



UNIVERSIDADE
ESTADUAL de LONDRINA

THIAGO AUGUSTO BRUZA ALVES

**EFICIÊNCIA E DESEMPENHO DE SISTEMAS DE MÚLTIPLO
ACESSO NÃO-ORTOGONAL MIMO MASSIVO E ULTRA
MASSIVO**



Universidade Estadual de Londrina
Centro de Tecnologia e Urbanismo
Departamento de Engenharia Elétrica
Programa de Doutorado em Engenharia Elétrica

Thiago Augusto Bruza Alves

Eficiência e Desempenho de Sistemas de Múltiplo Acesso Não-Ortogonal
MIMO Massivo e Ultra Massivo

Londrina,
3 de junho de 2024

Thiago Augusto Bruza Alves

Eficiência e Desempenho de Sistemas de Múltiplo Acesso Não-Ortogonal
MIMO Massivo e Ultra Massivo

Tese de Doutorado submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Estadual de Londrina como parte dos requisitos para adquirir o grau de Doutor em Engenharia Elétrica.

Área: Sistemas de Telecomunicações

Orientador: Prof. Dr. Taufik Abrão

Londrina,
3 de junho de 2024

Alves, Thiago Augusto Bruza

Eficiência e Desempenho de Sistemas de Múltiplo Acesso Não-Ortogonal MIMO Massivo e Ultra Massivo. Londrina, 2024. [158](#) p.

Orientador: Taufik Abrão

Thesis (Doctor of Philosophy) – Departamento de Engenharia Elétrica – Universidade Estadual de Londrina

1. Sistemas MIMO Massivo, 2. Múltiplo Acesso Não Ortogonal, 3. Comunicação do tipo máquina, 4. Acesso aleatório, 5. Eficiência de Recursos.

Thiago Augusto Bruza Alves

Eficiência e Desempenho de Sistemas de Múltiplo Acesso Não-Ortogonal
MIMO Massivo e Ultra Massivo

Tese de Doutorado submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Estadual de Londrina como parte dos requisitos para adquirir o grau de Doutor em Engenharia Elétrica.

Área: Sistemas de Telecomunicações

Comissão Examinadora

Prof. Dr. Daniel Costa Araújo
Departamento de Engenharia Eletrônica
Faculdade UnB Gama - FGA
Universidade de Brasília
Brasília - DF

Prof. Dr. Ivan Roberto Santana Casella
Centro de Engenharia, Modelagem e Ciências
Sociais Aplicadas (CECS)
Universidade Federal do ABC
Santo André - SP

Prof. Dr. João Luiz Rebelatto
Departamento Acadêmico de Eletrônica
Universidade Tecnológica Federal do Paraná
Campus Curitiba - Pr.

Prof. Dr. José Carlos Marinello Filho
Departamento de Engenharia Elétrica
Universidade Tecnológica Federal do Paraná
Cornélio Procópio - Pr.

Prof. Dr. Taufik Abrão
Departamento de Engenharia Elétrica
Universidade Estadual de Londrina
Orientador

Londrina,
3 de junho de 2024

Agradecimentos

Agradeço primeiramente a Deus, fonte da minha vida, que de forma singular me conduziu até aqui. Agradeço também aos meus pais Fani e José que me ensinaram desde cedo o caminho a seguir e me guiaram por este dando todo o apoio, sustento, direção e amor. Aos meus irmãos Rodrigo e Roberta pelo carinho. Aos meus sogros Luiz e Izabel pela motivação e apoio.

Agradeço, de forma especial, a minha esposa Graziela, pelo incentivo e apoio incondicional nesse período. E por meus filhos Rafael e Mateus, presentes de Deus na minha vida.

Agradeço ao professor Doutor Taufik Abrão, pela grande paciência, apoio e brilhante orientação nesses anos.

Aos amigos de laboratório: Professor Doutor Jaime, Layhon, João Henrique, Herman, Gabriel, David, Wilson, muito faltaria neste trabalho se não fosse por esta turma. A vocês minha eterna gratidão.

”Humildade é vermo-nos como somos, sem paliativos, com a verdade.”

(São Josemaria Escrivá)

“O maior inimigo do conhecimento não é a ignorância, é a ilusão do conhecimento.”

(Stephen Hawking)

Resumo

A demanda por recursos de rede de comunicações cresce em escala exponencial. A sociedade vive um momento de grande necessidade de acesso rápido aos serviços de *internet*, controle da informação e entretenimento. A evolução das técnicas de telecomunicação é fundamental para atender às demandas crescentes da sociedade e permitir o desenvolvimento de novas aplicações e serviços. É importante garantir que essa evolução seja sustentável, focada em segurança e privacidade dos dados, para permitir a comunicação eficiente e segura.

Nesta Tese de Doutorado, as técnicas de múltiplas-entradas e múltiplas-saídas massiva (mMIMO – *multiple-input multiple-output*) e suas variações, juntamente com a aplicação de técnicas de acesso múltiplo não ortogonal (NOMA – *non-orthogonal multiple access*), são investigadas visando atender aos desafios e às necessidades de evolução das redes sem fio de quinta geração (5G – *fifth generation*). Desenvolvemos uma análise comparativa para a eficiência energética (EE – *energy efficiency*) e a eficiência espectral (SE – *spectral efficiency*) em termos de bits recebidos por antena, comparando as duas técnicas, com destaque aos resultados superiores obtidos pelo esquema de transmissão NOMA no domínio das potências (PD-NOMA – *power domain NOMA*) com $0 < \rho \leq 2$, sendo $\rho = \frac{K}{M}$, K o número de dispositivos e M o número de antenas na estação rádio-base (BS – *base station*).

Na segunda parte da Tese, propomos um método de acesso aleatório baseado em concessão apropriado para operar em sistemas MIMO ultra massivos (XL-MIMO – *extra-large scale massive MIMO*) NOMA. As figuras de mérito analisadas incluem número de tentativas de acesso e soma das taxas de dados alcançada pelo sistema quando o número de sub-arranjos de antenas (SA – *subarrays*) é incrementado.

Por fim, apresentamos o estágio atual do desenvolvimento envolvendo as técnicas MIMO, NOMA, protocolos de acesso aleatórios e destacamos as oportunidades de trabalhos futuros.

Palavras-chave: mMIMO, XL-MIMO, NOMA, eficiência energética, eficiência espectral, acesso aleatório.

Abstract

The demand for communication network resources is increasing at an exponential rate. Society is in desperate need of speedy access to internet services, information control, and entertainment. The growth of telecommunication technology is critical to meeting society's expanding needs and enabling the creation of new applications and services. To enable efficient and secure communication, it is critical to guarantee that this progression is sustainable and focuses on data security and privacy.

In this Doctoral thesis, the massive multiple-input and multiple-output (mMIMO) techniques and their variations, together with the application of non-orthogonal multiple access (NOMA) techniques, are investigated in order to meet the challenges and evolution needs of fifth generation wireless networks fifth generation (5G). We developed a comparative analysis for energy efficiency (EE) and spectral efficiency (SE) in terms of bits received per antenna, comparing the two techniques, highlighting the superior results obtained by the NOMA transmission scheme in the power domain NOMA (PD-NOMA) with $0 < \rho < 2$, where $\rho = \frac{K}{M}$, K the number of devices and M the number of antennas in the base station (BS). In the second part of the thesis, we propose a grant-based random access method suitable for operating in extra-large scale massive MIMO (XL-MIMO) NOMA systems. Figures of merit analyzed include number of access attempts and sum of data rates achieved by the system when the number of antenna subarrays (SA – subarrays) is increased.

Finally, we present the current stage of development involving MIMO, NOMA, random access protocols and highlight opportunities for future work.

Palavras-chave: *mMIMO, XL-MIMO, NOMA, energy efficiency, spectral efficiency, random access.*

Lista de figuras

Figura 1.1 – Classificação dos esquemas de múltiplo acesso.	24
Figura 1.2 – Receptor SIC do UE ₁ no DL PD-NOMA com 2 usuários.	28
Figura 1.3 – Receptor PIC multiestágio do UE ₁ no DL PD-NOMA com K usuários: os sinais dos K são estimados no primeiro estágio do cancelador, nos estágios seguintes os sinais interferentes são cancelados e uma versão estimada com mais qualidade do sinal de interesse é produzida.	29
Figura 1.4 – Esquema ilustrando o DL PD-NOMA: UE ₁ decodifica o sinal do UE ₂ (interferente), aplica o SIC e decodifica seu sinal UE ₁ (interesse), UE ₂ decodifica seu sinal de interesse com a interferência do sinal do UE ₁	30
Figura 1.5 – Taxas alcançadas na comparação das técnicas NOMA e OMA no cenário da Figura 1.4. As figuras consideram o SIC perfeito, mas atendendo a Ξ em 3 níveis a) $\Xi = 10$ dB, b) $\Xi = 6$ dB e c) $\Xi = 3$ dB.	32
Figura 1.6 – Taxa alcançada por dispositivo UE ₁ e UE ₂ , no cenário representado na Figura 1.4. a) Sistema NOMA indicando as Ξ . b) Sistema OMA.	32
Figura 1.7 – Esquema representando o UL NOMA básico, UE ₁ e UE ₂ experimentam condições de canal distintas e transmitem com a mesma potência. A recepção ocorre na BS via receptor SIC.	33
Figura 1.8 – Esquema representando CD-NOMA básico, UEs compartilham o mesmo RB (tempo x frequência), com sequência de espalhamento (código) diferentes.	35
Figura 1.9 – Esquema representando PDMA básico, onde 6 UEs compartilham o mesmo RB (tempo x frequência), com potência e divisão espacial (<i>beamforming</i>) diferentes.	35
Figura 1.10 – Exemplo de célula única mMIMO com M antenas e K dispositivos.	36
Figura 1.11 – Protocolo de acesso aleatório baseado em concessão (SUCRe).	45
Figura 2.1 – Ilustração de exemplo de clusterização de 12 UEs no DL e no UL, quando o <i>cluster</i> é formado por 2 UEs.	58
Figura 2.2 – Representação dos possíveis cenários NOMA combinando canal LOS e NLOS, sendo UEs pertencentes ao mesmo <i>cluster</i> são atendidos pela mesmo <i>beamforming</i>	60
Figura 2.3 – Estrutura de quadro temporal baseado no tempo de coerência do canal: a estrutura de treinamento e transmissão de dados para os esquemas mMIMO e NOMA sob configuração TDD NLOS. Nesta configuração é obtida uma combinação linear dos canais dos UEs pertencentes ao <i>cluster</i>	67

Figura 2.4 – Estrutura de quadro temporal tendo como base o tempo de coerência do canal: a estrutura de treinamento e transmissão de dados para esquemas mMIMO e NOMA sob configuração TDD NLOS. Nesta segunda configuração é obtido somente os canais dos UEs do conjunto $\mathcal{K}_H$	68
Figura 3.1 – Diagrama representando as distâncias do k -ésimo usuário ao XL-MIMO URA, com ângulo de elevação θ e do ângulo de azimute ϕ . Sendo $M = M_y \times M_z$ igualmente espaçados horizontal e verticalmente com espaçamento d_m	70
Figura 3.2 – Diagrama da arquitetura da BS. O arranjo de antenas na BS é composto por Q SAs contendo um número fixo de elementos de antena M_q . Cada SA possui uma RPU e todas estão conectadas ao EGC na CPU.	70
Figura 3.3 – Sistema de comunicação XL-MIMO URA composto por Q -SAs contendo M_q antenas, as regiões coloridas em cada SA definem UEs visíveis, a probabilidade de sucesso de visibilidade P_b para cada SA.	72
Figura 3.4 – Modelo esquemático que demonstra a relação entre a abertura do arranjo de antenas e a distância de Rayleigh.	73
Figura 4.1 – Modelo do sistema indicando a formação de pares no sistema NOMA.	76
Figura 4.2 – Média da soma das taxas para o carregamento de usuários $0 < \rho \leq 2$, considerando os quatro métodos de alocação de potência: EPA, PCIPA, WF e Δ -WF. Sendo a média obtida após 1000 distribuições aleatórias de usuários.	80
Figura 4.3 – Média de dispositivos ativos após PA <i>versus</i> carregamento na faixa $0 < \rho \leq 2$. Sendo a média obtida após 1000 distribuições aleatórias de usuários.	82
Figura 4.4 – Média da soma das taxas pelo carregamento ρ e M	82
Figura 4.5 – A justiça dos sistemas NOMA e do mMIMO nas quatro políticas alocação de potência: (a) EPA; b) PCIPA; (c) WF and Δ -WF. Sendo a média obtida após 1000 distribuições aleatórias de usuários.	84
Figura 4.6 – EE para NOMA e mMIMO nas quatro políticas alocação de potência: (a) EPA; b) PCIPA; (c) WF and Δ -WF. Sendo a média obtida após 1000 distribuições aleatórias de usuários.	86
Figura 4.7 – EE com carregamento ρ e M	87
Figura 4.8 – Área sob a curva sistemas NOMA e mMIMO nas quatro políticas alocação de potência: (a) Curvas de SE; b) Curvas de EE. Sendo a média obtida após 1000 distribuições aleatórias de usuários. O eixo da esquerda é referente as barras e o eixo da direita referente as linhas.	88

Figura 4.9 – Figuras representando a faixa do melhor compromisso SE-EE para diferentes carregamentos com $M = 64$; com potência de transmissão $P_{TX} = 1W$ /antena. a) As curvas em vermelho contínuas são referentes a SE (eixo da esquerda) curvas em azul tracejado são referentes a EE (eixo da direita). b) Frente de Pareto, dentro deste fronte se encontre o ponto do melhor compromisso SE-EE.	89
Figura 4.10–Figuras representando a faixa do melhor compromisso SE-EE para diferentes carregamentos com $M = 128$; com potência de transmissão $P_{TX} = 1W$ /antena. a) As curvas em vermelho contínuas são referentes a SE (eixo da esquerda) curvas em azul tracejado são referentes a EE (eixo da direita). b) Frente de Pareto, dentro deste fronte se encontre o ponto do melhor compromisso SE-EE.	90
Figura 4.11–Figuras representando a faixa do melhor compromisso SE-EE para diferentes carregamentos com $M = 256$; com potência de transmissão $P_{TX} = 1W$ /antena. a) As curvas em vermelho contínuas são referentes a SE (eixo da esquerda) curvas em azul tracejado são referentes a EE (eixo da direita). b) Frente de Pareto, dentro deste fronte se encontre o ponto do melhor compromisso SE-EE.	90
Figura 5.1 – Diagrama do protocolo proposta NVR-XL, linhas tracejadas demonstram etapas idênticas ao SUCRe-XL, "Passo 0" e "Transmissão de Dados" estão presentes na representação.	93
Figura 5.2 – δ_{NVR-XL}^* ótimo em termos de maximização da soma das taxas, obtido numericamente para cada carregamento de usuários, nas duas configurações propostas e para as diferentes configurações de <i>array</i> $Q \in \{5; 10\}$	100
Figura 5.3 – Média da soma das taxas em função do número de iUEs. Obtido após ajuste do δ_{NVR-XL}^*	100
Figura 5.4 – Média da soma das taxas em função do número de iUEs, $Q = 10$ na configuração 1 no NVR-XL, com variações em torno δ^* ótimo.	101
Figura 5.5 – Performance de RA dos sistemas. A linha base SUCRe-XL e NVR-XL. (a) Quantidade média de tentativas de RA. (b) Probabilidade de falhas nas tentativas de acesso.	102
Figura 5.6 – Quantidade normalizada de usuários aceitos, obtido com δ_{NVR-XL}^*	102
Figura 5.7 – Porcentagem de colisões resolvidas pelo SIC e a quantidade média de usuários aceitos em NVR-XL.	104
Figura 5.8 – Fração de casos <i>falso positivos</i> resolvidos pelo SIC no NVR-XL. a) Configuração 1 e b) Configuração 2.	105

Lista de tabelas

Tabela 1.1 – Tabela de parâmetros para simulação NOMA vs OMA.	31
Tabela 2.1 – Informações prévias necessárias para a realização da alocação de potência em sistemas mMIMO x PD-NOMA.	66
Tabela 3.1 – Relação entre abertura do arranjo de antenas e distância de Rayleigh. .	73
Tabela 4.1 – Parâmetros de simulação sistema mMIMO-NOMA.	79
Tabela 4.2 – Valores dos parâmetros adotados na análise de EE (Björnson et al., 2014).	83
Tabela 4.3 – Compromisso SE-EE.	91
Tabela 5.1 – Parâmetros de simulação RA-GB NOMA XL-MIMO.	99

Lista de Acrônimos

1G	Redes de comunicações móveis da primeira geração
2G	Redes de comunicações móveis da segunda geração
3G	Redes de comunicações móveis da terceira geração
4G	Redes de comunicações móveis da quarta geração
5G	Redes de comunicações móveis da quinta geração
6G	Redes de comunicações móveis da sexta geração
B5G	Redes de comunicações móveis além da quinta geração (<i>beyond 5G</i>)
BS	Estação base (<i>base station</i>)
CDMA	Múltiplo acesso por divisão de código (<i>code division multiple access</i>)
CD-NOMA	Domínio do Código - múltiplo acesso não ortogonal (<i>code domain non orthogonal multiple access</i>)
CPU	Unidade de processamento central (<i>central processing unit</i>)
CSI	Informação do estado do canal (<i>channel state information</i>)
DL	Comunicação no sentido descendente (BS para UE) (<i>downlink</i>)
EE	Eficiência energética (<i>energy efficiency</i>)
EGC	Combinador de ganho igual (<i>equal gain combiner</i>)
eMBB	Banda larga móvel aprimorada (<i>enhanced mobile broadband</i>)
EPA	Alocação de potência igualitária (<i>equal power allocation</i>)
FDD	Duplex por divisão de frequência (<i>frequency division duplex</i>)
FDMA	Múltiplo acesso por divisão de frequência (<i>frequency division multiple access</i>)
FM	Modulação em frequência (<i>frequency modulation</i>)
GB	Baseado em concessão (<i>grant based</i>)
GF	Livre de concessão (<i>grant free</i>)

LDS	Espalhamento de baixa densidade (<i>low-density spreading</i>)
LU	Inferior superior (<i>lower upper</i>)
mMTC	Comunicação tipo máquina massiva (<i>massive machine type communication</i>)
MIMO	Múltiplas entradas e múltiplas saídas (<i>multiple-input multiple-output</i>)
mMIMO	MIMO massivo (<i>massive MIMO</i>)
NLOS	Sem linha de visada (<i>non-line-of-sight</i>)
NOMA	Múltiplo acesso não ortogonal (<i>non orthogonal multiple access</i>)
OFDM	Multiplexação de divisão de frequência ortogonal (<i>orthogonal frequency division multiplexing</i>)
OFDMA	Múltiplo acesso por divisão de frequências ortogonais (<i>orthogonal frequency division multiple access</i>)
OMA	Múltiplo acesso ortogonal (<i>orthogonal multiple access</i>)
PDMA	Acesso múltiplo por divisão de padrão (<i>pattern division multiple access</i>)
PD-NOMA	Domínio da potência NOMA (<i>pattern division multiple access</i>)
PIC	Cancelamento de interferência paralelo (<i>parallel interference cancelation</i>)
PCIPA	Alocação de potência pelo inverso do canal (<i>proportional channel inversion power allocation</i>)
RA	Acesso aleatório (<i>random access</i>)
RB	Bloco de recurso (<i>resource block</i>)
RF	Radio frequência (<i>radio frequency</i>)
RPU	Unidade de processamento remoto (<i>remote processing unit</i>)
ROC	Característica de operação do receptor (<i>receiver operating characteristic</i>)
SA	Sub-arranjo de antenas (<i>subarray</i>)
SE	Eficiência espectral (<i>spectral efficiency</i>)
SIC	Cancelamento de interferência sucessivo (<i>successive interference cancelation</i>)
SINR	Relação sinal ruído mais interferência (<i>signal to noise plus interference ratio</i>)

SVD	Decomposição em valores singulares (<i>singular value decomposition</i>)
SUCRe	Resolução de colisão de usuários mais fortes (<i>strongest users collision resolution</i>)
TAS	Seleção de antena transmissora (<i>transmitter antenna selection</i>)
TDD	Duplex por divisão de tempo (<i>time division duplex</i>)
TDMA	Múltiplo acesso por divisão de tempo (<i>time division multiple access</i>)
UE	Equipamento de usuário (<i>user equipment</i>)
UL	Comunicação no sentido ascendente (UE para BS) (<i>uplink</i>)
URA	Arranjo de antena retangular (<i>uniforme rectangular array</i>)
uRLLC	Comunicação ultra confiável de baixa latência (<i>ultra reliable low latency communication</i>)
VR	Região de visibilidade (<i>visible region</i>)
WF	Enchimento de água (<i>water filling</i>)
XL-MIMO	MIMO ultra massivo (<i>extra large MIMO</i>)
ZF	Forçamento a zero (<i>zero forcing</i>)

Notações Matemáticas

Definição de variáveis

k, λ, M	escalar, caractere itálico;
$\mathbf{a}$	vetor, caractere minúscula negrito;
$\mathbf{A}$	matriz, caractere maiúscula negrito;
$\mathcal{A}$	conjunto, caractere maiúscula caligrafico;

Operadores

$\max(\cdot)$	retorna o maior elemento do argumento;
$\min(\cdot)$	retorna o menor elemento do argumento;
$\operatorname{argmax}_{a \in \mathcal{A}}(\cdot)$	elemento do conjunto $\mathcal{A}$ que maximiza o argumento;
$\operatorname{argmin}_{a \in \mathcal{A}}(\cdot)$	elemento do conjunto $\mathcal{A}$ que minimiza o argumento;
$\mathbb{E}\{\cdot\}$	operador esperança estatística;
$ a $	operador de magnitude (valor absoluto);
$ \mathcal{A} $	cardinalidade do conjunto;
$\sum_{a=1}^N$	somatória considerando a soma dos elementos 1 até o N;
$\sum_{a \in \mathcal{N}}$	somatória considerando os elementos de $\mathcal{N}$;
$\Re\{\cdot\}$	parte real de um número complexo;
$\mathcal{CN}(0, \sigma^2 \mathbf{I}_{M_b})$	distribuição normal complexa com média 0 e variância $\sigma^2 \mathbf{I}_{M_b}$;
$\mathcal{B}\left(K, \frac{P_a}{\tau_{\text{RA}}}\right)$	distribuição binomial com número de rodadas K e probabilidade de sucesso $\frac{P_a}{\tau_{\text{RA}}}$;
$\ \cdot\ _\ell$	norma ℓ do argumento;
$\mathcal{A} \subset \mathcal{C}$	operador subconjunto, $\mathcal{A}$ é subconjunto de $\mathcal{C}$;
$\mathcal{D} = \mathcal{A} \cup \mathcal{C}$	operador união de conjuntos, $\mathcal{D}$ é conjunto resultante da união dos conjunto $\mathcal{A}$ e $\mathcal{C}$;
$\mathcal{D} = \mathcal{A} \setminus \mathcal{C}$	operador subtração de conjunto, $\mathcal{D}$ é conjunto resultante da diferença dos conjuntos $\mathcal{A}$ e $\mathcal{C}$;

Índices sobrescritos

$\mathbf{A}^{\text{H}}$ transposta conjugada, ou transposta hermitiana, da matriz $\mathbf{A}$;

$\mathbf{A}^{\text{T}}$ transposta da matriz $\mathbf{A}$;

$\mathbf{A}^{-1}$ inversa da matriz $\mathbf{A}$;

Índices subscritos

$\mathbf{a}_i$ i -ésimo elemento do vetor $\mathbf{a}$;

$\mathbf{A}_{i,j}$ elemento da i -ésima linha da j -ésima coluna da matriz $\mathbf{A}$;

$\mathbf{A}_k$ Vetor representado pela k -ésima coluna da matriz $\mathbf{A}$;

Lista de símbolos

$\mathcal{F}_K$	índice de justiça de Jain;
K	número considerado de terminais na célula;
$\mathcal{K}$	conjunto de K usuários;
$\mathcal{K}_H$	conjunto dos $K/2$ usuários com melhores condições de canal;
$\mathcal{K}_L$	conjunto dos $K/2$ usuários com piores condições de canal;
M	número de antenas na estação rádio-base;
$\bar{M}$	ganho de <i>array</i> obtido no NOMA;
M_q	número de antenas no SA;
L	número de <i>clusters</i> NOMA;
$\mathcal{L}$	conjunto de <i>clusters</i> NOMA;
D^{SIC}	Faixa dinâmica de operação do SIC;
B	largura de banda disponível;
I	número de operações por Joule;
$\mathbf{y}$	vetor sinais contendo a informação recebida;
$\mathbf{x}_k$	vetor sinais contendo a informação transmitida do usuário k ;
$\mathbf{h}_k$	vetor coeficiente de canal do usuário k ;
$\mathbf{g}_k$	vetor de pré-codificação usuário k ;
$\mathbf{n}_{0,k}$	vetor ruído do aditivo referente ao usuário k ;
s_k	sequência de espalhamento do usuário k ;
N_0	densidade espectral de potência na entrada do receptor;
β_k	perda de percurso usuário k ;
β_0	perda de percurso à distância de referência;
ξ	coeficiente de perda de percurso;
R	taxa alcançável;
T_c	tempo de coerência do canal;
Q	número de divisões do arranjo de antenas;
$\mathbf{H}$	matriz de coeficientes de canal ($M \times K$);
$\bar{\mathbf{H}}$	matriz de canal equivalente para BF-ZF;
$\mathbf{H}_l$	matriz de canal do <i>cluster</i> l ;

- ϱ constante de proporção da combinação linear;
- $\mathbf{G}^{ZF}$ matriz de pré-codificação ZF;
- φ perda de percurso - termo médio (*shadowing*);
- ϱ fração do canal do usuário de borda no canal equivalente;
- τ_p tamanho das sequências piloto em símbolos;
- P_a probabilidade de tentativa de acesso (primeira tentativa);
- P_{max} potência máxima a ser processada antes da saturação do receptor;
- P_{min} potência mínima ou potência de sensibilidade do receptor;
- P_{na} probabilidade de tentativa de acesso (novas tentativas);
- P_T potência total disponível para rádio frequência;
- ϑ eficiência do amplificador de potência;
- p_k potência do usuário k ;
- p_{BS} potência de transmissão da BS;
- α_k fração de potência alocada ao usuário k ;
- p_{cl} potência alocada ao *cluster* l ;
- $p_{l,k}$ potência alocada ao usuário k do *cluster* l ;
- Ξ relação mínima de potências necessária para realização do SIC;
- τ_c tamanho do bloco de coerência do canal;
- τ_{ra} número de sequências piloto disponível;
- pos_k vetor de posição 3D do UE $_k$;
- λ peso do regularizador;
- ζ número de dispositivos por *cluster*;

Sumário

	Lista de acrônimos	9
	Lista de notações matemáticas	12
	Lista de símbolos	14
	Lista de figuras	16
	Lista de tabelas	19
	Sumário	20
1	INTRODUÇÃO E FUNDAMENTAÇÃO CONCEITUAL . .	23
1.1	NOMA no Domínio da Potência (PD-NOMA)	25
1.1.1	Desvantagens de sistemas PD-NOMA	26
1.1.2	Cancelamento de interferência sucessivo – SIC	27
1.1.3	Cancelamento de interferência paralelo – PIC	28
1.1.4	Exemplo – DL PD-NOMA	29
1.1.5	Exemplo – UL PD-NOMA	31
1.2	Domínio do Código – CD-NOMA	34
1.3	Divisão de Padrão – PDMA	34
1.4	MIMO massivo	36
1.4.1	Modelo de sistema mMIMO	36
1.4.2	Modelagem do Canal sem fio	37
1.4.3	Bloco de coerência do canal	38
1.4.4	Reciprocidade do canal	39
1.4.5	Enrijecimento do canal e propagação favorável em mMIMO .	39
1.5	Figuras de mérito	40
1.5.1	Eficiência Espectral	41
1.5.2	Eficiência Energética	41
1.5.3	Área sobre as Curvas SE e EE	43
1.5.4	Índice de Justiça de Jain	43
1.5.5	Métricas para protocolos de acesso	44
1.6	Acesso Aleatório	44
1.6.1	Por concessão	44
1.6.2	Livre de concessão	46
1.7	Objetivos	46

1.8	Estrutura do Trabalho de Tese	47
1.9	Contribuições	47
2	MMIMO-NOMA	55
2.1	Modelo de sistema mMIMO-NOMA	56
2.2	Agrupamento de usuários (<i>Clustering</i>)	57
2.3	Formatação de feixe	59
2.4	Políticas de alocação de potência	62
2.4.1	Alocação de potência igualitária	63
2.4.2	Alocação de potência proporcional ao inverso do canal	63
2.4.3	Alocação de potência <i>water-filling</i>	64
2.4.4	WF Modificado para NOMA	64
2.4.5	Análise de complexidade	65
2.5	Informações do estado do canal	66
2.6	Conclusão do capítulo	68
3	MIMO ULTRA MASSIVO	69
3.1	Modelo de sistema XL-MIMO	69
3.2	Modelo de canal XL-MIMO	71
3.3	Validação do modelo de canal	72
3.4	Conclusão do Capítulo	73
4	EFICIÊNCIA DE RECURSOS EM MMIMO-NOMA SOB DIFERENTES POLÍTICAS DE ALOCAÇÃO DE POTÊNCIA	74
4.1	Modelo de Sistema	74
4.2	SE-EE em sistemas DL NOMA e mMIMO	75
4.2.1	Taxa de dados no DL NOMA via ZF	75
4.2.2	Taxa de dados DL mMIMO com ZF	78
4.3	Resultados Numéricos	78
4.3.1	Eficiência Espectral	78
4.3.2	Índice de Justiça de Jain	83
4.3.3	Eficiência Energética	83
4.3.4	Área sobre as Curvas SE e EE	85
4.3.5	Eficiência de Recurso em função do carregamento e o compro- misso SE-EE	87
4.4	Conclusão do Capítulo	89
5	ACESSO ALEATÓRIO COM CONCESSÃO PARA SISTE- MAS NOMA XL-MIMO	92
5.1	Protocolo NVR-XL	92
5.2	Resultados Numéricos	98

5.2.1	Soma das Taxas do Sistema	98
5.2.2	Quantidade média e Fração de falhas em tentativas de acesso	101
5.2.3	Quantidade normalizada de usuários aceitos	101
5.2.4	Resolução de colisões	103
5.2.5	<i>Overhead</i> , atraso no tempo de acesso e Complexidade	103
5.3	Conclusão do capítulo	105
6	CONCLUSÕES E TRABALHOS FUTUROS	106
6.1	Conclusões Gerais	106
6.2	Conclusões I - Introdução e fundamentação conceitual	106
6.3	Conclusões II - Sistema mMIMO-NOMA	106
6.4	Conclusões III - MIMO Ultra Massivo	106
6.5	Conclusões IV - Eficiência de recurso em mMIMO-NOMA sob diferentes políticas de alocação de potência	107
6.6	Conclusões V - Acesso Aleatório com concessão baseado em NOMA XL-MIMO	108
6.7	Perspectivas e trabalhos futuros	108
	REFERÊNCIAS	109
	ANEXOS	116
	ANEXO A – MASSIVE MIMO AND NOMA BITS-PER- ANTENNA EFFICIENCY UNDER POWER ALLOCATION POLICIES	117
	ANEXO B – ON THE SE AND EE IN MASSIVE MIMO AND NOMA	130
	ANEXO C – NOMA-BASED RANDOM ACCESS IN MMTC XL-MIMO	136
	ANEXO D – CROWDED MTC RANDOM ACCESS IN NOMA XL-MIMO	148
	ANEXO E – IMPROVING RANDOM ACCESS WITH NOMA IN MMTC XL-MIMO	153

1 Introdução e Fundamentação Conceitual

Conectar o crescente número de dispositivos à rede mundial de computadores, fornecendo diferentes tipo de serviço com impactos nas áreas de saúde, educação, segurança, transporte e na indústria é um dos principais desafios das redes modernas de comunicação sem fio. Analisando a evolução histórica destas redes é possível constatar que as técnicas de acesso múltiplo são os principais alicerces para essa evolução.

A rede de comunicação móvel de primeira geração (1G) que surgiu na década de 1980, acomodava múltiplos usuários com serviços de voz com modulação em frequência (FM – *frequency modulation*) analógica via múltiplo acesso por divisão de frequência (FDMA – *frequency division multiple access*) (Vannithamby; Talwar, 2017). A digitalização das transmissões sem fio se inicia na rede de comunicação móvel de segunda geração (2G) na década seguinte e tendo como fundamental avanço a aplicação da técnica de múltiplo acesso por divisão de tempo (TDMA – *time division multiple access*), com o passar dos anos, na virada do século, surge a necessidade de entregar vídeo e voz via rede sem fio, o que proporcionou a evolução para a rede de comunicação móvel de terceira geração (3G) que operou predominantemente com múltiplo acesso por divisão de código (CDMA – *code division multiple access*) (Liu et al., 2020). Para viabilizar aplicações de *internet* em dispositivos móveis surgiram no final da primeira década dos anos 2000, as redes de comunicação móveis de quarta geração (4G), as quais entre outros avanços, utilizam (ambas as tecnologias LTE e WiMAX) a multiplexação por divisão de frequência ortogonal (OFDM – *orthogonal frequency division multiplexing*) (Liu; Jiang, 2016). Estas técnicas compartilham, de modo a explorar a ortogonalidade de algum dos recursos de rede disponíveis (tempo, frequência ou código), reduzindo a interferência e possibilitando receptores mais eficientes e de baixa complexidade. Esse grupo de técnicas de múltiplo acesso é classificado como múltiplo acesso ortogonal (OMA – *orthogonal multiple access*).

Nesta sucinta cronologia é possível constatar que cada geração de comunicação móvel possui um ciclo de aproximadamente 10 anos. Em 2023, a rede de comunicação móvel de quinta geração (5G) tornou-se realidade em muitos países ao redor do mundo; nessas redes, cenários típicos projetados já estão em operação. As novas redes sem fio estão padronizados em três modos de uso: a) banda larga móvel aprimorada (eMBB – *enhanced mobile broadband*), b) comunicação de tipo de máquina massiva (mMTC – *massive machine type communication*) e c) comunicação ultra confiável de baixa latência (uRLLC – *ultra reliable low latency communication*) (Hong et al., 2022). Com isso, novos desafios têm surgido na implementação e evolução de cenários de aplicação, caracterizando as redes de comunicações móveis além da 5G (B5G – *beyond 5G*).

Outro aspecto a ser considerado é que as técnicas OMA descritas até aqui não são suficientes para atender os novos requisitos de evolução dessas redes sem fio; por exemplo, a alta conectividade, devido à limitação dos recursos ortogonais disponíveis. Para suprir essa deficiência, técnicas de múltiplo acesso não ortogonais (NOMA – *non orthogonal multiple access*) estão sendo estudadas e obtendo resultados promissores.

Nesta Tese, destacamos o NOMA no domínio da potência que foi proposto como alternativa ao LTE adotado nos sistemas 4G (Saito et al., 2013; Benjebbour et al., 2013; Benjebbour et al., 2014; Benjebbour et al., 2015). Este esquema de transmissão consiste em multiplexar usuários no domínio da potência e aplicar o cancelamento de interferência sucessivo (SIC – *successive interference cancellation*) na recepção para realizar a decodificação dos sinais de múltiplos usuários com potências desbalanceadas. Esta técnica apresenta melhoria significativa na soma das taxas dos usuários quando comparadas às técnicas OMA, além de acomodar mais de um usuários na mesma porção de banda de comunicação.

Na Figura 1.1 classificamos as principais técnicas de múltiplo acesso propostas na literatura, dentro do contexto desta Tese, ou seja, um ambiente com número massivo de antenas. Quando o número de antenas na estação rádio-base (BS – *base station*) cresce assintoticamente, os vetores de canal para diferentes UEs tornam-se mutuamente ortogonais. Neste contexto, a estratégia de múltiplo acesso por divisão espacial (SDMA – *spatial division multiple access*) pode ser classificada como um tipo de OMA, já que os sinais de diferentes UEs podem ser multiplexados espacialmente sem que haja interferência mútua. As técnicas são brevemente discutidas nas seções subsequentes, com especial ênfase na técnica NOMA no domínio das potências (PD-NOMA – *power domain NOMA*), foco desta Tese.

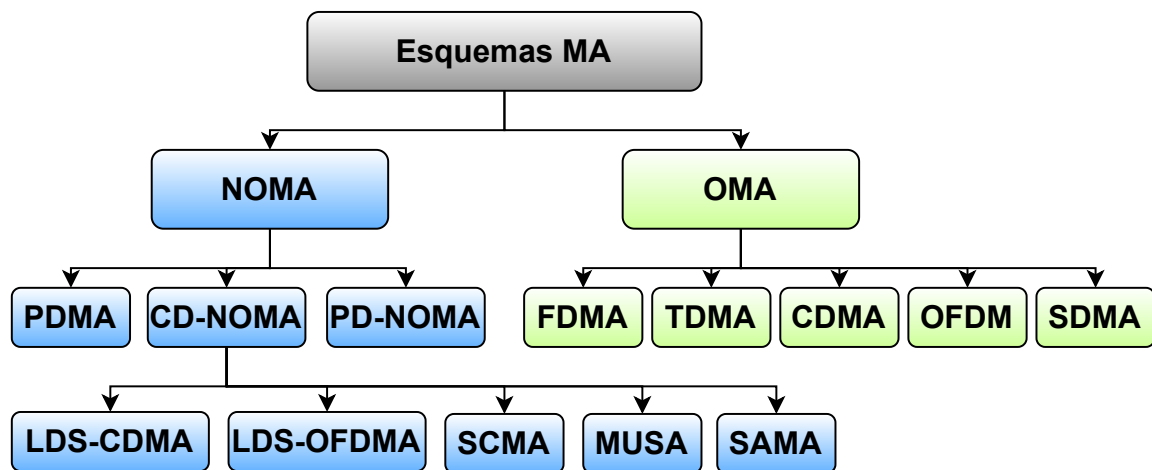


Figura 1.1 – Classificação dos esquemas de múltiplo acesso.

1.1 NOMA no Domínio da Potência (PD-NOMA)

A motivação para estudar as técnicas NOMA surge da possibilidade de melhoria na soma das taxas do sistema e pela capacidade intrínseca de acomodar mais de um usuário no mesmo bloco de recurso (RB – *resource block*).

Uma das técnicas NOMA em destaque nesta Tese é a de multiplexação no domínio da potência, ou seja, neste esquema é realizada a transmissão de sinais distintos com diferentes níveis de potência, explorando os coeficientes de canal distintos dos dispositivos (UE – *user equipment*), em um mesmo RB (tempo, frequência, código ou espaço). Para a detecção e decodificação multiusuário no receptor PD-NOMA deve-se adotar o SIC.

Um modelo de sistema PD-NOMA conceitual é descrito a seguir. Para exemplificar a descrição dos sinais envolvidos neste esquema, examinaremos o *downlink* (DL) do PD-NOMA, no qual uma BS equipada com uma antena ($M = 1$), K usuários estão presentes na célula e podem ser divididos em L *clusters* com largura de banda disponível B .

Por simplicidade, assumimos nesta descrição que a BS transmite para um único *cluster* ($L = 1$), os sinais contendo as informações $x_k \in \mathbb{C}$, destinados para cada um dos K UEs, sobrepostos no DL, sendo $\mathbb{E}[|x_k|^2] = 1$. Este sinal pode ser descrito como:

$$x = \sum_{k=1}^K \sqrt{\alpha_k P_T} x_k, \quad (1.1)$$

sendo P_T a potência total disponível para rádio frequência na BS e $\alpha_k P_T$ a fração de potência alocada para o k -ésimo UE, com $\sum_{k=1}^K \alpha_k = 1$. Este sinal x , Eq. (1.1), é recebido pelo k -ésimo UE presente na célula (e *cluster*), sendo descrito como:

$$y_k = h_k x + n_k, \quad \forall k, \quad (1.2)$$

sendo $h_k \in \mathbb{C}$ o coeficiente de canal complexo entre UEs e a BS, formado pelas parcelas de larga escala (perda de percurso/*path loss*) e pequena escala (desvanecimento/*fading*)¹ e n_k representa o ruído Gaussiano com densidade espectral de potência $N_{0,k}$. Destaque-se que para o PD-NOMA com SIC operar adequadamente, três condições devem ser atendidas:

1. Cada UE receptor deve ser capaz de processar o sinal recebido satisfazendo seguinte condição:

$$P_{\min} \leq \alpha_k P_T |h_k| \leq P_{\max}, \quad (1.3)$$

com $P_{\min}$ representando a potência de sensibilidade do receptor, $\alpha_k P_T$ a fração de potência definida para transmissão, P_T a potência total disponível para rádio frequência no transmissor e $P_{\max}$ sendo a potência máxima possível de ser processada antes da saturação no receptor. Com isso, podemos definir a faixa dinâmica de operação do SIC como $D^{\text{sic}} = P_{\max} - P_{\min}$.

¹A Seção 1.4.2 trata exclusivamente sobre o modelo de canal utilizada nesta Tese.

2. Os ganhos de canal dos K UEs que formam o conjunto $\mathcal{K}$ precisam ser previamente conhecidos e ordenados de modo que o UE₁ possua o melhor ganho (condição) de canal h_1 e o UE _{K} o pior coeficiente canal h_K , este ordenamento é utilizado no procedimento de clusterização dos usuários:

$$|h_1| > |h_k| > \dots > |h_{K-1}| > |h_K|. \quad (1.4)$$

3. Explorando o coeficiente de canal dos UEs através do procedimento de clusterização, detalhado na Subseção 2.2, os usuários são agrupados baseado em algum critério de disparidade do coeficiente de canal, sendo a potência total disponível adequadamente distribuída entre os ζ UEs do mesmo *cluster*, de modo que exista uma relação mínima entre as potências, Ξ , ou seja, entre a potência do sinal recebido interferente predominante e a soma das potências dos demais usuários, levando a em conta a faixa dinâmica de operação do SIC, D^{sic} . Supondo $\zeta = 3$ UEs por *cluster* é necessário atender as condições expressa pelas seguintes razões de potência, no UE₁:

$$\begin{aligned} \frac{\alpha_3 P_T}{(\alpha_1 P_T + \alpha_2 P_T)} &\geq \Xi, \quad 1^{\text{a}} \text{ etapa SIC} \\ \frac{\alpha_2 P_T}{\alpha_1 P_T} &\geq \Xi, \quad 2^{\text{a}} \text{ etapa SIC} \end{aligned} \quad (1.5)$$

sendo esta disparidade de potência relativa Ξ adimensional e necessária para o sucesso da aplicação do SIC, discutido na Subseção 1.1.3.

Observe que ao combinarmos Eq. (1.5) com a condição de ordenamento dos canais no *cluster*, Eq. (1.4), a ordem de decodificação/cancelamento não é alterada, i.e., o UE₁ realiza 1^a e 2^a etapas de cancelamento SIC (UE₂ e UE₃); o UE₂ realiza apenas o cancelamento do sinal UE₃ e considera o sinal do UE₁ como ruído; e o UE₃ trata os sinais da UE₁ e UE₂ como ruído. Isto se deve ao fato do ganho do canal ser dominado, na média pelo termo longo (*path loss*).

1.1.1 Desvantagens de sistemas PD-NOMA

Vale ressaltar que os sistemas PD-NOMA possuem algumas desvantagens, tais como: complexidade de *hardware* (incluindo SIC), dependência de conhecimento preciso da condição do canal (CSI – *channel state information*) no transmissor e receptor, design do receptor e política de alocação de potência e de piloto (Makki et al., 2020; Maraqa et al., 2020; Zafari et al., 2019).

A aplicação do PD-NOMA com o objetivo de acomodar múltiplos UEs dentro do mesmo RB, operando no *downlink*, exige que a capacidade de processamento dos dispositivos seja elevada. Acreditamos que devido ao avanço contínuo da tecnologia e ao

aumento da capacidade de processamento dos UEs, as próximas gerações de UEs serão capazes de processar o SIC (Maraqa et al., 2020).

A dependência do NOMA em conhecer precisamente a condição do canal é uma desvantagem de sua aplicação. Técnicas avançadas de estimação precisam ser implementadas, mas dependem de sinais piloto, o que torna o *overhead* excessivo à comunicação e reduz substancialmente a SE (Zafari et al., 2019).

A alocação de potência ótima em termo da SE para o sistema NOMA depende da formação do *cluster* e da função objetivo de cada sistema (Maraqa et al., 2020). Ao longo desta Tese, tendo o objetivo explorar o potencial da técnica, NOMA, adotamos algumas simplificações.

1.1.2 Cancelamento de interferência sucessivo – SIC

O SIC é um recurso de processamento de sinais da camada física bem conhecido que permite a um receptor reconstruir pacotes de dados de diferentes usuários que chegam simultaneamente e estão sobrepostos na mesma banda e operando em banda base, (Verdu, 1998), ou seja, é a capacidade de um receptor detectar, tratar e decodificar dois ou mais sinais sobrepostos no tempo e na frequência. O SIC é viável desde que os sinais a serem detectados, reconstruídos e cancelados sejam recebidos no mesmo receptor com consideráveis disparidades de potências entre si e dentro da faixa dinâmica de operação do SIC. Deste modo, o receptor poderá decodificar o sinal mais forte, reconstruí-lo, subtraí-lo do sinal original com sobreposição e então extrair o sinal mais fraco do resíduo. A Figura 1.2 apresenta o UE₁ recebendo dois sinais sobrepostos com potência de transmissão adequadamente distintas e aplicando o SIC.

O UE₁ demodula o sinal recebido e reconstrói uma versão estimada do sinal recebido $\tilde{x}_2$ (interferente), recebido com maior potência, utiliza essa versão estimada para cancelar o sinal interferente do sinal recebido e reconstrói uma versão estimada do sinal $\tilde{x}_1$ (interesse), o cancelamento será perfeito a caso disparidade de potência relativa Ξ seja alcançada, caso contrário, teremos ϖ representa a parcela do sinal x_2 que não foi cancelada corretamente, também chamada resíduo e $\tilde{n}_1$ é o ruído Gaussiano com densidade espectral de potência $\tilde{N}_{0,1}$. Para que o processo SIC opere com sucesso, o resíduo deve ser mantido em níveis muito baixos

Este processo se repete até que o sinal do usuário mais fraco (maior beneficiado neste processo) seja reconstruído. É comum apenas o sinal mais forte, dentro do sinal recebido, ser decodificado, tratando restante do sinal como interferência. No entanto, o SIC propicia a recuperação até mesmo do sinal mais fraco. Para isso, os bits do sinal mais forte são decodificados inicialmente. Este sinal, o mais forte, é então reconstruído a partir desses bits e subtraído (ou seja, cancelado) do sinal combinado (recebido). Os bits do pacote mais fraco são então decodificados desse resíduo. Isso pode ser um processo iterativo para recuperar vários pacotes de diferentes usuários sucessivamente, ordenados em potências recebidas decrescentes; e portanto, é denominado cancelamento de interferência sucessivo.

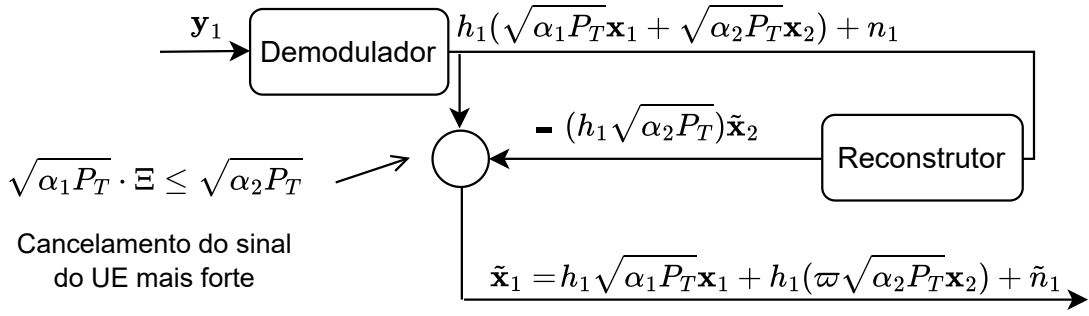


Figura 1.2 – Receptor SIC do UE₁ no DL PD-NOMA com 2 usuários.

Quantificar quão mais forte deve ser o sinal para ser decodificado, ou qual a diferença de potência entre os sinais, Eq. (1.5), não é o tema central desta Tese, mas sabemos que, quanto maior a diferença entre o sinal mais forte e a soma dos demais sinais interferentes mais eficiente será o cancelamento da interferência (menor o resíduo). A presença do resíduo após o SIC, $0 \leq \varpi \leq 1$, se deve ao cancelamento imperfeito do sinal interferente.

No trabalho desenvolvido por (Ramatryana; Shin, 2022) o SIC perfeito é aplicado em um modelo de sistema NOMA onde a relação sinal-ruído mais interferência (SINR – *signal interference plus noise ratio*) do sinal mais forte, sendo este sinal a ser decodificado primeiro e depois subtraído do sinal recebido, é o dobro da SINR dos sinais interferentes mais o ruído.

1.1.3 Cancelamento de interferência paralelo – PIC

O cancelamento de interferência paralelo (PIC – *parallel interference cancellation*) (Varanasi; Aazhang, 1990) é uma estrutura de cancelamento de interferência que opera de forma diferente do SIC. Está técnica foi empregada no CDMA nos sistemas onde existe controle de potência para transmissão dos sinais aos usuários ativos. No CDMA cada sinal é transmitido utilizando uma sequência de espalhamento diferente, e cada dispositivo receptor, conhecendo essas sequências, pode decodificar/reconstruir e cancelar os sinais interferentes. Este tipo de cancelador opera em estágios (dois ou mais estágios - PIC multiestágio), sendo que, no primeiro estágio são geradas estimativas dos sinais $\hat{x}_k$ dos K UEs e nos demais N estágios ocorrem o cancelamento paralelo dos sinais interferentes, Figura 1.3.

Deste modo, toda interferência, gerada por outros usuários é cancelada paralelamente a cada estágio e o sinal de interesse pode ser reconstruído. No contexto NOMA não é comum a aplicação da técnica PIC, no entanto é possível adapta-la para o contexto NOMA, assumindo algumas hipóteses, como a de incluir no sinal transmitido, x^{PIC} , uma sequências de espalhamento quasi-ortogonais para cada UE, $\mathbf{s}_k$, deste modo, podemos utilizar o PIC no contexto NOMA e realizamos uma análise comparativa descrita na Seção 4 desta Tese.

Com isso, descrevemos o sinal transmitido como:

$$x^{\text{PIC}} = \sum_{k=1}^K \sqrt{\alpha_k P_T} \mathbf{s}_k x_k, \quad (1.6)$$

sendo P_T a potência total disponível para rádio frequência na BS e $\alpha_k P_T$ a fração de potência alocada para o k -ésimo UE, com $\sum_{k=1}^K \alpha_k = 1$. A Figura 1.3 demonstra a operação de um receptor PIC para K sinais sobreposto recebidos potência aproximadamente iguais, neste caso, podemos representar o sinal recebido pelo UE₁ como:

$$y_1^{\text{PIC}} = h_1 \sum_{k=1}^K \sqrt{\alpha_k P_T} \mathbf{s}_k x_k + n_1, \quad (1.7)$$

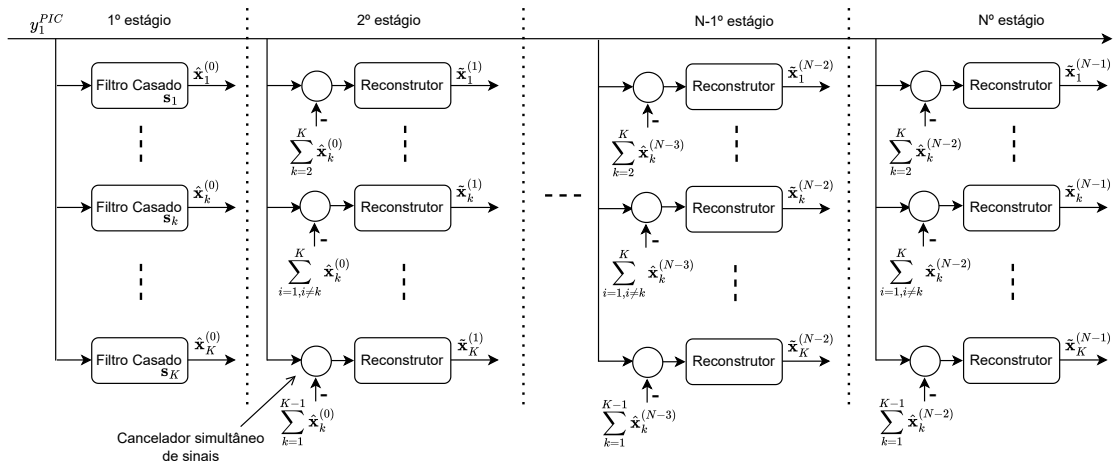


Figura 1.3 – Receptor PIC multiestágio do UE₁ no DL PD-NOMA com K usuários: os sinais dos K são estimados no primeiro estágio do cancelador, nos estágios seguintes os sinais interferentes são cancelados e uma versão estimada com mais qualidade do sinal de interesse é produzida.

1.1.4 Exemplo – DL PD-NOMA

Considerando a descrição da Subseção 1.1, nesta Subseção demonstramos o funcionamento do *downlink* PD-NOMA em uma célula onde a BS possui uma única antena e estão presentes 2 UEs, também equipados com uma única antena, e com ganhos de canal distintos, definidos pelo *pathloss*. Neste cenário a BS transmite o sinal x , utilizando o mesmo RB disponível, para os 2 UEs sobrepostos com potências de transmissão adequadamente alocadas e portanto distintas. Deste modo, os UEs recebem seus sinais de interesse adicionados da interferência causada pelo sinal do outro UE mais o ruído, conforme Eq. (1.2).

No PD-NOMA, as maiores frações de potência disponíveis são alocadas para usuários com as piores condições de canal. Uma vez que o processo de cancelamento SIC é implementado no equipamento do usuário que sofre maior interferência, e levando-se

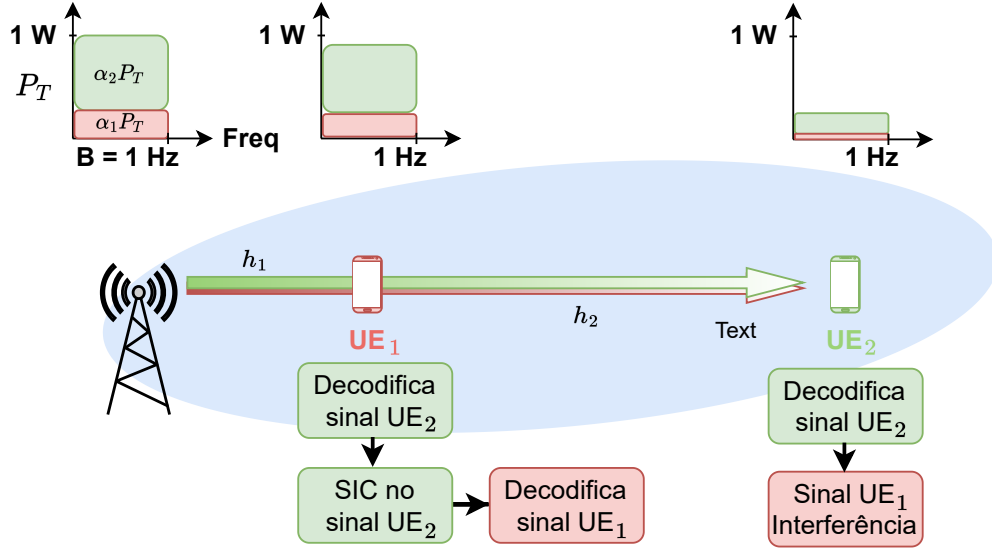


Figura 1.4 – Esquema ilustrando o DL PD-NOMA: UE₁ decodifica o sinal do UE₂ (interferente), aplica o SIC e decodifica seu sinal UE₁ (interesse), UE₂ decodifica seu sinal de interesse com a interferência do sinal do UE₁.

em consideração as 3 condições apresentadas na Subseção 1.1, a ordem de decodificação deve ser decrescente em relação às potências recebidas em cada usuário (Ali et al., 2018). Utilizando a Figura 1.4 como referência, podemos constatar que:

$$\alpha_2 P_T > \alpha_1 P_T, \quad (1.8)$$

sendo $\alpha_1 + \alpha_2 = 1$. Os sinais recebidos por cada usuário UE₁ e UE₂, com $|h_1| > |h_2|$, podem ser descritos como:

$$y_1 = h_1(\sqrt{\alpha_1 P_T} x_1 + \sqrt{\alpha_2 P_T} x_2) + n_1, \quad (1.9)$$

$$y_2 = h_2(\sqrt{\alpha_1 P_T} x_1 + \sqrt{\alpha_2 P_T} x_2) + n_2. \quad (1.10)$$

Seguindo a ordem de decodificação na Eq. (1.8), em nosso exemplo, o UE₁ primeiramente decodifica o sinal x_2 e o subtrai do sinal recebido y_1 . Com isso, o UE₁ pode decodificar o sinal de interesse x_1 sem a interferência do sinal x_2 . Entretanto, para a realização do SIC é necessário atender a condição 3:

$$\frac{(\sqrt{\alpha_2 P_T} |h_1|)^2}{(\sqrt{\alpha_1 P_T} |h_1|)^2} \Rightarrow \frac{\alpha_2}{\alpha_1} \geq \Xi. \quad (1.11)$$

Sendo assim, o UE₂ não realiza o cancelamento de interferência, já que o seu sinal de interesse x_2 possui potência predominante no sinal total y_2 , e decodifica o seu sinal considerando a interferência x_1 como ruído, pois é o primeiro na ordem de decodificação.

De acordo com a definição da capacidade de Shannon e assumindo cancelamento do sinal UE₂ no terminal UE₁, a taxa de dados do UE₁, $R_1^{\text{DL-NOMA}}$ pode ser descrita como:

$$R_1^{\text{DL-NOMA}} = B \log_2 \left(1 + \frac{\alpha_1 P_T |h_1|^2}{BN_{0,1}} \right). \quad (1.12)$$

Por sua vez, o terminal UE_2 não realiza o cancelamento da interferência (apenas dois usuários no *cluster*). O UE_2 decodifica diretamente o seu sinal de interesse x_2 do sinal recebido y_2 que contém o sinal interferente x_1 bastante atenuado devido a fração de potência alocada α_1 . Com isso, a equação da taxa $R_2^{\text{DL-NOMA}}$ pode ser descrita como:

$$R_2^{\text{DL-NOMA}} = B \log_2 \left(1 + \frac{\alpha_2 P_T |h_2|^2}{\alpha_1 P_T |h_2|^2 + B N_{0,2}} \right). \quad (1.13)$$

Observando as Eq. (1.12) e (1.13) é possível constatar que o controle das frações de potência de transmissão, α_1 e α_2 , afeta o desempenho do sistema, pois altera as taxas de transferência de dados de cada UE. Com isso, fica evidente que a soma das taxas no sistema PD-NOMA pode ser regulada otimizando a distribuição de potência.

Como forma de simples comparação, para demonstrar o ganho do PD-NOMA sobre a técnica OMA, utilizaremos o modelo de sistema da Figura 1.4, mas com a aplicação do FDMA. Sendo assim, assumindo ortogonalidade perfeita no OMA, a banda disponível normalizada ($0 \leq B \leq 1$) Hz é disponibilizada ao UE_1 e a banda restante $(1 - B)$ Hz é atribuída ao UE_2 . Com isso, as equações das taxas no OMA podem ser definidas como:

$$R_1^{\text{OMA}} = B \log_2 \left(1 + \frac{\alpha_1 P_T |h_1|^2}{B N_{0,1}} \right), \quad R_2^{\text{OMA}} = (1 - B) \log_2 \left(1 + \frac{\alpha_2 P_T |h_2|^2}{(1 - B) N_{0,2}} \right). \quad (1.14)$$

A seguir apresentamos alguns resultados de taxa alcançada pelos usuários UE_1 e UE_2 em determinadas condições de sistema. Os parâmetros utilizados nas simulações estão reunidos na tabela 1.1.

	NOMA	OMA
$ h_1 ^2/ h_2 ^2$	20 dB	
P_T	1 W	
α_1	$[0.01, 0.5[$	0.5
α_2	$1 - \alpha_1$	0.5
B	1 Hz	$[0, 1]$ Hz
Ξ	(3 dB; 6 dB; 10 dB)	—

Tabela 1.1 – Tabela de parâmetros para simulação NOMA vs OMA.

A Figura 1.5 apresenta o resultado de soma das taxas alcançadas pelos sistemas NOMA e OMA. No sistema NOMA quanto maior a potência relativa, Ξ , entre os sinais recebidos, menor será a faixa de operação do NOMA.

Na Figura 1.6 podemos observar as taxas individuais alcançada pelos dispositivos em operação em cada um dos sistemas. Destaque para as taxas alcançadas pelo usuário de borda do NOMA, sempre superiores quando comparadas ao OMA.

1.1.5 Exemplo – UL PD-NOMA

No UL PD-NOMA, o usuário k transmite o sinal x_k com potência de transmissão igualitária p_k . No cenário com dois UEs de antena única, sendo $p_1 = p_2 = p$, o sinal

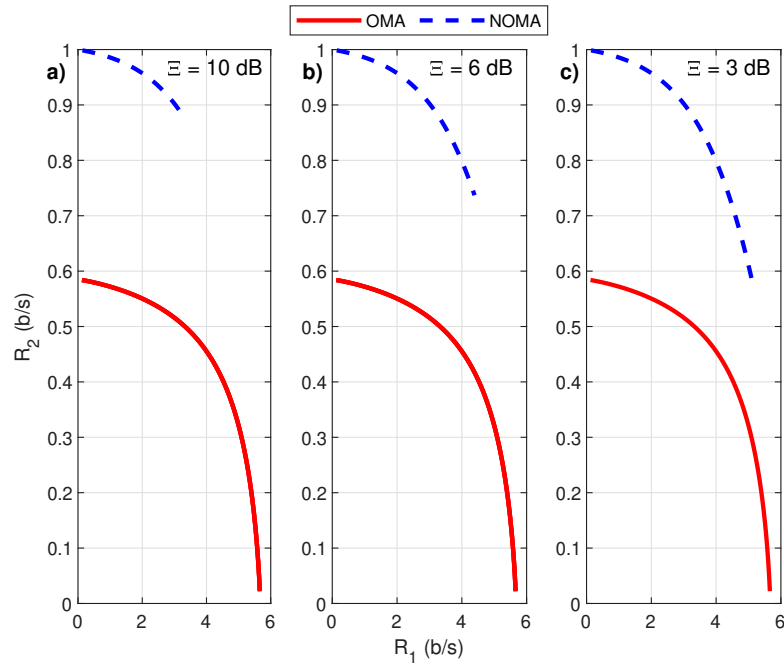


Figura 1.5 – Taxas alcançadas na comparação das técnicas NOMA e OMA no cenário da Figura 1.4. As figuras consideram o SIC perfeito, mas atendendo a Ξ em 3 níveis **a)** $\Xi = 10$ dB, **b)** $\Xi = 6$ dB e **c)** $\Xi = 3$ dB.

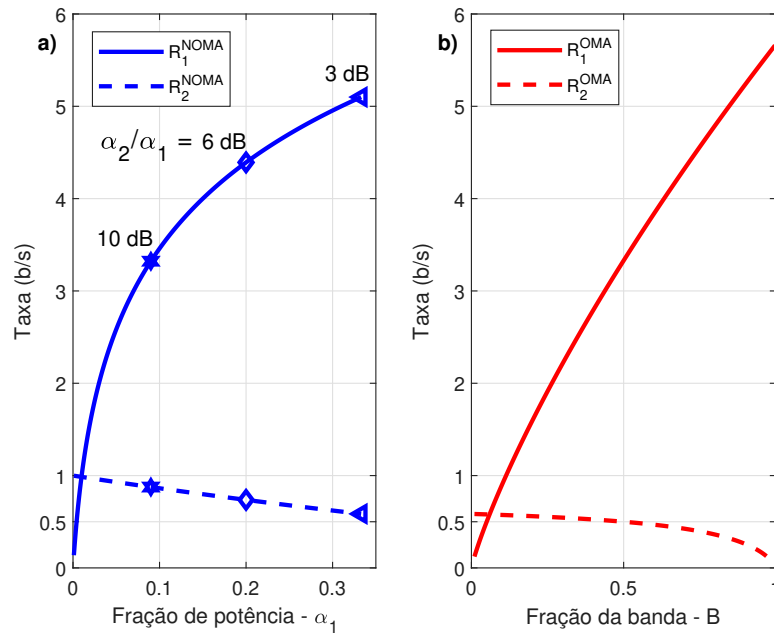


Figura 1.6 – Taxa alcançada por dispositivo UE₁ e UE₂, no cenário representado na Figura 1.4. **a)** Sistema NOMA indicando as Ξ . **b)** Sistema OMA.

recebido na BS de antenna única pode ser descrito como:

$$\mathbf{y} = \sqrt{p}h_1x_1 + \sqrt{p}h_2x_2 + n_0, \quad (1.15)$$

sendo n_0 representa amostras de ruído Gaussiano com densidade espectral de potência à entrada do receptor na BS igual a N_0 .

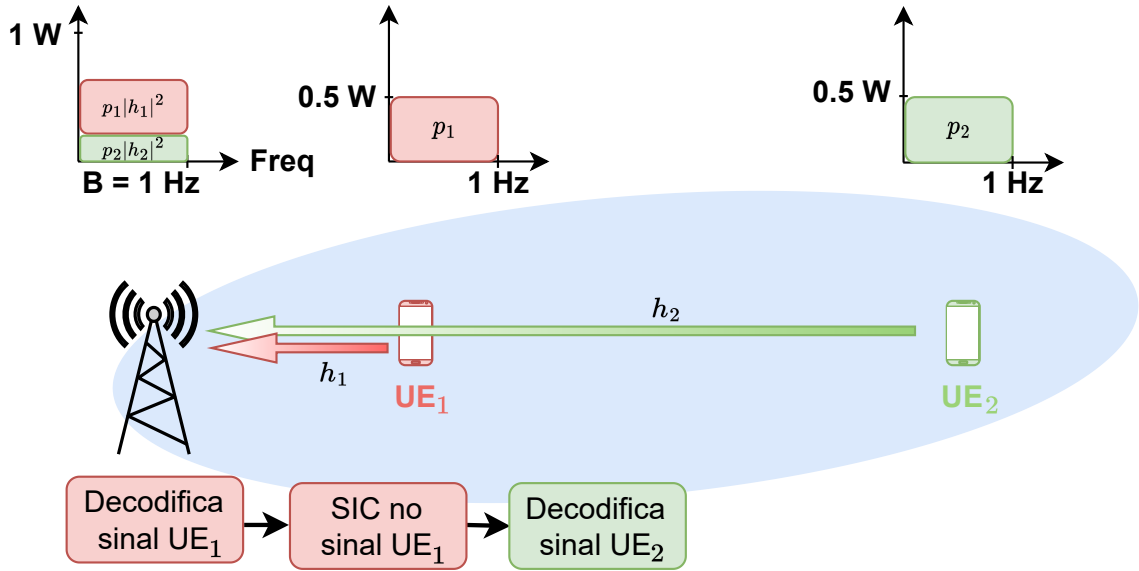


Figura 1.7 – Esquema representando o UL NOMA básico, UE_1 e UE_2 experimentam condições de canal distintas e transmitem com a mesma potência. A recepção ocorre na BS via receptor SIC.

No cenário representado na Figura 1.7, os sinais x_1 e x_2 são transmitidos usando a mesma banda do espectro, causando interferência mútua. No UL, o SIC é implementado no receptor da BS, respeitando uma P_{tol} , conforme a condição 2.

$$\frac{p|h_1|^2}{p|h_2|^2} \geq \Xi. \quad (1.16)$$

O receptor decodifica x_1 e x_2 em dois estágios, conforme representado na Figura 1.7. No primeiro estágio, o receptor decodifica o sinal predominante x_1 , e trata o sinal interferente x_2 como ruído térmico. Uma vez que o receptor decodifica x_1 , ele pode ser reconstruído e subtraído do sinal recebido y e então decodificar o x_2 . Como agora resta apenas o ruído Gaussiano no receptor da BS, a taxa de dados máxima com alocação de potência igualitária é alcançada pelo UE_2 . Neste caso, a equação das taxas pode ser descrita como:

$$R_1^{\text{UP-NOMA}} = \log_2 \left(1 + \frac{p_1|h_1|^2}{p_2|h_2|^2 + N_0} \right), \quad R_2^{\text{UP-NOMA}} = \log_2 \left(1 + \frac{p_2|h_2|^2}{N_0} \right). \quad (1.17)$$

Para o cenário UL OMA básico, sua operação ocorre de forma similar que no DL OMA e sua equação de capacidade pode ser considerada da mesma forma, ou seja, conforme Eq. (1.14). Supondo que a SNR dos sinais recebidos sejam $p_1|h_1|^2/N_0 = 20$

dB e $p_2|h_2|/N_0 = 0$ dB, a capacidade alcançada no NOMA é, $R_1^{\text{UP-NOMA}} = 5.67$ (b/s) e $R_2^{\text{UP-NOMA}} = 1.00$ (b/s). Em contrapartida, a capacidade obtida do OMA é, $R_1^{\text{OMA}} = 3.70$ (b/s) e $R_2^{\text{OMA}} = 0.80$ (b/s). Tendo como base a análise acima, o PD-NOMA *uplink* e *downlink* tem potencial de produzir melhores resultados em termos de soma das taxas que as técnicas OMA.

1.2 Domínio do Código – CD-NOMA

Para o múltiplo acesso no domínio do código (CD-NOMA – *code domain-NOMA*), diferentes usuários são alocados com diferentes sequências de códigos Gaussianos aleatórios e são multiplexados no mesmo RB e explora a detecção compressiva no receptor para maximizar a detecção do usuário e minimizar as taxas de erro de símbolo (Duchemin et al., 2018). O CD-NOMA é classificado em várias subcategorias, devido às variações de sua aplicação, entre as quais destacam-se:

LDS-CDMA: este esquema substitui sequências de espalhamento tradicionais por sequências de espalhamento de baixa densidade (LDS – *low-density spreading*). O esquema LDS-CDMA foi primeiramente discutido em (Hoshyar et al., 2008).

LDS-OFDMA: tal esquema foi proposto por (Al-Imari et al., 2011), combinando o LDS-CDMA com múltiplo acesso por divisão de frequências ortogonais (OFDMA – *orthogonal frequency division multiple access*); com isso os símbolos de cada usuário são espalhados por subportadoras selecionadas de forma inteligente e sobrepostos no domínio da frequência.

SCMA – Acesso Múltiplo de Código Esperso – (*sparse code multiple access*): aplicação de um caso especial do LDS no CDMA foi proposto por (Nikopour; Baligh, 2013).

MUSA – Acesso Compartilhado Multiusuário – (*multi user shared access*): outra aplicação de LDS desta vez no chamado CDMA melhorado (onde há baixa correlação cruzada entre as sequências de espalhamento, que facilitam o cancelamento de interferência), garante justiça (*fairness*) entre os usuários e não há perda de capacidade (Yuan et al., 2016).

SAMA – Acesso múltiplo assistido por SIC – (*SIC aided multiple access*): esta técnica determina as matrizes de espalhamento para facilitar cancelamento de interferência entre os usuários. No SAMA, o símbolo do primeiro usuário é o mais importante e pode ser determinado em algumas iterações. É vantajoso porque pode convergir detectando os símbolos de diferentes usuários tendo ordens de diversidade mais baixas (Dai et al., 2014).

1.3 Divisão de Padrão – PDMA

A técnica de acesso múltiplo por divisão de padrão (PDMA – *pattern division multiple access*) inicialmente proposto por (Zeng et al., 2015), apresenta vantagens significativas, entre elas: a) supressão de interferência co-canal; b) atinge baixo consumo de potência; c) alta eficiência espectral; d) baixa latência.

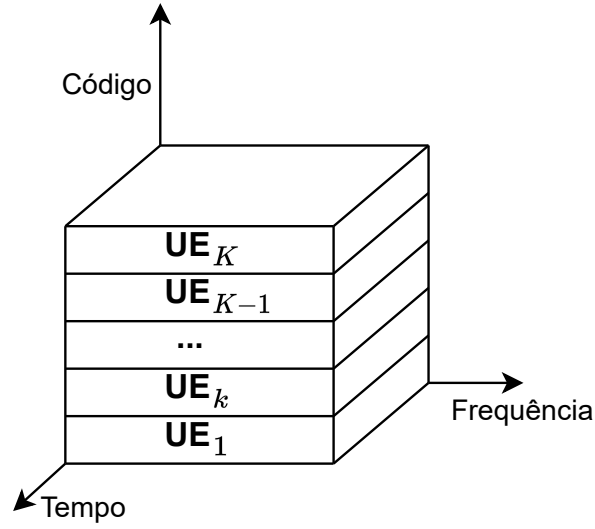


Figura 1.8 – Esquema representando CD-NOMA básico, UEs compartilham o mesmo RB (tempo x frequência), com sequência de espalhamento (código) diferentes.

O PDMA foi recentemente introduzido nos cenários 5G e B5G (Dai et al., 2018; Li et al., 2021). Sua principal diferença em relação ao PD-NOMA é a adoção de segmentação de padrões para separar os sinais dos usuários, mas tem em comum a dependência do SIC no receptor (Li et al., 2021). O modo de operação do PDMA, utiliza diferentes domínios de recurso (Jiang et al., 2019; Sun et al., 2020) para multiplexar os sinais dos dispositivos.

Para exemplificar a aplicação desta técnica, representamos a seguir um esquema de 6 UEs, compartilhando 4 RB multiplexados nos domínios espacial/potência em cada RB.

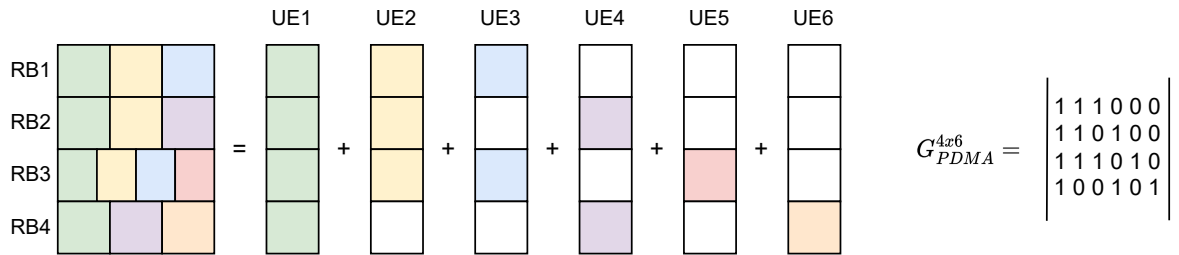


Figura 1.9 – Esquema representando PDMA básico, onde 6 UEs compartilham o mesmo RB (tempo x frequência), com potência e divisão espacial (*beamforming*) diferentes.

No PDMA, o processo de modulação torna-se simples, mas o escalonamento de potência e a mudança de fase precisam ser estudados com antecedência, formando uma matriz padrão. Assim, o valor '1' na matriz de padrão PDMA é substituído por um valor complexo que reflete a escala de potência e a mudança de fase dos sinais de transmissão (Sun et al., 2020).

1.4 MIMO massivo

Nesta Seção apresentamos conceitos fundamentais associados à operação do sistema mMIMO unicelular convencional, incluindo características e benefícios práticos desta técnica que se tornou essencial para os sistemas de comunicação sem fio 5G.

1.4.1 Modelo de sistema mMIMO

Assumimos o modelo mais simples, sendo uma única célula representada na Figura 1.10, sendo uma BS está equipada com M antenas servindo K usuários (ou dispositivos) com uma única antena cada. Adotamos o bloco de coerência do canal τ_c ² como um bloco de tempo-frequência grande o suficiente, de modo que podemos considerar aproximadamente constante o ganho de canal no período de símbolo entre a m -ésima antena da BS e o k -ésimo usuário neste intervalo.

O canal entre uma das antenas da BS m e a antena do usuário k que abrange a propagação multi-percurso (espalhamento, difusão e reflexão (Rappaport, 2002; Heath Jr; Lozano, 2018)), atenuação e sombreamento, pode ser representado por dois valores complexos (Goldsmith, 2005), um quando a comunicação é no sentido descendente, ou seja, a BS transmite e o UE recebe - *downlink* e outro quando a comunicação é no sentido ascendente, ou seja, o UE transmite e a BS recebe - *uplink*. Agrupamos as amostras do canal em vetores e chamamos de vetores de canal de *downlink* $\mathbf{h}_k^{\text{DL}} \in \mathbb{C}^M$ e *uplink* $\mathbf{h}_k^{\text{UL}} \in \mathbb{C}^M$ do UE _{k} .

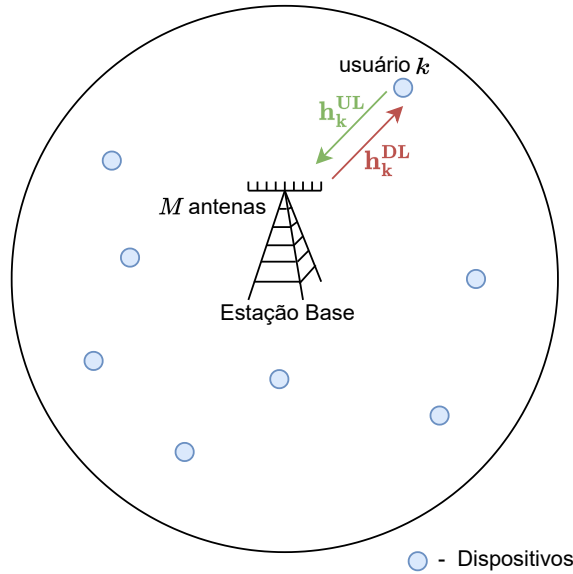


Figura 1.10 – Exemplo de célula única mMIMO com M antenas e K dispositivos.

O mMIMO convencional opera no modo duplexagem por divisão de tempo (TDD – *time division duplex*), o que quer dizer que transmissões no *downlink* e no *uplink* são

²conforme descrito na Subseção 1.4.3

separadas no tempo, que é diferente do TDMA, onde a informação de diferentes usuários são temporalmente separadas. A outra opção de operação é o modo duplexagem por divisão de frequência (FDD – *frequency division duplex*), onde as transmissões são individualizadas na frequência. Adotamos o TDD como padrão nesta Tese, pois é a estratégia de acesso mMIMO mais difundida na literatura.

1.4.2 Modelagem do Canal sem fio

Sabemos que o meio de propagação do sinal de comunicação é de fundamental relevância e determina o potencial do sistema de comunicação. A modelagem precisa do canal é imprescindível ao projeto de sistemas de comunicação confiáveis.

Na inviabilidade de realizar experimentos com os sistemas de comunicação no mundo real, os quais muitas vezes ainda não estão disponíveis ou mesmo operacionais, nós utilizamos os modelos de canal como ferramenta matemática e de simulação para avaliarmos o desempenho, a confiabilidade e a potencialidade desses sistemas. Com modelos de canal que descrevem com precisão diferentes cenários de propagação física do canal móvel, podemos projetar algoritmos de processamento de sinais que alcancem melhores resultados.

Em um canal multi-percurso, várias cópias atenuadas e atrasadas de um sinal transmitido chegarão ao receptor. Como o canal está em constante mudança devido a objetos que se movem no ambiente de propagação, não é possível rastrear todos os atrasos e atenuação do canal; portanto, é comum empregar um modelo estatístico do canal. Neste contexto, assumimos que as estatísticas do canal móvel serão constantes por um tempo muito maior que tempo de coerência do canal, permitindo assim avaliarmos o desempenho médio do sistema sujeito àquele tipo de canal.

Neste trabalho de Tese utilizamos uma das modelagens mais comum e amplamente aceita para a descrição estocástica do canal; o modelo de desvanecimento de canal *Rayleigh* que é baseado na teoria do limite central (Kobayashi et al., 2012), ou seja, a soma de vários percursos do sinal independentes e aleatórios deve aproximar-se de uma distribuição normal ou Gaussiana complexa (Björnson et al., 2017b). O canal de desvanecimento *Rayleigh* recebe esse nome porque o canal induz uma amplitude que segue uma distribuição estatística de *Rayleigh* (Rappaport, 2002).

No modelo de desvanecimento *Rayleigh* descorrelacionado (sem correlação espacial, temporal ou em frequência), cada bloco de coerência do canal segue uma distribuição Gaussiana complexa circularmente simétrica. Portanto, para o usuário k , seu canal é distribuído como:

$$\mathbf{h}_k = \mathcal{CN}(0, \beta_k \mathbf{I}_k), \quad (1.18)$$

sendo β_k o desvanecimento em larga escala que representa a atenuação do canal do k -ésimo

usuário e dependente da distância entre o k -ésimo usuário e a BS. Quando o dispositivo está na região *far-field* o mesmo β_k pode ser considerado para todas as antenas do arranjo mMIMO ou XL-MIMO.

Neste trabalho modelamos o desvanecimento em larga escala, ou seja, a perda de percurso de duas formas, a primeira considerando somente a parcela determinística:

$$\beta_k = \frac{\beta_0}{d_k^\xi}. \quad (1.19)$$

Em decibel, tem-se: $\beta_k^{[dB]} = \beta_0^{[dB]} - 10 \cdot \xi \cdot \log_{10}(d_k)$ [dB] sendo β_0 a perda de percurso na distância de referência $d_0 = 1m$, ξ é o coeficiente de percurso e d_k a distância do usuário até o arranjo de antenas em metros. O outro modelo utilizado neste trabalho, baseado no modelo de microcenário urbano disponível em (3GPP, 2018), considera também o sombreamento, que é estocástico:

$$\beta_k^{[dB]} = \beta_0^{[dB]} + \varphi^{[dB]} - 10 \cdot \xi \cdot \log_{10}(d_k) \text{ [dB]}, \quad (1.20)$$

sendo $\varphi \sim \mathcal{N}(0, \sigma_{sf}^2)$ a perda por sombreamento que segue uma distribuição aleatória log-normal com variância σ_{sf}^2 .

Consideramos no contexto desta Tese, que não existe correlação entre as antenas, deste modo buscamos explorar todo o potencial do PD-NOMA de modo a colaborar com o modelo de trabalho exploratório desta Tese.

1.4.3 Bloco de coerência do canal

Neste trabalho, adotamos o modelo de canal por desvanecimento em bloco; assumimos que as amostras de canal permanecem constante por um intervalo de tempo-frequência, identificado pelo bloco de coerência ou intervalo de coerência do canal. O bloco de coerência do canal é o produto de duas quantidades, o tempo de coerência do canal, que é o intervalo de tempo onde podemos considerar o canal como invariante no tempo, e a banda de coerência do canal, que é o intervalo do espectro de frequências onde podemos considerar o canal plano em frequência (Björnson et al., 2017b; Marzetta et al., 2016; Goldsmith, 2005).

O tempo de coerência T_c de um canal depende da propagação Doppler do canal, que por sua vez é proporcional ao comprimento de onda e inversamente proporcional à velocidade do transmissor (Goldsmith, 2005). Na prática significa que a velocidade de movimentação do usuário está inversamente ligada a duração do bloco de coerência do canal. Assim, um terminal em deslocamento com um usuário caminhando possui um tempo de coerência do canal maior que esse mesmo dispositivo se movimentando dentro de um carro na estrada.

A largura de banda de coerência B_c de um canal descreve o intervalo de frequência onde a resposta do canal é aproximadamente constante (Björnson et al., 2017b), a qual

depende da diferença entre o caminho mais curto e o mais longo dentro do ambiente de propagação, associado ao espalhamento temporal do canal.

O tamanho do bloco de coerência do canal é um produto do tempo de coerência do canal e da largura de banda de coerência do canal, $\tau_c = T_c B_c$, e é diferente para cada usuário. Como precisamos de um bloco de coerência do canal adequado para todo sistema, adotamos o menor intervalo detectado como padrão, sabendo que, cada link UE-BS muito provavelmente apresentará bloco de coerência do canal diferente, devido às diferenças de mobilidade. O tamanho do bloco é geralmente da ordem de centenas de símbolos de informação. No entanto, o tamanho pode aumentar para vários milhares em caso de usuários quase-estacionários ou estacionários (Björnson et al., 2017b; Marzetta et al., 2016).

1.4.4 Reciprocidade do canal

Adotamos o princípio da reciprocidade do canal em sistemas MIMO massivo operando no modo TDD. Este conceito foi fundamental para consolidar o mMIMO, peça-chave no desenvolvimento das novas tecnologias de rede 5G e B5G. Reciprocidade significa que o canal de *downlink* e *uplink* para um mesmo par de antenas (UE-BS) ou (BS-UE) respectivamente, são considerados simétricos ou equivalentes (Vieira et al., 2017; Björnson et al., 2017b), sendo mais precisa essa equivalência quando o tempo de coerência do canal é mais curto. Assim, quando representamos o vetor de canais entre o k -ésimo usuário e as antenas da BS, obtemos:

$$\mathbf{h}_k^{\text{UL}} \cong \mathbf{h}_k^{\text{DL}}, \quad (1.21)$$

isto é, são aproximadamente equivalentes e independentes. Ou seja, as estimativas de canal realizadas na BS (*uplink*) podem ser utilizadas no receptor, dentro do tempo de coerência do canal T_c . Com o propósito de padronizar a notação, agrupamos o vetor de canal de cada usuário na seguinte matriz de canal:

$$\mathbf{H}_k \in \mathbb{C}^{M_k} \sim \mathcal{CN}(\mathbf{0}, \beta_k \cdot \mathbf{I}), \quad (1.22)$$

sendo β_k o *path-loss* médio de cada UE presente no sistema.

1.4.5 Enrijecimento do canal e propagação favorável em mMIMO

O sistema mMIMO operando sob o modelo de desvanecimento de canal *Rayleigh* obtém duas propriedades importantes que tornam mais expedito lidar com um grande número de antenas na estação rádio-base. A primeira propriedade é chamada de enrijecimento do canal e é matematicamente definido como (Björnson et al., 2017b; Marzetta et al., 2016):

$$\frac{\|\mathbf{h}_k\|^2}{M} \xrightarrow{M \rightarrow \infty} \beta_k \quad (1.23)$$

sendo o vetor de canal $\mathbf{h}_k \in \mathbb{C}^M$.

A outra propriedade denominada propagação favorável assintótica é matematicamente definida como (Björnson et al., 2017b; Marzetta et al., 2016):

$$\frac{\mathbf{h}_k^H \mathbf{h}_{k'}}{M} \xrightarrow{M \rightarrow \infty} 0 \quad (1.24)$$

onde os canais $\mathbf{h}_k \in \mathbb{C}^M$ e $\mathbf{h}_{k'} \in \mathbb{C}^M$ para usuários distintos $k \neq k'$ seguem, respectivamente, distribuições estatísticas independentes. Assim, com um número suficientemente grande de antenas na BS, os respectivos canais de dois usuários resultam quase ortogonais, facilitando a separação dos sinais destes usuários na BS em um mesmo bloco de coerência do canal.

1.5 Figuras de mérito

As figuras de mérito são utilizadas para analisar e comparar sistemas de comunicação. Novas propostas de cenários/sistemas são exaustivamente testadas e monitoradas e seus resultados comparados com cenários já em operação, utilizando métricas generalistas como: SE, EE, índice de Jain, ou específicas como: média de tentativas de acesso, fração de falhas de tentativa de acessos, número normalizado de usuários aceitos e fração de resolução de colisão, sempre com o objetivo de verificar possíveis ganhos obtidos.

Tanto SE como a EE são métricas de desempenho muito relevantes e populares, sendo a primeira uma medida da eficiência com que um dado sistema de comunicação usa o espectro disponível, e a segunda é empregada para analisar o equilíbrio entre o consumo de potência e a taxa de dados atingível nas redes de comunicação.

Existe uma vasta literatura demonstrando o desempenho superior em termos de SE do NOMA quando comparado às técnicas OMA (Anwar et al., 2019). Existe também em grande número, estudos de novas topologias de comunicação envolvendo o mMIMO que já avaliaram o problema da EE, obtendo resultados ótimos, onde foram encontrados: número ótimo de antenas no *array*, número ótimo de dispositivos atendidos em uma célula e o ponto de máxima EE (Björnson et al., 2014; Björnson; Larsson, 2019).

A análise da EE em redes operando sistema NOMA é realizada em (Zhang et al., 2017), e sua superioridade foi demonstrada quando comparada com sistemas convencionais OMA. Outras pesquisas recentes buscam melhorar o desempenho NOMA, por exemplo, em (Mounchili; Hamouda, 2020), a distância mínima de pareamento foi encontrada e o desempenho da rede comparada com uma rede OMA, já em (Rezaei et al., 2020) é apresentada uma comparação entre o OMA e o sistema livre de células equipado com mMIMO-NOMA.

Uma análise da EE em Terahertz (THz)-NOMA-MIMO foi proposto em (Zhang et al., 2020), onde o número de dispositivos ativos é muito menor do que o número de antenas na BS. Por sua vez, em (Maraqa et al., 2020) um levantamento sobre PD-NOMA

foi realizado, evidenciando as poucas análises de EE e da comparação NOMA com muitas antenas versus mMIMO.

Em (Jacob; Abrão, 2020) é apresentada uma análise demonstrando as frações de potência necessárias para obter um índice de Jain, $\mathcal{F}$, o mais próximo possível de 1, para diferentes números de usuários no *cluster* PD-NOMA.

Alguns trabalhos recentes propõem a implantação do NOMA combinado a outras técnicas para um esquema de transmissão mais eficaz, por exemplo: em (Rajoria et al., 2021), NOMA e mMIMO são considerados operando conjuntamente em dois níveis de rede para acomodar um grande tráfego de dados. Em (Kim; Choi, 2021), aplica-se o PD-NOMA em Sistemas de Antenas Distribuídas (DAS – *Distributed Antennas System*) visando obter melhor desempenho quando comparado as Técnicas convencionais NOMA ou DAS isoladamente. Enquanto em (Budhiraja et al., 2021) mostra uma visão aprofundada do estado da arte das variantes NOMA de domínio de potência; além disso, apresenta várias questões em aberto e desafios de pesquisa baseados em NOMA e algumas propostas de aplicações são sistematizadas.

As métricas específicas, como as citadas acima, tem por objetivo medir determinadas etapas do processo de comunicação. Diversos trabalhos propondo melhorias no protocolo de acesso utilizam essas métricas, *e.g.*, (Carvalho et al., 2017; Nishimura et al., 2020; Pereira et al., 2021; Alves; Abrão, 2023).

1.5.1 Eficiência Espectral

A métrica de SE muito utilizada na literatura foi investigada na aplicação do NOMA a um sistema com um grande número de antenas na BS (Cheng et al., 2018; Senel et al., 2019). Esta métrica depende dos parâmetros de configuração do sistema, mas se resume, em obter a soma das taxas atingidas por todo sistema:

$$C_{\Sigma} = \sum_{k=1}^K C_k = \sum_{k=1}^K \log_2(1 + \text{SINR}_k), \quad \left[\frac{\text{bits/s}}{\text{Hz}} \right] \text{ ou } [\text{bits}], \quad (1.25)$$

sendo K o número de UEs ativos no sistema.

1.5.2 Eficiência Energética

A métrica EE é definida como a razão do número de bits de informação efetivamente recebidos pela potência total consumida pelo sistema para transmitir e receber/decodificar tal informação, incluindo componentes instantâneos e estáticos. O número de bits de informação efetivamente recebidos no destino pode ser determinado pela taxa de dados do sistema. Em alguns trabalhos na literatura o consumo de potência necessário para o processamento do sinal no lado do transmissor e do receptor é negligenciado, sendo somente calculado proporcionalmente a potência transmitida. O crescimento do número de antenas na BS e o aumento do número de dispositivos nos sistemas 5G podem afetar as

metas de EE dos sistemas modernos. Com ela, é possível avaliar a eficiência com que um sistema usa o recurso de potência, que é limitado, para comunicação de dados e também otimizar essa proporção. Em geral, como descrito em (Liu et al., 2020), o EE médio pode ser definido, considerando a banda como:

$$EE = \frac{B \sum_{k=1}^K C_k}{P_{\text{TOT}}}, \quad \left[\frac{\text{bits/Joule}}{\text{Hz}} \right] \quad \text{ou} \quad \left[\frac{\text{bits}}{\text{W}} \right], \quad (1.26)$$

sendo este o modelo adotado nesta Tese. Todavia, é bastante comum na literatura a eficiência energética considerar a banda unitária.

$$\overline{EE} = \frac{\sum_{k=1}^K C_k}{P_{\text{TOT}}}, \quad \left[\frac{\text{bits}}{\text{Joule}} \right], \quad (1.27)$$

sendo P_{TOT} é potência consumida total em todo sistema de comunicação. Deve-se considerar o consumo de potência de transmissão, a ineficiência do amplificador de potência de RF, o processamento de sinal de banda base e resfriamento, entre outros. Portanto, é necessário um modelo de consumo de potência mais realista e detalhado.

Com base em (Björnson et al., 2014), o modelo de consumo de potência adotado considera dois termos de potência: *a)* termo fixo; *b)* termos proporcional ao número de antenas M e ao número de dispositivos K . Os termos proporcionais ocorrem por causa das cadeias de transceptores, codificação/decodificação, estimativa de canal e pré-codificação. Seja a eficiência computacional I operações por joule na BS. Sendo descrito da seguinte forma:

Potência de Rádio Frequência: P_T é o total de potência total consumida para transmitir o sinal para os usuários ativos alcançarem a SINR objetivo e $0 < \vartheta \leq 1$ é a eficiência do amplificador de potência.

Consumo fixo de Potência: P_{FIXO} é a potência consumida na BS que não depende do número de antenas transmissoras e do número de dispositivos na célula, é formada pelo termo P_0 incluindo o consumo de potência da infraestrutura de *backhaul*, sinalização de controle, processador de banda base e o termo P_{OSC} de um oscilador simples.

$$P_{\text{FIXO}} = P_0 + P_{\text{OSC}}.$$

Potência dependente de K : P_K é formado pelo consumo para codificação e modulação de símbolos de informação para os dispositivos, representado por P_{COD} , o consumo para BS decodificar as sequências dos K símbolos de informação, definido por P_{DEC} , e a potência dos circuitos receptores, representada por P_{RX} , ainda compõe este termo, multiplicada também por K . Além disso, uma parte do custo de pré-codificação ZF depende apenas de K^3 .

$$P_K = K(P_{\text{COD}} + P_{\text{DEC}} + P_{\text{RX}}) + K^3 \frac{2}{3LT}.$$

Potência dependente de M . O termo P_M representa a potência de transmissão total, o termo (P_{TX}) é a potência de cada antena.

$$P_M = MP_{TX}.$$

Potência dependente de K e M : o termo P_{KM} é o custo da pré-codificação ZF (devido à inversão de matriz baseada em LU) (Boyd; Vandenberghe, 2018), que depende do número de dispositivos, do número de antenas e do vetor símbolo de informação.

$$P_{KM} = MK \frac{3+T}{TL} + MK^2 \frac{2}{TL},$$

com soma de cada porção, nós obtemos o consumo total de potência do sistema

$$P_{TOT} = \frac{P_T}{\vartheta} + P_{FIXO} + P_K + P_M + P_{KM} \quad [W]. \quad (1.28)$$

1.5.3 Área sobre as Curvas SE e EE

Motivados pela possibilidade de considerar toda a faixa de carregamento de UEs, e com isso, obter uma comparação mais justa entre os sistemas, pode-se considerar toda a faixa de carregamento de dispositivos dos resultados de SE e EE médios obtidos pelos sistemas PD-NOMA e mMIMO e normalizar por antena. Ademais, tais áreas sob as curvas de SE e EE, em [bits/antena] e [bits/antena/W], respectivamente fornecem valores de eficiência espectral e energética médias para uma ampla faixa de carregamento, permitindo comparar diretamente a SE e EE média para sistemas mMIMO versus PD-NOMA com diferentes limitações de carregamento. Portanto, sugere-se considerar a seguinte metodologia para uma comparação mais justa. Com isso podemos definir:

$$\mathcal{S}_M^{\text{SYST}} = \frac{1}{M} \cdot \int_0^{\rho=2} \overline{\text{SE}}(\rho) \, d\rho \quad \left[\frac{\text{bits/antena}}{\text{s} \cdot \text{Hz}} \right] \quad (1.29)$$

e

$$\mathcal{E}_M^{\text{SYST}} = \frac{1}{M} \cdot \int_0^{\rho=2} \overline{\text{EE}}(\rho) \, d\rho \quad \left[\frac{\text{bits/antena}}{\text{Joule} \cdot \text{Hz}} \right], \quad (1.30)$$

respectivamente, onde $\overline{\text{SE}}(\rho)$ é a média da soma das taxas do sistema, e $\overline{\text{EE}}(\rho)$ é a eficiência energética média geral do sistema alcançada sob carregamento de dispositivos ρ .

1.5.4 Índice de Justiça de Jain

O índice de justiça entre os K usuários ativos descrito em (Jain et al., 1984) é conhecido como índice de Jain, este índice é capaz de representar a diferença na taxa de transmissão alcançada pelos dispositivos ativos na célula e pode ser definido como:

$$\mathcal{F}_K^{\text{SYST}} = \frac{\left(\sum_{k=1}^K R_k \right)^2}{K \sum_{k=1}^K R_k^2}. \quad (1.31)$$

Este índice mede a "igualdade" da alocação do usuário k . Se todos os usuários obtiverem a mesma taxa, ou seja, R_k são todos iguais, então o índice de justiça de Jain é 1 e o sistema é 100% justo. À medida que a disparidade aumenta, a imparcialidade diminui e um esquema que favorece apenas alguns poucos usuários selecionados tem um índice de imparcialidade próximo a 0 (Jain et al., 1984).

O índice de Jain pode ser entendido como uma forma de medir o desvio padrão das taxas, mas representado por um valor numérico para o grupo. Sendo, quanto mais próximos de 1 são os resultados do índice de Jain, mais igualitariamente o sistema transfere taxa aos UEs e quanto mais próximo de 0 mais desigual é a taxa dos UEs.

1.5.5 Métricas para protocolos de acesso

Dentre as métricas utilizadas para medir o desempenho dos protocolos de acesso destacamos a *média de tentativas de acesso* que apresenta o número médio de tentativas de acesso realizadas pelos UEs para acessar o sistema, a *fração de falhas de tentativa de acessos* que monitora o número de UEs que falharam ao tentar acessar a rede, o *número normalizado de usuários aceitos* que mostra do total de UEs que tentaram acesso quantos foram aceitos e por último a fração de resolução de colisão que apresenta a *razão de colisões resolvidas* pelo protocolo analisado.

1.6 Acesso Aleatório

As técnicas de acesso são necessárias pois não é possível termos um link de comunicação dedicado entre transmissor e receptor, com isso, o link de acesso precisa ser compartilhado. Diversos protocolos de acesso foram propostos na literatura com o objetivo de evitar colisões, ou seja, de serem responsáveis por coordenar o acesso ao meio não dedicado. Os protocolos de acesso aleatório (RA – *random access*) são um dos grupos de acesso propostos que possuem como principais características, não estabelecer prioridades, onde as necessidades de acesso são aleatórias e não sequenciais. Apresentamos a seguir uma divisão dos protocolos de acesso aleatório: Baseado em concessão (GB - *grant-based*) e livre de concessão (GF - *grant-free*).

1.6.1 Por concessão

Diversos estudos publicados na literatura propõe técnicas de controle de acesso aleatório por concessão (Björnson et al., 2017a; Han et al., 2017a; Han et al., 2017b). Dentre estes, destacamos o importante protocolo RA baseado em concessão para sistemas MIMO massivos espacialmente estacionários que foi proposto em (Björnson et al., 2017a), identificado como protocolo de resolução de colisão de usuário mais forte (SUCRe - *Strongest User Collision Resolution*). Este protocolo possui 4 etapas e visa resolver colisões de pilotos

RA selecionando o UE com maior coeficiente de sinal para repetir seu piloto após a colisão, de forma totalmente descentralizada e descoordenada. Para ser possível esse procedimento BS e UEs, em uma etapa preliminar a transmissão de dados, trocam informações a fim de possibilitar a transmissão de dados, conforme figura 1.11.

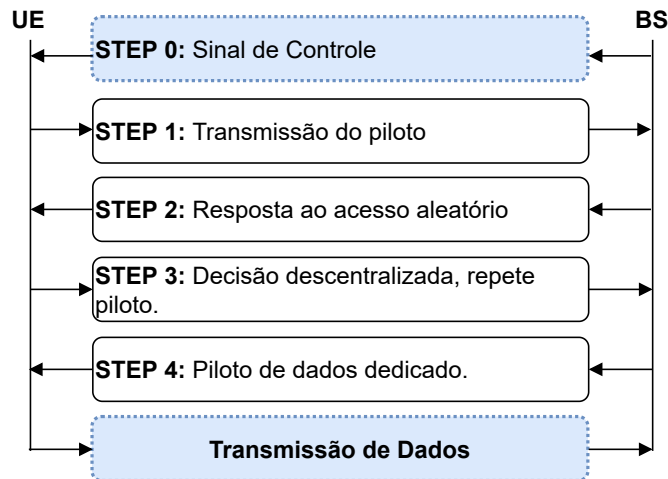


Figura 1.11 – Protocolo de acesso aleatório baseado em concessão (SUCRe).

1. **Transmissão Piloto:** Cada dispositivo ativo escolhe uma sequência piloto aleatória, de um conjunto pré-definido de pilotos ortogonais, e a envia para a BS para informar que tem dados para transmitir.
2. **Resposta ao acesso aleatório:** A BS responde a cada dispositivo ativo, um sinal precodificado.
3. **Decisão descentralizada, repete piloto:** De forma individual, ou seja, descentralizada, cada dispositivo decide, baseado no sinal precodificado que foi recebido, se continua ou não nesta tentativa de acesso.
4. **Piloto de dados dedicado:** A BS informa qual o piloto de dados que deverá ser utilizada para a transmissão de dados.

O processamento simples na BS é a principal vantagem do protocolo de acesso aleatório baseado em concessão. No entanto, apresenta algumas deficiências no contexto do acesso massivo, como por exemplo: a) elevada sobrecarga de sinalização, que pode ser proibitiva para volumes de dados de carga baixa. Assim, quando o volume de dados transmitido é baixo, é mais conveniente utilizar protocolos de acesso GF. Se uma sequência piloto for selecionada por mais de um dispositivo, ocorrerá uma colisão, portanto em acesso massivo os dispositivos podem sofrer com uma grande probabilidade de falha de acesso devido a colisões, aumentando consideravelmente a latência média de acesso, às vezes além do tolerável.

1.6.2 Livre de concessão

Em propostas mais recentes técnicas de acesso ao meio livres de concessão foram estudadas em (Liu et al., 2018; Elkourdi et al., 2018; Ding et al., 2019; Choi, 2021; Chen et al., 2021; Ramatryana; Shin, 2021; Ramatryana; Shin, 2022; Souza; Abrão, 2023), nestes protocolos, cada dispositivo ativo transmite diretamente sem qualquer concessão. Além disso, a transmissão livre de concessões usando OMA geralmente não é viável devido ao grande número de colisões, uma vez que a sequência piloto ortogonal disponível é limitada. Assim, o acesso aleatório sem concessão é geralmente associado ao NOMA, devido ao grande número de dispositivos (Liu et al., 2018; Elkourdi et al., 2018; Chen et al., 2021; Ramatryana; Shin, 2021; Ramatryana; Shin, 2022; Souza; Abrão, 2023).

A idéia principal do acesso aleatório sem concessão é detectar os dispositivos ativos, que pode ocorrer com base nas sequências piloto ou com estimativa de tráfego. A linha de protocolos GF que utilizam pilotos, ainda podem ser subdivididas em pilotos ortogonais e não ortogonais. Técnicas de RA GF com recurso ortogonal, procuram melhorar a capacidade de resolver as colisões, com repetição de pacotes ou previsão de tráfego (Elkourdi et al., 2018; Ramatryana; Shin, 2021; Ramatryana; Shin, 2022).

Quanto utilizando recursos não ortogonais, os dispositivos que possuem dados a transmitir, enviam suas sequências piloto e os dados em conjunto para a BS. A BS primeiro detecta qual sequência piloto foi utilizada, detectando o dispositivo ativo. Em seguida, se a sequência piloto for detectada corretamente, a BS estima o canal do dispositivo com base nessa sequência piloto recebida e então decodifica os dados (Souza; Abrão, 2023).

No entanto, esses protocolos enfrentam desafios importantes, com um problema notável sendo a detecção de dispositivos ativos. A detecção de dispositivos ativos demanda um grande processamento na BS, pois não é possível atribuir uma sequência piloto ortogonal a todos os dispositivos (Ding; Choi, 2020; Han et al., 2020). Uma vez que o número de dispositivos IoT e antenas na BS é muito grande nos sistemas 5G e além, pode resultar em uma matriz de estado do dispositivo de alta dimensão, implicando em enorme complexidade computacional (Chen et al., 2021).

1.7 Objetivos

Os objetivos desta Tese de doutorado podem ser compreendidos em objetivos geral e específicos:

1. **Geral:** Analisar os aspectos da implantação de técnicas e esquemas de transmissão que envolvem NOMA e propor soluções aplicando esse método em mMIMO e/ou XL-MIMO;
2. **Específicos:**

- a) Extensiva revisão bibliográfica relativa às aplicações de técnicas NOMA, sua integração em cenários de comunicações sem fio futuros (sistemas B5G e 6G) e problemas em aberto;
- b) Propor técnicas para resolução de problemas associados ao tema de Tese, especificamente: *i)* fornecer evidências que na aplicação do NOMA é possível atender uma quantidade maior de dispositivos e *ii)* estabelecer um método de acesso aleatório baseado em concessão GB em cenário XL-MIMO lotado;
- c) Desenvolver soluções algorítmicas corroboradas por simulações numéricas tendo em vista demonstrar a efetividade das soluções encontradas, comparando as mesmas ao estado-da-arte da literatura;
- d) Disseminar as soluções propostas, as análises sistemáticas (numéricas e/ou analíticas), algoritmos e outros produtos da pesquisa através de submissão de artigos científicos para revistas de alto fator de impacto (IF).

1.8 Estrutura do Trabalho de Tese

Esta Tese de Doutorado está organizada da seguinte forma: os capítulos 2 e 3 fornecem informações detalhadas sobre os modelos de sistema estudados; os capítulos 4 e 5 apresentam os resultados e a análise das simulações que foram realizadas e por fim no capítulo 6, apresentamos as considerações finais do trabalho de Tese, bem como traçamos possíveis caminhos para trabalhos futuros envolvendo os temas discutidos.

1.9 Contribuições

Este trabalho de pesquisa científica e tecnológica gerou as seguintes publicações e/ou submissões em periódicos indexados ou conferências internacionais relevantes para a área de sistemas de comunicações:

Artigos Completos em Periódicos

- [A] **Thiago Augusto Bruza Alves**, Taufik Abrao. *NOMA-Based Random Access in mMTC XL-MIMO*. IEEE Access, (IF=3.476), Vol. 11, pp. 1944-1954, **2023**, DOI: [10.1109/ACCESS.2022.3233942](https://doi.org/10.1109/ACCESS.2022.3233942).

Abstract: In machine-type communication (MTC), massive access attempts are generated and the massive MIMO is the key technology to support this demand. To support massive MTC (mMTC), the recent extra-large scale massive multiple-input multiple-output (XL-MIMO) architecture has been seen as a promising technology for providing very high-data rates in high-user density scenarios. Therefore, the large dimension is of the same order as the distances to the user equipment (UE) causing spatial non-stationarities and visibility regions (VRs) to occur across the huge XL array extension. In this work, we investigate

the random access (RA) problem in crowded mMTC XL-MIMO scenarios; the proposed grant-based random access (GB-RA) protocol combining the advantage of non-orthogonal multiple access (NOMA) and the strongest user collision resolution in extra-large arrays (SUCRe-XL), namely NOVR-XL scheme can allow access of multiple colliding users in the same XL subarray (SA) selecting the same pilot sequence. The proposed NOVR-XL GB-RA protocol is able to provide a reduction in the number of attempts to access the network, while improving the average sum-rate, as the number of SA increases.

Keywords: NOMA; Machine Type Communication; XL-MIMO; Random Access Protocols; Crowded scenarios; Uniform Rectangular Array (URA).

Resumo: Na comunicação do tipo máquina (MTC), são geradas massivas tentativas de acesso e o mMIMO é a tecnologia chave para suportar essa demanda. Para dar suporte ao MTC massivo (mMTC), a recente escala ultra massiva (XL-MIMO) foi identificada como uma tecnologia promissora para fornecer taxas de dados muito altas em cenários de alta densidade de usuários. Portanto, a grande abertura do conjunto de antenas é da mesma ordem das distâncias do conjunto até o equipamento do usuário (UE), causando não estacionariedades espaciais e regiões de visibilidade (VRs) ao longo da enorme extensão da matriz XL. Neste trabalho, investigamos o problema de acesso aleatório (RA) em cenários lotados mMTC XL-MIMO; o protocolo de acesso aleatório baseado em concessão (GB-RA) proposto combina a vantagem do acesso múltiplo não ortogonal (NOMA) e a resolução de colisão de usuário mais forte em matrizes extra grandes (SUCRe-XL), nomeado de protocolo NOVR-XL pode permitir o acesso de vários usuários em colisão no mesmo subarray XL (SA) selecionando a mesma sequência piloto. O protocolo NOVR-XL GB-RA proposto é capaz de reduzir o número de tentativas de acesso à rede, ao mesmo tempo em que melhora a taxa média de soma, à medida que o número de SA aumenta.

Palavras-chaves: NOMA; Comunicação do tipo máquina; XL-MIMO; Protocolos de acesso aleatórios; Cenários lotados; Conjunto de antenas retangular uniforme.

- [B] **Thiago Augusto Bruza Alves** and **Taufik Abrão**, *Massive MIMO and NOMA bits-per-antenna efficiency under power allocation policies*. Physical Communication, Elsevier (IF=2.379), Vol. 51, **2022**, 101588 DOI: [10.1016/j.phycom.2021.101588](https://doi.org/10.1016/j.phycom.2021.101588).

Abstract: A comparative resource allocation analysis in terms of received bits-per-antenna spectral efficiency (SE) and energy efficiency (EE) in downlink (DL) single-cell massive multiple-input multiple-output (mMIMO) and non-orthogonal multiple access (NOMA) systems considering a BS equipped with many (M) antennas, while K devices operate with a single-antenna, and the loading of devices $\rho = K/M$ ranging in $0 < \rho \leq 2$ is carried out under three different (PA) strategies: the inverse of the channel power allocation (PICPA), a modified water-filling (Δ -WF) allocation method, and the equal power allocation (EPA) reference method. Since in the NOMA system, the two devices per cluster are overlapped in the power-domain, the channel matrix requires transformation to perform the zero-forcing

(ZF) precoding as adopted in mMIMO. Hence, NOMA operating under many antennas can favor a group of devices with higher array gain, overcoming the mMIMO and operating conveniently in the higher loading range $0.6 < \rho < 2.0$. In such scenario, a more realistic and useful metric consists in evaluating the area under SE and EE curves, by measuring the bit-per-antenna and bit-per-antenna-per-watt efficiency, respectively. Our numerical results confirm a superiority of NOMA w.r.t. mMIMO of an order of 3x for the SE-area and 2x for the EE-area metric.

Keywords: Non-orthogonal multiple access (NOMA); Massive multiple-input multiple-output (mMIMO); Energy efficiency (EE); Spectral efficiency (SE).

Resumo: Uma análise comparativa alocação de recursos em termos de bits por antena recebidos da eficiência espectral (SE) e da eficiência energética (EE) no downlink (DL), em célula única nos sistemas massiva múltipla entradas e múltiplas saídas (mMIMO) e acesso múltiplo não ortogonal (NOMA) considerando uma estação base (BS) equipada com (M) antenas, enquanto os dispositivos K operam com uma única antena, e o carregamento dos dispositivos $\rho = K/M$ variando em $0 < \rho \leq 2$ é realizado sob três políticas de alocação de potência diferentes: alocação de potência pelo inverso do canal (PICPA), alocação de potência enchimento de água modificado (Δ -WF) e o método simplista de referência de alocação de potência igualitária (EPA). Como no sistema NOMA, os sinais dos dois dispositivos no cluster são sobrepostos no domínio de potência, a matriz do canal requer transformação para realizar a pré-codificação zero-forcing (ZF) adotada no mMIMO. Assim, o NOMA operando sob muitas antenas pode favorecer um grupo de dispositivos com maior ganho de array, superando o mMIMO e operando convenientemente na faixa de carregamento mais alta $0.6 < \rho < 2.0$. Nesse cenário, uma métrica mais realista e útil consiste em avaliar a área sob as curvas SE e EE, medindo a eficiência bit-por-antena e bit-por-antena-por-watt, respectivamente. Nossos resultados numéricos confirmam a superioridade de NOMA *versus* mMIMO nesta métrica na ordem de 3x para a área SE e 2x para a área EE.

Palavras-chaves: acesso múltiplo não ortogonal (NOMA); massiva múltipla entradas e múltiplas saídas (mMIMO); eficiência energética (EE); eficiência espectral (SE).

Artigos Completos em Conferências

- [C] **Thiago Augusto Bruza Alves** and Taufik Abrao. *On the SE and EE in Massive MIMO and NOMA*. 2021 IEEE URUCON (24-26 Nov. 2021), Montevideo - URUGUAI (<https://http://urucon2021.org/>) (Virtual conference).

Abstract: A comparative analysis on the Spectral Efficiency (SE) and Energy Efficiency (EE) in Downlink (DL) single-cell massive Multiple-Input Multiple-Output (mMIMO) and Non-Orthogonal Multiple Access (NOMA) considering a Base Station (BS) equipped with M antennas, while K devices operates with a single-antenna, and devices loading $\rho = K/M$ ranging in $0 < \rho \leq 2$ is carried out. In NOMA system, devices are overlapped

in the power-domain; hence, NOMA is able to serve a greater number of devices. Under certain devices loading conditions, the NOMA overcame the mMIMO in terms of SE-EE trade-off, while continuing operate conveniently in the extended range $0.6 < \rho < 2.0$. We demonstrate that NOMA system is capable to operate suitably under higher system loading. The proposed analysis of the areas under SE and EE curves reveals that NOMA achieves highest area results. *Keywords:* Non-Orthogonal Multiple Access (NOMA); massive Multiple-Input Multiple-Output (mMIMO); Energy Efficiency (EE); Spectral Efficiency (SE).

Resumo: Uma análise comparativa entre a eficiência espectral (SE) e a eficiência energética (EE) no *downlink* (DL) entre massivo múltiplas entradas e múltiplas saídas (mMIMO) de célula única e o acesso múltiplo não ortogonal (NOMA) considerando uma estação base (BS) equipada com M antenas, enquanto os K dispositivos operam com uma única antena, e com carregamento de dispositivos $\rho = K/M$ variando em $0 < \rho \leq 2$. No sistema NOMA, o sinal dos dispositivos são sobrepostos no domínio da potência; portanto, o NOMA é capaz de atender a um número maior de dispositivos. Sob certas condições de carregamento de dispositivos, o NOMA superou o mMIMO em termos de SE-EE, enquanto continuou a operar convenientemente na faixa estendida $0,6 < \rho < 2,0$. Demonstramos que o sistema NOMA é capaz de operar adequadamente sob carga de sistema mais alta. A análise proposta das áreas sob as curvas SE e EE revela que o NOMA alcança os maiores resultados de área.

Palavras-chaves: acesso múltiplo não ortogonal (NOMA); massiva múltipla entradas e múltiplas saídas (mMIMO); eficiência energética (EE); eficiência espectral (SE).

- [D] **Thiago Augusto Bruza Alves**, Taufik Abrao. ***Crowded MTC Random Access in NOMA XL-MIMO***. 2022 IOT Forum (24-28 Oct. 2022), São Paulo, SP - BRASIL <https://iotbrasil.org.br/congresso/2022/>. (Hybrid Event).

Abstract: Massive MIMO is one of the key technologies to support the growth of massive access attempts by devices, such as in massive machine type communication (mMTC). The evolution of antenna array technology brought the recent extra-large scale massive multiple input multiple output (XL-MIMO) systems, seen as a promising technology for providing very high-data rates in highuser density scenarios. Spatial non stationarities and visibility regions (VRs) occur across all huge XL array extension, since its large dimension is of the same order of the distances to the user equipment (UE). We investigate the random access (RA) problem in crowded XL-MIMO scenarios. The proposed nonorthogonal multiple access (NOMA) visible region extra large array (NVR-XL) protocol takes advantage of the power domain NOMA to allow access of two or more users colliding in the same XL sub-array (SA) selecting the same pilot sequence. The NVRXL provides a reduction in the number of attempts to access the network, while improving the average sum-rate, as the number of SA increases.

Keywords: NOMA; Machine Type Communication, XL-MIMO, Crowded scenarios.

Resumo: MIMO massivo é uma das principais tecnologias para suportar o crescimento do número de tentativas de acesso por dispositivos, como em comunicação de tipo de máquina massiva (mMTC). A evolução da tecnologia de arranjo de antenas trouxe os recentes sistemas massivos de múltiplas entradas e múltiplas saídas de escala extra grande (XL-MIMO), vistos como uma tecnologia promissora para fornecer taxas de dados muito altas em cenários de alta densidade de usuários. As não estacionariedades espaciais e as regiões de visibilidade (VRs) ocorrem em toda a extensão da grande matriz XL, uma vez que sua grande dimensão é da mesma ordem das distâncias ao equipamento do usuário (UE). Investigamos o problema de acesso aleatório (RA) em cenários lotados em XL-MIMO. O protocolo proposto de acesso múltiplo não ortogonal (NOMA) em região visível de escala extra grande (NVR-XL) aproveita o domínio de potência NOMA para permitir o acesso de dois ou mais usuários colidindo no mesmo subarray (SA) XL selecionando a mesma sequência piloto. O NVR-XL proporciona uma redução no número de tentativas de acesso à rede, ao mesmo tempo em que melhora a taxa média de soma, à medida que o número de SA aumenta.

Palavras-chaves: NOMA; comunicação tipo máquina; XL-MIMO; cenários lotados.

Artigo Completo Apresentado em Conferência

[E] **Thiago Augusto Bruza Alves**, Taufik Abrao. *Improving Random Access with NOMA in mMTC XL-MIMO*. VTC2023-Spring (20-23 Jun. 2023), Florence, Italy <https://events.vtsociety.org/vtc2023-spring/>. (Hybrid Event).

Abstract: The extra-large multiple-input multiple-output (XL-MIMO) architecture has been recognized as a technology for giving support for the massive MTC (mMTC), providing very high-data rates in high-user density scenarios. However, the large dimension of the array increases the Rayleigh distance (d_{Rayl}), in addition to obstacles and scatters causing spatial non-stationarities and distinct visibility regions (VRs) across the XL array extension. We investigate the random access (RA) problem in crowded XL-MIMO scenarios; the proposed grant-based random access (GB-RA) protocol combining the advantage of non-orthogonal multiple access (NOMA) and strongest user collision resolutions in extra-large arrays (SUCRe-XL) named NOMA-XL can allow access of two or three colliding users in the same XL sub-array (SA) selecting the same pilot sequence. The received signal processing in a SA basis changes the d_{Rayl} , enabling the far-field planar wavefront propagation condition, while improving the system performance. The proposed NOMA-XL GB-RA protocol is able to provide a reduction in the number of attempts to access the mMTC network while improving the average sum rate, as the number of SA increases.

Keywords: NOMA; machine type communication; XL-MIMO; crowded scenarios; random access; Rayleigh distance.

Resumo: A arquitetura extra grande de múltiplas entradas e múltiplas saídas (XL-MIMO) foi reconhecida como uma tecnologia para dar suporte ao MTC massivo (mMTC), forne-

cendo taxas de dados muito altas em cenários de alta densidade de usuários. No entanto, a grande dimensão da matriz de antenas aumenta a distância de Rayleigh (d_{Rayl}), além de obstáculos e dispersões causando não estacionariedades espaciais e regiões de visibilidade distintas (VRs) em toda a extensão da matriz XL. Investigamos o problema de acesso aleatório (RA) em cenários lotados de XL-MIMO; o protocolo proposto de acesso aleatório baseado em concessão (GB-RA) que combina a vantagem do acesso múltiplo não ortogonal (NOMA) e resoluções de colisão de usuário mais fortes em matrizes extragrandes (SUCRe-XL) denominadas NOMA-XL pode permitir o acesso de dois ou três usuários em colisão na mesma submatriz XL (SA) selecionando a mesma sequência piloto. O possibilidade de processamento do sinal em uma base SA muda o d_{Rayl} , permitindo a condição de propagação planar da frente de onda de campo distante, enquanto melhora o desempenho do sistema. O protocolo NOMA-XL GB-RA proposto é capaz de proporcionar uma redução no número de tentativas de acesso à rede enquanto melhora a taxa média de soma, à medida que o número de SA aumenta.

Palavras-chaves: NOMA; comunicação tipo máquina; XL-MIMO; cenários lotados; acesso aleatório; distância de Rayleigh.

Artigo Completo em Periódico como Co-autor

- [F] Wilson de Souza Junior, **Thiago Augusto Bruza Alves**, Taufik Abrao. ***Antenna selection in nonorthogonal multiple access multiple-input multiple-output systems aided by machine learning***. Transactions on Emerging Telecommunications Technologies, (IF=3.31), Vol.32, Pub. 10, **2021**, DOI: [10.1002/ett.4283](https://doi.org/10.1002/ett.4283). *Abstract:* This work proposes a transmitter antenna selection (TAS) method for multiple-input multiple-out (MIMO) nonorthogonal multiple access (NOMA) that is a promising multiple access technique for the fifth generation mobile communications systems. Specifically, we propose to activate the more suitable subset of base station antennas while allocating users into appropriate NOMA clusters such that the system operates in energy efficiency mode, selecting such appropriate antenna subset indexes that maximize the sum-rate (SR) of the NOMA-MIMO system. Once that the TAS based on exhaustive-search is very complex to be implemented in real communication systems, we propose an effective TAS method based on machine learning while keeping very promising performance. A convolutional neural network-based transmitter antenna selection (CNN-TAS) method is proposed for efficiently select antennas aiming at maximizing the system SR. Hence, extensive numerical results demonstrate that the CNN-TAS can suitably learn the problem, performing with very high accuracy choosing properly the antenna subset that maximizes the system spectral efficiency while reducing substantially the processing time of real-time operations.

Keywords: MIMO-NOMA; transmitter antenna selection.

Resumo: Este trabalho propõe um método de seleção de antena transmissora (TAS) para

acesso múltiplo não ortogonal (NOMA) de múltiplas entradas e múltiplas saídas (MIMO), que é uma técnica de acesso múltiplo promissora para os sistemas de comunicações móveis de quinta geração. Especificamente, propomos ativar o subconjunto mais adequado de antenas de estação base enquanto alocamos usuários em *clusters* NOMA apropriados, de modo que o sistema opere no modo de maior eficiência energética, selecionando índices de subconjunto de antena apropriados que maximizam a taxa de soma do Sistema MIMO-NOMA. Uma vez que o TAS baseado em busca exaustiva é muito complexo para ser implementado em sistemas de comunicação reais, propomos um método TAS eficaz baseado em aprendizado de máquina, mantendo um desempenho muito promissor. Um método de seleção de antena transmissora baseado em rede neural convolucional (CNN-TAS) é proposto para selecionar antenas de forma eficiente visando maximizar o soma das taxas do sistema. Portanto, resultados numéricos extensos demonstram que o CNN-TAS pode aprender adequadamente o problema, realizando com precisão muito alta, escolhendo adequadamente o subconjunto de antenas que maximiza a eficiência espectral do sistema enquanto reduz substancialmente o tempo de processamento de operações em tempo real.

Palavras-chaves: MIMO-NOMA; seleção de antena transmissora.

No contexto desse trabalho de Tese, a contribuição [A] trata de um aprimoramento do protocolo SUCRe-XL recentemente proposto, em cenários mMTC com *array* XL-MIMO, incluindo a possibilidade de processamento individualizado por *subarray* via RPU e posterior processamento em EGC na CPU e a resolução das colisões de sinais sobrepostos no domínio da potência (NOMA) com a aplicação do SIC, denominado protocolo de acesso aleatório com concessão (NOVR-XL). Neste trabalho, encontramos o *fator de escala* ótimo de modo a maximizar a média da soma das taxas do sistema, com substancial melhoria nos indicadores de acesso aleatório em cenários mMTC lotados.

Avançamos com o trabalho de Tese, ao publicarmos a contribuição [B]. Em [B] sistematizamos uma extensa análise comparativa entre mMIMO versus NOMA com múltiplas antenas e carregamento de usuários chegando em até o dobro do número de antenas na BS. Nossos resultados comprovam que mMIMO não opera adequadamente quando o número de usuários supera $0,6 \cdot M$ sendo M o número de antenas na BS. Sob a aplicação de diferentes políticas de alocação de potência e $M = [64, 128 \text{ e } 256]$, analisamos diversas métricas com destaque ao número de bits por antena sob as curvas de SE e EE, nas quais o NOMA apresenta resultados superiores significativos.

A contribuição [C] propõe uma análise comparativa entre os sistemas mMIMO e o PD-NOMA com múltiplas antenas sob a perspectiva da EE e SE, fornecendo evidências da habilidade do NOMA em operar em ambientes com um elevado número de usuários.

A contribuição [D] desenvolve uma ideia inicial de melhoria proposta ao protocolo SUCRe-XL, com a perspectiva de bons resultados nos indicadores de acesso aleatório.

Finalmente, a contribuição [E] analisa e sistematiza a efetividade da proposta de melhoria apresentada pelo protocolo NOVR-XL, observando a correta modelagem do canal

sem fio. Uma vez que as dimensões (*array aperture*) do arranjo de antenas influencia na distância de Rayleigh, podem ocorrer algumas situações de projeto: a) a modelagem em campo distante pode se tornar inválida; b) por sua vez, no XL-MIMO a divisão do *array* completo em SAs pode reduzir significativamente a distância de Rayleigh, o que valida a modelagem do canal em campo distante.

Adicionalmente a contribuição [F] apresenta uma técnica de seleção de antenas via rede neural convolucional, que seleciona uma parte das antenas disponível para atender usuários alocados adequadamente em agrupamentos *clusters* NOMA de modo que o sistema possa operar com máxima eficiência espectral.

2 mMIMO-NOMA

Os novos requisitos das redes B5G deverão exigir conexões ultradensas com requisitos amplamente heterogêneos. Os desafios nas redes persistem, incluindo a melhoria conjunta da eficiência espectral (SE – *spectral efficiency*) e da eficiência energética (EE – *energy efficiency*), ou seja encontrar o melhor *trade-off* SE–EE, sempre visando atender o crescente número de dispositivos conectados à rede. Entre as propostas para solucionar esses desafios, o sistema mMIMO é o principal esquema de transmissão proposto para 5G, o qual permite o aumento da capacidade do link, explorando a propagação de múltiplos caminhos com o uso de um grande número de antenas BS (Larsson et al., 2014; Björnson et al., 2014). Outra tecnologia habilitadora relevante é o PD-NOMA, que entre outras variantes, explora o domínio da potência e o SIC, como uma forma alternativa em termos de tecnologia de acesso múltiplo, ajudando a mitigar o problema de esgotamento do espectro e atendendo a mais de um dispositivo por bloco de coerência no mesmo canal (Ding; Poor, 2016). O NOMA é uma tecnologia candidata para 5G e B5G, a qual permite operar em redes densas (Makki et al., 2020) e sujeitas a interferência multiusuário.

Existe uma vasta literatura demonstrando o desempenho superior em termos de SE do NOMA quando comparado às técnicas OMA (Anwar et al., 2019). Pesquisas anteriores propondo melhorar o desempenho do sistema de comunicação combinando MIMO com um pequeno número de antenas e NOMA foram discutidas em (Ali et al., 2017; Zeng et al., 2017; Islam et al., 2018; Chen et al., 2018). Estudos comparando NOMA e mMIMO em uma única célula foram propostos em (Senel et al., 2019; Cheng et al., 2018; Dai et al., 2019). A estimação do estado do canal (CSI – *channel state information*) através da aquisição piloto compartilhada no sistema NOMA foi proposta em (Cheng et al., 2018). Em (Dai et al., 2019) a aplicação do NOMA no esquema mMIMO é proposta e melhores resultados são alcançados na análise comparativa. Além disso, em (Senel et al., 2019) é analisado o desempenho do esquema NOMA com múltiplas antenas e mMIMO em cenários com e sem linha de visada.

O mMIMO canônico refere-se aos sistemas com BS formada por um grande número de antenas M quando comparado ao número de dispositivos ativos K , sucintamente $M \gg K$ é considerado uma configuração mMIMO canônico. O NOMA típico melhora a SE sobrepondo os sinais dos dispositivos selecionados que formam um *cluster* no domínio da potência, multiplexando-os sobre o mesmo sinal e sendo servido pelo mesmo *beamforming*. No entanto, o sucesso do NOMA depende da habilidade do SIC em separar/detectar corretamente usuários do mesmo *cluster* multiplexados com níveis de potência distintos. Contribui para o desempenho do SIC o modo de agrupamento dos usuários nos diferentes *clusters* (regra de formação), bem como da metodologia de alocação de potência dos usuários pertencentes ao mesmo agrupamento.

2.1 Modelo de sistema mMIMO-NOMA

Consideramos uma transmissão DL de uma única célula multiusuário operando em TDD onde K UEs ativos de antena única (K será sempre um número par) se comunicam com uma BS equipada com M antenas, modelamos o canal como campo distante (*far-field*) sem linha de visada (NLOS – *no line of sight*), conforme Subseção 1.4.2.

Os K UEs formam o conjunto de usuários $\mathcal{K}$, esses dispositivos estão distribuídos aleatoriamente em um disco de raio d_{max} . O vetor que modela o ganho de canal do dispositivo k pode ser descrito como:

$$\mathbf{h}_k = \sqrt{\beta_k} \mathbf{h}'_k, \quad k = 1, \dots, K, \quad (2.1)$$

onde β_k é o desvanecimento em larga escala, definido conforme descrito na Eq. (1.19). Em cada RB, $\mathbf{h}'_k$, na Eq. (2.1) é um vetor onde cada posição representa uma realização independente e aleatória do desvanecimento em pequena escala que segue uma distribuição de *Rayleigh*, $\mathbf{h}'_k \sim \mathcal{CN}(0, \mathbf{I}_M)$, $k = 1, \dots, K$. O sinal transmitido $\mathbf{x}_k \in \mathbb{C}^M$ é o símbolo de dados pré-codificado do dispositivo k :

$$\mathbf{x}_k = \mathbf{g}_k \sqrt{\alpha_k P_T} s_k, \quad (2.2)$$

onde $\mathbf{g}_k$ é o vetor de feixe (pré-codificação) normalizado, $\alpha_k P_T$ é a potência de transmissão alocada para o UE $_k$ e $s_k \sim \mathcal{CN}(0, 1)$ é o símbolo de dados do dispositivo k . No DL o sinal recebido pelo k -ésimo dispositivo pode ser definido como:

$$\begin{aligned} y_k &= \mathbf{h}_k^T \sum_{k'=1}^K \mathbf{x}_{k'} + n_k, \\ &= \sqrt{\beta_k} \mathbf{h}'_k^T \sum_{k'=1}^K \mathbf{g}_{k'} \sqrt{\alpha_{k'} P_T} s_{k'} + n_k, \\ &= \sqrt{\beta_k \alpha_k P_T} \mathbf{h}'_k^T \mathbf{g}_k s_k + \sqrt{\beta_k} \mathbf{h}'_k^T \sum_{\substack{k' \neq k \\ k'=1}}^K \mathbf{g}_{k'} \sqrt{\alpha_{k'} P_T} s_{k'} + n_k, \end{aligned} \quad (2.3)$$

onde $n_k \sim \mathcal{CN}(0, 1)$ é o ruído aditivo. Observe que, até esta etapa, esta modelagem se aplica a ambos os sistemas NOMA e mMIMO, mas o feixe é formado de maneira diferente e este tópico será abordado na Subseção 2.3.

Importante: Em todos os esquemas de transmissão, a BS requer informações cruciais *a priori*; estas informações podem estar relacionadas ao conhecimento do coeficiente de canal, à localização dos dispositivos, e à taxa requerida; entretanto dependem decisivamente do esquema de múltiplo acesso em questão (Makki et al., 2020). Assim, neste trabalho de Tese, ao analisarmos os sistemas mMIMO e NOMA assumimos que as informações *a priori*; CSI, localização, taxa de dados mínima requerida, entre outras, mesmo que não descritas de forma direta, consideramos que podem ser obtidas com sucesso.

2.2 Agrupamento de usuários (*Clustering*)

No PD-NOMA, explorando a diferença entre os ganhos de canal dos UEs, na transmissão, os sinais com a informação, são multiplexados no domínio da potência e de forma não ortogonal compartilham o mesmo RB; na recepção a informação é decodificada em um receptor SIC multiusuário.

Ao explorar o NOMA e o mMIMO torna-se possível combinar os ganhos espaciais proporcionados pelo mMIMO e os ganhos oriundos da acomodação de múltiplos UEs no mesmo RB proporcionados pelo PD-NOMA. Baseado no que foi apresentado sobre o PD-NOMA nas Subseções 1.1.4 e 1.1.5, é possível estender a análise para um cenário com mais de 2 UEs. Conforme já mencionado anteriormente, o NOMA opera com o agrupamento de usuários, sendo este um processo crucial para seu correto funcionamento.

Considerando as três condições descritas na Seção 1.1 e ainda as restrições de taxa mínima para os usuários, concluímos que o processo de clusterização ótimo que maximiza a soma das taxas do sistema PD-NOMA, é de natureza combinatória e somente resolvida por busca exaustiva (Li et al., 2015).

Assim, adotamos nesta Tese um modelo de agrupamento de usuários sub-ótimo de baixa complexidade, proposto por (Ali et al., 2016). Nesta proposta de clusterização DL e UL são tratados de forma distinta. Em comum temos a necessidade de ordenação dos UEs. Partindo do conhecimento *a priori* dos ganhos de canal complexo dos UEs, $\mathcal{K}$, estes são ordenados de forma decrescente em relação ao ganho de canal complexo, conforme Eq. (1.4), ou seja, o primeiro possui a melhor condição de canal e o último a pior condição de canal.

No DL, após o SIC, a taxa do usuário de maior ganho de canal de um *cluster* não está sujeito à interferência de outro *cluster* (inter-*cluster*); em vez disso, sua taxa de transferência depende de seu próprio ganho de canal e da potência de transmissão. Sendo assim, é benéfico distribuir os usuários de alto ganho de canal em uma célula em diferentes *clusters* NOMA, pois eles podem contribuir significativamente para a soma da taxa de transferência de um *cluster*.

Entretanto, para aumentar o desempenho dos usuários com baixo ganho de canal, é útil emparelhá-los com usuários de alto ganho de canal. A razão é que os usuários de alto ganho de canal podem atingir uma taxa mais alta mesmo com baixos níveis de potência, ao mesmo tempo em que disponibilizam uma grande fração de potência para usuários de canal fraco. Sendo assim, no DL é vantajoso emparelhar o usuário com maior ganho de canal e o usuário com menor ganho de canal no mesmo *cluster* PD-NOMA, enquanto o segundo usuário com maior ganho de canal e o segundo usuário com menor ganho de canal em outro *cluster* NOMA, e assim por diante, Figura 2.1.

É conveniente adotar que os usuários são clusterizados em pares, pois assim somente uma etapa de SIC é realizada pelo usuário com melhor condição de canal. Sendo assim o

conjunto $\mathcal{K}$ precisa ser dividido em dois subgrupos:

$$\begin{aligned}\mathcal{K} &= \mathcal{K}_H \cup \mathcal{K}_L, \quad \text{sendo,} \\ \mathcal{K}_H &= \{\text{UE}_1, \text{UE}_k, \dots, \text{UE}_{K/2}\} \quad \text{e} \\ \mathcal{K}_L &= \{\text{UE}_{K/2+1}, \dots, \text{UE}_{K-1}, \text{UE}_K\}.\end{aligned}\tag{2.4}$$

No UL, os sinais transmitidos pelos UEs experimentam ganhos de canal distintos. Para realizar o SIC na BS, precisamos manter a distinção em potência dos sinais recebidos. Sendo assim, o controle de potência de transmissão convencional não é viável e a não transmissão em máxima potência resultará na degradação da soma das taxas. A distinção entre os canais dos UEs dentro de um *cluster* NOMA é crucial para cancelar a interferência e assim, maximizar a taxa de transferência.

Com isso, no uplink NOMA, o usuário de maior ganho de canal não interfere nos usuários de canal fraco, pois sua interferência é cancelada pelo SIC. Portanto, este usuário pode transmitir com sua potência máxima para obter uma taxa maior. Sendo assim, é benéfico incluir os usuários de alto ganho de canal transmitindo com suas potências máximas em cada *cluster* NOMA, pois eles podem contribuir significativamente para a taxa de transferência deste *cluster*.

Com base no exposto acima, fica claro que a forma sub-ótima de baixa complexidade de clusterização para os usuários no NOMA é diferente no DL e no UL. Evidencia-se também a necessidade do sistema NOMA conhecer previamente os ganhos de canal dos UEs, realizar o ordenamento, fazer a divisão em grupos e agrupá-los adotando-se diferentes regras conforme mostrado na Figura 2.1, dependendo do sentido da transmissão do sinal (DL ou UL). Assim, ainda de modo a possibilitar a realização de somente uma etapa de SIC, definimos que o *cluster* será composto por 2 UEs (Ding et al., 2016), ou seja, serão divididos em dois grupos: a Figura 2.1 apresenta um exemplo em uma célula com 12 UEs a forma adotada para a clusterização no DL e no UL.

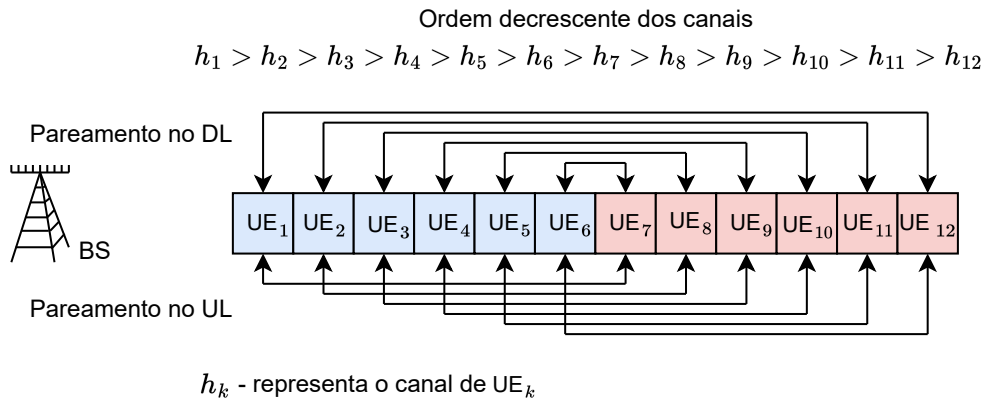


Figura 2.1 – Ilustração de exemplo de clusterização de 12 UEs no DL e no UL, quando o *cluster* é formado por 2 UEs.

Suposição 1: Assumimos nesta etapa que os UEs clusterizados dentro do mesmo *cluster* são atingidos pelo feixe formado após processamento e aplicação da técnica de pré-codificação. Para garantir que isso ocorra, assumirmos uma etapa de identificação da localização dos UEs em paralelo à etapa de clusterização nos sistemas PD-NOMA, implementada a partir do uso de técnicas de localização de UEs descritas, por exemplo, em (Zafari et al., 2019; Mohamed, 2020), com taxa de realização dependente da mobilidade dos dispositivos.

Na Figura 2.2 apresentamos 3 possíveis cenários de clusterização e formatação de feixe. A Figura 2.2 a) apresenta um cenário onde os UEs clusterização estão na condição LOS puramente AWGN. A Figura 2.2 b) apresenta um cenário misto onde um UE está na condição LOS e o outro está na condição NLOS. A Figura 2.2 c) apresenta um cenário onde os UEs estão na condição NLOS. No restante desta Tese, utilizamos a condição mais obstruída, conforme a Figura 2.2 c), por possuir a maior atenuação dos sinais e exigir o maior consumo de potência, escolha unicamente por ser uma condição mais complexa.

2.3 Formatação de feixe

Neste trabalho de Tese, consideramos o cenário de uma única célula, onde a BS possui M elementos de antena e adota a técnica de pré-codificação para criação de feixe (BF – *beamforming*) de forçamento a zero (ZF – *zero forcing*), técnica bastante difundida na literatura para o mMIMO, que funciona de modo a “forçar a zero” a interferência entre os usuários (Björnson et al., 2017b). Supondo *a priori* conhecimento perfeito do canal dos UEs ativos na célula, a BS cria a matriz de *beamforming* BF-ZF utilizando a seguinte regra:

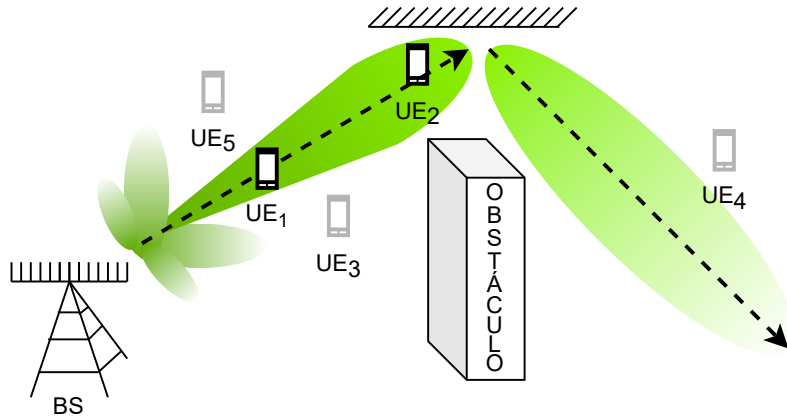
$$\mathbf{G}^{\text{ZF}} = \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1}, \quad (2.5)$$

que é a pseudo-inversa da matriz de canal $\mathbf{H}^H$.

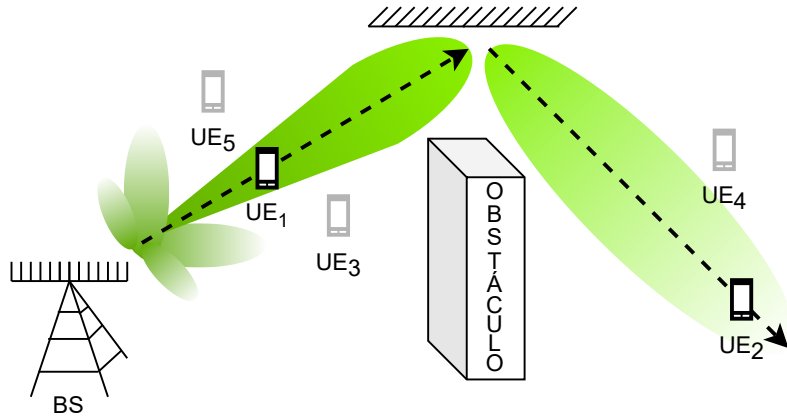
Assim, se calcularmos $\mathbf{H}^H \mathbf{G}^{\text{ZF}} = \mathbf{H}^H \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} = \mathbf{I}_K$, com $M > K$, ou seja, toda a interferência intracelular de outros UEs sendo transmitidos simultaneamente pelas M antenas da BS, será zero, enquanto os sinais transmitidos permanecem diferentes de zero. Sabendo-se que $\mathbf{H} \in \mathbb{C}^{M \times K}$, a aplicação da matriz de *precoding* linear ZF na BS (DL) resultará em K feixes de sinais exclusivos, ou seja, sem interferência intra-celular. Garantindo $M \geq 50$ ¹, a supressão da interferência é facilitada pela característica de propagação favorável no mMIMO; e em casos de canais não correlacionados, os sinais serão ainda favorecidos pelo ganho de *array*.

No NOMA, a formação do feixe funciona de forma diferente. Como o sinal a ser transmitido é uma sobreposição de sinais dos UEs reunidos em *clusters*, a pré-codificação

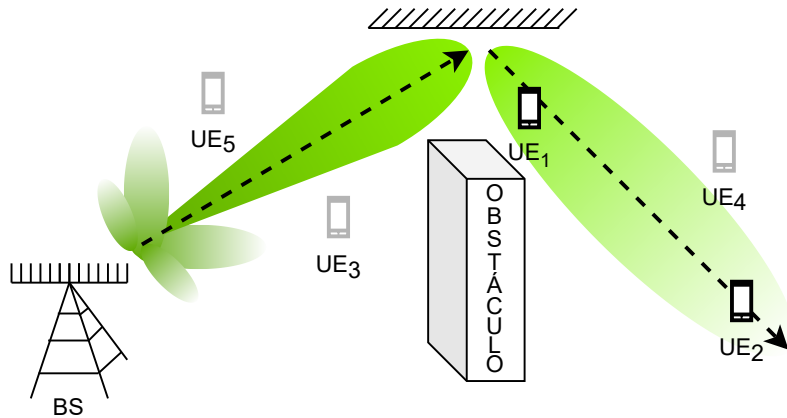
¹Quanto maior o número de antenas envolvidas na recepção dos sinais, mais eficiente será a supressão da interferência. Conforme demonstrado em (Björnson et al., 2017b) garantir esse número mínimo de antenas é suficiente para obter os ganhos das propriedades do mMIMO.



a) Condição de canal LOS (AWGN), para os UEs clusterizados.



b) Condição de canal LOS + NLOS, para os UEs clusterizados.



c) Condição de canal NLOS, para os UEs clusterizados.

Figura 2.2 – Representação dos possíveis cenários NOMA combinando canal LOS e NLOS, sendo UEs pertencentes ao mesmo *cluster* são atendidos pela mesmo *beam-forming*.

ZF deve funcionar para zerar a interferência sobre cada *cluster*, para isso, deve-se levar em consideração o conjunto de canais dos usuários pertencentes a cada *cluster*. Utilizando o exemplo na Figura 2.1, a matriz de canal é definida como:

$$\mathbf{H} \in \mathbb{C}^{M \times 12}, \quad (2.6)$$

considerando 2 UEs por *cluster* no DL, seguindo o que foi apresentado na Subseção 2.2, teremos $L = K/2$ *clusters*. O *cluster* l será composto por UE $_l$ e UE $_{K+1-l}$, sendo representado da seguinte forma:

$$\mathbf{H}_l \in \mathbb{C}^{M \times 2}, \quad \mathbf{H}_l = [\mathbf{h}_l, \mathbf{h}_{K+1-l}]. \quad (2.7)$$

Foram propostas na literatura algumas estratégias para aplicação do BF-ZF no PD-NOMA, sendo a sobreposição dos sinais no domínio da potência de cada *cluster* resulta na eliminação da interferência dos outros *clusters*. Em (Kim et al., 2013; Senel et al., 2019) foi demonstrado que é preciso escolher um canal para representar o *cluster* e criar a matriz de ZF; pode-se escolher o canal de um dos UEs integrantes do *cluster* para ser o canal equivalente do *cluster*. Nestes casos, o UE que usa o vetor de BF baseado no seu próprio canal não receberá nenhuma interferência dos outros feixes, ou seja, interferência entre *clusters*. No entanto, o outro UE do mesmo *cluster* precisa lidar com a interferência *inter-cluster*. Para maximização da taxa a escolha dos UEs dentro do conjunto $\mathcal{K}_H$, como o *cluster head*, é a mais adequada. Sendo assim, a matriz utilizada para o BF-ZF NOMA será:

$$\bar{\mathbf{H}} = \{\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4, \mathbf{h}_5, \mathbf{h}_6\}. \quad (2.8)$$

A interferência *inter-cluster* reduz a SINR e dificulta a decodificação do sinal recebido e uma vez que usuário com a melhor condição de canal executa o SIC, torna-se inútil escolher outra opção, senão aquela com maior ganho de canal (UEs dentro do conjunto $\mathcal{K}_H$) para criar a matriz de ZF para o *cluster*.

Outro projeto de BF-ZF baseado em *cluster* NOMA é proposto em (Ali et al., 2017). Este projeto permite a existência parcial de interferência *inter-cluster*, mas nesta abordagem a pré-codificação é realizada considerando o ganho de canal equivalente de cada *cluster*, em vez do ganho de canal de usuário específico. Para definir o canal equivalente é utilizada a decomposição em valores singulares (SVD – *singular value decomposition*) da matriz formada pelos UEs de cada *cluster* e depois essa matriz é multiplicada pela primeira coluna da matriz de vetores singulares à esquerda, obtendo um vetor de canal equivalente. Considerando o l -ésimo *cluster* com 2 UEs, aplicamos a decomposição SVD sobre a matriz de canal relativa a este l -ésimo *cluster*:

$$\mathbf{H}_l^H = \mathbf{U}_l \mathbf{\Sigma}_l \mathbf{V}_l^H. \quad (2.9)$$

No sistema mMIMO-NOMA cada BF-ZF atende a um *cluster*. Portanto, o ganho de canal equivalente da matriz $\mathbf{H}_l \in \mathbb{C}^{M \times 2}$, referente ao *cluster* l , será obtida como:

$$\bar{\mathbf{H}}_l = \mathbf{U}_l^{(1)H} \mathbf{H}_l^H, \quad (2.10)$$

sendo $\mathbf{U}_l^{(1)} \in \mathbb{C}^{K \times 1}$ a primeira coluna de $\mathbf{U}_l$ e $\bar{\mathbf{H}}_l \in \mathbb{C}^{1 \times M}$. Para este caso a matriz utilizada para o BF-ZF NOMA será:

$$\bar{\mathbf{H}} = [\bar{\mathbf{H}}_1^T, \bar{\mathbf{H}}_2^T, \dots, \bar{\mathbf{H}}_L^T]. \quad (2.11)$$

Como alternativa, (Cheng et al., 2018) propôs utilizar uma combinação linear dos canais para definir o canal equivalente do *cluster*. Ou seja:

$$\bar{\mathbf{H}}_l = (1 - \varrho) \cdot \mathbf{h}_l + \varrho \cdot \mathbf{h}_{K+1-l}, \quad (2.12)$$

caso, ϱ é constante de proporção da combinação linear, quando $\varrho = 0$ a matriz de canal equivalente será como a matriz Eq. (2.11).

Suposição 2: No sistema PD-NOMA, consideramos que será eliminada completamente a interferência *inter-clusters* de um *cluster* sobre o outro, devido à propriedade de propagação favorável do mMIMO, sob condições de CSI perfeitas.

2.4 Políticas de alocação de potência

As políticas de alocação de potência para o NOMA possuem algumas especificidades, caso seja proposto um problema de otimização para maximizar a soma das taxas de um sistema com L *clusters* compostos por 2 UEs cada:

$$\begin{aligned} \max_{\mathbf{p} \succeq 0} \quad & \sum_{l=1}^L R_l \\ \text{s.a.} \quad & \sum_{k=1}^K p_k \leq P_T, \end{aligned} \quad (2.13)$$

sendo, $R_l = R_{l,1} + R_{l,2}$ a soma das taxas do *cluster-l* e $\mathbf{p} = [p_1, \dots, p_k]$ o vetor das potências atribuídas a cada UE. Caso a única restrição seja a potência disponível, a alocação ótima de potências para cada UE no problema (2.13) consiste na solução via algoritmo *water-filling*, aplicável ao grupo de usuários com melhores condições de canal. Com isso, para maximização da soma das taxas do NOMA é necessário definir restrições de taxas mínimas, além de seguir as condições descritas na Subseção 1.1.

A seguir, apresentamos três políticas conhecidas e frequentemente aplicáveis ao problema de alocação de potência, mas devido às características inerentes ao NOMA, propomos algumas modificações ou considerações.

Ressaltamos que os sistemas mMIMO e NOMA exploram o mesmo número massivo de antenas na estação base, M . Portanto, devido ao efeito de enrijecimento do canal (Larsson et al., 2014; Marzetta et al., 2016) inerente à configuração MIMO massivo, o desvanecimento em pequena escala desaparece nas M antenas equipadas com um precodificador ZF linear com vetor definido na eq. (2.5). Assim, pode-se considerar apenas o desvanecimento em larga escala β_k como o principal parâmetro nas políticas de alocação de potência de sistemas baseados em número massivo de antenas.

2.4.1 Alocação de potência igualitária

A alocação de potência igualitária (EPA – *equal power allocation*) é implementada como uma estratégia simples e ingênua, onde são alocadas a todos os dispositivos a mesma potência. No mMIMO, todos os dispositivos são servidos com a mesma potência de transmissão independentemente de sua distância à BS. No sistema PD-NOMA, esta política de alocação de potência requer considerações adicionais. No UL PD-NOMA, se todos os dispositivos transmitem com a mesma potência e possuem coeficientes de canal distintos o SIC pode operar adequadamente, quando a P_{tol} é atingida.

No entanto, ao examinarmos o DL PD-NOMA, o cancelamento de interferência será realizado no UE com a melhor condição de canal e só será possível se $\Xi = 0$. Caso a Ξ não seja atingida outro método de cancelamento de interferência pode ser aplicado, como por exemplo o cancelamento paralelo de interferência (PIC – *parallel interference cancellation*). Levando-se em consideração a situação acima a potência total disponível para rádio frequência P_T é dividida igualmente entre os UEs no mMIMO e no NOMA e é definida como:

$$p_k = \frac{P_T}{K} \quad [\text{W}], \quad \forall k \in \mathcal{K}. \quad (2.14)$$

2.4.2 Alocação de potência proporcional ao inverso do canal

Ao contrário da técnica EPA, que aplica a mesma potência a todos os dispositivos, a alocação de potência pelo inverso do canal (PCIPA – *proportional channel inversion power allocation*) aplica mais potência aos dispositivos com as piores condições de canal, favorecendo a imparcialidade entre os dispositivos. Tal alocação de potência penaliza a média da soma das taxas em favor da justiça entre os usuários, de modo que, a P_{tol} deve ser sempre atendida.

A estratégia PCIPA aplicada ao mMIMO pode ser definida como:

$$p_k = P_T \frac{\beta_k^{-1}}{\sum_{k=1}^K \beta_k^{-1}} \quad [\text{W}], \quad \forall k \in \mathcal{K}, \quad (2.15)$$

enquanto o procedimento PCIPA aplicado ao NOMA segue duas etapas; na primeira etapa, a potência é alocada igualmente entre os L clusters:

$$p_{cl} = \frac{P_T}{L} \quad [\text{W}], \quad \forall l \in \mathcal{L}. \quad (2.16)$$

Em seguida, a potência de cada UE dentro de cada cluster l , contendo dois UEs por clusters é definida conforme:

$$p_{l,k} = p_{cl} \frac{\beta_{K+1-k}}{\beta_k - \beta_{K+1-k}}, \quad \text{e} \quad p_{l,K+1-k} = p_{cl} - p_{l,k}, \quad (2.17)$$

sendo $p_{l,k}$ é a potência alocada para o dispositivo k no cluster l , e $p_{l,K+1-k}$ é a potência alocada para o dispositivo $(K + 1 - k)$, também pertencente ao l -ésimo cluster.

2.4.3 Alocação de potência *water-filling*

Clássico: a aplicação do algoritmo (WF – *water filling*) no sistema mMIMO resulta em uma solução ótima (maximizada) da soma das taxas do sistema. No entanto, alguns dispositivos são retirados do serviço devido à condição de canal deteriorada. A estratégia de alocação de potência WF para mMIMO é descrita como (Khalighi et al., 2001):

$$\begin{aligned} \mu &= \frac{1}{|\mathcal{K}|} \left(P_T + \sum_{k \in \mathcal{K}} \frac{1}{\beta_k} \right), \\ P_T &= \sum_{k=1}^{K=|\mathcal{K}|} p_k, \quad \text{sendo} \quad p_k = \left(\mu - \frac{1}{\beta_k} \right)^+, \forall k \in \mathcal{K} \\ \text{e} \quad \mathbf{p} &= [p_1, p_2, \dots, p_{|\mathcal{K}|}], \end{aligned} \quad (2.18)$$

sendo o operador $(z)^+ = \max(0, z)$; μ é o nível d'água, p_k é a potência alocada para o k -ésimo UE e $|\mathcal{K}|$ é a cardinalidade do conjunto de usuários atendidos, i.e., com $p_k > 0$ é determinado pelo nível d'água, o qual depende da quantidade de potência total disponível, P_T . Observe que o valor máximo para a potência total disponível é definido como P_T [W]. O Algoritmo 1 descreve o procedimento WF clássico.

Algoritmo 1: WF clássico para mMIMO

Entrada: $\mathcal{K}$, P_T
1 $\mathcal{NP} = \{1\}$;
2 **enquanto** ($\mathcal{NP} \neq \emptyset$) **faça**
3 resolve Eq. (2.18) $\rightarrow \mathbf{p}$;
4 $\mathcal{NP} \leftarrow$ identifica posições nulas em $\mathbf{p}$;
5 $\mathcal{K} = \mathcal{K}/\mathcal{NP} \leftarrow$ exclui de $\mathcal{K}$ dispositivos rotulados em $\mathcal{NP}$;
6 **fim**
Saída: $\mathbf{p} = [p_1, p_2, \dots, p_{|\mathcal{K}|}]$

Em contrapartida, a aplicação direta do algoritmo WF no sistema NOMA implica em prejudicar a formação dos *clusters*, ou seja, os dispositivos presentes no conjunto $\mathcal{K}_L$ são efetivamente retirado do serviço, desfazendo o *cluster*. Para permitir alguma comparação, propomos uma modificação no WF clássico como veremos a seguir.

2.4.4 WF Modificado para NOMA

A aplicação do WF clássico no NOMA da mesma forma que é aplicado no mMIMO faz com que alguns pareamentos formados sejam desfeitos, pois após a aplicação do algoritmo WF, alguns dispositivos são retirados do serviço. Portanto, sugerimos modificar o procedimento WF clássico para ser aplicado ao PD-NOMA. As etapas do algoritmo Δ -WF são descritas a seguir.

No NOMA, a alocação de potência via Δ -WF tem duas etapas, na primeira etapa, a alocação de potência se dá entre os *clusters*. Assim, para evitar que o pareamento

seja desfeito, propomos a aplicação de WF com base nas diferenças de desvanecimento em larga escala dos dispositivos pareados, $\Delta\beta_l = (\beta_{l,k} - \beta_{l,K+1-k})$, desta forma estamos alocando mais potência aos pares com maior diferença de canal, o que favorece a soma das taxas do sistema. Na segunda etapa do procedimento, a potência é alocada entre os dispositivos dentro do mesmo *cluster* ou grupo; assume-se também que é possível o cancelamento de interferência, mesmo com $P_{\text{tol}} = 0$, quando pode ser aplicado o PIC. Sob essas considerações, o novo nível d'água na estratégia modificada de alocação de potência Δ -WF para NOMA é definido por:

$$\begin{aligned} \mu &= \frac{2}{|\mathcal{L}|} \left(P_T + \sum_{l \in \mathcal{L}} \frac{1}{\Delta\beta_l} \right), \\ P_T &= \sum_{l=1}^{L=|\mathcal{L}|} p_{cl}, \quad \forall l \in \mathcal{L} \\ \text{sendo } p_{cl} &= \left(\mu - \frac{1}{\Delta\beta_l} \right)^+, \\ \text{e } \mathbf{p}_L &= [p_{c1}, p_{c2}, \dots, p_{c|\mathcal{L}|}], \end{aligned} \tag{2.19}$$

sendo $\mathbf{p}_L$ o vetor de potência alocada por *cluster*. No segundo passo, a alocação de potência para ambos os dispositivos no l -ésimo *cluster* é simples e definida como:

$$p_{l,K+1-k} = p_{l,k} = \frac{p_{cl}}{2} \tag{2.20}$$

O algoritmo 2 sintetiza o procedimento de alocação de potência Δ -WF proposto, visando melhorar o SE do sistema NOMA.

Algoritmo 2: Δ -WF (modificado) para sistemas NOMA

Entrada: $\mathcal{K}_H, \mathcal{K}_L, P_T, \mathcal{L}$
1 $\mathcal{NP} = \{1\};$
2 **enquanto** ($\mathcal{NP} \neq \emptyset$) **faça**
3 resolve Eq. (2.19) $\rightarrow \mathbf{p}_L;$
4 $\mathcal{NP} \leftarrow$ identifica posições nulas em $\mathbf{p}_L;$
5 $\mathcal{L} = \mathcal{L}/\mathcal{NP} \leftarrow$ exclui de $\mathcal{L}$ dispositivos rotulados em $\mathcal{NP};$
6 **fim**
Saída: $\mathbf{p}_l = [p_{c1}, p_{c2}, \dots, p_{c|\mathcal{L}|}]$

2.4.5 Análise de complexidade

Em uma análise comparativa de complexidade, o algoritmo Δ -WF para alocação de energia no sistema NOMA (Algoritmo 2) executa duas operações adicionais simples em comparação com o procedimento WF clássico (Algoritmo 1): a) na eq. (2.19) a subtração $(\beta_{l,k} - \beta_{l,K+1-k})$; e b) a divisão por dois em (2.20). Além disso, os sistemas

NOMA *vs.* mMIMO requerem diferentes informações *a priori* para proceder de acordo com o procedimento PA.

Informações necessárias para Alocação de Potência: Para implementação das políticas de alocação de potência são necessárias informações prévias na BS, conforme definido na Tabela 2.1. Algumas dessas informações necessárias são obtidas através da etapa de transmissão piloto dedicada. Além disso, uma etapa anterior é assumida conhecida, na qual a estimativa da localização espacial² do *pathloss* de todos os dispositivos deve ser realizada. Com essa disponibilidade de informações *a priori*, a etapa de clusterização pode ser executada.

Tabela 2.1 – Informações prévias necessárias para a realização da alocação de potência em sistemas mMIMO x PD-NOMA.

PA		β_k	h'_k	K	$\mathcal{K}_H$	$\mathcal{K}_L$	p_k , eq.
EPA	mMIMO	—	$\sim$	$\blacktriangle^*$	—	—	(2.14)
	NOMA	—	$\sim$	$\blacktriangle$	—	—	(2.14)
PCIPA	mMIMO	$\blacktriangle^*$	$\sim$	$\blacktriangle$	$\blacktriangle^*$	$\blacktriangle^*$	(2.15)
	NOMA	$\blacktriangle$	$\sim$	$\blacktriangle$	$\blacktriangle^*$	$\blacktriangle$	(2.16), (2.17)
WF	mMIMO	$\blacktriangle^*$	$\sim$	$\blacktriangle$	$\blacktriangle^*$	$\blacktriangle^*$	(2.18), Alg. 1
Δ -WF	NOMA	$\blacktriangle$	$\sim$	$\blacktriangle$	$\blacktriangle^*$	$\blacktriangle$	(2.19, 2.20), Alg. 2

$\blacktriangle$ Informação necessária previamente; $*$ obtida via piloto; — Informação não necessária.

$\sim$ Informação importante, mas não necessária para alocação de potência.

h'_k - desvanecimento em pequena escala, Eq. (2.1).

2.5 Informações do estado do canal

Para operação adequada do NOMA em cenário de múltiplas antenas, onde se propõe aplicação de BF para eliminar interferência *intra-cluster* é necessário o conhecimento do canal dos usuários. A seguir apresentaremos duas propostas da literatura, sendo que, na primeira se estima uma combinação linear do canal dos UEs do *cluster* e o BF é criado utilizando essa combinação linear dos canais (Cheng et al., 2018), e na segunda se utiliza o canal do UE com melhor condição para o BF (Senel et al., 2019). Detalhamos as propostas a seguir.

Em (Cheng et al., 2018), foi proposto a reutilização do mesmo piloto para vários terminais na mesma célula. Em particular, a BS aloca o mesmo piloto para os UEs no mesmo *cluster*, sendo um desses previamente classificado no $\mathcal{K}_H$ e o outro no $\mathcal{K}_L$. Uma vez que os dois UEs estão usando o mesmo piloto e têm a mesma estatísticas de desvanecimento em pequena escala, a BS não consegue distinguir suas respostas de canal. No entanto, a

²Para o PD-NOMA operar corretamente nesta proposta, é necessário o alinhamento, ou minimamente que UEs do mesmo *cluster* sejam atingidos pelo feixe.

BS pode estimar uma combinação linear dos canais para ambos os terminais a partir desta transmissão de piloto.

Esta estimativa fornece uma descrição útil do canal combinado, particularmente, se o controle de potência for usado para nivelar as intensidades do sinal piloto dos dois terminais. Assim, o usuário de borda da célula, pertencente ao conjunto $\mathcal{K}_L$, realiza a decodificação no DL tratando a interferência de outros UEs como ruído, enquanto o usuário do centro da célula, pertencente ao conjunto $\mathcal{K}_H$, decodifica os dados do outro UE primeiro e realiza o SIC antes de decodificar seus próprios dados. Como os BFs são criados utilizando os canais de todos os usuários do *cluster*, o esquema proposto pode fornecer boas taxas de dados para todos os UEs, mas não para o sistema. A Figura 2.3 mostra a diferença estrutural entre os RBs dos sistemas mMIMO e esta proposta para o PD-NOMA (Cheng et al., 2018).

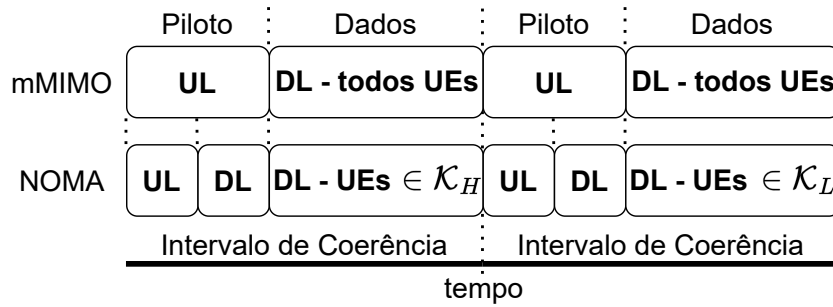


Figura 2.3 – Estrutura de quadro temporal baseado no tempo de coerência do canal: a estrutura de treinamento e transmissão de dados para os esquemas mMIMO e NOMA sob configuração TDD NLOS. Nesta configuração é obtida uma combinação linear dos canais dos UEs pertencentes ao *cluster*

A outra proposta em (Kim et al., 2013) e aplicada em (Senel et al., 2019), o BF é criado utilizando somente o canal dos UEs com melhores condições de canal, do conjunto $\mathcal{K}_H$, também denominado canal equivalente para o l -ésimo cluster. A Figura 2.4 compara a estrutura de transmissão de dados e de piloto ao longo do bloco de coerência do canal para ambos os sistemas, mMIMO e esta proposta de canal equivalente para o PD-NOMA. Sabendo-se que T_s é o tempo necessário para transmitir um símbolo de dados, com isso o intervalo de tempo de coerência do canal T é assumido como um múltiplo do período de símbolo de dados, $T = \iota \cdot T_s$.

A potência alocada para cada piloto na etapa de treinamento é suficiente, enquanto o número de pilotos e a porção dedicada do bloco de coerência para transmissão de dados são assumidos iguais em ambos os sistemas.

Como é possível observar na Figura 2.4, no sistema NOMA, a etapa de transmissão do piloto é dividida em duas partes, metade para os pilotos de UL e metade para pilotos DL; isso acontece porque para realizar o SIC, os dispositivos de centro de célula precisam conhecer os canais efetivos que são utilizados para o BF. Além disso, os vetores de formação

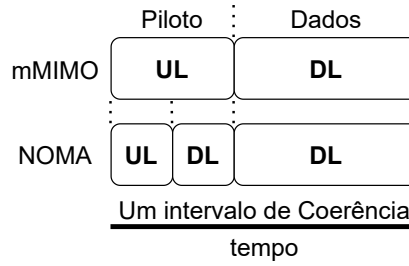


Figura 2.4 – Estrutura de quadro temporal tendo como base o tempo de coerência do canal: a estrutura de treinamento e transmissão de dados para esquemas mMIMO e NOMA sob configuração TDD NLOS. Nesta segunda configuração é obtido somente os canais dos UEs do conjunto $\mathcal{K}_H$.

de feixe são baseados nos dispositivos de centro de célula, produzindo taxas limitadas nos dispositivos de borda de célula. Em contrapartida, no esquema mMIMO, uma vantagem importante é que não há necessidade de pilotos DL, pois os canais efetivos utilizados para o BF são altamente previsíveis, ou seja, ganho e fase quase determinísticos devido ao efeito de endurecimento do canal (Senel et al., 2019). Para a operação adequada do NOMA a seguinte suposição deve ser atendida.

Suposição 3: No sistema NOMA, a potência alocada nos pilotos de DL é suficiente para o piloto ser recebido e detectado pelos UEs.

2.6 Conclusão do capítulo

Neste capítulo, apresentamos uma descrição dos principais conceitos de sistemas utilizando o esquema de comunicação mMIMO-NOMA. Discutimos procedimentos que envolvem o agrupamento de usuários inerentes ao NOMA, bem como sua diferença na aplicação em comunicação DL e UL. Constatamos que o agrupamento ótimo é computacionalmente difícil e apresentamos uma proposta sub-ótima para alcançar bons resultados e com baixa complexidade. A aplicação da precodificação linear ZF foi detalhada, demonstrando que a matriz de canais precisa ser reestruturada para aplicação do BF-ZF em sistema mMIMO-NOMA, discutimos que em cada estratégia de definição da matriz de canal equivalente um resultado diferente pode ser alcançado: a) maximizar a soma das taxas do sistema, ou b) melhorar a taxa recebida pelos usuários de borda de célula. Definimos algumas políticas de alocação de potência, entre elas, destacamos a estratégia WF, que foi modificada para a aplicação eficiente em sistemas mMIMO-NOMA. Apresentamos duas estratégias de aquisição da CSI, procedimento fundamental para o funcionamento adequado do sistema. As publicações [A] e [B] apresentam comparações e análises para este modelo de sistema, onde diversos conceitos apresentado aqui são discutidos.

3 MIMO Ultra Massivo

O mMIMO é uma tecnologia chave em sistemas de comunicação celular para aumentar a eficiência espectral, conforme Seção 1.4. Os maiores ganhos com MIMO massivo são alcançados quando a dimensão do arranjo de antenas é muito grande (Martinez et al., 2014). Isso tem motivado a introdução de novos tipos de implantação, onde *arrays* de dimensões extremamente grandes são implantados como parte de uma grande infraestrutura, por exemplo ao longo das paredes de edifícios em uma megacidade, em aeroportos, grandes shoppings ou ao longo da estrutura de um estádio (Martinez et al., 2014). Da mesma forma, surgiram grandes superfícies inteligentes envolvendo grandes superfícies eletromagnéticas (Hu et al., 2018). Tal sistema MIMO massivo com matrizes de antenas de dimensões extremamente grandes é denotado como XL-MIMO.

Com dimensões aumentadas do arranjo de antenas, as não estacionaridades espaciais aparecem em todo o arranjo devido a atributos de propagação eletromagnética, bem como a distância entre os usuários e o arranjo que se torna menor que a distância de Rayleigh. Em tais sistemas XL-MIMO, diferentes modelos de canal e algoritmos de receptor são necessários para modelar essa não estacionaridade.

Consideramos as propriedades não estacionárias através do conceito de região de visibilidade (VR - *visible region*). Uma VR está associada a um determinado usuário e é definida como a porção da matriz que um determinado usuário vê, ou seja, que é capaz de receber sinais do usuário. Esse comportamento do canal introduz uma dispersão inerente ao modelo do sistema, significando que o sinal transmitido de um usuário existe apenas em uma pequena porção do arranjo de antenas (Amiri et al., 2018).

Assim, em contraste com os modelos mMIMO comuns, a detecção do usuário pode ser feita processando apenas a VR de cada usuário. Usando essa importante propriedade do sistema, reduzindo os custos computacionais, realizando um processamento reduzido.

3.1 Modelo de sistema XL-MIMO

Consideramos uma configuração XL-MIMO em que a BS é formada por um arranjo de antenas extragrande, operando em TDD. Sem perda de generalidade, assumimos que uma matriz retangular uniforme (URA – *uniform rectangular array*) está instalada no plano y - z com o total de M elementos de antena, ou seja, $M_y \times M_z$, onde M_y e M_z denotam o número de elementos de antena ao longo dos eixos y e z , respectivamente, com a primeira linha do *array* começando na origem O e implantados ao longo do eixo de ordenadas, como na Figura 3.1.

Da mesma forma, como (Nishimura et al., 2020; Marinello F. et al., 2022), consideramos um modelo de grafo bipartido simplificado em XL-MIMO, o *array* é dividido em Q

$\mathcal{K}$ não possuem pilotos dedicados e caso desejem estar ativos devem ser atribuídos a um. Portanto, o número de iUEs na célula é $|\mathcal{K}| = K$. A BS apenas aloca pilotos para UE ativos e recupera os pilotos quando necessário. Os iUEs fazem uma tentativa de RA com probabilidade P_a .

Seja $\mathcal{M}$ o conjunto de todas as SAs da BS e $\mathcal{V}_k \subset \mathcal{M}$ o subconjunto de SAs visíveis para o k -ésimo dispositivo. Para compor a modelagem matemática, assumimos que o subconjunto $\mathcal{V}_k$ é gerado aleatoriamente, onde cada SA é visível com probabilidade P_b pelo equipamento k , como em (Nishimura et al., 2020; Marinello F. et al., 2022). Essa probabilidade simula a influência de obstáculos aleatórios e espalhadores no ambiente que interagem com os sinais transmitidos pelos/para os dispositivos, resultando em VRs. A existência de VRs é evidenciada por medições em (Carvalho et al., 2020).

Consideramos um sistema de comunicação multiusuário baseada em blocos de coerência tempo-frequência e os UEs inativos possuem blocos de acesso aleatório (RA – *random access*) para realizar tentativas de RA. Definimos τ_{RA} como o número de sequências piloto mutuamente ortogonais de RA que são possíveis $\mathbf{s}_1, \dots, \mathbf{s}_{\tau_{\text{RA}}} \in \mathbb{C}^{\tau_{\text{RA}} \times 1}$, cada um com comprimento τ_{RA} , e $\|\mathbf{s}_{\tau_{\text{RA}}}\|^2 = \tau_{\text{RA}}$.

Como a visibilidade de cada SA segue uma distribuição de Bernoulli com probabilidade de sucesso P_b , é importante observar que quando um SA está visível para um usuário, isso indica que o usuário pode transmitir/receber sinais para/deste SA com um valor de ganho de canal diferente de zero (Marinello F. et al., 2022).

Assumimos um cenário superlotado ($K \gg M$) onde K iUEs são distribuídos aleatoriamente dentro de uma célula quadrada com distância máxima d_{max} e distância mínima d_{min} , respectivamente, como na Figura 3.3.

3.2 Modelo de canal XL-MIMO

A seguir, modelamos o canal NLOS com base na posição de cada elemento de antena $m_{z,y}$ na URA, como na Fig 3.1. A localização específica do $m_{z,y}$ -ésimo elemento do *array* é representada pelo vetor 3D $\mathbf{w}_{m_{z,y}} = [0, (m_y - 1)d_m, h_{\text{array}} + (m_z - 1)d_m]$, onde $1 \leq m_y \leq M_y$, $1 \leq m_z \leq M_z$, e h_{array} é a altura onde o *array* está instalado. As dimensões físicas da URA ao longo dos eixos y e z são $D_y = (M_y)d_m$ e $D_z = (M_z)d_m$, onde d_m é a distância entre os elementos. A posição k do usuário é definida pelo vetor 3D $\mathbf{pos}_k$, a distância do k -th UE ao elemento de antena específico $m_{z,y}$ é determinada por $r_{k,m_y,m_z} = \|\mathbf{w}_{m_{z,y}} - \mathbf{pos}_k\|$ (Lu; Zeng, 2022).

A distância em linha reta perpendicular ao eixo y do k -ésimo UE até a superfície onde o *array* está instalado r_k (linha tracejada vermelha), e o ângulo de elevação θ_k , formado por r_k e a linha entre UE $_k$ e a altura de instalação do *array* (linha tracejada vinho), e o ângulo de azimuth ϕ_k , formado pela linha que liga UE- k ao alinhamento de $m_{x,y}$ (linha tracejada verde). Então a localização de UE- k é $\mathbf{pos}_k = [r_k \Psi_k, r_k \Phi_k, r_k \Omega_k]^T$

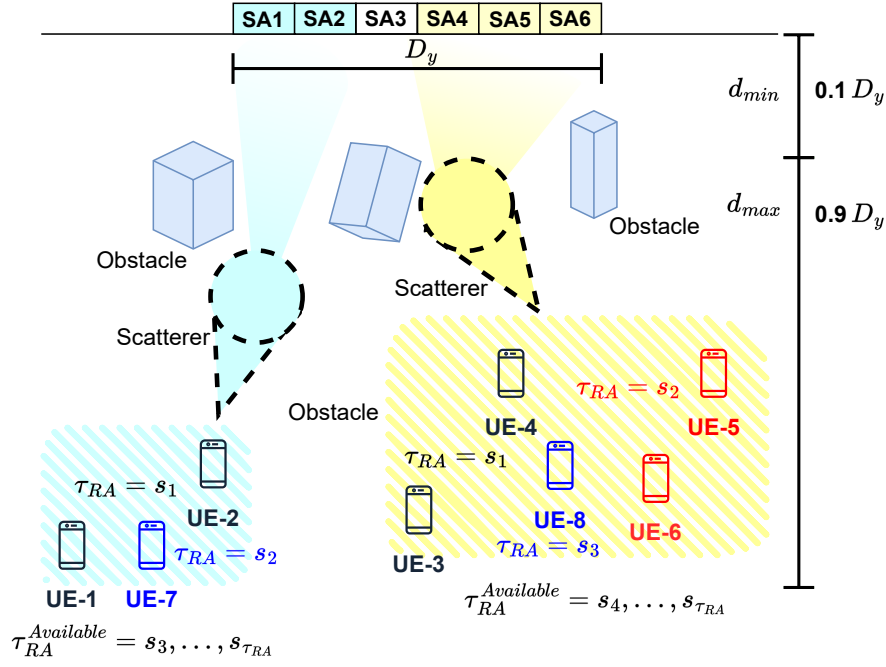


Figura 3.3 – Sistema de comunicação XL-MIMO URA composto por Q -SAs contendo M_q antenas, as regiões coloridas em cada SA definem UEs visíveis, a probabilidade de sucesso de visibilidade P_b para cada SA.

com $\Psi_k \triangleq \sin \theta_k \cos \phi_k$, $\Phi_k \triangleq \sin \theta_k \sin \phi_k$ e $\Omega_k \triangleq \cos \theta_k$, $1 \leq k \leq K$.

O modelo de canal segue o descrito na Seção 1.4.2 e que da mesma forma que (Nishimura et al., 2020; Marinello F. et al., 2022), adotamos a média do desvanecimento de larga escala para o k -ésimo usuário no q -ésimo SA, $\beta_k^{(q)} = \frac{1}{M_q} \sum_{m=1}^{M_q} \beta_{k,m}$. Seja $\mathbf{h}_k^{(q)} \in \mathbb{C}^{M_q \times 1}$ o vetor do canal de desvanecimento de Rayleigh entre UE $k \in \mathcal{K}$ e SA q , seguindo $\mathbf{h}_k^{(q)} \sim \mathcal{CN}(0, \beta_k^{(q)} \mathbf{I}_{M_q})$.

3.3 Validação do modelo de canal

A dimensão aumentada da abertura do arranjo de antenas XL-MIMO, faz com que a distância de Rayleigh, limite teórico onde a modelagem do canal muda de campo perto para campo distante, esteja a uma distância bastante grande do *array*.

Este limite teórico é calculado pelo quadrado da abertura do *array*, conforme Eq. (3.1), o canal de comunicação sem fio dos usuários dentro deste limite precisam ser modelados como *campo próximo*:

$$r_{Rayl} = \frac{2D^2}{\lambda}, \quad (3.1)$$

onde D é a abertura do arranjo de antenas e λ é o comprimento de onda do sinal.

A divisão do arranjo de antenas em SA, conceito aplicado em (Alves; Abrão, 2023), reduz a abertura considerada do arranjo de antenas, e consequentemente a distância de Rayleigh se torna menor o que pode validar, quando garantida a localização do usuário

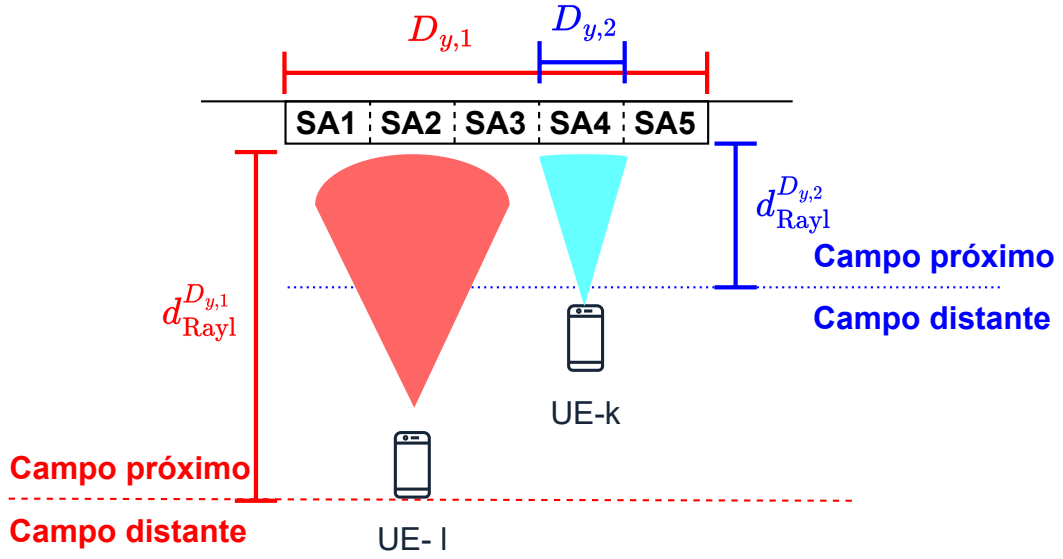


Figura 3.4 – Modelo esquemático que demonstra a relação entre a abertura do arranjo de antenas e a distância de Rayleigh.

após esse limite, considerarmos a modelagem do canal como campo distante (*far field*). A Tabela 3.1 apresenta alguns exemplos de abertura do arranjo de antenas e dependendo do número de divisões em SA (Q), qual a distância teórica de Rayleigh.

	$Q = 1$	$Q = 2$	$Q = 5$	$Q = 10$
$D_y = 100\text{m}$	160.000m	40.000m	6.400m	1.600m
$D_y = 10\text{m}$	1.600m	400m	64m	16m

Tabela 3.1 – Relação entre abertura do arranjo de antenas e distância de Rayleigh.

3.4 Conclusão do Capítulo

Neste capítulo, discutimos o sistema MIMO massivo com matrizes de antenas com dimensões extra grandes, o que chamamos de XL-MIMO. Apresentamos o conceito de VR e a possibilidade de processamento semi-distribuído em SA.

Verificamos também, que a grande dimensão do arranjo de antenas faz com que o limite teórico da distância de Rayleigh, relacionado a modelagem do canal, atinja valores extremamente altos. A divisão em SA, além de possibilitar o processamento distribuído, diminui o limite teórico da distância de Rayleigh e em alguns casos valida a adoção da modelagem de canal de campo distante. Os conceitos apresentados neste Capítulo são utilizados no desenvolvimento no Capítulo 5, e nas Contribuições [C], [D] e [E]. Estes conceitos são fundamentais para o entendimento dos resultados obtidos.

4 Eficiência de Recursos em mMIMO-NOMA sob diferentes Políticas de Alocação de Potência

Neste capítulo desenvolvemos uma análise extensa e comparativa sobre o desempenho da SE do sistema mMIMO em relação ao sistema NOMA com múltiplas antenas, variando a carga de usuários do sistema sob três diferentes métodos de alocação de potência bastante conhecidos na literatura e fazendo uso da área sob a curva de SE ($\mathcal{S}^{\text{SYST}}$) dos sistemas como uma métrica eficaz, útil e justa de desempenho e eficiência. Em paralelo desenvolvemos uma análise de EE utilizando um modelo detalhado de consumo de energia, conforme descrito na Seção 1.5.2, com parcelas de consumo de energia fixas e variáveis relacionadas ao circuito e levando em conta o número de antenas e dispositivos presentes no sistema, fornecendo uma análise extensa e comparativa tanto do NOMA quanto do sistema mMIMO em cenários de operação realistas e fazendo uso da área sob a curva de EE ($\mathcal{E}^{\text{SYST}}$) destes sistemas. Por último uma análise sobre a faixa de *trade-off* SE/EE é desenvolvida considerando vários carregamentos de dispositivos ($\rho = \frac{K}{M}$), sendo possível verificar a diferença entre as taxas alcançadas pelos dispositivos, fornecendo evidências da capacidade de operação do NOMA para atender a um número maior de dispositivos do que o sistema mMIMO. Foram feitas algumas suposições em favor da operabilidade do sistema NOMA, o que reforça a superioridade do mMIMO em certos cenários de aplicação. Por outro lado, melhores resultados do NOMA precisam ser analisados e compreendidos mais profundamente em cenários mais específicos.

4.1 Modelo de Sistema

As Figuras 4.1.(a) e (b) são representações geométricas genéricas de uma célula circular, onde no centro, a BS está equipada com um conjunto de M antenas, em um cenário típico de propagação NLOS. Os K usuários estão distribuídos aleatoriamente pelo interior da célula e estão divididos espacialmente em dois subgrupos $\mathcal{K}_L$ and $\mathcal{K}_H$, Eq. (2.4).

A Figura 4.1.(a) representa um cenário de UL. Observe que ao incluirmos uma região em anel ("zona-morta") separando, UEs de borda dos UEs de centro de célula, garante-se uma disparidade mínima de potências recebidas, onde mesmo sob uma política de alocação de potência igualitária (EPA), garante-se uma disparidade mínima nas potências recebidas dos UEs no mesmo *cluster*. Caso existam UEs presentes na região "zona-morta", estes devem ser atendidos por outro método de múltiplo acesso, ou então, são retirados do serviço.

A Figura 4.1.(b) representa um cenário de DL, que em contrapartida, na política EPA impossibilita a aplicação do SIC, para esses casos lançamos mão do PIC, conforme demonstrado na Subseção 1.1.2. Em todas as simulações consideramos um modelo de desvanecimento de bloco onde os recursos de tempo-frequência são divididos em intervalos de tempo de coerência (T), nos quais os canais permanecem constantes e a resposta em frequência plana, e são medidos em múltiplos períodos de transmissão de símbolo (T_s). Os sistemas mMIMO e NOMA considerados neste cenário exploram os M elementos de antenas omnidirecionais da BS.

Suposição 4: Assumimos que o cancelamento é perfeito, sendo que apenas uma etapa de cancelamento por *cluster* é executada, pois apenas 2 UEs por *cluster* são admitidos.

4.2 SE-EE em sistemas DL NOMA e mMIMO

Nesta Subseção, discutiremos as figuras de mérito SE e EE nos sistemas NOMA e mMIMO. A operação do sistema NOMA requer um processo de pareamento de UEs de forma que os ganhos de canal dos dispositivos de um mesmo *cluster* sejam adequadamente diferentes, possibilitando o uso do domínio de potência, conforme descrito na Subseção 2.2. Enfatizando o que está descrito na Subseção 2.3, verificaremos que o processo de cancelamento de interferência via BF no NOMA ocorre de forma diferente que no mMIMO.

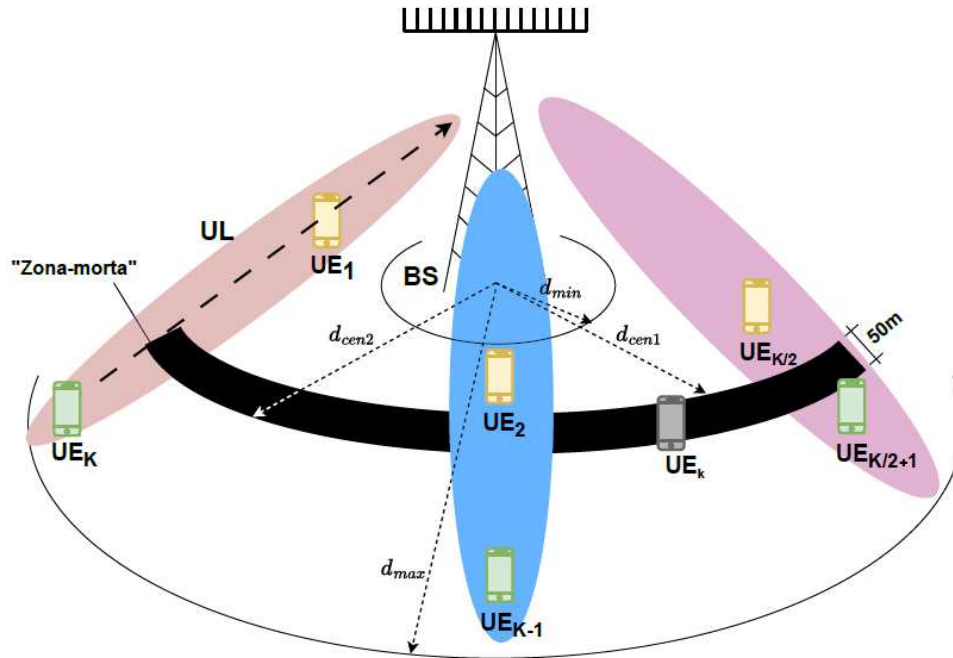
4.2.1 Taxa de dados no DL NOMA via ZF

Os dispositivos são divididos em dois conjuntos, conforme descrito na Eq. (2.4) e esses grupos são representados, Eq. (4.1), por seu coeficiente de desvanecimento em larga escala (perda de percurso) e são agrupados em pares formando um *cluster*, como discutido na Figura 4.1. O *cluster* l é formado por um dispositivo do conjunto de centro de célula $\mathcal{K}_H$ e um dispositivo do conjunto de borda de célula $\mathcal{K}_L$. Sendo assim, podemos representar os dois conjuntos como:

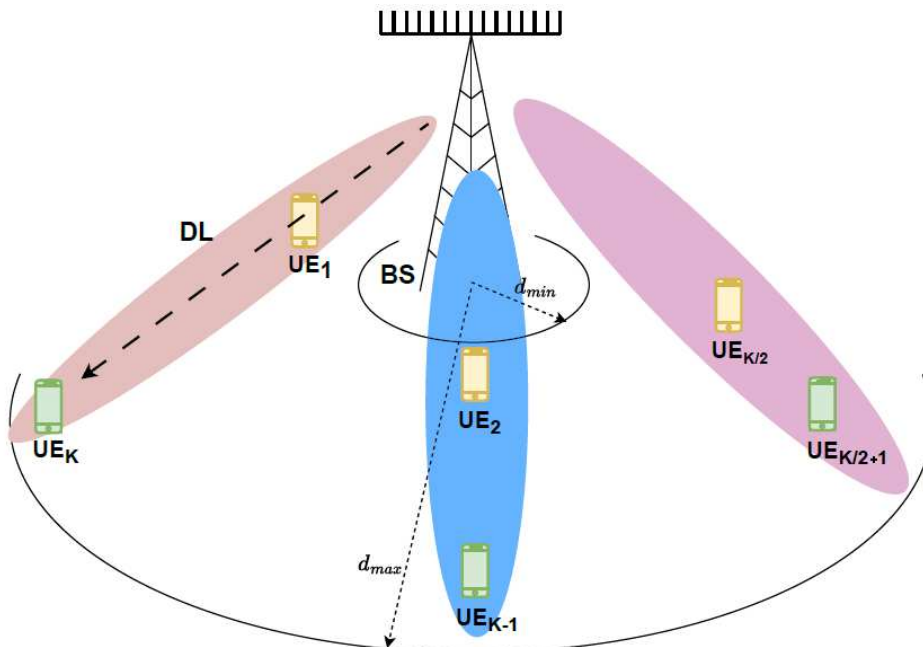
$$\begin{aligned}\mathcal{K}_H &= \{\beta_1 > \beta_2 > \dots > \beta_{K/2}\}, \quad (\text{conjunto de UEs - centro da célula}), \\ \mathcal{K}_L &= \{\beta_K < \beta_{K-1} < \dots < \beta_{K/2+1}\}, \quad (\text{conjunto de UEs - borda da célula}).\end{aligned}\tag{4.1}$$

O pareamento de usuários adotado na Eq. (4.1) foi descrito na Seção 2.2; este método forma *clusters* com a maior diferença possível nos coeficientes de canal entre os dispositivos que ainda não foram pareados. Portanto, seguindo os procedimentos de clusterização definidos, as SINRs do k -ésimo e $K + 1 - k$ -ésimo UEs pertencentes ao *cluster* l , são definidas como:

$$\text{SINR}_{l,k} = \frac{\beta_k p_k |\mathbf{h}_k^T \mathbf{g}_k|^2}{\beta_k \sum_{k' \neq k}^K p_{k'} |\mathbf{h}_k^T \mathbf{g}_{k'}|^2 + 1},\tag{4.2}$$



a) Cenário mMIMO-NOMA UL, região de "Zona-morta" está em destaque.



b) Cenário mMIMO-NOMA DL, na política EPA o NOMA opera com PIC.

Figura 4.1 – Modelo do sistema indicando a formação de pares no sistema NOMA.

$$\text{SINR}_{l,K+1-k} = \frac{\beta_{K+1-k} p_{K+1-k} |\mathbf{h}'_{K+1-k} \mathbf{g}_k|^2}{\beta_{K+1-k} \sum_{k' \neq k}^K p_{k'} |\mathbf{h}'_{K+1-k} \mathbf{g}_{k'}|^2 + 1}, \quad (4.3)$$

sendo $\mathbf{h}'_k$ o desvanecimento (termo curto) do canal, $\mathbf{g}_k$ o vetor de *precoding* ZF, discutido na Eq. (2.5) e β_k representa o ganho de percurso, Eq. (1.19).

Optamos pelo BF-ZF, e vamos utilizar os vetores de canal dos UEs presentes no conjunto $\mathcal{K}_H$, conforme descrito na Eq. (2.8). Esta opção é a que obtém melhor resultado de soma das taxas para o DL do sistema. Com isso, em cada *cluster*, os dispositivos de borda de célula tratam toda a interferência como ruído e decodificam seus símbolos de dados, não obtendo o ganho de *array*, enquanto o dispositivo de centro de célula pode decodificar os símbolos de dados do dispositivo de borda de célula e executar SIC, removendo efetivamente a interferência do dispositivo de borda de célula sob a **Suposição 2**.

Para realizar o cancelamento de interferência, o UE de centro da célula precisa ser capaz de decodificar o sinal de dados destinado ao UE da borda da célula, dentro do mesmo *cluster*, ou seja, o SINR ergódica do sinal para o dispositivo da borda da célula, $\text{SINR}_{l,K+1-k}$, recebida pelo dispositivo k , dentro do mesmo *cluster*, definido como $\text{SINR}_{l,(k,K+1-k)}$, deve ser maior ou igual ao SINR ergódico do k -ésimo dispositivo de centro de célula, ou seja, estamos considerando até $\Xi = 0$ dB, conforme descrito em Subseção 2.4.1. Portanto, dado a configuração do piloto *uplink* (mMIMO e NOMA) e *downlink* (NOMA) e assumindo CSI perfeito em todos os receptores, e admitindo a **Suposição 4**, a seguinte condição deve ser satisfeita (Senel et al., 2019; Cheng et al., 2018):

$$\mathbb{E}[\text{SINR}_{l,(k,K+1-k)}] \geq \mathbb{E}[\text{SINR}_{l,k}], \quad (4.4)$$

onde,

$$\text{SINR}_{l,(k,K+1-k)} = \frac{\beta_k p_{K+1-k} |\mathbf{h}'_k \mathbf{g}_k|^2}{\beta_k \sum_{k' \neq K+1-k}^K p_{k'} |\mathbf{h}'_k \mathbf{g}_{k'}|^2 + 1}. \quad (4.5)$$

Neste caso, a condição em (4.4) deve ser satisfeita selecionando as potências de transmissão apropriadamente, pois quanto maior a diferença da SINR melhor ocorrerá o cancelamento da interferência. Nos casos em que os sinais dos usuários estejam nos mesmos patamares, ou seja $\Xi = 0$ dB, o PIC será utilizado.

A taxa ergódica alcançável de dispositivos no cluster l , isto é, dispositivo k do subconjunto $\mathcal{K}_H$ e dispositivo $K+1-k$ do subconjunto $\mathcal{K}_L$, sob as **Suposições 1–4**, é dado pela taxa de contribuição ergódica do dispositivo mais próximo a BS:

$$R_k^{\text{NOMA}} = \tau \mathbb{E} [\log_2 (1 + \text{SINR}_k)], \quad \forall k \in \mathcal{K}_H \quad (4.6)$$

em [bits/s/Hz], e para o dispositivo mais afastado:

$$R_{K+1-k}^{\text{NOMA}} = \tau \mathbb{E} [\log_2 (1 + \text{SINR}_{K+1-k})], \quad \forall k \in \mathcal{K}_L \quad (4.7)$$

onde $\tau = (1 - \frac{K \cdot T_s}{T})$, é a porção de cada bloco de coerência do canal (T) que é utilizada para transmissão de dados.

Assumindo CSI perfeito, pré-codificação ZF para eliminação total da interferência *inter-clusters*, e usando os resultados da teoria de matrizes aleatórias (Marzetta et al., 2016), no l -ésimo *cluster* NOMA, a soma das taxas alcançável no l -ésimo *cluster* é obtida conectando a eq. (4.2), (4.6) e (4.7):

$$R_{\text{cl-}l}^{\text{NOMA}} = \tau \mathbb{E} \left[\log_2 \left(1 + \bar{M} \beta_k p_k \right) \right] + \tau \mathbb{E} \left[\log_2 \left(1 + \frac{\beta_{K+1-k} p_{K+1-k}}{\beta_{K+1-k} p_k + 1} \right) \right], \quad (4.8)$$

$\forall k \in \mathcal{K}$ e $\bar{M} = M + 1 - K/2$ e por simplicidade, esta expressão desconsidera possível amplificação de ruído do precodificador ZF. Assim, o sistema NOMA pode operar até $K < 2M - 1$ e o usuário de borda, por não conhecer o vetor de BF, não possui ganho de *array*. Por fim, a soma das taxas do sistema é definida por:

$$\mathcal{R}^{\text{NOMA}} = \sum_{l=1}^L R_{\text{cl-}l}^{\text{NOMA}}.$$

4.2.2 Taxa de dados DL mMIMO com ZF

No sistema mMIMO com pré-codificação ZF, a taxa alcançável ergódica para o dispositivo k é dada por:

$$R_k^{\text{M-MIMO}} = \tau \mathbb{E} [\log_2 (1 + \text{SINR}_k)], \quad [\text{bits/s/Hz}], \quad (4.9)$$

onde SINR_k é definido em (4.2). Portanto, a equação de taxa alcançável com a aplicação do BF-ZF no mMIMO acima torna-se:

$$R_k^{\text{M-MIMO}} = \tau \mathbb{E} [\log_2 (1 + (M - K) p_k \beta_k)], \quad [\text{bits/s/Hz}], \quad (4.10)$$

onde $(M - K)$ é obtido usando a teoria da matriz aleatória, representando o ganho de matriz coerente do sinal recebido (Marzetta et al., 2016). Sob pré-codificação linear e combinadores, o sistema mMIMO opera consistentemente quando $K < M$. Finalmente, a soma das taxas do sistema é definida simplesmente por:

$$\mathcal{R}^{\text{M-MIMO}} = \sum_{k=1}^K R_k^{\text{M-MIMO}}.$$

As equações do sistema mMIMO foram exaustivamente investigadas na literatura e podem ser encontradas em (Yang; Marzetta, 2013; Marzetta et al., 2016).

4.3 Resultados Numéricos

4.3.1 Eficiência Espectral

A análise de desempenho da SE do mMIMO e do NOMA foi realizada aumentando o número de dispositivos dois a dois até o limite de carregamento ($\rho = \frac{K}{M} = 2$), com a

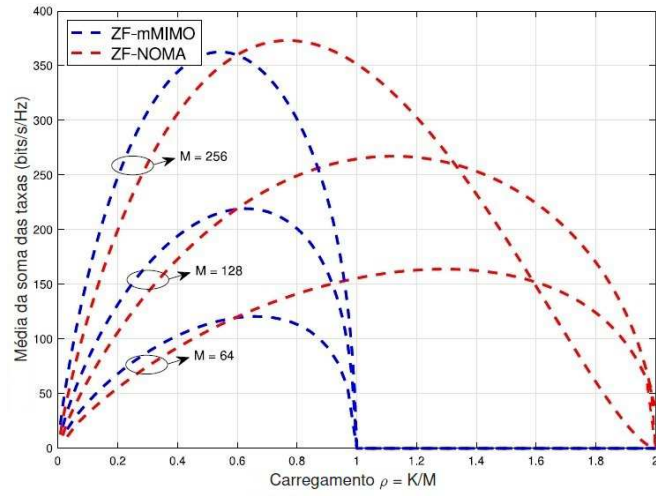
ressalva do mMIMO com ZF só operar até $K < M$. Os resultados consideram $M = 64, 128$ e 256 antenas BS. Os parâmetros de simulação de sistema e de canal estão descritos na Tabela 4.1, utilizando as Eqs. (4.8) e (4.10).

Na Figura 4.2.(a) os resultados de SE são obtidos quando a potência de RF disponível é alocada seguindo a estratégia EPA, onde são alocadas a mesma potência de rádio frequência para todos os dispositivos. Nesta estratégia, operando em DL, os sinais dos usuários chegam aos dispositivos no mesmo nível de potência o que impossibilita a aplicação do SIC, para esses casos, admitimos o cancelamento via PIC. O sistema mMIMO superou o NOMA operando com PIC em todas as situações em que o carregamento $\rho < 0,6$. No entanto, o sistema NOMA com PIC atinge um SE maior do que o mMIMO para cada cenário M quando o carregamento dos dispositivos aumenta, $\rho > 0,6$. O SE-médio máximo é 373 [bits/s/Hz], sendo obtido com ZF-NOMA $M = 256$ antenas e $\rho \approx 0,76$. Além disso, pode-se inferir que o mMIMO não funciona com carregamento maior que 1, devido ao ganho de *array* chegar a 0 no carregamento total dos dispositivos, enquanto o NOMA, dadas as simplificações e considerações adotadas no modelo, funciona adequadamente até que o carregamento atinja $M \cdot \zeta$, onde ζ é o número de dispositivos por *cluster*.

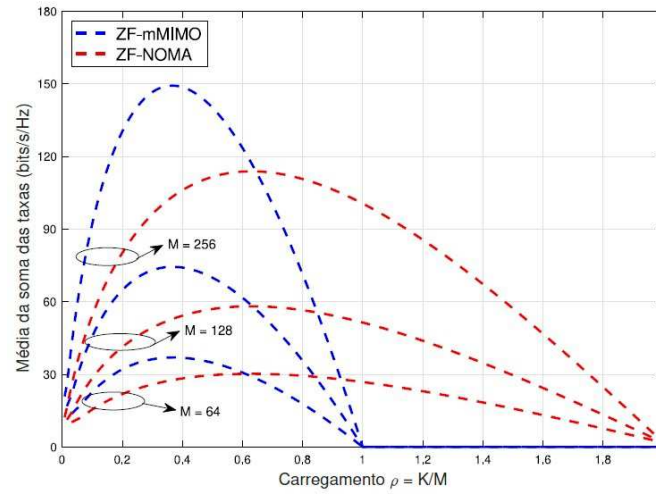
Tabela 4.1 – Parâmetros de simulação sistema mMIMO-NOMA.

Parameter	Value
BS antenas	$M = 64, 128$ e 256
Max. # dispositivos na célula	$K = \zeta \cdot M$ (NOMA) $K = M$ (mMIMO)
Carregamento da célula	$\rho = K/M$
RF Total - Potência disponível	$P_{\text{RF}} = 1\text{W}$
Pares NOMA / <i>Clusters</i>	$L = K/\zeta = K/2$
NOMA dispositivos por <i>cluster</i>	$\zeta = 2$
# antenas por dispositivos	1
Distância da borda da célula	$d_{\text{max}} = 350$ m
Posição dispositivos no $\mathcal{K}_H$	$[d_{\text{min}}; d_1] \in [50; 100]$ m
Posição dispositivos no $\mathcal{K}_L$	$[d_2; d_{\text{max}}] \in [150; 350]$ m
Ganho de array mMIMO	$M - K$
Ganho de array NOMA k_H	$M + 1 - K/2$
Duração do símbolo de dados	T_s
Tempo de coerência	$T = \iota \cdot T_s, \quad \iota = 512$
Channel	
Expoente de Pathloss	$\xi = 3,78$
Atenuação na referência, $d_0 = 1\text{m}$	$\beta_0 = 90$ [dB]
# Realizações de Monte-Carlo	1000

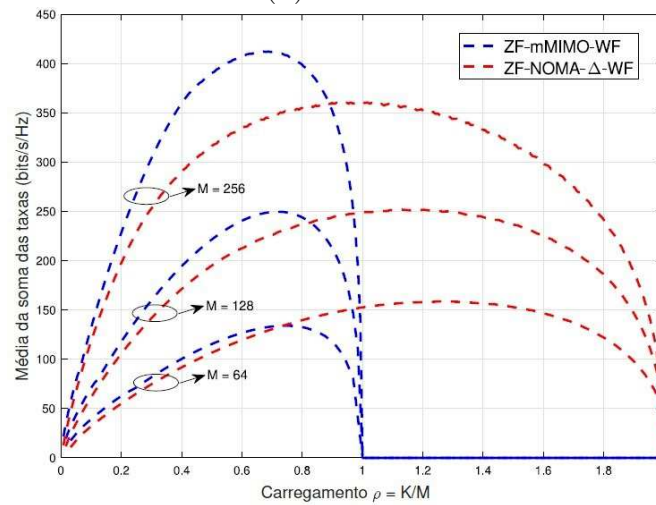
A Figura 4.2.(b) mostra os resultados de SE, alcançados nos sistemas mMIMO e NOMA quando o método PCIPA é aplicado para alocar a potência de RF disponível por dispositivo ao longo das antenas BS. O SE-médio máximo no sistema mMIMO supera a contrapartida NOMA até que o carregamento ρ exceda $\approx 0,62$ para as três configurações



(a) EPA



(b) PCIPA



(c) WF clássico no ZF-mMIMO-WF e Δ -WF no ZF-NOMA

Figura 4.2 – Média da soma das taxas para o carregamento de usuários $0 < \rho \leq 2$, considerando os quatro métodos de alocação de potência: EPA, PCIPA, WF e Δ -WF. Sendo a média obtida após 1000 distribuições aleatórias de usuários.

de antena BS, $M = 64, 128$ e 256 . Esta técnica PA fornece mais potência para dispositivos com a pior condição de canal, tornando o resultado de SE com valores máximos abaixo da EPA.

A Figura 4.2.(c) apresenta o algoritmo WF convencional aplicado ao mMIMO. Sob tal abordagem de alocação de potência, destacamos que a formação de pares é inviável no sistema NOMA. De fato, o algoritmo WF é capaz de maximizar o SE do sistema, pois aloca mais potência para dispositivos com melhores condições de canal, enquanto dispositivos com condições de canal ruins (abaixo do nível d'água) são retirados do serviço.

O algoritmo WF clássico foi adaptado para o sistema NOMA atuando na retirada de serviço um par de dispositivos do mesmo *cluster* por vez, até que o nível d'água seja positivo (vide Seção 2.4.3). Essa adaptação revela melhorias substanciais para a média da soma das taxas quando M é baixo em comparação com WF clássico em mMIMO. O procedimento de alocação de potência Δ -WF preserva a formação do agrupamento de pares no sistema NOMA, alocando mais potência para o agrupamento com maior diferença entre os coeficientes de desvanecimento em larga escala.

A Figura 4.2.(c) retrata o SE médio máximo ≈ 361 [bits/s/Hz], que é alcançado sob $\rho = 1$ ($K \approx 256$ dispositivos) quando o WF modificado é implantado no sistema NOMA. Além disso, quando o número de antenas BS é menor ($M = 64$ ou 128), o NOMA atingiu o maior pico de SE comparado ao mMIMO, por exemplo, para $M = 64$ antenas, o pico do SE mMIMO ocorre no carregamento $\rho \approx 0,7$, enquanto o NOMA SE atinge o pico em $\rho \approx 1,2$. No entanto, à medida que o número de antenas BS cresce, a vantagem do NOMA SE diminui.

Número de dispositivos ativos após o procedimento PA. A Figura 4.3 mostra o número de dispositivos ativos após a aplicação dos métodos PA: nos algoritmos EPA e PCIPA, todos os dispositivos permanecem ativados. No entanto, no sistema WF mMIMO clássico quando o número de dispositivos aumenta além de $\rho \approx 0,25$, a metade dos dispositivos é descartada; enquanto em Δ -WF NOMA PA, a porcentagem de dispositivos ativos é sempre maior, ex. superior a 70% para antenas $\rho \approx 1,1$ e $M = 256$, o pior caso.

A Figura 4.4 apresenta as superfícies da média da soma das taxa em termos de resultados $SE \times \rho \times M$ alcançados por NOMA com EPA, mMIMO com WF e NOMA com Δ -WF modificado. Na parte inicial do carregamento, $\rho < 0,65$, o clássico WF em ZF-mMIMO alcança os melhores resultados até o carregamento (superfície rosa). Quando o número de antenas é baixo como $M = 64$ e ρ está entre $0,7$ e $1,8$, o EPA aplicado ao NOMA (superfície verde) obteve resultados superiores, com a ressalva de operar com PIC. Além disso, quando $M = 128$ e $0,8 < \rho < 1,6$, o ZF-NOMA-EPA alcança resultados superiores (superfície verde). Para um maior número de antenas na BS, ou seja, $M = 256$ apenas em um curto intervalo de carregamento de dispositivos, $0,86 < \rho < 0,97$, o ZF-NOMA-EPA atinge resultados superiores de SE. Finalmente, quando $\rho > 0,97$ o Δ -WF modificado atinge resultados competitivos (superfície azul).

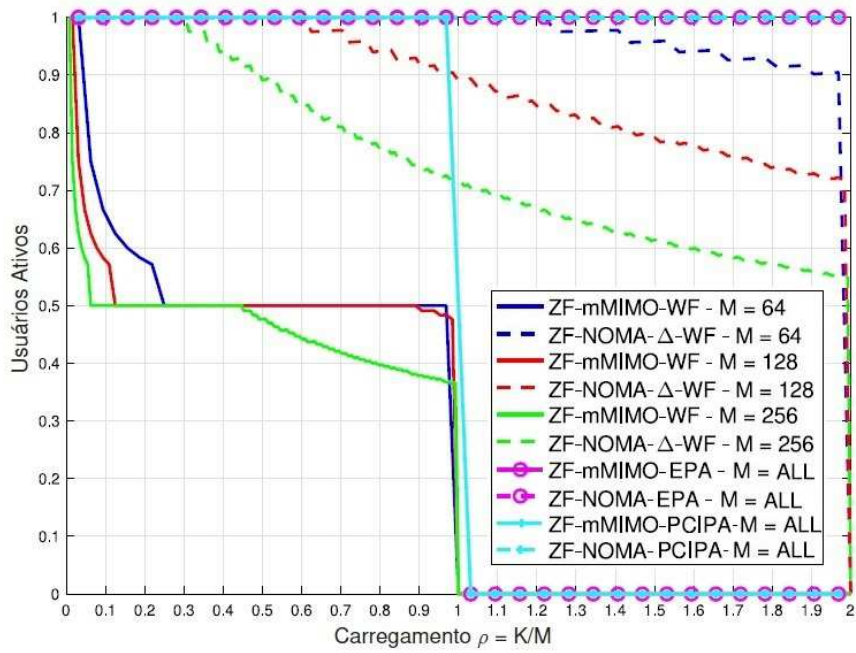


Figura 4.3 – Média de dispositivos ativos após PA *versus* carregamento na faixa $0 < \rho \leq 2$. Sendo a média obtida após 1000 distribuições aleatórias de usuários.

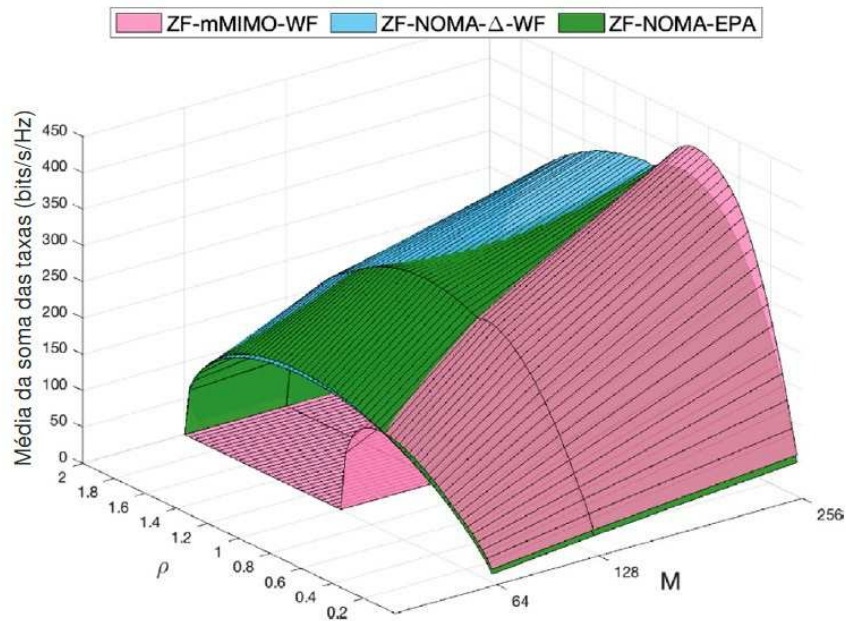


Figura 4.4 – Média da soma das taxas pelo carregamento ρ e M .

4.3.2 Índice de Justiça de Jain

A Figura 4.5. mostra as curvas de justiça (*fairness*) atingidas pelos sistemas NOMA e mMIMO com as políticas EPA, PCIPA, WF e Δ -WF quando o carregamento de dispositivos cresce até $\rho = 2$. Neste índice valores próximos a 1 indicam taxa próximas a igualdade, quanto mais distante de 1 e próximas a 0 maior a diferença entre as taxas alcançadas. Figura 4.5.(a) mostra o índice de justiça de Jain quando a política EPA é usada, o sistema mMIMO apresenta melhores resultados $\mathcal{F}$ do que NOMA para $\rho < 1$, por outro lado, o NOMA pode atingir $\mathcal{F}_M^{\text{NOMA}} \approx 0,5$ em quase todos os carregamentos de dispositivos, independentemente do M .

Figura 4.5.(b) revela o índice de Jain quando aplicado o método PCIPA, apesar do resultado SE ser menor neste método PA, o mMIMO obtém o melhor resultado de justiça, mantendo o $\mathcal{F}_M^{\text{mMIMO}}$ sempre acima de 0,85, já no NOMA fica abaixo de 0,5.

O índice de justiça de Jain quando WF e δ -WF são representados na Figura 4.5(c), é intrínseco a esses algoritmos alocar mais potência para o dispositivo com melhores condições de canal, o que causa parcialidade entre dispositivos. Neste método é possível observar uma influência significativa do número de antenas M na BS e o resultado de equidade.

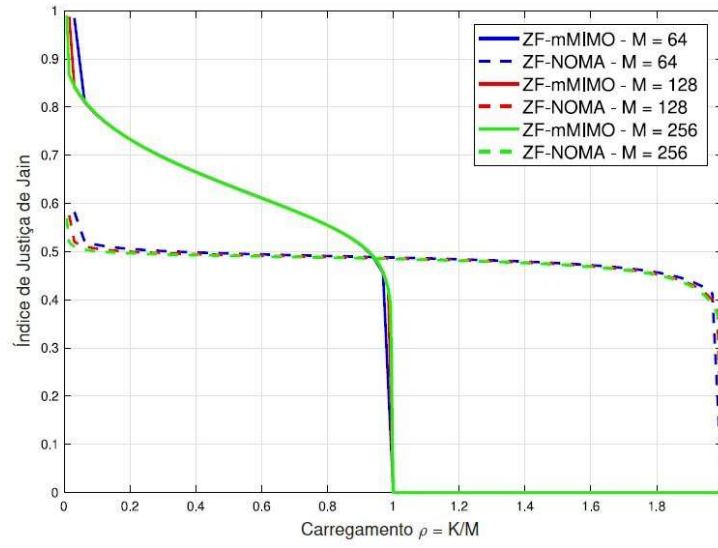
4.3.3 Eficiência Energética

A eficiência energética (EE) é outra importante figura de mérito usada para comparar o desempenho dos sistemas. Nesta Seção, adotaremos o modelo de consumo de potência descrito na Subseção 1.5.2, seguindo a Eq. 1.28. A Tabela 4.2 apresenta os valores dos parâmetros adotados para a análise e comparação do EE discutidos nesta Subseção. A Figura 4.6 mostra o desempenho do EE com os procedimentos EPA, PCIPA, WF e Δ -WF PA, considerando as mesmas três quantidades de antenas.

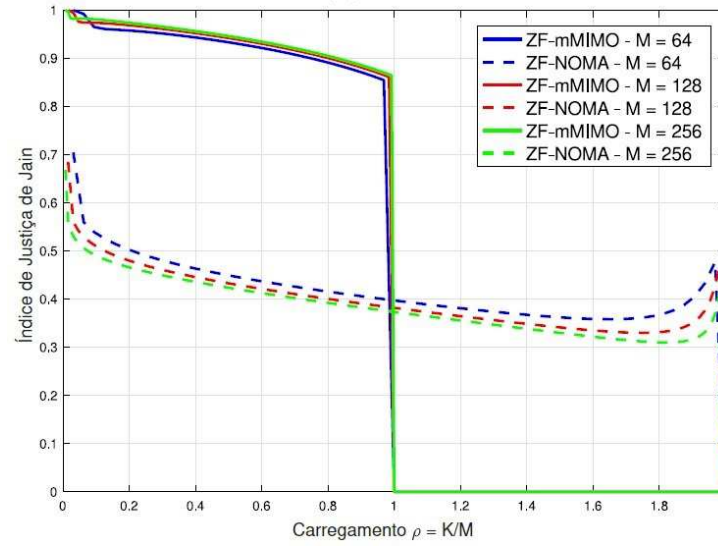
Tabela 4.2 – Valores dos parâmetros adotados na análise de EE (Björnson et al., 2014).

Parametro	Valores
Infraestrutura de Suporte	$P_0 = 2 \text{ W}$
Oscilador simples	$P_{OSC} = 2 \text{ W}$
Codificação e modulação	$P_{COD} = 4 \text{ W}$ por dispositivo
Decodificação e demodulação	$P_{DEC} = 0,5 \text{ W}$ por dispositivo
Potência do receptor	$P_{RX} = 0,3 \text{ W}$ por dispositivo
Potência transmitida	$P_{TX} = 1 \text{ W}$ por antena
Eficiência do amplificador de Potência	$\vartheta = 0,3$
Operações/Joule	$I = 10^9$ oper. por joule

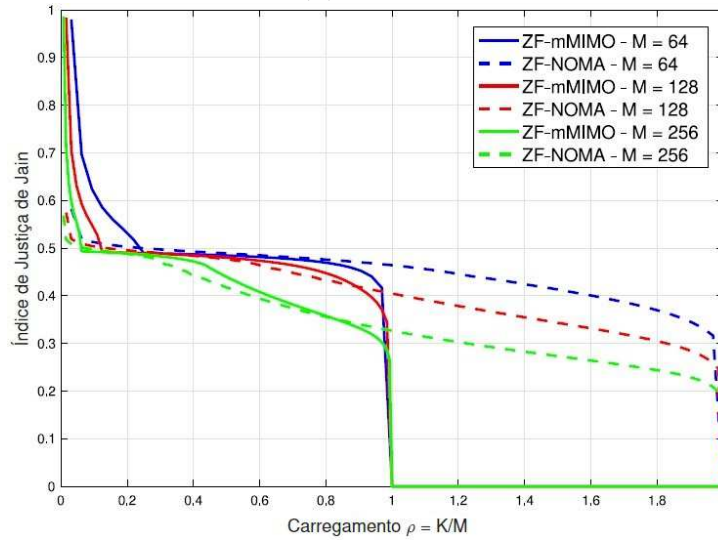
A Figura 4.6.(a) mostra o desempenho do EE com EPA, onde o NOMA opera com PIC. Neste método todos os sinais são transmitidos com a mesma potência. O EE-médio no sistema mMIMO supera o NOMA em torno de 13% a 20%. Foi possível observar que



(a) EPA



(b) PCIPA



(c) WF e Δ -WF

Figura 4.5 – A justiça dos sistemas NOMA e do mMIMO nas quatro políticas alocação de potência: (a) EPA; b) PCIPA; (c) WF and Δ -WF. Sendo a média obtida após 1000 distribuições aleatórias de usuários.

a adição de antenas na BS aumenta o consumo de potência, prejudicando o resultado de EE. Novamente, sob elevada carga de dispositivos $0,6 < \rho < 2,0$, o NOMA com o cancelamento paralelo PIC supera o sistema mMIMO.

A Figura 4.6.(b) revela o desempenho de EE com PCIPA. Nesse método, mais potência é alocada para dispositivos com coeficiente de canal ruim, resultando em desempenho de EE ruim/reduzido para os sistemas NOMA e mMIMO, atingindo um máximo de 0,25 e 0,16 [bits/W] para mMIMO e NOMA, respectivamente. Em geral, o EE máximo alcançado pelo mMIMO é cerca de 50% maior que o NOMA. No entanto, para carregamento de dispositivos $\rho \geq 0,65$, o NOMA supera o desempenho de EE do mMIMO.

A Figura 4.6.(c) mostra o desempenho do EE no mMIMO com WF clássico e no NOMA com algoritmo Δ -WF. É possível confirmar a superioridade em termos de eficiência energética do mMIMO dentro da faixa em que ele opera consistentemente, ou seja, $0 < \rho < 1$. Entre todos os cenários de sistemas analisados, o mMIMO operando com o WF clássico atinge a maior EE. Em geral, o mMIMO alcança melhores resultados de EE do que o NOMA para $\rho \leq 1$, mas por outro lado, o NOMA pode atender o dobro de dispositivos atendidos pelo sistema mMIMO.

A Figura 4.7 resume os melhores resultados de EE em uma plotagem de superfície para o mMIMO com WF superando o NOMA em toda a faixa de carregamento onde ele opera de forma consistente, ou seja, $0 < \rho < 1$. Para carregamento de dispositivos $\rho > 1$, o NOMA opera sob menor EE até o carregamento $\rho = 2$. Além disso, considerando o menor número de antenas na BS, o NOMA com EPA, operando com PIC, superou o NOMA com Δ -WF modificado; apesar disso, à medida que o número de antenas na BS aumenta, o NOMA com Δ -WF atinge resultados de EE marginalmente superiores.

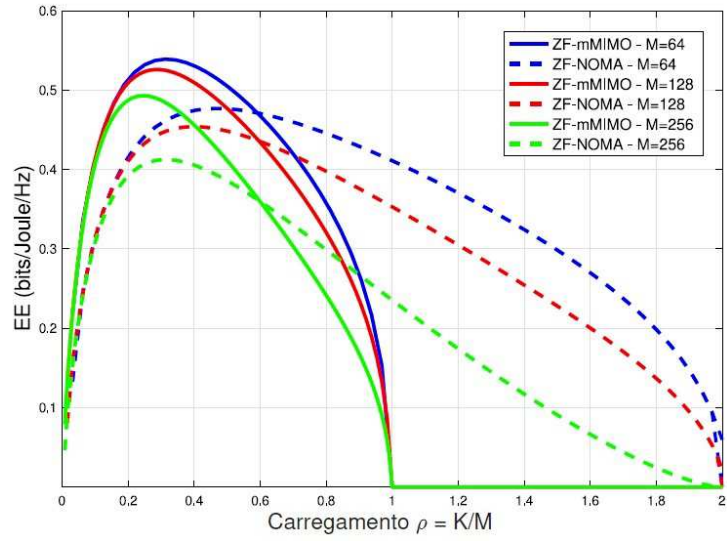
4.3.4 Área sobre as Curvas SE e EE

Outra perspectiva de análise é utilizar conforme descrito na Subseção 1.5.3, a área sob as curvas de SE e EE normalizada pela quantidade de antenas na BS. Considerando as áreas correspondentes sob as curvas SE e EE das Figuras 4.2 e 4.6 e aplicando as Eqs. (1.29) e (1.30) obtivemos a Figura 4.8.

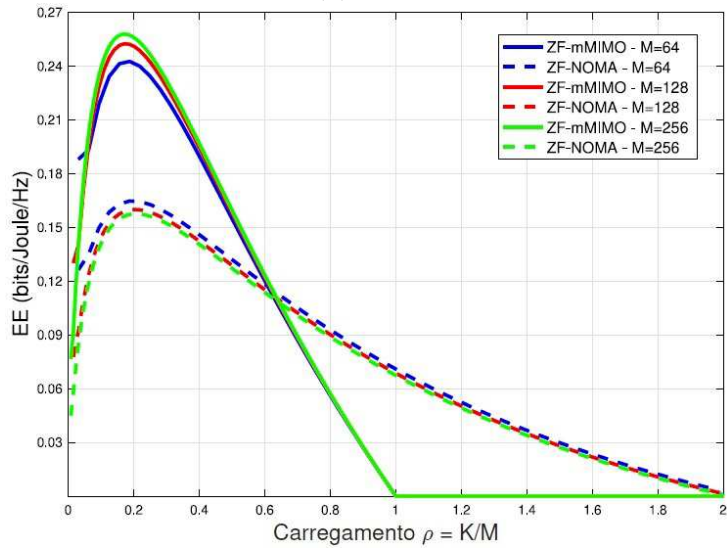
Ao analisarmos a Figura 4.8. (a), no ponto de vista SE, e considerando a política EPA, caso onde propomos o NOMA operando com PIC, a maior diferença entre as razões de área sob a curva de SE é alcançada quando o número de antenas BS é $M = 64$:

$$\mathcal{S}_{M=64}^{\text{NOMA}} \approx 2,7 \cdot \mathcal{S}_{M=64}^{\text{mMIMO}}.$$

Observe que quando o número de antenas M cresce a relação acima diminui. Da mesma forma, considerando a política WF a tendência de ganho se mantém. Em contraste, considerando a política PCIPA, a proporção permanece praticamente a mesma.



(a) EPA



(b) PCIPA

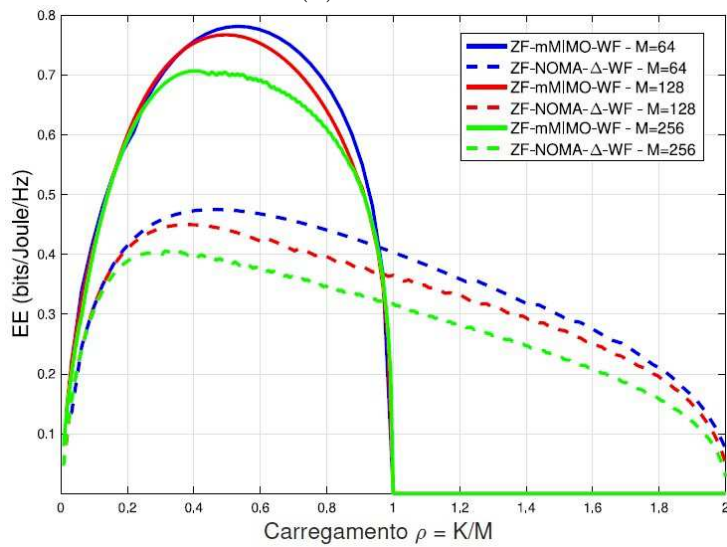

 (c) WF e Δ -WF

Figura 4.6 – EE para NOMA e mMIMO nas quatro políticas alocação de potência: (a) EPA; b) PCIPA; (c) WF and Δ -WF. Sendo a média obtida após 1000 distribuições aleatórias de usuários.

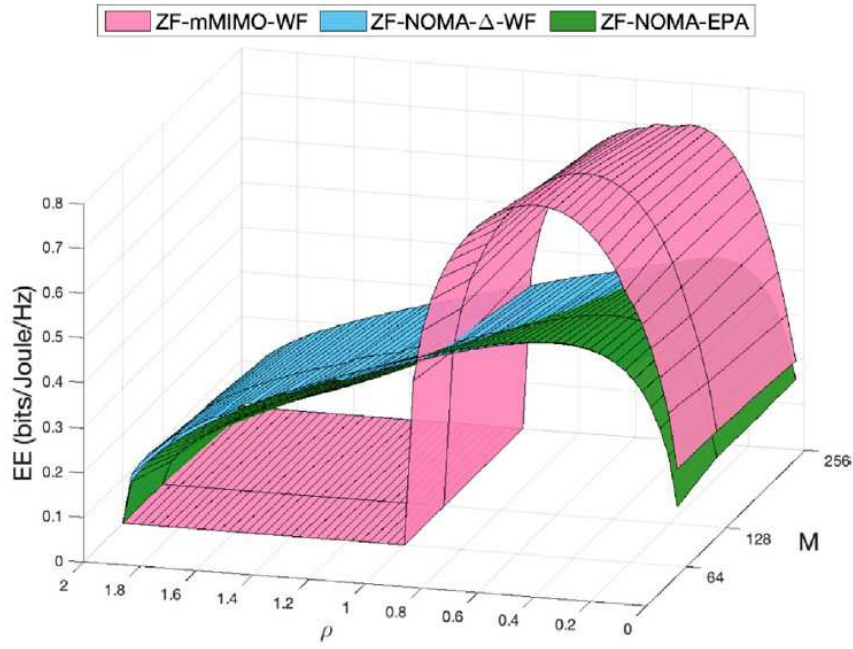


Figura 4.7 – EE com carregamento ρ e M .

Além disso, na Figura 4.8. (b), considerando agora a perspectiva de EE, sob a política da EPA, a taxa mais alta é alcançada quando a BS está equipada com $M = 64$ antenas:

$$\mathcal{E}_{M=64}^{\text{NOMA}} \approx 1,8 \cdot \mathcal{E}_{M=64}^{\text{mMIMO}}.$$

Como se pode concluir, o NOMA em quase todos os cenários é mais eficiente espectral e energeticamente do que o mMIMO em uma faixa muito ampla de carregamento de dispositivos $0 > \rho \geq 2$, aproximadamente, em média, 80% em termos de eficiência energética, e 170% mais eficiente em termos de eficiência espectral.

4.3.5 Eficiência de Recurso em função do carregamento e o compromisso SE-EE

Os sistemas NOMA e mMIMO são analisados em termos do compromisso entre SE e EE, ou seja, eficiência de recursos (RE - *resource efficiency*), considerando carregamento de dispositivos aumentando até $\rho = 2$. Nas Figuras 4.9, 4.10 e 4.11 definimos graficamente a faixa de carregamento de dispositivos entre a máxima EE e a máxima SE. Com isso, plotamos o fronte de Pareto, região onde se encontra o melhor compromisso SE-EE para cada configuração de antena da BS M . Nas Figuras da esquerda, o eixo y da esquerda (cor vermelha) representa o SE-médio, o eixo y da direita (cor azul) mostra o EE-médio e o eixo x é o carregamento de usuários na célula. Nas Figuras da direita plotamos o fronte de Pareto para as curvas apresentadas a esquerda.

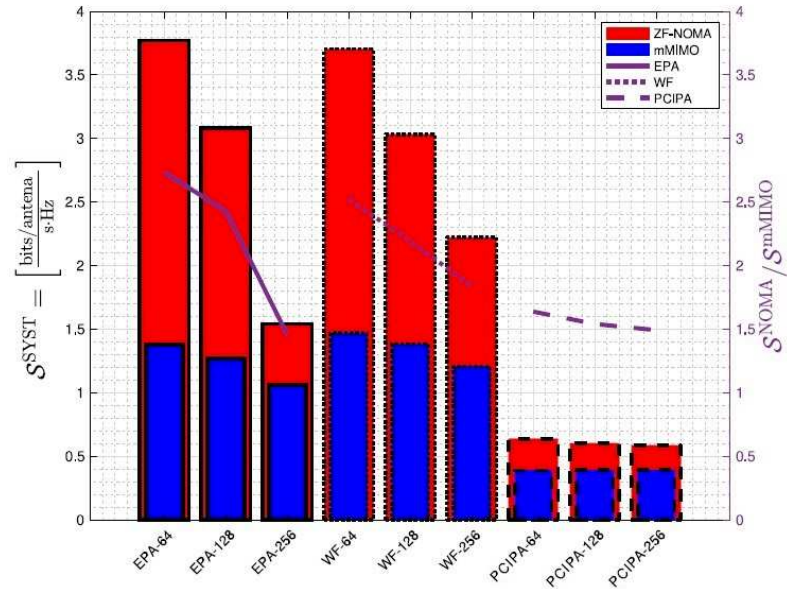
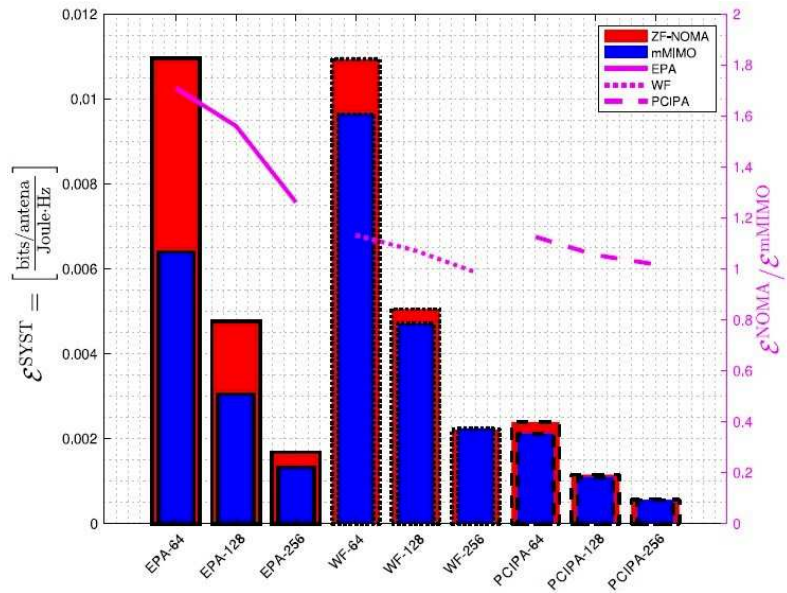

 (a) $\mathcal{S}$ - Gráfico de barras das Área sob curvas SE.

 (b) $\mathcal{E}$ - Gráfico de barras das Área sob curvas EE.

Figura 4.8 – Área sob a curva sistemas NOMA e mMIMO nas quatro políticas alocação de potência: (a) Curvas de SE; b) Curvas de EE. Sendo a média obtida após 1000 distribuições aleatórias de usuários. O eixo da esquerda é referente as barras e o eixo da direita referente as linhas.

A Tabela 4.3 resume os valores aproximados do ponto de cruzamento entre as curvas EE-SE e apresenta o ponto de carregamento de dispositivos deste cruzamento com a porcentagem de usuários ativos após alocação de potência e o índice de justiça de Jain. A Figura 4.9 revela os resultados quando $M = 64$, o NOMA com a política EPA, com as simplificações consideradas no modelo, incluindo a proposta de cancelamento de interferência via PIC, alcançou o maior ρ na faixa de compromisso SE-EE e também o maior SE, ou seja, pode operar com maior carregamento e garantir um bom compromisso SE-EE; por outro lado, mMIMO com WF clássico obteve no cruzamento das curvas o melhor EE, porém a porcentagem de dispositivos ativos é em torno de 0,65. A Figura 4.10 descreve os resultados quando $M = 128$, o NOMA com a política EPA, com as simplificações adotadas no modelo, incluindo a proposta de cancelamento de interferência via PIC, alcançou o maior na faixa de operação com bom compromisso SE-EE, em contrapartida, o mMIMO com WF clássico que alcançou o melhor resultado de EE no cruzamento das curvas EE-SE mas o carregamento de dispositivos ficou em torno de 0,62. A Figura 4.11 mostra os resultados quando $M = 256$, o NOMA com Δ -WF alcançou o maior ρ na faixa de compromisso SE-EE, e mais uma vez o mMIMO com WF alcançou melhor resultado de EE e de SE mas com ρ abaixo 0,6. É possível demonstrar que o aumento de antenas na BS melhora o resultado SE, por outro lado piora o resultado EE.

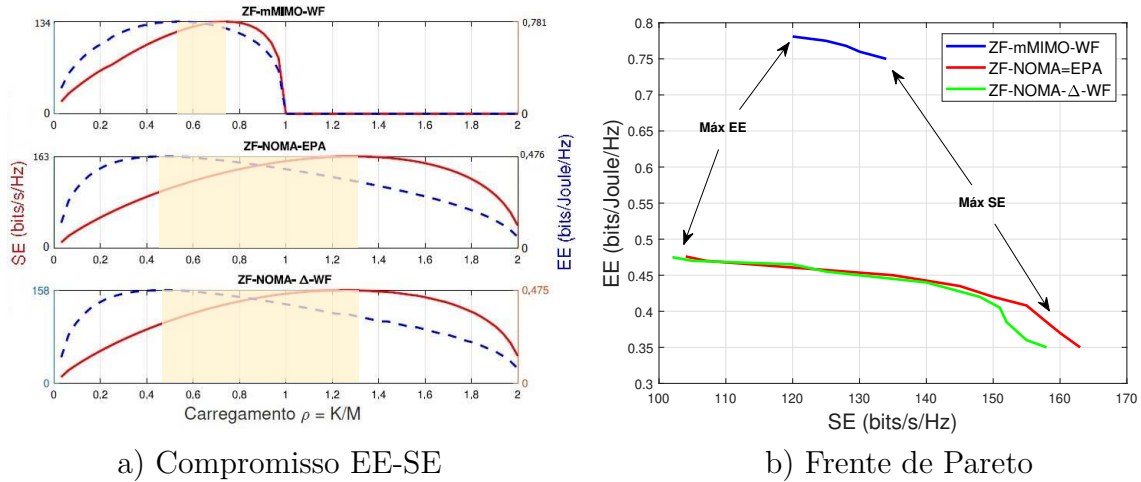


Figura 4.9 – Figuras representando a faixa do melhor compromisso SE-EE para diferentes carregamentos com $M = 64$; com potência de transmissão $P_{TX} = 1\text{W}/\text{antena}$. a) As curvas em vermelho contínuas são referentes a SE (eixo da esquerda) curvas em azul tracejado são referentes a EE (eixo da direita). b) Frente de Pareto, dentro deste fronte se encontra o ponto do melhor compromisso SE-EE.

4.4 Conclusão do Capítulo

Neste capítulo, nos propomos a realizar uma análise comparativa entre SE e EE em uma única célula DL entre sistemas mMIMO e NOMA com a BS equipada com três

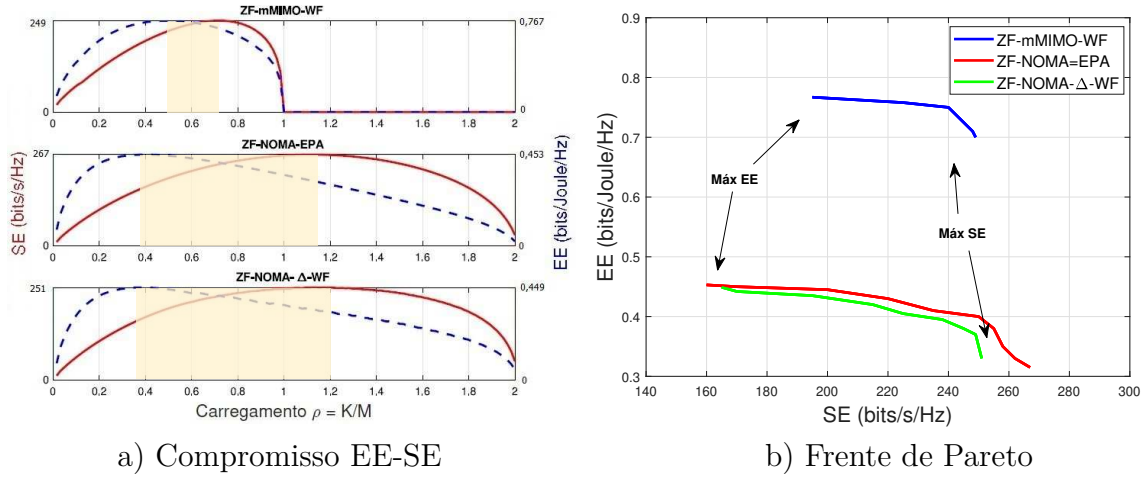


Figura 4.10 – Figuras representando a faixa do melhor compromisso SE-EE para diferentes carregamentos com $M = 128$; com potência de transmissão $P_{TX} = 1\text{W}/\text{antena}$. a) As curvas em vermelho contínuas são referentes a SE (eixo da esquerda) curvas em azul tracejado são referentes a EE (eixo da direita). b) Frente de Pareto, dentro deste fronte se encontra o ponto do melhor compromisso SE-EE.

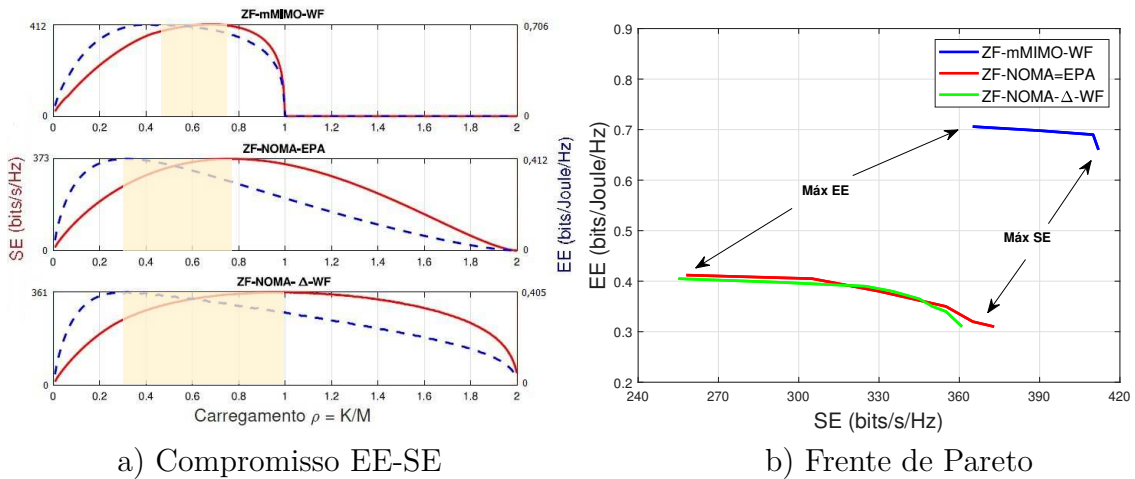


Figura 4.11 – Figuras representando a faixa do melhor compromisso SE-EE para diferentes carregamentos com $M = 256$; com potência de transmissão $P_{TX} = 1\text{W}/\text{antena}$. a) As curvas em vermelho contínuas são referentes a SE (eixo da esquerda) curvas em azul tracejado são referentes a EE (eixo da direita). b) Frente de Pareto, dentro deste fronte se encontra o ponto do melhor compromisso SE-EE.

Tabela 4.3 – Compromisso SE-EE.

M = 64					
	ρ	SE	EE	Usuários ativos	$\mathcal{F}$
mMIMO-WF	0.652	131.15	0.762	.50	.485
NOMA-EPA	0.875	147.45	0.428	1.0	.485
NOMA- Δ -WF	0.844	143.48	0.441	1.0	.475
M = 128					
	ρ	SE	EE	Usuários ativos	$\mathcal{F}$
mMIMO-WF	0.625	243.21	0.745	.50	.475
NOMA-EPA	0.734	241.54	0.411	1.0	.49
NOMA- Δ -WF	0.703	226.32	0.405	.98	.45
M = 256					
	ρ	SE	EE	Usuários ativos	$\mathcal{F}$
mMIMO-WF	0.578	404.84	0.691	.47	.43
NOMA-EPA	0.523	344.70	0.380	1.0	.495
NOMA- Δ -WF	0.594	333.31	0.375	.85	.398

diferentes configurações de antenas $M = [64, 128, 264]$. Sob a perspectiva da SE, mMIMO com algoritmo clássico de WF obteve melhores resultados sob baixo e médio carregamento de usuários. Por outro lado, quando a carga do sistema é maior que $\rho > 0,6$ o NOMA obteve melhores resultados até $\rho = 2$, com as simplificações adotadas no modelo.

Os métodos PA analisados aplicados ao sistema NOMA, (EPA, PCIPA e Δ -WF) resultam em diferentes desempenhos do SE. De fato, quando a condição de enrijecimento do canal é totalmente atingida, e a quantidade de antenas BS aumenta ($M = 128$ e 256), os melhores resultados SE são obtidos com o algoritmo Δ -WF proposto, mas, como esperado, o índice de equidade é prejudicado.

Sob a perspectiva da EE, o mMIMO obteve melhores resultados quando empregou os três métodos EPA, PCIPA e WF PA sob $K < M$. No entanto, o NOMA pode operar sob carga de sistema mais alta, *i.e.*, $K < 2M - 1$.

Em termos de métricas de área sob a curva SE e curva EE, $\mathcal{S}$ e $\mathcal{E}$, respectivamente, o sistema PD-NOMA obteve melhores resultados, devido à sua capacidade de atender a um número maior de usuários do que mMIMO. Tais resultados numéricos confirmam a capacidade do NOMA de operar com alta carga de dispositivos. Por outro lado, não é possível alcançar alta imparcialidade com o NOMA.

Do ponto de vista do compromisso SE-EE, o mMIMO obteve os melhores resultados, devido à superioridade em EE; que são comprovados nos frontes de Pareto, sempre alcançado em carregamento de dispositivos $\rho = 0.6$ em todas as configurações M . Os sistemas NOMA apresentam características interessantes e têm sido intensamente investigados como uma técnica promissora na criação de futuras gerações sem fio.

5 Acesso aleatório com concessão para sistemas NOMA XL-MIMO

Neste capítulo desenvolvemos um protocolo RA baseado em concessão para sistemas XL-MIMO explorando o NOMA no domínio da potência e aproveitamos a sobreposição de VR de sistemas XL-MIMO (NVR-XL), combinando as vantagens dos sistemas XL-MIMO e NOMA, permitindo vários usuários selecionando a mesma sequência piloto em pelo menos um SA.

5.1 Protocolo NVR-XL

O protocolo NVR-XL proposto explora os VRs sobrepostos de UEs específicos ao longo dos SAs com o objetivo de melhorar a operação do protocolo SUCRe-XL (Nishimura et al., 2020). Propomos modificar a regra de decisão de ativação proposta no protocolo SUCRe-XL (Nishimura et al., 2020). Utilizamos uma abordagem similar a proposta (Pereira et al., 2021), otimizamos numericamente o fator de escala, ou seja, $\delta_{\text{NVR-XL}}^{(K,Q)*}$ dentro do parâmetro de bias ϵ_k , visando maximizar a soma das taxas do sistema, fazendo com que mais de um usuário se considere capaz de retransmitir o piloto selecionado (no Passo 1); portanto, aumentando o número de possíveis falsos positivos, ou seja, vários UEs se autodenominam o vencedor da disputa. Isso é feito de forma distribuída, sem cooperação com outros UEs. Esses possíveis falsos positivos podem ser resolvidos com a implantação de um esquema baseado em NOMA, sendo capaz de processar no lado do receptor vários usuários compartilhando o piloto em VRs sobrepostos e processamento semi-distribuído em nível de SA.

A Figura 5.1 descreve os quatro Passos do protocolo de acesso aleatório proposto NOMA XL-MIMO baseado em concessão e destaca as diferenças em relação ao SUCRe-XL. Além disso, há uma Etapa preliminar (Passo 0), normalmente assumida pelos protocolos RA, na qual a BS transmite um sinal de controle; este sinal contém informações suficientes para gerar sincronização de tempo, enviar os valores do fator de escala otimizado $\delta_{\text{NVR-XL}}^{(K,Q)*}$ e conhecer o canal dos dispositivos, o que é crucial para o bom funcionamento do protocolo NVR-XL proposto. De fato, cada cenário que combina diferentes valores de K -iUEs e Q -SAs requer um *fator de escala* otimizado, esse fator é enviado aos usuários no Passo 0 e a BS mantém uma tabela de referência com esses fatores otimizados; tal processo é melhor explicado no Passo III. Os quatro Passos do protocolo NVR-XL proposto são descritos no procedimento a seguir.

PASSO I: *Transmissão de Sequência Piloto Aleatória:* Usuários no conjunto $\mathcal{K}$ fazem uma tentativa de RA com uma probabilidade P_a , sincronizada e com δ^* atualizado pelo sinal de

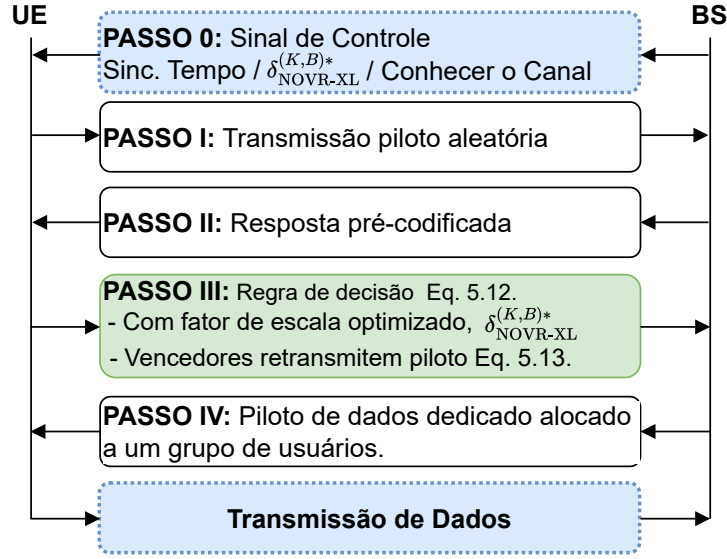


Figura 5.1 – Diagrama do protocolo proposta NVR-XL, linhas tracejadas demonstram etapas idênticas ao SUCRe-XL, "Passo 0" e "Transmissão de Dados" estão presentes na representação.

controle no Passo 0 anterior, cada UE que deseja estar ativo seleciona uniformemente ao acaso uma sequência piloto RA $\mathbf{s}_t$, $t = 1, \dots, \tau_{\text{RA}}$ e transmite com a potência de transmissão do usuário p_k . Portanto, seja $\mathcal{N}_t \in \mathcal{K}$ o conjunto de índices de UEs que selecionou o piloto t . Como em (Björnson et al., 2016), segue uma distribuição binomial.

$$|\mathcal{N}_t| \sim \mathcal{B}\left(K, \frac{P_a}{\tau_{\text{RA}}}\right). \quad (5.1)$$

Todos os iUEs que decidem se conectar transmitem sequências piloto RA UL, fazendo com que SA q receba o sinal:

$$\mathbf{Y}^{(q)} = \sum_{k \in \mathcal{K}} \sqrt{p_k} \mathbf{h}_k^{(q)} \mathbf{s}_t^T + \mathbf{N}^{(q)}, \quad (5.2)$$

onde $\mathbf{Y}^{(q)} \in \mathbb{C}^{M_q \times \tau_{\text{RA}}}$ e $\mathbf{N}^{(q)} \in \mathbb{C}^{M_q \times \tau_{\text{RA}}}$, seguindo $\mathcal{CN}(0, \sigma^2)$, é o ruído do receptor. Assim, cada SA correlaciona (5.2) usando um piloto normalizado arbitrário $\mathbf{s}_t$, e pode estimar o canal de UEs no mesmo conjunto $\mathcal{N}_t$ resultando em:

$$\mathbf{y}_t^{(q)} = \mathbf{Y}^{(q)} \frac{\mathbf{s}_t^*}{\|\mathbf{s}_t\|} = \sum_{i \in \mathcal{N}_t} \sqrt{p_i \tau_{\text{RA}}} \mathbf{h}_i^{(q)} + \mathbf{n}_t, \quad q = 1, \dots, Q, \quad (5.3)$$

onde $\mathbf{n}_t = \mathbf{N} \frac{\mathbf{s}_t^*}{\|\mathbf{s}_t\|} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{M_q})$ é o ruído efetivo do receptor. Devido às propriedades de MIMO massivo, *endurecimento do canal* e *propagação favorável assintótica*, temos em cada SA:

$$\frac{\left\| \sum_{q \in \mathcal{M}} \mathbf{y}_t^{(q)} \right\|^2}{M_q} \xrightarrow{M_q \rightarrow \infty} \alpha_t + Q\sigma^2, \quad (5.4)$$

onde α_t é a soma dos ganhos de sinal associadas ao piloto t em cada q -SA:

$$\alpha_t = \sum_{q \in \mathcal{M}} \sum_{i \in \mathcal{N}_t} p_i \beta_i^{(q)} \tau_{\text{RA}}. \quad (5.5)$$

PASSO II: *Resposta ao acesso aleatório pré-codificado:* No modo SA coordenado; SAs transmitem pilotos DL pré-codificados para UEs aquele piloto selecionado t no Passo I.

$$\mathbf{V}_{\text{XL}} = \sqrt{\frac{p_{\text{BS}}}{Q}} \sum_{t=1}^{\tau_{\text{RA}}} \frac{\sum_{q \in \mathcal{M}} \mathbf{y}_t^{(q)*}}{\left\| \sum_{q \in \mathcal{M}} \mathbf{y}_t^{(q)} \right\|} \mathbf{s}_t^H, \quad (5.6)$$

onde $\mathbf{V}_{\text{XL}} \in \mathbb{C}^{M_q \times \tau_{\text{RA}}}$ e p_{BS} é a potência de transmissão da BS. Neste caso UEs selecionando a sequencia piloto t recebem o sinal $\mathbf{z}_k^T \in \mathbb{C}^{\tau_{\text{RA}}}$ no canal reciproco com o ruído de recepção $\boldsymbol{\eta}_k \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{\tau_{\text{RA}}})$:

$$\mathbf{z}_k^T = \sum_{m \in \mathcal{V}_k} \mathbf{h}_k^{(m)T} \mathbf{V}_{\text{XL}} + \boldsymbol{\eta}_k^T. \quad (5.7)$$

de (5.7), cada UE correlaciona o sinal recebido com o piloto de RA selecionado $\mathbf{s}_t$,

$$z_k = \mathbf{z}_k^T \frac{\mathbf{s}_t}{\|\mathbf{s}_t\|} = \sqrt{\frac{p_{\text{BS}} \tau_{\text{RA}}}{Q}} \sum_{m \in \mathcal{V}_k} \mathbf{h}_k^{(m)T} \frac{\sum_{q \in \mathcal{M}} \mathbf{y}_t^{(q)*}}{\left\| \sum_{q \in \mathcal{M}} \mathbf{y}_t^{(q)} \right\|} + \eta_k \quad (5.8)$$

onde $\eta_k = \boldsymbol{\eta}_k \frac{\mathbf{s}_t}{\|\mathbf{s}_t\|} \sim \mathcal{CN}(0, \sigma^2)$ é efetivamente o ruído recebido. Para obter (5.4) do lado do UE, divide a eq. (5.8) por $\sqrt{M_q}$ e considerando as propriedade inerentes ao MIMO massivo *endurecimento do canal e propagação favorável assintótica*, teremos:

$$\frac{z_k}{\sqrt{M_q}} = \sqrt{\frac{p_{\text{BS}} \tau_p}{Q}} \frac{\left(\sum_{m \in \mathcal{V}_k} \mathbf{h}_k^{(m)T} \sum_{q \in \mathcal{M}} \mathbf{y}_t^{(q)} \right)^*}{M_q \sqrt{\frac{1}{M_q} \left\| \sum_{q \in \mathcal{M}} \mathbf{y}_t^{(q)} \right\|^2}} + \frac{\eta_k}{\sqrt{M_q}} \xrightarrow{M_q \rightarrow \infty} \sqrt{\frac{p_{\text{BS}} p_k}{Q}} \frac{\sum_{m \in \mathcal{V}_k} \beta_k^{(m)}}{\sqrt{\alpha_t + Q \sigma^2}}. \quad (5.9)$$

Assim, usando a função $\Re(\cdot)$ dá a parte real de sua entrada e descartando a parte imaginária, onde neste caso mitiga ruído e erros de estimativa da eq. (5.9), resultando:

$$\frac{\Re(z_k)}{\sqrt{M_q}} \approx \tau_{\text{RA}} \sqrt{\frac{p_{\text{BS}} p_k}{Q}} \frac{\sum_{m \in \mathcal{V}_k} \beta_k^{(m)}}{\sqrt{\alpha_t + Q \sigma^2}} \quad (5.10)$$

desta forma cada UE- k em $\mathcal{N}_t$ pode estimar α_t , de modo a :

$$\hat{\alpha}_{t,k} = \max \left[p_k \sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \tau_{\text{RA}}, \left(\frac{\Gamma(M_q + 1/2)}{\Gamma(M_q)} \right)^2 \frac{p_{\text{BS}} p_k \tau_{\text{RA}}^2 \left(\sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \right)^2}{Q [\Re(z_k)]^2} - Q \sigma^2 \right]. \quad (5.11)$$

PASSO III: *Resolução de Disputa com NOMA e Repetição do Piloto:* Ao explorar o NOMA, os usuários que colidem a mesma sequência piloto podem ser atendidos no mesmo recurso, ou seja, espacial (SA), frequência e tempo.

A resolução da colisão segue a regra de decisão descentralizada, ou seja, cada usuário decide se continua na tentativa de acesso, seguindo a regra de decisão (Pereira et al., 2021):

$$p_k \sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \tau_{\text{RA}} > \frac{\hat{\alpha}_{t,k}}{2} + \epsilon_k. \quad (5.12)$$

Se a desigualdade na Eq. (5.12) ocorre, o usuário repete sua transmissão piloto e forma o novo conjunto $\mathcal{S}_t$; caso contrário, permanece em silêncio e tenta se comunicar na próxima tentativa de RA com probabilidade P_{na} . A cardinalidade de $\mathcal{S}_t$ define o nível da disputa; de fato, se $|\mathcal{S}_t| = 1$ a colisão foi resolvida pela regra de decisão; caso contrário, para $|\mathcal{S}_t| > 1$ (múltiplos UEs) representando o possível caso *falso positivo*, será resolvido pelo domínio de potência NOMA:

$$\mathcal{S}_t = \left\{ k : p_k \sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \tau_{\text{RA}} > \frac{\hat{\alpha}_{t,k}}{2} + \epsilon_k \right\}. \quad (5.13)$$

A regra de decisão da Eq. (5.12), possui o termo de compensação ϵ_k , dado por:

$$\epsilon_k = \frac{\delta}{\sqrt{M_q} \cdot \sum_{q \in \mathcal{V}_k} \beta_k^{(q)}}, \quad (5.14)$$

onde δ é um *fator de escala* que pode ser ajustado para melhorar as métricas do sistema, como *número médio de tentativas de acesso* (ANAA), *fração de falhas de tentativas de acesso* (FFAA), *número normalizado de UEs aceitos* (NNAU), *média da soma das taxas* (ASR), e probabilidade de resolução de colisão.

Neste estudo, sugerimos otimizar o *fator de escala* δ numericamente, permitindo que múltiplos usuários no protocolo NVR-XL RA poderiam se considerar os vencedores no processo de resolução da disputa, gerando o *falso positivo* do protocolo SUCRe-XL, que pode ser resolvido no Passo IV com a aplicação do SIC, pois adotamos o prévio conhecimento perfeito do canal realizado no Passo 0. De fato, o sinal é processado em cada SA via RPU através da implantação do esquema NOMA; portanto, resolvendo sinais sobrepostos via estratégia SIC, com o objetivo de decodificar, reconstruir e cancelar os sinais do usuário compartilhando a mesma sequência piloto e VR. Finalmente, os sinais de saída da RPU são igualmente combinados em um EGC centralizado na CPU. O objetivo do ajuste no *fator de escala* é atingir a maximização da soma das taxas do sistema. A Seção de resultados numéricos explora a otimização do *fator de escala* δ .

Otimização do fator de escala: o valor ótimo $\delta_{\text{NVR-XL}}^{(K,Q)*}$ é baseado em (Pereira et al., 2021), implantando uma abordagem de pesquisa exaustiva. O procedimento é numericamente ótimo no sentido de maximização da média da soma das taxas; portanto, cada K UEs inativos e cada configuração Q-SA implica um valor de parâmetro ideal. Para obter o δ ótimo para o sistema NOMA XL-MIMO prático e cenários de canal, o seguinte procedimento foi adotado:

1. A BS conhece o número de SAs Q e possui uma estimativa do número de usuários que pretendem se tornar ativos KP_a ¹ presente na célula.
2. Começando com o valor ótimo obtido para o SUCRe, $\delta_{\text{SUCRe-XL}}^* = -1$, como em (Björnson et al., 2017b), o *fator de escala* é ajustado iterativamente para melhorar a soma do sistema. Esporadicamente, motivada pelas mudanças no cenário físico, a BS pode re-otimizar e transmitir o valor do *fator de escala* em um procedimento *look-up-table*².
3. Selecione o $\delta_{\text{NVR-XL}}^{(K,Q)*}$ ideal no sentido de maximizar a soma das taxas do sistema, como:

$$\delta_{\text{NVR-XL}}^{(K,Q)*} = \underset{\delta}{\operatorname{argmax}} R_{\sum(K,Q)}^{\text{NVR-XL}} \quad (5.15)$$

e registre o *Fator de escala* ótimo na tabela de referência. Observe que o *fator de escala* δ na Eq. (5.14) afeta a regra de decisão definida na Eq. (5.12), bem como a soma das taxas do sistema $R_{\sum(K,Q)}^{\text{NVR-XL}}$ conforme definido na sequência, Eq. (5.20).

4. Em uma etapa do preâmbulo do protocolo, ou seja, o Passo 0, levando em consideração o cenário atual do sistema e do canal, a BS consulta a tabela e transmite o correspondente valor otimizado $\delta_{\text{NVR-XL}}^{(K,Q)*}$.

Sem perda de generalidade, na etapa de resolução de disputa o protocolo aceita múltiplos usuários no *cluster* NOMA; portanto, a BS prossegue com o SIC de acordo com a ordem decrescente dos ganhos do canal ($\|\mathbf{h}_1\| > \|\mathbf{h}_k\| > \|\mathbf{h}_{|\mathcal{S}_t|}\|$), isto é, o sinal UE-1 mais forte é decodificado primeiro sem aplicação SIC, seguido pelo sinal UE-2 com uma etapa do SIC para remover o sinal UE-1 e permanecendo a interferência do UE-3; na sequência decodifica o sinal do UE-3 após dois passos SIC no sinal do UE-1 e UE-2, e assim por diante até o último usuário do conjunto $|\mathcal{S}_t|$.

A SINR (*signal-to-interference-plus-noise ratio*) baseado em SA do usuário k selecionando a mesma sequência piloto t é definida como:

$$\gamma_{t,k}^{(q)} = \frac{p_k \beta_k^{(q)}}{\varpi \sum_{j=1}^{k-1} p_j \beta_j^{(q)} + \sum_{i=k+1}^{|\mathcal{S}_t|} p_i \beta_i^{(q)} + \sigma^2}, \quad (5.16)$$

onde ϖ é o fator de resíduo de interferência após o SIC, assumindo que o cancelamento de interferência não é perfeito. A taxa alcançada por cada UE por SA pode ser definida como:

$$R_{t,k}^{(q)} = \log_2(1 + \gamma_{t,k}^{(q)}). \quad [\text{bpcu}] \quad (5.17)$$

¹Este processo de estimativa de usuários tentando se tornar ativo, pode ser obtido por uma ligeira modificação do Passo 0, incluindo um tom ou sequência piloto transmitido por cada usuário tentando se tornar ativo.

²Look-up-table é uma tabela de referência onde a BS mantém os valores otimizados do fator de escala, permitindo que cada UE possa recebe-los e utilizá-los rapidamente.

Em nosso protocolo RA proposto, quando a *regra de decisão* distribuída resulta em possíveis *falsos positivos* ou seja, vários usuários repetindo o piloto selecionado após a *regra de decisão* Eq. (5.12), esses usuários são agrupados no conjunto $\mathcal{S}_t$ e são tratados aplicando a estratégia NOMA-SIC, independentemente da cardinalidade desse conjunto. No entanto, para garantir que o SIC possa operar adequadamente, diferenças mínimas de potência dos sinais dos UEs em colisão devem ser garantidas; isso pode ser visto como um SINR mínimo entre os sinais em colisão no mesmo SA. Para fazer uma comparação realista, adotamos uma taxa de dados limiar $R_{\text{th}}^{(q)}$; portanto, os usuários no conjunto $\mathcal{S}_t$ que não satisfazem a taxa limite em uma base SA ainda são considerados *falsos positivos*. Portanto, pode-se definir um novo subconjunto de usuários com tentativa de RA bem-sucedida $\mathcal{S}_t^{\text{th}}$, formado por usuários colididos atingindo a taxa limite como:

$$\mathcal{S}_t^{\text{th}} = \{k : R_{t,k}^{(q)} > R_{\text{th}}^{(q)}\}, \quad (5.18)$$

que representa aqueles sinais de UE em colisão resolvidos pelo SIC-NOMA. Como consequência, o subconjunto de usuários *falsos positivos* $\mathcal{S}_t^{\text{fp}}$ é descrito como:

$$\mathcal{S}_t^{\text{fp}} = \mathcal{S}_t \setminus \mathcal{S}_t^{\text{th}}. \quad (5.19)$$

Observação: nomeamos possíveis usuários *falsos positivos* no conjunto $\mathcal{S}_t$, após a operação do SIC, aqueles usuários que cumprem a regra Eq. (5.18), forma o conjunto de usuários $\mathcal{S}_t^{\text{th}}$ que obtiveram sucesso na tentativa de RA, e formam o subconjunto $\mathcal{S}_t^{\text{fp}}$ os usuários *falsos positivos* que é a diferença dos subconjunto de $\mathcal{S}_t$ e $\mathcal{S}_t^{\text{th}}$.

PASSO IV: A BS estima o canal³ do usuário, ou usuários, considerando o sinal piloto enviado no Passo III, reconstrói o sinal e aplica recorrentemente o procedimento SIC. Depois disso, a mensagem correspondente é decodificada e os usuários são identificados. Após a decodificação ser bem-sucedida, uma alocação subsequente dos blocos de coerência dos pilotos de carga útil é realizada. UEs malsucedidos são instruídos a tentar novamente após um intervalo aleatório, limitado a no máximo 10 tentativas, então ele para de enviar a sequência piloto e o pacote é considerado perdido.

Na etapa de transmissão de dados do protocolo SUCRe-XL, apenas um usuário é aceito por sequência piloto em VR não sobreposta, por outro lado, no protocolo NVR-XL também haverá vários usuários que se consideram vencedores na etapa de disputa. A soma das taxas do protocolo NVR-XL em uma base SA pode ser definida usando (5.16) como:

$$R_{\sum(K,Q)}^{\text{NVR-XL}} = \sum_{q \in \mathcal{V}_k} \sum_{t=1}^{\tau_{\text{RA}}} \sum_{k=1}^{|\mathcal{S}_t^{\text{th}}|} \log_2 \left(1 + \gamma_{t,k}^{(q)} \right), \quad [\text{bpcu}] \quad (5.20)$$

Enquanto no SUCRe-XL a soma das taxas pode ser definida como:

$$R_{\sum(K,Q)}^{\text{SUCRe-XL}} = \sum_{q \in \mathcal{V}_k} \sum_{t=1}^{\tau_{\text{RA}}} \log_2 \left(1 + \frac{p_k \beta_k^{(q)}}{\sigma^2} \right), \quad [\text{bpcu}] \quad (5.21)$$

³Por simplicidade, e sem perda de generalidade, consideramos estimativas de CSI perfeitas.

assumindo que somente um usuário é aceito por piloto e o $\delta_{\text{SUCRe-XL}}^{(K,Q)*} = -1$, conforme definido em (Nishimura et al., 2020).

5.2 Resultados Numéricos

Assume-se uma URA com $M = 500$ antenas em uma área de célula de 200×100 m² com K (iUEs) distribuídos aleatoriamente conforme ilustrado na Figura 3.1, com a probabilidade de acesso P_a , o número de pilotos disponíveis τ_{RA} , e $q = p_k = 1\text{W}$. Os principais parâmetros de simulação estão listados na Tabela 5.1. A URA está dividida em Q SAs e a probabilidade de sucesso de visibilidade $P_b = 0,5$ para cada SA. Para fins de simulação, $|\mathcal{V}_k| > 0$, $\forall k$, ou seja, todos os usuários possuem pelo menos um SA em seu VR.

Nas simulações numéricas, duas configurações são propostas para avaliar numericamente o desempenho do protocolo NVR-XL: na configuração 1, assumimos o SIC funcionando perfeitamente, ou seja, $\varpi = 0,00$, e uma taxa limite $R_{\text{th}_1}^{(q)} = 1,585$ [bpcu] é adotado, o que representa o sinal de interesse tem uma diferença de potência w.r.t. a interferência o suficiente para ser decodificada corretamente sendo, $\Delta P \geq 2$. Enquanto na configurações 2, devido a falha em qualquer processo crucial anterior, o SIC é admitido imperfeito com fator de interferência residual $\varpi = 0,10$; este é um valor típico do fator residual adotado na literatura NOMA (Pereira et al., 2021; Wang et al., 2019), e uma taxa limite $R_{\text{th}_2}^{(q)} = 0,585$ [bpcu], portanto, o sinal de interesse pode não ter diferença de potência suficiente para ser decodificado perfeitamente, neste caso, adotamos $0,5 \leq \Delta P < 2$.

5.2.1 Soma das Taxas do Sistema

O valor ótimo de *fator de escala* $\delta_{\text{NVR-XL}}^{(K,Q)*}$ é obtido numericamente com base em (Pereira et al., 2021), por abordagem de busca exaustiva; é ótimo no sentido de maximização da soma das taxas; portanto, cada configuração K -iUE e Q -SA requer uma pesquisa para encontrar o valor ideal do parâmetro.

A Figura 5.2 revela os resultados para o parâmetro *fator de escala* ótimo obtido numericamente, seguindo o procedimento descrito na **PASSO III**, enquanto a soma das taxas NVR-XL maximizada correspondente é mostrada na Figura 5.3, nos propusemos a considerar duas configurações diferentes 1 e 2. Conforme evidenciado, o *fator de escala* ótimo diminui quando a interferência aumenta; seja quando o número de UEs escolhendo determinado piloto aumenta ou quando o número de SAs diminui.

Observe que a linha de base (SUCRe-XL) requer que o VR não sobreposto aceite o dispositivo; portanto, apresenta um ganho modesto em termos de somatória mesmo aumentando o número de SA. Os maiores valores de soma das taxas ocorrem antes da saturação do uso dos pilotos disponíveis. Por outro lado, o protocolo NVR-XL proposto

Tabela 5.1 – Parâmetros de simulação RA-GB NOMA XL-MIMO.

Parâmetro		Valor
Distância do usuário ao solo		$h_{user} \in [1; 1.7\text{m}]$
Elementos de antena eixo-y, eixo-z		$M_y = 100, M_z = 5$
Número de antenas na BS (URA)		$M = M_y \times M_z = 500$
Número de <i>Subarrays</i>		$Q \in \{5; 10\}$
Distância entre antenas		$d_m = 0.1 \text{ m}$
Altura do conjunto de antenas		$h_{array} = 12 \text{ m}$
Comprimento do conjunto de antenas		$L_y = M_y \times d_m$
Potência de transmissão usuário		$p_k = 1\text{W}$
Potência de transmissão BS		$p_{BS} = 1\text{W}$
Potência de ruído		$\sigma^2 = -98.65\text{dBm}$
Número de iUEs		$K \in [0; 20000]$
Probabilidade de acesso (1ª tentativa)		$P_a = 0, 01$
Probabilidade de acesso (novas tentativas)		$P_{na} = 0, 5$
Número de sequências piloto disponível		$\tau_{RA} = 10$
Realizações Monte Carlo		5000
# SAs	Abertura <i>Subarray</i> , D_{sa}	Dist. Rayleigh, d_{Rayl}
$Q = 1$	10m	1600m
$Q = 5$	2m	64m
$Q = 10$	1m	16m
Configuração 1		
Resíduo de SIC		$\varpi = 0, 00$
Taxa limitante 1		$R_{th1}^{(q)} = 1, 585 \text{ [bpcu]}$
Configuração 2		
Resíduo de SIC		$\varpi = 0, 10$
Taxa limitante 2		$R_{th2}^{(q)} = 0, 585 \text{ [bpcu]}$

processa o sinal sobreposto em cada SA, cancelando o sinal mais robusto do usuário, em algumas situações, com uma suposta interferência residual do SIC da ordem de ϖ , obtendo melhores taxas de soma resultados quando Q aumenta. Além disso, um limite de soma das taxas maior é alcançado ao ajustar *fator de escala* δ , estabelecendo o número máximo de usuários selecionando o mesmo piloto na etapa de transmissão de dados; além disso, quando o número de iUEs aumenta o *fator de escala* valor diminui. Observe que o valor do *fator de escala* δ é relativamente sensível à interferência.

Na Configuração 1, a soma das taxas máxima do sistema foi alcançada em um carregamento de $10(K \times P_a)/\tau_{RA}$. Na Configuração 2, a soma das taxas máxima foi alcançada no carregamento $5(K \times P_a)/\tau_{RA}$, aumentando a carga de usuários desses pontos, observe que reduz a soma das taxas do sistema.

Para verificar a sensibilidade do *fator de escala* δ^* , a Figura 5.4 mostra a soma das taxas do sistema com $Q = 10$ SAs. Quando a interferência aumenta, aumentando o número de iUEs K , a sensibilidade do *fator de escala* δ na soma das taxas aumenta

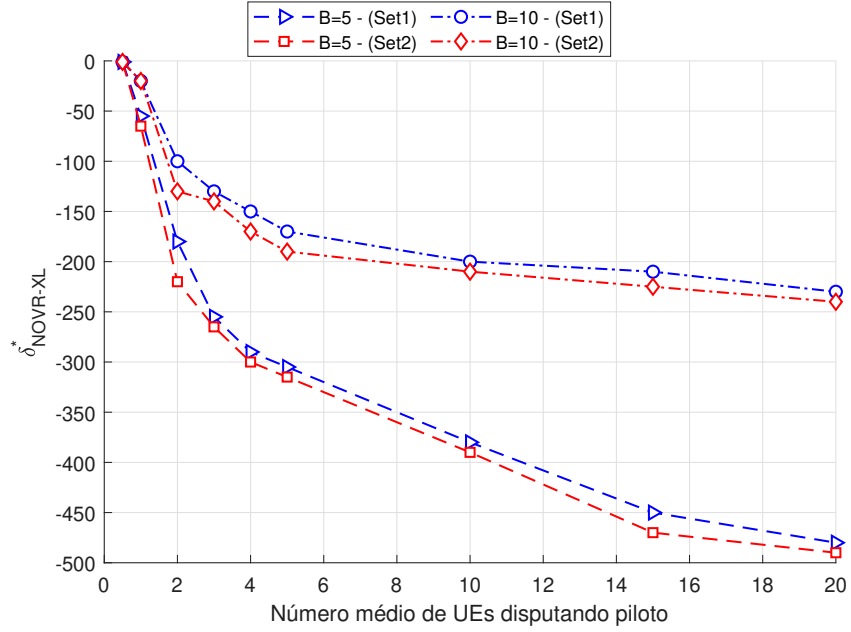


Figura 5.2 – $\delta_{\text{NOVR-XL}}^*$ ótimo em termos de maximização da soma das taxas, obtido numericamente para cada carregamento de usuários, nas duas configurações propostas e para as diferentes configurações de array $Q \in \{5; 10\}$.

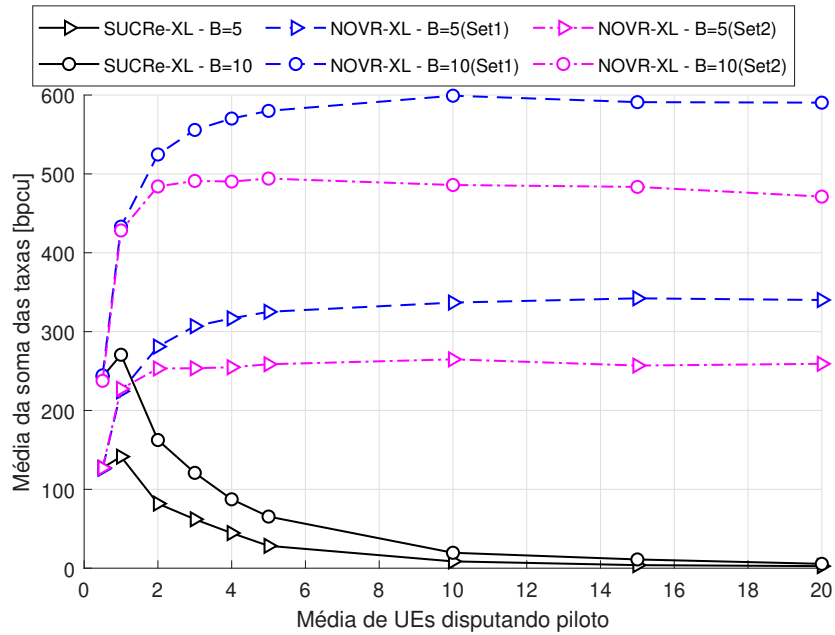


Figura 5.3 – Média da soma das taxas em função do numero de iUEs. Obtido após ajuste do $\delta_{\text{NOVR-XL}}^*$.

substancialmente, revelando a importância de estimar com precisão o *fator de escala* parâmetro no contexto do protocolo NVR-XL.

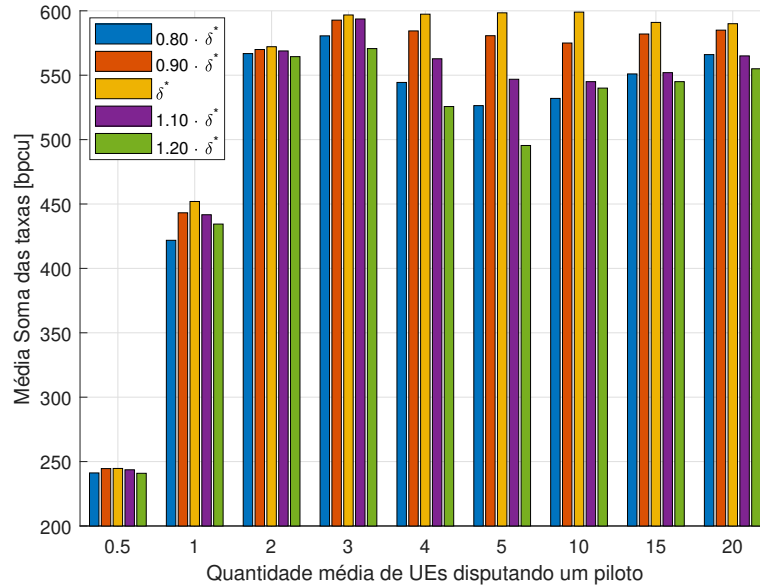


Figura 5.4 – Média da soma das taxas em função do número de iUEs, $Q = 10$ na configuração 1 no NVR-XL, com variações em torno δ^* ótimo.

5.2.2 Quantidade média e Fração de falhas em tentativas de acesso

A Figura 5.5 mostra que o protocolo proposto NVR-XL em cenários lotados ($1 > K \times P_a < 20$) alcança melhor desempenho de RA. A modificação na resolução de colisão descentralizada onde o NVR-XL aceita múltiplos usuários colidindo no VR sobreposto, graças à diversidade de potência; portanto, o sinal de superposição pode ser processado após o processo SIC em cada SA. Com o aumento no número de subarrays Q , mas limitado pelo pelas propriedades de *endurecimento do canal* e *propagação favoráveis*, a probabilidade de colisão em cada subarray pelo compartilhamento de VRs sobrepostos em cada SA aumenta. Tal configuração de resolução de colisão é impossível de tratar com SUCRe-XL devido à ausência da etapa de cancelamento de interferência. Como resultado, o NVR-XL permite diminuir o número de tentativas de acesso mesmo sob interferência sólida no mesmo SA.

5.2.3 Quantidade normalizada de usuários aceitos

A quantidade normalizada de UEs aceitos, calculado pelo número de usuários aceitos pelo número de usuários tentando acessar, dado na Figura 5.6 corrobora com os resultados melhorados no protocolo NVR-XL proposto para aceitar UEs para transmitir devido ao sinal de superposição poder ser processado pelo SIC em cada SA. Na configuração em que

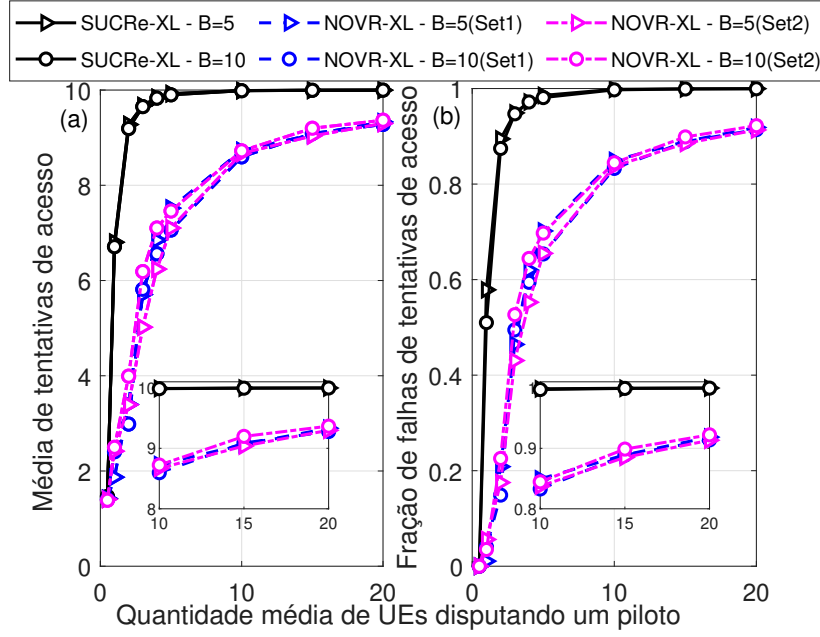


Figura 5.5 – Performance de RA dos sistemas. A linha base SUCRe-XL e NVR-XL. (a) Quantidade média de tentativas de RA. (b) Probabilidade de falhas nas tentativas de acesso.

a taxa limite é menor configuração 2, o SIC residual reduz a taxa dos usuários. Como o fator de escala foi otimizado para aumentar a soma das taxas, interferências altas reduzem o número médio de usuários aceitos.

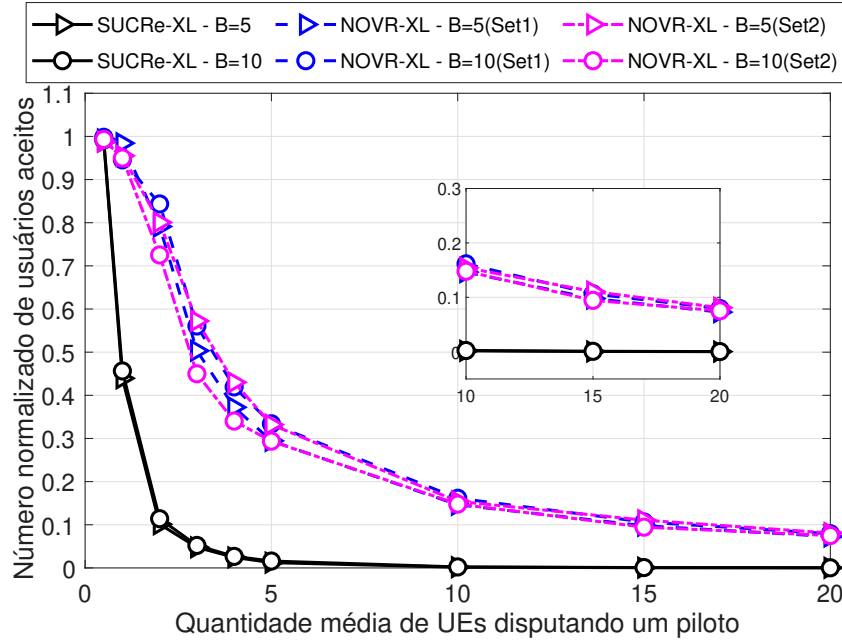


Figura 5.6 – Quantidade normalizada de usuários aceitos, obtido com $\delta_{\text{NVR-XL}}^*$.

5.2.4 Resolução de colisões

Nossa abordagem envolve ajustar numericamente o *fator de escala* para maximizar a média da soma das taxas do sistema, mas as duas configurações propostas têm respostas diferentes. Consideramos aceitos usuários que venceram a disputa ($|\mathcal{S}_t| = 1$) e usuários tratados pela NOMA-SIC ($\mathcal{S}_t^{\text{th}}$). Na configuração 1, o fato do SIC ser considerado perfeito, o número médio de usuários aceitos diminui mais lentamente e é maior em *arrays* com mais SAs, conforme mostrado na Figura 5.7 a), a taxa de limite mais alta aumenta a probabilidade de falsos positivos, consequentemente, uma média da soma das taxas maior conforme Figura 5.3. Por outro lado, na configuração 2 há um resíduo de interferência devido ao SIC ser considerado mais realista ou imperfeito que deteriora a soma das taxas do PD-NOMA. Para manter o nível mais alto da soma das taxas, mais colisões são resolvidas sem aplicar o NOMA, deteriorando o número médio de UEs aceitos, conforme mostrado na Figura 5.7 b), a taxa limite inferior diminui a probabilidade de falsos positivos em um limite inferior na média da soma das taxas como na Figura 5.3.

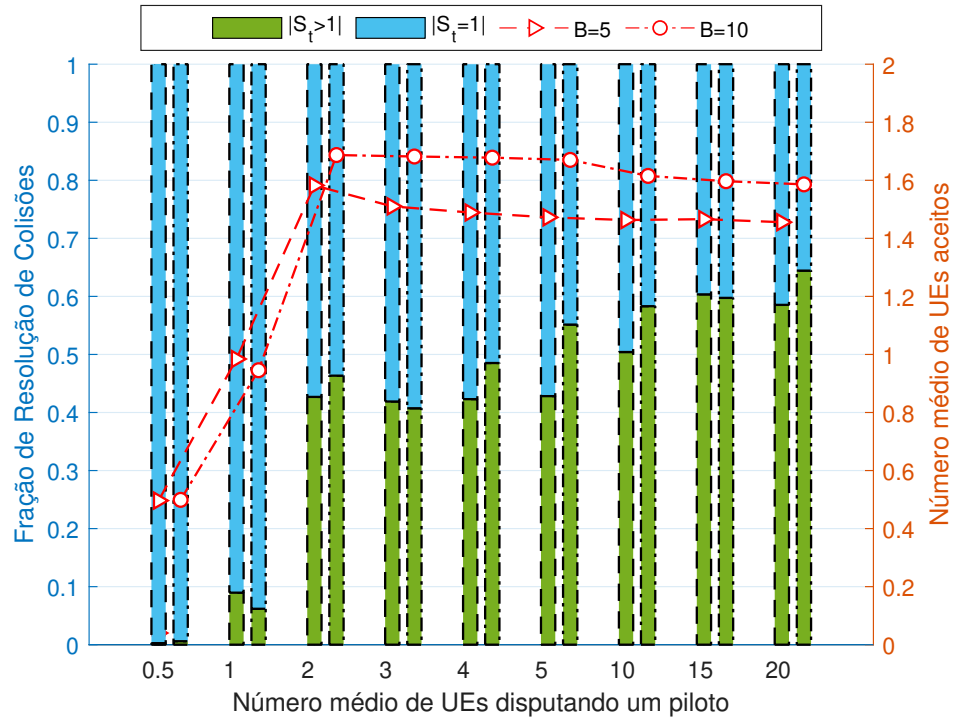
Aprofundando, a Figura 5.8 analisa a fração do número de usuários no conjunto $\mathcal{S}_t$ de casos *falsos positivos* resolvidos pelo SIC-NOMA. A figura mostra que o maior número de falsos positivos ocorre entre 2 usuários, sendo raros os casos com mais de 3 usuários. As configurações propostas alteram substancialmente o tamanho do conjunto $\mathcal{S}_t$ ao alterar o número de *subarrays* Q .

5.2.5 Overhead, atraso no tempo de acesso e Complexidade

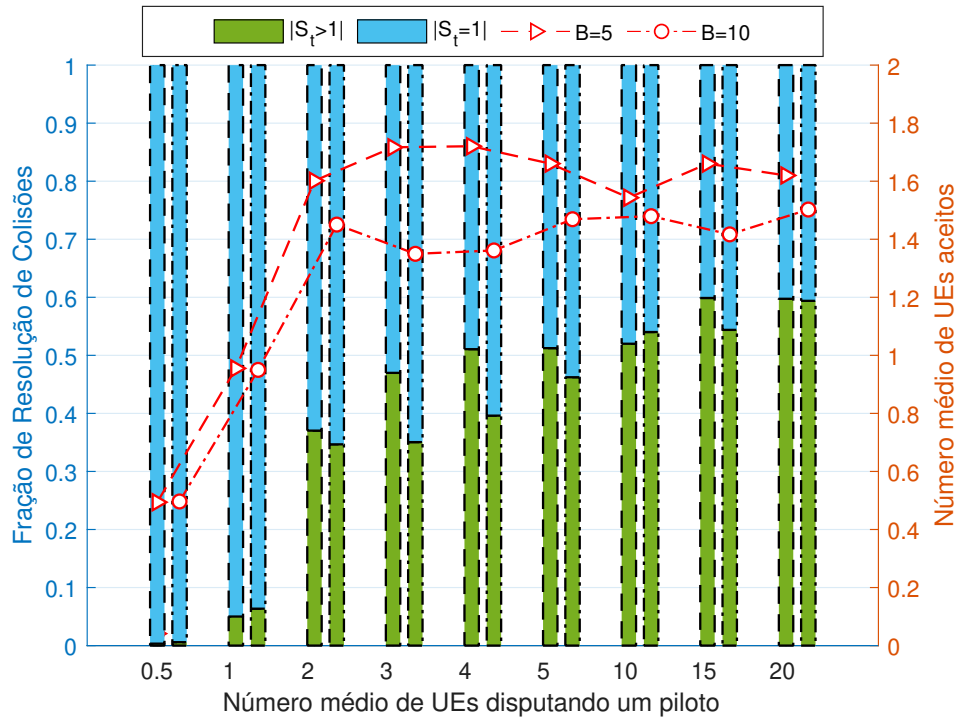
O protocolo baseado em concessão tem um *overhead* maior porque precisa de quatro etapas para conectar dispositivos à rede. Nossa proposta de protocolo NVR-XL é baseada em uma modificação do SUCRe (Björnson et al., 2017b), para operar em XL-MIMO (Nishimura et al., 2020). Tal procedimento não causa nenhuma sobrecarga ou atraso de tempo adicional em comparação com o protocolo SUCRe original (Nishimura et al., 2020). A otimização numérica e a atualização da tabela de referência ocorrem *online* na operação do sistema e não adicionam sobrecarga ao núcleo do protocolo (Passos I–IV).

Neste cenário de aplicação, o protocolo proposto melhora o tempo de atraso de acesso em todos os cenários estudados, quando comparado ao SUCRe-XL (Nishimura et al., 2020), como pode ser visto na Figura 5.5.

A complexidade do NVR-XL é comparável ao SUCRe-XL, a forma como é calculada a resposta de acesso aleatório pré-codificado Eq. (5.6) é a mesma. O *fator de escala* $\delta_{\text{NVR-XL}}^{(K,Q)*}$ otimizado eventualmente compartilhado no Passo 0 não adiciona complexidade ao protocolo proposto.



a) Configuração 1



b) Configuração 2.

Figura 5.7 – Porcentagem de colisões resolvidas pelo SIC e a quantidade média de usuários aceitos em NVR-XL.

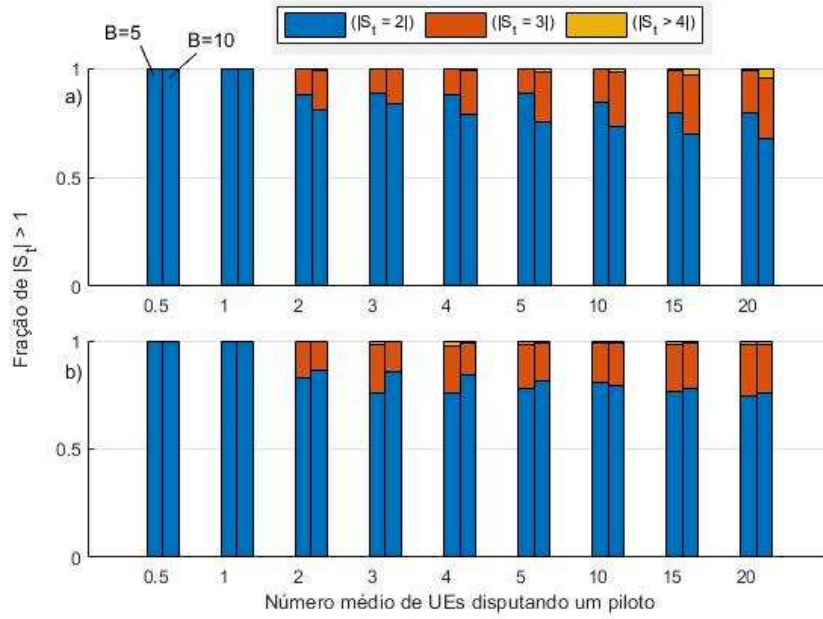


Figura 5.8 – Fração de casos *falso positivos* resolvidos pelo SIC no NVR-XL. a) Configuração 1 e b) Configuração 2.

5.3 Conclusão do capítulo

Ao explorar o NOMA no domínio da potência e o processamento de sinal em cada SA, o desempenho do protocolo RA baseado em concessão pode ser melhorado. Com efeito, o XL-MIMO acrescenta um novo grau de liberdade com o sinal recebido a ser processado por diferentes SAs com suporte de RPU, juntamente com o grau de liberdade introduzido pelo NOMA no domínio da potência, permitindo melhorar a resolução de colisões de múltiplos usuários usando a mesma sequência piloto nos mesmos SA de VR sobrepostos e aplicando o estágio SIC em cada SA no lado BS. Apoiados por resultados numéricos, encontramos uma substancial melhoria de desempenho do protocolo RA proposto em relação ao método SUCRe-XL da literatura recente, com a otimização numérica do *fator de escala* δ , para maximizar a soma das taxas do sistema, fazendo com que vários usuários colidam selecionando a mesma sequência piloto.

A principal distinção em relação aos esquemas SUCRe-XL e NVR-XL consiste no processamento independente por RPU e no processo de resolução de colisão. Além disso, revelamos a sensibilidade do *fator de escala* δ no esquema NVR-XL, principalmente em cenários com alta interferência; portanto, a seleção adequada do valor δ é fundamental para melhorar a resolução de colisão e a soma das taxas do sistema. De fato, no protocolo NVR-XL proposto, os dados de sinal sobrepostos recebidos podem ser adequadamente decodificados com etapas SIC em uma base SA, mesmo com uma restrição de limite de taxa, e capaz de alcançar tanto uma redução no número de tentativas de acesso e simultaneamente um aumento significativo na soma das taxas, aumentando o número médio de usuários aceitos nos cenários mMTC lotados que foram analisados ($1 < K \times P_a < 20$).

6 Conclusões e trabalhos futuros

6.1 Conclusões Gerais

Nesta Tese de doutorado, apresentamos extensa revisão bibliográfica relativa às aplicações de técnicas PD-NOMA, fornecendo evidências de que a tecnologia habilitadora NOMA reúne os atributos para atender uma imensa quantidade de dispositivos e terminais móveis em cenários de aplicações novos criados a partir do padrão 5G. Mesmo sob as condições de simplificação adotadas favorecendo a operação NOMA de múltiplas antenas mMIMO e XL-MIMO.

6.2 Conclusões I - Introdução e fundamentação conceitual

No Capítulo 1 realizou-se a apresentação e a fundamentação conceitual dos temas abordados nesta Tese. Apresenta-se uma descrição detalhada de operação do sistema PD-NOMA, com destaque para os ganhos de taxa e compartilhamento do mesmo bloco de recurso, possibilitado pelo SIC. Adicionalmente detalhamos os objetivos pretendidos na concepção da Tese. Concluí-se este capítulo com uma descrição da estrutura do texto de Tese e elencamos as contribuições que são resultado deste trabalho.

6.3 Conclusões II - Sistema mMIMO-NOMA

O Capítulo 2 apresenta o modelo de sistema mMIMO-NOMA, discute-se alguns conceitos inerentes a esse sistema: agrupamento de usuários, formatação de feixe, políticas de alocação de potência e informação do estado do canal. Na Subseção de agrupamento de usuário demonstrou-se uma solução quasi-ótima de baixa complexidade, para formatação de feixe e sugerimos a formatação da matriz de canal para maximizar a soma das taxas do sistema, quanto as políticas de alocação de potência descrevemos 2 políticas conhecidas e propomos uma alteração no WF clássico para o mMIMO-NOMA o que possibilita conhecer somente os canais dos usuários próximos a BS. Os conceitos apresentados são a base para os resultados apresentados no Capítulo 4.

6.4 Conclusões III - MIMO Ultra Massivo

No Capítulo 3 é apresentado o conceito de MIMO ultra massivo, sistema onde o arranjo de antenas é muito maior que o mMIMO, cujo processamento do sinal recebido é

organizado em *subarrays*. Esta topologia apresenta algumas particularidades em relação ao mMIMO, que são melhor entendidas com os conceitos de regiões de visibilidade e no processamento semi-distribuído de sinais. Constatamos que devido ao grande tamanho do arranjo de antenas o limite teórico de Rayleigh, que define qual a modelagem do canal sem fio é mais adequada, campo próximo ou campo distante, está bastante distante do arranjo de antenas. Em contrapartida, a divisão em *subarrays* altera este limite e valida a escolha da modelagem em campo distante. Estas soluções são implementadas e os resultados são apresentados no Capítulo 5.

6.5 Conclusões IV - Eficiência de recurso em mMIMO-NOMA sob diferentes políticas de alocação de potência

O Capítulo 4 apresenta os resultados de desempenho dos sistemas mMIMO clássico e o sistema mMIMO-NOMA. Constata-se que o modelo de sistema que explora o PD-NOMA pode operar com uma quantidade maior de dispositivos e apresenta em alguns resultados melhores resultados de soma das taxas. Por outro lado, ao analisarmos os resultados de eficiência energética o mMIMO alcança melhores resultados em toda faixa de carregamento onde ele opera de forma consistente, mesmo com as simplificações adotadas no sistema mMIMONOMA, o que comprova o potencial deste esquema.

A análise na perspectiva da área sob as curvas de SE revelam os melhores resultados obtidos com o PD-NOMA com a política de alocação de potência EPA considerando $M = 64$. Em contrapartida, existe uma vantagem de *fairness* obtido com o mMIMO. Destacamos que a modificação proposta no Δ -WF desativa uma quantidade menor de dispositivos e obtém uma área sob a curva maior.

De forma análoga, ao analisarmos a área sob a curva de EE verifica-se que o PD-NOMA supera o mMIMO com maior vantagem quando o número de antenas na BS é menor. Com isso, pode-se concluir que o PD-NOMA em quase todos os cenários é mais eficiente espectral e energeticamente do que o mMIMO em uma faixa muito ampla de carregamento de dispositivos $0 > \rho \leq 2$, aproximadamente, em média, 80% em termos de eficiência energética, e 170% mais eficiente em termos de eficiência espectral.

6.6 Conclusões V - Acesso Aleatório com concessão baseado em NOMA XL-MIMO

Tendo em vista as simplificações adotadas, propomos uma melhoria no protocolo SUCRe-XL, onde as colisões podem ser resolvidas aplicando o SIC, explorando o domínio da potência e o processamento de sinal em cada SA. Devido ao processamento de sinais ser realizado isolado por RPU e depois gerenciado por um CPU, somado a definição do fator de escala δ adequado, constatamos uma redução no número médio de tentativas de acesso possibilitando atender um número maior de dispositivos e com isso, um aumento na soma das taxas e no número médio de usuários aceitos em cenários lotados.

6.7 Perspectivas e trabalhos futuros

O trabalho de pesquisa apresentado nesta tese de doutorado teve como objetivo analisar os aspectos e da aplicação de técnicas NOMA nas futuras redes de comunicação sem fio. No entanto para aplicação do NOMA, ainda é necessário desenvolver técnicas avançadas de processamento de sinais NOMA com complexidade reduzida.

A principal característica do NOMA é permitir que mais de um dispositivo ocupe o mesmo bloco de recurso de canal, no entanto lidar com a interferência ainda é um dos principais problemas. A técnica amplamente utilizada para lidar com a interferência gerada pelo NOMA é o SIC, mas que muitas vezes é considerado perfeito. Note-se que para o SIC ter sucesso é necessário implementar uma adequada política de alocação de potência e desenvolver técnicas avançadas de processamento de sinais NOMA com baixa complexidade. As principais vias para a concretização de trabalhos futuros são listados a seguir:

Protocolos de acesso GF para sistemas NOMA: Aprofundar os estudos em protocolos de acesso GF aplicando técnicas NOMA, com foco principal na alocação de recursos no UL. Sugere-se a aplicação técnicas de detecção de atividade de usuários, possibilitando conhecimento prévio do número de sinais sobrepostos.

Análise das matrizes PDMA: Na técnica PDMA é definido de forma prévia uma matriz de padrões de transmissão; deste modo, cenários com alta densidade de usuários podem ser atendidos, devida a redução de complexidade de processamento de sinais da técnica. Sugere-se a proposição e a análise e propor técnicas de formação de matrizes padrão de modo a melhorar a performance do sistema e simultaneamente reduzir a complexidade do processamento de sinais.

Referências

3GPP. Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations, Version 15.0.0. document TR 25.996, Jun. 2018. Cited on page 38.

Al-Imari, M.; Imran, M. A.; Tafazolli, R.; Chen, D. Subcarrier and Power Allocation for LDS-OFDM System. In: *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*. [S.l.: s.n.], 2011. p. 1–5. Cited on page 34.

Ali, M. S.; Hossain, E.; Kim, D. I. Non-Orthogonal Multiple Access (NOMA) for Downlink Multiuser MIMO Systems: User Clustering, Beamforming, and Power Allocation. *IEEE Access*, v. 5, p. 565–577, 2017. ISSN 2169-3536. Available on: <http://ieeexplore.ieee.org/document/7802615/>. Cited 2 times on page(s) 55 e 61.

Ali, M. S.; Hossain, E.; Kim, D. I. Coordinated Multipoint Transmission in Downlink Multi-Cell NOMA Systems: Models and Spectral Efficiency Performance. *IEEE Wireless Communications*, v. 25, n. 2, p. 24–31, 2018. Cited on page 30.

Ali, M. S.; Tabassum, H.; Hossain, E. Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems. *IEEE Access*, v. 4, p. 6325–6343, 2016. Cited on page 57.

Alves, T. A. B.; Abrão, T. NOMA-Based Random Access in mMTC XL-MIMO. *IEEE Access*, v. 11, p. 1944–1954, 2023. Cited 2 times on page(s) 41 e 72.

Amiri, A.; Angelichinoski, M.; de Carvalho, E.; Heath, R. W. Extremely Large Aperture Massive MIMO: Low Complexity Receiver Architectures. In: *2018 IEEE Globecom Workshops (GC Wkshps)*. [S.l.: s.n.], 2018. p. 1–6. Cited on page 69.

Anwar, A.; Seet, B.-C.; Hasan, M. A.; Li, X. J. A Survey on Application of Non-Orthogonal Multiple Access to Different Wireless Networks. *Electronics*, v. 8, n. 11, p. 1355, nov 2019. ISSN 2079-9292. Available on: <https://www.mdpi.com/2079-9292/8/11/1355>. Cited 2 times on page(s) 40 e 55.

Benjebbour, A.; Li, A.; Kishiyama, Y.; Jiang, H.; Nakamura, T. System-level performance of downlink NOMA combined with SU-MIMO for future LTE enhancements. In: *2014 IEEE Globecom Workshops (GC Wkshps)*. [S.l.: s.n.], 2014. p. 706–710. Cited on page 24.

Benjebbour, A.; Li, A.; Saito, K.; Saito, Y.; Kishiyama, Y.; Nakamura, T. NOMA: From concept to standardization. In: *2015 IEEE Conference on Standards for Communications and Networking (CSCN)*. [S.l.: s.n.], 2015. p. 18–23. Cited on page 24.

Benjebbour, A.; Saito, Y.; Kishiyama, Y.; Li, A.; Harada, A.; Nakamura, T. Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access. In: *2013 International Symposium on Intelligent Signal Processing and Communication Systems*. [S.l.: s.n.], 2013. p. 770–774. Cited on page 24.

Björnson, E.; de Carvalho, E.; Larsson, E. G.; Popovski, P. Random access protocol for massive MIMO: Strongest-user collision resolution (SUCR). In: *2016 IEEE International Conference on Communications (ICC)*. [S.l.: s.n.], 2016. p. 1–6. Cited on page 93.

- Björnson, E.; de Carvalho, E.; Sørensen, J. H.; Larsson, E. G.; Popovski, P. A Random Access Protocol for Pilot Allocation in Crowded Massive MIMO Systems. *IEEE Transactions on Wireless Communications*, v. 16, n. 4, p. 2220–2234, 2017. Cited on page 44.
- Björnson, E.; Hoydis, J.; Sanguinetti, L. Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency. *Foundations and Trends® in Signal Processing*, v. 11, n. 3-4, p. 154–655, 2017. Cited 8 times on page(s) 37, 38, 39, 40, 59, 70, 96 e 103.
- Björnson, E.; Larsson, E. G. How Energy-Efficient Can a Wireless Communication System Become? *Conference Record - Asilomar Conference on Signals, Systems and Computers*, IEEE, v. 2018-Octob, p. 1252–1256, 2019. ISSN 10586393. Cited on page 40.
- Björnson, E.; Sanguinetti, L.; Hoydis, J.; Debbah, M. Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer? *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, n. January, p. 242–247, apr 2014. ISSN 15253511. Available on: (<http://ieeexplore.ieee.org/document/6951974/>). Cited 5 times on page(s) 19, 40, 42, 55 e 83.
- Boyd, S.; Vandenberghe, L. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. [S.l.]: Cambridge University Press, 2018. ISBN 978-1-316-51896-0. Cited on page 43.
- Budhiraja, I.; Kumar, N.; Tyagi, S.; Tanwar, S.; Han, Z.; Suh, D. Y.; Piran, M. J. A Systematic Review on NOMA Variants for 5G and Beyond. *IEEE*, v. 9, p. 85573–85644, 2021. ISSN 21693536. Cited on page 41.
- Carvalho, E. de; Ali, A.; Amiri, A.; Angjelichinoski, M.; Heath, R. W. Non-Stationarities in Extra-Large-Scale Massive MIMO. *IEEE Wireless Communications*, v. 27, n. 4, p. 74–80, 2020. Cited on page 71.
- Carvalho, E. de; Björnson, E.; Sørensen, J. H.; Larsson, E. G.; Popovski, P. Random Pilot and Data Access in Massive MIMO for Machine-Type Communications. *IEEE Transactions on Wireless Communications*, v. 16, n. 12, p. 7703–7717, 2017. Cited on page 41.
- Chen, X.; Gong, F.-K.; Li, G.; Zhang, H.; Song, P. User Pairing and Pair Scheduling in Massive MIMO-NOMA Systems. *IEEE Communications Letters*, v. 22, n. 4, p. 788–791, apr 2018. ISSN 1089-7798. Available on: (<https://ieeexplore.ieee.org/document/8116612/>). Cited on page 55.
- Chen, X.; Ng, D. W. K.; Yu, W.; Larsson, E. G.; Al-Dhahir, N.; Schober, R. Massive Access for 5G and Beyond. *IEEE Journal on Selected Areas in Communications*, v. 39, n. 3, p. 615–637, 2021. Cited on page 46.
- Cheng, H. V.; Björnson, E.; Larsson, E. G. Performance Analysis of NOMA in Training-Based Multiuser MIMO Systems. *IEEE Transactions on Wireless Communications*, IEEE, v. 17, n. 1, p. 372–385, jan 2018. ISSN 1536-1276. Cited 6 times on page(s) 41, 55, 62, 66, 67 e 77.
- Choi, J. An Approach to Preamble Collision Reduction in Grant-Free Random Access With Massive MIMO. *IEEE Transactions on Wireless Communications*, v. 20, n. 3, p. 1557–1566, 2021. Cited on page 46.

- Dai, L.; Wang, B.; Peng, M.; Chen, S. Hybrid Precoding-Based Millimeter-Wave Massive MIMO-NOMA With Simultaneous Wireless Information and Power Transfer. *IEEE Journal on Selected Areas in Communications*, IEEE, v. 37, n. 1, p. 131–141, jan 2019. ISSN 0733-8716. Available on: <https://ieeexplore.ieee.org/document/8485639/>. Cited on page 55.
- Dai, X.; Chen, S.; Sun, S.; Kang, S.; Wang, Y.; Shen, Z.; Xu, J. Successive interference cancelation amenable multiple access (SAMA) for future wireless communications. In: *2014 IEEE International Conference on Communication Systems*. [S.l.: s.n.], 2014. p. 222–226. Cited on page 34.
- Dai, X.; Zhang, Z.; Bai, B.; Chen, S.; Sun, S. Pattern Division Multiple Access: A New Multiple Access Technology for 5G. *IEEE Wireless Communications*, v. 25, n. 2, p. 54–60, 2018. Cited on page 35.
- Ding, J.; Choi, J. Preamble-Data Superposition Random Access in Massive MIMO Systems. *IEEE Wireless Communications Letters*, v. 9, n. 6, p. 906–910, 2020. Cited on page 46.
- Ding, J.; Qu, D.; Jiang, H.; Jiang, T. Success Probability of Grant-Free Random Access With Massive MIMO. *IEEE Internet of Things Journal*, v. 6, n. 1, p. 506–516, 2019. Cited on page 46.
- Ding, Z.; Fan, P.; Poor, H. V. Impact of User Pairing on 5G Nonorthogonal Multiple-Access Downlink Transmissions. *IEEE Transactions on Vehicular Technology*, v. 65, n. 8, p. 6010–6023, 2016. Cited on page 58.
- Ding, Z.; Poor, H. V. Design of Massive-MIMO-NOMA with Limited Feedback. *IEEE Signal Processing Letters*, v. 23, n. 5, p. 629–633, may 2016. ISSN 10709908. Available on: <http://ieeexplore.ieee.org/document/7434594/>. Cited on page 55.
- Duchemin, D.; Gorce, J.-M.; Goursaud, C. Code Domain Non Orthogonal Multiple Access versus ALOHA: A simulation based study. In: *2018 25th International Conference on Telecommunications (ICT)*. [S.l.: s.n.], 2018. p. 445–450. Cited on page 34.
- Elkourdi, M.; Mazin, A.; Balevi, E.; Gitlin, R. D. Enabling slotted Aloha-NOMA for massive M2M communication in IoT networks. In: *2018 IEEE 19th Wireless and Microwave Technology Conference (WAMICON)*. [S.l.: s.n.], 2018. p. 1–4. Cited on page 46.
- Goldsmith, A. *Wireless Communications*. [S.l.]: Cambridge University Press, 2005. ISBN 978-0-521-83716-3. Cited 2 times on page(s) 36 e 38.
- Han, H.; Guo, X.; Li, Y. A High Throughput Pilot Allocation for M2M Communication in Crowded Massive MIMO Systems. *IEEE Transactions on Vehicular Technology*, v. 66, n. 10, p. 9572–9576, 2017. Cited on page 44.
- Han, H.; Li, Y.; Guo, X. A Graph-Based Random Access Protocol for Crowded Massive MIMO Systems. *IEEE Transactions on Wireless Communications*, v. 16, n. 11, p. 7348–7361, 2017. Cited on page 44.
- Han, H.; Li, Y.; Zhai, W.; Qian, L. A Grant-Free Random Access Scheme for M2M Communication in Massive MIMO Systems. *IEEE Internet of Things Journal*, v. 7, n. 4, p. 3602–3613, 2020. Cited on page 46.

- Heath Jr, R. W.; Lozano, A. *Foundations of MIMO communication*. [S.l.]: Cambridge University Press, 2018. ISBN 978-0521762281. Cited on page 36.
- Hong, E.-K.; Lee, I.; Shim, B.; KO, Y.-C.; Kim, S.-H.; Pack, S.; Lee, K.; Kim, S.; Kim, J.-H.; Shin, Y.; Kim, Y.; Jung, H. 6G R&D vision: Requirements and candidate technologies. *Journal of Communications and Networks*, v. 24, n. 2, p. 232–245, 2022. Available on: <http://doi.org/10.23919/JCN.2022.000015>. Cited on page 23.
- Hoshyar, R.; Wathan, F. P.; Tafazolli, R. Novel Low-Density Signature for Synchronous CDMA Systems Over AWGN Channel. *IEEE Transactions on Signal Processing*, v. 56, n. 4, p. 1616–1626, 2008. Cited on page 34.
- Hu, S.; Rusek, F.; Edfors, O. Beyond Massive MIMO: The Potential of Positioning With Large Intelligent Surfaces. *IEEE Transactions on Signal Processing*, v. 66, n. 7, p. 1761–1774, 2018. Cited on page 69.
- Islam, S. M. R.; Zeng, M.; Dobre, O. A.; Kwak, K.-S. Resource Allocation for Downlink NOMA Systems: Key Techniques and Open Issues. *IEEE Wireless Communications*, v. 25, n. 2, p. 40–47, apr 2018. ISSN 1536-1284. Available on: <https://ieeexplore.ieee.org/document/8352621/>. Cited on page 55.
- Jacob, J. L.; Abrão, T. Nonorthogonal multiple access systems optimization to ensure maximum fairness to users. *Transactions on Emerging Telecommunications Technologies*, v. 31, n. 4, p. 1–14, 2020. ISSN 21613915. Cited on page 41.
- Jain, R.; Chiu, D. M.; Hawe, W. R. A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems. *DEC Research Report TR-301*, cs.NI/9809099, 1984. Cited 2 times on page(s) 43 e 44.
- Jiang, Y.; Li, P.; Ding, Z.; Zheng, F.-C.; Ma, M.; You, X. Joint Transmitter and Receiver Design for Pattern Division Multiple Access. *IEEE Transactions on Mobile Computing*, v. 18, n. 4, p. 885–895, 2019. Cited on page 35.
- Khalighi, M.; Brossier, J.-M.; Jourdain, G.; Raoof, K. Water filling capacity of Rayleigh MIMO channels. In: *12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications. PIMRC 2001. Proceedings (Cat. No.01TH8598)*. [S.l.: s.n.], 2001. v. 1, p. A–A. Cited on page 64.
- Kim, B.; Lim, S.; Kim, H.; Suh, S.; Kwun, J.; Choi, S.; Lee, C.; Lee, S.; Hong, D. Non-orthogonal Multiple Access in a Downlink Multiuser Beamforming System. In: *MILCOM 2013 - 2013 IEEE Military Communications Conference*. [S.l.: s.n.], 2013. p. 1278–1283. Cited 2 times on page(s) 61 e 67.
- Kim, D.; Choi, M. Non-Orthogonal Multiple Access in Distributed Antenna Systems for Max-Min Fairness and Max-Sum-Rate. *IEEE Access*, v. 9, p. 69467–69480, 2021. ISSN 21693536. Cited on page 41.
- Kobayashi, H.; Mark, B. L.; Turin, W. *Probability, Random Processes, and Statistical Analysis*. [S.l.]: Cambridge University Press, 2012. ISBN 978-0-521-89544-6. Cited on page 37.
- Larsson, E. G.; Edfors, O.; Tufvesson, F.; Marzetta, T. L. Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, IEEE, v. 52, n. 2, p. 186–195, 2014. ISSN 01636804. Cited 2 times on page(s) 55 e 62.

- Li, A.; Benjebbour, A.; Chen, X.; Jiang, H.; Kayama, H. Uplink non-orthogonal multiple access (NOMA) with single-carrier frequency division multiple access (SC-FDMA) for 5G systems. *IEICE Transactions on Communications*, E98-B, n. 8, p. 1426–1436, 2015. Cited on page 57.
- Li, S.; Su, B.; Sun, Y.; Sun, S.; Jin, L. Uplink grant-free pattern division multiple access transmission scheme by exploiting poly complementary sequence. In: *Transactions on Emerging Telecommunications Technologies*. [S.l.: s.n.], 2021. p. 1–18. Cited on page 35.
- Liu, G.; Jiang, D. 5G: Vision and Requirements for Mobile Communication System towards Year 2020. *Chinese Journal of Engineering*, v. 2016, n. ID 5974586, 2016. Available on: <http://dx.doi.org/10.1155/2016/5974586>. Cited on page 23.
- Liu, L.; Larsson, E. G.; Yu, W.; Popovski, P.; Stefanovic, C.; de Carvalho, E. Sparse Signal Processing for Grant-Free Massive Connectivity: A Future Paradigm for Random Access Protocols in the Internet of Things. *IEEE Signal Processing Magazine*, v. 35, n. 5, p. 88–99, 2018. Cited on page 46.
- Liu, Y.; Qin, Z.; Ding, Z. *Non-Orthogonal Multiple Access for Massive Connectivity*. Springer International Publishing, 2020. ISBN 978-3-030-30975-6. Available on: https://doi.org/10.1007/978-3-030-30975-6_1. Cited 2 times on page(s) 23 e 42.
- Lu, H.; Zeng, Y. Near-Field Modeling and Performance Analysis for Multi-User Extremely Large-Scale MIMO Communication. *IEEE Communications Letters*, v. 26, n. 2, p. 277–281, 2022. Cited on page 71.
- Makki, B.; Chitti, K.; Behravan, A.; Alouini, M.-S. A Survey of NOMA: Current Status and Open Research Challenges. *IEEE Open Journal of the Communications Society*, IEEE, v. 1, n. December 2019, p. 179–189, 2020. ISSN 2644-125X. Cited 3 times on page(s) 26, 55 e 56.
- Maraqa, O.; Rajasekaran, A. S.; Al-Ahmadi, S.; Yanikomeroğlu, H.; Sait, S. M. A Survey of Rate-Optimal Power Domain NOMA with Enabling Technologies of Future Wireless Networks. *IEEE Communications Surveys and Tutorials*, v. 22, n. 4, p. 2192–2235, 2020. ISSN 1553877X. Cited 3 times on page(s) 26, 27 e 40.
- Marinello F., J. C.; Brante, G.; Souza, R. D.; Abrão, T. Exploring the Non-Overlapping Visibility Regions in XL-MIMO Random Access and Scheduling. *IEEE Transactions on Wireless Communications*, p. 1–1, 2022. Cited 3 times on page(s) 69, 71 e 72.
- Martinez, A. O.; De Carvalho, E.; Nielsen, J. O. Towards very large aperture massive MIMO: A measurement based study. In: *2014 IEEE Globecom Workshops (GC Wkshps)*. [S.l.: s.n.], 2014. p. 281–286. Cited on page 69.
- Marzetta, T. L.; Larsson, E.; Yang, H.; Ngo, Q. *Fundamentals of Massive MIMO*. [S.l.]: Cambridge University Press, 2016. ISBN 978-1-107-17557-0. Cited 5 times on page(s) 38, 39, 40, 62 e 78.
- Mohamed, E. M. Joint users selection and beamforming in downlink millimetre-wave NOMA based on users positioning. *IET Communications*, v. 14, n. 8, p. 1234–1240, 2020. Available on: <https://emea.mitsubishielectric.com/ar/products-solutions/factory-automation/index.html>. Cited on page 59.

- Mounchili, S.; Hamouda, S. Pairing Distance Resolution and Power Control for Massive Connectivity Improvement in NOMA Systems. *IEEE Transactions on Vehicular Technology*, IEEE, v. 69, n. 4, p. 4093–4103, 2020. ISSN 19399359. Cited on page 40.
- Nikopour, H.; Baligh, H. Sparse code multiple access. In: *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. [S.l.: s.n.], 2013. p. 332–336. Cited on page 34.
- Nishimura, O. S.; Marinello, J. C.; Abrão, T. A Grant-Based Random Access Protocol in Extra-Large Massive MIMO System. *IEEE Communications Letters*, v. 24, n. 11, p. 2478–2482, 2020. Cited 7 times on page(s) 41, 69, 71, 72, 92, 98 e 103.
- Pereira, H. D.; Brante, G.; Farhat, J.; Souza, R. D.; Filho, J. C. M.; Abrão, T. On the Sum-Rate of Contention Resolution in Massive MIMO With NOMA. *IEEE Access*, v. 9, p. 24965–24974, 2021. Cited 4 times on page(s) 41, 92, 95 e 98.
- Rajoria, S.; Trivedi, A.; Godfrey, W. W. Sum-rate optimization for NOMA based two-tier hetnets with massive MIMO enabled wireless backhauling. *AEU - International Journal of Electronics and Communications*, Elsevier GmbH, v. 132, n. October 2020, p. 153626, 2021. ISSN 16180399. Available on: <https://doi.org/10.1016/j.aeue.2021.153626>. Cited on page 41.
- Ramatryana, I. N. A.; Shin, S. Y. NOMA-Based CRDSA with Access Control for Next-Generation IoT Networks. In: *2021 International Conference on Information and Communication Technology Convergence (ICTC)*. [S.l.: s.n.], 2021. p. 997–1001. Cited on page 46.
- Ramatryana, I. N. A.; Shin, S. Y. Priority Access in NOMA-Based Slotted ALOHA for Overload 6G Massive IoT. *IEEE Communications Letters*, v. 26, n. 12, p. 3064–3068, 2022. Cited 2 times on page(s) 28 e 46.
- Rappaport, T. S. *Wireless Communications: Principles and Practice*. [S.l.]: Prentice Hall, 2002. ISBN 978-0130422323. Cited 2 times on page(s) 36 e 37.
- Rezaei, F.; Tellambura, C.; Tadaion, A.; Heidarpour, A. R. Rate analysis of cell-free massive MIMO-NOMA with three linear precoders. *IEEE Transactions on Communications*, v. 68, n. 6, p. 3480–3494, 2020. ISSN 23318422. Cited on page 40.
- Saito, Y.; Kishiyama, Y.; Benjebbour, A.; Nakamura, T.; Li, A.; Higuchi, K. Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access. In: *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*. [S.l.: s.n.], 2013. p. 1–5. Cited on page 24.
- Senel, K.; Cheng, H. V.; Björnson, E.; Larsson, E. G. What Role Can NOMA Play in Massive MIMO? *IEEE Journal on Selected Topics in Signal Processing*, v. 4553, n. c, p. 1–16, 2019. ISSN 19324553. Cited 7 times on page(s) 41, 55, 61, 66, 67, 68 e 77.
- Souza, J. H. I. de; Abrão, T. Deep Learning-Based Activity Detection for Grant-Free Random Access. *IEEE Systems Journal*, v. 17, n. 1, p. 940–951, 2023. Cited on page 46.
- Sun, J.; Wang, C.; Zeng, J.; Su, X.; Lv, T. Design of PDMA Pattern Matrix in 5G Scenarios. In: *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. [S.l.: s.n.], 2020. p. 1–6. Cited on page 35.

- Vannithamby, R.; Talwar, S. *Towards 5G Applications: Requirements and Candidate Technologies*. John Wiley & Sons, 2017. ISBN 978-1-118-97984-6. Available on: <https://doi.org/10.1002/978-1-118-97984-6>. Cited on page 23.
- Varanasi, M. K.; Aazhang, B. Multistage Detection in Asynchronous Code-Division Multiple Access Communications. *IEEE Transactions on Communications*, v. 38, n. 4, p. 509–519, April 1990. Cited on page 28.
- Verdu, S. *Multiuser Detection*. [S.l.]: Cambridge University Press, 1998. ISBN 0 521 59373 5 hardback. Cited on page 27.
- Vieira, J.; Rusek, F.; Edfors, O.; Malkowsky, S.; Liu, L.; Tufvesson, F. Reciprocity Calibration for Massive MIMO: Proposal, Modeling, and Validation. *IEEE Transactions on Wireless Communications*, v. 16, n. 5, p. 3042–3056, 2017. Cited on page 39.
- Wang, X.; Chen, R.; Xu, Y.; Meng, Q. Low-Complexity Power Allocation in NOMA Systems With Imperfect SIC for Maximizing Weighted Sum-Rate. *IEEE Access*, v. 7, p. 94238–94253, 2019. Cited on page 98.
- Yang, H.; Marzetta, T. L. Performance of conjugate and zero-forcing beamforming in large-scale antenna systems. *IEEE Journal on Selected Areas in Communications*, v. 31, n. 2, p. 172–179, 2013. ISSN 07338716. Cited on page 78.
- Yuan, Z.; Yu, G.; Li, W.; Yuan, Y.; Wang, X.; Xu, J. Multi-User Shared Access for Internet of Things. In: *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*. [S.l.: s.n.], 2016. p. 1–5. Cited on page 34.
- Zafari, F.; Gkelias, A.; Leung, K. K. A Survey of Indoor Localization Systems and Technologies. *IEEE Communications Surveys and Tutorials*, IEEE, v. 21, n. 3, p. 2568–2599, 2019. ISSN 1553877X. Cited 3 times on page(s) 26, 27 e 59.
- Zeng, J.; Li, B.; Su, X.; Rong, L.; Xing, R. Pattern division multiple access (PDMA) for cellular future radio access. In: *2015 International Conference on Wireless Communications & Signal Processing (WCSP)*. [S.l.: s.n.], 2015. p. 1–5. Cited on page 34.
- Zeng, M.; Yadav, A.; Dobre, O. A.; Tsiropoulos, G. I.; Poor, H. V. Capacity Comparison Between MIMO-NOMA and MIMO-OMA With Multiple Users in a Cluster. *IEEE Journal on Selected Areas in Communications*, v. 35, n. 10, p. 2413–2424, oct 2017. ISSN 0733-8716. Available on: <http://ieeexplore.ieee.org/document/7974731/>. Cited on page 55.
- Zhang, H.; Zhang, H.; Liu, W.; Long, K.; Dong, J.; Leung, V. C. Energy Efficient User Clustering, Hybrid Precoding and Power Optimization in Terahertz MIMO-NOMA Systems. *arXiv*, v. 38, n. 9, p. 2074–2085, 2020. ISSN 23318422. Cited on page 40.
- Zhang, Y.; Wang, H. M.; Zheng, T. X.; Yang, Q. Energy-Efficient Transmission Design in Non-orthogonal Multiple Access. *IEEE Transactions on Vehicular Technology*, v. 66, n. 3, p. 2852–2857, 2017. ISSN 00189545. Cited on page 40.

Anexos

ANEXO A – Massive MIMO and NOMA bits-per-antenna efficiency under power allocation policies



Full length article

Massive MIMO and NOMA bits-per-antenna efficiency under power allocation policies[☆]Thiago A. Bruza Alves, Taufik Abrão^{*}

State University of Londrina (UEL), Department of Electrical Engineering, Londrina-PR, Brazil



ARTICLE INFO

Article history:

Received 25 September 2021

Received in revised form 20 December 2021

Accepted 22 December 2021

Available online 29 December 2021

Keywords:

Non-orthogonal multiple access (NOMA)

Massive multiple-input multiple-output (mMIMO)

Energy efficiency (EE)

Spectral efficiency (SE)

ABSTRACT

A comparative resource allocation analysis in terms of received bits-per-antenna spectral efficiency (SE) and energy efficiency (EE) in downlink (DL) single-cell massive multiple-input multiple-output (mMIMO) and non-orthogonal multiple access (NOMA) systems considering a BS equipped with many (M) antennas, while K devices operate with a single-antenna, and the loading of devices $\rho = \frac{K}{M}$ ranging in $0 < \rho \leq 2$ is carried out under three different (PA) strategies: the inverse of the channel power allocation (PICPA), a modified water-filling (Δ -WF) allocation method, and the equal power allocation (EPA) reference method. Since in the NOMA system, the two devices per cluster are overlapped in the power-domain, the channel matrix requires transformation to perform the zero-forcing (ZF) precoding as adopted in mMIMO. Hence, NOMA operating under many antennas can favor a group of devices with higher array gain, overcoming the mMIMO and operating conveniently in the higher loading range $0.6 < \rho < 2.0$. In such scenario, a more realistic and useful metric consists in evaluating the area under SE and EE curves, by measuring the bit-per-antenna and bit-per-antenna-per-watt efficiency, respectively. Our numerical results confirm a superiority of NOMA w.r.t. mMIMO of an order of 3x for the SE-area and 2x for the EE-area metric.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The beyond Fifth Generation (5G) of wireless communication systems must allow ultra-dense connections with vastly heterogeneous requirements. The challenges in networks persist, including the Spectral Efficiency (SE) and the Energy Efficiency (EE) joint improvement, the increase in the SE-EE trade-off and Quality of Service (QoS), always aiming to meet the growing number of devices connected to the network. Among the proposals to solve these challenges, the massive Multiple-Input Multiple-Output (mMIMO) system is the main proposed system that allows the increase of the link capacity, exploring the propagation of multiple paths with the use of a large number of antennas at the Base Station (BS) [1,2]. Another relevant enabling technology is the Non-Orthogonal Multiple Access (NOMA), which explores the power-domain as an alternative way in terms of multiple access technology, helping to mitigate the spectrum exhaustion problem and serving more than one device per resource block [3].

Although in many works mMIMO is classified as an orthogonal technique, the fact of allocating the signal from devices in the same resource block, possible by spatial diversity, allows us to classify it as a non-orthogonal technique too [4]. There is a vast literature demonstrating the superior performance of the Spectral Efficiency of NOMA when compared to Orthogonal Multiple Access (OMA) techniques [5]. Previous aiming to improve the communication system performance by combining MIMO (with a small number of antennas M) and NOMA have been discussed in [6–9].

Studies comparing NOMA and mMIMO in a single cell are proposed in [4,10,11]. The acquisition of Channel State Information (CSI) through pilot acquisition to NOMA system is proposed in [10]. In [11] the application of NOMA in mMIMO scheme is proposed and better results are achieved in the proposed comparative. Moreover, in [4] is analyzed the performance of NOMA and mMIMO in line of sight and non-line of sight.

The canonical mMIMO refers to the systems with BSs formed by a large number of antennas M when compared to the number of active devices, K , succinctly $M \gg K$ is considered a mMIMO setup. The typical NOMA improves the SE by superposing the signals of the selected devices to form a cluster in the power domain multiplexing it over the same signal and are served by the same beamforming. Nonetheless, the success of NOMA depends on the Successive Interference Cancellation (SIC).

[☆] This work was supported in part by the National Council for Scientific and Technological Development (CNPq) of Brazil under Grant 310681/2019-7.

^{*} Corresponding author.

E-mail addresses: thiago.bruzaalves@uel.br (T.A.B. Alves), taufik@uel.br (T. Abrão).

Power-domain NOMA can be a candidate technology in dense networks [12]. To improve performance and minimize the impact to assume the perfect SIC [13], devices are divided into two groups and after grouped in pairs, forming a cluster, to have a high difference between channel conditions, each pair forms a cluster, the device with higher channel condition can decode the signal sent to the device with the lower channel condition, and the interference can thus be eliminated by SIC. The use of NOMA in BS equipped with a large number of antennas was investigated in terms of SE [4,10] we propose in a similar configuration system increase the loading up to 2 times the number of antennas in BS and analyze the SE, EE and SE–EE trade-off.

The EE metric is a very popular method employed to analyze the balance between power consumption and data rate. The EE is defined as the ratio between effectively transmitted data rate and the total power expended during the transmission process, including instantaneous and static components. With the EE metric, it is possible to evaluate the efficiency with which a system uses the limited energy resource to communicate effective data, and also optimize this ratio. Can show the tendency of energy consumption in the case of seeking justice among devices.

The Zero Forcing (ZF) is simple and popular alternative interference suppression beamforming under perfect CSI condition and achieving a satisfactory condition in real situations when imperfect CSI, in this work we adopt perfect CSI, for that the pilots are needed. The adoption of NOMA system with a large number of antennas requires a defined equivalent channel to be deployed for interference mitigation; and according to the NOMA principle, makes the equivalent channel matrix smaller than the original one due to the exploration of power domain in NOMA.

Various transmission topologies already deal with the EE problem in mMIMO, finding the optimal number of antennas, number of devices in a cell and the maximal EE [2,14]. The EE analysis in NOMA system is carried out in [15], and its superiority is demonstrated when compared with conventional orthogonal multiple access (OMA) systems. Recent researches seek to improve the NOMA performance, e.g. in [16], the minimum pairing distance is defined and compared to the OMA, while in [17] it is presented a comparison between OMA and cell-free system equipped with mMIMO-NOMA. An EE analysis in Terahertz (THz)-NOMA-Multiple-Input Multiple-Output (MIMO) was proposed in [18] but the number of active devices is much smaller than the number of antennas in the BS and [19] is a survey about Power Domain NOMA and make clear the vacuum of EE analysis and comparison between NOMA with many antennas and mMIMO.

Recent works propose the deployment of NOMA combined with other techniques a more effective transmission scheme; e.g., in [20] NOMA and mMIMO are jointly considered in a two-tier network for accommodating huge traffic. Furthermore, in [21] authors apply NOMA in *Distributed Antenna Systems* (DAS) aiming to achieve better performance when compared to the conventional NOMA or DAS technique alone. While [22] shows an in-depth survey of the state-of-the-art of power-domain NOMA variants; moreover, several open issues and research challenges of NOMA-based applications are systematized. The NOMA system presents drawbacks, such as: hardware (including SIC) complexity, channel feedback, receiver design, and careful power and pilot allocation strategies [12,19,23].

This work focus on revealing the advantages of applying the mMIMO scheme *versus* NOMA scheme with a massive number of BS antennas, and varying the loading of devices, i.e., the ratio of the number of mobile devices to the number of BS antennas, $\rho = \frac{K}{M}$, while we change the Power Allocations (PA) strategy. Besides, we adopt a realistic model for the system's power consumption

as in [2] but adapted to our needs, aiming at providing a suitable analysis of the system resource allocation.

Contributions: the contributions of this work are fourfold. (a) an extensive and comparative analysis on the spectral efficiency (SE) performance of mMIMO system against NOMA system, varying the system loading under specific (three different) power allocation methods and making use of the area under the SE ($\mathcal{S}^{\text{SYST}}$) curve of the system as an effective, useful and fair metric of performance and efficiency; (b) we develop an energy efficiency (EE) analysis using a detailed model of energy consumption, with fixed and variable terms related to circuitry power consumption with number of antennas and devices, respectively, providing an extensive and comparative analysis on both the NOMA and mMIMO systems under realistic operation scenarios and making use of the area under the EE ($\mathcal{E}^{\text{SYST}}$) curve of system; (c) an analysis on the SE–EE trade-off is developed considering a wide range of loading of devices, verifying the fairness between devices; (d) finally, under mild conditions, we provide evidences for the NOMA's ability to serve a greater number of devices than mMIMO system.

The remainder of the paper is organized as follows. Section 2 describes the system models for NOMA and mMIMO adopted in this work. In Section 3 we present the proposed EE–SE formulation for NOMA and massive MIMO systems. Numerical results are analyzed in Section 4. Section 5 concludes the paper.

Notation. In this work, bold face lower case and upper case characters denote vectors and matrices, respectively. The operator $(x)^+ = \max(0, x)$. The operators $[\cdot]^T$, $\mathbb{E}[\cdot]$ and $|\cdot|$ denote transpose, expectation and cardinality, respectively. A random vector $\mathbf{x} \sim \mathcal{CN}\{0, \mathbf{I}_m\}$ is circularly symmetric Gaussian distributed with mean 0 and covariance matrix $\mathbf{I}_m$. $\mathbf{I}_m$ is $m \times m$ identity matrix.

2. System models

Let us consider a multi-user single-cell Downlink (DL) transmission operating in a Time Division Duplex (TDD) with K single-antenna active devices, communicating with one BS, which is equipped with M transmit antennas in Non-Line-Of-Sight (NLOS). The set $\mathcal{K}$ is formed by K devices, these devices are randomly distributed in a radius disk $d_{\max}$, the disk area is formed by two sub-disk with the same number of devices in each sub-area; both subsets are identified as $\mathcal{K}_H$ and $\mathcal{K}_L$. In the first subset, $\mathcal{K}_H$ represents devices' indexes having the higher channel coefficient and sort in descending order, while the other subset $\mathcal{K}_L$ are formed by the devices with lower channel coefficient and sort in ascending order; the indexes $k \in \mathcal{K}_H$ and $k \in \mathcal{K}_L$ such that:

$$\mathcal{K} = \mathcal{K}_H \cup \mathcal{K}_L, \quad \text{where} \quad (1)$$

$$\mathcal{K}_H = \{1, \dots, K/2\} \quad \text{and} \quad \mathcal{K}_L = \{K/2 + 1, \dots, K\}.$$

The channel vector modeling of device k can be described liked as:

$$\mathbf{h}_k = \sqrt{\beta_k} \mathbf{h}'_k, \quad k = 1, \dots, K, \quad (2)$$

where β_k is the large-scale fading coefficient and satisfy

$$\beta_j > \beta_i, \quad \forall j \in \mathcal{K}_H, \quad \forall i \in \mathcal{K}_L. \quad (3)$$

Herein, the pathloss model in [dB] is defined as:

$$\beta_k = \beta_0 + 10 \cdot \xi \cdot \log_{10}(d_k), \quad (4)$$

where d_k is the distance of user k to BS, ξ is the pathloss coefficient, and β_0 is the attenuation at the distance of reference.

In each coherence interval, $\mathbf{h}'_k$ in (2) for device k is an independent random small-scale fading realization from an independent Rayleigh fading distribution, $\mathbf{h}'_k \sim \mathcal{CN}(0, \mathbf{I}_M)$, $k = 1, \dots, K$. The

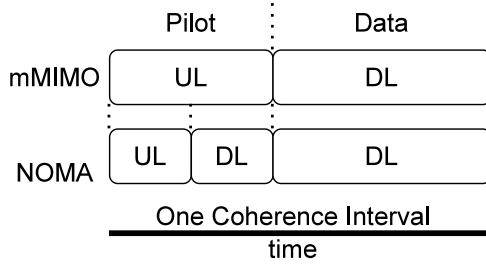


Fig. 1. Coherence time interval structure: the training and data transmission structure for mMIMO and NOMA schemes under TDD NLOS setup.

transmitted signal $\mathbf{x}_k \in \mathbb{C}^M$ is the beamformed data symbol of device k :

$$\mathbf{x}_k = \mathbf{g}_k \sqrt{p_k} s_k, \quad (5)$$

where $\mathbf{g}_k$ is a normalized beamforming vector, p_k normalized transmission power and $s_k \sim \mathcal{CN}(0, 1)$ is the data symbol of device k , and period T_s . The signal received at the k th device:

$$\begin{aligned} y_k &= \sum_{k'=1}^K \mathbf{h}_k^T \mathbf{x}_{k'} + n_k, \\ &= \sqrt{\beta_k} \mathbf{h}_k^T \sum_{k'=1}^K \mathbf{g}_{k'} \sqrt{p_{k'}} s_{k'} + n_k, \\ &= \sqrt{\beta_k p_k} \mathbf{h}_k^T \mathbf{g}_k s_k + \sqrt{\beta_k} \mathbf{h}_k^T \sum_{\substack{k' \neq k \\ k'=1}}^K \mathbf{g}_{k'} \sqrt{p_{k'}} s_{k'} + n_k, \end{aligned} \quad (6)$$

where $n_k \sim \mathcal{CN}(0, 1)$ is the additive noise. Notice that so far, this modeling applies to both NOMA and mMIMO systems, but beamforming are selected differently and this topic will be addressed in the next sections.

2.1. Prior actions

Because the BS needs to know *a priori* crucial information related to the channel and devices distributed in the cell, including device location, rate demanded, channel coefficient, such required *a priori* information may differ depending on the multiple access scheme considered [12].

The option for the mMIMO and NOMA systems was carried out with the guarantee that the needs of the devices would be met and this prior step is carried out successfully. Section 3.5 briefly discusses the prior information required to proceed with different Power Allocation (PA) procedures in both mMIMO and NOMA systems.

2.2. Pilot overhead for channel state information

Fig. 1 compares the pilot-data transmission structure along the one channel coherence interval for both mMIMO and NOMA systems considered. Notice that T_s is the time required to transmit a data symbol (Data). Moreover, the channel coherence time interval T is assumed a multiple of Data symbol period, $T = \iota \cdot T_s$. The power allocated to each pilot in training step is enough, while the number of pilots and the dedicated portion of coherence interval for data transmission are assumed the same in both systems.

Notice that in NOMA transmission, the pilot transmission step is split into two portions, half for the UL transmission pilots and receive the DL pilot confirmation; this happens because to

perform SIC, the cell-center devices need to learn the effective channels that are established by the beamforming. Additionally, the beamforming vectors are based on cell-center devices, producing limited rates achieved in cell-edge devices. On the other hand, in the mMIMO scheme, an important advantage is that there is no need for DL pilots since the effective channels created by the beamforming are highly predictable, i.e. nearly deterministic gain and phase due to channel hardening effect [4].

Assumption 1. In the NOMA system, the power allocated to each downlink pilot is sufficient to reach the destination device.

2.3. Beamforming for NOMA and mMIMO systems

At the mMIMO system, each device is served by a single beamforming vector. The ZF technique is a popular interference suppressing beamforming scheme in the mMIMO system since it eliminates all inter-user interference using individual beamforming for each device, while such interference suppressing is facilitated by the favorable propagation in massive MIMO configurations. Besides, to perform ZF precoding in NOMA system, it is essential to understand the NOMA *user-pairing* concept.

User-pairing: Inherent to the NOMA system, user clustering can be performed in several ways after the *user-sorting* and the user classification in center-users and edge-users subsets. Because we know that the SE of NOMA is directly proportional to the difference between the pathloss of the users, a natural choice consists in pairing users with as higher as possible pathloss differences [6]:

$$\Delta\beta_k = \beta_k - \beta_{K+1-k}, \quad (7)$$

forming the cluster k for $k = 1, \dots, K/2$. With the pair formed, carefully beamforming vectors selection is required. Hence, in NOMA we assume that the *beamforming vector* for paired users is the same, i.e., $\mathbf{g}_k = \mathbf{g}_{K+1-k}$ for all $k = 1, \dots, K/2$.

Assumption 2. In *user-pairing* procedure we assume that the paired users are aligned with the BS so that the same beamforming can serve all paired users simultaneously. Hence, by admitting that each pair of devices is spatially aligned with the BS, and using localizing tools described, for instance, in [23,24], one should assume further *a priori user-pairing* step in NOMA systems.

Assumption 3. In NOMA system, beamforming serves more than one aligned device simultaneously; specifically, in this paper, two aligned devices per cluster are admitted according to the *user-pairing* step, while eliminating the inter-cluster interference completely (favorable propagation) under adopted perfect CSI conditions.

In this work we adopt the linear ZF precoding as defined by the vector:

$$\mathbf{g}_k = \mathbf{h}_k' (\mathbf{h}_k'^T \mathbf{h}_k')^{-1}, \quad (8)$$

and satisfying $\mathbf{h}_i^T \mathbf{g}_k = 0, \forall i \neq k$, i.e., the favorable propagation effect between users belonging to distinct clusters.

3. SE-EE in NOMA and mMIMO systems

In this section, we discuss the SE and the EE configurations in the NOMA and mMIMO systems. The operation of NOMA system requires a process of pairing devices so that the channel coefficients of the devices in the same cluster must be appropriately different, enabling the power domain usage. As already mentioned, the interference cancellation process via beamforming presents problems that we will demonstrate below.

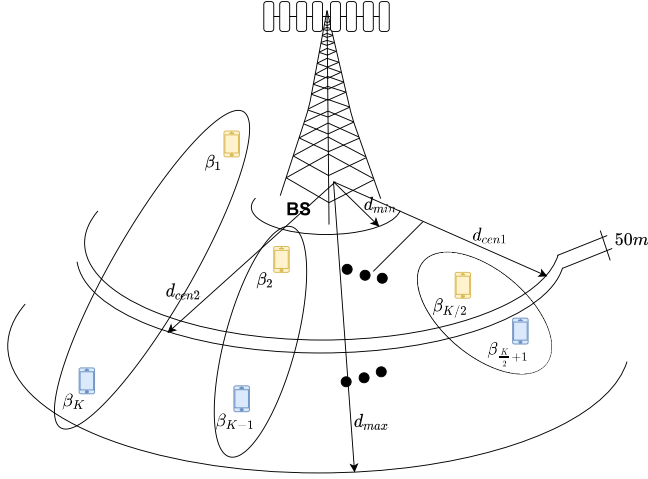


Fig. 2. System model indicating the pairing formation in NOMA system. Both mMIMO and NOMA systems deploy the same massive number of antennas at base-station, M .

3.1. Data rates in NOMA with ZF

Devices are divided into two sets like described in Eq. (1) and these groups are represented by Eq. (9) by their large-scale fading coefficient and are grouped to pairs forming a cluster as Fig. 2, the cluster k is formed by one device in cell center set $\mathcal{K}_H$ and one device in cell edge set $\mathcal{K}_L$. Hence, the devices are grouped into two subsets:

$$\begin{aligned} \mathcal{K}_H &= \{\beta_1 > \beta_2 > \dots > \beta_{K/2}\}, \quad (\text{center devices set}) \\ \mathcal{K}_L &= \{\beta_K < \beta_{K-1} < \dots < \beta_{K/2+1}\}. \quad (\text{edge devices set}) \end{aligned} \quad (9)$$

The *user-pairing* adopted in Eq. (9) is the same as proposed in [6], creating the largest possible difference in channel coefficients for devices not yet paired.

Assumption 4. In this paper, we assume perfect SIC, and only one perfect SIC stage per cluster is performed, since just 2 devices per cluster are admitted.

The instantaneous Signal to Interference plus Noise Ratio (SINR) of devices in cluster k is defined as:

$$\text{SINR}_k = \frac{\beta_k p_k |\mathbf{h}_k^T \mathbf{g}_k|^2}{\beta_k \sum_{k' \neq k}^K p_{k'} |\mathbf{h}_k^T \mathbf{g}_{k'}|^2 + 1}. \quad (10)$$

In each cluster, the cell-edge devices treat the interference as noise and decode their data symbols, whereas the cell-center device can decode the data symbols of the cell-edge device and perform SIC, hence effectively removing the interference due to the cell-edge device under **Assumption 3**.

To perform SIC, the cell-center device needs to be able to decode data signal intended for the cell-edge device, *i.e.*, the ergodic SINR of the cell-edge device, SINR_{K+1-k} , at device k , defined as $\text{SINR}_{k,K+1-k}$, must be greater than or equal to the ergodic SINR of the k th cell-center device. Hence, given the uplink (mMIMO and NOMA) and downlink (NOMA) pilot overhead and assuming perfect CSI in all receivers, and admitting **Assumption 4**, the following condition must be satisfied [4,10]:

$$\mathbb{E}[\text{SINR}_{k,K+1-k}] \geq \mathbb{E}[\text{SINR}_k], \quad (11)$$

where

$$\text{SINR}_{k,K+1-k} = \frac{\beta_k p_{K+1-k} |\mathbf{h}_k^T \mathbf{g}_k|^2}{\beta_k \sum_{k' \neq K+1-k}^K p_{k'} |\mathbf{h}_k^T \mathbf{g}_{k'}|^2 + 1}. \quad (12)$$

Herein, the condition in (11) must be satisfied by selecting the transmit powers appropriately.

The achievable ergodic rate of devices in cluster k , *i.e.* device k in $\mathcal{K}_H$ subset and device $K+1-k$ in $\mathcal{K}_L$ subset, under **Assumptions 1–4**, is given by the ergodic rate contribution of user-center device:

$$\mathcal{R}_k^{\text{NOMA}} = \tau \mathbb{E} [\log_2 (1 + \text{SINR}_k)], \quad \forall k \in \mathcal{K}_H \quad (13)$$

in [bits/s/Hz], and for the user-edge device:

$$\mathcal{R}_{K+1-k}^{\text{NOMA}} = \tau \mathbb{E} [\log_2 (1 + \text{SINR}_{K+1-k})], \quad \forall k \in \mathcal{K}_L \quad (14)$$

where $\tau = (1 - \frac{K T_s}{T})$, is the portion of each channel coherence interval (T) that is used for data transmission.

Assuming perfect channel state information, ZF precoding for inter-clusters interference elimination, and using random matrix theory results [25], the k th cluster NOMA achievable rate is obtained plugging Eqs. (10), (13) and (14):

$$\begin{aligned} \mathcal{R}_{\text{cl-}k}^{\text{NOMA}} &= \tau \mathbb{E} [\log_2 (1 + \bar{M} \beta_k p_k)] + \\ &\tau \mathbb{E} \left[\log_2 \left(1 + \frac{\beta_{K+1-k} p_{K+1-k}}{\beta_{K+1-k} p_k + 1} \right) \right], \end{aligned} \quad (15)$$

$\forall k \in \mathcal{K}$ and $\bar{M} = M + 1 - K/2$. Hence, the NOMA system can operate until $K < 2M - 1$. A detailed derivation of the expressions on this section can be found in [4] and [10].

3.2. Data rates in mMIMO with ZF

In the mMIMO system with ZF precoding the ergodic achievable rate for device k is given by:

$$\mathcal{R}_k^{\text{M-MIMO}} = \tau \mathbb{E} [\log_2 (1 + \text{SINR}_k)], \quad [\text{bits/s/Hz}] \quad (16)$$

where SINR_k is defined in (10). Hence, the above mMIMO achievable rate equation becomes:

$$\mathcal{R}_k^{\text{M-MIMO}} = \tau \mathbb{E} [\log_2 (1 + (M - K) p_k \beta_k)], \quad [\text{bits/s/Hz}], \quad (17)$$

where $(M - K)$ is obtained using random matrix theory, representing the coherent array gain of the received signal [25]. Under linear precoding and combiners, the mMIMO system operates consistently when $K < M$. Finally, the *average system sum-rate* (avg-sum-rate) is defined simply by:

$$\mathcal{R}^{\text{M-MIMO}} = \sum_{k=1}^K \mathcal{R}_k^{\text{M-MIMO}} \quad \text{and} \quad \mathcal{R}^{\text{NOMA}} = \sum_{k=1}^{\mathcal{K}_H} \mathcal{R}_{\text{cl-}k}^{\text{NOMA}}.$$

The mMIMO system equations have been thoroughly investigated in literature and can be found in [25,26].

3.3. Energy efficiency

EE metric is defined as the ratio of the number of effective bits of information received over the total energy consumed by the overall system to transmit and receive/decode such information. The number of effective information bits received at the destination can be determined by the system data rate. Power consumption required for processing the signal at the transmitter and receiver side is often neglected; in this sense, it is calculated just as proportional to the radiated transmitted power. The growth in the number of antennas in the BS and the increase number of devices in 5G systems can lead to unattainable EE goals. In general, the average EE can be expressed as:

$$\text{EE} = \frac{\sum_{k=1}^K \mathcal{R}_k}{P_{\text{TOT}}}, \quad [\text{bits/Joule/Hz}], \quad (18)$$

where P_{TOT} is the total power consumption across the entire communication system. It should consider transmission power

consumption, such as RF power amplifier inefficiency, baseband signal processing, and cooling, among others. Therefore, a more realistic and detailed energy consumption model is required.

Based on [2], the adopted power consumption model in our work considers two power terms: (a) fixed-term; (b) terms scaled with the number of antennas M and the number of devices K . The scaled terms occur because of the transceiver chains, coding/decoding, channel estimation, and precoding. Let the computational efficiency be L operations per joule in BS. We describe it as follows:

RF Power: P_{RF} is the power consumed to transmit the signal to active devices achieved the SINR target and $0 < \varpi \leq 1$ is the efficiency of the power amplifier.

Fixed consumption: P_{FIXED} is the power consumed at the BS which is independent of the number of transmit antennas and devices in the cell, is formed of term P_0 including the power consumption of backhaul infrastructure, control signaling, baseband processor, and term P_{SYN} a single oscillator used in all BS.

$$P_{\text{FIXED}} = P_0 + P_{\text{SYN}}$$

Dependence only on K : P_K is formed by the consumption to coding and modulation of information symbols to devices, represented by P_{COD} , the consumption to BS decoded the K sequences of information symbols, defined by P_{DEC} , and the received power, represented by P_{RX} , still composes this term, multiplied by K as well. In addition, a portion of the ZF precoding cost [27] depends only on K^3 .

$$P_K = K(P_{\text{COD}} + P_{\text{DEC}} + P_{\text{RX}}) + K^3 \frac{2}{3LT}$$

Dependence only on M : The term P_M represents the transmitted power (P_{TX}), hence

$$P_M = MP_{\text{TX}}$$

Dependence on K and M : the term P_{KM} is the cost of the ZF precoding (due to LU-based matrix inversion) [27], which depends on the number of devices, the number of antennas, and the vector information symbol.

$$P_{\text{KM}} = MK \frac{3+T}{TL} + MK^2 \frac{2}{TL}$$

Adding the portions, we obtain the overall power consumption of the system:

$$P_{\text{TOT}} = \frac{P_{\text{RF}}}{\varpi} + P_{\text{FIXED}} + P_K + P_M + P_{\text{KM}} \quad [\text{W}]. \quad (19)$$

3.4. Power allocation strategies

In the sequel, we present three well-known and frequently applied strategies for power allocation, but due to the inherent characteristics of NOMA, we propose modifications on the classical water-filling (WF) algorithm to enable application in the NOMA system. Such modifications, namely Δ -WF, ensure that none of the paired devices are dropped-out without undoing the pairing of devices. To guarantee a certain level of power disparities in each paired device, the power allocation Δ -WF procedure in the NOMA system has two steps: (a) first we allocated power for the clusters; (b) we allocate power between paired devices. Thereby we could analyze the behavior of the systems and compare their results.

Notice that both mMIMO and NOMA systems deploy the same massive number of antennas at base-station, M . Hence, due to the *channel hardening* effect [1,25] inherent to massive MIMO

configuration, the small-scale fading vanishes across the M antennas equipped with a linear ZF precoding with vector as Eq. (8). Hence, one can consider just the *pathloss coefficients* β_k as the main parameter in the power allocation policies of systems based on massive number of antennas.

3.4.1. Equal Power Allocation (EPA)

Equal Power Allocation (EPA) power allocation is deployed as a simple, naive strategy, where all devices are served with the same power. In mMIMO, all devices are served with the same transmission power regardless of their distance from the BS. In NOMA, power allocation has two steps. In the first step, the power is allocated equally between the pairs, and then we allocate the power of each device equally too.

The EPA strategy applied to mMIMO can be defined by:

$$p_k = \frac{P_{\text{RF}}}{K} \quad [\text{W}], \quad \forall k \in \mathcal{K}. \quad (20)$$

In the case of EPA procedure applied to NOMA, it is composed of two steps: in the first step, power reference to each cluster can be defined simply as:

$$p_{\text{ref}}^{\text{cl}} = \frac{2 \cdot P_{\text{RF}}}{K} \quad [\text{W}], \quad \forall k \in \mathcal{K}_H. \quad (21)$$

In the second step, the power allocation among the devices in the same cluster is defined as:

$$p_k^{\text{cl-k}} = p_{K+1-k}^{\text{cl-k}} = \frac{p_{\text{ref}}^{\text{cl}}}{2}. \quad (22)$$

3.4.2. Proportional Channel Inversion Power Allocation (PICPA)

is another power allocation technique adopted in this analysis. Unlike the EPA technique, which applies the same power to all devices, PICPA applies more power to devices with the worst channel conditions, favoring fairness across the devices. Such power allocation penalizes the average sum rate in favor of fairness among all users.

The Proportional to the Inverse of the Channel Power Allocation (PICPA) strategy applied to mMIMO can be defined as:

$$p_k = P_{\text{RF}} \frac{\beta_k^{-1}}{\sum_{k=1}^K \beta_k^{-1}} \quad [\text{W}], \quad \forall k \in \mathcal{K}, \quad (23)$$

while the PICPA procedure applied to NOMA follows two steps; in the first step, the power is allocated equally among the $K/2$ clusters:

$$p_{\text{ref}}^{\text{cl}} = \frac{2 \cdot P_{\text{RF}}}{K} \quad [\text{W}], \quad \forall k \in \mathcal{K}_H, \quad (24)$$

after that, the power of each device within the k th cluster is defined by allocating more power to the device with smaller large-scale fading β_k :

$$p_k^{\text{cl-k}} = p_{\text{ref}}^{\text{cl}} \frac{\beta_{K+1-k}}{\beta_k - \beta_{K+1-k}}, \quad \text{and} \quad p_{K+1-k}^{\text{cl-k}} = p_{\text{ref}}^{\text{cl}} - p_k^{\text{cl-k}}, \quad (25)$$

where $p_k^{\text{cl-k}}$ is the power allocated to the device k in the cl-k cluster, and $p_{K+1-k}^{\text{cl-k}}$ is the power allocated to the device $(K+1-k)$, also belonging to the k th cluster.

3.4.3. Classical Water-Filling (WF) algorithm

The application of the Water-Filling (WF) algorithm in mMIMO system results in an optimal (maximum) system sum-rate solution. However, some devices are dropped-out from the service

due to they experiment deteriorated channel condition. The WF power allocation strategy for mMIMO is described as:

$$\mu = \frac{1}{|\mathcal{K}|} \left(P_{\text{RF}} + \sum_{k=1, k \in \mathcal{K}} \frac{1}{\beta_k} \right), \quad (26)$$

$$P_{\text{RF}} = \sum_{k=1, k \in \mathcal{K}} p_k, \quad \text{where } p_k = \left(\mu - \frac{1}{\beta_k} \right)^+, \quad \forall k \in \mathcal{K}$$

$$\text{and } \mathbf{p} = [p_1, p_2, \dots, p_{|\mathcal{K}|}],$$

with the operator $(z)^+ = \max(0, z)$. Notice that the constrained value for the total power available is set to P_{RF} [W]. The Algorithm 1 describes the classical WF procedure.

Algorithm 1: Classical Water Filling (WF) for mMIMO

Input: $\mathcal{K}, P_{\text{RF}}, K = |\mathcal{K}|$
1 $\text{NP} \neq \emptyset$;
2 **while** ($\text{NP} \neq \emptyset$) **do**
3 solve Eq. (26) $\rightarrow \mathbf{p}$;
4 $\text{NP} \leftarrow$ identify null positions in $\mathbf{p}$;
5 $\mathcal{K}/\{k\}_{\text{NP}} \leftarrow$ exclude from $\mathbf{p}$ devices labeled as NP
6 **end**
Output: $\mathbf{p} = [p_1, p_2, \dots, p_{|\mathcal{K}|}]$

On the other hand, the direct application of WF algorithm in the NOMA system implies harming the pair formation, i.e., devices present in the $\mathcal{K}_L$ set are effectively dropped-out of the service, undoing the pair. To allow some comparison, we propose a modification in classical WF like the following.

3.4.4. Δ -WF for NOMA

The application of classical WF in NOMA in the same way as it is applied to mMIMO causes some formed pairings to be broken, since after WF algorithm application, some devices are dropped-out from the system, making the NOMA power difference (Δ) in the devices of the same cluster vanish. Hence, we suggest modifying the classical WF procedure to be applied to NOMA accordingly. The steps of Δ -WF algorithm are described as follows.

In NOMA, the power allocation has two steps, in the first step, the allocation is between clusters. Hence, to prevent the formed pairs be broken, we propose the application of WF based on the large-scale fading differences of the paired devices, as defined in Eq. (7): $\Delta\beta_k = (\beta_k - \beta_{K+1-k})$ inside each $\mathcal{K}_L$ and $\mathcal{K}_H$ subsets, Eq. (9). In the second step of the procedure, the power is allocated between the devices inside the group, assuming perfect successive interference cancellation (SIC); for this to be possible the condition in Eq. (11) must be satisfied. The new water-level in the modified Δ -WF power allocation strategy for NOMA is defined by

$$\mu = \frac{2}{\mathcal{K}_H} \left(P_{\text{RF}} + \sum_{k=1, k \in \mathcal{K}_H} \frac{1}{\Delta\beta_k} \right), \quad (27)$$

$$P_{\text{RF}} = \sum_{k=1, k \in \mathcal{K}_H} p_{\text{cl-}k}, \quad \forall k \in \mathcal{K}_H$$

$$\text{where } p_{\text{cl-}k} = \left(\mu - \frac{1}{\Delta\beta_k} \right)^+,$$

$$\text{and } \mathbf{p}_{\text{cl-}k} = [p_1, p_2, \dots, p_{|\mathcal{K}_H|}],$$

In the second step, the power allocation to the both devices in the k th cluster is simply defined as:

$$p_{K+1-k}^{\text{cl-}k} = p_k^{\text{cl-}k} = \frac{p_{\text{cl-}k}}{2} \quad (28)$$

Algorithm 2 summarize the proposed Δ -WF power allocation procedure, aiming to improve the SE of NOMA systems.

Algorithm 2: Δ -WF (modified) for NOMA systems

Input: $\mathcal{K}_H, \mathcal{K}_L, P_{\text{RF}}$
1 $\text{NP} \neq \emptyset$;
2 **while** ($\text{NP} \neq \emptyset$) **do**
3 solve Eq. (27) $\rightarrow \mathbf{p}_{\text{cl-}k}$;
4 $\text{NP} \leftarrow$ identify null position in $\mathbf{p}_{\text{cl-}k}$;
5 $\mathcal{K}_H/\{k\}_{\text{NP}} \leftarrow$ exclude from $\mathbf{p}_{\text{cl-}k}$ devices labeled as NP
6 **end**
Output: $\mathbf{p}_{\text{cl-}k} = [p_1, p_2, \dots, p_{|\mathcal{K}_H|}]$

Complexity analysis: In a comparative analysis of complexity, the Δ -WF algorithm for power allocation in NOMA system (Algorithm 2) performs two simple additional operations compared to the classical WF procedure (Algorithm 1): (a) in Eq. (27) the subtraction in $(\beta_k - \beta_{K+1-k})$; and (b) the division by two in (28). Besides, NOMA vs. mMIMO systems require different *a priori* information to proceed accordingly with the PA procedure.

3.5. Prior information for power allocation step

For implementing the PA policies, prior information are required at the BS, as defined in Table 1. Some of these necessary information can be obtained through the dedicated *pilot* transmission step, at cost of some overhead, as described in Section 2.2. Moreover, a prior step is assumed known, in which the *spatial localization* and *path loss* estimation of all devices must be realized. With such *a priori* information availability, the *user-sorting* and *user-pairing* steps can be performed.

4. Numerical results

The numerical evaluations for the proposed analyses of NOMA and mMIMO systems are presented in this section. The simulation system and channel parameter values deployed along this section are depicted in Table 2. The BS is located at the center of cell and equipped with a massive M BS transmit antennas in typical non-line-of-sight (NLOS) channel propagation scenario, while the devices are randomly distributed in the cell area and split into two subsets, $\mathcal{K}_L$ and $\mathcal{K}_H$, as illustrated in Fig. 2. In all simulations we consider a block fading model where the time-frequency resources are divided into coherence time intervals (T), in which the channels remain constant and frequency flat, and it is measured in multiple of symbol transmit period (T_s). The system and channel scenarios have been simulated using Matlab 2019 software running under one Intel HD Graphics 6000 GPU, Intel(R) Dual-Core(TM) i5 CPU @ 1.6 GHz and 8 GB RAM.

4.1. Spectral efficiency comparison

The mMIMO and NOMA SE performance analysis is carried out in this subsection, by increasing the number of devices two by two until the loading limit $\rho = 2$. The results consider $M = 64, 128$ and 256 BS antennas. In Fig. 3(a) the results of SE achieved when the RF power available is allocated following EPA strategy, where each device receive the same PA values. The mMIMO system overcame NOMA in all situations when the loading $\rho < 0.6$. However, the NOMA system achieves a higher SE than the mMIMO for each M scenario when the loading of devices increases, $\rho > 0.6$. The maximum avg-SE is 373 [bits/s/Hz], being attained with ZF-NOMA $M = 256$ antennas and $\rho \approx 0.76$. Besides, one can infer that the mMIMO does not work with a loading higher than 1, due to the array gain reaching 0 at full

Table 1

Prior information required to PA procedure.

PA		β_k	h'_k	K	κ_H	κ_L	p_k , eq.
EPA	mMIMO	—	—	▲*	—	—	(20)
	NOMA	—	—	▲	—	—	(21), (22)
PICPA	mMIMO	▲*	—	—	▲*	▲*	(23)
	NOMA	▲*	—	—	▲*	▲*	(24), (25)
WF	mMIMO	▲*	—	—	▲*	▲*	(26), Alg. 1
Δ -WF	NOMA	▲	—	—	▲*	▲	(27) and (28), Alg. 2

▲ Information needed a prior, * Obtained via pilot overhead.

— Information not needed.

Table 2

Simulation parameters.

Parameter	Value
BS antennas	$M = 64, 128$ and 256
Max. # Devices in the cell	$K = \zeta \cdot M$ (NOMA) $K = M$ (mMIMO)
Cell loading	$\rho = K/M$
Total RF power available	$P_{RF} = 1$ W
Pairs NOMA/Clusters	$N = K/\zeta = K/2$
NOMA devices per cluster	$\zeta = 2$
# antennas per device	1
Cell edge length	$d_{max} = 350$ m
Strong device position	$[d_{min}; d_1] \in [50; 100]$ m
Weak device position	$[d_2; d_{max}] \in [150; 350]$ m
Array gain mMIMO device	$M - K$
Array gain NOMA k_H	$M + 1 - K/2$
Data symbol period	T_s
Coherence time interval	$T = 512 \cdot T_s$, $\ell = 512$
Channel	
Pathloss exponent	$\xi = 3.78$
Attenuation at a d_0 reference	$\beta_0 = 130$ [dB]
# Monte-Carlo realizations	1000

loading of devices, while NOMA works suitably until the loading attains $M \cdot \zeta$, where ζ is number of devices per cluster.

Fig. 3(b) depicts the results of SE achieved in the mMIMO and NOMA systems when the PICPA method is applied to allocate the available RF power per device along the BS antennas. The maximum avg-SE in mMIMO system overcomes the NOMA counterpart until the loading ρ exceeds ≈ 0.62 for the three BS antenna configurations, $M = 64, 128$ and 256 . This PA technique provides more power to devices with the worst channel condition, making the SE result reach maximum values below the EPA.

Fig. 3(c) depicts the conventional WF algorithm applied to mMIMO. Under such power allocation approach, we highlight that the formation of pairs is unfeasible in the NOMA system. Indeed, the WF algorithm is able to maximize the system SE since it allocates more power to devices with better channel conditions, while devices under bad channel conditions (below the water level) are dropped-out from the service.

The classical WF algorithm has been adapted to the NOMA system dropped-out always a pair of devices. Such adaptation reveals substantial improvements of avg-sum-rate when M is low in comparison with classical WF PA in mMIMO. The Δ -WF power allocation procedure preserves the pairs clustering formation in the NOMA system, allocating more power to the cluster with higher difference between coefficients of large-scale fading. Fig. 3(c) shows that the maximum avg-SE ≈ 361 [bits/s/Hz], which is achieved under $\rho = 1$ ($K \approx 256$ devices) when the modified WF is deployed in NOMA system. Moreover, when the number of BS antennas is lower ($M = 64$ or 128), the NOMA achieved a peak higher than mMIMO, e.g., for $M = 64$ antennas, the peak of SE mMIMO occurs at loading $\rho \approx 0.7$, while the

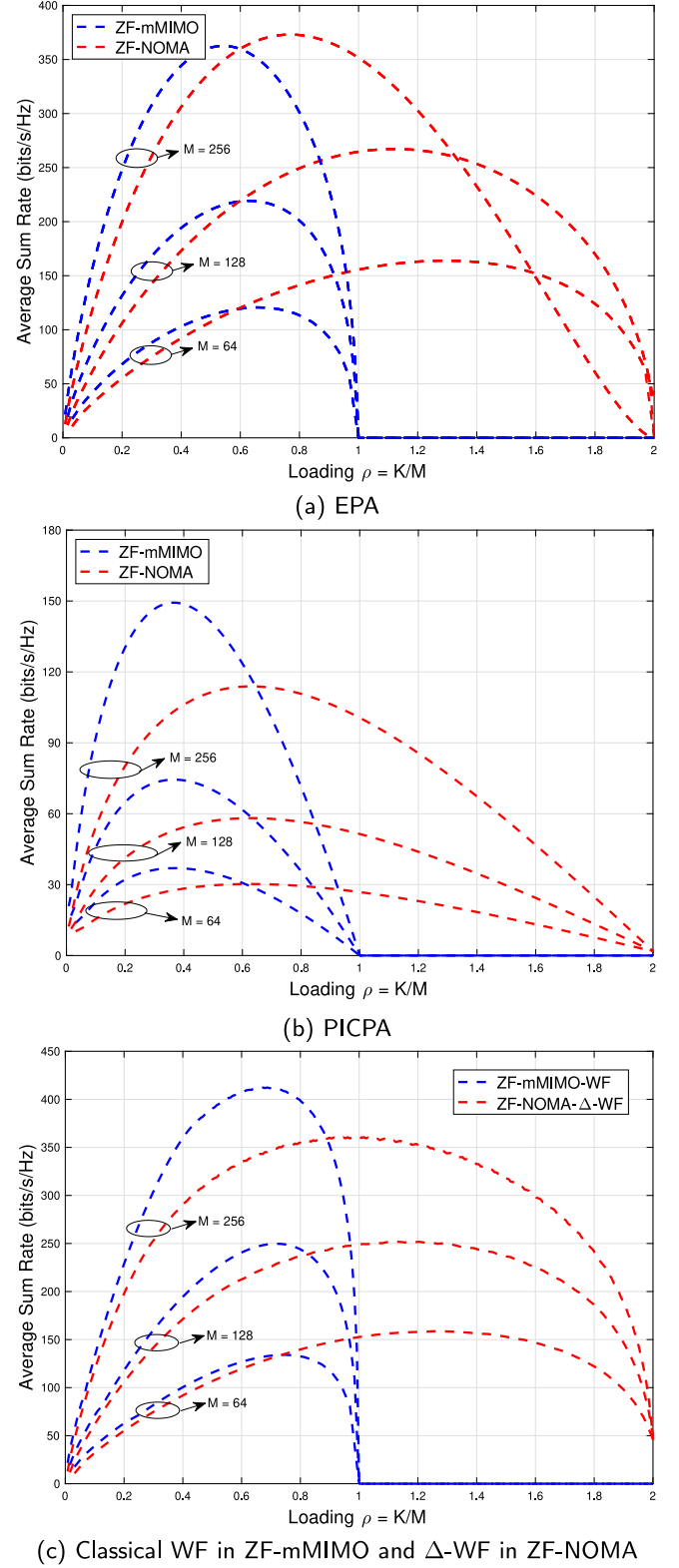


Fig. 3. The average sum-rate with loading of devices $0 < \rho \leq 2$, considering four power allocation methods: EPA, PICPA, WF, and Δ -WF. The average is obtained over 1000 random devices locations.

NOMA SE peaks at $\rho \approx 1.2$. However, as the number of BS antennas grows, the NOMA SE advantage decreases.

Number of active devices after PA procedure. Fig. 4 shows the number of active device after applying PA methods: in the EPA

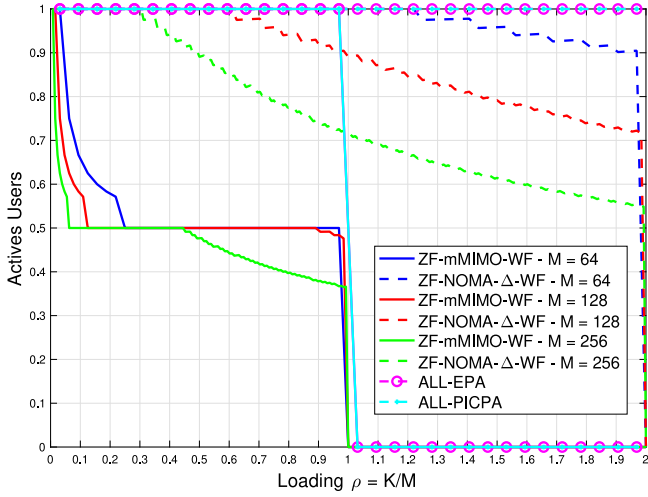


Fig. 4. The average of active devices after PA procedure versus loading of devices in the range $0 < \rho \leq 2$. The average is obtained over 1000 random devices locations.

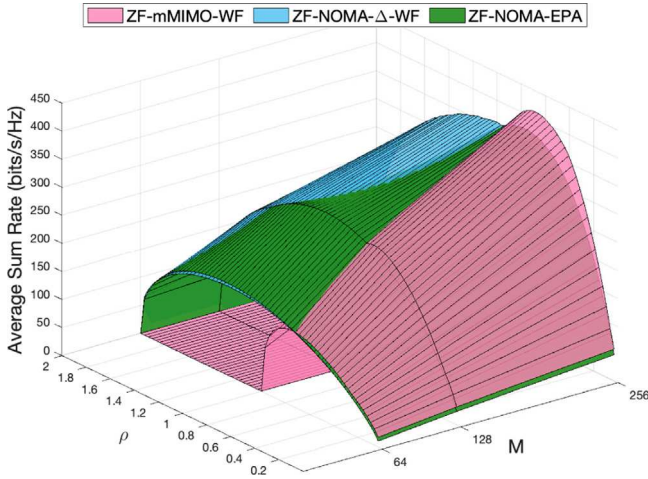


Fig. 5. Average sum rate with loading ρ and M .

and PICPA algorithms, all devices remain activated. However, in classical WF mMIMO system when the number of device increases beyond $\rho \approx 0.25$, the half of devices are dropped-out; while in Δ -WF NOMA PA, the percentage of active devices is always higher, e.g. higher than 70% for $\rho \approx 1.1$ and $M = 256$ antennas, the worst case.

Fig. 5 summarizes avg-sum-rate surfaces in terms of $SE \times \rho \times M$ results achieved by NOMA with EPA, mMIMO with WF, and NOMA with modified Δ -WF. In the initial loading part, $\rho < 0.65$, the classical WF PA in ZF-mMIMO achieves the better results until the loading (pink surface). When the number of antennas is low as $M = 64$ and ρ is in between 0.7 and 1.8, the EPA PA applied to NOMA (green surface) achieved superior results. Moreover, when $M = 128$ and $0.8 < \rho < 1.6$, the ZF-NOMA-EPA achieve superior results (green surface). For a higher number of antennas in BS, i.e., $M = 256$ only in a short loading of devices range, $0.86 < \rho < 0.97$, the ZF-NOMA-EPA achieves superior SE results. Finally, when $\rho > 0.97$ the modified Δ -WF achieves competitive results (blue surface).

4.2. Jain's fairness index

Another important analysis developed was to analyze the fairness between the devices, i.e., to know the difference in the

Table 3

Adopted parameters values for EE analysis [2].

Parameter	Value
Backhaul infrastructure	$P_0 = 2$ W
Single oscillator	$P_{syn} = 2$ W
Coding and modulation	$P_{COD} = 4$ W per device
Decoding and demodulation	$P_{DEC} = 0.5$ W per device
Receive power	$P_{RX} = 0.3$ W per device
Transmitted power	$P_{TX} = 1$ W per antenna
Efficiency of power amplifier	$\sigma = 0.3$
Operations/Joule	$L = 10^9$ oper. per joule

transmission rate achieved by active devices in the cell. For this measure we use the Jain's Fairness index, like described in [28] and can be defined as:

$$\mathcal{F}_M^{\text{SYST}} = \frac{\left(\sum_{k=1}^M R_k\right)^2}{M \sum_{k=1}^M R_k^2}. \quad (29)$$

The Fig. 6 depicts the fairness curves attainable by NOMA and mMIMO systems with EPA, PICPA, WF and Δ -WF PA procedures when the loading of devices grow until $\rho = 2$. Fig. 6(a) shows the Jain's Fairness Index when EPA policy is used, the mMIMO system performs better $\mathcal{F}$ results than NOMA for $\rho < 1$, on the other hand, NOMA can attains $\mathcal{F}_M^{\text{NOMA}} \approx 0.5$ in almost every loading of devices, independent of M .

Fig. 6(b) reveal the Jain's Fairness Index when PICPA method is applied, despite the SE result being lower in this PA method, the mMIMO obtains the best fairness result, keeping the $\mathcal{F}_M^{\text{mMIMO}}$ always above 0.85, still the NOMA under 0.5.

The Jain's Fairness Index when WF and δ -WF are depicted in Fig. 6(c), it is intrinsic to these algorithms to allocate more power to device with better channel conditions, which causes fairness between devices to be impaired. In this method it is possible to observe a significant influence of the number of antennas M in the BS and the result of fairness.

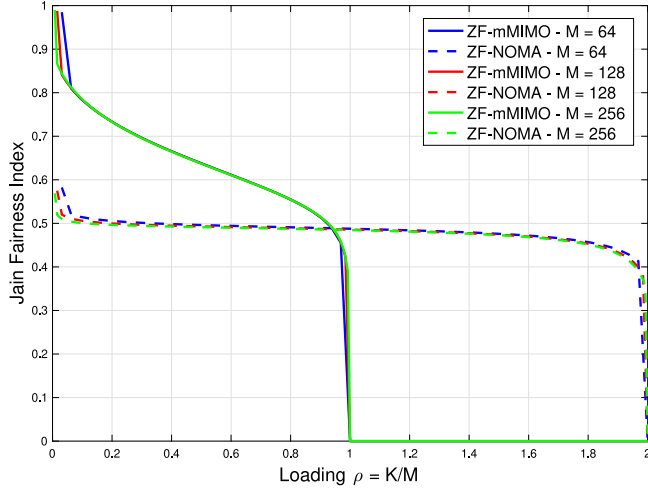
4.3. Energy efficiency comparison

Energy efficiency (EE) is another important figure of merit used to compare systems performance. In this section, power consumption model based on fixed circuitry power part and that one varying according to the number of antennas M and the number of devices K has been adopted, following Eq. (19).

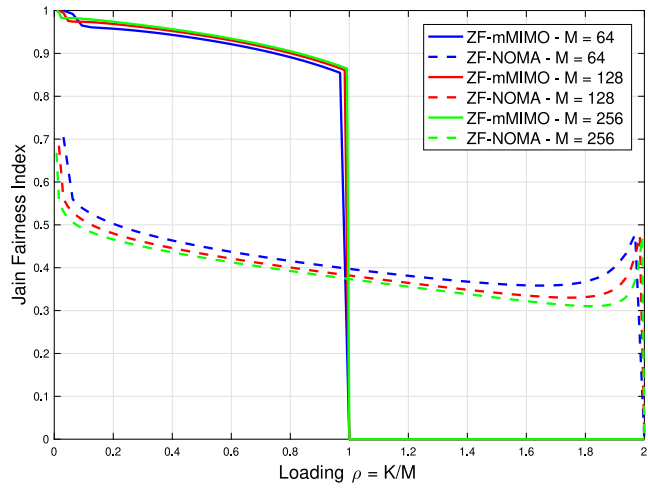
Table 3 [2] presents the adopted parameter values for the EE analysis and comparison discussed in this subsection. Fig. 7 depicts performance of EE with EPA, PICPA, WF and Δ -WF PA procedures, considering the same three quantities of antennas. Fig. 7(a) shows the EE performance with EPA. In this method all devices receive the same power. The avg-EE mMIMO overcomes the NOMA around 13% to 20%. It was possible to observe that the addition of antennas in the BS increases the power consumption, harming the EE result. Again, under loading of devices $0.6 < \rho \leq 2.0$, the NOMA overcomes the mMIMO system.

Fig. 7(b) reveals the EE performance with PICPA. In this method more power is allocated to devices with poor channel coefficient, resulting in reduced poor EE performance for both NOMA and M-MIMO system, attaining a maximum of 0.25 and 0.16 [bits/W] for mMIMO and NOMA, respectively. In general, the maximum EE attained by mMIMO is around 50% higher than NOMA. However, for loading of devices $\rho \geq 0.65$, NOMA overcomes mMIMO EE performance.

Fig. 7(c) depicts the EE performance in the mMIMO with classical WF and in the NOMA with Δ -WF algorithm. It is possible to confirm the superiority in terms of energy efficiency of mMIMO within the range where it operates consistently, i.e., $0 < \rho < 1$.



(a) EPA



(b) PICPA

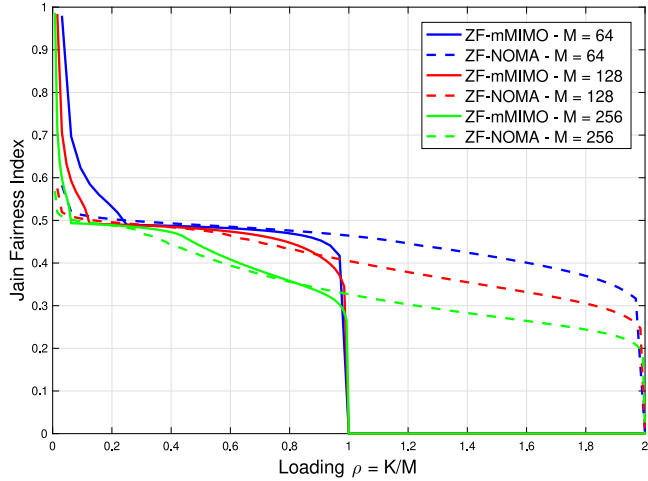
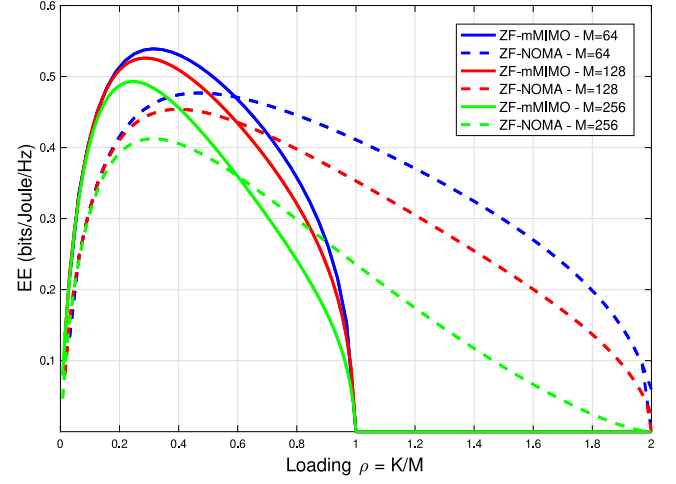
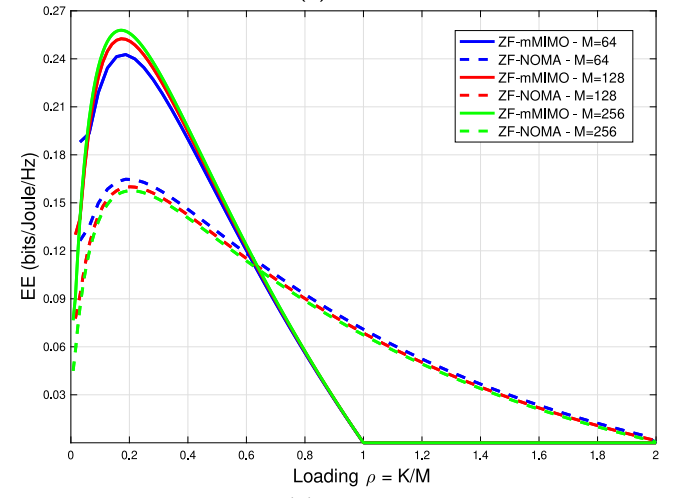
(c) WF and Δ -WF

Fig. 6. The fairness of NOMA and mMIMO system under three power allocation procedures: (a) EPA; (b) PICPA; (c) WF and Δ -WF. The average is obtained over 1000 random devices locations.

Notice that the maximum EE achieved by mMIMO are about 70% higher than NOMA for different BS antennas. Finally, NOMA becomes more energy efficient than mMIMO only when the loading of devices is high, $\rho > 0.95$.



(a) EPA



(b) PICPA

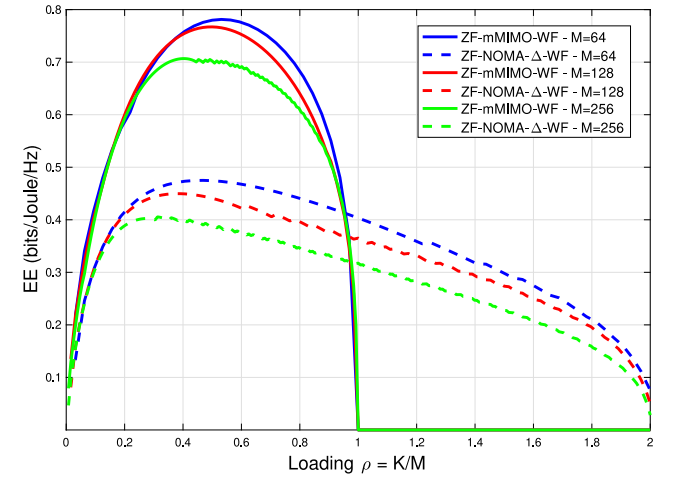
(c) WF and Δ -WF

Fig. 7. EE for NOMA vs. mMIMO under three power allocation procedures: (a) EPA; (b) PICPA; (c) WF and Δ -WF. Average EE obtained over 1000 random devices locations.

In all system scenario analyzed the mMIMO equipped with classical WF PA procedure achieves higher maximum EE. In general, the mMIMO attains better EE results than NOMA for $\rho < 1$.

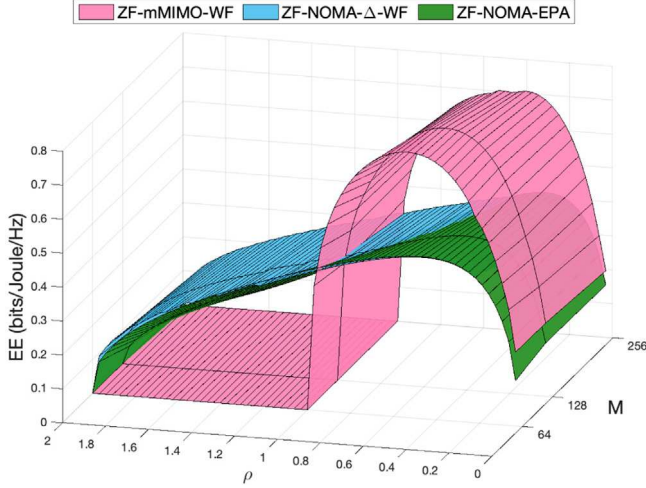


Fig. 8. EE with loading ρ and M .

On the other hand, NOMA can serve a larger number of devices (twice) than mMIMO.

Fig. 8 summarizes the best EE results in a surface plotting for the mMIMO with WF overcoming NOMA across the entire loading range where it operates consistently. For loading of devices $\rho > 1$, the NOMA operates under lower EE until the loading $\rho = 2$. Moreover, considering the smallest number of antennas in the BS, the NOMA with EPA overcame the NOMA with modified Δ -WF; in spite of that, as the number of antennas in the BS increases, the NOMA with Δ -WF achieves marginal superior EE results.

4.4. Area under curves SE and EE

For a fair comparison, one can consider a wide range of average SE and EE along the loading of devices, and normalized per antenna, which can be attainable by NOMA and mMIMO systems. Hence, let us consider the corresponding areas under the SE and EE curves in Figs. 3 and 7, such that:

$$\mathcal{S}_M^{\text{SYST}} = \frac{1}{M} \cdot \int_0^{\rho=2} \overline{\text{SE}}(\rho) d\rho \quad \left[\frac{\text{bits/antenna}}{\text{s Hz}} \right]$$

and

$$\mathcal{E}_M^{\text{SYST}} = \frac{1}{M} \cdot \int_0^{\rho=2} \overline{\text{EE}}(\rho) d\rho \quad \left[\frac{\text{bits/antenna}}{\text{J Hz}} \right],$$

respectively, where $\overline{\text{SE}}(\rho)$ is the average overall system sum-rate, and $\overline{\text{EE}}(\rho)$ is the average overall system energy efficiency achieved under specific loading of devices ρ . Hence, comparing the values of corresponding areas under the SE and EE curves of Figs. 3 and 7, we obtained Fig. 9.

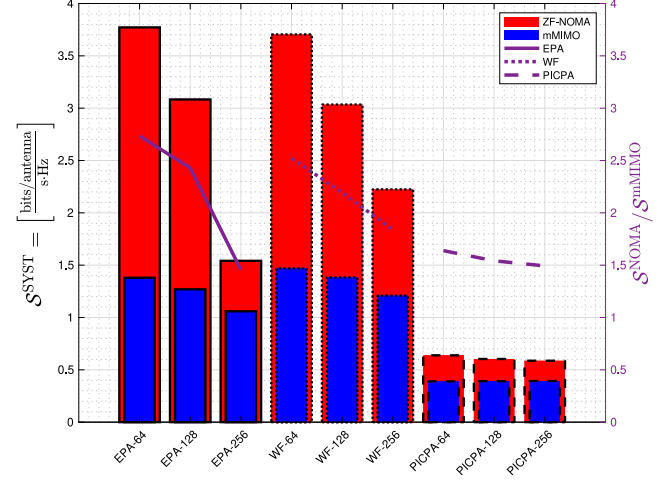
From the SE perspective, and considering EPA policy, the higher area-under-SE-curve ratio gain is achieved when the number of BS antennas is $M = 64$:

$$\mathcal{S}_{M=64}^{\text{NOMA}} \approx 2.7 \cdot \mathcal{S}_{M=64}^{\text{mMIMO}}.$$

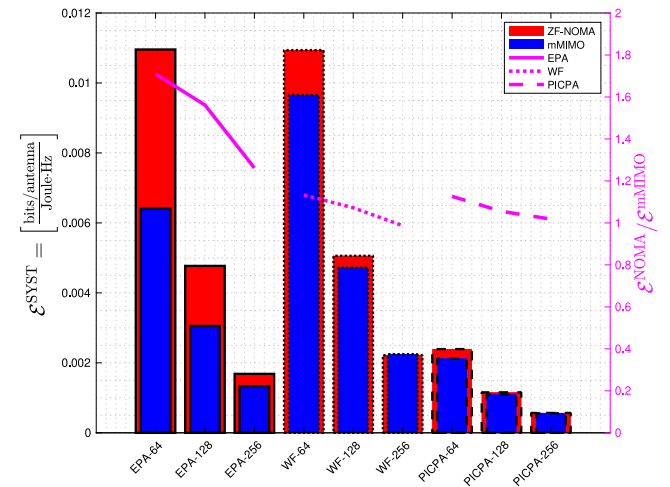
Notice that when the number of antennas M grow the ratio above decrease. In the same way, considering WF policy the gain trend remains. In contrast, considering PICPA policy, the ratio practically remains the same.

Furthermore, considering now the EE perspective, under EPA policy, the higher ratio is achieved when the BS is equipped with $M = 256$ antennas:

$$\mathcal{E}_{M=256}^{\text{NOMA}} \approx 1.8 \cdot \mathcal{E}_{M=256}^{\text{mMIMO}}.$$



(a) $\mathcal{S}$ - Bar chart of Area Under SE curves.



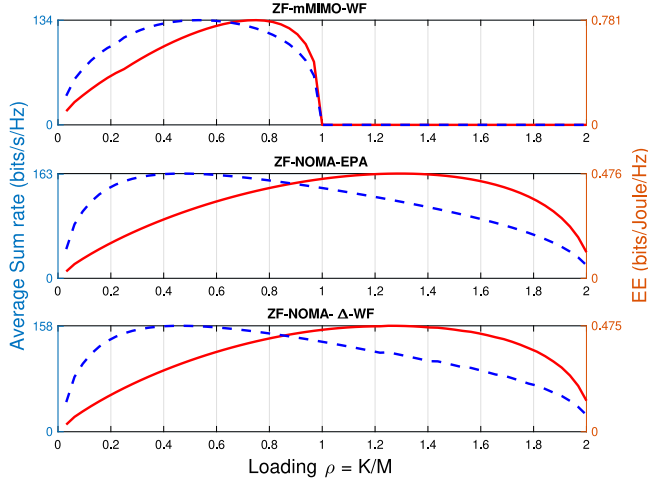
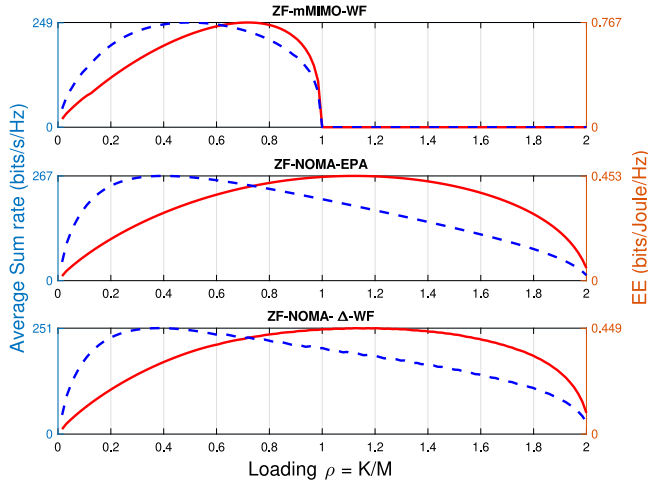
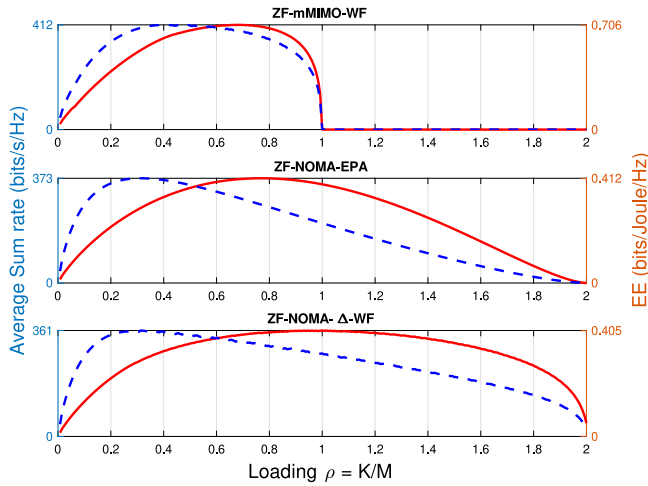
(b) $\mathcal{E}$ - Bar chart of Area Under EE curves.

Fig. 9. The Area Under curve of NOMA and mMIMO system under three power allocation procedures: (a) SE curves; (b) EE curves. The average is obtained over 1000 random devices locations.

As one can conclude, NOMA in almost all scenarios, is more spectrally and energetically efficient than mMIMO over a very wide range of loading of devices $0 < \rho \leq 2$, roughly, in average, 80% in terms of energy efficiency, and 170% more efficient in terms of spectral efficiency.

4.5. Resource efficiency (SE-EE trade-off)

The NOMA and mMIMO are analyzed in terms of SE and EE trade-off, namely resource efficiency (RE), considering loading of devices increasing up to 2. From Fig. 10, one can find graphically the best loading of devices range that maximizes the SE-EE trade-off for each BS antenna configuration M . The left y-axis depicts avg-SE and the right y-axis shows the avg-EE. Table 4 summarizes the optimal loading of devices that maximize the SE-EE trade-off and show the percentage of active users after Power Allocations and the Jain's Fairness Index. Fig. 10(a) reveals the results when $M = 64$, NOMA with EPA achieved SE-EE trade-off with the highest loading of devices and the highest SE in trade-off; on the other hand, mMIMO with classical WF achieved the highest SE-EE trade-off, however the percentage of active devices is around 0.5. Fig. 10(b) depicts results when $M = 128$, NOMA with EPA achieved SE-EE trade-off with the highest loading of

(a) $M = 64$ BS antennas(b) $M = 128$ BS antennas(c) $M = 256$ BS antennasFig. 10. SE-EE trade-off points when $M = 64, 128$ and 256 .

devices, in contrast mMIMO with classical WF with lower loading of devices achieved higher values of SE and EE in trade-off with a half of active devices. Fig. 10(c) shows the results when $M =$

Table 4
SE-EE trade-off.

$M = 64$					
	ρ	SE	EE	Actives users	$\mathcal{F}$
mMIMO-WF	0.652	131.15	0.762	.50	.485
NOMA-EPA	0.875	147.45	0.428	1.0	.485
NOMA- Δ -WF	0.844	143.48	0.441	1.0	.475
$M = 128$					
	ρ	SE	EE	Actives users	$\mathcal{F}$
mMIMO-WF	0.625	243.21	0.745	.50	.475
NOMA-EPA	0.734	241.54	0.411	1.0	.49
NOMA- Δ -WF	0.703	226.32	0.405	.98	.45
$M = 256$					
	ρ	SE	EE	Actives users	$\mathcal{F}$
mMIMO-WF	0.578	404.84	0.691	.47	.43
NOMA-EPA	0.523	344.70	0.380	1.0	.495
NOMA- Δ -WF	0.594	333.31	0.375	.85	.398

256, NOMA with Δ -WF achieved SE-EE trade-off with the highest loading of devices, and one more time mMIMO with WF achieved higher SE-EE trade-off with 47% of active devices. It is possible to demonstrate that the increase of antennas in the BS improves the SE result, on the other hand it worsens the EE result.

5. Conclusion and future works

This work proposes a comparative SE and EE analyze in DL single-cell between mMIMO and NOMA with BS equipped with three different antennas configurations. Under SE perspective mMIMO with classic WF algorithm achieved better results under low and medium loading. On the other hand, when the system loading is higher $\rho > 0.6$ the NOMA achieved better results until $\rho = 2$.

The analyzed PA methods applied to NOMA system, (EPA, PICPA, and Δ -WF) result in different SE performance. Indeed, when the channel hardening condition is fully attained, and the amount of BS antennas increases ($M = 128$ and 256), the best SE results are attained with the proposed Δ -WF algorithm, but, as expected, the fairness index is harmed.

Under EE perspective, the mMIMO achieved better results when employed the three EPA, PICPA, and WF PA methods under $K < M$. However, the NOMA can operate under higher system loading, i.e., $K < 2M - 1$.

In terms of area-under-SE-curve and EE-curve metrics, $\mathcal{S}$ and $\mathcal{E}$, respectively, the NOMA system attained better results, due to its ability to serve a larger number of users than mMIMO. Such numerical results confirm NOMA's ability to operate with high loading of devices. On the other hand, it is not possible to achieve high fairness with NOMA.

From the perspective of SE-EE trade-off, mMIMO achieved the best results, because of the superiority in EE; always achieved in loading of devices $\rho = 0.6$ in all M setups.

NOMA systems present interesting features and has been intensively investigated as a promising technique in devising future wireless generations. As future works, hybrid NOMA systems, as well as alternatives techniques such as *rate-splitting multiple access* (RSMA) can improve the overall EE of massive MIMO systems.

CRedit authorship contribution statement

Thiago A. Bruza Alves: Conception and design of study, Acquisition of data, Drafting the manuscript. **Taufik Abrão:** Conception and design of study, Analysis and/or interpretation of data, Drafting the manuscript, Revising the manuscript critically for important intellectual content.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

All authors approved the version of the manuscript to be published.

References

- [1] E.G. Larsson, O. Edfors, F. Tufvesson, T.L. Marzetta, Massive MIMO for next generation wireless systems, *IEEE Commun. Mag.* 52 (2) (2014) 186–195.
- [2] E. Bjornson, L. Sanguinetti, J. Hoydis, M. Debbah, Designing multi-user MIMO for energy efficiency: When is massive mimo the answer? in: 2014 IEEE Wireless Communications and Networking Conference (WCNC), IEEE, 2014, pp. 242–247.
- [3] Z. Ding, H.V. Poor, Design of massive-MIMO-noma with limited feedback, *IEEE Signal Process. Lett.* 23 (5) (2016) 629–633.
- [4] K. Senel, H.V. Cheng, E. Bjornson, E.G. Larsson, What role can NOMA play in massive mimo? *IEEE J. Sel. Top. Sign. Proces.* 4553 (c) (2019) 1–16.
- [5] A. Anwar, B.-C. Seet, M.A. Hasan, X.J. Li, A survey on application of non-orthogonal multiple access to different wireless networks, *Electronics* 8 (11) (2019) 1355.
- [6] M.S. Ali, E. Hossain, D.I. Kim, Non-orthogonal multiple access (NOMA) for downlink multiuser mimo systems: User clustering, beamforming, and power allocation, *IEEE Access* 5 (2017) 565–577.
- [7] M. Zeng, A. Yadav, O.A. Dobre, G.I. Tsiropoulos, H.V. Poor, Capacity comparison between MIMO-noma and MIMO-OMA with multiple users in a cluster, *IEEE J. Sel. Areas Commun.* 35 (10) (2017) 2413–2424.
- [8] S.M.R. Islam, M. Zeng, O.A. Dobre, K.-S. Kwak, Resource allocation for downlink NOMA systems: Key techniques and open issues, *IEEE Wirel. Commun.* 25 (2) (2018) 40–47.
- [9] X. Chen, F.-K. Gong, G. Li, H. Zhang, P. Song, User pairing and pair scheduling in massive MIMO-noma systems, *IEEE Commun. Lett.* 22 (4) (2018) 788–791.
- [10] H.V. Cheng, E. Bjornson, E.G. Larsson, Performance analysis of NOMA in training-based multiuser mimo systems, *IEEE Trans. Wireless Commun.* 17 (1) (2018) 372–385.
- [11] L. Dai, B. Wang, M. Peng, S. Chen, Hybrid precoding-based millimeter-wave massive MIMO-noma with simultaneous wireless information and power transfer, *IEEE J. Sel. Areas Commun.* 37 (1) (2019) 131–141.
- [12] B. Makki, K. Chitti, A. Behravan, M.-S. Alouini, A survey of NOMA: Current status and open research challenges, *IEEE Open Journal of the Communications Society* 1 (December 2019) (2020) 179–189.
- [13] M.R. Usman, A. Khan, M.A. Usman, Y.S. Jang, S.Y. Shin, On the performance of perfect and imperfect SIC in downlink non orthogonal multiple access (NOMA), in: 2016 International Conference on Smart Green Technology in Electrical and Information Systems (ICSGTEIS), IEEE, 2016, pp. 102–106.
- [14] E. Bjornson, E.G. Larsson, How energy-efficient can a wireless communication system become? in: Conference Record - Asilomar Conference on Signals, Systems and Computers, Vol. 2018-October, IEEE, 2019, pp. 1252–1256.
- [15] Y. Zhang, H.M. Wang, T.X. Zheng, Q. Yang, Energy-efficient transmission design in non-orthogonal multiple access, *IEEE Trans. Veh. Technol.* 66 (3) (2017) 2852–2857.
- [16] S. Mounchili, S. Hamouda, Pairing distance resolution and power control for massive connectivity improvement in NOMA systems, *IEEE Trans. Veh. Technol.* 69 (4) (2020) 4093–4103.
- [17] F. Rezaei, C. Tellambura, A. Tadaion, A.R. Heidarpour, Rate analysis of cell-free massive MIMO-noma with three linear precoders, *IEEE Trans. Commun.* 68 (6) (2020) 3480–3494.
- [18] H. Zhang, H. Zhang, W. Liu, K. Long, J. Dong, V.C. Leung, Energy efficient user clustering, hybrid precoding and power optimization in terahertz MIMO-noma systems, *ArXiv* 38 (9) (2020) 2074–2085.
- [19] O. Maraqa, A.S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, S.M. Sait, A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks, *IEEE Commun. Surv. Tutor.* 22 (4) (2020) 2192–2235.
- [20] S. Rajoria, A. Trivedi, W.W. Godfrey, Sum-rate optimization for NOMA based two-tier hetnets with massive MIMO enabled wireless backhauling, *AEU - Int. J. Electron. Commun.* 132 (October 2020) (2021) 153626.
- [21] D. Kim, M. Choi, Non-orthogonal multiple access in distributed antenna systems for max-min fairness and max-sum-rate, *IEEE Access* 9 (2021) 69467–69480.
- [22] I. Budhiraja, N. Kumar, S. Tyagi, S. Tanwar, Z. Han, D.Y. Suh, M.J. Piran, A Systematic Review on NOMA Variants for 5G and Beyond, Vol. 9, *IEEE*, 2021, pp. 85573–85644.
- [23] F. Zafari, A. Gkelias, K.K. Leung, A survey of indoor localization systems and technologies, *IEEE Commun. Surv. Tutor.* 21 (3) (2019) 2568–2599.
- [24] E.M. Mohamed, Joint users selection and beamforming in downlink millimetre-wave noma based on users positioning, *IET Commun.* 14 (8) (2020) 1234–1240.
- [25] T.L. Marzetta, E.G. Larsson, H. Yang, H.Q. Ngo, *Fundamentals of Massive MIMO*, Cambridge University Press, 2016.
- [26] H. Yang, T.L. Marzetta, Performance of conjugate and zero-forcing beamforming in large-scale antenna systems, *IEEE J. Sel. Areas Commun.* 31 (2) (2013) 172–179, <http://dx.doi.org/10.1109/JSAC.2013.130206>.
- [27] S. Boyd, L. Vandenberghe, Numerical linear algebra background, [on Line] - Available: <https://Web.Stanford.Edu/Class/Ee364a/Lectures/Num-Lin-Alg.Pdf>.
- [28] J.L. Jacob, T. Abrão, Nonorthogonal multiple access systems optimization to ensure maximum fairness to users, *Trans. Emerg. Telecommun. Technol.* 31 (4) (2020) 1–14.



Thiago Augusto Bruza Alves received the M.Sc. degree in Electrical Engineering at Universidade Estadual de Londrina (UEL), Paraná, Brazil, in 2017. Now he is student in electrical engineering at UEL/UTFPR and an assistant professor at Department of engineering Centro Universitario Filadélfia (UNIFIL), Londrina, Brazil. His research interest includes, machine learning and optimization problem applied to telecommunications, especially resource allocation for massive MIMO and NOMA systems.



Taufik Abrão (SM'12, SM-SBrT) received the Ph.D. degree in electrical engineering from the Polytechnic School of the University of São Paulo, São Paulo, Brazil, in 2001. Since March 1997, he has been with the Department of Electrical Engineering, Londrina State University, Paraná, Brazil, where he is currently an Associate Professor in Telecommunications and the Head of the Telecomm. & Signal Processing Lab. He is a Productivity Researcher from the CNPq Brazilian Agency. From July–October 2018 he was with the Connectivity section, Aalborg University as a Guest Researcher. In 2012, he was an Academic Visitor with the Southampton Wireless Research Group, University of Southampton, Southampton, U.K. He has served as Associate Editor for the *IEEE Systems Journal* (2020), the *IEEE Access* (2016–2018), *IEEE Communication Surveys & Tutorials* (2013–2017), the *AEU-Elsevier* (2020), the *IET Signal Processing* (2019), and *JCIS-SBrT* (2018–2020). Prof. Abrão has been served as Executive Editor of the *ETT-Wiley* journal since 2016. His current research interests include communications and signal processing, massive MIMO, XL-MIMO, URLLC, mMTC, optimization methods, random access protocols, detection and estimation, multicarrier systems, cooperative communication and relaying, resource allocation.

ANEXO B – On the SE and EE in Massive MIMO and NOMA

On the SE and EE in Massive MIMO and NOMA

Thiago Augusto Bruza Alves
Department of Electrical Engineering
State University of Londrina (UEL)
Londrina-PR, Brazil
Email: thiagobruza@hotmail.com

Taufik Abrão
Department of Electrical Engineering
State University of Londrina (UEL)
Londrina-PR, Brazil
Email: taufik@uel.br

Abstract—A comparative analysis on the Spectral Efficiency (SE) and Energy Efficiency (EE) in Downlink (DL) single-cell massive Multiple-Input Multiple-Output (mMIMO) and Non-Orthogonal Multiple Access (NOMA) considering a Base Station (BS) equipped with M antennas, while K devices operates with a single-antenna, and devices loading $\rho = \frac{K}{M}$ ranging in $0 < \rho \leq 2$ is carried out. In NOMA system, devices are overlapped in the power-domain; hence, NOMA is able to serve a greater number of devices. Under certain devices loading conditions, the NOMA overcame the mMIMO in terms of SE-EE trade-off, while continuing operate conveniently in the extended range $0.6 < \rho < 2.0$. We demonstrate that NOMA system is capable to operate suitably under higher system loading. The proposed analysis of the areas under SE and EE curves reveals that NOMA achieves highest area results.

Index Terms—Non-Orthogonal Multiple Access (NOMA); massive Multiple-Input Multiple-Output (mMIMO); Energy Efficiency (EE); Spectral Efficiency (SE).

I. INTRODUCTION

The beyond Fifth Generation (5G) of wireless communication systems must allows ultra-dense connections with vastly heterogeneous requirements. Among the proposals to solve this challenge, the massive Multiple-Input Multiple-Output (mMIMO) system is the main proposed system that allows the increase of the link capacity, exploring the propagation of multiple paths with the use of a large number of antennas at the Base Station (BS) [1], [2].

Another proposal technology is the power-domain *non-orthogonal multiple access* (NOMA), a good choice of multiple access technology, helping to mitigate the spectrum exhaustion problem and serving more than one device by resource block [3].

Studies analyzing NOMA and mMIMO in single-cell scenarios are proposed in [4]–[6]. Due to its unique characteristics, NOMA is a natural candidate technology to operate in dense wireless networks [6]. The use of NOMA in BS equipped with a large number of antennas is investigated in terms of SE in [5]; under a similar system configuration, herein, we analyse the networks increasing the loading up to 2 times the number of antennas in BS, evaluating the corresponding areas under the SE and EE curves, as well as the SE-EE trade-off.

Recent publications investigate enabling the NOMA application in different dense network scenarios; e.g., [7] proposes to apply NOMA with massive MIMO to a two tier network for accommodating huge traffic. Besides, a framework using NOMA in *distributed antenna systems*

was recently proposed in [8], while [9] addresses variant NOMA techniques and challenges beyond 5G.

This paper focus on the EE-SE performance comparison of the mMIMO vs. multiple antennas NOMA system varying the devices loading, i.e., the ratio of the number of devices to the number of BS antennas, $\rho = \frac{K}{M}$. More specifically, we adopt the areas under the SE and EE curves to quantify the SE-EE trade-off, aiming at providing a suitable and fair comparison analysis of both systems.

The *contributions* of this work are twofold: **a)** analyses on the SE-EE trade-off, by determining the region of loading devices trade-off for each system. **b)** making use of the area under the curves of system SE ($\mathcal{S}^{\text{SYST}}$) and EE ($\mathcal{E}^{\text{SYST}}$), we compare the systems' performance under three power allocation methods.

II. SYSTEM MODELS

Let us consider a multi-user single-cell Downlink (DL) transmission operating in a Time Division Duplex (TDD) with K single-antenna actives devices, communicating with one BS, which is equipped with M transmit antennas in Non-Line-Of-Sight (NLOS). The set $\mathcal{K}$ is formed by K devices, this devices are randomly distributed in a radius disk $d_{\max}$, the disk area is formed by two sub-disk with the same number of devices in each sub-area; both subsets are identified as $\mathcal{K}_H$ and $\mathcal{K}_L$. The indexes $k \in \mathcal{K}$ such that:

$$\mathcal{K} = \mathcal{K}_H \cup \mathcal{K}_L, \quad \text{where} \quad (1)$$

$$\mathcal{K}_H = \{1, \dots, K/2\} \text{ and } \mathcal{K}_L = \{K/2 + 1, \dots, K\}.$$

The channel vector modeling of device k can be described liked as:

$$\mathbf{h}_k = \sqrt{\beta_k} \mathbf{h}'_k, \quad k = 1, \dots, K, \quad (2)$$

where β_k is the large-scale fading coefficient and satisfy

$$\beta_j > \beta_i, \quad \forall j \in \mathcal{K}_H, \quad \forall i \in \mathcal{K}_L. \quad (3)$$

Herein, the pathloss model in [dB] is defined as:

$$\beta_k = \beta_0 + 10 \cdot \xi \cdot \log_{10}(d_k), \quad (4)$$

where d_k is the distance of user k to BS, ξ is the pathloss coefficient, and β_0 is the attenuation at the distance of reference.

In each coherence interval, $\mathbf{h}'_k$ in (2) for device k is an independent random small-scale fading realization from an independent Rayleigh fading distribution, $\mathbf{h}'_k \sim$

$\mathcal{CN}(0, \mathbf{I}_M), k = 1, \dots, K$. The transmitted signal $\mathbf{x}_k \in \mathbb{C}^M$ is the beamformed data symbol of device k :

$$\mathbf{x}_k = \mathbf{g}_k \sqrt{p_k} s_k, \quad (5)$$

where $\mathbf{g}_k$ is a normalized beamforming vector, p_k normalized transmission power and $s_k \sim \mathcal{CN}(0, 1)$ is the data symbol of device k , and period T_s . The signal received at the k -th device:

$$y_k = \sqrt{\beta_k p_k} \mathbf{h}'_k \mathbf{g}_k s_k + \sqrt{\beta_k} \mathbf{h}'_k \sum_{\substack{k'=1 \\ k' \neq k}}^K \mathbf{g}_{k'} \sqrt{p_{k'}} s_{k'} + n_k, \quad (6)$$

where $n_k \sim \mathcal{CN}(0, 1)$ is the additive noise. Notice that so far, this modelling applies to both NOMA and mMIMO systems, but beamforming are selected differently and this topic will be addressed in the next sections.

III. SE-EE IN NOMA AND mMIMO SYSTEMS

In this section we define the SE and the EE for both NOMA and mMIMO systems. In the mMIMO system, each device is served by a single beamforming vector; while in the NOMA, two devices are grouped in a cluster k and served by the same beamforming, sharing the same channel $\mathbf{g}_k = \mathbf{g}_{K/2+1}$ for all $k = 1, \dots, K/2$, and using the interference cancellation process to detect the weaker user in the k -th cluster.

A. Data Rates in NOMA with ZF

Devices are divided in two sets like described in Eq. (1) and these groups are represented by Eq. (7) by their large-scale fading coefficient and are grouped in pairs forming a cluster as Fig. 1, the cluster k is formed by one device in cell center set $\mathcal{K}_H$ and one device in cell edge set $\mathcal{K}_L$. Hence, the devices are grouped in two subsets:

$$\mathcal{K}_H = \{\beta_1 > \beta_2 > \dots > \beta_{K/2}\}, \quad (\text{center devices set}) \quad (7)$$

$$\mathcal{K}_L = \{\beta_K < \beta_{K-1} < \dots < \beta_{K/2+1}\}. \quad (\text{edge devices set})$$

The pairing rule adopted in (7) creates the largest possible difference in channel coefficients for devices not yet paired.

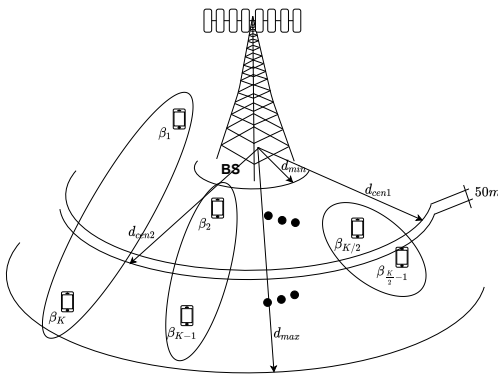


Figure 1: System Model indicating the Pairing formation in NOMA system.

The instantaneous Signal to Interference plus Noise Ratio (SINR) of devices in cluster k is defined as:

$$\text{SINR}_k = \frac{\beta_k p_k |\mathbf{h}'_k \mathbf{g}_k|^2}{\beta_k \sum_{k' \neq k}^K p_{k'} |\mathbf{h}'_k \mathbf{g}_{k'}|^2 + 1}. \quad (8)$$

In each cluster, the cell-edge devices treat the interference as noise and decodes its own data symbols, whereas the cell-center device can decode the data symbols of the cell-edge device and perform Successive Interference Cancellation (SIC), hence effectively removing the interference due to the cell-edge device.

In order to perform SIC, the cell-center device needs to be able to decode data signal intended for the cell-edge device, hence, in view of the uplink (mMIMO and NOMA) and downlink (NOMA) pilot overhead and assuming perfect {Channel State Information (CSI) in all receivers, this condition must be satisfied [5]:

$$\mathbb{E}[\text{SINR}_{k, K/2+k}] \geq \mathbb{E}[\text{SINR}_k], \quad (9)$$

where

$$\text{SINR}_{k, K/2+k} = \frac{\beta_k p_{K/2+k} |\mathbf{h}'_k \mathbf{g}_{K/2+k}|^2}{\beta_k \sum_{k' \neq K/2+k}^K p_{k'} |\mathbf{h}'_k \mathbf{g}_{k'}|^2 + 1}. \quad (10)$$

Herein, the condition in (9) must be satisfied by selecting the transmit powers appropriately. The achievable ergodic rate of devices in cluster k , i.e. device k in $\mathcal{K}_H$ set and k in $\mathcal{K}_L$ set, in this work the CSI is considered perfect and the Zero Forcing (ZF) eliminates all inter-clusters interference, the k -th cluster NOMA achievable rate, measured in [bits/s/Hz]:

$$\begin{aligned} R_{\text{cl-}k}^{\text{NOMA}} &= \tau \mathbb{E} [\log_2 (1 + \bar{M} \beta_k p_k)] \\ &+ \tau \mathbb{E} \left[\log_2 \left(1 + \frac{\beta_{K/2+k} p_{K/2+k}}{\beta_{K/2+k} p_k + 1} \right) \right]. \end{aligned} \quad (11)$$

$\forall k \in \mathcal{K}$ and $\bar{M} = M + 1 - K/2$ and $\tau = (1 - \frac{K}{T})$ is the portion of each channel coherence interval (T) that is used for data transmission, like proposed in [5] because of the need for a pilot overhead. Hence, the NOMA system can operate until $K < 2M - 1$.

B. Data Rates in mMIMO with ZF

In the mMIMO system with ZF precoding the ergodic achievable rate for device k is given by:

$$R_k^{\text{M-MIMO}} = \tau \mathbb{E} [\log_2 (1 + (M - K) p_k \beta_k)]. \quad (12)$$

Under linear precoding and combiners, the mMIMO system operates consistently when $K < M$. Finally, the average system sum-rate (avg-sum-rate) is defined simply by:

$$\mathcal{R}^{\text{M-MIMO}} = \sum_{k=1}^K R_k^{\text{M-MIMO}} \quad \text{and} \quad \mathcal{R}^{\text{NOMA}} = \sum_{k=1}^{\mathcal{K}_H} R_{\text{cl-}k}^{\text{NOMA}}.$$

C. Energy Efficiency

The Energy Efficiency (EE) metric is defined as the ratio of the number of effective information bits received over the total energy consumed by the system to transmit that information, as:

$$EE = \frac{\sum_{k=1}^K R_k}{P_{\text{TOT}}} \quad [\text{bits/Joule/Hz}]. \quad (13)$$

Based on the power model proposed in [2] the total power consumption is composed of fixed term and terms scaled with the number of antennas M and number of devices K :

$$P_{\text{TOT}} = \frac{P_{\text{RF}}}{\varpi} + P_{\text{FIXED}} + P_K + P_M + P_{KM} \quad [\text{W}]. \quad (14)$$

where $P_{\text{FIXED}} = P_0 + P_{\text{SYN}}$, $P_K = K(P_{\text{COD}} + P_{\text{DEC}} + P_{\text{RX}}) + K^3 \frac{2}{3LT}$, $P_M = M \cdot P_{\text{TX}}$ and $P_{KM} = MK \frac{3+T}{TL} + MK^2 \frac{2}{TL}$.

D. Power Allocation Strategies

In the sequel, we present three strategies for the Power Allocations (PA), aiming to increase the system sum-rate.

1) *Equal Power Allocation (EPA)*: all devices are served with same power. The EPA strategy applied to mMIMO can be defined by:

$$p_k = \frac{P_{\text{RF}}}{K} \quad [\text{W}], \quad \forall k \in \mathcal{K}, \quad (15)$$

EPA to NOMA, it is composed by two steps: in the first, power reference to each cluster can be defined simply as:

$$p_{\text{ref}}^{\text{cl}} = \frac{2 \cdot P_{\text{RF}}}{K} \quad [\text{W}], \quad (16)$$

next the PA among the devices in the same cluster is defined as:

$$p_k^{\text{cl-k}} = p_{K/2+k}^{\text{cl-k}} = \frac{p_{\text{ref}}^{\text{cl}}}{2} \quad [\text{W}]. \quad (17)$$

2) *Classical Water-Filling (WF) Algorithm*: The application of the WF algorithm in mMIMO system results in optimal (maximum) system sum-rate solution.

3) *Δ -WF for NOMA*: We suggest modify the classical WF procedure to be applied to NOMA accordingly, avoiding the breaking of formed pairs, keeping NOMA power differences at least as equal to Δ . The steps of Δ -WF algorithm are described as follows. *Step 1*: the PA in NOMA occurs initially among the clusters. Hence, we propose the application of WF based on the large-scale fading differences $\Delta\beta_k = \beta_k - \beta_{K/2+k}$ of devices inside each $\mathcal{K}_L$ and $\mathcal{K}_H$ subsets, defined in (7). *Step 2*: the power is allocated between the devices inside the group, hence Eq. (9) must be satisfied. Eq. (18) and Algorithm

1 summarize the proposed Δ -WF PA procedure, aiming to improve the SE of NOMA systems.

$$\mu = \frac{2}{\mathcal{K}_H} \left(P_{\text{RF}} + \sum_{k=1, k \in \mathcal{K}_H}^{|\mathcal{K}_H|} \frac{1}{(\beta_k - \beta_{K/2+k})} \right); \quad (18)$$

$$P_{\text{RF}} = \sum_{k=1, k \in \mathcal{K}_H}^{|\mathcal{K}_H|} p_{\text{cl-k}}, \quad \forall k \in \mathcal{K}_H,$$

$$\text{where } p_{\text{cl-k}} = \left(\mu - \frac{1}{(\beta_k - \beta_{K/2+k})} \right)^+$$

$$\text{and } \mathbf{p}_{\text{cl-k}} = [p_1, p_2, \dots, p_{|\mathcal{K}_H|}].$$

Notice that in the Step 2, the power allocation for both devices belong the k -th cluster is simply defined as:

$$p_{K/2+k}^{\text{cl-k}} = p_k^{\text{cl-k}} = \frac{p_{\text{cl-k}}}{2} \quad [\text{W}]. \quad (19)$$

Algorithm 1: Δ -WF (modified) for NOMA systems

Input: $\mathcal{K}_H, \mathcal{K}_L, P_{\text{RF}}$

```

1 NP  $\leftarrow \emptyset$ ;
2 while (NP  $\neq \emptyset$ ) do
3   solve Eq. (18)  $\rightarrow \mathbf{p}_{\text{cl-k}}$ ;
4   NP  $\leftarrow$  identify null position in  $\mathbf{p}_{\text{cl-k}}$ ;
5    $\mathcal{K}_H/\{k\}_{\text{NP}} \leftarrow$  exclude from  $\mathbf{p}_{\text{cl-k}}$  devices
   labeled as NP.
6 end
```

Output: $\mathbf{p}_{\text{cl-k}} = [p_1, p_2, \dots, p_{|\mathcal{K}_H|}]$

IV. NUMERICAL RESULTS

The numerical evaluations for the proposed analyses of NOMA vs mMIMO systems are presented in this section. The simulation system channel parameter and power consumption values deployed along this section are depicted in Table I. The scenario has simulated in Matlab 2019 utilizing one Intel HD Graphics 6000 GPU, Intel(R) Dual-Core(TM) i5 CPU @ 1.6 GHz and 8 GB RAM.

A. SE-EE Trade-off (Resource Efficiency - RE)

The NOMA and mMIMO are analyzed in terms of SE-EE trade-off, namely *resource efficiency* (RE), considering device loading increasing up to 2. From Fig. 2, shows the best results, one can find graphically the best device loading that maximizes the SE-EE trade-off for each BS antenna configuration (M). The left y-axis depicts avg-SE and the right y-axis shows the avg-EE. (a) when $M = 64$, NOMA with EPA achieved SE-EE trade-off with the highest device loading $\rho = 0.875$. (b) $M = 128$, NOMA with EPA achieved SE-EE trade-off with the highest device loading $\rho = 0.734$. (c) $M=256$, NOMA with Δ -WF achieved SE-EE trade-off with the highest device loading $\rho = 0.594$.

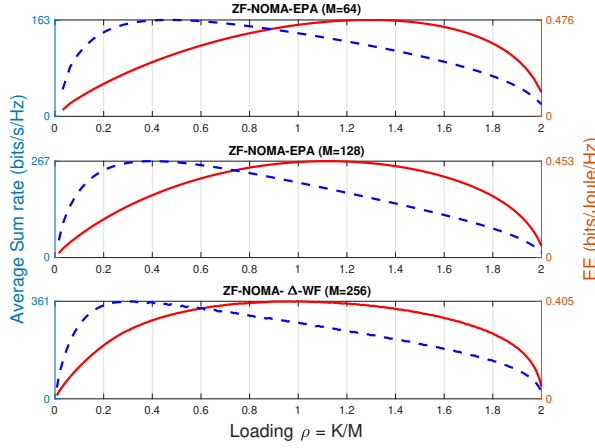


Figure 2: Best results in trade-off SE-EE in each M antennas in BS.

Table I:
Simulation Parameters

Parameter	Value
Pairs NOMA / Clusters	$N = K/\zeta = K/2$
NOMA devices per cluster	$\zeta = 2$
Strong device position	$[d_{\min}; d_1] \in [50; 100]$ m
Weak device position	$[d_2; d_{\max}] \in [150; 350]$ m
Data symbol period	T_s
Coherence time interval	$T = 512 \cdot T_s, \quad \ell = 512$
Parameters values for EE analysis	
Backhaul Infrastructure	$P_0 = 2$ W
Total RF power available	$P_{\text{RF}} = 1$ W
Single oscillator	$P_{\text{syn}} = 2$ W
Coding and modulation	$P_{\text{COD}} = 4$ W per device
Decoding and demodulation	$P_{\text{DEC}} = 0.5$ W per device
Receive power	$P_{\text{RX}} = 0.3$ W per device
Transmitted power	$P_{\text{TX}} = 1$ W per antenna
Efficiency of Power Amplifier	$\varpi = 0.3$
Operations/Joule	$L = 10^9$ oper. per joule
Channel	
Pathloss exponent	$\xi = 3.78$
Attenuation at a d_0 reference	$\beta_0 = 130$ [dB]
# Monte-Carlo realizations	1000

B. Area Under Curves SE and EE

Let us consider the corresponding areas under the SE and EE curves in Fig. 2, such that:

$$\mathcal{S}^{\text{SYST}} = \frac{1}{M} \cdot \int_0^{\rho=2} \overline{\text{SE}}(\rho) d\rho \quad \left[\frac{\text{bits/antenna}}{\text{s} \cdot \text{Hz}} \right]$$

and

$$\mathcal{E}^{\text{SYST}} = \frac{1}{M} \cdot \int_0^{\rho=2} \overline{\text{EE}}(\rho) d\rho \quad \left[\frac{\text{bits/antenna}}{\text{Joule} \cdot \text{Hz}} \right],$$

respectively, where $\overline{\text{SE}}(\rho)$ is the average overall system sum-rate, and $\overline{\text{EE}}(\rho)$ is the average overall system energy efficiency achieved under specific device loading ρ .

Fig. 3 reveals the values of such integrals considering the range of average SE and EE along the device loading ($K = 64, 128$, and 256), and normalized per BS antenna, attainable by NOMA and mMIMO systems. Hence, comparing such values of corresponding areas under the SE and EE curves (from Fig. 2), one can infer that considering EPA policy, in average, we have roughly:

$$\mathcal{S}_{M=64}^{\text{NOMA}} \approx 2.7 \cdot \mathcal{S}_{M=64}^{\text{mMIMO}}, \quad \text{and} \quad \mathcal{E}_{M=256}^{\text{NOMA}} \approx 1.8 \cdot \mathcal{E}_{M=256}^{\text{mMIMO}},$$

i.e., NOMA is more spectrally and energetically efficient than mMIMO over a wide range of device loading $0 < \rho \leq 2$, roughly 170% more efficient in terms of SE, and roughly 80% w.r.t. EE.

V. CONCLUSION AND REMARKS

This work proposes a comparative SE and EE analysis in DL single-cell between mMIMO and NOMA with BS equipped with different antennas configurations. Regards to PA methods applied to NOMA system, the EPA and the proposed Δ -WF results different SE performance. Indeed, the EPA strategy achieved marginally best SE for all M .

Under the EE perspective, the mMIMO achieved better results when employed the EPA and WF PA methods under $K < M$. However, the NOMA can operate under higher system loading, i.e., $K < 2M - 1$.

From the perspective of SE-EE trade-off, the NOMA attains superior trade-offs for all set of system configuration analysed, achieving the highest device loading to all M setups. Finally, NOMA achieved the highest ratio $\frac{\mathcal{S}^{\text{NOMA}}}{\mathcal{S}^{\text{mMIMO}}} \approx 2.75$ when $M = 64$ and for the EE analysis when $M = 256$, attaining $\frac{\mathcal{E}^{\text{NOMA}}}{\mathcal{E}^{\text{mMIMO}}} \approx 1.8$. These results corroborate the NOMA's ability to operate suitably under high loading systems, $1 < \rho < 2$.

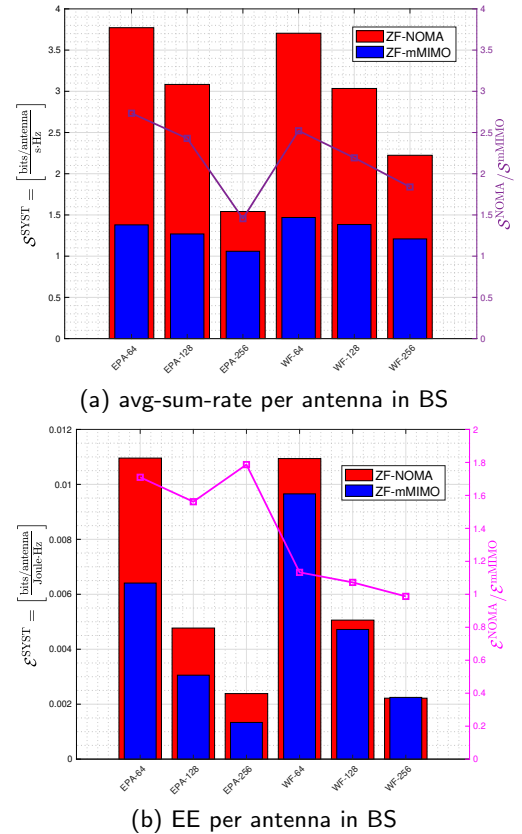


Figure 3: Area under SE and EE curves per antenna in BS.

REFERENCES

- [1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

- [2] E. Bjornson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?," *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 242–247, apr 2014.
- [3] Z. Ding and H. V. Poor, "Design of Massive-MIMO-NOMA with Limited Feedback," *IEEE Signal Processing Letters*, vol. 23, pp. 629–633, may 2016.
- [4] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid Precoding-Based Millimeter-Wave Massive MIMO-NOMA With Simultaneous Wireless Information and Power Transfer," *IEEE Journal on Selected Areas in Communications*, vol. 37, pp. 131–141, jan 2019.
- [5] K. Senel, H. V. Cheng, E. Bjornson, and E. G. Larsson, "What Role Can NOMA Play in Massive MIMO?," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4553, no. c, pp. 1–16, 2019.
- [6] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A Survey of NOMA: Current Status and Open Research Challenges," *IEEE Open Journal of the Communications Society*, vol. 1, no. December 2019, pp. 179–189, 2020.
- [7] S. Rajoria, A. Trivedi, and W. W. Godfrey, "Sum-rate optimization for NOMA based two-tier hetnets with massive MIMO enabled wireless backhauling," *AEU - International Journal of Electronics and Communications*, vol. 132, no. October 2020, p. 153626, 2021.
- [8] D. Kim and M. Choi, "Non-Orthogonal Multiple Access in Distributed Antenna Systems for Max-Min Fairness and Max-Sum-Rate," *IEEE Access*, vol. 9, pp. 69467–69480, 2021.
- [9] I. Budhiraja, N. Kumar, S. Tyagi, S. Tanwar, Z. Han, D. Y. Suh, and M. J. Piran, *A Systematic Review on NOMA Variants for 5G and Beyond*, vol. 9. IEEE, 2021.

ANEXO C – NOMA-Based Random Access in mMTC XL-MIMO

RESEARCH ARTICLE

NOMA-Based Random Access in mMTC XL-MIMO

THIAGO AUGUSTO BRUZA ALVES¹ AND **TAUFIK ABRÃO²**, (Senior Member, IEEE)

Department of Electrical Engineering, State University of Londrina (UEL), Londrina 86057-970, Brazil

Corresponding author: Thiago Augusto Bruza Alves (thiagobruza@hotmail.com)

This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) under Grant 001, and in part by the National Council for Scientific and Technological Development (CNPq) of Brazil under Grant 310681/2019-7.

ABSTRACT In machine-type communication (MTC), massive access attempts are generated and the massive MIMO is the key technology to support this demand. To support massive MTC (mMTC), the recent extra-large scale massive multiple-input multiple-output (XL-MIMO) architecture has been seen as a promising technology for providing very high-data rates in high-user density scenarios. Therefore, the large dimension is of the same order as the distances to the user equipment (UE) causing spatial non-stationarities and visibility regions (VRs) to occur across the huge XL array extension. In this work, we investigate the random access (RA) problem in crowded mMTC XL-MIMO scenarios; the proposed grant-based random access (GB-RA) protocol combining the advantage of non-orthogonal multiple access (NOMA) and the strongest user collision resolution in extra-large arrays (SUCRe-XL), namely NOVR-XL scheme can allow access of multiple colliding users in the same XL subarray (SA) selecting the same pilot sequence. The proposed NOVR-XL GB-RA protocol is able to provide a reduction in the number of attempts to access the network, while improving the average sum-rate, as the number of SA increases.

INDEX TERMS NOMA, machine type communication, XL-MIMO, random access protocols, crowded scenarios, uniform rectangular array (URA).

I. INTRODUCTION

There are a large number of users/devices in massive machine-type communications (mMTC) intermittently transmitting modest quantities of data to the base station (BS). This means that the number of active users at any given moment is often substantially lower than the overall number of system users. Simple internet-of-things (IoT) sensors are an example of mMTC devices. Serving those massive connections while guaranteeing a quality of service (QoS) requirement is the prominent challenge in the next generation of multiple access protocols and systems.

The new use case and service classes in 5G systems require significant improvements in physical (PHY) and medium access control (MAC) layers to overcome such challenges. At the PHY layer, the non-orthogonal multiple access (NOMA) transmission schemes is a known and promising candidate for the next generations of multiple access systems, due to its ability to support multiple devices in the

same resource element (RE). When compared to the traditional orthogonal multiple access (OMA), the goal consists in attaining higher system throughput and reliability, improving the users' fairness, and reducing latency, while supporting massive connections [1], some types of NOMA schemes with preminent results are code-domain NOMA, artificial-intelligence NOMA, and power-domain NOMA.

Code-domain NOMA which multiplexing signals in the code-domain was recently explored in [2] to provide connections in overcrowded regime. Artificial-intelligence NOMA in [3] has applied computational learning techniques, such as deep learning in typical NOMA wireless communication systems. Power-domain NOMA is a transmitting structure, proposed initially to improve the spectral efficiency (SE) of wireless networks, sharing the same orthogonal resource, in time and frequency, by combining superposition coding in the transmitter side, and successive interference cancellation (SIC) in the receiver side [4]. In this work we deploy the power-domain NOMA.

A vast number of devices trying to access the network at the same time is a MAC challenge, traditional random

The associate editor coordinating the review of this manuscript and approving it for publication was Walid Al-Hussaini³.

access (RA) schemes are not able to handle such a large number of requests [5], pure RA schemes like ALOHA have severe performance limitation. The incorporation of NOMA significantly improves performance, admitting two users or more per time slot [6].

In massive scenarios, the number of user connections attempting the access the same resource block (RB) far outnumbers the number of available pilot sequences. As a consequence, the implementation of collision resolution protocols becomes essential to allow coherent communication. One well-known decentralized grant-based (RA) protocol for crowded massive multiple-input multiple-output (MIMO) systems is the strongest user collision resolution (SUCRe), which takes advantage of MIMO properties [7], giving preference to the users with good channel conditions, harming edge users.

Seeking to improve RA schemes, the SUCRe protocol [7] has undergone several analyses and improvements. For instance, in [8] a pilot collision reduction strategy was implemented, when compared to [7], making the failed user re-select their pilots by means of an access class barring (ACB) procedure, named idle pilot access (SUCR-IPA) protocol. Another extension is proposed in [9], where a graph-based pilot access (SUCR-GBPA) protocol is proposed, enabling all users who lost contention resolution to choose a new pilot at random. Furthermore, combining NOMA and SUCRe was an important evolution; the work in [10] compares the proposed NOMA-RA protocol to the classical SUCRe scheme, allowing two users trying to access the medium with the same pilot signal to resolve collisions by distinguishing them in the power-domain NOMA and attaining a superior sum-rate performance with reduced average latency.

The recent extra-large scale massive MIMO (XL-MIMO) technology distributes a huge number of antennas elements in a wide space, like facades of buildings, shopping malls, or stadiums; the distance between user equipments (UEs) and the BS antennas might become of the same order of the dimensions of the array [11]. Such arrangement configured special channel conditions, especially unusual channel conditions, such as spatial non-stationarities and *visibility regions* (VRs) [12].

Recently, XL-MIMO systems have received a lot of research interest. Recent papers analyse RA improvements in linear receiver schemes [13], [14], [15], appropriate channel models [12], channel estimation algorithms [16], non-linear receiver schemes [17], and efficient RA schemes in XL-MIMO [18], [19], [20].

In [13], a revamping in SUCRe protocol enabled its implementation in an XL-MIMO scenarios; the concept of non-overlapping VR was explored in the XL-MIMO context, by associating geographic area and the portion of the array that is visible from that area. Hence, considering that a certain UE is present in that region, the corresponding VR and the associated set of clusters become accessible. Hence, the probability of SA b being visible by UE k can be explored to improve the RA protocol performance in XL-MIMO.

In this work we propose a NOMA grant-based RA protocol for mMTC XL-MIMO scenarios by admitting users colliding, *i.e.*, more than one user equipment (UE) selecting the same pilot sequence, and transmitting simultaneously to the same XL subarray, by exploring the advantage of power-domain NOMA.

This work is an extension of our previous work [21]. The *Contribution* of this work is threefold:

- i. we propose a grant-based RA protocol for XL-MIMO systems equipped with power-domain NOMA and by exploring the overlapping VR of XL-MIMO systems (NOVR-XL), combining the advantages of XL-MIMO systems and NOMA, allowing multiple users selecting the same pilot sequence in at least one SA.
- ii. we develop a numerical optimization procedure for the *scale factor* $\delta_{\text{NOVR-XL}}$ in order to maximize the sum-rate of the system; by exploring the NOMA-SIC strategy allows multiple users selecting the same pilot sequence, in a case of possible *false positives*, be resolved converting them into successful attempts, improving the system performance metrics. Besides, aiming to expeditiously maximize the system sum-rate, we propose a look-at-table reference for different loading users and scenarios.
- iii. extensive numerical analyses on the superimposed received signal processed on the SA basis, considering the optimal scale factor $\delta_{\text{NOVR-XL}}^*$ and performing SIC to identify the signal of each user.

The rest of this paper is organized as follows. Section II is described the system model, including the array/subarray integration, while the proposed NOVR-XL is described in Section III. Numerical results are discussed in Section VI. The conclusions are presented in Section V.

A. NOTATIONS

Boldface low-case $\mathbf{a}$ and boldface capital letter $\mathbf{A}$ letters represent, respectively, vectors and matrices. Capital calligraphic letters $\mathcal{A}$ represent finite sets, and $|\mathcal{A}|$ denotes the cardinality. Besides, the operator $\mathcal{A} \subset \mathcal{D}$ denotes the set $\mathcal{A}$ is a subset of $\mathcal{D}$ and $\mathcal{C} = \mathcal{A} \setminus \mathcal{D}$ denotes the set $\mathcal{C}$ is composed by the difference of the set $\mathcal{A}$ and the set $\mathcal{D}$.

The transpose, conjugate-transpose, and conjugate of a matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^H$, and $\mathbf{A}^*$, respectively. We let $\mathbf{I}_M$ denote the $M \times M$ identity matrix, whereas $\|\cdot\|$ stands for the Euclidean norm of a vector. We use $\mathcal{CN}(\mu, \sigma^2)$ to denote circularly-symmetric complex Gaussian distribution with mean μ and variance σ^2 , $\mathcal{B}(\cdot, \cdot)$ to denote a binomial distribution, and $\mathbb{C}$ to denote spaces of complex-valued. The Gamma function is denoted by $\Gamma(\cdot)$.

II. SYSTEM MODEL

We consider an XL-MIMO setup where the BS deploys an extra-large antenna array, operating in a time-division-duplexing (TDD). Without loss of generality, we assume a uniform rectangular array (URA) is installed on the y - z plane with a total of M antenna-elements, *i.e.*, $M_y \times M_z$, where

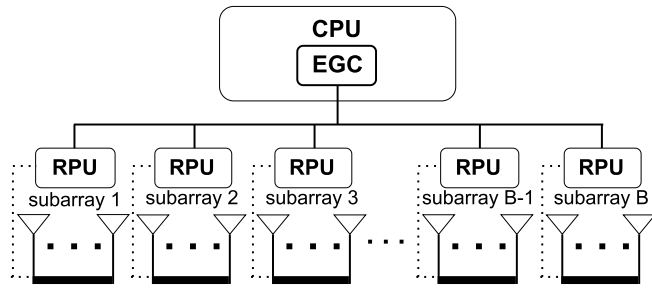


FIGURE 1. Diagram of BS architecture. The antenna array at the BS is composed by B SA containing a fixed number of antenna elements M_b . Each SA has an RPU, and all are connected to the EGC in the CPU.

M_y and M_z denote the number of antenna elements along the y -axis and z -axis, respectively, with the first line of the array starting at the origin O and deployed along the ordinate axis, as in Fig. 3. Similarly, as [13] and [15], we consider a simplified bipartite graph model in XL-MIMO, the array is divided into B subarrays (SAs), each composed by a fixed number of $M_b = M/B$ antennas, ensuring the minimum antennas elements ($M_b \geq 50$), to achieve the inherent properties of massive MIMO [22]. Each SA is equipped with a remote processing unit (RPU) that performs channel estimate, SIC, and resource allocation in a distributed manner, these RPUs are connected to a coordinating central processing unit (CPU), and decoded signals in RPU are sent to CPU to be combined in an equal gain combiner (EGC), where each branch has the same weight, integrated to the BS. All elements are visualized in Fig. 1.

All single-antenna user devices are randomly distributed and grouped in set $\mathcal{U}$, $\mathcal{A} \subset \mathcal{U}$ be the subset of active UEs, with temporarily dedicated data pilots, and $\mathcal{K} = \mathcal{U} \setminus \mathcal{A}$ be the set of inactive UEs (iUEs). The devices in set $\mathcal{K}$ do not have dedicated pilots and if they want to be active, they must be assigned to one. Hence, the number of iUEs in the cell is $|\mathcal{K}| = K$. The BS only allocates pilots to active devices and reclaims the pilots when needed. The iUEs make a RA attempt with probability P_a .

Let $\mathcal{M}$ denote the set of all BS SAs and $\mathcal{V}_k \subset \mathcal{M}$ the subset of SAs visible to the k -th UE. To compose the mathematical modeling, we assume that the subset $\mathcal{V}_k$ is generated at random, where each SA is visible with probability P_b by UE k , like in [13] and [15]. This probability simulates the influence of random obstacles and scatterers in the environment that interact with the signals transmitted by/to the UEs, resulting in VRs. The existence of VRs is evidenced by measurements in [11].

We consider a multiple-user communication systems based on time-frequency coherence blocks and inactive UEs have RA blocks to realize RA attempts. We define τ_{RA} as the number of RA mutually orthogonal pilot sequences that are possible $s_1, \dots, s_{\tau_{RA}} \in \mathbb{C}^{\tau_{RA} \times 1}$, each with length τ_{RA} , and $\|s_{\tau_{RA}}\|^2 = \sqrt{\tau_{RA}}$.

Since each SA visibility follows a Bernoulli distribution with success probability P_b , it is crucial to note that when

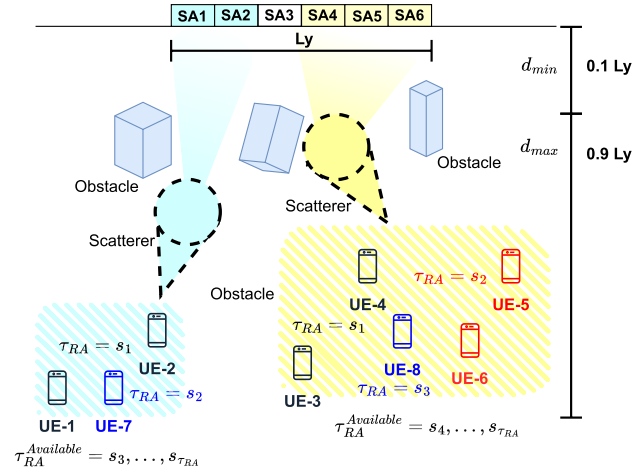


FIGURE 2. Massive MTC system with XL-MIMO URA composed by B -SAs containing M_b antennas, the colored regions in each SA define visible UEs, the visibility success probability $P_b = 0.5$ to each SA.

an antenna is visible to a user, it indicates that the user may transmit/receive signals to/from this antenna with a non-zero channel gain [15]. We assume an overcrowded scenario where K iUEs ($K \gg M$) are randomly distributed within a square cell with maximum distance $d_{\max}$ and minimum distance $d_{\min}$, respectively, as in Fig. 2.

Due to scatterers and obstacles, whose presence is statistically modeled by the probability of visibility P_b , Fig. 2 illustrates three UEs on the left side having an overlapping VR formed by SA1–SA2: users UE-1 and UE-2 colliding in the same pilot sequence s_1 , and UE-7 selects an exclusive pilot sequence s_2 , represented by the cyan color area; on the right side, five users having the same VR formed by the SA4–SA5–SA6 identified by the yellow color area; in that SAs, two collisions occur: in the first collision, the UE-3 and UE-4 colliding by sharing the pilot sequence s_1 , and UE-5 and UE-6 colliding with pilot sequence s_2 ; finally, UE-8 selects an exclusive pilot sequence s_3 .

In the next step of the proposed protocol NOVR-XL, the orthogonal message of UE-7 is decoded directly at each SA in VR, without interference; besides, the messages of UE-1 and UE-2 collide by sharing the same s_1 ; hence, such messages can be decoded in SA1 and SA2 by deploying a NOMA scheme, with an application of one SIC step at each SA. In the second SAs group (yellow area), the orthogonal message of UE-8 is decoded directly at each SA in the VR; also, the messages of UE-3 and UE-4 colliding by deploying the same pilot sequence s_1 , and the message of UE-5 and UE-6 colliding in the pilot sequence s_2 can be successfully decoded in SA4, SA5, and SA6 by using NOMA, by applying one SIC step. Finally, each decoded signal can be combined in the CPU via EGC combiner. It is important to notice that in the scenario of Fig. 2, SA3 is fully blocked, *i.e.* no UE has SA3 in its VR; hence, such SA3 can be turned off until the channel scenario changes, improving the overall system energy efficiency. Notice that each SA has a dedicated RPU, and the BS has a CPU with EGC and can coordinate the signal processing at each SA.

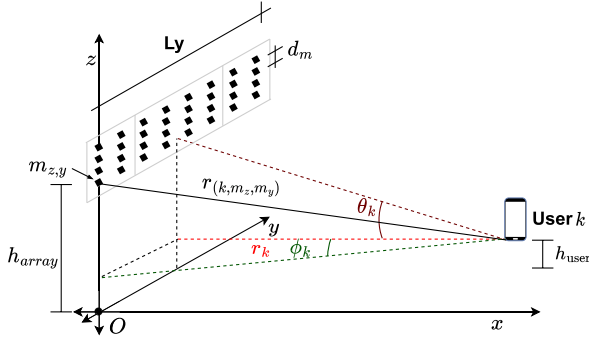


FIGURE 3. Diagram of the k th user at the URA XL-MIMO with an impinging sphere waveform from elevation angle θ and azimuth angle ϕ . The $M = M_y \times M_z$ are horizontally and vertically equally spaced with spacing d_m .

A. CHANNEL MODEL

In the next, we model the NLOS channel based on the position of each URA antenna element $m_{z,y}$, as in Fig 3. The specific localization of $m_{z,y}$ -th array element is represented by 3D vector $\mathbf{w}_{m_{z,y}} = [0, (m_y - 1)d_m, h_{array} + (m_z - 1)d_m]$, where $1 \leq m_y \leq M_y$, $1 \leq m_z \leq M_z$, and h_{array} is the height where the array is installed. The physical dimension of the URA along the y -axis and z -axis are $L_y = (M_y)d_m$ and $L_z = (M_z)d_m$, where d_m is the distance between elements. The user- k position is defined by 3D vector $\mathbf{q}_k$, the distance of the k -th UE to the specific antenna element $m_{z,y}$ is determined by $r_{k,m_y,m_z} = \|\mathbf{w}_{m_{z,y}} - \mathbf{q}_k\|$ [23].

The distance in a straight line perpendicular to the y -axis of the k -th UE of the wide space where the array is installed r_k , and the elevation angle θ_k , formed by r_k and the line between UE $_k$ and the height of array line (red dashed line), and the azimuth angle ϕ_k , formed by the line connecting UE- k to the aperture of $m_{x,y}$ position (green dashed line). Then the location of UE- k is $\mathbf{q}_k = [r_k \Psi_k, r_k \Phi_k, r_k \Omega_k]^T$ with $\Psi_k \triangleq \sin \theta_k \cos \phi_k$, $\Phi_k \triangleq \sin \theta_k \sin \phi_k$ and $\Omega_k \triangleq \cos \theta_k$, $1 \leq k \leq K$. The large-scale fading includes path loss and shadowing, on the basis of an urban micro scenario model provided by [24]:

$$\beta_{k,m_z,m_y} = 10^{-\kappa \log(r_{k,m_z,m_y}) + \frac{g_o + \varphi}{10}}, \quad (1)$$

where $g_o = -34.53$ dB is the path loss at the reference distance, the path loss exponent $\kappa = 3.8$, and $\varphi \sim \mathcal{N}(0, \sigma_{sf}^2)$ is the shadow fading, a log-normal random variable with standard deviation $\sigma_{sf} = 10$ dB.

Similarly as [13] and [15], we adopt the average large scale fading for the k -th user at the b -th SA, $\beta_k^{(b)} = \frac{1}{M_b} \sum_{m=1}^{M_b} \beta_{k,m}$. Let $\mathbf{h}_k^{(b)} \in \mathbb{C}^{M_b \times 1}$ be the Rayleigh fading channel vector between UE $k \in \mathcal{K}$ and SA b , following $\mathbf{h}_k^{(b)} \sim \mathcal{CN}(0, \beta_k^{(b)} I_{M_b})$.

III. NOVR-XL – RANDOM ACCESS FOR NOMA XL-MIMO

Initially, we explain the four Steps of the SUCRe-XL protocol as depicted in Fig. 4. The operations of the proposed protocol NOVR-XL are detailed in the sequel. Its operations can be

seen in Fig. 5. The main difference lies in obtaining and applying at each UE the optimal *scale factor* δ^* which is adjusted aiming to improve the system performance metrics.

A. SUCRe-XL PROTOCOL

The SUCRe-XL protocol [13] is a distributed method for resolving pilot contentions on the user side in XL-MIMO systems. The protocol functionality is illustrated in Fig. 4. Four steps define the SUCRe-XL protocol of [13]:

Step 1: KP_a users randomly choose an uplink (UL) RA pilot and transmit it;

Step 2: the BS receives such UE signals and responds with precoded downlink (DL) pilots;

Step 3: Using the received signal, UEs can estimate the sum of the channel gains of contending UEs for the same RA pilot, comparing the estimate with its own long-term channel gain (estimated in step 0) and retransmitting the chosen RA pilot if it judges it to be the strongest UE among the contenders;

Step 4: the BS allocates dedicated data pilots to the UEs transmitting in Step before. Users accepted to transmit must fit non-overlapping VRs condition.

After the SUCRe-XL steps, user k in the transmission stage with a dedicated data pilot attains the following SNR in each visible SA- b :

$$\text{SNR}_k^{(b)} = \frac{p_k \beta_k^{(b)}}{\sigma^2}, \quad (2)$$

in which p_k is the user' transmit power.

B. EXPLORING VRs, SubArrays, AND NOMA IN NOVR-XL PROTOCOL

The proposed NOVR-XL protocol deploys the non-overlapping VRs of specific UEs along the SAs aiming to improve the SUCRe-XL protocol [13] operation. Our proposed protocol modifies the *activation decision rule* proposed in the SUCRe-XL protocol [13]. As proposed in [10] we numerically optimize the *scale factor*, namely $\delta_{\text{NOVR-XL}}^{(K,B)*}$ within the bias parameter ϵ_k , aiming to maximize the system's sum-rate, making more than one user consider themselves able to retransmit the selected pilot (in Step 1); hence, increasing the number of possible *false positives*, i.e., multiple UEs appoint themselves as the contention winner. This is made in a distributed manner, without cooperation with other UEs. Such possible *false positives* can be solved by deploying a NOMA-based scheme, being able to process at the receiver side multiple users sharing the pilot in overlapping VRs, processing on a SA-basis.

Fig. 5 depicts the four steps of the proposed grant-based NOMA XL-MIMO random access protocol and highlights the differences regarding the SUCRe-XL. Also, there is a preliminary Step 0, typically assumed by RA protocols, in which the BS broadcasts a beacon signal; this signal contains enough information to generate time synchronization and send the optimized scale factor $\delta_{\text{NOVR-XL}}^{(K,B)*}$ values, which is crucial to the well functioning of the proposed NOVR-XL protocol. Indeed, each scenario combining different K -iUEs

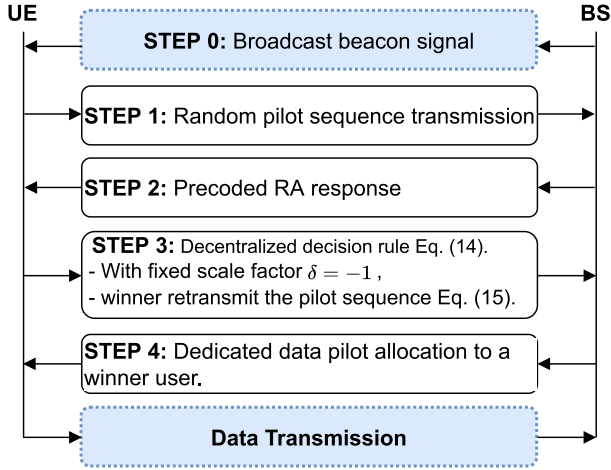


FIGURE 4. Diagram of the steps of the SUCRe-XL random access protocol.

and B values requires an optimized *scale factor*, this factor is sent to users in Step 0, and the BS maintains a look-at-table reference with these optimized factors; such a process is explained in Step III. The four steps of the proposed NOVR-XL protocol are described in the following procedure.

Step I: Random Pilot Sequence Transmission: Users in set $\mathcal{K}$ make an RA attempt with a probability P_a , synchronized and δ^* updated by beacon signal in previous Step 0, each UE that wants to be active select uniformly at random one RA pilot sequence $\mathbf{s}_t$, $t = 1, \dots, \tau_{\text{RA}}$ and transmits with the user's transmit power p_k . Therefore, let $\mathcal{N}_t \in \mathcal{K}$ be the set of UEs indices that selected pilot t . As in [25], it follows a binomial distribution.

$$|\mathcal{N}_t| \sim \mathcal{B}\left(K, \frac{P_a}{\tau_{\text{RA}}}\right). \quad (3)$$

All inactive UEs that decide to connect transmit RA UL pilot sequences, making SA b receive the signal:

$$\mathbf{Y}^{(b)} = \sum_{k \in \mathcal{K}} \sqrt{p_k} h_k^{(b)} \mathbf{s}_t^T + \mathbf{N}^{(b)}, \quad (4)$$

where $\mathbf{Y}^{(b)} \in \mathbb{C}^{M_b \times \tau_{\text{RA}}}$ and $\mathbf{N}^{(b)} \in \mathbb{C}^{M_b \times \tau_{\text{RA}}}$, with entries drawn from $\mathcal{CN}(0, \sigma^2)$, is the receiver noise. So each SA correlates (4) using an arbitrary normalized pilot $\mathbf{s}_t$, and can estimate the channel of UEs in the same set $\mathcal{N}_t$ yielding:

$$y_t^{(b)} = \mathbf{Y}^{(b)} \frac{\mathbf{s}_t^*}{\|\mathbf{s}_t\|} = \sum_{i \in \mathcal{N}_t} \sqrt{p_i \tau_{\text{RA}}} h_i^{(b)} + n_t, \quad b = 1, \dots, B, \quad (5)$$

where $\mathbf{n}_t = N \frac{\mathbf{s}_t^*}{\|\mathbf{s}_t\|} \sim \mathcal{CN}(0, \sigma^2 I_{M_b})$ is the effective receiver noise. Due the properties of massive MIMO, *channel hardening* and *asymptotic favorable propagation*, we have in each SA:

$$\frac{\left\| \sum_{b \in \mathcal{M}} y_t^{(b)} \right\|^2}{M_b} \xrightarrow{M_b \rightarrow \infty} \alpha_t + B\sigma^2, \quad (6)$$

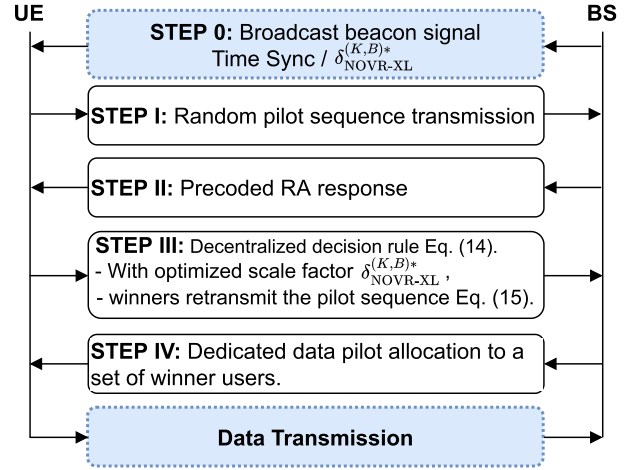


FIGURE 5. Diagram of proposed random access NOVR-XL protocol, the dashed Steps are the same as SUCRe-XL, "STEP 0" and "Data Transmission" phases were included.

where α_t is the *sum of the signal gains* associated with pilot t in each b -SA:

$$\alpha_t = \sum_{b \in \mathcal{M}} \sum_{i \in \mathcal{N}_t} p_i \beta_i^{(b)} \tau_{\text{RA}}. \quad (7)$$

Step II: Precoded Random Access Response: In SA coordinated mode; SAs transmit a precoded DL pilots to UEs that selected pilot t in Step I.

$$\mathbf{V}_{\text{XL}} = \sqrt{\frac{q}{B}} \sum_{t=1}^{\tau_{\text{RA}}} \frac{\sum_{b \in \mathcal{M}} y_t^{(b)*}}{\left\| \sum_{b \in \mathcal{M}} y_t^{(b)} \right\|} \mathbf{s}_t^H, \quad (8)$$

where $\mathbf{V}_{\text{XL}} \in \mathbb{C}^{M_b \times \tau_{\text{RA}}}$ and q is the BS transmit power. In this case UEs selecting pilot sequence t receive signal $\mathbf{z}_k^T \in \mathbb{C}^{\tau_{\text{RA}}}$ in the reciprocal channel with receiver noise $\eta_k \sim \mathcal{CN}(0, \sigma^2 I_{\tau_{\text{RA}}})$:

$$\mathbf{z}_k^T = \sum_{m \in \mathcal{V}_k} h_k^{(m)T} \mathbf{V}_{\text{XL}} + \eta_k^T. \quad (9)$$

From (9), each UE correlates the signal with its RA pilot $\mathbf{s}_t$,

$$z_k = \mathbf{z}_k^T \frac{\mathbf{s}_t}{\|\mathbf{s}_t\|} = \sqrt{\frac{q \tau_{\text{RA}}}{B}} \sum_{m \in \mathcal{V}_k} h_k^{(m)T} \frac{\sum_{b \in \mathcal{M}} y_t^{(b)*}}{\left\| \sum_{b \in \mathcal{M}} y_t^{(b)} \right\|} + \eta_k \quad (10)$$

where $\eta_k = \eta_k \frac{\mathbf{s}_t}{\|\mathbf{s}_t\|} \sim \mathcal{CN}(0, \sigma^2)$ is the effective receiver noise. To obtain (6) at the UE side, dividing the eq. (10) by $\sqrt{M_b}$ and considering that *asymptotic favorable propagation* and *channel hardening*, it follows that:

$$\frac{z_k}{\sqrt{M_b}} = \sqrt{\frac{q \tau_p}{B}} \frac{\left(\sum_{m \in \mathcal{V}_k} h_k^{(m)T} \sum_{b \in \mathcal{M}} y_t^{(b)*} \right)^*}{M_b \sqrt{\frac{1}{M_b} \left\| \sum_{b \in \mathcal{M}} y_t^{(b)} \right\|^2}} + \frac{\eta_k}{\sqrt{M_b}} \xrightarrow{M_b \rightarrow \infty} \sqrt{\frac{q p_k}{B}} \frac{\sum_{m \in \mathcal{V}_k} \beta_k^{(m)}}{\sqrt{\alpha_t + B\sigma^2}}. \quad (11)$$

Thus, using the function $\Re(\cdot)$ gives the real part of your input and discards the imaginary part, where in this case mitigates noise and estimation errors from eq. (11), resulting:

$$\frac{\Re(z_k)}{\sqrt{M_b}} \approx \tau_{\text{RA}} \sqrt{\frac{qp_k}{B}} \frac{\sum_{m \in \mathcal{V}_k} \beta_k^{(m)}}{\sqrt{\alpha_t + B\sigma^2}} \quad (12)$$

this way each UE- k in $\mathcal{N}_t$ can estimate α_t , so that

$$\hat{\alpha}_{t,k} = \max \left[p_k \sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \tau_{\text{RA}}, \left(\frac{\Gamma(M_b + 1/2)}{\Gamma(M_b)} \right)^2 \frac{qp_k \tau_{\text{RA}}^2 \left(\sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \right)^2}{B[\Re(z_k)]^2} - B\sigma^2 \right] \quad (13)$$

Step III: Contention Resolution with NOMA and Pilot Repetition: By exploring the NOMA, users colliding the same pilot sequence can be served at the same resource, *i.e.*, spatial (SA), frequency, and time.

The collision resolution follows the decentralized decision rule, *i.e.*, each user decides whether to continue in the access attempt, by following the decision rule [10]:

$$p_k \sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \tau_{\text{RA}} > \frac{\hat{\alpha}_{t,k}}{2} + \epsilon_k \quad (14)$$

If inequality in Eq. (14) holds, the user repeats its pilot transmission and forms the new set $\mathcal{S}_t$; otherwise, it remains silent and tries to communicate in the next RA attempt with probability P_{na} . The cardinality of $\mathcal{S}_t$ defines the contention level; indeed, if $|\mathcal{S}_t| = 1$ the collision was solved by the decision rule; otherwise, for $|\mathcal{S}_t| > 1$ (multiple UEs) representing the possible *false positive* case, it will be solved by power-domain NOMA.

$$\mathcal{S}_t = \left\{ k : p_k \sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \tau_{\text{RA}} > \frac{\hat{\alpha}_{t,k}}{2} + \epsilon_k \right\} \quad (15)$$

The decision rule in Eq. (14), presents a bias term ϵ_k , given by:

$$\epsilon_k = \frac{\delta}{\sqrt{M_b} \cdot \sum_{b \in \mathcal{V}_k} \beta_k^{(b)}}, \quad (16)$$

where δ is a *scale factor* that may be adjusted in order to improve the system metrics such as the *average number of access attempts* (ANAA), *fraction of failed access attempts* (FFAA), *normalized number of accepted UEs* (NNAU), *average sum-rate* (aSR), and probability of collision resolution.

Herein, we suggest optimizing the *scale factor* δ numerically, allowing multiple users in the NOVR-XL RA protocol could consider themselves the winners in the contention resolution process, generating *false positive* in SUCRe-XL, which can be resolved in the next step by SIC into the NOMA structure. Indeed, the signal is processed in each SA via RPU by deploying NOMA scheme; hence, resolving overlapping

signals via SIC strategy, aiming to decode, reconstruct and cancel the user's signals sharing the same pilot sequence and VR. Finally, the RPU output signals are equally combined in a centralized EGC at the CPU. The purpose of adjustment in the *scale factor* is to attain the system sum-rate maximization. Numerical results section explores the *scale factor* δ optimization.

1) SCALE FACTOR OPTIMIZATION

the optimal value $\delta_{\text{NOVR-XL}}^{(K,B)*}$ is based on [10], by deploying exhaustive search approach. The procedure is numerically optimal in the sense of sum-rate maximization; hence, each K inactive UEs and B -SA configuration implies an optimal parameter value. To obtain the optimal δ for practical NOMA XL-MIMO system and channel scenarios, the following procedure was adopted:

- 1) The BS knows the number of SAs B , and an estimation of the number of users intending to become active KP_a ¹ present in the cell.
- 2) Starting with the SUCRe value, $\delta_{\text{SUCRe-XL}}^* = -1$, as in [7], the *scale factor* is tuned iteratively to improve the system sum-rate. Sporadically, motivated by the changes in the physical scenario, BS could re-optimize and broadcast the *scale factor* value in a look-at-table procedure.²
- 3) Select the optimal $\delta_{\text{NOVR-XL}}^{(K,B)*}$ in the sense that maximizes the system sum-rate, such as:

$$\delta_{\text{NOVR-XL}}^{(K,B)*} = \underset{\delta}{\operatorname{argmax}} R_{\sum(K,B)}^{\text{NOVR-XL}} \quad (17)$$

and record the optimal scale factor in the reference table. Notice that the *scale factor* δ in eq. (16) affects the decision rule defined in Eq. (14), as well as the system sum-rate $R_{\sum(K,B)}^{\text{NOVR-XL}}$ as defined in the sequel, Eq. (22).

- 4) In a preamble step of the protocol, namely Step 0, taking into account the current system and channel scenario, the BS look-at-table and broadcast the correspondent optimized $\delta_{\text{NOVR-XL}}^{(K,B)*}$ value.

Without loss of generality, in the contention resolution step the protocol accepts multiple users in NOMA cluster; hence, the BS proceeds with the SIC according to the descending order of channel gains ($\|h_1\| > \|h_k\| > \|h_{|\mathcal{S}_t|}\|$), *i.e.*, the strongest UE-1 signal is decoded first without SIC application, followed by UE-2 signal with one step of SIC to remove UE-1 signal and remaining the interference of UE-3; in the sequel decodes the signal of UE-3 after the two steps SIC of the signal of UE-1 and UE-2, and so on until the last user in set $|\mathcal{S}_t|$.

¹This estimation process can be obtained by a slight modification of Step 0, by including a tone or pilot sequence transmitted by each user trying to become active.

²Look-at-table is a reference table where the BS keeps the optimized values of the scale factor, allowing that each UE could find and use them expeditiously.

The signal-to-interference-plus-noise ratio (SINR) in a SA basis of user k selecting the same pilot sequence t is defined as:

$$\gamma_{t,k}^{(b)} = \frac{p_k \beta_k^{(b)}}{\varpi \sum_{j=1}^{k-1} p_j \beta_j^{(b)} + \sum_{i=k+1}^{|\mathcal{S}_t|} p_i \beta_i^{(b)} + \sigma^2}, \quad (18)$$

where ϖ is the residual interference factor after step SIC assuming imperfect interference cancellation. The rate achieved by each user per SA can be defined as:

$$R_{t,k}^{(b)} = \log_2(1 + \gamma_{t,k}^{(b)}). \quad [bpcu] \quad (19)$$

In our proposed RA protocol, when the distributed *decision rule* results in possible *false positives* i.e., multiple users repeating the selected pilot after the *decision rule* Eq. (14), these users are grouped in the set $\mathcal{S}_t$ and are treated applying NOMA-SIC strategy, regardless of the cardinality of this set. However, to guarantee that SIC could operate suitably, minimal power differences from colliding UEs' signals must be guaranteed; this can be seen as a minimum SINR among the colliding signals in the same SA. To make a realistic comparison, we adopt a threshold data rate $R_{th}^{(b)}$; hence, users in the set $\mathcal{S}_t$ not satisfying the threshold rate in an SA basis are still deemed *false positives*. Therefore, one can define a new subset of users with successful RA attempt $\mathcal{S}_t^{th}$, formed by collided users achieving the threshold rate as:

$$\mathcal{S}_t^{th} = \left\{ k : R_{t,k}^{(b)} > R_{th}^{(b)} \right\}, \quad (20)$$

which represents those colliding UE' signals resolved by the SIC-NOMA. As a consequence, the subset of *false positive* users $\mathcal{S}_t^{fp}$ hold as:

$$\mathcal{S}_t^{fp} = \mathcal{S}_t \setminus \mathcal{S}_t^{th}. \quad (21)$$

Remark: we named possible *false positive* users in the set $\mathcal{S}_t$, after SIC operation, those users fulfilling the rule Eq. (20), forming the set of users $\mathcal{S}_t^{th}$ that attain successful RA attempt, and a subset $\mathcal{S}_t^{fp}$ named *false positive* users as the subset difference of $\mathcal{S}_t$ and $\mathcal{S}_t^{th}$.

Step IV: The BS estimates the channel³ to the user, or users, considering the pilot signal sent in Step III, it reconstructs the signal and recurrently applies SIC procedure. After that, the corresponding message is decoded, and users are identified. After decoding is successful a subsequent allocation of the dedicated payload pilots coherence blocks. Unsuccessful UEs are instructed to try again after a random interval, limited to a maximum of 10 attempts, then it stops sending pilot sequence and the packet is considered lost.

In the data transmission step of the SUCRe-XL protocol, only one user is accepted per pilot sequence in non-overlapping VR, on the other hand, in NOVR-XL protocol there will also be multiple users who consider themselves

³For simplicity, and without loss of generality, we consider perfect CSI estimates.

winners in the contention step. The sum-rate of NOVR-XL protocol in an SA basis can be defined using (18) as:

$$R_{\sum(K,B)}^{\text{NOVR-XL}} = \sum_{b \in \mathcal{V}_k} \sum_{t=1}^{\tau_{RA}} \sum_{k=1}^{|\mathcal{S}_t^{th}|} \log_2 \left(1 + \gamma_{t,k}^{(b)} \right), \quad [bpcu] \quad (22)$$

While in the SUCRe-XL the sum-rate can be defined as:

$$R_{\sum(K,B)}^{\text{SUCRe-XL}} = \sum_{b \in \mathcal{V}_k} \sum_{t=1}^{\tau_{RA}} \log_2 \left(1 + \frac{p_k \beta_k^{(b)}}{\sigma^2} \right), \quad [bpcu] \quad (23)$$

by assuming that just one user per pilot sequence is accepted, and $\delta_{\text{SUCRe-XL}}^{(K,B)*} = -1$, as in [13].

IV. NUMERICAL RESULTS

It is assumed an URA with $M = 500$ antennas in a $200 \times 100 \text{ m}^2$ cell area with K inactive users (iUEs) randomly distributed as illustrated in Fig. 3, the access probability P_a , the number of available pilots τ_{RA} , and $q = p_k = 1 \text{ W}$. The main simulation parameters are listed in Table 1. The URA has divided in B SA and the visibility success probability $P_b = 0.5$ to each SA. For simulation purposes, $|\mathcal{V}_k| > 0$, $\forall k$, i.e., all users have at least one SA in its VR.

In the numerical simulations, two settings are proposed to evaluate the performance of the NOVR-XL protocol numerically: in setting 1 we assume the SIC running perfectly i.e., $\varpi = 0.00$, and a threshold rate $R_{th1}^{(b)} = 1.585 [bpcu]$ is adopted, which represents the signal of interest has a power difference w.r.t. the interference enough to be decoded correctly i.e., $\Delta P \geq 2$. While in setting 2, due fail in any previous crucial process, the SIC is admitted imperfect with residual interference factor $\varpi = 0.10$; this is a typical value of residual factor adopted in NOMA literature [10], [26], and a threshold rate $R_{th2}^{(b)} = 0.585 [bpcu]$, hence, the signal of interest can lack enough power difference to be decoded perfectly, in this case, we have adopted $0.5 \leq \Delta P < 2$.

A. SYSTEM SUM-RATE

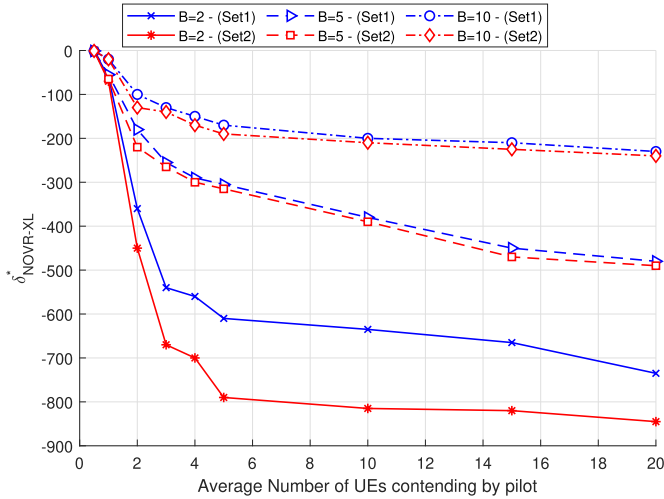
The optimal value of *scale factor* $\delta_{\text{NOVR-XL}}^{(K,B)*}$ is obtained numerically based on [10], by exhaustive search approach; it is optimal in the sense of sum-rate maximization; hence, each K -iUE and B -SA configuration requires a search to find the optimal parameter value.

Fig. 6 reveals results for the optimal *scale factor* parameter obtained numerically, following the procedure described in Step III, while it correspondent maximized NOVR-XL sum-rate is shown in Fig. 7, we proposed to consider two different configurations 1 and 2. As evinced, the optimal *scale factor* decreases when the interference increases; whether by the number of UEs contending the pilot grows or when the number of SA decreases.

Notice that the baseline (SUCRe-XL) requires the non-overlapping VR to accept the device; hence, presents a modest gain in terms of sum-rate even increasing the number of SA. The highest sum-rate values occur before the saturation of usage of available pilots. On the other hand, the proposed

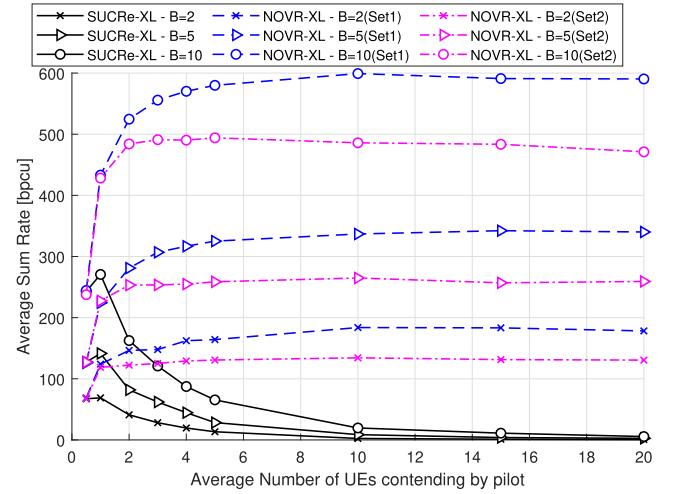
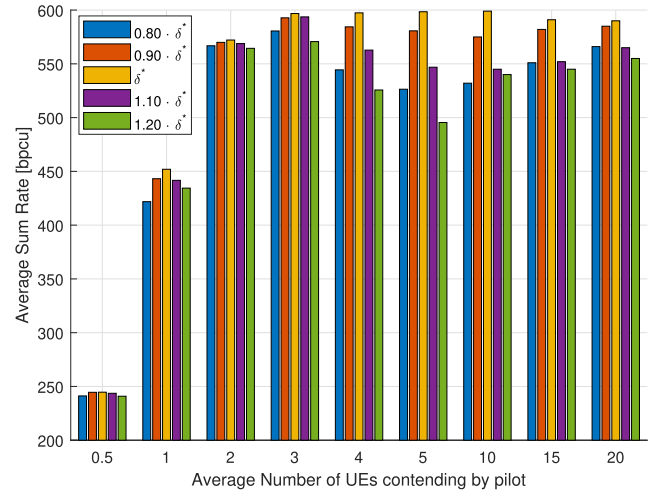
TABLE 1. Simulation parameters.

Parameter	Value
User distance to array	$r_{k,m_z,m_y} \in [10; 180\text{m}]$
User distance from the ground	$h_{user} \in [1; 1.7\text{m}]$
Antennas elements y-axis, z-axis	$M_y = 100, M_z = 5$
Number of BS antennas (URA)	$M = M_y \times M_z = 500$
Number of Subarrays	$B \in \{2; 5; 10\}$
Antenna element distance	$d_m = 1\text{m}$
Array height	$h_{array} = 12\text{m}$
Array length	$L_y = M_y \times d_m$
User transmit power	$p_k = 1\text{W}$
BS transmit power	$q = 1\text{W}$
Noise Power	$\sigma^2 = -98.65\text{dBm}$
Number of iUEs	$K \in [0; 20000]$
Access probability (1 st attempt)	$P_a = 0.01$
Access probability (new attempts)	$P_{na} = 0.5$
Number of available pilot sequences	$\tau_{RA} = 10$
Monte Carlo realizations	5000
Setting 1	
SIC residual	$\varpi = 0.00$
Threshold rate 1	$R_{th1}^{(b)} = 1.585 [\text{bpcu}]$
Setting 2	
SIC residual	$\varpi = 0.10$
Threshold rate 2	$R_{th2}^{(b)} = 0.585 [\text{bpcu}]$

**FIGURE 6. $\delta_{NOVR-XL}^*$ optimum in terms of sum-rate maximization, obtained numerically for each user loading, both setting 1 and 2, as well for different number of SAs, $B \in \{2; 5; 10\}$.**

NOVR-XL protocol processes the signal superimposed on each SA, canceling the more robust user signal, in some situations, with an assumed residual SIC interference on the order of ϖ , obtaining better sum-rates results when B increases. Also, a higher sum-rate limit is reached when adjusting *scale factor* δ , establishing the maximum number of users selecting the same pilot in the data transmission step; besides, when the number of iUEs grows the *scale factor* value decreases. Notice that the *scale factor* δ value is relatively sensitive to the interference.

In Setting 1, the maximum system sum rate was reached at a loading of $10(K \times P_a)/\tau_{RA}$. In Setting 2 the maximum sum-rate was reached at loading $5(K \times P_a)/\tau_{RA}$, increasing

**FIGURE 7. Average sum-rate as a function of a number of iUEs. Obtained after adjusting $\delta_{NOVR-XL}^*$.****FIGURE 8. Average sum-rate as function of number of iUEs, $B = 10$ in Setting 1 of NOVR-XL, and sum-rate achieved with δ values.**

the load of users from these points observe reduces the sum rate of the system.

In order to verify the sensitivity of the *scale factor* δ^* , Fig. 8 shows the system sum-rate with $B = 10$ SAs. When the interference increases by increasing the number of iUEs K , the sensitivity of the *scale factor* δ on the sum-rate increases substantially, revealing the importance in accurately estimating the *scale factor* parameter in the NOVR-XL protocol context.

B. AVERAGE NUMBER AND FRACTION OF FAILED ACCESS ATTEMPTS

Fig. 9 shows that the proposed protocol NOVR-XL in crowded scenarios ($1 < K \times P_a < 20$) achieves improved RA performance. The modification in decentralized collision resolution is that NOVR-XL accepts multiple users colliding on the overlapping VR, thanks to the power diversity; therefore, the superposition signal can be processed after SIC process

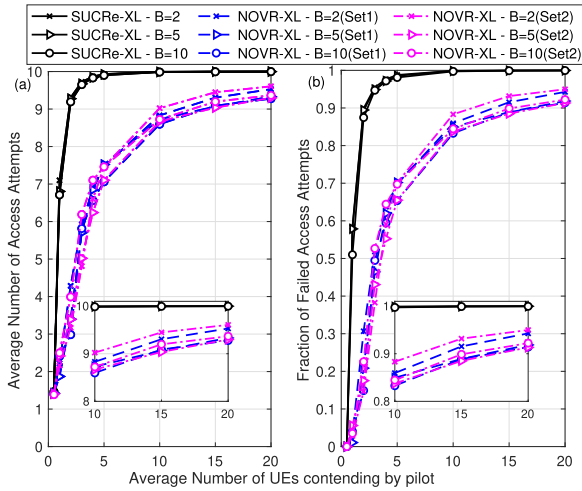


FIGURE 9. RA systems performance. The baseline SUCRe-XL and NOVR-XL. (a) Avg. number of RA attempts. (b) Prob. of failed access attempts.

in each SA. With the increase in the number of subarrays B , but constrained by the channel hardening and favorable propagation properties, the probability of collision in each subarray by sharing overlapping VRs in each SA increases. Such collision resolution configuration is impossible to treat with SUCRe-XL due to the absence of the interference cancellation step. As a result, the NOVR-XL allows decreasing the number of access attempts even under solid interference in the same SA.

C. NORMALIZED NUMBER OF ACCEPTED USERS

The normalized number of accepted UEs, calculated by the number of users accepted by the number of users trying to access, given in Fig. 10 corroborate with the improved results in NOVR-XL proposed protocol to accept UEs to transmit due to the superposition signal can be processed by SIC in each SA. In the configuration where the threshold rate is lower setting 2, the residual SIC reduces the users' rate. As the scale factor was optimized to increase the sum rate, high interference reduce the average number of accepted users.

D. COLLISION RESOLUTION

Our approach involves numerically adjusting the *scale factor* to maximize the system's average sum rate, but the two proposed settings have different responses. We consider accepted users who won the contention ($|S_t| = 1$) and users who are treated by NOMA-SIC (S_t^{th}). In setting 1 the fact that the SIC is perfect the average number of accepted users decreases more slowly and is higher in arrays with more subarrays as shown in Fig. 11 a), the higher threshold rate increase the probability of false positives, in consequence, a higher average sum rate as Fig. 7. On the other hand, in setting 2 there is a residue of interference due to the imperfect SIC which deteriorates the sum rate of NOMA. To maintain the higher level of the sum rate more collisions are resolved without applying NOMA, deteriorating the average number of Accepted UEs, as shown in Fig. 11 b), the lower threshold

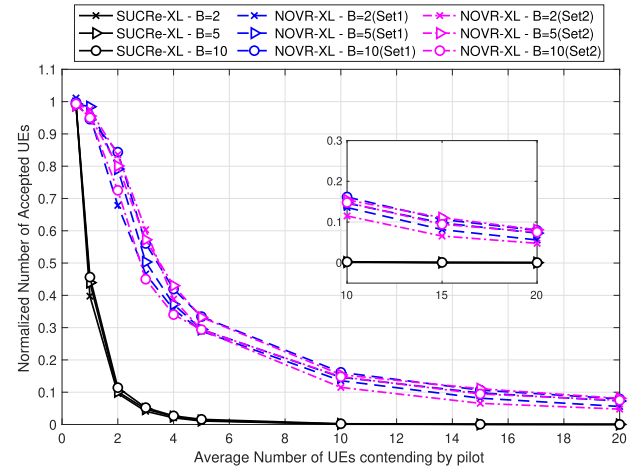


FIGURE 10. Normalized number of accepted UEs, obtained with $\delta^*_{\text{NOVR-XL}}$.

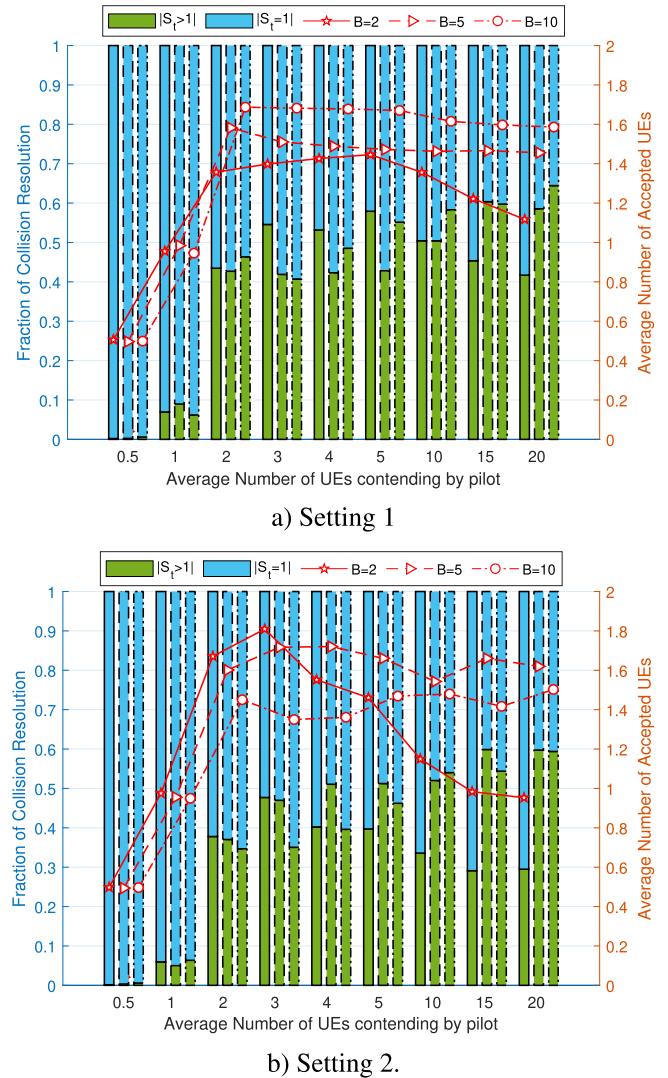


FIGURE 11. Percentage collision resolution by SIC and the average number of Accepted users in NOVR-XL.

rate decrease the probability of false positives in a lower bound in average sum rate as Fig. 7.

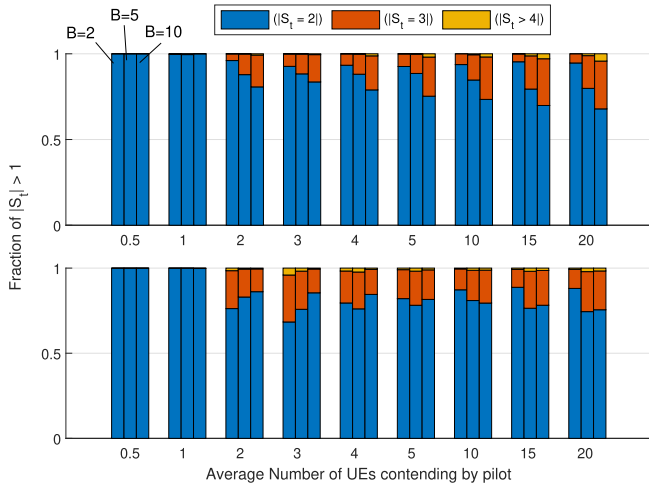


FIGURE 12. Fraction of false positive cases solved by SIC in NOVR-XL. a) Setting 1 and b) Setting 2.

Elaborating further, Fig. 12 analyses the fraction of the number of users in the set S_i of *false positives* cases resolved by the SIC-NOMA. The figure depicts that the highest number of false positives happens between 2 users, and cases greater than 3 users are rare. The proposed settings substantially change the size of the set S_i when changing the number of subarrays B .

E. OVERHEAD, ACCESS TIME DELAY AND COMPLEXITY

The grant-based protocol has a higher overhead because it needs four steps to connect devices to the network. Our proposed protocol NOVR-XL is based on a modification of SUCRe [7], to operate in XL-MIMO [13]. Such a procedure does not cause any overhead or additional time delay compared to the original SUCRe protocol [13]. The numerical optimization and the update of the reference table occur online in system operation and add no overhead to the core of protocol (steps I–IV).

In this application scenario, the proposed protocol improves the access time delay in all studied scenarios, when compared to SUCRe-XL [13], as can be seen in Figure 9.

NOVR-XL complexity that is comparable to SUCRe-XL, the way it is calculated the precoded Random Access Response (8) is the same. The optimized *scale factor* $\delta_{\text{NOVR-XL}}^{(K,B)*}$ eventually shared in Step 0, does not add complexity to the proposed protocol.

V. CONCLUSION

By exploiting power-domain NOMA and signal processing in each SA the grant-based RA protocol performance can be improved. Indeed, the XL-MIMO adds a new degree of freedom with the received signal being processed by different SAs with the support of RPU, together with the degree of freedom introduced by power-domain NOMA, allowing to improve the collision resolution of multiple users using the same pilot sequence in the same VR subarrays, and by applying the SIC stage in each SA at the BS side. Supported

by numerical results, we found a substantial performance improvement of the proposed RA protocol regarding the recent literature SUCRe-XL method, with the numerical optimization of *scale factor* δ , in order to maximize the system sum-rate, making multiple users collide deploying the same pilot sequence.

The primary distinction regarding the SUCRe-XL and the NOVR-XL schemes consists of the independent processing per RPU and the collision resolution process. In addition, we unveil the sensitivity of the *scale factor* δ in the NOVR-XL scheme, mainly in scenarios with high interference; hence, the suitable δ value selection is paramount to improve the collision resolution and the system sum-rate. Indeed, in the proposed NOVR-XL protocol, the received superimposed signal data can be properly decoded with SIC steps in a SA basis, even with a rate threshold restriction, and able to achieve both a reduction in the number of access attempts and simultaneously a significant increase in the sum-rate, increasing the average number of accepted users under crowded mMTC scenarios analyzed ($1 < K \times P_a < 20$).

REFERENCES

- [1] A. Akbar, S. Jangsher, and F. A. Bhatti, "NOMA and 5G emerging technologies: A survey on issues and solution techniques," *Comput. Netw.*, vol. 190, May 2021, Art. no. 107950.
- [2] M. T. P. Le, L. Sanguinetti, E. Björnson, and M.-G.-D. Benedetto, "Code-domain NOMA in massive MIMO: When is it needed?" *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4709–4723, May 2021.
- [3] F. Sun, K. Niu, and C. Dong, "Deep learning based joint detection and decoding of non-orthogonal multiple access systems," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–5.
- [4] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2192–2235, 4th Quart., 2020.
- [5] F. Clazzer, A. Munari, G. Liva, F. Lazaro, C. Stefanovic, and P. Popovski, "From 5G to 6G: Has the time for modern random access come?" 2019, *arXiv:1903.03063*.
- [6] M. V. da Silva, S. Montejo-Sánchez, R. D. Souza, H. Alves, and T. Abrão, "D2D assisted Q-learning random access for noma-based MTC networks," *IEEE Access*, vol. 10, pp. 30694–30706, 2022.
- [7] E. Björnson, E. de Carvalho, E. G. Larsson, J. H. Sørensen, and P. Popovski, "A random access protocol for pilot allocation in crowded massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2220–2234, Apr. 2017.
- [8] H. Han and Y. Li, "A high throughput pilot allocation for M2M communication in crowded massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9572–9576, Oct. 2017.
- [9] H. Han, Y. Li, and X. Guo, "A graph-based random access protocol for crowded massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7348–7361, Nov. 2017.
- [10] H. D. Pereira, G. Brante, J. Farhat, R. D. Souza, J. C. M. Filho, and T. Abrão, "On the sum-rate of contention resolution in massive MIMO with NOMA," *IEEE Access*, vol. 9, pp. 24965–24974, 2021.
- [11] E. D. Carvalho, A. Ali, A. Amiri, M. Angelichinoski, and R. W. Heath Jr., "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, Aug. 2020.
- [12] J. Flordelis, X. Li, O. Edfors, and F. Tufvesson, "Massive MIMO extensions to the COST 2100 channel model: Modeling and validation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 380–394, Jan. 2020.
- [13] O. S. Nishimura, J. C. Marinello, and T. Abrão, "A grant-based random access protocol in extra-large massive MIMO system," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2478–2482, Nov. 2020.
- [14] O. S. Nishimura, J. C. M. Filho, T. Abrão, and R. D. Souza, "Fairness in a class barring power control random access protocol for crowded XL-MIMO systems," *IEEE Syst. J.*, vol. 16, no. 3, pp. 4574–4582, Sep. 2022.

- [15] J. C. M. Filho, G. Brante, R. D. Souza, and T. Abrao, "Exploring the non-overlapping visibility regions in XL-MIMO random access and scheduling," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6597–6610, Aug. 2022.
- [16] Y. Han, S. Jin, C.-K. Wen, and X. Ma, "Channel estimation for extremely large-scale massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 633–637, May 2020.
- [17] V. Croisfelt, A. Amiri, T. Abrao, E. de Carvalho, and P. Popovski, "Accelerated randomized methods for receiver design in extra-large scale MIMO arrays," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6788–6799, Jul. 2021.
- [18] J. H. I. de Souza, A. Amiri, T. Abrao, E. de Carvalho, and P. Popovski, "Quasi-distributed antenna selection for spectral efficiency maximization in subarray switching XL-MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6713–6725, Jul. 2021.
- [19] L. M. Taniguchi, J. H. I. Souza, D. W. M. Guerra, and T. Abrão, "Resource efficiency and pilot decontamination in XL-MIMO double-scattering correlated channels," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 12, p. e4365, Dec. 2021.
- [20] J. C. Marinello, T. Abrao, A. Amiri, E. de Carvalho, and P. Popovski, "Antenna selection for improving energy efficiency in XL-MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13305–13318, Nov. 2020.
- [21] T. A. B. Alves and T. Abrão, "Crowded MTC random access in NOMA XL-MIMO," in *Proc. IEEE SIoT LABCIoT*, Oct. 2022.
- [22] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, Nov. 2017.
- [23] H. Lu and Y. Zeng, "Near-field modeling and performance analysis for multi-user extremely large-scale MIMO communication," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 277–281, Feb. 2022.
- [24] M. Fallgren and B. Timus, *Deliverable D1.1: Scenarios, Requirements and KPIs for 5G Mobile and Wireless System*, document ICT-317669-METIS, 2013.
- [25] E. Björnson, E. de Carvalho, E. G. Larsson, and P. Popovski, "Random access protocol for massive MIMO: Strongest-user collision resolution (SUCR)," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
- [26] X. Wang, R. Chen, Y. Xu, and Q. Meng, "Low-complexity power allocation in NOMA systems with imperfect SIC for maximizing weighted sum-rate," *IEEE Access*, vol. 7, pp. 94238–94253, 2019.



THIAGO AUGUSTO BRUZA ALVES received the B.S. degree in electrical engineering from the Universidade Norte do Paraná (UNOPAR), and the M.Sc. degree in electrical engineering from the Universidade Estadual de Londrina (UEL), Paraná, Brazil, in 2017. He is currently pursuing the joint Ph.D. degree in electrical engineering with UEL and UTFPR Program. His research interests include machine learning and resource allocation applied to telecommunications, extra large massive MIMO, access protocol for mMTC systems, and application of NOMA technologies.



TAUFIK ABRÃO (Senior Member, IEEE) received the B.S., M.Sc., and Ph.D. degrees in electrical engineering from the Polytechnic School, University of São Paulo, São Paulo, Brazil, in 1992, 1996, and 2001, respectively. Since March 1997, he has been with the Communications Group, Department of Electrical Engineering, Londrina State University, Paraná, Brazil, where he is currently an Associate Professor in telecommunications and the Head of the Telecommunications and Signal Processing Laboratory. From July 2018 to October 2018, he was with the Connectivity Section, Aalborg University, as a Guest Researcher. In 2012, he was an Academic Visitor with the Southampton Wireless Research Group, University of Southampton, U.K. He has also served as an Associate Editor for the IEEE SYSTEMS JOURNAL, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the *IET Electronics Letters*, the *ETT* journal (Wiley), and the *AEUe* (Elsevier). Previously, he served as an Associate Editor for the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, from 2013 to 2017, IEEE ACCESS, from 2016 to 2019, and the *IET Signal Processing*, from 2018 to 2020. His research interests include massive and XL-MIMO, RIS-aided communications, URLLC, mMTC, random access protocols, detection and estimation, NOMA systems, cooperative communication, resource allocation, machine learning-aided communications, and optimization techniques for 5G and 6G wireless systems. He is also a Pq-1D CNPq.

...

ANEXO D – Crowded MTC Random Access in NOMA XL-MIMO

Crowded MTC Random Access in NOMA XL-MIMO

Thiago Augusto Bruza Alves
Electrical Eng. Dept. - UEL
Londrina, Brazil.
thiagobruza@outlook.com

Taufik Abrão
Electrical Eng. Dept. - UEL
Londrina, Brazil.
taufik@uel.br

Abstract—Massive MIMO is one of the key technologies to support the growth of massive access attempts by devices, such as in massive machine type communication (mMTC). The evolution of antenna array technology brought the recent extra-large scale massive multiple input multiple output (XL-MIMO) systems, seen as a promising technology for providing very high-data rates in high-user density scenarios. Spatial non stationarities and visibility regions (VRs) occur across all huge XL array extension, since its large dimension is of the same order of the distances to the user equipment (UE). We investigate the random access (RA) problem in crowded XL-MIMO scenarios. The proposed non-orthogonal multiple access (NOMA) visible region extra large array (NVR-XL) protocol takes advantage of the power domain NOMA to allow access of two or more users colliding in the same XL sub-array (SA) selecting the same pilot sequence. The NVR-XL provides a reduction in the number of attempts to access the network, while improving the average sum-rate, as the number of SA increases.

Index Terms—NOMA; Machine Type Communication; XL-MIMO; Crowded scenarios.

I. INTRODUCTION

The number of active devices keeps growing jointly with the number of access requests for different types of services (e.g. vehicles, sensors, mobiles, etc.), and in applications for 5G and beyond, as massive machine type communication (mMTC) and crowded mobile broadband (cMBB) use cases, the base station (BS) will need to attend simultaneously devices with a range of capabilities and deployments.

Power-domain NOMA was proposed initially to improve the spectral efficiency of wireless networks, sharing the same orthogonal resource (time and frequency), by superposition coding in transmitter side and successive interference cancellation (SIC) in receiver side [1].

A vast number of devices trying to access the network at the same time is a medium access control challenge. Traditional random access (RA) schemes are not able to handle such a large number of requests [2]. Pure RA schemes, like ALOHA, have severe performance limitation. The incorporation of NOMA significantly improves the system performance, by admitting two or more users per time slot at same frequency [3].

In a massive RA scenario, the number of user connection attempts far outnumbers the number of available pilot sequences.

This work was supported in part by the CAPES (Financial Code 001) and National Council for Scientific and Technological Development (CNPq) of Brazil under Grant 310681/2019-7.

As a consequence, the implementation of a collision resolution protocols became essential to allow coherent communication. One well-known decentralized grant-based random access protocol for crowded massive multiple-input multiple-output (MIMO) systems is the strongest user collision resolution (SUCRe), which takes advantage of MIMO properties [4], giving preference to users with good channel conditions.

Combining NOMA and SUCRe was an important evolution. In [5] compare to the classic SUCRe scheme, the proposed NOMA-RA protocol allows users that try to access the medium with the same pilot signal to resolve collisions by distinguishing them in the power domain and attaining a superior sum-rate performance with reduced average latency.

In [6], an adaptation of SUCRe in extra large scenarios is proposed. The concept of VR is explored in XL-MIMO context, since there is a relationship between the geographic area and the portion of the array that is visible from that area. When a certain UE is present in that region, the corresponding visibility region is accessible; on the other hand, when UE exits the VR, it encounters a new set of clusters. The probability that SA b is visible by UE k can be explored to improve the RA.

The contribution of this paper is twofold: **a)** explore the overlapping VR in the proposed NVR-XL RA protocol, by combining the advantages of XL-MIMO scheme and NOMA, allowing two users to select the same pilot sequence at the same SA. The superimposed signal is processed on a SA basis, which is performed SIC to decode each user's information. **b)** carry out a comprehensive numerical simulation analysis by comparing the proposed NVR-XL scheme with the literature SUCRe-XL protocol by deploying three different metrics.

II. SYSTEM MODEL

We consider an XL-MIMO setup where the BS deploys an extra-large antenna array, operating in a time-division-duplexing (TDD). Without loss of generality, we assume a uniform rectangular array (URA) is placed on the y - z plane with a total of M antenna elements, i.e., $M_y \times M_z$, where M_y and M_z denote the number of antenna elements along the y - and z -axis, respectively, with the first line of an array starting at the origin O and deployed along the ordinate axis, as in Fig. 1. Similarly to [6], [7], we consider a simplified bipartite graph model in XL-MIMO, the array is divided into B subarrays (SAs), each composed of a fixed number

of $M_b = M/B$ antennas, ensuring the minimum antennas elements ($M_b \geq 50$), to achieve the inherent properties of massive MIMO.

All single-antenna user devices are randomly distributed and grouped in set $\mathcal{U}$, $\mathcal{A} \subset \mathcal{U}$ be the subset of active UEs, with temporarily dedicated data pilots, and $\mathcal{K} = \mathcal{U} \setminus \mathcal{A}$ be the set of inactive UEs (iUEs). The devices in set $\mathcal{K}$ do not have dedicated pilots; if they want to be active, they must be assigned one. Hence, the number of iUEs in the cell is $|\mathcal{K}| = K$. The BS only allocates pilots to active devices and reclaims the pilots when needed. The iUEs make a RA attempt with probability P_a .

Let $\mathcal{M}$ denote the set of all BS SAs and $\mathcal{V}_k \subset \mathcal{M}$ the subset of SAs visible to the k -th UE. To compose the mathematical modeling, we assume that the subset $\mathcal{V}_k$ is generated at random, where each SA is visible with probability P_b by UE k , like in [6], [7]. This probability simulates the influence of random obstacles and scatterers in the environment interacting with the signals transmitted by/to the UEs, resulting in VRs. The existence of VRs as evidenced by measurements in [8].

We consider a multi-user communication system based on time-frequency coherence blocks and inactive UEs have RA blocks to realize RA attempts. We define τ_{RA} as the number of RA mutually orthogonal pilot sequences that are possible $\mathbf{s}_1, \dots, \mathbf{s}_{\tau_{RA}} \in \mathbb{C}^{\tau_{RA} \times 1}$, each with length τ_{RA} , and $\|\mathbf{s}_{\tau_{RA}}\| = \sqrt{\tau_{RA}}$.

Since each SA visibility follows a Bernoulli distribution with success probability P_b , it is crucial to note that when an antenna is visible to a user, it indicates that the user may transmit/receive signals to/from this antenna with a non-zero channel gain [7]. We assume an overcrowded scenario where K iUEs ($K \gg M$) are randomly distributed within a square cell with maximum distance d_{max} and minimum distance d_{min} , respectively, as in Fig. 2.

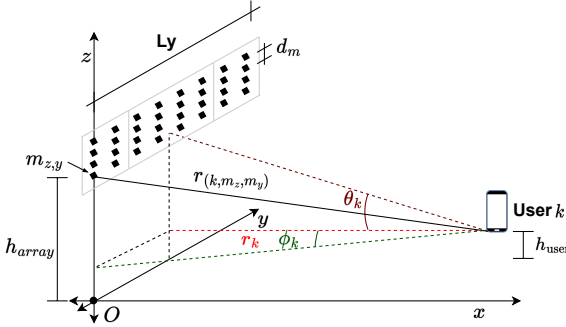


Fig. 1. Diagram of the k th user at the URA XL-MIMO with an impinging sphere waveform from elevation angle θ and azimuth angle ϕ . The $M = M_y \times M_z$ are horizontally and vertically equally spaced with spacing d_m .

A. Channel Model

In next, we model the NLOS channel based on the position of each URA antenna element $m_{z,y}$, as in Fig 1. The specific localization of $m_{z,y}$ -th array element is represented by 3D vector $\mathbf{w}_{m_{z,y}}$. The physical dimension of the URA along the y - and z -axis are $L_y = (M_y)d_m$ and $L_z = (M_z)d_m$, where d_m is the distance between elements. The user- k position is defined by 3D vector $\mathbf{q}_k$, the distance of the k -th UE to the

specific antenna element $m_{z,y}$ is determined by $r_{k,m_y,m_z} = \|\mathbf{w}_{m_{z,y}} - \mathbf{q}_k\|$ [9]. The large-scale fading includes path loss and shadowing, on the basis of an urban micro scenario model provided by:

$$\beta_{k,m_z,m_y} = 10^{-\kappa \log(r_{k,m_z,m_y}) + \frac{g+\varphi}{10}}, \quad (1)$$

where $g = -34.53$ dB is the path loss at the reference distance, the path loss exponent $\kappa = 3.8$, and $\varphi \sim \mathcal{N}(0, \sigma_{sf}^2)$ is the shadow fading, a log-normal random variable with standard deviation $\sigma_{sf} = 10$ dB.

Similarly as [6], [7], we adopt the average large-scale fading for the k -th user at the b -th SA, $\beta_k^{(b)} = \frac{1}{M_b} \sum_{m=1}^{M_b} \beta_{k,m}$. Let $\mathbf{h}_k^{(b)} \in \mathbb{C}^{M_b \times 1}$ be the Rayleigh fading channel vector between UE $k \in \mathcal{K}$ and SA b , following $\mathbf{h}_k^{(b)} \sim \mathcal{CN}(0, \beta_k^{(b)} \mathbf{I}_{M_b})$.

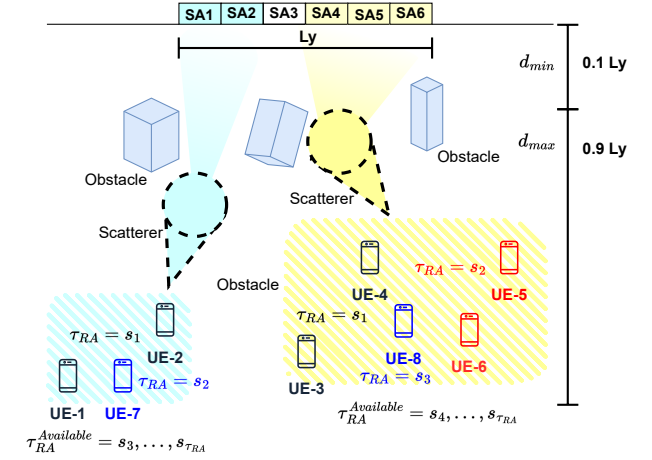


Fig. 2. Crowded MTC system with URA divided into a 6-SA; the colored regions in each SA define visible UEs, P_b is visible success probability.

III. RANDOM ACCESS FOR NOMA XL-MIMO

A. SUCRe protocol

The SUCRe protocol [4] is a distributed method for resolving pilot collisions on the user side. The protocol consists of the following steps: **a)** KP_a users randomly choose an uplink (UL) RA pilot and transmit it; **b)** the BS receives such UE signals and responds with precoded downlink (DL) pilots; **c)** Using the received signal, UEs can estimate the total of the signal strengths of UEs competing for the same RA pilot chosen by them, compare such estimate to its signal gain, and retransmit the chosen RA pilot if it judges it is the strongest UE among the contenders; **d)** in conclusion, the BS allocates dedicated data pilots to the UEs not colliding in Step before. In mMTC scenarios K is large when compared to the number of antennas M ($K \gg M$), in contrast $P_a \ll 1$.

B. Exploring VRs and NOMA in SUCRe Protocol

The proposed NVR-XL protocol combines the improvement of non-overlapping VRs singular UEs along the SAs and overlapping VRs when two users can select the same pilot along the SAs. The Steps 1 and 2 are similar to the proposed in the SUCRe-XL protocol [6].

Step I: KP_a UEs make an RA attempt.

Step II: SAs send a precoded DL pilots to UEs from Step I.

Step III: By exploring the power domain NOMA, users colliding pilots can be served at the same resource (spatial (SA), frequency, time, pilot). Due to imperfections in the SIC process, in this work, we accept collisions of up to two users, *i.e.*, performing up to one SIC step. However, in collisions between three or more users in at least one SA, the protocol fails to resolve the collision, and unsuccessful UEs are instructed to try again on the next RA with probability P_{na} , limited to a maximum of 10 attempts.

The collision resolution follows the distributed decision rule; if $\rho_k \sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \tau_{RA} > \hat{\alpha}_{t,k}/2 + \epsilon_k$ is true, the user decides to repeat its pilot transmission, otherwise, the user decides to remain inactive by pulling out the RA attempt and trying to communicate in next RA stage with probability P_{na} .

In this decision rule, the *bias* term ϵ_k is given by :

$$\epsilon_k = \frac{\delta}{\sqrt{M_b} \cdot \sum_{b \in \mathcal{V}_k} \beta_k^{(b)}}, \quad (2)$$

where δ is a scale factor that may be adjusted in order to improve some system metrics. We adjust the scale factor δ numerically as in [5], by using an exhaustive search approach so that it maximizes the sum-rate; hence, each K -iUE and B -SA configuration results in an optimal δ_{NVR-XL}^{K,B^*} .

The main feature proposed in our NVR-XL protocol is the modification in the contention process of [6], wherein a pilot collision in the set $\mathcal{S}_t$ and occurring overlaps in VRs, the signal is processed in each SA to decode superimposed signals. Hence, a SIC strategy is evoked to cancel and decode users' signals sharing the same pilot sequence.

Without loss of generality, we assume, in a two-UEs case $\|\mathbf{h}_1\| > \|\mathbf{h}_2\|$, on a SA basis, the BS conducts the SIC according to the descending order of channel gains, *i.e.*, the strongest UE-1 signal is decoded first without SIC application followed by UE-1 signal reconstruction and cancellation from the weakest UE-2 signal. Therefore, the signal-to-interference-plus-noise ratio (SINR) on each SA basis considering the two overlapping users $\gamma_1^{(b)}$ and $\gamma_2^{(b)}$ selecting the same pilot sequence τ_{RA} can be defined as:

$$\gamma_1^{(b)} = \frac{\rho_1 \beta_1^{(b)}}{\rho_2 \beta_2^{(b)} \sigma^2}, \quad \gamma_2^{(b)} = \frac{\rho_2 \beta_2^{(b)}}{\varpi_1 (\rho_1 \beta_1^{(b)}) + \sigma^2}. \quad (3)$$

where ϖ_1 is the residual interference factor after first step SIC assuming imperfect signal reconstruction.

Step IV: The BS estimate the channel to the user, or users, using the pilot signal sent in Step 3, and tries to decode the corresponding message [5]. If the decoding is successful, the BS has identified users in set $\mathcal{S}_t$ and admits its to the payload coherence blocks by allocating a dedicated data pilot sequence. If the decoding fails, the protocol has failed to resolve the collision and the unsuccessful UE is instructed to try again after a random interval, limited in a maximal of 10 attempts, then it stops sending pilot sequence and the packet is considered lost.

C. System Sum-Rate

Defining the sum rate metric to compare systems. In SUCRe-XL only one user is accepted per pilot sequence in

non-overlapping VR, on the other hand, in NVR-XL there will also be two users who consider themselves winners in the contender. The sum rate of NVR-XL on SA basis can be defined using (3) as:

$$R_{\Sigma}^{NVR-XL} = \sum_{b=1}^B \sum_{t=1}^{\tau_{RA}} \sum_{k=1}^{|\mathcal{S}_t|} \log_2 \left(1 + \gamma_{t,k}^{(b)} \right), \quad [\text{bps}] \quad (4)$$

The sum rate of SUCRe-XL can be defined as:

$$R_{\Sigma}^{SUCRe-XL} = \sum_{b \in \mathcal{V}_k} \sum_{t=1}^{\tau_{RA}} \log_2 \left(1 + \frac{\rho_k \beta_k^{(b)}}{\sigma^2} \right), \quad [\text{bps}] \quad (5)$$

assuming that just one user per pilot sequence is accepted.

IV. NUMERICAL RESULTS

It is assumed a 100 meter URA with $M = 500$ antennas in a $200 \times 100 \text{ m}^2$ cell with K inactive users (iUES) randomly distributed as illustrated in Fig. 1, a number of available pilots $\tau_{RA} = 10$, and normalized transmit power $\rho_k = 1\text{W}$. The main simulation parameters is listed in Table I. The scenario has simulated in Matlab 2019 utilizing one Intel HD Graphics 6000 GPU, Intel(R) Dual-Core(TM) I5 CPU @ 1.6 GHz and 8 GB RAM.

TABLE I
SIMULATION PARAMETERS.

Parameter	Value
User distance to array	$r_{k,m_z,m_y} \in [10; 180\text{m}]$
User length	$h_{user} \in [1; 1.7\text{m}]$
Antennas elements y-axis, z-axis	$M_y = 100, M_z = 5$
Number of BS antennas (URA)	$M = M_y \times M_z = 500$
Antenna element distance	$d_m = 1\text{m}$
Array height	$h_{array} = 12\text{m}$
Array length	$L_y = M_y \times d_m$
User transmit power	$\rho_k = 1\text{W}$
Noise Power	$\sigma^2 = 1\text{W}$
Number of iUES	$K \in [0; 5000]$
Access probability (1 st attempt)	$P_a = 0.01$
Access probability (new attempts)	$P_{na} = 0.5$
Number of available pilot sequences	$\tau_{RA} = 10$
SIC imperfection factor	$\varpi = 0.1$
Monte Carlo realizations	5000

A. Average Number of Access Attempts

Fig. 3 shows that the proposed protocol NVR-XL in crowded scenarios ($1000 < K < 5000$) achieves improved RA performance. The modification in collision resolution is that NVR-XL accepts two users colliding on the overlapping VR, thanks to the power diversity; therefore, the superposition signal can be processed by SIC in each SA. With the increase in the number of subarrays B , but constrained by the channel hardening and favorable propagation properties, the probability of collision in each subarray by sharing overlapping VRs in each SA increases. Such collision resolution configuration is impossible to treat with SUCRe-XL due to the absence of the interference cancellation step. As a result, the NVR-XL allows decreasing the number of access attempts even under solid interference in the same SA.

The normalized number of accepted UEs, calculated by the number of users accepted by the number of users trying to access, given in Fig. 4 corroborate with the improved results

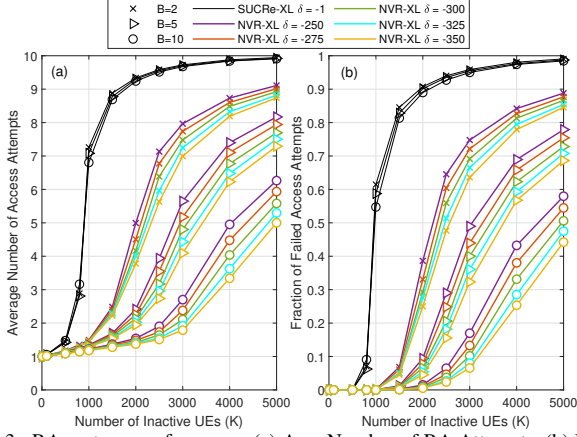


Fig. 3. RA systems performance. (a) Avg. Number of RA Attempts. (b) Prob. of failed access attempts.

in NVR-XL to accept UEs to transmit due the superposition signal can be processed by SIC in each SA.

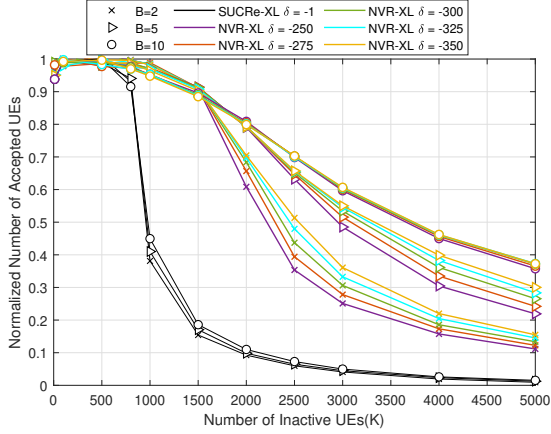


Fig. 4. Normalized number of accepted UEs.

B. Systems sum-rate

To adjust numerically the best value of scale factor, $\delta_{\text{NVR-XL}}^{K,B*}$, we chose the scenario with $B = 10$ SAs and deploys an exhaustive search approach to find the best δ that maximizes the sum-rate; we have found the interval $\delta_{\text{NVR-XL}}^{(K,10)*} = [-350 : -250]$. Fig. 5 reveals the sum-rate performance for both protocols; the SUCRe-XL requires the non-overlapping VR em each SA to accept the device; hence, presents a modest gain even increasing the number of SA. The highest sum rate result occurs before the saturation of usage of available pilots. On the other hand, the NVR-XL processes the signal superimposed in each SA, canceling the more robust user's signal, with a residual interference in order of $\varpi_1 = 0.1$, and achieving improved results when B increases. Also, a maximum sum rate limit is reached when a pair of users is processed in each SA in each RA attempt, as the SIC is able to cancel just one interfering signal, in the growing crowded scenario interference in each SA continues to increase (three or more users accessing the same SA with the same pilot).

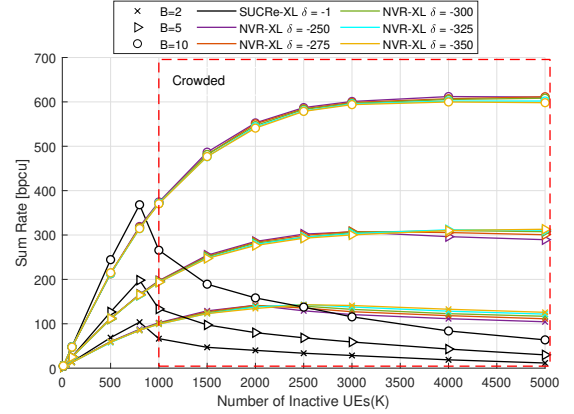


Fig. 5. Avg sum-rate as function of number of iUEs.

V. CONCLUSION

This work proposes to explore the power domain to improve the performance of SUCRe-XL grant-based RA protocol; indeed, the XL-MIMO adds a new degree of freedom with the received signal being processed by different SAs, together with a new degree of freedom introduced by NOMA, allows to resolve the collision of two users using the same pilot sequence, with application of a SIC in each SA at the BS side. Our numerical results reveal a substantial performance improvement of the proposed RA protocol regarding the recent literature SUCRe-XL method, attaining simultaneously a decreasing of the number of access attempts and a substantial increment in the sum-rate for the crowded scenarios analyzed ($1000 < K < 5000$).

REFERENCES

- [1] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroğlu, and S. M. Sait, "A survey of rate-optimal power domain noma with enabling technologies of future wireless networks," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2192–2235, 2020.
- [2] F. Clazzer, A. Munari, G. Liva, F. Lazaro, C. Stefanovic, and P. Popovski, "From 5g to 6g: Has the time for modern random access come?" *arXiv:1903.03063*, 2019.
- [3] M. V. da Silva, S. Montejo-Sánchez, R. D. Souza, H. Alves, and T. Abrão, "D2d assisted q-learning random access for noma-based mtc networks," *IEEE Access*, vol. 10, pp. 30 694–30 706, 2022.
- [4] E. Björnson, E. de Carvalho, J. H. Sørensen, E. G. Larsson, and P. Popovski, "A random access protocol for pilot allocation in crowded massive mimo systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2220–2234, 2017.
- [5] H. D. Pereira, G. Brante, J. Farhat, R. D. Souza, J. C. M. Filho, and T. Abrão, "On the sum-rate of contention resolution in massive mimo with noma," *IEEE Access*, vol. 9, pp. 24 965–24 974, 2021.
- [6] O. S. Nishimura, J. C. Marinello, and T. Abrão, "A grant-based random access protocol in extra-large massive mimo system," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2478–2482, 2020.
- [7] J. C. M. Filho, G. Brante, R. D. Souza, and T. Abrão, "Exploring the non-overlapping visibility regions in xl-mimo random access and scheduling," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.
- [8] E. D. Carvalho, A. Ali, A. Amiri, M. Angelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive mimo," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 74–80, 2020.
- [9] H. Lu and Y. Zeng, "Near-field modeling and performance analysis for multi-user extremely large-scale mimo communication," *IEEE Communications Letters*, vol. 26, no. 2, pp. 277–281, 2022.

ANEXO E – Improving Random Access with NOMA in mMTC XL-MIMO

Improving Random Access with NOMA in mMTC XL-MIMO

Thiago Augusto Bruza Alves
Department of Electrical Engineering
State University of Londrina (UEL).
86057-970, Londrina, PR, Brazil.
thiagobruza@hotmail.com

Taufik Abrão
Department of Electrical Engineering
State University of Londrina (UEL).
86057-970, Londrina, PR, Brazil.
taufik@uel.br

Abstract—The extra-large multiple-input multiple-output (XL-MIMO) architecture has been recognized as a technology for giving support for the massive MTC (mMTC), providing very high-data rates in high-user density scenarios. However, the large dimension of the array increases the Rayleigh distance (d_{Rayl}), in addition to obstacles and scatters causing spatial non-stationarities and distinct visibility regions (VRs) across the XL array extension. We investigate the random access (RA) problem in crowded XL-MIMO scenarios; the proposed grant-based random access (GB-RA) protocol combining the advantage of non-orthogonal multiple access (NOMA) and strongest user collision resolutions in extra-large arrays (SUCRe-XL) named NOMA-XL can allow access of two or three colliding users in the same XL sub-array (SA) selecting the same pilot sequence. The received signal processing in a SA basis changes the d_{Rayl} , enabling the far-field planar wavefront propagation condition, while improving the system performance. The proposed NOMA-XL GB-RA protocol is able to provide a reduction in the number of attempts to access the mMTC network while improving the average sum rate, as the number of SA increases.

Index Terms—NOMA; Machine Type Communication; XL-MIMO; Random Access Protocols; Crowded scenarios; Uniform Rectangular Array (URA).

I. INTRODUCTION

In B5G, the base station (BS) needs to simultaneously support devices with a variety of capabilities and deployments as the number of active devices continues to increase as well as access requests for multiple services, *e.g.*, vehicles, sensors, mobiles, etc., and applications in 5G and B5G, such as massive machine-type communication (mMTC) and crowded mobile broadband (cMBB) [1]. Hence, to support massive connections a promising candidate for the next generation of multiple access techniques, non-orthogonal multiple access (NOMA) has the ability to serve in the same resource element (RE), more than one device. In comparison to standard orthogonal multiple access (OMA), the NOMA increases the system throughput, improves user fairness, reduces latency, and enables large connections [2], sharing the same orthogonal resource (time and frequency) by coding superposition in transmitter side and successive interference cancellation (SIC) in receiver side [3].

A high number of devices attempting to connect to the network at the same time is a medium access control prob-

lem; standard random access (RA) systems are incapable of handling such a huge number of requests [4], and pure RA methods such as ALOHA have serious performance limitations. The adoption of NOMA and SIC considerably enhances performance, allowing two or more users per time slot [5], [6].

In (over-)crowded scenarios, the number of user connection attempts considerably outnumbers the number of available pilot sequences. As a result, establishing collision resolution techniques became crucial for enabling efficient communication. The strongest user collision resolution (SUCRe) protocol is a well-known decentralized grant-based random access (RA) protocol for crowded massive multiple-input multiple-output (MIMO) systems, which takes advantage of MIMO properties [7], giving preference to users with good channel conditions while harming edge users. Hence, the SUCRe protocol has undergone several evolutions; *e.g.*, in [8] a graph-based pilot access (SUCR-GBPA) protocol is proposed, enabling all users who lost contention resolution to choose a new pilot at random. Moreover, improving SUCRe by adding NOMA, the work in [9] compares the NOMA-RA protocol to the classical SUCRe scheme, allowing users who try to access the medium with the same pilot signal to resolve collisions by distinguishing them in the power domain and achieving superior sum rate performance with reduced average latency.

The large-scale MIMO technology (XL-MIMO) distributes the antennas elements in a wide space, such as facades of buildings, shopping malls, or stadiums; therefore, typically, the devices, scatterers, and obstacles are located inside the Rayleigh distance of the entire uniform rectangular array (URA) in Fig. 1, which in this case is formed by five SAs, resulting in different channel conditions, like *near-field propagation* and *spatial non-stationarity*. As a result, the received signal energy of a given UE varies significantly across the entire array; in addition, different regions of the array see different sets of scattered, obstacles, and UEs, based on the concept of visibility regions (VRs) [10].

In this work, we adopt the far-field model for XL-MIMO scenarios where the UE-array distance is typically higher than the Rayleigh distance, $d_{\text{UE}_k} > d_{\text{Rayl}}$, since the received signal processing is accomplished in a SA basis, reducing the array aperture D . The Rayleigh distance is quadratically proportional to the array aperture D as:

This work was supported in part by the CAPES (Financial Code 001) and the National Council for Scientific and Technological Development (CNPq) of Brazil under Grant 310681/2019-7.

$$d_{\text{Rayl}} = 2 \frac{D^2}{\lambda_c}, \quad (1)$$

where λ_c is the carrier wavelength. Hence, splitting the entire array in SAs changes considerable the limit of near-field and far-field, by changing the array aperture to $D_{\text{SA}} = \frac{D}{B}$, where B is the number of SAs. Therefore, users in the near-field area when considering the entire array can be seen on a SA basis as in far-field planar wave propagation.

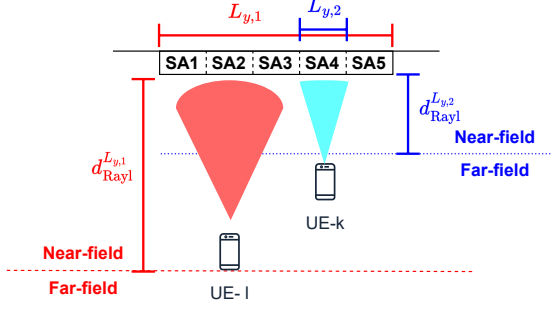


Fig. 1. URA XL-MIMO with a k -th user, illustrating how a change in array aperture D to D' (subarray SA_i) also changes the Rayleigh distance d_{Rayl} . In our random access XL-MIMO system, mMTC devices are distributed at random, located in a far-field area, i.e. $d_{\text{UE}_k} > d_{\text{Rayl}}$.

The non-stationarity can be exploited by applying the concept of VR, where an association with the geographic area of the cell and the visible portion of the array or SA, as demonstrated in [11] where SUCRe was adapted to the XL-MIMO context. With that, a certain UE is present in a region and has an associated VR due to the associated scatterers. For simplicity, we consider that an entire SA can be visible to the device, i.e., is part of its VR. Thus, the probability of the b -th SA being visible by the k -th UE can be exploited to improve the performance of the RA protocol in XL-MIMO.

In [12], a solution for the random access problem is addressed by jointly estimating the activity and the channels of UEs in XL-MIMO subjecting to non-stationarity in the far-field propagation scenarios; the effectiveness of the proposed algorithm is demonstrated in terms of the *normalized mean square error* (NMSE) and *activity error rate* (AER) performance indicators. Differently from [13] and [14], in this work, by optimizing the scale factor δ_{NOMA}^* , we admit more than two users (up to three users) colliding by the same pilot sequence under NOMA scheme, maximizing the sum-rate for different loading users and the number of SAs. Notice that herein, we treat and process the received signal at SA level; hence, by dividing the array into SA one can manipulate the Rayleigh distance d_{Rayl} , so users still experience the far-field condition.

The *Contribution* of this paper is twofold: **a)** a grant-based RA protocol is proposed, based on power-domain NOMA and by exploring the overlapping VR of XL-MIMO systems (NOMA-XL), which combines the benefits of XL-MIMO systems and NOMA, allowing until three of users selecting the same pilot sequence in at least one SA. **b)** the composition of the array into SAs, change the Rayleigh distance, allowing UEs to stay in the far-field area; in this condition, numerical

analyses are carried out taking into account the superimposed received signal processed in an SA basis, the optimal scale factor δ_{NOMA}^* , and performing SIC to identify the signal of each user.

II. SYSTEM MODEL

A time-division-duplexing (TDD) system is considered, in which a BS deploys an extra-large URA placed on the y - z plane with a total of M antenna elements, i.e., $M_y \times M_z$, where M_y and M_z denote the number of antenna elements along the y - and z -axis, respectively, with the first line of the array starting at the origin O and deployed along the ordinate axis. In the same way, as [11], [15], we consider a simplified bipartite graph model in XL-MIMO, the array is divided into B subarrays (SAs), each composed by a fixed number of $M_b = M/B$ antennas, ensuring the minimum antennas elements ($M_b \geq 50$).

UEs equipped with one antenna are randomly distributed in a square cell and distanced beyond the Rayleigh distance (d_{Rayl}) limit from the array, and grouped in a set $\mathcal{U}$, being $\mathcal{A} \subset \mathcal{U}$ the subset of active UEs, with temporarily dedicated data pilots, and $\mathcal{K} = \mathcal{U} \setminus \mathcal{A}$ be the set of inactive UEs (iUEs). The cardinality $|\mathcal{K}| = K$ represents the number of iUEs in the cell, and it has no dedicated pilot, in case the UE needs to become active, they must be assigned to one. The BS only allocates pilots to active devices and reclaims the pilots when needed. The iUEs try a RA attempt with probability P_a .

Let $\mathcal{M}$ denote the set of all BS SAs and $\mathcal{V}_k \subset \mathcal{M}$ the subset of SAs visible to the k -th UE. To compose the mathematical modeling, we assume that the subset $\mathcal{V}_k$ is generated randomly, where each SA is visible with probability P_b by UE k , like in [11], [15]. This probability simulates the influence of random obstacles and scatterers in the environment interacting with the signals transmitted by/to the UEs, resulting in VRs. We assume that each UE previously knows its VR $\mathcal{V}_k$, by deploying the same measurement procedure as in [16].

We define τ_{RA} as the number of RA mutually orthogonal pilot sequences that are possible $\mathbf{s}_1, \dots, \mathbf{s}_{\tau_{\text{RA}}} \in \mathbb{C}^{\tau_{\text{RA}}} \times 1$, each with length τ_{RA} , and $\|\mathbf{s}_{\tau_{\text{RA}}}\| = \sqrt{\tau_{\text{RA}}}$.

Since each SA visibility follows a Bernoulli distribution with success probability P_b , it is crucial to note that when an antenna is visible to a user, it indicates that the user may transmit/receive signals to/from this antenna with a non-zero channel gain [15]. We assume an overcrowded scenario where K iUEs ($K \gg M$) are randomly distributed within a square cell in front of the XL-MIMO entire array. Modifications in the array aperture, e.g. the SA, impact the Rayleigh distance, since the array aperture scale down from $D = L_y$ to $D_{\text{SA}} = \frac{L_y}{B}$. As a result, in terms of SA level, the XL-MIMO channel can be modeled as a far-field propagation considering typical application scenarios, as illustrated in Fig. 2.

A. Channel Model

We model the NLOS channel propagation, and the location of the elements follows a 3D representation [17], ensuring the mMTC devices are physically far away from the Rayleigh

distance. The physical dimension of the URA along the y - and z -axis are $L_y = (M_y)d_m$ and $L_z = (M_z)d_m$, where M_y and M_z denotes the number of antenna elements along y - and z -axis, and d_m is the distance between elements. We adopt the far-field planar wavefront propagation model based on the Rayleigh distance, which is directly linked to the array aperture as defined in eq. (1). Thus, we adopted the far-field model where the location of device k is $\mathbf{q}_k$ and the location of each URA antenna element $m_{z,y}$ is $\mathbf{w}_{m_{z,y}}$, hence, the distance of the k -th UE to the specific antenna element $m_{z,y}$ is determined by [17]:

$$r_{k,m_{z,y}} = \|\mathbf{w}_{m_{z,y}} - \mathbf{q}_k\| > d_{\text{Rayl}} \quad (2)$$

Considering an urban micro scenario, the large-scale fading model includes path loss and shadowing [1]:

$$\beta_{k,m_{z,y}} = 10^{-\kappa \log(r_{k,m_{z,y}}) + \frac{g+\varphi}{10}}, \quad (3)$$

where $g = -34.53$ dB is the path-loss at the reference distance, the path-loss exponent $\kappa = 3.8$, and $\varphi \sim \mathcal{N}(0, \sigma_{\text{sf}}^2)$ is the shadow fading, a log-normal random variable with standard deviation $\sigma_{\text{sf}} = 10$ dB. In same way as [11], [15], we adopt the average large scale fading for the k -th user at the b -th SA, $\beta_k^{(b)} = \frac{1}{M_b} \sum_{m=1}^{M_b} \beta_{k,m}$. Let $\mathbf{h}_k^{(b)} \in \mathbb{C}^{M_b \times 1}$ be the Rayleigh fading channel vector between UE $k \in \mathcal{K}$ and SA b , following $\mathbf{h}_k^{(b)} \sim \mathcal{CN}(0, \beta_k^{(b)} \mathbf{I}_{M_b})$.

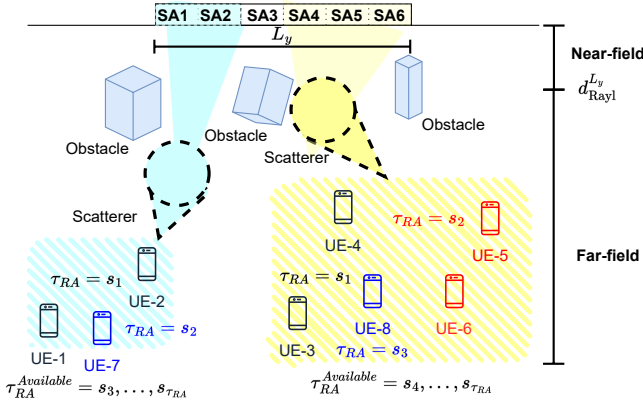


Fig. 2. Crowded MTC system with URA divided into a 6-SA; the colored regions in each SA define visible UEs in far-field area, P_b is visible success probability.

III. RANDOM ACCESS FOR NOMA XL-MIMO

The **SUCRe protocol** [7] is a decentralized RA method for resolving pilot collisions on the user side. The protocol consists of following steps: **a)** KP_a users randomly choose an uplink (UL) RA pilot and transmit it; **b)** the BS receives such UE signals and responds with precoded downlink (DL) pilots; **c)** Using the received signal, UEs can estimate the total of the signal strengths of UEs competing for the same RA pilot chosen by them, compare the such estimate to its signal gain, and retransmit the chosen RA pilot if it judges it is the strongest UE among the contenders; **d)** in conclusion, the BS allocates dedicated data pilots to the UEs not colliding in Step before. In mMTC scenarios, K is large compared to the number of antennas M ($K \gg M$), in contrast, $P_a \ll 1$.

A. Exploring VRs, SAs and NOMA in NOMA-XL Protocol

The proposed NOMA-XL protocol combines the improvement of non-overlapping VRs of specific UEs along the SAs operating under the SUCRe-XL protocol [11]. Besides, our proposed protocol can proceed with up to two or three users in overlapping VRs (in SA-based) selecting the same pilot and physically in a far-field area. Steps I and II are similar to the SUCRe protocol in [7].

Step I: KP_a UEs make an RA attempt.

Step II: SAs send a precoded DL pilot to UEs from Step I.

Step III: By exploring the power domain NOMA, users colliding pilots can be served at the same resource (SA, frequency, time, pilot). Due to imperfections in the SIC process, in this work, the proposed protocol can accept collisions of up to three users, *i.e.*, performing up to two SIC steps. However, in collisions between four or more users in at least one SA, the protocol could fail to resolve the collision, and unsuccessful UEs are instructed to try again on the next RA with probability P_{na} , limited to a maximum of 10 attempts.

The collision resolution follows the decentralized decision rule, *i.e.* each users decides whether to continue in access attempt; following the decision rule [11]:

$$\rho_k \sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \tau_{\text{RA}} > \frac{\hat{\alpha}_{t,k}}{2} + \epsilon_k, \quad (4)$$

where $\hat{\alpha}_{t,k}$ is the estimation of the sum of signal gains of contending UEs with its own $\rho_k \sum_{m \in \mathcal{V}_k} \beta_k^{(m)} \tau_{\text{RA}}$ by the same pilot τ_{RA} , and ϵ_k is the bias term. If inequality in Eq. (4) holds, the k -th user repeats its pilot transmission and forms the new set $\mathcal{S}_t$; otherwise, it remains in silence and tries to communicate in the next RA stage with probability P_{na} . The *bias term* is given by:

$$\epsilon_k = \frac{\delta_{\text{NOMA}}^*}{\sqrt{M_b} \cdot \sum_{b \in \mathcal{V}_k} \beta_k^{(b)}}, \quad (5)$$

where δ_{NOMA}^* is the optimal *scale factor* that should be adjusted to improve the average sum rate of the system.

The main feature of our NOMA-XL RA protocol is the modification in the contention resolution process proposed in [11]. The signal is processed in each SA to decode superimposed signals; hence, a SIC strategy is evoked to cancel and decode users' signals sharing the same pilot sequence. Without loss of generality, in the contention resolution step, the protocol accepts up to three UEs in the NOMA cluster assuming power disparities on a SA basis; hence, the BS proceeds with the SIC according to the descending order of channel gains ($\|\mathbf{h}_1\| > \|\mathbf{h}_2\| > \|\mathbf{h}_3\|$), *i.e.*, the strongest UE-1 signal is decoded first without SIC application, followed by UE-2 signal with one step of SIC to partially removed UE-1 signal and interference of UE-3 and in the sequel decodes the signal of UE3 after a two steps SIC to partially cancel signal of UE1 and UE2.

The signal-to-interference-plus-noise ratio (SINR) in each SA, selecting the same pilot sequence t is defined as [14]:

$$\gamma_{t,k}^{(b)} = \frac{p_k \beta_k^{(b)}}{\varpi_1 \sum_{j=1}^{k-1} p_j \beta_j^{(b)} + \sum_{i=k+1}^{|S_t|} p_i \beta_i^{(b)} + \sigma^2} \quad (6)$$

where ϖ_1 is the residual interference factor after applying SIC, assuming imperfect signal reconstruction.

Step IV: We consider that the BS perfectly estimates the channel to the user, or users, using the pilot signal sent in Step III. After that, the corresponding message is decoded. If the decoding is successful, the BS has identified users in set S_t and admits it to the payload coherence blocks by allocating a dedicated data pilot sequence. If the decoding fails, the protocol also fails to resolve the collision. The unsuccessful UE is instructed to try again after a random interval, limited to a maximum of 10 attempts. It stops sending the pilot sequence, and the packet is considered lost.

B. System Sum-rate

To compare the RA protocols, we define the sum-rate metric. In SUCRe-XL, only one user is accepted per pilot sequence in non-overlapping VR; on the other hand, in NOMA-XL, there will also be two or three users who consider themselves winners in the contender. The sum rate of NOMA-XL in SA basis can be defined using Eq. (6) as:

$$R_{\Sigma}^{\text{NOMA-XL}} = \sum_{b=1}^B \sum_{t=1}^{\tau_{\text{RA}}} \sum_{k=1}^{|S_t|} \log_2 \left(1 + \gamma_{t,k}^{(b)} \right), \quad [\text{bpcu}] \quad (7)$$

The sum rate of SUCRe-XL can be defined as:

$$R_{\Sigma}^{\text{SUCRe-XL}} = \sum_{b \in \mathcal{V}_k} \sum_{t=1}^{\tau_{\text{RA}}} \log_2 \left(1 + \frac{\rho_k \beta_k^{(b)}}{\sigma^2} \right), \quad [\text{bpcu}] \quad (8)$$

assuming that just one user per pilot sequence is accepted, and $\delta_{\text{SUCRe-XL}}^{(K,B)*} = -1$.

IV. NUMERICAL RESULTS

It is assumed a $L_y = 10$ meters URA with $M = 500$ antennas in a $100 \times 100 \text{ m}^2$ cell with K inactive users (iUES) randomly distributed in far-field area of the cell, and the number of available pilots is $\tau_{\text{RA}} = 10$ sequences. The normalized transmit power $\rho_k = 1\text{W}$, other simulation parameters are listed in Table I.

A. System sum rate

The optimal value of *scale factor*, $\delta_{\text{NOMA-XL}}^{(K,B)*}$ is obtained numerically, by exhaustive search approach, as in [9]; so, it is optimal in the sense of sum rate maximization; hence, each K -iUE and B -SA configuration results in an optimal parameter.

Fig. 3 a) shows the optimal *scale factor* parameter obtained numerically for each scenario of K and B by exhaustive search approach and it correspondent maximum NOMA-XL sum rate in Fig. 3 b), notice that the baseline (SUCRe-XL) requires the non-overlapping VR em each SA to accept the device; hence, presents a modest gain even increasing the number of SA. The highest sum rate values occur before the saturation of usage of available pilots. On the other hand, the NOMA-XL processes

the signal superimposed in each SA, canceling the more robust user's signal, with a SIC residual interference assumed in order of $\varpi_1 = 0.1$, and achieving improved results when B increases. Also, a maximum sum rate limit is reached when adjusting the *scale factor* δ , limiting the maximum number of users selecting the same pilot to three users, when the number of iUEs grows the *scale factor* decreases. This decay can be observed when the number of SA is lower, corroborating that the greater the interference of users, the smaller the δ factor.

TABLE I
SIMULATION PARAMETERS.

Parameter	Value
Antennas elements y-axis, z-axis	$M_y = 100, M_z = 5$
Number of BS antennas (URA)	$M = M_y \times M_z = 500$
Antenna element distance	$d_m = 0.1\text{m}$
Array aperture	$D = M_y \times d_m$
User transmit power	$\rho_k = 1\text{W}$
Noise Power	$\sigma^2 = 1\text{W}$
Number of iUEs	$K \in [0; 15000]$
Carrier wavelength	$\lambda_c = 0.125\text{m}$
Access probability (1 st attempt)	$P_a = 0.01$
Access probability (new attempts)	$P_{na} = 0.5$
Success probability of each SA being visible	$P_b = 0.5$
Number of available pilot sequences	$\tau_{\text{RA}} = 10$
SIC imperfection factor	$\varpi_1 = 0.1$
Monte Carlo realizations	5000
# SAs	Sub-Array Aperture, D_{SA}
$B = 1$	10m
$B = 5$	2m
$B = 10$	1m
	Rayleigh dist., d_{Rayl}
	1600m
	64m
	16m

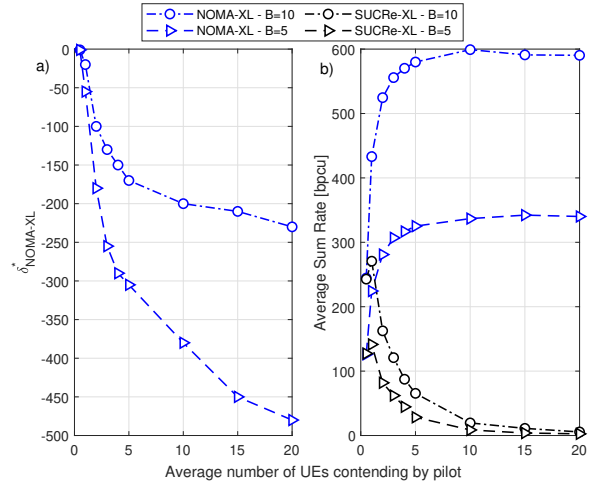


Fig. 3. **a)** Optimal *scale factor* for average sum-rate maximization for each $\{K; B\}$ scenario, obtained via exhaustive search. **b)** Average maximum sum-rate in each K and B scenario, attained after adjusting $\delta_{\text{NOMA-XL}}^*$.

B. Average Number of Access Attempts

Fig. 4 shows that the proposed protocol NOMA-XL in crowded scenarios ($1000 < K < 15000$) achieves improved RA performance. The modification in decentralized collision resolution is that NOMA-XL accepts two or three users colliding on the overlapping VR, thanks to the power diversity; therefore, the superposition signal can be processed after SIC process in each SA. With the increase in the number of SAs

B , but constrained by the channel hardening and favorable propagation properties, the probability of collision in each subarray by sharing overlapping VRs in each SA increases. Such collision resolution configuration is impossible to treat with SUCRe-XL due to the absence of the interference cancellation step. As a result, the NOMA-XL allows decreasing the number of access attempts even under solid interference in the same SA.

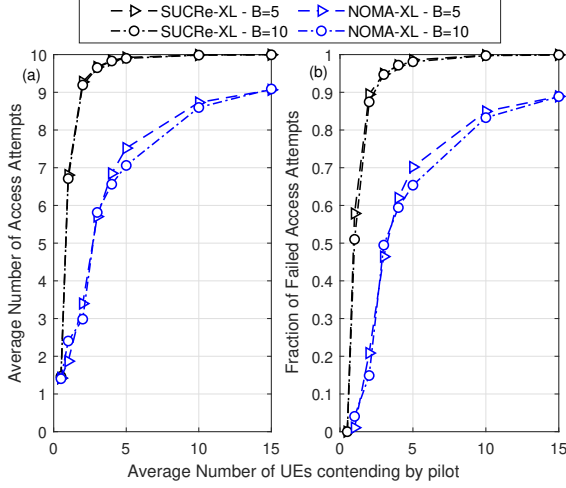


Fig. 4. RA systems performance. The Baseline SUCRe-XL and NOMA-XL. (a) Avg. Number of RA Attempts. (b) Prob. of failed access attempts.

The normalized number of accepted UEs, calculated by the number of users accepted by the number of users trying to access, given in Fig. 5 corroborate with the improved results in NOMA-XL to accept UEs to transmit due to the superposition signal can be processed by SIC in each SA.

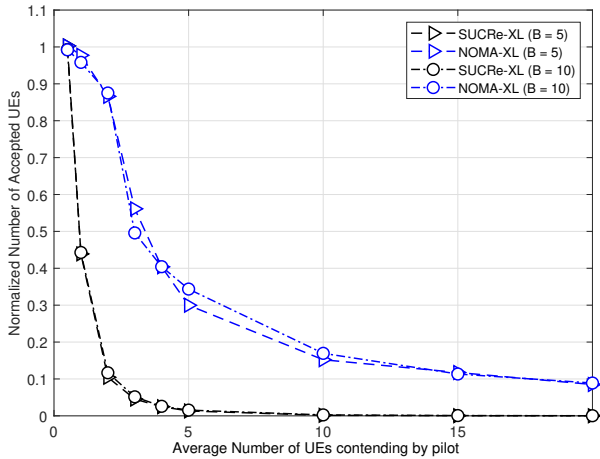


Fig. 5. Normalized number of accepted UEs, obtained with $\delta_{\text{NOMA-XL}}^*$.

V. CONCLUSION

By exploiting NOMA and signal processing in each SA the grant-based RA protocol performance can be improved. Indeed, the XL-MIMO adds a new degree of freedom with the received signal being processed by different SAs, together

with the degree of freedom introduced by power domain NOMA, allowing to resolve of the collision until three users using the same pilot sequence in the same VR subarray, and by applying the SIC stage in each SA at the BS side. We demonstrate that by processing the received signal on a SA basis can change the Rayleigh distance and place users in the far-field area. Anchored in our numerical results, we found a substantial improvement of the proposed RA protocol regarding the recent literature SUCRe-XL method, achieving both a reduction in the number of access attempts and a significant increase in the sum rate under crowded scenarios analyzed ($1000 < K < 15000$).

REFERENCES

- [1] M. Fallgren and B. Timus, *Deliverable D1.1: Scenarios, requirements and KPIs for 5G mobile and wireless system*. document ICT-317669-METIS, 2013.
- [2] A. Akbar, S. Jangsher, and F. A. Bhatti, "Noma and 5g emerging technologies: A survey on issues and solution techniques," *Computer Networks*, vol. 190, no. 107950, 2021.
- [3] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroğlu, and S. M. Sait, "A survey of rate-optimal power domain noma with enabling technologies of future wireless networks," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2192–2235, 2020.
- [4] F. Clazzer, A. Munari, G. Liva, F. Lazaro, C. Stefanovic, and P. Popovski, "From 5g to 6g: Has the time for modern random access come?," *arXiv:1903.03063*, 2019.
- [5] D. Jakovetić, D. Bajović, D. Vukobratović, and V. Crnojević, "Cooperative slotted aloha for multi-base station systems," *IEEE Transactions on Communications*, vol. 63, no. 4, pp. 1443–1456, 2015.
- [6] M. V. da Silva, S. Montejo-Sánchez, R. D. Souza, H. Alves, and T. Abrão, "D2d assisted q-learning random access for noma-based mtc networks," *IEEE Access*, vol. 10, pp. 30694–30706, 2022.
- [7] E. Björnson, E. de Carvalho, J. H. Sørensen, E. G. Larsson, and P. Popovski, "A random access protocol for pilot allocation in crowded massive mimo systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2220–2234, 2017.
- [8] H. Han, Y. Li, and X. Guo, "A graph-based random access protocol for crowded massive mimo systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7348–7361, 2017.
- [9] H. D. Pereira, G. Brante, J. Farhat, R. D. Souza, J. C. M. Filho, and T. Abrão, "On the sum-rate of contention resolution in massive mimo with noma," *IEEE Access*, vol. 9, pp. 24965–24974, 2021.
- [10] J. Flordelis, X. Li, O. Edfors, and F. Tufvesson, "Massive mimo extensions to the cost 2100 channel model: Modeling and validation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 380–394, 2020.
- [11] O. S. Nishimura, J. C. Marinello, and T. Abrão, "A grant-based random access protocol in extra-large massive mimo system," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2478–2482, 2020.
- [12] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, D. González G., and O. Gonsa, "Joint activity and channel estimation for extra-large mimo systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7253–7270, 2022.
- [13] T. A. Bruza Alves and T. Abrão, "Crowded mtc random access in noma xl-mimo," *2022 Symposium on Internet of Things (SIoT)*, pp. 1–4, 2022.
- [14] T. A. B. Alves and T. Abrão, "Noma-based random access in mmte xl-mimo," *IEEE Access*, vol. 11, pp. 1944–1954, 2023.
- [15] J. C. M. Filho, G. Brante, R. D. Souza, and T. Abrão, "Exploring the non-overlapping visibility regions in xl-mimo random access and scheduling," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.
- [16] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive mimo," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 74–80, 2020.
- [17] H. Lu and Y. Zeng, "Near-field modeling and performance analysis for multi-user extremely large-scale mimo communication," *IEEE Communications Letters*, vol. 26, no. 2, pp. 277–281, 2022.