

Projeto Brasil 6G

Tecnologias Habilitadoras e Estudo Exploratório de Algoritmos para Camada física e de Enlace para Redes 6G



30/11/2022



Histórico de Atualizações:

Versão	Data	Autor(es)	Notas		
1	15/11/2022	Aldebaro Klautau Cleverson Nahum Danilo Gaspar Davi da Silva Brilhante Joanna Manjarres Luan Gonçalves Luiz Fernando Gontijo José Ferreira de Rezende Luciano Leonel Mendes Mariana Mello Michelle Facina Paulo Cardieri Pedro H. C. de Souza Plínio Santini Dester Richard Demo Souza Victoria Dala Pegorara Souto	Elaboração de conteúdo		
2	30/11/2022	Daniely Gomes Silva José Cândido Silveira Santos Filho Luciano Leonel Mendes Paulo Cardieri Richard Demo Souza	Revisão de texto		

Lista de Figuras

1	Diagrama simplificado do cenário de Internet of Things (IoT), em que diversos dispositivos transmitem para a Base Station (BS) empregando o mesmo recurso de tempo-frequência. Os canais h_i são considerados independentes entre si devido ao distanciamento entre os dispositivos.	3
2	Exemplo de árvore de busca para detectores <i>Sphere Detector</i> (SD). Os nós cheios representam segmentos dos ramos que estão dentro do raio de busca, enquanto	0
3	que os nós vazios representam os segmentos que estão fora do raio de busca Desempenho dos detectores PDA, LRA-PDA, V-BLAST e MV. O cenário ilus- trado considera constelações (a)-(b) 16-QAM e (c)-(d) 64-QAM respectivamente	5
4	(para $N_{\rm t} = N_{\rm r} = 4$, 16, 32 MIMO)	13
	dulação Binary Phase-Shift Keying (BPSK), $S = 5$, $\mathcal{P} = 4$, $v_t = 0, 8$, $v_f = 1$, K = 5, $M = 5$ e filtro protótipo Dirichlet. (a) Canal Additive White Gaussian Noise (AWGN) (b) Canal Time-Invariant Frequency-Selective (TIFS) (c) Ca-	
F	nal <i>Time-Variant Frequency-Selective</i> (TVFS)	20
9	modulação BPSK, $S = 5$, $\mathcal{P} = 4$, $v_t = 1$, $v_f = 0.8$, $K = 6$, $M = 4$ e filtro	20
6	BER para o sistema FTN-GFDM com compressão no tempo e frequência, con- siderando modulação BPSK, $S = 5$, $\mathcal{P} = 4$, $v_t = 0.8$, $v_f = 0.75$, $K = 6$, $M = 5$	20
	e filtro protótipo <i>Root-Raised-Cosine</i> (RRC) com fator de <i>roll-off</i> $\alpha = 0.5$. (a) Canal AWGN. (b) Canal TIFS. (c) Canal TVFS	21
7	As curvas sólidas são os limitantes superior e inferior; a linha tracejada é a aproximação em (61); e as barras de erro são os desvios padrões das simulações de Monte Carlo. A parte inferior de barra de erro pão aparece cuando e desvio	
0	padrão é maior que a média.	27
8	Sistema de comunicação de <i>downlink</i> com um usuario assistido por uma Intelligent Reflective Surfaces (IRS)	30
9	Fluxograma da solução proposta.	32
10 11	Configuração do sistema	33
	$(x_{\rm u}, y_{\rm u}) = (95, 8) \text{ m}, \text{ ou seja}, d_{\rm IU} \sim 5 \text{ m}. \dots \dots$	34
12	Taxa alcançável versus distância entre BS e UE, para $M = 40. \dots \dots \dots$	34
13	Taxa alcançável no UE versus o número de elementos (M) na IRS	35
14	<i>Overhead</i> de treinamento da solução proposta, considerando $(x_u, y_u) = (95,8)$ m,	26
15	Modelo do sistema com $K = 2$ $N = 12$ e $M = 4$ (superior) e estrutura do	30
10	protocolo de treinamento/carregamento (inferior).	38
16	Principais regiões de operação. Para fins de ilustração, a posição dos dispositivos Energy Harvesting (EH) é dada por: $\boldsymbol{\zeta}_{k}^{\text{ue}} = [Lk/32, Lk/32, 2^{k/7-1}]$ para k impar,	
17	enquanto $\boldsymbol{\zeta}_{k}^{\text{ue}} = [L(9-k)/32, Lk/36, 2^{1-k/\ell}]$ para k par, com $K = 8, \ldots, \ldots$	39
11	de otimização (direita) em função do número de UEs	41



18	<i>Time Division Multiple Access</i> (TDMA) vs <i>Space Division Multiple Access</i> (SDMA) em termos de a) potência de transmissão total média (topo) e b) tempo médio de execução de otimização (fundo) ao servir um número crescente de UEs. Definimos	
	$\xi = 1/k$ para que a potência total coletada permaneça em 1 W, independente-	10
10	mente do número de usuários.	42
19	Usuarios calculam a atualização de cada modelo e enviam o modelo com maior	
	usuário acossar o mosmo canal há uma colisão. Um usuário podo tor ponhuma	
	uma ou várias atualizações para enviar em cada iteração	11
20	Norma de erro $ u (t) - u $ do modelo mais lento para diferentes abordagens:	44
20	Compartilhado (Shared) Uniforme (Uniform) Proporcional a τ_r (Compute) e o	
	Esquema Proposto (<i>Proposed</i>)	48
21	Norma de erro $ \boldsymbol{w}_n(t) - \boldsymbol{w}_n $ com $t = 1000$ do modelo mais lento para diferentes	10
	abordagens: Compartilhado (<i>Shared</i>), Uniforme (<i>Uniform</i>), Proporcional a τ_n	
	(Compute) e o Esquema Proposto (Proposed).	48
22	Exemplo de um processo de handover. A Reference Signal Received Power	
	(RSRP) da BS de serviço (em vermelho) fica abaixo da RSRP da BS alvo (em	
	azul) pelo período de <i>Time to Trigger</i> (TTT)	52
23	Representação de um sistema V2I com enlace LOS beam verde e enlace NLOS	
	beam vermelho e amarelo. Sendo que o beam amarelo representa a difração do	
	sinal chegando no receptor alvo com desfase e menor potencia.	60
24	Representação da seleção de feixe num sistema Multiple-Input Multiple-Output	
~ ~	(MIMO) millimeter wave (mmWave)	60
25 26	Exemplo de um discriminador da letra E na fase de treinamento \ldots	62
26	Exemplo de classificação de um numero usando a rede <i>Wilkes, Stonham and</i>	69
97	Aleksander Recognition Device (WISARD)	03 62
21 28	Exemple do uso do mecanismo <i>oleaciting</i>	05 65
20 20	Acurácia em função do hiper-parâmetro tamanho da memória da rede WISABD	65
30	Processo de schedulina em um cenário sem slices de rede no qual os <i>Badio</i>	00
00	Resource Groups (RBGs) disponíveis na BS precisam ser distribuídos para as	
	User Equipments (UEs) de forma a atender os objetivos para o qual o scheduler	
	foi projetado.	69
31	Processo de <i>scheduling</i> em um cenário com <i>slices</i> na rede de acesso em que	
	os RBGs disponíveis na BS precisam ser distribuídos para os slices (inter-slice	
	scheduling), e posteriormente cada slice distribui os seus RBGs disponíveis para	
	as UEs (<i>intra-slice scheduling</i>) de forma a atender os objetivos para o qual o	
	scheduler foi projetado	70
32	Diagrama de fluxo do funcionamento da classe principal Communication env. e	
	da classe BS. O <i>scheduler</i> , recebe as informações da rede e baseado nelas envia	
	a decisão de alocação de RBGs que será utilizada pelo simulador para gerar as	P 1
<u></u>	Diamanda da funcionamente de classe Clisse UE O de del de si	(1
აპ	Diagrama de fluxo do funcionamento das classes Slice e UE. O scheduler primeiro	
	distribuídos para as UEs associadas a ele	79
	unitinativo para as C_{12} associatas a cie	12



34	Quantidade de pacotes recebidos em cada passo n para cada UE. Os pacotes recebidos são armazenados nos <i>buffers</i> das UEs e são enviados de acordo com a	
	alocação do <i>scheduler</i> .	75
35	Recompensas acumuladas obtidas por cada agente de <i>scheduler</i> ao longo do	
	pacotes, onde o melhor agente é definido como o agente que faz o sistema perder	
	menos pacotes.	76
36	Diferenças entre os sistemas de comunicação convencional, semântico e baseado	
	em tarefa. No sistema convencional, a foto é transmitida integralmente, sem	
	qualquer atenção ao seu conteúdo; no sistema semântico, o transmissor envia	
	apenas a parte da foto que interessa ao destino; na comunicação baseada em	
	tarefa, o transmissor enviar o comando que causará a mudança do estado do	
	destino.	78
37	Estrutura típica de uma sistema de comunicação semântica fim-a-fim. Baseada	
	na Fig. 2 de [1]	80

Lista de Tabelas

1	Alocação de recursos em <i>Field-Programable Gate Array</i> (FPGA) para os detec-	
	tores SD convencional e SD-Affine Transform (AFT) considerando as técnicas	
	de implementação FX18 e FP	8
2	Complexidade computacional dos detectores cujo desempenhos são avaliados neste capítulo. Observe que os detectores estão dispostos em ordem crescente,	
	ou seja, do menos para o mais complexo a partir da primeira linha da tabela	12
3	Complexidade dos detectores Maximum Likelihood Sequence Estimation (MLSE),	
	SD, SSSgbKSE, SSSSE e FDE.	19
4	Valores dos parâmetros do canal usados nas simulações	19
5	Impacto de múltiplos modelos	49
6	Eventos de handover definidos pelo 3rd Generation Partnership Project (3GPP)	
	[2]	51
7	Abordagens para seleção de feixes assistido por contexto separados em três gru- pos. O grupo I composto por diversos tipos de entradas, o grupo II apresenta	
	dados de LiDAR complementado por dados de posição e imagens, e o grupo III	
	estrategias com entradas apenas de LiDAR	67
8	Parâmetros da simulação definidos no arquivo de configuração utilizado para	
	obtenção dos resultados.	74

Acrônimos

3D tridimensional **3GPP** 3rd Generation Partnership Project 4G Quarta Geração 5G Quinta Geração 6G Sexta Geração **AFT** Affine Transform **AP** Access Point AWGN Additive White Gaussian Noise ALS varredura a laser no ar AMCW onda contínua modulada em amplitude pulsada **B5G** Beyond Fifth Generation of Mobile Network **BER** Bit Error Rate **BPSK** Binary Phase-Shift Keying **BS** Base Station **CF-mMIMO** Cell-Free Massive MIMO **CNN** Convolutional Neural Network **C-RAN** Cloud-Radio Access Network **CSI** Channel-State Information **C-RAN** Cloud-Radio Area Network **DFT** Discrete Fourier Transformer **DNN** Deep Neural Network **D2D** Device-to-Device **DMI** Indicador de Medição de Distância **Deep RL** Deep reinforcement learning **DL** Deep Learning **EH** Energy Harvesting **eMBB** Enhanced Mobile Broadband eRAC Enhanced Remote Area Communications ECU unidade de computação eletrônica **FDE** Frequency-Domain Equalization FDP Função Densidade de Probabilidade **FLOP** Floating-Point Operation **FMCW** Frequency Modulated Continuous-Wave **FPGA** Field-Programable Gate Array



FTN Faster-than-Nyquist **FL** Federated Learning FMCW onda contínua modulada em frequência **FNN** Feedforward neural network **GD** Gradient Descent GFDM Generalized Frequency Division Multiplexing **GNSS** Global Navigation Satellite System **GPS** Global Positioning System **GA** Genetic Algorithms **GSN** Goal Seeking Neuron **GRAM** Generalization RAM **HIT** Handover Interruption Time IAI Inter Antenna Interference **ICI** Inter Carrier Interference **IoT** Internet of Things **IRS** Intelligent Reflective Surfaces **ISI** Inter Symbol Interference I-CSI Instantaneous Channel State Information IMU Unidade de Medição Inercial **IFML** improved fast machine learning **KPI** Key Performance Indicators **LiDAR** Light Detecting and Ranging LLL Lenstra, Lenstra, Lovász LR Lattice Reduction LTE Long Term Evolution LoS Line-of-Sight LRA-PDA Lattice Reduction Aided Probability Data Association **LSTM** long-short term memory **LNN** lightweight Neural Network MAB Multi-Armed Bandit **MF** Matched Filter **MIMO** Multiple-Input Multiple-Output **MISO** Multiple-Input Single-Output **mMIMO** Massive MIMO ML Machine Learning



MLSE Maximum Likelihood Sequence Estimation **MMSE** Minimum Mean Square Error **mMTC** massive Machine Type Communications **MR** Measurement Report MRT Maximum Ratio Transmission **MToS** Minimal Time of Stay **MV** Máxima Verossimilhança **MPE** Maximum Permissible Exposure **mmWave** millimeter wave MILP Mixed Integer Linear Programming **NLoS** Non-Line-of-Sight **NR** New Radio **OFDM** Orthogonal Frequency Division Multiplexing **PB** Power Beacon **PDA** Probability Data Association **PIC** Parallel Interference Cancellation **PPP** Processo Pontual de Poisson **QAM** Quadrature Amplitude Modulation **QoS** Quality of Service **RACH** Random Access Shared Channel **RAT** Radio Access Technology **RBG** Radio Resource Group **RL** Reinforcement Learning **RF** Rádio Frequência **RRC** Root-Raised-Cosine **RRM** Radio Resource Management **RRS** Radio Resource Scheduling **RSRP** Reference Signal Received Power **RA** Random Access **RNSP** Rede Neural Sem Peso **RNN** Recurrent Neural Networks **RSSI** Received Signal Strength Indicator **SAC** Soft Actor-Critic SCA Successive Convex Approximation **SD** Sphere Detector



SDP Semidefinite Programs **SDMA** Space Division Multiple Access **SER** Symbol Error Rate SINR Signal-to-Interference-and-Noise Ratio **SM-MIMO** Space Multiplexing Multiplo Input Multiple Output **SNR** Signal-to-Noise Ratio SSSgbKSE Successive Symbol-by-Symbol with go-back K Sequence Estimator **SSSSE** Successive Symbol-by-Symbol Sequence Estimator S-CSI Statistical Channel State Information **TDMA** Time Division Multiple Access THz Terahertz **TIFS** Time-Invariant Frequency-Selective **ToS** Time of Stay **TTI** Transmission Time Interval **TTT** Time to Trigger **TVFS** Time-Variant Frequency-Selective **ToF** *Time-of-Flight* **TLS** varredura a laser terrestre **VoIP** Voice over Internet Protocol **UAV** Unmanned Aerial Vehicle **UE** User Equipment **UHF** Ultra High Frequency **URLLC** Ultra Reliable Low Latency Communications **ULA** Uniform Linear Array **UPA** Uniform Planar Array **VHF** Very High Frequency **V2I** Vehicle-to-Infrastructure **VG-RAM** Virtual Generalization RAM **V2I** Vehicle to Infrastructure **WET** Wireless Energy Transfer **WSN** Wireless Sensors Network WiSARD Wilkes, Stonham and Aleksander Recognition Device **ZF** Zero Forcing

Sumário

1	Introdução
2	Utilização da Transformada Affine para Redução de Complexidade do Sphere Detector 2.1 Introdução 2.2 Modelo de Sistema 2.3 Detectores MIMO Convencionais 2.4 Integração da Transformada Affine com o Sphere Detector 2.5 Análise de Complexidade 2.6 Conclusão
3	Lattice Reduction Integrado ao PDA para Melhoria de Desempenho em Sistemas MIMO de Baixa Ordem e Constelações Densas
	3.1Introdução93.2Modelo de Sistema93.3Detectores PDA e LRA-PDA93.4Análise de Complexidade e de Desempenho em termos de SER163.5Conclusão14
4	Utilização do Sphere Detector para Detecção de Sinais FTN-GFDM emCanais Duplamente Dispersivos184.1 Introdução184.2 Descrição do FTN-GFDM184.3 SD para Detecção de Sinais FTN-GFDM164.4 Análise de Complexidade174.5 Análise de Desempenho194.6 Conclusão20
5	Geometria Estocástica e Redes Cell-Free225.1Introdução225.2Resultados da Literatura de Geometria Estocástica225.3Modelo do Sistema235.4Resultados Espaciais em Redes Cell-Free235.5Conclusão24
6	Projeto de Beamforming na IRS baseado em S-CSI296.1Modelo do Sistema296.2Formulação do Problema316.3Projeto de Beamforming com S-CSI326.4Resultados326.5Conclusão36
7	Transferência de Energia Sem Fio usando Radio Stripes377.1Modelo do Sistema377.2Formulação do Problema387.3Otimização com Restrições de Radiação39

Brasil 65

	$7.4 \\ 7.5$	Resultados	40 42
8	Fedd 8.1 8.2 8.3 8.4 8.5	erated Learning usando ALOHA Multicanal Modelo do Sistema	43 43 44 45 46 48
9	Han 9.1 9.2 9.3 9.4	adover e Mobilidade em Redes 6G Introdução	50 50 51 54 56
10	<i>Bea</i> 10.1 10.2 10.3 10.4 10.5 10.6 10.7	m Selection com o Uso de LiDARs Introdução LiDAR Aprendizado de Máquina Aplicado a Seleção de Feixe Seleção de Feixe Usando Lidar Rede Neural Sem Peso Resultados Parciais 10.6.1 Adquisição dos dados No.2 Pre-processamento dos dados Resultados Parciais Conclusão	57 57 59 61 64 64 64 64 64 64 65
11	Aml diza 11.1 11.2 11.3	biente de Simulação para Alocação de Recursos de Rádio para Apren- do por Reforço Simulador	68 68 74 75
12	Com 12.1 12.2 12.3 12.4 12.5 12.6 Con	nunicação Semântica Introdução	 77 77 79 80 81 81 82 84



1 Introdução

Estudos recentes, como aquele apresentado no relatório "Estado da Arte da Camada Física para Redes de Acesso 6G" [3], elaborado na Fase I desse projeto, mostram que os casos de uso esperados para as redes de Sexta Geração (6G) exigirão sistemas de comunicação com requisitos mais rigorosos, como taxa de dados de até 1 Tbps, baixa latência, da ordem de décimos de milissegundos, e atendimento de até 100 dispositivos por m³. É um consenso na comunidade científica que tais requisitos somente serão atendidos como o emprego de tecnologias de comunicação inovadoras, que explorem de forma mais eficientes os recursos rádio.

Em continuidade aos estudos iniciados na Fase I e reportados em [3], trabalhos de pesquisa estão sendo realizados na Fase II do projeto com o objetivo de explorar tecnologias habilitadoras inovadoras nas camadas física e de enlace para as redes 6G. O presente relatório apresenta os resultados dessas pesquisas obtidos em 2022. Uma breve descrição de conteúdo de cada capítulo do relatório é apresentada a seguir.

O **Capítulo 2** trata do problema da complexidade da detecção de sinais em sistemas empregando a tecnologia *Space Multiplexing Multiplo Input Multiple Output* (SM-MIMO). Essa tecnologia possibilita atingir alta vazão em um grande número de conexões, mas às custas de alta complexidade na recepção. Os autores avaliam técnicas para reduzir essa complexidade, de forma a tornar a técnica SM-MIMO atrativa em cenários com dispositivos com baixa capacidade de processamento.

A complexidade de detectores é também o foco no **Capítulo 3**, mas agora em sistemas empregando a tecnologia MIMO massivo. O uso intenso dessa tecnologia degrada o desempenho dos detectores pela perda da ortogonalidade das transmissões. Nesse capítulo, os autores estudam o uso da técnica *Lattice Reduction* em conjunto com o detector *Probability Data As*sociation (PDA), que mostrou ter desempenho satisfatório e baixa complexidade.

Diversos casos de uso das redes 6G exigirão a provisão de altas taxas de dados em regiões com alta densidade de terminais. No **Capítulo 4**, os autores estudam o uso da forma de onda *Faster-than-Nyquist – Generalized Frequency Division Multiplexing* combinada com a detecção *Sphere Detector* para atender esses casos de uso, e mostram que o desempenho em termos de BER desse esquema em ambientes dispersivos é quase ótimo, com complexidade reduzida.

O **Capítulo 5** apresenta um estudo sobre as redes de comunicação sem fio empregando a arquitetura *cell-free*, em que a configuração dos enlaces de comunicação é baseada na posição conjunta dos usuários. Usando ferramentas de geometria estocástica, os autores propõem a definição de *usuários vizinhos* em um processo pontual no contexto de *cell-free*, e caracterizam as distribuições envolvendo a vizinhanças entre terminais de usuários e pontos de acesso.

No **Capítulo 6**, os autores consideram um sistema de comunicação MISO auxiliado por uma *Intelligent Reflective Surface* (IRS) e propõem uma nova abordagem para projetar a formação dos feixes na IRS. Um aspecto inovador da abordagem proposta é a não exigência do conhecimento explícito do canal. Os resultados mostram que a solução apresentada pode atingir um desempenho próximo ao da solução ideal, mas exigindo baixa sobrecarga de treinamento.

O **Capítulo 7** investiga uma tecnologia utilizada para a transferência de energia sem fio via rádio frequência, usando esquemas de *energy beamforming* ideais e sub-ótimos, avaliados para um sistema *cell-free* com *radio stripes*. Também foram apresentadas as vantagens e desvantagens de diferentes técnicas de múltiplo acesso empregadas nesse sistema.

O Capítulo 8 propõe um protocolo ALOHA multicanal combinado à técnica de esparsifica-

Capítulo escrito por Daniely Gomes Silva, José Cândido Silveira Santos Filho, Luciano Leonel Mendes, Paulo Cardieri e Richard Demo Souza.



ção a fim de reduzir e equalizar o tempo de convergência em um sistema *Federated Learning* (FL) multimodelo. O esquema proposto apresenta uma convergência significativamente mais rápida que a de esquemas de referência na literatura.

O **Capítulo 9** descreve a operação de mobilidade em redes de geração avançada por meio de *handover*, bem como seus desafios e abordagens da literatura para superá-los. Em particular, apresenta um esquema via aprendizado de máquina para reduzir a taxa de *handover* e o número de *handovers* causados pelo efeito *ping-pong*, considerando um cenário de mobilidade na banda de mmWave com o efeito de obstáculos estáticos causando enlaces *Non-Line-of-Sight* (NLoS).

No **Capítulo 10**, o uso de dados de sensores *Light Detecting And Ranging* (LiDAR), em conjunto com técnicas de aprendizado de máquina, é explorado para a seleção de feixes de sistemas de múltiplas antenas em enlaces com e sem linha de visada, permitindo maior eficiência espectral e redução da latência.

O **Capítulo 11** apresenta o desenvolvimento de um ambiente de simulação para redes 6G, utilizando aprendizado por reforço. O objetivo aqui foi fornecer um simulador simples e abrangente, que se integre com ferramentas e bibliotecas externas, e que possa ser explorado para comparação justa de diferentes métodos de gestão de recursos de rádio.

O **Capítulo 12** apresenta conceitos relacionados à comunicação semântica, que tem recebido considerável atenção nos últimos anos dado o seu potencial de levar a um uso mais eficiente dos recursos rádios. Essa forma de comunicação explora o contexto e o propósito da transmissão de uma mensagem, visando diminuir a quantidade de bits que precisam ser transmitidos e, consequentemente, aumentando assim a eficiência no uso dos recursos rádio.

O Capítulo 13 apresenta as conclusões gerais do relatório.



2 Utilização da Transformada Affine para Redução de Complexidade do *Sphere Detector*

2.1 Introdução

O advento da IoT trouxe uma alta demanda por conectividade para as redes móveis e a Rede Quinta Geração (5G) já prevê um modo de operação específico para este fim, denominado de massive Machine Type Communications (mMTC). No entanto, a necessidade de obter informações e de atuar nos mais diversificados sistemas de forma dinâmica vêm impulsionando um crescimento mais acentuado por conectividade IoT nos últimos anos. Para as Redes 6G, espera-se que esse elevado número de conexões esteja associado ao aumento de vazão por cada nó da rede, em função da maior complexidade das aplicações previstas para as próximas redes móveis.

Uma das abordagens previstas para o aumento da vazão e do número de conexões nas futuras redes móveis consiste em utilizar *Space Multiplexing Multiplo Input Multiple Output* (SM-MIMO), onde um mesmo recurso tempo-frequência, como, por exemplo, uma dada subportadora de um símbolo *Orthogonal Frequency Division Multiplexing* (OFDM) nas redes Quarta Geração (4G) e 5G, é simultaneamente compartilhado por diversos dispositivos localizados em diferentes pontos espaciais [4, 5]. A Fig. 1 apresenta uma diagrama simplificado do enlace de subida de um sistema empregando SM-MIMO. Maiores detalhes sobre o modelo de sistema serão apresentados na subseção a seguir.



Figura 1: Diagrama simplificado do cenário de IoT, em que diversos dispositivos transmitem para a BS empregando o mesmo recurso de tempo-frequência. Os canais h_i são considerados independentes entre si devido ao distanciamento entre os dispositivos.

O conteúdo deste capítulo foi publicado em D. Gaspar, L. L. Mendes, and T. C. Pimenta, "Sphere detector with low complexity Euclidean distance computation based on affine transform modulation," in Electronics Letters, vol. 58, no. 20, pp. 779-782, 2022, https://doi.org/10.1049/ell2.12594.



2.2 Modelo de Sistema

No cenário apresentado na Fig. 1, assume-se que a BS possui um número de antenas, $N_{\rm R}$, maior do que o número de dispositivos transmitindo simultaneamente, $N_{\rm T}$, sendo que cada dispositivo emite um símbolo *Quadrature Amplitude Modulation* (QAM), d, de uma constelação contendo M símbolos. Desta forma, o vetor recebido pela BS no k-ésimo recurso tempo-frequência (por exemplo, na k-ésima subportadora de um símbolo OFDM), é dado por

$$\mathbf{Y}^{(k)} = \mathbf{H}^{(k)} \mathbf{d}^{(k)} + \mathbf{W}^{(k)},\tag{1}$$

em que $\mathbf{H}^{(k)}$ é a matriz de dimensão $N_{\mathrm{R}} \times N_{\mathrm{T}}$ contendo os ganhos do canal entre as antenas transmissoras (dispositivos) e as antenas receptoras da BS, considerando o k-ésimo recurso de tempo-frequência, $\mathbf{d}^{(k)}$ é o vetor de dados transmitidos pelos N_{T} dispositivos no k-ésimo recurso de tempo-frequência e $\mathbf{W}^{(k)}$ é o vetor de ruído AWGN. Por simplicidade, a notação referente ao recurso tempo-frequência será suprimida no restante desta seção.

2.3 Detectores MIMO Convencionais

O desafio no cenário descrito na Fig. 1 consiste em separar os dados enviados pelos $N_{\rm T}$ dispositivos, eliminando a *Inter Antenna Interference* (IAI) [6]. Existem diversas propostas de detectores que podem ser empregados para se obter uma estimativa dos dados transmitidos em um sistema SM-MIMO [7]. O detetor por Máxima Verossimilhança (MV) apresenta o desempenho ótimo em termos de probabilidade de erro, avaliando todas as possibilidades do vetor **d** e assumindo o conhecimento do canal, sendo que o vetor estimado é dado por

$$\hat{\mathbf{d}} = \arg\min_{\mathbf{d}\in\mathbf{M}^{N_{\mathrm{T}}}} ||\mathbf{Y} - \mathbf{H}\mathbf{d}||_{2}^{2}, \tag{2}$$

em que $\mathbf{M}^{N_{\mathrm{T}}}$ é o conjunto de todas as possibilidades para o vetor \mathbf{d} .

Embora atinja o desempenho ótimo em termos de *Symbol Error Rate* (SER), o detector MV tem uma complexidade de implementação que cresce exponencialmente com o número de dispositivos e com a ordem de modulação, tornando a sua aplicação prática proibitiva. Detectores lineares de baixa complexidade podem ser obtidos através da inversão da matriz do canal, sendo os detectores *Zero Forcing* (ZF) e *Minimum Mean Square Error* (MMSE) os mais comuns. No entanto, estes detectores impõem uma perda significativa de desempenho em termos de SER, principalmente porque estes detectores são incapazes de aproveitar a diversidade presente no canal de comunicação.

O Sphere Detector (SD) é um detector proposto para atingir o desempenho do detector MV em termos de SER, porém com uma complexidade que seja tolerável para uma implementação prática. O princípio por trás do SD consiste em restringir o espaço de busca dentro de uma esfera multidimensional do universo de possibilidade de $\mathbf{M}^{N_{\rm T}}$, de modo a não ser necessário realizar a busca entre todos os possíveis vetores **d** para se determinar o vetor transmitido com a maior probabilidade. Assim, pode-se considerar que o SD restringe a busca aos vetores hipóteses que estejam a uma distância do vetor recebido que seja menor do que o raio da hiper esfera, sendo que cada vetor hipótese pode ser visto como um caminho de nós sequencialmente conectados em uma estrutura de árvore finita, como mostra a Fig. 2 para o caso de $N_{\rm T} = 3$ e M = 2.

Apesar de a complexidade do SD ser menor do que aquela do detector MV, o primeiro ainda apresenta uma complexidade com crescimento exponencial. De fato, se o raio de busca permanecer constante e for suficientemente grande, a complexidade do SD tende para a complexidade





Figura 2: Exemplo de árvore de busca para detectores SD. Os nós cheios representam segmentos dos ramos que estão dentro do raio de busca, enquanto que os nós vazios representam os segmentos que estão fora do raio de busca.

do MV. Por outro lado, se o raio de busca for muito pequeno, pode acontecer de nenhum vetor válido estar presente dentro da hiper esfera, demandando que o algoritmo seja reiniciado com um raio de busca maior. Uma solução prática para este problema consiste em iniciar o raio de busca com base em um vetor válido obtido a partir de um detector de baixa complexidade (ZF ou MMSE) e atualizar esse raio de busca toda vez que uma hipótese mais favorável for encontrada no processo de busca. Isso significa que o cálculo do novo raio de busca é executado de forma intensa no algoritmo do SD e uma simplificação neste cômputo pode levar a uma redução de complexidade significativa para o algoritmo.

2.4 Integração da Transformada Affine com o Sphere Detector

A proposta apresentada pelos pesquisadores do Projeto Brasil 6G [8] visa buscar exatamente esta simplificação através do uso da transformada Affine [9]. Essa transformada permite que, em dadas circunstâncias, o mapeamento dos bits em símbolos M_u -QAM possa ser realizado através de uma operação linear. Neste caso, os símbolos QAM podem ser expressos em função do vetor **b** que traz a pilha de bits a serem transmitidos, tendo o bit menos significativo no topo do vetor, de modo que

$$\mathbf{d} = \mathbf{A}_M \mathbf{b} + \beta_M,\tag{3}$$

em que \mathbf{A}_M é uma matriz esparsa de dimensão $N_T \times \mu N_T$, sendo $\mu = \log_2(M)$, e β_M é um escalar complexo. É importante observar que tanto \mathbf{A}_M quanto β_M dependem da ordem da modulação. Assumindo baixas ordens de modulação, adequadas para as aplicações de IoT, tem-se

$$\beta_M = \begin{cases} -1 & \text{para } M = 2, \\ -1 + j & \text{para } M = 4, \\ -3 + j & \text{para } M = 8, \end{cases}$$
(4)

enquanto que

$$\mathbf{A}_{M} = \begin{cases} 2 & \text{para } M = 2, \\ 2(\mathbf{G}_{1} - j\mathbf{G}_{0}) & \text{para } M = 4, \\ 4\mathbf{G}_{2} + 2\mathbf{G}_{1} - j(2\mathbf{G}_{1} - 4\mathbf{G}_{0}) & \text{para } M = 8, \end{cases}$$
(5)

onde assumiu-se o uso de uma constelação não quadrada com mape
amento Gray para o caso em que M = 8.



A matriz de seleção \mathbf{G}_i , com $i \in \{0, 1, \dots, \mu - 1\}$, é esparsa, com dimensão $N_{\mathrm{T}} \times \mu N_{\mathrm{T}}$ e possui um único elemento binário em cada linha, seguindo a lei de formação dada por

$$\mathbf{G}_{i_{l,k}} = \begin{cases} 1 & k = \mu(l-1) + i + 1, \text{ para } l = 1, 2, \dots, N_{\mathrm{T}}, \\ 0 & \text{caso contrário.} \end{cases}$$
(6)

Por exemplo, a matriz de seleção para o bit 0, \mathbf{G}_0 , assumindo $N_{\rm T} = 3$ e $\mu = 2$, é dada por

Aplicando (3) em (1) resulta em

$$\mathbf{Y} = \mathbf{H}\mathbf{A}_M\mathbf{b} + \mathbf{H}\beta_M + \mathbf{W}.$$
 (8)

Fazendo $\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{H}\boldsymbol{\beta}$, tem-se

$$\tilde{\mathbf{Y}} = \mathbf{H}\mathbf{A}_M\mathbf{b} + \mathbf{W},\tag{9}$$

que representa os sinais recebidos no k-ésimo recurso tempo-frequência nas $N_{\rm R}$ antenas receptoras. Fica claro que (9) é uma representação linear dos bits transmitidos usando o mesmo recurso tempo-frequência.

Aplicando a fatorização QR de Householder na matriz do canal, tem-se que

$$\mathbf{QR} = \mathrm{HQR}(\mathbf{H}). \tag{10}$$

Aplicando (10) em (9) e multiplicando ambos lados da equação resultante por \mathbf{Q}^{H} , tem-se

$$\mathbf{Q}^{\mathrm{H}}\tilde{\mathbf{Y}} = \mathbf{Q}^{\mathrm{H}}\mathbf{Q}\mathbf{R}\mathbf{A}_{M}\mathbf{b} + \mathbf{Q}^{\mathrm{H}}\mathbf{W}$$
$$\hat{\mathbf{Y}} = \hat{\mathbf{R}}\mathbf{b} + \hat{\mathbf{W}}.$$
(11)

Como \mathbf{Q} é uma matriz ortonormal, a solução de (11) equivale à solução de (8). Desta forma, o SD para este caso precisa encontrar a solução para o o problema de otimização

$$\hat{\mathbf{b}}_{\mathrm{SD}} = \operatorname{argmin}_{\tilde{\mathbf{b}}_i = \mathcal{M}^{-1}(\tilde{\mathbf{d}}_i)} || \hat{\mathbf{Y}} - \hat{\mathbf{R}} \tilde{\mathbf{b}}_i ||_2^2,$$
(12)

em que $\mathcal{M}^{-1}(\cdot)$ é a função que faz o mapeamento inverso de símbolos QAM em bits. Como $\hat{\mathbf{R}}$ mantém a sua estrutura triangular superior, a abordagem de busca através dos nós de uma árvore continua sendo aplicável para a solução de (12). Observe que $\hat{\mathbf{R}}$ possui $(N_{\rm T} + 1 - n)\mu$ entradas não nulas em sua *n*-ésima linha, o que significa que uma sequência de μ bits é detectada através do procedimento de busca.

O Algoritmo 1 descreve os passos necessários para realizar a detecção da sequência de bit transmitida pelas $N_{\rm T}$ antenas, assumindo a abordagem proposta nesta seção.

2.5 Análise de Complexidade

A complexidade do SD integrado com a Transformada Affine depende primordialmente da complexidade das funções presentes na rotina SD *function*, descritas pelas linhas 7 à 20 no Algoritmo 1. Dentre essas funções, o cálculo da distância Euclidiana nas linhas 10 e 11 se



Algorithm 1 SD empregando a Transformada Affine.

Result: $\hat{\mathbf{b}}_{sd}$ **QR** = HQR(**H**); $\tilde{\mathbf{Y}} = \mathbf{Q}^{\mathrm{H}} (\mathbf{Y} - \mathbf{H}\beta);$ $\underline{\mathbf{R}} = \mathbf{R}\boldsymbol{\alpha};$ $\rho = ||\tilde{\mathbf{Y}} - \underline{\mathbf{R}}\hat{\mathbf{b}}||^2;$ $\hat{\mathbf{b}}_{\mathrm{sd}} = \hat{\mathbf{b}};\\ \ell = N_{\mathrm{T}};$ function SD (ℓ) : 7 for $s \leftarrow 1$ to M_c do 8 $\mathbf{b}_{\mu(\ell-1)+1:\mu\ell} = \mathcal{M}^{-1}(d_s);$ 9 $\hat{\rho}_{\ell} = \left| \tilde{y}_{\ell} - \sum_{i=\mu(\ell-1)+1}^{\mu N_{\mathrm{T}}} \underline{r}_{\ell,i|\tilde{b}_i=1} \right|^2;$ 10 if $(\sum_{i=\ell}^{l} \hat{\rho}_i < \rho)$ then 11 $\mathbf{i}\mathbf{f}^{i}(\ell == 1)$ then 12 13 $\mathbf{b}_{sd} = \mathbf{b};$ $\rho = \sum_{i=1}^{N_{\mathrm{T}}} \hat{\rho}_i;$ 14 15 else 16 SD $(\ell - 1)$; end 17 18 end end 19 end function 20

destacam. Como a linha 11 é a mesma tanto para a proposta apresentada nesta seção quanto para o algoritmo SD convencional, pode-se empregar apenas a linha 10 na análise comparativa de complexidade entre estas duas abordagens.

Considerando o pior cenário possível, onde todos os ramos da árvore de busca são percorridos para encontrar a sequência de bits transmitida com maior probabilidade, a complexidade do SD é dada por

$$\mathcal{O}_{t} = M^{N_{T}} \sum_{l=1}^{N_{T}} \mathcal{O}_{l}, \qquad (13)$$

em que \mathcal{O}_l é a complexidade para analisar um nó do *l*-ésimo ramo da árvore de busca, dada por

$$\mathcal{O}_{l} = \begin{cases} \sum_{i=l}^{N_{\mathrm{T}}} 8(N_{\mathrm{T}} - i) + 11, & \text{para o SD convencional,} \\ \sum_{i=l}^{N_{\mathrm{T}}} 2\mu(N_{\mathrm{T}} - i + 1) + 3 & \text{para o SD com transformada Affine.} \end{cases}$$
(14)

Encontrando a convergência dos somatórios em (14) e aplicando os resultados em (13), temse que

$$\mathcal{O}_{t}^{SD} = \frac{M^{N_{T}} N_{T} \left(8N_{T}^{2} + 33N_{T} + 25\right)}{6} \\ \mathcal{O}_{t}^{SD-AFT} = \frac{M^{N_{T}} N_{T} \left(2\mu N_{T}^{2} + (6\mu + 9)N_{T} + 4\mu + 9\right)}{6},$$
(15)

para o SD convencional e o SD com transformada Affine (SD-AFT), respectivamente.

A redução de complexidade obtida pela transformada Affine pode ser melhor analisada através da segunda derivada em relação a $N_{\rm T}$ da razão entre as complexidades apresentadas em (15), ou seja,

$$\eta = \frac{\partial^2 \left(\mathcal{O}_{\rm t}^{\rm SD} / \mathcal{O}_{\rm t}^{\rm SD-AFT} \right)}{\partial^2 N_{\rm T}} = \frac{2\mu}{8}.$$
(16)



O resultado apresentado em (16) mostra que a transformada Affine apresenta redução de complexidade quando as modulações com 2, 4 e 8 símbolos são empregadas. Para M = 16 não observa-se redução de complexidade e para ordens de modulação maiores, o uso da transformada Affine resulta em maior complexidade do que aquela do SD convencional.

Os algoritmos para o SD convencional e para o SD-AFT foram implementados utilizando linguagem descritiva de hardware para o FPGA Intel Cyclone V modelo 5CSEMA6F31C7, considerando $NT = N_{\rm R} = 4$ e apenas uma subportadora do símbolo OFDM.

A Tabela 1 compara os recursos do FPGA empregados nas suas implementações, considerando $\mu \in \{1, 2, 3\}$. Duas formas de implementação foram consideradas: a primeira emprega notação de ponto fixo com 18 bits (FX18), enquanto que a segunda emprega notação de ponto flutuante de 32 bits (FP). Observando os resultados apresentados na Tabela 1, pode-se concluir

Tabela 1: Alocação de recursos em FPGA para os detectores SD convencional e SD-AFT considerando as técnicas de implementação FX18 e FP.

Inner product entities		Area paran	Latency parameters		$\hat{\eta}_{\mu} = \frac{1}{2} \frac{(\sum \text{area} \times \text{Latency}/\text{F}_{\text{max}})_{\text{SD-ATM}}}{(\sum \text{area} \times \text{Latency}/\text{F}_{\text{max}})_{\text{conv. SD}}}$				
	ALMs	ALUTs	DLRs	DSPs	F _{max} [MHz]	Latency [T _{clk}]	$\mu = 1$	$\mu = 2$	$\mu = 3$
$\mathbf{R} \times \mathbf{d} \text{ FXP18} \text{ (conv. SD)}$	677.5 (1.62%)	1182 (1.41%)	261 (0.22%)	0 (0%)	77.975	$4(N_{\rm T}-\ell+1)$	0.158	0.268	0.378
$\mathbf{R} \times \mathbf{b} \text{ FXP18} (\text{SD-ATM})$	45.0 (0.11%)	44 (0.05%)	116 (0.01%)	0 (0%)	177.465	$2\mu(N_{\rm T}-\ell+1)$	0.150	0.200	0.570
$\mathbf{R} \times \mathbf{d} \operatorname{FP}(\operatorname{conv.} \operatorname{SD})$	1310.0 (3.12%)	2032 (2.42%)	1014 (0.87%)	4 (3.57%)	71.925	$10(N_{\rm T}-\ell+1)$	0.300	0.571	0.752
$\mathbf{R} \times \mathbf{b} \operatorname{FP}(\operatorname{SD-ATM})$	567.5 (1.35%)	876 (1.05%)	377 (0.32%)	0 (0%)	79.430	$4\mu(N_{\rm T}-\ell+1)$	0.390	0.571	0.752

que a área consumida para a implementação do SD-AFT é consideravelmente menor do que aquela consumida para a implementação do SD convencional. Reduções de áreas mais expressivas são observadas para a implementação com ponto fixo onde, no pior caso analisado ($\mu = 3$), a área empregada pelo SD-AFT foi de apenas 38% da área ocupada pelo SD convencional. Já para o caso da implementação com ponto flutuante, as reduções de complexidade são menos evidentes e se restringem aos valores observados em (16), com o SD-AFT consumindo 75% da área ocupada pelo SD convencional. A redução de complexidade pelo uso da transformada Affine também resulta em uma menor latência e maior frequência de operação do algoritmo implementação com ponto fixo é utilizada. Já com a implementação em ponto flutuante, o aumento da frequência de operação é marginal, sendo de apenas 10%.

2.6 Conclusão

Os resultados apresentados nesta seção mostram que o uso da transformada Affine permite uma redução de complexidade na implementação do algoritmo SD e um aumento significativo da frequência de operação, principalmente quando a implementação em hardware emprega a notação de ponto fixo e a ordem de modulação utilizada é baixa. Essa solução torna-se atrativa para o cenário em que os dispositivos empregados na rede possuem restrições de complexidade e não requerem o uso de modulações muito densas, o que está alinhado com as demandas associadas ao uso de dispositivos IoT em redes móveis.



3 *Lattice Reduction* Integrado ao PDA para Melhoria de Desempenho em Sistemas MIMO de Baixa Ordem e Constelações Densas

3.1 Introdução

O uso de sistemas MIMO aumentou substancialmente em sistemas celulares a partir das redes 4G, sobretudo nas redes 5G, com o MIMO massivo [7, 10]. No entanto, esse uso intenso de sistemas MIMO força a mitigação da *Inter Antenna Interference* (IAI) a fim de aumentar a ordem de diversidade espacial obtida [10]. A presença de IAI implica que os coeficientes da matriz do canal não pertencentes à diagonal principal não serão mutualmente ortogonais. Nesse sentido, a técnica *Lattice Reduction* (LR) [11] pode ser utilizada para aumentar o grau de ortogonalidade entre esses coeficientes e, assim, melhorar o desempenho de detecção em sistemas MIMO.

O detector *Probability Data Association* (PDA) [12], por sua vez, demonstra uma relação de compromisso satisfatória entre o desempenho de detecção e a complexidade computacional, mas também é penalizado pela IAI. Nesta seção, demonstra-se o detector baseado na integração do LR ao PDA [13]. Desta forma, é possível obter uma ordem de diversidade similar àquela do detector ótimo de MV, mas com uma complexidade comparativamente menor, pois o LR não altera significativamente a complexidade nativa do detector PDA. As subseções a seguir apresentam maiores detalhes sobre o modelo de sistema, o detector PDA e os resultados de desempenho encontrados.

3.2 Modelo de Sistema

Suponha que em um sistema de múltiplas antenas existam N_t antenas de transmissão e N_r antenas de recepção, constituindo um sistema MIMO $N_r \times N_t$. Bits de informação são demultiplexados em N_t feixes de bits que, em seguida, são mapeados em uma sequência de símbolos complexos pertencentes a uma constelação M-QAM de energia unitária. Os símbolos são transmitidos pela sua respectiva antena de transmissão usando OFDM, assumindo que o comprimento do prefixo cíclico supere o espalhamento temporal para todos os N_rN_t canais. Desta forma, temos um sistema MIMO ponto-a-ponto centralizado, cuja representação em banda básica do sinal recebido na k-ésima subportadora OFDM, após a realização da Discrete Fourier Transformer (DFT), pode ser escrita como

$$\tilde{\mathbf{r}}_k = \mathbf{H}_k \tilde{\mathbf{a}}_k + \tilde{\mathbf{n}},\tag{17}$$

em que $\tilde{\mathbf{H}}_k \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$ é a matriz de coeficientes da resposta em frequência do canal para a késima subportadora OFDM, $\tilde{\mathbf{a}}_k \in \mathbb{C}^{N_{\mathrm{t}}}$ representa o vetor de símbolos transmitidos na k-ésima subportadora, e $\tilde{\mathbf{n}} \in \mathbb{C}^{N_{\mathrm{r}}}$ é o ruído AWGN complexo no domínio da frequência, com média nula e matriz de covariância dada por $\sigma^2 \mathbf{I}_{N_{\mathrm{r}}}$.

Note que na sequência desta seção a representação em números reais será utilizada, de maneira que o sinal recebido indicado na expressão (17) seja representado pela concatenação

O conteúdo deste capítulo foi publicado em P. H. C. de Souza and L. L. Mendes, "Lattice Reduction Aided Probability Data Association Detector for MIMO Systems," in IEEE Communications Letters, vol. 26, no. 10, pp. 2375-2379, Oct. 2022, doi: 10.1109/LCOMM.2022.3190722.



das respectivas partes reais e imaginárias, tal que

$$\mathbf{r}_k = \mathbf{H}_k \mathbf{a}_k + \mathbf{n},\tag{18}$$

sendo que

$$\mathbf{r}_{k} = \left[\Re(\tilde{\mathbf{r}}_{k})^{T} \ \Im(\tilde{\mathbf{r}}_{k})^{T}\right]^{T} \in \mathbb{R}^{2N_{r}}, \ \forall \ k,$$
(19)

$$\mathbf{H}_{k} = \begin{bmatrix} \Re(\tilde{\mathbf{H}}_{k}) & -\Im(\tilde{\mathbf{H}}_{k}) \\ \Im(\tilde{\mathbf{H}}_{k}) & \Re(\tilde{\mathbf{H}}_{k}) \end{bmatrix} \in \mathbb{R}^{2N_{r} \times 2N_{t}}, \ \forall \ k,$$
(20)

como os vetores \mathbf{a}_k e \mathbf{n} seguindo a mesma forma de concatenação que aquela apresentada em (19).

3.3 Detectores PDA e LRA-PDA

No detector PDA, o sinal recebido \mathbf{r}_k , dado por (18), é primeiramente equalizado segundo o princípio do ZF, em que

$$\mathbf{z}_k = \mathbf{H}_k^{\dagger} \mathbf{r}_k = \mathbf{a}_k + \mathbf{v},\tag{21}$$

em que $\mathbf{H}_{k}^{\dagger} = (\mathbf{H}_{k}^{T}\mathbf{H}_{k})^{-1}\mathbf{H}_{k}^{T}$ representa a pseudo-inversa à esquerda de Moore-Penrose e $\mathbf{v} = \mathbf{H}_{k}^{\dagger}\mathbf{n}$ é o ruído AWGN com *noise enhancing*. Assim, ao reescrever (21) temos

$$\mathbf{z}_{k} = \mathbf{e}_{i} a_{k}(i) + \underbrace{\sum_{j \neq i} \mathbf{e}_{j} a_{k}(j) + \mathbf{v}}_{\mathcal{V}_{i}}, \forall i, j \in \{0, 1, \dots, 2N_{t} - 1\},$$
(22)

em que \mathbf{e}_i é o vetor com 1 (um) na *i*-ésima posição e 0 (zero) no restante, e \mathcal{V}_i é uma variável aleatória que pode ser interpretada como sendo o ruído efetivo contaminando $a_k(i)$ [12].

Observe que para detectar o símbolo transmitido pela *i*-ésima antena, pode-se assumir que todos os outros $j \neq i$ símbolos transmitidos são interferência somada ao ruído AWGN, o que é descrito por \mathcal{V}_i . Sendo assim, o detector PDA associa o vetor de probabilidades $\mathbf{p}_i \in \mathbb{R}^{\sqrt{M}}$ para cada símbolo transmitido $a_k(i)$, o que é calculado de acordo com $P_m(a_k(i) = q_m | \mathbf{z}_k, {\mathbf{p}_j}_{\forall j \neq i})$, para o qual $q_m \in \mathbb{S}$ é uma coordenada da constelação M-QAM e $m \in \{0, 1, \ldots, \sqrt{M} - 1\}$.

Considerando que \mathcal{V}_i possua distribuição Gaussiana, então as probabilidades são calculadas a posteriori de acordo com [14]

$$p_{i}(m) = \frac{\exp\left(\alpha_{m}(i)\right)}{\sum_{m=0}^{\sqrt{M}-1}\exp\left(\alpha_{m}(i)\right)},$$
(23)

em que

$$\alpha_m(i) = \left(\mathbf{z} - \boldsymbol{\mu}_i - 0.5\mathbf{e}_i q_m\right)^T \boldsymbol{\Omega}_i^{-1} \mathbf{e}_i q_m, \tag{24}$$

е

$$\boldsymbol{\mu}_{i} = \sum_{j \neq i} \mathbf{e}_{j} \left(\mathbf{q}^{T} \mathbf{p}_{j} \right), \text{ where } \mathbf{q} = \left[q_{0} \ q_{1} \ \dots \ q_{\sqrt{M}-1} \right]^{T}, \tag{25}$$

$$\mathbf{\Omega}_{i} = \sum_{j \neq i} \mathbf{e}_{j} \mathbf{e}_{j}^{T} \left(\left[\mathbf{q}^{2} \right]^{T} \mathbf{p}_{j} - \boldsymbol{\mu}_{j}^{2} \right) + 0.5 \sigma^{2} \mathbf{G}^{-1}.$$
(26)



Note que $\mathbf{G}^{-1} = (\mathbf{H}^T \mathbf{H})^{-1}$ é a inversa da matriz denominada Gram [7], que em seus elementos têm quantificados o grau de *noise enhancing*. Além disso, observe que o subscrito $(\cdot)_k$, indicando a *k*-ésima subportadora, é omitido com o intuito simplificar a notação.

Quanto mais a matriz \mathbf{G} se aproxima de uma matriz identidade, melhor é o desempenho de detecção de sistemas MIMO. Em outras palavras, a medida que os elementos de \mathbf{G} fora da diagonal principal se aproximam de zero, menor é a IAI e, consequentemente, melhor se torna o desempenho de detecção. Portanto, em seguida é apresentado o detector *Lattice Reduction Aided Probability Data Association* (LRA-PDA), em que o LR é utilizado para aprimorar as propriedades da matriz do canal, reduzindo, assim, a IAI na detecção realizada pelo PDA.

No entanto, primeiramente devem ser observadas algumas definições e conceitos sobre o LR. Um *lattice* ou reticulado é representado por uma base como, por exemplo, $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m)$, de modo que qualquer ponto do reticulado possa ser descrito por uma superposição de múltiplos inteiros dos vetores base \mathbf{b}_i . Sendo assim, considere $\mathbf{B} = \mathbf{H}$ e que o sinal recebido em cada subportadora não é contaminado pelo ruído, tal que

$$\mathcal{L}(\mathbf{H}) = \mathcal{L}(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{2N_t}) = \sum_{j=1}^{2N_t} \mathbf{h}_j a \, (j-1),$$
(27)

em que $\mathbf{h}_j \in \mathbb{R}^{2N_r}$ representa a *j*-ésima coluna da matriz do canal \mathbf{H} , a(j-1) denota a (j-1)ésima posição do vetor de símbolos transmitidos e $\mathcal{L}(\mathbf{H})$ representa o reticulado de números reais. Desta forma, o sinal recebido sem ruído pode ser visto como um arranjo periódico de pontos, isto é, um reticulado, para o qual os vetores base são as colunas da matriz do canal [11, 15].

O reticulado aprimorado, denotado por $\mathcal{L}(\dot{\mathbf{H}})$, é obtido por meio da transformação dos vetores base (ou colunas da matriz do canal) pela matriz unimodular \mathbf{T} [15], como segue:

$$\mathcal{L}(\dot{\mathbf{H}}) = \underbrace{\mathbf{HT}}_{\dot{\mathbf{H}}} \underbrace{\mathbf{T}^{-1} \mathbf{a}}_{\mathbf{h}}.$$
(28)

A matriz **T** realiza a LR e pode ser gerada por meio de algoritmos como o proposto por Lenstra, Lenstra, Lovász (LLL) [11], por exemplo. Contudo, para lançar mão do LR, é necessário que as coordenadas da constelação dos símbolos transmitidos permitam ser descritas em termos de números inteiros \mathbb{Z} . Isso é feito pela transladação e escalonamento dos pontos da constelação M-QAM [13, 15], por exemplo.

Desta forma, a equalização ZF considerando o reticulado aprimorado fica

$$\mathbf{z} = \dot{\mathbf{H}}^{\dagger} \mathbf{r} = \mathbf{u} + \dot{\mathbf{H}}^{\dagger} \mathbf{n}.$$
 (29)

Observe que a pseudo-inversa $\dot{\mathbf{H}}^{\dagger}$ está referenciada ao reticulado aprimorado. Isso implica na diminuição do noise enhancing por causa do maior grau de ortogonalidade entre as colunas de $\dot{\mathbf{H}}$. Sendo assim, o detector LRA-PDA calcula as probabilidades considerando $P_m(\bar{u}(i) = \bar{q}_m | \bar{\mathbf{z}}, \{\bar{\mathbf{p}}_j\}_{\forall j \neq i})$, para o qual $\bar{q}_m \in \mathbb{Z}_K$ é um elemento do subconjunto dos K números inteiros mais próximos de $\bar{z}(i)$, considerando também que $m \in \{0, 1, \ldots, K-1\}$ e $\bar{\mathbf{u}} \subset \mathbb{Z}^{2N_t}$. Note também que os possíveis pontos transmitidos $\mathbf{u} = \mathbf{T}^{-1}\mathbf{a}$ são agora desconhecidos, o que demanda a realização das seguintes operações após a equalização em (29):

$$\bar{\mathbf{z}} = \left[\frac{\mathbf{z}}{2E_0} - 0.5\mathbf{T}^{-1}\mathbf{1}_{2N_t} \right],\tag{30}$$



Tabela 2: Complexidade computacional dos detectores cujo desempenhos são avaliados neste capítulo. Observe que os detectores estão dispostos em ordem crescente, ou seja, do menos para o mais complexo a partir da primeira linha da tabela.

Detector	Complexidade Computacional
V-BLAST (PIC)	$\mathcal{O}(2N_{\rm t}^3 + \frac{5}{2}N_{\rm t}^2N_{\rm r} + 2N_{\rm t}N_{\rm r})$
PDA	$\mathcal{O}(N_{\rm t}^4 + \sqrt{M}N_{\rm t}^3 + N_{\rm t}^2N_{\rm r} + N_{\rm t}N_{\rm r})$
LRA-PDA	$\mathcal{O}(N_{\rm t}^4 + K N_{\rm t}^3 + N_{\rm t}^2 N_{\rm r} + N_{\rm t} N_{\rm r})$
MV	$\mathcal{O}(M^{N_{\mathrm{t}}})$

em que $E_0 = \sqrt{\frac{3}{2(M-1)}}$ e $\mathbf{1}_{2N_t}$ é o vetor com 1 (um) em todas as posições. Desta forma, as probabilidades calculadas pelo detector LRA-PDA são associadas ao conjunto finito de números inteiros \bar{q}_m , ao passo que o PDA as associa diretamente aos possíveis símbolos da constelação M-QAM, dado que $q_m \in \mathbb{S}$. O Algoritmo 2 apresenta os passos executados pelo detector LRA-PDA na detecção dos símbolos.

Algorithm 2 O detector LRA-PDA

```
 \begin{array}{l} \mathbf{Entrada:} \ \mathcal{L}(\dot{\mathbf{H}}) \ \text{via} \ (28) \ \text{and} \ \bar{\mathbf{z}} \ \text{como em} \ (30) \\ \mathbf{Entrada:} \ k_i \ (\text{veja} \ [16, \ \$II-C, \ p. \ 222]), \ \epsilon > 0 \\ \mathbf{Entrada:} \ \bar{p}_i \ (m) \leftarrow \frac{1}{K}, \ \forall \ m \ \forall \ i \\ \\ \mathbf{do} \\ \left| \begin{array}{c} \mathbf{for} \ i = 1, 2, \ldots, 2N_t \ \mathbf{do} \\ & \left| \begin{array}{c} \bar{\mathbf{p}}_i' \leftarrow \bar{\mathbf{p}}_i \\ \text{Compute} \ \boldsymbol{\mu}_{k_i} \ (25) \in \mathbf{\Omega}_{k_i} \ (26) \ \text{com} \ \{\bar{\mathbf{p}}_j\}_{\forall j \neq k_i} \ \mathrm{e} \ \bar{\mathbf{q}} \\ & \left| \begin{array}{c} \mathbf{for} \ m = 1, 2, \ldots, K \ \mathbf{do} \\ & \left| \begin{array}{c} \mathrm{Calcule} \ \alpha_m \ (k_i) \ (24) \\ & \mathrm{Calcule} \ P_m(\bar{u}(i) = \bar{q}_m | \bar{\mathbf{z}}, \{\bar{\mathbf{p}}_j\}_{\forall j \neq i}) \approx \bar{p}_{k_i} \ (m) \\ & \mathbf{end} \\ & \mathbf{end} \\ \end{array} \right| \\ \mathbf{end} \\ \mathbf{while} \ | \bar{\mathbf{p}}_i - \bar{\mathbf{p}}_i' | > \epsilon, \ \forall \ i \\ l_i \leftarrow \arg \max\{\bar{p}_i \ (m)\}, \ \forall \ i \\ \bar{z}(i) \leftarrow \overline{q}_{l_i}, \ \forall \ i \end{array} \right.
```

3.4 Análise de Complexidade e de Desempenho em termos de SER

A complexidade computacional do detector LRA-PDA, bem como de outros detectores analisados, incluindo o V-BLAST *Parallel Interference Cancellation* (PIC), é descrita na Tabela 2. Nesta tabela observa-se que a complexidade do detector LRA-PDA cresce a uma taxa polinomial com o número de antenas de transmissão N_t . Isso faz com que, como esperado, a sua complexidade seja significativamente menor que a apresentada pelo MV, que, por sua vez, cresce exponencialmente com N_t . É importante mencionar que a matriz **T** é computada no transmissor, antes da transmissão dos símbolos, não adicionando complexidade extra na detecção. Isso explica em boa parte a similaridade da complexidade entre o LRA-PDA e PDA.

A Figura 3 traz os resultados de desempenho de detecção para o LRA-PDA e outros detectores (veja Tabela 2) para fins de comparação. A métrica de desempenho é a taxa de erros de





Figura 3: Desempenho dos detectores PDA, LRA-PDA, V-BLAST e MV. O cenário ilustrado considera constelações (a)-(b) 16-QAM e (c)-(d) 64-QAM, respectivamente (para $N_{\rm t} = N_{\rm r} = 4, 16, 32$ MIMO).

vetores símbolos, $P(\hat{\mathbf{a}} \neq \mathbf{a})$, que é obtida através de múltiplos experimentos de *Monte Carlo*, considerando uma faixa de valores de *Signal-to-Noise Ratio* (SNR) definidos por

$$\Gamma_k = \frac{\mathrm{E}\left[\|\mathbf{H}_k \mathbf{a}_k\|_2^2\right]}{N_{\mathrm{r}} \sqrt{M} \sigma^2}, \ \forall \ k.$$
(31)

Note que $\tilde{H}_{i,j} \sim C\mathcal{N}(0, 1/N_r)$, $\forall i, j$, de modo que assume-se um canal Rayleigh plano para cada subportadora OFDM e que o ganho do canal é constante durante a transmissão do *frame* OFDM. Note também que o cômputo de **T** é feito com $\delta = 3/4$ configurado no algoritmo LLL [11, 15] e que K = 8 [13] elementos são definidos no Algoritmo 2.

Sendo assim, observa-se na Figura 3 que a inclinação da curva de desempenho para o detector LRA-PDA é similar àquela do detector MV, seja qual for o cenário MIMO analisado. Isso é uma consequência da redução da IAI e, portanto, do aumento da diversidade espacial obtida pelo detector LRA-PDA, como discutido anteriormente. Adicionalmente, note que o desempenho do detector LRA-PDA para o MIMO de baixa ordem 4×4 , na Figura 3 (a) e (c), é o mais atrativo, visto que o PDA é superado em praticamente toda a faixa de SNR analisada. Além disso, pode ser verificado também que o LRA-PDA é melhor que o PDA em



quase todos os cenários MIMO restantes, quando se considera as taxas de erro $P(\hat{\mathbf{a}} \neq \mathbf{a}) < 10^{-2}$ e, principalmente, a constelação densa 64-QAM.

3.5 Conclusão

Os esquemas MIMO são essenciais para que os requisitos de eficiência espectral, vazão e robustez sejam atingidos nas futuras redes de comunicação móvel. A solução clássica do detector SD possui uma complexidade elevada que não é suportada por dispositivos com capacidade computacional limitada. O detector PDA visa reduzir a complexidade do receptor, oferecendo um desempenho melhor que os esquemas mais simples, mas sem atingir a ordem de diversidade do SD. Esta seção apresentou uma proposta de integração da técnica LR ao detector PDA que resulta em uma melhoria do ganho de diversidade sem aumento significativo da complexidade. Este detector, denominado de LRA-PDA demonstrou ser uma solução interessante pois apresenta uma relação de compromisso entre complexidade e desempenho em termos de taxa de erro de símbolo que é atrativa para diversos cenários esperados para as Redes 6G.



4 Utilização do *Sphere Detector* para Detecção de Sinais FTN-GFDM em Canais Duplamente Dispersivos

4.1 Introdução

A eficiência espectral das redes 6G precisará aumentar significativamente para permitir o alto volume de tráfego de dados previsto para aplicações futuras. Abordagens que podem empregadas para aumentar a eficiência espectral e alcançar altas taxas em áreas densamente povoadas incluem (i) o uso da faixa de ondas milimétricas do espectro para permitir a alocação de larguras de banda muito maiores, (ii) a implantação de células com raio de cobertura pequeno, e (iii) o uso de MIMO massivo. No entanto, essas abordagens não atendem todos os cenários previstos para redes móveis futuras. Um exemplo importante é o cenário Enhanced Remote Area Communications (eRAC), que tem como objetivo fornecer serviços de banda larga e de IoT em áreas rurais [17]. Nestas áreas, células com grande raio de cobertura devem ser implantadas usando as bandas Very High Frequency (VHF) ou Ultra High Frequency (UHF) disponíveis, por meio de alocação de frequência oportunista. Neste caso, uma solução razoável para aumentar a taxa de dados do sistema é aumentar a eficiência espectral da forma de onda.

O objetivo desta seção consiste em apresentar uma forma de onda que comprime as subportadoras e/ou os subsímbolos do *Generalized Frequency Division Multiplexing* (GFDM), aumentando a eficiência espectral, sem que haja penalidades em termos de *Bit Error Rate* (BER). Esta seção também apresenta uma modificação do SD que viabiliza a recuperação dos dados transmitidos de forma quase ótima e com complexidade inferior àquela apresentada pelo detector MLSE.

4.2 Descrição do FTN-GFDM

Pequenas modificações na definição da grade tempo-frequência permitem que o GFDM inclua a sinalização *Faster-than-Nyquist* (FTN) [18, 19], como discutido brevemente a seguir. Considere que o filtro protótipo possui \mathcal{N} amostras divididas em \mathcal{P} períodos com \mathcal{S} amostras cada, e que os sub-símbolos sejam definidos em intervalos de tempo de \mathcal{K} amostras e subportadoras em intervalos de frequência de \mathcal{M} amostras. O sinal GFDM de transmissão pode então ser escrito como

$$x[n] = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} s_{k,m} g[\langle n - m\mathcal{K} \rangle_{\mathcal{N}}] \exp\left(j2\pi \frac{k\mathcal{M}}{\mathcal{N}}n\right),\tag{32}$$

em que $n = 0, \ldots, \mathcal{N} - 1$, $s_{k,m}$ é o símbolo QAM para a k-ésima subportadora e m-ésimo sub-símbolo, e g[n] é a resposta ao impulso do filtro protótipo. Para incluir a sinalização FTN no GFDM é necessário que $\mathcal{K} < \mathcal{S}$ e/ou $\mathcal{M} < \mathcal{P}$. Os fatores de compressão no tempo e na frequência são $v_t = \mathcal{K}/\mathcal{S}$ e $v_f = \mathcal{M}/\mathcal{P}$, respectivamente. Os valores de K e M são dados por

$$K = \frac{\mathcal{PS}}{\mathcal{M}} = \frac{\mathcal{S}}{v_f} = \left\lfloor \frac{\mathcal{N}}{\mathcal{M}} \right\rfloor,\tag{33}$$

$$M = \frac{\mathcal{PS}}{\mathcal{K}} = \frac{\mathcal{P}}{v_t} = \left\lfloor \frac{\mathcal{N}}{\mathcal{K}} \right\rfloor.$$
(34)

O conteúdo deste capítulo foi publicado em M. B. Mello and L. L. Mendes, "Low-Complexity Detection Algorithms Applied to FTN-GFDM Systems," in IEEE Access, vol. 10, pp. 101683-101696, 2022, doi: 10.1109/AC-CESS.2022.3208878.



Quando $\frac{N}{M} \in \frac{N}{K}$ não são números inteiros, os fatores de compressão devem ser ajustados para $\bar{v}_t = \mathcal{P}/M \in \bar{v}_f = \mathcal{S}/K$, respectivamente, e o sinal GFDM deve ser ajustado adequadamente para

$$\bar{x}[n] = \sqrt{\bar{v}_t \bar{v}_f} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} s_{k,m} g[\langle n - m v_t \mathcal{S} \rangle_{\mathcal{N}}] \exp\left(j2\pi \frac{k v_f}{\mathcal{S}} n\right),$$
(35)

em que $\sqrt{\overline{v}_t \overline{v}_f}$ é usado para manter a potência de transmissão inalterada.

O vetor $\mathbf{\bar{x}} = [\bar{x}[0], \bar{x}[1], \dots, \bar{x}[\mathcal{N}-1]]^{\mathrm{T}}$ pode ser escrito como

$$\bar{\mathbf{x}} = \mathbf{As},\tag{36}$$

em que $\mathbf{s} = [s_{0,0}, s_{1,0}, \cdots, s_{K-1,0}, s_{0,1}, \cdots, s_{K-1,M-1}]^{\mathrm{T}}$ é o vetor que contém os símbolos QAM e \mathbf{A} é a matriz de transmissão contendo as N = KM versões do filtro protótipo \mathbf{g} deslocado circularmente no tempo e na frequência, ou seja, $\mathbf{A} = [\mathbf{g}_{0,0} \ \mathbf{g}_{1,0} \cdots \mathbf{g}_{K-1,0} \ \mathbf{g}_{0,1} \cdots \mathbf{g}_{K-1,M-1}]$. O vetor coluna

$$\mathbf{g}_{k,m} = \sqrt{\bar{v}_t \bar{v}_f} g[\langle n - m v_t \mathcal{S} \rangle_{\mathcal{N}}] \exp\left(j 2\pi k v_f n / \mathcal{S}\right), \qquad (37)$$

contém as amostras do filtro protótipo deslocadas para a k-ésima subportadora e o m-ésimo sub-símbolo, com $n = 0, \ldots, N - 1$.

4.3 SD para Detecção de Sinais FTN-GFDM

O vetor do sinal recebido é dado por

$$\mathbf{y} = \mathbf{H}\bar{\mathbf{x}} + \mathbf{w},\tag{38}$$

em que \mathbf{H} contém versões deslocadas circularmente da resposta ao impulso do canal \mathbf{h} , e \mathbf{w} é o vetor de ruído AWGN.

No lado do receptor, um *Matched Filter* (MF) maximiza a *Signal-to-Interference-and-Noise Ratio* (SINR), permitindo a detecção adequada. Assumindo o conhecimento da resposta ao impulso do canal, o sinal discreto na saída MF é dado por

$$\mathbf{r} = \mathbf{A}^{\mathrm{H}}\mathbf{H}^{-1}\mathbf{H}\mathbf{A}\mathbf{s} + \mathbf{A}^{\mathrm{H}}\mathbf{H}^{-1}\mathbf{w} = \mathbf{G}\mathbf{s} + \bar{\mathbf{w}},\tag{39}$$

em que $\mathbf{G} = \mathbf{A}^{\mathrm{H}}\mathbf{A}$ é a matriz de coeficientes de correlação, $\mathbf{\bar{w}} = \mathbf{A}^{\mathrm{H}}\mathbf{H}^{-1}\mathbf{w}$, e (.)^H representa o operador Hermitiano. Como a operação linear $\mathbf{A}^{\mathrm{H}}\mathbf{H}^{-1}$ correlaciona as amostras de \mathbf{w} , as amostras filtradas $\mathbf{\bar{w}}$ são coloridas.

Para reduzir a perda de desempenho em termos de BER, o receptor deve mitigar a interferência inerente introduzida pela sinalização FTN. O detector MLSE com um filtro branqueador pode mitigar os efeitos da interferência sem causar penalidades na BER, levando a

$$\arg\min_{\hat{\mathbf{s}}\in\xi^{N}}\left\| (\mathbf{A}^{H})^{-1}(\mathbf{r}-\mathbf{G}\hat{\mathbf{s}}) \right\|^{2},\tag{40}$$

em que $\|.\|$ é a norma euclidiana e ξ^N é o conjunto de todas as sequências de símbolos QAM possíveis. No entanto, esse tipo de detector exige uma busca exaustiva entre todas as sequências possíveis de símbolos transmitidos, levando uma complexidade proibitiva para aplicações práticas.

O Sphere Detector (SD) simplifica o esforço computacional e se aproxima do desempenho ótimo. Portanto, a proposta apresentada pelos pesquisadores do Projeto Brasil 6G considera o SD [20] um detector eficiente para resolver o problema da alta complexidade de (40).



O princípio de operação do algoritmo SD é limitar o espaço de busca em uma esfera multidimensional, procurando apenas por soluções candidatas que estejam dentro da esfera. Como o candidato mais próximo dentro da esfera é também o candidato mais próximo em todo o espaço de busca, então o SD alcança a solução ótima com menor custo computacional que aquele do detector MLSE [21].

Para que uma sequência candidata $\hat{\mathbf{s}}$ esteja dentro da esfera de raio ρ , a seguinte condição deve ser satisfeita:

$$\arg\min_{\hat{\mathbf{s}}\in\xi^{N}}\left\| (\mathbf{A}^{\mathrm{H}})^{-1}(\mathbf{r}-\mathbf{G}\hat{\mathbf{s}}) \right\|^{2} \le \rho^{2}.$$
(41)

Usando $\zeta = (\mathbf{A}^{H}\mathbf{A})^{-1}\mathbf{A}^{H}\mathbf{r}$, o raio SD pode ser expresso por

$$(\zeta - \hat{\mathbf{s}})^{\mathrm{H}} \mathbf{G}^{\mathrm{H}} \mathbf{A}^{-1} (\mathbf{A}^{-1})^{\mathrm{H}} \mathbf{G} (\zeta - \hat{\mathbf{s}}) = (\zeta - \hat{\mathbf{s}})^{\mathrm{H}} \mathbf{G}^{\mathrm{H}} (\zeta - \hat{\mathbf{s}}) = \rho.$$
(42)

Para simplificar o cálculo da norma euclidiana em (41), a matriz **G** deve ser decomposta por Cholesky [22] de acordo com

$$\operatorname{chol}\{\mathbf{G}^{\mathrm{H}}\} = \mathbf{L}^{\mathrm{H}}\mathbf{L},\tag{43}$$

em que \mathbf{L} é uma matriz triangular superior. Então, a condição em (41) pode ser expressa como

$$\arg\min_{\hat{\mathbf{s}}\in\xi^{N}}\left\|\mathbf{L}(\zeta-\hat{\mathbf{s}})\right\|^{2}\leq\rho^{2}.$$
(44)

A estrutura triangular superior da matriz \mathbf{L} permite que a desigualdade (44) seja expressa como

$$\rho^{2} \ge \sum_{i=1}^{N} \left| \sum_{j=i}^{N} L_{i,j}(\zeta_{j} - s_{j}) \right|^{2}.$$
(45)

Observe que o N-ésimo elemento de (45) depende apenas de \hat{s}_N , o (N-1)-ésimo elemento depende apenas de $\{\hat{s}_N, \hat{s}_{N-1}\}$, e assim por diante [23] [24]. Desta forma, a busca por sequências dentro da esfera torna-se uma busca em árvore. As etapas para implementar o detector SD proposto nesta seção estão resumidas no Algoritmo 3.

4.4 Análise de Complexidade

A complexidade do algoritmo SD é ditada pelo cálculo da distância euclidiana parcial ρ_n^2 , definida na linha 4 no Algoritmo 3. O número de operações necessárias para calcular ρ_n^2 aumenta à medida que *n* diminui devido à estrutura de **L**. Assim, o número de *Floating-Point* Operations (FLOPs) necessários para calcular ρ_n^2 em um nó no nível *n* é

$$\mathcal{O}_n = 10(N-n) + 12, \tag{46}$$

em que \mathcal{O}_n representa a complexidade de inspeção de um único nó de nível n. Diante disso, a complexidade do SD pode ser definida como a soma de nós visitados ponderada pelo número de FLOPs necessários para calcular a distância euclidiana parcial em cada nível. O número de nós visitados depende do raio de busca inicial. Assumindo o pior cenário, onde toda a árvore de busca é inspecionada, a complexidade do SD em termos de FLOPs é igual a

$$\sum_{n=1}^{N} J^{(N+1-n)} \mathcal{O}_n,\tag{47}$$



Algorithm 3 Detector SD proposto para o sinal FTN-GFDM.

1: Entrada: Matriz triangular superior L, estimativa de máxima verossimilhança ζ , raio de busca inicial ρ_{initial} , ordem de modulação J e constelação \mathcal{D} 2: $\rho^2 = \rho_{\text{initial}}^2$; $\rho_{N+1}^2 = 0$; n = N; $\varsigma_{N+1} = 0$; $\varsigma_N = 1$; 3: while $\varsigma_{N+1} < 1$ do $\rho_n^2 = \rho_{n+1}^2 + \left| \sum_{j=n}^N L_{n,j} \left(\zeta_j - \mathcal{D}(\varsigma_j) \right) \right|^2;$ if n > 1 then 4: 5: $\begin{array}{l} \mbox{if } \rho^2 \geq \rho_n^2 \mbox{ then } \\ n=n-1; \ \varsigma_n=\varsigma_n+1; \end{array}$ 6: 7: else 8: 9: if $\varsigma_n \geq J$ then $\varsigma_{1:n} = 0; n = n + 1; \varsigma_n = \varsigma_n + 1;$ 10:while $\varsigma_n > J$ do 11: $\varsigma_{1:n} = 0; n = n + 1; \varsigma_n = \varsigma_n + 1;$ 12:else 13:14: $\varsigma_n = \varsigma_n + 1;$ 15:else if $\rho^2 \ge \rho_n^2$ then 16: $\hat{\mathbf{d}} = \mathcal{D}(\varsigma_{n:N}); \ \rho^2 = \rho_n^2;$ 17:if $\varsigma_n \geq J$ then 18: $\varsigma_{1:n} = 0; n = n + 1; \varsigma_n = \varsigma_n + 1;$ 19:while $\varsigma_n > J$ do 20: $\varsigma_{1:n} = 0; n = n + 1; \varsigma_n = \varsigma_n + 1;$ 21: 22: else 23: $\varsigma_n = \varsigma_n + 1;$ 24: Saída: Vetor de dados estimados $\hat{\mathbf{s}}$

em que J é a ordem de modulação e $J^{(N+1-i)}$ é o número de nós no *i*-ésimo nível da árvore de busca. Aplicando algumas identidades de soma, (47) pode ser reescrito como

$$\frac{2J\left((5N+1)J^{N+1} - (5N+6)J^N - J + 6\right)}{(J-1)^2}.$$
(48)

Da mesma forma, assumindo o melhor cenário, onde apenas um único caminho da raiz até o primeiro nó folha é inspecionado e os demais nós são descartados, o limite inferior da complexidade SD em termos de FLOPs é dado por

$$\sum_{n=1}^{N} J\mathcal{O}_n. \tag{49}$$

Neste caso, apenas J nós são visitados em cada nível da árvore de busca, onde um nó visitado corresponde ao caminho com a distância euclidiana mínima (processo de ida até a folha), e os demais nós representam o processo de poda (processo de volta à raiz). Aplicando as identidades de soma em (49), o limite inferior de complexidade pode ser definido como

$$J(5N^2 + 7N). (50)$$

Portanto, a complexidade do SD depende do raio de busca inicial e, consequentemente, da estimativa inicial. Como a estimativa inicial é baseada na detecção de ZF, a complexidade e



Complexidade Big - \mathcal{O} dos detectores propostos na literatura							
MLSE	SD (limite superior)	SSSgbKSE	SD (lower bound)	SSSSE	FDE		
$\mathcal{O}(N^2 J^N)$	$\mathcal{O}(NJ^N)$	$\mathcal{O}(KN^2 + K^3)$	$\mathcal{O}(JN^2)$	$\mathcal{O}(N^2)$	$\mathcal{O}(N \log N)$		

Tabela 3:	Complexidade	dos	detectores	MLSE,	SD,	SSSgbKSE.	$SSSE \epsilon$	FDE.
	1			/		0		

o desempenho dependerão da SNR e da matriz **G**. Do ponto de vista da árvore de busca, a complexidade também depende da ordem de modulação J e de N. No pior cenário, o detetor SD atinge complexidade exponencial.

Para fins de comparação, a Tabela 3 apresenta a complexidade na notação big-O de alguns detectores baseados em estimativa de sequências propostos na literatura para a sinalização FTN. Entre eles estão o Successive Symbol-by-Symbol Sequence Estimator (SSSSE) [25], Successive Symbol-by-Symbol with go-back K Sequence Estimator (SSSgbKSE) [25] e Frequency-Domain Equalization (FDE) [26]. Observe-se que as complexidades são classificadas em ordem decrescente, ou seja, da maior para a menor complexidade. Considerando a notação assintótica, o MLSE apresenta a maior complexidade, seguido do SD no pior caso (limite superior da complexidade). Note-se que, nesses dois casos a complexidade é exponencial. No entanto, ao considerarmos o melhor cenário, o limite inferior da complexidade do SD é quadrática, equivalente à complexidade do SSSSE.

4.5 Análise de Desempenho

O desempenho do detector SD em termos de BER foi avaliado para os canais AWGN, TIFS e TVFS. O canal TVFS possui quatro coeficientes com distribuição Rayleigh, que mudam a cada bloco FTN-GFDM e apresentam resposta de frequência plana por subportadora. A Tabela 4 mostra os valores dos parâmetros do canal empregados nas simulações. CSI perfeito é assumido no receptor. O raio inicial do detector SD foi definido para ser a distância da estimativa obtida por um detector ZF até os símbolos transmitidos com maior probabilidade.

Tabela 4: Valores dos parâmetros do canal usados nas simulações.

Canal	Resposta ao impulso
AWGN	$\mathbf{h}_{AWGN} = 1$
TIFS	$\mathbf{n}_{\text{TIFS}} = \begin{bmatrix} 1 & 0.4 & 0.2 & 0.08 \end{bmatrix}^{1}$
TVFS	$\mathbf{h}_{\text{TVFS}} = \begin{bmatrix} e^0 h_0 & e^{-4} h_1 & e^{-8} h_2 & e^{-12} h_3 \end{bmatrix}^{\text{T}}, h_n \sim \mathbb{CN}(0, 1)$

As Figs. 4, 5 e 6 apresentam a BER dos detectores para o sistema FTN-GFDM com compressão de tempo, frequência e tempo-frequência, respectivamente, nos três canais propostos: (a) AWGN, (b) TIFS e (c) TVFS. Na Fig. 4, Inter Symbol Interference (ISI) é inserida no sinal devido a $v_t = 0.8$ e, consequentemente, a taxa de dados aumenta 25%. Para os parâmetros considerados na Fig. 5, o fator de compressão real é $\bar{v}_f = \frac{\bar{S}}{K} = \frac{5}{6} = 0,8333$ e, portanto, o aumento na taxa de dados é de 20%. Na Fig. 6, os fatores de compressão v_t e v_f possuem ganhos independentes que podem ser combinados para alcançar maior eficiência espectral. Neste caso, é possível aumentar a taxa de dados em 50% sem penalidades de desempenho em termos de BER usando o detector SD. Apesar de os demais detectores analisados possuírem menor complexidade computacional que aquele do SD, os desempenhos em termos de BER são subótimos. O detector FDE tem um desempenho ruim em todos os cenários avaliados, e não pode





Figura 4: BER para o sistema FTN-GFDM com compressão no tempo, considerando modulação BPSK, S = 5, $\mathcal{P} = 4$, $v_t = 0, 8$, $v_f = 1$, K = 5, M = 5 e filtro protótipo Dirichlet. (a) Canal AWGN. (b) Canal TIFS. (c) Canal TVFS.



Figura 5: BER para o sistema FTN-GFDM com compressão na frequência, considerando modulação BPSK, S = 5, $\mathcal{P} = 4$, $v_t = 1$, $v_f = 0.8$, K = 6, M = 4 e filtro protótipo rect. (a) Canal AWGN. (b) Canal TIFS. (c) Canal TVFS.

ser considerado uma técnica de detecção viável para FTN-GFDM. O receptor SSSSE possui a menor complexidade computacional entre os não-lineares, mas seu desempenho é afetado pelo condicionamento de \mathbf{G} . O SSSgbKSE pode ter um desempenho muito próximo daquele do MLSE, desde que o condicionamento de \mathbf{G} seja aceitável e que K seja suficientemente grande. Assim, este detector pode ser uma solução interessante em cenários onde os parâmetros do sistema levam a condições de operação apropriadas e a complexidade do SD não pode ser suportada.

4.6 Conclusão

Os resultados apresentado nesta seção mostram que o SD pode garantir um desempenho em termos de BER quase ótimo ótimo para sistemas FTN-GFDM, mesmo em cenários de altos





Figura 6: BER para o sistema FTN-GFDM com compressão no tempo e frequência, considerando modulação BPSK, S = 5, $\mathcal{P} = 4$, $v_t = 0.8$, $v_f = 0.75$, K = 6, M = 5 e filtro protótipo RRC com fator de *roll-off* $\alpha = 0.5$. (a) Canal AWGN. (b) Canal TIFS. (c) Canal TVFS.

ISI e *Inter Carrier Interference* (ICI), com uma complexidade consideravelmente menor do que aquela apresentada pelo MLSE. Assi, pode-se concluir que o FTN-GFDM pode aumentar a eficiência espectral do sistema sem penalidades em termos de BER e complexidade aceitável para as modernas técnicas de implementação, mostrando que é uma solução interessante em futuros cenários de comunicação móvel onde outras técnicas para aumentar eficiência espectral não são suportadas.



5 Geometria Estocástica e Redes Cell-Free

5.1 Introdução

Em redes de comunicação sem fio *cell-free*, a divisão espacial da rede em células não é realizada. Ou seja, a configuração dos enlaces de comunicação é baseada na posição conjunta dos usuários, ao invés de depender apenas da posição de um determinado usuário pertencer ou não à uma célula. Essa abordagem *cell-free* é uma arquitetura promissora para ser utilizada nas redes sem fio 6G. Destaca-se a tecnologia *Cell-Free Massive MIMO* (CF-mMIMO), na qual cada UE pode ser servido por mais de um *Access Point* (AP), que por sua vez serve mais de um UE até um número máximo determinado pelo número de sequências piloto. As sequências piloto têm o propósito de possibilitar a estimação do canal, o que permite otimizar a transmissão e a recepção dos sinais transmitidos através do meio [27].

A distribuição dos UEs e dos APs por uma determinada região depende de diversos fatores e pode variar significativamente de local para local. Dessa forma, para analisar as propriedades estatísticas relacionadas ao posicionamento dessas entidades de forma geral, é natural supor o processo estocástico que exige a menor quantidade de informação do sistema, isto é, o Processo Pontual de Poisson (PPP) [28].

A teoria que estuda processos pontuais é a geometria estocástica. Alguns resultados importantes juntamente com contribuições para essa teoria serão apresentadas, focando em aplicações para as redes de comunicação *cell-free*.

5.2 Resultados da Literatura de Geometria Estocástica

Nesta subseção, os pesquisadores do Projeto Brasil 6G apresentam alguns resultados da geometria estocástica que serão úteis para demonstrar os resultados obtidos.

Considere um PPP homogêneo Φ de densidade constante $\lambda > 0$, no espaço euclideano \mathbb{R}^2 . A probabilidade de que um conjunto mensurável $A \subset \mathbb{R}^2$ contenha exatamente $n \in \mathbb{N}$ pontos é dada por $(\lambda \mid A \mid)^n$

$$\mathbb{P}(\Phi(A) = n) = \frac{(\lambda|A|)^n}{n!} e^{-\lambda|A|},$$
(51)

onde |A| é a medida de Lebesgue do conjunto A, que para o caso do \mathbb{R}^2 é a área de A. Ademais, denotamos por $B_{x_0}(r)$ um disco aberto de raio r > 0 centrado em $x_0 \in \mathbb{R}^2$, ou seja, $B_{x_0}(r) \triangleq \{x \in \mathbb{R}^2 : ||x - x_0|| < r\}$. Logo, $|B_x(r)| = \pi r^2$. Seja a variável aleatória R_n^{Φ} a distância do *n*-ésimo ponto de Φ mais próximo à origem ou

Seja a variável aleatória R_n^{Φ} a distância do *n*-ésimo ponto de Φ mais próximo à origem ou à um ponto típico de Φ (pelo Teorema de Slivnyak–Mecke). Então, a Função Densidade de Probabilidade (FDP) de R_n^{Φ} é dada por [29]

$$f_{R_n^{\Phi}}(r) = \lambda 2\pi r \frac{(\lambda \pi r^2)^{n-1}}{(n-1)!} e^{-\lambda \pi r^2}, \quad r \ge 0.$$
(52)

Enquanto que a FDP conjunta de $(R_1^{\Phi}, R_2^{\Phi}, \dots, R_n^{\Phi})$ é dada por

$$f_{R_1^{\Phi},\dots,R_n^{\Phi}}(r_1,r_2,\dots,r_n) = r_1 r_2 \cdots r_n (\lambda 2\pi)^n e^{-\lambda \pi r_n^2}, \quad 0 \le r_1 \le r_2 \le \dots \le r_n.$$
(53)

 $f_{R_n^{\Phi}} \in f_{R_1^{\Phi},...,R_n^{\Phi}}$ serão denotados como $f_{R_n}^{\Phi} \in f_{R_1,...,R_n}^{\Phi}$, respectivamente, para simplificar a notação.

É possível ultrapassar esse número, porém o fenômeno de contaminação de pilotos se intensifica, prejudicando significativamente a estimação do canal [27].

Capítulo escrito por Plínio Santini Dester e Paulo Cardieri.



5.3 Modelo do Sistema

A princípio, foi considerado dois PPPs homogêneos $\Phi_1 e \Phi_2$, com densidades $\lambda_1 > 0 e \lambda_2 > 0$, e sem nenhuma especificação sobre o que eles representam fisicamente. Com isso é possível alcançar uma apresentação mais limpa e geral dos resultados, uma vez que eles não estão limitados às redes de comunicação, e podem ter aplicações em outros sistemas que sejam modelados através de PPPs.

Uma rede CF-mMIMO será considerada para as aplicações. Dessa forma, um dos PPPs será tomado como a posição espacial dos UEs no \mathbb{R}^2 e o outro PPP como a posição dos APs. Esses PPPs são denotados como $\Phi_{\rm UE}$ e $\Phi_{\rm AP}$ e têm densidades espaciais $\lambda_{\rm UE} > 0$ e $\lambda_{\rm AP} > 0$, respectivamente. Ademais, o número de sequências piloto ortogonais é denotado por $\tau_p \in \mathbb{N}$. Neste trabalho, será assumido que cada AP serve no máximo os τ_p UEs mais próximos.

5.4 Resultados Espaciais em Redes Cell-Free

Primeiramente, os pesquisadores propõe uma definição para Φ -vizinhos e outra definição que relaciona pontos vizinhos de dois PPPs. Em seguida, um lema é provado, auxiliando na caracterização da relação de vizinhança entre pontos de PPPs e sendo útil em outras demonstrações.

Definição 1. Seja Φ um processo pontual. O n-ésimo Φ -vizinho de um ponto $x \in \mathbb{R}^2$ é definido como o ponto de Φ que seja o n-ésimo mais próximo de x.

Definição 2. Sejam os PPPs $\Phi_1 e \Phi_2$ independentes e com densidades $\lambda_1 > 0 e \lambda_2 > 0$, respectivamente, no \mathbb{R}^2 . Um ponto típico $x_1 de \Phi_1$ é tomado. Seja, x_2 o k-ésimo Φ_2 -vizinho de x_1 . Se x_2 é o n-ésimo Φ_1 -vizinho de x_1 , então a variável aleatória $N_k^{\Phi_1,\Phi_2}$ é definida como n.

É importante atentar que os índices 1 e 2 estão sendo usados ao invés de UE e AP para que se alcance a liberdade de trocá-los entre si e, também porque esses resultados não estão limitados a CF-mMIMO, facilitando a aplicação em outras tecnologias ou áreas.

Lema 1. A probabilidade de que um ponto típico de Φ_1 seja o n-ésimo Φ_1 -vizinho do ponto que é k-ésimo Φ_2 -vizinho do ponto típico, com $n, k \in \mathbb{N}^* \triangleq \mathbb{N} \setminus \{0\}$, é dada por

$$\mathbb{P}(N_k^{\Phi_1,\Phi_2}=n) = \binom{k+n-2}{k-1} \left(1+\frac{\lambda_1}{\lambda_2}\right)^{-k} \left(1+\frac{\lambda_2}{\lambda_1}\right)^{-(n-1)}.$$
(54)

A esperança matemática de $N_k^{\Phi_1,\Phi_2}$ é

$$\mathbb{E}[N_k^{\Phi_1,\Phi_2}] = 1 + k \frac{\lambda_1}{\lambda_2}.$$
(55)

Ademais, a máxima probabilidade ocorre para o n^* -ésimo Φ_1 -vizinho, em que

$$n^* = \arg\max_{n \in \mathbb{N}^*} \mathbb{P}(N_k^{\Phi_1, \Phi_2} = n) = \left\lfloor (k-1)\frac{\lambda_1}{\lambda_2} \right\rfloor,\tag{56}$$

onde $|\cdot|$ é a função floor que retorna o maior inteiro menor ou igual a \cdot .



Demonstração. Sem perda de generalidade, pode-se considerar que o ponto típico se encontra na origem pelo Teorema de Slivnyak–Mecke. A probabilidade descrita é equivalente a existir exatamente n - 1 pontos de Φ_1 no disco com raio igual à distância ao k-ésimo Φ_2 -vizinho e centrado no mesmo. Descondicionando em relação à essa distância tem-se que

$$\mathbb{P}(N_k^{\Phi_1,\Phi_2} = n) = \int_0^\infty \mathbb{P}(\Phi_1(B_{x_k}(r)) = n - 1) f_{R_k^{\Phi_2}}(r) \,\mathrm{d}r$$
$$= \int_0^\infty \frac{(\lambda_1 \pi r^2)^{n-1}}{(n-1)!} \,\mathrm{e}^{-\lambda_1 \pi r^2} \lambda_2 2\pi r \frac{(\lambda_2 \pi r^2)^{k-1}}{(k-1)!} \,\mathrm{e}^{-\lambda_2 \pi r^2} \,\mathrm{d}r.$$
(57)

Resolvendo a integral (57), a solução desejada é obtida.

Para mostrar que n^* é o ponto de máximo, pode-se verificar que $\mathbb{P}(N_k^{\Phi_1,\Phi_2} = n) \geq \mathbb{P}(N_k^{\Phi_1,\Phi_2} = n+1)$ se, e somente se, $n \geq (k-1)\lambda_1/\lambda_2$.

Essa é uma ótima oportunidade para ilustrar que o k-ésimo ponto mais próximo em um PPP não é um ponto típico, uma vez que $\mathbb{P}(N_k^{\Phi_1,\Phi_2}=n) \neq \mathbb{P}(N_n^{\Phi_2,\Phi_1}=k)$. Ou seja, resultados diferentes são obtidos se as relações entre $\Phi_1 \in \Phi_2$ forem invertidas.

Dessa forma, o número esperado e o número que maximiza a probabilidade de $N_k^{\Phi_1,\Phi_2}$ crescem com a densidade espacial de Φ_1 e decrescem com Φ_2 .

Nesse momento, uma pergunta interessante a se responder seria qual a distribuição de probabilidade da distância do k-ésimo Φ_2 -vizinho de um ponto típico de Φ_1 , dado que esse é n-ésimo Φ_1 vizinho daquele. Definindo essa distância como a variável aleatória tem-se

$$R_{n,k}^{\Phi_1,\Phi_2} \triangleq \left(R_{N_k^{\Phi_1,\Phi_2}}^{\Phi_2} \mid N_k^{\Phi_1,\Phi_2} = n \right).$$
(58)

Essa pergunta é respondida através do teorema seguinte. Mas, antes de enunciá-lo é importante recordar que duas variáveis aleatórias $X_1 \in X_2$ são iguais em distribuição quando $\mathbb{P}(X_1 \leq x) = \mathbb{P}(X_2 \leq x)$ para todo x. Nesse caso, tem-se que $X_1 \stackrel{d}{=} X_2$. Ademais, o PPP resultante da sobreposição de $\Phi_1 \in \Phi_2$ é denotado por $\Phi_1 + \Phi_2$. Essa notação de soma (+) faz sentido, pois processos pontuais são definidos como *medidas* (aleatórias) [30]. Segue o teorema.

Teorema 1. $R_{n,k}^{\Phi_1,\Phi_2} \stackrel{d}{=} R_{n+k-1}^{\Phi_1+\Phi_2}$

Demonstração. Seguindo a linha de raciocínio da demonstração do Lema 1 e usando o resultado do mesmo, tem-se que a FDP de $R_{n,k}^{\Phi_1,\Phi_2}$ é dada por

$$f_{R_{n,k}^{\Phi_{1},\Phi_{2}}}(r) = \frac{\mathbb{P}(\Phi_{1}(B_{x_{k}}(r)) = n - 1) f_{R_{k}^{\Phi_{2}}}(r)}{\mathbb{P}(N_{k}^{\Phi_{1},\Phi_{2}} = n)} \qquad (r > 0)$$

$$= \frac{\frac{(\lambda_{1}\pi r^{2})^{n-1}}{(n-1)!} e^{-\lambda_{1}\pi r^{2}} \lambda_{2} 2\pi r \frac{(\lambda_{2}\pi r^{2})^{k-1}}{(k-1)!} e^{-\lambda_{2}\pi r^{2}}}{\binom{k+n-2}{k-1} \left(1 + \frac{\lambda_{1}}{\lambda_{2}}\right)^{-k} \left(1 + \frac{\lambda_{2}}{\lambda_{1}}\right)^{-(n-1)}}$$

$$= (\lambda_{1} + \lambda_{2}) 2\pi r \frac{\left((\lambda_{1} + \lambda_{2})\pi r^{2}\right)^{n+k-2}}{(n+k-2)!} e^{-(\lambda_{1} + \lambda_{2})\pi r^{2}} = f_{R_{n+k-1}}^{\Phi_{1}+\Phi_{2}}(r), \qquad (59)$$

de onde a última igualdade vem de (52) e completa a demonstração.


É importante notar que apesar de $R_{n,k}^{\Phi_1,\Phi_2}$ ser igual à $R_{n+k-1}^{\Phi_1+\Phi_2}$ em distribuição, essas variáveis aleatórias não são iguais *almost-surely*.

O Teorema 1 diz que a distribuição da distância entre dois pontos tais que um é *n*-ésimo Φ_1 vizinho e o outro é *k*-ésimo Φ_2 -vizinho um do outro é a mesma que a distribuição da distância do (k+n-1)-ésimo $(\Phi_1+\Phi_2)$ -vizinho em relação à origem. Isso é inesperado e contra-intuitivo, pois, por exemplo, é possível fazer a densidade λ_2 de Φ_2 tender à 0 pela direita e aumentar a densidade λ_1 de Φ_1 tal que $\lambda_1 + \lambda_2$ se mantenha constante. De acordo com o teorema, a distribuição da distância vai ser a mesma, sendo que intuitivamente a distância deveria aumentar. Isso ocorre pois a quantidade de pontos em Φ_2 é muito mais escassa quando $\lambda_2 \rightarrow 0^+$, o que levaria à distâncias maiores entre os pontos e, consequentemente, a uma distribuição de probabilidade diferente.

Outra consequência direta do Teorema 1 é a distribuição da distância entre o 2-ésimo Φ_1 vizinho e o 2-ésimo Φ_2 -vizinho um do outro ser a mesma que a distribuição da distância do 1-ésimo Φ_1 -vizinho ao 3-ésimo Φ_2 -vizinho um do outro. Isso também não é esperado ou intuitivo quando as densidades espaciais de Φ_1 e Φ_2 são diferentes.

Considerando cenários mais especificamente voltados ao CF-mMIMO, a probabilidade de ocorrer a situação indesejada em que um AP não possa servir todos os UEs mais próximos será analisada. Para isso, será deduzido um resultado interessante, que é a probabilidade do AP típico ter mais de n UEs mais próximos a ele do que qualquer outro AP.

Definição 3. Considere um AP típico e os n UEs mais próximos a ele. Definimos $\tilde{P}_{AP}(n)$ como a probabilidade desses UEs serem mais próximos ao AP típico do que em relação aos demais APs.

De forma exata, $\tilde{P}_{AP}(n)$ é equivalente à probabilidade da união dos discos ao redor dos nUEs mais próximos não conter nenhum AP em seu interior, ou seja,

$$\tilde{P}_{\rm AP}(n) = \int_{\mathcal{S}_n} \int_{[0,2\pi]^n} f_{R_1,\dots,R_n}^{\Phi_{\rm UE}}(r_1, r_2,\dots,r_n) \mathbb{P}\left(\Phi_{\rm AP}\left(\bigcup_{k=1}^n B_{x_k}(r_k)\right) = 0\right) \frac{\mathrm{d}\boldsymbol{\varphi}\mathrm{d}\boldsymbol{r}}{(2\pi)^n},\tag{60}$$

onde $f_{R_1,\ldots,R_n}^{\Phi_{\text{UE}}}$ é a FDP de $(R_1^{\Phi_{\text{UE}}},\ldots,R_n^{\Phi_{\text{UE}}})$, $S_n = \{\mathbf{r} \in \mathbb{R}^n \mid 0 \leq r_1 \leq r_2 \leq \cdots \leq r_n\}$ e $x_k = (r_k \cos \varphi_k, r_k \sin \varphi_k) \in \mathbb{R}^2$ é a posição do k-ésimo UE mais próximo. O cálculo de $\tilde{P}_{\text{AP}}(n)$ é deveras complexo para $n \geq 2$ e não é esperado que exista fórmula fechada para tal. Porém, é interessante obter uma aproximação. A complexidade dos cálculos é a união de todos os discos ao redor dos UEs. Portanto, essa união pode ser aproximada por algo mais simples e, assim, obter uma boa aproximação para $\tilde{P}_{\text{AP}}(n)$. Se apenas os dois discos maiores forem considerados e não existindo intersecção entre eles, essa integral pode ser resolvida. Dessa forma, a seguinte aproximação é obtida

$$\tilde{P}_{\rm AP}(n) \approx \left(1 + \frac{\lambda_{\rm AP}}{\lambda_{\rm UE}}\right)^{-1} \left(1 + 2\frac{\lambda_{\rm AP}}{\lambda_{\rm UE}}\right)^{-(n-1)},\tag{61}$$

Vale destacar que para n = 1 a aproximação é exata.

Mais rigoroso que usar aproximações, é usar limitantes inferiores e superiores. Segue abaixo uma proposição nesse sentido.

Duas variáveis aleatórias $X_1 \in X_2$ são iguais almost-surely se $\mathbb{P}(X_1 \neq X_2) = 0$.



Proposição 1. A probabilidade $\tilde{P}_{AP}(n)$ pode ser limitada inferior e superiormente da seguinte forma

$$\left(1+4\frac{\lambda_{\rm AP}}{\lambda_{\rm UE}}\right)^{-n} \le \tilde{P}_{\rm AP}(n) \le \left(1+\frac{\lambda_{\rm AP}}{\lambda_{\rm UE}}\right)^{-n}.$$
(62)

Outro limitante inferior que é melhor para $n \leq 8 \ e \ \lambda_{AP} \geq \lambda_{UE} \ é$

$$\tilde{P}_{\rm AP}(n) \ge \prod_{k=1}^{n} \left(1 + k \frac{\lambda_{\rm AP}}{\lambda_{\rm UE}} \right)^{-1}.$$
(63)

Demonstração. Para o limitante inferior, um cenário conservador é assumido, para o qual pelo menos n UEs estão dentro de um disco centrado no AP típico e cujo raio é igual à metade do raio de um disco centrado no AP típico. Se o disco maior não contiver APs além do AP típico, então ele será o AP mais próximo de todos os UEs. Esse cenário ocorre com probabilidade menor ou igual à $\tilde{P}_{AP}(n)$, ou seja,

$$\tilde{P}_{AP}(n) \geq \int_{0}^{\infty} \mathbb{P}(\Phi_{UE}(B_{o}(r/2)) > n) f_{R_{1}}^{\Phi_{AP}}(r) dr$$

$$= \int_{0}^{\infty} \sum_{k=n}^{\infty} \frac{(\lambda_{UE} \pi (r/2)^{2})^{k}}{k!} e^{-\lambda_{UE} \pi (r/2)^{2}} 2\pi \lambda_{AP} r e^{-\lambda_{AP} \pi r^{2}} dr$$

$$= \left(1 + 4 \frac{\lambda_{AP}}{\lambda_{UE}}\right)^{-n}.$$
(64)

Para o limitante superior, considera-se que o disco centrado no UE mais longe e raio igual à distância ao AP típico não contém nenhum AP em seu interior. Esse cenário ocorre com probabilidade maior ou igual à $\tilde{P}_{AP}(n)$, ou seja,

$$\tilde{P}_{AP}(n) \leq \int_{0}^{\infty} \mathbb{P}(\Phi_{UE}(B_{x_n}(r)) = 0) f_{R_n}^{\Phi_{UE}}(r) dr$$

$$= \int_{0}^{\infty} e^{-\lambda_{AP}\pi r^2} \lambda 2\pi r \frac{(\lambda \pi r^2)^{n-1}}{(n-1)!} e^{-\lambda \pi r^2} dr$$

$$= \left(1 + \frac{\lambda_{AP}}{\lambda_{UE}}\right)^{-n},$$
(65)

onde $x_n \in \mathbb{R}^2$ é a posição espacial do *n*-ésimo UE e não influencia nos cálculos.

Para encontrar o último limitante inferior, basta usar em (60) a seguinte inequação

$$\mathbb{P}\left(\Phi_{\mathrm{AP}}\left(\bigcup_{k=1}^{n} B_{x_{k}}(r_{k})\right) = 0\right) \leq \prod_{k=1}^{n} \mathbb{P}\left(\Phi_{\mathrm{AP}}\left(B_{x_{k}}(r_{k})\right) = 0\right),$$

pois $|\bigcup B| \leq \sum |B|$. Dessa forma, a integral pode ser resolvida em forma fechada.

A Figura 7 mostra o gráfico das equações obtidas em comparação com simulações de Monte Carlo para $\lambda_{AP}/\lambda_{UE} = 4$.

Em sistemas CF-mMIMO é interessante que $\tilde{P}_{AP}(\tau_p+1)$ seja muito pequena, pois no cenário que o AP tem mais de τ_p UEs próximos, o fenômeno de contaminação de pilotos se intensifica.





Figura 7: As curvas sólidas são os limitantes superior e inferior; a linha tracejada é a aproximação em (61); e as barras de erro são os desvios padrões das simulações de Monte Carlo. A parte inferior da barra de erro não aparece quando o desvio padrão é maior que a média.

Proposição 2. Considere um UE típico e τ_p sequências piloto. Seja C_k a variável aleatória em $\{0,1\}$ que indica se o k-ésimo AP mais próximo desse UE pode servi-lo, $k \in \mathbb{N}$; 1 se possível e 0 caso contrário. Então,

$$\mathbb{P}(C_k = 1) = \sum_{n=1}^{\tau_p - 1} \binom{k+n-2}{k-1} \left(1 + \frac{\lambda_{\mathrm{UE}}}{\lambda_{\mathrm{AP}}}\right)^{-k} \left(1 + \frac{\lambda_{\mathrm{AP}}}{\lambda_{\mathrm{UE}}}\right)^{-(n-1)}, \tag{66}$$
$$\mathbb{E}\left[\sum_{k=1}^{\infty} C_k\right] = \tau_p \frac{\lambda_{\mathrm{AP}}}{\lambda_{\mathrm{UE}}}, \qquad (n \text{úmero esperado de APs que podem servir o UE típico}). \tag{67}$$

Demonstração. O *k*-ésimo AP mais próximo estará livre se houver menos que τ_p UEs mais próximos dele. Portanto, a primeira equação é uma aplicação direta do Lema 1 com $\lambda_1 = \lambda_{\rm UE}$ e $\lambda_2 = \lambda_{\rm AP}$, dado que $\tau_p - 1$

$$\mathbb{P}(C_k = 1) = \sum_{n=1}^{\gamma_p - 1} \mathbb{P}(N_k^{\Phi_{\mathrm{UE}}, \Phi_{\mathrm{AP}}} = n).$$
(68)

Para mostrar o segundo resultado, utiliza-se que a esperança é um operador linear, assim

$$\mathbb{E}\left[\sum_{k=1}^{\infty} C_k\right] = \sum_{k=1}^{\infty} \mathbb{E}[C_n] = \sum_{k=1}^{\infty} \mathbb{P}(C_k = 1) = \sum_{k=1}^{\infty} \sum_{n=0}^{\tau_p - 1} \binom{k+n-1}{n} \left(1 + \frac{\lambda_{\rm AP}}{\lambda_{\rm UE}}\right)^{-n} \left(1 + \frac{\lambda_{\rm UE}}{\lambda_{\rm AP}}\right)^{-k}.$$
(69)

Novamente, é possível trocar os somatórios de lugar pelo teorema de Tonelli. Logo,

$$\mathbb{E}\left[\sum_{k=1}^{\infty} C_k\right] = \sum_{n=0}^{\tau_p - 1} \frac{\lambda_{\rm AP}}{\lambda_{\rm UE}} = \tau_p \frac{\lambda_{\rm AP}}{\lambda_{\rm UE}}.$$
(70)



O valor esperado do número de APs que servem o UE típico era previsível, uma vez que cada AP serve τ_p UEs.

É importante notar que as variáveis aleatórias C_1, C_2, \ldots não são independentes, pois as regiões cobertas pelos APs podem se intersectar. Portanto, encontrar a probabilidade de pelo menos um AP servir o UE típico é muito difícil e requer aproximações.

Uma probabilidade relevante é a probabilidade do AP mais próximo do UE típico não poder servi-lo, dada por

$$1 - \mathbb{P}(C_1 = 1) = \left(1 + \frac{\lambda_{\rm AP}}{\lambda_{\rm UE}}\right)^{-\tau_p},\tag{71}$$

que é uma fórmula semelhante ao limitante superior na Proposição 1. A probabilidade do segundo AP mais próximo não poder servi-lo é

$$1 - \mathbb{P}(C_2 = 1) = \left(1 + (1 + \tau_p)\frac{\lambda_{\rm AP}}{\lambda_{\rm UE}}\right) \left(1 + \frac{\lambda_{\rm AP}}{\lambda_{\rm UE}}\right)^{-(\tau_p + 1)}.$$
(72)

Vale destacar que a probabilidade dos dois APs mais próximos de um dado UE não poderem servi-lo não é igual ao produto $(1-\mathbb{P}(C_1=1))(1-\mathbb{P}(C_2=1))$, pois $C_1 \in C_2$ conforme mencionado anteriormente não são independentes. Apesar desse produto ser um limitante superior da probabilidade, verifica-se numericamente que não é considerado uma boa aproximação.

5.5 Conclusão

Este trabalho propôs a definição de Φ -vizinhos em um processo pontual Φ (Definição 1). Partindo desta definição, diversos resultados relativos à posição espacial de entidades distribuídas de acordo com PPPs foram provados. Esses resultados se aplicam no contexto das redes *cell-free* quando UEs e APs estão distribuídos no \mathbb{R}^2 de acordo com PPPs. Foram caraterizadas a distribuição da relação de vizinhança entre UEs e APs (Lema 1), da relação de distâncias (Teorema 1), da probabilidade de um AP estar saturado (Proposição 1), e da probabilidade de um UE ser servido por um dado AP vizinho (Proposição 2).



6 Projeto de *Beamforming* na IRS baseado em S-CSI

Intelligent Reflective Surfaces (IRS) [31] são meta-superfícies compostas por um grande número de elementos capazes de refletir o sinal incidente a partir de uma determinada mudança de fase/amplitude. A partir da implantação massiva de IRSs em redes de comunicação sem fio e da coordenação de seus elementos, os canais podem ser controlados de forma intencional para melhorar a qualidade do sinal no receptor e aumentar a capacidade da rede [32, 33, 34, 35]. Neste sentido, a otimização do *beamforming* na IRS e na BS têm sido estudada em diversos trabalhos [36, 34, 37, 38, 39, 40, 41, 32, 42, 43]. No entanto, tais trabalhos consideram o conhecimento perfeito da Instantaneous Channel State Information (I-CSI) na BS, o que é muito difícil de se obter na prática devido ao grande número de elementos na IRS e da necessidade de implantação de inúmeras cadeias de Rádio Frequência (RF) na IRS. Porém, recentemente, investigamos a otimização conjunta do *beamforming* na BS e na IRS com o objetivo de minimizar a potência de transmissão na BS sem qualquer conhecimento prévio do CSI [36]. No entanto, em [36], foram assumidas fases contínuas na IRS, o que não é viável na prática devido a limitações de hardware.

Motivados pelos trabalhos acima, consideramos no presente trabalho um sistema de comunicação Multiple-Input Single-Output (MISO) auxiliado por uma IRS. Em seguida, propomos uma nova abordagem para projetar o beamforming na IRS e na BS considerando métodos baseados em Genetic Algorithms (GA). Nosso objetivo é maximizar a taxa alcançável do sistema enquanto atendemos a um requisito máximo de potência de transmissão na BS através da otimização do *beamforming* na IRS e na BS. Para atingir o objetivo principal deste trabalho, a solução proposta foi dividida em duas etapas. Primeiro, o *beamforming* na BS e o desfasamento de cada elemento da IRS são projetados com base apenas no Statistical Channel State Information (S-CSI). Segundo, o par de *beamforming* previamente projetado é utilizado como solução inicial do método proposto a fim de melhorar sua convergência. Nesta etapa, o projeto do beamforming na BS e na IRS é realizado com base apenas na realimentação da SNR no UE. Aqui, é importante notar que a solução proposta não requer nenhum processo explícito de estimativa de canal. Portanto, as principais contribuições deste trabalho são: (i) desenvolvemos um novo framework baseado em métodos meta-heurísticos para projetar o beamforming na BS e na IRS, sem a necessidade do conhecimento do I-CSI; e (ii) revelamos que a solução proposta pode atingir um desempenho próximo ao ideal considerando apenas o conhecimento do S-CSI, com baixa sobrecarga de treinamento.

6.1 Modelo do Sistema

Neste trabalho consideramos um sistema de comunicação sem fio onde uma IRS é implantada para auxiliar a comunicação entre uma BS equipada com N antenas e um UE equipado com apenas uma antena, conforme mostrado na Figura 8. Assume-se que a IRS possui $M = M_{\rm h} \times M_{\rm v}$ elementos refletores. Além disso, consideramos que os canais mudam lentamente, pois assumimos mobilidade limitada. Logo, o sinal recebido no UE é dado por [38]

$$y = \left(\sqrt{\beta_{\rm C}} \mathbf{h}_{\rm iu}^{\rm H} \boldsymbol{\Theta} \mathbf{H}_{\rm bi} + \sqrt{\beta_{\rm d}} \mathbf{h}_{\rm bu}^{\rm H}\right) \mathbf{w} s + n, \tag{73}$$

O conteúdo deste capítulo foi publicado em Souto, V.D.P.; Souza, R.D.; Uchôa-Filho, B.F.; Li, Y. Intelligent Reflecting Surfaces Beamforming Optimization with Statistical Channel Knowledge. Sensors 2022, 22, 2390. https://doi.org/10.3390/s22062390.





Figura 8: Sistema de comunicação de *downlink* com um usuário assistido por uma IRS.

em que β_d , e β_c denotam a perda de percurso do link direto (BS-UE) e do link em cascata (BS-IRS-UE), respectivamente, $\mathbf{h}_{iu}^{\mathrm{H}} \in \mathbb{C}^{1 \times M}$ é o vetor de canal entre a IRS e o UE (IRS-UE), $\mathbf{H}_{\mathrm{bi}} \in \mathbb{C}^{M \times N}$ denota a matriz do canal entre a BS e a IRS (BS-IRS), $\mathbf{h}_{\mathrm{bu}}^{\mathrm{H}} \in \mathbb{C}^{1 \times N}$ denota o vetor do canal BS-UE, $\mathbf{w} \in \mathbb{C}^{N \times 1}$ é o vetor de *beamforming* na BS, *s* são os dados transmitidos, definidos como uma variável aleatória independente com média zero e variância unitária, e $n \sim \mathcal{CN}(0, \sigma^2)$ é ruído gaussiano branco aditivo. Além disso, $\boldsymbol{\Theta} = \mathrm{diag}([\xi]_1 \mathrm{e}^{j[\theta]_1}, \ldots, [\xi]_M \mathrm{e}^{j[\theta]_M})$ onde $[\xi]_m \in [0, 1]$ (com $m = 1, \ldots, M$) e $\boldsymbol{\theta} = [[\theta]_1, \ldots, [\theta]_M]$ denotam o vetor de amplitudes dos coeficientes de reflexão e o vetor de deslocamentos de fase discretos na IRS, respectivamente. A fase discreta $[\theta]_m$ assume valores no conjunto $\mathcal{T} = \left\{0, \frac{2\pi}{K}, \ldots, \frac{2\pi(K-1)}{K}\right\}$, em que $K = 2^B$, com B sendo o número de bits para cada elemento da IRS. Além disso, de acordo com [37, 38], consideramos $[\xi]_m = 1$ para $m = 1, \ldots, M$.

Na abordagem proposta, consideramos dois modelos diferentes de perda de percurso. O primeiro modelo representa a perda de percurso no link direto BS-UE (em dB) e é dado por [44]

$$\beta_{\rm d} = G_{\rm t} + G_{\rm r} - 22 \log_{10}(d_{\rm BU}) - \beta_0, \tag{74}$$

em que G_t e G_r são os ganhos de antena na BS e UE (em dBi), respectivamente, d_{BU} é a distância entre a BS e o UE, e β_0 denota o fator de perda de percurso. O segundo modelo representa a perda de percurso do link BS-IRS-UE (em dB), e é dada por [45]

$$\beta_{\rm c} = G_{\rm t} + G_{\rm r} + 10 \log_{10}(D) - \beta_1, \tag{75}$$

em que β_1 é o fator de perda de percurso, $D = \left(\frac{ab}{d_{\rm BI}d_{\rm IU}}\right)^2 \cos^2 \phi_{\rm c}$, $d_{\rm BI}$ é a distância entre a BS e a IRS, $d_{\rm IU}$ é a distância entre a IRS e o UE, $a = M_{\rm h}d_{\rm IRS}\lambda$ e $b = M_{\rm v}d_{\rm IRS}\lambda$ representam as dimensões da IRS, λ é o comprimento de onda e $\phi_{\rm c} = \arctan(y_{\rm b}/x_{\rm b})$ é o ângulo de chegada na IRS, em que $(x_{\rm b}, y_{\rm b})$ denota a posição da BS [45].

Além disso, para o desvanecimento em pequena escala, consideramos o modelo Rice para os links BS-UE e IRS-UE, os quais são definidos por

$$\mathbf{h}_{i} = \left(\sqrt{\frac{\kappa_{i}}{1+\kappa_{i}}}\mathbf{h}_{i}^{\mathrm{LoS}} + \sqrt{\frac{1}{1+\kappa_{i}}}\mathbf{h}_{i}^{\mathrm{NLoS}}\right),\tag{76}$$



em que $i \in \{iu, bu\}$ representa o link IRS-UE ou BS-UE, e κ_i é o fator Rice. Para simplificar, neste trabalho consideramos $\kappa_i = \kappa \forall i$. Além disso, $\mathbf{h}_i^{\text{LoS}}$ representa os componentes determinísticos *Line-of-Sight* (LoS) que definem o S-CSI e $\mathbf{h}_i^{\text{NLoS}}$ denota o desvanecimento Rayleigh dos links. É importante salientar que, de acordo com [46, 47, 48, 49], consideramos um link LoS entre a BS e a IRS ($\mathbf{H}_{\text{bi}} = \mathbf{H}_{\text{bi}}^{\text{LoS}}$). Além disso, consideramos links com LoS parcial entre a BS e o UE (link BS-UE), e entre a IRS e o UE (link IRS-UE), ambos definidos por (76). Essa é uma formulação geral e está de acordo com [50, 37]. Portanto, neste trabalho, definimos o S-CSI como ($\mathbf{H}_{\text{bi}}^{\text{LoS}}, \mathbf{h}_{\text{bu}}^{\text{LoS}}, \mathbf{h}_{\text{iu}}^{\text{LoS}}$).

Os componentes LoS de tódos os links são expressos pelas respostas do arranjo de antenas na IRS e na BS que são dependentes da geometria do arranjo. Neste trabalho, consideramos que a BS está equipada com uma Uniform Linear Array (ULA), como em [51, 52], e a IRS está equipada com uma Uniform Planar Array (UPA). Assim, os componentes LoS dos links BS-IRS, BS-UE e IRS-UE são, respectivamente, dados por $\mathbf{H}_{bi}^{\text{LoS}} = \mathbf{a}_{\text{IRS}} (\varphi_{\text{IRS}}, \theta_{IRS})^{\text{H}} \mathbf{a}_{\text{BS}} (\theta_{\text{BS}}) e^{j\phi_{\text{bi}}}$, $\mathbf{h}_{bu}^{\text{LoS}} = \mathbf{a}_{\text{BS}} (\theta_{\text{BS}}) e^{j\phi_{\text{bu}}}$, e $\mathbf{h}_{iu}^{\text{LoS}} = \mathbf{a}_{\text{IRS}} (\varphi_{\text{IRS}}, \theta_{IRS}) e^{j\phi_{\text{iu}}}$, em que φ_{IRS} e $(\theta_{\text{IRS}}, \theta_{\text{BS}})$ denotam o ângulo azimute e de elevação, respectivamente, e $\phi_{\text{bi}}, \phi_{\text{bu}}$ e ϕ_{iu} denotam a fase aleatória nos componentes LoS dos links BS-IRS, BS-UE e IRS-UE, respectivamente. Por sua vez, $\mathbf{a}_{\text{BS}} (\theta_{\text{BS}})$ e $\mathbf{a}_{\text{IRS}} (\varphi_{\text{IRS}}, \theta_{IRS})$ denotam os fatores de arranjo de uma ULA e de uma UPA, respectivamente, e são dados por [31]

$$\mathbf{a}_{\mathrm{BS}}(\theta_{\mathrm{BS}}) = \frac{1}{N} \Big[1, \mathrm{e}^{jkd_{\mathrm{BS}}\cos(\theta_{\mathrm{BS}})}, \dots, \mathrm{e}^{jkd_{\mathrm{BS}}(N-1)\cos(\theta_{\mathrm{BS}})} \Big],\tag{77}$$

$$\mathbf{a}_{\mathrm{IRS}}(\varphi_{\mathrm{IRS}},\theta_{IRS}) = \left[\mathrm{e}^{j\mathbf{\Lambda}(\varphi_{\mathrm{IRS}},\theta_{\mathrm{IRS}})^{\mathrm{T}}\mathbf{u}_{1}},\ldots,\mathrm{e}^{j\mathbf{\Lambda}(\varphi_{\mathrm{IRS}},\theta_{\mathrm{IRS}})^{\mathrm{T}}\mathbf{u}_{\mathrm{N}}}\right],\tag{78}$$

em que $k = \frac{2\pi}{\lambda}$ é o número de onda, $\mathbf{u}_m = [0, \mathcal{I}(m)d_{\text{IRS}}\lambda, \mathcal{J}(m)d_{\text{IRS}}\lambda]$, para $m = 1, \ldots, M$, em que $\mathcal{I}(m) = \mod(m-1, M_{\text{h}}), \mathcal{J}(m) = \lfloor (m-1)/M_{\text{h}} \rfloor, d_{\text{BS}}$ e d_{IRS} são as distâncias entre os elementos na ULA da BS e entre os elementos de antenas na UPA da IRS, respectivamente, e $\Lambda(\varphi_{\text{IRS}})$ denota o vetor de onda, dado por

$$\Lambda(\varphi_{\rm IRS}, \theta_{\rm IRS}) = \frac{2\pi}{\lambda} \Big[\cos(\varphi_{\rm IRS}) \cos(\theta_{\rm IRS}), \ \sin(\varphi_{\rm IRS}) \cos(\theta_{\rm IRS}), \sin(\theta_{\rm IRS}) \Big]^{\rm T}.$$
(79)

Finalmente, a taxa alcançável em bps/Hz no UE é

$$\mathcal{R} = \log_2 \left(1 + \text{SNR} \right), \tag{80}$$

em que SNR é a relação sinal-ruído no UE, e pode ser escrita como

$$SNR = \frac{\left| \left(\sqrt{\beta_{c}} \mathbf{h}_{iu}^{H} \Theta \mathbf{H}_{bi} + \sqrt{\beta_{d}} \mathbf{h}_{bu}^{H} \right) \mathbf{w} \right|^{2}}{\sigma^{2}}.$$
(81)

6.2 Formulação do Problema

O objetivo deste trabalho é otimizar o *beamforming* na BS e os deslocamentos de fase na IRS para maximizar a taxa alcançável no usuário, mas considerando uma restrição de potência de transmissão máxima. Logo, este problema de otimização é dado por





Figura 9: Fluxograma da solução proposta.

$$\begin{array}{ll} \underset{\mathbf{w},\boldsymbol{\theta}}{\operatorname{Maximizar}} & \mathcal{R} \\ \text{Sujeito à} & ||\mathbf{w}||^2 \leq P_{\mathrm{t}}, \\ & [\boldsymbol{\theta}]_m \in \mathcal{T}, \ m = 1, \dots, M, \end{array}$$

$$(82)$$

em que P_t é a potência de transmissão na BS. Pode-se verificar que o problema de otimização acima é não convexo e não pode ser resolvido por um método de otimização padrão. Assim, para superar este problema, propomos uma nova solução baseada em GA, descrita a seguir.

6.3 Projeto de Beamforming com S-CSI

Na abordagem proposta, exploramos o S-CSI no início do processo de projeto de *beamforming*. Dividimos nossa solução em três fases, conforme ilustrado na Figura 9. Na primeira fase, o S-CSI é adquirido usando técnicas de estimativa padrão [53, 54] e realimentado para a BS. Na segunda fase, considerando apenas o S-CSI, os vetores de *beamforming* na BS e IRS são determinados na BS. O par de *beamforming* sub-ótimo é avaliado em termos da taxa instantânea alcançável definida em (80). É importante ressaltar que, nesta fase, consideramos apenas os componentes LoS de todos os links de comunicação. Assim, para cada par de *beamforming*, a BS recebe o *feedback* da SNR observada no UE, e o par de *beamforming* atual é avaliado com base em (80). Portanto, maximizamos a taxa alcançável instantânea. Este processo é realizado para todos os pares de *beamforming*. Observe que, nesta fase, o I-CSI não precisa ser explicitamente estimado, o que reduz consideravelmente o *overhead* de treinamento. Para finalizar, as principais etapas da abordagem proposta são descritas a seguir.

- 1. Estimação do S-CSI: Nesta fase, consideramos que a IRS está no modo *sensing* e os S-CSIs de todos os links ($\mathbf{H}_{\mathrm{bi}}^{\mathrm{LoS}}, \mathbf{h}_{\mathrm{bu}}^{\mathrm{LoS}}, \mathbf{h}_{\mathrm{iu}}^{\mathrm{LoS}}$) são estimados usando técnicas de estimativa padrão [53, 54].
- 2. Determinação de $\overline{\mathbf{w}}$ e $\overline{\boldsymbol{\theta}}$: Com base no S-CSI estimado, computamos $\overline{\mathbf{w}}$ e $\overline{\boldsymbol{\theta}}$.
- 3. Determinação de w e θ: Nesta fase, os vetores de beamforming na BS e os deslocamentos de fase na IRS a serem testados são definidos na BS usando GA e enviados para a IRS através do controlador ilustrado na Figura 8. Para cada par de beamforming, a BS recebe o feedback da SNR no UE, e o par de beamforming é avaliado com base em (80). Este processo é repetido para todos os pares de beamforming. Além disso, visando acelerar a convergência da solução proposta e reduzir o overhead, uma vez que assumimos um cenário de mobilidade limitada, consideramos no processo de geração da primeira população o conhecimento tanto do par de beamforming baseado em S-CSI calculado na Fase 2 quanto do melhor par de beamforming calculado na realização do canal anterior. Observe que, nesta abordagem, o I-CSI não precisa ser estimado.

Brasil 65



Figura 10: Configuração do sistema.

6.4 Resultados

Nesta seção são apresentados os resultados obtidos considerando a configuração ilustrada na Figura 10, em que $(x_{\rm b}, y_{\rm b})$, $(x_{\rm i}, y_{\rm i})$ e $(x_{\rm u}, y_{\rm u})$ denotam a localização da BS, IRS e UE, respectivamente. Portanto, $d_{\rm BI}$, $d_{\rm BU}$ e $d_{\rm IU}$ são as distâncias entre BS-IRS, BS-UE e IRS-UE, dadas por $d_{\rm BI} = \sqrt{(y_{\rm b} - y_{\rm i})^2 + (x_{\rm b} - x_{\rm i})^2}$, $d_{\rm BU} = \sqrt{(y_{\rm b} - y_{\rm u})^2 + (x_{\rm b} - x_{\rm u})^2}$, e $d_{\rm IU} = \sqrt{(y_{\rm i} - y_{\rm u})^2 + (x_{\rm i} - x_{\rm u})^2}$, respectivamente.

Os resultados numéricos foram obtidos para três soluções de *beamforming*: (i) Upper Bound, denota o *beamforming* ótimo considerando deslocamentos de fase contínuos na IRS e *beamforming* ótimo na BS, obtido através da resolução do problema de otimização em [37, Eq. 12] usando o CVX [55]. Neste *benchmark*, a taxa alcançável resultante é um limite superior para o problema de otimização em (82); (ii) GA - Without CSI, o beamforming na BS e os deslocamentos de fase na IRS são obtidos sem qualquer conhecimento do S-CSI; (iii) GA - With S-CSI, denota a solução proposta apresentada na Seção 6.3, explorando o conhecimento do S-CSI.

Para os resultados apresentados nesta seção foram considerados os seguintes parâmetros de simulação: N = 10, $M_v = 10$, $N_{it} = 10^3$, $N_f = 2$, $N_{tourn} = 2$, $p_{mut} = 8\%$ (para o benchmark GA - With S-CSI), $p_{mut} = 5\%$ (para o benchmark GA - Without S-CSI), $p_c = 90\%$ (para ambos os benchmarks), $d_{IRS} = 0.5$, $G_t = G_r = 3$ dBi, $f_c = 4$ GHz, $d_{BS} = \lambda/2$, $\kappa = 1$ (se não for especificado de outra forma), $(x_b, y_b) = (50, 10)$ m, $(x_i, y_i) = (100, 10)$ m, $\sigma^2 = -80$ dBm [38], $P_t = 20$ dBm, $\beta_0 = 41.98$ e $\beta_1 = 21.98$. Além disso, todas as curvas apresentam a média de 10^3 realizações de canais independentes.

Inicialmente consideramos deslocamentos de fase discretos em cada elemento da IRS. Logo, a Figura 11 apresenta a convergência dos métodos avaliados para diferentes valores de B, considerando M = 40 e $(x_u, y_u) = (95, 8)$ m $(d_{IU} \approx 5 \text{ m}, \text{ ou seja}, \text{ UE próximo da IRS})$. A partir dos resultados, pode-se verificar que, quando consideramos o conhecimento do S-CSI, a solução proposta atinge um desempenho sub-ótimo com apenas $B \geq 3$ bits (Figura 11). Logo, podemos concluir que a solução proposta pode ser aplicada em cenários práticos com deslocamentos de fase de baixa resolução. Portanto, no restante dos resultados apresentados nesta seção, consideramos B = 3 bits.

A Figura 12 compara as taxas alcançáveis do sistema para as três soluções de *beamforming* em função da distância entre BS e UE, considerando M = 40. A partir dos resultados obtidos é possível concluir que a solução proposta alcança um desempenho próximo ao ótimo para pequenos valores de d_{IU} , ou seja, quando o UE está próximo da IRS. Além disso, verifica-se





Figura 11: Análise do número de bits de controle em cada elemento da IRS, com e sem o conhecimento do CSI, considerando M = 40 e $(x_u, y_u) = (95, 8)$ m, ou seja, $d_{IU} \sim 5$ m.

que, quando $d_{\rm IU}$ diminui, a influência do conhecimento do S-CSI aumenta consideravelmente.



Figura 12: Taxa alcançável versus distância entre BS e UE, para M = 40.

A Figura 13 mostra a taxa alcançável versus o número de elementos na IRS (M) para $d_{IU} \in \{45, 25, 5\}$ m. As seguintes observações podem ser feitas: i) Para $d_{IU} = 45$ m na Figura 13a, a taxa alcançável permanece quase constante. Isso pode ser explicado pelo fato de que, como o UE está longe da IRS, o aumento de M não aumenta a taxa alcançável no UE. Assim, neste caso, a taxa alcançável é mais dependente do *beamforming* na BS; (ii) Para $d_{IU} \leq 25$ m nas Figuras 13b e 13c, a taxa alcançável aumenta consideravelmente com M e a solução proposta atinge um desempenho próximo ao ideal com fases discretas em cada elemento da IRS e apenas um pequeno número de bits de controle. Isso pode ser explicado pelo fato de que, como o UE está próximo da IRS, o UE pode explorar o ganho de *beamforming* na IRS adicionado ao sistema. Além disso, podemos verificar que a posição do UE tem uma influência



considerável; isto pode estar relacionado com o ganho do *beamforming* na BS ou na IRS que o UE pode explorar quando está próximo da BS ou da IRS. Esta conclusão é ratificada na Figura 12, onde podemos notar a importância da posição do UE.



Figura 13: Taxa alcançável no UE versus o número de elementos (M) na IRS.

Em seguida, avaliamos a quantidade de feedback do usuário para a BS, $N_{it}L$. Se a quantidade de feedback não for maior do que o número de pilotos exigido para realizar a estimativa explícita de canal, (MN + 1) [56], então a solução proposta tem uma clara vantagem. A Figura 14 mostra a sobrecarga de treinamento de nossa solução em termos de número de pilotos que seriam necessários para estimar o canal. Podemos ver que o método proposto, mesmo usando apenas 30% do overhead de treinamento necessário para estimar o canal, tem um desempenho próximo ao Upper Bound. Além disso, lembre-se de que nossa solução não requer cadeias de RF adicionais nos elementos da IRS. Por fim, é importante ressaltar que o Upper Bound considera fases contínuas na IRS, enquanto que a solução proposta considera um cenário mais realista, com fases discretas. Logo, podemos verificar que é possível obter uma solução próxima ao ideal com uma quantidade reduzida de feedback do usuário.





Figura 14: Overhead de treinamento da solução proposta, considerando $(x_u, y_u) = (95,8)$ m, ou seja, $d_{IU} \sim 5$ m.

6.5 Conclusão

Neste trabalho, avaliamos duas soluções diferentes baseadas em GA visando projetar o *beamforming* na BS e na IRS, considerando restrição de potência de transmissão máxima e deslocamentos de fase discretos na IRS. Primeiramente, propusemos uma solução que não considera nenhum conhecimento do CSI na BS e realizamos o projeto subótimo do *beamforming* na BS e IRS. Em seguida, propusemos outra solução para resolver o problema de *beamforming* apenas explorando o conhecimento do S-CSI. As novas soluções foram avaliadas considerando diferentes configurações, e a partir dos resultados podemos concluir que elas alcançam um desempenho próximo ao ótimo considerando fases discretas com poucos bits de controle em cada elemento da IRS e com uma quantidade reduzida de *feedback* do UE. Isso mostra que as soluções propostas são muito atrativas na prática.



7 Transferência de Energia Sem Fio usando Radio Stripes

A transferência sem fio de energia via RF, ou *Wireless Energy Transfer* (WET), é uma tecnologia que tem sido explorada para alimentar aplicações de IoT [57, 58, 59, 60]. De fato, a indústria de IoT já aposta fortemente nesta tecnologia. Prova disso é a variedade de empresas emergentes com um grande portfólio de soluções WET, como *Powercast, TransferFi* e *Ossia*.

Neste trabalho, pesquisadores do Projeto Brasil 6G investigaram WET usando esquemas de energy beamforming ideais e sub-ótimos, os quais foram avaliados para um sistema Massive MIMO (mMIMO) distribuído com radio stripes em um ambiente interno, conforme ilustrado na Figura 15, considerando restrições de radiação eletromagnética. Especificamente, a radiação deve ser estritamente limitada na presença de seres vivos. Preocupações sobre as consequências adversas da exposição à radiação resultaram no estabelecimento de limites de exposição como a Maximum Permissible Exposure (MPE), também chamada de densidade de potência, medida em [W/m²]. Nenhum trabalho apresentado na literatura abordou tais restrições no contexto do MIMO-WET e também não considerou um sistema radio stripes.

As principais contribuições alcançadas neste trabalho são:

- Um sistema cell-free com radio stripes para WET foi apresentado. O sistema é composto por N > 1 Power Beacons (PBs) multi-antena sujeitos a restrições de potência de transmissão máxima. Pré-codificadores foram projetados com o objetivo de minimizar a potência de transmissão total.
- 2. Um *framework* analítico foi proposto para avaliar a densidade de potência de RF nas proximidades dos UEs e também em um ponto aleatório da rede.
- 3. Evidências numéricas de que o treinamento não é um fator crítico para o sistema foram apresentadas, pois apenas uma pequena quantidade de energia é necessária para atingir um desempenho semelhante ao de um sistema com CSI ideal.
- 4. Os principais *trade-offs* entre o SDMA e o TDMA foram discutidos. Foi mostrado que o TDMA é mais eficiente quando atende a um pequeno número de UEs, enquanto o SDMA pode ser preferível quando o número de UEs é grande.

7.1 Modelo do Sistema

Como na Figura 15, neste trabalho foi considerada uma rede *cell-free* mMIMO com *radio* stripes composta por um conjunto \mathcal{N} de PBs para $|\mathcal{N}| = N$, cada um com M antenas, alimentando $K \leq MN$ UEs de antena única no downlink. Na prática, $K \ll MN$ deve ser mantido e pode ser necessário para que o sistema de *radio stripes* possa energizar dispositivos que consomem muita energia, como smartphones, controladores de console, etc.

Os PBs são igualmente distribuídos em uma radio stripe de comprimento L que é instalada em torno de um perímetro quadrado de mesmo comprimento, por exemplo, uma sala ou escritório, a uma altura G do nível do piso. Então, o número de PBs que podem ser implantados é

O conteúdo deste capítulo foi publicado em O. L. A. López, D. Kumar, R. D. Souza, P. Popovski, A. Tölli and M. Latva-Aho, "Massive MIMO With Radio Stripes for Indoor Wireless Energy Transfer," in IEEE Transactions on Wireless Communications, vol. 21, no. 9, pp. 7088-7104, Sept. 2022, doi: 10.1109/TWC.2022.3154428.





Figura 15: Modelo do sistema com K = 2, N = 12 e M = 4 (superior) e estrutura do protocolo de treinamento/carregamento (inferior).

limitado por

$$N < \frac{L}{M\lambda/2} = \frac{2L}{\lambda} \tag{83}$$

considerando elementos de antena espaçados de meio comprimento de onda (λ). Além disso, consideramos o desvanecimento Rice, tal que o canal entre PB_n e UE_k é distribuído como

$$\mathbf{h}_{kn} \sim \mathcal{CN}(\bar{\mathbf{h}}_{kn}, \mathbf{R}_{kn}), \tag{84}$$

onde $\bar{\mathbf{h}}_{kn} \in \mathbb{C}^M$ é o componente LoS e $\mathbf{R}_{kn} \in \mathbb{C}^{M \times M}$ é a matriz de covariância. Além disso, o bloco de coerência do canal consiste em τ usos. Antes do *downlink* WET, há uma fase de estimativa de canal de *uplink* que consiste em τ_p usos do canal para transmissões de pilotos dos UEs. Portanto, o *downlink* WET ocorre sobre os $\tau - \tau_p$ usos restantes do canal.

7.2 Formulação do Problema

O objetivo do problema proposto é minimizar o consumo de energia do sistema projetando/definindo adequadamente os vetores de pré-codificação $\{\mathbf{v}_{kn}\}$. O sistema está sujeito a restrições de energia por PB e de EH por UE em cada fase WET. Isto é, cada PB está sujeito a uma restrição de potência total p_{max} e cada UE_k tem um requisito de energia alvo ξ_k por bloco. Com um conhecimento adequado das características de *hardware* do dispositivo de EH, dadas pela função $\{g_k\}$, tal restrição pode ser transformada em

$$\hat{E}_k \ge \xi_k \to \hat{P}_k \ge g_k^{-1} \left(\frac{\xi_k}{1 - \tau_p/\tau}\right) \triangleq \delta_k, \tag{85}$$

onde o acento circunflexo denota estimativa, uma vez que não é possível considerar a estimativa perfeita do canal. Na prática, uma restrição na forma de (85) não impede que a energia coletada





Figura 16: Principais regiões de operação. Para fins de ilustração, a posição dos dispositivos EH é dada por: $\boldsymbol{\zeta}_{k}^{\mathrm{ue}} = [Lk/32, Lk/32, 2^{k/7-1}]$ para k ímpar, enquanto $\boldsymbol{\zeta}_{k}^{\mathrm{ue}} = [L(9 - k)/32, Lk/36, 2^{1-k/7}]$ para k par, com K = 8.

caia abaixo do limite δ_k para estimativas imperfeitas do canal. Além disso, pequenas flutuações em torno de δ_k não afetarão o desempenho da WET na prática.

O problema de otimização pode ser formulado como

$$\mathbf{P1}: \min_{\mathbf{v}_{k'n} \in \mathbb{C}^{M}, \ \forall k', n} \quad P_{T} = \sum_{n=1}^{N} \sum_{k'=1}^{K'} ||\mathbf{v}_{k'n}||^{2}$$
(86a)

subject to $\hat{P}_k(\{\mathbf{v}_{k'n}\}) \ge \delta_k, \qquad \forall k,$ (86b)

$$\sum_{k'=1}^{K} ||\mathbf{v}_{k'n}||^2 \le p_{\max}, \qquad \forall n, \qquad (86c)$$

o qual não é convexo devido à restrição (86b). Neste trabalho são apresentadas diferentes abordagens de otimização para resolver **P1**. Mais especificamente, foram propostas as seguintes abordagens de projeto: (i) Pré-codificador baseado em *Semidefinite Programs* (SDP); (ii) Précodificador ótimo baseado em *Successive Convex Approximation* (SCA); e (iii) Pré-codificador baseado em *Maximum Ratio Transmission* (MRT).

Além disso, analisando **P1**, podemos observar que a topologia de implantação de *radio stripes* permite um *beamforming* tridimensional (3D) distribuído apontando para os locais dos UEs. No entanto, em sua forma atual, **P1** não impede o aumento dos níveis de radiação em outros pontos. Neste trabalho tais restrições de radiação foram consideradas e serão apresentadas a seguir.

7.3 Otimização com Restrições de Radiação

Nessa seção, uma estrutura analítica para determinar, analisar e limitar os níveis de exposição a campos eletromagnéticos são apresentadas. Cada PB das radio stripes (Fig. 15) é assumido estar equipado com um número relativamente pequeno de antenas, tal que $p_{\max}M$ é limitado a um nível desejado, e impomos uma restrição de separação $l_0 > \lambda/2$ entre PBs



consecutivos. Para facilitar a implantação e o projeto do sistema, é assumido que nenhum PB se expande sobre quaisquer dois lados consecutivos do perímetro.

Observe que o foco do trabalho é ambientes *indoor* onde os canais LoS são predominantes. Então, considerando canais LoS, é possível definir facilmente as restrições de sistema para evitar alta exposição à radiação, não necessariamente na proximidade dos UEs. Seja a energia em um ponto de medição virtual $\boldsymbol{\zeta} = [x, y, z]$ dada por P(x, y, z). Nesta abordagem, a seguinte restrição probabilística para exposição à radiação será adotada

$$\mathbb{P}\big[P(x, y, z) \ge \Theta_2\big] \le \varepsilon,\tag{87}$$

onde $\varepsilon \ll 1$ é a probabilidade de violação tolerável e x, y, z são variáveis aleatórias uniformemente distribuídas na região de interesse. Esse tipo de restrição é geralmente difícil de resolver, pois a distribuição de P(x, y, z) depende dos parâmetros de configuração do sistema. Portanto, P(x, y, z) será transformada em uma restrição média usando a desigualdade de Markov como

$$\mathbb{P}\big[P(x,y,z) \ge \Theta_2\big] \le \mathbb{E}\big[P(x,y,z)\big]/\Theta_2 \le \varepsilon \to \mathbb{E}\big[P(x,y,z)\big] \le \varepsilon \Theta_2.$$
(88)

Para prosseguir, duas regiões principais são distinguidas como:

- R_0 : proximidade das *radio stripes*, onde a probabilidade de tecido vivo tende a 0;
- R_1 : área de implantação restante, onde espécies vivas podem andar livremente, excluindo todas as r_k regiões próximas.

Na abordagem proposta, as discussões são concentradas em R_1 , que é fundamental, e a restrição de radiação previamente definida no problema de otimização é inserida. A seguir são apresentados os resultados obtidos para diferentes abordagens de projeto de pré-codificador definidas neste trabalho, considerando as restrições de radiação.

7.4 Resultados

Nesta seção, o desempenho do sistema com *radio stripes* para WET é apresentado, considerando as restrições relacionadas a radiação. Salvo indicação em contrário, os parâmetros do sistema são definidos conforme indicado a seguir.

- Geometria do Cenário: Uma sala com dimensões $6 \times 6 \times 3$ m³ é considerada, onde o sistema de *radio stripes* é assumido no nível do teto, ou seja, $z^{ap} = G = 3$ m. Além disso, é definido G' = 2, 5 m, e salvo indicação em contrário, f = 4 GHz é a frequência central de operação, para a qual $\lambda = 7, 5$ cm. A restrição de separação mínima entre PBs consecutivos é definida como $l_0 = 10$ cm.
- Parâmetros do Transmissor/Receptor/Canal: Os PBs são equipados com M = 8antenas. Além disso, os UEs transmitem os símbolos piloto com $p_k = 10$ mW em cada intervalo de tempo de coerência, enquanto os circuitos de recepção dos PBs experimentam um ruído AWGN com $\sigma^2 = -100$ dBm. Os canais permanecem estáticos para $\tau = 1000$ usos de canal e $\tau_p = K$. Os requisitos de EH dos UEs são considerados homogêneos e iguais a 1 W (potência e energia são equivalentes assumindo tempo de bloco normalizado). Adicionalmente, as matrizes de correlação espacial { \mathbf{R}_{kn} } são geradas, usando o modelo de espalhamento local Gaussiano [61, Cap. 2] com desvio padrão angular de 10°.





Figura 17: a) Potência de transmissão total média (esquerda) e b) tempo médio de execução de otimização (direita) em função do número de UEs.

• Função de Transferência de EH: Considerando a função [62]

$$g_k(x) = \nu_k \left(\frac{1 + e^{ab}}{1 + e^{-a(x-b)}} - 1\right) e^{-ab}, \ \forall x \ge 0,$$
(89)

que descreve a não linearidade de circuitos EH, ajustando adequadamente os parâmetros $a, b \in \mathbb{R}^+$. Para essa função, as igualdades a = 1, b = 4 e $\nu_k = 4$ W são consideradas.

Os resultados apresentados nesta seção consideram vários UEs em diferentes locais. Mais especificamente, oito localizações diferentes para os UEs são consideradas, conforme ilustrado na Fig. 16. Salvo indicação em contrário, foram definidos $p_{\rm max} = 10$ dB e os parâmetros relacionados às restrições de radiação como $r_k = 1,5$ cm, $\Theta_1 = 1000$ W/m² [63], $\varepsilon = 10^{-2}$ e $\Theta_2 = 0,25$ W.

Inicialmente, avaliou-se o desempenho em termos de consumo médio de potência de transmissão e do tempo de execução necessário em função do número de UEs. Alimentar k UEs significa alimentar UE_1, \dots, UE_k com posições ilustradas na Fig. 16. Como um sistema multiusuário é considerado, algumas restrições foram relaxadas para favorecer a viabilidade de otimização, ou seja, $r_k = 2$ cm e $p_{max} = 20$ dB. A Fig. 17 evidencia que o esquema baseado em SDP leva ao consumo mais eficiente de energia, porém a complexidade aumenta linearmente com o número de UEs. O método baseado em SCA deveria ter levado à solução ótima global. No entanto, esse não foi o caso ao servir 3 ou mais UEs e o procedimento iterativo do SCA antes de atingir a convergência foi interrompido (o *solver* usado no pacote CVX deparou-se com erros de aproximação numérica). A estratégia baseada em MRT funciona extremamente bem ao atender um pequeno número de UEs. No entanto, na configuração considerada não é possível atender a cinco ou mais UEs respeitando as restrições de radiação.

No caso de cenários multi-usuário, a Fig. 18 compara o uso de TDMA e SDMA. Observe que não há distinção de funcionamento entre TDMA e SDMA para k = 1. Além disso, o TDMA mostra-se mais eficiente em termos de consumo médio de potência de transmissão quando atende a um número relativamente pequeno de UEs. Isso ocorre principalmente porque, independentemente do número de UEs, a eficiência do circuito de EH é mantida alta. Enquanto isso, quando o número de UEs é significativamente grande. No caso do SDMA, tem-se uma redução da energia de RF mínima incidente em cada UE. Isso facilita o atendimento das restrições relacionadas à exposição a radiação, podendo superar o desempenho do TDMA. Em





Figura 18: TDMA vs SDMA em termos de a) potência de transmissão total média (topo) e b) tempo médio de execução de otimização (fundo) ao servir um número crescente de UEs. Definimos $\xi = 1/k$ para que a potência total coletada permaneça em 1 W, independentemente do número de usuários.

termos de complexidade, o TDMA mostra-se mais caro computacionalmente do que o SDMA ao usar SDP, enquanto o oposto ocorre ao usar a pré-codificação baseada em SCA/MRT. Além disso, o TDMA permite o uso de estruturas baseadas em SCA e até mesmo baseadas em MRT, que funcionam bem para atender a um único UE. Em geral, os *trade-offs* de desempenho entre SDMA e TDMA são estritamente dominados pela não linearidade de EH e pelas restrições de radiação e estudos adicionais são necessários.

7.5 Conclusão

Neste trabalho, foi abordado um sistema *cell-free* mMIMO com *radio stripes* para carregamento sem fio de UEs que consomem muita energia. Pré-codificadores ótimos e sub-ótimos baseados em SDP, SCA e MRT foram apresentados. Esses pré-codificadores visam minimizar a potência de transmissão das *radio stripes* sujeitas a requisitos de EH por UE. Os pré-codificadores baseados em MRT/SCA são particularmente atraentes para atender um número pequeno de UEs. Foram discutidos os principais *trade-offs* entre escalonamento SDMA e TDMA. O último mostrou-se mais eficiente em termos de consumo médio de potência de transmissão ao servir um número relativamente pequeno de UEs, enquanto o SDMA pode ser preferível quando o número de UEs é relativamente grande.



8 Federated Learning usando ALOHA Multicanal

A popularização da IoT trouxe um volume sem precedentes de dados gerados por sensores e dispositivos inteligentes. No entanto, coletar e processar todos esses dados de maneira centralizada para tarefas de aprendizado de máquina pode levar a altos riscos de vazamento de dados e privacidade. Alternativamente, FL é uma solução de *Machine Learning* (ML) descentralizada na qual um modelo é treinado localmente em cada dispositivo e agregado em uma BS sem a necessidade de trocar dados brutos [64].

No FL padrão, cada dispositivo geralmente treina um único modelo usando um único conjunto de dados. Por outro lado, inspirados nas arquiteturas e características de Wireless Sensors Network (WSN), neste trabalho consideramos um cenário FL onde cada dispositivo treina N diferentes modelos usando N conjuntos de dados. No FL padrão, cada modelo estaria associado a um enlace de comunicação para a BS. No entanto, seguir essa regra resulta em muitos enlaces em FL multimodelo, o que é impraticável para aplicações com largura de banda limitada. Para contornar este problema, pode-se considerar a técnica de esparsificação (sparsification) [65], onde os enlaces são ativados esporadicamente apenas para a atualização do modelo. Além disso, a esparsificação também pode ser integrada com outras técnicas, como compressão [66] e controle de potência de transmissão [67], visando otimizar a eficiência espectral e energética do sistema.

Para um determinado número de enlaces de comunicação, o acesso múltiplo pode ser controlado por meio de agendamento para melhorar a velocidade de convergência e justiça [68]. Embora eficaz, o agendamento requer o *upload* de informações secundárias adicionais para a BS e sinalização de controle de *downlink*, o que pode se tornar caro para dispositivos com recursos limitados. Nesse sentido, os protocolos de *Random Access* (RA) [69, 70] são alternativas ao escalonamento, ao custo de colisões. Para reduzir as colisões, *slotted* ALOHA com vários canais foi aplicado com sucesso ao FL de modelo único em [71], demonstrando grande potencial.

Motivados pelos trabalhos previamente apresentados, propomos um novo protocolo ALOHA multicanal combinado com a técnica de esparsificação, visando reduzir o tempo de convergência em um sistema FL multimodelo, garantindo a equidade da taxa de convergência de diferentes modelos. O problema proposto é otimizado em relação à alocação de canais usando o método da bisseção. Os resultados numéricos corroboram a eficácia do protocolo ALOHA multicanal proposto, alcançando uma velocidade de convergência superior à dos esquemas de referência.

8.1 Modelo do Sistema

Neste trabalho, consideramos um cenário com M canais ortogonais, uma BS e K usuários. Cada usuário tem N conjuntos de dados diferentes e N modelos diferentes. Para o usuário k, o conjunto de dados n está associado ao modelo n (ou seu vetor de peso \boldsymbol{w}_n), $\mathcal{D}_{n,k} = \{\boldsymbol{x}_{n,k}, \boldsymbol{y}_{n,k}\}$, onde $k \in \mathcal{K} = \{1, 2, \dots, K\}$, $\boldsymbol{x}_{n,k}$ é o vetor de amostra não rotulado e $\boldsymbol{y}_{n,k}$ é o vetor de saída. Para treinar NK modelos de K usuários com comunicação eficiente, propomos uma abordagem multimodelo usando esparsificação e *slotted* ALOHA multi-canal com quatro etapas, conforme ilustrado na Figura 19 e descrito a seguir:

1. Atualização dos modelos locais: Cada usuário treina seus N modelos locais independentemente. Para o modelo n, ele minimiza a função custo $f_{n,k}(\boldsymbol{w}_n, \boldsymbol{x}_{n,k})$ usando o

O conteúdo deste capítulo foi publicado em R. V. da Silva, J. Choi, J. Park, G. Brante and R. D. Souza, "Multichannel ALOHA Optimization for Federated Learning With Multiple Models," in IEEE Wireless Communications Letters, vol. 11, no. 10, pp. 2180-2184, Oct. 2022, doi: 10.1109/LWC.2022.3196241.





Figura 19: Usuários calculam a atualização de cada modelo e enviam o modelo com maior norma através de um dos M canais usando *slotted* ALOHA. Se mais de um usuário acessar o mesmo canal, há uma colisão. Um usuário pode ter nenhuma, uma ou várias atualizações para enviar em cada iteração.

algoritmo *Gradient Descent* (GD), produzindo o gradiente $\boldsymbol{g}_{n,k}(t) = \nabla_{\boldsymbol{w}_n} f_{n,k}(\boldsymbol{w}_n, \boldsymbol{x}_{n,k})$ na rodada de comunicação t.

- 2. ALOHA multi-canal esparso: A BS aloca M_n canais para os *uplinks* do modelo n dos K usuários, de modo que $M = \sum_{n=1}^{N} M_n$. Cada usuário seleciona a maior norma de gradiente $||\boldsymbol{g}_{n,k}(t)||$ dos N modelos locais, e transmite apenas o gradiente $\boldsymbol{g}_{n,k}(t)$. A transmissão ocorre num canal escolhido aleatoriamente entre os M_n canais.
- 3. Atualização dos modelos globais: Seguindo o método FedAvg [64], a BS agrega as atualizações dos modelos locais de K usuários, construindo o modelo global $\Delta \boldsymbol{w}_n(t)$

$$\Delta \boldsymbol{w}_{n}(t) = \mu \sum_{k \in \mathcal{K}} d_{n,k} \boldsymbol{g}_{n,k}(t) \delta_{n,k}^{S}(t) \delta_{n,k}^{C}(t), \qquad (90)$$

onde $\mu > 0$ é a taxa de aprendizado e $d_{n,k} = |\mathcal{D}_{n,k}| / \sum_{k=1}^{K} |\mathcal{D}_{n,k}|$. A função indicadora $\delta_{n,k}^{S}(t)$ retorna 1 se $\mathbf{g}_{n,k}(t)$ for transmitido, enquanto a função indicadora $\delta_{n,k}^{C}$ torna-se 1 se a transmissão for livre de colisões.

4. Reinicialização dos modelos locais: Cada usuário baixa as atualizações do modelo global $\Delta \boldsymbol{w}_n(t)$ para os N modelos. Para a próxima rodada de comunicação, cada usuário reinicializa o modelo local como $\boldsymbol{w}_n(t+1) = \boldsymbol{w}_n(t) + \Delta \boldsymbol{w}_n(t)$.

8.2 Formulação do Problema

Em (90), note que $\delta_{n,k}^S(t)$ e $\delta_{n,k}^C(t)$ desempenham papéis fundamentais na melhoria da eficiência da comunicação, por não permitir atualizações dos modelos menos importantes e modelar



colisões entre os K usuários, respectivamente. Isso, no entanto, tem o custo de se desviar da atualização desejada do modelo global sob o processo de FL padrão, $\Delta \boldsymbol{w}_n^*(t) = \mu \sum_{k \in \mathcal{K}} d_{n,k} \boldsymbol{g}_{n,k}(t).$

Assim, $\delta_{n,k}^{S}(t) \in \delta_{n,k}^{C}(t)$ podem incorrer em mudanças indesejadas na taxa de aprendizagem efetiva e/ou ameaçar a equidade do treinamento entre os modelos. Por exemplo, na Etapa 2, a alocação igual de M_n canais para todos os N modelos prioriza o modelo de treinamento mais lento que provavelmente produzirá a maior norma de gradiente. Esta operação está longe de ser um treinamento paralelo de N modelos. Para resolver esse problema de justiça, primeiro consideramos as atualizações médias do modelo global:

$$\mathbb{E}\left[\Delta \boldsymbol{w}_{n}(t)\right] = \mu \sum_{k \in \mathcal{K}} \mathbb{E}\left[d_{n,k}\boldsymbol{g}_{n,k}(t)\delta_{n,k}^{S}(t)\delta_{n,k}^{C}(t)\right]$$
(91)

$$\approx \mu p_n^S p_n^C \sum_{k \in \mathcal{K}} \mathbb{E} \left[d_{n,k} \boldsymbol{g}_{n,k}(t) \right], \qquad (92)$$

onde a média é tomada em relação a K. Aqui, assumimos que $\delta_{n,k}^S(t)$ e $\delta_{n,k}^C(t)$ têm covariância zero, ou seja, uma relação não linear. Com base em nossas observações experimentais, consideramos adicionalmente que $\mathbb{E}[\delta_{n,k}^S(t)]$ não é sensível a t, o que é válido sob funções de perda e/ou conjuntos de dados suficientemente diferentes, e o mesmo vale para $\delta_{n,k}^{C}(t)$. Isso leva à aproximação (92), ou seja, $p_n^{S} \approx p_n^{S}(t) = \mathbb{E}[\delta_{n,k}^{S}(t)] \in p_n^{C} \approx p_n^{C}(t) = \mathbb{E}[\delta_{n,k}^{C}(t)]$. Então, temos que garantir a equidade de treinamento multimodelo na média, satisfazendo

 $p_n^S p_n^C = p_m^S p_m^C$ para todos os $m \neq n \in \mathcal{N}$. Assim, formulamos o seguinte problema:

$$\min_{\boldsymbol{w}_n, M_n} \quad \frac{1}{K} \sum_{k \in \mathcal{K}} f_{n,k}(\boldsymbol{w}_n, \boldsymbol{x}_{n,k})$$
s.t. $p_n^S p_n^C = p^S p^C \quad \forall n \in \mathcal{N}$

$$\sum_{n=1}^N M_n = M.$$
(93)

Como a alocação de M_n canais é assumida como independente de t, podemos resolver o problema com relação a \boldsymbol{w}_n e M_n separadamente. Primeiro, temos que minimizar (93) em relação a \boldsymbol{w}_n , que é resolvido pelo Passo 1) descrito na Seção 8.1. Segundo, temos que satisfazer as restrições em (93) ajustando M_n , como descrito na Seção 8.3. Aqui, a probabilidade p_n^C de transmissão livre de colisão de cada usuário caracteriza o desempenho do ALOHA que depende da probabilidade p_n^S de cada usuário enviar sua atualização local, o número de canais (M_n) e o número de usuários (K).

Alocação de Canal Multimodelo Ideal 8.3

Primeiro, observe que o número médio de atualizações locais por usuário recebidas com sucesso na BS em relação ao modelo n é dado por

$$\eta_n = p_n^S p_n^C = p_n^S \left(1 - \frac{p_n^S}{M_n} \right)^{(K-1)} \approx p_n^S \exp\left(-\frac{Kp_n^S}{M_n}\right),\tag{94}$$

onde a aproximação vem de $(1+x) \approx \exp(x)$ para $x \to 0$. Se, em média, as atualizações locais impactarem os modelos globais com a mesma intensidade, então suas taxas de convergência seriam iguais se $\eta = \eta_n$, $\forall n$, respeitando a restrição inter-modelo em (93).



É difícil resolver o problema acima para N > 2, pois ele envolve um sistema não linear com N variáveis e N equações. Assim, recorremos ao método da bisseção [72] para fornecer numericamente a alocação ótima de M_n canais. O objetivo do problema proposto é equalizar as taxas de convergência, $\eta = \eta_n$, $\forall n$. Esta condição pode ser formalmente escrita como

$$\ln p_n^S - \frac{K p_n^S}{\alpha_n M} = c, \quad n \in \mathcal{N},$$
(95)

onde $\alpha_n = \frac{M_n}{M}$ e *c* são constantes. Como α_n é proporcional a *M*, as seguintes condições devem ser verdadeiras: $\sum_{n=1}^{N} \alpha_n = 1$ e $0 \le \alpha_n \le 1$. De (95), pode-se mostrar que

$$\alpha_n = \frac{p_n^S K}{(\ln p_n^S - c)M}.$$
(96)

Com
o $\alpha_n \geq 0,$ então $(\ln p_n^S - c) \geq 0.$ Além disso, com
o $\alpha_n \leq 1,$ temos que

$$c \le \min_{n} \left[\ln p_n^S - \frac{p_n^S K}{M} \right].$$
(97)

Partindo de (96), e como $0 < p_n^S \le 1$, é possível mostrar que α_n é uma função crescente contínua de c, dentro dos limites em (97).

Nosso objetivo é encontrar o c^* ótimo que torna a condição em (95) verdadeira. Com c^* em mãos, então α_n pode ser obtido em (96). Assim, podemos definir a função

$$f(c) = \frac{M}{K} \sum_{n=1}^{N} \alpha_n = \sum_{n=1}^{N} \frac{p_n^S}{(\ln p_n^S - c)}.$$
(98)

Dado que $\sum_{n=1}^{N} \alpha_n = 1$, então

$$f(c^{\star}) = \sum_{n=1}^{N} \frac{p_n^S}{\ln p_n^S - c^{\star}} = \frac{M}{K}.$$
(99)

Para determinar c^* utilizamos o método da bisseção descrito pelo Algoritmo 4.

8.4 Resultados

Nesta seção, investigamos numericamente o desempenho do método proposto por meio de simulações considerando N = 3, K = 100, M = 21, $\{\tau_1, \tau_2, \tau_3\} = \{0.04, 0.08, 0.17\}$, L = 10, $\mu = 0.01$ e $\epsilon = 10^{-5}$, salvo indicação em contrário. Além da tolerância ϵ do método da bisseção, os outros parâmetros configuram a carga de tráfego do sistema. Consideramos um problema de regressão linear com a seguinte função custo [71]:

$$f_{n,k} = \frac{1}{2} |\boldsymbol{x}_{n,k}^T \boldsymbol{w}_n(t) - y_{n,k}|^2,$$
(100)

onde $\boldsymbol{x}_{k,n} \in \boldsymbol{w}_n \in \mathbb{C}^{L \times 1}$ denotam os vetores Gaussianos aleatórios independentes, ou seja, $\boldsymbol{x}_{k,n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \in \boldsymbol{w}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, respectivamente, enquanto $y_{n,k} = \boldsymbol{x}_{k,n}^T \boldsymbol{w}_n$. Para efeito de comparação, consideramos os seguintes *benchmarks*: i) Alocação Uniforme (*Uniform*), em que



Algorithm 4 Método da Bisseção para computar c^* .

Input : K, M, ϵ, N, p_n^S para $n \in \mathcal{N}$ $a = \min_n \left[\ln p_n^S \right] - \max_n \left[\frac{p_n^S K N}{M} \right];$ $b = \min_n \left[\ln p_n^S - \frac{p_n^S K}{M} \right].$ while $|f(c) - M/K| \ge \epsilon$ do c = (a + b)/2if f(c) > M/K then | b = celse | a = cend $c^* = c$ return c^*

os canais são exclusivos e distribuídos igualmente; ii) Canais Compartilhados (*Shared*), em que todos os canais são compartilhados pelos N modelos; e iii) Probabilidade de Cálculo (τ_n) (*Compute*), em que os canais são alocados proporcionalmente à probabilidade de cálculo de cada modelo. As curvas mostradas a seguir são uma média de 100 execuções.

O desempenho em termos de aprendizado de modelo é apresentado na Figura 20. Aqui, o desempenho é definido pela norma de erro, $||\boldsymbol{w}_n(t) - \boldsymbol{w}_n||$. As curvas mostram o desempenho do modelo mais lento para cada esquema (no esquema proposto os N modelos aprendem no mesmo ritmo). O esquema proposto tem uma taxa de convergência muito mais rápida do que os métodos alternativos. Por exemplo, para uma norma de erro de 10^{-3} , o método proposto requer 28,4% menos iterações que a alocação uniforme, 48,6% menos que a estratégia de canais compartilhados e 66% menos que a abordagem proporcional a τ_n .

Em seguida, analisamos o sistema sob diferentes cargas de tráfego. Na Figura 21 à esquerda, variamos o número de usuários de 25 a 300. Se a carga de tráfego for muito alta, o desempenho das três abordagens tende a ser o mesmo, já que quase nenhum treinamento é realizado devido ao número excessivo de colisões. Além disso, se o número de usuários for muito baixo, em algumas iterações não há atualizações locais para enviar, levando a um aprendizado muito lento e ineficiente. Em condições normais de tráfego, o método proposto tem um desempenho significativamente melhor do que as outras abordagens. O lado direito da Figura 21 apresenta os resultados quando corrigimos o número de usuários, K = 100, enquanto o número de canais varia. Novamente, sob condições de carga de tráfego regular, o esquema proposto tem um desempenho consideravelmente melhor; à medida que o número de canais aumenta, a carga de tráfego do sistema torna-se mais leve e o desempenho do esquema proposto e de alocação uniforme tende a ser o mesmo, pois há excesso de canais disponíveis.

Para investigar o impacto do número de modelos sendo treinados, consideramos uma configuração com K = 100 usuários, M = 21 canais e $N \in \{2, 3, 4\}$ modelos. A Tabela 5 apresenta o número de iterações necessárias pelo modelo mais lento em cada método para atingir uma norma de erro de 10^{-3} . Quanto mais modelos aprenderem em paralelo, mais lenta é a convergência, enquanto o método proposto apresenta a convergência mais rápida em todos os casos.





Figura 20: Norma de erro $||\boldsymbol{w}_n(t) - \boldsymbol{w}_n||$ do modelo mais lento para diferentes abordagens: Compartilhado (*Shared*), Uniforme (*Uniform*), Proporcional a τ_n (*Compute*) e o Esquema Proposto (*Proposed*).



Figura 21: Norma de erro $||\boldsymbol{w}_n(t) - \boldsymbol{w}_n|| \mod t = 1000$ do modelo mais lento para diferentes abordagens: Compartilhado (*Shared*), Uniforme (*Uniform*), Proporcional a τ_n (*Compute*) e o Esquema Proposto (*Proposed*).

Para N = 2, há mais canais por modelo, de modo que o desempenho melhora com as probabilidades de computação.

8.5 Conclusão

Neste trabalho foi proposto um método para alocar vários canais em um sistema FL multimodelo baseado em ALOHA. O método proposto aloca canais com o objetivo de equalizar as taxas de convergência. Através dos resultados obtidos verificou-se que o método proposto supera consideravelmente outros esquemas de alocação de canais.



		$ t \text{ para } \boldsymbol{w}_n(t) - \boldsymbol{w}_n \le 10^{-3}$			
N	$ au_n$	Proposed	Uniform	Shared	Compute
2	$\{0.04, 0.08\}$	467	515	610	601
	$\{0.04, 0.17\}$	496	517	956	914
	$\{0.04, 0.12\}$	491	523	956	722
	$\{0.08, 0.17\}$	374	402	546	549
	$\{0.08, 0.12\}$	363	367	439	437
	$\{0.17, 0.12\}$	374	382	395	416
3	$\{0.04, 0.08, 0.17\}$	610	852	1187	1373
	$\{0.04, 0.08, 0.12\}$	575	595	968	1041
	$\{0.04, 0.17, 0.12\}$	641	869	1335	1616
	$\{0.08, 0.17, 0.12\}$	593	822	796	948
4	$\{0.04, 0.08, 0.17, 0.12\}$	903	2348	1877	2202

Tabela 5: Impacto de múltiplos modelos.



9 Handover e Mobilidade em Redes 6G

9.1 Introdução

As previsões para as redes *Beyond Fifth Generation of Mobile Network* (B5G) e 6G contemplam diversas aplicações com mobilidade, como veículos autônomos e Internet tátil. Contudo, devido às interrupções de serviço e congestionamentos das interfaces de rede, torna-se difícil manter os indicadores de desempenho, como latência abaixo de 1 ms e confiabilidade de 99.999%, em cenários de mobilidade [73]. Enlaces na banda mmWave e Terahertz (THz) são particularmente afetados pela mobilidade, devido à alta direcionalidade das antenas usadas nessas faixas de frequência e à baixa capacidade de difração que leva à dependência de enlaces LoS. Haja vista a capacidade que essas bandas oferecem devido a disponibilidade de espectro, é essencial encontrar métodos que habilitem a mobilidade e suportem os requisitos de desempenho exigidos pelas aplicações recentes, como jogos em realidade aumentada e redes veículo-infraestrutura (do inglês, *Vehicle to Infrastructure* (V2I)).

O principal mecanismo de mobilidade é o *handover*. Para o 3GPP, o *handover* compreende o processo de escolha de uma nova BS para receber um UE, a transferência dos dados e o sincronismo do UE com a nova BS. Em geral, um *handover* acontece quando um usuário se encontra na borda da área de cobertura de uma BS e se move em direção à área de cobertura de outra BS. O *handover* deve ser transparente para o UE em mobilidade, ou seja, o *handover* não deve interferir na experiência do usuário. Assim, os principais problemas a serem atacados em *handover* são as falhas de *handover*, o efeito *ping-pong*, o congestionamento das BSs e reduzir os tempos de transferência dos dados e de sincronismo do UE [74].

Em situações altamente dinâmicas, os mecanismos de *handover* tradicionais, baseados apenas em potência recebida, não são capazes de lidar com os índices de capacidade, latência e confiabilidade, como exigido por aplicações de Ultra Reliable Low Latency Communications (URLLC), por exemplo. Entre os fatores que impedem esses mecanismos de handover de atingir o desempenho desejado para as futuras tecnologias, destaca-se o efeito *ping-pong*, que é quando um UE faz handover para uma BS, mas não permanece nela por mais que um intervalo de permanência mínimo e retorna para a BS anterior. Os handovers resultantes do efeito ping-pong são considerados desnecessários, causando congestionamento das BS e overhead de sinalização. Somado a isso, os *handovers* podem falhar, causando a perda total do enlace entre UE e BS. Essas falhas podem ser causadas por flutuações de duração prolongada no canal de rádio, induzindo os mecanismos de handover a antecipar ou retardar o handover. Logo, os mecanismos de handover são um ponto de melhoria para as futuras gerações de comunicação sem fio para que as funcionalidades da rede sejam mantidas com indicadores de desempenho dentro da margem desejada. Neste capítulo, iremos nos referir à BS à qual o UE está associado como BS de serviço (do inglês, service BS) e à BS que está recebendo o UE em handover designaremos por BS alvo (do inglês, *target BS*).

Por consequência do efeito *ping-pong* ou mau planejamento da rede, os *handovers* em excesso podem causar outros problemas para a rede, como o congestionamento das BS ou falhas de *handover* excessivas. O congestionamento das BS ocorre quando muitas UE fazem *handover* para uma mesma BS, causando sobrecarga de sinalização e excedência do número de UE suportado. Uma forma de controlar o congestionamento é realizar o balanceamento de carga através de técnicas de controle de admissão, aceitando ou rejeitando nas fases preliminares ao *handover* [74]. As falhas de *handover*, por sua vez, são causadas pela antecipação ou atraso de

Capítulo escrito por Davi da Silva Brilhante e José Ferreira de Rezende.



Evento	Condição	Aplicação
A1	RSRP da BS de serviço fica acima de um limiar	
A2	RSRP da BS de serviço fica abaixo de um limiar	
Λ 3	RSRP da BS alvo fica uma margem	Mesma RAT
AJ	acima da RSRP da BS de serviço	
A4	RSRP da BS alvo fica acima de um limiar	
Δ.5	RSRP da BS de serviço fica abaixo de um limiar L_1 e a	
AJ	RSRP da BS alvo fica acima de um limiar L_2	
46	RSRP da BS alvo fica uma margem	
	acima da RSRP da BS secundária	
B1	RSRP da BS alvo inter-RAT	
DI	fica acima de um limiar	RAT diferente
B9	RSRP da BS de serviço fica abaixo de um limiar L_1 e a	
	RSRP da BS alvo inter-RAT fica acima de um limiar L_2	

Tabela 6: Eventos de handover definidos pelo 3GPP [2].

um *handover*, que levam a uma falha do enlace de rádio devido a o UE estar ainda distante da BS alvo, em caso de antecipação, ou da BS de serviço, em caso de atraso. O resultado das falhas de *handover* é a perda de serviço e o reestabelecimento posterior da conexão, causando períodos longos de quebra na conexão. Para mitigar as falhas de *handover* são necessários o planejamento cuidadoso da cobertura da rede e o ajuste dos parâmetros de *handover*.

Especificamente no contexto de mmWave e THz, o handover deve ser acompanhado da mobilidade dos feixes usados na comunicação direcional entre BS e UE. Para exemplificar, um UE com N feixes e uma BS com M feixes, resultam em MN combinações de pares de feixes a serem testadas até que o melhor par seja determinado. Como os vetores de antenas possuem muitos elementos gerando feixes estreitos, o alcance dos sinais nessas bandas é reduzido e a penetração em obstáculos é baixa, o que implica em mais BS sendo implantadas para cobrir uma mesma área. Por consequência, esse aumento no tamanho e complexidade tornará os UE mais propensos a fazer handovers [75]. Assim, é ainda mais necessário que os handovers sejam eficientes junto a mecanismos de gerenciamento de feixes que reduzam o overhead do processo de seleção de feixes.

9.2 3GPP 5G NR Handover

O mecanismo de handover adotado pelo 3GPP para o 5G New Radio (NR) é semelhante ao utilizado no Long Term Evolution (LTE). Uma vez que o UE está em CONNECTED_MODE, ele recebe sequências de referência das BS em seu entorno. A partir de medições realizadas nestas sequências de referência, o UE decide a qual BS irá se associar e também se é necessário fazer handover da BS de serviço. Portanto, um handover acontece quando a RSRP da BS de serviço e da BS alvo atendem a uma condição determinada pelo evento de handover. Os diferentes eventos de handover estão descritos na Tabela 6, mas usaremos como exemplo nesta seção os eventos do tipo A3 para handovers na mesma Radio Access Technology (RAT).

A condição para que um *handover* do tipo A3 ocorra é de que a RSRP de uma BS vizinha se torne maior do que a RSRP da BS de serviço por uma margem composta por um *offset* e uma faixa de histerese. O *offset* e a histerese são definidos pelo padrão, mas podem ser configurados de acordo com o provedor de serviço. Uma vez que essa condição é satisfeita, ela





Figura 22: Exemplo de um processo de *handover*. A RSRP da BS de serviço (em vermelho) fica abaixo da RSRP da BS alvo (em azul) pelo período de TTT.

deve ser mantida por um intervalo chamado TTT. Durante esse intervalo o UE mantém-se medindo as duas BS de interesse, ou seja, a BS de serviço e a BS alvo. Ao fim do TTT, uma vez que a condição que disparou o evento de *handover* tenha sido mantida, o usuário envia à BS de serviço uma mensagem de *Measurement Report* (MR), contendo o resultado das medições realizadas. Caso contrário, se a condição não se mantiver verdadeira até o término do TTT, então o *handover* é interrompido. O TTT, bem como as margens offset e a histerese, previnem que flutuações rápidas do canal ocasionem *handovers* desnecessários ou que possam gerar falhas. A Figura 22 ilustra o processo de *handover*.

Depois que a BS de serviço recebe a mensagem MR, é iniciada a negociação com a BS alvo para receber o UE em *handover*. Primeiro, a BS alvo realiza um controle de admissão para que sua capacidade não seja excedida ao receber um novo UE. Caso aceite o novo UE, então a BS de serviço recebe uma resposta à sua requisição de *handover* e é realizada a notificação do UE e a transferência do contexto do usuário de uma BS para a outra. Uma vez que o UE e a BS alvo terminam suas interações com a BS de serviço, o processo de sincronização entre o UE e a BS alvo se inicia e, ao concluí-lo, a BS alvo será a nova BS de serviço.

Um dos problemas com o método adotado pelo 3GPP para *handover* é a perda de serviço durante a sincronização do UE com a BS alvo. Esse período de interrupção é chamado de *Handover Interruption Time* (HIT). Na banda de mmWave a sincronização do *uplink* é necessária para seleção dos feixes (do inglês, *beam selection*), também chamado de acesso inicial (do inglês, *initial access*), portanto não deve ser ignorada. A depender do número de feixes, o HIT pode aumentar, chegando a 65 ms. Outra questão é o ajuste do TTT de acordo com a velocidade do usuário. Quanto maior a velocidade, menor deve ser o TTT, para que o UE responda a tempo às mudanças de canal devido à alta mobilidade. O inverso também é verdadeiro, ou seja, UE em baixa velocidade deve ter TTT alto para evitar *handovers* desnecessários.

Para minimizar o impacto do HIT busca-se minimizar a duração do processo de sincronização entre UE e BS alvo. Em [76] é proposto um método make-before-break, ou seja, realizar todo o procedimento de handover antes que a associação do UE seja transferida para a BS alvo. Esse trabalho apresenta uma evolução do handover RA-less, caso em que não há sincronização no uplink entre UE e BS. O handover RA-less depende da latência da interface X2 e do ajuste do tempo para o UE se reconfigurar à nova BS. Portanto, no método proposto, o UE permanece associada às duas BS, alvo e de serviço, até que a BS alvo possa atender integralmente à UE. Apesar de reduzir o HIT, esse método aumenta a complexidade do UE para o caso inter-RAT e aumenta a ocupação dos buffers das duas BS. Além disso, para os cenários de mmWave, o



Random Access Shared Channel (RACH) é essencial para a descoberta de quais feixes serão empregados no *uplink*.

Ainda que seja minimizado, o HIT pode ainda degradar o desempenho da rede se handovers ocorrerem com muita frequência. Logo, deve-se minimizar os handovers desnecessários, também conhecidos como efeito ping-pong. O 3GPP classifica como ping-pong todo handover cujo tempo de permanência (do inglês, Time of Stay (ToS)) for menor do que o parâmetro Minimal Time of Stay (MToS) e então retornar para a BS de serviço anterior [77]. Então, handovers causados pelo efeito ping-pong causam interrupções no serviço devido ao HIT, sem contudo serem necessários para a manutenção do serviço prestado ao UE.

Além do HIT, as falhas de handover também prejudicam o desempenho de UE em mobilidade. Um handover é considerado falho quando uma falha de enlace de rádio é detectada. Uma falha de enlace de rádio ocorre quando N310 amostras de RSRP ficam abaixo de um limiar Q_{out} , disparando o temporizador T310. Se houver N311 amostras de RSRP acima de um limiar Q_{in} antes que o temporizador T310 encerre, então o enlace é considerado recuperado. Caso contrário, o enlace é considerado interrompido e o usuário deve refazer o processo de associação com uma nova BS. Assim, as falhas de handover acarretam um aumento nas mensagens de controle, somando as mensagens do handover e da associação para recuperar a conexão do UE à rede.

Na banda de mmWave e THz há mais um fator a ser considerado: o fato de os enlaces serem em sua maioria LoS e de obstáculos presentes no ambiente tornarem esses enlaces NLoS, bloqueando-os. Com a instabilidade de RSRP devido aos bloqueios, os UE ficam mais propensos ao efeito *ping-pong* e aumento da taxa de *handover*. Isso se deve ao fato de os enlaces NLoS resultarem em RSRP muito baixas, tornando elegíveis para *handover* BS distantes ou que também estão em situação de bloqueio, ou seja, situações propensas a mais *handovers*. Portanto, em mmWave os bloqueios frequentes inviabilizam aplicações que utilizam URLLC [78], como veículos autônomos, *Unmanned Aerial Vehicle* (UAV) e *smart* ambulâncias [79].

Para contornar esses problemas, diversas abordagens surgiram na literatura. Em [80] e [81], os autores sugerem o uso de conexão dupla (do inglês, *dual connection*), explorando a diversidade espacial, para reduzir a taxa de *handovers* e assim diminuir o *overhead* gerado pelas mensagens de controle. De fato, a taxa de *handover* diminui em comparação com o *handover* em conexão simples, aumenta a taxa de sucesso no *download* de arquivos e reduz a probabilidade de o UE ficar fora de cobertura. A literatura também sugere o *handover* condicional como forma de reduzir a taxa de falhas de *handover* [82, 83]. O *handover* condicional consiste em aguardar a ocorrência de um evento de execução posterior ao evento de *handover*, chamado então de evento de preparação. Na literatura, também encontramos trabalhos que tratam do ajuste dos parâmetros de *handover*, por exemplo, o TTT e a margem de *offset* de RSRP [84, 85]. Desta forma, as alterações na rede são mínimas, mantendo compatibilidade com o padrão definido pelo 3GPP.

Nesta seção descrevemos a operação de mobilidade do 5G NR através de *handover*, bem como seus desafios e abordagens da literatura para superá-los. Na seção a seguir, iremos descrever métodos para reduzir a taxa de *handover* e o número de *handovers* causados pelo efeito *ping-pong*, considerando um cenário de mobilidade em uma rede 5G NR na banda de mmWave com o efeito de obstáculos estáticos causando enlaces NLoS. Para otimizar o *handover*, será utilizada uma abordagem por meio de aprendizagem de máquina.



9.3 Mecanismo de Handover baseado em Aprendizagem de Máquina

O desempenho das redes sem fio em cenário de mobilidade é degradado pela carga excedente de mensagens de controle e interrupções de serviço causadas pelo *handover*, como evidenciado na Seção 9.2. Nas bandas de mmWave e THz, devido ao alinhamento das antenas, a necessidade de eficiência do *handover* é ainda mais crítica. Portanto, torna-se oportuno o emprego de técnicas de aprendizagem de máquina, que devido à sua capacidade de generalização, levam a desempenho quase-ótimo mesmo em problemas complexos e difíceis de serem solucionados analiticamente.

Na literatura, encontramos algumas abordagens heurísticas para o problema de handover. As heurísticas destacam-se por sua baixa complexidade computacional e oferecerem solução subótima. Em [86], os autores desenvolvem um problema de otimização bi-objetivo para garantir o balanceamento de carga ótimo e a minimização do efeito dos bloqueios nos enlaces mmWave. Contudo, devido à complexidade para resolver o problema de maneira ótima, é proposto um algoritmo para solucionar o dual do problema em tempo polinomial e de forma semi-distribuída, ou seja, realizando parte das operações na Cloud-Radio Area Network (C-RAN). As heurísticas propostas apresentam desempenho superior às demais heurísticas analisadas, mas ainda inferior aos métodos ótimos propostos. Em [87], os autores modelam um esquema de handover, beamforming e relay Device-to-Device (D2D) usando uma formulação Mixed Integer Linear Programming (MILP). Semelhante ao trabalho anterior, é proposto um algoritmo de decisão que, mediante as condições do canal em relação à taxa de dados alcançada, encontra a melhor das formas de manutenção do enlace: relay D2D, realizar handover ou fazer beamforming com a BS. Assim, o desempenho obtido com a heurística fica acima de usar só um dos modos isolados, ou seja, beamforming, handover ou D2D, e próximo à solução ótima resultante da solução do MILP.

As heurísticas, apesar da baixa complexidade computacional, têm a contrapartida de serem métodos específicos para a solução de um certo problema e dependem do conhecimento dos modelos. Por isso, técnicas de aprendizagem de máquina têm tido protagonismo na literatura de redes sem fio nos últimos anos. O *handover* é comumente modelado como um problema de tomada de decisão, e predominam abordagens de aprendizado por reforço, ou seja, um agente interage com o cenário com *estados* definidos através de *ações* e recebe *recompensas* de acordo com o sucesso de cada ação. Existem também modelagens usando classificação e predição, empregando técnicas de aprendizado supervisionado e de aprendizado profundo, por exemplo.

As técnica mais popular de aprendizado por reforço é o *Q-learning*, como empregado em [88] e [89]. O *Q-learning* se baseia em uma tabela que é preenchida para todos os pares estado-ação com a probabilidade aquela ação ser tomada quando o agente se encontra naquele estado. Em ambos trabalhos, as ações são a decisão de *handover* e os estados são indicadores de desempenho, como RSRP e localização. O aprendizado é dividido em duas fases: exploração e intensificação. Durante a primeira, as ações são tomadas de formas mais aleatórias e a tabela é atualizada, enquanto na segunda as ações tomadas correspondem às que maximizam a recompensa com maior probabilidade. A eficiência do *Q-learning* depende do balanceamento das duas fases, para permitir que o modelo corresponda ao cenário sem que retarde a inicialização do sistema. De maneira geral, as demais técnicas de aprendizado por reforço funcionam de maneira similar, mas com maior complexidade, de modo a evitar vieses e permitir convergência mais rápida.

De modo a serem mais robustas em comparação com o *Q-learning*, outras técnicas de aprendizado por reforço modificam a forma como os estados, ações e recompensas se relacionam. Um exemplo é a técnica *Multi-Armed Bandit* (MAB) [90, 91, 92], que permite a criação de contextos



para auxiliar na tomada de decisão em vez de usar apenas as recompensas acumuladas por cada ação. O modelo *actor-critic* emprega múltiplos agentes para aprimorar o treinamento através de uma função de valor que ajusta as recompensas recebidas por cada ação e que é aprendida simultaneamente à função de tomada de decisão. Em [89], as ações são um par de BS para fazer *handover* imediato e uma BS reserva, para a realização do handover em caso de falha, e em [93] o *handover* é realizado com base na BS e o *beam* que será usado na associação com essa BS. Essas técnicas possuem a vantagem de poder operar de modo *on-line*, contudo possuem maior complexidade computacional e também de modelagem.

Com maior capacidade de generalização e de lidar com bases de dados mais volumosas e abrangentes, têm crescido o número de aplicações utilizando o aprendizado profundo. Por exemplo, utilizando reforço profundo é possível predizer eventos com base em dados do passado, ou seja, para os dados de entrada fornecidos é possível mapear qual o resultado mais provável de acontecer em um intervalo de tempo. Em [94] e [95, 96], o handover ocorre de modo proativo, com base em predição de bloqueio usando histórico de *beamforming* e imagens, respectivamente, evitando que o usuário passe longos períodos sem serviço devido a bloqueios de longa duração. Além da predição de bloqueio, outra forma de predição que auxilia a mobilidade é a predição da posição do usuário, como realizado em [97], usando redes neurais profundas com camadas convolucionais, que obteve erro máximo de 4 m na predição da posição e redução de até 70% na frequência dos handovers. Contudo, dois aspectos são passíveis de crítica no uso de redes de aprendizado profundo em aplicações como o handover. Em primeiro lugar, o treinamento pode demandar uma capacidade de processamento acima do que está disponíveis nos dispositivos móveis, exigindo em alguns casos o offloading para a nuvem. Em segundo lugar, redes de aprendizado profundo dependem de muitos ajustes e parâmetros, dificultando sua ampla adoção.

Podemos destacar outras abordagens além de aprendizado por reforço e aprendizado profundo. Em [98], a escolha da melhor BS para fazer handover é baseada em uma classificação pelo histórico de RSRP usando diferentes algoritmos de aprendizado de máquina, como gradient boosting e random forest. Também é frequente na literatura o emprego de redes neurais, realizando as funções de predição, tomada de decisão e reconhecimento de padrões. A técnica de fingerprint é usada em [99] e [100], através de uma rede neural long-short term memory (LSTM), aproveitando dados passados de Received Signal Strength Indicator (RSSI) para predizer as condições futuras do canal e reduzir a latência gerada pelo handover, antecipando a transferência do UE para a BS alvo apropriada. Outra forma de realizar o handover é atacar o problema de forma multi-objetiva, ou seja, visando não apenas que o handover leve uma métricas complementares, por exemplo, latência e a carga de usuários nas BS, que podem melhorar o desempenho da rede em outras aplicações, como Voice over Internet Protocol (VoIP) e streaming. Isso foi mostrado em [101], reduzindo-se a latência, o jitter e a perda de pacotes com uso de uma Recurrent Neural Networks (RNN).

Outro aspecto importante para a definição e aplicação das técnicas de aprendizado de máquina são os dados de entrada. O primeiro argumento é o da disponibilidade e plausibilidade dos dados. Por isso, os tipos de dados mais frequentes nos artigos são os *Key Performance Indicators* (KPI) já disponíveis ao usuário e à rede, como por exemplo, RSRP, SINR ou RSSI [102]. Esse tipo de dado é facilmente obtido através de modelos de canal já consolidados e conhecidos da indústria e academia, mas também podem ser facilmente medidos nos dispositivos de usuário e da rede. Além disso, a posição dos usuários e das BS, usualmente em forma de coordenadas de *Global Positioning System* (GPS), aparece como informação de contexto,



adicionando dados que auxiliam as técnicas de aprendizado de máquina a alcançarem maior acurácia em suas funções [101]. Para o *handover*, esses dois tipos de dado são usados com maior frequência por estarem ligados diretamente à mobilidade do usuário e ao método empregado por padrão pela rede, que é o *handover* baseado em RSRP.

Um tipo de dados menos comum são imagens, devido à quantidade de informação, à sobrecarga na rede que o seu tráfego pode gerar e à complexidade para tornar esses dados privados. Contudo, torna-se cada vez mais factível o emprego de imagens, devido à evolução dos dispositivos móveis e ao aumento da taxa suportada pelas redes sem fio. O uso de aprendizado profundo é um habilitador do processamento de imagens auxiliando as redes sem fio, resultado da capacidade das múltiplas camadas das redes neurais usadas no aprendizado profundo de lidar com os canais de cor e a quantidade de informação existente nas imagens. No que diz respeito ao aprimoramento das técnicas de *handover*, o processamento de imagens pode ser usado para antecipar a ocorrência de bloqueios e sua duração [96], prevenindo interrupções de serviço e melhorando a qualidade da experiência do usuário.

9.4 Conclusão

Ao longo deste capítulo, descrevemos como o *handover* é feito nos padrões de comunicação atuais, em particular no 5G NR, e os seus aspectos positivos e negativos. Para a próxima geração, contudo, é importante reduzir a latência gerada pelo *handover*, mesmo em casos de alta mobilidade e cenários extremos, como locais com densidade de bloqueio alta em redes nas bandas de mmWave e THz. Portanto, o aprendizado de máquina, com sua capacidade de generalização, agrega a possibilidade de reduzir o HIT ou ainda otimizar outros aspectos da rede, por exemplo, balanceamento de carga e *Quality of Service* (QoS). Contudo, as abordagens na literatura frequentemente não estabelecem os limites superior e inferior para as métricas testadas ou ainda a composição de arcabouços, envolvendo técnicas de mineração de dados e *big data*, uso da nuvem para processamento e *federated learning*, que tem um forte apelo quanto à privacidade dos dados dos usuários e em redes com alta densidade de usuários.



10 Beam Selection com o Uso de LiDARs

10.1 Introdução

O crescimento exponencial da demanda por serviços multimídia móveis motivou uma extensa pesquisa para melhorar a eficiência do espectro usando técnicas que aumentam a reutilização do espectro geográfico. Bandas de ondas milimétricas, dado a promessa de 2 GHz de largura de banda e subutilização do espectro na banda de 57 a 72 GHz, são uma solução possível para o problema de fornecer *backhaul* de pequenas células e acesso em redes celulares em camadas. Dentre as vantagens da utilização de bandas de ondas milimétricas, podemos citar a disponibilidade de vários gigahertz de espectro não utilizado [103] e a natureza de comunicações de visada direta que ajuda no controle de interferência entre sistemas. Por outro lado, sistemas de ondas milimétricas dependem de um alto ganho direcional para contornar o inconveniente da alta perda de propagação por percurso quando comparada com os sistemas de comunicação baseados em ondas de radio frequência de maior comprimento de onda. Visando minimizar estes inconvenientes o uso de sistemas de múltiplas entradas e múltiplas saídas (MIMO) são utilizados a fim de viabilizar o uso de arrays de antenas de grande escala permitindo uma geração de feixes estreitos e direcionados com maior ganho e diminuição na perda de propagação, provocando uma maximização na vazão de dados e menor de latência [104][105]. No entanto, este aumento na resolução da formação de feixe faz com que o sistema ganhe uma alta sobrecarga no processo de seleção de feixes.

Com o avanço dos recursos computacionais, estratégias de aprendizado de máquina tornaram-se opções viáveis a fim de reduzir a latência em sistemas de pré-codificação, seleção/rastreamento de feixes, etc. Tais estratégias buscam alcançar um redução de latência através da exploração de informações do ambiente, evitando assim a necessidade de estimar as condições do canal de comunicação através de sinais pilotos.

10.2 LiDAR

Os sensores LiDAR fazem parte das tecnologias de geração de imagens 3D, as quais fornecem informações do ambiente através de ondas óticas. Estes sistemas tem apresentado maior eficiência energética, maior precisão em detecção de alcance, tornaram-se compactos, podem trabalhar em condições ambientais variadas (como chuva, neve e nevoa), tornando-se uma escolha adequada para imageamento 3D [106][107].

Ao longo dos anos, as tecnologias LiDAR tornaram-se um dos meios de aquisição de informações geoespaciais mais utilizados. Embora existam muitos tipos de sistemas LiDAR a maioria deles é composto por 5 componentes principais: 1) uma plataforma móvel; 2) um sistema de navegação que integra antenas de sistemas de navegação global por satélite *Global Navigation Satellite System* (GNSS), unidades de medidas incerciais Unidade de Medição Inercial (IMU) e um indicador de medidas de distância Indicador de Medição de Distância (DMI); 3) scanner(s) a laser; 4) câmera(s); e 5) um sistema de controle incluindo armazenamento de dados e integrando a função de todos os sensores [108] e dependendo da configuração do sistema é possível que o mesmo consiga gerar vários Giba bits por segundo, logo é necessário o uso de técnicas que permitam comprimir este volume de dados em nuvem de pontos 3D para transferir a informação do sensor a uma unidade de computação eletrônica (ECU) [109].

Capítulo escrito por Joanna Manjarres, Luan Gonçalves, José Ferreira de Rezende e Aldebaro Klautau.



Os sensores Laser, emitem feixes de laser contínuos em um incremento angular fixo ou definido pelo usuário para medir as distâncias dos objetos. Para isto, tipicamente duas abordagens são usadas, a medição de pulso e a medição de mudança de fase. A abordagem de medição de pulso possui maior alcance, porém com precisões relativamente baixas, variando de subcentímetro a centímetro e menores taxas de dados (<500.000 pontos/s em varredura a laser terrestre (TLS)/LiDAR móvel e <20.000 pontos/s em varredura a laser no ar (ALS)). Nesta abordagem é usada a técnica *Time-of-Flight* (ToF), a qual prove informações de distancia a partir do tempo de voo da luz emitida. É tipicamente usada em sistemas LiDAR móveis e podem ser encontrados em sistemas comerciais como Velodyne , Optech Lynx , RIEGL , Sick , Leica Trimble e DYNASCAN MDL . Para esta abordagem, duas arquiteturas foram encontradas;

- Scanning LiDAR: o qual consiste em um ou mais emissores de luz (TX) e um ou mais detetores de luz refletida (RX) e
- Pulsed LiDAR: pode atingir longas distancias, pois, usa pulsos de pequena duração (na ordem de nanossegundo) com alta potência de pico instantânea mantendo a potencia media abaixo de limite seguro para os olhos [109][110].

Na abordagem da medição de mudança de fase, as ondas podem ser moduladas em amplitude ou em frequência, esta abordagem é caracterizada por ter altas taxas de dados (até 1.000.000 pontos/s) e maior precisão variando de sub-milímetro a sub-centímetro, porém de curto alcance [110]. Alguns dos sistemas LiDAR comerciais que usam esta abordagem são Z+F e Faro . Esta abordagem pode ser categorizada nos esquemas onda contínua modulada em amplitude pulsada (AMCW) e onda contínua modulada em frequência (FMCW) [109]

- Nos esquemas de AMCW são usados lasers de ondas quase continuas no transmissor, e sua intensidade pode ser variada. O alcance é medido através da diferença de fase entre o feixe de laser emitido e a luz de laser retornada [111]. Este esquema não oferece resolução de faixa fina com capacidade de detecção de vários alvos, mas pode alcançar uma precisão inferior a um centímetro. As principais aplicações deste tipo de abordagem são aplicações *indoor*, como jogos e robótica [109] e,
- nos esquemas de FMCW, a intensidade de luz é fixa e o alcance é medido explorando as frequências de batimento [111]. Estes esquemas são usados para alcançar precisão submicrométrica em aplicações de médio e longo alcance, como maior sensibilidade e robustez a variações no ambiente como mudanças de temperatura ou ruido [109].

Os principais mecanismos de varredura dos sensores LiDAR encontrados no estado da arte podem ser categorizados em sistemas optomecânicos, eletromecânicos, microeletromecânicos (MEMS) e varredura de estado sólido. Em [107] os autores apresentam um detalhamento de

http://www.riegl.com

https://velodynelidar.com

https://www.geo3d.hr/3d-laser-scanners/teledyne-optech/optech-lynx-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/teledyne-sg-mobile-mapper/scanners/tel

https://www.sick.com/br/es/soluciones-de-medicion-y-deteccion/sensores-3d-lidar/c/g282752

https://geodetics.com

https://geospatial.trimble.com/products-and-solutions/laser-scanning

https://www.laserscanning-europe.com/en/news/mdl-dynascan--new-mobile-mapping-scanner-interval and the second se

https://www.zofre.de/en/laser-scanners

https://www.faro.com/pt-BR/Products/Hardware/Vantage-Laser-Trackers



cada categoria e qual deles seria a melhor escolha de acordo com os requisitos demandados por diversas aplicações.

Sistemas LiDAR podem ser aplicados em diversas áreas como carros autônomos, navegação, drones, robótica, sensoriamento remoto, sistemas avançados de assistência à condução (ADAS), aplicações médicas, industriais, etc. Atualmente, no mundo científico, estes sensores tornaramse uma importante ferramenta no âmbito das redes de telecomunicações, especificamente no cenário de seleção de feixe em sistemas MIMO na faixa de ondas milimétricas, pois o uso das imagens 3D coletadas pelo sistema LiDAR junto a técnicas de inteligencia artificial apresenta resultados promissores principalmente nas tarefas de *precoding* e seleção de feixes ao evitar a necessidade de se estimar o canal, através de pilotos, e e aproveitar informações do ambiente para realizar tais tarefas.

10.3 Aprendizado de Máquina Aplicado a Seleção de Feixe

Usar o aprendizado de máquina para melhorar a telecomunicação é uma das propostas do 5G, mas para isso é necessário usar uma grande quantidade de dados para testar algoritmos e avaliar o desempenho dos sistemas. Ao levar isso em consideração, junto ao fato da coleta real de informações ser custosa por conta dos equipamentos usados e uma análise elabora de área de testes [112], fazer uso de simulações computacionais é considerada uma opção viável. Essas simulações podem ser feitas com os mais diversos tipos de dados, como LiDAR, coordenadas dos alvos/antenas, imagens, entre outras.

Em [113] é apresentada uma metodologia para geração de dados de canal em onda milimétrica em cenários com sistemas de comunicação MIMO. O objetivo era facilitar a investigação de problemas de aprendizado de máquina relacionados a telecomunicações, como a seleção do melhor feixe de comunicação, em cenários com relativo grau de realismo. Assim, a partir dessa investigação, concluiu-se que os dados de LiDAR, que foram simulados na pesquisa, pode ser usado para detecção do melhor canal de comunicação entre as antenas e redução da sobrecarga de seleção de feixe mmWave. O caráter realístico da simulação é sustentado na construção do cenário, uma vez que a seleção da região que o compõe é feita a partir da extração de locais reais[114].

Em [115] foi desenvolvida uma estratégia para seleção de feixe baseada no aprendizado de máquina para um sistema de comunicação entre veículo e infraestrutura V2I mmWave, onde o veículo conectado é equipado com o sensor LiDAR. A Figura 23 representa esse tipo de sistema; têm-se a antena transmissora da estação base interagindo com os veículos equipados com antenas receptoras. A seta verde representa uma comunicação direta entre a estação base e o receptor e as demais representam uma comunicação através de reflexões causadas por bloqueadores. Neste trabalho, a proposta do aprendizado de máquina para seleção de feixe considera que a estação base pode transmitir seu valor absoluto por meio de um canal de controle de baixa frequência, o que possibilita que todo o processo de aprendizado seja feito pelo veículo conectado, descentralizando a simulação e diminuindo a sobrecarga do processo. Por causa autonomia de processamento dos receptores, eles usam suas informações de LiDAR, o valor da própria posição e da posição da antena transmissora para fazer a identificação de todos os raios de comunicação possíveis entre transmissor e receptor, essa técnica é conhecida como beam sweeping (varredura de raios), em outras palavras, seria a cobertura de uma área espacial com um conjunto de feixes transmitidos e recebidos de acordo com intervalos e direções pré-especificadas [116].

A Figura 24 exemplifica essa técnica com um conjunto de feixes que contem oito feixes





Figura 23: Representação de um sistema V2I com enlace LOS beam verde e enlace NLOS beam vermelho e amarelo. Sendo que o beam amarelo representa a difração do sinal chegando no receptor alvo com desfase e menor potencia.

diferentes com uma direção de irradiação diferente para cada. Nesse exemplo a gNB (antena transmissora) é configurada para fazer a varredura da quantidade total de raios de comunicação formados. Ao ser feita a varredura, é estimada uma quantidade de melhores raios de comunicação que são informados e treinados na estação base, onde a rede escolhe a melhor combinação possível de dados.



Figura 24: Representação da seleção de feixe num sistema MIMO mmWave

A Tab. 7 apresenta alguns trabalhos relacionados a tarefa de seleção de feixe com diferentes entradas, diferentes estratégias de ML e divididos em três grupos de estratégias: I) monomodal sem LiDAR; II) multimodais com LiDAR e III) monomodal com LiDAR. Trabalhos que baseiam-se em informações do canal, como CSI e RSS, geralmente possuem uma alta latência pelo fato de haver a necessidade de estimar o canal. Por este motivo, atualmente métodos que exploram informações do ambiente são o maior foco no contexto de seleção de feixes. Por tanto, quanto mais informações a respeito do meio, maior as chances de ser bem sucedido na tarefa. Por sim, observou-se que os sensores que mais fornecem informações são o radar e LiDAR, com


maior atenção para o último devido sua utilização em carros autônomos.

10.4 Seleção de Feixe Usando Lidar

As redes de comunicação baseadas em ondas milimétricas são caracterizadas pela degradação dos sinais ante a penetração de diferentes tipos de materiais, fazendo com que esta tecnologia seja mais dependente de enlaces LoS para que a potência recebida seja suficiente. Quando o enlace é obstruído ocorrem fluatações na qualidade da comunicação ou até mesmo uma perda na conexão, podendo comprometer a confiabilidade e a latência das redes [117]. Uma abordagem de previsão proativa de bloqueadores poderia possibilitar que a rede pudesse se antecipar ao evento e tomar decisões como realizar um *hand-off* oportunístico ou até fazer uma mudança de troca de feixe. As informações de contexto obtidas por sensores como LiDAR, Radar, etc. em conjunto com o uso de ferramentas de *machine learning* poderiam auxiliar na previsão de bloqueio.

O estado da arte apresenta diferentes abordagens para realizar a predição proativa de bloqueio. Em [122], [117] e [135] os autores usaram a base de dados *DeepSense*, composta por uma unidade trasnmissora e uma receptora, Rx e TX. O Rx é composto por uma antena linear uniforme de 16 elementos e o Tx por uma antena omnidirecional. Em [122] os autores usaram um radar na estação base com o intuito de identificar objetos em movimento para antecipar um possível bloqueio no enlace de comunicação através da utlização de uma rede neural profunda. Como resultado deste trabalho foi obtida uma acurácia acima do 90% com uma capacidade de previsão de bloqueio com um segundo de antecedência. Em [117] os autores usaram câmeras RGB como sensores para mapear o ambiente e realizar a previsão de bloqueio. Uma rede neural convolucional recorrente foi utilizada para realizar a previsão. Os resultados apresentam uma acurácia de 90% para um tempo de predição de bloqueio inferior a um segundo e 80% para um segundo. Em [135] é apresentado o problema de previsão de bloqueio auxiliado por LiDAR, Os resultados apresentam uma acuaria de 95% para previsões de 100 ms e 80 para previsões de 1 seg.

Em [133], foi usada uma rede neural profunda para para detecção de LoS e seleção de feixes. Neste trabalho, os autores utilizaram dados de sensores LiDAR de forma centralizada e distribuída. No caso de detecção de LoS a arquitetura de LiDAR distribuído conseguiu o melhor desempenho com uma acurácia de 91% enquanto a arquitetura centralizada alcançou uma acurácia de 79%. No caso de seleção de feixes, todas as arquiteturas apresentaram bons comportamentos, em termos de acurácia, contudo, a solução LiDAR distribuída oferece melhor desempenho, ao custo de exigir um LIDAR por veículo.

As técnicas de *machine learning* mencionadas, são conhecidas por serem computacionalmente custosas. Devido a isto, neste trabalho é proposto a exploração do uso de una rede neural sem peso, também conhecida como Redes Neurais Baseadas em RAM. Esta técnica é caracterizada por operar com versões binarizadas das entradas de interesse, em razão disso, estes modelos são altamente eficazes e de custo operacional relativamente baixo, permitindo a construção de algoritmos flexíveis e aprendizado rápido.

10.5 Rede Neural Sem Peso

A predição de obstaculo foi realizada usando uma Rede Neural Sem Peso (RNSP). Este tipo de redes faz uma representação da sinalização excitatória e inibitória realizada pela árvore dendrítica do neurônio ao mapear a sinalização de forma *booleana* como entradas binarias



artificiais [137], esta quantização dos dados permite uma rápida e ajustável implementação em hardware.

Nas redes Neurais Sem Peso, os neurônios são representados por memorias RAMs e o aprendizado neste tipo de rede depende das mudanças nos conteúdos das memórias RAM, evidenciando aqui a principal diferença com respeito a redes neurais artificiais tradicionais que localizam seu aprendizado nos pesos das conexões, proporcionando uma alta velocidade de aprendizado.

As RNSP datam desde os anos 50 com o trabalho apresentado pelos autores *Bledsoe* e *Brow*ning no reconhecimento de caracteres alfanuméricos [138], nos anos 70 foi patenteada e usada comercialmente a primeira rede neural artificial sem peso, a rede WiSARD e posteriormente surgiram outros modelos como PLN, *Goal Seeking Neuron* (GSN), *Generalization* RAM (GRAM) e Virtual Generalization RAM (VG-RAM) [139].

Para o desenvolvimento deste trabalho foi usada a rede WiSARD. A arquitetura desta rede baseia-se em discriminadores, sendo eles responsáveis pela categorização em problemas de classificação. Cada discriminador é composto por k memórias RAM, e cada uma dessas memórias possuem n entradas responsáveis por endereçar a memória RAM e uma saída binária.

No início da fase do treinamento, todas as posições das memórias RAMs são inicializadas em zero e as entradas são dividas em n tuplas de forma pseudo-aleatória. No entanto, depois que as tuplas são formadas, as mesmas permanecem constantes por todo o ciclo de vida da rede neural. O padrão de uma categoria é apresentado ao discriminador e as memórias RAM que o compõem são acessadas pelo endereço formado pelas tuplas, escrevendo o valor 1 na posição da memória que dito endereço indica, conforme apresentado na Figura 25.



Figura 25: Exemplo de um discriminador da letra E na fase de treinamento

Na fase de classificação da rede WiSARD, os novos dados a serem avaliados usam o mesmo mapeamento de entrada que na fase de treinamento. A entrada irá endereçar as memorias RAM agora para leitura do conteúdo das mesmas, e em caso desta leitura apontar para um '1', a RAM ira gerar como saída esse valor armazenado. A partir disto, o discriminador ira fornecer um grau de similaridade ao somar todos os bits gerados por cada um dos nós RAM que o compõem, e o discriminador com maior valor na sua resposta, indicara que possui uma maior probabilidade dessa entrada pertencer a classe que o discriminador representa. Na figura 26 pode ser observado um exemplo de reconhecimento de carácter alfanumérico na etapa de classificação.





Figura 26: Exemplo de classificação de um número usando a rede WiSARD

Diante da situação em que duas classes tiveram a mesma resposta de saída, em outras palavras, os discriminadores ficarem empatados, é implementada a técnica de *bleaching* [140]. Se trata de adicionar um contador de frequência nas memorias RAM, quando as memorias forem lidas, um contador sera incrementado se o valor acessado na memoria for '1', e a resposta será retornada se valor no contador for maior ou igual ao valor de *bleaching* designado. No caso em que o empate persista, o valor *bleaching* irá aumentando geralmente em 1 até não haver mais empates. Assim, o discriminador selecionado sera aquele que tiver o maior valor na saída [139][141], na figura 27 pode ser observado um exemplo de uso do *bleaching*.



Figura 27: Exemplo do uso do mecanismo bleaching



10.6 Resultados Parciais

10.6.1 Adquisição dos dados

Para o desenvolvimento deste trabalho foram usados os dados disponíveis no *dataset* Raymobtime . Este *dataset* possui uma metodologia de simulação com o intuito de gerar canais de comunicação realistas. Para isto foi necessária a integração de diversos softwares como o Cadmapper e OpenStreetMaps (responsáveis de modelar os cenários realistas 3D *outdoor*), o SUMO (encarregado de dar mobilidade aos veículos), o Blensor (biblioteca do Blender que simula sensores, como LiDAR e as imagens das câmeras) e o Wireless InSite para geração dos raios dos canais de comunicação. Por fim, o *dataset* é composto por dados de *ray-tracing*, LiDAR, canais de comunicação, imagens, e informação de GPS do usuário [142]. Neste *dataset* encontram-se 9 cenários diferentes. Cada cenário esta composto por episódios e estes a sua vez por cenas, existindo uma relação temporal entre as cenas mas não necessariamente entre os episódios. Este trabalho foi realizado a partir dos dados encontrados no cenário s008, o qual é caracterizado por representar uma rua da cidade Rosslyn. Neste cenário existem 10 usuários moveis por cena e apenas uma cena ou "*snapshot*" por episodio, 11.000 canais válidos de comunicação usando uma frequência de 60 GHz nos 2086 episódios. Para realizar a predição de obstáculos e a seleção de feixes foram usados os dados de posição e LiDAR deste cenário.

10.6.2 Pre-processamento dos dados

No *dataset* do Raymobtime os dados do LiDAR estão compostos por uma nuvem de pontos, os quais posteriormente foram quantizados em um histograma, *array* de 3 dimensões, onde cada elemento do histograma corresponde a uma região fixa da área de cobertura. Por fim, estes dados indicam a presença de obstáculos na área escaneada pelo LiDAR, e obedecem a uma arquitetura centralizada onde o LiDAR se encontra na estação base. Na Figura 28, é apresentado no episodio 8 do LiDAR correspondente ao *dataset* 008.

Devido ao fato de que a rede WiSARD tem como restrição o uso de dados binários como entrada, os dados do LiDAR sofreram um preprocessamento. No *dataset* as posições do receptor e transmissor são indicadas pelos valores -1 e -2, sendo assim necessário tratar estes valores. A estrategia usada foi binarizar a informação de LiDAR de forma que todos os valores diferentes de 0 sejam substituídos por 1, e acrescentar das enrtadas novas informando apenas a posição do receptor.

No *Raymobtime* o posicionamento dos veículos no cenário é realizado através do software SUMO e, com o intuito de tornar os dados mais realistas, a simulação do posicionamento dos receptores a través de coordenadas GPS em que foram adicionadas um ruido gaussiano independente e identicamente distribuído [143].

10.7 Resultados Parciais

A predição de bloqueio no contexto de redes veiculares a partir de uma rede neural WiSARD, tem como intuito definir se um enlace de comunicação possui ou não linha de visado a partir

https://www.lasse.ufpa.br/raymobtime/

https://cadmapper.com

https://www.openstreetmap.org/#map=4/-15.13/-53.19

https://sumo.dlr.de/userdoc/

https://www.blensor.org

https://www.remcom.com/wireless-insite-em-propagation-software/





Figura 28: Episódio 8 do LiDAR quantizado do dataset 008

de dados oferecidos por sensores LiDAR.

Para avaliar o desempenho da rede WiSARD, o *dataset* s008 foi divido em dois conjuntos, 80% do *dataset* para treinamento e 20% para teste. O tamanho das memorias RAMs foi variado desde o tamanho 6 até 64, com incrementos de 6, onde cada variação deste parâmetro foi executada 10 vezes. Como resultado parcial deste trabalho, a rede WISARD apresentou um aumento linear na acurácia superando o 70% na melhor configuração, conforme apresentado na Figura 29.



Figura 29: Acurácia em função do hiper-parâmetro tamanho da memória da rede WISARD

10.8 Conclusão

O uso de informações de contexto como posição, imagens, dados provenientes de sensores inerciais ou sensores LiDAR, permitem aos sistemas de comunicação 6G e além a diminuir a sobrecarga dos enlaces o que consequentemente irá melhorar o desempenho da eficiência



espectral e redução da latência. Técnicas de *Machine Learning* possibilitam o uso de grandes quantidades de dados que permitam a seleção do *Beam* ótimo ao identificar o tipo de enlace de comunicação LoS ou NLoS. O uso das redes RNSP viabilizam que aplicações praticas possam ser implementadas a partir de componentes de baixo custo em dispositivos capazes de operar em tempo real. No entanto, o trabalho desenvolvido até agora não tem apresentado uma performance significativa quando comparada com outro tipo de técnicas de ML como uma rede perceptron e uma CNN alcançando uma acurácia de 85%



Grupo	Entrada	Dataset	Algoritmo ML	Ref
I	Posição UE	Raymobtime	Linear SVM, Ada- Boost, Decision tree, Random Fo- rest, Deep Neural Network (DNN), Deep reinforcement learning (Deep RL)	[113]
	câmera RGB	DeepSense 6G	Deep Learning (DL)	[117]
	pilotos	DeepMIMO	DL	[118]
	informações da BS LTE	-	improved fast machine learning (IFML)	[119]
	CSI	DeepMIMO	LSTM	[120]
	RSS	Gerado pelos autores	DNN - Feedforward neural network (FNN)	[121]
	Radar	DeepSense6G	DNN	[122]
	Radar	Gerado pelos autores usando Wireless insite	-	[123]
	Posição	-	DNN	[124]
	Câmera	vision- communication- dataset	DNN	[125]
	Posição e Orientação	Gerado pelos autores	DNN	[126]
	LiDAR, Posição UE e BS	Raymobtime	DL	[115]
II ——	LiDAR, imagens e GPS	Raymobtime e NEU dataset	DL	[127]
	LiDAR e GPS	Raymobtime	lightweight Neural Network (LNN) Convolutional Neural Network (CNN)	[128]
	LiDAR e GNSS	Raymobtime	MLP CNN	[129]
	LiDAR, Posição e imagens	Raymobtime	DNN	[130]
	LiDAR e posição	Raymobtime	FL	[131]
	LiDAR, Cena 3D e COL- MAP	Gerado pelos autores usando Wireless In- Site	DNN	[132]
	LiDAR	Raymobtime	DL	[133]
III —	LiDAR	Raymobtime	CNN	[134]
	LiDAR	DeepSense6G	CNN	[135]
	LiDAR	Raymobtime	CNN	[128]
	LiDAR	DeepSense 6G	RNN	[136]

Tabela 7. Abordagens para seleção de feixes assistido por contexto separados em três grupos

=



11 Ambiente de Simulação para Alocação de Recursos de Rádio para Aprendizado por Reforço

Radio Resource Management (RRM) é um sistema de gerenciamento e coordenação para que dispositivos trabalhem de forma eficiente. Ele engloba processos, estratégias e algoritmos para monitoramento, alocação de usuários, taxas de dados, critérios de transferência, esquema de modulação e codificação de erros. De forma a garantir a qualidade de serviço para determinado usuário móvel, seu objetivo é alocar de forma flexível e dinâmica, os recursos da rede disponíveis na transmissão de dados [144]. Em outras palavras, o módulo RRM inclui principalmente o algoritmo relativo às seguintes funções: controle de potência, alocação de canais, *radio scheduling, handover*, controle de acesso e de carga.

Mas essa é uma tarefa que tem se tornado cada vez mais complexa à medida que as redes móveis evoluem, proporcionando um aumento exponencial do número de usuários e novos dispositivos conectados. Nesse novo mundo, algoritmos convencionais baseados em modelos estáticos não serão mais aplicáveis [145]. Além das características aleatórias de desvanecimento e interferências do canal sem fio, as redes de sexta geração prometem ser ainda mais heterogêneas. Um grande número de dispositivos conectados exigirão uma variedade de serviços com diferentes requisitos. E é aí que surge a ideia de fatiamento da rede. Os chamados *slices* permitem a multiplexação de redes lógicas virtualizadas e independentes numa mesma infraestrutura de rede física.

Nesta etapa do projeto, o objetivo foi desenvolver um ambiente de simulação de RRM adaptado às redes 6G cujo agente de decisão fosse baseado em aprendizado por reforço. Já existem na literatura diferentes propostas de *radio schedulers* [146, 147, 148, 149, 150] mas as suas respectivas simulações costumam assumir cenários e plataformas diferentes, tornando difícil a comparação de métodos para *schedulers*. Por isso, é importante desenvolver um simulador simples e mais abrangente que se integre facilmente a ferramentas de geração de mobilidade e canal externas, assim como a bibliotecas de aprendizado de máquina (como aprendizado por reforço e árvores de decisão) de forma que o principal foco do desenvolvimento seja o método de *scheduler* a ser desenvolvido e a comparação com baselines.

O trabalho desenvolvido até agora está organizado da seguinte forma: a Seção 11.1 descreve em detalhes o simulador enquanto a Seção 11.2 apresenta os resultados obtidos ao se utilizar as técnicas de Round-Robin e aprendizado por reforço em *schedulers*.

11.1 Simulador

O ambiente de simulação 6G Radio Management (sixg_radio_mgmt) foi desenvolvido na linguagem de programação Python devido ao vasto número de ferramentas e bibliotecas disponíveis nessa linguagem para o desenvolvimento de soluções utilizando aprendizado de máquina [151]. Os dois principais cenários de foco do ambiente de simulação são: a) alocação de RBG de cada BS para as UEs associadas, b) alocação de RBGs para slices em cada BS (interslice scheduling) e alocação de RBGs de cada slice para UEs (intra-slice scheduling) [152].

A Figura 30 demonstra o cenário a) em que os RBGs disponíveis na BS são alocados pelo *scheduler* que, por sua vez, pode usar as informações da rede para tomar a decisão de quais RBGs alocar para cada UE. Por exemplo, um algoritmo de *scheduler* baseado na maximização

Capítulo escrito por Cleverson Nahum, Michelle Facina e Aldebaro Klautau.

 $https://github.com/lasseufpa/sixg_radio_mgmt/$



da taxa de rede pode alocar os RBGs para as UEs com maiores valores de eficiência espectral e com mais dados no *buffer* [153]. Além disso o *scheduler* pode vir a atender um número variável de UEs devido a admissão de novas UEs na rede ou mesmo a saída delas.



Figura 30: Processo de *scheduling* em um cenário sem *slices* de rede, no qual os RBGs disponíveis na BS precisam ser distribuídos para as UEs de forma a atender os objetivos para o qual o *scheduler* foi projetado.

A Figura 31 demonstra o cenário b), no qual os RBGs são primeiro distribuídos para os *slices (inter-slice scheduling)*, e posteriormente são distribuídos para as UEs associadas a cada slice (*intra-slice scheduling*). A principal diferença em relação ao cenário sem *slices* é o foco em atender as necessidades dos grupos de UEs em cada slice ao invés de atender as necessidades de cada UE individualmente [154]. Portanto, o objetivo do *scheduler* nesse cenário é atender os requisitos de cada slice de rede.

De forma geral, o ambiente de simulação apresenta uma classe principal responsável por implementar o cenário de comunicação móvel e executar as BSs, *slices* e UEs, levando em consideração a decisão do *scheduler* para alocar os RBGs para as UEs ou *slices*, conforme pode ser visto no diagrama de fluxo da classe da Figura 32. Primeiramente, a classe *Communication Env.* cria o máximo número de BSs, *slices* e UEs que existirão durante a simulação. Em seguida, são instanciadas as classes responsáveis pela geração da informação de canais, mobilidade e tráfego de cada UE (essas classes serão explicadas posteriormente). São realizadas as associações entre BSs, *slices* e UEs, de forma a permitir a alocação de recursos apenas entre as entidades que possuem relação entre si, isto é, o *scheduler* só conseguirá alocar um dado RBG de uma *basestation* ou *slice* para uma UE caso ela esteja conectada a BS ou slice no qual os recursos estão disponíveis. Nesse caso, a principal função das BSs e *slices* é delimitar as ações do *scheduler* de forma a permitir apenas ações condizentes com o cenário. No diagrama de fluxo da classe BS é possível perceber ainda que os RBG são alocados para os *slices* (caso a rede possua *slices* na rede de acesso) ou então diretamente para as UEs de acordo com a decisão do

Brasil 65



Figura 31: Processo de *scheduling* em um cenário com *slices* na rede de acesso em que os RBGs disponíveis na BS precisam ser distribuídos para os *slices* (*inter-slice scheduling*), e posteriormente cada *slice* distribui os seus RBGs disponíveis para as UEs (*intra-slice scheduling*) de forma a atender os objetivos para o qual o *scheduler* foi projetado.

scheduler.

A Figura 33 apresenta o diagrama de fluxo para as classes do *slice* e UE. A partir de sua análise, podemos verificar que quando o cenário a ser implementado utiliza *slices*, o *scheduler* primeiro aloca os RBG para os *slices*, e, posteriormente, os recursos de cada *slice* são distribuídos para as UEs associadas a ele. A UE cria um *buffer* de acordo com as definições da simulação. Em cada passo da simulação, ela recebe uma determinada quantidade de pacotes do gerador de tráfego e os aloca no *buffer*. Vale a pena ressaltar que a quantidade de pacotes a serem retirados do *buffer* para então serem enviados através da rede são determinados pela quantidade de RBGs alocados para a UE, pelo *scheduler* e de acordo com os valores de eficiência espectral determinados pelo gerador de canal do simulador.

É possível definir um vetor de alocação de RBGs para cada *slice* como

$$\boldsymbol{R}_{n} = [R_{1}(n), R_{2}(n), \dots, R_{S}(n)],$$
(101)

em que $R_s(n)$ representa o número de RBGs alocados para a fatia de rede s no passo n. Para um cenário sem *slices*, podemos considerar S = 1, de forma que todos os recursos da BS vão para o único *slice* no sistema. O processo de alocação de RBGs obedece a

$$\sum_{s=1}^{S} R_s(n) \le R,\tag{102}$$

em que a soma de todos os RBGs distribuídos entre os *slices* é menor ou igual ao número de RBGs disponíveis, R, na BS. Portanto, a principal função do *scheduler* é definir $R_s(n)$ para





Figura 32: Diagrama de fluxo do funcionamento da classe principal *Communication env.* e da classe BS. O *scheduler*, recebe as informações da rede e baseado nelas envia a decisão de alocação de RBGs que será utilizada pelo simulador para gerar as dinâmicas da simulação.

cada *slice* s no passo n e posteriormente a quantidade de RBGs dos *slices* que serão alocados para cada uma das UEs, de forma a alcançar os objetivos determinados para o agente de *Radio Resource Scheduling* (RRS).

O simulador permite a implementação de diferentes métricas de rede para avaliação dos cenários simulados, mas já possui algumas métricas implementadas por padrão para a avaliação da performance dos agentes. Por exemplo, taxa de transferência servida pela rede a cada UE $r_{\rm u}(n)$ é a máxima taxa em bits por passo n que uma UE u pode obter, considerando o número

Brasil 65



Figura 33: Diagrama de fluxo do funcionamento das classes Slice e UE. O *scheduler* primeiro aloca os RBG para os *slices*, e, posteriormente, os recursos de cada *slice* são distribuídos para as UEs associadas a ele.

de RBGs alocados para ela e a sua eficiência espectral. Ou seja,

$$r_{\rm u}(n) = \left\lfloor \frac{(R_s^u(n)/R) \text{BWSE}_u(n)}{\text{PS}} \right\rfloor \text{PS},$$
(103)

em que $R_s^u(n)$ representa o número de RBGs alocados para a UE u pelo *instra-slice scheduling*, $SE_u(n)$ a eficiência espectral para a UE u no passo n, e PS o tamanho dos pacotes que são a unidade de transferência miníma da rede.

Dessa forma, a taxa de transferência efetiva, $r_{\rm u}^{\rm eff}(n)$, representa os dados enviados pela rede que estavam disponíveis no *buffer*. e é definido como

$$r_{\rm u}^{\rm eff}(n) = \min(r_{\rm u}(n), b_{\rm u}(n)),$$
 (104)

em que $b_{\mathbf{u}}(n)$ representa a quantidade de dados disponível no *buffer* da UE u no passo de simulação n. Consequentemente, a taxa de transferência efetiva é sempre $r_{\mathbf{u}}^{\text{eff}}(n) \leq r_{\mathbf{u}}(n)$ com $r_{\mathbf{u}}^{\text{eff}}(n) = r_{\mathbf{u}}(n)$ quando $b_{\mathbf{u}}(n) \geq r_{\mathbf{u}}(n)$.

Logo, a taxa de ocupação do buffer $b_{\rm u}^{\rm occ}(n)$ é definida como

$$b_{\rm u}^{\rm occ}(n) = \frac{b_{\rm u}(n)}{b_{\rm max}},\tag{105}$$



onde b_{\max} é a maxima capacidade do *buffer* da UE. Pacotes são descartados toda vez que o *buffer* está cheio ou que a latência de um pacote ultrapassa a latência máxima permitida pelo *buffer* l_{\max} . Então, esses pacotes são contabilizados no número de pacotes perdidos, $d_u(n)$, que representam a quantidade em bits de pacotes perdidos por passo n.

Por sua vez, $r_{u}^{rqt}(n)$ é a taxa de dados requisitada por cada UE para ser atendido pela rede móvel no passo n. Ela depende da implementação realizada para gerar os tráfegos na simulação.

Já a latência média do *buffer*, $\ell_u(n)$, representa o tempo médio que cada pacote esperou antes de ser enviado ou perdido, e é definido como

$$\ell_{\rm u}(n) = \frac{\sum_{i=0}^{l_{\rm max}} i \boldsymbol{l}_{\rm n}^{\rm u}(i)}{\sum_{i=0}^{l_{\rm max}} \boldsymbol{l}_{\rm n}^{\rm u}(i)},\tag{106}$$

sendo que $l_n^u = [l_0, l_1, \ldots, l_{l_{\max}}]$ é um vetor de dimensão $l_{\max} + 1$ representando a latência de pacotes no *buffer* da UE *u* no passo *n*. $l_{l_{\max}}$ representa o número de pacotes que esperaram um total de l_{\max} Transmission Time Intervals (TTIs) no *buffer*.

Vale a pena mencionar que o processo de funcionamento das classes de ambiente de comunicação, BS, *slice* e UE são estáticos. A variabilidade de um cenário de simulação é garantida pelos geradores de tráfego, canal, mobilidade e o agente de *scheduler*. Ou seja, é possível permitir que o usuário defina funções próprias para cada um dos geradores e *scheduler*, desde que sejam obedecidas as interfaces definidas para cada um. Por exemplo, ferramentas externas de geração de canais como o QuaDRiGa [155] e Wireless InSite [156], de mobilidade como o software SUMO [157] podem ser facilmente integradas ao simulador.

O funcionamento do gerenciador de mobilidade é definido pelo usuário obedecendo a interface do simulador que espera os valores de mobilidade mob $(n) = (x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_U, y_U, z_U)$, onde x_u, y_u, z_u representa as coordenadas tridimensionais da UE no cenário. O gerador de canal implementado pelo usuário pode utilizar as informações retornadas pelo gerador de mobilidade para o calculo dos canais das UEs. O gerador de canais fornece ao simulador uma matriz tridimensional com valores de eficiência espectrais com dimensões definidas como $B \times U \times R$ especificando o valor da eficiência espectral na BS b, RBG r para o UE u. Por fim, o gerador de tráfego é um vetor $tr(n) = [r_1^{rqt}(n), r_2^{rqt}(n), \ldots, r_U^{rqt}(n)],$ onde $r_u^{rqt}(n)$ representa a taxa de dados que a UE u requisita da rede de acordo com o que foi definido pelo gerador de tráfego.

Como conclusão, podemos citar que o ambiente de simulação desenvolvido permite a implementação de agentes de scheduler fornecidos pelo usuário que serão responsáveis por definir quais RBGs serão alocados para cada UE, ou então slices e posteriormente para as UEs em caso de cenários com slices. O simulador fornece por padrão informações sobre as UEs presentes na rede móvel, como latência, perda de pacote e etc, mas outras métricas podem ser criadas na implementação do agente de scheduler. Além disso, o ambiente de simulação possui um padrão de formatação para a alocação dos RBGs definido por uma matriz binária alloc(n) de dimensão $B \times U \times R$, em que o valor 1 representa que o RBG r foi alocado para a UE u na BS b. Caso a rede possua slices, o próprio agente estabelece os processos de inter-slice e intra-slice scheduler e gera uma decisão de alocação que seja compatível com o simulador. Por fim, as características do cenário como o número máximo de BSs, UEs e slices, quantidade de RBGs disponíveis por BS são definidas em um arquivo de configuração da simulação, permitindo a mudança de cenários de forma mais dinâmica. Alguns exemplos de agente de scheduler foram implementados com geradores de canais, tráfego e mobilidade simples para ilustrar o funcionamento do simulador em repositório público no Github.

 $https://github.com/lasseufpa/sixg_scheduling_example$



11.2 Exemplo de *scheduler* Utilizando Round-Robin e Aprendizado por Reforço

Um Jupyter notebook foi criado para demonstrar as funções da plataforma de simulação durante o Minicurso "Inteligência Artificial e Aprendizado de Máquina Aplicados a Redes 5G e 6G" apresentado no evento XL SBrT. A Tabela 8 mostra os parâmetros da simulação utilizados no arquivo de configuração. Há apenas uma BS, um único slice (equivalente a um cenário sem slices) e 4 UEss associados. Cada UE possui um buffer que suporta até 5 pacotes e uma latência máxima por pacote de 5 ms. Por sua vez, a BS possui 8 RBGs que serão distribuídos dentre as 4 UEs presentes no sistema de acordo com as decisões do scheduler. A escolha de uma largura de banda de apenas 8 Hz e tamanhos de pacotes de apenas 1 bit torna o exemplo mais simples. Substituindo os valores escolhidos para os parâmetros, a equação (103) é reduzida para

$$r_{\rm u}(n) = \left\lfloor \frac{R_s^u(n)\mathsf{SE}_u(n)}{\mathsf{PS}} \right\rfloor \mathsf{PS},\tag{107}$$

em que a taxa de transferência servida pela rede para cada UE u passa a ser definida principalmente em torno do número de RBGs atribuídos pelo *scheduler* e a eficiência espectral da UE.

Parâmetros:	Valores:	
Máx. Número de BSs (B)	1	
Máx. Número de UEs (U)	4	
Max. Número de $Slices$ (S)	1	
Tamanho do <i>buffer</i> (b_{\max})	5 Pacotes	
Tamanho dos pacotes (PS)	1 bit	
Máxima latência $buffer$ (l_{max})	5 ms	
Número de RBGs disponíveis (R)	8	
Largura de banda (BW)	8 Hz	

Tabela 8: Parâmetros da simulação definidos no arquivo de configuração utilizado para obtenção dos resultados.

Para interagir com o ambiente simulado, foram implementados geradores de canal, tráfego de rede e mobilidade simples. Com o intuito de facilitar a compreensão do funcionamento do simulador, definimos o gerador de canal com eficiência espectral constante de 1 bit/Hz para cada UE independente de qual RBG foi alocado. Dessa forma, a Equação 107 com PS = 1 se torna $r_u(n) = R_s^u(n)$, ou seja, para cada RBG alocado é possível enviar um pacote de dados. Por exemplo, caso o *scheduler* aloque 2 RBGs para a UE u, ela será capaz de enviar 2 pacotes através da rede móvel, tornando a interpretação do sistema de simulação mais simples. Além disso, o gerador de mobilidade retorna posições constantes para cada UE e as informações de mobilidade não são utilizadas pelo gerador de canal, pois ele retorna valores contantes de eficiência espectral.

Já o gerador de tráfego define a quantidade de pacotes que cada UE recebe para ser armazenado em seu *buffer* em cada passo n da simulação. A implementação do gerador de tráfego define que as UEs 0 e 1 recebem uma quantidade de pacotes definida por uma distribuição uniforme $r_{\rm U}^{\rm rqt} = \mathcal{U}(2,4)$, enquanto que as UEs 2 e 3 seguem uma distribuição uniforme $r_{\rm U}^{\rm rqt} = \mathcal{U}(0,2)$. Logo, as UEs 0 e 1 requisitam da rede um fluxo de tráfego médio maior que as

 $https://github.com/aldebaro/ai6g/blob/main/09_rl_based_radio_resource_allocation.ipynb$



UEs 2 e 3. A Figura 34 demonstra a quantidade de pacotes recebidos no *buffer* de cada UE gerados de acordo com as distribuições explicadas anteriormente.



Figura 34: Quantidade de pacotes recebidos em cada passo n para cada UE. Os pacotes recebidos são armazenados nos *buffers* das UEs e são enviados de acordo com a alocação do *scheduler*.

Para avaliação do papel dos *schedulers*, foram implementados dois agentes de *scheduler*. O primeiro utiliza o método de Round-Robin [153], em que os RBGs são distribuídos igualmente dentre as UEs, logo em cada passo o agente de aloca 2 RBGs para cada UE, e consequentemente 2 pacotes são retirados do *buffer* para serem enviados pela rede.

O segundo agente utiliza da técnica de Soft Actor-Critic (SAC) Reinforcement Learning (RL) [158], com espaço de observação composto pelo número de pacotes perdidos e a taxa de ocupação de buffers em cada UE, totalizando 8 variáveis. O espaço de ação do agente de aprendizado supervisionado é composto por 8 variáveis de saída (uma para cada RBG) com valores no intervalo [-1, 1], de forma que quando a saída apresenta um valor entre [-1, -0.5[o RBG é alocado para a UE 0, de [-0.5, 0[para UE 1, de [0, 0.5[para a UE 2, e de [0.5, 1] para a UE 3. A função de recompensa que guia o aprendizado do agente RL, treinado com 10 mil passos, é dada em função da perda de pacotes em cada um deles, $\mathsf{RW}(n) = -\sum_{u=1}^{U} d_u(n)$. Portanto, o agente aprenderá a evitar a perda de pacotes.

Posteriormente, ambos os agentes (Round-Robin e SAC) são testados em 10 passos. Como pode ser visto na Figura 35, o agente de RL obteve um resultado superior ao agente de Round-Robin, conseguindo perder menos pacotes devido ao seu treinamento focado nessa métrica. Como esperado, o agente de Round-Robin perdeu mais pacotes por não conseguir diferenciar o tratamento para cada UE, não levando em consideração os diferentes tráfegos de rede requisitados. Apesar disso, o desempenho do agente de RL poderia ser superior normalizando ou mesmo selecionando melhor as variáveis do espaço de observação.

11.3 Conclusão

De forma geral, podemos citar que a plataforma de simulação de alocação de recursos de rádio demonstrou grande flexibilidade para a implementação de diferentes geradores de canal,





Figura 35: Recompensas acumuladas obtidas por cada agente de *scheduler* ao longo do testas usando a simulação. As recompensas são definidas em função da perda de pacotes, onde o melhor agente é definido como o agente que faz o sistema perder menos pacotes.

mobilidade, tráfego e *schedulers*, incluindo agentes de RL. Apesar da simplicidade do cenário de exemplo demonstrado nos resultados, cenários mais complexos podem ser implementados no simulador para testar o desempenho de diferentes *schedulers*. Em [159], por exemplo, o simulador foi utilizado para a comparação de *schedulers* utilizando métodos tabulares, aprendizado por reforço e programação inteirapara alocação de recursos de rádio em cenários com *slices* de rede *Enhanced Mobile Broadband* (eMBB), URLLC e *best effort*.



12 Comunicação Semântica

12.1 Introdução

Uma análise rápida da evolução dos padrões de sistemas de comunicação sem fio, desde a primeira geração, introduzida na década de 1970, até os estudos iniciais dos sistemas 6G, mostra uma característica comum nas passagens de uma geração para a próxima: a busca por taxas de transmissão cada vez mais altas, disponibilidade de serviço cada vez maior, atraso baixo e alta confiabilidade. Essa busca por redes de alto desempenho tem sido motivada pelo uso cada vez mais intenso dessas redes nas mais variadas atividades humanas. O atendimento a requisitos mais rigorosos exige o uso mais eficiente dos recursos de rádio.

Esse cenário de uso cada vez mais intenso dos recursos de rádio tem motivado uma revisão da forma de projetar e operar um sistema de comunicação, com o contexto da comunicação e o objetivo da transmissão de uma mensagem adquirindo um papel no relevante no processo de comunicação. Essa revisão tem como base os três níveis estabelecidos por Weaver para o problema da comunicação [160] :

- ▷ Nível Técnico: Trata da precisão na transmissão dos símbolos da mensagem. Nesse nível encontra-se a Teoria Clássica de Informação, iniciada com os estudos e proposições de Shannon.
- Nível Semântico: Refere-se à precisão da transmissão da informação transportada pela mensagem.
- Nível da Efetividade: Trata da eficácia da transmissão da mensagem no processo de mudança do comportamento do destino.

O **Nível Técnico** é regido pela chamada Teoria de Comunicação Clássica, que trata da transmissão dos símbolos de uma mensagem de forma rápida e confiável, sem, no entanto, se preocupar com o significado da mensagem ou com o motivo da transmissão daquela mensagem.

Por outro lado, nos **Níveis Semântico** e **da Efetividade**, o processo de comunicação leva em conta o conteúdo da mensagem a ser transmitida. O enfoque dado nesses dois níveis está relacionado ao conceito de informação proposto por Dretske: aquilo capaz de produzir conhecimento [161] [162]. A informação, quando recebida pelo destinatário, pode (*i*) aumentar o grau de conhecimento do destinatário a respeito de uma situação, ou (*ii*) alterar seu estado ou comportamento de alguma forma. É nesse contexto que surge o conceito de *comunicação semântica*, que se estabelece no **Nível Semântico**. Em um sistema de comunicação semântica, busca-se transmitir apenas a *semântica das mensagens*, ou seja, aquilo que interessa ao receptor para aumentar seu conhecimento ou alterar o seu estado. O transmissor deverá possuir, portanto, um *codificador semântico* para a extração da semântica das mensagens. No receptor, um *decodificador semântico* recupera o significado da mensagem transmitida. Note-se, portanto, que o advento da comunicação semântica traz uma mudança essencial na forma de comunicação, passando da forma *transmitir antes de entender*, para *entender antes de transmitir*.

Os processos de codificação e decodificação semântica são baseados em conhecimentos *locais* prévios existentes na fonte e no destino. Para que o objetivo da comunicação sob o ponto de vista semântico seja atingido, isto é, para que a transmissão do significado da mensagem seja bem sucedida, é necessário que esses conhecimentos prévios sejam comuns e compartilhados.

Capítulo escrito por Luiz Fernando Gontijo e Paulo Cardieri.



Muitos autores dessa área de comunicação semântica tem observado que o problema do **Nível Semântico** foi desconsiderado por Shannon em seus trabalhos quando afirmou [163]: "Frequently the messages have meaning; that is, they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design."

No terceiro nível, o **da Efetividade**, o foco é garantir que o objetivo da transmissão de uma mensagem seja atingido no destino, o que leva à chamada *comunicação orientada a tarefas ou objetivos*. Esse tipo de comunicação envolve extrair da mensagem bruta a informação relevante ao objetivo ou tarefa pretendidos no destino [164]. Como no caso da comunicação semântica, na comunicação orientada a tarefa os dois lados do enlace precisam manter as suas bases de conhecimento compartilhadas e atualizadas, para evitar erros na execução da tarefa alvo.

A Figura 36 exemplifica a diferença entre as formas de comunicação convencional, semântica, e aquela baseada em tarefa. No cenário considerado na figura, se deseja alertar o destino (p.ex., um carro com algum grau de autonomia) sobre a existência de uma placa de trânsito PARE em um cruzamento à frente, usando uma foto do cruzamento. Na comunicação convencional, a



Figura 36: Diferenças entre os sistemas de comunicação convencional, semântico e baseado em tarefa. No sistema convencional, a foto é transmitida integralmente, sem qualquer atenção ao seu conteúdo; no sistema semântico, o transmissor envia apenas a parte da foto que interessa ao destino; na comunicação baseada em tarefa, o transmissor enviar o comando que causará a mudança do estado do destino.

foto capturada do local é transmitida ao destino sem qualquer atenção ao conteúdo da imagem, o que certamente consumirá muitos recursos do sistema de comunicação. Por outro lado, na



comunicação semântica o objetivo é alertar o destino sobre a presença da placa, de forma que o codificador semântico buscará na foto a placa, transmitindo ao destino apenas a parcela da foto pertinente ao objetivo da comunicação. Por fim, na comunicação orientada a tarefa, o objetivo é fazer com que o carro pare quando a foto contiver a imagem de uma placa PARE. Desta forma, a mensagem transmitida e recuperada no destino consiste apenas na orientação para frear o carro, o que consome ainda menos recursos de comunicação. Note-se que as formas de comunicação semântica e por tarefas exigem que o transmissor e o destino compartilhem uma "linguagem", o que permite que o transmissor sabia o que deve ser buscado na foto e que o destino saiba o significado da mensagem transmitida. Portanto, o objetivo de um sistema de comunicação semântica é fornecer a parte útil da informação ao ponto exato de atuação (destino) e no tempo correto [165], o que exige a redefinição dos processos de geração e transmissão da informação. Note-se que essa definição de um sistema de comunicação não está alinhada com o objetivo atual dos sistemas de comunicação, qual seja, transmitir o mais rápido possível e com a maior precisão possível toda a informação gerada.

Apesar de a retomada pelo interesse em comunicação semântica ser recente, já se pode observar o surgimento de focos de pesquisa dedicados a questões centrais. Segundo Wheeler e Natarajan em [166], os principais focos das pesquisas em comunicação semântica são:

- ▷ Comunicação Semântica Clássica,
- ▷ Comunicação Semântica baseada em base de conhecimento,
- ▷ Comunicação Semântica baseada em Machine Learning,
- ▷ Comunicação Semântica baseada na importância da informação.

Esses focos serão discutidos nas seções a seguir.

12.2 Comunicação Semântica Clássica

Os primeiros estudos sobre sistemas de comunicação semântica trataram de aspectos mais teóricos, buscando uma formulação semelhante àquela proposta por Shannon para as questões relacionadas aos problemas do Nível Técnico. Um dos primeiros trabalhos foi de R. Carnap e Y. Bar-Hillel [167], que propuseram uma estrutura para comunicação semântica baseada em um modelo de linguagem ideal, com números finitos de *nomes* e *adjetivos*. A quantidade de informação semântica de uma palavra é então definida como uma função do número de sentenças que essa palavra pode implicar.

Mais recentemente, Bao et al. [168] discutiram formas de medir a quantidade de informação semântica em uma mensagem. Nesse sentido, eles argumentam que a entropia de uma mensagem depende da chamada *probabilidade lógica*, que está relacionada à probabilidade de ocorrência do cenário em que aquela mensagem é verdadeira. Usando o exemplo apresentado por Bao et al. em [168], a frase "Rex não é um tiranossauro" é menos surpreendente hoje em dia do que a frase "Rex não é um cão", pois tiranossauros são menos comuns hoje em dia. Portanto, não é surpresa que Rex não seja um tiranossauro. Dito de outra forma, o "mundo" em que a afirmação "Rex não é um tiranossauro" é verdade é mais provável.

Diversos outros trabalhos buscaram também estabelecer as bases para uma teoria de comunicação semântica (ver, por exemplo Floridi [169], D'Alfonso [170], Basu et al. [171]), mas parece haver um entendimento na comunidade científica de que há ainda inúmeras questões



abertas no campo de comunicação semântica [168]: (i) Como o uso da semântica da informação pode ajudar na compressão da informação e na confiabilidade da transmissão? (ii) Como deve operar a codificação semântico e quais são os limites dessa codificação?

12.3 Comunicação Semântica usando Base de Conhecimento

Uma outra frente de estudo é aquela em que a comunicação semântica é baseada no conhecimento compartilhado entre as partes que se comunicam. Exemplos desse conhecimento compartilhado incluem uma linguagem e regras lógicas que permitam que mensagens recebidas pelo destino de forma incompleta ou ruidosa possam ser corrigidas pelo receptor [1]. Essa é uma situação familiar na comunicação entre duas pessoas: conseguimos entender a mensagem emitida pelo nosso interlocutor mesmo quando não compreendemos exatamente todas as palavras ditas. Esse entendimento é baseado na contextualização (isto é, o assunto da conversa) e em algum conhecimento prévio comum entre as duas partes, como, por exemplo, o idioma usado. Nesse sentido, Strinati e Barbarossa propõem em [1] um sistema, mostrado na Figura 37, composto por um codificar semântico, que gera mensagens m contendo a informação semântica que se deseja transmitir. Essas mensagens são codificadas em uma sequência de símbolos x, que é enviada ao destino através de um canal. No destino, a sequência de símbolos y recebida é decodificada sintaticamente, isto é, sem levar em conta o seu significado, resultando na sequência x'. Essa sequência recuperada é, então, interpretada com base no Knowledge Base do destino, denotado por KB_D, resultando na mensagem m'. Erros na interpretação de x' podem decorrer devido a diferenças entre as bases KB_D e KB_S ou a erros de interpretação.



Figura 37: Estrutura típica de uma sistema de comunicação semântica fim-a-fim. Baseada na Fig. 2 de [1].

Portanto, a comunicação semântica como apresentada por Strinati e Barbarossa permite que mensagens sejam transmitidas através do canal com uma menor quantidade de detalhes ou menos protegida contra perturbações e ruído e, ainda assim, serem entendidas no destino, graças ao conhecimento compartilhado entre a fonte e o destino. Tudo isso, no entanto, vem

A base KB_D compartilha conhecimento com a respectiva base na fonte, denotada por KB_S .



às custas de uma maior complexidade na fonte e no destino. Essa situação abre espaço para o uso de técnicas de *deep learning* na implementação de sistemas de comunicação semântica, como investigado por diversos autores, como Qin et al. [172].

12.4 Comunicação Semântica baseada em Machine Learning

As primeiras implementações de sistemas de comunicação ponta-a-ponta com aplicação de técnicas de *deep learning* na camada física datam de 2017, como aquela apresentada por O'Shea et al. [173]. Em sua proposta, os autores introduzem módulos de codificação e decodificação implementados usando *deep learning*, para o envio de símbolos por um canal de comunicação AWGN. O primeiro módulo é composto por camadas de neurônios que possuem a função de representar o símbolo em forma compacta para o seu envio pelo canal de comunicação. O módulo de decodificação recebe o código, passa por camadas de neurônios até a entrega à uma função *softmax*, a qual atribui um valor ao símbolo recebido resultante da probabilidade de pertencer à alguma representação pela base de dados. Nota-se que a proposta de comunicação com base na semântica da mensagem ainda não é abordada nesse trabalho de O'Shea et al., mas o uso de *deep learning* para codificação e decodificação de mensagens já é introduzido.

Com o avanço dos algoritmos e de novas arquiteturas baseadas em *deep learning*, como as redes neurais convolucionais [174], as redes neurais recorrentes [175] e os mecanismos de atenção [176], o contexto de cada símbolo enviado em uma mensagem passou a ser considerado para a sua codificação e posterior decodificação. Dessa forma, a obtenção da informação semântica pode ser alcançada em outras implementações.

As aplicações e implementações de comunicação semântica vem sendo dividida na literatura em três áreas: transmissão de texto, de sinais de áudio e de imagens [162]. Aplicações em todas essas áreas foram impulsionadas de forma expressiva com o desenvolvimento de algoritmos baseados em deep learning. Tomando como foco a transmissão de mensagens de texto, podemos considerar como exemplo a implementação de Xie et al. apresentada em [177]. Assim como em [173], Xie et al. propõem módulos de codificação e decodificação, mas usando a arquitetura de Transformers, que é baseada em mecanismo de auto-atenção com múltiplas cabeças [178]. A proposta de sistema de comunicação ponta-a-ponta apresentada por Xie et al. em [177] divide a etapa de codificação em duas partes. A primeira etapa consiste na extração da informação semântica dos símbolos por meio do codificador da arquitetura de Transformers, enquanto que a segunda consiste na codificação de canal com uso de camadas de neurônios, visando a sua transmissão pelo canal. A etapa de decodificação também é dividida em duas partes: primeiro, camadas de neurônios são usadas para a detecção de símbolos recebidos; em seguida, um decodificador da arquitetura de Transformers correspondente é usado para a estimação do texto enviado. O sistema proposto foi testado em diferentes cenários de comunicação (modelos de canal), e taxa de acerto de recebimento das mensagens, avaliada pela métrica BLEU (Bilingual *Evaluation Understudy*), superou os sistemas convencionais.

12.5 Comunicação Semântica baseada na Importância da Informação

Essa quarta frente de pesquisa em comunicação semântica está relacionada aos problemas do Nível de Efetividade proposto por Weaver [160]. O tratamento desse tipo de problema requer a caracterização da *importância* de uma mensagem no cumprimento de um objetivo ou tarefa no destino. Portanto, neste caso, a semântica de uma mensagem está relacionada à sua importância na execução da tarefa pelo destino, e o processo de comunicação é dito *orientado*



ao objetivo ou à tarefa.

Um dos desafios na concepção de um sistema de comunicação orientado ao objetivo ou tarefa é a definição da forma de medir a importância de uma informação. Uysal et al. [165] apresentam três formas, considerando o cenário de um processo alimentado por amostras de uma grandeza física coletadas por sensores: (*i*) **frescor da informação**, ou seja, o quanto a informação é nova, quantificando a utilidade da amostra para a finalidade do processo; (*ii*) **valor de uma amostra**, que quantifica o compromisso entre a importância daquela amostra para o processo fim e o custo de transmitir aquela amostra; (*iii*) **relevância da informação**, que está associada à quantidade de "novidade" que uma amostra carrega, comparada com as amostras anteriores. Portanto, Uysal et al. definem "semântica da informação" não como o significado da informação, mas sim como a utilidade daquela mensagem para o contexto considerado.

Os conceitos de Age of Information e Value of Information, relacionados, respectivamente, aos casos (i) e (ii) citados acima, são discutidos a seguir.

Age of Information

O conceito de Age of Information (AoI), recentemente proposto no contexto de redes de sensores e redes para sistemas de controles, é considerado como o primeiro passo no desenvolvido de sistemas de comunicação que levam em conta o significado e a importância da mensagem. Aqe of Information é uma métrica fim-a-fim que quantifica a idade da última informação recebida no destino a respeito do processo monitorado. Como exemplo de aplicação dessa métrica, considere a monitoração do estado de um processo industrial por meio de amostragem de alguma grandeza física, com o envio dessas amostras a uma central controladora. Suponha que para manter o funcionamento adequado do processo, a controladora precise ser mantida informada sobre o estado do processo. A AoI da controladora indica o tempo passado desde a última atualização sobre o estado do processo. Claramente, a AoI depende não apenas do atraso na rede, mas também da taxa de amostragem do processo monitorada. Assim, quando a AoI da controladora se aproxima do valor máximo tolerável, maior prioridade deve ser dada à transmissão da mensagem com a nova atualização do processo. Ou, quando a transmissão da atualização falha, o sistema também precisa se adaptar para garantir a entrega rápida da próxima atualização. Portanto, como mencionado por Kountouris e Pappas, a AoI pode ser vista com um substituto de semântica [179].

Value of Information

Outro conceito que tem sido associado à comunicação semântica é o de Value of Information (VoI). Os primeiros trabalhos sobre esse conceito surgiram logo após os estudos de Shannon sobre teoria da informação, visando incluir a ideia de valor (p.ex., econômico) de uma mensagem em um processo de decisão [180]. Recentemente, o conceito de value of information de uma amostra de uma grandeza foi proposto, definido com a diferença entre o benefício de ter aquela amostra e o custo de sua transmissão [181].

12.6 Conclusão

As aplicações previstas para as redes 6G exigirão o uso mais eficiente dos recursos rádio, o que pode ser conseguido por uma mudança de foco na operação das redes: ao invés de prover altas taxas de transmissão e alta confiabilidade, os sistemas de comunicação devem ser projetados



para garantir a efetividade da comunicação, ou seja, o cumprimento do objetivo da transmissão de uma mensagem. Uma maior efetividade da comunicação requer a contextualização da comunicação, tanto no transmissor como no receptor. Ou seja, os dois lados do enlace precisam conhecer os motivos que levaram a transmissão de uma mensagem. Essa contextualização permite que apenas "fragmentos" da mensagem sejam de fato transmitidos, dado que o receptor conhece o contexto e poderá inferir a mensagem completa. A transmissão de fragmentos da mensagem certamente exige menos recursos rádio para a transmissão. Essas são as bases da comunicação semântica, que vem sendo considerada uma das técnicas promissoras para uso nas redes 6G.



13 Conclusão

Os casos de uso esperados para as redes 6G exigirão das redes sem fio requisitos muito mais rigorosos do que aqueles oferecidos pelas redes de gerações anteriores, como altas taxas de dados, baixíssimo atraso, e provisão de serviço a altas densidades de terminais. O atendimento a tais requisitos somente será conseguido com o uso de tecnologias inovadoras que explorem os recursos rádio com uma eficiência não ainda vista nas gerações anteriores. Nesse sentido, inúmeras técnicas estão sendo investigadas pela comunidade científica como candidatas a compor as redes 6G.

O presente relatório trouxe resultados preliminares de investigações sobre técnicas inovadoras que poderão ser usadas na redes 6G. Foram estudadas técnicas relacionadas à camada física, como mecanismos de detecção de sinais e de seleção de feixes em sistemas de múltiplas antenas, e técnicas de formação de feixes em superfícies refletoras inteligentes. Foram também estudadas questões relacionadas à camada de enlace, como o acesso ao meio para aplicações de IoT, o *handover* para tratar a mobilidade de terminais, e a modelagem matemática de redes do tipo *cell-free*. Ainda no contexto da camada de enlace, um simulador para o estudo de alocação de recursos foi proposto. Por fim, foram apresentados os conceitos básicos da comunicação semântica, que aparece como uma estratégia de comunicação promissora para as redes 6G.

Um aspecto comum entre muitas das técnicas estudadas nesse relatório é o uso de inteligência artificial na implementação dessas técnicas. De fato, o alto desempenho esperado das redes 6G não será atingido apenas com o refinamento das técnicas de comunicação que já vem sendo usadas na geração atual das redes sem fio. Serão necessários saltos significativos na concepção das estratégias de comunicação e de uso dos recursos rádio. Estudos tem mostrado que esses saltos poderão ser dados por meio da provisão de inteligência às entidades que compõem a rede de comunicação e através do uso massivo de dados sobre o comportamento dos usuários e do meio de comunicação. Dessa forma, espera-se que as técnicas de inteligência artificial terão papel central nas redes 6G, como já pode ser observado nesse relatório.



Referências

- E. Calvanese Strinati e S. Barbarossa, "6G networks: Beyond shannon towards semantic and goal-oriented communications," *Computer Networks*, v. 190, p. 107930, 2021. [Online]. Disponível em: https://www.sciencedirect.com/science/article/pii/S1389128621000773
- [2] 3GPP, "5G; NR; Radio Resource Control (RRC); Protocol specification," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.331, 10 2018, Version 15.3.0. [Online]. Disponível em: https://www.etsi.org/deliver/etsi_ts/138300_138399/ 138331/15.03.00_60/ts_138331v150300p.pdf
- [3] D. G. Silva, F. A. P. Figueiredo, G. L. Tejerina, J. S. Ferreira, L. L. Mendes, M. S. P. Facina, P. Cardieri, R. D. Souza, e V. Souto, "Estado da arte da camada física para redes de acesso 6G," Projeto Brasil 6G, Tech. Rep., 2021.
- [4] V. Angeline Beulah e S. Markkandan, "Performance analysis of precoding techniques for Massive MU-MIMO systems," in 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015, p. 1–5.
- [5] S. Domouchtsidis, C. G. Tsinos, S. Chatzinotas, e B. Ottersten, "Symbol-Level Precoding for Low Complexity Transmitter Architectures in Large-Scale Antenna Array Systems," *IEEE Transactions on Wireless Communications*, v. 18, n. 2, p. 852–863, 2019.
- [6] J. Ketonen, J. Karjalainen, M. Juntti, e T. Hänninen, "MIMO detection in single carrier systems," in 2011 19th European Signal Processing Conference, 2011, p. 654–658.
- [7] M. A. Albreem, M. Juntti, e S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Communications Surveys & Tutorials*, v. 21, n. 4, p. 3109–3132, 2019.
- [8] D. Gaspar, L. Mendes, e T. Pimenta, "Sphere detector with low complexity euclidean distance computation based on affine transform modulation," *Electronics Letters*, v. 58, n. 20, p. 779–782, 2022. [Online]. Disponível em: https: //ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ell2.12594
- [9] R. Ma, J. Cao, T. Wang, e X. Lai, "Affine transformation based hierarchical extreme learning machine," in 2020 IEEE International Symposium on Circuits and Systems (IS-CAS), 2020, p. 1–5.
- [10] S. Yang e L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Communications Surveys & Tutorials*, v. 17, n. 4, p. 1941–1988, 2015.
- [11] D. Wübben, D. Seethaler, J. Jaldén, e G. Matz, "Lattice reduction," *IEEE Signal Processing Magazine*, v. 28, n. 3, p. 70–91, 2011.
- [12] D. Pham, K. Pattipati, P. Willett, e J. Luo, "A generalized probabilistic data association detector for multiple antenna systems," *IEEE Communications Letters*, v. 8, n. 4, p. 205– 207, 2004.
- [13] P. H. C. De Souza e L. L. Mendes, "Lattice reduction aided probability data association detector for MIMO systems," *IEEE Communications Letters*, p. 1–1, 2022.



- [14] S. Yang, T. Lv, R. G. Maunder, e L. Hanzo, "From Nominal to True A Posteriori Probabilities: An Exact Bayesian Theorem Based Probabilistic Data Association Approach for Iterative MIMO Detection and Decoding," *IEEE Transactions on Communications*, v. 61, n. 7, p. 2782–2793, 2013.
- [15] D. Wubben, R. Bohnke, V. Kuhn, e K.-D. Kammeyer, "MMSE-based Lattice-reduction for Near-ML Detection of MIMO Systems," in *ITG Workshop on Smart Antennas (IEEE Cat. No.04EX802)*, 2004, p. 106–113.
- [16] M. Varanasi, "Decision feedback multiuser detection: a systematic approach," *IEEE Transactions on Information Theory*, v. 45, n. 1, p. 219–240, 1999.
- [17] L. Mendes e et al, "Enhanced remote areas communications: The missing scenario for 5G and beyond 5G networks," *IEEE Access*, v. 8, p. 219859–219880, 2020.
- [18] M. Mello, L. Mendes, e T. Barbosa, "Spectrum efficient GFDM based on faster than Nyquist signaling," *JCIS*, v. 35, n. 1, p. 349–356, Dec 2020.
- [19] M. Matthé, L. Mendes, I. Gaspar, N. Michailow, D. Zhang, e G. Fettweis, "Precoded GFDM transceiver with low complexity time domain processing," *EURASIP Journal on Wireless Communications and Networking*, v. 2016, n. 1, p. 138, 2016.
- [20] U. Fincke e M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Mathematics of Computation*, v. 44, n. 170, p. 463–471, 1985. [Online]. Disponível em: http://www.jstor.org/stable/2007966
- [21] B. Hassibi e H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Transactions on Signal Processing*, v. 53, n. 8, p. 2806–2818, Aug 2005.
- [22] N. J. Higham, "Cholesky factorization," Wiley Interdisciplinary Reviews: Computational Statistics, v. 1, p. 251 – 254, 09 2009.
- [23] B. Hassibi e H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Transactions on Signal Processing*, v. 53, n. 8, p. 2806–2818, Aug 2005.
- [24] E. Bedeer, H. Yanikomeroglu, e M. H. Ahmed, "Reduced complexity optimal detection of binary faster-than-Nyquist signaling," in *International Conference on Communications* (ICC), May 2017, p. 1–6.
- [25] E. Bedeer, M. H. Ahmed, e H. Yanikomeroglu, "A very low complexity successive symbolby-symbol sequence estimator for faster-than-Nyquist signaling," *IEEE Access*, v. 5, p. 7414–7422, 2017.
- [26] S. Sugiura, "Frequency-domain equalization of faster-than-Nyquist signaling," IEEE Wireless Communications Letters, v. 2, n. 5, p. 555–558, 2013.
- [27] Özlem Tugfe Demir, E. Björnson, e L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," Foundations and Trends[®] in Signal Processing, v. 14, n. 3-4, p. 162–472, 2021.
- [28] M. Haenggi, *Stochastic geometry for wireless networks*. Cambridge University Press, 2012.



- [29] —, "On distances in uniformly random networks," *IEEE Transactions on Information Theory*, v. 51, n. 10, p. 3584–3586, 2005.
- [30] F. Baccelli, B. Błaszczyszyn et al., "Stochastic geometry and wireless networks," Foundations and Trends (R) in Networking, v. 4, n. 1–2, p. 1–312, 2010.
- [31] Ö. Özdoğan e E. Björnson, "Deep learning-based phase reconfiguration for intelligent reflecting surfaces," in 2020 54th Asilomar Conference on Signals, Systems, and Computers, 2020, p. 707–711.
- [32] Q. Wu e R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, p. 7830–7833.
- [33] Q. Wu e R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, v. 58, n. 1, p. 106–112, 2019.
- [34] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, e C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions* on Wireless Communications, v. 18, n. 8, p. 4157–4170, 2019.
- [35] M. Di Renzo, A. Zappone, M. Debbah, M. S. Alouini, C. Yuen, J. de Rosny, e S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, v. 38, n. 11, p. 2450–2525, 2020.
- [36] V. D. P. Souto, R. D. Souza, B. F. Uchôa-Filho, A. Li, e Y. Li, "Beamforming optimization for intelligent reflecting surfaces without CSI," *IEEE Wireless Communications Letters*, v. 9, n. 9, p. 1476–1480, 2020.
- [37] Q. Wu e R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in 2018 IEEE Global Communications Conference (GLOBECOM), 2018, p. 1–6.
- [38] —, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, v. 18, n. 11, p. 5394– 5409, 2019.
- [39] Y. Yang, S. Zhang, e R. Zhang, "IRS-Enhanced OFDM: power allocation and passive array optimization," in 2019 IEEE Global Communications Conference (GLOBECOM), 2019, p. 1–6.
- [40] Q. Wu e R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Transactions on Communications*, v. 68, n. 3, p. 1838–1851, 2019.
- [41] S. Abeywickrama, R. Zhang, Q. Wu, e C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Transactions on Communicati*ons, v. 68, n. 9, p. 5849–5863, 2020.



- [42] H. Guo, Y. Liang, J. Chen, e E. G. Larsson, "Weighted sum-rate maximization for intelligent reflecting surface enhanced wireless networks," in 2019 IEEE Global Communications Conference (GLOBECOM), 2019, p. 1–6.
- [43] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, e H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE Journal on Selected Areas in Communications*, v. 38, n. 8, p. 1809–1822, 2020.
- [44] "Further advancements for E-UTRA physical layer aspects (release 9)," 3GPP, Tech. Rep., Mar 2010.
- [45] O. Ozdogan, E. Bjornson, e E. G. Larsson, "Using intelligent reflecting surfaces for rank improvement in MIMO communications," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, p. 9160–9164.
- [46] Y. Cheng, K. H. Li, Y. Liu, K. C. Teh, e H. Vincent Poor, "Downlink and uplink intelligent reflecting surface aided networks: NOMA and OMA," *IEEE Transactions on Wireless Communications*, v. 20, n. 6, p. 3988–4000, 2021.
- [47] M. Zeng, X. Li, G. Li, W. Hao, e O. A. Dobre, "Sum rate maximization for IRS-Assisted Uplink NOMA," *IEEE Communications Letters*, v. 25, n. 1, p. 234–238, 2021.
- [48] Z. Ding e H. Vincent Poor, "A simple design of IRS-NOMA transmission," IEEE Communications Letters, v. 24, n. 5, p. 1119–1123, 2020.
- [49] B. Zheng, Q. Wu, e R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?" *IEEE Communications Letters*, v. 24, n. 4, p. 753–757, 2020.
- [50] J. Zhang, J. Liu, S. Ma, C. Wen, e S. Jin, "Transmitter design for large intelligent surfaceassisted MIMO wireless communication with statistical CSI," in 2020 IEEE International Conference on Communications Workshops (ICC Workshops), 2020, p. 1–5.
- [51] Y. Han, W. Tang, S. Jin, C. Wen, e X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Transactions on Vehicular Technology*, v. 68, n. 8, p. 8238–8242, 2019.
- [52] Z. Li, W. Chen, Q. Wu, K. Wang, e J. Li, "Joint beamforming design and power splitting optimization in IRS-Assisted SWIPT NOMA networks," *IEEE Transactions on Wireless Communications*, p. 1–1, 2021.
- [53] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Transactions on Information Theory*, v. 54, n. 11, p. 5113–5129, 2008.
- [54] K. Werner, M. Jansson, e P. Stoica, "On estimation of covariance matrices with kronecker product structure," *IEEE Transactions on Signal Processing*, v. 56, n. 2, p. 478–491, 2008.
- [55] M. Grant e S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014. [Online]. Disponível em: http://cvxr.com/cvx/



- [56] A. Zappone, M. Di Renzo, F. Shams, X. Qian, e M. Debbah, "Overhead-aware design of reconfigurable intelligent surfaces in smart radio environments," *IEEE Transactions on Wireless Communications*, v. 20, n. 1, p. 126–141, 2021.
- [57] O. L. A. López et al., "Massive wireless energy transfer: Enabling sustainable IoT toward 6G era," *IEEE Internet Things J.*, v. 8, n. 11, p. 8816–8835, 2021.
- [58] N. H. Mahmood *et al.*, "White paper on critical and massive machine type communication towards 6G," *6G Research Visions*, n. 11, 2020, http://jultika.oulu.fi/files/ isbn9789526226781.pdf.
- [59] K. Huang e X. Zhou, "Cutting the last wires for mobile communications by microwave power transfer," *IEEE Commun. Mag.*, v. 53, n. 6, p. 86–93, 2015.
- [60] B. Clerckx *et al.*, "Wireless power transfer for future networks: Signal processing, machine learning, computing, and sensing," *arXiv:2101.04810*, 2021.
- [61] E. Björnson et al., "Massive MIMO networks: Spectral, energy, and hardware efficiency," Found. Trends Signal Process., v. 11, n. 3-4, p. 154–655, 2017.
- [62] E. Boshkovska et al., "Practical non-linear energy harvesting model and resource allocation for SWIPT systems," *IEEE Commun. Lett.*, v. 19, n. 12, p. 2082–2085, 2015.
- [63] L.-G. Tran et al., "RF power harvesting: a review on designing methodologies and applications," Micro Nano Syst. Lett., v. 5, n. 1, p. 1–16, 2017.
- [64] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, e D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016. [Online]. Disponível em: https://arxiv.org/abs/1610.05492
- [65] J. Park, S. Samarakoon, A. Elgabli, J. Kim, M. Bennis, S.-L. Kim, e M. Debbah, "Communication-efficient and distributed learning over wireless networks: Principles and applications," *Proceedings of the IEEE*, v. 109, n. 5, p. 796–819, 2021.
- [66] J. Sun, T. Chen, G. Giannakis, e Z. Yang, "Communication-efficient distributed learning via lazily aggregated quantized gradients," in Advances in Neural Information Processing Systems, Canada, 2019.
- [67] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, e S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Transactions* on Wireless Communications, v. 20, n. 1, p. 269–283, 2021.
- [68] H. H. Yang, Z. Liu, T. Q. S. Quek, e H. V. Poor, "Scheduling policies for federated learning in wireless networks," *IEEE Transactions on Communications*, v. 68, n. 1, p. 317–333, 2020.
- [69] N. Abramson, "The ALOHA system: another alternative for computer communications," in AFIPS '70 (Fall): Proceedings of the November 17-19, 1970, fall joint computer conference. New York, NY, USA: ACM, 1970, p. 281–285.
- [70] M. Berioli, G. Cocco, G. Liva, e A. Munari, "Modern random access protocols," Foundations and Trends (R) in Networking, v. 10, n. 4, p. 317–446, 2016.



- [71] J. Choi e S. R. Pokhrel, "Federated learning with multichannel ALOHA," *IEEE Wireless Communications Letters*, v. 9, n. 4, p. 499–502, 2020.
- [72] J. McNamee e V. Pan, "Chapter 7 bisection and interpolation methods," in Numerical Methods for Roots of Polynomials - Part II, ser. Studies in Computational Mathematics, J. McNamee e V. Pan, Eds. Elsevier, 2013, v. 16, p. 1–138.
- [73] C. She, C. Sun, Z. Gu, Y. Li, C. Yang, H. V. Poor, e B. Vucetic, "A tutorial on ultrareliable and low-latency communications in 6G: Integrating domain knowledge into deep learning," *Proceedings of the IEEE*, v. 109, n. 3, p. 204–246, 2021.
- [74] Tayyab, Muhammad and Gelabert, Xavier and Jäntti, Riku, "A survey on handover management: From LTE to NR," *IEEE Access*, v. 7, p. 118907–118930, 2019.
- [75] M. S. Mollel, A. I. Abubakar, M. Ozturk, S. F. Kaijage, M. Kisangiri, S. Hussain, M. A. Imran, e Q. H. Abbasi, "A survey of machine learning applications to handover management in 5G and beyond," *IEEE Access*, v. 9, p. 45770–45802, 2021.
- [76] L. C. Gimenez, P. H. Michaelsen, K. I. Pedersen, T. E. Kolding, e H. C. Nguyen, "Towards zero data interruption time with enhanced synchronous handover," in 2017 IEEE 85th Vehicular Technology Conference (VTC Spring). IEEE, 2017, p. 1–6.
- [77] 3GPP, "Evaluation and design aspects for NR mobility enhancement," 3rd Generation Partnership Project (3GPP), Technical Document (TDoc) 168852, 11 2016, R2. [Online]. Disponível em: https://www.etsi.org/deliver/etsi_ts/138300_138399/138331/15.03.00_ 60/ts_138331v150300p.pdf
- [78] I. K. Jain, R. Kumar, e S. S. Panwar, "The impact of mobile blockers on millimeter wave cellular systems," *IEEE Journal on Selected Areas in Communications*, v. 37, n. 4, p. 854–868, 2019.
- [79] M. A. Usman, N. Y. Philip, e C. Politis, "5G enabled mobile healthcare for ambulances," in 2019 IEEE Globecom Workshops (GC Wkshps). IEEE, 2019, p. 1–6.
- [80] S. Kang, S. Choi, G. Lee, e S. Bahk, "A dual-connection based handover scheme for ultra-dense millimeter-wave cellular networks," in 2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019, p. 1–6.
- [81] K. Qi, T. Liu, C. Yang, S. Suo, e Y. Huang, "Dual connectivity-aided proactive handover and resource reservation for mobile users," *IEEE Access*, v. 9, p. 36100–36113, 2021.
- [82] C. Lee, H. Cho, S. Song, e J.-M. Chung, "Prediction-based conditional handover for 5G mm-wave networks: A deep-learning approach," *IEEE Vehicular Technology Magazine*, v. 15, n. 1, p. 54–62, 2020.
- [83] H.-S. Park, Y. Lee, T.-J. Kim, B.-C. Kim, e J.-Y. Lee, "ZEUS: Handover algorithm for 5G to achieve zero handover failure," *ETRI Journal*, v. 44, n. 3, p. 361–378, 2022.
- [84] D. Castro-Hernandez e R. Paranjape, "Optimization of handover parameters for LTE/LTE-A in-building systems," *IEEE Transactions on Vehicular Technology*, v. 67, n. 6, p. 5260–5273, 2017.



- [85] Farooq, Muhammad Umar Bin and Manalastas, Marvin and Zaidi, Syed Muhammad Asad and Abu-Dayya, Adnan and Imran, Ali, "Machine Learning Aided Holistic Handover Optimization for Emerging Networks," arXiv preprint arXiv:2202.02851, 2022.
- [86] M. Zarifneshat, P. Roy, e L. Xiao, "Multi-Objective Approach for User Association to Improve Load Balancing and Blockage in Millimeter Wave Cellular Networks," *IEEE Transactions on Mobile Computing*, 2021.
- [87] M. Feng, S. Mao, e T. Jiang, "Dealing with Link Blockage in mmWave Networks: A Combination of D2D Relaying, Multi-beam Reflection, and Handover," *IEEE Transactions on Wireless Communications*, 2022.
- [88] V. Yajnanarayana, H. Rydén, e L. Hévizi, "5G handover using reinforcement learning," in 2020 IEEE 3rd 5G World Forum (5GWF). IEEE, 2020, p. 349–354.
- [89] S. Khosravi, H. Shokri-Ghadikolaei, e M. Petrova, "Learning-based handover in mobile millimeter-wave networks," *IEEE Transactions on Cognitive Communications and Networking*, v. 7, n. 2, p. 663–674, 2020.
- [90] Sun, Yao and Feng, Gang and Qin, Shuang and Liang, Ying-Chang and Yum, Tak-Shing Peter, "The SMART handoff policy for millimeter wave heterogeneous cellular networks," *IEEE Transactions on Mobile Computing*, v. 17, n. 6, p. 1456–1468, 2017.
- [91] L. Sun, J. Hou, e T. Shu, "Optimal handover policy for mmWave cellular networks: A multi-armed bandit approach," in 2019 IEEE Global Communications Conference (GLO-BECOM). IEEE, 2019, p. 1–6.
- [92] —, "Spatial and temporal contextual multi-armed bandit handovers in ultra-dense mmWave cellular networks," *IEEE Transactions on Mobile Computing*, v. 20, n. 12, p. 3423–3438, 2020.
- [93] C. Xu, S. Liu, C. Zhang, Y. Huang, e L. Yang, "Joint user scheduling and beam selection in mmWave networks based on multi-agent reinforcement learning," in 2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM). IEEE, 2020, p. 1–5.
- [94] A. Alkhateeb, I. Beltagy, e S. Alex, "Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff," in 2018 IEEE Global conference on signal and information processing (GlobalSIP). IEEE, 2018, p. 1055–1059.
- [95] Y. Koda, K. Nakashima, K. Yamamoto, T. Nishio, e M. Morikura, "Handover management for mmwave networks with proactive performance prediction using camera images and deep reinforcement learning," *IEEE Transactions on Cognitive Communications and Networking*, v. 6, n. 2, p. 802–816, 2019.
- [96] —, "Cooperative sensing in deep RL-based image-to-decision proactive handover for mmWave networks," in 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC). IEEE, 2020, p. 1–6.
- [97] R. Klus, L. Klus, D. Solomitckii, M. Valkama, e J. Talvitie, "Deep learning based localization and HO optimization in 5G NR networks," in 2020 International Conference on Localization and GNSS (ICL-GNSS). IEEE, 2020, p. 1–6.



- [98] R. Zhohov, A. Palaios, H. Rydén, R. Moosavi, e J. Berglund, "Reducing Latency: Improving Handover Procedure Using Machine Learning," in 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring). IEEE, 2021, p. 1–5.
- [99] A. Masri, T. Veijalainen, H. Martikainen, S. Mwanje, J. Ali-Tolppa, e M. Kajó, "Machinelearning-based predictive handover," in 2021 IFIP/IEEE International Symposium on Integrated Network Management (IM). IEEE, 2021, p. 648–652.
- [100] N. Aljeri e A. Boukerche, "A two-tier machine learning-based handover management scheme for intelligent vehicular networks," Ad Hoc Networks, v. 94, p. 101930, 2019.
- [101] E. Zeljković, N. Slamnik-Kriještorac, S. Latré, e J. M. Marquez-Barja, "ABRAHAM: machine learning backed proactive handover algorithm using SDN," *IEEE Transactions* on Network and Service Management, v. 16, n. 4, p. 1522–1536, 2019.
- [102] N. Nayakwadi e R. Fatima, "Machine learning based handover execution algorithm for heterogeneous wireless networks," in 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). IEEE, 2020, p. 54–58.
- [103] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, e F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5g cellular communications: Theoretical feasibility and prototype results," *IEEE communications magazine*, v. 52, n. 2, p. 106–113, 2014.
- [104] J. Wang, H. Zhu, L. Dai, N. J. Gomes, e J. Wang, "Low-complexity beam allocation for switched-beam based multiuser massive mimo systems," *IEEE Transactions on Wireless Communications*, v. 15, n. 12, p. 8236–8248, 2016.
- [105] Y. Long, Z. Chen, J. Fang, e C. Tellambura, "Data-driven-based analog beam selection for hybrid beamforming under mm-wave channels," *IEEE Journal of Selected Topics in Signal Processing*, v. 12, n. 2, p. 340–352, 2018.
- [106] I. Maksymova, C. Steger, e N. Druml, "Review of lidar sensor data acquisition and compression for automotive applications," *Multidisciplinary Digital Publishing Institute Proceedings*, v. 2, n. 13, p. 852, 2018.
- [107] T. Raj, F. H. Hashim, A. B. Huddin, M. F. Ibrahim, e A. Hussain, "A survey on lidar scanning mechanisms," *Electronics*, v. 9, n. 5, p. 741, 2020.
- [108] I. Puente, H. González-Jorge, J. Martínez-Sánchez, e P. Arias, "Review of mobile mapping and surveying technologies," *Measurement*, v. 46, n. 7, p. 2127–2145, 2013.
- [109] B. Behroozpour, P. A. M. Sandborn, M. C. Wu, e B. E. Boser, "Lidar system architectures and circuits," *IEEE Communications Magazine*, v. 55, n. 10, p. 135–142, 2017.
- [110] H. Guan, J. Li, S. Cao, e Y. Yu, "Use of mobile lidar in road information inventory: A review," *International Journal of Image and Data Fusion*, v. 7, n. 3, p. 219–242, 2016.
- [111] J. Beraldin, F. Blais, e U. Lohr, "Laser scanning technology," Airborne and terrestrial laser scanning, p. 1–42, 2010.



- [112] G. R. MacCartney, H. Yan, S. Sun, e T. S. Rappaport, "A flexible wideband millimeterwave channel sounder with local area and nlos to los transition measurements," in 2017 IEEE International Conference on Communications (ICC). IEEE, 2017, p. 1–7.
- [113] A. Klautau, P. Batista, N. González-Prelcic, Y. Wang, e 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, p. 1–9.
- [114] A. Oliveira, M. Dias, I. Trindade, e A. Klautau, "Ray-tracing 5g channels from scenarios with mobility control of vehicles and pedestrians," *arXiv preprint arXiv:2001.09964*, 2020.
- [115] A. Klautau, N. González-Prelcic, e R. W. Heath, "Lidar data for deep learning-based mmwave beam-selection," *IEEE Wireless Communications Letters*, v. 8, n. 3, p. 909–912, 2019.
- [116] M. Giordani, M. Polese, A. Roy, D. Castor, e M. Zorzi, "Standalone and non-standalone beam management for 3gpp nr at mmwaves," *IEEE Communications Magazine*, v. 57, n. 4, p. 123–129, 2019.
- [117] G. Charan e A. Alkhateeb, "Computer vision aided blockage prediction in real-world millimeter wave deployments," arXiv preprint arXiv:2203.01907, 2022.
- [118] X. Li e A. Alkhateeb, "Deep learning for direct hybrid precoding in millimeter wave massive mimo systems," in 2019 53rd Asilomar Conference on Signals, Systems, and Computers. IEEE, 2019, p. 800–805.
- [119] J. Gui, Y. Liu, X. Deng, e B. Liu, "Network capacity optimization for cellular-assisted vehicular systems by online learning-based mmwave beam selection," Wireless Communications and Mobile Computing, v. 2021, 2021.
- [120] T. T. Nguyen e K.-K. Nguyen, "A deep learning framework for beam selection and power control in massive mimo-millimeter-wave communications," *IEEE Transactions on Mobile Computing*, 2022.
- [121] T. S. Cousik, V. K. Shah, T. Erpek, Y. E. Sagduyu, e J. H. Reed, "Deep learning for fast and reliable initial access in ai-driven 6g mm wave networks," *IEEE Transactions on Network Science and Engineering*, 2022.
- [122] U. Demirhan e A. Alkhateeb, "Radar aided proactive blockage prediction in real-world millimeter wave systems," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, p. 4547–4552.
- [123] N. Gonzalez-Prelcic, R. Méndez-Rial, e R. W. Heath, "Radar aided beam alignment in mmwave v2i communications supporting antenna diversity," in 2016 Information Theory and Applications Workshop (ITA). IEEE, 2016, p. 1–7.
- [124] H. P. Tauqir e A. Habib, "Deep learning based beam allocation in switched-beam multiuser massive mimo systems," in 2019 Second International Conference on Latest trends in Electrical Engineering and Computing Technologies (INTELLECT). IEEE, 2019, p. 1– 5.



- [125] W. Xu, F. Gao, X. Tao, J. Zhang, e A. Alkhateeb, "Computer vision aided mmwave beam alignment in v2x communications," 2022. [Online]. Disponível em: https://arxiv.org/abs/2207.11409
- [126] S. Rezaie, C. N. Manchón, e E. de Carvalho, "Location- and orientation-aided millimeter wave beam selection using deep learning," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, p. 1–6.
- [127] B. Salehi, G. Reus-Muns, D. Roy, Z. Wang, T. Jian, J. Dy, S. Ioannidis, e K. Chowdhury, "Deep learning on multimodal sensor data at the wireless edge for vehicular network," *IEEE Transactions on Vehicular Technology*, v. 71, n. 7, p. 7639–7655, 2022.
- [128] M. Zecchin, M. B. Mashhadi, M. Jankowski, D. Gündüz, M. Kountouris, e D. Gesbert, "Lidar and position-aided mmwave beam selection with non-local cnns and curriculum training," *IEEE Transactions on Vehicular Technology*, v. 71, n. 3, p. 2979–2990, 2022.
- [129] J. Ruseckas, G. Molis, e H. Bogucka, "Mimo beam selection in 5g using neural networks," International Journal of Electronics and Telecommunications, v. 67, n. 4, p. 693–698, 2021.
- [130] A. Klautau, A. de Oliveira, I. P. Trindade, e W. Alves, "Generating mimo channels for 6g virtual worlds using ray-tracing simulations," in 2021 IEEE Statistical Signal Processing Workshop (SSP). IEEE, 2021, p. 595–599.
- [131] M. B. Mashhadi, M. Jankowski, T.-Y. Tung, S. Kobus, e D. Gündüz, "Federated mmwave beam selection utilizing lidar data," *IEEE Wireless Communications Letters*, v. 10, n. 10, p. 2269–2273, 2021.
- [132] W. Xu, F. Gao, S. Jin, e A. Alkhateeb, "3d scene-based beam selection for mmwave communications," *IEEE Wireless Communications Letters*, v. 9, n. 11, p. 1850–1854, 2020.
- [133] M. Dias, A. Klautau, N. González-Prelcic, e R. W. Heath, "Position and lidar-aided mmwave beam selection using deep learning," in 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2019, p. 1–5.
- [134] Y. Zheng, S. Chen, e R. Zhao, "A deep learning-based mmwave beam selection framework by using lidar data," in 2021 33rd Chinese Control and Decision Conference (CCDC). IEEE, 2021, p. 915–920.
- [135] S. Wu, C. Chakrabarti, e A. Alkhateeb, "Lidar-aided mobile blockage prediction in realworld millimeter wave systems." in WCNC, 2022, p. 2631–2636.
- [136] S. Jiang, G. Charan, e A. Alkhateeb, "Lidar aided future beam prediction in real-world millimeter wave v2i communications," 2022. [Online]. Disponível em: https://arxiv.org/abs/2203.05548
- [137] I. Aleksander, M. De Gregorio, F. M. G. França, P. M. V. Lima, e H. Morton, "A brief introduction to weightless neural systems." in *ESANN*. Citeseer, 2009, p. 299–305.
- [138] W. W. Bledsoe e I. Browning, "Pattern recognition and reading by machine," in IRE-AIEE-ACM '59 (Eastern), 1959.



- [139] F. França, M. De Gregorio, P. Lima, e W. de Oliveira, "Advances in weightless neural systems," in European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2014, p. 497–504.
- [140] B. P. Grieco, P. M. Lima, M. De Gregorio, e F. M. França, "Producing pattern examples from "mental" images," *Neurocomputing*, v. 73, n. 7-9, p. 1057–1064, 2010.
- [141] R. Lacerda Queiroz, F. Ferrentini Sampaio, C. Lima, e P. Machado Vieira Lima, "Ai from concrete to abstract: Demystifying artificial intelligence to the general public," arXiv eprints, p. arXiv-2006, 2020.
- [142] A. Klautau, P. Batista, N. González-Prelcic, Y. Wang, e R. W. Heath, "5g mimo data for machine learning: Application to beam-selection using deep learning," in 2018 Information Theory and Applications Workshop (ITA), 2018, p. 1–9.
- [143] A. Klautau, N. González-Prelcic, e R. W. Heath, "Lidar data for deep learning-based mmwave beam-selection," *IEEE Wireless Communications Letters*, v. 8, n. 3, p. 909–912, 2019.
- [144] L. Li, W. Shao, e X. Zhou, "A flexible scheduling algorithm for the 5th-generation networks," *Intelligent and Converged Networks*, v. 2, n. 2, p. 101–107, 2021.
- [145] W. Guan, H. Zhang, e V. C. M. Leung, "Customized slicing for 6g: Enforcing artificial intelligence on resource management," 2021. [Online]. Disponível em: https://arxiv.org/abs/2102.10498
- [146] M. Elsayed e M. Erol-Kantarci, "AI-enabled radio resource allocation in 5G for URLLC and eMBB users," in 2019 IEEE 2nd 5G World Forum (5GWF). IEEE, 2019, p. 590– 595.
- [147] M. Polese, L. Bonati, S. D'Oro, S. Basagni, e T. Melodia, "ColO-RAN: Developing Machine Learning-based xApps for Open RAN Closed-loop Control on Programmable Experimental Platforms," arXiv preprint arXiv:2112.09559, 2021.
- [148] J. Mei, X. Wang, K. Zheng, G. Boudreau, A. B. Sediq, e H. Abou-Zeid, "Intelligent radio access network slicing for service provisioning in 6G: A hierarchical deep reinforcement learning approach," *IEEE Transactions on Communications*, v. 69, n. 9, p. 6063–6078, 2021, publisher: IEEE.
- [149] I.-S. Comsa, A. De-Domenico, e D. Ktenas, "Qos-driven scheduling in 5g radio access networks-a reinforcement learning approach," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, p. 1–7.
- [150] J. S. Shekhawat, R. Agrawal, K. G. Shenoy, e R. Shashidhara, "A reinforcement learning framework for QoS-driven radio resource scheduler," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, p. 1–7.
- [151] F. Chollet, *Deep Learning with Python*. Manning, Nov. 2017.
- [152] X. Li, R. Ni, J. Chen, Y. Lyu, Z. Rong, e R. Du, "End-to-end network slicing in radio access network, transport network and core network domains," *IEEE Access*, v. 8, p. 29525–29537, 2020.



- [153] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, e P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE communications surveys* & tutorials, v. 15, n. 2, p. 678–700, 2012.
- [154] B. Khodapanah, A. Awada, I. Viering, J. Francis, M. Simsek, e G. P. Fettweis, "Radio resource management in context of network slicing: What is missing in existing mechanisms?" in *Proc. of IEEE Wireless Communications and Networking Conference* (WCNC). IEEE, 2019, p. 1–7.
- [155] S. Jaeckel, L. Raschkowski, K. Börner, e L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Transactions on Antennas and Propagation*, v. 62, n. 6, p. 3242–3256, 2014.
- [156] P. Medeđović, M. Veletić, e Ž. Blagojević, "Wireless insite software verification via analysis and comparison of simulation and measurement results," in 2012 Proceedings of the 35th International Convention MIPRO. IEEE, 2012, p. 776–781.
- [157] D. Krajzewicz, "Traffic simulation with sumo-simulation of urban mobility," in Fundamentals of traffic simulation. Springer, 2010, p. 269–293.
- [158] T. Haarnoja, A. Zhou, P. Abbeel, e S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proc. of International* conference on machine learning. PMLR, 2018, p. 1861–1870.
- [159] R. Aben-Athar, C. Nahum, D. Brilhante, J. Rezende, L. Mendes, e A. Klautau, "User Scheduling and Beam-Selection with Tabular and Deep Reinforcement Learning," in *Proc. of XL Brazilian Symposium on Telecommunications and Signal Processing (SBrT)*. SBrT, 2022.
- [160] W. Weaver, "Recent contributions to the mathematical theory of communication," *ETC:* a review of general semantics, p. 261–281, 1953.
- [161] F. I. Dretske, Knowledge and the Flow of Information. MIT Press, 1981.
- [162] W. Yang, H. Du, Z. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. S. Shen, e C. Miao, "Semantic communications for 6g future internet: Fundamentals, applications, and challenges," 2022. [Online]. Disponível em: https://arxiv.org/abs/2207.00427
- [163] C. E. Shannon, "A mathematical theory of communication," The Bell System Technical Journal, v. 27, n. 3, p. 379–423, 1948.
- [164] Y. Zhong, "A theory of semantic information," China communications, v. 14, n. 1, p. 1–17, 2017.
- [165] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad, G. Liva, A. Munari, T. Soleymani, B. Soret, e K. H. Johansson, "Semantic communications in networked systems," 2021. [Online]. Disponível em: https://arxiv.org/abs/2103.05391
- [166] D. Wheeler e B. Natarajan, "Engineering semantic communication: A survey," 2022. [Online]. Disponível em: https://arxiv.org/abs/2208.06314


- [167] R. Carnap e Y. Bar-Hillel, "An outline of a theory of semantic information," Journal of Symbolic Logic, v. 19, n. 3, p. 230–232, 1954.
- [168] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, e J. A. Hendler, "Towards a theory of semantic communication," in 2011 IEEE Network Science Workshop, 2011, p. 110–117.
- [169] L. Floridi, "Outline of a theory of strongly semantic information," Minds and Machines, v. 14, 05 2004.
- [170] S. D'Alfonso, "On quantifying semantic information," Information, v. 2, 12 2011.
- [171] P. Basu, J. Bao, M. Dean, e J. Hendler, "Preserving quality of information by using semantic relationships," in 2012 IEEE International Conference on Pervasive Computing and Communications Workshops, 2012, p. 58–63.
- [172] Z. Qin, H. Ye, G. Y. Li, e B.-H. F. Juang, "Deep learning in physical layer communications," *IEEE Wireless Communications*, v. 26, n. 2, p. 93–99, 2019.
- [173] T. O'shea e J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, v. 3, n. 4, p. 563–575, 2017.
- [174] A. Krizhevsky, I. Sutskever, e G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, v. 60, n. 6, p. 84–90, 2017.
- [175] T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, e S. Khudanpur, "Recurrent neural network based language model." in *Interspeech*, v. 2, n. 3. Makuhari, 2010, p. 1045– 1048.
- [176] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, e I. Polosukhin, "Attention is all you need," Advances in neural information processing systems, v. 30, 2017.
- [177] H. Xie, Z. Qin, G. Y. Li, e B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, v. 69, p. 2663–2675, 2021.
- [178] Y. Tay, M. Dehghani, D. Bahri, e D. Metzler, "Efficient transformers: A survey," ACM Computing Surveys (CSUR), 2020.
- [179] M. Kountouris e N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, v. 59, n. 6, p. 96–102, 2021.
- [180] Wikipedia, "Value of Information Wikipedia, the free encyclopedia," http://https: //en.wikipedia.org/wiki/Value_of_information, 2022, [Online; accessed 25-September-2022].
- [181] E. Uysal, O. Kaya, A. Ephremides, K. Johansson, M. Codreanu, B. Soret, M. Assad, G. Liva, e A. Munari, "Semantic communications in networked systems: A data significance perspective," *IEEE Network*, 01 2022.