Tabela 1: Parâmetros dos sistemas FTN-GFDM e GFDM. Modulação BPSK, filtro RRC, $\mathcal{P} = 4$, e $\mathcal{S} = 5$.

Parâmetros	**FTN-GFDM-I**	**FTN-GFDM-II**	**GFDM**
Espaçamento temporal (Δ_K)	4	4	5
Espaçamento em frequência (Δ_M)	4	3	4
Fator de sobreposição temporal (τ)	0.8	0.8	1
Fator de sobreposição espectral (ϕ)	1	0.75	1
Subportadoras por bloco (K)	5	6	5
Subsímbolos por bloco (M)	5	5	4
Densidade de símbolos por amostra ($N/\mathcal{N}$)	1.25	1.5	1

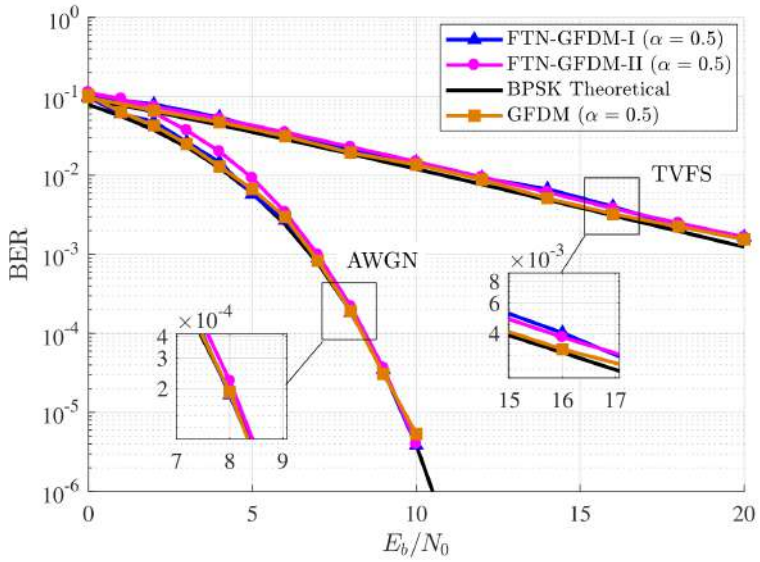


Figura 5: Desempenho de BER dos sistemas FTN-GFDM utilizando BPSK e filtro RRC com fator de roll-off $\alpha = 0.5$.

Tabela 2: Modelos de canal para avaliação dos sistemas FTN-GFDM e GFDM.

Canal	**Resposta impulsiva**
AWGN	$h = 1$
TVFS	$\mathbf{h} = [\mathrm{e}^{0}h_0; \mathrm{e}^{-3}h_1; \mathrm{e}^{-6}h_2; \mathrm{e}^{-9}h_3]^{\mathrm{T}}$, $h_n \sim \mathbb{C}\mathcal{N}(0,1)$

Consequentemente, integrar o FTN-GFDMA com sistemas A-RoF melhora a eficiência espectral e a capacidade de transmissão em cenários nos quais o espectro é limitado, além de fornecer uma solução flexível e adaptável para a ampliação da cobertura em regiões rurais, enfrentando a crescente demanda por conectividade de qualidade.

Apesar dos benefícios mencionados, a combinação do FTN-GFDMA com sistemas A-RoF impõe uma série de desafios técnicos que precisam ser cuidadosamente considerados. Um dos principais entraves é o aumento do PAPR, resultante da sobreposição temporal e espectral dos pulsos e da utilização de filtros protótipos que, embora melhorem o comportamento espectral, também contribuem para variações na amplitude do sinal transmitido.

Esse aumento no PAPR agrava os efeitos não-lineares nos componentes ópticos e eletrônicos do sistema, especialmente no MZM e no *High Power Amplifier* (HPA), que operam próximos à saturação para maximizar a eficiência energética. A operação do HPA na região de saturação pode provocar distorções severas, degradando a qualidade do sinal e comprometendo o desempenho do sistema.

A Figura 6 ilustra a *Complementary Cumulative Distribution Function* (CCDF) do PAPR para os sistemas FTN-GFDMA e *Generalized Frequency Division Multiplexing* (GFDM), evidenciando que a sobreposição dos pulsos e o aumento do nível de ISI e ICI resultam em uma maior probabilidade de ocorrência de picos de potência.

Além das distorções provocadas pelo alto PAPR, a integração do FTN-GFDMA em sistemas A-RoF também amplia as distorções inerentes ao ISI e ICI do próprio sinal. Essas interferências são ainda mais pronunciadas em canais sem fio dispersivos e variantes no tempo, onde a equalização se torna mais complexa. Em cenários com altas frequências de Doppler, por exemplo, as rápidas variações do canal dificultam a manutenção de um modelo de canal invariante durante

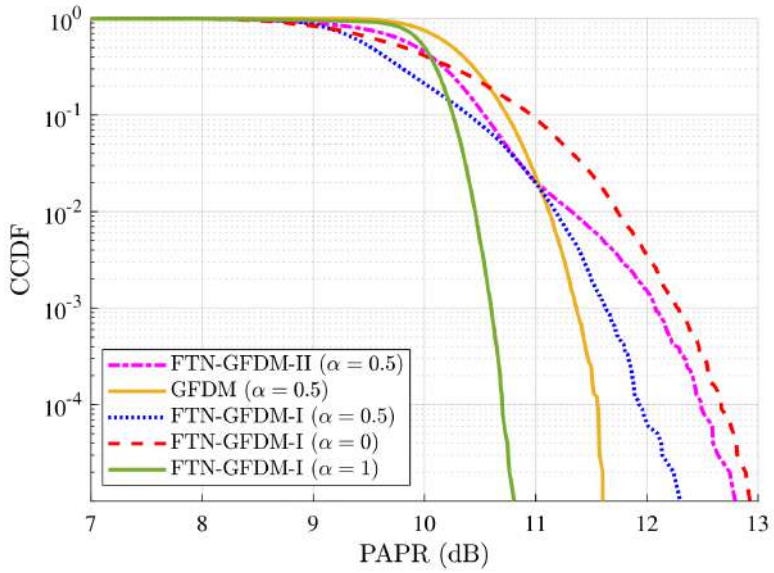


Figura 6: PAPR dos sistemas FTN-GFDM e GFDM, com CCDF obtida empiricamente a partir de 10^5 quadros, cada um contendo 40 blocos de FTN-GFDM, e fator de superamostragem de 4.

a equalização e decodificação, impactando negativamente o desempenho em termos de BER.

A necessidade de estimativas precisas da CSI torna-se, portanto, ainda mais crítica nesse contexto, pois desvios em relação ao canal real resultam em uma equalização subótima, prejudicando a confiabilidade e o desempenho do sistema. Além disso, os efeitos não-lineares e a dispersão cromática em sistemas A-RoF introduzem distorções adicionais, que degradam ainda mais a integridade do sinal, especialmente em aplicações de longa distância.

Outro desafio importante é a redução da eficiência espectral devido ao aumento do comprimento do CP. Conforme discutido anteriormente, o CP deve ser suficientemente longo para abranger a resposta impulsiva combinada dos canais óptico e sem fio, evitando a interferência inter-blocos. No entanto, esse aumento implica maior sobrecarga e redução do tempo útil de transmissão, comprometendo a eficiência espectral, um aspecto crítico para sistemas que operam em bandas de frequência limitadas.

Para superar esses desafios, torna-se desejável o desenvolvimento de esquemas abrangentes de linearização capazes de mitigar as diversas fontes de distorção. Idealmente, essas soluções devem ser implementadas no lado da CO, visando simplificar a infraestrutura da rede e reduzir custos. Nesse contexto, técnicas como *Digital Pre-Distortion* (DPD) e estratégias de redução de PAPR são fundamentais e podem ser potencializadas por soluções baseadas em AI, que oferecem a capacidade de compensar simultaneamente efeitos com e sem memória, além de mitigar a CD das fibras ópticas.

Por fim, destaca-se a complexidade do receptor como um desafio adicional. A necessidade de lidar simultaneamente com distorções por ICI, ISI, e por efeitos não-lineares e dispersivos dos canais ópticos e sem fio impõe a adoção de algoritmos avançados de recepção, que demandam elevado poder computacional. Tal complexidade pode acarretar não apenas maiores custos de hardware e software, mas também aumentar a latência na detecção e decodificação do sinal. Assim, o desenvolvimento de soluções eficientes e de baixa complexidade para a detecção de dados em sistemas FTN-GFDMA é imprescindível para viabilizar a implementação prática e assegurar um desempenho robusto em ambientes de comunicação desafiadores.

4.4 Conclusão

Este trabalho explorou a integração do FTN-GFDMA com sistemas A-RoF, com foco em sua aplicação no contexto de eRAC. A combinação dessas tecnologias apresenta um potencial significativo para melhorar a eficiência espectral, a capacidade de transmissão e a flexibilidade de cobertura, oferecendo uma solução promissora para a expansão da conectividade em regiões remotas e rurais.

O estudo também destacou os principais desafios relacionados à integração dessa técnica de múltiplo acesso a sistemas A-RoF voltados a áreas remotas. Em particular, é necessário gerenciar de forma eficaz o elevado PAPR, bem como as interferências inerentes ao FTN-GFDMA, como ISI e ICI. Além disso, as não-linearidades e a CD típicas dos sistemas A-RoF requerem técnicas apropriadas de compensação.

Superar esses desafios demandará o desenvolvimento de técnicas avançadas de linearização e redução de PAPR, possivelmente por meio de abordagens baseadas em AI. Essas soluções devem buscar um equilíbrio entre desempenho e complexidade de implementação, de modo a atender aos requisitos das aplicações eRAC.

Por fim, este trabalho reforça a necessidade de esforços dedicados para otimizar a integração entre FTN-GFDMA e sistemas A-RoF, visando explorar plenamente o potencial dessas tecnologias nas redes sem fio de próxima geração.

5 Sensoriamento Espectral para Redes 6G

Lucas dos Santos Costa
Inatel

Nos últimos anos, a rápida evolução das redes de comunicação levou à revolução dos meios de acesso à informação e das interações interpessoais através da Internet e da conectividade entre dispositivos eletrônicos. Os adventos recentes da Internet das coisas (IoT) e da rede 5G a propósito levaram a conectividade a um novo patamar, impulsionando um aumento sem precedentes na demanda por novos serviços e atividades de telecomunicações.

Mesmo apesar do curto intervalo de tempo desde que as novas tecnologias advindas da IoT e do 5G começaram a ser implementadas, as pesquisas atuais sobre as redes 6G, bem como as conjecturas para as Próximas Gerações de Redes Móveis Celulares (xGs), já se iniciaram mundo afora, e os objetivos almejados indicam mais um grande aumento de demanda, pois vislumbram um mundo totalmente conectado e com conectividade sem fio para todos [13]. Para conseguir acompanhar toda essa evolução e assim prover conectividade a um número realmente massivo de transceptores de forma harmoniosa (com interoperabilidade), antes há que se prover meios de acesso ao espectro radioelétrico que possam mitigar os atuais problemas da escassez e subutilização espectral [14].

O rádio cognitivo (*Cognitive Radio* (CR)) surgiu nesse contexto de escassez e subutilização espectral como uma solução promissora para um uso mais eficiente dos recursos espectrais por meio do chamado acesso dinâmico ao espectro (DSA) [15, 16]. Uma das tarefas mais fundamentais de um CR é a realização do sensoriamento espectral (SS), por meio do qual usuários secundários (*Secundary User*s (SUs)) são equipados com uma cognição tal que os capacita a detectar oportunidades de acesso à faixas espectrais sem causar interferências prejudiciais aos usuários primários (*Primary User*s (PUs)).

Uma grande quantidade de técnicas de SS pode ser encontrada na literatura [17, 18, 19, 20, 21, 22, 23, 24, 25]. Algumas das mais conhecidas incluem a detecção por filtro casado (*Matched Filter Detection* (MFD)), a detecção pelas propriedades ou características cicloestacionárias do sinal primário (*Cyclostationary Feature Detection* (CFD)), e a detecção de energia (*Energy Detection* (ED)). A ED é uma das mais exploradas principalmente por conta da simplicidade e baixa complexidade de implementação. Porém, a vulnerabilidade à incerteza de ruído (NU) é uma desvantagem significativa desta técnica, que pode comprometer drasticamente o desempenho de detecção e ainda levar ao fenômeno conhecido como relação sinal-ruído de barreira (*Signal-to-Noise Ratio Wall* (SNRW)) [26]. A SNRW representa um limite fundamental para o SS e pode ser definida como o valor de limiar para a relação sinal-ruído (SNR) do sinal recebido abaixo do qual não se pode garantir detecção confiável independentemente do quão longo possa ser o tempo de sensoriamento. Se a SNR do sinal recebido cair abaixo desse limiar, pode tornar a detecção via ED ineficiente ou até mesmo inutilizar o SS.

Dada: i) a necessidade de se prover convivência harmoniosa entre os inúmeros transceptores tanto na geração atual quanto nas gerações futuras das redes de comunicação através de um recurso natural intrinsecamente escasso, isto é, as faixas do espectro eletromagnético, ii) a importância do acesso dinâmico e oportunista às faixas primárias via SS nessa tarefa, e iii) a relevância da ED e do potencial de impacto da NU e da SNRW sobre sua eficiência de detecção, é essencial estudar esses fenômenos em detalhes [27, 28, 29]. Por conseguinte, o objetivo desta seção é apresentar os avanços de pesquisa envolvendo a NU e a SNRW da ED tanto no SS não cooperativo (*Non-Cooperative Spectrum Sensing* (NCSS)) quanto no SS

cooperativo (*Cooperative Spectrum Sensing* (CSS)) com fusão de decisão suave (*Soft-Decision* (SD)) e fusão de decisão abrupta (*Hard-Decision* (HD)) [16, 30].

5.1 Fundamentos sobre sensoriamento espectral

A escassez espectral é resultado da política de alocação fixa de banda adotada até então pelos órgãos reguladores de acesso ao espectro de radiofrequências; política na qual o direito de uso de uma determinada faixa é dado de forma exclusiva ao usuário incumbente, ou PU, por todo o período por ele contratado [31]. O problema é que tal política se tornou obsoleta com a evolução dos sistemas de comunicação, não podendo suprir o aumento explosivo da demanda por novas alocações de espectro visto que o espectro radioelétrico é um recurso naturalmente escasso. No entanto, por outro lado, essa parte do espectro eletromagnético também se encontra paradoxalmente subtilizada: uma constatação feita a partir das pesquisas voltadas para o desenvolvimento de soluções para o problema da escassez espectral que revela a ocorrência frequente de canais ociosos em determinadas regiões geográficas durante o período de outorga concedido aos respectivos PUs [32, 33]. Como esses canais são das faixas de radiofrequência, onde são compreendidas as frequências muito altas (VHF) e as frequências ultra altas (UHF), utilizadas para a prestação dos serviços primários de radiodifusão de televisão (*Television* (TV)), eles são chamados espaços em branco de televisão (TVWS).

Ultimamente, políticas DSA têm sido consideradas pelos órgãos reguladores como uma alternativa viável à política de acesso fixo ao espectro radioelétrico para mitigar a escassez e a subutilização espectral. A ideia central dessa nova abordagem de acesso é o compartilhamento das faixas primárias com usuários não incumbentes, ou SUs, por meio de permissão de acesso oportunista mediante à detecção de canais ocasionalmente vagos.

O conceito de SS via CR surgiu nessa conjuntura de acesso oportunista e políticas DSA como uma solução promissora para reduzir a escassez e a subutilização espectral. A nova estratégia de acesso tem como premissa tornar usuários não incumbentes em CRs, ou SUs inteligentes, equipando-os com a cognição necessária para que sejam capazes de detectar ociosidades nas faixas espectrais e ocupar canais ociosos através do SS. O SS é, por conseguinte, a atividade de monitoramento dos canais de radiofrequência, o que é feito por meio da coleta de N amostras do sinal recebido durante certo período de sensoriamento, do cômputo de uma estatística de teste T a partir dessas N amostras, da comparação dessa estatística com um limiar de decisão τ predefinido, e da tomada de uma decisão sobre o estado de ocupação do canal sensoriado; chamada de decisão de ocupação. Uma decisão de ocupação é tomada em favor da hipótese do canal sensoriado estar ocioso, representada por Hipótese $\mathcal{H}_0$ (sinal ausente), se $T \leq \tau$; caso contrário tal decisão é tomada em favor da hipótese alternativa de canal ocupado, representada por Hipótese $\mathcal{H}_1$ (sinal presente). Por isso a forma mais utilizada para sintetizar o SS é o teste binário de hipóteses, dado por $\mathcal{H}_0$ e $\mathcal{H}_1$, do qual a n-ésima amostra de sensoriamento recebida em um SU pode ser descrita como

$$y(n) = \begin{cases} v(n) & , \text{sob } \mathcal{H}_0 \\ s(n) + v(n) & , \text{sob } \mathcal{H}_1, \end{cases} \tag{23}$$

em que $s(n)$ e $v(n)$ são variáveis aleatórias complexas com distribuição Gaussiana de média zero e variância σ_{s}^2 e σ_{v}^2, ou seja, $s(n) \sim \mathcal{CN}(0,\sigma_{\text{s}}^2)$ e $v(n) \sim \mathcal{CN}(0,\sigma_{\text{v}}^2)$, e denotam a n-ésima amostra do sinal de um usuário presente na banda sensoriada e a n-ésima amostra do ruído aditivo Gaussiano branco (AWGN) presente na entrada desse SU, respectivamente. Portanto a estatística de teste T é computada a partir das amostras em $y(n)$, com $n = 1,2,\ldots,N$.

O SS pode ser classificado como NCSS ou CSS. No NCSS, cada SU autonomamente realiza o sensoriamento em um dado canal e toma uma decisão de ocupação também autônoma, enquanto que no CSS um conjunto de SUs compartilham as informações de sensoriamento para melhorar a acurácia das decisões de ocupação. O tipo das informações de sensoriamento compartilhadas para a fusão dos dados de todos os SUs cooperadores definem dois esquemas de fusão: a fusão SD e a fusão HD, cada qual com sua relação de compromisso entre complexidade e desempenho de detecção. Na fusão SD, as informações de sensoriamento podem ser as próprias amostras do sinal recebido, coletadas em um período de sensoriamento, ou alguma quantidade derivada dessas amostras; como a energia, por exemplo. Alternativamente, na fusão HD as informações de sensoriamento são as decisões de ocupação tomadas localmente em cada SU. Neste caso a decisão de ocupação cooperativa é alcançada lançando mão da regra k-em-M, na qual as M decisões locais compartilhadas, após recebidas, são combinadas para se chegar a uma decisão final que é tomada em favor de $\mathcal{H}_1$ se pelo menos k desse conjunto também forem em favor de $\mathcal{H}_1$. Dessa regra geral derivam-se três regras de fusão particulares conforme o valor determinado para a variável k [30]; isto é, a fusão de decisão pela regra OU (*OR Decision Fusion Rule* (OR)), com $k = 1$, a fusão de decisão pela regra majoritária (*Majority Decision Fusion Rule* (MAJ)), com $k = \lfloor M/2 + 1 \rfloor$, e a fusão de decisão pela regra E (*AND Decision Fusion Rule* (AND)), com $k = M$, em que $\lfloor x \rfloor$ representa o maior valor inteiro menor ou igual a x.

Os desempenhos do SS são majoritariamente medidos avaliando-se a acurácia das decisões de ocupação por meio da probabilidade de detecção, P_{d}, e da probabilidade de falso alarme, P_{fa}. $P_{\mathrm{d}} = \Pr(\text{decisão} = \mathcal{H}_1|\mathcal{H}_1) = \Pr(T > \tau|\mathcal{H}_1)$ é a probabilidade de haver uma decisão em favor de $\mathcal{H}_1$ quando o canal sensoriado está de fato ocupado, sendo equivalente à probabilidade da estatística de teste ser maior que o limiar de decisão quando tal canal está realmente sob a hipótese $\mathcal{H}_1$. Por outro lado, $P_{\mathrm{fa}} = \Pr(\text{decisão} = \mathcal{H}_1|\mathcal{H}_0) = \Pr(T > \tau|\mathcal{H}_0)$ é a probabilidade de haver uma decisão em favor de $\mathcal{H}_1$ quando o canal sensoriado está na verdade desocupado, ou seja, sob $\mathcal{H}_0$, correspondendo à probabilidade de se ter $T > \tau$ dado que o canal sensoriado está sob a hipótese $\mathcal{H}_0$. As análises de desempenhos são geralmente feitas de forma gráfica lançando mão da curva característica de operação do receptor (*Receiver Operating Characteristic* (ROC)). Cada ponto em uma curva ROC representa um par de P_{fa} versus P_{d} computado a partir de um dado valor predefinido para o limiar de decisão. Outra medida frequentemente utilizada é a área abaixo da curva ROC, comumente denotada apenas como área sob a curva (*Area Under the Curve* (AUC)). A Figura 7 sintetiza essas formas de análises por meio das curvas ROC e respectivas AUCs de duas estatísticas de teste hipotéticas, T_1 e T_2. O maior valor de AUC (AUC = 0,9647) indica a estatística de teste com melhor desempenho médio de detecção do SS; ou seja, a estatística de teste hipotética T_2 nestes exemplos.

No NCSS ou CSS com fusão HD, a estatística de teste T é computada localmente nos SUs, enquanto que no CSS com fusão SD computa-se uma estatística de teste T global após a combinação das informações de sensoriamento recebidas. No caso do SS via ED, o cômputo de T ou do limiar de decisão τ requer o conhecimento *a priori* do valor exato da variância de ruído σ_{v}^2 dos SUs para que se possa garantir o alcance do desempenho de detecção desejado em termos de P_{fa}, P_{d}, ou ambos. Na prática, no entanto, o que se tem é um valor aproximado da variância de ruído, $\hat{\sigma}_{\mathrm{v}}^2$, pois esse é um parâmetro que precisa ser estimado. O desempenho de detecção da ED é criticamente sensível ao menor grau de incerteza dessa estimativa em relação ao valor exato da variância de ruído, e os erros de estimativa caracterizam a NU. Por isso a NU vem sendo indispensavelmente considerada nos estudos do fenômeno da SNRW no SS via ED. Além de ser definida como o valor limite para a SNR do sinal recebido igual ou abaixo do qual não é possível garantir detecção robusta de um sinal não importando o tempo de sensoriamento, em

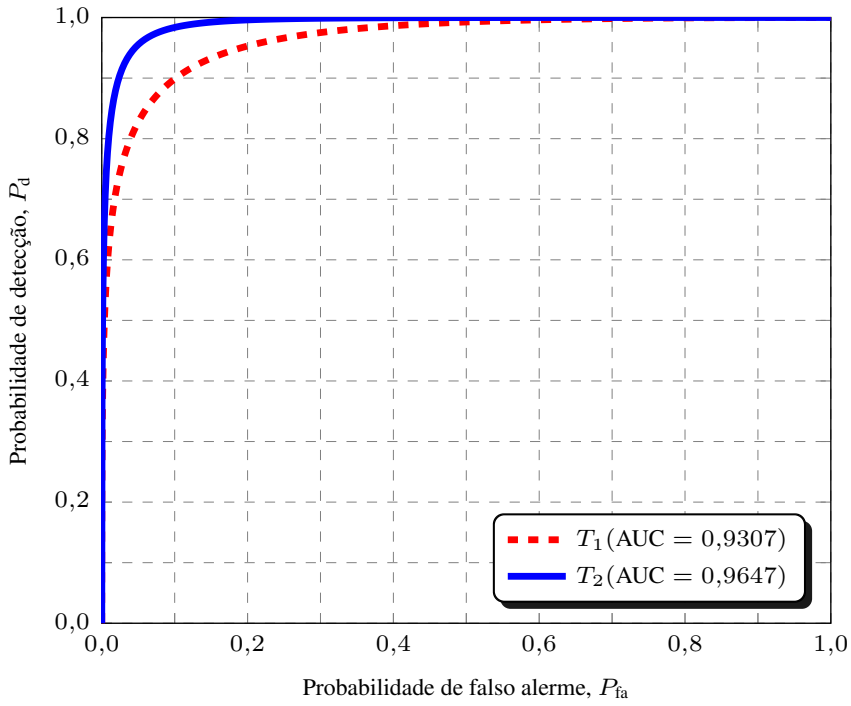


Figura 7: Curvas ROC das estatísticas de teste T_1 e T_2 e respectivas AUCs.

uma definição mais formal a SNRW também pode ser definida como sendo o limiar para a SNR abaixo do qual não se pode garantir o alcance de uma detecção ilimitadamente confiável, o que corresponde à $P_{\text{fa}} = 0$ e $P_{\text{d}} = 1$, mesmo que o número de amostras de sensoriamento tenda ao infinito, ou seja, $N \to \infty$ [27]. Por fim, em uma última interpretação mais ampla, a SNRW também pode ser entendida como o valor limite para SNR do sinal recebido igual ou abaixo do qual não se poder assegurar o alcance de quaisquer desempenhos em termos de P_{fa} e P_{d} dentro de seus limites aceitáveis, que são $P_{\text{fa}} \in [0; 0{,}5)$ e $P_{\text{d}} \in (0{,}5; 1]$, mesmo com $N \to \infty$ [34].

5.1.1 Comparação entre o Modelo Tradicional e o Novo Modelo Proposto para Incerteza de Ruído

Tendo em vista a importância do SS para aliviar os problemas da escassez e subutilização de canais de radiocomunicação e promover convivência harmoniosa entre os inúmeros transceptores tanto no contexto atual da IoT e do 5G quanto no contexto do 6G e das gerações futuras dos sistemas de comunicação, xG, o objetivo aqui é apresentar as contribuições da pesquisa sobre a SNRW da ED como consequência da NU no NCSS e CSS com fusão SD e fusão HD via regra k-em-M. Tais contribuições incluem investigações sobre o impacto de diferentes modelamentos da NU sobre os desempenhos de detecção da ED e deduções de expressões para o cômputo das métricas de desempenho P_{fa} e P_{d}, do número de amostras de sensoriamento, N, e da SNRW em cada modelamento. Mais especificamente sobre os modelamentos da NU, as investigações se baseiam em combinações envolvendo o modelamento tradicional de fonte e intervalo de NU adotado na literatura e o novo modelamento de fonte e intervalo de NU proposto com o intuito de prover uma análise mais extensiva e realista em termos práticos de implementação do SS [35].

O modelamento tradicional de NU pressupõe o conhecimento *a priori* do valor exato da variância de ruído σ_{v}^2 dos SUs no cômputo do limiar de decisão, $\tau \equiv \tau(\sigma_{\text{v}}^2)$. Por isso o efeito da NU é levado em conta, respectivamente, sobre as médias, $\hat{\mu}_0 \equiv \hat{\mu}_0(\hat{\sigma}_{\text{v}}^2)$ e $\hat{\mu}_1 \equiv \hat{\mu}_1(\hat{\sigma}_{\text{v}}^2)$, e

variâncias, $\hat{\sigma}_0^2 \equiv \hat{\sigma}_0^2(\hat{\sigma}_v^2)$ e $\hat{\sigma}_1^2 \equiv \hat{\sigma}_1^2(\hat{\sigma}_v^2)$, da estatística de teste T sob $\mathcal{H}_0$ e $\mathcal{H}_1$, que claramente oscilam com as estimativas da variância de ruído. Uma vez que neste caso não se tem qualquer controle sobre a NU, visto que essa abordagem equivale à considerar sinais interferentes (que se combinam com o sinal alvo dentro da banda sensoriada) como fonte de NU, também não é possível modelá-la, e nem mitigá-la. Além disso, nesse modelamento tradicional as estimativas em $\hat{\sigma}_v^2$ são comumente modeladas como uma variável aleatória com distribuição uniforme em dB, gerando valores em escala linear que variam em relação à σ_v^2 conforme o intervalo $[a,b]$ tradicional de NU dado por $[a = (1/\rho)\sigma_v^2, b = \rho\sigma_v^2]$ [26]. Logo, especificamente, tem-se $\hat{\sigma}_v^2 \in [a,b] = [(1/\rho)\sigma_v^2, \rho\sigma_v^2]$, sendo $\rho \geq 1$ definido como um fator de incerteza que indica o grau de incerteza da estimativa em $\hat{\sigma}_v^2$ em relação ao valor exato em σ_v^2. Por fim, é fácil notar que esse modelo de NU é polarizado, pois $(a+b)/2 \neq \sigma_v^2$ para $\rho > 1$. Por estas características, distinguir-se-á essa abordagem como modelo tradicional de fonte e intervalo de NU.

Ao contrário desse modelamento tradicional de NU, o novo modelamento proposto pressupõe o conhecimento *a priori* do valor exato da variância de ruído σ_v^2 dos SUs respectivamente nos cômputos das médias, $\mu_0 \equiv \mu_0(\sigma_v^2)$ e $\mu_1 \equiv \mu_1(\sigma_v^2)$, e das variâncias, $\sigma_0^2 \equiv \sigma_0^2(\sigma_v^2)$ e $\sigma_1^2 \equiv \sigma_1^2(\sigma_v^2)$, da estatística de teste T sob $\mathcal{H}_0$ e $\mathcal{H}_1$, que claramente não mais oscilam com as estimativas da variância de ruído em $\hat{\sigma}_v^2$. A implicação é que o valor estimado da variância de ruído é agora embutido no cômputo do limiar de decisão, que por isso passa a ser descrito como $\hat{\tau} \equiv \hat{\tau}(\hat{\sigma}_v^2)$. Uma vez que neste caso se tem controle sobre a NU, visto que essa nova abordagem equivale à considerar os erros de estimativa de um estimador real implementado nos SUs como fonte de NU, também é possível modelá-la e ainda mitigá-la pelo controle da qualidade da estimativa dos estimadores. Adicionalmente, nesse novo modelamento as estimativas em $\hat{\sigma}_v^2$ são modeladas como uma variável aleatória com distribuição Gaussiana truncada, em escala linear, gerando valores em escala linear que variam em relação à σ_v^2 conforme o novo intervalo $[a,b]$ de NU dado por $[a = (1-\rho)\sigma_v^2, b = (1+\rho)\sigma_v^2]$ [35]. Logo, especificamente, tem-se $\hat{\sigma}_v^2 \in [a,b] = [(1-\rho)\sigma_v^2, (1+\rho)\sigma_v^2]$, sendo $0 \geq \rho < 1$ o fator de incerteza que indica o grau de incerteza da estimativa em $\hat{\sigma}_v^2$ em relação ao valor exato em σ_v^2. É fácil notar que essa nova abordagem se trada de um modelo de NU não polarizado, pois, agora, veja que $(a+b)/2 = \sigma_v^2$ para $\rho \neq 0$. Por estas novas características, distinguir-se-á essa abordagem como modelo novo de fonte e intervalo de NU.

As combinações entre os modelos tradicional e novo de fonte e intervalo de NU adotadas nas análises são:

i) a combinação da fonte tradicional de NU com o intervalo tradicional de NU,

ii) a combinação da fonte tradicional de NU com o novo intervalo de NU,

iii) a combinação da nova fonte de NU com o intervalo tradicional de NU, e

iv) a combinação da nova fonte de NU com o novo intervalo de NU.

Admitir essas combinações nas referidas análises levam à uma investigação exaustiva dos efeitos da NU sobre o SS via ED. Em se tratando do CSS com fusão SD, as análises também contemplam um novo computo para a estatística de teste T além do cômputo tradicional. O resultado desse novo cômputo é que não apenas melhores desempenhos de detecção podem ser alcançados em cenários práticos, onde o esperado é que os SUs sejam heterogêneos (possuindo variâncias de ruído distintas), mas ele também leva à uma SNRW mais conservativa (maior) sob a combinação iv) quando os SUs estão operando sob diferentes níveis de NU. Por último, as análises também adotam um esquema de mapeamento entre os intervalos tradicional e novo de

NU. Esse esquema foi proposto para facilitar as transições entre as novas deduções e as deduções tradicionais reportadas na literatura, bem como para favorecer comparações e o levantamento das vantagens e desvantagens de se considerar um ou outro modelo de fonte e intervalo de NU.

5.2 Detecção de energia

5.2.1 Sensoriamento não cooperativo

Como o próprio nome sugere, na ED o cômputo da estatística de teste entrega uma medida da energia presente nas amostras de sensoriamento coletadas do sinal recebido. Logo, a partir das amostras em (7), pode-se descrevê-la para o cômputo no i-ésimo SU como [28, 27]

$$T_i = \frac{1}{N}\sum_{n=1}^{N} |y_i(n)|^2, \tag{24}$$

em que $|x|$ indica o valor absoluto de x. De acordo com o teorema do limite central, se N for suficientemente elevado, a distribuição de T_i pode ser aproximada por uma Gaussiana [28, 27] de média e variância respectivamente dadas por

$$\mu_{0_i} = \sigma^2_{\mathrm{v}_i} \text{ e } \sigma^2_{0_i} = \sigma^4_{\mathrm{v}_i}/N, \tag{25}$$

sob a hipótese $\mathcal{H}_0$, ou por

$$\mu_{1_i} = \sigma^2_{\mathrm{s}_i} + \sigma^2_{\mathrm{v}_i} \text{ e } \sigma^2_{1_i} = \sigma^4_{\mathrm{v}_i}\,(1+\bar{\gamma}_i)^2/N, \tag{26}$$

sob a hipótese $\mathcal{H}_1$, em que $\bar{\gamma}_i = \sigma^2_{\mathrm{s}_i}/\sigma^2_{\mathrm{v}_i}$ representa a SNR média do sinal recebido no i-ésimo SU. Por isso, P_{fa_i} e P_{d_i} podem ser escritas, respectivamente, como

$$P_{\mathrm{fa}_i} = Q\left(\frac{\tau_i - \mu_{0_i}}{\sigma_{0_i}}\right) = Q\left(\frac{\tau_i - \sigma^2_{\mathrm{v}_i}}{\sigma^2_{\mathrm{v}_i}/\sqrt{N}}\right) \tag{27}$$

e

$$P_{\mathrm{d}_i} = Q\left(\frac{\tau_i - \mu_{1_i}}{\sigma_{1_i}}\right) = Q\left(\frac{\tau_i - \sigma^2_{\mathrm{v}_i}(1+\bar{\gamma}_i)}{\sigma^2_{\mathrm{v}_i}(1+\bar{\gamma}_i)/\sqrt{N}}\right), \tag{28}$$

sendo $Q(t) = (1/\sqrt{2\pi})\int_t^\infty \exp\left(-x^2/2\right) dx$ a função-Q de uma Gaussiana padrão [36], $\tau_i = \lambda\sigma^2_{\mathrm{v}_i}$ o limiar de decisão e $\lambda > 0$ uma constante ajustada em função do valor alvo para P_{fa_i}, P_{d_i} ou ambas as métricas. O número de amostras necessário para satisfazer os valores alvos para essas métricas pode ser obtido extraindo τ_i de (27) e (28) e igualando os resultados, o que leva à

$$N_i = \left[Q^{-1}(P_{\mathrm{fa}_i}) - Q^{-1}(P_{\mathrm{d}_i})(1+\bar{\gamma}_i)\right]^2/\bar{\gamma}_i^2. \tag{29}$$

5.2.2 Sensoriamento cooperativo

5.2.2.1 Fusão de decisão abrupta No caso do CSS com fusão HD, as métricas de desempenhos globais Q_{fa} e Q_{d} podem ser dadas, respectivamente, em função das métricas locais em (27) e (28), como [37]

$$Q_{\mathrm{fa}} = \sum_{\ell=k}^{M} \sum_{\mathbf{d}:w_{\mathrm{H}}(\mathbf{d})=\ell} \prod_{i=1}^{M} P_{\mathrm{fa}_i}^{d_i}(1-P_{\mathrm{fa}_i})^{1-d_i} \tag{30}$$

e

$$Q_{\mathrm{d}} = \sum_{\ell=k}^{M} \sum_{\mathbf{d}:w_{\mathrm{H}}(\mathbf{d})=\ell} \prod_{i=1}^{M} P_{\mathrm{d}_i}^{d_i}(1-P_{\mathrm{d}_i})^{1-d_i}, \tag{31}$$

em que $\mathbf{d} = (d_1, \ldots, d_M)$ é um vetor binário de comprimento M cujo elemento da i-ésima entrada é a decisão local do i-ésimo SU à favor da hipótese $\mathcal{H}_0$, dada por $d_i = 0$, ou da hipótese $\mathcal{H}_1$, dada por $d_i = 1$, e $w_{\mathrm{H}}(\cdot)$ denota a distância de Hamming [35], dada por $w_{\mathrm{H}}(\mathbf{d}) = \sum_{i=1}^{M} d_i$.

5.2.2.2 Fusão de decisão suave

Cômputo tradicional da estatística de teste A estatística de teste global tradicional na fusão SD é geralmente computada a partir da estatística local em (24) como

$$T = \frac{1}{M} \sum_{i=1}^{M} T_i. \tag{32}$$

Neste caso, se N em T_i for suficientemente elevado, a distribuição de T pode ser representada por uma Gaussiana de média e variância respectivamente dadas por

$$\mu_0 = \frac{1}{M} \sum_{i=1}^{M} \sigma_{\mathrm{v}_i}^2 \text{ e } \sigma_0^2 = \frac{1}{M^2 N} \sum_{i=1}^{M} \sigma_{\mathrm{v}_i}^4 \tag{33}$$

sob a hipótese $\mathcal{H}_0$, ou por

$$\mu_1 = \frac{1}{M} \sum_{i=1}^{M} \sigma_{\mathrm{v}_i}^2 (1+\bar{\gamma}_i) \text{ e } \sigma_1^2 = \frac{1}{M^2 N} \sum_{i=1}^{M} \sigma_{\mathrm{v}_i}^4 (1+\bar{\gamma}_i)^2 \tag{34}$$

sob a hipótese $\mathcal{H}_1$. Então, Q_{fa} e Q_{d} podem ser computadas, respectivamente, como

$$Q_{\mathrm{fa}} = Q\left(\frac{\tau-\mu_0}{\sigma_0}\right) = Q\left(\frac{M\tau - \sum_{i=1}^{M} \sigma_{\mathrm{v}_i}^2}{\frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^{M} \sigma_{\mathrm{v}_i}^4}}\right) \tag{35}$$

e

$$Q_{\mathrm{d}} = Q\left(\frac{\tau-\mu_1}{\sigma_1}\right) = Q\left(\frac{M\tau - \sum_{i=1}^{M} \sigma_{\mathrm{v}_i}^2 (1+\bar{\gamma}_i)}{\frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^{M} \sigma_{\mathrm{v}_i}^4 (1+\bar{\gamma}_i)^2}}\right), \tag{36}$$

em que $\tau = (\lambda/M) \sum_{i=1}^{M} \sigma_{\mathrm{v}_i}^2$ denota o limiar de decisão global preestabelecido. O número de amostras necessário para satisfazer os valores alvos para Q_{fa} e Q_{d} pode ser verificado extraindo τ de (35) e (36) e igualando os resultados. Fazendo isso, e admitindo SUs homogêneos (operando sob mesmas condições de sensoriamento, isto é, com $\sigma_{\mathrm{v}_i}^2 = \sigma_{\mathrm{v}}^2$ e $\bar{\gamma}_i = \bar{\gamma}$ para todo i), tem-se

$$N = \frac{1}{M} \left[Q^{-1}(Q_{\mathrm{fa}}) - Q^{-1}(Q_{\mathrm{d}})(1+\bar{\gamma})\right]^2 / \bar{\gamma}^2, \tag{37}$$

significando que nestas condições, para dado desempenho alvo em termos das probabilidades de falso alarme e detecção, tal desempenho pode ser alcançado no CSS com um número de amostras M vezes menor que no NCSS (compare (37) com (29)).

Novo cômputo da estatística de teste De forma opcional ao cômputo tradicional em (32), a estatística de teste ED pode ser computada a partir de T_i em (24), como

$$T = \frac{1}{M}\sum_{i=1}^{M}\frac{T_i}{\sigma^2_{\mathrm{v}_i}} = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{N\sigma^2_{\mathrm{v}_i}}\sum_{n=1}^{N}|y_i(n)|^2. \tag{38}$$

Veja que neste novo cômputo a estatística global é uma soma ponderada das energias parciais em $\{T_i\}$, em que o i-ésimo peso é inversamente proporcional à variância do ruído presente no i-ésimo SU. Tal ponderação permite que as métricas de desempenhos alcancem melhores resultados em cenários práticos de sensoriamento [35] (quando os SUs estão possivelmente operando sob diferentes níveis de ruído). Em comparação com o cômputo tradicional, esse novo cômputo altera respectivamente a média e a variância da estatística de teste para

$$\mu_0 = 1 \text{ e } \sigma_0^2 = \frac{1}{MN} \tag{39}$$

sob a hipótese $\mathcal{H}_0$, ou para

$$\mu_1 = \frac{1}{M}\sum_{i=1}^{M}(1+\bar{\gamma}_i) \text{ e } \sigma_1^2 = \frac{1}{M^2N}\sum_{i=1}^{M}(1+\bar{\gamma}_i)^2 \tag{40}$$

sob a hipótese $\mathcal{H}_1$. Consequentemente, as métricas globais Q_{fa} e Q_{d} tornam-se, respectivamente,

$$Q_{\mathrm{fa}} = Q\left(\frac{\tau-\mu_0}{\sigma_0}\right) = Q\left[(\tau-1)\sqrt{MN}\right] \tag{41}$$

e

$$Q_{\mathrm{d}} = Q\left(\frac{\tau-\mu_1}{\sigma_1}\right) = Q\left(\frac{M\tau - \sum_{i=1}^{M}(1+\bar{\gamma}_i)}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}(1+\bar{\gamma}_i)^2}}\right), \tag{42}$$

com $\tau = \lambda$ [35]. Executando o mesmo procedimento adotado em relação ao cômputo tradicional de T em (32) para a dedução de N em (37), veja que o novo cômputo em (38) não altera o número de amostras de sensoriamento necessário para satisfazer certo desempenho de detecção alvo em termos de Q_{fa} e Q_{d}, pois a expressão resultante para N extraída a partir de (41) e (42), com $\bar{\gamma}_i = \bar{\gamma}$ para todo i, permanece idêntica àquela apresentada em (37).

5.3 Modelamento da incerteza de ruído

O modelamento da NU adota o mapeamento entre os limites/intervalos $[a_i,b_i]$ tradicional e novo, proposto como $\left[a_i = (1-\rho_{1_i})\sigma^2_{\mathrm{v}_i}, b_i = (1+\rho_{2_i})\sigma^2_{\mathrm{v}_i}\right]$. Logo, uma estimativa da variância de ruído referente ao i-ésimo SU pode ser modelada como

$$\hat{\sigma}^2_{\mathrm{v}_i} \in \left[a_i = (1-\rho_{1_i})\sigma^2_{\mathrm{v}_i}, b_i = (1+\rho_{2_i})\sigma^2_{\mathrm{v}_i}\right], \tag{43}$$

Com isso, basta substituir ρ_{1_i} por $(\rho_i-1)/\rho_i$ e ρ_{2_i} por $\rho_i - 1$, $\rho_i \geq 1$, em (43) para configurar o intervalo tradicional, ou $\rho_{1_i} = \rho_{2_i}$ por ρ_i, $0 \leq \rho_i < 1$, para configurar o novo intervalo de NU. Esse mapeamento facilita a dedução de expressões para a SNRW e favorece comparações

entre os modelos tradicional e novo de fonte e intervalo de NU. A distribuição das estimativas em $\hat{\sigma}^2_{\mathrm{v}_i}$ segue uma função densidade de probabilidade (*Probability Density Function* (PDF)) Gaussiana, em escala linear, truncada no intervalo $[a_i,b_i]$, e pode ser descrita como

$$f_{\hat{\sigma}^2_{\mathrm{v}_i}}(x_i) = \begin{cases} 0 & , x_i \leq a_i, \\ \frac{\phi(\mu_i,\sigma_i^2;x_i)}{\Phi(\mu_i,\sigma_i^2;b_i)-\Phi(\mu_i,\sigma_i^2;a_i)} & , a_i < x_i < b_i, \\ 0 & , x_i \geq b_i. \end{cases} \tag{44}$$

Em (44), o símbolo $\phi(\mu_i,\sigma_i^2;x_i)$ denota a função Gaussiana, de média $\mu_i = \sigma^2_{\mathrm{v}_i}$ e variância $\sigma_i^2 = [(\rho_{1_i}+\rho_{2_i})\sigma^2_{\mathrm{v}_i}/6]^2$, aplicada à variável $x_i = \hat{\sigma}^2_{\mathrm{v}_i}$. Já $\Phi(\mu_i,\sigma_i^2;a_i)$ ou $\Phi(\mu_i,\sigma_i^2;b_i)$ denota uma constante em $x_i = a_i$ ou $x_i = b_i$ associada à função de distribuição cumulativa (*Cumulative Distribution Function* (CDF)) Gaussiana, respectivamente [35].

5.4 Detecção de energia sob incerteza de ruído

Esta seção trata do SS via ED sob NU. No entanto, primeiramente é importante notar que na ausência de NU, sob $\mathcal{H}_0$ ou $\mathcal{H}_1$, a variância da estatística de teste ED tende para zero à medida que $N \to \infty$, mas a média é independente de N (vide (25), (26), (33), (34), (39), e (40)). Por isso, uma vez que a média sob $\mathcal{H}_1$ é sempre maior que a média sob $\mathcal{H}_0$, para qualquer nível de SNR maior que zero ($\{\bar{\gamma}_i,\bar{\gamma}\} > 0$) é possível alcançar qualquer desempenho de detecção, inclusive o nível de desempenho ideal, que corresponde à probabilidades de falso alarme e detecção iguais a zero e um, respectivamente, desde que o número de amostras de sensoriamento possa ser suficientemente elevado. Matematicamente, pode-se escrever que

$$\lim_{N\to\infty} P_{\mathrm{fa}_i} = 0 \text{ e } \lim_{N\to\infty} P_{\mathrm{d}_i} = 1 \tag{45}$$

e que

$$\lim_{N\to\infty} Q_{\mathrm{fa}} = 0 \text{ e } \lim_{N\to\infty} Q_{\mathrm{d}} = 1. \tag{46}$$

Uma vez que as métricas de desempenhos são computadas a partir da função-Q da Gaussiana padrão, também é importante destacar que

$$\lim_{N\to\infty} Q\left(g\sqrt{N}\right) = \begin{cases} 0 & , \text{se } g > 0 \\ 1 & , \text{se } g < 0 \end{cases}. \tag{47}$$

Essas informações serão utilizadas posteriormente nas deduções das expressões para a SNRW.

5.4.1 Sensoriamento não cooperativo

5.4.1.1 Fonte tradicional combinada com intervalo tradicional/novo de NU Como discutido na Seção 5.1.1, quando a fonte tradicional de NU é considerada, o valor exato da variância de ruído é incorporado no cômputo do limiar de decisão e o valor estimado é incorporado no cômputo das médias e variâncias da estatística de teste sob as hipóteses $\mathcal{H}_0$ e $\mathcal{H}_1$. Especificamente, o limiar de decisão é dado por $\tau_i = \lambda\sigma^2_{\mathrm{v}_i}$, e a média e a variância de T_i em (24) por $\hat{\mu}_{0_i} = \hat{\sigma}^2_{\mathrm{v}_i}$ e $\hat{\sigma}^2_{0_i} = \hat{\sigma}^4_{\mathrm{v}_i}/N$ sob a hipótese $\mathcal{H}_0$, ou por $\hat{\mu}_{1_i} = \hat{\sigma}^2_{\mathrm{v}_i} + \sigma^2_{\mathrm{s}_i}$ e $\hat{\sigma}^2_{1_i} = (\hat{\sigma}^2_{\mathrm{v}_i} + \sigma^2_{\mathrm{s}_i})^2/N$ sob a hipótese $\mathcal{H}_1$, respectivamente. Com isso, para o dado valor específico da estimativa em $\hat{\sigma}^2_{\mathrm{v}_i}$, as métricas de desempenho P_{fa_i} e P_{d_i} tornam-se, respectivamente,

$$P_{\mathrm{fa}_i} = Q\left(\frac{\tau - \hat{\mu}_{0_i}}{\hat{\sigma}_{0_i}}\right) = Q\left(\frac{\tau_i - \hat{\sigma}^2_{\mathrm{v}_i}}{\hat{\sigma}^2_{\mathrm{v}_i}/\sqrt{N}}\right) \tag{48}$$

e

$$P_{\mathrm{d}_i} = Q\left(\frac{\tau - \hat{\mu}_{1_i}}{\hat{\sigma}_{1_i}}\right) = Q\left(\frac{\tau_i - (\hat{\sigma}^2_{\mathrm{v}_i} + \sigma^2_{\mathrm{s}_i})}{(\hat{\sigma}^2_{\mathrm{v}_i} + \sigma^2_{\mathrm{s}_i})/\sqrt{N}}\right). \tag{49}$$

Porém, para levar em conta todo o intervalo de variação das estimativa da variância de ruído, deve-se computar as métricas de desempenho médias, integrando-as conforme a PDF da NU apresentada em (44), o que, convenientemente substituindo x_i por $\hat{\sigma}^2_{\mathrm{v}_i}$ em P_{fa_i} e P_{d_i}, resulta em

$$\bar{P}_{\mathrm{fa}_i} = \int_{a_i}^{b_i} Q\left(\frac{\tau_i - x_i}{x_i/\sqrt{N}}\right) f_{\hat{\sigma}^2_{\mathrm{v}_i}}(x_i)dx_i \tag{50}$$

e

$$\bar{P}_{\mathrm{d}_i} = \int_{a_i}^{b_i} Q\left(\frac{\tau_i - (x_i + \sigma^2_{\mathrm{s}_i})}{(x_i + \sigma^2_{\mathrm{s}_i})/\sqrt{N}}\right) f_{\hat{\sigma}^2_{\mathrm{v}_i}}(x_i)dx_i. \tag{51}$$

Analisando os argumentos das funções-Q em (50) e (51) a fim de satisfazer (45) levando em conta (47) para $N \to \infty$, é necessário que $\lambda > 1 + \rho_{2_i}$ para que $\bar{P}_{\mathrm{fa}_i} = 0$ e que $\lambda < 1 - \rho_{1_i} + \bar{\gamma}_i$ para que $\bar{P}_{\mathrm{d}_i} = 1$ [35]. Relacionar essas duas inequações resulta em $\bar{\gamma}_i > \rho_{1_i} + \rho_{2_i}$. Uma vez que se a inequação $\bar{\gamma}_i > \rho_{1_i} + \rho_{2_i}$ não for satisfeita, ou seja, se $\bar{\gamma}_i \leq \rho_{1_i} + \rho_{2_i}$, também não se pode satisfazer (45), uma SNRW generalizada pode ser definida a partir do esquema de mapeamento em (43) como $\gamma_{\mathrm{w}_i} = \rho_{1_i} + \rho_{2_i}$. Logo, substituindo $\rho_{1_i} = \rho_{2_i} = \rho_i$, $0 \leq \rho_i < 1$, em γ_{w_i}, a SNRW particularizada para a combinação da fonte tradicional de NU com o novo intervalo de NU fica como

$$\gamma_{\mathrm{w}_i} = 2\rho_i. \tag{52}$$

Por outro lado, substituindo ρ_{1_i} por $(\rho_i - 1)/\rho_i$ e ρ_{2_i} por $\rho_i - 1$, $\rho_i \geq 1$, em γ_{w_i}, a SNRW particularizada para a combinação da fonte tradicional de NU com o intervalo tradicional de NU resulta em

$$\gamma_{\mathrm{w}_i} = (\rho_i^2 - 1)/\rho_i. \tag{53}$$

O número de amostras necessário para satisfazer o desejado para as métricas de desempenhos pode ser determinado extraindo τ_i de (48) e (49), admitindo-se a pior situação em termos de NU em cada caso, isto é, admitindo-se $\hat{\sigma}^2_{\mathrm{v}_i} = (1 + \rho_{2_i})\sigma^2_{\mathrm{v}_i}$ em (48), o que aumenta a P_{fa}, e $\hat{\sigma}^2_{\mathrm{v}_i} = (1 - \rho_{1_i})\sigma^2_{\mathrm{v}_i}$ em (49), o que reduz a P_{d}, e igualando os resultados. Fazendo isso, tem-se

$$N_i = \frac{\left[Q^{-1}(P_{\mathrm{fa}_i})(1+\rho_{2_i}) - Q^{-1}(P_{\mathrm{d}_i})\,(1-\rho_{1_i}+\bar{\gamma}_i)\right]^2}{\left[\bar{\gamma}_i - (\rho_{1_i} + \rho_{2_i})\right]^2}. \tag{54}$$

5.4.1.2 Nova fonte combinada com intervalo tradicional/novo de NU Como apresentado na Seção 5.1.1, neste caso, ao contrário do que acontece com a fonte tradicional de NU, com a nova fonte de NU o valor estimado da variância de ruído é incorporado no cômputo do limiar de decisão, e o valor exato no cômputo das médias e variâncias da estatística de teste sob $\mathcal{H}_0$ e $\mathcal{H}_1$. Então, o limiar de decisão é dado por $\hat{\tau}_i = \lambda\hat{\sigma}^2_{\mathrm{v}_i}$, e a média e a variância de T_i em (24) por $\mu_{0_i} = \sigma^2_{\mathrm{v}_i}$ e $\sigma^2_{0_i} = \sigma^4_{\mathrm{v}_i}/N$ sob a hipótese $\mathcal{H}_0$, ou por $\mu_{1_i} = \sigma^2_{\mathrm{v}_i} + \sigma^2_{\mathrm{s}_i}$ e $\sigma^2_{1_i} = (\sigma^2_{\mathrm{v}_i} + \sigma^2_{\mathrm{s}_i})^2/N$ sob a hipótese $\mathcal{H}_1$, respectivamente. Com isso, para o dado valor específico da estimativa em $\hat{\sigma}^2_{\mathrm{v}_i}$, as métricas de desempenho P_{fa_i} e P_{d_i} tornam-se, respectivamente,

$$P_{\mathrm{fa}_i} = Q\left(\frac{\hat{\tau}_i - \mu_{0_i}}{\sigma_{0_i}}\right) = Q\left(\frac{\hat{\tau}_i - \sigma^2_{\mathrm{v}_i}}{\sigma^2_{\mathrm{v}_i}/\sqrt{N}}\right) \tag{55}$$

e

$$P_{\mathrm{d}_i} = Q\left(\frac{\hat{\tau}_i - \mu_{1_i}}{\sigma_{1_i}}\right) = Q\left(\frac{\hat{\tau}_i - (\sigma^2_{\mathrm{v}_i} + \sigma^2_{\mathrm{s}_i})}{(\sigma^2_{\mathrm{v}_i} + \sigma^2_{\mathrm{s}_i})/\sqrt{N}}\right). \tag{56}$$

Seguindo o mesmo procedimento adotado com a fonte tradicional de NU, considerar todo o intervalo de variação da estimativa da variância de ruído implica em integrar as métricas de desempenho de acordo com a PDF da NU apresentada em (44). Com isso essas métricas podem ser reescritas, respectivamente, como

$$\bar{P}_{\mathrm{fa}_i} = \int_{a_i}^{b_i} Q\left(\frac{\lambda x_i - \sigma^2_{\mathrm{v}_i}}{\sigma^2_{\mathrm{v}_i}/\sqrt{N}}\right) f_{\hat{\sigma}^2_{\mathrm{v}_i}}(x_i)dx_i \tag{57}$$

e

$$\bar{P}_{\mathrm{d}_i} = \int_{a_i}^{b_i} Q\left(\frac{\lambda x_i - (\sigma^2_{\mathrm{v}_i} + \sigma^2_{\mathrm{s}_i})}{(\sigma^2_{\mathrm{v}_i} + \sigma^2_{\mathrm{s}_i})/\sqrt{N}}\right) f_{\hat{\sigma}^2_{\mathrm{v}_i}}(x_i)dx_i. \tag{58}$$

Verificando os argumentos das funções-Q em (57) e (58) a fim de satisfazer (45) levando em conta (47) para $N \to \infty$, desta vez é necessário que $\lambda > (1-\rho_{1_i})^{-1}$ para que $\bar{P}_{\mathrm{fa}_i} = 0$ e que $\lambda < (1+\bar{\gamma}_i)/(1+\rho_{2_i})$ para que $\bar{P}_{\mathrm{d}_i} = 1$ [35], o que leva à $\bar{\gamma}_i > (\rho_{1_i} + \rho_{2_i})/(1-\rho_{1_i})$. Portanto, não se pode satisfazer (45) se $\bar{\gamma}_i \leq (\rho_{1_i} + \rho_{2_i})/(1-\rho_{1_i})$. Por isso uma SNRW generalizada pode ser definida a partir do esquema de mapeamento em (43) como $\gamma_{\mathrm{w}_i} = (\rho_{1_i} + \rho_{2_i})/(1-\rho_{1_i})$. Consequentemente, substituindo ρ_{1_i} por $(\rho_i - 1)/\rho_i$ e ρ_{2_i} por $\rho_i - 1$, $\rho_i \geq 1$, em γ_{w_i}, a SNRW particularizada para a combinação da nova fonte de NU com o intervalo tradicional de NU é computada como

$$\gamma_{\mathrm{w}_i} = \rho_i^2 - 1. \tag{59}$$

Por outro lado, substituindo $\rho_{1_i} = \rho_{2_i} = \rho_i$, $0 \leq \rho_i < 1$, em γ_{w_i}, a SNRW particularizada para a abordagem da combinação da nova fonte de NU com o novo intervalo de NU pode ser escrita como

$$\gamma_{\mathrm{w}_i} = \frac{2\rho_i}{1-\rho_i}. \tag{60}$$

Finalmente, o número de amostras requerido para satisfazer os valores alvos para as métricas de desempenho pode ser obtido extraindo τ_i de (57) e (58) admitindo os piores casos em termos de NU, isto é, admitindo $\hat{\sigma}^2_{\mathrm{v}_i} = (1-\rho_{1_i})\sigma^2_{\mathrm{v}_i}$ em (48) e $\hat{\sigma}^2_{\mathrm{v}_i} = (1+\rho_{2_i})\sigma^2_{\mathrm{v}_i}$ em (49), e igualando os resultados. Essas operações levam à

$$N_i = \frac{\left[Q^{-1}(P_{\mathrm{fa}_i})\frac{1+\rho_{2_i}}{1-\rho_{1_i}} - Q^{-1}(P_{\mathrm{d}_i})\,(1+\bar{\gamma}_i)\right]^2}{\left[\bar{\gamma}_i - \frac{\rho_{1_i}+\rho_{2_i}}{1-\rho_{1_i}}\right]^2}. \tag{61}$$

A Tabela 3 agrupa as expressões particularizadas deduzidas para a SNRW da ED no NCSS para as quatro combinações envolvendo os modelos tradicional e novo de fonte e intervalo de NU.

5.4.2 Sensoriamento cooperativo

5.4.2.1 Fusão de decisão abrupta Para o CSS com fusão HD sob NU, as métricas globais $\bar{Q}_{\mathrm{fa}}$ e $\bar{Q}_{\mathrm{d}}$ podem ser descritas lançando-se mão das métricas locais sob NU, apresentadas em (50)

Tabela 3: SNRW da ED no i-ésimo SU, γ_{w_i}, no NCSS para as quatro combinações envolvendo os modelos tradicional e novo de fonte e intervalo de NU.

	Fonte Tradicional	Nova Fonte
Intervalo Tradicional	$\gamma_{\mathrm{w}_i} = \frac{\rho_i^2-1}{\rho_i}, \rho_i \geq 1,$ @(53)	$\gamma_{\mathrm{w}_i} = \rho_i^2 - 1, \rho_i \geq 1,$ @(59)
Novo Intervalo	$\gamma_{\mathrm{w}_i} = 2\rho_i, 0 \leq \rho_i < 1,$ @(52)	$\gamma_{\mathrm{w}_i} = \frac{2\rho_i}{1-\rho_i}, 0 \leq \rho_i < 1,$ @(60)

e (51), e substituindo-as em (30) e (31), respectivamente. Portanto, fazendo isso, tem-se, respectivamente [37]

$$\bar{Q}_{\mathrm{fa}} = \sum_{\ell=k}^{M} \sum_{\mathbf{d}:w_{\mathrm{H}}(\mathbf{d})=\ell} \prod_{i=1}^{M} \bar{P}_{\mathrm{fa}_i}^{d_i}(1-\bar{P}_{\mathrm{fa}_i})^{1-d_i} \tag{62}$$

e

$$\bar{Q}_{\mathrm{d}} = \sum_{\ell=k}^{M} \sum_{\mathbf{d}:w_{\mathrm{H}}(\mathbf{d})=\ell} \prod_{i=1}^{M} \bar{P}_{\mathrm{d}_i}^{d_i}(1-\bar{P}_{\mathrm{d}_i})^{1-d_i}, \tag{63}$$

Um análise combinatória aplicada em (62) e (63) revela que pelo menos $\binom{M}{M-k+1}$ combinações de $M-k+1$ SUs devem satisfazer $\bar{P}_{\mathrm{fa}} = 0$ para que se possa alcançar $\bar{Q}_{\mathrm{fa}} = 0$ e que pelo menos $\binom{M}{k}$ combinações de k SUs devem satisfazer $\bar{P}_{\mathrm{d}} = 1$ para que se possa alcançar $\bar{Q}_{\mathrm{d}} = 1$ [35]. Estas informações serão úteis para as deduções de expressões para o cálculo da SNRW a seguir.

Fonte tradicional combinada com intervalo tradicional/novo de NU No que diz respeito à probabilidade de falso alarme, foi mostrado que para se alcançar $\bar{P}_{\mathrm{fa}_i} = 0$ em (55) é necessário que $\lambda > 1+\rho_{2_i}$. Portanto, para se ter $\bar{Q}_{\mathrm{fa}} = 0$ em (62), observe que para $i = 1, \ldots, M$ as seguintes inequações devem ser satisfeitas:

$$\lambda > \begin{cases} 1+\rho_{2_1}, \mathrm{e}, \ldots, \mathrm{e}\ 1+\rho_{2_{M-k+1}} & , \text{ ou} \\ \vdots & , \text{ ou} \\ 1+\rho_{2_k}, \mathrm{e}, \ldots, \mathrm{e}\ 1+\rho_{2_M}. & \end{cases} \tag{64}$$

Considerando que é necessário satisfazer pelo menos uma das $\binom{M}{M-k+1}$ linhas de (64), com $M-k+1$ condições em cada linha, para que se possa ter $\bar{Q}_{\mathrm{fa}} = 0$, pode-se reescrever λ em função do mínimo valor resultante do conjunto dos máximos valores extraídos de (64) (cada valor máximo proveniente de uma das linhas de (64)). Fazendo isso, tem-se

$$\lambda > \min \begin{pmatrix} \max\left(1+\rho_{2_1}, \ldots, 1+\rho_{2_{M-k+1}}\right) & , \\ \vdots & , \\ \max\left(1+\rho_{2_k}, \ldots, 1+\rho_{2_M}\right) & \end{pmatrix}. \tag{65}$$

Por outro lado, em relação à probabilidade de detecção, foi mostrado que para se alcançar $\bar{P}_{\mathrm{d}_i} = 1$ em (56) é necessário que $\lambda > 1+\rho_{1_i}+\bar{\gamma}_i$. Logo, para se ter $\bar{Q}_{\mathrm{d}} = 1$ em (63), observe que para $i = 1, \ldots, M$ as seguintes inequações devem ser satisfeitas:

$$\lambda < \begin{cases} 1-\rho_{1_1}+\bar{\gamma}_1, \mathrm{e}, \ldots, \mathrm{e}\ 1-\rho_{1_k}+\bar{\gamma}_k & , \text{ ou} \\ \vdots & , \text{ ou} \\ 1-\rho_{1_{M-k+1}}+\bar{\gamma}_{M-k+1}, \mathrm{e}, \ldots, \mathrm{e}, 1-\rho_{1_M}+\bar{\gamma}_M & . \end{cases} \tag{66}$$

Levando em conta que é necessário satisfazer pelo menos uma das $\binom{M}{k}$ linhas de (66), com k condições cada linha, para que se possa ter $\bar{Q}_{\mathrm{d}} = 1$, pode-se reescrever λ em (66) como

$$\lambda < \begin{cases} \min\left(1-\rho_{1_1}+\bar{\gamma}_1,\ldots,1-\rho_{1_k}+\bar{\gamma}_k\right) & \text{, ou} \\ \vdots & \text{, ou} \\ \min\left(1-\rho_{1_{M-k+1}}+\bar{\gamma}_{M-k+1},\ldots,1-\rho_{1_M}+\bar{\gamma}_M\right). \end{cases} \tag{67}$$

Desconsiderando por ora a situação de igualdade entre os fatores de NU, e organizando-os em ordem decrescente, isto é, $\rho_j > \rho_{j+1}$ para $j = 1,\ldots,M-1$, note a partir de (65) e (67) que

$$1+\rho_{2_k} < \begin{cases} \min(1-\rho_{1_1}+\bar{\gamma}_1,\ldots,1-\rho_{1_k}+\bar{\gamma}_k) & \text{, ou} \\ \vdots & \text{, ou} \\ \min\left(1-\rho_{1_{M-k+1}}+\bar{\gamma}_{M-k+1},\ldots,1-\rho_{1_M}+\bar{\gamma}_M\right). \end{cases} \tag{68}$$

Para determinar o valor mínimo em cada linha de (68), note que é necessário conhecer o valor de $\bar{\gamma}_i$ se $k \neq 1$. Então, generalizadamente, para $k \neq 1$ pode-se escrever que

$$\begin{array}{ll} \bar{\gamma}_1 > \rho_{2_k}+\rho_{1_1},\ldots,\bar{\gamma}_k > \rho_{2_k}+\rho_{1_k} & \text{, ou} \\ \vdots & \text{, ou} \\ \bar{\gamma}_{M-k+1} > \rho_{2_k}+\rho_{1_{M-k+1}},\ldots,\bar{\gamma}_M > \rho_{2_k}+\rho_{1_M}. & \end{array} \tag{69}$$

Foi mencionada a necessidade de se satisfazer pelo menos uma das $\binom{M}{M-k+1}$ linhas de (64), cada uma com $M-k+1$ condições a serem atendidas, para se ter $\bar{Q}_{\mathrm{fa}} = 0$ em (62). Da mesma forma, também foi mencionada a necessidade de se satisfazer pelo menos uma das $\binom{M}{k}$ linhas de (66), cada uma com k condições a serem atendidas, para se ter $\bar{Q}_{\mathrm{d}} = 1$ em (63). Porém, veja que atender a essas restrições, e então ter $\bar{Q}_{\mathrm{fa}} = 0$ em (62) e $\bar{Q}_{\mathrm{d}} = 1$ em (63), equivale a atender pelo menos k condições em pelo menos uma das linhas de (69).

Portanto, a SNRW generalizada, referente ao i-ésimo SU, pode ser obtida admitindo-se a igualdade em cada inequação mostrada em (69), o que leva à

$$\gamma_{\mathrm{w}_i} = \rho_{2_k}+\rho_{1_i}. \tag{70}$$

Se todos os SUs tiverem SNR iguais, isto é, se $\bar{\gamma}_i = \bar{\gamma}$ para todo i, e se $k = 1$, de (68) tem-se

$$\bar{\gamma} > \rho_{2_1}+\rho_{1_1}\text{,ou } \bar{\gamma} > \rho_{2_1}+\rho_{1_2}\text{,ou,}\ldots\text{,ou } \bar{\gamma} > \rho_{2_1}+\rho_{1_{M-1}}\text{,ou } \bar{\gamma} > \rho_{2_1}+\rho_{1_M}. \tag{71}$$

Já que basta que uma das inequações em (71) seja satisfeita para que a restrição sobre a SNR seja satisfeita, (71) pode ser simplificada para $\bar{\gamma} > \rho_{2_1}+\rho_{1_M}$ para que se selecione a menor SNRW generalizada. Assim, com $k = 1$, essa SNRW generalizada pode ser escrita como

$$\gamma_{\mathrm{w}} = \rho_{2_1}+\rho_{1_M}. \tag{72}$$

Da mesma forma, isto é, admitindo-se $\bar{\gamma}_i = \bar{\gamma}$ para todo i, para $1 < k < M$, de (68) tem-se

$$\bar{\gamma} > \rho_{2_k}+\rho_{1_1}\text{,ou } \bar{\gamma} > \rho_{2_k}+\rho_{1_2}\text{,ou,}\ldots\text{,ou } \bar{\gamma} > \rho_{2_k}+\rho_{1_{M-k+1}}, \tag{73}$$

o que leva a simplificação para $\bar{\gamma} > \rho_{2_k}+\rho_{1_{M-k+1}}$ se $k < M-k+1$ ou para $\bar{\gamma} > \rho_{2_k}+\rho_{1_k}$ se $k \geq M-k+1$ para se obter a menor SNRW generalizada. Por isso, para $1 < k < M$, tem-se

$$\gamma_{\mathrm{w}} = \begin{cases} \rho_{2_k}+\rho_{1_{M-k+1}} \text{ se } k < M-k+1 \text{ ou} & (74) \\ \rho_{2_k}+\rho_{1_k} \text{ se } k \geq M-k+1. & (75) \end{cases}$$

Desta vez, para $\bar{\gamma}_i = \bar{\gamma}$ para todo i e $k = M$, de (68) tem-se $\bar{\gamma} > \rho_{2_M} + \rho_{1_1}$, resultando em

$$\gamma_{\mathrm{w}} = \rho_{2_M} + \rho_{1_1}. \tag{76}$$

Por sua vez, admitindo-se $\bar{\gamma}_i = \bar{\gamma}$ para todo i e $k = \lfloor M/2+1 \rfloor$ (regra MAJ), veja que $M-k+1 = k-1 = M/2$ para M par ou $M-k+1 = \lfloor M/2+1 \rfloor$ para M ímpar. Por isso a simplificação a partir de (73) fica como $\bar{\gamma} > \rho_{2_{\lfloor M/2+1 \rfloor}} + \rho_{1_{M/2}}$ para M par ou como $\bar{\gamma} > \rho_{2_{\lfloor M/2+1 \rfloor}} + \rho_{1_{\lfloor M/2+1 \rfloor}}$ para M ímpar. Então,

$$\gamma_{\mathrm{w}} = \begin{cases} \rho_{2_{M/2+1}} + \rho_{1_{M/2}} \text{ se } M \text{ for par, ou} & (77) \\ \rho_{2_{\lfloor M/2+1 \rfloor}} + \rho_{1_{\lfloor M/2+1 \rfloor}} \text{ se } M \text{ for ímpar.} & (78) \end{cases}$$

Agora, considerando a situação de igualdade entre os fatores de NU, isto é, $\rho_i = \rho$ para todo i, as expressões anteriores para a SNRW generalizada se convertem para

$$\gamma_{\mathrm{w}} = \rho_2 + \rho_1. \tag{79}$$

As expressões particularizadas para a SNRW conforme as combinações entre os modelos tradicional e novo de fonte e intervalo de NU são computadas via SNRW generalizadas. Por exemplo, substituindo ρ_{2_k} por ρ_k e ρ_{1_i} por ρ_i em (70), com $0 \leq \{\rho_k,\rho_i\} < 1$, tem-se a SNRW do i-ésimo SU na combinação da fonte tradicional de NU com o novo intervalo de NU como

$$\gamma_{\mathrm{w}_i} = \rho_k + \rho_i. \tag{80}$$

Para $\rho_i = \rho$ para todo i, essa SNRW se torna $\gamma_{\mathrm{w}_i} = 2\rho_i$, sendo consistente com (52). Por outro lado, para o caso particular da fonte e intervalo tradicionais de NU, substituindo ρ_{2_k} por $\rho_k - 1$ e ρ_{1_i} por $(\rho_i - 1)/\rho_i$ em (70), com $\{\rho_k,\rho_i\} \geq 1$, tem-se

$$\gamma_{\mathrm{w}_i} = \frac{\rho_k\rho_i - 1}{\rho_i}, \tag{81}$$

que se torna a SNRW do i-ésimo SU no NCSS na combinação da fonte tradicional de NU com o intervalo tradicional de NU quando $\rho_i = \rho$ para todo i, ou seja, $\gamma_{\mathrm{w}_i} = (\rho_i^2 - 1)/\rho_i$ (vide (53), ou Tabela 3).

Nova fonte combinada com intervalo tradicional/novo de NU No diz respeito à probabilidade de falso alarme, para se ter $\bar{Q}_{\mathrm{fa}} = 0$ em (62), observe com a ajuda da inequação $\lambda > (1-\rho_{1_i})^{-1}$, definida acerca de (57), que para $i = 1,\ldots,M$,

$$\lambda > \begin{cases} (1-\rho_{1_i})^{-1},\mathrm{e},\ldots,\mathrm{e}\ (1-\rho_{1_{M-k+1}})^{-1} & , \text{ ou} \\ \vdots & , \text{ ou} \\ (1-\rho_{1_k})^{-1},\mathrm{e},\ldots,\mathrm{e}\ (1-\rho_{1_M})^{-1}. & \end{cases} \tag{82}$$

Visto que só se pode alcançar $\bar{Q}_{\mathrm{fa}} = 0$ em (62) se pelo menos uma das $\binom{M}{M-k+1}$ linhas de (82), com $M-k+1$ condições em cada linha, for satisfeita, pode-se reescrever λ em função do resultado mínimo entre os máximos das linhas de (82) (cada máximo advindo de uma das linhas). Fazendo isso, tem-se

$$\lambda > \min \begin{pmatrix} \max\left[(1-\rho_{1_i})^{-1},\ldots,(1-\rho_{1_{M-k+1}})^{-1}\right] & , \\ \vdots & , \\ \max\left[(1-\rho_{1_k})^{-1},\ldots,(1-\rho_{1_M})^{-1}\right] & \end{pmatrix}. \tag{83}$$

Da mesma forma, no que diz respeito à probabilidade de detecção, para se ter $\bar{Q}_{\mathrm{d}} = 1$ em (63), observe com a ajuda da inequação $\lambda < (1+\bar{\gamma}_1)/(1+\rho_{2_1})$, definida acerca de (58), que para $i = 1, \ldots, M$,

$$\lambda < \begin{cases} \frac{1+\bar{\gamma}_1}{1+\rho_{2_1}}, \text{e}, \ldots, \text{e } \frac{1+\bar{\gamma}_k}{1+\rho_{2_k}} & , \text{ ou} \\ \vdots & , \text{ ou} \\ \frac{1+\bar{\gamma}_{M-k+1}}{1+\rho_{2_{M-k+1}}}, \text{e}, \ldots, \text{e } \frac{1+\bar{\gamma}_M}{1+\rho_{2_M}} & . \end{cases} \tag{84}$$

Visto que só se pode alcançar $\bar{Q}_{\mathrm{d}} = 1$ em (63) se pelo menos uma das $\binom{M}{k}$ linhas de (84), com k condições em cada linha, for satisfeita, pode-se reescrever λ em (84) como

$$\lambda < \begin{cases} \min\left(\frac{1+\bar{\gamma}_1}{1+\rho_{2_1}}, \ldots, \frac{1+\bar{\gamma}_k}{1+\rho_{2_k}}\right) & , \text{ ou} \\ \vdots & , \text{ ou} \\ \min\left(\frac{1+\bar{\gamma}_{M-k+1}}{1+\rho_{2_{M-k+1}}}, \ldots, \frac{1+\bar{\gamma}_M}{1+\rho_{2_M}}\right). \end{cases} \tag{85}$$

Admitindo por ora apenas a disposição dos fatores de NU em ordem decrescente, isto é, $\rho_j > \rho_{j+1}$ para $j = 1, \ldots, M-1$, note a partir de (83) e (85) que

$$\frac{1}{1-\rho_{1_k}} < \begin{cases} \min\left(\frac{1+\bar{\gamma}_1}{1+\rho_{2_1}}, \ldots, \frac{1+\bar{\gamma}_k}{1+\rho_{2_k}}\right) & , \text{ ou} \\ \vdots & , \text{ ou} \\ \min\left(\frac{1+\bar{\gamma}_{M-k+1}}{1+\rho_{2_{M-k+1}}}, \ldots, \frac{1+\bar{\gamma}_M}{1+\rho_{2_M}}\right). \end{cases} \tag{86}$$

Para determinar o valor mínimo em cada linha de (86), note que é necessário conhecer o valor de $\bar{\gamma}_i$ se $k \neq 1$. Então, generalizadamente, para $k \neq 1$ pode-se escrever que

$$\begin{cases} \bar{\gamma}_1 > \frac{\rho_{1_k}+\rho_{2_1}}{1-\rho_{1_k}}, \ldots, \bar{\gamma}_k > \frac{\rho_{1_k}+\rho_{2_k}}{1-\rho_{1_k}} & , \text{ ou} \\ \vdots & , \text{ ou} \\ \bar{\gamma}_{M-k+1} > \frac{\rho_{1_k}+\rho_{2_{M-k+1}}}{1-\rho_{1_k}}, \ldots, \bar{\gamma}_M > \frac{\rho_{1_k}+\rho_{2_M}}{1-\rho_{1_k}}. \end{cases} \tag{87}$$

Aquelas mesmas interpretações feitas acerca das $\binom{M}{M-k+1}$ linhas de (64), com $M-k+1$ condições em cada linha, acerca das $\binom{M}{k}$ linhas de (66), com k condições em cada linha, e acerca de (69) para se atingir $\bar{Q}_{\mathrm{fa}} = 0$ em (62) e $\bar{Q}_{\mathrm{d}} = 1$ em (63) também se aplicam aqui. Especificamente, satisfazer ambos os conjuntos de restrições acerca de (82) e (84) para se atingir $\bar{Q}_{\mathrm{fa}} = 0$ em (62) e $\bar{Q}_{\mathrm{d}} = 1$ em (63) é equivalente a satisfazer pelo menos um grupo de k condições em pelo menos uma das linhas de (87).

Agora, considerando a igualdade entre os fatores de NU, isto é, $\rho_i = \rho$ para todo i, em (87), uma SNRW generalizada para o i-ésimo SU no caso da nova fonte de NU pode ser dada por

$$\gamma_{\mathrm{w}_i} = \frac{\rho_{1_k} + \rho_{2_i}}{1-\rho_{1_k}}. \tag{88}$$

Considerando-se SUs com SNR idênticas, isto é, $\bar{\gamma}_i = \bar{\gamma}$ para todo i, e $k = 1$, a partir de (86) pode-se escrever que

$$\bar{\gamma} > \frac{\rho_{1_1}+\rho_{2_1}}{1-\rho_{1_1}}, \text{ou } \bar{\gamma} > \frac{\rho_{1_1}+\rho_{2_2}}{1-\rho_{1_1}}, \text{ou}, \ldots, \text{ou } \bar{\gamma} > \frac{\rho_{1_1}+\rho_{2_{M-1}}}{1-\rho_{1_1}}, \text{ou } \bar{\gamma} > \frac{\rho_{1_1}+\rho_{2_M}}{1-\rho_{1_1}}. \tag{89}$$

Note que basta que uma das inequações em (89) seja satisfeita para que a condição sobre a SNR também seja satisfeita. Por isso, com o intuito de levar à menor SNRW generalizada possível, (89) pode ser simplificada como $\bar{\gamma} > (\rho_{1_1} + \rho_{2_M})/(1 - \rho_{1_1})$. Logo, para $k = 1$ tem-se a SNRW generalizada dada por

$$\gamma_{\mathrm{w}} = \frac{\rho_{1_1} + \rho_{2_M}}{1 - \rho_{1_1}}. \tag{90}$$

Para $1 < k < M$ as inequações em (86) podem ser reescritas como

$$\bar{\gamma} > \tfrac{\rho_{1_k}+\rho_{2_1}}{1-\rho_{1_k}}, \text{ou } \bar{\gamma} > \tfrac{\rho_{1_k}+\rho_{2_2}}{1-\rho_{1_k}}, \text{ou}, \ldots, \text{ou } \bar{\gamma} > \tfrac{\rho_{1_k}+\rho_{2_{M-k+1}}}{1-\rho_{1_k}}, \tag{91}$$

de onde se pode estabelecer a seguinte simplificação para que seja possível levar à menor SNRW generalizada: $\bar{\gamma} > (\rho_{1_k} + \rho_{2_{M-k+1}})/(1 - \rho_{1_k})$ para $k < M - k + 1$ ou $\bar{\gamma} > (\rho_{1_k} + \rho_{2_k})/(1 - \rho_{1_k})$ para $k \geq M - k + 1$. Assim, para $1 < k < M$, tem-se

$$\gamma_{\mathrm{w}} = \begin{cases} (\rho_{1_k} + \rho_{2_{M-k+1}})(1 - \rho_{1_k})^{-1}, \text{ se } k < M - k + 1 & (92) \\ (\rho_{1_k} + \rho_{2_k})(1 - \rho_{1_k})^{-1}, \text{ se } k \geq M - k + 1. & (93) \end{cases}$$

Por fim, veja que para $k = M$, (86) se torna $\bar{\gamma} > (\rho_{1_M} + \rho_{2_1})(1 - \rho_{1_M})^{-1}$, resultando em

$$\gamma_{\mathrm{w}} = \frac{\rho_{1_M} + \rho_{2_1}}{1 - \rho_{1_M}}. \tag{94}$$

Semelhante ao que foi feito no caso da fonte tradicional de NU, no caso da nova fonte de NU, para simplificar (73) na regra MAJ, (91) pode ser escrita como $\bar{\gamma} > (\rho_{1_{M/2+1}}+\rho_{2_{M/2}})/(1-\rho_{1_{M/2+1}})$ quando M for par, ou como $\bar{\gamma} > (\rho_{1_{\lfloor M/2+1 \rfloor}} + \rho_{2_{\lfloor M/2+1 \rfloor}})/(1 - \rho_{1_{\lfloor M/2+1 \rfloor}})$ quando M for ímpar, o que implica em

$$\gamma_{\mathrm{w}} = \begin{cases} (\rho_{1_{M/2+1}} + \rho_{2_{M/2}})(1 - \rho_{1_{M/2+1}})^{-1} \text{ se } M \text{ for par, ou} & (95) \\ (\rho_{1_{\lfloor M/2+1 \rfloor}} + \rho_{2_{\lfloor M/2+1 \rfloor}})(1 - \rho_{1_{\lfloor M/2+1 \rfloor}})^{-1} \text{ se } M \text{ for ímpar.} & (96) \end{cases}$$

Se $\rho_i = \rho$ para todo i, as expressões anteriores para a SNRW generalizada se tornam

$$\gamma_{\mathrm{w}} = \frac{\rho_1 + \rho_2}{1 - \rho_1}. \tag{97}$$

Admitindo-se $\rho_{2_i} = \rho_i - 1$ e $\rho_{1_k} = (\rho_k - 1)/\rho_k$ em (88), $\{\rho_k,\rho_i\} \geq 1$, tem-se

$$\gamma_{\mathrm{w}_i} = \rho_k \rho_i - 1, \tag{98}$$

que corresponde ao caso particular da SNRW no i-ésimo SU quando a nova fonte de NU é combinada com o intervalo tradicional de NU. Por outro lado, com $\rho_{2_k} = \rho_k$ e $\rho_{1_i} = \rho_i$ em (88), $0 \leq \{\rho_k,\rho_i\} < 1$, veja que a SNRW do i-ésimo SU quando a nova fonte de NU é combinada com o novo intervalo de NU pode ser escrita como

$$\gamma_{\mathrm{w}_i} = \frac{\rho_k + \rho_i}{1 - \rho_i}. \tag{99}$$

A Tabela 4 reúne as expressões para a SNRW no i-ésimo SU, γ_{w_i}, no CSS via HD sob a regra k-em-M para todas as combinações envolvendo os modelos tradicional e novo de fonte e intervalo de NU.

Tabela 4: SNRW da ED no i-ésimo SU, γ_{w_i}, no CSS via HD sob a regra k-em-M para as quatro combinações envolvendo os modelos tradicional e novo de fonte e intervalo de NU.

	Fonte Tradicional	Nova Fonte
Intervalo Tradicional	$\gamma_{\mathrm{w}_i} = \frac{\rho_k\rho_i - 1}{\rho_i}, \{\rho_i,\rho_k\} \geq 1,$ @(81)	$\gamma_{\mathrm{w}_i} = \rho_k\rho_i - 1, \{\rho_i,\rho_k\} \geq 1,$ @(98)
Novo Intervalo	$\gamma_{\mathrm{w}_i} = \rho_k + \rho_i, 0 \leq \{\rho_i,\rho_k\} < 1,$ @(80)	$\gamma_{\mathrm{w}_i} = \frac{\rho_k + \rho_i}{1-\rho_i}, 0 \leq \{\rho_i,\rho_k\} < 1,$ @(99)

5.4.2.2 Fusão de decisão suave

Cômputo tradicional da estatística de teste

Fonte tradicional combinada com intervalo tradicional/novo de NU Quando não há NU, as médias, variâncias e métricas de desempenho global referentes ao cômputo tradicional da estatística de teste T em (32) são dadas por (33), (34), (35) e (36), respectivamente. No entanto, quando há NU, deve-se substituir nessas expressões o valor nominal da variância de ruído (σ_{v}^2) pelo valor real (estimado) ($\hat{\sigma}_{\mathrm{v}}^2$). Fazendo isso, essas médias e variâncias se tornam, respectivamente,

$$\hat{\mu}_0 = \frac{1}{M}\sum_{i=1}^{M} \hat{\sigma}_{\mathrm{v}_i}^2 \text{ e } \hat{\sigma}_0^2 = \frac{1}{M^2 N}\sum_{i=1}^{M} \hat{\sigma}_{\mathrm{v}_i}^4 \tag{100}$$

sob a hipótese $\mathcal{H}_0$, ou por

$$\hat{\mu}_1 = \frac{1}{M}\sum_{i=1}^{M} \left(\hat{\sigma}_{\mathrm{v}_i}^2 + \sigma_{\mathrm{s}_i}^2\right) \text{ e } \hat{\sigma}_1^2 = \frac{1}{M^2 N}\sum_{i=1}^{M} \left(\hat{\sigma}_{\mathrm{v}_i}^2 + \sigma_{\mathrm{s}_i}^2\right)^2 \tag{101}$$

sob a hipótese $\mathcal{H}_1$.

Para um dado valor específico de $\hat{\sigma}_{\mathrm{v}_i}^2$, as métricas de desempenho global, Q_{fa} e Q_{d}, podem ser descritas como

$$Q_{\mathrm{fa}} = Q\left(\frac{\tau - \hat{\mu}_0}{\hat{\sigma}_0}\right) = Q\left(\frac{M\tau - \sum_{i=1}^{M} \hat{\sigma}_{\mathrm{v}_i}^2}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M} \hat{\sigma}_{\mathrm{v}_i}^4}}\right) \tag{102}$$

e

$$Q_{\mathrm{d}} = Q\left(\frac{\tau - \hat{\mu}_1}{\hat{\sigma}_1}\right) = Q\left(\frac{M\tau - \sum_{i=1}^{M} \left(\hat{\sigma}_{\mathrm{v}_i}^2 + \sigma_{\mathrm{s}_i}^2\right)}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M} \left(\hat{\sigma}_{\mathrm{v}_i}^2 + \sigma_{\mathrm{s}_i}^2\right)^2}}\right), \tag{103}$$

com $\tau = (\lambda/M)\sum_{i=1}^{M} \sigma_{\mathrm{v}_i}^2$.

Integrando Q_{fa} e Q_{d} de acordo com a PDF da NU em (44), têm-se[1]

$$\bar{Q}_{\text{fa}} = \int_{\mathbf{a}}^{\mathbf{b}} Q\left(\frac{M\tau - \sum_{i=1}^{M} x_i}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M} x_i^2}}\right) f_{\hat{\sigma}^2_{\text{v}_M},\ldots,\hat{\sigma}^2_{\text{v}_1}}(\mathbf{x}) d^M\mathbf{x} \tag{104}$$

e

$$\bar{Q}_{\text{d}} = \int_{\mathbf{a}}^{\mathbf{b}} Q\left(\frac{M\tau - \sum_{i=1}^{M}(x_i + \sigma^2_{\text{s}_i})}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}(x_i + \sigma^2_{\text{s}_i})^2}}\right) f_{\hat{\sigma}^2_{\text{v}_M},\ldots,\hat{\sigma}^2_{\text{v}_1}}(\mathbf{x}) d^M\mathbf{x}. \tag{105}$$

Ao aplicar (47) em (104) e em (105) para garantir $\bar{Q}_{\text{fa}} = 0$ e $\bar{Q}_{\text{d}} = 1$ em todo o intervalo de NU, veja que as seguintes condições devem ser atendidas

$$\sum_{i=1}^{M}(1 + \rho_{2_i})\sigma^2_{\text{v}_i} < \lambda < \sum_{i=1}^{M}(1 - \rho_{1_i} + \bar{\gamma}_i)\sigma^2_{\text{v}_i}. \tag{106}$$

Para $\sigma^2_{\text{v}_i} = \sigma^2_{\text{v}}$ para todo i, pode-se escrever $\sum_{i=1}^{M}\bar{\gamma}_i > \sum_{i=1}^{M}(\rho_{2_i} + \rho_{1_i})$ a partir de (106). Ainda, para $\sigma^2_{\text{v}_i} = \sigma^2_{\text{v}}$ e $\bar{\gamma}_i = \bar{\gamma}$ para todo i em (106), a SNRW generalizada pode ser escrita como

$$\gamma_{\text{w}} = \frac{1}{M}\sum_{i=1}^{M}(\rho_{2_i} + \rho_{1_i}). \tag{107}$$

Agora, particularizando o resultado de acordo com o intervalo de NU, veja que substituindo $\rho_{1_i} = \rho_{2_i} = \rho_i$, $0 \leq \rho_i < 1$, em (107), resulta em

$$\gamma_{\text{w}} = \frac{2}{M}\sum_{i=1}^{M}\rho_i, \tag{108}$$

sendo esta a SNRW no CSS via SD quando a fonte tradicional de NU é combinada com o novo intervalo de NU. Por outro lado, com $\rho_{1_i} = (\rho_i - 1)/\rho_i$ e $\rho_{2_i} = \rho_i - 1$, $\rho_i \geq 1$, em (107), a SNRW quando a fonte tradicional de NU é combinada com o intervalo tradicional de NU é dada por

$$\gamma_{\text{w}} = \frac{1}{M}\sum_{i=1}^{M}\frac{\rho_i^2 - 1}{\rho_i}. \tag{109}$$

O número de amostras N requerido para se atingir dado desempenho global em termos de Q_{fa} e Q_{d} é deduzido isolando-se o limiar de decisão global, τ, em (102) e substituindo o resultado em (103) admitindo os piores casos em termos de NU, ou seja, admitindo $\hat{\sigma}^2_{\text{v}_i} = (1 + \rho_{2_i})\sigma^2_{\text{v}_i}$ em (102) e $\hat{\sigma}^2_{\text{v}_i} = (1 - \rho_{1_i})\sigma^2_{\text{v}_i}$ em (103). Então, com $\bar{\gamma}_i = \bar{\gamma}$ para todo i, tem-se

$$N = \frac{\left[Q^{-1}(Q_{\text{fa}})\sqrt{\sum_{i=1}^{M}(1 + \rho_{2_i})^2} - Q^{-1}(Q_{\text{d}})\sqrt{\sum_{i=1}^{M}(1 - \rho_{1_i} + \bar{\gamma})^2}\right]^2}{M^2\left[\bar{\gamma} - \frac{1}{M}\sum_{i=1}^{M}(\rho_{2_i} + \rho_{1_i})\right]^2}, \tag{110}$$

[1]Dada a função de um vetor $f(\cdot)$ e os vetores de constantes $\mathbf{a} = (a_1, \ldots, a_M)$ e $\mathbf{b} = (b_1, \ldots, b_M)$, utiliza-se a notação simplificada $\int_{\mathbf{a}}^{\mathbf{b}} f(\mathbf{x}) d^M\mathbf{x}$ para a escrita da integral múltipla $\int_{a_1}^{b_1}\cdots\int_{a_M}^{b_M} f(x_1, \ldots, x_M) dx_M \cdots dx_1$.

em que $N \to \infty$ à medida que $\bar{\gamma} \to (1/M)\sum_{i=1}^{M}(\rho_{2_i}+\rho_{1_i})$, o que está de acordo com a expressão para a SNRW apresentada em (107). Por isso, é claro que aqui a condição estabelecida pela inequação dada por $\bar{\gamma} > (1/M)\sum_{i=1}^{M}(\rho_{2_i}+\rho_{1_i})$ deve ser atendida.

Nova fonte combinada com intervalo tradicional/novo de NU Fazendo as análises com a nova fonte de NU, as médias e variâncias da estatística de teste sob $\mathcal{H}_0$ e $\mathcal{H}_1$ são exatamente aquelas em (33) e (34), respectivamente, e o valor estimado da variância de ruído é incorporado no cômputo do limiar de decisão, que fica como $\hat{\tau} = (\lambda/M)\sum_{i=1}^{M}\hat{\sigma}^2_{\mathrm{v}_i}$. Portanto, para um dado valor específico de $\hat{\sigma}^2_{\mathrm{v}_i}$, Q_{fa} e Q_{d} podem ser respectivamente escritas como

$$Q_{\mathrm{fa}} = Q\left(\frac{\hat{\tau}-\mu_0}{\sigma_0}\right) = Q\left(\frac{M\hat{\tau} - \sum_{i=1}^{M}\sigma^2_{\mathrm{v}_i}}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}\sigma^4_{\mathrm{v}_i}}}\right) \tag{111}$$

e

$$Q_{\mathrm{d}} = Q\left(\frac{\hat{\tau}-\mu_1}{\sigma_1}\right) = Q\left(\frac{M\hat{\tau} - \sum_{i=1}^{M}(\sigma^2_{\mathrm{v}_i}+\sigma^2_{\mathrm{s}_i})}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}(\sigma^2_{\mathrm{v}_i}+\sigma^2_{\mathrm{s}_i})^2}}\right). \tag{112}$$

Fazendo a integração de Q_{fa} e Q_{d} conforme a PDF da NU em (44), pode-se escrever

$$\bar{Q}_{\mathrm{fa}} = \int_{\mathbf{a}}^{\mathbf{b}} Q\left(\frac{\lambda\sum_{i=1}^{M}x_i - \sum_{i=1}^{M}\sigma^2_{\mathrm{v}_i}}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}\sigma^4_{\mathrm{v}_i}}}\right) f_{\hat{\sigma}^2_{\mathrm{v}_M},\ldots,\hat{\sigma}^2_{\mathrm{v}_1}}(\mathbf{x})d^M\mathbf{x} \tag{113}$$

e

$$\bar{Q}_{\mathrm{d}} = \int_{\mathbf{a}}^{\mathbf{b}} Q\left(\frac{\lambda\sum_{i=1}^{M}x_i - \sum_{i=1}^{M}(\sigma^2_{\mathrm{v}_i}+\sigma^2_{\mathrm{s}_i})}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}(\sigma^2_{\mathrm{v}_i}+\sigma^2_{\mathrm{s}_i})^2}}\right) f_{\hat{\sigma}^2_{\mathrm{v}_M},\ldots,\hat{\sigma}^2_{\mathrm{v}_1}}(\mathbf{x})d^M\mathbf{x}. \tag{114}$$

Aplicando (47) em (113) e em (114) para garantir $\bar{Q}_{\mathrm{fa}} = 0$ e $\bar{Q}_{\mathrm{d}} = 1$ em todo o intervalo de NU, veja que $\left(\sum_{i=1}^{M}\sigma^2_{\mathrm{v}_i}\right)/\left(\sum_{i=1}^{M}(1-\rho_{1_i})\sigma^2_{\mathrm{v}_i}\right) < \lambda < \left(\sum_{i=1}^{M}(1+\bar{\gamma}_i)\sigma^2_{\mathrm{v}_i}\right)/\left(\sum_{i=1}^{M}(1+\rho_{2_i})\sigma^2_{\mathrm{v}_i}\right)$, o que leva à

$$\sum_{i=1}^{M}(1+\bar{\gamma}_i)\sigma^2_{\mathrm{v}_i} > \frac{\sum_{i=1}^{M}\sigma^2_{\mathrm{v}_i}\sum_{i=1}^{M}(1+\rho_{2_i})\sigma^2_{\mathrm{v}_i}}{\sum_{i=1}^{M}(1-\rho_{1_i})\sigma^2_{\mathrm{v}_i}}. \tag{115}$$

Se $\sigma^2_{\mathrm{v}_i} = \sigma^2_{\mathrm{v}}$ para todo i, pode-se escrever $\sum_{i=1}^{M}\bar{\gamma}_i = M\sum_{i=1}^{M}(\rho_{2_i}+\rho_{1_i})/\sum_{i=1}^{M}(1-\rho_{1_i})$ a partir de (115). Ainda, se $\sigma^2_{\mathrm{v}_i} = \sigma^2_{\mathrm{v}}$ e $\bar{\gamma}_i = \bar{\gamma}$ para todo i em (115), $\bar{\gamma} > \sum_{i=1}^{M}(\rho_{2_i}+\rho_{1_i})/\sum_{i=1}^{M}(1-\rho_{1_i})$.

Assim, a SNRW generalizada em cada SU é dada por

$$\gamma_{\mathrm{w}} = \frac{\sum_{i=1}^{M}(\rho_{2_i} + \rho_{1_i})}{\sum_{i=1}^{M}(1 - \rho_{1_i})}. \tag{116}$$

Substituindo $\rho_{1_i} = (\rho_i - 1)/\rho_i$ e $\rho_{2_i} = \rho_i - 1$ em (116), $\rho_i \geq 1$, essa SNRW generalizada torna-se para o caso particular quando a nova fonte de NU é combinada com o intervalo tradicional de NU, resultando em

$$\gamma_{\mathrm{w}} = \frac{\sum_{i=1}^{M} \frac{\rho_i^2 - 1}{\rho_i}}{\sum_{i=1}^{M} \frac{1}{\rho_i}}. \tag{117}$$

Por outro lado, com $\rho_{1_i} = \rho_{2_i} = \rho_i$ em (116), $0 \leq \rho_i < 1$, essa SNRW generalizada torna-se para o caso particular quando a nova fonte de NU é combinada com o novo intervalo de NU, isto é,

$$\gamma_{\mathrm{w}} = \frac{2\sum_{i=1}^{M} \rho_i}{\sum_{i=1}^{M}(1 - \rho_i)}. \tag{118}$$

O número de amostras requerido para se atingir o desejado para as métricas de desempenho pode ser obtido extraindo λ de (111) e (112) admitindo os piores casos em termos de NU, isto é, admitindo $\hat{\sigma}^2_{\mathrm{v}_i} = (1 - \rho_{1_i})\sigma^2_{\mathrm{v}_i}$ em (111) e $\hat{\sigma}^2_{\mathrm{v}_i} = (1 + \rho_{2_i})\sigma^2_{\mathrm{v}_i}$ em (112), e igualando os resultados. Fazendo isso, para $\sigma^2_{\mathrm{v}_i} = \sigma^2_{\mathrm{v}}$ e $\bar{\gamma}_i = \bar{\gamma}$ para todo i, essas operações levam à

$$N = \frac{\left[Q^{-1}(Q_{\mathrm{fa}})\frac{\sum_{i=1}^{M}(1+\rho_{2_i})}{\sum_{i=1}^{M}(1-\rho_{1_i})} - Q^{-1}(Q_{\mathrm{d}})(1+\bar{\gamma})\right]^2}{M\left[\bar{\gamma} - \frac{\sum_{i=1}^{M}(\rho_{2_i}+\rho_{1_i})}{\sum_{i=1}^{M}(1-\rho_{1_i})}\right]^2}, \tag{119}$$

em que $N \to \infty$ à medida que $\bar{\gamma} \to \sum_{i=1}^{M}(\rho_{2_i} + \rho_{1_i})/\sum_{i=1}^{M}(1 - \rho_{1_i})$, o que é consistente com o resultado apresentado em (116). É importante notar que é necessário satisfazer a condição dada pela inequação $\bar{\gamma} > \sum_{i=1}^{M}(\rho_{2_i} + \rho_{1_i})/\sum_{i=1}^{M}(1 - \rho_{1_i})$ neste caso.

A Tabela 5 reúne as expressões para a SNRW em cada SU, γ_{w}, no CSS via SD para todas as combinações envolvendo os modelos tradicional e novo de fonte e intervalo de NU.

Novo cômputo da estatística de teste

Fonte tradicional combinada com intervalo tradicional/novo de NU Com o novo cômputo para a estatística de teste para o CSS via SD em (38), a variância de ruído que irá compor o i-ésimo peso da soma ponderada das energias parciais $\{T_i\}$ provém do limiar de decisão, que é dado por $\tau = (\lambda/M)\sum_{i=1}^{M}\sigma^2_{\mathrm{v}_i}$ no cômputo tradicional da estatística de teste em (32), mas passa a ser dado por $\tau = \lambda$ neste caso. Note então que a i-ésima variância que

Tabela 5: ANRW da ED em cada SU, $\gamma_{\rm w}$, no CSS via SD e cômputo tradicional de T em (32) para as quatro combinações envolvendo as fontes e intervalos tradicional e novo de NU.

	Fonte Tradicional	Nova Fonte
Intervalo Tradicional	$\gamma_{\rm w}=\frac{1}{M}\sum_{i=1}^{M}\frac{\rho_i^2-1}{\rho_i}, \rho_i\geq 1,$ @(109)	$\gamma_{\rm w}=\frac{\sum_{i=1}^{M}(\rho_i^2-1)/\rho_i}{\sum_{i=1}^{M}1/\rho_i}, \rho_i\geq 1,$ @(117)
Novo Intervalo	$\gamma_{\rm w}=\frac{2}{M}\sum_{i=1}^{M}\rho_i, 0\leq\rho_i<1,$ @(108)	$\gamma_{\rm w}=\frac{2\sum_{i=1}^{M}\rho_i}{\sum_{i=1}^{M}(1-\rho_i)}, 0\leq\rho_i<1,$ @(118)

estava presente no limiar de decisão é aquela que se encontra nos denominadores de (38). Note ainda que essa variância representa o valor exato (esperado), e não o valor estimado, o que está em concordância com a admissão da fonte tradicional de NU. Portanto, sob NU, as médias e variâncias da nova estatística de teste em (38) são computadas, respectivamente, como

$$\hat{\mu}_0=\frac{1}{M}\sum_{i=1}^{M}\frac{\hat{\sigma}^2_{{\rm v}_i}}{\sigma^2_{{\rm v}_i}} \text{ e } \hat{\sigma}_0^2=\frac{1}{M^2N}\sum_{i=1}^{M}\frac{\hat{\sigma}^4_{{\rm v}_i}}{\sigma^4_{{\rm v}_i}}, \tag{120}$$

sob a hipótese $\mathcal{H}_0$, ou como

$$\hat{\mu}_1=\frac{1}{M}\sum_{i=1}^{M}\left(\frac{\hat{\sigma}^2_{{\rm v}_i}}{\sigma^2_{{\rm v}_i}}+\bar{\gamma}_i\right) \text{ e } \hat{\sigma}_1^2=\frac{1}{M^2N}\sum_{i=1}^{M}\left(\frac{\hat{\sigma}^2_{{\rm v}_i}}{\sigma^2_{{\rm v}_i}}+\bar{\gamma}_i\right)^2 \tag{121}$$

sob a hipótese $\mathcal{H}_1$.

Com isso, para um dado valor específico de $\hat{\sigma}^2_{{\rm v}_i}$ as métricas de desempenho global, $Q_{\rm fa}$ e $Q_{\rm d}$, são dadas, respectivamente, por

$$Q_{\rm fa}=Q\left(\frac{\tau-\hat{\mu}_0}{\hat{\sigma}_0}\right)=Q\left(\frac{M\tau-\sum_{i=1}^{M}\frac{\hat{\sigma}^2_{{\rm v}_i}}{\sigma^2_{{\rm v}_i}}}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}\frac{\hat{\sigma}^4_{{\rm v}_i}}{\sigma^4_{{\rm v}_i}}}}\right) \tag{122}$$

e

$$Q_{\rm d}=Q\left(\frac{\tau-\hat{\mu}_1}{\hat{\sigma}_1}\right)=Q\left(\frac{M\tau-\sum_{i=1}^{M}\left(\frac{\hat{\sigma}^2_{{\rm v}_i}}{\sigma^2_{{\rm v}_i}}+\bar{\gamma}_i\right)}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}\left(\frac{\hat{\sigma}^2_{{\rm v}_i}}{\sigma^2_{{\rm v}_i}}+\bar{\gamma}_i\right)^2}}\right). \tag{123}$$

Fazendo a integração dessas métricas conforme a PDF da NU em (44) para se ter os valores médios, leva, respectivamente, à

$$\bar{Q}_{\rm fa}=\int_{\mathbf{a}}^{\mathbf{b}}Q\left(\frac{M\lambda-\sum_{i=1}^{M}\frac{x_i}{\sigma^2_{{\rm v}_i}}}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}\frac{x_i^2}{\sigma^4_{{\rm v}_i}}}}\right)f_{\hat{\sigma}^2_{{\rm v}_M},\ldots,\hat{\sigma}^2_{{\rm v}_1}}(\mathbf{x})d^M\mathbf{x} \tag{124}$$

e

$$\bar{Q}_{\mathrm{d}} = \int_{\mathbf{a}}^{\mathbf{b}} Q\left(\frac{M\lambda - \sum_{i=1}^{M}\left(\frac{x_i}{\sigma^2_{\mathrm{v}_i}} + \bar{\gamma}_i\right)}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}\left(\frac{x_i}{\sigma^2_{\mathrm{v}_i}} + \bar{\gamma}_i\right)^2}} \right) f_{\hat{\sigma}^2_{\mathrm{v}_M},\ldots,\hat{\sigma}^2_{\mathrm{v}_1}}(\mathbf{x}) d^M\mathbf{x}. \tag{125}$$

Aplicando (47) em (124), a inequação $M\lambda - \sum_{i=1}^{M} x_i/\sigma^2_{\mathrm{v}_i} > 0$ deve ser respeitada para se ter $\bar{Q}_{\mathrm{fa}_i} = 0$. Logo, $\lambda > \frac{1}{M}\sum_{i=1}^{M} x_i/\sigma^2_{\mathrm{v}_i}$. Note a partir de (43) a necessidade de se adotar $x_i = b_i = (1+\rho_{2_i})\sigma^2_{\mathrm{v}_i}$ para garantir $\bar{Q}_{\mathrm{fa}} = 0$ em todo o intervalo de NU. Similarmente, aplicando (47) em (125), a inequação $M\lambda - \sum_{i=1}^{M}(x_i/\sigma^2_{\mathrm{v}_i} + \bar{\gamma}_i) < 0$ deve ser respeitada para se ter $\bar{Q}_{\mathrm{d}} = 1$. Logo, $\lambda < \frac{1}{M}\sum_{i=1}^{M}(x_i/\sigma^2_{\mathrm{v}_i} + \bar{\gamma}_i)$ com $x_i = a_i = (1-\rho_{1_i})\sigma^2_{\mathrm{v}_i}$ para assegurar $\bar{Q}_{\mathrm{d}} = 1$ em todo o intervalo de NU. Então, para assegurar $\bar{Q}_{\mathrm{fa}} = 0$ e $\bar{Q}_{\mathrm{d}} = 1$, veja que as inequações $\lambda > (1/M)\sum_{i=1}^{M}(1+\rho_{2_i})$ e $\lambda < (1/M)\sum_{i=1}^{M}(1-\rho_{1_i}+\bar{\gamma}_i)$ devem ser respectivamente respeitadas, o que resulta em satisfazer $\sum_{i=1}^{M}(1+\rho_{2_i}) < \lambda < \sum_{i=1}^{M}(1-\rho_{1_i}+\bar{\gamma}_i)$, ou

$$\sum_{i=1}^{M}\bar{\gamma}_i > \sum_{i=1}^{M}(\rho_{2_i}+\rho_{1_i}), \tag{126}$$

a partir do que a SNRW generalizada, γ_{w}, para o i-ésimo SU é obtida trocando o sinal da inequação por um sinal de igualdade. Então, $\gamma_{\mathrm{w}} = \sum_{i=1}^{M}(\rho_{2_i} + \rho_{1_i})$.

Interpretando (126), pode-se concluir que no CSS via ED com fusão SD alguns SUs podem até mesmo operar com sua SNR abaixo de respectiva SNRW contanto que a soma de todas as SNRs seja maior que γ_{w}, o que é uma importante vantagem dessa abordagem.

Admitindo-se $\bar{\gamma}_i = \bar{\gamma}$ para todo i em (126) leva à $\bar{\gamma} > \frac{1}{M}\sum_{i=1}^{M}(\rho_{2_i} + \rho_{1_i})$, a partir do que a SNRW generalizada relativa ao i-ésimo SU pode ser escrita como

$$\gamma_{\mathrm{w}} = \frac{1}{M}\sum_{i=1}^{M}(\rho_{2_i} + \rho_{1_i}), \tag{127}$$

resultando em $\gamma_{\mathrm{w}} = \rho_2 + \rho_1$ se $\rho_i = \rho$ para todo i, o que está de acordo com (79). Note que a expressão em (127) é idêntica àquela em (107). Logo, as deduções restantes também são idênticas às correspondentes expressões apresentadas posteriormente àquela em (127).

Por exemplo, admitindo-se $\rho_{1_i} = \rho_{2_i} = \rho_i$, $0 \leq \rho_i < 1$, em (127), essa SNRW generalizada é particularizada para a SNRW quando a fonte tradicional de NU é combinada com o novo intervalo de NU. Isto é,

$$\gamma_{\mathrm{w}} = \frac{2}{M}\sum_{i=1}^{M}\rho_i, \tag{128}$$

sendo convertida para a expressão em (52) se $M = 1$ ou $\rho_i = \rho$ para todo i. Por outro lado, utilizando-se $\rho_{1_i} = (\rho_i - 1)/\rho_i$ e $\rho_{2_i} = \rho_i - 1$, com $\rho_i \geq 1$, em (127), resulta em

$$\gamma_{\mathrm{w}} = \frac{1}{M}\sum_{i=1}^{M}\frac{\rho_i^2 - 1}{\rho_i}, \tag{129}$$

sendo esta a SNRW quando a fonte tradicional de NU é combinada com o intervalo tradicional de NU. Note que (129) se torna idêntica à expressão em (53) se $M = 1$ ou $\rho_i = \rho$ para todo i.

O número de amostras requerido para se atingir o desejado para as métricas de desempenho pode ser obtido extraindo τ de (122) e (123) admitindo os piores casos em termos de NU, isto é, admitindo $\hat{\sigma}^2_{\mathrm{v}_i} = (1+\rho_{2_i})\sigma^2_{\mathrm{v}_i}$ em (122) e $\hat{\sigma}^2_{\mathrm{v}_i} = (1-\rho_{1_i})\sigma^2_{\mathrm{v}_i}$ em (123), e igualando os resultados. Fazendo isso, para $\bar{\gamma}_i = \bar{\gamma}$ para todo i, essas operações levam à

$$N = \frac{\left[Q^{-1}(Q_{\mathrm{fa}})\sqrt{\sum_{i=1}^{M}(1+\rho_{2_i})^2} - Q^{-1}(Q_{\mathrm{d}})\sqrt{\sum_{i=1}^{M}(1-\rho_{1_i}+\bar{\gamma})^2}\right]^2}{M^2\left[\bar{\gamma} - \frac{1}{M}\sum_{i=1}^{M}(\rho_{2_i}+\rho_{1_i})\right]^2}, \tag{130}$$

em que $N \to \infty$ à medida que $\bar{\gamma} \to (1/M)\sum_{i=1}^{M}(\rho_{2_i}+\rho_{1_i})$, o que está em conformidade com (127). Logo, é claro que $\bar{\gamma} > (1/M)\sum_{i=1}^{M}(\rho_{2_i}+\rho_{1_i})$ é uma restrição que precisa ser atendida.

Para $\rho_i = \rho$ para todo i, veja que a expressão em (130) é convertida em

$$N = \frac{\left[Q^{-1}(Q_{\mathrm{fa}})(1+\rho_2) - Q^{-1}(Q_{\mathrm{d}})\left(1-\rho_1+\bar{\gamma}\right)\right]^2}{M\left[\bar{\gamma} - (\rho_1+\rho_2)\right]^2}. \tag{131}$$

Comparando (131) com (54), conclui-se que o CSS via SD não reduz a SNRW com $\rho_i = \rho$ e $\bar{\gamma}_i = \bar{\gamma}$ para todo i, mas reduz N em M vezes para se ter os valores alvos para Q_{fa} e Q_{d} [38]. Como esperado, note que (131) se iguala à (54) com $M = 1$.

Nova fonte combinada com intervalo tradicional/novo de NU Ao contrário do que acontece com a fonte tradicional de NU, com a nova fonte de NU o valor estimado da variância de ruído $\hat{\sigma}^2_{\mathrm{v}_i}$ é incorporado no cômputo do limiar de decisão. Por isso, para o novo cômputo da estatística de teste T em (38), $\hat{\sigma}^2_{\mathrm{v}_i}$ irá aparecer nos denominadores de (38). Logo, a média e a variância de T em (38) sob $\mathcal{H}_0$ ou $\mathcal{H}_1$ são computadas, respectivamente, como

$$\hat{\mu}_0 = \frac{1}{M}\sum_{i=1}^{M}\frac{\sigma^2_{\mathrm{v}_i}}{\hat{\sigma}^2_{\mathrm{v}_i}} \text{ e } \hat{\sigma}^2_0 = \frac{1}{M^2N}\sum_{i=1}^{M}\frac{\sigma^4_{\mathrm{v}_i}}{\hat{\sigma}^4_{\mathrm{v}_i}} \tag{132}$$

sob a hipótese $\mathcal{H}_0$, ou

$$\hat{\mu}_1 = \frac{1}{M}\sum_{i=1}^{M}\frac{\sigma^2_{\mathrm{v}_i}}{\hat{\sigma}^2_{\mathrm{v}_i}}\left(1+\bar{\gamma}_i\right) \text{ e } \hat{\sigma}^2_1 = \frac{1}{M^2N}\sum_{i=1}^{M}\frac{\sigma^4_{\mathrm{v}_i}}{\hat{\sigma}^4_{\mathrm{v}_i}}\left(1+\bar{\gamma}_i\right)^2 \tag{133}$$

sob a hipótese $\mathcal{H}_1$.

Então, dado um valor específico de $\hat{\sigma}^2_{\mathrm{v}_i}$, as métricas globais podem ser escritas como

$$Q_{\mathrm{fa}} = Q\left(\frac{\tau-\hat{\mu}_0}{\hat{\sigma}_0}\right) = Q\left(\frac{M\tau - \sum_{i=1}^{M}\frac{\sigma^2_{\mathrm{v}_i}}{\hat{\sigma}^2_{\mathrm{v}_i}}}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}\frac{\sigma^4_{\mathrm{v}_i}}{\hat{\sigma}^4_{\mathrm{v}_i}}}}\right) \tag{134}$$

e

$$Q_{\mathrm{d}} = Q\left(\frac{\tau-\hat{\mu}_1}{\hat{\sigma}_1}\right) = Q\left(\frac{M\tau - \sum_{i=1}^{M}\frac{\sigma^2_{\mathrm{v}_i}}{\hat{\sigma}^2_{\mathrm{v}_i}}\left(1+\bar{\gamma}_i\right)}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M}\frac{\sigma^4_{\mathrm{v}_i}}{\hat{\sigma}^4_{\mathrm{v}_i}}\left(1+\bar{\gamma}_i\right)^2}}\right), \tag{135}$$

com $\tau = \lambda$.

Integrando (134) e (135) de acordo com a PDF da NU em (44), os valores médios para essas métricas globais ficam, respectivamente, como

$$\bar{Q}_{\text{fa}} = \int_{\mathbf{a}}^{\mathbf{b}} Q\left(\frac{M\lambda - \sum_{i=1}^{M} \frac{\sigma_{\text{v}_i}^2}{x_i}}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M} \frac{\sigma_{\text{v}_i}^4}{x_i^2}}}\right) f_{\hat{\sigma}_{\text{v}_M}^2,\ldots,\hat{\sigma}_{\text{v}_1}^2}(\mathbf{x}) d^M\mathbf{x} \tag{136}$$

e

$$\bar{Q}_{\text{d}} = \int_{\mathbf{a}}^{\mathbf{b}} Q\left(\frac{M\lambda - \sum_{i=1}^{M} \frac{\sigma_{\text{v}_i}^2(1+\bar{\gamma}_i)}{x_i}}{\frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{M} \frac{\sigma_{\text{v}_i}^4(1+\bar{\gamma}_i)^2}{x_i^2}}}\right) f_{\hat{\sigma}_{\text{v}_M}^2,\ldots,\hat{\sigma}_{\text{v}_1}^2}(\mathbf{x}) d^M\mathbf{x}. \tag{137}$$

Aplicando (47) em (136), a inequação $M\lambda - \sum_{i=1}^{M} \sigma_{\text{v}_i}^2/x_i > 0$ deve ser respeitada para se ter $\bar{Q}_{\text{fa}_i} = 0$. Logo, note que $\lambda > \frac{1}{M}\sum_{i=1}^{M} \sigma_{\text{v}_i}^2/x_i$, com $x_i = a_i = (1-\rho_{1_i})\sigma_{\text{v}_i}^2$ para garantir $\bar{Q}_{\text{fa}} = 0$ em todo o intervalo de NU. Da mesma forma, aplicando (47) em (137), a inequação $M\lambda - \sum_{i=1}^{M} \sigma_{\text{v}_i}^2 (1+\bar{\gamma}_i)/x_i < 0$ deve ser respeitada para se ter $\bar{Q}_{\text{d}} = 1$. Consequentemente, é necessário que $\lambda < \frac{1}{M}\sum_{i=1}^{M} \sigma_{\text{v}_i}^2 (1+\bar{\gamma}_i)/x_i$, com $x_i = b_i = (1+\rho_{2_i})\sigma_{\text{v}_i}^2$, para que seja possível garantir $\bar{Q}_{\text{d}} = 1$ em todo o intervalo de NU. Portanto, atingir $\bar{Q}_{\text{fa}} = 0$ e $\bar{Q}_{\text{d}} = 1$ implica em satisfazer $\lambda > (1/M)\sum_{i=1}^{M}(1-\rho_{1_i})^{-1}$ e $\lambda < (1/M)\sum_{i=1}^{M}(1+\bar{\gamma}_i)(1+\rho_{2_i})^{-1}$ conjuntamente. Por isso, pode-se estabelecer que $\sum_{i=1}^{M}(1-\rho_{1_i})^{-1} < \lambda < \sum_{i=1}^{M}(1+\bar{\gamma}_i)(1+\rho_{2_i})^{-1}$, ou que

$$\sum_{i=1}^{M} \bar{\gamma}_i(1+\rho_{2_i})^{-1} > \sum_{i=1}^{M}(1-\rho_{1_i})^{-1} - \sum_{i=1}^{M}(1+\rho_{2_i})^{-1}. \tag{138}$$

Explorando a situação em que se tem $\bar{\gamma}_i = \bar{\gamma}$ para todo i, a expressão em (138) pode ser simplificada para $\bar{\gamma} > \left(\sum_{i=1}^{M}(1-\rho_{1_i})^{-1} - \sum_{i=1}^{M}(1+\rho_{2_i})^{-1}\right)\left(\sum_{i=1}^{M}(1+\rho_{2_i})^{-1}\right)^{-1}$. Portanto, a SNRW generalizada referente ao i-ésimo SU pode ser dada como

$$\gamma_{\text{w}} = \frac{\sum_{i=1}^{M}(1-\rho_{1_i})^{-1} - \sum_{i=1}^{M}(1+\rho_{2_i})^{-1}}{\sum_{i=1}^{M}(1+\rho_{2_i})^{-1}}. \tag{139}$$

Admitindo-se $\rho_i = \rho$ para todo i em (139), tal SNRW se converte para

$$\gamma_{\text{w}} = \frac{\rho_2 + \rho_1}{1-\rho_1}, \tag{140}$$

o que está em concordância com (97).

Substituindo $\rho_{1_i} = (\rho_i - 1)/\rho_i$ e $\rho_{2_i} = \rho_i - 1$ em (139), $\rho_i \geq 1$, a particularização para a SNRW quando a nova fonte de NU é combinada com o intervalo tradicional de NU fica como

$$\gamma_{\text{w}} = \frac{\sum_{i=1}^{M} \frac{\rho_i^2 - 1}{\rho_i}}{\sum_{i=1}^{M} \frac{1}{\rho_i}}, \tag{141}$$

o que se iguala à expressão em (59) se $M = 1$ ou $\rho_i = \rho$ para todo i, isto é, se iguala à $\gamma_{\mathrm{w}} = \rho^2 - 1$. De forma alternativa, substituindo $\rho_{1_i} = \rho_{2_i} = \rho_i$ em (139), $0 \leq \rho_i < 1$, quanto a nova fonte de NU é combinada com o novo intervalo de NU, a particularização dessa SNRW generalizada fica como

$$\gamma_{\mathrm{w}} = \frac{2\sum\limits_{i=1}^{M} \frac{\rho_i}{1-\rho_i^2}}{\sum\limits_{i=1}^{M} \frac{1}{1+\rho_i}}. \tag{142}$$

Observe que se $M = 1$ ou $\rho_i = \rho$ para todo i, a expressão em (142) se iguala àquela em (60).

Por fim, o número de amostras requerido para se atingir os valores alvos para as métricas de desempenho pode ser obtido extraindo τ de (134) e (135) admitindo os piores casos em termos de NU, isto é, admitindo $\hat{\sigma}^2_{\mathrm{v}_i} = (1-\rho_{1_i})\sigma^2_{\mathrm{v}_i}$ em (134) e $\hat{\sigma}^2_{\mathrm{v}_i} = (1+\rho_{2_i})\sigma^2_{\mathrm{v}_i}$ em (135), e igualando os resultados. Fazendo isso com $\bar{\gamma}_i = \bar{\gamma}$ para todo i,

$$N = \frac{\left[Q^{-1}(Q_{\mathrm{fa}})\sqrt{\frac{\sum\limits_{i=1}^{M}(1-\rho_{1_i})^{-2}}{\left[\sum\limits_{i=1}^{M}(1+\rho_{2_i})^{-1}\right]^2}} - Q^{-1}(Q_{\mathrm{d}})(1+\bar{\gamma})\sqrt{\frac{\sum\limits_{i=1}^{M}(1+\rho_{2_i})^{-2}}{\left[\sum\limits_{i=1}^{M}(1+\rho_{2_i})^{-1}\right]^2}}\right]^2}{\left[\bar{\gamma} - \frac{\sum\limits_{i=1}^{M}(1-\rho_{1_i})^{-1} - \sum\limits_{i=1}^{M}(1+\rho_{2_i})^{-1}}{\sum\limits_{i=1}^{M}(1+\rho_{2_i})^{-1}}\right]^2}, \tag{143}$$

sendo fácil notar que $N \to \infty$ à medida que $\bar{\gamma} \to \sum_{i=1}^{M}(1-\rho_{1_i})^{-1}/\sum_{i=1}^{M}(1+\rho_{2_i})^{-1} - 1$, e que esse resultado está em conformidade com (139). Logo, quando a nova fonte de NU é considerada no CSS via SD, deve-se garantir que $\bar{\gamma} > \sum_{i=1}^{M}(1-\rho_{1_i})^{-1}/\sum_{i=1}^{M}(1+\rho_{2_i})^{-1} - 1$.

Sob a condição de $\rho_i = \rho$ para todo i, a expressão em (143) pode ser escrita como

$$N = \frac{\left[Q^{-1}(Q_{\mathrm{fa}})\frac{1+\rho_2}{1-\rho_1} - Q^{-1}(Q_{\mathrm{d}})\,(1+\bar{\gamma})\right]^2}{M\left[\bar{\gamma} - (\rho_1+\rho_2)(1-\rho_1)^{-1}\right]^2}, \tag{144}$$

a partir do que uma comparação com (61) leva à mesma conclusão feita acerca de (131): ou seja, o CSS via SD não reduz a SNRW quando $\rho_i = \rho$ e $\bar{\gamma}_i = \bar{\gamma}$ para todo i, mas reduz N em M vezes para se atingir um dado desempenho alvo em termos de Q_{fa} e Q_{d} [38].

A Tabela 6 agrupa as expressões particulares deduzidas para a SNRW da ED em cada SU, γ_{w}, no CSS via fusão SD quando a fonte tradicional ou nova de NU é combinada com o intervalo tradicional ou novo de NU. Note que em termos de resultado mais conservativo, conclusões idênticas àquelas acerca dos resultados das Tabelas 3 e 4 também podem ser feitas aqui. Especificamente, observe, por exemplo, que (142) $\geq$ (128), (141) $\geq$ (129) e (128) $\geq$ (129), para $0 \leq \rho_i < 1$ e $\rho_i \geq 1$, respectivamente. Além do mais, observe ainda que (141) $\geq$ (128) para $\rho_i \geq 1$ e $0 \leq \rho_i < 1$, respectivamente.

Uma comparação entre os resultados das Tabelas 6 e 5 mostra que as deduções para a SNRW levam à expressões idênticas quando feitas para a estatística de teste tradicional apresentada em (32) (ver Seção 5.4.2.2) ou para a nova estatística de teste apresentada em (38), exceto para o caso da combinação entre a nova fonte e o novo intervalo de NU. Para essa combinação, note que (142) $\geq$ (118), e que a igualdade é atingida se $\rho_i = \rho$ para todo i. É por esse motivo que o novo cômputo da estatística de teste em (38) leva a uma SNRW mais conservativa quando quando os SUs estão operando sob níveis de NU distintos na combinação da nova fonte de NU

Tabela 6: SNRW da ED referente ao i-ésimo SU, γ_{w}, no CSS via SD com o novo cômputo da estatística de teste em (38) para as quatro combinações envolvendo as fontes e intervalos tradicional e novo de NU.

	Fonte Tradicional	Nova Fonte
Intervalo Tradicional	$\gamma_{\mathrm{w}} = \frac{1}{M}\sum_{i=1}^{M} \frac{\rho_i^2-1}{\rho_i}, \rho_i \geq 1,$ @(129)	$\gamma_{\mathrm{w}} = \frac{\sum_{i=1}^{M}(\rho_i^2-1)/\rho_i}{\sum_{i=1}^{M} 1/\rho_i}, \rho_i \geq 1,$ @(141)
Novo Intervalo	$\gamma_{\mathrm{w}} = \frac{2}{M}\sum_{i=1}^{M} \rho_i, 0 \leq \rho_i < 1,$ @(128)	$\gamma_{\mathrm{w}} = \frac{2\sum_{i=1}^{M} \frac{\rho_i}{1-\rho_i^2}}{\sum_{i=1}^{M} \frac{1}{1+\rho_i}}, 0 \leq \rho_i < 1,$ @(142)

com o novo intervalo de NU. Ainda, como as expressões para a SNRW deduzidas via (38) são válidas para $\bar{\gamma}_i = \bar{\gamma}$ para todo i enquanto que aquelas deduzidas via (32) são válidas para $\sigma^2_{\mathrm{v}_i} = \sigma^2_{\mathrm{v}}$ e $\bar{\gamma}_i = \bar{\gamma}$ para todo i, pode-se concluir que a nova estatística de teste em (38) permite deduções mais generalizadas para a SNRW da ED no CSS via SD do que a estatística de teste tradicional em (32).

5.5 Resultados Numéricos

Esta seção apresenta os resultados numéricos via simulações de Monte Carlo e/ou teóricos i) para os desempenhos de detecção em termos das probabilidades de detecção e falso alarme alcançados pela ED sob NU e ii) para a SNRW da ED devido ao efeito da NU. Os resultados são para o NCSS e para o CSS com fusão SD e fusão HD via regra k-em-M a partir das quatro combinações envolvendo as abordagens tradicional e nova entre fonte e intervalo de NU [35]. As análises contaram com uma rede primária formada por um único PU e uma rede secundária formada por $M = 5$ SUs no CSS. A configuração para a potência de sinal primário recebida no i-ésimo SU, $\sigma^2_{\mathrm{s}_i}$, $i = 1,2,\ldots,M$, foi arbitrariamente estabelecida para se ter uma rede secundária heterogênea, significando que $\sigma^2_{\mathrm{s}_i} \neq \sigma^2_{\mathrm{s}_j}$ para $i,j = 1,2,\ldots,M$, $i \neq j$, com $\{\sigma^2_{\mathrm{s}_1}; \sigma^2_{\mathrm{s}_2}; \ldots; \sigma^2_{\mathrm{s}_M}\} = \{0{,}9153; 1{,}0310; 0{,}9552; 1{,}0819; 1{,}0166\}$. Por isso também o valor esperado para a potência de ruído térmico, AWGN, no i-ésimo SU foi configurado para ser computado a partir $\sigma^2_{\mathrm{s}_i}$ e do valor desejado para a respectiva SNR média, $\bar{\gamma}_i$, como $\sigma^2_{\mathrm{v}_i} = \sigma^2_{\mathrm{s}_i}/\bar{\gamma}_i$, com $\bar{\gamma}_i \neq \bar{\gamma}_j$ para $i,j = 1,2,\ldots,M$, $i \neq j$. As simulações de Monte Carlo contaram com 50.000 realizações, equivalendo à 50.000 períodos de sensoriamento. A atividade do PU foi simulada como uma variável aleatória de Bernoulli nessas 50.000 realizações, com 50% sob a hipótese $\mathcal{H}_1$ e 50% sob a hipótese $\mathcal{H}_0$ para os cômputos das probabilidades de detecção e falso alarme, respectivamente, contando com a coleta de $N = 2.000$ amostras do sinal recebido em cada período por cada SU. Os resultados são analisados via curvas ROC, AUCs, tabelas, e curvas que relacional N versus SNR, e foram obtidos via *software Wolfram Mathematica*, versão 13.

Na Figura 8 apresenta-se um primeiro conjunto de resultados. Eles são exclusivamente para avaliar o impacto de diferentes níveis de NU sobre as probabilidades de detecção e falso alarme por meio das curvas ROC alcançadas por um SU configurado com SNR média de $\bar{\gamma} = -12$ dB considerando os intervalos tradicional e novo de NU. Portanto, os resultados são para o NCSS, com fatores de NU dados por $\rho = \{1; 1{,}06; 1{,}09; 1{,}12; 1{,}15\}$, para a combinação entre fonte e intervalo tradicionais de NU, ou por $\rho = \{0; 0{,}06; 0{,}09; 0{,}12; 0{,}15\}$ para a combinação entre nova fonte e novo intervalo de NU. Adicionalmente, a figura ainda apresenta uma curva ROC admitindo a ausência de NU, com $\rho = 0$ para o novo intervalo de NU, ou com $\rho = 1$ para o intervalo tradicional de NU, para fins de comparação com os casos sob a influência da NU. Os resultados confirmam a vulnerabilidade da ED à NU, mesmo com valores de ρ próximos daqueles

que representam os casos livres de NU, isto é, $\rho \approx 1$ para o intervalo tradicional de NU, e $\rho \approx 0$ para o novo intervalo de NU. Eles ainda revelam que a combinação da nova fonte com o novo intervalo de NU leva a desempenhos de detecção mais conservativos (inferiores) comparados àqueles alcançados com a combinação da fonte tradicional com o intervalo tradicional de NU, especialmente para valores mais elevados de ρ em cada caso de intervalo de NU. Nota-se que os resultados teóricos e simulados estão em perfeito alinhamento, o que contribui para a validação das expressões deduzidas.

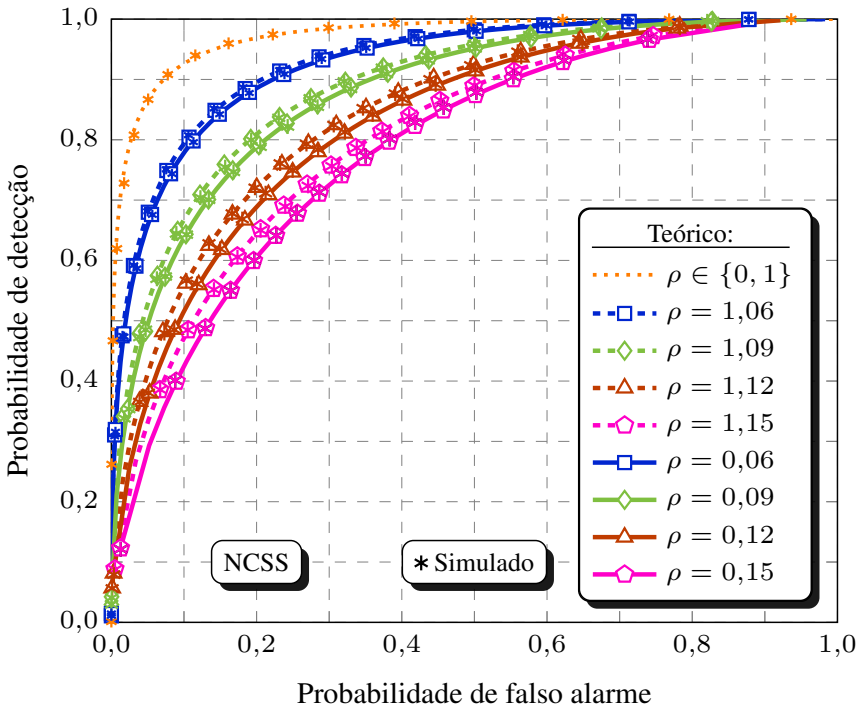


Figura 8: Curvas ROC ED no NCSS; com SNR média de $\bar{\gamma} = -12$ dB, na ausência de NU (linha pontilhada, com $\rho = 0$) e sob NU ($\rho \neq \{0{,}1\}$), com $\rho = \{1{,}06; 1{,}09; 1{,}12; 1{,}15\}$ na combinação entre o intervalo e a fonte tradicional de NU (linhas tracejadas) e $\rho = \{0{,}06; 0{,}09; 0{,}12; 0{,}15\}$ na combinação entre a nova fonte e o novo intervalo de NU (linhas sólidas).

A Tabela 7 complementa os resultados apresentados na Figura 8 com os respectivos valores de AUC e de SNRW, γ_{w}, para os dois conjuntos de fatores de NU, $\rho = \{1; 1{,}06; 1{,}09; 1{,}12; 1{,}15\}$ e $\rho = \{0; 0{,}06; 0{,}09; 0{,}12; 0{,}15\}$, de acordo com a combinação entre fonte e intervalo de NU tradicional e novo, respectivamente. Adicionalmente, a tabela também inclui os resultados em termos de AUC e SNRW referentes às duas combinações restantes, que são a combinação da nova fonte de NU com o intervalo tradicional de NU e a combinação da fonte tradicional de NU com o novo intervalo de NU. O valores referentes às abordagens apresentadas na Figura 8 são aqueles destacados nas células sombreadas. Uma comparação de resultados entre células mostra que a combinação da nova fonte com o novo intervalo de NU é a mais conservativa, pois leva aos maiores valores de SNRW. Por essa mesma comparação entre células mostra-se também que a combinação menos conservativa é aquela entre a fonte tradicional de NU e o intervalo tradicional de NU, pois é a que leva aos menores valores de SNRW.

Abordando um segundo conjunto de resultados, a Figura 9 apresenta as curvas ROC para uma rede secundária contendo $M = 5$ SUs no NCSS e CSS com fusão SD e fusão HD via regra k-em-M com $k = 1$ (regra OR), $k = \lfloor M/2 + 1 \rfloor$ (regra MAJ), e $k = M$ (regra AND). A fusão

Tabela 7: SNRWs, γ_{w}, em dB, e AUCs para a ED no NCSS, com SNR média de $\bar{\gamma} = -12$ dB e para as quatro combinações entre as fontes e intervalos tradicional e novo de NU, com $\rho = \{1{,}06; 1{,}09; 1{,}12; 1{,}15\}$ para o intervalo tradicional e $\rho = \{0{,}06; 0{,}09; 0{,}12; 0{,}15\}$ para o novo intervalo de NU.

	Fonte Tradicional	Nova Fonte
Intervalo Tradicional	$\gamma_{\mathrm{w}} = -5{,}5217; \mathrm{AUC} = 0{,}9143$	$\gamma_{\mathrm{w}} = -4{,}9147; \mathrm{AUC} = 0{,}9126$
	$\gamma_{\mathrm{w}} = -6{,}4370; \mathrm{AUC} = 0{,}8456$	$\gamma_{\mathrm{w}} = -5{,}9448; \mathrm{AUC} = 0{,}8446$
	$\gamma_{\mathrm{w}} = -7{,}6304; \mathrm{AUC} = 0{,}7663$	$\gamma_{\mathrm{w}} = -7{,}2561; \mathrm{AUC} = 0{,}7674$
	$\gamma_{\mathrm{w}} = -9{,}3329; \mathrm{AUC} = 0{,}6902$	$\gamma_{\mathrm{w}} = -9{,}0798; \mathrm{AUC} = 0{,}6931$
Novo Intervalo	$\gamma_{\mathrm{w}} = -5{,}2288; \mathrm{AUC} = 0{,}9105$	$\gamma_{\mathrm{w}} = -4{,}5230; \mathrm{AUC} = 0{,}9092$
	$\gamma_{\mathrm{w}} = -6{,}1979; \mathrm{AUC} = 0{,}8344$	$\gamma_{\mathrm{w}} = -5{,}6427; \mathrm{AUC} = 0{,}8346$
	$\gamma_{\mathrm{w}} = -7{,}4473; \mathrm{AUC} = 0{,}7463$	$\gamma_{\mathrm{w}} = -7{,}0377; \mathrm{AUC} = 0{,}7494$
	$\gamma_{\mathrm{w}} = -9{,}2082; \mathrm{AUC} = 0{,}6624$	$\gamma_{\mathrm{w}} = -8{,}9395; \mathrm{AUC} = 0{,}6676$

SD ainda leva em contra o cômputo tradicional da estatística de teste T em (32) (ROC teórica com linha pontilhada) e o novo cômputo de T em (38) (ROC teórica com linha contínua). Esses resultados foram obtidos com uma SNR média para o i-ésimo SU configurada como $\{\bar{\gamma}_1; \bar{\gamma}_2; \ldots; \bar{\gamma}_M\} = \{-13; -14; -15; -16; -17\}$ dB, na combinação da nova fonte de NU com o novo intervalo de NU, com $\{\rho_1; \rho_2; \ldots; \rho_M\} = \{0{,}02; 0{,}015; 0{,}012; 0{,}01; 0{,}0075\}$. Os resultados evidenciam o melhor desempenho de detecção alcançado pelo CSS em comparação com o NCSS, como se deveria esperar. Observando o conjunto de curvas do gráfico destinado à apresentação dos resultados do CSS, nota-se que dentre os desempenhos alcançados pela fusão HD, a regra MAJ obteve os melhores resultados, seguida pelas regras OR e AND, respectivamente. Nota-se também que a nova estatística de teste em (38) de fato pode oferecer desempenhos de detecção consideravelmente melhores quando os SUs estão operando sob diferentes níveis de potência de ruído, $\sigma^2_{\mathrm{v}_i} \neq \sigma^2_{\mathrm{v}_j}$, $i,j = 1, \ldots, M$, $i \neq j$, comparado aos desempenhos de detecção oferecidos pela estatística de teste tradicional em (32). Veja que os resultados teóricos e simulados continuam em perfeito alinhamento.

Tal como a Tabela 7 complementa os resultados apresentados na Figura 8, a complementação da Figura 9 é feita por meio da Tabela 8 com os respectivos valores de AUC e de SNRW, γ_{w}, para a combinação da nova fonte com o novo intervalo de NU. Adicionalmente, ela também inclui os resultados em termos de AUC e SNRW para às combinações restantes, que são combinação entre a fonte tradicional de NU e o intervalo tradicional de NU, a combinação da nova fonte de NU com o intervalo tradicional de NU e a combinação da fonte tradicional de NU com o novo intervalo de NU. O resultados relacionados com a abordagem apresentadas na Figura 8 são aqueles destacados nas células sombreadas. O conjunto de valores para o fator de incerteza de NU foi configurado como $\{\rho_1; \rho_2; \ldots; \rho_M\} = \{1{,}02; 1{,}015; 1{,}012; 1{,}01; 1{,}0075\}$ no caso do intervalo tradicional de NU. Para o caso do novo intervalo de NU, os valores são aqueles utilizados na Figura 9, ou seja, $\{\rho_1; \rho_2; \ldots; \rho_M\} = \{0{,}02; 0{,}015; 0{,}012; 0{,}01; 0{,}0075\}$. Observando os resultados, pode-se fazer as mesmas conclusões feitas acerca dos resultados da Tabela 7; a saber, a combinação mais conservativa é a da nova fonte de NU com o novo intervalo de NU, pois leva aos maiores valores de SNRW; e a menos conservativa é a combinação da fonte tradicional com o intervalo tradicional de NU, visto que leva aos menores valores de SNRW. Também pode-se concluir, tanto pelos resultados da Tabela 7 quanto pelos resultados da Tabela 8, que no geral a combinação mais conservativa também leva a desempenhos de detecção também mais

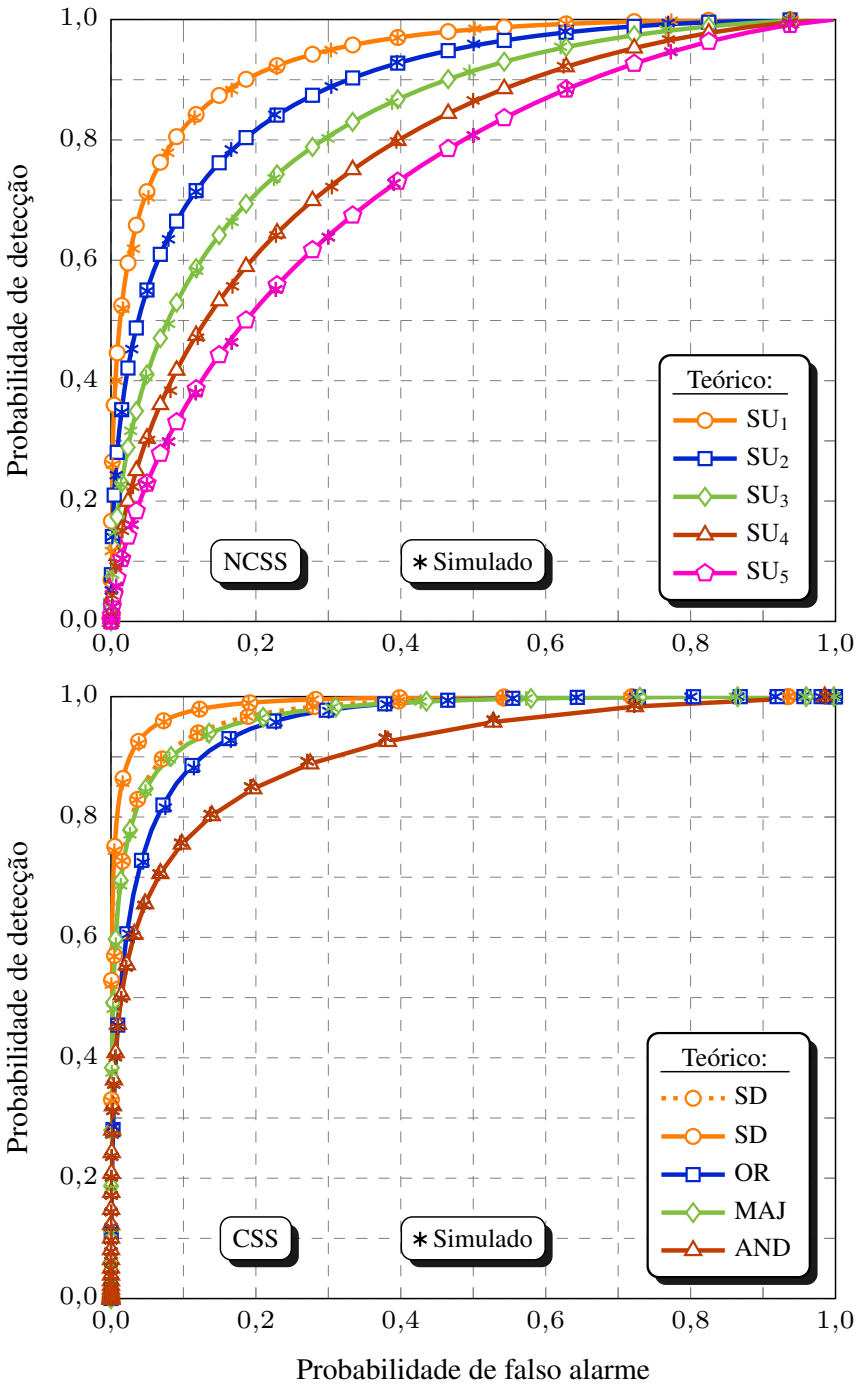


Figura 9: Curvas ROC (teóricas e simuladas) da estatística de teste ED no NCSS e CSS (com $M = 5$ SUs) via SD com os cômputos tradicional (32) (linha pontilhada) e novo (38) (linha sólida) da estatística de teste, e via HD sob as regras OR ($k = 1$), MAJ ($k = \lfloor M/2 + 1 \rfloor = 3$), e AND ($k = 5$), com SNR média $\{\bar{\gamma}_1; \bar{\gamma}_2; \ldots; \bar{\gamma}_M\} = \{-13; -14; -15; -16; -17\}$ dB, combinação da nova fonte com o novo intervalo de NU e fatores de incerteza de ruído $\{\rho_1; \rho_2; \ldots; \rho_M\} = \{0{,}02; 0{,}015; 0{,}012; 0{,}01; 0{,}0075\}$.

conservativos, implicando nos menores valores de AUC. Já a combinação menos conservativa leva à desempenhos de detecção também menos conservativos, implicando nos maiores valores de AUC. Os resultados da Tabela 8 mostram que essa conclusão só não se mantém na fusão

HD via regra OR.

Tabela 8: SNRWs, γ_{w}, em dB, e AUCs para a ED no NCSS, CSS via SD com $M = 5$ SU para os cômputos tradicional (32) e novo (38) da estatística de teste e SNRWs e AUCs no CSS via HD sob as regras OR ($k = 1$), MAJ ($k = \lfloor M/2+1 \rfloor = 3$) e AND ($k = M = 5$), com fatores de NU $\{\rho_1; \ldots; \rho_M\} = \{0{,}02; 0{,}015; 0{,}012; 0{,}01; 0{,}0075\}$ e $\{\rho_1; \ldots; \rho_M\} = \{1{,}02; 1{,}015; 1{,}012; 1{,}01; 1{,}0075\}$ conforme as quatro combinações entre as fontes e intervalos tradicional e novo de NU, e SNR média $\{\bar{\gamma}_1; \ldots; \bar{\gamma}_M\} = \{-13; -14; -15; -16; -17\}$ dB.

	nCSS	
	Fonte Tradicional	Nova Fonte
Intervalo Tradicional	$\gamma_{\mathrm{w}_1} = -14{,}0222$; $\mathrm{AUC}_1 = 0{,}927322$ $\gamma_{\mathrm{w}_2} = -15{,}2610$; $\mathrm{AUC}_2 = 0{,}882498$ $\gamma_{\mathrm{w}_3} = -16{,}2237$; $\mathrm{AUC}_3 = 0{,}830432$ $\gamma_{\mathrm{w}_4} = -17{,}0113$; $\mathrm{AUC}_4 = 0{,}777786$ $\gamma_{\mathrm{w}_5} = -18{,}2553$; $\mathrm{AUC}_5 = 0{,}729369$	$\gamma_{\mathrm{w}_1} = -13{,}9362$; $\mathrm{AUC}_1 = 0{,}926977$ $\gamma_{\mathrm{w}_2} = -15{,}1963$; $\mathrm{AUC}_2 = 0{,}882303$ $\gamma_{\mathrm{w}_3} = -16{,}1719$; $\mathrm{AUC}_3 = 0{,}830325$ $\gamma_{\mathrm{w}_4} = -16{,}9680$; $\mathrm{AUC}_4 = 0{,}777730$ $\gamma_{\mathrm{w}_5} = -18{,}2228$; $\mathrm{AUC}_5 = 0{,}729346$
Novo Intervalo	$\gamma_{\mathrm{w}_1} = -13{,}9794$; $\mathrm{AUC}_1 = 0{,}927150$ $\gamma_{\mathrm{w}_2} = -15{,}2288$; $\mathrm{AUC}_2 = 0{,}882411$ $\gamma_{\mathrm{w}_3} = -16{,}1979$; $\mathrm{AUC}_3 = 0{,}830385$ $\gamma_{\mathrm{w}_4} = -16{,}9897$; $\mathrm{AUC}_4 = 0{,}777760$ $\gamma_{\mathrm{w}_5} = -18{,}2391$; $\mathrm{AUC}_5 = 0{,}729358$	$\gamma_{\mathrm{w}_1} = -13{,}8917$; $\mathrm{AUC}_1 = 0{,}926805$ $\gamma_{\mathrm{w}_2} = -15{,}1631$; $\mathrm{AUC}_2 = 0{,}882216$ $\gamma_{\mathrm{w}_3} = -16{,}1455$; $\mathrm{AUC}_3 = 0{,}830278$ $\gamma_{\mathrm{w}_4} = -16{,}9461$; $\mathrm{AUC}_4 = 0{,}777704$ $\gamma_{\mathrm{w}_5} = -18{,}2064$; $\mathrm{AUC}_5 = 0{,}729336$

	CSS via HD sob as regras OR ($k = 1$), MAJ ($k = \lfloor M/2+1 \rfloor = 3$) e AND ($k = M = 5$)					
	Fonte Tradicional			Nova Fonte		
	$k = 1$	$k = \lfloor M/2+1 \rfloor$	$k = M$	$k = 1$	$k = \lfloor M/2+1 \rfloor$	$k = M$
Intervalo Tradicional	$\gamma_{\mathrm{w}_1} = -14.0222$ $\gamma_{\mathrm{w}_2} = -14.5869$ $\gamma_{\mathrm{w}_3} = -14.9679$ $\gamma_{\mathrm{w}_4} = -15.2431$ $\gamma_{\mathrm{w}_5} = -15.6155$ $\mathrm{AUC}_1 = 0.95203$	$\gamma_{\mathrm{w}_1} = -15{,}0021$ $\gamma_{\mathrm{w}_2} = -15{,}7222$ $\gamma_{\mathrm{w}_3} = -16{,}2237$ $\gamma_{\mathrm{w}_4} = -16{,}5954$ $\gamma_{\mathrm{w}_5} = -17{,}1121$ $\mathrm{AUC}_{\lfloor M/2+1 \rfloor} = 0{,}96734$	$\gamma_{\mathrm{w}_1} = -15{,}6691$ $\gamma_{\mathrm{w}_2} = -16{,}5212$ $\gamma_{\mathrm{w}_3} = -17{,}1315$ $\gamma_{\mathrm{w}_4} = -17{,}5943$ $\gamma_{\mathrm{w}_5} = -18{,}2553$ $\mathrm{AUC}_M = 0{,}88880$	$\gamma_{\mathrm{w}_1} = -13{,}9362$ $\gamma_{\mathrm{w}_2} = -14{,}5009$ $\gamma_{\mathrm{w}_3} = -14{,}8819$ $\gamma_{\mathrm{w}_4} = -15{,}1571$ $\gamma_{\mathrm{w}_5} = -15{,}5295$ $\mathrm{AUC}_1 = 0{,}95204$	$\gamma_{\mathrm{w}_1} = -14{,}9502$ $\gamma_{\mathrm{w}_2} = -15{,}6704$ $\gamma_{\mathrm{w}_3} = -16{,}1719$ $\gamma_{\mathrm{w}_4} = -16{,}5436$ $\gamma_{\mathrm{w}_5} = -17{,}0603$ $\mathrm{AUC}_{\lfloor M/2+1 \rfloor} = 0{,}96722$	$\gamma_{\mathrm{w}_1} = -15{,}6366$ $\gamma_{\mathrm{w}_2} = -16{,}4887$ $\gamma_{\mathrm{w}_3} = -17{,}0990$ $\gamma_{\mathrm{w}_4} = -17{,}5618$ $\gamma_{\mathrm{w}_5} = -18{,}2228$ $\mathrm{AUC}_M = 0{,}88859$
Novo Intervalo	$\gamma_{\mathrm{w}_1} = -13{,}9794$ $\gamma_{\mathrm{w}_2} = -14{,}5593$ $\gamma_{\mathrm{w}_3} = -14{,}9485$ $\gamma_{\mathrm{w}_4} = -15{,}2288$ $\gamma_{\mathrm{w}_5} = -15{,}6067$ $\mathrm{AUC}_1 = 0{,}95197$	$\gamma_{\mathrm{w}_1} = -14{,}9485$ $\gamma_{\mathrm{w}_2} = -15{,}6864$ $\gamma_{\mathrm{w}_3} = -16{,}1979$ $\gamma_{\mathrm{w}_4} = -16{,}5758$ $\gamma_{\mathrm{w}_5} = -17{,}0997$ $\mathrm{AUC}_{\lfloor M/2+1 \rfloor} = 0{,}96730$	$\gamma_{\mathrm{w}_1} = -15{,}6067$ $\gamma_{\mathrm{w}_2} = -16{,}4782$ $\gamma_{\mathrm{w}_3} = -17{,}0997$ $\gamma_{\mathrm{w}_4} = -17{,}5696$ $\gamma_{\mathrm{w}_5} = -18{,}2391$ $\mathrm{AUC}_M = 0{,}88874$	$\gamma_{\mathrm{w}_1} = -13{,}8917$ $\gamma_{\mathrm{w}_2} = -14{,}4716$ $\gamma_{\mathrm{w}_3} = -14{,}8608$ $\gamma_{\mathrm{w}_4} = -15{,}1410$ $\gamma_{\mathrm{w}_5} = -15{,}5189$ $\mathrm{AUC}_1 = 0{,}95198$	$\gamma_{\mathrm{w}_1} = -14{,}8961$ $\gamma_{\mathrm{w}_2} = -15{,}6339$ $\gamma_{\mathrm{w}_3} = -16{,}1455$ $\gamma_{\mathrm{w}_4} = -16{,}5233$ $\gamma_{\mathrm{w}_5} = -17{,}0472$ $\mathrm{AUC}_{\lfloor M/2+1 \rfloor} = 0{,}96718$	$\gamma_{\mathrm{w}_1} = -15{,}5740$ $\gamma_{\mathrm{w}_2} = -16{,}4455$ $\gamma_{\mathrm{w}_3} = -17{,}0670$ $\gamma_{\mathrm{w}_4} = -17{,}5369$ $\gamma_{\mathrm{w}_5} = -18{,}2064$ $\mathrm{AUC}_M = 0{,}88854$

	CSS via SD considerando a estatística de teste T tradicional	
	Fonte Tradicional	Nova Fonte
Intervalo Tradicional	$\gamma_{\mathrm{w}} = -15{,}9146$; $\mathrm{AUC} = 0{,}969912$	$\gamma_{\mathrm{w}} = -15{,}8590$; $\mathrm{AUC} = 0{,}969872$
Novo Intervalo	$\gamma_{\mathrm{w}} = -15{,}8838$; $\mathrm{AUC} = 0{,}969889$	$\gamma_{\mathrm{w}} = -15{,}8272$; $\mathrm{AUC} = 0{,}969851$

	CSS via SD considerando a nova estatística de teste T	
	Fonte Tradicional	Nova Fonte
Intervalo Tradicional	$\gamma_{\mathrm{w}} = -15{,}9180$; $\mathrm{AUC} = 0{,}984992$	$\gamma_{\mathrm{w}} = -15{,}8594$; $\mathrm{AUC} = 0{,}984957$
Novo Intervalo	$\gamma_{\mathrm{w}} = -15{,}8789$; $\mathrm{AUC} = 0{,}984961$	$\gamma_{\mathrm{w}} = -15{,}8252$; $\mathrm{AUC} = 0{,}984932$

5.6 Conclusões

Todo o conteúdo que compõe os tópicos abordados nesta Seção 5 deste documento é sobre os avanços das pesquisas que vêm sendo desenvolvidas no contexto do SS baseado em CR como parte das soluções de interoperabilidade entre os sistemas de comunicação legados e secundários, projetados para as aplicações de TVWS. Esses avanços contribuem para a concepção das redes 6G, e mais especificamente para a concepção da rede de acesso para as redes 6G, bem como servem de base para as redes futuras, xG, pois colaboram para a convivência harmoniosa entre os sistemas independentemente das demandas de conectividade e serviços atuais e previstos de acordo com as conjecturas já elaboradas na literatura. Os resultados numéricos apresentados na seção anterior, a Seção 5.5, mostram a importância de se considerar modelos mais realistas a fim de se analisar o comportamento das redes secundárias do ponto de vista da identificação de canais ociosos via SS, pois pôde-se notar as respostas mais ou menos realistas sobre os

desempenhos de detecção da ED sob NU e sobre a SNRW da ED devido à NU frente às combinações mais ou menos realistas envolvendo os modelos tradicional e novo de fonte e intervalo de NU, respectivamente. Os resultados mostram que a combinação mais realista é aquela referente à proposta para o novo modelo de fonte e intervalo de NU descrito na Seção 5.1.1. Além do mais, com base nesses resultados, esse novo modelo revela que o limiar de SNR requerido para que o SS possa ser realizado a um nível de SNR acima da SNRW, e assim seja possível alcançar uma detecção robusta de sinais na banda de interesse, pode ser mais restrito do que aqueles até então reportados na literatura. Já a combinação menos realista é justamente aquela tradicionalmente adotada na literatura, que leva em conta a combinação da fonte tradicional de NU com o intervalo tradicional de NU. Ela corresponde ao modelamento tradicional da NU, também descrito na Seção 5.1.1.

A capacidade do SS como ferramenta auxiliar para prover convivência harmoniosa entre as distintas aplicações, atividades, dispositivos, tecnologias e gerações – como IoT, 5G, 6G e além – é de grande importância para possibilitar uma conectividade inclusiva, resiliente e escalável. Portanto, as pesquisas que vem sendo executadas sobre SS no contexto do projeto Brasil 6G podem contribuir de forma específica para a concepção das redes 6G no Brasil uma vez que podem oferecer subsídios técnicos e metodológicos para apoiar o desenvolvimento de mecanismos inteligentes e adaptativos de acesso ao espectro. Tais mecanismos são essenciais para atender às demandas nacionais por conectividade eficiente em áreas urbanas e em regiões remotas, promovendo inclusão digital, inovação tecnológica e autossuficiência no planejamento das evoluções das infraestruturas de comunicação. Portanto, as contribuições sobre SS podem fortalecer as bases para a construção de uma arquitetura de rede 6G alinhada às necessidades e particularidades do Brasil.

6 Análise do Slotted ALOHA com Múltiplos *Age Thresholds*

Plínio Santini Dester
UNICAMP

6.1 Introdução

O avanço de aplicações ciberfísicas de baixa latência, como veículos autônomos, redes de sensores, redes industriais e sistemas de realidade aumentada, têm intensificado o interesse por métricas que caracterizam a atualidade da informação. Nesses sistemas, é essencial que os dados recebidos por um monitor reflitam, de forma fiel e oportuna, o estado atual do processo monitorado. Essa necessidade motivou o surgimento da métrica AoI, que quantifica o tempo decorrido desde a geração da última atualização recebida com sucesso [39].

Diferentemente de métricas tradicionais como atraso médio ou *throughput*, o AoI captura diretamente o frescor da informação disponível no destino. Estudos recentes mostram que estratégias de transmissão que maximizam a utilização do canal ou minimizam o atraso nem sempre resultam em um AoI baixo [40]. Isso ocorre pois o congestionamento de pacotes ou a obsolescência de atualizações podem levar a um monitor desinformado, mesmo em sistemas com alta taxa de transmissão.

Nesse contexto, políticas de acesso ao meio baseadas no valor atual do AoI têm se mostrado eficazes para reduzir a idade média das informações. Em particular, o protocolo *Age-Threshold Slotted ALOHA* (TSA) introduz uma probabilidade de acesso condicionada ao AoI local de cada usuário, ou seja, a transmissão ocorre apenas quando o AoI ultrapassa um limiar predefinido. Essa abordagem, simples e *distribuída*, já demonstrou ganhos expressivos em termos de AoI médio. De fato, em [41], foi provado que para um número de nós suficientemente grande, o AoI médio pode reduzir quase pela metade, com uma diminuição marginal no *throughput*, em relação ao *slotted* ALOHA.

Nesta pesquisa, é explorada uma extensão do TSA denominada preliminarmente como *Multiple Age Threshold Slotted ALOHA* (M-TSA), na qual M limiares de AoI são definidos, cada um associado a uma distinta probabilidade de acesso. A ideia central é permitir uma resposta mais adaptativa e granular ao crescimento do AoI, sem comprometer o aspecto distribuído do algoritmo, refinando o comportamento de acesso e explorando melhor as oportunidades de transmissão em ambientes altamente congestionados, como aqueles previstos em cenários futuros das redes 6G.

Trabalhos Relacionados

Alguns trabalhos recentes (2024 e 2025) envolvendo o tema apresentado (TSA, ALOHA, AoI) são brevemente mencionados a seguir. Em [42], considera-se um TSA com tráfego exógeno e um limite máximo de retransmissões. Em [43], os autores comparam uma estratégia baseada em prioridade em função do AoI. Em [44], compara-se uma estratégia reativa e proativa para o tempo de espera em retransmissões no *slotted* ALOHA considerando o AoI médio. Em [45], é proposto um modelo matemático para comparar a energia gasta, *throughput*, e AoI em sistemas baseados em ALOHA. Em [46], os autores analisam o esquema *slotted* ALOHA considerando colisões e probabilidade de erro na transmissão, no contexto de *short-packet communications*. Em [47], foram obtidos resultados exatos da distribuição do AoI considerando *slotted* ALOHA. Em [48], são apresentadas expressões analíticas para a probabilidade de sucesso e AoI médio

considerando TSA e as posições dos nós (geometria estocástica). Nenhum dos trabalhos acima explora uma estratégia como a M-TSA proposta neste documento.

6.2 Modelo do Sistema

Considera-se uma rede sem fio com n nós (ou usuários) que compartilham um canal comum e se comunicam com um AP. Cada nó deseja transmitir atualizações de informações por meio do AP. Adota-se o modelo *generate-at-will* [49], no qual cada nó gera um novo pacote imediatamente antes de cada tentativa de transmissão.

O tempo é dividido em intervalos discretos (*timeslots*) e os nós operam de forma sincronizada. O canal segue um modelo de colisão clássico: se dois ou mais nós transmitem simultaneamente, todas as transmissões são perdidas; caso apenas um nó transmita, a entrega é bem-sucedida. Após uma transmissão bem-sucedida, o nó recebe um sinal de confirmação (*acknowledgment*) de forma imediata e sem erros, através de um canal separado.

O AoI do nó $i \in \{1, \ldots, n\}$ no instante $t \in \mathbb{N}$ é definido como

$$\Delta_i(t) = t - U_i(t),$$

onde $U_i(t)$ representa o tempo de geração do último pacote recebido com sucesso até o instante t associado ao nó i. A evolução do processo de AoI é dada por

$$\Delta_i(t) = \begin{cases} 1, & \text{se o nó } i \text{ transmitiu com sucesso em } t-1; \\ \Delta_i(t-1) + 1, & \text{caso contrário.} \end{cases} \tag{145}$$

O AoI médio do nó i é definido por

$$\overline{\Delta}_i = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \Delta_i(t),$$

caso o limite exista.

Algoritmo de Acesso ao Meio

Neste trabalho, propomos o M-TSA, uma generalização do TSA. Ao invés de um único limiar, define-se uma sequência de M limiares $A_1 < A_2 < \cdots < A_M$. Cada faixa de AoI $[A_k, A_{k+1})$ está associada uma probabilidade de acesso $p_k \in (0,1)$, para $k \in \{1, \ldots, M\}$. Formalmente, para cada nó i e tempo t, seja $B_i(t)$ a variável de acesso ao meio

$$B_i(t) \sim \mathcal{B}(p_k), \quad \text{se } A_k \leq \Delta_i(t) < A_{k+1},$$

com $A_{M+1} = \infty$, sendo $\mathcal{B}(p)$ a distribuição de Bernoulli com parâmetro p. Condicionadas aos seus respectivos AoIs, as variáveis $B_1(t), \ldots, B_n(t)$ são independentes.

6.3 Resultados

Para avaliar os possíveis ganhos do M-TSA, comparamos o caso específico do 2-TSA com o TSA original. A análise foi realizada por meio de simulações de Monte Carlo, uma vez que ainda não foram obtidas expressões analíticas para o desempenho do 2-TSA.

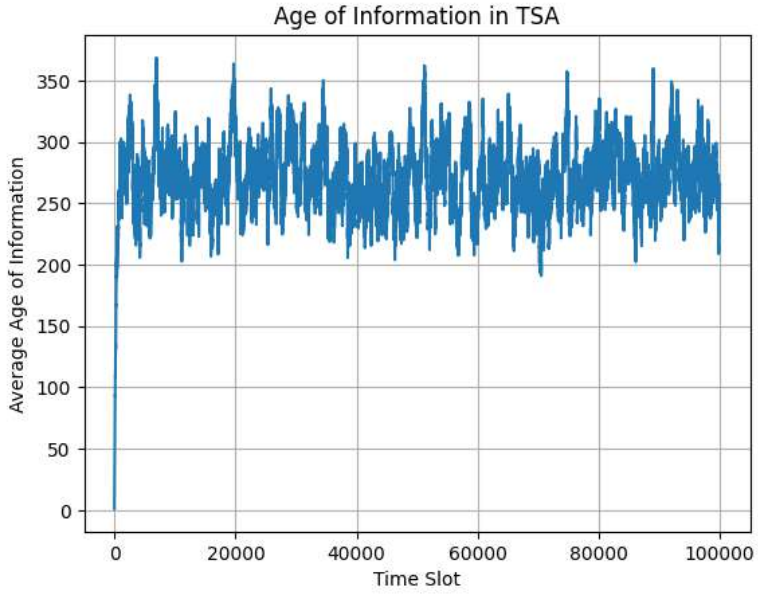


Figura 10: 1-TSA com $n = 100$, $A_1 = 0$, $p_1 = 0.01$, correspondente ao *slotted* Aloha clássico. *Throughput* médio: 0.3704, AoI médio: 270.63, desvio padrão: 26.33.

As simulações buscaram verificar se a introdução de um limiar adicional, que permite um controle mais refinado das decisões de acesso, pode influenciar positivamente no desempenho da rede. Ademais, para simular uma situação de congestionamento, iniciamos a simulação com todos os nós da rede com o mesmo AoI. Dessa forma, todos ficarão ativos simultaneamente, colocando a rede numa situação de "pior caso". Para todas as simulações, foram utilizados $n = 100$ nós, 100 mil *timeslots*, e as médias foram calculadas a partir de 20 mil *timeslots* para minimizar o efeito do transitório (em particular, das condições iniciais de congestionamento).

Na Figura 10, é mostrado o AoI médio dos nós da rede em cada instante de tempo ao simular o *slotted* ALOHA clássico com parâmetros ótimos, ou seja, $M = 1$, $A_1 = 0$, e $p_1 = 1/n$. Como esperado, as métricas de performance foram próximas de $1/e$ para o *throughput* e $e \cdot n$ para o AoI médio. Na Figura 11, é mostrado AoI médio dos nós da rede em cada instante de tempo ao simular o TSA com parâmetros ótimos de acordo com [41], ou seja, $M = 1$, $A_1 = 2.17\,n$, e $p_1 = 4.43/n$. Em comparação com o *slotted* ALOHA, o TSA lida mal com congestionamento (vide início da simulação), pois a probabilidade de acesso é quatro vezes mais alta do que o ideal quando quase todos os nós estão ativos. Entretanto, em momentos em que não há congestionamento, ou seja, apenas uma parte dos nós está com o AoI acima do limiar, o TSA possui AoI médio de aproximadamente metade do *slotted* ALOHA.

Na Figura 12, é mostrado o AoI médio dos nós da rede em cada instante de tempo ao simular o 2-TSA ($M = 2$) com os mesmos parâmetros do TSA, exceto pelo fato de ter um limiar adicional $A_2 \gg A_1$ com probabilidade de acesso $p_2 < p_1$. Sob estas condições conseguimos um sistema que possui desempenho semelhante ao TSA ótimo, porém claramente lida melhor com congestionamento. Esse resultado, é um indicativo que a estratégia de múltiplos limiares (*thresholds*) pode ser promissora nesse sentido. Em termos simples, quando o AoI de um nó ultrapassa A_2, que é muito maior que A_1, é um indicativo que o sistema está em uma situação de congestionamento, então diminuir a probabilidade de acesso parece uma abordagem adequada. Também foram testados extensivamente casos $M = 2$ com $p_2 > p_1$ e $M = 3$ com $p_3 > p_2 > p_1$, porém não foi possível encontrar pontos de operação interessantes ou

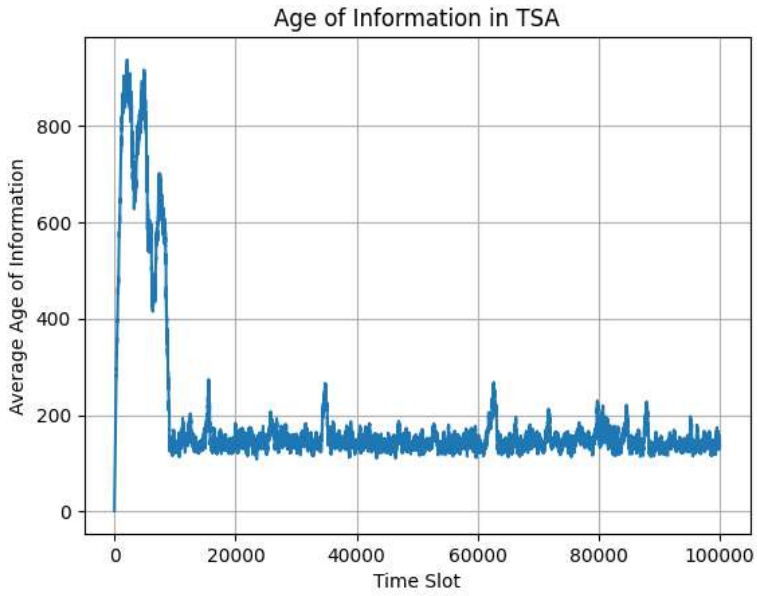


Figura 11: 1-TSA com $n = 100$, $A_1 = 217$, $p_1 = 0.0443$. *Throughput* médio: 0.3379, AoI médio: 147.85, desvio padrão: 9.42.

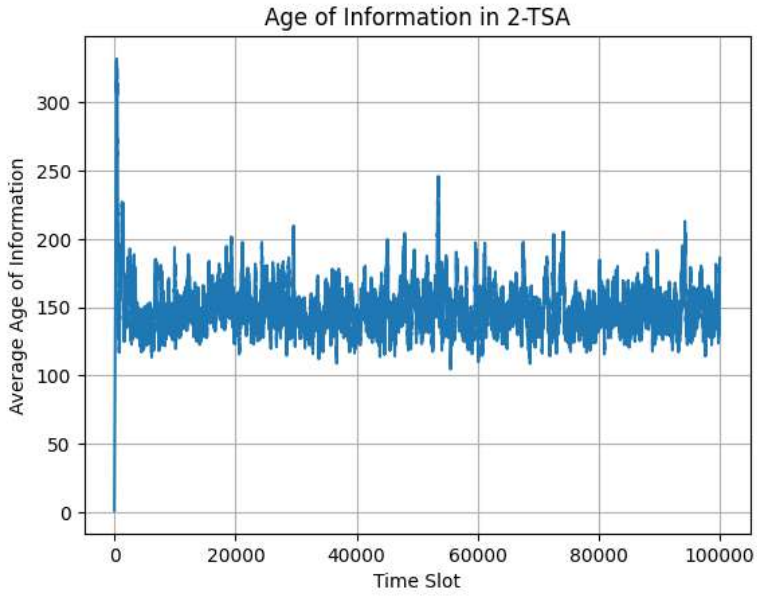


Figura 12: 2-TSA com $n = 100$, $A_1 = 217$, $p_1 = 0.0443$, $A_2 = 300$, $p_2 = 0.03$. *Throughput* médio: 0.3594, AoI médio: 148.48, desvio padrão: 9.84.

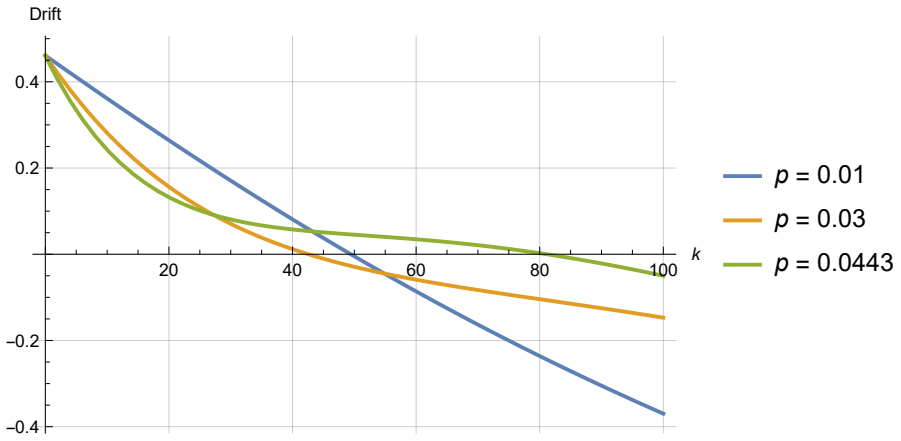


Figura 13: Drift do TSA com $n = 100$, $A = 217$, e vários valores de p.

que aprimorassem significativamente o AoI médio. De fato, as simulações e testes feitos até o momento, apontam que a escolha ótima satisfaz $p_2 < p_1$, o que de certa forma é contra-intuitivo em uma primeira análise, pois são os nós com maior AoI que deveriam ter mais chances de transmitir.

Ademais, uma análise analítica simples também aponta para o mesmo caminho. O *drift*[2] do TSA ($M = 1, A_1 = A, p_1 = p$) é dado por

$$D_k = \frac{n-k}{A} - k\,p(1-p)^{k-1}\left(1 - \frac{n-k}{A}\right),$$

sendo $k \in \{0,1,\ldots,n\}$ o número de usuários ativos, ou seja, que alcançaram ou ultrapassaram o limiar A de AoI. A Figura 13 mostra o *drift* para o TSA em função do número de usuários ativos k com $n = 100$ usuários, valor ótimo de *threshold* $A = 217$, e diversas probabilidades de acesso p. Tomando $p = 0.0443$, que minimiza o AoI médio [41], verifica-se o menor *drift* para um número pequeno de usuários ativos, mas o maior *drift* em condições de congestionamento; de fato, o ponto de equilíbrio é em torno de 80 usuários ativos. Tomando $p = 0.01$, que otimiza o *slotted* ALOHA clássico, verifica-se o menor *drift* em condições extremas, em que todos os 100 usuários estão ativos, porém a probabilidade de se aproximar desse estado é baixa; o ponto de equilíbrio é em torno de 50, um ponto mais interessante que o anterior. Por fim, tomando $p = 0.03$ tem-se o ponto de equilíbrio em torno de 40 usuários, que é um ponto mais interessante que o anterior, e ainda, para valores pequenos do número de usuários ativos, apresenta um *drift* menor que a situação $p = 0.01$, mas ligeiramente maior que $p = 0.0443$. Dessa forma, utilizar $p = 0.03$ quando o sistema está levemente congestionado (e.g. $k > 50$) faz o sistema retornar (em média) muito mais rápido para uma situação mais interessante (e.g $k < 25$) do que $p = 0.0443$. Portanto, combinar $p = 0.0443$ em situações pouco congestionadas com $p = 0.03$ para as demais situações através do 2-TSA se mostra uma estratégia interessante.

6.4 Conclusão

A extensão do TSA para múltiplos limiares se mostrou promissora, principalmente no sentido de lidar mais eficientemente com o estado da rede (e.g. situações de congestionamento),

[2]O *drift* corresponde ao valor esperado do aumento do número de usuários ativos [50].

oferecendo uma política de acesso mais flexível e potencialmente mais eficiente para redes distribuídas que priorizam o frescor das informações. Embora os resultados apresentados nesta seção sejam baseados em simulações e modelos analíticos simples, eles motivam futuras investigações teóricas sobre o comportamento do M-TSA e sua aplicação em redes da próxima geração.

Nesse primeiro momento, ficou claro que o aspecto estocástico, discreto, e sutil (sensível à variação dos parâmetros) do problema torna a otimização através de simulações inviável. Dessa forma, os próximos passos consistem em derivar resultados analíticos para guiar a otimização dos *thresholds* e suas respectivas probabilidades de acesso, além de contribuir com uma compreensão mais profunda e fundamental desse tipo de sistema.

7 Métricas para Comunicação Orientada a Tarefa

Isaac Andrés Balbuca Guachizaca e Paulo Cardieri
UNICAMP

7.1 Introdução

A evolução dos sistemas de comunicação sem fio, desde a introdução da primeira geração na década de 1970 até as pesquisas iniciais sobre 6G, revela uma busca constante por taxas de transmissão mais altas, baixa latência e alta confiabilidade, de forma a atender às crescentes demandas de aplicações que dependem de comunicação sem fio. O atendimento a essas demandas requer maior uso de recursos de rádio, motivando pesquisas para repensar as estratégias de comunicação sem fio.

Inspirada pela Teoria de Comunicação de Shannon, que aborda problemas no **nível técnico** (a transmissão precisa de símbolos), a comunidade acadêmica tem explorado recentemente os níveis semântico e de efetividade da comunicação, propostos por Shannon e Weaver em [51], na busca de novas estratégias de comunicação. No **nível semântico**, o foco está na transmissão do significado da mensagem transmitida, enquanto no **nível de efetividade**, o objetivo é avaliar o impacto da mensagem no cumprimento de uma tarefa. O interesse recente nas abordagens baseadas nos níveis semântico e de efetividade, denominadas de uma forma geral por *comunicação semântica*, decorre do seu potencial para levar a uma maior eficiência no uso dos recursos de rádio.

Dentre os diversos desafios relacionados à essas novas abordagens de comunicação, destaca-se a caracterização de métricas adequadas para o projeto e avaliação de sistemas de comunicação empregando tais abordagens. As métricas da comunicação convencional, como taxa de transmissão ou probabilidade de erro de bit, não são mais adequadas quando o objetivo é, por exemplo, projetar um sistema que maximize a probabilidade de cumprimento de uma tarefa no terminal de destino, isto é, com foco no nível de efetividade. Este capítulo discute as principais métricas apropriadas ao nível de efetividade, destacando seus fundamentos e aplicações. Ao final, é apresentada a aplicação que será adotada na continuidade desta linha de pesquisa.

7.2 Comunicação orientada a tarefas

O nível da efetividade, descrito como o terceiro nível da proposta de Shannon e Weaver, foca no impacto da informação recebida na realização da tarefa associada à comunicação [52]. Essa abordagem é frequentemente chamada de **orientada a tarefas** (*task-oriented communications*).

No desenvolvimento de sistemas de comunicação baseados nessa abordagem, o objetivo principal é otimizar as estratégias de comunicação para assegurar a conclusão bem-sucedida de um tarefa predefinida. Enquanto os modelos baseados na comunicação convencional buscam transmitir todas as informações brutas, os sistemas orientados a tarefas visam transmitir apenas os dados mais relevantes para a realização da tarefa. Essa abordagem, entretanto, requer algum grau de cooperação entre os terminais fonte e destino, além do conhecimento, por parte do terminal fonte, dos objetivos específicos da comunicação.

Exemplo de comunicação orientada a tarefa

Para ilustrar a abordagem de comunicação orientada a tarefas, vamos considerar o caso do rastreamento remoto de uma fonte Markoviana discreta para a execução de uma tarefa no terminal remoto (ver [53] para mais detalhes sobre esse exemplo). Trata-se, portanto, de um problema de reconstrução, no terminal destino, de um processo gerado por uma fonte (a Seção 7.5 traz uma descrição mais detalhada desse problema).

Em uma abordagem de **comunicação convencional**, que é agnóstica em relação ao objetivo da comunicação, as amostras do processo da fonte seriam coletadas em uma taxa de amostragem suficientemente alta para minimizar a perda de qualquer transição no processo de Markov. Nesse caso, o foco está na transmissão contínua de todas as amostras, independentemente de sua relevância específica para a tarefa.

Por outro lado, na abordagem **orientada a tarefa**, as amostras do processo na fonte seriam coletadas, por exemplo, apenas quando ocorressem transições significativas no estado do processo. Avançando nessa abordagem, as transmissões poderiam ocorrer somente quando fosse detectada uma discrepância entre o estado atual do processo na fonte e o estado estimado no terminal destino. Essa última abordagem resultaria em transmissões ainda menos frequentes, mas exigiria um maior nível de interação entre o terminal destino e a fonte, para a identificação dos instantes necessários de transmissão.

Essa diferença entre as abordagens ressalta a eficiência potencial da comunicação orientada a tarefa, que transmite apenas informações relevantes para a realização da tarefa no terminal remoto, otimizando o uso de recursos críticos como energia e largura de banda.

A avaliação da importância da informação contida nas mensagens ou amostras em sistemas de comunicação orientada a tarefa requer a definição de métricas coerentes com o contexto da tarefa. Diversas métricas têm sido propostas na literatura, que podem ser classificadas da seguinte forma: métricas que associam a importância de uma amostra à sua **idade** e métricas que buscam caracterizar de forma mais explícita a importância, ou **valor**, da informação para a tarefa associada à transmissão. Essas duas classes serão discutidas em maior detalhe nas seções 7.3 e 7.4.

7.3 A Métrica Idade da Informação e suas variantes

Idade da Informação

Em muitas aplicações, a importância de uma mensagem ou amostra de um processo físico está diretamente relacionada ao tempo decorrido desde o "nascimento" da mensagem ou amostra – ou seja, quanto mais nova for a mensagem, mais importante ela é. Com base nessa ideia, Kaul et al. propuseram a métrica AoI [54], que quantifica o frescor das mensagens recebidas no terminal destino. Essa métrica, denotada aqui por $\Delta(t)$, é definida como o intervalo de tempo decorrido desde a geração do último pacote recebido[3], isto é, $\Delta(t) = t - t_0$, em que t é o tempo atual e t_0 é o instante de tempo em que a última mensagem recebida no receptor foi gerada.

Como exemplo de aplicação dessa métrica, considere o monitoramento do estado de um processo industrial por meio de amostragem de alguma grandeza física e envio dessas amostras a uma central controladora. Suponha que a amostra 1 tenha sido tomada no instante t_1 e recebida na central controladora somente no instante t_1' (ver Fig. 14). Nesse instante, o AoI vale, portanto, $t_1' - t_1$. Quando, no instante t_2', a amostra seguinte, 2, chega à controladora, o AoI passa de $t_2' - t_1$ para $t_2' - t_2$. Assim, o AoI quantifica o frescor da última amostra recebida.

[3]Uma definição mais geral e formal de AoI pode ser encontrada em [55].

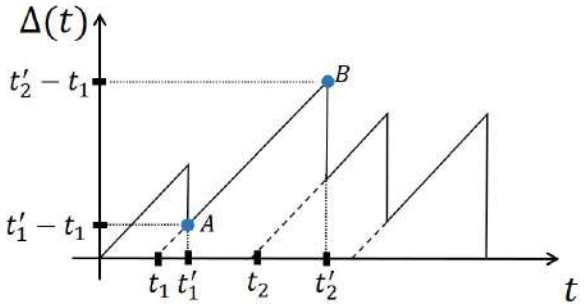


Figura 14: Exemplo de realização de $\Delta(t)$. As marcas t_i e t'_i representam os instantes de geração e de recepção de i-ésima amostra; os pontos A e B indicam os valores de AoI da amostra 1 quando recebida e quando substituída pela amostra 2, respectivamente.

Em aplicações em que a importância da mensagem está relacionada ao seu frescor, o AoI é uma métrica adequada para ser usada na implementação de estratégias de comunicação orientada a tarefa. Idealmente, busca-se manter baixo o AoI no receptor.

Note-se, no entanto, que o AoI é insensível ao conteúdo das mensagens e ao contexto específico do receptor, limitando sua aplicação em sistemas orientados a tarefa em que o conteúdo da mensagem é relevante.

Idade da Informação Incorreta

As limitações da métrica idade da informação, que não leva em conta a importância da amostra para a tarefa associada à comunicação, têm motivado a proposta de diversas métricas alternativas. Entre essas alternativas, uma métrica que vem ganhando destaque é a Idade da Informação Incorreta (*Age of Incorrect Information*, AoII). A AoII combina aspectos da AoI com uma medida que reflete a gravidade dos desvios entre o estado real de um sistema e o estado percebido [56]. No contexto de rastreamento remoto de um processo $X(t)$, AoII é definida como

$$\Delta_{\text{AoII}}(t) = f(t) \times g(X(t),\hat{X}(t)), \tag{146}$$

em que $\hat{X}(t)$ é a estimativa de $X(t)$ no terminal destino, $f(t)$ é uma função crescente que representa a penalidade a ser paga pelo destino por não estar ciente do estado correto na fonte, e $g(t)$ representa a diferença entre o processo na fonte e no destino.

Existem várias opções para as funções f e g. Por exemplo, o AoII pode ser calculado como

$$\Delta_{\text{AoII}}(t) = (t - V(t)) \times \mathbb{1}\{\hat{X}(t) \neq X(t)\},$$

em que $V(t)$ é o último instante de tempo em que o destino tinha informação correta sobre o estado da fonte e $\mathbb{1}\{.\}$ é a função indicadora. Essa métrica se destaca por permitir introduzir funções de penalidade não lineares, representando de forma mais precisa os impactos de erros e desvios.

Outras métricas relevantes

As limitações da métrica AoI e os desafios no uso da métrica VoI motivaram a proposição de diversas outras métricas no contexto de comunicação orientada a tarefa. A seguir, são apresentadas algumas delas.

Idade da Informação Incorreta (*Age of Incorrect Information* - AoII): A AoII combina aspectos do AoI e uma medida da gravidade dos desvios entre o status real e o status percebido [56]. Essa métrica introduz funções de penalidade não lineares para medir erros e desvios no contexto de sistemas orientados a objetivos.

Urgência da Informação (*Urgency of Information* - UoI): A UoI incorpora fatores ambientais para determinar a relevância da informação [57]. Essa métrica atribui pesos com base na urgência do ambiente, permitindo que sistemas inteligentes priorizem pacotes que impactam diretamente na segurança ou eficiência operacional.

Custo do Erro de Atuação: Ao contrário das métricas anteriores, que medem a efetividade indiretamente, o Custo do Erro de Atuação quantifica diretamente as consequências de erros em decisões baseadas em dados incorretos [58]. Essa métrica é especialmente útil em aplicações onde o impacto das decisões é imediato e significativo.

7.4 Valor da Informação

O conceito de **valor da informação** (*Value of Information* - VoI) é bem conhecido no contexto de estratégias de tomada de decisão, sendo definido como o preço que o tomador de decisão está disposto a pagar para aperfeiçoar a sua decisão usando aquela informação [59]. Nesses casos, o valor da informação é dito **esperado**. No contexto de comunicação orientada a tarefa, o VoI de uma informação é geralmente definido como uma medida da redução de incerteza que o uso de uma informação (p.ex., uma amostra já coletada) proporciona na realização da tarefa. Em outras palavras, o VoI representa a contribuição dessa informação para o sucesso da tarefa associada ao processo de comunicação [60][61]. Nesse caso, o valor da informação é dito **observado**. Na próxima seção, essas duas facetas do conceito de VoI serão exploradas.

Valor da Informação observado e Valor da Informação esperado

O *valor da informação observado* é avaliado *a posteriori*, depois que uma amostra foi coletada. Nessa situação, o conteúdo da amostra está disponível, de modo que o cálculo do valor pode explorar as características intrínsecas da amostra. Algumas abordagens atribuem pesos a atributos da amostra como relevância, precisão e frescor (p.ex., AoI). Outras abordagens quantificam o ganho de conhecimento decorrente da nova observação em relação a um estado de crença anterior. Na prática, o valor da informação observado orienta decisões *reativas*: transmitir com prioridade, descartar, armazenar, retransmitir, etc.

O valor da informação esperado, por outro lado, é avaliado *a priori*, isto é, antes que a medição seja realizada. O agente de decisão trabalha apenas com modelos e distribuições de probabilidade conhecidas ou inferidas que descrevem o que *poderá* ser observado, buscando mensurar o quanto se espera que uma observação reduza a incerteza ou aumente o valor de alguma métrica de utilidade. Por exemplo, em monitoramento de estruturas físicas, como viadutos e pontes, o valor da informação esperado pode ser usado para determinar a posição dos sensores a serem instalado para a detecção eficiente de falhas [62].

Valor da Informação no contexto de comunicação orientada a tarefa

No âmbito das comunicações orientadas a tarefa, o conceito de valor da informação mais pertinente é o do valor observado, que quantifica o benefício concreto que uma amostra pro-

porciona para a execução da tarefa em questão.

Embora a definição teórica do VoI seja clara, seu cálculo em cenários práticos pode ser desafiador. Considere, por exemplo, um cenário em que a tarefa a ser executada no terminal de destino dependa do estado de uma fonte, cujas amostras são enviadas a esse terminal. Para estimar o valor da informação, é fundamental compreender tanto o processo aleatório gerado pela fonte quanto a tarefa em si, além de entender a relação entre ambos. A partir desse conhecimento, deve-se avaliar em que medida cada amostra do processo da fonte efetivamente contribui para a execução da tarefa. Desse modo, em um sistema de comunicação orientada a tarefa, a amostragem e a transmissão devem priorizar apenas as amostras que agregam valor à realização da tarefa, em vez de buscar a reconstrução completa do sinal da fonte no terminal de destino.

Algumas estratégias de medida do Valor da Informação

Analytic Hierarchy Process

Desenvolvido por Saaty [63], o *Analytic Hierarchy Process* (AHP) é uma técnica de tomada de decisão que permite organizar e priorizar alternativas com base em múltiplos critérios, estruturando-os hierarquicamente. Para isso, o método decompõe problemas complexos em subproblemas menores, nos quais são feitas comparações entre **atributos** e alternativas. Cada atributo recebe um peso relativo, determinado a partir de sua relevância para o objetivo geral. Tais pesos são calculados por meio de julgamentos qualitativos e quantitativos, que são então combinados em uma matriz de comparação.

O método AHP já foi aplicado por Giordani et al. [64] no contexto de comunicação veicular, em que diversos sensores coletam informações do ambiente. Nessas condições, a técnica AHP foi usada para quantificar o valor da informação presente no conjunto de amostras dos sensores e, com base nesse valor, decidir se as informações deveriam ou não ser transmitidas aos veículos vizinhos.

Divergência de Kullback-Leibler

Uma outra forma que quantificar o valor de uma amostra é através da divergência de Kullback-Leibler (KL), definida como uma medida da diferença entre duas funções densidade de probabilidade. Para uma variável aleatória x, se uma amostra x_0 altera a função densidade *a priori* $f_{prior}(x)$ para a função densidade *a posteriori* $f_{post}(x)$, o valor da informação de x_0 baseado na divergência KL pode ser calculado como

$$VoI_{KL}(x_0) = \int_{-\infty}^{\infty} f_{post}(x) \log \frac{f_{post}(x)}{f_{prior}(x)} dx.$$

Padhy et al. [65] propuseram mecanismos de gerenciamento de coleta de amostras de dados ambientais, e usaram a divergência de Kullback-Leibler para quantificar o valor de cada amostra e aperfeiçoar um modelo de predição dessas amostras. Neste estudo, $f_{prior}(x)$ é a distribuição a priori para estimar a próxima observação, enquanto que $f_{post}(x)$ é a distribuição a posteriori para a próxima observação. Somente amostras com divergência KL acima de um dado limite são coletadas e usadas para ajustar o modelo de previsão.

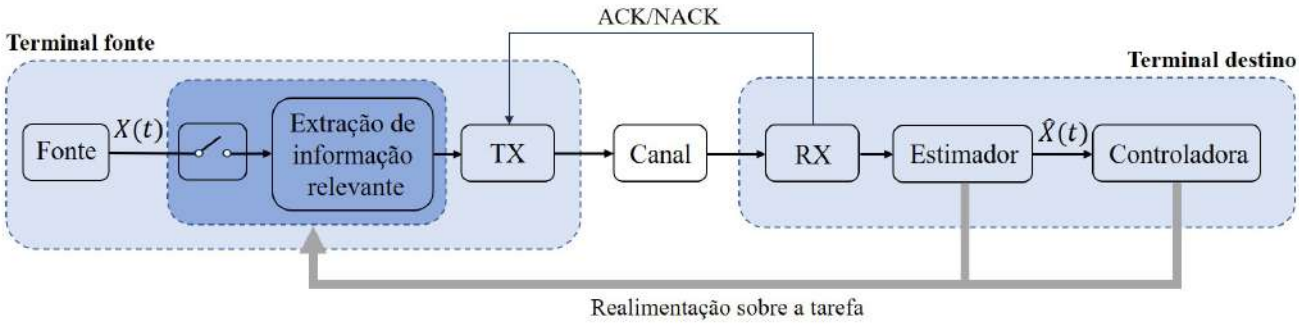


Figura 15: Sistema de rastreamento remoto de processos empregando a estratégia de comunicação orientada a tarefa.

7.5 O problema do rastreamento remoto

Como discutido nas seções anteriores, um dos principais desafios na aplicação dos conceitos de comunicação orientada a tarefa é a forte dependência entre a caracterização das métricas e as tarefas consideradas, em especial no caso do valor da informação. Com o objetivo de ter-se um cenário concreto de aplicação e reduzir essas dificuldades, a continuidade deste projeto concentrar-se-á no cenário de rastreamento remoto de processos. Esse cenário é representativo de diversas aplicações em que a comunicação orientada a tarefa se mostra adequada, tais como IoT industrial, sistemas de controle em rede, veículos aéreos não tripulados, veículos autônomos e redes de comunicação auto-organizáveis [52].

Um sistema de rastreamento remoto e estimação de processos baseado na abordagem de comunicação orientada a tarefa pode ser caracterizado como mostra a Fig. 15. Amostras do processo $X(t)$ são tomadas e avaliadas quanto à sua relevância para a realização das tarefa, e somente as consideradas contendo informação significativa são transmitidas. No terminal receptor, as amostras recebidas são utilizadas para estimar $X(t)$, sendo essa estimativa, por sua vez, empregada pela entidade controladora na execução das tarefas.

Em muitas situações, o receptor pode conter um mecanismo de envio de mensagens do tipo ACK (*Acknowledgement*), a fim de confirmar o recebimento correto dos dados, e NACK (*Negative Acknowledgement*), para indicar erros na recepção. Essas mensagens permitem ao transmissor obter informações sobre a estimativa $\hat{X}(t)$ no receptor. Por exemplo, ao receber uma mensagem ACK, o terminal fonte sabe que o terminal destino está sincronizado com a fonte, ou seja, $\hat{X}(t) = X(t)$. Esse conhecimento pode ser usado pelo transmissor para controlar o envio de novas amostras do processo $X(t)$.

A Fig. 15 mostra também um canal de realimentação, denominado "Realimentação sobre a tarefa", que conecta o Estimador/Controladora às etapas de amostragem e extração de informações relevantes no transmissor. Esse canal de realimentação pode ser empregado em técnicas de cálculo do valor da informação de uma amostra considerando o impacto da amostra na execução de uma tarefa, e não apenas na reconstrução de $X(t)$.

A estrutura apresentada na Fig. 15 tem sido utilizada como modelo de sistema em diversos trabalhos sobre técnicas de rastreamento remoto e estimação de processos baseadas em comunicação orientada a tarefa, e será empregada na continuação das atividades desse projeto.

7.6 Conclusões

Este capítulo apresentou uma síntese dos principais conceitos relacionados à comunicação orientada a tarefa. Foram discutidas métricas relevantes propostas na literatura para esse contexto, classificadas em idade da informação e valor da informação. Embora a métrica idade da informação permaneça útil como indicador de frescor, seu caráter agnóstico ao conteúdo das amostras limita sua aplicação em cenários em que o significado da mensagem é determinante. Essa limitação tem impulsionado o desenvolvimento de métricas mais expressivas, como AoII e UoI. Abordou-se também a noção de valor da informação, destacando as dificuldades de se obter uma formulação geral para seu cálculo, dada a forte dependência da métrica à tarefa considerada. Ainda assim, a literatura apresenta diversas caracterizações práticas do VoI, como as baseadas em AHP e na divergência de Kullback–Leibler. Por fim, foi apresentada a arquitetura de rastreamento remoto que será examinada em maior profundidade ao longo deste projeto.

8 Sistemas de Comunicação com Abordagens Semânticas

Luiz Fernando Gontijo
UNICAMP

8.1 Introdução

Sistemas de comunicação baseados em semântica representam uma nova abordagem no campo das telecomunicações, cujo foco vai além da transmissão fiel de bits. Em vez de priorizar a precisão bit a bit, esses sistemas visam garantir que o significado ou a informação essencial da mensagem seja preservada ao longo do canal de comunicação. Essa abordagem é inspirada pela forma como os humanos se comunicam, em que o entendimento do contexto e da intenção é mais importante do que a transmissão exata de cada palavra. Assim, a comunicação semântica busca otimizar a eficiência da transmissão, reduzindo redundâncias e transmitindo apenas as informações mais relevantes para o entendimento da mensagem no receptor.

Para implementar a comunicação semântica, técnicas avançadas de processamento, como aprendizado profundo e redes neurais, são usadas para codificar e decodificar o conteúdo transmitido. Modelos como autocodificadores e Transformers podem ser aplicados para capturar as características semânticas das mensagens, possibilitando uma compressão mais eficiente e, ao mesmo tempo, preservando o conteúdo essencial em situações de ruído ou restrições de largura de banda. Dessa forma, sistemas de comunicação semântica demonstram grande potencial para aplicações em ambientes desafiadores, como redes de IoT, comunicação por satélite e veículos autônomos, onde a preservação do significado é fundamental para a tomada de decisões e o desempenho geral do sistema.

Este estudo tem por objetivo investigar e conceber soluções avançadas para o desenvolvimento de sistemas de comunicação baseados em semântica. Inicialmente, o estudo aqui proposto tem o objetivo de confirmar o possível uso desses sistemas para o desenvolvimento das próximas gerações de sistemas de telecomunicação, além de induzir a discussão sobre a aplicabilidade das inovações propostas a partir da noção de comunicação baseada em semântica.

As primeiras noções de sistemas de comunicação baseados em semântica foram introduzidas por Shannon ainda em 1948 [66]. Em seu artigo contendo as bases da teoria da informação, foram propostos três problemas básicos para categorizar as comunicações: problema técnico, problema semântico e problema da efetividade. O primeiro problema visa promover uma sistema em que o receptor obtenha de forma acurada as informações enviadas pelo transmissor. Já no segundo, o transmissor utiliza o conhecimento local e compartilhado com o receptor para codificar a mensagem a ser transmitida (problema semântico). No terceiro problema, a preocupação é se a mensagem recebida afeta a conduta esperada e da maneira desejada.

Diversas abordagens foram propostas para atingir a comunicação semântica, como apresentado em [67]. No entanto, alternativas que consideram o uso de Transformers, redes adversárias e redes convolucionais foram estudadas com maior profundidade nesse projeto. A escolha por essa abordagem justifica-se pelos avanços alcançados com o uso de arquiteturas de aprendizado profundo. Em 2021, por exemplo, o uso de Transformers, baseados no uso de camadas de atenção [68], colocou o modelo DeepSC [1] como uma dos grandes avanços nesse campo.

Em [1], camadas de codificação e decodificação semânticas são desenvolvidas a partir de Transformers e treinadas de forma síncrona para reduzir o erro a partir da redução da função de entropia cruzada. Outras propostas foram desenvolvidas com base no DeepSC, como em [69]

- modelo para a transmissão de sinais de voz - e [70], em que foi proposto o uso de quantização vetorial juntamente com camadas de Transformers para a transmissão de imagens.

Sistemas de comunicação baseados em semântica tiveram grandes avanços quando considerada a transmissão de mensagens de texto. Os primeiros trabalhos com esse objetivo foram ainda com o uso de LSTM - arquitetura definida com maiores detalhes em [71] - como apresentado em [72]. Além disso, com os avanços apresentados em [1] a partir do uso de Transformers, outros modelos foram propostos, como em [73], que introduz um modelo compacto para o uso de IoT.

Com base nos últimos artigos publicados em 2024, deve-se observar o desenvolvimento do uso de modelos pré-treinados na tarefa de comunicação semântica, como o modelo GenSC [74] baseado no modelo BART [75]. Além disso, técnicas de empregam maior robustez para a transmissão dos códigos semânticos, como em [76], que usa a quantização vetorial [77], e [78], com o uso de uma camada extra para a correção de erros, são exemplos desse foco de estudo.

O projeto em desenvolvimento e aqui apresentado coloca como foco, assim, o desenvolvimento de sistemas de comunicação baseados em semântica utilizando arquiteturas de Transformers e, sobretudo, modelos de linguagem pré-treinados. Além disso, algumas outras propostas de modulação do sinal codificado para explorar a dimensão dos códigos semânticos também foram exploradas e têm seus resultados iniciais apresentados no próximo tópico.

8.2 Propostas de Sistemas de Comunicação Baseadas em Semântica

Para a apresentação dos sistemas propostos, assumimos que as sentenças de entrada são representadas por $\mathbf{s} = [w_1, w_2, w_3, \ldots, w_k]$, em que w_l representa a l-ésima palavra da frase. O sinal codificado pode ser representado por

$$\mathbf{x} = C_\alpha(S_\beta(\mathbf{s})), \tag{147}$$

em que $\mathbf{x}$ são símbolos complexos necessários para transmissão, $S_\beta(\cdot)$ representa a rede de codificação semântica com o conjunto de parâmetros β e $C_\alpha(\cdot)$ é a rede de codificação do canal com o conjunto de parâmetros α.

Os símbolos $\mathbf{x}$ são transmitidos através de um canal de comunicação com ruído $\mathbf{n}$ e desvanecimento em pequena escala h. Assim, a saída do canal é

$$\mathbf{y} = h\mathbf{x} + \mathbf{n}. \tag{148}$$

Já na recepção o sinal codificado é representado por

$$\hat{\mathbf{s}} = S_\chi^{-1}(C_\sigma^{-1}(\mathbf{y})), \tag{149}$$

em que $\hat{\mathbf{s}}$ é o vetor com os *tokens* decodificados (menores unidades de significado, como palavras e números). Além disso, $S_\chi^{-1}(\cdot)$ representa o decodificador semântico com parâmetros χ e $C_\sigma^{-1}(\cdot)$ o decodificador de canal com parâmetros σ.

O sistema geral das propostas apresentadas nesse documento está ilustrado na Figura 16. Na imagem estão presentes as etapas de codificação e decodificação semântica além das etapas de codificação e decodificação de canal. Além disso, também está presente o canal de comunicação sem fio simulado para a transmissão das mensagens codificadas. As etapas de codificação e decodificação do canal foram implementadas utilizando redes neurais, conforme proposto em [1].

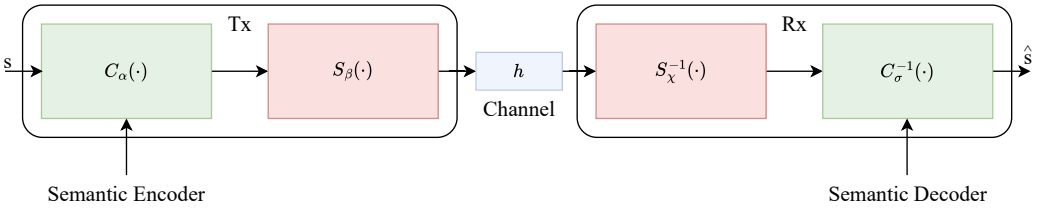


Figura 16: Diagrama do sistema de comunicação semântica baseado em [1].

As etapas de codificação e decodificação semântica são comparáveis, por analogia, às etapas de codificação e decodificação de fonte, em que símbolos de entrada são codificados de acordo com algum algoritmo, como Huffman. No entanto, quando utilizados algoritmos para a codificação semântica, são obtidos símbolos que se alocam em um espaço de acordo com a similaridade entre si.

Para o treinamento dos sistemas de comunicação baseados em semântica propostos, foi considerado o Pseudocódigo 8.1. Pelo Pseudocódigo 8.1, nota-se que o sinal codificado $\mathbf{x}$ é obtido pela codificação semântica e de canal dos símbolos $\mathbf{s}$. Em seguida, os símbolos codificados são transmitidos pelo canal de comunicação e é aplicada a decodificação de canal e a decodificação semântica para obter as sentenças $\hat{\mathbf{s}}$. Além disso, o canal de comunicação usado durante todo o treinamento foi puramente AWGN com 12 dB de SNR - fato esse que pode contribuir para análise dos resultados em outros modelos de propagação.

Algorithm 8.1 Algoritmo geral de treinamento.

Inicialização: Pesos iniciais α, β, σ e χ de passos anteriores.
Inputs: *Tokens* de entrada $\mathbf{s}$.
Transmissão:
    $\mathbf{x} \leftarrow C_\alpha(S_\beta(\mathbf{s}))$,
    Transmite $\mathbf{x}$ pelo canal de comunicação.
Recepção:
    Detecta $\hat{\mathbf{y}}$ no receptor,
    $\hat{\mathbf{x}} \leftarrow S_\chi^{-1}(C_\sigma^{-1}(\hat{\mathbf{y}}))$,
    Calcula $\mathcal{L}_{CE}$ pela Equação (150) comparando $\mathbf{s}$ e $\hat{\mathbf{s}}$,
    Treina S_α, C_β, C_σ^{-1} e $S_\chi^{-1} \leftarrow$ Gradiente descendente com $\mathcal{L}_{CE}$.
Saída: Atualiza os pesos das redes S_α, C_β, C_σ^{-1} e S_χ^{-1}.

Para atualizar os pesos α, β, σ, χ, uma boa escolha para a equação de perda é a entropia cruzada dada por

$$\mathcal{L}_{CE} = -\sum q(w_l)\log(p(w_l)) + (1 - q(w_l))\log(1 - p(w_l)), \tag{150}$$

em que $q(w_l)$ é a probabilidade de que a l-ésima palavra, w_l, apareça na sentença transmitida $\mathbf{s}$, e $p(w_l)$ é a probabilidade prevista de que a i-ésima palavra, w_i, apareça na sentença decodificada $\hat{\mathbf{s}}$. A CE foi mantida como função de perda para todas as outras propostas indicadas nessa seção.

Por fim, as métricas para análise dos resultados foram escolhidas de acordo com a sua frequência em outros estudos da área. A primeira delas é a BLEU (Bilingual Evaluation Undeststudy), proposta inicialmente em [79], e compara a frase decodificada com a original de

acordo com a frequência de grupos de palavras (ou n-gramas). Já a segunda métrica considerada é o BERTScore (ou similaridade de sentença), proposta pela primeira vez em [80], e considera os *embeddings* contextuais pré-treinados do modelo BERT [81]; é calculada, assim, a similaridade de cossenos entre o vetor de *embeddings* decodificado e original para definir a proximidade entre eles. Essa última métrica apresenta maior vantagem para analisar a similaridade semântica entre frases diferentes.

Uso de modelos de linguagem pré-treinados

Como proposta inicial de estudo, foi introduzido o uso de modelos de linguagem em larga escala pré-treinados (LLM) para a realização da codificação e decodificação semântica. Os modelos escolhidos devem ter camadas de codificação e decodificação e não ter dimensões que o tornem incompatíveis com um sistema de comunicação.

Ao revisar alguns repositórios de modelos, foram considerados, inicialmente, os modelos BART [75] e T5 [82]. Ambos possuem camadas de codificação e decodificação e sucessivas versões com menores quantidades de parâmetros. Foram escolhidas as versões BART-BASE, com 139 M de parâmetros, e T5-SMALL, com 60.5 M de parâmetros. Como comparação, o modelo de base tem aproximadamente 14 M de parâmetros.

O treinamento dos sistemas com modelos pré-treinados necessitou somente cinco épocas para a convergência a um valor de função de perda mínimo. Já para os modelos sem pré-treinamento houve a necessidade de ao menos 50 épocas. Tal ponto coloca os modelos com uso de LLM em evidência.

A Figura 17 apresenta os resultados das métrica BLEU ambientes Rayleigh propostos e a Figura 18 apresenta a métrica BERTScore para os três canais considerando. Tais resultados foram publicados em [83].

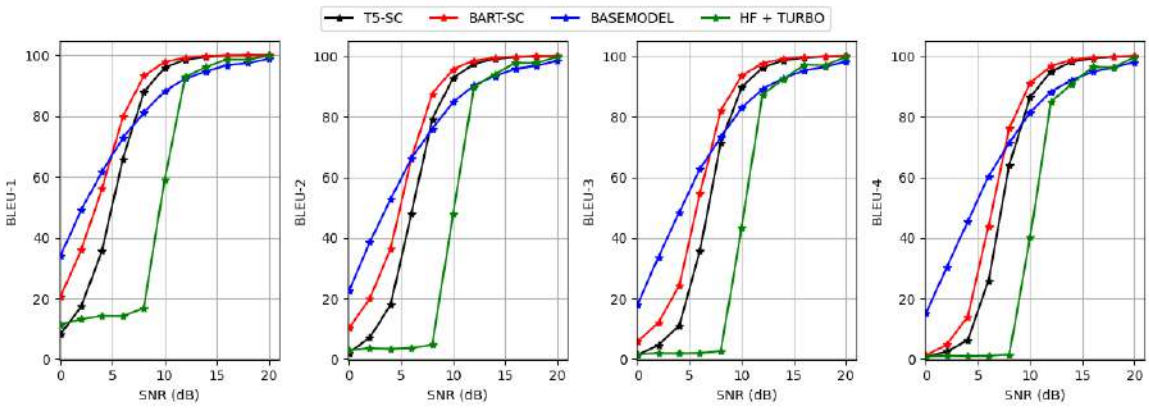


Figura 17: Resultados de BLEU para o canal Rayleigh.

Como uma análise primária, nota-se na Figura 17 que o modelo de base (BASEMODEL), obtido como base do desenvolvimento em [1], apresenta métricas de BLEU comparáveis aos modelos T5-SC e BART-SC. Pode-se observar que, em um ambiente com desvanecimento do tipo Rayleigh, o BASEMODEL tem melhor desempenho em condições piores de canal (menor SNR), enquanto os modelos com uso de LLM passam a ter melhor desempenho à medida que a SNR aumenta.

Ainda considerando a Figura 17, verifica-se que o uso do modelo BART-BASE com aproximadamente o dobro de parâmetros do modelo T5-SMALL, não resultou em ganhos proporcio-

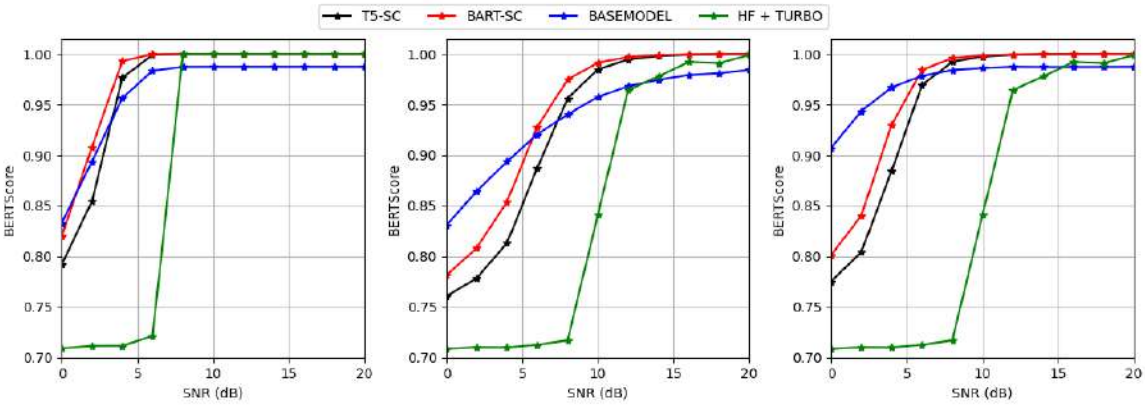


Figura 18: Resultados de BERTScore para os canais AWGN, Rayleigh e Rice.

nais de desempenho. Escalar ainda mais o número de parâmetros (e complexidade) talvez não seja, assim, uma alternativa eficiente para o desenvolvimento dessa proposta.

Para a métrica de BERTScore os resultados apresentados da Figura 18 estão de acordo com o desempenho dos resultados para a BLEU. Além disso, o modelo de base tem desempenho melhor em regiões de baixa SNR e pior em regiões de alto SNR quando considera-se o canal Rayleigh e, por fim, nota-se que o BASEMODEL não atinge valores de BERTScore unitários, como nos outros casos. Nem mesmo para altos valores de SNR isso ocorre.

Mesmo em cenários distintos daqueles utilizados no treinamento, os sistemas mantiveram desempenho satisfatório. Isso implica que os modelos tiverem uma boa capacidade de realizar o treinamento síncrono entre o codificador e o decodificador. A proposta de uso de modelos de linguagem, assim, mostra resultados promissores.

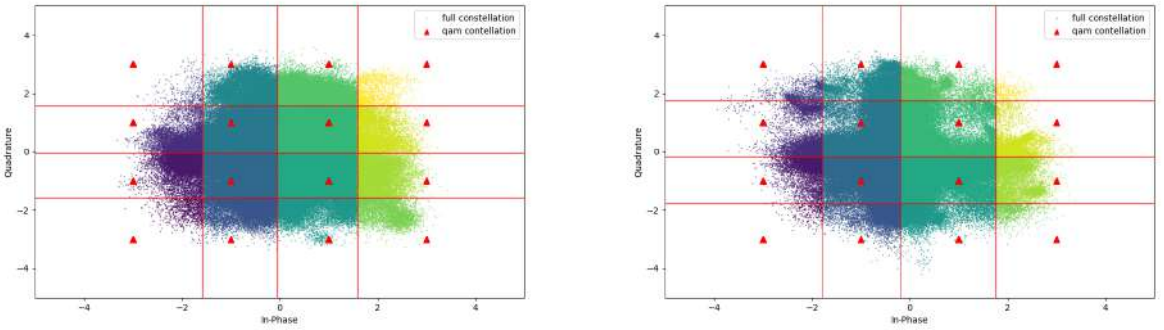
Modelos com etapa de modulação QAM não uniforme

A busca por um sistema de comunicação baseado em semântica que se aproxime mais da realidade induziu o grupo de pesquisa a considerar uma etapa de modulação dos símbolos codificados. Foi realizada uma revisão da literatura em busca de alternativas de modulação que se adequassem ao plano semântico gerado pelos símbolos após a codificação semântica. A referência [84] destacou-se como a mais promissora para esse objetivo.

Em [84] são consideradas duas formas de modulação QAM não uniformes: Mapping to Regular Constellation (MRC) e Mapping to Irregular Constellation (MIC). Ambas possuem parâmetros que são adaptados de acordo com o treinamento do modelo. A primeira possui como parâmetro a ser atualizado, por meio do *backpropagation*, os limites de quantização de definem a que ponto da constelação cada símbolo semânticos será modulado - nesse caso, os pontos da constelação são fixos. Já a segunda proposta implica na atualização da posição dos pontos da constelação a partir da movimentação dos símbolos semânticos.

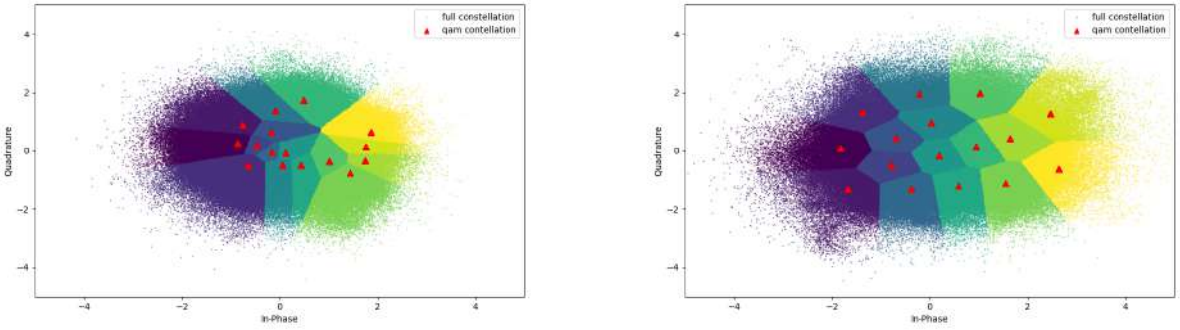
Durante o treinamento dos modelos, foram geradas imagens das constelações para monitoramento. O objetivo foi observar como as constelações nos mapeamentos mudam conforme os símbolos codificados se ajustam. As Figuras 19 e 20 mostram, respectivamente, os resultados do algoritmo MRC e MIC.

Pela Figura 19, observa-se que as áreas destinadas a cada símbolo QAM, destacadas em vermelho, ajustam-se conforme a densidade de pontos. Como observado na Figura 20, os pontos vermelhos da constelação QAM se posicionam conforme o algoritmo apresentado em [84], movendo-se mais ao centro da nuvem de símbolos codificados, onde a densidade é maior.



(a) Resultado após 1000 batches. (b) Resultado após 228000 batches.

Figura 19: Resultados preliminares de treinamento para o algoritmo MRC.



(a) Resultado após 1000 batches. (b) Resultado após 228000 batches.

Figura 20: Resultados preliminares de treinamento para o algoritmo MIC.

Como consideração final acerca dessa etapa da pesquisa, os resultados de BLEU e BERTScore para o uso dos algoritmos MRC e MIC estão sendo processados. Suas divulgações serão realizadas em momentos oportunos.

8.3 Conclusão

Mesmo com algumas propostas em ainda em etapa de aquisição de resultados, conclusões podem ser obtidas. A primeira delas é a boa alternativa de uso de modelos de linguagem pré-treinados. Isso é observado quando analisadas as Figuras 17 e 18. São notados valores maiores de BLEU e BERTScore quando consideradas variações de SNR propostas nos canais AWGN, Rayleigh e Rice. No entanto, nota-se que o modelo de base proposto possui melhores resultados em regimes de SNR menores, ou seja, com maior degradação do sinal enviado.

Além disso, o uso de algoritmos de modulação não lineares mostra-se promissor ao explorar o espaço gerado pelos símbolos semânticos. Com a alocação adaptada de pontos de modulação de acordo com a disposição dos símbolos, no decorrer do treinamento dos modelos, obtêm-se constelações e limiares de modulação mais bem adaptados às características do sistema proposto. Essa adaptação tem potencial para melhorar o desempenho em ambientes com baixa qualidade de canal, especialmente sob regimes de baixa SNR.

Tomando-se como base os avanços apresentados até aqui, nota-se que outras abordagens

devem ser consideradas. Sendo assim, o grupo de pesquisa pretende explorar abordagens alternativas às propostas apresentadas. Dentre elas estão o uso de modelos de quantização vetorial, a continuação de estudos de modelos de linguagem pré-treinados — com a aplicação de outros modelos e alternativas de testes — e, por fim, a aplicação de testes em canais de comunicação sem CSI.

Por fim, o grupo de estudos mantém o foco em apresentar propostas de sistemas de comunicação baseados em semântica que demonstrem a efetividade e robustez desses modelos. Os resultados futuros serão documentados em publicações científicas, estimulando debates relevantes sobre a aplicabilidade dessas propostas em cenários reais.

9 Modelos de Fontes Semânticas para Relações Taxa-distorção Não-separáveis

Eduardo Nunes Velloso e José Cândido Silveira Santos Filho
UNICAMP

9.1 Introdução

Sistemas de comunicação digital são projetados com o objetivo de garantir a transmissão de informação por um canal ruidoso. Embora essa abordagem seja bem sucedida para aquilo que se propõe a fazer, isto é, viabilizar aplicações de transferência de dados, grande parte das aplicações emergindo atualmente envolvem tomadas de decisão automáticas baseadas em grandes volumes de dados continuamente coletados por sensores. Essas aplicações motivam o desenvolvimento de um novo paradigma que prioriza a preservação do conteúdo semântico de uma mensagem, ao invés da mensagem em si [85, 86].

Um elemento chave para essa abordagem de comunicação semântica é que o codificador tem acesso a uma observação extrínseca, mas não ao estado intrínseco que deve ser preservado no receptor. A teoria de informação de Shannon ainda se aplica a essa nova classe de problemas, apesar de explicitamente desconsiderar a semântica das mensagens na sua formulação do problema central da comunicação. De fato, esse problema pode ser entendido como um caso de transmissão indireta, o qual pode ser analisado pela teoria de Shannon conforme [87]. Para isso, a medida de distorção semântica é transformada em uma medida análoga entre a observação e o estado reconstruído.

Essa formulação de comunicações semânticas é também aplicável ao cenário de compressão de dados com perdas. A demanda de aplicações emergentes por taxas de compressão extremas reduz a importância do objetivo usual de minimização do erro de reconstrução. Esse problema de compressão orientada a tarefa é fundamentalmente limitado por uma relação taxa-distorção, cuja compreensão é essencial para o projeto de sistemas práticos.

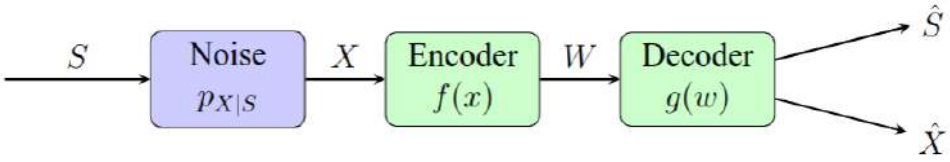


Figura 21: Modelo de compressão para uma fonte semântica.

Um modelo de fonte semântica recente proposto envolve a combinação de um problema de transmissão indireta com um problema de múltiplas fontes, com o objetivo de satisfazer simultaneamente condições de distorção sintática e semântica [88, 89]. Nesse caso, assumindo uma distribuição conjunta conhecida entre o estado S e a observação X, o objetivo final é projetar um codificador e um decodificador com um taxa de compressão R tal que as reconstruções $\hat{S}$ e $\hat{X}$ estejam suficiente próximas dos sinais originais de acordo com alguma métrica de distorção adequada para a aplicação. Esse modelo de compressão para uma fonte semântica desse tipo é ilustrado na Figura 21.

Entender a função taxa-distorção semântica é um passo essencial para o projeto de sistemas práticos. Por exemplo, códigos com desempenhos semelhantes na reconstrução da observação

X podem ser significativamente diferentes em termos da capacidade de reconstruir fielmente o estado S. Por esse motivo, estudos teóricos a respeito dos limites fundamentais desse modelo têm sido desenvolvidos ao longo dos últimos anos. Nesse trabalho, métodos teóricos e soluções numéricas serão investigados para o caso em que as duas métricas de distorção estão acopladas de maneira não-separável. As condições de otimalidade encontradas terão consequências para possíveis estratégias de projeto.

9.2 Formulação do Problema

O problema de compressão com perdas tradicional tem como solução uma relação taxa-distorção definida pelo seguinte problema de otimização com restrições:

$$R(D) = \inf_{q(\hat{x}|x)} I(X;\hat{X}) \tag{151}$$

$$\text{s.a. } \mathbb{E}_{X,\hat{X}} d(x,\hat{x}) \leq D \tag{152}$$

em que $R(D)$ é a menor taxa de compressão R para a qual existe algum código com valor esperado da distorção entre a observação X e sua reconstrução $\hat{X}$ abaixo da tolerância D. A minimização ocorre para todos os possíveis mapeamentos probabilísticos entre entrada e saída $q(\hat{x}|x)$.

Quando estendemos o problema para o modelo apresentado na Figura 21, os mesmos princípios ainda podem ser aplicados, desde que a superfície taxa-distorção $R(D_s,D_x)$, descrita pelas duas tolerâncias D_s e D_x, seja agora dada pela solução desse novo problema de otimização:

$$R(D_s,D_x) = \inf_{q(\hat{s},\hat{x}|x)} I(X;\hat{S},\hat{X}) \tag{153}$$

$$\text{s.a. } \mathbb{E}_{X,\hat{S}} \tilde{d}_s(x,\hat{s}) \leq D_s \tag{154}$$

$$\mathbb{E}_{X,\hat{X}} d_x(x,\hat{x}) \leq D_x \tag{155}$$

onde $\tilde{d}_s(x,\hat{s}) = \mathbb{E}_{S|x} d(s,\hat{s})$ é uma modificação da distorção semântica para torná-la função da única entrada observável.

Para resolver o problema acima, assumimos que conhecemos a fonte em termos da distribuição conjunta $p(s,x)$ e dos alfabetos de possíveis valores dos sinais S e X. Além disso, conhecemos as métricas de distorção d_s e d_x definidas pela tarefa no receptor. No caso discreto, podemos enumerar as possíveis combinações dos alfabetos de entrada e representar os valores esperados por somatórios. A lagrangeana do problema primal nesse cenário é, portanto:

$$\mathcal{L}(q,\lambda_s,\lambda_x,\mu) = \sum_{ijk} \log\left(\frac{q_{ijk}}{\nu_{jk}}\right) q_{ijk} p_i - \lambda_s \left(\sum_{ijk} \tilde{d}_{sij} q_{ijk} p_i - D_s\right) - \tag{156}$$

$$\lambda_x \left(\sum_{ijk} d_{xik} q_{ijk} p_i - D_x\right) + \sum_i \mu_i \left(\sum_{jk} q_{ijk} - 1\right) \tag{157}$$

em que λ_s, λ_x e μ são multiplicadores de Lagrange, e a expressão para a informação mútua foi explicitada.

A solução para esse problema dual é dada pelo $q(\hat{s},\hat{x}|x)$ que maximiza a expressão acima, isto é, que torna a derivada igual a zero:

$$\frac{\partial \mathcal{L}}{\partial q} = 0 \Leftrightarrow \begin{cases} q_{ijk} & = \dfrac{\exp(\lambda_s \tilde{d}_{sij} + \lambda_x d_{xik})\nu_{jk}}{\sum_{jk} \exp(\lambda_s \tilde{d}_{sij} + \lambda_x d_{xik})\nu_{jk}} \\ \nu_{jk} & = \sum_i q_{ijk} p_i \end{cases} \tag{158}$$

Essa equação se aplica para o caso geral, e técnicas de otimização alternada podem ser empregadas para encontrar os valores de q_{ijk} e ν_{jk} iterativamente.

Métricas de Distorção Não-separáveis

Os dois exemplos de métricas de distorção acopladas que investigaremos são baseados em casos de uso de sistemas realistas. O primeiro cenário é o de um sistema médico, no qual a observação consiste em um sinal medido por um exame (e.g. imagens de tomografia ou sinais de eletrocardiograma) e o estado consiste no diagnóstico associado a tal exame (rótulos de classes discretas). Podemos imaginar que a distorção do exame comprimido é irrelevante se o diagnóstico for recuperado com exatidão, e deve ser minimizada em termos de mínimos quadrados do ruído de compressão caso contrário. Em outras palavras, uma métrica para representar essa tarefa seria

$$d(s,x,\hat{s},\hat{x}) = \mathbb{1}\{s \neq \hat{s}\}||x - \hat{x}||^2 \tag{159}$$

O segundo exemplo trata de tarefas que exigem confiabilidade na coerência entre as duas reconstruções. Por exemplo, um sistema de câmeras de segurança pode requerer que, além de baixos erros na imagem e seu rótulo, que qualquer erro afete ambos de maneira similar, permitindo interpretação e intervenção humana. Outro exemplo são sistemas em que o estado são legendas textuais que descrevem imagens (observação), caso no qual também é importante que pequenas variações na imagem reconstruída sejam acompanhadas por variações coerentes na sua legenda. De maneira genérica, essa métrica de distorção que leva em conta a coerência pode ser escrita como

$$d(s,x,\hat{s},\hat{x}) = \lambda_s d_s(s,\hat{s}) + \lambda_x d_x(x,\hat{x}) + \lambda_{sx} d_{sx}(\hat{s},\hat{x}) \tag{160}$$

9.3 Conclusão

Embora o trabalho esteja em andamento, os resultados teóricos obtidos estabelecem uma base sólida para futuras simulações e análises numéricas, especialmente no contexto de aplicações práticas como diagnóstico médico assistido e interpretação de imagens com suporte semântico. Próximos passos incluirão a implementação computacional dos métodos propostos e a investigação empírica do comportamento das soluções sob diferentes métricas de distorção.

10 Estratégias de Acesso Não-Ortogonal para IoT Direto ao Satélite

Richard Demo Souza
UFSC

10.1 Introdução

A dinâmica da vida contemporânea impõe um desafio extraordinário aos futuros sistemas de comunicação sem fio.Desde o surgimento da Primeira Geração de Rede Móvel Celular (1G) até a atual era da 5G, pesquisadores e a indústria desempenharam papéis essenciais na interconexão global e na elevação dos níveis de desempenho e qualidade de serviço. No entanto, um marco notável foi recentemente alcançado: o número de dispositivos conectados superou o número de humanos conectados [90]. Esse avanço significativo abre caminho para a economia das coisas [90], possibilitando uma diversidade de aplicações com cerca de 30 bilhões de dispositivos IoT conectados[4] .

A tecnologia conhecida como *Machine Type Communications* (MTC) massiva é crucial para as demandas de conectividade em cidades inteligentes, fábricas e logística, promovendo conectividade global [91]. Ainda assim, aplicações remotas, como monitoramento climático e logística marítima, enfrentam dificuldades devido à infraestrutura limitada de comunicação, uma lacuna que pode ser preenchida com soluções de redes NTN e tecnologias IoT de longo alcance [92, 93]. O *3rd Generation Partnership Project* (3GPP) visa a 6G no *Release* 20, focando na integração contínua entre soluções terrestres e NTN com uso de tecnologias avançadas para habilitar novas aplicações como teleoperação, gêmeos digitais e veículos autônomos [94]. Além disso, o conceito de DtS-IoT lidera as inovações em NTN utilizando *gateways* embarcados em satélites, o que reduz os custos de infraestrutura, mas enfrenta desafios como alta instabilidade de canal e perdas por multipercurso ao longo da trajetória orbital [95]. Enquanto satélites geoestacionários se movem com a rotação da Terra e oferecem conectividade fixa em uma área, os satélites LEO se movem a cerca de 7 km/s, atendendo diferentes regiões de acordo com seu movimento. Apesar da dinâmica mais complexa, os satélites LEO são muito mais baratos que os geoestacionários e crescem rapidamente em número.

As tecnologias de *Low Power Wide Area Network* (LPWAN) terrestres surgem como alternativa promissora [96] para conectar dispositivos terrestres a satélites. Seu baixo consumo energético, longo alcance e reduzida necessidade de infraestrutura tornam as LPWAN atraentes para aplicações DtS-IoT. Em particular, a tecnologia *Long Range Wide Area Network* (LoRaWAN) [97] emprega a modulação *Long Range* (LoRa) na camada física, sendo a Lacuna Space uma empresa pioneira oferecendo DtS-IoT com tal tecnologia. A Lacuna Space inclusive recentemente expandiu sua cobertura com o lançamento bem-sucedido da segunda geração do Lacuna Space, embarcado na plataforma modular da NanoAvionics e lançado via Space-X Transporter−7 [98].

Apesar da simplicidade, protocolos de acesso aleatório, como os utilizados em LoRaWAN, sofrem com baixa escalabilidade e alta taxa de colisões em implantações densas [99]. Em NTN, o cenário se agrava devido a fatores dinâmicos e restrições temporais de visibilidade impostas pela órbita do satélite. Estratégias com NOMA surgem como soluções candidatas para reduzir

[4]O conteúdo desta seção é um resumo do artigo submetido F. A. Tondo, J. M. de S. Sant'Ana, S. Montejo-Sánchez, O. L. Alcaraz López, S. Céspedes, and R. D. Souza, "Non-Orthogonal Multiple-Access Strategies for Direct-to-Satellite IoT Networks," https://arxiv.org/abs/2409.02748

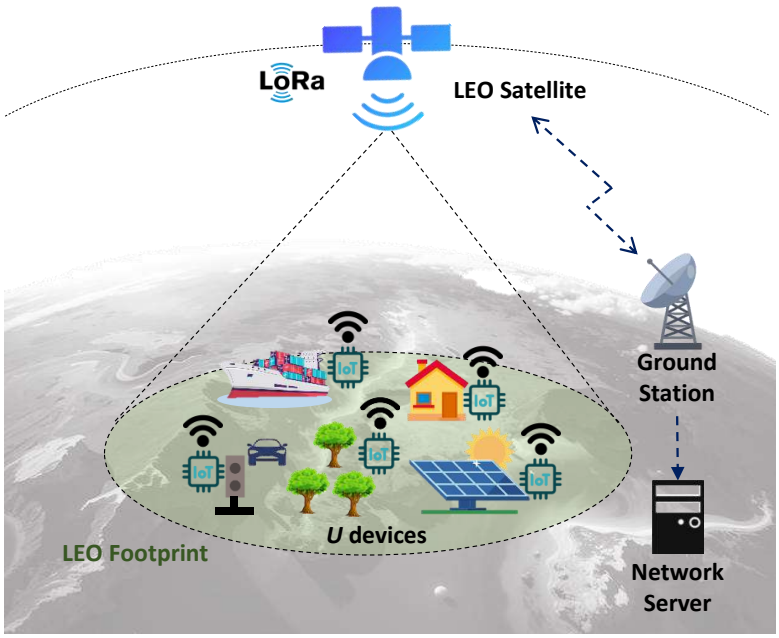


Fig. 22. A arquitetura DtS-IoT é composta por um *gateway* embarcado em um satélite LEO, dispositivos IoT distribuídos sobre a área de interesse, uma estação terrestre e um servidor de rede.

colisões não resolvidas no receptor e aumentar a eficiência do sistema [100]. Diversos trabalhos investigam o NOMA no domínio da potência em cenários DtS-IoT, porém com foco predominante no enlace de descida [101, 102]. Outros estudos buscam melhorar o desempenho do enlace de subida com aprendizado de máquina [103] ou técnicas cooperativas [104]. Apesar dos avanços, tais abordagens podem ser impraticáveis devido à necessidade de *gateways* sofisticados ou informações extras de sincronização dos dispositivos.

Neste trabalho, propomos novas abordagens de enlace de subida para redes DtS-IoT baseadas em LoRaWAN, utilizando NOMA no domínio da potência e explorando o conhecimento da trajetória do satélite. Nosso objetivo é melhorar a escalabilidade do enlace de subida sem comprometer o consumo de energia e sem assumir sincronização estrita entre os dispositivos.

10.2 Desenvolvimento

Nesta seção, introduzimos duas estratégias de transmissão para o enlace de subida em um cenário de DtS-IoT, como ilustrado na Figura 22. Assumimos que os dispositivos possuem um preditor da órbita do satélite, algo que já é comum inclusive em aplicações comerciais com as da Lacuna Space. Dado esse conhecimento, uma estratégia direta seria usar ALOHA puro, permitindo que os dispositivos transmitam livremente em suas janelas de visibilidade, com potência fixa. Embora simples de implementar, essa abordagem pode levar a muitas colisões, já que diversas janelas de visibilidade podem se sobrepor. Outra alternativa seria transmitir em um instante específico, como no momento do ângulo de elevação máximo (*Max Angle*), correspondente à menor distância até o satélite e maior potência recebida. Contudo, isso reduz a janela efetiva de visibilidade do satélite e aumenta a probabilidade de colisão.

Uma alternativa promissora seria algo entre ALOHA puro e uma única oportunidade de transmissão. Por exemplo, permitir que os dispositivos transmitam em diferentes posições da

janela de visibilidade, reduzindo colisões devido às localizações distintas. Outra possibilidade interessante é aplicar NOMA no domínio da potência com SIC no *gateway*. Assim, mais de um dispositivo pode transmitir simultaneamente com sucesso, aliviando os problemas do ALOHA puro. A seguir, apresentamos duas estratégias inéditas de enlace de subida baseadas nessas ideias.

FTP: Na estratégia FTP, os dispositivos utilizam uma potência de transmissão fixa, escolhendo o instante de transmissão na janela de visibilidade de forma que a potência média recebida no *gateway* seja $\mathcal{L}_l$, para $l \in 1, 2, \ldots, L$, onde L é o número de níveis predefinidos de potência. Isso é feito invertendo a perda de percursos com potência fixa P, resultando em d^{FTP}. Assim, com potência fixa P, o dispositivo escolhe o instante de transmissão em que sua distância ao satélite d^{FTP} produza a perda de percurso desejada, atingindo o nível de potência recebido alvo. Como a órbita do satélite é simétrica sobre uma região, há no máximo $2 \times L$ posições possíveis por dispositivo: L durante a ascensão e L durante a descida. A posição usada deve ser escolhida aleatoriamente entre essas. Assumimos que os níveis de potência são suficientemente distintos para permitir aplicação de NOMA com SIC e que cada nível possui um piloto ortogonal.

CTP: A estratégia FTP limita as transmissões a instantes específicos, ajudando a reduzir colisões, mas é vulnerável a dispositivos com distâncias muito semelhantes ao satélite. Para contornar isso, a estratégia CTP permite transmissões a qualquer instante da órbita, ajustando a potência de transmissão conforme a posição na janela de visibilidade, potência esta definida como P^{CTP}, respeitando uma potência máxima.

Note que existe relação entre distância e ângulo de elevação. Para órbitas com ângulo de elevação máximo baixo, certos níveis de potência podem ser inalcançáveis por alguns dispositivos. Adicionalmente, desvanecimento e sombreamento podem afetar a potência recebida, comprometendo o sucesso do SIC caso os níveis reais estejam muito próximos.

Para a obtenção de resultados numéricos, consideramos o uso de tecnologia LoRa e que os dispositivos estão uniformemente distribuídos sobre a França. Além disso, consideramos a órbita real do satélite LEO LacunaSat-3, posicionado a uma altitude de 500 a 600 km, e que o ângulo mínimo de elevação é 30°.

A Fig. 23 mostra o *goodput* dos esquemas propostos FTP (vermelho, marcador x) e CTP (verde claro, quadrado), além do ALOHA (marrom, triângulo) e *Max Angle* (roxo, triângulo invertido), em função do número de dispositivos, com carga útil de $b = 20$ bytes. Para $U = 100$ dispositivos, FTP apresenta ganho de 28% (992 bytes/volta) sobre o ALOHA (773 bytes/volta), enquanto CTP melhora em 18% (913 bytes/volta). Respectivamente, FTP e CTP também superam o *Max Angle* em 234% e 207% (297 bytes/volta). FTP tem a vantagem de menos colisões com poucos dispositivos, já que transmite em intervalos específicos. No entanto, com mais dispositivos, crescem as sobreposições temporais, elevando a taxa de colisão, cenário onde CTP passa a ter melhor desempenho. No exemplo da Fig. 23, isso ocorre para $U > 400$. Com $U = 600$, CTP supera ALOHA em 60% (341 bytes/volta vs. 213 bytes/volta). Além disso, o desempenho do FTP se aproxima do ALOHA devido ao efeito de captura. Em cenários com tráfego intenso, muitos dispositivos compartilham janelas de visibilidade semelhantes, aumentando a chance de escolherem os mesmos pontos de transmissão. Esse ponto de transição é visível na Fig.23 em $U = 600$, quando o momento da transmissão deixa de fazer diferença. Por fim, confirmamos a expectativa de haver um alto número de mensagens malsucedidas na estratégia *Max Angle*.

Para avaliar os esquemas propostos em termos de eficiência energética (bytes/Joule), a

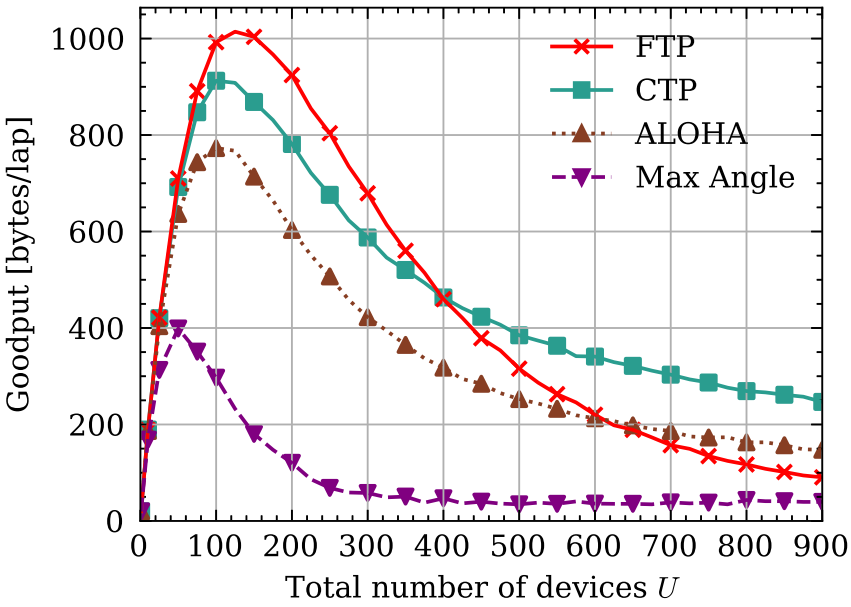


Fig. 23. Número médio de bytes recebidos com sucesso por volta, em função do número de dispositivos IoT, para os esquemas propostos FTP e CTP, assim como para o ALOHA e a estratégia *Max Angle*.

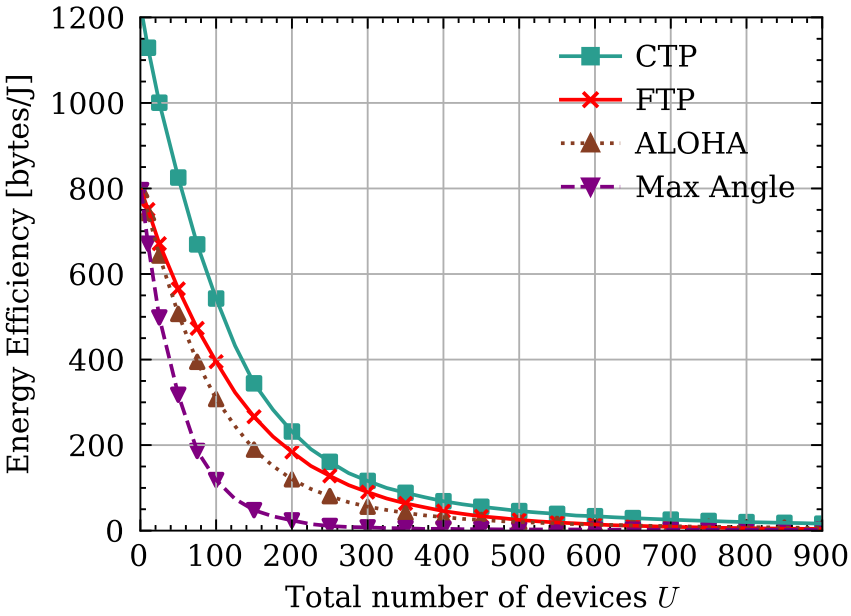


Fig. 24. Eficiência energética em função do número de dispositivos IoT para os esquemas propostos FTP (em vermelho, marcador x) e CTP (em verde claro, marcador quadrado), bem como para as abordagens ALOHA (em marrom, marcador triangular) e Max Angle (em roxo, triângulo invertido).

Fig. 24 apresenta essa métrica em função do número de dispositivos IoT U. Nota-se que o método CTP oferece uma vantagem significativa em eficiência energética em relação ao FTP,

ALOHA e à estratégia *Max Angle*. Com $U = 100$ dispositivos, o CTP proporciona um ganho de 37,50% em eficiência energética em relação ao FTP, enquanto esse ganho é de 76,30% em relação ao ALOHA. Além disso, o CTP supera as abordagens FTP, ALOHA e *Max Angle* em eficiência energética para todas as quantidades de dispositivos IoT avaliadas, mas em termos de *goodput*, o CTP é superado pelo FTP em redes de baixa densidade.

10.3 Conclusão

Este trabalho apresentou dois novos esquemas de múltiplo acesso DtS-IoT baseados em NOMA no domínio da potência. Para explorar posições específicas dentro da janela de visibilidade, propusemos as estratégias FTP e CTP com potência fixa e controlada, respectivamente. Avaliamos o goodput e a eficiência energética, comparando com ALOHA e *Max Angle*. Os métodos propostos superam amplamente as abordagens de referência em *goodput*, com FTP se destacando em redes menos densas e CTP em cenários com muitos dispositivos. Como esperado, a estratégia CTP apresenta maior eficiência energética ao evitar colisões, independentemente do número de dispositivos.

11 Técnicas de Transmissão Alternativas para LR-FHSS

Richard Demo Souza
UFSC

11.1 Introdução

As LPWAN tornaram-se cada vez mais relevantes nos últimos anos devido à sua capacidade de fornecer conectividade de longo alcance, baixo consumo de energia e baixo custo para aplicações de IoT. Tecnologias LPWAN podem ser implantadas em áreas remotas e desconectadas para viabilizar diferentes aplicações [105, 106]. Dentre as tecnologias LPWAN, o LoRaWAN [107] destaca-se como uma solução de longo alcance e baixo consumo de energia. Por operar em espectro não licenciado, o LoRaWAN representa uma solução de conexão de baixo custo e tem sido amplamente reconhecido como um padrão global de LPWAN, com reconhecimento formal pela União Internacional de Telecomunicações (UIT) [108]. Essa padronização garante interoperabilidade, conformidade regulatória e ampla adoção em diferentes setores[5].

O protocolo de rede LoRaWAN opera tipicamente sobre a tecnologia de camada física LoRa [109]. Apesar de os fatores de espalhamento quase ortogonais do LoRa permitirem a coexistência de múltiplos dispositivos, é difícil alocar de forma ideal os fatores de espalhamento em ambientes de rede dinâmicos, limitando a capacidade total da rede. Por exemplo, selecionar o fator de espalhamento ideal é particularmente desafiador em comunicações DtS, especialmente em órbitas baixas, onde os satélites estão em constante movimento ao redor da Terra. Esses desafios motivaram o desenvolvimento e a introdução recente da camada física *Long Range - Frequency Hopping Spread Spectrum* (LR-FHSS) nas especificações do LoRaWAN [110]. O LR-FHSS adota modulação de faixa estreita e salto em frequência, e foi projetado para melhorar o desempenho em redes massivas, como aquelas utilizadas em soluções DtS-IoT com alta densidade de dispositivos na região de cobertura do satélite. É importante ressaltar que o LR-FHSS oferece suporte apenas à comunicação no enlace de subida, enquanto LoRa pode ser utilizado tanto nos enlaces de subida quanto de descida.

Estudos teóricos e experimentos recentes em DtS confirmaram a viabilidade do conceito de comunicação direta ao satélite e identificaram desafios e oportunidades em aberto [96, 95, 111], entre eles a integração de satélites LEO com LR-FHSS, que tem atraído significativa atenção acadêmica e industrial devido à sua relevância para soluções IoT de larga escala [106, 105]. Apesar da viabilidade demonstrada do LR-FHSS para comunicações DtS, a confiabilidade permanece uma questão crítica, especialmente com o aumento da densidade da rede.

Na literatura existente, técnicas de transmissão redundante têm se mostrado eficazes para melhorar a confiabilidade e a escalabilidade das LPWAN [99]. Por exemplo, a replicação de mensagens melhora a escalabilidade da rede ao transmitir múltiplas cópias da mesma mensagem. Essa abordagem explora a diversidade temporal para aumentar a probabilidade de que pelo menos uma cópia seja recebida com sucesso pelo *gateway*. A replicação tem sido amplamente estudada na literatura e até incorporada a padrões de LPWAN [112, 113]. Um conceito semelhante também é empregado no Release 13 do 3GPP, que inclui transmissões repetitivas para melhorar a confiabilidade e a cobertura de IoT via rede celular [114].

[5]O conteúdo desta seção é um resumo de um artigo que está em preparação: L. B. Rocha, J. M. S. Sant'Ana, M. A. Ullah, S. Montejo-Sánchez, J. L. Rebelatto, and R. D. Souza, "Towards Improving the Scalability of LR-FHSS: Transmission Policies and Performance Analysis".

De forma similar, o LR-FHSS também adota parcialmente o conceito de transmissão redundante ao transmitir múltiplas cópias de cabeçalhos, com o objetivo de melhorar a confiabilidade e, consequentemente, a escalabilidade da rede. Isso pode ser feito individualmente (replicação), conforme o modo de operação padrão do LR-FHSS [110]. No entanto, essa redundância também acarreta maior congestionamento da rede, aumentando o número de colisões e limitando os ganhos de desempenho. Assim, ainda existe a necessidade de otimizar as técnicas de transmissão de mensagens para melhorar a escalabilidade do LR-FHSS.

Este trabalho apresenta a proposta conceitual de três técnicas de transmissão de mensagens para comunicações LR-FHSS, concatenação, retransmissão embutida e replicação, assim como uma avaliação preliminar de desempenho.

11.2 Desenvolvimento

Esta seção apresenta conceitualmente as técnicas de transmissão de mensagens empregadas neste trabalho. Termos-chave são definidos para garantir clareza na discussão:

- *Mensagem*: Unidade básica de informação gerada pelos dispositivos, como leituras de sensores (por exemplo, temperatura, umidade, pressão).
- *Payload*: Parte do pacote LR-FHSS que contém mensagens codificadas. Um único *payload* pode conter uma ou mais mensagens geradas pelo mesmo dispositivo final, dependendo da técnica de transmissão empregada.
- *Pacote*: Unidade completa de transmissão LR-FHSS, composta por todos os cabeçalhos replicados e fragmentos de *payload*.

Concatenação: A concatenação de mensagens funciona armazenando mensagens geradas por um único dispositivo até que se alcance um número predefinido de c mensagens. Essas mensagens são então concatenadas em um único *payload* e transmitidas como um único pacote, conforme ilustrado na Figura 25. Essa técnica aumenta tanto o intervalo médio entre transmissões quanto o tamanho do *payload* por transmissão, reduzindo o número total de transmissões por um fator c. Considerando que redes LR-FHSS tendem a ser densas, o tráfego de rede tem papel essencial na probabilidade de colisão e no desempenho global. Assim, espera-se que a concatenação de mensagens seja vantajosa na redução da ocupação do canal e na melhora do desempenho da rede. Para que a concatenação seja eficaz, as mensagens individuais devem ser suficientemente curtas para que múltiplas delas caibam dentro dos limites de tamanho de *payload*. Além disso, essa técnica aumenta o risco de perda de informação: a perda de um único pacote pode acarretar na perda de várias mensagens, o que é particularmente problemático em ambientes com condições adversas de canal.

Retransmissão Embutida: A técnica de retransmissão embutida introduz redundância transmitindo cada mensagem p vezes ao longo de p pacotes sucessivos, garantindo que a falha de um único pacote não resulte necessariamente na perda da mensagem. Isso é feito concatenando-se as p mensagens mais recentes em cada *payload* transmitido, conforme ilustrado na Figura 26. Em comparação com a concatenação simples, essa abordagem oferece o benefício adicional da redundância, mantendo um atraso médio mais baixo, já que cada nova mensagem é transmitida tão logo esteja disponível. Contudo, essa técnica aumenta a ocupação da rede devido ao maior tamanho do *payload* por transmissão. Portanto, os benefícios da redundância devem superar os custos de congestionamento para que a técnica seja eficaz. Assim como na concatenação, a

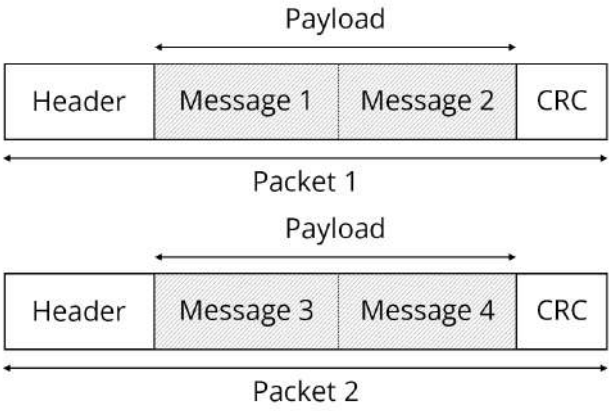


Figura 25: Descrição da estrutura do pacote LR-FHSS quando a concatenação é aplicada com $c = 2$. O padrão hachurado indica que as mensagens foram codificadas em um único *payload*.

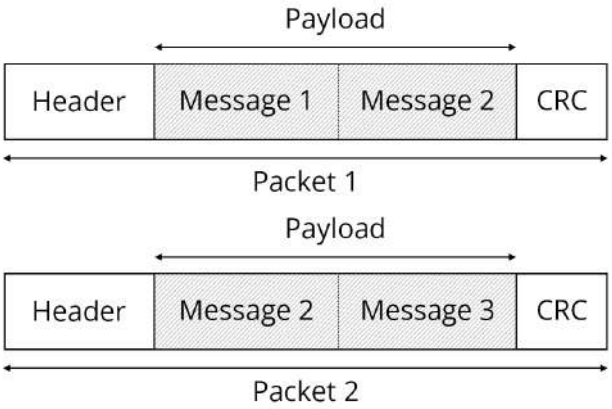


Figura 26: Estrutura do pacote LR-FHSS quando a retransmissão embutida é aplicada com $p = 2$. O padrão hachurado indica que as mensagens foram codificadas em um único *payload*.

viabilidade da retransmissão embutida depende das mensagens serem curtas o suficiente para que múltiplas caibam no *payload* máximo do LR-FHSS.

Replicação: Diferentemente da estratégia anterior, que incorpora múltiplas mensagens em um único *payload*, a replicação introduz redundância no nível de pacote, garantindo que cada pacote transporte apenas uma mensagem por *payload*. Como ilustrado na Figura 27, essa técnica consiste em transmitir cada mensagem n vezes em n pacotes distintos. Isso aumenta a chance de recepção bem-sucedida e assegura independência entre transmissões. A principal vantagem é que cada réplica enfrenta diferentes condições de canal e cenários de colisão, reduzindo o risco de perdas sistemáticas por interferência de transmissões de outros dispositivos ou por degradação do canal. Vale destacar que essa estratégia é amplamente utilizada em tecnologias LPWAN, para compensar a ausência de confirmações no *downlink* e aumentar a confiabilidade em canais de longa distância sujeitos a interferências. A desvantagem da replicação é o aumento da carga de rede, pois ela reduz o intervalo médio entre transmissões e aumenta o número total

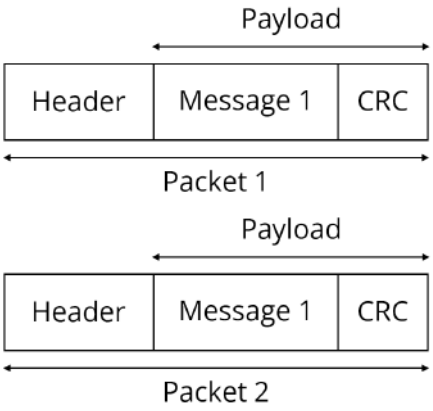


Figura 27: Descrição da estrutura do pacote LR-FHSS quando a replicação é aplicada com $n = 2$.

de pacotes enviados. Esse tráfego adicional pode intensificar colisões, especialmente em redes densas.

Vale ressaltar que essas técnicas operam totalmente no nível do servidor de aplicação, sem necessidade de alterações no *gateway* ou no servidor de rede. Isso significa que *payloads* contendo múltiplas mensagens (como na concatenação e na retransmissão embutida) ou mensagens replicadas (como na replicação) são interpretados e processados pelo servidor de aplicação, sem impactar a infraestrutura da rede. Essa escolha de projeto simplifica a implementação e amplia a escalabilidade dessas técnicas em implantações reais de IoT.

Para avaliar o desempenho dos métodos discutidos anteriormente, foram feitas simulações computacionais utilizando o LR-FHSS-Sim [115], um simulador gratuito e de código aberto escrito em Python para redes LR-FHSS. Novas funcionalidades foram incluídas no simulador, de modo a considerar as três técnicas de transmissão investigadas neste trabalho. Consideramos mensagens com $L = 10$ bytes e a configuração DR8 do LR-FHSS, caracterizada por sua elevada robustez, devido a uma taxa de codificação de $1/3$ no *payload* e $N = 3$ repetições de cabeçalho por pacote, o que a torna adequada para cenários que exigem comunicação confiável em ambientes densos ou com alta interferência. Além disso, assumimos que os dispositivos geram tráfego segundo uma distribuição exponencial, com intervalo médio de 900 segundos entre transmissões, enquanto os resultados consideram um total de 18 000 segundos de tempo de simulação. Ademais, assumimos que o canal tem uma probabilidade de apagamento de mensagens de 10%, um valor coerente para redes LPWAN, de acordo com fabricantes e a literatura. Uma comparação da probabilidade de sucesso média das três técnicas de transmissão é apresentada na Fig. 28, considerando apenas o melhor valor de parâmetro para cada técnica e uma combinação das técnicas redundantes com concatenação. A concatenação supera os demais métodos em redes de alta densidade, que é o ponto de operação mais relevante para redes LR-FHSS. Além disso, a retransmissão embutida é superior à replicação, pois equilibra melhor a redundância e a carga da rede. Quando combinadas com a concatenação, ambas as técnicas redundantes apresentam melhores resultados do que quando utilizadas isoladamente, beneficiando-se da redução da sobrecarga de transmissão. Por exemplo, $(c = 2, p = 2)$, em que duas mensagens são concatenadas em um único pacote e essas mensagens concatenadas são retransmitidas utilizando a técnica de retransmissão embutida em um pacote subsequente, apresenta melhor desempenho do que

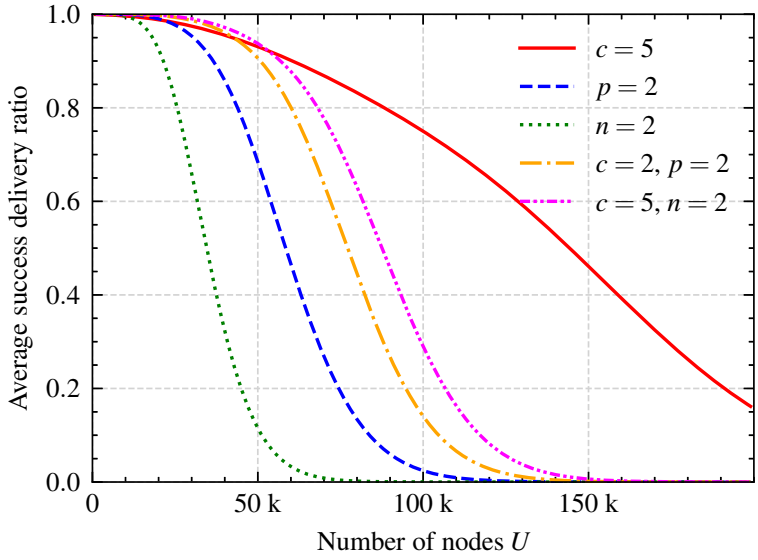


Figura 28: Probabilidade média de sucesso, considerando o melhor valor de parâmetro para cada técnica. Os resultados para combinações das três técnicas também são apresentados. Por exemplo, $(c = 2, p = 2)$ significa que duas mensagens são concatenadas em um único pacote e que as mensagens concatenadas são retransmitidas utilizando a técnica de retransmissão embutida em um pacote subsequente.

o caso de apenas retransmissão embutida com $p = 2$. Notavelmente, a replicação combinada com concatenação supera a retransmissão embutida combinada com concatenação, pois permite maiores valores de c sem exceder o comprimento máximo do *payload*.

11.3 Conclusão

Este trabalho apresentou conceitualmente três técnicas de transmissão de mensagens para LR-FHSS, uma tecnologia empregada, por exemplo, em cenários de DtS-IoT. Os próximos passos incluem as seguintes tarefas:

- Finalização da avaliação analítica do desempenho das técnicas investigadas, em termos de *goodput*, atraso e consumo energético. Essa avaliação encontra-se em estágio bastante avançado, com validação em andamento.
- Simulações adicionais do desempenho das técnicas, de modo a corroborar o desenvolvimento analítico.

12 Escalabilidade e Alocação de Recursos em Cell-free Massive MIMO usando múltiplas CPUs

Jonathan N. Gois
UFRJ

12.1 Introdução

O avanço contínuo da tecnologia tem impulsionado os sistemas de comunicação a oferecerem serviços cada vez mais eficientes, confiáveis e de alta capacidade. Em particular, a crescente demanda por taxas de dados elevadas é um dos principais critérios na avaliação e evolução das tecnologias móveis modernas. Esse aumento na demanda reflete diretamente no desenvolvimento de arquiteturas de rede mais sofisticadas, capazes de atender às exigências de cobertura e qualidade de serviço em ambientes densamente povoados. Nesse cenário, o uso de múltiplas antenas se destaca como uma abordagem promissora, oferecendo diversas vantagens tanto no nível físico quanto no nível de rede.

Um dos pilares fundamentais para essa evolução tecnológica é o uso de sistemas Múltiplas Entradas e Múltiplas Saídas (MIMO, do inglês *Multiple Input and Multiple Output*). Quando múltiplas antenas são utilizadas simultaneamente para transmitir sinais a um conjunto de receptores, surgem benefícios relevantes como melhoria na direcionalidade do feixe, aumento na taxa de transmissão, redução no consumo de energia e maior robustez contra interferências [116]. Essas vantagens podem ser ainda mais significativas quando o número de antenas transmissoras excede, em grande escala, o número de equipamentos de usuário (UEs, do inglês *User Equipment*). O conceito de MIMO massivo (mMIMO) surge nesse contexto como uma solução que supera a barreira de escalabilidade, apoiando-se em três premissas principais: (i) apenas a estação base estima o canal de *downlink*; (ii) o número de antenas é várias vezes maior que o número de UEs; e (iii) utiliza-se processamento linear simples na etapa de transmissão. Essas premissas permitem manter a eficiência do sistema mesmo com um grande número de usuários ativos, simplificando o design de hardware e reduzindo os requisitos de sinalização.

Apesar dos benefícios proporcionados pelo mMIMO, os recursos espectrais disponíveis não acompanham o mesmo ritmo de crescimento da demanda. Uma das alternativas para expandir a largura de banda é elevar a frequência da portadora. No entanto, essa abordagem impõe desafios significativos de propagação, como maior atenuação e menor penetração em ambientes internos, o que limita a cobertura efetiva. Para enfrentar esses obstáculos, surge como alternativa o MIMO Massivo Livre de Células (CF-mMIMO, do inglês *Cell-Free Massive MIMO*), uma arquitetura que substitui a estrutura tradicional baseada em células por uma rede cooperativa composta por vários Pontos de Acesso (APs) distribuídos geograficamente. Nesse modelo, não há atribuição fixa de UEs a células, o que permite uma distribuição mais uniforme do sinal e uma mitigação mais eficaz de interferências.

Embora o conceito de CF-mMIMO apresente diversas vantagens teóricas, a sua implementação prática ainda enfrenta importantes desafios. Um deles diz respeito à necessidade de múltiplas Unidades Centrais de Processamento (CPUs) operando de forma coordenada para gerenciar os APs distribuídos. Uma arquitetura centralizada, na qual uma única CPU controla todos os APs, pode facilitar a coordenação, mas tende a sofrer com problemas de escalabilidade, alta latência e vulnerabilidade a falhas. Por outro lado, o uso de múltiplas CPUs, distribuídas na rede, permite aliviar os gargalos de tráfego de dados e reduzir a sobrecarga de processamento. No entanto, essa abordagem exige estratégias robustas de cooperação entre CPUs para

garantir a consistência no controle dos recursos, na alocação de potência e na sincronização entre as entidades distribuídas. A complexidade dessas interações cresce consideravelmente com o número de UEs e a mobilidade dos usuários, exigindo soluções eficientes para preservar a estabilidade do sistema.

Diversos autores têm proposto modelos hierárquicos para estruturar os níveis de colaboração entre CPUs em ambientes CF-mMIMO. Kim et al. [117] definem quatro níveis distintos de cooperação. No nível mais básico, Nível 1, cada AP opera de maneira totalmente independente, realizando localmente a estimativa de canal e o controle de potência. No Nível 2, a CPU gerencia grupos disjuntos de APs, centralizando o processamento dentro de cada grupo. O Nível 3 introduz uma troca limitada de informações ambientais entre CPUs, possibilitando uma cooperação parcial. Já no Nível 4, ocorre uma troca completa de informações, inclusive do estado instantâneo do canal (CSI), entre todas as CPUs, o que impõe uma demanda significativa de *backhaul*. De modo complementar, Li et al. [118] propuseram um modelo hierárquico baseado na quantidade e tipo de informação trocada entre CPUs. Os níveis mais baixos operam sem compartilhamento de informações, enquanto os mais altos requerem troca de CSI estatístico ou instantâneo, com diferentes graus de centralização computacional.

A escalabilidade permanece sendo uma preocupação central no CF-mMIMO, especialmente em cenários de alta densidade de usuários. Assim, compreender os obstáculos à colaboração entre unidades de controle é fundamental para viabilizar aplicações práticas dessa arquitetura. Um avanço relevante nesse sentido é o uso da técnica de Agrupamento com Cooperação Dinâmica (DCC, do inglês *Dynamic Cooperation Clustering*). Em um cenário ideal, cada UE teria acesso simultâneo a todos os APs da rede, maximizando a eficiência espectral. Entretanto, esse modelo é inviável do ponto de vista computacional e de sinalização. O DCC explora o fato de que nem todos os APs contribuem igualmente para a qualidade do enlace com um determinado UE. Assim, a ideia central é restringir o número de conexões, concentrando o atendimento de cada usuário em um subconjunto mais relevante de APs, o que permite reduzir sobrecarga e manter alto desempenho. Essa estratégia também facilita o gerenciamento de recursos e a aplicação de políticas de alocação de potência localizadas.

A adoção dessa técnica, contudo, traz consigo um custo computacional considerável. O CSI precisa ser atualizado constantemente, e os agrupamentos de APs devem ser reavaliados de forma dinâmica. Esse processo se torna ainda mais complexo à medida que a quantidade de UEs e APs aumenta, acarretando uma escalada na carga de processamento associada à formação de *clusters* e à alocação de recursos. Esse desafio é agravado em ambientes com canais que variam rapidamente no tempo, onde são necessárias reconfigurações frequentes para manter o desempenho estável.

Diante desse contexto, a atual pesquisa propõe o desenvolvimento de ferramentas para o aumento da escalabilidade de todo sistema CF-mMIMO e da alocação de recursos. Como resultado parcial do trabalho, será apresentado aqui um algoritmo de baixa complexidade computacional que visa mitigar os custos computacionais do processo de DCC sem grandes prejuízos na eficiência espectral de todo o sistema. A abordagem comum, baseada apenas na ordenação dos coeficientes de larga escala avaliados no CSI, pode ser insuficiente quando realizada de modo paralelo em várias CPUs de forma não cooperativa. A proposta é utilizar algoritmo de aprendizado não supervisionado para a formação dos agrupamentos virtuais de APs alinhados às fronteiras de controle das CPUs, mesmo na ausência de colaboração entre elas. Essa estratégia visa a centralização do processamento do DCC em uma única CPU por UE, mantendo a consistência e reduzindo o tráfego de sinalização.

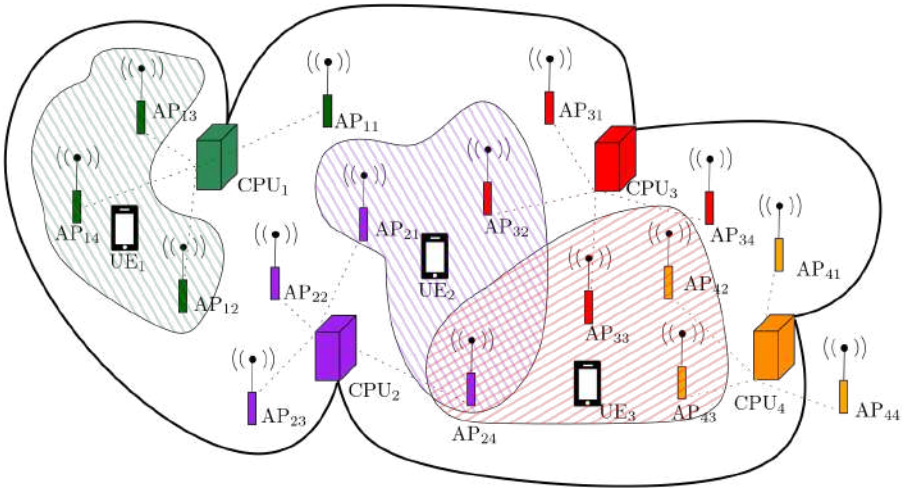


Figura 29: CF-mMIMO em um ambiente com múltiplas CPUs. Esta figura ilustra o agrupamento virtual formado pela seleção de APs por cada UE e o serviço compartilhado fornecido por diferentes APs controlados por diferentes CPUs.

12.2 Modelo do Sistema

Como no uso tradicional do CF-mMIMO, consideramos todo o esquema operando no modo duplex por divisão no tempo (TDD) e utilizando pilotos ortogonais para a estimação de canal de todos os UEs na transmissão em *downlink*. Aqui, assumimos que as CPUs não compartilham nenhuma informação e que cada CPU realiza sua própria estimação de CSI. Essa hipótese representa um caso prático onde não se pode depender de um *backhaul* com alta capacidade e baixa latência. Outra consideração é a possibilidade de que um dado UE selecione APs controlados por mais de uma CPU para servi-lo.

O cenário estudado consiste em K UEs com uma única antena, L APs, cada um equipado com M antenas, e Q CPUs. Usando a Figura 29 como referência, definimos o *cluster real* como o conjunto de índices de APs que estabelecem uma conexão direta e sem erros com uma determinada CPU, denotado como $\mathcal{R}_q$, onde q é o índice da CPU. As linhas tracejadas na figura ilustram os enlaces conectando cada AP do *cluster* real à CPU correspondente. Do ponto de vista do k-ésimo UE, o conjunto de CPUs que controlam pelo menos um AP que o atende é denotado por $\mathcal{Q}_k$.

Em contraste, o *cluster virtual* é o subconjunto de índices de APs que um dado UE seleciona para conexão. Esse subconjunto é representado como $\mathcal{V}_{q,k}$, onde k denota o índice do UE. O motivo da indexação por q é que a CPU q é responsável pela alocação de recursos do UE k. Na Figura 29, as áreas hachuradas representam os *clusters* virtuais de cada CPU, diferenciados por cor. O conjunto $\mathcal{V}_{q,k}$ será um dos alvos desta pesquisa. Por ora, assume-se que ele é conhecido.

Para estimar o canal, consideramos que cada CPU estima o canal de *downlink* entre os APs e os UEs pertencentes ao seu *cluster* real. Assim, o canal real da m-ésima antena do AP l até o UE k é dado por

$$\mathbf{g}_{l,m,k} = \sqrt{\beta_{l,m,k}}\mathbf{h}_{l,m,k} \ \in \mathbb{C}^{M\times 1}, \tag{161}$$

em que $\beta_{l,m,k}$ é o coeficiente de desvanecimento em larga escala (LSFC) e $\mathbf{h}_{l,m,k}$ representa o desvanecimento em pequena escala. Neste trabalho, consideramos um canal Rayleigh correlacionado e uma distribuição Gaussiana dos ângulos de chegada dentro de um modelo de espalhamento local [119].

Durante o processo de estimação de canal, cada UE k envia ao AP l um piloto $\boldsymbol{\varphi}_k \in \mathbb{C}^{\tau_p \times 1}$ que, neste trabalho, assume-se ortogonal aos demais pilotos ($\boldsymbol{\varphi}_k^{\mathrm{H}} \cdot \boldsymbol{\varphi}_{k'} = 0$, para $k' \neq k$). Para garantir a ortogonalidade, o comprimento da sequência piloto τ_p deve ser maior ou igual a K.

Para construir a estimação de canal a partir dos sinais recebidos, é necessário considerar o *cluster* virtual previamente mencionado [117]:

$$\hat{\mathbf{g}}_{l,m,k} = (\tau_p \beta_{l,m,k}) \boldsymbol{\rho} \left(\sum_{k' \in \mathcal{V}_{q,k'}} \mathbf{g}_{l,k'} \boldsymbol{\varphi}_{k'}^{\mathrm{H}} + \mathbf{Y} + \mathbf{N} \right) \boldsymbol{\varphi}_k. \tag{162}$$

Na expressão acima, o valor $\boldsymbol{\rho}$ representa a potência normalizada de transmissão dos pilotos, a matriz $\mathbf{Y}$ denota a interferência proveniente de UEs fora do *cluster* virtual, e $\mathbf{N}$ é uma matriz de ruído.

No *downlink*, o sinal de dados transmitido por cada AP l é dado por:

$$\mathbf{x}_l = \sum_{k \in \mathcal{V}_{q,k}} \sqrt{\rho_{l,k}} \mathbf{w}_{l,k}^{\mathrm{H}} q_k, \tag{163}$$

em que $\rho_{l,k}$ é a potência normalizada de transmissão para o UE k, $\mathbf{w}_{l,k}$ é o vetor de pré-codificação e q_k é o símbolo de dado transmitido.

O sinal recebido por cada UE é expresso como:

$$y_k = \mathrm{DS}_k q_k + \mathrm{BU}_k q_k + \mathrm{UI}_{k,k'} q_{k'}, \tag{164}$$

conforme apresentado por Kim et al. [117].

Na equação acima, o sinal desejado para o k-ésimo UE é representado por DS_k. A incerteza no ganho de *beamforming* (BU_k) e a interferência causada pelo k'-ésimo UE ($\mathrm{UI}_{k,k'}$) degradam o sinal recebido correspondente ao símbolo transmitido q_k. Esses efeitos ocorrem principalmente devido à natureza distribuída da arquitetura e às imperfeições na estimação de canal.

Cada componente do sinal y_k é descrito como:

$$\mathrm{DS}_k \triangleq \sum_{q \in \mathcal{Q}_k} \sum_{q \in \mathcal{V}_k} \sqrt{\rho_{l,m,k}} \mathbb{E}\{\mathbf{g}_{l,m,k}^{\mathrm{H}} \mathbf{w}_{l,m,k}\}, \tag{165}$$

$$\mathrm{BU}_k \triangleq \sum_{q \in \mathcal{Q}_k} \sum_{q \in \mathcal{V}_k} \sqrt{\rho_{l,m,k}} (\mathbf{g}_{l,m,k}^{\mathrm{H}} \mathbf{w}_{l,m,k} - \mathbb{E}\{\mathbf{g}_{l,m,k}^{\mathrm{H}} \mathbf{w}_{l,m,k}\}), \tag{166}$$

$$\mathrm{UI}_{k,k'} \triangleq \sum_{q \in \mathcal{Q}_{k'}} \sum_{q \in \mathcal{V}'_k} \sqrt{\rho_{l,m,k}} \mathbb{E}\{\mathbf{g}_{l,m,k}^{\mathrm{H}} \mathbf{w}_{l,m,k'}\}. \tag{167}$$

Utilizando as definições anteriores, a razão sinal-interferência-mais-ruído (SINR) para o k-ésimo UE é dada por:

$$\mathrm{SINR}_k = \frac{|\mathrm{DS}_k|^2}{\mathbb{E}[|\mathrm{BU}_k|^2] + \sum\limits_{\substack{k'=1 \\ k' \neq k}}^{K} \mathbb{E}[|\mathrm{UI}_{kk'}|^2] + 1}. \tag{168}$$

Neste trabalho, aplicamos a pré-codificação por máxima razão (MR), definida como:

$$\mathbf{w}_{l,k} = \frac{\hat{\mathbf{g}}_{l,k}}{\sqrt{\mathbb{E}\left\{ \|\hat{\mathbf{g}}_{l,k}\|^2 \right\}}}. \tag{169}$$

Esta pesquisa concentra-se na etapa de DCC com o objetivo de melhorar a seleção de APs para cada UE. Cada UE pode ser atendido por múltiplas CPUs, mas cada AP é controlado por apenas uma CPU. Considerando que o UE pode se mover por toda a área, ele não deve estar associado a uma única CPU fixa. No entanto, para calcular a alocação de potência, é necessário que uma (ou mais) CPUs sejam responsáveis por esse processo.

12.3 Método Proposto

Inicialmente, é possível definir diversas estratégias para a atribuição de APs às CPUs. Para a q-ésima CPU, o conjunto correspondente de APs é denotado como $\mathcal{R}_q$. Esse conjunto pode ser determinado com base em diferentes critérios, tais como qualidade do sinal, balanceamento de carga ou restrições de conectividade. Neste trabalho, adotamos um critério baseado na distância, em que cada CPU seleciona seus APs associados com base na proximidade geográfica.

Para alcançar uma distribuição equilibrada de APs entre as CPUs, implementamos uma estratégia de seleção sequencial: a cada rodada, uma CPU seleciona o AP disponível mais próximo, seguindo-se um esquema em turnos (round-robin) para as CPUs subsequentes. Essa abordagem modela um cenário generalizado em que qualquer AP pode ser associado a qualquer CPU, independentemente de limitações físicas.

Entretanto, em implantações práticas, essa suposição pode não ser válida. Os APs são normalmente conectados às CPUs por meio de enlaces cabeados, e a associação está limitada às conexões fisicamente viáveis. Portanto, esquemas de associação realistas devem levar em conta as restrições de infraestrutura, como topologia de rede e alcance de conectividade, para garantir a viabilidade da associação proposta.

Diversos trabalhos na literatura abordaram o DCC sob restrições específicas. Por exemplo, em [120], os autores propõem uma abordagem em que cada UE k inicialmente seleciona um AP mestre (*Master* AP) com base no coeficiente de desvanecimento em larga escala (LSFC) mais forte, $\beta_{l,k}$, que representa a melhor condição de canal. Esse mesmo AP mestre também é responsável por alocar o piloto ao UE, de modo a minimizar a contaminação de pilotos.

A regra para associar APs adicionais ao conjunto de atendimento do UE considera se o AP não está atualmente servindo outro UE com o mesmo piloto. Alternativamente, se o AP já estiver associado a um UE com aquele piloto, mas o novo UE apresentar um canal mais forte, o AP troca sua associação para atender o novo UE. Sob essa solução, podemos limitar o número de APs em cada *cluster*, porém não há garantia de que todos os *clusters* terão a mesma quantidade de APs.

Outros métodos vêm sendo explorados na literatura, como aprendizado por reforço com otimização de baleias [121], aprendizado por reforço clássico [122] e aprendizado profundo [123], geralmente considerando apenas uma CPU.

Ao utilizar mais de uma CPU e considerando um *backhaul* restrito, o problema torna-se mais desafiador. Primeiramente, a alocação de pilotos não pode ser garantida conforme descrito nos modelos escaláveis tradicionais do CF-mMIMO. É necessária certa coordenação para garantir a estimação de canal e a alocação de potência; portanto, o uso de múltiplas CPUs exige o compartilhamento de algumas informações auxiliares.

Aqui, consideramos que o número de UEs, K, é igual ao número de pilotos disponíveis, operando, portanto, em um ambiente livre de contaminação de pilotos. No entanto, a alocação de potência para cada UE é baseada exclusivamente no *cluster* real ao qual o UE pertence. A CPU responsável pelo UE k determina a alocação de potência necessária. A política de alocação de potência deve ser a mesma para todas as CPUs, mas apenas uma CPU calcula o

valor da potência para cada UE em seu *cluster*. Essa é uma característica importante, e os valores computados ainda precisam ser compartilhados.

Entretanto, sob a mesma política de alocação de potência, a solução calculada por apenas uma CPU pode não ser ótima. Para ilustrar a gravidade desse problema, considere o caso em que cada CPU é responsável por apenas um UE — nesse caso, ela alocaria toda a potência disponível a esse único usuário. Se todas as CPUs agirem da mesma forma, o total de potência alocada para todos os UEs seria igual, mas não necessariamente atenderia a critérios de justiça como o *max-min fairness* (MMPFA) ou outras políticas globais.

Considerando o cenário com coordenação mínima entre CPUs, é preciso otimizar o *cluster* virtual $\mathcal{V}_{q,k}$ para alcançar um resultado global melhor. Utilizar apenas o melhor canal com base no LSFC pode não ser ideal, pois o canal mais forte entre o AP l e o UE k pode pertencer a um AP localizado no domínio de outra CPU. Em outras palavras, o melhor canal não implica necessariamente o melhor enlace.

Uma abordagem alternativa é considerar a distância entre o AP e o UE. Essa solução é plausível, pois os APs podem ser associados a cada UE com base na proximidade. No entanto, o AP mais próximo nem sempre resulta na melhor qualidade de canal, o que pode penalizar a taxa de dados do UE e a capacidade geral do ambiente.

Propomos[6] o uso de uma característica que combina distância e resposta de canal para criar o *cluster* virtual. Define-se um vetor $\mathbf{s}_{l,k}$ como resultado da ponderação entre $\beta_{l,k}$ e a distância $d_{l,k}$ entre o AP l e o UE k:

$$\mathbf{s}_{l,k} = ((1-\alpha)\beta_{l,k},\ \ \alpha d_{l,k},\ \alpha(1-\alpha)\beta_{l,k}d_{l,k})\,, \tag{170}$$

em que α é o parâmetro que define a importância relativa entre distância e LSFC. O vetor $\mathbf{s}_{l,k}$ fornece uma representação contínua entre os métodos baseados em LSFC e os baseados em distância, por meio do ajuste do parâmetro $\alpha \in [0, 1]$, permitindo assim reconstruir estratégias de clusterização tradicionais.

No entanto, esse novo vetor $\mathbf{s}_{l,k}$ não suporta, por si só, uma ordenação direta com base nas métricas de LSFC ou distância. Para superar essa limitação, emprega-se o algoritmo de clusterização *K-means* [124]. Como técnica de aprendizado não supervisionado, ele permite o agrupamento dos usuários com base na estrutura subjacente de $\mathbf{s}_{l,k}$, possibilitando a recuperação de características espaciais que podem não ser observáveis por simples critérios de ordenação. O uso do *K-means* torna o processo de formação de *clusters* mais robusto a variações espaciais complexas. O terceiro componente do vetor evidencia padrões que não poderiam ser identificados na representação bidimensional baseada apenas em distância e LSFC.

12.4 Resultados Parciais

Realizamos simulações para avaliar o desempenho do método proposto. O ambiente simulado abrange uma área quadrada com lados de 2 quilômetros, na qual CPUs, APs e UEs são distribuídos. Para aumentar o realismo do cenário, é considerada uma topologia *wraparound*, que fornece condições de contorno periódicas. Nesse espaço, são posicionadas quatro CPUs, 400 APs equipados com quatro antenas cada e 20 UEs com antena única. A avaliação de desempenho foi conduzida ao longo de 50 cenários independentes, cada um repetido 20 vezes, a fim de capturar a variabilidade estatística dos resultados.

[6]Esse método é apresentado em artigo submetido ao XLIII Simpósio Brasileiro de Telecomunicações (SBrT 2025), que será realizado entre os dias 29 de setembro e 03 de outubro de 2025.

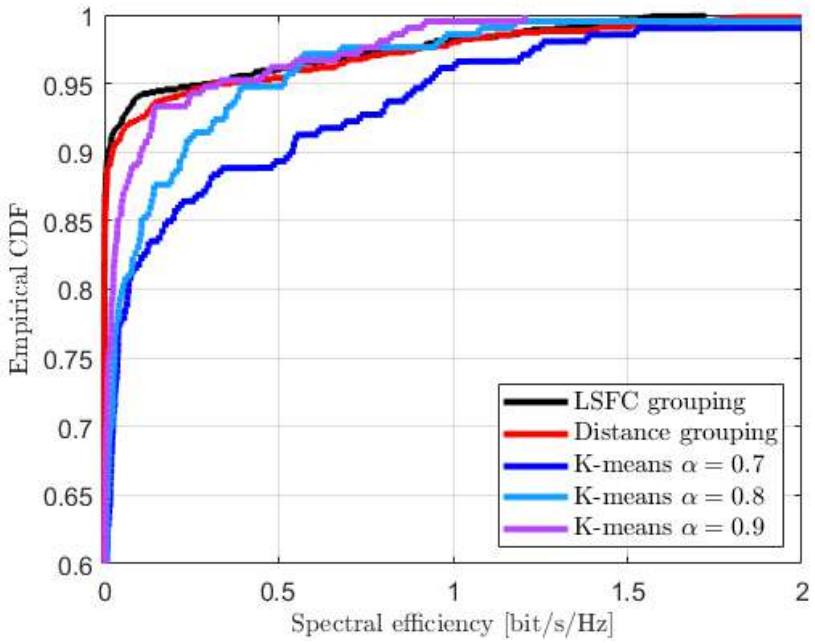


Figura 30: Função de distribuição acumulada empírica (CDF) da eficiência espectral para diferentes métodos de DCC no ambiente simulado. Entre as configurações testadas, o melhor desempenho foi alcançado com $\alpha = 0{,}7$.

As quatro CPUs são posicionadas nos centros de cada quadrante da área de simulação. Tanto os APs quanto os UEs são distribuídos de forma uniforme e aleatória em todo o espaço. Para introduzir estrutura na atribuição de APs às CPUs, é adotado um procedimento de alocação sequencial: cada CPU, em sua vez, seleciona o melhor AP disponível com base em um critério pré-definido; em seguida, a próxima CPU realiza sua escolha. Esse processo é repetido até que todos os APs estejam alocados.

A tarefa de cálculo dos recursos para cada UE é simplificada ao considerar a distância entre o UE e a CPU. Essa estratégia está alinhada com a topologia baseada em distância ou no coeficiente de desvanecimento em larga escala (LSFC) implementada no sistema. A potência de transmissão por UE é fixada em aproximadamente 100 mW, equivalente a 20 dBm no *downlink*. O modelo de canal utilizado nesta simulação segue a definição de Björnson e Sanguinetti [120], considerando perda de percurso geométrica, sombreamento e correlação espacial do canal.

No método proposto, utilizamos quatro grupos para clusterizar os vetores de características associados aos enlaces entre APs e UEs. No grupo que contém o "melhor" AP — definido como o AP mestre em [120] — selecionamos cinco APs para compor o cluster virtual de cada UE. Em nossos experimentos, variamos o parâmetro α dentro do conjunto $\{0{,}7, 0{,}8, 0{,}9\}$ para produzir os resultados apresentados na Figura 30. Adicionalmente, conduzimos experimentos de menor escala com α variando de 0,50 a 0,99, com o objetivo de identificar valores apropriados para as simulações principais.

A Figura 30 apresenta a função de distribuição acumulada empírica (CDF) da eficiência espectral para diferentes estratégias de DCC. O método proposto apresenta desempenho superior em comparação com abordagens tradicionais baseadas apenas em distância ou LSFC, alcançando maior eficiência espectral média entre os usuários.

A curva azul, correspondente ao caso $\alpha = 0{,}7$, proporciona um ganho marginal em eficiência

espectral em relação aos demais valores de α testados. Comparando os métodos com o limite superior definido pela máxima taxa de dados observada nas simulações, a abordagem tradicional atinge 1,7259 bits/s/Hz, a seleção baseada em distância alcança 2,2363 bits/s/Hz e o método proposto atinge o valor de 2,3376 bits/s/Hz.

12.5 Conclusão

Este trabalho abordou o uso da técnica de Agrupamento com Cooperação Dinâmica (DCC) em sistemas CF-mMIMO com múltiplas CPUs. O método proposto realiza a formação eficiente de *clusters* de APs para atendimento aos UEs por meio de uma abordagem de aprendizado não supervisionado, reduzindo a necessidade de comunicação entre CPUs. Os resultados de simulação indicam que a estratégia proposta proporciona ganhos em eficiência espectral em comparação com abordagens convencionais, especialmente quando $\alpha = 0{,}7$. Tais ganhos foram observados mesmo em cenários com coordenação limitada entre as unidades de processamento.

Esses resultados parciais sugerem que *trade-offs* favoráveis entre desempenho, complexidade computacional e escalabilidade podem ser alcançados, o que torna o método promissor para aplicações práticas em implantações de larga escala de CF-mMIMO. Vale destacar que a simplicidade da solução é compatível com ambientes reais onde o *backhaul* é limitado e a troca de informações entre CPUs deve ser mínima.

Como trabalho futuro, pretende-se expandir o número de cenários simulados e investigar diferentes políticas de alocação de potência entre as CPUs. Além disso, será avaliado o impacto da mobilidade dos UEs e da presença de bloqueios no ambiente, buscando tornar a avaliação mais próxima de cenários operacionais reais. O caso analisado neste relatório representa um cenário de pior caso, em que as CPUs operam com cooperação mínima, o que aumenta consideravelmente a complexidade do problema e serve como referência para validar a robustez do algoritmo proposto.

Na etapa seguinte da pesquisa, espera-se refinar os modelos e validar os resultados por meio de métricas adicionais, consolidando a viabilidade da proposta.

13 Algoritmos de AI/ML para Seleção de Feixe na Transmissão em mmWaves

Joanna Manjarres
UFRJ

13.1 Introdução

O estabelecimento de enlaces em sistemas de ondas milimétricas (mmWave) é amplamente explorado e investigado na literatura devido aos desafios associados, como atenuação de sinal e perdas por penetração. Técnicas como *beamforming* desempenham um papel crucial ao direcionar o sinal para o receptor, maximizando a energia de sinais de radiofrequência (RF) recebida e garantindo a eficiência no uso do canal de comunicação.

A seleção de feixes *(beam selection)* permite que transmissores (Tx) e receptores (Rx) explorem o espaço de feixes, uniformemente particionado em setores discretos, até convergir para o par ideal. Métodos tradicionais utilizam sinais de RF em banda ou fora de banda, onde sinais pilotos são testados em feixes largos antes de restringir a busca aos feixes contidos nesses setores iniciais.

Nos últimos anos, tem-se intensificado o interesse por abordagens que integram dados contextuais — como informações de posicionamento (GPS) e sensoriamento ambiental (LiDAR) — ao processo de seleção de feixe, viabilizando a aplicação de algoritmos de aprendizado de máquina (ML), incluindo redes convolucionais (CNNs) e redes neurais profundas (DNNs). Neste contexto, propõe-se uma alternativa inovadora: o uso de Redes Neurais Sem Peso (RNSP), especificamente a rede WiSARD, como uma solução leve, eficiente e compatível com restrições operacionais de sistemas veiculares e de borda.

Os experimentos desenvolvidos na Atividade 3.2 do Projeto Brasil 6G - Fase III demonstraram que a rede WiSARD, ao utilizar dados LiDAR e coordenadas espaciais como entrada, apresentou desempenho superior a modelos clássicos da literatura (Salehi Batool [125] e Ruseckas[126]) em cenários desafiadores, como no caso da acurácia top-1, com ganhos de até 17,49%. Além disso, a técnica mostrou-se extremamente eficiente em termos computacionais, sendo cerca de 99,78% mais rápida no tempo de treinamento, caracterizando-se como uma solução altamente viável para aplicações em tempo real.

Com base nesses resultados promissores, esta Atividade 2.2 tem como foco principal avaliar a robustez e capacidade de generalização dos modelos de ML para seleção de feixe, em cenários representativos da rede de acesso. A partir de variações no ambiente de propagação — como condições de linha de visada (LOS) e não linha de visada (NLOS) — e da troca de domínios entre conjuntos de dados, esta etapa busca verificar a estabilidade dos modelos, identificar limitações e apoiar decisões de projeto para futuras redes 6G.

13.2 Fundamentação Teórica

A chegada da quinta geração (5G) trouxe requisitos mais elevados, como altas taxas de transferência e maior capacidade em cenários densos. Embora muitas implementações iniciais tenham sido feitas utilizando bandas abaixo de 6 GHz, o uso de frequências mais altas, acima de 24 ou 60 GHz, tem sido considerado como um complemento. Essas bandas oferecem acesso a um maior espectro, reduzindo a carga sobre as frequências mais baixas. A tecnologia mmWave tem-se tornado um componente-chave que permite velocidades de dados mais rápidas e com

maior conectividade; trata-se do uso de bandas no espectro de frequências que vão de 30 GHz a 300 GHz. Esta parte do espectro é conhecida pelos seus comprimentos de onda curtos, ou ondas milimétricas.

As ondas milimétricas são mais suscetíveis à interferência de obstáculos e enfrentam maior perda de propagação, o que exige uma rede mais densa de estações base para garantir uma cobertura consistente e feixes focados para compensar essa atenuação. Ao focar o sinal em dispositivos específicos, a formação de feixes reduz a interferência e melhora a eficiência. Apesar dos desafios colocados pelas barreiras físicas, a operação do mmWave é otimizada através de tecnologias avançadas e projeto estratégico de rede.

O uso de múltiplas antenas tanto no transmissor quanto no receptor (MIMO) melhora a capacidade, cobertura, taxa de dados e confiabilidade da rede. Essa técnica oferece diversidade contra desvanecimento e possibilita o aumento da relação sinal-ruído (SNR), permitindo a combinação construtiva dos sinais transmitidos.

Esta combinação construtiva pode ser alcançada ao ajustar a fase e amplitude do sinal de cada elemento da antena, o que possibilita a modificação do padrão de radiação de um *array* de antenas, direcionando o sinal para aumentar a potência na direção desejada (*beamforming*).

Além disso, a presença de múltiplas antenas viabiliza a multiplexação espacial, permitindo a transmissão de múltiplos fluxos simultâneos usando os mesmos recursos de tempo e frequência. A escolha da arquitetura para formação de feixe (banda base, intermediária ou híbrida) e o tipo de implementação (analógica, digital ou híbrida) são fatores essenciais para essa técnica.

Na formação de feixe analógica, o mesmo sinal é transmitido por cada antena, mas com fases ajustadas de maneira independente, direcionando o feixe na direção desejada. O processamento é realizado com deslocadores de fase analógicos em cada elemento da antena. Esse tipo de processamento, geralmente realizado por portadora, exige que os feixes para dispositivos localizados em direções diferentes sejam temporariamente separados. No receptor, o processamento multi-antena ocorre antes da conversão analógico-digital, limitando a formação de feixes a uma direção por vez. Isso reduz a capacidade de multiplexação espacial, já que a recepção de múltiplas direções ocorre em momentos distintos.

Em uma arquitetura digital, cada elemento de antena possui sua própria cadeia de RF e conversores ADC e DAC. Isso permite maior flexibilidade no processamento multi-antena, com ajustes tanto na fase quanto na potência de cada elemento. Além disso, o processamento independente de múltiplas antenas possibilita a transmissão simultânea de sinais de diferentes dispositivos, localizados em direções distintas, dentro da mesma portadora [127]. No receptor, a recepção simultânea de feixes de múltiplos sinais oriundos de diferentes direções é possível, aumentando a flexibilidade. Embora essa arquitetura favoreça a multiplexação espacial, ela envolve altos custos, maior complexidade e consumo de energia.

A arquitetura híbrida combina a formação de feixe digital e analógica, aprimorando o desempenho do *beamforming* analógico ao permitir maior quantidade de fluxos. Além disso, reduz a complexidade da arquitetura digital, pois exige um menor número de cadeias de RF em comparação ao número de antenas. Como resultado, há uma diminuição na carga computacional, no consumo de energia e na complexidade da arquitetura.

A técnica MIMO visa combater ou explorar a dispersão de múltiplos caminhos no canal de comunicação [128]. Num sistema MIMO com N_t antenas transmissoras e N_r antenas receptoras, é denominado sistema MIMO $N_t \times N_r$.

Para mitigar o desvanecimento, utiliza-se o conceito de diversidade, no qual réplicas do mesmo sinal são transmitidas de forma independente. Isso reduz o impacto do desvanecimento, pois as réplicas possuem diferentes características, como amplitude, fase, atraso, ângulos de

partida (AoD) e chegada (AoA). As formas de gerar diversidade incluem frequência, tempo, polarização e espacial. No receptor, as réplicas podem ser combinadas de diversas maneiras, como combinação seletiva, combinação de ganho igual e *maximal ratio combining* (MRC) [129, 127].

Além disso, a dispersão temporal do sinal, causada por reflexões, refrações, atenuações e difrações, resulta na chegada de múltiplas réplicas do sinal no receptor, em diferentes momentos, melhorando a confiabilidade e a capacidade do sistema.

Assim, matematicamente, um sinal s transmitido é multiplicado por um valor complexo, chamado resposta do canal e denotado por $h_i, i = 1, ...,N_r$, em que N_R representa o número de antenas de recepção.

O canal de propagação em um sistema MIMO é caracterizado pela matriz de de canal $\mathbf{H}$ $(N_R \times N_T)$,

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N_T} \\ h_{21} & h_{22} & \cdots & h_{2N_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R1} & h_{N_R2} & \cdots & h_{N_RN_T} \end{bmatrix}, \tag{171}$$

em que h_{ij} representa o modelo de canal entre a antena i no transmissor e antena j no receptor. Essa matriz, portanto, modela como os sinais transmitidos por um conjunto de N_T antenas são recebidos em um *array* de N_R antenas, capturando os efeitos de propagação e interação do sinal com o meio.

Com o intuito de maximizar o desempenho da rede, o sinal enviado também é afetado por ajustes tanto na transmissão quanto na recepção. Portanto, são aplicadas técnicas de processamento no sinal de pre-codificação ou *precoder*[7] no transmissor e combinação no *combiner*[8] no receptor. Assim, uma completa representação do sinal recebido num *array* com N_R elementos pode ser modelado por:

$$\boldsymbol{y} = \boldsymbol{w}_r^* \boldsymbol{H} \boldsymbol{s} \boldsymbol{w}_t + \boldsymbol{w}_r^* \boldsymbol{n} \tag{172}$$

em que:

$\boldsymbol{y}$ contém os sinais das antenas receptoras,

$\boldsymbol{w}_r$ é a matriz de combinação que processa o sinal no receptor (*combiner*)

$\boldsymbol{H}$ representa a matriz de canal complexa

$\boldsymbol{w}_t$ é a matriz que ajusta os sinais antes da transmissão (*precoder*)

$\boldsymbol{n}$: o ruído aditivo

$\boldsymbol{s}$: o vetor de sinais transmitidos.

Este canal de comunicação causa uma atenuação no sinal transmitido e introduz dois tipos de desfasamento: um devido à distância entre o Tx-Rx, e outro devido ao ângulo de chegada dos sinais no *array* de antenas receptoras, como representado na Figura 31.

[7] *Precoder*: Define como o sinal s é distribuído entre as antenas transmissoras para maximizar a eficiência espectral e minimizar interferências

[8] *Combiner*: Processa os sinais recebidos no receptor para combinar os sinais úteis e rejeitar interferências ou ruído.

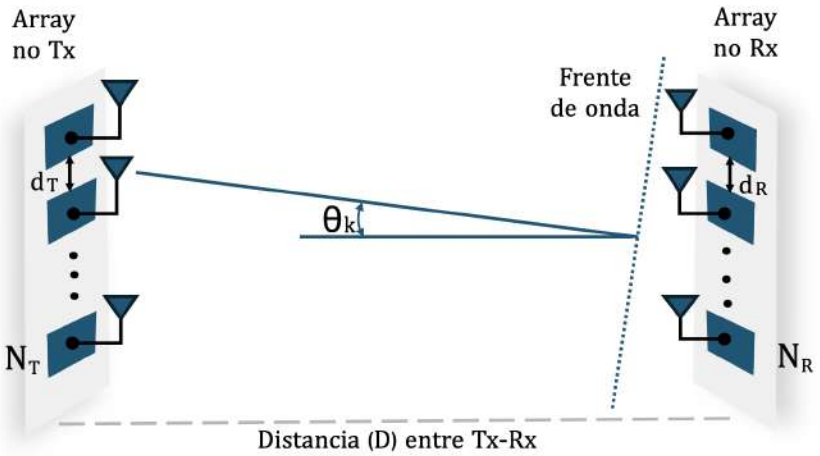


Figura 31: Representação do sinal transmitido num enlace com linha de visada direta (LoS).

Portanto, cada elemento h_{ij} da matriz de canal $\boldsymbol{H}$ representa a resposta do canal entre a j-ésima antena transmissora e a i-ésima antena receptora, calculada como:

$$h_{ij} = \alpha e^{j\phi_{ij}}, \tag{173}$$

em que:

- α é a atenuação do sinal, incluindo a perda de espaço livre dado por:

$$\alpha = \left(\frac{4\pi f D}{c}\right)^2 \tag{174}$$

  com f sendo a frequência da portadora, D a distância entre transmissor e receptor, e c a velocidade da luz.

- ϕ_{ij} que representa o deslocamento de fase, composto pelo deslocamento devido à propagação do sinal $(2\pi D/\lambda)$ e pelo deslocamento devido à fase relativa entre antenas $-2\pi(N_R - 1)sen(\theta_k)d_r/\lambda$

Com o intuito de evidenciar a influência dos vetores de direção previamente ajustados pelo *precoder* e *combiner*, a matriz na $\mathbf{H}$ pode ser escrita como:

$$\mathbf{H} = \alpha\, \mathbf{a}_R(\theta_k)\, \mathbf{a}_T^H(\theta_k), \tag{175}$$

em que $\mathbf{a}_T$ e $\mathbf{a}_R$ são os vetores de direção associados aos *arrays* de transmissão e recepção, respectivamente. Note-se, portanto, que a matriz de canal $\mathbf{H}$ é determinada pela geometria do canal (posição das antenas e ângulos de chegada). É importante mencionar que nesta geometria podem ser usados os *arrays* de antenas uniformes lineares (ULA), planares (UPA), ou circulares (UCA), onde cada uma delas afeta na modelagem da matriz do canal $\mathbf{H}$.

Se for assumido o uso de uma geometria de *arrays* linear (ULA), e também que a distância D entre os *arrays* TX e RX é muito maior que as distâncias entre os elementos dos *arrays* de transmissão e recepção, d_T e d_R, os vetores de direção são modelados como:

$\mathbf{a_T}(\theta_k)$ é o vetor de direção associado ao *array* de transmissão. Supondo N_T antenas transmissoras, um espaçamento d_T entre os elementos das antenas no *array* do TX, comprimento de onda λ e ângulo de partida do sinal (AoD) θ, $\mathbf{a_T}(\theta_k)$ é dado por:

$$\mathbf{a}_T(\theta) = \frac{1}{\sqrt{N_T}} \begin{bmatrix} 1 \\ e^{j\frac{2\pi d}{\lambda} d_T \sin(\theta)} \\ \vdots \\ e^{j\frac{2\pi d}{\lambda}(N_T-1) d_T \sin(\theta)} \end{bmatrix}. \tag{176}$$

$\mathbf{a_R}(\theta_k)$ é o vetor de direção associado ao *array* de recepção. Com N_R antenas receptoras, um espaçamento d_R entre os elementos das antenas no RX, comprimento de onda λ e ângulo de chegada do sinal (AoA) θ, $\mathbf{a_R}(\theta_k)$ é dado por:

$$\mathbf{a}_R(\theta) = \frac{1}{\sqrt{N_R}} \begin{bmatrix} 1 \\ e^{j\frac{2\pi d}{\lambda} d_R \sin(\theta)} \\ \vdots \\ e^{j\frac{2\pi d}{\lambda}(N_R-1) d_R \sin(\theta)} \end{bmatrix}. \tag{177}$$

Por outro lado, as condições de propagação dos enlaces de comunicações desempenham um papel crítico na operação destes sistemas, causando impedimentos como a atenuação, o sombreamento e o multipercurso.

Desse modo, nos modelos de propagação é geralmente assumido, propagações com linha de visada direta (LoS) ou sem linha de visada (NLoS). No cenário em que o sistema possui canais com linha de visada direta, o caminho direto entre transmissor e receptor é predominantemente desobstruído, resultando em menor atenuação do sinal, como representado pela equação 175. Isso permite uma propagação mais eficiente da energia transmitida, reduzindo a perda de potência ao longo do percurso. Além disso, os canais LoS possuem menor dispersão temporal e espectral, o que contribui para uma maior coerência do canal. Essa característica favorece um aumento na relação sinal-ruído, facilitando a decodificação precisa da informação. Em sistemas MIMO, a presença de um componente dominante (direto) também melhora a condição da matriz de canal, tornando o sistema mais eficiente para técnicas como *beamforming* e seleção de feixes. Na Figura 31, é assumido que o sinal chega como uma onda plana com um ângulo de incidência θ_k. Neste caso, o *precoder* W_T é projetado para concentrar a energia do feixe no ângulo de saída desejado (AoD) θ_k e o *combiner* W_R é configurado para maximizar o ganho de recepção no ângulo correspondente (AoA).

Para cenários sem linha de visada (NLoS), considera-se a contribuição de múltiplos trajetos devido a reflexões, difrações e espalhamento, provocados por obstáculos no ambiente de propagação. A matriz de canal $\mathbf{H}$ observada na equação (178), que modela os cenários NLoS é expressa como uma soma de componentes de múltiplos caminhos onde cada um deles possui amplitude, atraso, mudança de fase, ângulo de partida e ângulo de chegada distintos.

$$\mathbf{H} = \sum_{l=1}^{L} \alpha_l \mathbf{a}_R(\theta_{R,l}) \mathbf{a}_T^H(\theta_{T,l}), \tag{178}$$

em que:

L: número de trajetos multi-percursos;

α_l: ganho complexo associado ao l-ésimo trajeto;

$\mathbf{a}_R(\theta_{R,l})$: vetor de direção no receptor para o ângulo $\theta_{R,l}$

$\mathbf{a}_T(\theta_{T,l})$: vetor de direção no transmissor para o ângulo $\theta_{T,l}$

Como foi mencionado anteriormente, cada elemento h_{ij} da matriz $\mathbf{H}$ representa a ganho complexo do canal entre a antena j do transmissor e a antena i do receptor. Esse ganho é composto por:

- Um módulo (amplitude), que representa a atenuação do sinal,
- Um ângulo (fase), que representa o deslocamento de fase causado pela propagação.

Com o objetivo de analisar visualmente os efeitos das condições de propagação, nas Figuras 32 e 33 são apresentadas as magnitudes e fases de cada elemento h_{ij} da matriz $\mathbf{H}$. Considerando vetores de steering com base em antenas uniformemente espaçadas (ULA), onde é simulado um sistema MIMO assimétrico com 32 elementos de antenas no transmissor e 8 no receptor.

Na Figura 32, é mostrada a magnitude da matriz $\mathbf{H}$ nos dois cenários, LoS e NLoS. Através desta simulação, é possível ilustrar as diferenças estruturais em um canal sparse (LoS) versus rich-scattering (NLoS). Isto permite entender a distribuição de energia entre os pares de antenas. Onde a matriz $\mathbf{H}_{LoS}$ tende a mostrar uma estrutura altamente correlacionada (baixa variação espacial), pois há apenas um caminho dominante, lado esquerdo da Figura 32. Por outro lado, a matriz $\mathbf{H}_{NLoS}$ é mais dispersa, com ganhos aleatórios espalhados entre diferentes pares Tx/Rx, sem padrão visível, indicando múltiplos caminhos com diferentes amplitudes, lado direito da Figura 32. Embora o canal LoS seja forte em uma direção, o canal NLoS oferece maior diversidade espacial, permitindo maior capacidade e mais graus de liberdade — algo desejável para multiplexação espacial em MIMO.

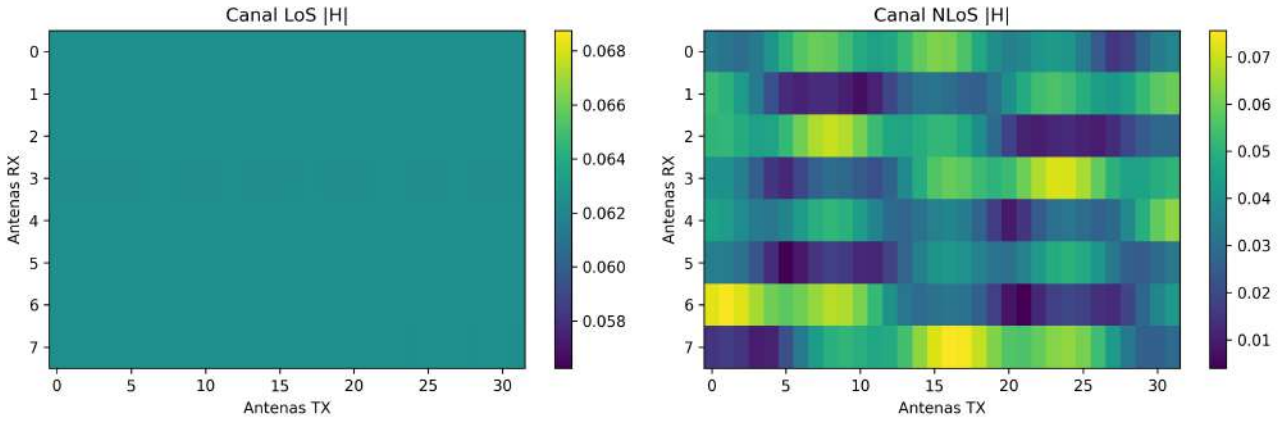


Figura 32: Comparação da magnitude da matriz de canal H (LoS vs NLoS)

Na Figura 33, é possível observar como cada ponto da matriz $\mathbf{H}(i,j)$ representa a contribuição de caminho entre a antena j no transmissor e a antena i no receptor. A fase desses elementos (representada pelas cores no gráfico) mostra como os sinais são afetados por múltiplos caminhos com diferentes direções e atrasos. No cenário LoS, a fase é mais estruturada e suave. Isso ocorre porque há um único caminho dominante entre Tx e Rx. A variação da fase segue um padrão quase linear, típico de canais com propagação direta. Em ambientes NLoS, a

fase apresenta variação mais aleatória e complexa devido à reflexão, difração e espalhamento, o que indica um canal mais dispersivo, comparado ao cenário com linha de visada (LoS).

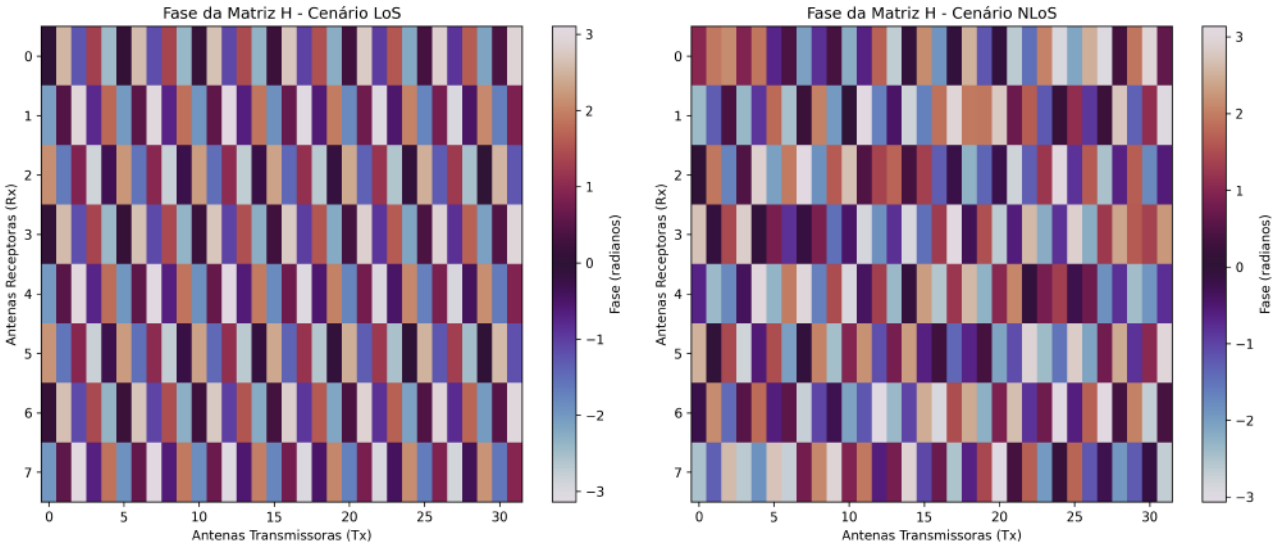


Figura 33: Mapas de fase da matriz de canal $\mathbf{H}$ para os cenários LoS e NLoS

Descrição do problema Como foi previamente descrito, o sinal recebido $\boldsymbol{y}$ é afetado por um vetor *precoder* $\boldsymbol{W}_T$ no transmissor, um vetor *combiner* $\boldsymbol{W}_R$ no receptor, uma matriz do canal $\mathbf{H}$, e ruído n, ou seja,

$$\boldsymbol{y}_{(t,r)} = \boldsymbol{w}_r^* \boldsymbol{H} \boldsymbol{s} \boldsymbol{w}_t + \boldsymbol{w}_r^* \boldsymbol{n} \tag{179}$$

Com o intuito de criar feixes altamente direcionados, o objetivo do *beamforming* é otimizar o *precoder* no transmissor e o *combiner* no receptor. A seleção do feixe escolhe um subconjunto de vetores dentre os contidos nos *codebooks* $C_T = \left\{\boldsymbol{w}_{t_1}, \ldots, \boldsymbol{w}_{t_{|c_t|}}\right\}$ e $C_R = \left\{\boldsymbol{w}_{r_1}, \ldots, \boldsymbol{w}_{r_{|c_r|}}\right\}$, em que C_T e C_R podem ser calculados a partir de técnicas baseadas em transformações matemáticas, como a Transformada discreta de Fourier (DFT) [130], Grassmannian ou Gram-Schmidt. Assim como também, os *codebooks* podem ser projetados a partir de modelos estatísticos, como os *codebook* baseados em canal ou por aprendizado de máquina.

Considerando o conjunto de todos os pares de feixes possíveis (t,r) em um único índice $i \in \{1,2,3,...,m\}$, em que $m \leq |C_T||C_R|$, o objetivo do problema é selecionar o par de feixes ótimos i que maximize o sinal recebido, isto é:

$$i = \underset{i \in \{1,2,3,\ldots,m\}}{\operatorname{argmax}} (\boldsymbol{y}_i). \tag{180}$$

Algumas técnicas propostas para resolver o problema de seleção de feixe incluem: busca exaustiva, métodos baseados em heurísticas e técnicas baseadas em aprendizado de máquina (ML). No entanto, a seleção de feixe de um único par enfrenta desafios técnicos, como alta complexidade computacional, dinâmica do ambiente, entre outros. Com o intuito de reduzir esta complexidade, um subconjunto k de feixes será selecionado com o objetivo de identificar o melhor par (t,r).

Este trabalho propõe o uso de técnicas baseadas em aprendizado de máquina para prever o subconjunto k, tal que contém o melhor par de feixes com base em informações de contexto.

13.3 Método de ML proposto

Ao longo deste trabalho, tem sido estudado o uso de redes neurais sem peso (RNSP) para o problema de seleção de feixes. De forma geral, as RNSP diferem das redes neurais tradicionais no uso de memórias RAMs como uma representação dos neurônios artificiais. O aprendizado das RNSPs depende das mudanças nos conteúdos das memórias, sendo esta a principal diferença com as redes neurais tradicionais que localizam seu aprendizado nos pesos das conexões, proporcionando uma alta velocidade de aprendizado.

Neste trabalho, foi usada a rede WiSARD. Este modelo foi introduzido em 1981 em [131] na Brunel University, e foi a primeira máquina de rede neural artificial a ser patenteada e produzida comercialmente 1984 [132]. Trata-se de um modelo de aprendizado de máquina inspirado em sistemas neuromórficos e baseado na lógica distribuída. A WiSARD é uma implementação prática de um classificador que utiliza memórias associativas distribuídas para realizar tarefas de classificação, dispensando o uso de pesos contínuos, como nas redes neurais tradicionais.

Em fases anteriores do projeto Brasil 6G, foram apresentadas diversas técnicas da literatura que propõem soluções ao problema de seleção de feixes. Contudo, até o momento não foi encontrada nenhuma referência na literatura do uso de redes neurais sem peso no contexto de seleção de feixe para redes celulares.

As redes neurais sem peso, especificamente a rede WISARD, destacam-se por sua eficiência e simplicidade em relação a outras redes neurais. Ao longo do desenvolvimento deste trabalho, foi-se descobrindo a utilidade destas características no problema de seleção de feixe. Dentre elas, foi evidenciado:

- Baixo custo computacional: A ausência de pesos contínuos elimina a necessidade de cálculos intensivos, como multiplicações de matrizes, pois o treinamento consiste apenas em atualizar as RAMs com base nos endereços observados, o que é altamente eficiente. Além disso, a inferência é baseada em uma soma discreta, o que reduz o custo computacional.

- Treinamento rápido: O aprendizado ocorre de maneira incremental e não envolve processos iterativos de ajuste de pesos, como o *backpropagation*.

- Robustez a ruídos: A estrutura distribuída permite lidar bem com ruídos nos dados de entrada e permite a generalização.

- Interpretação direta: A rede permite inspecionar diretamente os padrões armazenados na memória, facilitando a interpretabilidade, principalmente se a técnica DRASiW for usada, dado que esta extensão do modelo WiSARD tem a capacidade de armazenar as frequências de padrões vistos durante a fase de treinamento em uma estrutura de dados interna, chamada "imagem mental" (MI).

13.4 Análise do *Dataset*

Com o intuito de validar o desempenho do método de ML proposto para seleção de feixe, foi usado o *dataset* Raymobtime[9], amplamente reconhecido na literatura por sua aplicação em diversas técnicas. Este *dataset* foi desenvolvido com o objetivo de criar simulações realistas de sistemas de comunicação em redes veiculares através da técnica traçado de raios ou *ray-tracing*. O *ray-tracing* é uma ferramenta computacional usada para modelar a propagação de ondas

[9]https://www.lasse.ufpa.br/raymobtime/

eletromagnéticas na faixa de ondas milimétricas. Esta técnica simula como as ondas de rádio interagem com o ambiente, considerando efeitos como reflexão, difração e espalhamento em superfícies de edifícios, veículos e outros obstáculos.

Por fim, este conjunto de dados é composto por informações de traçado de raios, nuvem de pontos providas por sensores LiDAR, imagens de câmeras e informações de posição (x,y,z) dos receptores [133]. A Figura 34 mostra um diagrama de blocos da composição do Raymobtime.

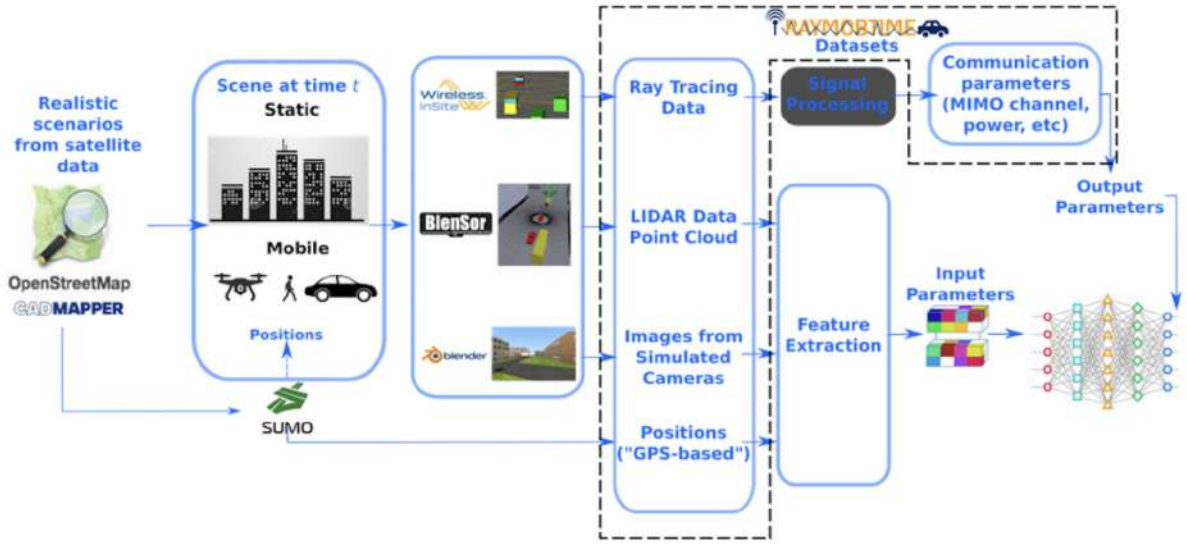


Figura 34: Diagrama de blocos do *dataset* Raymobtime tomado de [2].

O *dataset* Raymobtime contém nove cenários distintos, identificados como s000 a s009. Para este trabalho, foram selecionados os subconjuntos s008 e s009, nos quais os dados foram coletados em um ambiente que simula uma rua tipo cânion, localizada em Rosslyn, Virginia. Por tratar-se de uma simulação realista, nestes subconjuntos se encontram situações de comunicação dos tipos LoS e NLoS entre um par TX-RX. Deve ser lembrado que o tipo de conexão NLoS pode aumentar o grau de dificuldade da seleção do índice do feixe por se tratar de um enlace de comunicação composto por múltiplos trajetos.

Na Tabela 9 estão resumidas as estatísticas gerais dos *datasets* em relação às amostras LoS e NLoS.

Dataset	**# Amostras**	**Canais LOS**	**Canais NLOS**
s008	11194	6482	4712
s009	9638	1473	8165

Tabela 9: Estatísticas dos *dataset* s008 e s009.

Apesar de estes subconjuntos serem compostos pelos dados multimodais LiDAR, imagens de câmeras, dados de posição, e traçado de raios, para o desenvolvimento deste trabalho não foram usados os dados das imagens das câmeras.

13.5 Metodologia proposta

A avaliação do desempenho da rede WiSARD nesta etapa foi organizada em duas frentes complementares: (i) análise por condição de visada (LOS/NLOS) e (ii) análise de generalização entre cenários distintos. Ambas foram conduzidas utilizando os subconjuntos s008 e s009 do *dataset* Raymobtime, previamente utilizados na Atividade 3.2.

a) Separação do *Dataset* por Condição de Propagação (LOS/NLOS)

Com o intuito de isolar os efeitos das condições de visada direta (LOS) e não visada (NLOS), os subconjuntos s008 e s009 foram particionados em dois grupos distintos com base nas etiquetas fornecidas pelo *dataset*, as quais indicam o tipo de canal para cada enlace. As etapas incluíram:

- Extração de amostras LOS e NLOS: seleção individual das amostras rotuladas como LOS ou NLOS.
- Treinamento e teste em condições homogêneas: Cenário 1, treinamento e teste exclusivamente com dados LOS. Cenário 2, treinamento e teste exclusivamente com dados NLOS.

Essa separação permitiu investigar a sensibilidade do modelo WiSARD a cada regime de propagação, analisando a variação de desempenho (acurácia top-1/top-k) em ambientes com maior previsibilidade (LOS) e em ambientes de alta dispersão multipercurso (NLOS).

b) Avaliação Cruzada entre Cenários por Condição de Propagação (LOS/NLOS)

Com o intuito de avaliar a sensibilidade e generalização da rede WiSARD frente a mudanças do ambiente, serão repetidos os experimentos da proposta do item *a*, porém realizando uma inversão dos *datasets* para treinamento e para teste. Desta forma, será possível investigar como o modelo responde frente aos ambientes considerados.

c) Avaliação de Generalização entre Cenários (Troca de Domínio)

Para avaliar a capacidade de generalização da rede WiSARD e dos modelos de *baseline* frente à mudança de ambiente, foram conduzidos experimentos de *cross-scenario testing*, nos quais os conjuntos de treinamento e teste pertencem a diferentes subconjuntos (s008 e s009), agora sem considerar o tipo de conexão. As configurações consideradas foram:

Configuração A: treinamento com s008 e teste com s009.

Configuração B: treinamento com s009 e teste com s008.

Essa abordagem simula a implantação de um modelo treinado em um cenário urbano e sua posterior aplicação em um ambiente distinto, sem adaptação.

O objetivo é verificar se o modelo mantém desempenho satisfatório quando exposto a dados com distribuição espacial e estatística diferente da observada no treinamento.

A métrica utilizada para avaliação dos modelos foi a acurácia top-k (percentual de acertos em que a resposta correta está entre o conjunto k das melhores predições).

13.6 Resultados parciais

a) Avaliação por condição de propagação (LoS/NLoS)

Nesta seção, apresentamos os resultados obtidos pela rede WiSARD ao ser avaliada separadamente em cenários com visada direta (LoS) e sem visada direta (NLoS), considerando diferentes tipos de entrada (coordenadas, LiDAR e dados multimodais) em função de diferentes tamanhos de endereçamento das memórias RAM (hiper-parâmetro deste modelo).

A Figura 35 mostra a acurácia de predição top-1 em função do tamanho do endereço de memória (eixo horizontal), variando de 6 a 64 bits, para os três tipos de entrada: Coordenadas (esquerda), LiDAR (centro), LiDAR + Coordenadas (direita). Cada curva representa o desempenho da rede WiSARD em um tipo de cenário: LOS (azul), NLoS (laranja) e todos os dados combinados (verde).

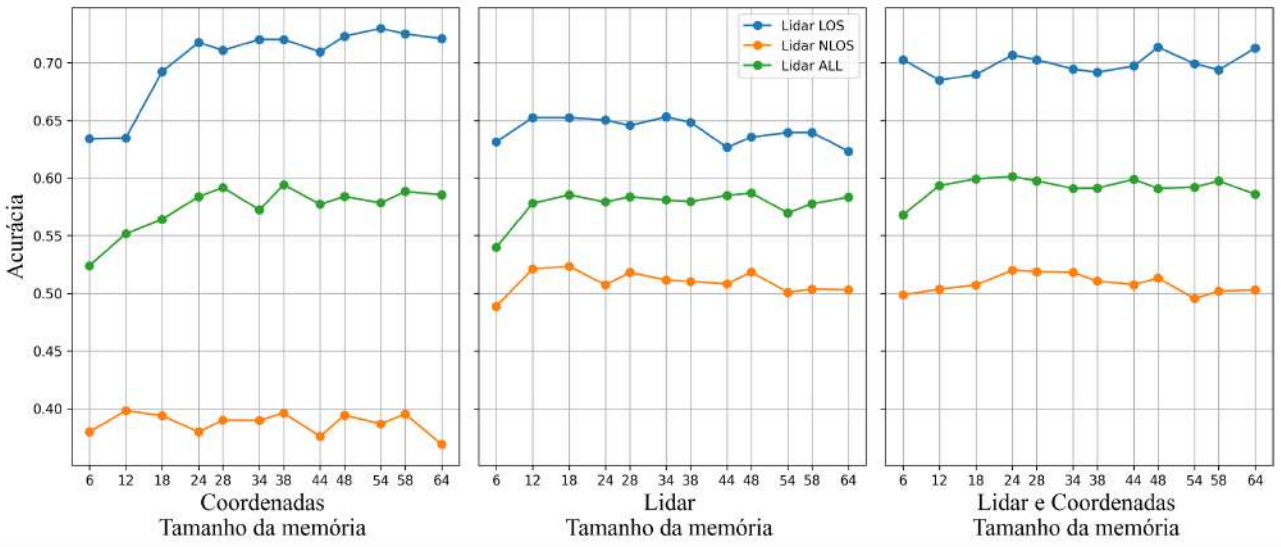


Figura 35: Acurácia da rede WiSARD em função do tamanho de memória, para diferentes tipos de entrada e cenários LOS/NLOS.

- Entradas com coordenadas

  A rede WiSARD apresentou melhor desempenho com entradas baseadas em coordenadas no cenário LOS, atingindo acurácia superior a 71% com endereços de memória entre 28 e 44 bits. Por outro lado, o desempenho em NLOS foi significativamente inferior, mantendo-se abaixo de 40% em todos os casos. Isso evidencia a limitação da informação posicional em ambientes com bloqueios e reflexões predominantes.

- Entradas com LiDAR

  Ao utilizar apenas dados LiDAR, a acurácia apresentou comportamento mais estável entre os diferentes tamanhos de memória. No cenário LoS, a acurácia situou-se em torno de 65%, enquanto no cenário NLoS foi ligeiramente superior a 50%. Essa diferença reflete a maior capacidade dos dados LiDAR de representar obstáculos e estrutura do ambiente, mesmo em situações não lineares.

- Entradas multimodais (LiDAR + Coordenadas)

A combinação dos dados de LiDAR e coordenadas resultou em uma melhora geral de desempenho. O modelo alcançou acurácia de até 73% no cenário LoS, com desempenho mais consistente no NLoS (em torno de 51%), superando os resultados obtidos com entradas individuais. Isso demonstra que os dados complementares oferecem vantagens significativas, principalmente em ambientes heterogêneos.

b) Avaliação cruzada entre cenários (*Cross-Domain* LoS/NLoS)

Para investigar a capacidade de generalização da rede WiSARD frente à variação de cenário, foi conduzida uma avaliação cruzada, onde os subconjuntos s008 e s009 foram utilizados de forma invertida para treinamento e teste. A análise considerou separadamente os canais com visada direta (LoS) e sem visada (NLoS), além do conjunto completo (ALL). Os resultados são apresentados na Figura 36.

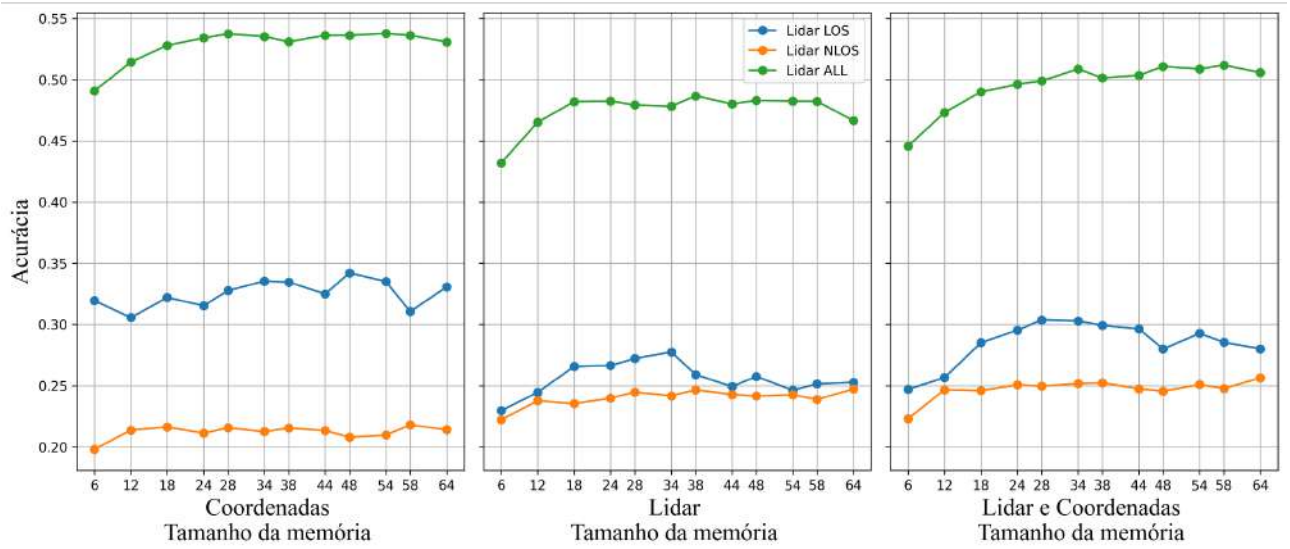


Figura 36: Acurácia do modelo WiSARD com inversão dos conjuntos de treinamento e teste (cross-domain), considerando separação entre cenários LOS, NLOS e ALL.

- Desempenho geral. De forma geral, observa-se uma queda expressiva nas acurácias quando comparadas à avaliação intra-cenário. Isso evidencia a sensibilidade do modelo WiSARD à distribuição espacial dos dados, especialmente sob mudanças topológicas e estatísticas entre os cenários.
- Entradas com Coordenadas. Quando utilizadas apenas as coordenadas como entrada, o modelo apresentou desempenho satisfatório para o conjunto completo, alcançando acurácias acima de 53% a partir de 24 bits. No entanto, nos cenários LoS e NLoS isoladamente, os valores de acurácia foram limitados a aproximadamente 33% e 22%, respectivamente. A baixa performance em NLOS reforça a limitação das informações de posição em ambientes com múltiplos caminhos e bloqueios.
- Entradas com LiDAR. Com este tipo de entrada, o modelo manteve maior estabilidade, atingindo cerca de 48% no conjunto completo. Em cenários LoS e NLoS, a acurácia variou entre 25% e 20%, com pouca variação em relação ao tamanho da memória. Ainda que inferiores aos resultados obtidos em avaliação intra-cenário, os dados LiDAR contribuíram para uma generalização ligeiramente superior em NLoS.

- Entradas Multimodais (LiDAR + Coordenadas). A combinação das entradas resultou no melhor desempenho geral, atingindo aproximadamente 52% de acurácia para o conjunto completo. No entanto, os ganhos em NLoS foram modestos (até 25%), o que mostra que, mesmo com entradas enriquecidas, o modelo enfrenta dificuldades para generalizar nesse tipo de cenário. O desempenho em LoS, por outro lado, foi ligeiramente superior ao obtido com entradas isoladas.

c) Avaliação comparativa de generalização dos modelos de ML estudados entre cenários

A inversão dos conjuntos de treinamento e teste entre os *datasets* s008 e s009 permitiu avaliar a capacidade de generalização dos modelos de ML sob mudança de domínio, evidenciando seu comportamento frente a variações estruturais e contextuais entre ambientes distintos. Esta análise é crucial em cenários reais, onde a distribuição dos dados em produção pode divergir significativamente daquela observada durante o treinamento. Nas Figuras 37, 38 e 39, é possível observar o desempenho do modelo WiSARD (curvas em verde) e dos modelos de *baseline* apresentados por Ruseckas [126](vermelho) e Batool [125](azul). Nelas, são resumidos os resultados alcançados por todos os modelos nas duas configurações A e B. As linhas contínuas mostram o desempenho dos modelos usando a configuração A (treinamento com o *dataset* s008 e teste com o s009), já as linhas pontilhadas apresentam o desempenho da configuração B (treinamento com o *dataset* s009 e teste com o s008).

c.1) Comparação de desempenho – Entrada: Coordenadas

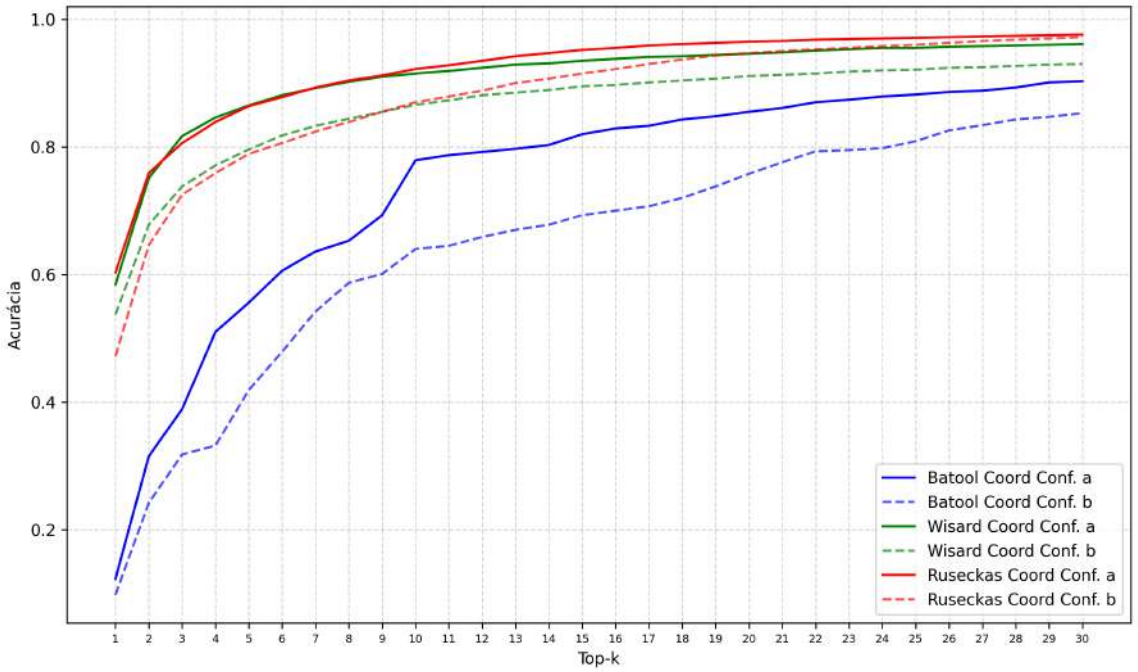


Figura 37: Comparação de desempenho dos modelos de ML com e sem inversão dos conjuntos de treinamento e teste *(cross-domain)* usando dados de coordenadas.

De forma geral, é possível observar na Figura 37, que todos os modelos apresentam melhor desempenho na configuração A (s008→s009) do que na configuração B (s009→s008), indicando que o cenário com treino em s008 e teste em s009 é mais

favorável à generalização. Esse padrão é mais acentuado no modelo da Batool, o que evidencia maior sensibilidade deste modelo à troca de domínio.

O desempenho do modelo proposto pela *Batool* na configuração A começou com uma acurácia top-1 baixa (12%) e cresceu gradualmente até cerca de 78% no top-10. Na configuração B, a acurácia top-1 é ainda mais baixa (10%) e sobe mais lentamente, atingindo 65% no top-10. Diante este resultado, podemos observar que este modelo é fortemente impactado por mudança de domínio; aprendizado altamente dependente do contexto espacial do treinamento.

O desempenho do modelo proposto pelo *Ruseckas* na configuração A, apresentou um bom desempenho já no top-1 (60%), superando 80% no top-3 e atingindo 92% no top-10. Na configuração B, embora haja redução, a acurácia top-10 ainda fica em torno de 87%, similar à WiSARD. Este modelo apresenta robustez à troca de cenário — menos sensível que *Batool*, com desempenho mais consistente entre domínios.

O desempenho do modelo proposto neste trabalho usando a *RNSP-WiSARD*, apresentou uma boa acurácia na configuração A, já no top-1 (58%), superando 85% no top-3 e chegando a 91% no top-10. Na configuração B, apesar da queda de desempenho, mantém acurácia acima de 53% no top-1 e 87% no top-10. Em suma, a WiSARD apresenta robustez à troca de cenário — é, portanto, menos sensível. Também robusto à mudança de domínio, com leve vantagem sobre o modelo de *Ruseckas* em top-k inferiores a k=10.

c.2) Comparação de desempenho – Entrada: LiDAR

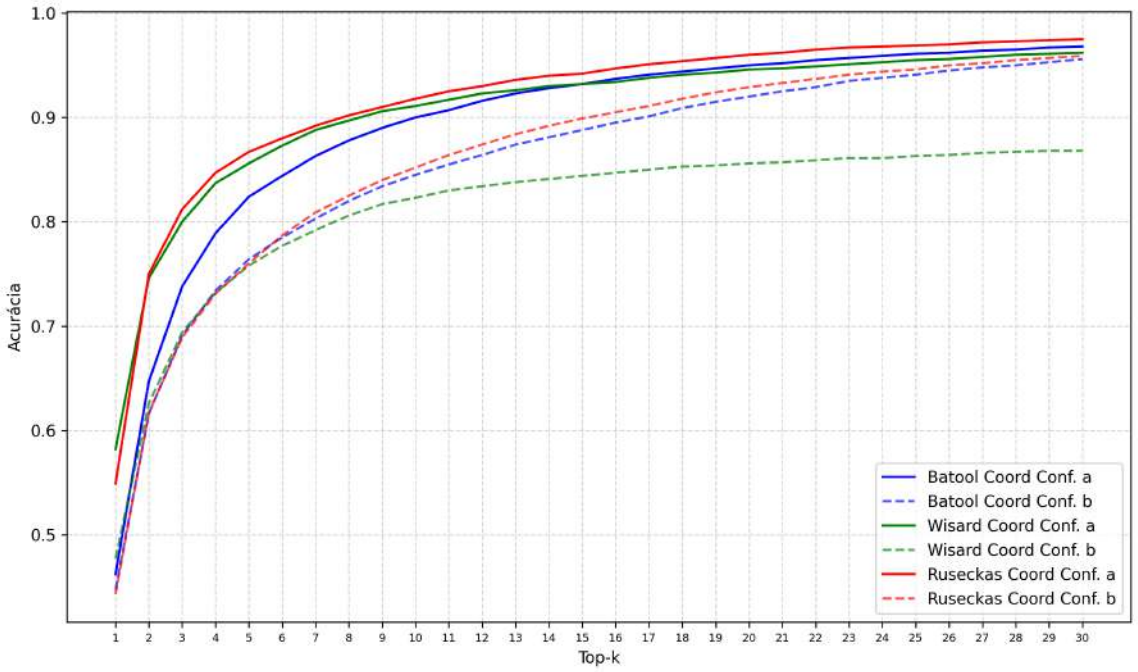


Figura 38: comparação de desempenho dos modelos de ML com e sem inversão dos conjuntos de treinamento e teste *(cross-domain)* usando dados de LiDAR.

Com o uso de dados LiDAR, todos os modelos apresentaram maior estabilidade entre os dois cenários em comparação ao uso de coordenadas, como evidenciado pela Figura 38. A diferença entre a configuração A (s008→s009) e B (s009→s008)

se reduziu substancialmente. O ganho em acurácia ao aumentar o top-k é mais consistente e rápido, especialmente para WiSARD e Ruseckas.

Embora o modelo da *Batool* na configuração A tenha iniciado com menor acurácia top-1 (47%), o comportamento ascendente da curva até top-10 revela bom aproveitamento das características espaciais capturadas pelo LiDAR, o que suaviza sua fragilidade contextual identificada com o uso exclusivo de coordenadas.

Na configuração B, o resultado foi muito próximo, no top-1 45%, top-10 83%. Isto evidencia um desempenho moderado, mas com boa generalização entre os domínios. O uso do LiDAR claramente compensa a fragilidade observada com coordenadas.

O modelo proposto pelo *Ruseckas* apresentou um desempenho comparável com a rede *WiSARD*, na configuração A, o Top-1 alcançou uma acurácia de 54%, top-10 92%. Já na configuração B, o Top-1 diminuiu para 44%, e o top-10 85%. Estes resultados mostram que este modelo continua robusto à mudança de domínio, mas menos impactado pelo tipo de entrada do que *Batool*.

A rede *WiSARD*, na configuração A apresentou um Top-1 já alto (58%), atingindo 91% no top-10. Na configuração B, foi evidenciada uma pequena queda de desempenho — top-1 53%, top-10 86%. Assim, o modelo WiSARD apresentou um excelente desempenho com LiDAR. Ainda é o modelo mais eficaz mesmo em configuração B, com mínima perda de acurácia.

c.3) Comparação de desempenho – Entrada: Coordenadas + LiDAR

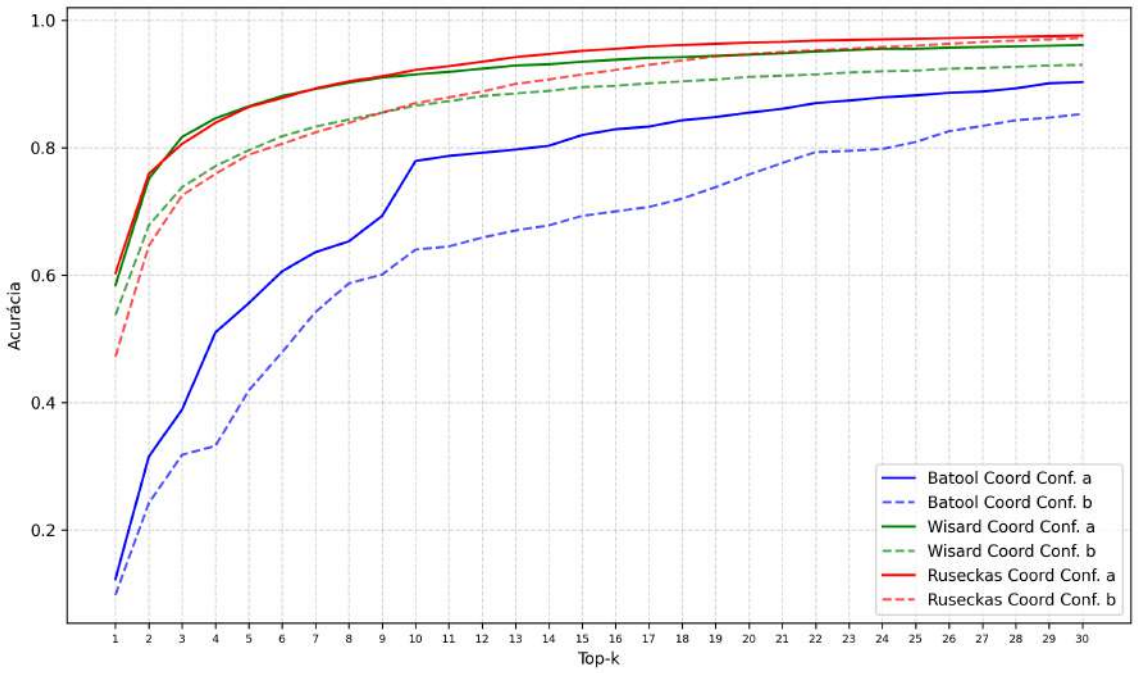


Figura 39: Comparação de desempenho dos modelos de ML com e sem inversão dos conjuntos de treinamento e teste (cross-domain) usando dados de LiDAR + coordenadas.

Os resultados da junção de coordenadas + LiDAR sao apresentados na Figura 39. Esta junção evidencia os ganhos claros de desempenho para todos os modelos. A acurácia top-1 é mais alta em relação às entradas isoladas e cresce de forma mais

acentuada até o top-5. A diferença entre os cenários (configuração A vs. B) se mantém pequena, indicando robustez na generalização com esse tipo de entrada.

O modelo da *Batool* apresentou um ganho claro sobre as entradas separadas (especialmente coordenadas), mas o desempenho ainda está abaixo dos demais modelos. O modelo *WiSARD*, continuou sendo o modelo com melhor desempenho global. A fusão de entradas fez com que a *Ruseckas* superasse ou empatasse com *WiSARD* até o top-5 no cenário A. Embora a diferença frente eles tenha diminuído. A curva mostra alta precisão e estabilidade, com diferença modesta entre os cenários.

13.7 Conclusão e próximos passos

Resultados Gerais

Em geral, os modelos apresentaram uma queda de desempenho na configuração B (s009→s008), indicando menor capacidade de generalização neste sentido de domínio.

A análise comparativa entre os modelos Batool, WiSARD e Ruseckas, considerando diferentes tipos de entrada (coordenadas, LiDAR e a combinação de ambos), revela padrões importantes em termos de desempenho e robustez, como mostrado no resumo da Tabela 10. A entrada baseada exclusivamente em coordenadas demonstrou maior sensibilidade à inversão dos domínios de treino e teste (configurações A e B), impactando principalmente o modelo Batool. Por outro lado, o uso exclusivo de dados LiDAR proporcionou maior estabilidade entre os cenários, beneficiando especialmente o modelo WiSARD e o modelo proposto por Ruseckas. A combinação de coordenadas com LiDAR, por sua vez, resultou nos melhores desempenhos globais para todos os modelos, reduzindo as diferenças entre eles e evidenciando o valor da fusão multimodal. Nesse cenário, WiSARD se mostrou altamente competitivo, alcançando desempenho muito próximo ao de Ruseckas com menor variação entre os cenários. Assim, observa-se que a entrada combinada não apenas eleva o teto de desempenho, mas também fortalece a generalização entre domínios, favorecendo aplicações em ambientes heterogêneos.

Próximos Passos e Considerações Futuras

Devido à acentuada queda de desempenho observada na avaliação cruzada entre cenários de propagação, especialmente nas condições de linha de visada direta, torna-se necessário aprofundar a investigação sobre os fatores que influenciam essa sensibilidade nos modelos de aprendizado de máquina, incluindo a rede WiSARD.

Esse comportamento foi identificado de forma recorrente em todos os tipos de entrada (coordenadas, LiDAR e fusão de ambos), indicando que a limitação não está apenas no tipo de dado, mas possivelmente relacionada à maneira como as características espaciais e estruturais dos ambientes são codificadas e interpretadas pelos modelos.

Nos canais LOS, a resposta do canal tende a ser fortemente dominada por um único caminho direto, o que resulta em menor diversidade espacial e maior dependência da geometria local (por exemplo, ângulos de chegada e distâncias relativas exatas). Como consequência, pequenas variações entre os conjuntos de treinamento e teste — como mudanças na topologia urbana, posição dos edifícios ou nas trajetórias dos veículos — podem resultar em grandes discrepâncias nas estatísticas do canal. Isso afeta diretamente a generalização dos modelos, especialmente se eles se baseiam fortemente nessas características determinísticas.

Em contraste, canais NLOS apresentam maior diversidade de trajetos de propagação e múltiplos componentes refletidos e difratados, o que pode levar a uma distribuição estatística

Modelo	Entrada	Conf. A		Conf. B		Observações
		Top-1	Top-10	Top-1	Top-10	
Batool	Coordenadas	12%	78%	9,7%	64%	Alta sensibilidade ao domínio; desempenho inferior.
	LiDAR	46%	89%	44%	84%	Melhora significativa; maior estabilidade entre domínios.
	LiDAR + Coord	42%	82%	38%	83%	Fusão ajuda, mas ainda inferior aos demais.
Ruseckas	Coordenadas	60%	92%	47%	87%	Robusto à troca de domínio; liderança em coordenadas.
	LiDAR	54%	92%	44%	85%	Curvas consistentes; ótima generalização.
	LiDAR + Coord	58%	91%	53%	86%	Melhora geral; empate com WiSARD até top-5.
WiSARD	Coordenadas	58%	91%	53%	86%	Melhor que Batool; similar ao Ruseckas até top-10.
	LiDAR	58%	91%	47%	82%	Curvas consistentes; ótima generalização.
	LiDAR + Coord	60%	92%	50%	87%	Maior estabilidade entre os modelos avaliados.

Tabela 10: Resumo comparativo da acurácia dos modelos nas configurações a (s008→s009) e b (s009→s008), considerando diferentes tipos de entrada e top-k.

mais rica e menos dependente de um caminho específico. Tal propriedade parece contribuir para uma generalização mais robusta quando os modelos são treinados e testados em domínios distintos.

Dessa forma, os próximos passos sugeridos incluem:

- Caracterização do espaço de feixes ótimos (*codebook indices*) em cada domínio, investigando a distribuição dos rótulos e sua variabilidade intra-classe;
- Estudo da sensibilidade dos modelos à geometria do cenário: avaliar a correlação entre pequenas mudanças espaciais (por exemplo, deslocamento do receptor) e grandes alterações nos rótulos de feixe ótimo;

14 Agrupamento de Usuários em Redes *Cell-Free* com TDD Dinâmico

Samuel S. Silva, Yuri C. B. Silva, Igor M. Guerreiro, Victor F. Monteiro, Matheus D. Carneiro e Abraão C. Albuquerque
UFC

14.1 Introdução

Os sistemas de comunicação sem fio vêm evoluindo para acomodar a crescente densidade de usuários ativos, fornecer maiores taxas de transmissão de dados e baixa latência, além de suportar tráfego intenso de dados [134]. Visando atingir esses objetivos, diversos estudos indicaram que a arquitetura de rede sem célula, do inglês *cell-free*, que combina uma implantação distribuída com técnicas avançadas de processamento de sinais, oferece maior eficiência e escalabilidade para atender a esses requisitos [135]. Uma rede *cell-free* é composta por um grande número de pontos de acesso, do inglês *Access Point*s (APs), distribuídos pela área de cobertura para fornecer uma qualidade de sinal mais uniforme e aprimorada. Esses APs são conectados a uma unidade central de processamento, do inglês *Central Processing Unit* (CPU), por meio de enlaces de *fronthaul* com fio, o que lhes permite compartilhar informações e realizar transmissões conjuntas para cada equipamento de usuário, do inglês *User Equipment* (UE), melhorando assim a recepção do sinal e a qualidade geral do serviço [136].

No entanto, apesar das vantagens conhecidas dessa arquitetura de rede, a maioria dos estudos que abordam cenários com redes *cell-free* assume que a alocação da direção de transmissão é realizada estaticamente, o que significa que todos os APs do sistema transmitem simultaneamente no mesmo modo e compartilham a mesma configuração de quadros, que consiste em uma divisão fixa dos recursos de tempo entre enlace reverso, do inglês *Uplink* (UL), e o enlace direto, do inglês *Downlink* (DL). Como resultado da crescente heterogeneidade do tráfego de dados e da crescente demanda por taxas de transmissão mais altas, sistemas que empregam duplexação por divisão do tempo, do inglês *Time Division Duplex* (TDD), de forma estática tendem a apresentar degradação de desempenho e capacidade reduzida para suportar um grande número de UEs, exigindo, portanto, novas estratégias para lidar com essa limitação e gerenciar os recursos temporais de forma mais eficiente.

Considerando esse conjunto de fatores, a implementação de TDD dinâmico surgiu como uma solução promissora para gerenciar as demandas de tráfego heterogêneas e melhorar a qualidade de serviço, do inglês *Quality of Service* (QoS), do UE por meio da utilização mais eficiente dos recursos de tempo. A ideia fundamental por trás do TDD dinâmico é configurar o quadro de transmissão do AP da maneira que melhor se adapte às demandas instantâneas de tráfego do sistema. Em cenários celulares tradicionais com TDD dinâmico, a configuração do quadro de cada célula é normalmente ajustada individualmente, com base apenas no tráfego gerado pelo UE dentro daquela célula. No entanto, em um cenário *cell-free*, empregando uma abordagem centrada no usuário, é possível que um quadro de transmissão do UE seja configurado de maneira mais individualizada, adaptada às demandas específicas de tráfego de cada usuário. Isso é possível por meio de técnicas de agrupamento que permitem que cada UE seja atendido

O conteúdo desta seção é uma adaptação do artigo submetido ao SBrT 2025: Samuel S. Silva, Yuri C. B. Silva, Igor M. Guerreiro, Victor F. Monteiro, Matheus D. Carneiro e Abraão C. Albuquerque, "User-Centric Clustering in Cell-Free Networks with Dynamic TDD".

por um *cluster* separado de APs dedicado exclusivamente a esse UE, considerando um sistema no qual o AP opera em modo *half-duplex*.

No entanto, a implementação de TDD dinâmico introduz novos componentes de interferência devido à coexistência de transmissões UL e DL, como a interferências entre APs, do inglês *Inter-AP Interference* (IAI), e a interferência entre usuários, do inglês *Inter-User Interference* (IUI). Diante dessa questão, tornou-se claro que técnicas eficazes de cancelamento de interferência são essenciais para permitir a implantação prática de TDD dinâmico, do inglês *Dynamic TDD* (D-TDD). Alguns estudos abordaram o cancelamento de interferência entre pontos de acesso explorando a comunicação entre APs por meio de enlaces de *fronthaul*, uma técnica conhecida como *full-duplex* assistido por rede [137]. Outros estudos exploraram a possibilidade de mitigar parte da interferência entre UEs por meio de uma técnica de agrupamento baseada em *clusters*, considerando simultaneamente o cancelamento de IAI mencionado anteriormente [138].

A principal contribuição deste estudo é a análise de dois algoritmos de agrupamento que realizam uma etapa adicional de alocação de APs em um sistema *cell-free* com D-TDD. A motivação para a implementação desses algoritmos é permitir que o AP atenda a múltiplos UEs transmitindo na mesma direção, aumentando assim o número de antenas que atendem a cada UE e aprimorando a qualidade do serviço, sem aumentar significativamente os níveis de interferência.

14.2 Modelo do sistema

Neste estudo, consideramos um sistema *cell-free* de múltiplas entradas e múltiplas saídas, do inglês *Multiple-Input Multiple-Output* (MIMO), com D-TDD, composto por L APs e K UEs, ambos distribuídos aleatoriamente pela área de cobertura e equipados com uma única antena omnidirecional. Os APs são conectados a uma CPU central por meio de enlaces de *fronthaul* e operam em modo *half-duplex*, o que significa que podem transmitir apenas em uma direção por vez. A Figura 40 ilustra o sistema considerado.

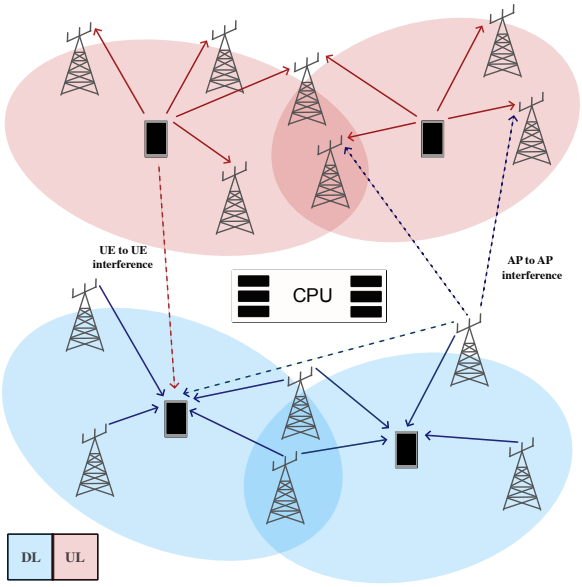


Figura 40: Sistema *cell-free* com TDD dinâmico.

O canal de propagação entre o k-ésimo transmissor e o l-ésimo receptor, denotado por $h_{k,l} \in \mathbb{C}$, é modelado seguindo uma distribuição de Rice, com componentes de linha de visada,

do inglês *Line of Sight* (LOS), e sem linha de visada, do inglês *Non-Line of Sight* (NLOS), o qual pode ser descrito da seguinte forma:

$$\mathrm{h}_{k,l} = \left[\sqrt{\frac{\bar{K}_{k,l}}{\bar{K}_{k,l}+1}} + \sqrt{\frac{1}{\bar{K}_{k,l}+1}} h_{k,l}^{(w)} \right] \beta_{k,l} \,, \tag{181}$$

onde $h_{k,l}^{(w)} \sim C\mathcal{N}\,(0{,}1)$ é o componente NLOS modelado seguindo uma distribuição não correlacionada de Rayleigh, $\beta_{k,l}$ representa o coeficiente de larga escala, que modela os efeitos de perda de percurso e sombreamento, e $\bar{K}_{k,l}$ é o fator K de Rice, como em [139]. Podemos omitir o vetor de direção na equação, já que cada transmissor e receptor possui apenas uma antena. A probabilidade de LOS é definida para cada tipo de enlace. Por exemplo, o enlace entre UEs e APs, bem como o enlace entre dois APs, está sujeito à probabilidade de LOS. Para ambos os tipos de enlace, a probabilidade de LOS pode ser definida como:

$$\mathrm{Prob}_{\mathrm{LOS}}(R) = \; 0.5 - \min\left(0.5, 5e^{-0.156/R}\right) + \min\left(0.5, 5e^{-R/0.03}\right) \,, \tag{182}$$

em que R é a distância entre o transmissor e o receptor em km. Dada a probabilidade de LOS, é possível determinar o fator K de Rice. Se o canal for LOS, $K \sim \mathcal{N}\,(9{,}5)$ dB para ambos os enlaces UE-AP ou AP-AP; se o canal for NLOS, então K é igual a zero, de acordo com [140]. Consideramos que o enlace entre dois UEs não tem probabilidade de ser LOS.

O coeficiente de desvanecimento de larga escala $\beta_{k,l}$ é definido como:

$$\beta_{k,l} = 10^{\frac{PL_{k,l}+\sigma_{sh}}{10}} \,, \tag{183}$$

onde $\sigma_{sh} \sim \mathcal{N}\,(0, \sigma^2)$ e $PL_{k,l}$ representam os ganhos de sombreamento e perda de percurso em dB. O cálculo da perda de percurso varia de acordo com o tipo de enlace modelado e se o canal é ou não LOS. O cálculo da perda de percurso para todos os tipos de enlace e condições de LOS é implementado como em [141].

Como mencionado anteriormente, a coexistência de transmissões em direções opostas dá origem a novos tipos de interferência, como IAI e IUI. Além disso, os níveis de interferência no sistema variam continuamente conforme a demanda de tráfego flutua, o que exige o recálculo constante da eficiência espectral em cada intervalo de tempo.

Denotamos o vetor de canal entre o k-ésimo UE e os APs que compõem o *cluster* C_k como $\mathbf{h}_{(C_k,k)} \in \mathbb{C}^{|C_k|\times 1}$, o vetor de canal entre o j-ésimo UE interferente e o *cluster* de APs C_k como $\mathbf{h}_{(C_k,j)} \in \mathbb{C}^{|C_k|\times 1}$, a matriz de canal entre os clusters de APs C_k e C_m como $\mathbf{H}_{(C_k,C_m)} \in \mathbb{C}^{|C_k|\times|C_m|}$ e o coeficiente de canal entre o k-ésimo UE e o j-ésimo UE como $h_{k,j} \in \mathbb{C}$.

O sinal recebido pelo *cluster* de APs C_k operando em UL no n-ésimo *slot* é dado por:

$$\begin{aligned}\mathbf{y}_k^{\mathrm{UL}} = & \underbrace{\sqrt{\rho_{\mathrm{ul}}}\mathbf{h}_{(C_k,k)} x_k T_{\mathrm{ul}(k)}}_{\text{sinal desejado}} + \sum_{\substack{j=1\\ j\neq k}}^{K} \underbrace{\sqrt{\rho_{\mathrm{ul}}}\mathbf{h}_{(C_k,j)} x_j T_{\mathrm{ul}(j)}(1-\delta_{(j)})}_{\text{Interferência UE-AP}} \\ & + \sum_{\substack{m=1\\ m\neq k}}^{K} \underbrace{\mathbf{H}_{(C_k,C_m)}\mathbf{P}_{\mathrm{dl}(m)}\mathbf{w}_m s_m T_{\mathrm{dl}(m)}\delta_{(m)}}_{\text{Interferência AP-AP}} + \underbrace{\mathbf{z}_{\mathrm{ul}(k)}}_{\text{ruído}},\end{aligned} \tag{184}$$

onde $\rho_{\mathrm{ul}} \in \mathbb{R}_+$ é a potência de transmissão no UL; $\mathbf{P}_{\mathrm{dl}(m)} \in \mathbb{R}_+^{|C_m|\times|C_m|}$ é uma matriz diagonal de alocação de potência no DL relativa ao UE m, com cada elemento na diagonal indicando

a raiz quadrada da potência alocada ao UE m em cada AP do *cluster* de APs C_m; δ_k indica a direção de transmissão do UE k, com $\delta_k = 0$ indicando UL e $\delta_k = 1$ indicando DL; $T_{\mathrm{ul}(k)}$ e $T_{\mathrm{dl}(k)}$ denotam a presença ou ausência de dados a serem transmitidos na fila de UL e DL do UE k; $\mathbf{w}_m \in \mathbb{C}^{|C_m|\times 1}$ denota o vetor de formação de feixe de transmissão do *cluster* C_m para o UE m, e x_k e s_m denotam os símbolos de transmissão de potência unitária no UL e DL, respectivamente. Os índices temporais de cada componente da equação foram omitidos para que possam ser melhor visualizados; no entanto, é extremamente importante mencionar que alguns desses componentes variam ao longo do tempo. Em um intervalo de tempo, a quantidade de bits transmitidos pelo k-ésimo UE operando no UL é dada por:

$$R_k^{\mathrm{ul}} = B\,\log_2\left(1 + \frac{\rho_{\mathrm{ul}}\left|\mathbf{u}_k^T\mathbf{h}_{(C_k,k)}\right|^2 T_{\mathrm{ul}(k)}}{\gamma_k^{UL}}\right) \times \tau_s, \tag{185}$$

onde $\mathbf{u}_k \in \mathbb{C}^{|C_k|\times 1}$ representa o vetor de formação de feixe de recepção de norma unitária para o *cluster* de APs C_k, τ_s é a duração do *slot* e γ_k^{UL} representa a soma dos componentes de interferência e ruído:

$$\gamma_k^{\mathrm{UL}} = \sum_{\substack{j=1\\ j\neq k}}^{K} \rho_{\mathrm{ul}}\left|\mathbf{u}_k^T\mathbf{h}_{(C_k,j)}\right|^2 T_{\mathrm{ul}(j)}\left(1-\delta_{(j)}\right) + \sum_{\substack{m=1\\ m\neq k}}^{K}\left|\mathbf{u}_k^T\mathbf{H}_{(C_k,C_m)}\mathbf{P}_{\mathrm{dl}(m)}\mathbf{w}_m\right|^2 T_{\mathrm{dl}(m)}\delta_{(m)} + \sigma_{\mathrm{ul}}^2. \tag{186}$$

O sinal recebido pelo k-ésimo UE operando em DL no n-ésimo slot é dado por:

$$\begin{aligned} y_k^{\mathrm{DL}} = &\underbrace{\mathbf{h}_{(C_k,k)}^T\mathbf{P}_{\mathrm{dl}(k)}\mathbf{w}_k s_k T_{\mathrm{dl}(k)}}_{\text{sinal desejado}} + \sum_{\substack{j=1\\ j\neq k}}^{K}\underbrace{\mathbf{h}_{(C_j,k)}^T\mathbf{P}_{\mathrm{dl}(j)}\mathbf{w}_j s_j T_{\mathrm{dl}(j)}\delta_{(j)}}_{\text{Interferência AP-UE}} \\ &+ \sum_{\substack{m=1\\ m\neq k}}^{K}\underbrace{\sqrt{\rho_{\mathrm{ul}}}\mathrm{h}_{m,k}x_m T_{\mathrm{ul}(m)}\left(1-\delta_{(m)}\right)}_{\text{Interferência UE-UE}} + \underbrace{z_{\mathrm{dl}(k)}}_{\text{ruído}}. \end{aligned} \tag{187}$$

Em um *slot* de tempo, a quantidade de bits transmitidos para o k-ésimo UE operando em DL é dada por:

$$R_k^{\mathrm{dl}} = B\,\log_2\left(1 + \frac{|\mathbf{h}_{(C_k,k)}^T\mathbf{P}_{\mathrm{dl}(k)}\mathbf{w}_k|^2 T_{\mathrm{dl}(k)}}{\gamma_k^{dl}}\right) \times \tau_s, \tag{188}$$

onde γ_k^{dl} representa a soma dos componentes de interferência e ruído percebidos pelo UE k:

$$\gamma_k^{\mathrm{DL}} = \sum_{\substack{j=1\\ j\neq k}}^{K}|\mathbf{h}_{(C_j,k)}^T\mathbf{P}_{\mathrm{dl}(j)}\mathbf{w}_j|^2 T_{\mathrm{dl}(j)}\delta_{(j)} + \sum_{\substack{m=1\\ m\neq k}}^{K}\rho_{\mathrm{ul}}|\mathrm{h}_{k,m}|^2 T_{\mathrm{ul}(m)}\left(1-\delta_{(m)}\right) + \sigma_{\mathrm{dl}}^2. \tag{189}$$

Ressalta-se que os termos de interferência cruzada, do inglês *Cross Link Interference* (CLI), são inexistentes quando todos os UEs são escalonados na mesma direção de transmissão, seja UL ou DL.

14.3 Configuração de quadro

Nesta seção, descrevemos como o quadro é configurado individualmente para cada UE, de acordo com [138]. Para esta etapa, o número de *slots* em um quadro que serão reservados para

transmissões de UL é calculado primeiro, conforme descrito a seguir:

$$\tau_{\mathrm{ul}(k)} = \mathrm{round}\left\{\frac{Q_{\mathrm{ul}(k)}}{Q_{\mathrm{dl}(k)} + Q_{\mathrm{ul}(k)}} \times N\right\}, \tag{190}$$

onde round é uma função que retorna o número inteiro mais próximo, $Q_{\mathrm{ul}(k)}$ e $Q_{\mathrm{dl}(k)}$ denotam a quantidade de dados de UL e DL, respectivamente, que são enfileirados para o UE k, e N é o número de *slots* que compõem um quadro completo. Após isso, o quadro para um determinado UE é configurado da seguinte forma:

$$\delta_k[n] = \begin{cases} 0, & 1 \leq n \leq \tau_{\mathrm{ul}(k)} \\ 1, & \tau_{\mathrm{ul}(k)} < n \leq N \end{cases}. \tag{191}$$

14.4 Algoritmos de agrupamento

Neste trabalho, consideramos uma estratégia de agrupamento de referência, que consiste em alocar um número fixo e igual de APs para cada UE, garantindo que sejam mantidos em *clusters* disjuntos. A motivação para esta primeira etapa é garantir que os UEs possam ser atendidos com uma configuração de tráfego adaptada às suas necessidades específicas. No entanto, a imposição de *clusters* disjuntos requer a definição de um critério de prioridade para a seleção de APs. Nesta fase inicial de alocação, a seleção de APs é baseada no desvanecimento em larga escala: o UE com o maior ganho médio de larga escala em todos os APs recebe a maior prioridade e seleciona os APs para o qual apresenta os maiores ganhos. Além disso, é importante destacar que esse esquema baseado em prioridade pode resultar na atribuição de alguns UEs a APs com ganhos de canal comparativamente menores.

A fim de aprimorar o desempenho do sistema e introduzir maior flexibilidade na alocação de APs de acordo com a direção do tráfego, este trabalho propõe a implementação de algoritmos que permitem que APs transmitindo na mesma direção possam atender múltiplos UEs como uma etapa complementar à estratégia de agrupamento de referência. Dada a variabilidade das direções de transmissão e do tráfego, essa etapa de reagrupamento é realizada recorrentemente no sistema e se torna viável na prática, pois as configurações de quadros de cada UE são definidas antecipadamente, antes da transmissão. Isso permite que a CPU calcule a alocação dinâmica de APs para todo o quadro.

Para esta etapa adicional, dois algoritmos de agrupamento originalmente projetados para redes *cell-free* com TDD estático, do inglês *Static TDD* (S-TDD), foram selecionados e adaptados para se adequarem ao cenário considerado. Os dois algoritmos de agrupamento compartilham o uso do desvanecimento em larga escala como métrica de seleção principal, embora também apresentem variações individuais em seu projeto. O agrupamento baseado em efeitos de desvanecimento em larga escala leva em consideração os efeitos de perda de percurso e sombreamento, mas não considera variações rápidas no canal de propagação, o que permite certa tolerância a variações nos níveis de sinal e maior flexibilidade na alocação de recursos de rede [142].

Algoritmo baseado em cooperação dinâmica:

Este algoritmo, baseado no método de agrupamento cooperativo dinâmico, do inglês *Dynamic Cooperation Clustering* (DCC), é caracterizado pela seleção de APs usando a relação sinal-ruído, do inglês *Signal-to-Noise Ratio* (SNR), como critério principal. A primeira etapa deste algoritmo identifica o AP já atribuído ao UE com o maior ganho em larga escala, que será designado como AP mestre [120]. Posteriormente, é realizada uma busca por APs que não pertencem ao

cluster atual do UE, mas que satisfazem à condição de SNR. Para que a condição seja satisfeita, a diferença entre a SNR do AP candidato e a do AP mestre, expressa em escala logarítmica, deve ser maior ou igual a um limiar predefinido. A variação deste limiar pode tornar a seleção do AP mais restritiva ou mais permissiva.

Algoritmo baseado em desvanecimento de larga escala

Na primeira etapa do algoritmo, assumindo que a CPU conhece os efeitos do desvanecimento de larga escala, do inglês *Large Scale Fading* (LSF), para todos os enlaces possíveis entre os APs e UEs, são calculados os coeficientes de larga escala entre todos os APs e UEs transmitindo no mesmo modo. Em seguida, todos os valores de ganho são agrupados para calcular um fator de limiar α_{lsf} da seguinte forma:

$$\alpha_{lsf} = \frac{1}{K'}\frac{1}{L'}\sum_{k=1}^{K'}\sum_{l=1}^{L'}\beta_{k,l}, \tag{192}$$

em que K' denota o número de UEs que pretende transmitir na mesma direção e L' denota a quantidade de APs que estão servindo esses UEs. Se o coeficiente de larga escala entre o k-ésimo UE e o l-ésimo AP $\beta_{k,l}$ for maior ou igual a α_{lsf}, o l-ésimo AP será alocado para servi-lo.

14.5 Resultados de simulação

O cenário simulado *cell-free* com D-TDD consiste em $K = 30$ UEs e $L = 120$ APs distribuídos em uma área de cobertura de 1.200 metros quadrados. O processamento de máxima razão, do inglês *Maximum Ratio* (MR), é considerado para vetores de formação de feixe de transmissão e recepção. Assumimos que todos os APs estão conectados a uma CPU central por meio de enlaces de *fronthaul*, o que permite o compartilhamento de informações. A potência de transmissão é constante em cada UE e em cada AP. Ao aplicar a etapa de reagrupamento, no caso de um AP servir a vários UEs no DL, a potência total do AP é dividida igualmente entre os fluxos desses UEs.

O desempenho do sistema é avaliado usando a função de distribuição acumulada, do inglês *Cumulative Distribution Function* (CDF), do atraso do pacote como métrica, com análises separadas para o desempenho de UL e DL. O tráfego de pacotes é modelado como um sistema de enfileiramento de tempo discreto, onde as chegadas de pacotes seguem um processo de Poisson, e tanto o enfileiramento quanto o desenfileiramento de pacotes são restritos ao início e ao fim de cada *slot*, respectivamente. Consideramos que o tempo médio entre chegadas dos pacotes varia de acordo com o tipo de pacote e a demanda de tráfego do UE, sendo menor para pacotes correspondentes à direção em que o UE tem maior demanda e maior para aqueles associados à direção de menor demanda. Para UEs com demanda de tráfego simétrica, o tempo entre chegadas é considerado igual para ambas as direções. O atraso do pacote é definido como o número de *slots* de tempo em que ele permanece na fila, contados do momento em que é enfileirado até o momento em que é desenfileirado. Os parâmetros de simulação, composição do sistema e tráfego de pacotes estão listados na Tabela 11.

A Figura 41 apresenta os resultados das estratégias de agrupamento implementadas em termos da CDF do atraso dos pacotes no UL: o agrupamento de referência (*No Re-Cluster*), os dois algoritmos adicionais (DCC e LSF) e as variações do DCC. A partir desses resultados, pode-se observar que as estratégias de reagrupamento, baseadas nos algoritmos apresentados, reduziram o atraso dos pacotes no UL. Esse resultado pode ser explicado pelo fato de que

Tabela 11: Parâmetros de simulação das estratégias de agrupamento de usuários.

Parâmetro	**Valor**
Altura do AP	11.5 m
Altura do UE	1.5 m
Largura de banda (B)	20 MHz
Potência UL	100 mW
Potência DL	1 W
Potência do ruído	-96 dBm
Variância do sombreamento (σ^2_{sh})	4 dB
Duração do *slot*	1 ms
Tamanho do pacote DL	50 kbytes
Tamanho do pacote UL	50 kbytes
Tempo de chegada entre pacotes	(UL, DL)
UE com carga elevada no UL:	(8, 16) ms
UE com carga elevada no DL:	(16, 8) ms
UE com carga simétrica:	(12, 12) ms

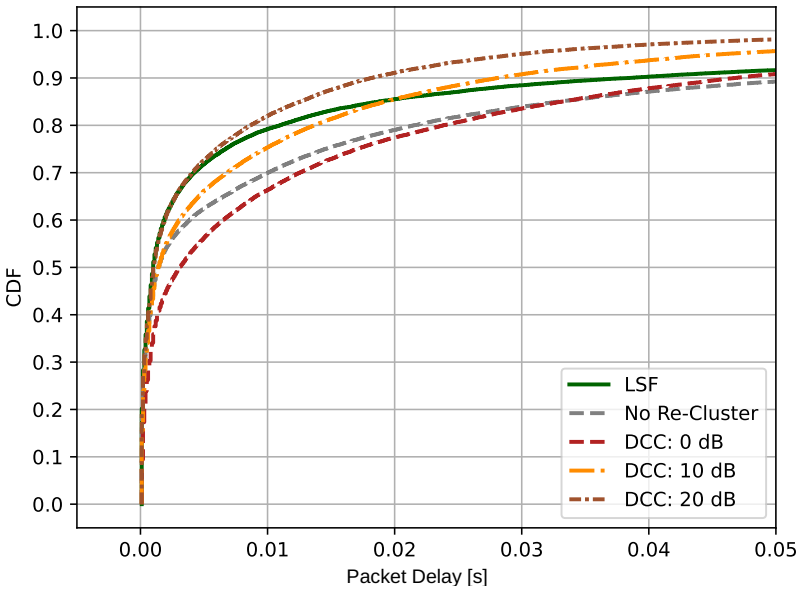


Figura 41: CDF do atraso de pacote no UL.

a estratégia de agrupamento de referência, que impõe *clusters* disjuntos, resulta em justiça limitada na alocação de APs. A etapa adicional de agrupamento atenua parcialmente esse problema, permitindo que um AP atenda a múltiplos UEs, melhorando assim o desempenho dos UEs em condições desfavoráveis.

Dentre as estratégias de reagrupamento implementadas, os métodos com melhor desempenho foram o algoritmo DCC com limiar de 20 dB e a abordagem baseada LSF, gerando reduções de 59,6% e 58,6% no atraso de pacotes no 70º percentil, respectivamente, em relação ao agrupamento de referência (*No Re-Cluster*). Para o mesmo valor de percentil, o algoritmo DCC com limiar de 10 dB obteve uma redução de 33,33% no atraso de pacotes em comparação com

o cenário de agrupamento totalmente disjunto, enquanto o DCC com limiar de 0 dB resultou em um aumento no atraso. No 90º percentil, os algoritmos DCC com limiares de 20 dB e 10 dB obtiveram reduções de atraso de pacotes de 66,17% e 47,74%, respectivamente, enquanto o método baseado em LSF resultou em uma redução de 34,21%. Esses resultados revelam uma lacuna significativa de desempenho entre os algoritmos nesse percentil. Para valores percentuais mais baixos, seu desempenho se aproxima bastante daquele do cenário de *clusters* disjuntos.

A Figura 42 apresenta a CDF do atraso dos pacotes no DL para as estratégias de agrupamento consideradas. A partir desses resultados, pode-se observar que as estratégias de reagrupamento também levaram a ganhos de desempenho. No entanto, as reduções no atraso dos pacotes foram consideravelmente menores do que as observadas no UL. Este resultado pode ser explicado pelo fato de que, ao servir múltiplos UEs, a potência de transmissão de um AP é dividida entre os símbolos a serem transmitidos. Isso leva a uma redução na intensidade do sinal de interesse para cada UE e a um aumento nos níveis de interferência devido aos símbolos de outros UEs.

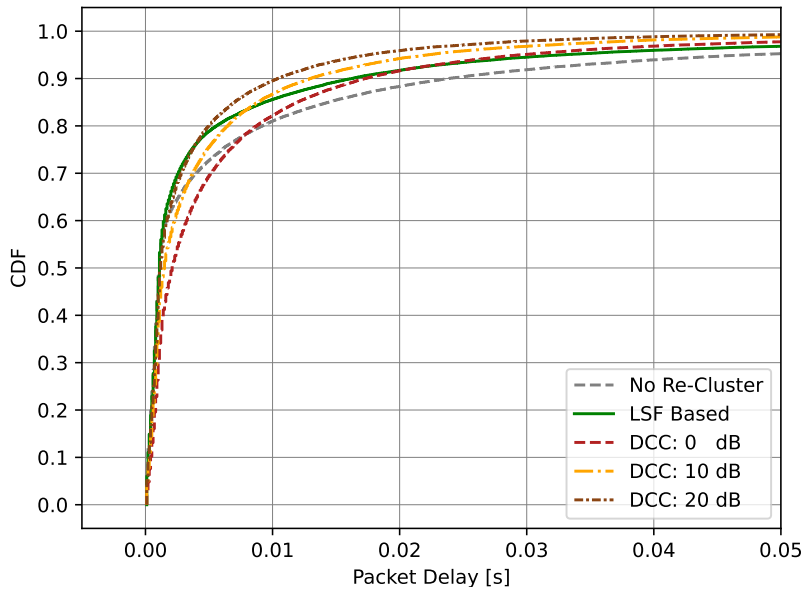


Figura 42: CDF do atraso de pacote no DL.

Para o caso no DL, entre os algoritmos implementados, o DCC com um limiar de 20 dB e o método baseado em LSF alcançaram novamente os melhores resultados no 70º percentil, mostrando reduções de 30% e 35%, respectivamente, no atraso do pacote no DL em comparação com o cenário de *clusters* disjuntos. Para o mesmo valor de percentil, o algoritmo DCC com um limiar de 10 dB mostrou uma redução de 7,5% no atraso do pacote, o que é substancialmente menor do que a melhoria correspondente no UL. Enquanto isso, o mesmo algoritmo com um limiar de 0 dB resultou novamente em um aumento no atraso neste percentil. No 90º percentil, os algoritmos DCC com limiares de 20 dB e 10 dB alcançaram reduções no atraso do pacote de 56,07% e 45,61%, respectivamente, enquanto o método baseado em LSF resultou em uma redução de 31,8% em comparação com o cenário de *clusters* disjuntos.

14.6 Conclusão

Neste trabalho, avaliamos o desempenho de um sistema MIMO *cell-free* com D-TDD sob diferentes estratégias adicionais de reagrupamento, considerando condições de propagação distintas para cada tipo de enlace no sistema. O sistema adotou inicialmente um método de agrupamento de referência que impõe *clusters* disjuntos.

Com base nos resultados obtidos, concluímos que a implementação da etapa de reagrupamento melhora o desempenho do sistema em termos do atraso de pacotes, em comparação com *clusters* de APs totalmente disjuntos. No entanto, dependendo do valor do limiar, o algoritmo DCC pode degradar o desempenho do sistema, pois o ganho de ter um AP adicional pode não compensar a interferência adicional. Para os atrasos de pacotes no UL, o algoritmo DCC com limiares de 10 e 20 dB e o algoritmo baseado em LSF obtiveram resultados satisfatórios e mostraram uma redução no atraso de pacotes em comparação com o cenário de *clusters* disjuntos.

15 Análise da Eficiência Energética na Alocação de Potência em Sistemas D-MIMO com TDD Dinâmico

Matheus D. Carneiro, Igor M. Guerreiro, Yuri C. B. Silva, Samuel S. Silva, Victor F. Monteiro e Abraão C. Albuquerque
UFC

15.1 Introdução

A próxima geração de redes móveis, conhecida como Sexta Geração de Rede Móvel Celular (6G), traz consigo novos paradigmas arquiteturais para lidar com demandas crescentes por conectividade, eficiência e robustez.Uma das propostas em destaque é a arquitetura distribuída de múltiplas entradas e saídas, do inglês *Distributed Multi-input Multi-output* (D-MIMO). Nessa configuração, cada equipamento do usuário, do inglês *User Equipment* (UE), é atendido de maneira conjunta e coerente por múltiplos pontos de acesso, do inglês *Access Point* (AP), localizados na mesma região geográfica. Esses APs compartilham o mesmo recurso espectral e atuam simultaneamente, formando um *cluster* colaborativo.

A coordenação entre os APs é realizada por uma unidade central de processamento, do inglês *Central Processing Unit* (CPU), por meio de enlaces de *fronthaul*. Essa unidade é responsável pelo gerenciamento de recursos de rádio, do inglês *Radio Resource Management* (RRM), possibilitando um desempenho mais homogêneo ao longo da área de cobertura. Com isso, mitigam-se os efeitos negativos de borda típicos de arquiteturas celulares tradicionais, nos quais os UEs localizados nas extremidades das células sofrem com interferências e com a perda de percurso [136].

Para melhor atender às exigências individuais de tráfego de cada UE, foi proposto o uso de duplexação por divisão dinâmica do tempo, do inglês *Dynamic TDD* (D-TDD). Diferentemente do esquema convencional de duplexação por divisão estática do tempo, do inglês *Static TDD* (S-TDD), no qual a direção de transmissão de cada recurso temporal é a mesma para todo o sistema, o D-TDD adota uma abordagem centrada no usuário. Nele, cada UE tem sua própria direção de transmissão.

De acordo com [138], em sistemas S-TDD, o sinal recebido é composto por três elementos principais: sinal desejado, interferência (entre UE e AP, conforme o modo de transmissão), e ruído. No entanto, o uso de D-TDD introduz a chamada interferência cruzada, do inglês *Cross Link Interference* (CLI), que compromete o desempenho do sistema. A CLI inclui dois componentes: a interferência entre pontos de acesso, do inglês *Inter-AP Interference* (IAI), que afeta o enlace reverso, do inglês *Uplink* (UL), e é causada por APs operando em enlace direto, do inglês *Downlink* (DL); e a interferência entre usuários, do inglês *Inter-User Interference* (IUI), que ocorre no DL, devido à atividade de UEs transmitindo no UL. Essa situação está ilustrada na Fig. 43.

Dado esse cenário, torna-se necessário adotar técnicas de mitigação da CLI. Em [137], os enlaces de *fronthaul* de alta capacidade são explorados para suprimir e cancelar parcialmente a IAI. Além disso, [138] propõe um algoritmo de agrupamento que visa alinhar *clusters* próximos em um mesmo modo de transmissão, reduzindo a interferência mútua.

O conteúdo desta seção é uma adaptação do artigo submetido ao SBrT 2025: Matheus D. Carneiro, Igor M. Guerreiro, Yuri C. B. Silva, Samuel S. Silva, Victor F. Monteiro, Abraão de C. Albuquerque, "Energy Efficiency Analysis of Power Allocation on D-MIMO with Dynamic TDD".

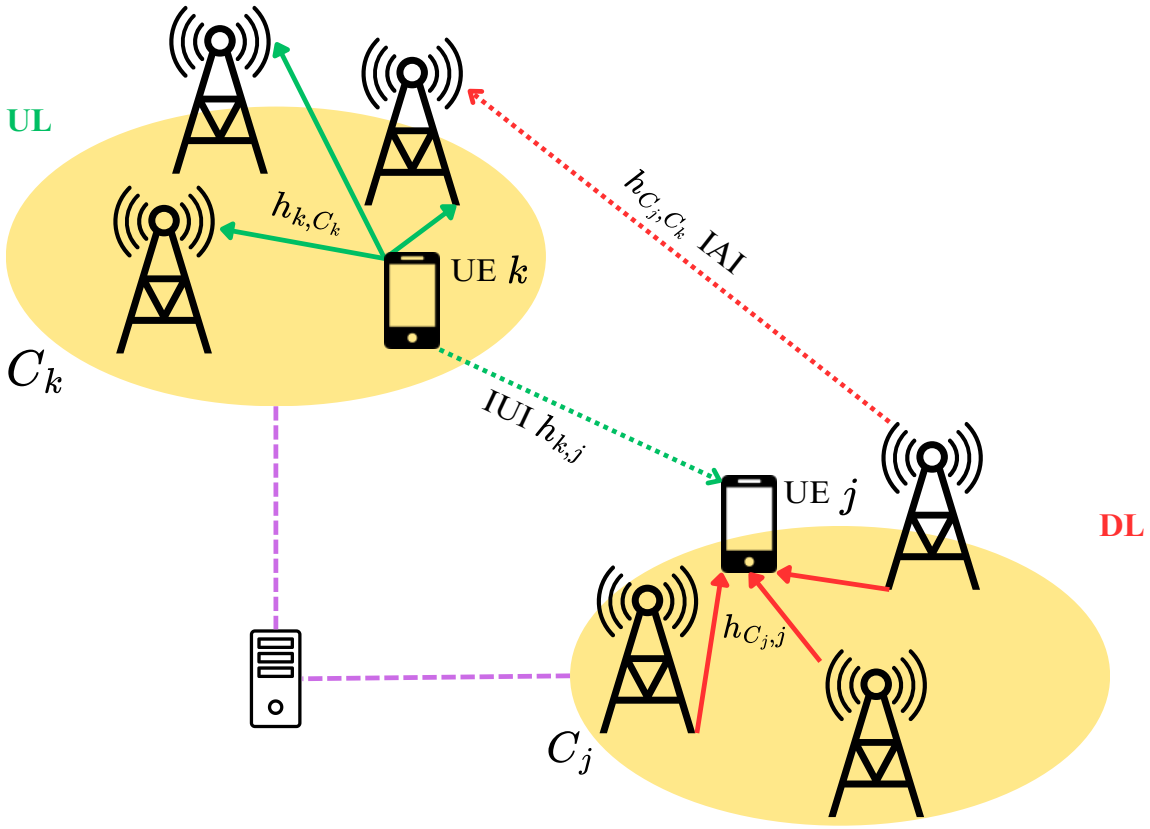


Figura 43: Arquitetura D-MIMO com D-TDD, na qual dois *clusters* disjuntos operando em modos de transmissão distintos sofrem interferência cruzada (CLI).

Outra técnica fundamental para lidar com interferências e, ao mesmo tempo, melhorar a eficiência energética, do inglês *Energy Efficiency* (EE), é o controle de potência. O estudo apresentado em [143] propõe uma otimização conjunta do controle de potência e do ponto de comutação da direção de transmissão para cada UE, utilizando aproximações convexas sucessivas, com o objetivo de maximizar a EE.

Entre as abordagens clássicas, destaca-se o controle de potência fracionário, do inglês *Fractional Power Control* (FPC), amplamente utilizado na literatura para compensar parcialmente os efeitos do desvanecimento em larga escala, do inglês *Large Scale Fading* (LSF). A aplicação do FPC em cenários com D-TDD é investigada em [144], e sua formulação é detalhada em [145].

Neste trabalho, analisa-se um cenário D-MIMO operando com D-TDD, conforme o modelo proposto em [138]. Implementa-se o algoritmo FPC com o objetivo de avaliar seu impacto sobre a eficiência espectral, do inglês *Spectral Efficiency* (SE), e a EE, considerando diferentes proporções de tráfego nos enlaces UL e DL.

Notação: Nesta seção, utiliza-se a seguinte convenção: letras minúsculas em negrito (por exemplo, $\mathbf{h}$) representam vetores; letras maiúsculas em negrito (por exemplo, $\mathbf{H}$), matrizes; e letras não negritadas em itálico (por exemplo, h), escalares. A notação $(\cdot)^T$ denota transposição e $|\cdot|$, o módulo de um número complexo. A expressão $\mathbf{h} \in \mathcal{A}^{N\times 1}$ indica que o vetor pertence ao conjunto $\mathcal{A}$ e possui dimensão $N \times 1$. Notação análoga aplica-se a escalares e matrizes. Os conjuntos dos números reais e complexos são representados por $\mathbb{R}$ e $\mathbb{C}$, respectivamente. A notação $\mathcal{N}(\mu, \sigma^2)$ refere-se a uma distribuição normal real com média μ e variância σ^2, enquanto

$\mathcal{CN}(\mu, \sigma^2)$ denota uma distribuição normal complexa circularmente simétrica. A notação $\mathcal{U}(a, b)$ representa uma distribuição uniforme contínua sobre o intervalo $[a, b]$.

15.2 Modelo do Sistema

Esta seção descreve o modelo de sistema adotado, abrangendo o modelo de propagação, o modelo de sinal e os *Key Performance Indicator*s (KPIs) considerados na avaliação.

Neste trabalho, é analisado um cenário D-MIMO composto por K UEs e M APs, todos com uma única antena. Os nós são distribuídos aleatoriamente de forma uniforme dentro de uma área quadrada de lado L, sendo os APs conectados a uma CPU por meio de enlaces de *fronthaul*. Considera-se que todos os UEs compartilham a mesma faixa de frequência, utilizada simultaneamente para transmissões no UL e no DL.

Conforme ilustrado na Fig. 43, cada UE k é atendido de forma conjunta e coerente por um cluster $\mathcal{C}_k$, composto por N APs selecionados dentre o conjunto total de M APs disponíveis na área. A seleção é baseada nos coeficientes de desvanecimento em larga escala $\beta_{k,m}$, sendo escolhidos os N APs com os maiores valores de $\beta_{k,m}$ para cada UE k. Essa associação é feita de modo que cada AP atenda exclusivamente a um único UE, resultando em clusters mutuamente disjuntos. Adota-se, ainda, uma configuração baseada em D-TDD, permitindo que a definição da direção de transmissão de cada UE seja feita de forma dinâmica de acordo com suas necessidades de tráfego [138].

Modelo de Propagação

Para fins de simplicidade, a perda de percurso representada por $PL_{k,m} \in \mathbb{R}$ (em dB), entre um nó k e um nó m, é modelada de forma independente do tipo de enlace (seja entre dois UEs, entre um UE e um AP, ou entre dois APs). A expressão da perda de percurso considera a distância $d_{k,m}$ entre os nós e segue o modelo descrito em [136]:

$$PL_{k,m} = -30.5 - 36.7 \log(d_{k,m} \text{ [em metros]}). \tag{193}$$

Já o efeito de sombreamento é modelado por uma variável aleatória $\eta_{sh} \sim \mathcal{N}(0, \sigma_{sh}^2) \in \mathbb{R}$, onde σ_{sh}^2 representa a variância do sombreamento. O LSF incorpora tanto a perda de percurso quanto o efeito de sombreamento, sendo expresso por:

$$\beta_{k,m} = 10^{(PL_{k,m} + \eta_{sh})/10}. \tag{194}$$

Adicionalmente, conforme o modelo apresentado em [141], cada enlace entre dois APs, ou entre um AP e um UE, pode estar sob condição de linha de visada, do inglês *Line of Sight* (LOS), ou sem linha de visada, do inglês *Non-Line of Sight* (NLOS). A probabilidade de ocorrência de um enlace LOS depende da distância d (em quilômetros) entre os nós e é dada por:

$$\text{Prob}_{\text{LOS}}(d) = 0.5 - \min\left(0.5, 5\exp\left(-0.156/d\right)\right) + \min\left(0.5, 5\exp\left(-d/0.03\right)\right). \tag{195}$$

Para cada enlace, é sorteado um limiar aleatório $U \sim \mathcal{U}(0, 1)$. Considera-se que o enlace está sob condição LOS sempre que $\text{Prob}_{\text{LOS}}(d) > U$. Nos enlaces LOS, aplica-se um fator de Rician médio $\overline{K}_{k,m} \sim \mathcal{N}(\mu_k, \sigma_k^2) \in \mathbb{R}$, conforme especificado em [140]. Caso contrário, para enlaces NLOS, o fator de Rician é definido como zero.

Adota-se, neste trabalho, o modelo de canal descrito em [139]. De acordo com essa referência, os coeficientes de canal entre pares de UEs são representados por $h_{k,k'} = \beta_{k,k'}\ w_{k,k'}$, onde

$w_{k,k'} \sim \mathcal{CN}(0,1) \in \mathbb{C}$ é uma variável complexa com distribuição Rayleigh, responsável por modelar o efeito de desvanecimento de pequena escala, do inglês *Small Scale Fading* (SSF).

Por outro lado, os enlaces entre APs e entre UE e AP apresentam coeficientes de canal com desvanecimento Rician, incorporando o fator $\overline{K}_{k,m}$ da seguinte forma:

$$h_{k,m} = \left[\sqrt{\frac{\overline{K}_{k,m}}{\overline{K}_{k,m}+1}} + \sqrt{\frac{1}{\overline{K}_{k,m}+1}} w_{k,m} \right] \beta_{k,m}. \tag{196}$$

Modelo de Sinal

A seguir descreve-se o modelo de sinal adotado para o UL e DL, bem como as expressões para a razão sinal-interferência-ruído, do inglês *Signal to Interference plus Noise Ratio* (SINR), e a SE, conforme apresentado em [138].

No modelo de sinal para o UL, considera-se um UE k transmitindo o sinal normalizado x_k para seu *cluster* $\mathcal{C}_k$, com potência de transmissão p_k e vetor de canal $\mathbf{h}_{\mathcal{C}_k,k} \in \mathbb{C}^{N\times 1}$, cujos elementos representam os coeficientes de canal entre o UE k e cada AP pertencente ao *cluster* $\mathcal{C}_k$. A mesma modelagem é aplicada para os UEs interferentes j também em UL.

Para interferências oriundas de APs no DL, o sinal normalizado transmitido é denotado por s_j, e $\mathbf{H}_{\mathcal{C}_k,\mathcal{C}_j} \in \mathbb{C}^{N\times N}$ representa a matriz de coeficientes de canal entre os APs dos *clusters* $\mathcal{C}_k$ e $\mathcal{C}_j$. A matriz $\mathbf{P}_{\mathcal{C}_j} \in \mathbb{R}^{N\times N}$ é diagonal e contém, em sua diagonal principal, a raiz quadrada da potência de transmissão de cada AP do *cluster* $\mathcal{C}_j$. O vetor $\mathbf{w}_j \in \mathbb{C}^{N\times 1}$ representa o vetor de pré-codificação de transmissão do *cluster* $\mathcal{C}_j$.

A variável L_j indica a direção de transmissão: 0 para UL e 1 para DL. O vetor $\mathbf{z}_k$ representa o ruído aditivo no UL. Assim, o sinal recebido no enlace reverso pelo *cluster* $\mathcal{C}_k$ associado ao UE k é dado por:

$$\mathbf{y}_k^{UL} = \underbrace{\sqrt{p_k}\ \mathbf{h}_{\mathcal{C}_k,k}\ x_k}_{\text{sinal desejado}} + \sum_{\substack{j=1 \\ j\neq k}}^{K} \Big(\underbrace{\sqrt{p_j}(1-L_j)\ \mathbf{h}_{\mathcal{C}_k,j}\ x_j}_{\text{interferência UE-para-AP}} + \underbrace{\mathbf{H}_{\mathcal{C}_k,\mathcal{C}_j}\ \mathbf{P}_{\mathcal{C}_j} L_j \mathbf{w}_j\ s_j}_{\text{interferência AP-para-AP}} \Big) + \underbrace{\mathbf{z}_k}_{\text{ruído}}.$$

A SE para o UE k em UL, considerando um vetor de recepção $\mathbf{u}_k \in \mathbb{C}^{N\times 1}$ no *cluster* $\mathcal{C}_k$, é dada por:

$$R_k^{UL} = \log_2 \left(1 + \frac{p_k \left| \mathbf{u}_k^T \mathbf{h}_{\mathcal{C}_k,k} \right|^2}{\gamma_k^{UL}} \right), \tag{197}$$

em que γ_k^{UL} representa a soma da interferência entre UEs e APs, da IAI entre APs, e do ruído térmico:

$$\gamma_k^{UL} = \sum_{\substack{j=1 \\ j\neq k}}^{K} p_j (1-L_j) \left| \mathbf{u}_k^T \mathbf{h}_{\mathcal{C}_k,j} \right|^2 + L_j \left| \mathbf{u}_k^T\ \mathbf{H}_{\mathcal{C}_k,\mathcal{C}_j} \mathbf{P}_{\mathcal{C}_j} \mathbf{w}_j \right|^2 + \sigma_{\text{UL}}^2, \tag{198}$$

onde σ_{UL}^2 representa a variância do ruído no UL.

A EE no UL é definida como a razão entre a soma das SEs dos UEs em UL e a soma das respectivas potências de transmissão:

$$EE^{UL} = \frac{\sum_{j=1}^{K} (1-L_j) R_j^{UL}}{\sum_{j=1}^{K} (1-L_j) p_j}. \tag{199}$$

Para o DL, o sinal recebido pelo UE k tem expressão similar à do UL, com notação correspondente ao contexto de transmissão dos APs:

$$y_k^{DL} = \underbrace{\mathbf{h}_{\mathcal{C}_k,k}^T \left(\mathbf{P}_{\mathcal{C}_k}\mathbf{w}_k\right) x_k}_{\text{sinal desejado}} + \sum_{\substack{j=1 \\ j\neq k}}^{K} \underbrace{\mathbf{h}_{\mathcal{C}_j,k}^T \left(\mathbf{P}_{\mathcal{C}_j} L_j \mathbf{w}_j\right) x_j}_{\text{interferência AP-para-UE}} + \underbrace{\sqrt{p_j}(1-L_j)h_{j,k}s_j}_{\text{interferência UE-para-UE}} + \underbrace{z_k}_{\text{ruído}}. \tag{200}$$

A SE do UE k no enlace direto é dada por:

$$R_k^{DL} = \log_2\left(1 + \frac{|\mathbf{h}_{\mathcal{C}_k,k}^T \left(\mathbf{P}_{\mathcal{C}_k}\mathbf{w}_k\right)|^2}{\gamma_k^{DL}}\right), \tag{201}$$

onde γ_k^{DL} representa a soma da interferência AP-para-UE, da IUI entre UEs, e do ruído:

$$\gamma_k^{DL} = \sum_{\substack{j=1 \\ j\neq k}}^{K} L_j \left|\mathbf{h}_{\mathcal{C}_j,k}^T \left(\mathbf{P}_{\mathcal{C}_j}\mathbf{w}_j\right)\right|^2 + p_j(1-L_j)\left|h_{k,j}\right|^2 + \sigma_{\text{DL}}^2. \tag{202}$$

A EE no DL é definida como a razão entre a soma das SEs dos UEs em DL e a soma das potências de transmissão dos APs correspondentes:

$$EE^{DL} = \frac{\sum_{j=1}^{K} L_j R_j^{DL}}{\sum_{j=1}^{K} L_j \text{tr}(\mathbf{P}_{\mathcal{C}_j}\mathbf{P}_{\mathcal{C}_j})}. \tag{203}$$

15.3 Algoritmo de Controle de Potência

Com o objetivo de mitigar os efeitos da CLI e melhorar tanto a SE quanto a EE, considera-se neste trabalho a implementação de uma abordagem amplamente explorada de alocação de potência. Esta seção apresenta o FPC e a estratégia de potência máxima, do inglês *Max Power* (MP). Ambas as abordagens são aplicadas aos UEs operando no UL.

Potência Máxima

A estratégia MP representa uma abordagem de referência direta, na qual cada UE transmite utilizando o máximo de potência disponível. Essa configuração é expressa por:

$$p_k = P_{\max}, \quad \forall\ k \in \{1, ..., K\} \text{ em modo UL.} \tag{204}$$

Essa abordagem busca maximizar o sinal transmitido por cada UE. No entanto, ela também resulta em um aumento da interferência no sistema e implica em um consumo excessivo de energia.

Controle de Potência Fracionário

Adota-se neste trabalho o esquema de FPC proposto em [139], o qual visa compensar parcialmente os efeitos da perda de percurso por meio de um fator de compensação $\alpha \in [0, 1]$ e de uma constante multiplicativa P_0. Este parâmetro P_0 é determinado empiricamente pela rede com o objetivo de normalizar os coeficientes de potência.

A potência de transmissão (em Watts) de um equipamento do usuário k é definida por:

$$p_k = \min(P_{\max},\, P_0\, \zeta_k^{-\alpha}), \tag{205}$$

em que $P_{\max}$ representa a potência máxima de transmissão do UE k, e ζ_k incorpora o LSF entre o UE k e seu cluster $\mathcal{C}_k$, conforme:

$$\zeta_k = \sqrt{\sum_{a \in \mathcal{C}_k} \beta_{k,a}}. \tag{206}$$

Diferentemente da formulação apresentada em [139], este trabalho não considera um cenário com múltiplas antenas. Dessa forma, o vetor de direcionamento de antenas, do inglês *array steering vector*, torna-se escalar, o que resulta diretamente na expressão da equação (206).

Ressalta-se que o FPC não leva em consideração os termos de interferência. Seu objetivo é garantir um desempenho arbitrariamente satisfatório para cada UE, com base em P_0, minimizando o consumo de potência. Caso o valor calculado ultrapasse o limite permitido, a potência é limitada pela estratégia de MP.

15.4 Resultados Numéricos

Esta seção apresenta os resultados obtidos por meio de simulações de Monte Carlo, nas quais o algoritmo de FPC é avaliado em um sistema D-MIMO operando sob o esquema D-TDD, em comparação com a abordagem de referência baseada em MP, aplicada a todos os UEs sob as mesmas condições.

Considera-se um cenário composto por $K = 30$ UEs e $M = 120$ APs distribuídos aleatoriamente em uma área quadrada de lado $L = 1{,}2$ km. A formação de clusters centrados no usuário é realizada com base no LSF, conforme descrito na Seção II, sendo adotado $N = M/K = 4$. Como vetores de pré-codificação, são utilizados o processamento de combinação com razão máxima, do inglês *Maximal Ratio Combining* (MRC), para recepção e transmissão com razão máxima para transmissão, do inglês *Maximal Ratio Transmission* (MRT).

Com relação à proporção de usuários em UL e DL, a porcentagem de usuários em UL é variada segundo o parâmetro $\rho_{UL} \in \{0,\ 0{,}25,\ 0{,}5,\ 0{,}75,\ 1\}$.

Para simplificar a modelagem, considera-se que cada realização da simulação representa um recurso temporal com sua própria configuração de transmissão. O modo de transmissão de cada UE, bem como o de seu respectivo *cluster*, é atribuído aleatoriamente a cada realização, respeitando o valor de ρ_{UL}. Os demais parâmetros relevantes para a simulação estão apresentados na Tabela 12.

Nas Figuras 44 e 45, é considerado um cenário em que o FPC é aplicado aos UEs em UL, variando-se o fator de compensação $\alpha \in \{0,\ 0{,}2,\ 0{,}4,\ 0{,}6,\ 0{,}8,\ 1\}$. O objetivo é identificar o valor mais adequado de α para este cenário. Também é avaliado um cenário de referência em que não se aplica controle de potência, adotando-se a MP. A constante multiplicativa P_0 é mantida fixa em -10 dBm em todas as simulações.

A Figura 44 apresenta a função de distribuição acumulada, do inglês *Cumulative Distribution Function* (CDF), da SE. Observa-se que, para os casos com $\alpha = 0{,}8$ e $\alpha = 1$, as curvas convergem para os resultados obtidos com MP. À medida que o valor de α diminui, a SE também se reduz, evidenciando menor desempenho espectral com maior compensação do desvanecimento.

Por outro lado, a Figura 45 exibe a CDF da EE. Nela, observa-se que $\alpha = 0{,}4$ apresenta melhor desempenho em termos de EE, quando comparado aos demais valores do fator de

Tabela 12: Parâmetros de simulação das estratégias de alocação de potência

Parâmetro	**Valor**
Altura dos APs	11,5 m
Altura dos UEs	1,5 m
Potência máxima no enlace UL	100 mW
Potência de transmissão de cada AP no DL	1 W
Largura de banda	10 MHz
Potência do ruído térmico	$\sigma^2_{UL} = \sigma^2_{DL} = -174$ dBm
Variância do sombreamento (σ^2_{sh})	4 dB
Fator de Rice (média, variância) (μ_k, σ^2_k)	9 e 5
Parâmetros do FPC	$P_0 = -10$ dBm, $\alpha = 0{,}4$

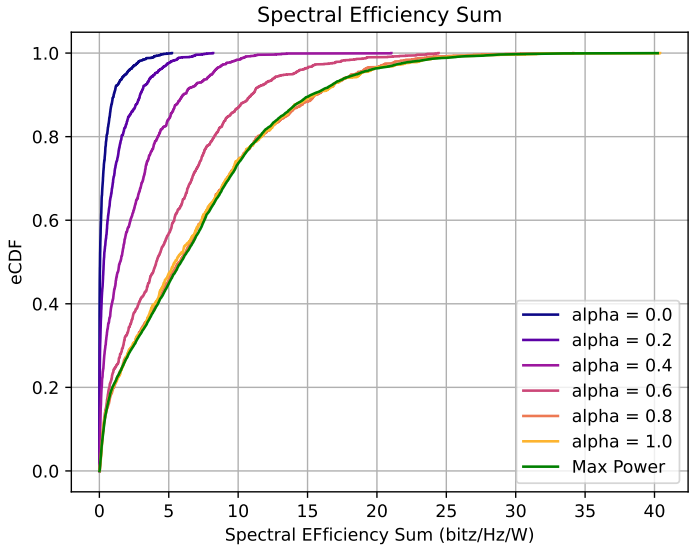


Figura 44: CDFs da SE para diferentes valores do parâmetro α do FPC.

compensação. Além disso, em termos de SE, esse valor está próximo dos melhores cenários, o que indica um equilíbrio satisfatório entre EE e SE.

Os valores $\alpha = 0{,}2$ e $\alpha = 0$ não são recomendados, pois apresentam valores elevados de EE apenas devido ao uso de níveis de potência próximos de zero. Isso compromete significativamente o desempenho espectral, tornando-os inadequados para aplicações práticas.

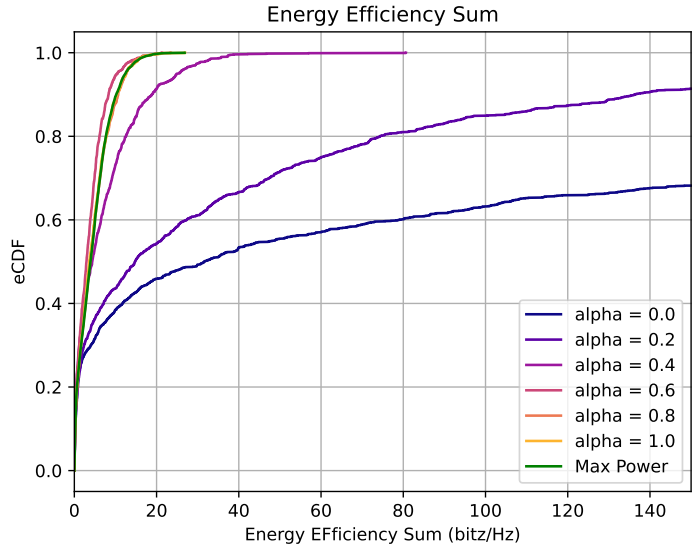


Figura 45: CDFs da EE para diferentes valores do parâmetro α do FPC.

A Figura 46 apresenta a CDF da SE para os seguintes cenários de configuração de rede: (i) aplicação da MP para todos os UEs no UL; e (ii) aplicação do FPC para todos os UEs em UL. Em ambos os casos, os APs em DL operam com potência fixa, e as curvas referentes ao DL representam os UEs nesse modo, em cada uma das configurações descritas.

Quando $\rho_{UL} = 0$ (ou seja, todos os UEs estão em DL), as curvas de SE são idênticas para os dois esquemas, pois tanto o FPC quanto o MP são estratégias de controle de potência específicas para o UL, e os APs em DL mantêm potência de transmissão fixa.

Para os casos com $\rho_{UL} = 0{,}25$ e $\rho_{UL} = 0{,}5$, os UEs em DL apresentam desempenho superior aos UEs em UL. No entanto, essa tendência se inverte em $\rho_{UL} = 0{,}75$, devido à redução da IAI, que afeta o UL, e ao aumento da IUI, que impacta o DL.

Com exceção do caso em que $\rho_{UL} = 0$, observa-se que os UEs em UL que utilizam FPC consistentemente apresentam valores de SE inferiores quando comparados àqueles que utilizam MP. Por outro lado, embora com impacto mais sutil, observa-se uma melhora no desempenho em DL quando o FPC é aplicado ao sistema, especialmente evidente no caso com $\rho_{UL} = 0{,}75$.

A Figura 47 apresenta a CDF da EE nos mesmos cenários de configuração de rede utilizados na Figura 46. De forma consistente, os UEs em DL apresentam menores valores de EE em comparação com os UEs em UL. Isso se deve, principalmente, à diferença significativa entre a potência de transmissão dos APs (1 W, conforme a Tabela 12) e a potência máxima de transmissão dos UEs em UL (100 mW).

À medida que a proporção de UEs em UL (ρ_{UL}) aumenta, observa-se um deslocamento das curvas de EE para a direita, indicando melhor desempenho energético. Nota-se, ainda, que o FPC proporciona ganhos consistentes de EE em relação ao MP, sendo esse ganho mais evidente à medida que ρ_{UL} aumenta.

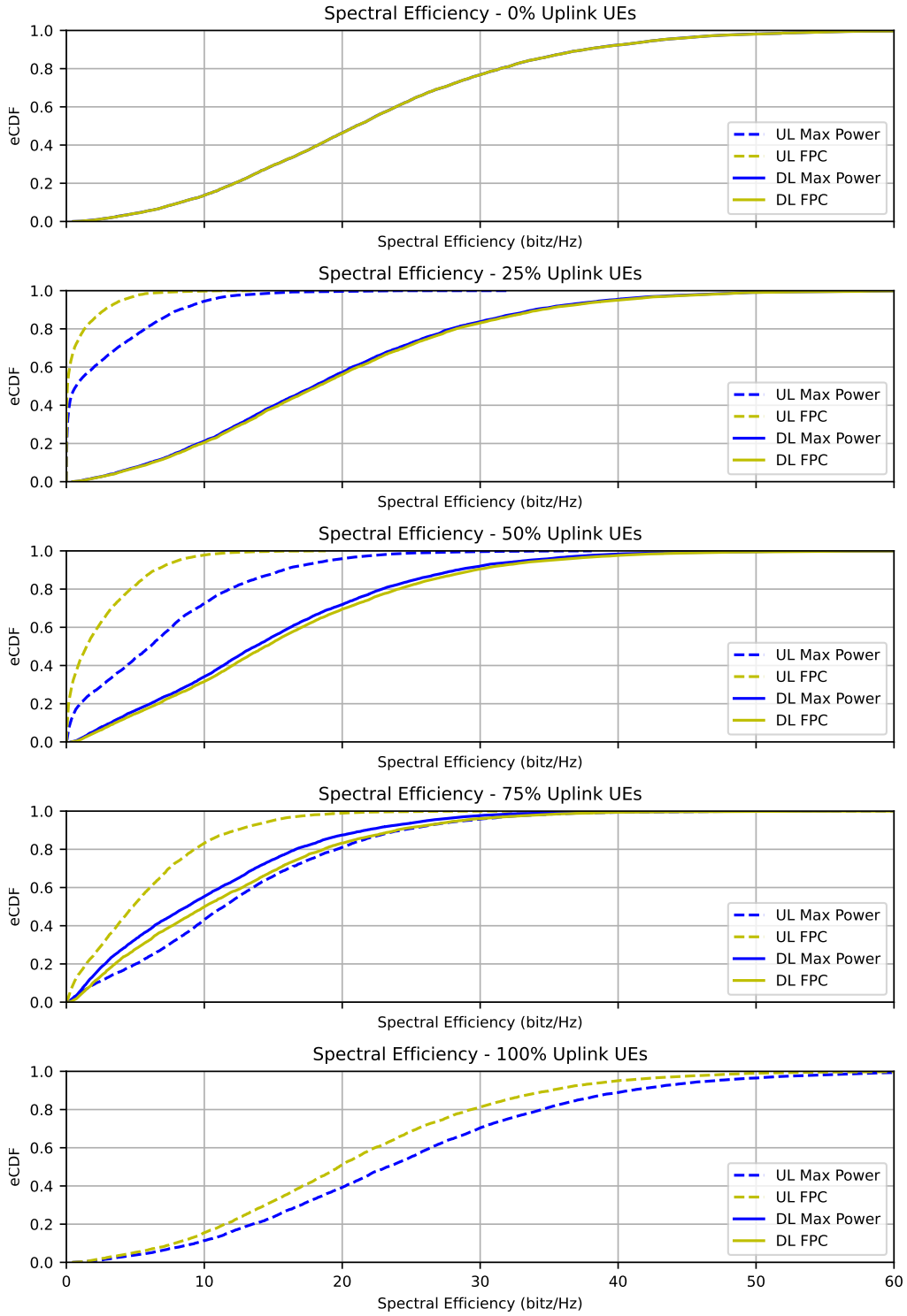


Figura 46: CDF da SE dos UEs em UL e DL variando ρ_{UL}.

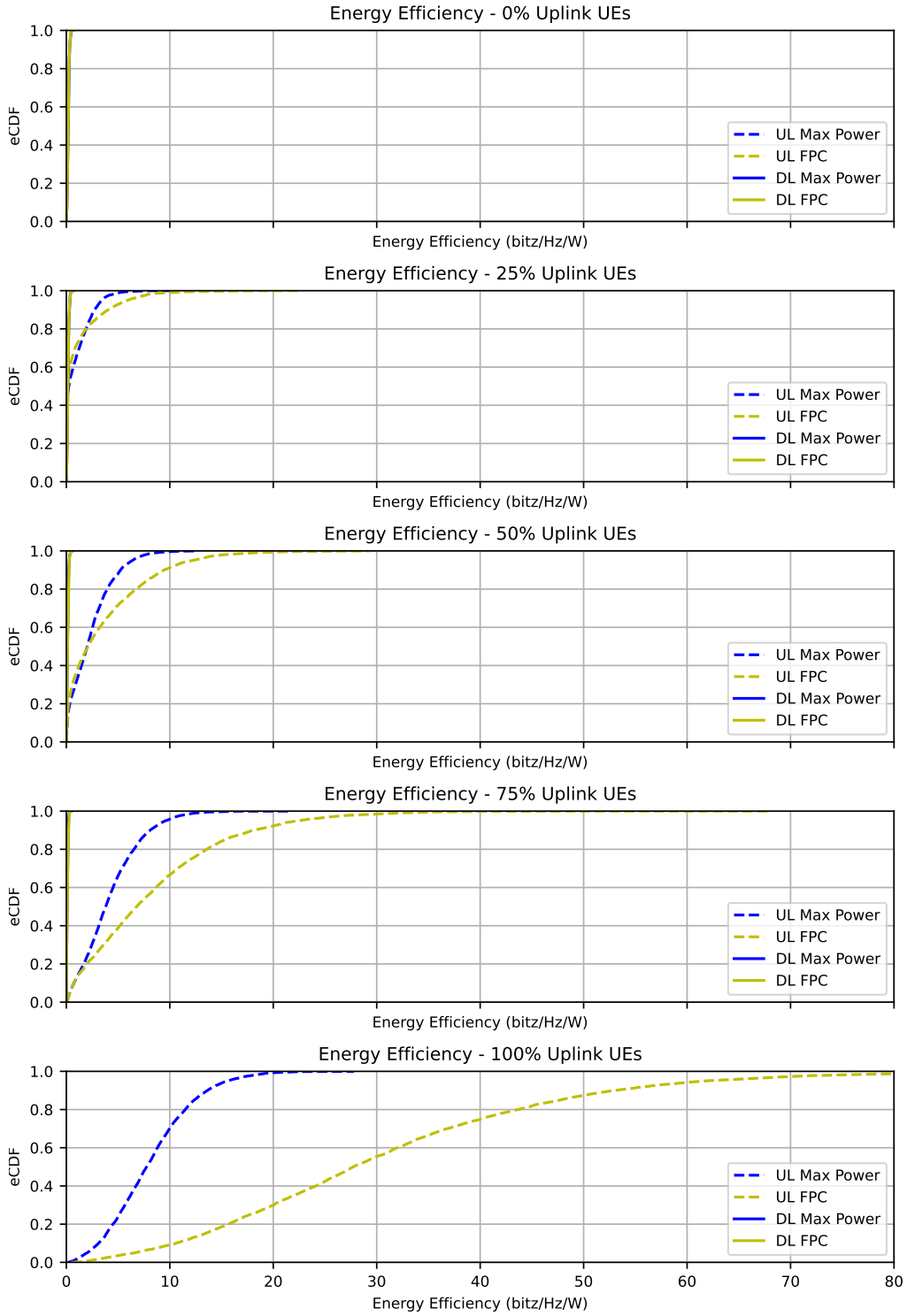


Figura 47: CDF da EE dos UEs em UL e DL variando ρ_{UL}.

15.5 Conclusão

Este trabalho teve como objetivo investigar a eficácia do FPC, uma abordagem clássica da literatura, na melhoria do desempenho de arquiteturas D-MIMO operando com D-TDD, especialmente diante dos desafios impostos pela CLI.

Considerou-se um modelo de canal no qual os coeficientes dos enlaces entre UEs seguem uma distribuição Rayleigh, enquanto os enlaces entre UEs e APs, bem como entre APs, seguem uma distribuição de Rice. O modelo de sinal incorporou explicitamente os efeitos de CLI, contemplando tanto a interferência entre APs, a IAI, quanto a interferência entre usuários, a IUI. A partir disso, foram formuladas expressões para a SE e a EE.

O algoritmo FPC foi descrito no contexto da arquitetura D-MIMO com D-TDD, e discutiram-se suas implicações no tratamento de interferências. Os resultados numéricos evidenciaram que a proporção de UEs em UL e DL influencia diretamente o desempenho dos usuários nos respectivos modos de operação. Em outras palavras, quando há predominância de usuários em UL, esses tendem a apresentar melhor desempenho em seus KPIs; o mesmo ocorre, de forma análoga, para o DL.

Além disso, observou-se que, embora o FPC apresente desempenho inferior ao da MP em termos de SE, ele supera o MP significativamente quando se avalia a EE. Esses resultados indicam que o FPC constitui uma alternativa viável para otimização do consumo energético em futuras redes de 6G, promovendo uma operação mais sustentável da rede e contribuindo para uma melhor experiência do usuário.

16 Métodos Baseados em NOMA

Fabbryccio Akkazzha Chaves Machado Cardoso e Francisco Raimundo Albuquerque Parente
Fundação CPqD

16.1 Introdução

O Non-Orthogonal Multiple Access (NOMA) [146][147] é uma técnica de acesso múltiplo que permite que múltiplos usuários compartilhem simultaneamente os mesmos recursos de alocação, como tempo e frequência, diferenciando-os por características como potência [148][149] e códigos de espalhamento [150]. Ao contrário dos esquemas ortogonais (OMA), como TDMA, WCDMA e OFDMA, onde cada usuário recebe um recurso exclusivo, o NOMA utiliza a superposição de sinais e técnicas avançadas de detecção, como o Successive Interference Cancellation (SIC), para separar e recuperar os dados transmitidos.

O Power-Domain NOMA (PD-NOMA) diferencia os usuários que compartilham o mesmo recurso de rádio por meio de níveis de potência distintos. A separação no receptor é realizada através do SIC, onde os sinais dos usuários com maior potência são decodificados primeiro e, posteriormente, subtraídos para recuperar os sinais de menor potência. Essa abordagem é conceitualmente simples e eficiente em termos de espectro quando há diferenças significativas de potência entre os usuários. No entanto, o PD-NOMA é sensível a erros de cancelamento de interferência e pode apresentar alta complexidade de decodificação em cenários mais densos.

Já o Code-Domain NOMA (CD-NOMA) diferencia os usuários por padrões de espalhamento específicos ou assinaturas únicas. A separação dos sinais no receptor é realizada por desespalhamento ou decodificação, oferecendo um ganho significativo de eficiência espectral, especialmente quando o número de usuários excede o comprimento dos códigos de espalhamento. Embora os métodos baseados em código possam suportar um maior número de usuários simultâneos, os algoritmos de detecção são mais complexos, podendo exigir otimização das assinaturas ou códigos.

16.2 Métodos baseados no PD-NOMA

Estratégias de alocação de potência no transmissor, coordenadas através de estações radio-base, definem as técnicas de Power-Domain NOMA (PD-NOMA) no que diz respeito à forma como a separação de usuários é feita. Idealmente, o PD-NOMA atribui diferentes níveis de potência com base nas condições de canal. Usuários com ganho de canal fraco recebem mais potência, para garantir a qualidade da transmissão. Por outro lado, usuários com ganho de canal forte recebem menos potência, evitando a saturação do receptor e minimizando a interferência excessiva.

No entanto, o levantamento constante dos estados de canal (Channel State Information - CSI) para todas as combinações possíveis de usuários que compartilham o mesmo recurso aumenta significativamente a complexidade computacional do sistema. Assim, a alocação de potência deve ser otimizada com base em critérios de desempenho, como vazão de dados e eficiência energética. Duas técnicas destacadas na literatura para otimização da alocação de potência são:

- Max-Min Fairness (MMF): Busca maximizar a menor taxa de transmissão entre os usuários, garantindo equidade no acesso aos recursos [148][151][152].

- Sum-Rate Maximization: Otimiza a soma das taxas de transmissão de todos os usuários, visando aumentar a eficiência espectral total [148][153][154].

No lado do receptor, a Successive Interference Cancellation (SIC) é a técnica empregada para cancelar a interferência dos sinais de alta potência antes de decodificar os sinais de baixa potência. No SIC, o receptor decodifica os sinais de forma iterativa, começando pelo usuário de maior potência ou de maior correlação em código [155] e removendo a interferência antes de passar ao próximo sinal. Este método é ideal quando o estado do canal é conhecido no receptor. Por outro lado, o SIC sensível a erros de decodificação nos estágios iniciais que podem se propagar e comprometer a recuperação dos sinais subsequentes. Esses erros de decodificação decorrem de um conhecimento imperfeito dos estados do canal, interferência residual e ruído pelo receptor [146]. Entretanto, vários estudos mostram que o emprego do SIC em sistemas NOMA é mais resiliente a erros de estimativa de canal do que os métodos tradicionais de detecção de sistemas OMA [146][156]. Também há trabalhos que propõem melhorias no SIC, como o uso de técnicas de cancelamento de interferência paralelo [146][156] e a combinação do SIC com técnicas de cancelamento de interferência multiusuário [146][156].

16.3 Métodos baseados no CD-NOMA

Há diversas linhas de pesquisa focadas no desenvolvimento de códigos customizados e assinaturas específicas para NOMA, com o objetivo de reduzir a complexidade de detecção e melhorar a robustez do sistema em diferentes cenários de comunicação. Essas abordagens se dividem em dois principais grupos: métodos baseados em códigos específicos e códigos customizados projetados exclusivamente para NOMA.

Os métodos baseados em códigos específicos utilizam códigos bem estabelecidos e adaptados para NOMA, nos quais cada usuário transmite seu sinal utilizando um código previamente definido. O receptor separa os sinais por meio de correlação cruzada, verificando qual código corresponde a cada usuário. Embora esses algoritmos de detecção possam ser complexos, eles são eficientes em cenários com interferência controlada e não exigem controle rigoroso de potência, já que os códigos ajudam a diferenciar os usuários. Contudo, o desempenho pode ser limitado pela correlação entre códigos, exigindo otimizações para minimizar a interferência. Entre esses métodos destacam-se:

- SCMA (Sparse Code Multiple Access): Utiliza alocação esparsa de símbolos em códigos únicos para cada usuário, mapeando os sinais em um espaço esparso para reduzir a interferência [157].
- PDMA (Pattern Division Multiple Access): Define padrões distintos de alocação de código e frequência, facilitando a separação dos usuários.
- MUSA (Multi-user Shared Access): Atribui assinaturas de código aleatórias para cada usuário, simplificando a implementação, mas aumentando a necessidade de algoritmos de detecção robustos.
- IGMA (Interleave-Grid Multiple Access): Implementa padrões de interleaving para diferenciar os usuários, garantindo baixa correlação entre os sinais transmitidos.
- CDMA (Code Division Multiple Access): Utiliza códigos ortogonais ou quase ortogonais para separar os usuários, sendo eficiente em cenários de baixa densidade de usuários, mas com limitações em ambientes mais densos devido à interferência residual.

Diferentemente dos códigos específicos, os códigos customizados são projetados para maximizar a eficiência do sistema NOMA. Eles exploram características matemáticas avançadas, como o limite de Welch (WB), que otimiza a separação dos usuários ao minimizar a interferência e melhorar a detecção no receptor. Esses códigos são ideais para cenários de alta densidade de usuários, proporcionando uma separação mais eficiente, embora exijam maior processamento para sua geração e alocação. Entre os principais métodos de códigos customizados, destacam-se os seguintes:

- WSMA (Welch-Bound Equality Spread Multiple Access): Técnica de espalhamento baseada no limite de Welch, que utiliza sequências de baixa correlação para separar usuários e minimizar a interferência em sistemas NOMA.
- LCRS (Low Code Rate Spreading): Método que usa códigos de baixa taxa de espalhamento para aumentar a densidade de usuários por recurso, melhorando a eficiência espectral em redes NOMA.
- NOCA (Non-Orthogonal Coded Access): Esquema de acesso múltiplo baseado em codificação não ortogonal, projetado para maximizar a separação de usuários e reduzir a interferência em sistemas de comunicação densa.
- UGMA (User Grouped Multiple Access): Técnica que agrupa usuários em subconjuntos e atribui códigos otimizados dentro de cada grupo, reduzindo a interferência intra-grupo para melhorar a robustez do sistema.

A principal motivação do NOMA é proporcionar maior eficiência espectral ao permitir que múltiplos usuários compartilhem os mesmos recursos simultaneamente. Essa característica torna o NOMA uma tecnologia promissora para redes densas e aplicações de Internet das Coisas (IoT), onde há um grande número de dispositivos competindo pelo espectro disponível. No entanto, essa abordagem traz consigo um custo de complexidade, principalmente na separação e decodificação dos sinais sobrepostos, o que pode resultar em maior latência, especialmente em sistemas que utilizam mecanismos de confirmação de recebimento e retransmissão de mensagens (ACK/NACK). Diante desse cenário, torna-se fundamental investigar técnicas que reduzam a complexidade do NOMA e aprimorem a eficiência do sistema.

16.4 Estratégias de integração sistêmica do NOMA

A implementação do NOMA enfrenta desafios significativos, como a alta complexidade computacional, a gestão de interferência e os elevados requisitos de processamento. Não há uma solução única que reduza essa complexidade de forma generalizada. Estudos indicam que a estratégia mais eficaz para mitigar esses desafios é explorar características específicas do sistema e da aplicação, priorizando aspectos como a coleta de informações para agrupamento de usuários e a otimização dos processos de retransmissão, de modo a equilibrar o desempenho e a complexidade do sistema.

Um exemplo crítico dessa complexidade está na estimativa dos estados de canal, utilizada para a alocação eficiente de usuários aos recursos compartilhados. A quantidade de informações necessárias para essa estimativa cresce proporcionalmente ao número de recursos disponíveis, ao número de usuários e às combinações possíveis de alocação. Em cenários de alta densidade, isso pode inviabilizar justamente as aplicações que mais se beneficiariam do NOMA, como as comunicações massivas no 6G. Portanto, o desenvolvimento de métodos mais eficientes para

gerenciamento de canal e alocação de recursos é prerequisito para viabilizar o NOMA em redes futuras de altíssima densidade.

Além disso, a retransmissão de mensagens representa outro desafio importante. Quanto maior o nível de compartilhamento de recursos, maior a probabilidade de falhas na decodificação, afetando todos os usuários que compartilham um mesmo canal. O aumento das retransmissões impacta negativamente a eficiência espectral e eleva a latência na entrega de pacotes. Se o sistema não for devidamente projetado, os ganhos iniciais de eficiência espectral proporcionados pelo NOMA podem ser anulados por falhas de decodificação. Isso pode inviabilizar aplicações sensíveis à latência, como o URLLC (Ultra-Reliable Low Latency Communications) no 5G e sua evolução para Hyper Reliable and Low-Latency Communication no 6G. Mesmo em cenários de IoT industrial, onde sensores e atuadores operam em malha fechada, a latência elevada pode comprometer a funcionalidade de sistemas críticos.

Diante desses desafios, diversas oportunidades de melhoria do NOMA estão sendo exploradas. A redução do custo computacional e do consumo de energia é um pré-requisito essencial para viabilizar o uso do NOMA em dispositivos IoT e sensores, que possuem restrições de processamento e energia. Isso permitiria que esses dispositivos adotassem o NOMA sem comprometer a vida útil da bateria, ao mesmo tempo em que a infraestrutura de rede se tornaria mais eficiente no gerenciamento da carga computacional em ambientes densamente conectados.

Soluções inovadoras têm sido propostas para lidar com esses desafios, incluindo pareamento simplificado de usuários e detecção iterativa de baixa complexidade, ambas voltadas para reduzir o processamento necessário no receptor, tanto em dispositivos quanto na estação radiobase. Essas técnicas prometem tornar o NOMA mais acessível e eficiente, consolidando seu papel nas redes móveis de próxima geração.

Nas aplicações URLLC, como em redes veiculares e automação industrial, a latência precisa ser mínima, e retransmissões excessivas devem ser evitadas. Para isso, customizações do protocolo HARQ (Hybrid Automatic Repeat Request) podem ser implementadas para ajustar dinamicamente as retransmissões e otimizar a recepção, reduzindo tanto a carga computacional do receptor quanto os atrasos na entrega de pacotes.

Já em redes densas, como aquelas presentes em cidades inteligentes ou eventos de grande porte, milhares de dispositivos podem disputar os mesmos recursos simultaneamente. Estratégias como agrupamento adaptativo de usuários e alocação eficiente de potência tornam possível escalar o NOMA sem aumentar excessivamente a complexidade. Isso garante um acesso eficiente para um grande número de dispositivos, além de melhorar a qualidade de serviço (QoS) e evitar a sobrecarga da rede.

Na busca por compatibilidade com sistemas de comunicação existentes, a implementação do NOMA não deve exigir mudanças radicais na infraestrutura das redes móveis atuais. Técnicas como o WSMA (Welch-bound Equality Spread Multiple Access) possibilitam a integração gradual do NOMA em redes 6G, com modificações mínimas nos receptores tradicionais de OMA, facilitando a adoção e reduzindo os custos de transição entre tecnologias OMA e NOMA.

Além disso, melhorar a confiabilidade e a qualidade do serviço (QoS) continua sendo uma prioridade. O NOMA tradicional pode ser vulnerável a erros no cancelamento de interferência (SIC), o que compromete a qualidade da transmissão e aumenta a taxa de falhas. Métodos de decodificação adaptativa surgem como uma alternativa eficaz, ajustando dinamicamente a estratégia de recepção conforme a qualidade do canal. Essa abordagem visa aumentar a taxa de sucesso na decodificação, reduzindo a perda de pacotes e melhorando significativamente a experiência do usuário final.

Os casos de uso mais promissores para o NOMA estão em ambientes de baixa mobilidade

e comunicação IoT. O NOMA mostra-se mais eficaz em cenários estáticos ou com pouca mobilidade, como fábricas automatizadas e redes de sensores. Nesses contextos, estratégias de pareamento eficiente de usuários e técnicas de alocação de potência de baixa complexidade otimizam a comunicação, reduzindo a sobrecarga de controle e evitando o desperdício de recursos, já que a mobilidade não é um fator crítico.

Ainda no contexto de métodos para reduzir a complexidade do NOMA, a literatura acadêmica apresenta diversas estratégias para mitigar esse desafio. Uma abordagem comum é focar na redução da complexidade de decodificação no receptor, onde o uso de técnicas de detecção iterativa com mínima interferência cruzada pode diminuir significativamente a carga computacional. Além disso, algoritmos de decodificação adaptativa são propostos para reduzir o número de operações de cancelamento de interferência (SIC), tornando o processo mais eficiente.

Outra linha de pesquisa explora a otimização do pareamento de usuários, utilizando técnicas como pareamento dinâmico de usuários e agrupamento de usuários com requisitos de canal similares, o que ajuda a reduzir a sobrecarga na coleta de informações do estado do canal (CSI - Channel State Information). Por exemplo, um usuário com canal ruim pode ser pareado com um usuário de canal bom para equilibrar a transmissão e melhorar a robustez.

No contexto de redes móveis, alguns estudos adaptam o uso do HARQ por meio de retransmissão seletiva baseada na condição do canal, o que reduz tanto a latência quanto o consumo energético. Estratégias mais avançadas propõem a adaptação do esquema de múltiplo acesso durante as retransmissões, minimizando o número de retransmissões necessárias e melhorando a eficiência espectral.

Ainda sobre redes móveis, trabalhos recentes investigam a operação sincronizada entre estações radiobase na RAN (Radio Access Network) para otimizar o uso do espectro. Essas operações sincronizadas podem melhorar a eficiência do NOMA, reduzindo os requisitos de equalização e simplificando o cancelamento de interferência. Além disso, a otimização dos recursos alocados para diferentes usuários contribui para reduzir a complexidade no gerenciamento de potência, garantindo melhor qualidade de serviço (QoS) com menor custo computacional.

A adaptação do NOMA para diferentes cenários de aplicação ainda é um campo aberto de pesquisa, com foco no desenvolvimento de métodos e algoritmos customizados para integração sistêmica do NOMA ou para sua combinação com outras tecnologias emergentes. Uma das combinações mais promissoras é o uso do NOMA em conjunto com o MU-MIMO (Multi-User Multiple-Input Multiple-Output).

O MU-MIMO, por sua natureza de suporte a múltiplos fluxos simultâneos em diferentes dimensões espaciais, complementa o NOMA ao proporcionar ganhos adicionais de eficiência espectral. A integração dessas duas tecnologias permite que mais usuários compartilhem o mesmo recurso de forma mais eficaz, diferenciando-os não apenas por potência (como no NOMA), mas também por direção de transmissão e espaço (como no MU-MIMO). Essa combinação é particularmente eficaz em ambientes de propagação complexos, como áreas urbanas densas, onde a dispersão do sinal é alta. Ao coordenar beamforming adaptativo e pareamento inteligente de usuários, o sistema pode otimizar o desempenho, reduzindo a interferência e aumentando a robustez da comunicação em cenários de alta densidade.

16.5 NOMA como uma tecnologia habilitadora para a IoT e o Agronegócio

A quinta geração (5G) das tecnologias de comunicação sem fio foi projetada para oferecer uma variedade de serviços amplamente classificados em três categorias: Banda Larga Móvel

Aprimorada (eMBB); Comunicação Ultraconfiável de Baixa Latência (URLLC); e Comunicação Massiva entre Máquinas (mMTC). Dentre essas categorias, o paradigma mMTC concentra-se no suporte a aplicações da Internet das Coisas (IoT) que toleram baixa taxa de transmissão de dados e atrasos, com alta densidade de conexões. Um dos padrões de IoT considerados para viabilizar o mMTC é o Narrowband IoT (NB-IoT), desenvolvido pelo projeto de parceria de terceira geração (3GPP). A versão atual do NB-IoT oferece suporte a redes IoT tradicionais. Devido aos requisitos de conectividade massiva do mMTC, o 3GPP está planejando aprimorar o padrão NB-IoT para atender dispositivos de baixo custo que demandam baixa taxa de transmissão de dados e baixo consumo de energia [158].

Os autores em [158] detalham um mecanismo de acesso utilizado no NB-IoT, denominado mecanismo de Acesso Aleatório (RA) em quatro etapas. Todo o processo após a transmissão do preâmbulo é baseado em concessão (grant-based). As desvantagens desse mecanismo baseado em concessão são mais acentuadas em cenários de comunicação massiva entre máquinas (mMTC), pois o número de usuários é elevado, tornando necessária a adoção de novos mecanismos de acesso à rede. Um desses mecanismos é o procedimento de acesso sem concessão (grant-free – GF), no qual as Unidades de Equipamento do Usuário (UEs) transmitem seus dados sem aguardar autorizações explícitas da Estação Base (BS), tornando o acesso baseado em contenção. Os esquemas GF são semelhantes aos esquemas de Acesso Aleatório (RA) utilizados para a transmissão de preâmbulo no enlace ascendente, porém envolvem tanto o preâmbulo RA quanto a transmissão de dados. O acesso GF para uma UE geralmente envolve a transmissão de um preâmbulo seguida do envio dos seus dados em uma única etapa, utilizando um conjunto predefinido de recursos. Especificamente, as UEs escolhem um preâmbulo de maneira aleatória e uniforme a partir de um conjunto disponível e selecionam uma sequência de espalhamento (ou assinatura) para a transmissão dos seus dados. Essa sequência pode ter uma relação de um-para-um ou de muitos-para-um em relação ao preâmbulo escolhido. O preâmbulo e a sequência de espalhamento (assinatura) identificam conjuntamente a UE. Quando essas sequências não são perfeitamente ortogonais entre si, o mecanismo resultante se torna um esquema de acesso múltiplo não ortogonal (NOMA), denominado GF-NOMA [158]. Neste estudo, os autores propuseram um mecanismo GF-NOMA compatível com o padrão NB-IoT existente e analisaram seu desempenho na camada PHY em termos de probabilidade de sucesso por usuário, latência e consumo de energia. A probabilidade de sucesso por usuário depende da detecção do pré-amble, decodificação de dados e recepção do ACK. Para isso, avaliaram a probabilidade de detecção e o desempenho da estimativa de canal (em termos de NMSE) para três sequências de pré-amble: 1) ZC, 2) RG e 3) Golay. A transmissão de dados no sistema GF-NOMA proposto foi baseada no canal de uplink (NPUSCH) do NB-IoT, com a única diferença de que utilizaram NOMA, onde os dados foram espalhados por meio da sequência de pré-amble. Compararam a capacidade de decodificação de dados do sistema GF-NOMA, considerando a presença ou ausência do SIC. Também foi proposta uma mensagem ACK compatível com o canal de downlink (NPDSCH) do NB-IoT, e demonstraram sua decodificação. Os resultados de desempenho de detecção e decodificação foram apresentados para dois cenários compatíveis com o NB-IoT, incluindo repetições e retransmissões. Além disso, calcularam a latência e o consumo de energia por usuário para esses cenários. Os autores concluíram que o mecanismo GF-NOMA proposto atende aos requisitos de mMTC da 3GPP, alcançando mais de 99% de probabilidade de sucesso por usuário, com redução no consumo de energia.

Em [159], é feito um estudo sobre comunicações por retroespalhamento habilitadas para NOMA. Devido aos benefícios incomparáveis da combinação das comunicações por retroespalhamento e NOMA, é importante investigar como as comunicações por retroespalhamento

habilitadas para NOMA podem evoluir para atender às necessidades das redes de IoT. Embora os estudos iniciais sobre comunicação por retroespalhamento tenham se limitado ao desenvolvimento e evolução da tecnologia de identificação por radiofrequência (RFID), trabalhos recentes têm se concentrado em melhorar o alcance da comunicação e a eficiência energética desses dispositivos. Nessas pesquisas, várias técnicas de processamento de sinal e modulação foram introduzidas, aumentando simultaneamente a confiabilidade e a segurança das etiquetas de retroespalhamento. No entanto, a avaliação da comunicação por retroespalhamento e das técnicas avançadas de acesso múltiplo não recebeu a mesma atenção. A combinação de NOMA e comunicação por retroespalhamento já é reconhecida como uma tecnologia eficiente em termos de energia, espectro e custo para redes massivas de IoT. Curiosamente, essa fusão traz benefícios duplos para os sistemas de comunicação convencionais. Quando uma estação base (BS) atende apenas os usuários celulares por meio de etiquetas de retroespalhamento no enlace ascendente, o sistema se transforma em um sistema de rádio simbiótico (SR). Em contraste, quando as etiquetas de retroespalhamento não estão se comunicando (ou estão colhendo energia), o sistema de comunicação se torna um sistema NOMA convencional. Além desses benefícios, o retroespalhamento habilitado para NOMA, que é inerentemente de baixa potência, pode ajudar a viabilizar a localização e mapeamento simultâneos (SLAM) para aplicações de mapeamento ambiental.

Diante dessa perspectiva, [159] fornece insights importantes sobre as tendências atuais e as pesquisas que levam a rápidas melhorias nas comunicações por retroespalhamento habilitadas para NOMA. Em seguida, é apresentada uma taxonomia detalhada que esclarece os aspectos menos desenvolvidos dessa tecnologia, além de vários casos de uso para comunicações por retroespalhamento habilitadas para NOMA. Isso é seguido por um estudo de caso sobre a maximização da taxa de soma em uma rede de comunicações por retroespalhamento que emprega NOMA como esquema de acesso múltiplo. Por fim, o trabalho destaca os diversos desafios e direções de pesquisa que podem contribuir para a viabilização da adoção das comunicações por retroespalhamento habilitadas para NOMA.

A comunicação de retrodispersão habilitada para NOMA marcará o início de uma nova era de conectividade perfeita para um grande número de dispositivos IoT. Neste contexto, o artigo apresentou importantes informações sobre os avanços recentes nessa área. Vários casos de uso e aplicações foram identificados, com base na natureza de baixa potência da comunicação de retrodispersão habilitada para NOMA. Também foi observada uma escassez significativa de estudos sobre comunicações de retrodispersão seguras utilizando técnicas NOMA. Além disso, foi apresentado um estudo de caso sobre dispositivos de detecção de retrodispersão habilitados para NOMA com alimentação sem fio, voltados para aplicações de IoT na agricultura inteligente. Embora o uso de técnicas de retrodispersão habilitadas para NOMA tenha mostrado grande potencial, diversos desafios impedem a plena realização dessas redes. Por exemplo, o gerenciamento da interferência em uma rede de retrodispersão NOMA de ultrabaixo consumo é complexo, pois as tags podem não ser capazes de se comunicar diretamente entre si, o que pode gerar alta interferência na rede. Além disso, o pequeno tamanho das fontes de energia desses dispositivos dificulta a implementação de técnicas de segurança no link. Espera-se que os insights fornecidos neste artigo sirvam como alicerces fundamentais para estudos futuros sobre comunicações de retrodispersão habilitadas por NOMA.

No conceito de agricultura inteligente, as Redes de Sensores Sem Fio (WSNs) desempenham um papel fundamental no monitoramento e coleta de dados de interesse nos campos agrícolas para diversas aplicações. Com base no modelo de crescimento das culturas e no uso de WSNs na agricultura, a agricultura inteligente busca viabilizar a irrigação, a aplicação de fertilizantes

e pesticidas de forma precisa, a fim de evitar o desperdício de água, a baixa fertilidade do solo, o uso excessivo de fertilizantes e o surgimento de doenças. Os nós sensores das WSNs, implantados nas áreas agrícolas, são capazes de transmitir periodicamente os dados coletados sobre o solo para o nó coletor por meio de comunicação sem fio [160].

Nas aplicações agrícolas das WSNs, a vida útil da rede depende principalmente do consumo de energia dos nós sensores. Diferentemente do uso de WSNs na indústria, onde os nós sensores são instalados em ambientes mais acessíveis e podem ter suas baterias recarregadas ou substituídas com facilidade a qualquer momento, nas aplicações agrícolas, esses sensores geralmente são implantados em campos agrícolas seguindo uma grade retangular e enterrados a uma profundidade de pelo menos 20 cm, dependendo exclusivamente das baterias durante toda a sua vida útil. Essa forma de implantação torna o consumo de energia um problema crítico, uma vez que a recarga ou substituição das baterias dos sensores é inviável no contexto agrícola. Dessa forma, os métodos de WSNs voltados para o monitoramento de áreas agrícolas na agricultura de precisão são categorizados e analisados com base em suas características de consumo energético [160].

Em [160], foi introduzido o uso de NOMA auxiliado por relay na transmissão de uplink em WSNs para aplicações na agricultura. Para transmissões periódicas e de rajada de curto alcance típicas de uplink, a técnica NOMA permite acomodar um maior número de nós sensores com o mesmo número de recursos de transmissão (REs). Foi proposto um esquema de transmissão híbrido, que inclui tanto uplink quanto downlink, para WSNs na agricultura, onde o OMA tradicional é utilizado na transmissão de downlink, e o esquema MONA auxiliado por relay é empregado na transmissão de uplink dos nós sensores para o nó sink. Com a conectividade em larga escala e a alta taxa de dados somada do NOMA, as WSNs que utilizam NOMA podem oferecer uma taxa de dados mais alta ao nó sink, diminuir o consumo de energia dos nós sensores e aumentar a vida útil das redes. O desempenho da taxa de dados total e da probabilidade de queda da aplicação do NOMA em uplink foi avaliado. Os resultados das simulações numéricas mostraram que as WSNs com NOMA auxiliado por relay apresentam uma menor probabilidade de queda e uma maior taxa de dados total, indicando que a técnica NOMA pode ser uma opção competitiva tanto para transmissões de uplink quanto de downlink em WSNs na agricultura.

16.6 Conclusão e Direcionamentos

O avanço das redes de comunicação para o 5G-Advanced e a transição para o 6G trazem novos desafios, especialmente relacionados à densificação de dispositivos e à otimização do uso espectral. Nesse contexto, o desenvolvimento e a aplicação da tecnologia NOMA emergem como soluções promissoras para atender às demandas de Massive Machine-Type Communications (mMTC) e futuras redes de Massive Communication no 6G.

Nossa linha de pesquisa e desenvolvimento busca implementar e avaliar o desempenho de algoritmos baseados inicialmente no Power Domain NOMA (PD-NOMA), proporcionando uma referência de desempenho para novas soluções NOMA. O cenário operacional escolhido concentra-se no mMTC, considerando a evolução para Massive Communication no 6G. Nesse cenário, planeja-se desenvolver algoritmos inovadores de alocação de potência e pareamento de usuários para otimizar a eficiência espectral e garantir a robustez dos sistemas.

Um dos pilares desta pesquisa é a integração de Inteligência Artificial (IA) e Machine Learning (ML) nos sistemas NOMA. Pretende-se empregar redes neurais profundas tanto no receptor quanto no transmissor. No receptor, redes neurais serão exploradas para criar mecanismos de Cancelamento Sucessivo de Interferência (SIC) inteligente, reduzindo a complexidade computa-

cional e melhorando a eficiência de detecção. No transmissor, as redes neurais serão aplicadas para otimizar o pareamento de usuários e a alocação de potência, buscando um equilíbrio entre desempenho e eficiência computacional. A adoção dessas técnicas visa não apenas reduzir a complexidade dos algoritmos convencionais, mas também melhorar o desempenho geral do sistema.

Além do PD-NOMA, a pesquisa inclui a exploração de mecanismos baseados no Code Domain NOMA (CD-NOMA), com o objetivo de permitir uma maior densificação de dispositivos IoT na rede de acesso. Mecanismos de codificação não ortogonais serão investigados para atender a requisitos específicos de conectividade massiva em redes futuras, como alta confiabilidade e baixa latência.

Outra vertente estratégica desta pesquisa é a integração do NOMA com tecnologias de MIMO, incluindo massive MIMO e MIMO 3D. Essa abordagem busca ampliar a eficiência espectral, melhorando a cobertura e a capacidade da rede, especialmente em cenários de alta densidade de dispositivos.

Em uma fase posterior, o foco será a integração de NOMA com tecnologias de transmissão em full duplex, permitindo comunicação simultânea no mesmo canal. Essa combinação promete aumentar significativamente a densidade de dispositivos suportados e a eficiência de uso do espectro, tornando-se uma solução candidata para o 6G.

Por meio dessas iniciativas, nossa pesquisa visa contribuir com soluções NOMA tecnicamente viáveis para a evolução das redes sem fio, propondo soluções práticas para os desafios de densificação massiva e definindo novos paradigmas para a tecnologia de acesso múltiplo não ortogonal.

17 Análise de Eficiência Espectral entre MIMO e NOMA no Domínio de Potência para Redes 6G OFDM

Fabbryccio Akkazzha Chaves Machado Cardoso e Maikon Renan Pereira da Silva
Fundação CPqD

Esta seção analisa a viabilidade de multiplexação não-ortogonal em domínio de potência (PD-NOMA) como complemento ao MIMO-OFDM em redes celulares 6G densas. Com base em um *framework* de código aberto desenvolvido sobre o Sionna 0.16, foram modelados sistemas MIMO 1×2/2×2/4×4 e pares de *streams* NOMA sobrepostos em *streams* MIMO sob perfis de canal CDL (TR 38.901), preservando a mesma eficiência espectral. As simulações mostram que o PD-NOMA duplica o número de *streams* (ou usuários) com sobrecarga baixa de E_b/N_o. A diversidade de recepção mitiga a interferência extra em cenários NLOS, enquanto ambientes LOS se beneficiam dos autovalores mais pronunciados para reduzir a cascata de erros. Além disso, quantifica-se o impacto da razão de potência entre os *streams* e da ordem MIMO sobre a BLER, fornecendo diretrizes práticas de ajuste de parâmetros para diferentes perfis de canal. Por fim, discutem-se limitações de interferência inter-*stream* após detecção MMSE e propõem-se métricas para estimar o *leakage* residual, apontando oportunidades de pesquisa em cancelamento de interferência avançado. Os resultados corroboram o PD-NOMA como estratégia de baixo custo para ampliar a capacidade espacial em rádios 6G, desde que o CSI seja suficientemente preciso e a decodificação iterativa permaneça viável em receptores comerciais.

17.1 Introdução

A crescente pressão sobre espectro e energia em redes móveis de próxima geração (5G-Advanced e 6G) [161] faz com que arquiteturas baseadas apenas em multiplexação ortogonal com MIMO puro atinjam rapidamente um ponto de saturação: adicionar mais *streams* espaciais exige cada vez mais antenas, aumenta a correlação entre canais e torna o processamento de sinais mais oneroso. A multiplexação não-ortogonal em domínio de potência (PD-NOMA) surge, assim, como complemento natural ao MIMO-OFDM [146, 148], permitindo a sobreposição controlada de duas camadas por *resource element* e a remoção da interferência via cancelamento sucessivo (SIC) [155]. Ao empregar superposição em vez de ortogonalidade, o PD-NOMA preserva a largura de banda e aloca potência de forma adaptativa, abrindo margem para ganhos de eficiência espectral [156] e flexibilidade de enlace.

Esta seção oferece quatro contribuições principais. Primeiro, apresenta um *framework* de simulação em código aberto, construído sobre o Sionna 0.16, capaz de modelar com precisão sistemas MIMO–OFDM–NOMA sob os perfis de canal CDL do 3GPP TR 38.901 [162], incluindo grade configurável de símbolos OFDM, tamanho da FFT e LDPC do 5GÑR. Segundo, realiza uma comparação exaustiva de desempenho entre MIMO $1 \times 2/2 \times 2/4 \times 4$ e PD-NOMA de duas camadas, mantendo a mesma eficiência espectral; demonstra-se que o PD-NOMA dobra potencialmente o número de *streams* (ou de usuários) com sobrecarga de apenas 0.5–1dB de SNR [153][154] quando $\alpha < 0.8$. Terceiro, quantifica o impacto da razão de potência (α, β) e da diversidade de recepção sobre o BLER, oferecendo *guidelines* práticos para ajuste de parâmetros em cenários NLOS e LOS. Quarto, discute limitações de interferência inter-stream após detecção MMSE e propõe métricas para avaliar o *leakage* residual [163] (híbrido SIC) e

[158] (combinações avançadas SIC-PIC) — informação essencial para projeto de receptores SIC escaláveis.

A adoção de PD-NOMA-MIMO-OFDM mostra-se especialmente vantajosa em contextos nos quais a multiplexação puramente ortogonal enfrenta limitações de eficiência espectral ou de custo. No *broadcast* de nova geração, por exemplo, o padrão ATSC 3.0 já incorpora o *Layered Division Multiplexing* (LDM) para sobrepor, num mesmo canal de 6 MHz, uma camada de alta potência destinada a receptores móveis robustos e outra de baixa potência voltada a transmissões UHD fixas [164]. Estudos de capacidade demonstram que essa abordagem supera sistematicamente esquemas baseados em TDM ou FDM [160], entregando taxas mais elevadas a ambos os serviços. Outro ambiente naturalmente propício é o dos sistemas MIMO sobrecarregados [165], em que o número de antenas da estação-base é inferior ao de usuários ativos. Nesses casos, algoritmos iterativos que combinam cancelamento sucessivo de interferência (SIC) e cancelamento paralelo (PIC), como o ISP [165][157], exploram a superposição de camadas PD-NOMA para recuperar boa parte do ganho de capacidade sem incorrer na cascata de erros típica do SIC convencional.

Redes celulares densas, sobretudo as arquiteturas *cell-free* [149] previstas para o 6G, também colhem benefícios diretos: ao dobrar o número de *streams* por portadora, o NOMA atende picos de tráfego e equilibra requisitos de QoS sem necessidade de expandir a grade OFDM, favorecendo enlaces de borda e *handovers* em alta mobilidade. Por fim, na Internet das Coisas massiva [159], dispositivos de baixa taxa podem ser alocados em uma camada secundária de potência reduzida, enquanto terminais de banda larga permanecem na camada primária; esse mapeamento facilita um agendamento unificado e aproveita os mecanismos de HARQ com maior eficiência. Essa abordagem pode ser utilizada caso o dispositivo tenha capacidade de processamento para implementar o SIC. Caso contrário, pode-se inverter a lógica e alocar o *stream* primário mais robusto e simples de decodificar.

Em redes celulares densas — sobretudo nas arquiteturas *cell-free* projetadas para o 6G — a superposição de camadas PD-NOMA emerge como solução elegante para elevar a capacidade sem ampliar a largura de banda. Ao possibilitar o dobro de *streams* por portadora, o esquema absorve picos de tráfego, ajusta dinamicamente requisitos de QoS e diminui a pressão sobre a grade OFDM, o que se traduz em *handovers* mais suaves e melhor atendimento aos usuários de borda. Em cenários de Internet das Coisas massiva, essa flexibilidade permite mapear dispositivos de baixo débito em uma camada secundária de potência reduzida, reservando a camada primária para terminais de banda larga; o resultado é um agendamento unificado que explora com maior eficiência os mecanismos de HARQ, mantendo a robustez do enlace. Convém salientar, contudo, que a sobreposição de camadas pressupõe que o terminal alocado na camada secundária disponha de poder de processamento suficiente para executar o cancelamento sucessivo de interferência (SIC). Nos casos em que tal capacidade é limitada, inverter a lógica preserva os benefícios do NOMA sem onerar equipamentos de baixo custo — isto é, atribuir o fluxo primário, mais robusto e simples de decodificar, ao dispositivo restrito e relegar a camada secundária ao terminal com maior poder computacional.

As evidências numéricas apresentadas nessa seção indicam que o PD-NOMA complementa o MIMO sempre que o ganho espacial se esgota ou torna-se economicamente inviável, oferecendo um caminho para aumentar capacidade e número de usuários com penalidade modesta de SNR e controle fino via alocação de potência [150]. Para sustentar essa conclusão, essa seção organiza-se da seguinte forma: a próxima subseção apresenta o modelo sistêmico detalhado, seguido da metodologia de simulação empregada; na sequência, discute-se os resultados comparativos entre MIMO-OFDM e PD-NOMA-MIMO-OFDM, explorando diferentes perfis de canal e razões de

potência; por fim, são analisadas as implicações práticas para o projeto de rádios 6G baseados em OFDM e sintetizamos as principais lições e oportunidades para trabalhos futuros.

17.2 Modelo Sistêmico

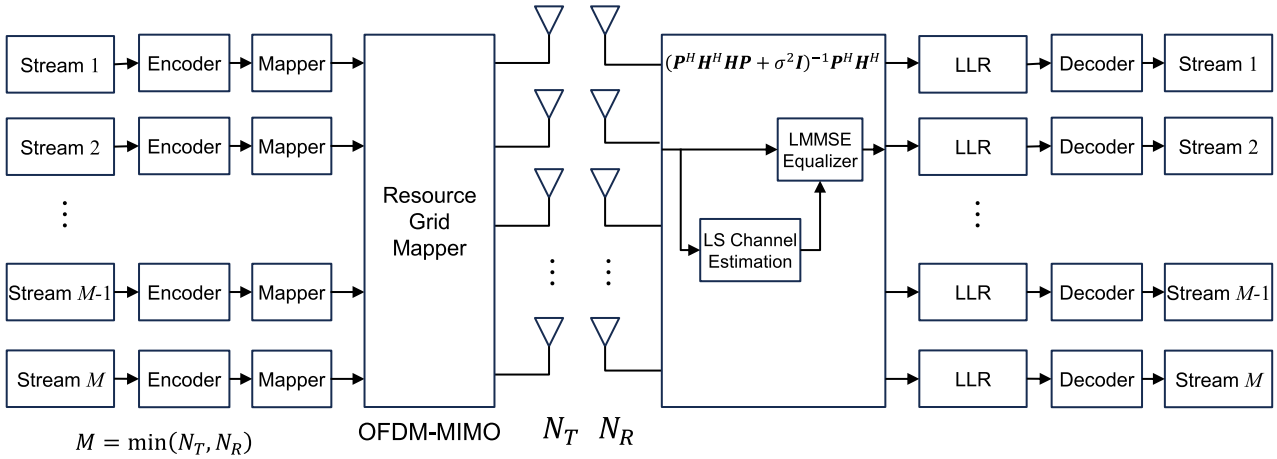


Figura 48: Modelo de um sistema OFDMA-MIMO.

Considere um sistema MIMO-OFDM como mostrado na Figura 48, operando em mobilidade tal que o tempo de coerência do canal T_c excede a duração do símbolo OFDM T_{sym}, isto é, $T_c \gg T_{\text{sym}}$. Em cenários de alta velocidade, isso equivale a garantir $f_D T_{\text{sym}} \ll 1$, suprimindo a interferência inter-portadora de origem Doppler. Além disso, o prefixo cíclico T_{CP} *é dimensionado* para ser maior que o atraso excessivo máximo $\tau_{\max}$, satisfazendo $\tau_{\max} < T_{\text{CP}}$. Como o espaçamento entre subportadoras Δf é selecionado de modo a ser menor do que a largura de banda de coerência B_c do canal, a resposta em frequência torna-se aproximadamente plana em cada subportadora, preservando a ortogonalidade dos símbolos OFDM. Nestas condições, a convolução no domínio do tempo reduz-se a multiplicações escalares independentes no domínio da frequência, permitindo a aplicação de algoritmos de detecção e equalização MIMO por subportadora. Sobre essa base, a multiplexação NOMA em *domínio de potência* acrescenta apenas a sobreposição de camadas de usuário ou de *streams* em cada subportadora, resolvida via cancelamento sucessivo de interferência (SIC) também formulado no domínio da frequência. Desse modo, as premissas adotadas possibilitam modelar o sistema MIMO-OFDM-NOMA diretamente na frequência, simplificando a análise, reduzindo a complexidade computacional e preservando toda a flexibilidade de alocação de potência característica do NOMA.

A Figura 49 mostra um modelo de transmissão NOMA-OFDMA-MIMO projetado para somar pares de *streams* em um mesmo recurso tempo-frequência de um *stream* MIMO, potencialmente dobrando a eficiência espectral de um MIMO convencional se esses *streams* forem processados com o mesmo esquema de modulação e codificação (MCS) através dos blocos funcionais Encoder e Mapper. A combinação dos *streams* é realizada no domínio da constelação de sinal com ganhos $\sqrt{\alpha}$ e $\sqrt{\beta}$ de modo que $\alpha+\beta = 1$ e $\alpha > \beta$, dividindo a potência de transmissão proporcionalmente. Essa estrutura visa aumentar a eficiência espectral de um mesmo usuário, dobrando o número de *streams* alocados para ele em cenários onde a ordem do MIMO não tem como ser aumentada.

No receptor da Figura 50, a detecção MIMO é realizada por subportadora k, incluindo estimação de canal e equalização para recuperar os *streams* MIMO. O algoritmo SIC é então

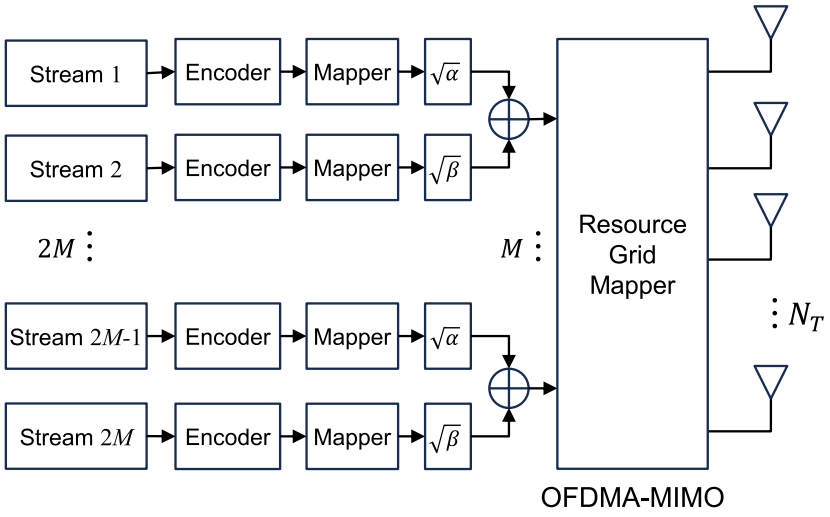


Figura 49: Modelo de transmissor para NOMA-OFDMA-MIMO.

aplicado aos *streams* MIMO equalizados. A primeira iteração do SIC é realizada para o *stream* NOMA principal de maior potência α, onde se calcula a LLR de bit e se realiza a decodificação suave dos bits. Na segunda iteração do SIC, os bits do *stream* NOMA principal são utilizados para regenerar o sinal do *stream* NOMA principal e então subtraí-lo dos *streams* MIMO, possibilitando decodificar o *stream* secundário de potência β através do cálculo da LLR de bit e da decodificação suave. No final da segunda iteração tem-se os bits transmitidos dos $2M$ *streams* NOMA de entrada. Embora a Figura 50 mostre como recuperar todos os *streams* NOMA transmitidos, na prática o receptor deverá decodificar apenas os *streams* alocados para o usuário.

Com base no discutido, os sinais transmitidos por antena em cada subportadora $\mathbf{y}_k^{(\text{ant})} \in \mathbb{C}^{N_T \times 1}$ nas N_T portas de antena estão relacionados aos M *streams* MIMO $\mathbf{x}_k \in \mathbb{C}^{M \times 1}$ pela matriz de *precoding* unitária $\mathbf{P}_k \in \mathbb{C}^{N_T \times M}$, onde $\mathbf{P}_k^{\text{H}}\mathbf{P}_k = \mathbf{I}_M$, de modo que

$$\mathbf{y}_k^{(\text{ant})} = \mathbf{P}_k \mathbf{x}_k. \tag{207}$$

Por outro lado, os $2M$ *streams* NOMA são divididos em M *streams* primários $\mathbf{s}_k^{\text{P}} \in \mathbb{C}^{M \times 1}$ e M secundários $\mathbf{s}_k^{\text{S}} \in \mathbb{C}^{M \times 1}$. Desta forma, a relação entre os M *streams* MIMO e os $2M$ *streams* NOMA é dada por

$$\mathbf{x}_k = \sqrt{\alpha}\mathbf{s}_k^{\text{P}} + \sqrt{\beta}\mathbf{s}_k^{\text{S}}, \tag{208}$$

em que α e β são definidos de modo a otimizar o processo de detecção SIC, sendo $\alpha > \beta$ e $\alpha + \beta = 1$, resultando que os *streams* primários possuem potência α e os secundários β.

Os N_T sinais $\mathbf{y}_k^{(\text{ant})}$ são transmitidos por subportadora k através de uma matriz de canal MIMO $\mathbf{H}_k \in \mathbb{C}^{N_R \times N_T}$, resultando em N_R sinais $\mathbf{z}_k^{(\text{ant})} \in \mathbb{C}^{N_R \times 1}$ por antena no lado do receptor, dado por

$$\mathbf{z}_k^{(\text{ant})} = \mathbf{H}_k \mathbf{y}_k^{(\text{ant})} + \mathbf{n}_k, \tag{209}$$

em que $\mathbf{n}_k \in \mathbb{C}^{N_R \times 1}$ representa os ruídos AWGN na cadeia de recepção por antena, associados a cada subportadora OFDM. Entretanto, pode-se considerar o canal efetivo diretamente no espaço de *streams* MIMO para se recuperar os sinais $\mathbf{z}_k^{(\text{ant})}$ dado por

$$\mathbf{z}_k^{(\text{ant})} = \mathbf{H}_k \mathbf{P}_k \mathbf{x}_k + \mathbf{n}_k. \tag{210}$$

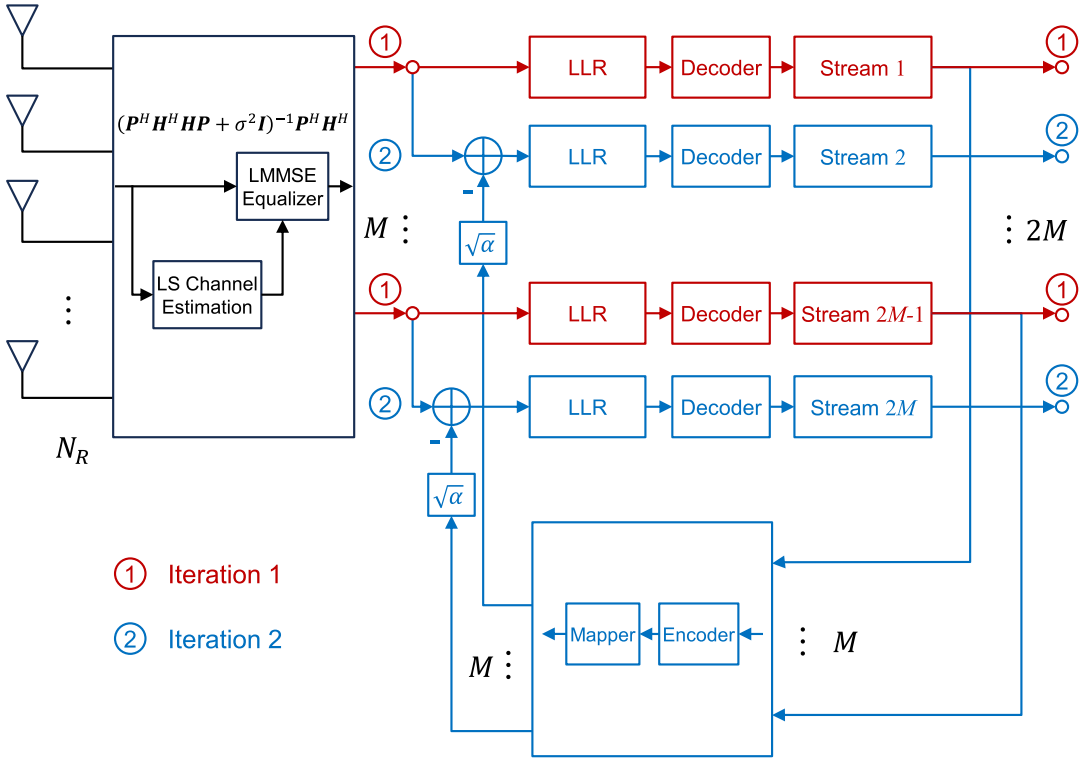


Figura 50: Modelo de receptor para NOMA-OFDMA-MIMO.

A equalização dos sinais é realizada no espaço dos *streams* MIMO através de estimativa da matriz de canal MIMO efetiva $\mathbf{H}_k\mathbf{P}_k$ com auxílio de subportadoras piloto e interpolação para subportadora de dados. No caso de equalização MMSE linear, obtém-se uma estimativa dos sinais por *stream* MIMO, dada por

$$\hat{\mathbf{x}}_k = \left(\mathbf{P}_k^{\mathrm{H}}\hat{\mathbf{H}}_k^{\mathrm{H}}\hat{\mathbf{H}}_k\mathbf{P}_k + \sigma^2\mathbf{I}_M\right)^{-1}\mathbf{P}_k^{\mathrm{H}}\hat{\mathbf{H}}_k^{\mathrm{H}}\mathbf{z}_k^{(\mathrm{ant})}, \tag{211}$$

em que σ^2 é a variância estimada do ruído. Note que se o número de portas de antena de transmissão for igual ao número de *streams* MIMO, $N_T = M$, então a matriz de precodificação degenera para a identidade $\mathbf{P}_k = \mathbf{I}_M$.

Além disso, dada a composição dos *streams* MIMO pelos *streams* NOMA dada por

$$\hat{\mathbf{x}}_k = \sqrt{\alpha}\hat{\mathbf{s}}_k^{\mathrm{P}} + \sqrt{\beta}\hat{\mathbf{s}}_k^{\mathrm{S}} + \tilde{\mathbf{n}}_k, \tag{212}$$

pode-se recuperar os *streams* primários $\mathbf{s}_k^{\mathrm{P}}$ e secundários $\mathbf{s}_k^{\mathrm{S}}$ do NOMA empregando SIC em duas iterações a partir da estimativa dos *streams* MIMO $\hat{\mathbf{x}}_k$:

1. Os bits de informação dos *streams* primários $\hat{\mathbf{s}}_k^{\mathrm{P}}$ são obtidos pela decodificação direta de $\hat{\mathbf{x}}_k$ considerando os *streams* NOMA secundários como ruído, ou seja,

$$\hat{\mathbf{x}}_k = \sqrt{\alpha}\hat{\mathbf{s}}_k^{\mathrm{P}} + \tilde{\mathbf{n}}_k^{\mathrm{P}}, \tag{213}$$

em que $\tilde{\mathbf{n}}_k^{\mathrm{P}} = \sqrt{\beta}\hat{\mathbf{s}}_k^{\mathrm{S}} + \tilde{\mathbf{n}}_k$. Neste caso, a decodificação envolve o cálculo da LLR e do Decoder.

2. Os bits de informação dos *streams* secundários $\mathbf{s}_k^{\mathrm{S}}$ são obtidos pela decodificação de

$$\hat{\mathbf{x}}_k - \sqrt{\alpha}\tilde{\mathbf{s}}_k^{\mathrm{P}} = \sqrt{\beta}\hat{\mathbf{s}}_k^{\mathrm{S}} + \tilde{\mathbf{n}}_k^{\mathrm{S}}, \tag{214}$$

em que $\tilde{\mathbf{s}}_k^{\mathrm{P}}$ correspondem aos *streams* NOMA primários regenerados a partir dos bits de informação do passo 1, sendo que o ruído resultante é dado por

$$\tilde{\mathbf{n}}_k^{\mathrm{S}} = \sqrt{\beta}\left(\hat{\mathbf{s}}_k^{\mathrm{P}} - \tilde{\mathbf{s}}_k^{\mathrm{P}}\right) + \tilde{\mathbf{n}}_k. \tag{215}$$

Essa abordagem de SIC considera relevantes apenas as interferências entre o par primário–secundário de cada recurso NOMA, pressupondo que a detecção MMSE dos *streams* MIMO resolve as interferências entre *streams*. Na prática, qualquer erro de CSI ou correlação espacial torna

$$\mathbf{P}_k^{\mathrm{H}}\hat{\mathbf{H}}_k^{\mathrm{H}}\hat{\mathbf{H}}_k\mathbf{P}_k \tag{216}$$

não diagonal, gerando componentes fora da diagonal que o MMSE atenua apenas parcialmente. Surge, então, interferência cruzada entre *streams* MIMO, limitando o desempenho do NOMA real, que cancela apenas a interferência **intra-NOMA** (primário × secundário). O termo de erro $\hat{\mathbf{s}}_k^{\mathrm{P}} - \tilde{\mathbf{s}}_k^{\mathrm{P}}$ evidencia que, quanto maior o vazamento inter-stream, maior será essa parcela — e menor o ganho do SIC sobre o *stream* secundário.

17.3 Resultados de Simulação

No 3GPP TR 38.901, há cinco perfis de *Clustered Delay Line* (CDL) estabelecidos para avaliações de enlace em frequências de 500 MHz a 100 GHz, conforme descritos pela Tabela 13. Os três primeiros perfis (A, B, C) cobrem situações *NLOS* de complexidade crescente, enquanto os dois últimos (D, E) cobrem cenários *LOS* com diferentes larguras de feixe e relações K de Rice. Todas essas definições e seus parâmetros detalhados (atrasos normalizados, potências, ângulos AOD/AOA/ZOD/ZOA, *spreads* etc.) encontram-se na Seção 7.7.1 do relatório [162].

Tabela 13: Modelos de canal CDL do 3GPP TR 38.901.

Modelo	**Descrição**
CDL-A	Representa um canal *NLOS* com dispersão de atraso curta/mediana.
CDL-B	Canal *NLOS* com dispersão de atraso intermediária.
CDL-C	Canal *NLOS* com dispersão de atraso longa, usado para cenários urbanos ou *indoor* densos.
CDL-D	Perfil *LOS*, contempla presença de trajetória direta mais componente especular pronunciada.
CDL-E	Outro perfil *LOS*, com dispersão angular e de atraso maior que o CDL-D, adequado a ambientes abertos com múltiplas reflexões.

As avaliações em nível de enlace foram conduzidas com um simulador próprio desenvolvido sobre o *framework Sionna 0.16*, cuja classe central `OFDMSystem` encapsula transmissor e receptor OFDM conforme modelo descrito na Seção 17.2. Salvo indicação em contrário, todas as simulações utilizam um quadro OFDM com 14 símbolos, FFT de 76 pontos, prefixo cíclico de 6 amostras e espaçamento de subportadora de 30 kHz, em conformidade com a numerologia 0 do

5G NR. Os códigos LDPC seguem o *base-graph* 2 do 5G NR com taxa de codificação $r = 1/2$ para todas as modulações analisadas, possibilitando que a comparação isole a influência do multiplexação espacial e da alocação de potência do NOMA sobre a taxa de erro de bloco (BLER).

A variância de ruído é obtida a partir do alvo E_b/N_0 via

$$N_o = \frac{E_s}{N_{\text{biteff}}} \cdot \frac{1}{E_b/N_o}, \tag{217}$$

em que E_s denota a energia total transmitida por RE (Resource Element - subportadora durante um símbolo OFDM) que é mantida unitária ($E_s = 1$) no simulador; e N_{biteff} é o número efetivo de bits de informação por subportadora de dados, dado por

$$N_{\text{biteff}} = \frac{K_{\text{data}}}{K_{\text{total}}} \cdot \frac{N_{\text{FFT}}}{N_{\text{FFT}} + N_{\text{CP}}} \sum_{l=0}^{L-1} \left(M_l^{\text{mod}} \, r_l \right), \tag{218}$$

que considera (i) a fração de subportadoras de dados no *grid* de recursos, (ii) a sobrecarga do prefixo cíclico e (iii) o número total de bits de informação em uma subportadora, resultante da composição dos L *streams* NOMA e MIMO, onde cada *stream* l possui taxa de codificação de canal r_l e modulação com M_l^{mod} bits na constelação de sinal. Há dois *streams* NOMA para cada *stream* MIMO, sendo M *streams* MIMO, totalizando $L = 2M$ *streams*.

Os parâmetros considerados são: o número total de subportadoras K_{total} no *grid* de recursos OFDM (slot), o número de subportadoras de dados K_{data} nesse mesmo *grid*, descontadas subportadoras piloto e de guarda, o comprimento da *FFT* N_{FFT} (número de subportadoras em um símbolo OFDM), o comprimento do prefixo cíclico N_{CP} em amostras, e o número total de amostras $N_{\text{CP}} + N_{\text{FFT}}$ por símbolo OFDM.

Nas simulações, o *grid* de recursos foi configurado como mostrado pela Figura 51, destacando a composição de 14 símbolos OFDM com 76 subportadoras cada, sendo dois símbolos dedicados para subportadoras piloto, onde pilotos são alocados de forma ortogonal entre *streams* MIMO.

As simulações utilizam o modelo de canal `sn.channel.tr38901.CDL` do Sionna, onde o desvanecimento segue os perfis *Clustered Delay Line* (CDL) definidos pelo 3GPP TR 38.901, conforme descrito pela Tabela 13, que resume suas características principais. O *Doppler* corresponde a um terminal móvel a 10 m/s numa portadora de 2.6 GHz. O canal é amostrado com período igual à duração do símbolo OFDM, de modo a seguir as premissas estabelecidas na Seção 17.2, que possibilitam modelar o sistema MIMO-OFDM-NOMA diretamente na frequência.

Efeito dos parâmetros de potência α e β

Os efeitos da divisão de potência entre os *streams* primário e secundário do NOMA são demonstrados pela Figura 52, com o modelo de canal CDL-B — *NLOS* com Atraso Intermediário e Sensibilidade à Potência. As simulações consideram um único *stream* MIMO e diversidade de recepção com duas antenas, constituindo um sistema MIMO 1×2. São utilizados como referência para a análise o desempenho do sistema MIMO OMA 1×2 com modulação QPSK e 16QAM. O sistema NOMA utiliza *stream* primário (PS) e secundário (SS) sobrepostos no *stream* MIMO OMA 1×2 com modulação QPSK cada, de modo que sua eficiência espectral é equivalente à do MIMO 1×2 com modulação 16QAM.

A análise considera como ponto de partida o MIMO OMA com a mesma eficiência espectral que o NOMA, ou seja, o MIMO 1×2 16QAM. Neste caso, o NOMA pode ser interpretado como

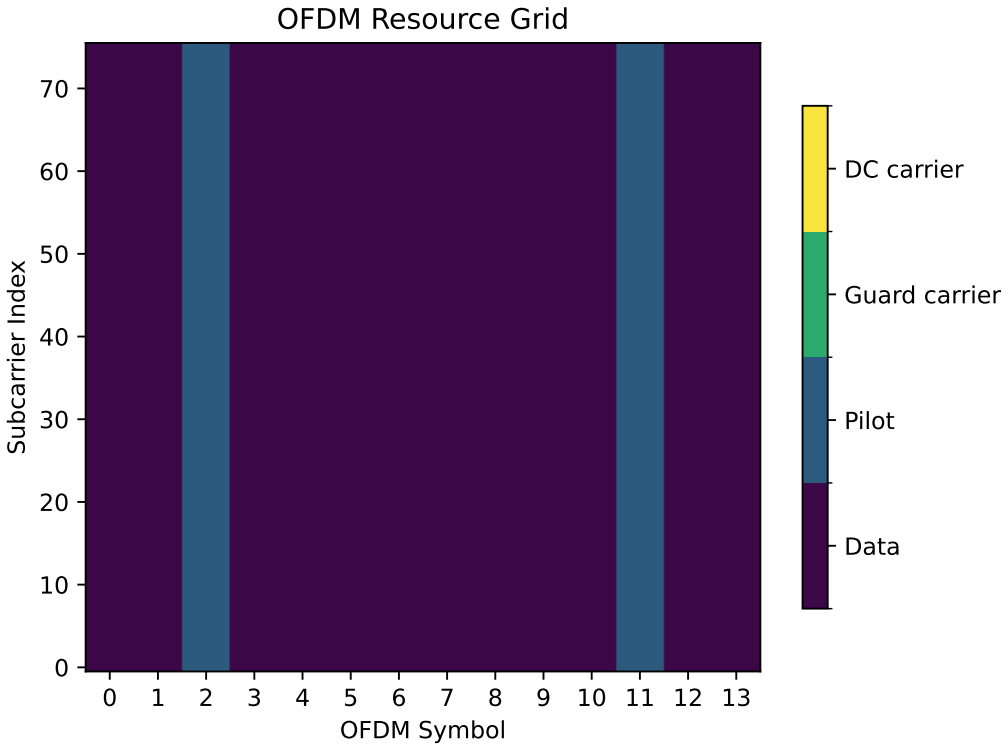


Figura 51: Grid de recursos OFDM para alocação dos *streams* MIMO e NOMA

sendo a divisão do *stream* MIMO 16QAM em dois *streams* QPSK sobrepostos, onde o *stream* PS tem desempenho superior, estando sua curva de BLER mais à esquerda que a do MIMO 16QAM em termos de E_b/N_o, e o *stream* SS com desempenho de BLER inferior com curva de desempenho mais à direita do MIMO 16QAM. O aumento relativo do parâmetro de potência α aumenta o espaçamento em termos de E_b/N_o entre as curvas dos *streams* PS e SS, com o MIMO 16QAM permanecendo entre elas. Para $\alpha = 0.7$, o desempenho de BLER dos *streams* PS e SS estão próximos do desempenho do MIMO 16QAM. Ao aumentar α para 0,75, o desempenho do PS se aproximou consideravelmente do MIMO QPSK e o do SS se afastou em torno de 1 dB do desempenho do MIMO 16QAM. Nos casos de $\alpha = 0{,}8$ e 0,9, o desempenho do *stream* PS supera o do MIMO QPSK, mas com uma piora de desempenho equivalente do *stream* SS.

Esses resultados da Figura 52 destacam o benefício do NOMA proposto em aumentar (dobrar) o número de *streams* disponíveis no sistema comparado com o MIMO de mesma eficiência espectral. O aumento no número de *streams* se traduz em suportar um maior número de usuários compartilhando o mesmo *grid* de recursos OFDM. Há ainda maiores oportunidades e flexibilidade para configurar transmissões associadas ao processo HARQ em caso de *NACK*. Além disso, pode-se ajustar o parâmetro α para configurar o pareamento de usuários conforme a distância da BS, *CSI* e demanda de tráfego. Por exemplo, usuários com SINR semelhantes (CQI semelhantes) poderiam ser pareados com $\alpha = 0{,}7$ de acordo com a Figura 52. Por outro lado, usuários com SINR distintas poderiam ser pareados com configurações de $\alpha > 0{,}7$.

O NOMA também provê maior flexibilidade ao sistema na adaptação do enlace. Assumindo que os *streams* PS e SS sejam alocados para um mesmo usuário, o sistema que esteja operando, por exemplo, com MIMO QPSK tem a opção de escolher aumentar a ordem de modulação para 16QAM ou adicionar um *stream* secundário com SINR ajustada pelo parâmetro α para refletir a componente CQI reportada pelo usuário.

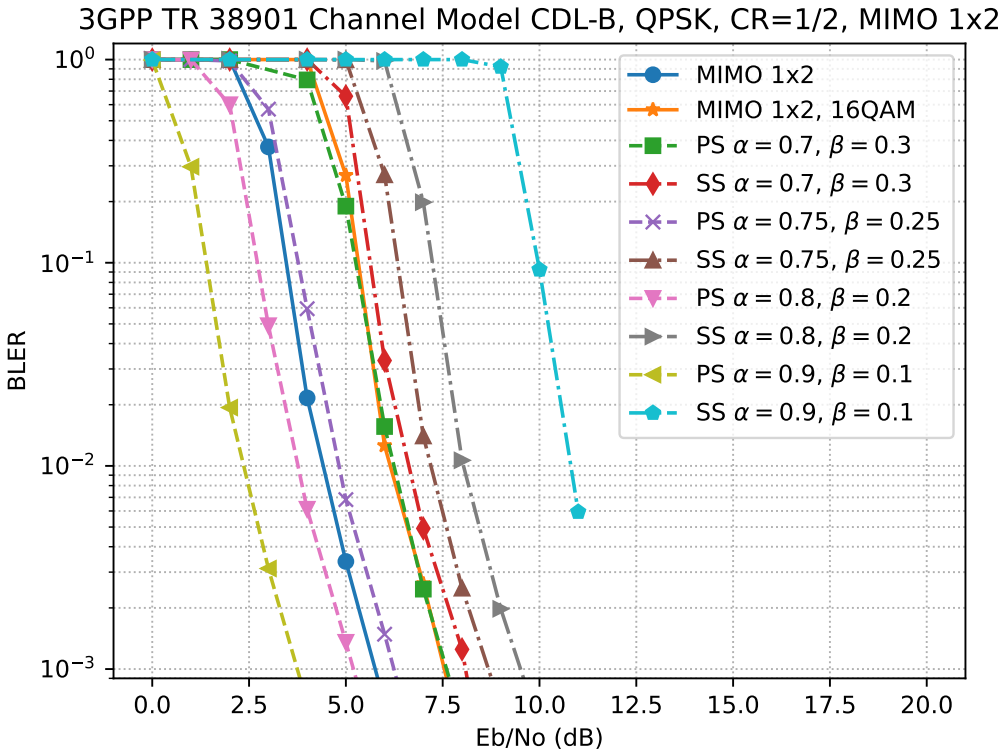


Figura 52: Comparação de desempenho do NOMA para diferentes valores de α e β.

Desempenho em Ambiente NLOS com modelos CDL-A e CDL-C

A Figura 53 compara SISO, MIMO 2×2 e NOMA 1×1 no perfil NLOS mais brando do CDL-A. O MIMO 2×2 emprega dois *streams* MIMO com QPSK e apresenta declive praticamente constante de −10 dB/década. Apesar de a BLER de 10% ocorrer para $E_b/N_o \approx 5$ dB, como no SISO QPSK, a BLER de 1% ocorre em torno de 15 dB, enquanto que para SISO QPSK ocorre em torno de 6 dB. O desempenho é pior inclusive para o 64QAM, que possui limiar de 11 dB para BLER de 1%. Isso é um indicativo de que adicionar um segundo *stream* ao sistema com MIMO pode ser restritivo e, neste caso, impossibilitaria adaptar o enlace com referência à BLER de 1%, sendo possível apenas com referência à BLER de 10%.

Por outro lado, o NOMA consegue adicionar um segundo *stream* de forma mais eficiente que o MIMO, mesmo sem empregar diversidade espacial. Ao utilizar dois fluxos QPSK, o NOMA alcança a mesma eficiência espectral do SISO 16QAM e do MIMO 2×2 QPSK. Com relação ao SISO 16QAM, o *stream* PS apresenta um ganho de aproximadamente 2 dB e o SS uma perda de cerca de 1 dB, empregando os parâmetros $\alpha = 0{,}75$ e $\beta = 0{,}25$. Mesmo com desempenho semelhante ao SISO 16QAM, o sistema dobra o número de *streams*, com potencial para suportar o dobro de usuários. Em relação ao SISO QPSK, o *stream* PS apresenta desempenho semelhante, com uma perda de aproximadamente 0.5 dB, enquanto o SS apresenta uma perda de cerca de 3,5 dB. Ainda assim, a entrega simultânea de duas palavras-código no NOMA QPSK/QPSK dobra a eficiência espectral por um pequeno custo de SNR.

Já o NOMA 16QAM/QPSK (PS/SS) utiliza parâmetros próximos do limite superior de $\alpha = 0{,}93$ e $\beta = 0{,}07$, com eficiência espectral semelhante à do SISO 64QAM. Nessa configuração, o *stream* PS apresenta um ganho de aproximadamente 2 dB e o SS uma perda de cerca de 2,5 dB. A flexibilidade é limitada no sentido de que se ganha um *stream* adicional de baixa

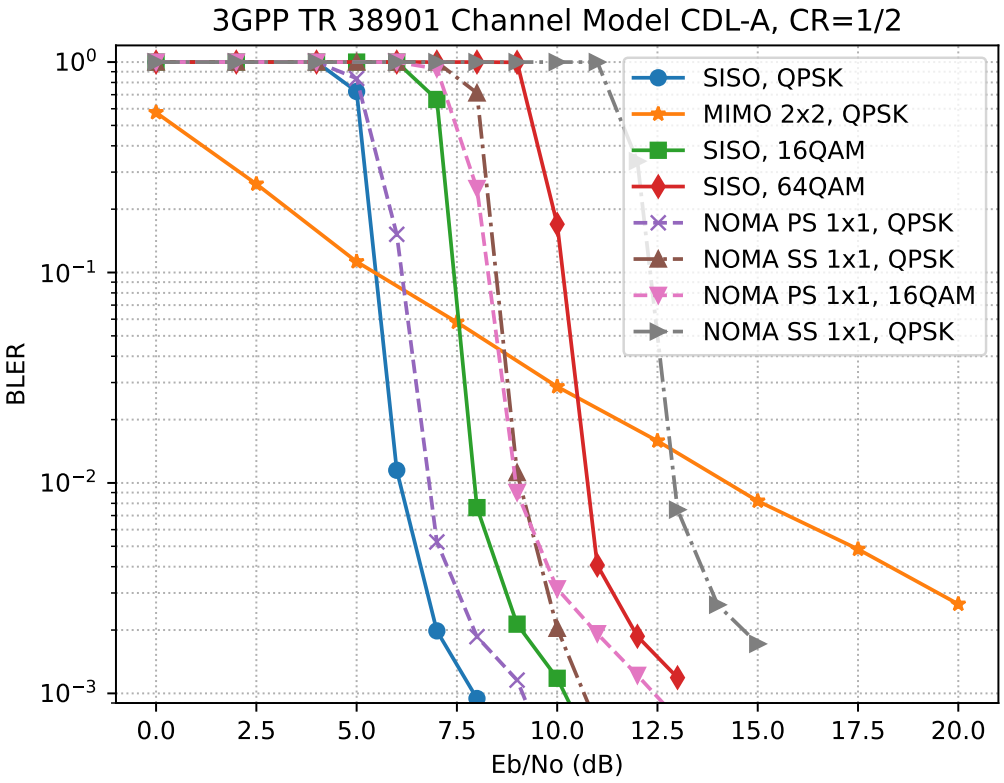


Figura 53: Comparação entre MIMO e NOMA para taxa de codificação 1/2 e modelo de canal CDL-A.

taxa (QPSK), mas com um alto limiar de E_b/N_o em torno de 12,5 dB. Em comparação ao SISO 16QAM, o PS apresenta uma perda de aproximadamente 0,8 dB e o SS uma perda de cerca de 5 dB, mas há um aumento na eficiência espectral equivalente ao do SISO 64QAM, que por sua vez apresenta uma perda de 2,5 dB em relação ao 16QAM. Embora a figura não mostre, ainda seria possível diminuir α para que o PS e o SS tenham desempenho semelhante ao 64QAM, com o benefício adicional do ganho no número de *streams*.

No ambiente urbano severo CDL-C da Figura 54, a diversidade de recepção torna-se decisiva: o MIMO 1×4 QPSK supera o 2×2 em mais de 6 dB em BLER de 10% graças às ramificações extras alimentando o equalizador LMMSE. Neste caso, dobra-se a eficiência espectral a um custo de 6 dB. Já o NOMA 1×4 QPSK/QPSK dobra a eficiência espectral com relação ao MIMO 1×4, mantendo o desempenho com o *stream* PS e obtendo uma perda de desempenho de 3 dB@10%. Neste caso, o NOMA utiliza a configuração de $\alpha = 0{,}75$ e $\beta = 0{,}25$. Em relação ao MIMO 1×4 16QAM, o *stream* SS perde apenas cerca de 0,8 dB, indicando que pode ser vantajoso comutar do MIMO 1×4 QPSK para o NOMA 1×4 QPSK/QPSK em vez do MIMO 1×4 16QAM, visto que o PS ganha cerca de 2,5 dB em relação ao 16QAM. Já o NOMA 16QAM/QPSK na configuração de $\alpha = 0{,}93$ e $\beta = 0{,}07$ aumenta a eficiência espectral com relação ao MIMO 1×4 16QAM com uma perda de apenas cerca de 0,5 dB no PS, mas com uma perda de aproximadamente 4 dB no SS. O SS ainda perde cerca de 2,5 dB em relação ao 64QAM. Entretanto, há margem para reduzir α visando diminuir a diferença do SS para o 64QAM, tornando mais atraente o aumento do número de *streams* proporcionado pelo NOMA.

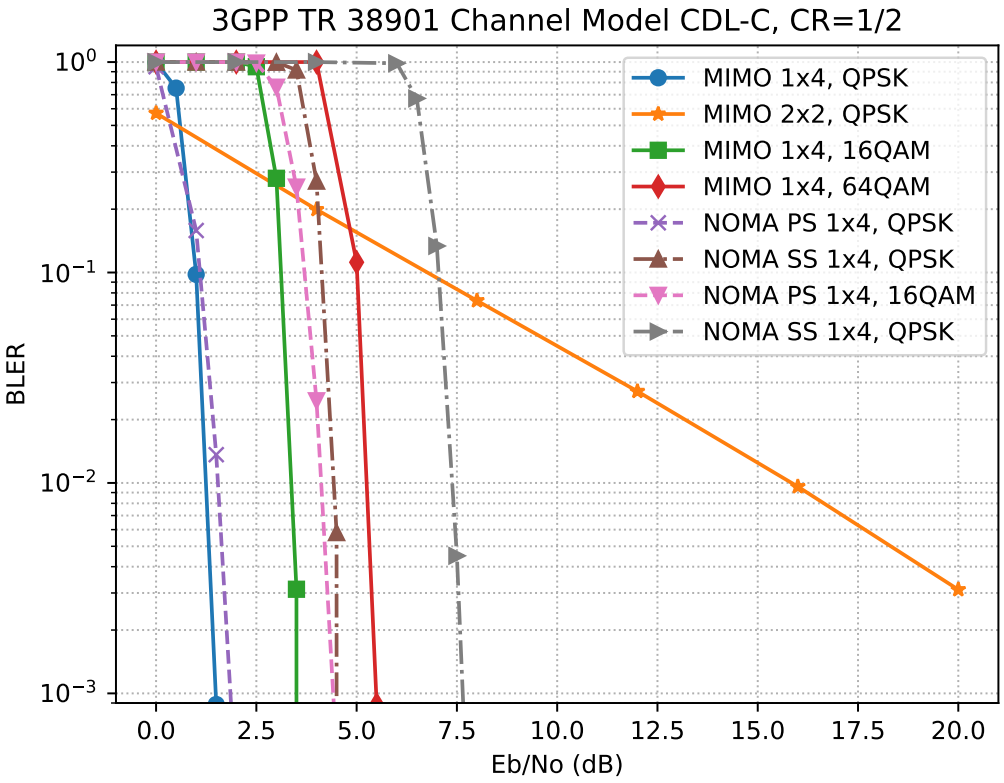


Figura 54: Comparação entre MIMO e NOMA para taxa de codificação 1/2 e modelo de canal CDL-C.

Desempenho em canal CDL-E, com LOS e Dispersão Angular

Por fim, a Figura 55 contrasta MIMO 2×2, MIMO 4×4 e NOMA de duas camadas com QPSK, 16-QAM e 64-QAM. Enquanto MIMO 2×2 desempenha bem para as modulações QPSK, 16QAM e 64QAM, o MIMO 4×4 QPSK demonstra não ser possível aumentar a eficiência espectral apenas pela adição de mais *streams*, obtendo BLER@10% em torno de 20 dB. Seria necessário aumentar também a diversidade de transmissão ou de recepção para reduzir esse limiar de BLER@10%. Em contraste com o MIMO 4×4, o NOMA 2×2 consegue dobrar o número de *streams* com uma perda de até 0,5 dB@10% para o PS QPSK e 16QAM. Observa-se o mesmo padrão de resultados observado na Figura 52, em que o desempenho do NOMA é equivalente ao do MIMO para mesma eficiência espectral, porém com o benefício de possibilitar dobrar o número de usuários no sistema. Por exemplo, o NOMA QPSK/QPSK equivale ao MIMO 16QAM, o NOMA 16QAM/QPSK equivale ao MIMO 64QAM e, embora não seja mostrado na Figura 55, o NOMA 64QAM/QPSK equivale a um MIMO 256QAM. Em todos esses casos há a possibilidade de ajustar o limiar de E_b/N_o @10% dos *streams* PS e SS através do parâmetro α. Os resultados para o NOMA QPSK/QPSK foram obtidos para $\alpha = 0{,}78$ e $\beta = 0{,}22$. O NOMA 16QAM/QPSK utiliza $\alpha = 0{,}93$ e $\beta = 0{,}07$. E o NOMA 64QAM/QPSK $\alpha = 0{,}97$ e $\beta = 0{,}03$.

17.4 Conclusão

Os resultados mostram que o NOMA em domínio de potência pode duplicar o *throughput* espacial QPSK com acréscimo de apenas 0,5–1 dB em SNR, desde que a divisão de potência

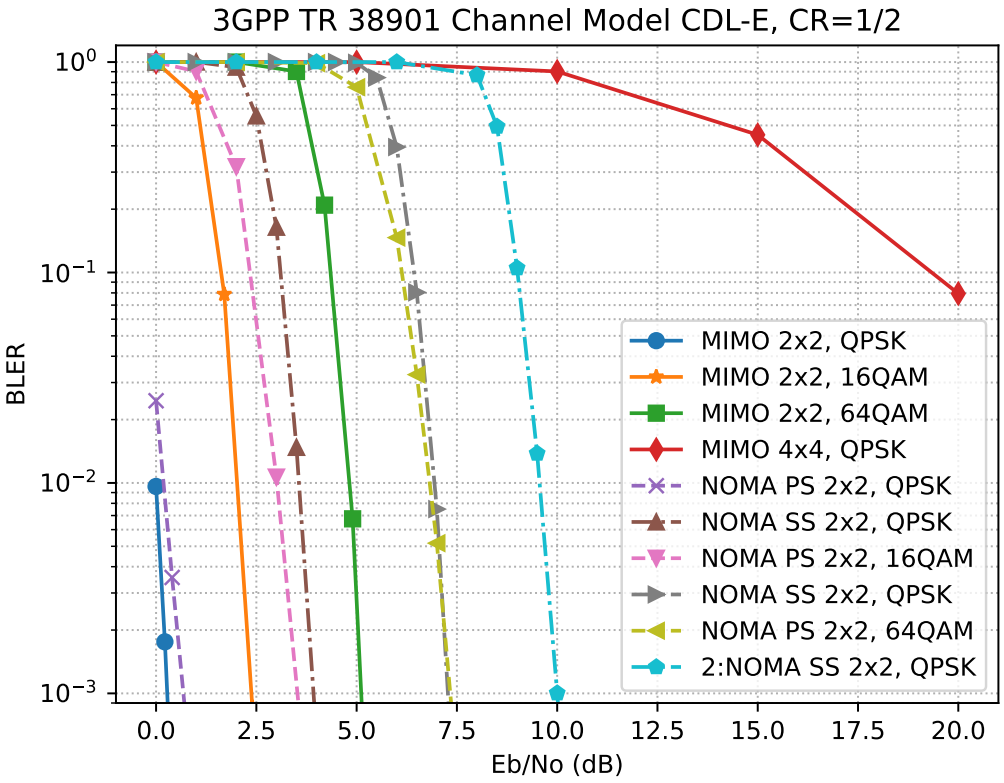


Figura 55: Comparação entre MIMO e NOMA para taxa de codificação 1/2 e modelo de canal CDL-E.

permaneça em $\alpha \leq 0{,}8$. A diversidade de recepção mitiga a interferência extra em cenários NLOS, enquanto ambientes LOS se aproveitam dos valores singulares mais altos para "encurtar" ainda mais a cascata de erros. Tais tendências validam o NOMA como complemento promissor ao MIMO convencional em redes 6G de borda e *cell-free*, desde que o CSI seja estimado com precisão e a decodificação iterativa se mantenha viável em receptores comerciais.

Como lições aprendidas, este estudo confirma que a superposição em domínio de potência amplia a capacidade espacial com uma penalidade modesta de SNR, desde que a razão de potência se mantenha controlada e a diversidade de recepção seja explorada; tal equilíbrio revela-se particularmente eficaz em cenários NLOS, enquanto ambientes LOS tiram proveito dos autovalores mais acentuados para reduzir a propagação de erros. Contudo, permanecem desafios e oportunidades atraentes para investigações futuras. Primeiro, a integração de algoritmos avançados de cancelamento de interferência inter-stream em arquiteturas MIMO — inclusive técnicas que combinem SIC e cancelamento paralelo sob CSI imperfeito — promete elevar ainda mais a robustez do enlace. Segundo, o próprio framework aqui apresentado pode ser estendido para um esquema de sobreposição em domínio de espalhamento espectral, no qual *streams* adicionais são multiplexados por códigos pseudo-ortogonais em vez de potência. Essa alternativa, inspirada em abordagens de code-domain NOMA como SCMA, tem potencial para multiplicar significativamente o número de *streams* disponíveis além do limite duplo imposto pelo PD-NOMA, preservando a largura de banda e aliviando a exigência de controle fino de potência. A análise comparativa entre estes métodos, aliada ao estudo do custo computacional em receptores de baixa complexidade, configura um caminho promissor para consolidar o NOMA como alicerce de rádios 6G flexíveis e escaláveis.

18 Conclusões Gerais

Este relatório técnico traz resultados obtidos no primeiro ano da Fase III do Projeto Brasil 6G, no contexto da Atividade 2.2. Foram desenvolvidos estudos inéditos em diversas frentes de pesquisa, abrangendo aspectos teóricos, metodológicos e aplicados. As contribuições abordaram temas como redução de PAPR, comunicação assistida por BD-RIS, acesso dinâmico ao espectro, integração de tecnologias ópticas e sem fio para conectividade em áreas remotas, comunicação orientada a tarefas, técnicas de múltiplo acesso para IoT massivo e comunicações via satélite.

De modo geral, as técnicas tratadas nesses estudos buscaram atender aos requisitos associados aos casos de uso previstos para os sistemas 6G, como maior eficiência espectral e energética, maior robustez, e soluções para lidar com os desafios de escalabilidade, latência e conectividade ubíqua. Destaca-se, ainda, o uso crescente de algoritmos de inteligência artificial e aprendizado de máquina como ferramentas para otimização de desempenho e adaptação a ambientes complexos e dinâmicos.

Referências

[1] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, 2021.

[2] L. P. LASSE, "Itu artificial intelligence/machine learning in 5g challenge," 2020. [Online]. Available: https://ai5gchallenge.ufpa.br/

[3] N. Rajatheva, I. Atzeni, E. Björnson, A. Bourdoux, S. Buzzi, J.-B. Doré, S. Erkucuk, M. Fuentes, K. Guan, Y. Hu, X. Huang, J. Hulkkonen, J. M. Jornet, M. Katz, R. Nilsson, E. Panayirci, K. Rabie, N. Rajapaksha, M. Salehi, and W. Xu, "White paper on broadband connectivity in 6g," 6G Research Visions, No. 10. University of Oulu, 2020.

[4] Anatel, "Resolução nº 190, 6 de outubro de 2021."

[5] M. Khalil, J. Qadir, O. Onireti, M. A. Imran, and S. Younis, "Feasibility, architecture and cost considerations of using tvws for rural internet access in 5g," in *2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN)*, 2017, pp. 23–30.

[6] B. S. C. da Silva, M. B. de Mello, and L. L. Mendes, "Papr reduction technique for mobile communications system based on pattern learning," in *2024 19th International Symposium on Wireless Communication Systems (ISWCS)*, 2024, pp. 1–6.

[7] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. Elkashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst, X. You, and L. Hanzo, "Reconfigurable Intelligent Surfaces for 6G Systems: Principles, Applications, and Research Directions," *IEEE Communications Magazine*, vol. 59, no. 6, pp. 14–20, 2021.

[8] H. Li, S. Shen, and B. Clerckx, "Beyond Diagonal Reconfigurable Intelligent Surfaces: From Transmitting and Reflecting Modes to Single-, Group-, and Fully-Connected Architectures," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2311–2324, 2023.

[9] S. Shen, B. Clerckx, and R. Murch, "Modeling and Architecture Design of Reconfigurable Intelligent Surfaces Using Scattering Parameter Network Analysis," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 1229–1243, 2022.

[10] Y. Liu, W. Yi, Z. Ding, X. Liu, O. A. Dobre, and N. Al-Dhahir, "Developing NOMA to Next Generation Multiple Access: Future Vision and Research Opportunities," *IEEE Wireless Communications*, vol. 29, no. 6, pp. 120–127, 2022.

[11] Q. Zhang, G. Luo, Z. Dong, F. Sun, X. Wang, and J. Liu, "Beyond-Diagonal Reconfigurable Intelligent Surface Enhanced NOMA Systems," *IEEE Wireless Communications Letters*, vol. 14, no. 1, pp. 118–122, 2025.

[12] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of Non-Orthogonal Multiple Access in LTE and 5G Networks," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 185–191, 2017.

[13] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, 2020.

[14] S. Majumdar, S. Hazra, S. Chattopadhyay, and S. Sardar, "Cooperative spectrum sensing for cognitive radio based industrial IoT applications," in *2023 International Conference on Computational Intelligence, Networks and Security (ICCINS)*, 2023, pp. 1–6.

[15] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," *IEEE personal communications*, vol. 6, no. 4, pp. 13–18, 1999.

[16] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 1, pp. 40–62, 2011.

[17] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[18] Y. Zeng and Y.-C. Liang, "Spectrum-sensing algorithms for cognitive radio based on statistical covariances," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1804–1815, 2009.

[19] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 814–823, 2010.

[20] R. López-Valcarce, G. Vazquez-Vilar, and J. Sala, "Multiantenna spectrum sensing for cognitive radio: overcoming noise uncertainty," in *2010 2nd International Workshop on Cognitive Information Processing*, 2010, pp. 310–315.

[21] J. K. Tugnait, "Multiple antenna spectrum sensing for cognitive radios in unknown noise," in *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*, 2011, pp. 431–435.

[22] B. Nadler, F. Penna, and R. Garello, "Performance of eigenvalue-based signal detectors with known and unknown noise level," in *2011 IEEE International Conference on Communications (ICC)*, 2011, pp. 1–5.

[23] D. Ramirez, J. Via, I. Santamaria, and L. L. Scharf, "Locally most powerful invariant tests for correlation and sphericity of Gaussian vectors," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2128–2141, 2013.

[24] R. Gao, Z. Li, P. Qi, and H. Li, "A robust cooperative spectrum sensing method in cognitive radio networks," *IEEE Communications Letters*, vol. 18, no. 11, pp. 1987–1990, 2014.

[25] I. Kakalou, D. Papadopoulou, T. Xifilidis, K. E. Psannis, K. Siakavara, and Y. Ishibashi, "A survey on spectrum sensing algorithms for cognitive radio networks," in *2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, 2018, pp. 1–4.

[26] R. Tandra and A. Sahai, "Snr walls for signal detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 4–17, 2008.

[27] K. M. Captain and M. V. Joshi, "SNR wall for cooperative spectrum sensing using generalized energy detector," in *2018 10th International Conference on Communication Systems Networks (COMSNETS)*, 2018, pp. 143–150.

[28] S. S. Kalamkar and A. Banerjee, "On the performance of generalized energy detector under noise uncertainty in cognitive radio," in *2013 National Conference on Communications (NCC)*, 2013, pp. 1–5.

[29] Z. Li, X. Su, J. Zeng, Y. Kuang, and H. Wang, "A study of SNR wall phenomenon under cooperative energy spectrum sensing," in *2013 22nd International Conference on Computer Communication and Networks (ICCCN)*, 2013, pp. 1–5.

[30] S. Chaudhari, J. Lunden, V. Koivunen, and H. V. Poor, "Cooperative sensing with imperfect reporting channels: Hard decisions or soft decisions?" *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 18–28, 2012.

[31] M. Sheharyar, A. Naushad, I. Khan, M. A. Hussain, and F. Khan Kakar, "Study on fixed and dynamic spectrum access models for cellular networks," *Journal of Telecommunications and the Digital Economy*, vol. 10, no. 2, pp. 267–287, 2022.

[32] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in *Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum*. New York, NY, USA: Association for Computing Machinery, 2006, p. 1–es.

[33] Y. Chen and H. S. Oh, "A survey of measurement-based spectrum occupancy modeling for cognitive radios," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 848–859, Firstquarter 2016.

[34] R. Tandra and A. Sahai, "SNR walls for feature detectors," in *2007 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, 2007, pp. 559–570.

[35] L. d. S. Costa, D. A. Guimarães, and B. F. Uchôa-Filho, "Influence of noise uncertainty source and model on the SNR wall of energy detection," *IEEE Transactions on Cognitive Communications and Networking*, vol. 10, no. 5, pp. 1897–1912, 2024.

[36] S. A. Dyer and J. S. Dyer, "Approximations to error functions," *IEEE Instrumentation & Measurement Magazine*, vol. 10, no. 6, pp. 45–48, 2007.

[37] T. An, D. Kim, I. Song, M. S. Yeu, H.-K. Min, S. Lee, and W. Lee, "Cooperative spectrum sensing based on generalized likelihood ratio test under impulsive noise circumstances," in *MILCOM 2012 - 2012 IEEE Military Communications Conference*, 2012, pp. 1–6.

[38] P. B. Gohain, S. Chaudhari, and V. Koivunen, "Cooperative energy detection with heterogeneous sensors under noise uncertainty: SNR wall and use of evidence theory," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 3, pp. 473–485, 2018.

[39] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A survey on low latency towards 5g: Ran, core network and caching solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3098–3130, 2018.

[40] J. Sachs, L. A. Andersson, J. Araújo, C. Curescu, J. Lundsjö, G. Rune, E. Steinbach, and G. Wikström, "Adaptive 5g low-latency communication for tactile internet services," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 325–349, 2018.

[41] O. T. Yavascan and E. Uysal, "Analysis of slotted aloha with an age threshold," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1456–1470, 2021.

[42] Q. Liang and Y. Wu, "Age of information optimization of random access networks with retry limit," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 5, pp. 7483–7496, 2025.

[43] H. Xie, S.-W. Jeon, and H. Jin, "Distributed real-time control for minimizing aoi in random access networks," *IEEE Internet of Things Journal*, vol. 12, no. 8, pp. 10 524–10 538, 2025.

[44] A. Buratto, A. Munari, and L. Badia, "Strategic backoff of slotted aloha for minimal age of information," *IEEE Communications Letters*, vol. 29, no. 1, pp. 155–159, 2025.

[45] P. S. Dester, M. Costa, P. H. J. Nardelli, P. E. Gória Silva, and J. M. Moualeu, "Novel energy-centric analysis for random access networks," *IEEE Wireless Communications Letters*, vol. 14, no. 1, pp. 133–137, 2025.

[46] Y. Kim, "Minimizing age of information in slotted aloha with short-packet communications," *IEEE Access*, vol. 12, pp. 186 779–186 793, 2024.

[47] D. Fiems and A. Vinel, "The age-of-information distribution in slotted aloha," *IEEE Access*, vol. 12, pp. 159 462–159 472, 2024.

[48] F. Zhao, N. Pappas, C. Ma, X. Sun, T. Q. S. Quek, and H. H. Yang, "Optimizing information freshness in mobile networks with age-threshold aloha," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 963–968.

[49] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, 2017.

[50] D. P. Bertsekas and R. G. Gallager, *Data Networks*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1992.

[51] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.

[52] A. Mostaani, T. X. Vu, S. K. Sharma, V.-D. Nguyen, Q. Liao, and S. Chatzinotas, "Task-oriented communication design in cyber-physical systems: A survey on theory and applications," *IEEE Access*, vol. 10, pp. 133 842–133 868, 2022.

[53] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, vol. 59, no. 6, pp. 96–102, 2021.

[54] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, 2012, pp. 1–6.

[55] T. M. Getu, G. Kaddoum, and M. Bennis, "Making sense of meaning: A survey on metrics for semantic and goal-oriented communication," *IEEE Access*, vol. 11, pp. 45 456–45 492, 2023.

[56] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The age of incorrect information: A new performance metric for status updates," *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, pp. 2215–2228, 2020.

[57] X. Zheng, S. Zhou, and Z. Niu, "Urgency of information for context-aware timely status updates in remote control systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7237–7250, 2020.

[58] E. Fountoulakis, N. Pappas, and M. Kountouris, "Goal-oriented policies for cost of actuation error minimization in wireless autonomous systems," *IEEE Communications Letters*, vol. 27, no. 9, pp. 2323–2327, 2023.

[59] R. A. Howard, "Information value theory," *IEEE Transactions on Systems Science and Cybernetics*, vol. 2, no. 1, pp. 22–26, 1966.

[60] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, "Age-of-information vs. value-of-information scheduling for cellular networked control systems," in *Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems*, ser. ICCPS '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 109–117. [Online]. Available: https://doi.org/10.1145/3302509.3311050

[61] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Shen, and C. Miao, "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 213–250, 2023.

[62] D. Zonta, B. Glisic, and S. Adriaenssens, "Value of information: Impact of monitoring on decision-making," *Structural Control and Health Monitoring*, vol. 21, 07 2014.

[63] T. L. Saaty, "What is the analytic hierarchy process?" in *Mathematical Models for Decision Support*, G. Mitra, H. J. Greenberg, F. A. Lootsma, M. J. Rijkaert, and H. J. Zimmermann, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1988, pp. 109–121.

[64] M. Giordani, T. Higuchi, A. Zanella, O. Altintas, and M. Zorzi, "A framework to assess value of information in future vehicular networks," in *Proceedings of the 1st ACM MobiHoc Workshop on Technologies, MOdels, and Protocols for Cooperative Connected Cars*, ser. TOP-Cars '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 31–36. [Online]. Available: https://doi.org/10.1145/3331054.3331551

[65] P. Padhy, R. K. Dash, K. Martinez, and N. R. Jennings, "A utility-based adaptive sensing and multihop communication protocol for wireless sensor networks," *ACM Trans. Sen. Netw.*, vol. 6, no. 3, Jun. 2010. [Online]. Available: https://doi.org/10.1145/1754414.1754423

[66] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[67] D. Wheeler and B. Natarajan, "Engineering semantic communication: A survey," *IEEE Access*, vol. 11, pp. 13 965–13 995, 2023.

[68] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[69] Z. Weng, Z. Qin, and G. Y. Li, "Semantic communications for speech signals," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.

[70] Q. Fu, H. Xie, Z. Qin, G. Slabaugh, and X. Tao, "Vector quantized semantic communication system," *IEEE Wireless Communications Letters*, vol. 12, no. 6, pp. 982–986, 2023.

[71] S. Hochreiter, "Long short-term memory," *Neural Computation MIT-Press*, 1997.

[72] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 2326–2330.

[73] H. Xie and Z. Qin, "A lite distributed semantic communication system for internet of things," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 142–153, 2020.

[74] M.-K. Chang, C.-T. Hsu, and G.-C. Yang, "Gensc: Generative semantic communication systems using bart-like model," *IEEE Communications Letters*, 2024.

[75] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *CoRR*, vol. abs/1910.13461, 2019. [Online]. Available: http://arxiv.org/abs/1910.13461

[76] Y. Miao, J. Yan, Y. Wang, Z. Li, and D. Hu, "A semantic communication system based on vector quantization and generative model," in *2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE)*. IEEE, 2024, pp. 46–50.

[77] R. Gray, "Vector quantization," *IEEE Assp Magazine*, vol. 1, no. 2, pp. 4–29, 1984.

[78] X. Peng, Z. Qin, X. Tao, J. Lu, and L. Hanzo, "A robust semantic text communication system," *IEEE Transactions on Wireless Communications*, 2024.

[79] K. Papineni, S. Roukos, T. Ward, and W. jing Zhu, "Bleu: a method for automatic evaluation of machine translation," 2002, pp. 311–318.

[80] T. Zhang*, V. Kishore*, F. Wu*, K. Q. Weinberger, and Y. Artzi, "Bertscore: Evaluating text generation with bert," in *International Conference on Learning Representations*, 2020. [Online]. Available: https://openreview.net/forum?id=SkeHuCVFDr

[81] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *CoRR*, vol. abs/1810.04805, 2018. [Online]. Available: http://arxiv.org/abs/1810.04805

[82] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020. [Online]. Available: http://jmlr.org/papers/v21/20-074.html

[83] L. F. Gontijo and P. Cardieri, "Pre-trained language models in semantic communication," in *XLII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT2024)*. SBRT, 2024.

[84] M. Wang, J. Li, M. Ma, and X. Fan, "Constellation design for deep joint source-channel coding," *IEEE Signal Processing Letters*, vol. 29, pp. 1442–1446, 2022.

[85] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," *arXiv preprint arXiv:2201.01389*, 2021.

[86] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, 2022.

[87] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293–304, 1962.

[88] J. Liu, S. Shao, W. Zhang, and H. V. Poor, "An indirect rate-distortion characterization for semantic sources: General model and the case of gaussian observation," *IEEE Transactions on Communications*, vol. 70, no. 9, pp. 5946–5959, 2022.

[89] P. A. Stavrou and M. Kountouris, "The role of fidelity in goal-oriented semantic communication: A rate distortion approach," *IEEE Transactions on Communications*, vol. 71, no. 7, pp. 3918–3931, 2023.

[90] Raquel Katigbak, IBM Distinguished Industry Leader. Accessed: Jun. 15, 2024. [Online]. Available: https://www.ibm.com/blog/the-economy-of-things-the-next-value-lever-for-telcos/

[91] J. Choi *et al.*, "Grant-Free Random Access in Machine-Type Communication: Approaches and Challenges," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 151–158, 2022.

[92] M. Asad Ullah, G. Pasolini, K. Mikhaylov, and H. Alves, "Understanding the Limits of LoRa Direct-to-Satellite: The Doppler Perspectives," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 51–63, 2024.

[93] O. Kodheli, E. Lagunas, N. Maturo, S. K. Sharma, B. Shankar, J. F. M. Montoya, J. C. M. Duncan, D. Spano, S. Chatzinotas, S. Kisseleff, J. Querol, L. Lei, T. X. Vu, and G. Goussetis, "Satellite Communications in the New Space Era: A Survey and Future Challenges," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 70–109, 2021.

[94] M. Harounabadi and T. Heyn, "Toward Integration of 6G-NTN to Terrestrial Mobile Networks: Research and Standardization Aspects," *IEEE Wireless Communications*, vol. 30, no. 6, pp. 20–26, 2023.

[95] J. A. Fraire, O. Iova, and F. Valois, "Space-Terrestrial Integrated Internet of Things: Challenges and Opportunities," *IEEE Communications Magazine*, vol. 60, no. 12, pp. 64–70, 2022.

[96] M. Centenaro *et al.*, "A Survey on Technologies, Standards and Open Challenges in Satellite IoT," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1693–1720, 2021.

[97] *LoRa Alliance Technical Committee Regional Parameters Workgroup, Version 1.0.4 and Regional Parameters RP002*, LoRa Alliance, Sept. 2022. [Online]. Available: https://resources.lora-alliance.org/technical-specifications/rp002-1-0-4-regional-parameters

[98] Nanosats Database. Lacuna Space launches latest LS2f satellite with NanoAvionics & Space-X. Accessed: Jun. 8, 2024. [Online]. Available: https://lacuna.space/lacuna-space-launches-latest-ls2f-satellite-with-nanoavionics-space-x/

[99] O. Georgiou and U. Raza, "Low Power Wide Area Network Analysis: Can LoRa Scale?" *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 162–165, 2017.

[100] X. Yan, K. An, T. Liang, G. Zheng, Z. Ding, S. Chatzinotas, and Y. Liu, "The Application of Power-Domain Non-Orthogonal Multiple Access in Satellite Communication Networks," *IEEE Access*, vol. 7, pp. 63 531–63 539, 2019.

[101] Z. Gao, A. Liu, and X. Liang, "The Performance Analysis of Downlink NOMA in LEO Satellite Communication System," *IEEE Access*, vol. 8, pp. 93 723–93 732, 2020.

[102] B. M. Elhalawany *et al.*, "Outage Analysis of Coordinated NOMA Transmission for LEO Satellite Constellations," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 2195–2202, 2022.

[103] D. A. Tubiana, J. Farhat, G. Brante, and R. D. Souza, "Q-Learning NOMA Random Access for IoT-Satellite Terrestrial Relay Networks," *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1619–1623, 2022.

[104] R. Ge *et al.*, "Performance Analysis of Cooperative Nonorthogonal Multiple Access Scheme in Two-Layer GEO/LEO Satellite Network," *IEEE Systems Journal*, vol. 16, no. 2, pp. 2300–2310, 2022.

[105] M. Asad Ullah, K. Mikhaylov, and H. Alves, "Massive machine-type communication and satellite integration for remote areas," *IEEE Wireless Communications*, vol. 28, no. 4, pp. 74–80, 2021.

[106] J. A. Fraire, O. Iova, and F. Valois, "Space-terrestrial integrated internet of things: Challenges and opportunities," *IEEE Communications Magazine*, vol. 60, no. 12, pp. 64–70, 2022.

[107] *LoRaWAN™ L2 1.0.4 Specification (TS001-1.0.4)*, LoRa Alliance, Inc., Oct. 2020, rev. 1.0.4. [Online]. Available: https://resources.lora-alliance.org/technical-specifications/ts001-1-0-4-lorawan-l2-1-0-4-specification

[108] International Telecommunication Union, "ITU-T Recommendation Y.4480: Low Power Protocol for Wide Area Wireless Networks," International Telecommunication Union, Tech. Rep. Y.4480, November 2021.

[109] *AN1200.22: LoRa® Modulation Basics*, Semtech Corp., 2015, Rev. 2.0.

[110] *AN1200.64: LR-FHSS System Performance*, Semtech Corp., 2022, Rev. 1.2.

[111] M. A. Ullah, K. Mikhaylov, and H. Alves, "An overview of direct-to-satellite IoT: Opportunities and open challenges," in *2023 IEEE 9th World Forum on Internet of Things (WF-IoT)*, 2023, pp. 1–8.

[112] A. Lavric, A. I. Petrariu, and V. Popa, "Sigfox communication protocol: The new era of IoT?" in *2019 International Conference on Sensing and Instrumentation in IoT Era (ISSI)*, 2019, pp. 1–4.

[113] M. Chen, Y. Miao, Y. Hao, and K. Hwang, "Narrow band internet of things," *IEEE Access*, vol. 5, pp. 20 557–20 577, 2017.

[114] R. Ratasuk, N. Mangalvedhe, Y. Zhang, M. Robert, and J.-P. Koskinen, "Overview of narrowband IoT in LTE rel-13," in *2016 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2016, pp. 1–7.

[115] J. M. de Souza Sant'Ana, A. Hoeller, H. Alves, and R. D. Souza, "LR-FHSS-Sim: A discrete-event simulator for LR-FHSS networks," in *2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 2024, pp. 700–705.

[116] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

[117] S. Kim, S. Ahn, J. Park, J. Youn, Y. Kwon, and S. Cho, "CPU-cooperative power control scheme for scalable cell-free massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 13 904–13 919, 2024.

[118] F. Li, Q. Sun, X. Ji, and X. Chen, "Scalable cell-free massive MIMO with multiple CPUs," *Mathematics*, vol. 10, no. 11, p. 1900, 2022.

[119] E. Björnson, J. Hoydis, L. Sanguinetti *et al.*, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3–4, pp. 154–655, 2017.

[120] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4247–4261, 2020.

[121] J. Jiang, J. Wang, H. Chu, Q. Gao, and J. Zhang, "Whale swarm reinforcement learning based dynamic cooperation clustering method for cell-free massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 3, pp. 4114–4118, 2022.

[122] R. Y. Chang, S.-F. Han, and F.-T. Chien, "Reinforcement learning-based joint cooperation clustering and content caching in cell-free massive MIMO networks," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*. IEEE, 2021, pp. 1–7.

[123] L. A. Iliadis, Z. D. Zaharis, S. Sotiroudis, P. Sarigiannidis, G. K. Karagiannidis, and S. K. Goudos, "The road to 6G: A comprehensive survey of deep learning applications in cell-free massive MIMO communications systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2022, no. 1, p. 68, 2022.

[124] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. University of California Press, 1965–6, pp. 281–297.

[125] B. Salehi, G. Reus-Muns, D. Roy, Z. Wang, T. Jian, J. Dy, S. Ioannidis, and K. Chowdhury, "Deep learning on multimodal sensor data at the wireless edge for vehicular network," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7639–7655, 2022.

[126] J. Ruseckas, G. Molis, and H. Bogucka, "Mimo beam selection in 5g using neural networks," *International Journal of Electronics and Telecommunications*, vol. 67, no. 4, 2021.

[127] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The next generation wireless access technology*. Academic Press, 2020.

[128] A. Zaidi, F. Athley, J. Medbo, U. Gustavsson, G. Durisi, and X. Chen, *5G Physical Layer: principles, models and technology components*. Academic Press, 2018.

[129] J. R. Hampton, *Introduction to MIMO communications*. Cambridge university press, 2013.

[130] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 20, pp. 5289–5304, 2017.

[131] I. Aleksander, W. Thomas, and P. Bowden, "Wisard· a radical step forward in image recognition," *Sensor review*, vol. 4, no. 3, pp. 120–124, 1984.

[132] I. Aleksander, M. De Gregorio, F. M. G. França, P. M. V. Lima, and H. Morton, "A brief introduction to weightless neural systems." in *ESANN*. Citeseer, 2009, pp. 299–305.

[133] A. Klautau, P. Batista, N. González-Prelcic, Y. Wang, and R. W. Heath, "5g mimo data for machine learning: Application to beam-selection using deep learning," in *2018 Information Theory and Applications Workshop (ITA)*. IEEE, 2018, pp. 1–9.

[134] A. Chowdhury and C. R. Murthy, "Half-duplex APs with dynamic TDD versus full-duplex APs in Cell-Free systems," *IEEE Transactions on Communications*, vol. 72, no. 7, pp. 3856–3872, 2024.

[135] S. Chen, J. Zhang, J. Zhang, E. Björnson, and B. Ai, "A survey on user-centric cell-free massive MIMO systems," *Digital Communications and Networks*, vol. 8, no. 5, pp. 695–719, 2022.

[136] O. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Foundations and Trends in Signal Processing*, vol. 14, no. 3–4, p. 162–472, 2021.

[137] D. Wang, M. Wang, P. Zhu, J. Li, J. Wang, and X. You, "Performance of network-assisted full-duplex for cell-free massive MIMO," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1464–1478, 2020.

[138] H. Kim, H. Lee, T. Kim, and D. Hong, "Cell-free mMIMO systems with dynamic TDD," in *Proc. IEEE 95th Vehicular Technology Conference (VTC2022-Spring)*, 2022, pp. 1–6.

[139] C. D'Andrea, A. Garcia-Rodriguez, G. Geraci, L. G. Giordano, and S. Buzzi, "Analysis of UAV communications in cell-free massive MIMO systems," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 133–147, 2020.

[140] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Technical Report TR 38.901, Mar. 2022, online. Available: https://www.3gpp.org/ftp/Specs/html-info/38901.htm.

[141] F. R. V. Guimarães, G. Fodor, W. C. Freitas, and Y. C. B. Silva, "Pricing-based distributed beamforming for dynamic time division duplexing systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3145–3157, 2018.

[142] S. Mashdour, S. Salehi, R. C. de Lamare, A. Schmeink, and J. P. S. H. Lima, "Clustering and scheduling with fairness based on information rates for cell-free MIMO networks," *IEEE Wireless Communications Letters*, vol. 13, no. 7, pp. 1798–1802, 2024.

[143] M. Andersson, T. T. Vu, P. Frenger, and E. G. Larsson, "Joint optimization of switching point and power control in dynamic tdd cell-free massive mimo," in *2023 57th Asilomar Conference on Signals, Systems, and Computers*, 2023, pp. 988–992.

[144] J. Rachad, R. Nasri, and L. Decreusefond, "Dynamic-tdd interference tractability approaches and performance analysis in macro-cell and small-cell deployments," *Annals of Telecommunications*, vol. 75, p. 739–753, 2020.

[145] R. Nikbakht and A. Lozano, "Uplink fractional power control for cell-free wireless networks," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–5.

[146] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, 2017.

[147] J. Choi, "Noma: Principles and recent results," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, 2017, pp. 349–354.

[148] P. N. Thakre and S. B. Pokle, "A survey on power allocation in pd-noma for 5g wireless communication systems," in *2022 10th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-22)*, 2022, pp. 1–5.

[149] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain noma with enabling technologies of future wireless networks," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2192–2235, 2020.

[150] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of noma: Current status and open research challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.

[151] R. Jiao, L. Dai, W. Wang, F. Lyu, N. Cheng, and X. Shen, "Max-min fairness for beamspace mimo-noma: From single-beam to multi-beam," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 739–752, 2022.

[152] L. Zhu, J. Zhang, Z. Xiao, and X. Cao, "User fairness non-orthogonal multiple access (noma) in millimeter-wave communications," in *2018 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2018, pp. 80–84.

[153] I. Abu Mahady, E. Bedeer, S. Ikki, and H. Yanikomeroglu, "Sum-rate maximization of noma systems under imperfect successive interference cancellation," *IEEE Communications Letters*, vol. 23, no. 3, pp. 474–477, 2019.

[154] X. Wang, R. Chen, Y. Xu, and Q. Meng, "Low-complexity power allocation in noma systems with imperfect sic for maximizing weighted sum-rate," *IEEE Access*, vol. 7, pp. 94 238–94 253, 2019.

[155] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a ds/cdma system," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 5, pp. 796–807, 1994.

[156] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 654–667, 2016.

[157] M. Moltafet, N. M. Yamchi, M. R. Javan, and P. Azmi, "Comparison study between pd-noma and scma," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1830–1834, 2018.

[158] H. Grama Srinath, S. Jha, and N. Mysore Balasubramanya, "An efficient nb-iot compatible gf-noma phy mechanism for mmtc," *IEEE Internet of Things Journal*, vol. 10, no. 20, pp. 17 949–17 963, 2023.

[159] F. Jameel, S. Zeb, W. U. Khan, S. A. Hassan, Z. Chang, and J. Liu, "Noma-enabled backscatter communications: Toward battery-free iot networks," *IEEE Internet of Things Magazine*, vol. 3, no. 4, pp. 95–101, 2020.

[160] Z. Hu, L. Xu, L. Cao, S. Liu, Z. Luo, J. Wang, X. Li, and L. Wang, "Application of non-orthogonal multiple access in wireless sensor networks for smart agriculture," *IEEE Access*, vol. 7, pp. 87 582–87 592, 2019.

[161] 3GPP TR 21.915, "Release 15 Description; Summary of Rel-15 Work Items," Tech. Rep. 21.915, 2019. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3389

[162] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Technical Report TR 38.901, 2020, release 16. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/138900_138999/138901/16.01.00_60/tr_138901v160100p.pdf

[163] Y. Sun, W. Cao, M. Zhou, and Z. Ding, "Hybrid successive interference cancellation and power adaptation: A win-win strategy for robust uplink noma transmission," *IEEE Transactions on Communications*, vol. 72, no. 2, pp. 771–785, 2024.

[164] L. Zhang, W. Li, Y. Wu, X. Wang, S.-I. Park, H. M. Kim, J.-Y. Lee, P. Angueira, and J. Montalban, "Layered-division-multiplexing: Theory and practice," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 216–232, 2016.

[165] C. Min and Y. Yongqin, "Iterative interference cancellation for non-orthogonal multiple access system," *Chinese Journal of Electronics*, vol. 29, no. 3, pp. 540–546, 2020. [Online]. Available: https://cje.ejournal.org.cn/en/article/doi/10.1049/cje.2020.03.009