Universidade Estadual de Londrina

GABRIEL AVANZI UBIALI

XL-MIMO FOR IMPROVED CONNECTIVITY: DETERMINISTIC SINR, USER SCHEDULING, PILOT ASSIGNMENT AND SUBARRAYS SELECTION

Londrina

2026

GABRIEL AVANZI UBIALI

XL-MIMO FOR IMPROVED CONNECTIVITY: DETERMINISTIC SINR, USER SCHEDULING, PILOT ASSIGNMENT AND SUBARRAYS SELECTION

Doctoral Thesis presented to the Doctorate Program of Universidade Estadual de Londrina for obtaining the degree of Doctor of Science in Electrical Engineering. Concentration area: Telecommunication Systems.

Supervisor: Dr. Taufik Abrão

Co-supervisor: Dr. José Carlos Marinello Filho

Londrina
2026

Ubiali, Gabriel Avanzi.

XL-MIMO Para Conectividade Melhorada: SINR Determinística, Agendamento de Usuários, Alocação de Pilotos e Seleção de Subarranjos / Gabriel Avanzi Ubiali. - Londrina, 2026. 206 p.

Orientador: Taufik Abrão

Coorientador: José Carlos Marinello Filho

Tese (Doutorado em Engenharia Elétrica) - Universidade Estadual de Londrina, Centro de Tecnologia e Urbanismo, Programa de Pós-Graduação em Engenharia Elétrica, 2026.

Inclui bibliografia.

1. Extra-large-scale antenna array. 2. User scheduling. 3. Pilot assignment. 4. Subarray selection. 5. Deterministic SINR. 6. Non-stationary channels. 7. Spectral efficiency. 8. Fairness. I. Abrão, Taufik. II. Marinello, José Carlos. III. Universidade Estadual de Londrina. Centro de Tecnologia e Urbanismo. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

GABRIEL AVANZI UBIALI

XL-MIMO FOR IMPROVED CONNECTIVITY: DETERMINISTIC SINR, USER SCHEDULING, PILOT ASSIGNMENT AND SUBARRAYS SELECTION

Doctoral Thesis presented to the Doctorate Program of Universidade Estadual de Londrina for obtaining the degree of Doctor of Science in Electrical Engineering. Concentration area: Telecommunication Systems.

EXAMINATION BOARD

Supervisor: Dr. Taufik Abrão
Universidade Estadual de Londrina

Dr. Richard Demo Souza
Federal University of Santa Catarina (UFSC), Florianópolis

Dr. Ramon Maia Borges
Federal University of Itajubá (UNIFEI)

Dr. Luis Carlos Mathias
Federal Technological University of Paraná (UTFPR), Toledo Campus

Dr. David William Marques Guerra
Federal University of Pernambuco (UFPE)

Londrina, January 19, 2026.

ACKNOWLEDGEMENTS

The completion of this doctoral work spanned a significant chapter of my life. It was a period marked by intense learning, the result of much effort and dedication, leading to substantial personal and professional growth and maturation. However, the conclusion of this journey was only possible thanks to the support of many people, to whom I hereby express my sincere acknowledgments.

I first lift my thanks to God, the author of life, the source of all grace, and the creator of all that exists, for enlightening and inspiring me at every moment of my life.

I am deeply and sincerely grateful to my parents, Rita and Wilson, who were essential throughout this stage. Thank you for your encouragement, unconditional support, daily assistance, prayers, affection, and for always cheering me on.

To my girlfriend, Vitória, with whom I have shared wonderful moments, I express my gratitude for her understanding and for bringing joy and balance to my days.

I thank my grandfather Nelson, who walked this academic path some decades before me, for his encouragement, sage advice, and for rooting for my success.

I am especially grateful to Professors Taufik and José Carlos for their supervision and co-supervision. Thank you for your trust and for the invaluable guidance, teachings, mentorship, and advice provided during the development of this work.

I am also grateful to the members of the examination committees—Professors David, Jaime, Luis, Ramon, and Richard—for their availability and for the insightful corrections, contributions, and suggestions offered during the qualification and defense stages.

I thank my laboratory colleagues for their companionship and daily interaction. You made this journey more enjoyable through our shared laughter and the exchange of ideas and knowledge.

Finally, I would like to thank all those who contributed, directly or indirectly, to the completion of this work.

'Pouco me importa que me digam que fulano é um bom filho meu — um bom cristão —, mas um mau sapateiro!' Se não se esforça por aprender bem o seu ofício ou por executá-lo com esmero, não poderá santificá-lo nem oferecê-lo ao Senhor. E a santificação do trabalho ordinário é como que o eixo da verdadeira espiritualidade para os que — imersos nas realidades temporais — estão decididos a ter uma vida de intimidade com Deus.

'What use is it telling me that so and so is a good son of mine — a good Christian — but a bad shoemaker?' If he doesn't try to learn his trade well, or doesn't give his full attention to it, he won't be able to sanctify it or offer it to Our Lord. The sanctification of ordinary work is, as it were, the hinge of true spirituality for people who, like us, have decided to come close to God while being at the same time fully involved in temporal affairs.

São Josemaria Escrivá de Balaguer
(Amigos de Deus, 61 / Friends of God, 61)

UBIALI, G. A.. **XL-MIMO Para Conectividade Melhorada: SINR Determinística, Agendamento de Usuários, Alocação de Pilotos e Seleção de Subarranjos**. 2026. 206f. Tese de Doutorado em Engenharia Elétrica – Universidade Estadual de Londrina, Londrina, 2026.

RESUMO

À medida que as redes sem fio evoluem para a sexta geração (6G), a demanda por conectividade melhorada de alta taxa e qualidade de serviço uniforme tornou-se primordial. Os sistemas de múltiplas entradas e múltiplas saídas extra-grandes (XL-MIMO, *eXtra-Large Multiple-Input Multiple-Output*) são um facilitador chave para essa visão, mas sua implementação prática é dificultada por características únicas da camada física e por uma imensa complexidade operacional. A motivação central para esta tese é a inadequação dos modelos de canal convencionais para XL-MIMO e a dificuldade de implementação de algoritmos de gerenciamento de recursos que dependem de informação de estado de canal (CSI, *Channel State Information*) instantânea. Para enfrentar esses desafios, esta tese primeiramente estabelece um modelo de canal XL-MIMO realista para cenários micro-urbanos que captura tanto a propagação com linha de visada (LoS, *Line-of-Sight*) dependente da distância quanto a propagação sem linha de visada (NLoS, *non-LoS*) com correlação espacial, e incorpora os conceitos de regiões de visibilidade e não-estacionaridades espaciais. Com base nesse fundamento, foram derivadas expressões determinísticas inéditas e em forma fechada para a razão sinal-interferência-mais-ruído (SINR, *Signal-to-Interference-plus-Noise Ratio*) no *uplink*, tanto em arquiteturas centralizadas quanto distribuídas. Válidas para combinadores baseados no erro quadrático médio mínimo (MMSE, *Minimum Mean Square Error*) e estimação de canal MMSE, essas expressões dependem apenas de estatísticas de longo prazo do canal, fornecendo uma alternativa tratável e eficiente a algoritmos de alocação de recursos computacionalmente dispendiosos. Para viabilizar a operação de rede escalável, é necessário solucionar três problemas interligados. Primeiro, o agendamento de usuários é importante em cenários de alta densidade, onde servir todos os usuários simultaneamente causaria interferência excessiva. Segundo, é necessário alocar pilotos de modo estratégico, visto que o seu inevitável reuso corrompe as estimativas de canal e limita a capacidade do sistema. Finalmente, a seleção de *subarrays* é fundamental tanto para a escalabilidade da rede quanto para a redução da complexidade, pois empregar todo o extenso arranjo de antenas para servir cada usuário agendado é computacionalmente dispendioso e também desnecessário, dado que cada usuário é tipicamente "visto" por apenas alguns *subarrays*. Para solucionar essas questões, as expressões de SINR determinísticas são utilizadas para projetar algoritmos eficientes, baseados em CSI estatístico, para o agendamento de usuários, alocação de pilotos e seleção de *subarrays*. Ao usar as expressões derivadas para estimar o desempenho do sistema, essas estratégias são projetadas para aumentar a equidade entre os usuários, maximizando a menor eficiência espectral (SE, *Spectral Efficiency*) dentre os usuários agendados, sem a sobrecarga do rastreamento de canal em tempo real. Resultados numéricos validam a alta precisão das aproximações de SINR propostas e a eficácia dos algoritmos de alocação de recursos, oferecendo um arcabouço abrangente e prático para a análise e alocação de recursos em redes sem fio da próxima geração.

Palavras-chave: MIMO de extra-larga escala, agendamento de usuários, alocação de pilotos, seleção de *subarrays*, SINR determinística, canais não-estacionários, eficiência espectral, equidade.

UBIALI, G. A.. **XL-MIMO For Improved Connectivity: Deterministic SINR, User Scheduling, Pilot Assignment and Subarrays Selection**. 2026. 206p. Thesis in Electrical Engineering – State University of Londrina, Londrina, 2026.

ABSTRACT

As wireless networks evolve towards its sixth-generation (6G), the demand for improved high-rate connectivity and uniform quality of service has become paramount. Extra-large multiple-input multiple-output (XL-MIMO) systems are a key enabler for this vision, yet their practical implementation is hindered by unique physical-layer characteristics and immense operational complexity. The central motivation for this work is the inadequacy of conventional channel models for XL-MIMO and the challenges associated with resource management algorithms that rely on instantaneous channel state information (CSI). To address these challenges, this thesis first establishes a realistic XL-MIMO channel model for micro-urban scenarios that captures the essential properties of distance-dependent line-of-sight (LoS) and spatially correlated non-LoS (NLoS) propagation, and incorporates the concepts of visibility regions (VRs) and spatial non-stationarities. Building on this foundation, novel closed-form deterministic expressions for the uplink signal-to-interference-plus-noise ratio (SINR) are derived for both centralized and distributed architectures. Valid for minimum mean square error (MMSE) receive combining and MMSE channel estimation, these expressions depend only on long-term channel statistics, providing a tractable and efficient alternative to computationally expensive resource allocation algorithms. To address the core challenge of scalable network operation, it is necessary to solve three intertwined problems. First, user scheduling is important in crowded scenarios where serving all users simultaneously would create unsustainable interference. Second, strategic pilot assignment is required because the inevitable reuse of pilots leads to pilot contamination, which corrupts channel estimates and limits system capacity. Finally, subarray selection is fundamental for both network scalability and complexity reduction, as employing the entire large-scale array to serve every scheduled user is computationally prohibitive and unnecessary, given that each user is typically "seen" by only a subset of subarrays. To solve these issues, the deterministic SINR expressions are leveraged to design efficient, statistical-CSI based algorithms for joint user scheduling, pilot assignment, and subarray selection. By using the derived expressions to estimate system performance, these strategies are designed to enhance user fairness by maximizing the minimum per-user spectral efficiency (SE), without the overhead of real-time channel tracking. Numerical results validate the high accuracy of the proposed SINR approximations and the effectiveness of the resource allocation algorithms, offering a comprehensive and practical framework for the analysis and resource allocation in next-generation wireless networks.

Keywords: Extra-large-scale antenna array, user scheduling, pilot assignment, subarray selection, deterministic SINR, non-stationary channels, spectral efficiency, fairness.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

5G	Fifth-generation
6G	Sixth-generation
AI	Artificial intelligence
AoA	Angle of arrival
AP	Access point
AWGN	Additive white Gaussian noise
BS	Base station
CDF	Cumulative density function
CoMP	Coordinated multipoint
C-RAN	Cloud radio access network
CPU	Central processing unit
CSI	Channel state information
DL	Downlink
DoA	Direction of arrival
DoD	Direction of departure
eMBB	Enhanced mobile broadband
ELAA	Extremely large aperture array
FDD	Frequency-division-duplexing
GA	Genetic algorithm
HD	Half-duplex
i.i.d.	Independent and identically distributed
ISAC	Integrated sensing and communication
KKT	Karush–Kuhn–Tucker
L-MMSE	Local MMSE

LoS	Line-of-sight
LS	Least-squares
LSF	Large-scale fading
LTE	Long term evolution
MIMO	Multiple-input multiple-output
MMSE	Minimum mean square error
mMTC	Massive machine type communications
MNAE	Mean NAE
MR	Maximum-ratio
MSE	Mean square error
NAE	Normalized absolute error
NLoS	Non-LoS
NMSE	Normalized mean square error
NUEW	Non-uniform evanescent wave
NUPW	Non-uniform plane wave
NUSW	Non-uniform spherical wave
PA	Pilot assignment
PDF	Probability density function
P-MMSE	Partial MMSE
PZF	Partial ZF
QoS	Quality of service
RF	Radio frequency
RIS	Reflective intelligent surface
RZF	Regularized ZF
SA	Subarray
SE	Spectral efficiency

SINR	Signal-to-interference-plus-noise ratio
SNR	Signal-to-noise ratio
SSF	Small-scale fading
TDD	Time-division-duplexing
UE	User equipment
ULA	Uniform linear array
UL	Uplink
UPA	Uniform planar array
UPBW	Uniform parabolic wave
UPD	Uniform power distance
UPW	Uniform plane wave
URLLC	Ultra-reliable and low latency communications
US	User scheduling
USW	Uniform spherical wave
VR	Visibility region
XL-MIMO	Extra-large scale MIMO
ZF	Zero-forcing

LIST OF SYMBOLS

$\mathbf{A}^{\mathrm{H}}$	Hermitian (conjugate transpose) of matrix $\mathbf{A}$
$\mathbf{A}^{\mathrm{T}}$	Transpose of matrix $\mathbf{A}$
c	Speed of light in vacuum
C	Number of scattering clusters
c_{iq}	Gain amplitude for scatterer (i, q) in cluster i
$\mathbf{C}_k$	Block-diagonal MMSE error covariance for UE k
$\mathbf{C}_{kl}$	Covariance of MMSE estimation error (UE k, subarray l)
d	Generic UE-array distance
d_{crit}	Critical distance (lower near-field)
d_{Fresnel}	Fresnel distance
d_{Rayleigh}	Rayleigh distance
d_{UPD}	Uniform-power distance
$d_{l,iq}$	Distance from the q-th scatterer in cluster i to the reference antenna of subarray l
$\dot{d}_{k,iq}$	Distance from scatterer (i, q) to reference antenna of UE k
$d_{lm,iq}$	Distance from scatterer (i, q) to the m-th subarray l antenna
$\dot{d}_{kn,iq}$	Distance from scatterer (i, q) to the n-th UE k antenna
$d_m(d, \varphi, \theta)$	Exact source-to-m-th antenna distance
$d_m^{\mathrm{parab}}(d, \varphi, \theta)$	Parabolic (2nd-order) approx. of propagation distance
$d_m^{\mathrm{plane}}(d, \varphi, \theta)$	Planar (1st-order) approx. of propagation distance
D	Subarray length or array aperture
D_{y}	Subarray's horizontal length
D_{z}	Subarray's vertical length
$\dot{D}$	UE's length

$\dot{D}_{\mathrm{y}}$	UE's horizontal length
$\dot{D}_{\mathrm{z}}$	UE's vertical length
$\mathcal{D}$	Set of subarray indices $\{1, \ldots, L\}$
$\mathcal{D}_k$	Set of subarrays serving UE k
$\mathbf{D}_k$	Block-diagonal matrix indicating antennas/subarrays serving UE k
$\mathbf{D}_{kl}$	Indicator matrix: $\mathbf{I}_N$ if subarray l serves UE k, $\mathbf{0}_N$ otherwise
$\mathrm{e}^x, \exp(x)$	Exponential function of x
$\mathbb{E}\{\cdot\}$	Expectation operator
$\mathbf{I}_M$	Identity matrix of order M
f	Carrier frequency
F_{kl}	Shadow fading (dB) between UE k and subarray l
$g_{kl,b}$	Gain of the b-th NLoS path
$\mathbf{h}_k$	Collective channel vector of UE k
$\overline{\mathbf{h}}_k$	LoS collective channel vector of UE k
$\check{\mathbf{h}}_k$	NLoS collective channel vector of UE k
$\widehat{\mathbf{h}}_k$	Collective channel estimate for UE k
$\widetilde{\mathbf{h}}_k$	Collective channel estimation error for UE k
$\mathbf{h}_{kl}$	Channel vector between UE k and subarray l
$\overline{\mathbf{h}}_{kl}$	LoS channel vector between UE k and subarray l
$\check{\mathbf{h}}_{kl}$	NLoS channel vector between UE k and subarray l
$\widehat{\mathbf{h}}_{kl}$	Channel estimate between UE k and subarray l
$\widetilde{\mathbf{h}}_{kl}$	Channel estimation error for UE k and subarray l
$\mathbf{H}_{kl}$	Channel matrix between subarray l and UE k
$\overline{\mathbf{H}}_{kl}$	LoS channel matrix between subarray l and UE k
$\check{\mathbf{H}}_{kl}$	NLoS channel matrix between subarray l and UE k
$\check{\mathbf{H}}_{kl,i}$	NLoS channel matrix between l, k, via cluster i

$\mathbf{H}_l$	Collection of the channel vectors at subarray l
$\widehat{\mathbf{H}}_l$	Collection of the channel estimate vectors at subarray l
$\widetilde{\mathbf{H}}_l$	Collection of the channel estimation error vectors at subarray l
$\mathbf{I}_N$	Identity matrix of order N
j	Imaginary unit: $j = \sqrt{-1}$
K	Number of UEs
$\mathcal{K}$	Set of UE indices $\{1, 2, \ldots, K\}$
$\mathbf{k}(\varphi, \theta)$	Unit vector direction based on azimuth and elevation
$\mathbf{k}_{kl}$	Unit vector pointing from the l-th subarray's reference antenna to the k-th UE's reference antenna
$\dot{\mathbf{k}}_{kl}$	Unit vector pointing from the k-th UE's reference antenna to the l-th subarray's reference antenna
L	Number of subarrays
L_k	Number of subarrays serving UE k
$\mathcal{L}$	Set of subarray indices $\{1, 2, \ldots, L\}$
M	Number of antennas in each subarray
M_{y}	Number of antennas per row in subarray UPA
M_{z}	Number of antennas per column in subarray UPA
m_0	Index of subarray reference antenna
$m_{\mathrm{y},0}$	Horizontal index of reference antenna in subarray
$m_{\mathrm{z},0}$	Vertical index of reference antenna in subarray
m_{y}	Horizontal index of the subarray's m-th antenna
m_{z}	Vertical index of the subarray's m-th antenna
MSE_k	Conditional mean squared error for UE k
N	Number of antennas in each UE
N_{y}	Number of antennas per row in UE UPA
N_{z}	Number of antennas per column in UE UPA

$\mathbf{n}_l^{\mathrm{ul}}$	AWGN noise vector at subarray l
n_0	Index of the UE's reference antenna
$n_{\mathrm{y},0}$	Horizontal index of the UE's reference antenna
$n_{\mathrm{z},0}$	Vertical index of the UE's reference antenna
n_{y}	Horizontal index of the UE's n-th antenna
n_{z}	Vertical index of the UE's n-th antenna
$\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$	Circularly symmetric complex Gaussian distribution
NMSE_{il}	Normalized MSE for UE i at subarray l
NMSE_i	Aggregate (user-level) NMSE for UE i
$\mathrm{NMSE}_{\mathrm{th}}$	NMSE threshold for user scheduling
$\mathbf{p}_{l,iq}$	Scatterer (i, q) position, subarray l frame
$\dot{\mathbf{p}}_{k,iq}$	Scatterer (i, q) position, UE k frame
$\mathcal{P}_k$	Set of UEs sharing the pilot with UE k
$\mathcal{P}_t$	Set of UEs using pilot sequence t
Q_i	Number of scatterers in cluster i
$\mathbf{Q}_{kl}$	Correlation matrix of the channel (UE k, subarray l)
$\mathbf{Q}_k$	Collective channel correlation for UE k
$\mathbf{r}_m$	Position of the m-th antenna of a subarray, relative to its reference antenna
$\mathbf{r}_0$	Position of the subarray's reference antenna
$\mathbf{R}_{kl}$	Covariance matrix of the channel (UE k, subarray l)
$\mathbf{R}_k$	Block-diagonal channel covariance for UE k
$\mathbf{s}_n$	Position of the n-th antenna of a UE, relative to its reference antenna
$\mathbf{s}_0$	Position of the UE's reference antenna, in its local coordinate system
$\mathbf{s}_{kl}$	Position of the k-th UE's reference antenna, relative to the l-th subarray's reference antenna
$\mathbf{s}_{kln}$	Position of the n-th antenna of the k UE w.r.t. subarray l's reference

η_k	Spectral efficiency of UE k
η_{th}	Minimum required per-user SE (threshold)
$\eta_{\min}$	Minimum SE among scheduled users
$\eta_{\min}^{(t)}$	Minimum SE if pilot t is assigned to candidate user
t_k	Index of the pilot sequence assigned to UE k
$\mathbf{T}_{kl}$	Rotation matrix of the k-th UE, relative to the l-th subarray
U	Number of scheduled UEs
$\mathcal{U}$	Set of scheduled UEs
$\mathcal{U}'$	Tentative scheduled user set during algorithm iterations
$\mathcal{U}_l$	Set of UEs served by subarray l
$\mathbf{v}_{kl}$	Uplink combining vector at subarray l for UE k
$\mathbf{v}_k$	Uplink collective combining vector for UE k
$\mathbf{V}_l$	Collection of the combining vectors at subarray l
$\mathbf{W}_l$	Interference-plus-noise covariance at subarray l (distributed)
x_k	Symbol transmitted by UE k
$\widehat{x}_k$	Final estimate of x_k
$\widehat{x}_{kl}$	Local estimate of x_k at subarray l
$\mathcal{X}_{kl}$	Shadow fading factor
$\mathbf{y}_l^{\mathrm{ul}}$	Uplink data at subarray l
$\mathbf{y}_l^{\mathrm{pilot}}$	Uplink pilot at subarray l
$\mathbf{Z}_k$	Interference-plus-noise covariance for UE k (centralized)
$\overline{\mathbf{Z}}_k$	Expectation of $\mathbf{Z}_k$
$\mathbf{Z}_{kl}$	Interference-plus-noise covariance for UE k at subarray l (distributed)
$\overline{\mathbf{Z}}_{kl}$	Expectation of $\mathbf{Z}_{kl}$
α_{kl}	Amplitude, UE k to subarray l
$\tilde{\alpha}_m$	Relative amplitude error for m-th antenna

$\tilde{\alpha}^*$	Maximum relative amplitude error across antennas
β_k	Average large-scale fading (LSF) channel gain of UE k
β_{kl}	LSF channel gain for LoS path
β_0	Channel gain at reference distance 1 m
Γ_k	Global SINR for UE k
$\hat{\Gamma}_k$	Approximated global SINR (distributed)
$\hat{\Gamma}_k^*$	Maximum global SINR w/ optimal weights
Γ_{kl}	Local SINR, subarray l, UE k
γ	NLoS path-loss exponent
δ	Decorrelation distance for shadow fading
δ_{ij}	Kronecker delta: 1 if $i = j$, 0 if $i \neq j$
Δ	Generic antenna spacing
Δ_{y}	Antenna spacing in horizontal dimension (subarray)
Δ_{z}	Antenna spacing in vertical dimension (subarray)
$\dot{\Delta}_{\mathrm{y}}$	Antenna spacing in horizontal dimension (UE)
$\dot{\Delta}_{\mathrm{z}}$	Antenna spacing in vertical dimension (UE)
η_k	SE of UE k
η_k^{erg}	Ergodic SE approx. (centralized op.)
η_k^{asy}	Asymptotic SE approx. (centralized op.)
$\hat{\eta}_k$	Approximate instantaneous SE (distributed op.)
$\hat{\eta}_k^{\mathrm{erg}}$	Ergodic SE approx. (distributed op.)
θ_{kl}	Elevation angle from UE k to subarray l
$\overline{\theta}_{kl}$	Elevation for LoS path
$\dot{\theta}_{kl}$	Elevation angle from subarray l to UE k
λ	Wavelength
λ_k	Lagrange multiplier (weight optimization for UE k)

μ_{kl}	Weight for combining subarray l estimates (distributed, UE k)
σ_{n}^2	AWGN noise variance
σ_{SF}	Shadow fading standard deviation
σ_{φ}	Azimuth angular std. dev. (NLoS)
σ_{θ}	Elevation angular std. dev. (NLoS)
τ_{p}	Number (length) of pilot sequences
Υ	Power variation ratio between the strongest and weakest antenna of the array
$\boldsymbol{\phi}_{t_k}$	Pilot sequence for UE k
ϕ_{kl}	Phase shift, UE k to subarray l
$\tilde{\phi}_m^{\mathrm{UPBW}}$	Phase error (UPBW approx), m-th antenna
$\tilde{\phi}_m^{\mathrm{UPW}}$	Phase error (UPW approx), m-th antenna
φ_{kl}	Azimuth angle from UE k to subarray l
$\overline{\varphi}_{kl}$	Azimuth for LoS path
$\dot{\varphi}_{kl}$	Azimuth angle from subarray l to UE k
$\boldsymbol{\Psi}_{t_k l}$	Covariance of pilot signal at subarray l
$\boldsymbol{\Psi}_{tl}$	Pilot signal covariance at subarray l for pilot t
ψ_{iq}	Phase of scatterer (i, q) in cluster i
ω_{kl}	Bernoulli r.v., LoS: whether UE k visible to subarray l
$\omega_{kl,i}$	Bernoulli r.v., LoS via cluster i
$\mathbf{0}_N$	Zero vector of size N
$\lvert\mathcal{A}\rvert$	Cardinality of set $\mathcal{A}$
$\lVert\cdot\rVert$	Euclidean (ℓ_2) norm of a vector

CONTENTS

1 INTRODUCTION

1.1 Contextualization

Since their widespread deployment began in 2019, fifth-generation (5G) wireless networks have been engineered to support demanding application scenarios such as enhanced mobile broadband (eMBB), massive machine type communications (mMTC), and ultra-reliable and low latency communications (URLLC). To meet these demands, 5G was designed with ambitious performance targets, including peak data rates of 10 Gb/s for uplink and 20 Gb/s for downlink, latency as low as 1 ms, and 99.999% reliability [1–4]. A cornerstone for achieving these high data rates is an evolution of multiple-input multiple output (MIMO) technology, known as massive MIMO (mMIMO). In the cellular mMIMO paradigm, each base station (BS) is equipped with a large antenna array—typically comprising tens or hundreds of elements—to capture a vast number of channel observations. This enhances micro-diversity, array gain, and spatial separability, making it a key enabler for 5G performance [5–7].[1]

In cellular massive MIMO, UEs located near a BS — within the cell-center area — achieve high data rates, which is mainly due to the strong received signal resulting from the short propagation distance and the minimal impact of pilot contamination. Typically, pilots are not reused among UEs within the same cell, and when pilot reuse is unavoidable, the BS strategically assigns pilots to reduce interference. Conversely, UEs at the cell edge often endure poor quality of service (QoS) as a result of low signal-to-noise ratio (SNR) and significant inter-cell interference.[2]

The sixth-generation (6G) wireless networks, expected to address communication requirements beyond 2030, have ignited substantial research interest due to emerging demands driven by the rapid evolution of wireless applications. Compared with 5G, 6G networks are projected to deliver a 100-fold increase in peak data rates (reaching terabit-per-second levels), a tenfold reduction in latency, and a reliability of 99.99999% [1, 10–12], in order to achieve immersive communication (eMBB+), massive communication (mMTC+) and hyper reliable and low-latency communication (URLLC+), which are expansions of the usage scenarios defined in 5G, and to comport new scenarios that will flourish in the new era of 6G, such as integrated sensing and communication (ISAC), integrated artificial intelligence (AI) and communication, and improved connectivity [13].

[1] Spatial separability is the system's ability to discriminate the desired signal from interfering signals in the uplink, and to transmit signals in a highly directional way in the downlink.

[2] Inter-cell interference comes mainly from UEs in the cell-edge of adjacent cells. Since each BS typically only estimates the channels of the UEs within its own cell, it cannot effectively mitigate interference caused by the UEs in neighboring cells. This problem is exacerbated by pilot contamination, which arises when pilot sequences are reused in adjacent cells [8, 9].

To meet these extraordinarily high performance targets, several promising technologies are garnering significant attention, such as extra-large scale MIMO (XL-MIMO) [1,13,14], reflective intelligent surface (RIS), and Terahertz-band communications.

XL-MIMO is a natural evolution of mMIMO technology, where the number of antennas is increased by at least an order of magnitude — up to several hundreds or even thousands [13, 15]. This expansion vastly enhances spectral efficiency (SE) and spatial resolution, which is critical for eMBB+ applications like augmented, virtual and mixed reality, as well as holographic displays [13, 16]. The extremely large antenna array also significantly benefits mMTC+ and high-precision localization and sensing [13,16]. Similar terminology include extremely large aperture array (ELAA) [15], ultra-massive MIMO (UM-MIMO) [17] and extremely large aperture massive MIMO (xMaMIMO) [18].

XL-MIMO can be deployed in four main architectures: collocated, sparse, modular, and distributed [13]. These architectures are discussed in more detail in Appendix E.

XL-MIMO panels have a versatile deployment, and can be integrated into walls and columns of a stadium or shopping center [19], into facades or indoor walls, and can support a wide range of future wireless use cases, such as [13]:

- **Immersive eMBB+ Experiences:** The massive throughput enables ultra high definition video, augmented, virtual and mixed reality, and holographic displays with true multi-sensory immersion.
- **Hotspot Coverage Enhancement:** The substantial increase in capacity supports ultra crowded scenarios, such as sports stadiums, concerts, or transit hubs, by boosting coverage and user data rates.
- **Ultra-Dense internet of things (IoT) and Sensing:** The fine spatial resolution of XL-MIMO empowers highly reliable, dense connectivity for smart homes, wearables, agriculture, and factory automation, as well as high-accuracy localization and navigation for robots. It also enables precise wireless sensing and object tracking, benefiting applications like advanced driver assistance and autonomous vehicles.

1.2 Motivation

This section outlines the key motivations for the choices made in this thesis:

1. Adopting an XL-MIMO architecture to exploit ultra-large apertures for enhanced spatial resolution and more uniform QoS;
2. Developing and utilizing a realistic XL-MIMO channel model that captures spherical wave propagation, non-stationarities, and correlated Rician fading;

3. Partitioning the antenna array into subarrays and performing local processing to improve scalability and reduce fronthaul load;

4. Dynamically selecting a subset of subarrays to serve each user;

5. Selecting which users will be served at a time to manage interference, enhance spectral efficiency and QoS, and guarantee network scalabity;

6. Designing a pilot-assignment scheme that minimizes interference among simultaneously served users;

7. Deriving closed-form deterministic signal-to-interference-plus-noise ratio (SINR) expressions to enable tractable performance analysis under a robust channel model.

1.2.1 XL-MIMO For Improved Connectivity And Huge Data Rates

Although current cellular massive MIMO networks can deliver high peak data rates due to high micro-diversity enabled by a massive number of antennas, the QoS often varies considerably between cell-center and cell-edge UEs. As wireless connectivity underpins services like payments, navigation, and entertainment, future networks must ensure consistently high data rates across the entire coverage area, rather than focusing solely on peak performance.

In this context, XL-MIMO emerges as a key enabler of improved connectivity, not merely because of its potential for a 10-fold increase in peak data rates compared to massive MIMO, but because of its ability to provide more uniform service quality. Its primary advantage lies in improving the experience of users with weaker connections, significantly increasing the likelihood of achieving acceptable data rates with high reliability (e.g., a 95% success probability) [20].

This enhanced performance stems from the extremely large array aperture, which boosts macro-diversity and ensures that each UE is likely to be close to at least part of the antenna array. This proximity reduces the average propagation distance to the nearest BS antenna [9,15,21], thereby not only increasing peak data rates but also equalizing QoS across users.

1.2.2 Realistic Channel Model For XL-MIMO

Extremely large array apertures give rise to propagation characteristics that differ significantly from those observed in conventional massive MIMO, such as spherical wavefronts, spatial non-stationarities, and a high probability of line-of-sight (LoS) links.

Near-field propagation and spherical wavefronts: In conventional massive MIMO, due to the small to moderate array apertures, the receivers are usually in the far-field region, where the electromagnetic wave can be simply modeled based on the plane wave assumption [14, 22], and all antennas embrace the same signal amplitude and angle of arrival (AoA). In XL-MIMO, due to the large array apertures, the propagation distances and angles between transmitter and receiver vary over the array [23]. In other words, the receiver is usually located in the near-field region, and the electromagnetic wave should be modeled based on the spherical wave assumption [1, 22–24].

LoS component: In XL-MIMO, since the UEs are probably located in close proximity to at least some antennas, the LoS component often contributes a dominant and deterministic portion of the received power. Thus, the Rayleigh fading, typically used to describe massive MIMO channels, is not accurate as an XL-MIMO channel model. Neglecting either LoS or non-LoS (NLoS) components leads to underestimating SNR and mispredicting performance. Several works have demonstrated that including the LoS component in the channel model improves channel estimation accuracy and system performance metrics [25, 26].

Spatial non-stationarities: Conventional massive MIMO presents spatially stationary channels, meaning that the channel gain is approximately uniform across all antennas. In contrast, the extremely large array apertures in XL-MIMO systems give rise to spatially non-wide-sense stationary channel characteristics [1, 18, 21, 27–31]. This occurs because different antennas may receive signals from distinct propagation paths or from the same paths but with different power levels, due to variations in the propagation distance across the array [1, 21]. Consequently, each UE is typically visible only to a specific portion of the array, and its transmitted power is predominantly captured within that region. This subset of the array that effectively "sees" a given UE is referred to as its visibility region (VR) [1,18,27,32,33]. While qualitatively described here as the portion of the array with a viable propagation path to the UE, this concept is mathematically formalized in Chapter 2.

Spatially correlated NLoS component: Due to the limited number of multipath components and the proximity between adjacent antennas, the conventional assumption of spatially uncorrelated channels is usually unrealistic. Moreover, numerical results in [34] demonstrate that spatial correlation can enhance the SE by improving channel estimation quality and enhancing favorable propagation. This occurs because spatial correlation increases the orthogonality between user channels and reduces multi-user interference. As noted in [8, Sec. 4.1], the primary benefit of strong spatial channel correlation is the reduction of interference between users whose spatial correlation matrices are sufficiently different.

An accurate channel model for XL-MIMO accounting for these propagation characteristics is introduced in Chapter 2 and employed throughout this work deriving deterministic SINR expressions and obtaining numerical results from simulations.

1.2.3 Local Processing (Distributed Operation)

As the number of antennas at the BS increases largely in XL-MIMO systems, the signal processing may become computationally prohibitive. To address this challenge and make spatially non-stationary XL-MIMO systems more tractable, a practical approach is to partition the array into multiple subarrays. Each subarray comprises a small to moderate number of antennas and typically serves only a subset of users located within its visibility region.

The most advanced uplink architecture is a fully centralized implementation, where each subarray simply forwards its received signals during both uplink pilot and data transmissions to a central processing unit (CPU) [9, Sec. 5.1]. The CPU then performs channel estimation and data detection based on the aggregated signals from all subarrays. This centralized approach achieves the highest SE by enabling joint interference suppression across subarrays [35]. However, its practical deployment is limited by the high fronthaul bandwidth and processing requirements.

To reduce these burdens, a more scalable alternative is distributed processing, in which each subarray performs signal processing tasks—such as channel estimation, receive combining, and transmit precoding—locally. Since each subarray can easily be equipped with a local baseband processor that is at least as powerful as those in the UEs, we are able to integrate new subarrays into the network without having to upgrade the CPUs [9, Sec. 5.2], enhancing scalability and implementation flexibility.

1.2.4 Subarrays Selection

Despite the clear advantages of XL-MIMO over conventional massive MIMO, realizing a practically feasible architecture remains daunting. A scalable implementation requires the complexity and resource requirements of signal processing to be finite as the number of users grows to infinity [9], which demands that each subarray serves a finite number of UEs. Network scalability is addressed in more details in Appendix H.

Using the entire antenna array to serve all UEs is both impractical and unnecessary, as each UE is typically in proximity to only a subset of the subarrays. Selecting only some subarrays to serve the UE inevitably leads to some degradation in service quality, but the performance loss is negligible if all the subarrays that can reach the UE with a non-negligible signal power are included in the serving set. Subarray selection offers significant practical advantages. First, it reduces fronthaul signaling, as only the selected subarrays must transmit the downlink data intended for the UE and send the estimates of the uplink data to the CPU. Second, it lowers computational complexity, since each subarray serves only a limited number of UEs [9, 36, 37].

1.2.5 User Scheduling

User scheduling refers to selecting a subset of UEs to be served within a given time slot or resource block. In crowded scenarios, it is infeasible to serve all UEs simultaneously due to constraints such as finite transmit power, a restricted number of resource blocks, and the limited spatial degrees of freedom imposed by the finite number of antennas. Additionally, the performance of linear precoding or combining schemes degrades when too many UEs are scheduled simultaneously, as channel orthogonality deteriorates. As a result, meeting QoS requirements (such as minimum data rate) under a limited power budget necessitates careful user scheduling [38].

Efficient user scheduling plays a key role in meeting QoS demands, including [39]: a) interference management: selecting UEs that can be jointly served with minimal mutual interference; b) computational efficiency: lowering the dimensionality of precoding and detection matrices by serving fewer UEs at a time; c) throughput enhancement: prioritizing UEs with favorable channel conditions to boost sum-rate; d) QoS and fairness assurance: balancing resource allocation to meet individual rate constraints and ensure fair access.

1.2.6 Pilot Assignment

Accurate channel estimation is critical for achieving high SE. Ideally, each UE is assigned a unique pilot sequence from an orthogonal set. However, in densely populated or high-mobility scenarios, the overhead associated with pilot transmission can significantly reduce the effective SE. When the coherence time is short, assigning mutually orthogonal pilots to all UEs is infeasible [9], and pilot reuse becomes necessary. As a result, interference-free channel estimation is no longer achievable, and pilot contamination becomes a limiting factor [40].

In such conditions, efficient pilot assignment is important to mitigate the resulting coherent interference. In XL-MIMO systems, however, the impact of pilot reuse can be less severe, since different UEs are typically served by distinct subsets of subarrays, and each UE is usually associated with the subarrays exhibiting the strongest channel gains.

1.2.7 Statistical CSI-Based Algorithms

In XL-MIMO systems, the design of algorithms for subarray selection, user scheduling, or pilot assignment based on instantaneous channel state information (CSI) is computationally prohibitive due to the extremely large number of antennas and users [41]. A more scalable approach is to rely on statistical CSI, which varies much more slowly over time. This enables algorithmic decisions to remain valid across multiple coherence blocks, significantly reducing computational complexity while preserving system performance.

1.2.8 Deterministic SINR Expressions

Closed-form deterministic SINR approximations offer an efficient alternative to Monte Carlo simulations by relying solely on channel statistics, such as the mean channel vectors and the covariance matrices, which vary much more slowly than the instantaneous channels. Since they depend only on large-scale fading (LSF), which typically changes only when user positions vary, these expressions require infrequent updates, significantly reducing computational complexity and fronthaul signaling overhead.

This enables resource allocation and scheduling algorithms to operate over many channel coherence blocks with a single evaluation, streamlining implementation in practical systems. Additionally, deterministic SINR formulations provide valuable analytical insights into system behavior, such as the impact of the XL-MIMO geometry, pilot contamination, estimation errors, and power control strategies. Moreover, due to the channel hardening effect in XL-MIMO, detection based purely on statistical knowledge becomes nearly optimal, further reinforcing the practical utility of deterministic expressions in realistic deployments.

Deterministic SINR expressions have been extensively investigated in the context of multi-user massive MIMO, XL-MIMO, and cell-free massive MIMO. However, the majority of them were derived based on spatially uncorrelated Rayleigh fading and maximum-ratio (MR) processing. In contrast, realistic XL-MIMO deployments in micro-urban environments are expected to experience both LoS and NLoS propagation conditions, which is well captured by the Rician fading channel model. Furthermore, minimum mean square error (MMSE)-based combining schemes perform better than MR irrespective of the level of cooperation among the subarrays and the number of antennas used at each one, which is the main conclusion of [35].

Deterministic SINR expressions for spatially uncorrelated Rician fading channels have been derived in [42–48]. Specifically, [42] considers uplink transmission with perfect CSI and zero-forcing (ZF) combining, while [43–48] derive SINR expressions for MR combining, accounting for channel estimation errors, *i.e.*, imperfect CSI. Additionally, [46] provides a deterministic expression valid for ZF combining, while [47] extends the analysis to include partial ZF (PZF) and partial MMSE (P-MMSE) combining schemes. However, all these studies assume spatially uncorrelated channels.

Deterministic SINR expressions under spatially correlated Rayleigh fading have been derived in [41, 49, 50] for perfect CSI and in [51–57] for imperfect CSI. However, most of them are valid only for MR processing. Since MR is known to perform poorly in interference-limited regimes, these assumptions significantly limit the applicability of the results in practical dense deployments. Furthermore, all of the expressions derived in these papers ignore the LoS channel component. Regarding to other processing schemes,

[41] also derives a deterministic SINR expression for regularized ZF (RZF) combining, assuming perfect CSI, while [57] also derives deterministic expressions for ZF and RZF under imperfect CSI. Deterministic SINR expressions under spatially correlated Rayleigh fading with VRs are derived in [50] for MR and ZF precoding, considering perfect channel estimation.

Finally, deterministic SINR expressions have been derived in [34, 48, 58–61] for spatially correlated Rician fading, accounting for channel estimation errors. Specifically, [34,48,58–60] works with cell-free massive MIMO, while [61] deals with downlink multicell massive MIMO. However, the derived expressions are only valid for MR processing, except [61], which considers RZF precoding. Nonetheless, the expression in [61] is not fully closed-form, as its evaluation requires solving a system of coupled fixed-point equations through numerical iteration, although the authors demonstrate existence and uniqueness of the solution.

Thus, this thesis presents novel closed-form deterministic SINR expressions for spatially correlated Rician fading channels, under imperfect CSI, where channel estimates are obtained via MMSE channel estimation, and signal detection is performed using MMSE combining.

1.3 Objectives

The general objective of this thesis is to develop strategies for achieving improved access to high data rate services in the context of 5G and 6G wireless networks employing XL-MIMO architectures. The specific objectives are:

- Conduct a comprehensive literature review on XL-MIMO, with particular emphasis on the derivation and applicability of deterministic SINR expressions.
- Develop a realistic channel model for XL-MIMO systems that captures spatial non-stationarities and includes both LoS and spatially correlated NLoS components.
- Derive accurate closed-form deterministic uplink SINR expressions for centralized and distributed XL-MIMO architectures under spatially correlated Rician fading, assuming MMSE receive combining and imperfect CSI obtained via MMSE channel estimation. Such expressions are currently not available in the existing literature.
- Design user scheduling, pilot assignment, and subarray selection strategies aimed at maximizing either the sum SE or the minimum per-user SE, by leveraging the derived deterministic SINR expressions as optimization metrics.
- Provide extensive numerical results to validate the accuracy of the proposed expressions and to demonstrate the performance benefits of the developed algorithms.

1.4 Contributions

Building on the motivations and objectives outlined above, the principal contributions of this thesis are:

1. **Derivation of novel deterministic SINR expressions for correlated Rician fading and MMSE combining:** Closed-form deterministic SINR expressions are derived for both centralized and distributed uplink operation modes in XL-MIMO systems. These expressions are deterministic, which means they do not depend on instantaneous CSI. Since they depend on statistical CSI, they are valid across multiple channel coherence blocks. Hence, they require computation only once per statistical channel state, significantly reducing complexity. Despite their statistical nature, numerical results confirm that the derived expressions closely approximate the true SINR. These expressions are original in that they account for spatially correlated Rician fading, MMSE receive combining, and imperfect CSI obtained via MMSE channel estimation.

2. **Design of efficient algorithms for joint subarray selection, user scheduling and pilot assignment in XL-MIMO:** User scheduling and pilot assignment algorithms are proposed to maximize the sum SE or the minimum SE in centralized and distributed uplink XL-MIMO scenarios. The algorithms rely on the derived deterministic SINR expressions, enabling decision-making based on long-term channel statistics. As a result, the algorithms remain valid across many coherence blocks, avoiding the need for per-block recomputation and thus offering significantly lower complexity compared to solutions based on instantaneous SINR. Furthermore, subarray selection strategies are integrated to ensure scalability and computational efficiency with negligible performance loss. Numerical evaluations demonstrate that the proposed methods closely match the performance of instantaneous-SINR-based approaches.

3. **Development of a realistic XL-MIMO channel model:** A robust channel model is proposed for XL-MIMO networks, capturing spatial non-stationarities and incorporating hybrid propagation composed of LoS and spatially correlated NLoS channels (correlated Rician fading).

1.5 Achievements (Scientific Papers)

This section outlines the produced papers regarding the contributions above.

1.5.1 Paper A

The paper "Improving spectral efficiency via pilot assignment and subarray selection under realistic XL-MIMO channels," was published in *Physical Communication* Journal (IF: 2.0), vol. 58, June 2023, doi: 10.1016/j.phycom.2023.102062 (Appendix A).

This paper introduces a robust channel model for XL-MIMO, which combines a deterministic LoS component with a spatially correlated Rayleigh fading component, corresponding to the NLoS propagation paths. This channel model incorporates the concepts of spatial non-stationarity and user-specific VRs. Building upon this model, the paper proposes a novel quasi-optimal and low-complexity iterative pilot assignment (PA) strategy based on genetic algorithm (GA). The GA-based approach improves QoS uniformity by minimizing the channel estimation normalized mean square error (NMSE) across UEs during pilot allocation. Then, a subarray selection procedure is applied to ensure that each subarray serves only a limited number of UEs, preserving system scalability without compromising QoS. Since UEs are generally visible to only a subset of subarrays, subarray selection maintains performance while reducing complexity. The subarray clusters are informed by the PA outcome and the resulting NMSE. Simulations confirm that the proposed PA strategy significantly improves the minimum per-user SE compared to random and greedy baseline algorithms.

1.5.2 Paper B

The paper "User Scheduling in Multi-State Los/NLoS Crowded XL-MIMO Channels," was published in the *proceedings of the 2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Istanbul, Turkiye, 2023, pp. 1-6, doi: 10.1109/WINCOM59760.2023.10322996 (Appendix B).

This paper introduces a novel asymptotic SINR expression for centralized uplink operation in XL-MIMO systems, considering a channel model composed of LoS and spatially correlated NLoS components. The derived expression is deterministic, which means it depends solely on channel statistics instead of on the instantaneous channels, and is valid under least-squares (LS) channel estimation, zero pilot reuse, and ZF receive combining. In addition, a user scheduling algorithm is proposed to determine the optimal number of active users that maximizes the system's sum SE. Numerical results demonstrate the accuracy of the proposed SINR expression, particularly for large amounts of antennas, and confirm the effectiveness of the scheduling algorithm.

1.5.3 Paper C

The paper "User Scheduling in Crowded XL-MIMO Under Multi-State LoS/NLoS Channels" was published in the *Journal of Network and Systems Management* (IF: 4.1), vol. 33, July 2025, doi: 10.1007/s10922-025-09961-w (Appendix C).

This paper also investigates deterministic SINR expressions for centralized uplink operation of XL-MIMO systems under correlated Rician fading. Unlike Paper B, it considers MMSE channel estimation, MMSE receive combining, and subarray selection, where each user is served by a subset of the available subarrays. The asymptotic SINR expression from Paper B is extended to this more general scenario. Numerical results confirm the accuracy of the new asymptotic expression in the absence of pilot reuse and when the number of antennas serving each user is sufficiently large.

To handle crowded scenarios—where the number of users surpasses the number of antennas allocated per user—a novel ergodic SINR expression is derived, which remains valid irrespective of the presence or absence of pilot reuse. Results show that the proposed expression yields accurate approximations in regimes where the asymptotic expression becomes unreliable, and vice versa, highlighting their complementary applicability across different operating conditions. Additionally, simplified forms of both expressions are provided to reduce computational complexity.

A joint user scheduling and pilot assignment algorithm is also proposed to select the set of users to be served and assign pilot sequences, aiming to maximize the system's sum SE. Since the algorithm relies on the deterministic SINR expressions, its decisions depend only on long-term channel statistics, eliminating the need for frequent recomputation across coherence blocks. Numerical results confirm the algorithm's effectiveness.

1.5.4 Paper D

This full paper was submitted to *Physical Communication* Journal and is available in Appendix D.

The paper investigates deterministic SINR expressions—one asymptotic and one ergodic—for centralized and distributed uplink operation in XL-MIMO systems under correlated Rician fading. The analysis considers MMSE channel estimation, local MMSE (L-MMSE) receive combining, and subarray selection. Distributed operation is more tractable and offers lower computational complexity than its centralized counterpart. The global signal estimate is formed as a weighted sum of the local estimates produced by the subarrays, and the optimal weights are derived in this paper to maximize the global SINR.

In contrast to Papers B and C, the asymptotic expression derived here is valid even in the presence of pilot reuse. Numerical results demonstrate that the proposed asymptotic and ergodic expressions offer complementary accuracy: each is reliable in operating regimes where the other tends to lose precision, depending on the number of antennas per subarray and on the number of scheduled users.

A joint user scheduling and pilot assignment algorithm is also proposed to select the set of users to be served and assign pilot sequences, aiming to maximize the system's minimum SE. Since the algorithm estimates the SE based on the derived deterministic SINR expressions, its decisions rely only on long-term channel statistics, avoiding frequent recomputation across coherence blocks. The algorithm automatically switches between the asymptotic and the ergodic SINR expression, based on the number of scheduled UEs, ensuring that the SE estimation is accurate. Numerical results validate the effectiveness of the proposed approach.

1.6 Thesis Structure

The remainder of this thesis is organized into three technical chapters, followed by concluding remarks that synthesize the findings and outline directions for future research.

Chapter 2 investigates propagation modeling for XL-MIMO systems operating in micro-urban scenarios, where both the transmitter and receiver employ uniform planar arrays. It begins with a rigorous analysis of electromagnetic field regions, revealing the limitations of traditional planar and parabolic wave approximations when applied to large-aperture arrays. A hybrid channel model is then developed, combining LoS and NLoS components. Within this framework, three wave models are examined: non-uniform spherical wave (NUSW), UPBW, and UPW. Comparative analyses quantify the accuracy of these approximations in terms of phase and amplitude errors. The insights and formulations presented in this chapter provide the physical foundation for the analytical developments that follow.

Chapter 3 derives novel deterministic SINR approximations for both centralized and distributed uplink operations. The analysis assumes Rician fading, MMSE channel estimation, and MMSE receive combining, advancing beyond existing models limited to Rayleigh channels or ideal CSI. Closed-form expressions for ergodic and asymptotic SINR approximations are developed for each operation mode, and their respective accuracy are evaluated in different system dimensions. The chapter further examines the impact of subarray selection, combining weights, and processing strategy on the SE.

Chapter 4 proposes and evaluates statistical-CSI-based algorithms for joint subarray selection, user scheduling, and pilot assignment in XL-MIMO systems. The algorithms are explicitly designed to exploit the deterministic SINR formulations developed in Chapter 3. Numerical simulations confirm that the proposed schemes significantly enhance SE and user fairness under practical constraints such as finite pilot resources.

Finally, Chapter 5 summarizes the main contributions, consolidates theoretical and simulation insights, and discusses potential research extensions.

2 XL-MIMO CHANNEL MODEL

This chapter investigates the channel characteristics of a wireless communication system comprising K multi-antenna UEs and a BS equipped with an XL-MIMO array, partitioned into L subarrays. Each subarray is connected to a CPU, which coordinates subarray cooperation. The antenna array of each subarray and each UE lies in the local y-z plane, where the z-axis denotes the vertical (height) direction and the y-axis spans horizontally. First, the array geometry adopted throughout this chapter is described, since all subsequent channel formulations build upon this configuration.

As depicted by Figure 2.1, the whole BS antenna array is divided into L planar subarrays, with M_{y} antennas in the horizontal direction (y-axis) and M_{z} antennas in the vertical direction (z-axis), such that $M = M_{\mathrm{y}}M_{\mathrm{z}}$ is the number of antennas in each subarray. The subarray length is D_{y} and D_{z} in the horizontal and vertical dimensions, respectively. Then, the array's largest physical dimension is $D = \sqrt{D_{\mathrm{y}}^2 + D_{\mathrm{z}}^2}$.

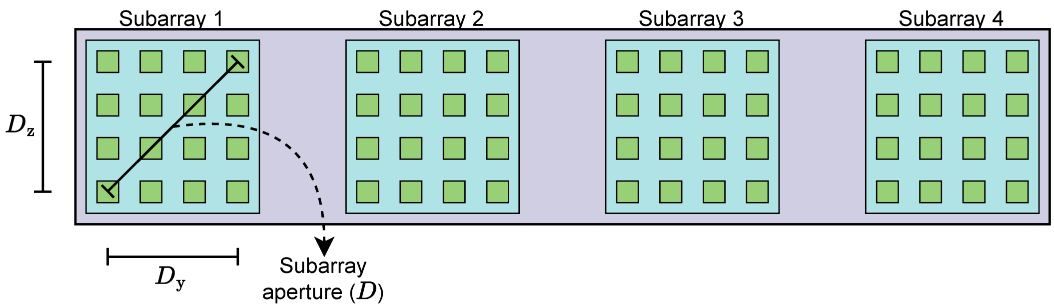


Figure 2.1 – Antenna array containing 4 subarrays; each subarray is a UPA with $M = 16$ antennas ($M_{\mathrm{y}} = M_{\mathrm{z}} = 4$).

The spacings between adjacent elements are Δ_{y} in the horizontal and Δ_{z} in the vertical direction, so that $D_{\mathrm{y}} = (M_{\mathrm{y}} - 1)\Delta_{\mathrm{y}}$ and $D_{\mathrm{z}} = (M_{\mathrm{z}} - 1)\Delta_{\mathrm{z}}$.

One of the most central antennas is selected as the subarray's reference antenna, and its horizontal and vertical indices are given by

$$m_{\mathrm{y},0} = \left\lceil \frac{M_{\mathrm{y}}}{2} \right\rceil \quad \text{and} \quad m_{\mathrm{z},0} = \left\lceil \frac{M_{\mathrm{z}}}{2} \right\rceil . \tag{2.1}$$

The position of the reference antenna, given by $\mathbf{r}_0 = [0, 0, 0]^{\mathrm{T}}$, is adopted as the origin of the subarray's local coordinate system. We can also refer to any of the M antennas using a unique index, by counting the antennas row by row. The unique index of the subarray's reference antenna is denoted by m_0, which can be obtained based on the horizontal and vertical indices as $m_0 = m_{\mathrm{y},0} + (m_{\mathrm{z},0} - 1)M_{\mathrm{y}}$. Similarly, the index of an arbitrary antenna $m \in \{1, \ldots, M\}$ is given by $m = m_\mathrm{y} + (m_\mathrm{z} - 1)M_\mathrm{y}$, where

$$m_\mathrm{y} = \mathrm{mod}(m-1, M_\mathrm{y}) + 1 \quad \text{and} \quad m_\mathrm{z} = \left\lceil \frac{m}{M_\mathrm{y}} \right\rceil, \tag{2.2}$$

are its horizontal and vertical indices. The position vector of the m-th antenna, within the local coordinate system, is expressed by

$$\mathbf{r}_m = \begin{bmatrix} 0 \\ (m_\mathrm{y} - m_{\mathrm{y},0})\Delta_\mathrm{y} \\ (m_\mathrm{z} - m_{\mathrm{z},0})\Delta_\mathrm{z} \end{bmatrix}. \tag{2.3}$$

Figure 2.2 illustrates a planar array geometry comprising $M = 12$ antennas, with $M_\mathrm{y} = 4$ and $M_\mathrm{z} = 3$. Each square represents an antenna, labeled by its unique index m. The y and z axes display the respective horizontal and vertical positions of each element, relative to the local coordinate system, while the outer axes indicate the two-dimensional indices of each antenna element. The reference antenna is at the origin, selected according to the centrality criterion given in (2.1). In this case, $m_0 = 6$, $m_{\mathrm{y},0} = 2$, and $m_{\mathrm{z},0} = 2$.

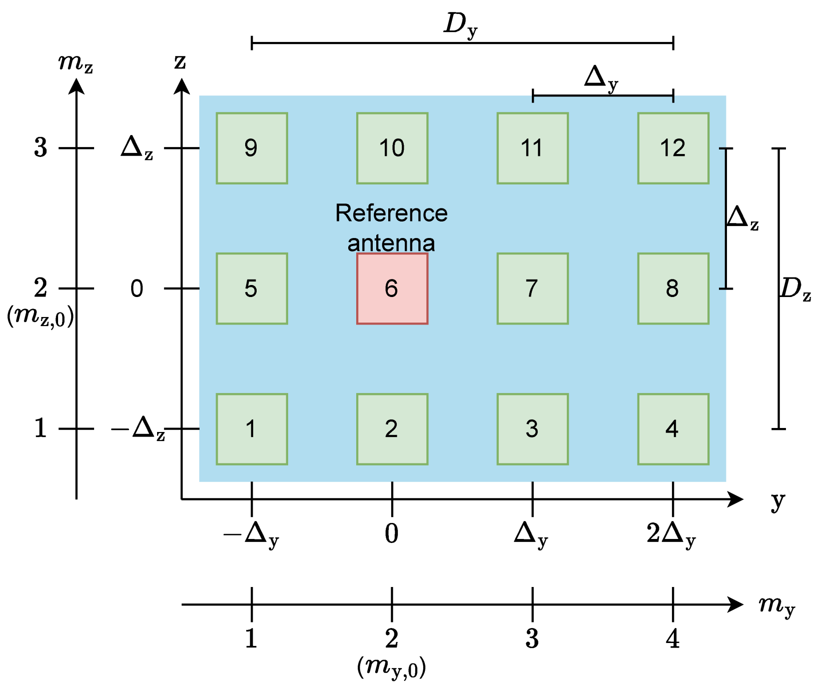


Figure 2.2 – Planar array geometry with reference element centralized for local coordinate system alignment.

Each UE is equipped with a uniform planar array (UPA) of $N = N_\mathrm{y}N_\mathrm{z}$ antennas, arranged in a grid with N_y elements along the horizontal y-axis of its local coordinate system and N_z elements along the vertical z-axis. The antenna spacings are $\dot{\Delta}_\mathrm{y}$ and $\dot{\Delta}_\mathrm{z}$ in the horizontal and vertical dimensions, respectively, so that the sides of the square array are $\dot{D}_\mathrm{y} = (N_\mathrm{y} - 1)\dot{\Delta}_\mathrm{y}$ and $\dot{D}_\mathrm{z} = (N_\mathrm{z} - 1)\dot{\Delta}_\mathrm{z}$, respectively. The aperture of the UE's antenna array is then $\dot{D} = \sqrt{\dot{\Delta}_\mathrm{y}^2 + \dot{\Delta}_\mathrm{z}^2}$.

The UE's reference antenna, whose index is denoted by n_0, is located at $\mathbf{s}_0 = [0,0,0]^{\mathrm{T}}$, in the UE's local coordinate system. The horizontal and vertical indices of this antenna are[1]

$$n_{\mathrm{y},0} = \left\lceil \frac{N_{\mathrm{y}}}{2} \right\rceil \quad \text{and} \quad n_{\mathrm{z},0} = \left\lceil \frac{N_{\mathrm{z}}}{2} \right\rceil . \tag{2.4}$$

The 2D indices for the n-th antenna of an arbitrary UE, where $n \in \{1,\ldots,N\}$, are given by

$$n_{\mathrm{y}} = \mathrm{mod}(n-1, N_{\mathrm{y}}) + 1 \quad \text{and} \quad n_{\mathrm{z}} = \left\lceil \frac{n}{N_{\mathrm{y}}} \right\rceil , \tag{2.5}$$

and its position in the UE's local coordinate system is

$$\mathbf{s}_n = \begin{bmatrix} 0 \\ (n_{\mathrm{y}} - n_{\mathrm{y},0})\dot{\Delta}_{\mathrm{y}} \\ (n_{\mathrm{z}} - n_{\mathrm{z},0})\dot{\Delta}_{\mathrm{z}} \end{bmatrix} . \tag{2.6}$$

The position of the k-th UE's reference antenna, relative to the l-th subarray's reference antenna, is given by

$$\mathbf{s}_{kl} = d_{kl}\mathbf{k}_{kl}, \tag{2.7}$$

where $d_{kl} = \|\mathbf{s}_{kl}\|$ is the distance between them, $\mathbf{k}_{kl} = \mathbf{k}(\varphi_{kl}, \theta_{kl})$ is the unit vector pointing from the l-th subarray's reference antenna to the k-th UE's reference antenna, φ_{kl} and θ_{kl} are respectively the azimuth and elevation angles of the incident wave at the subarray's reference antenna,[2] and

$$\mathbf{k}(\varphi, \theta) = \begin{bmatrix} \cos(\varphi)\cos(\theta) \\ \sin(\varphi)\cos(\theta) \\ \sin(\theta) \end{bmatrix} . \tag{2.8}$$

The position of the n-th antenna of the k-th UE, relative to the l-th subarray's reference antenna, is given by

$$\mathbf{s}_{kln} = \mathbf{s}_{kl} + \mathbf{T}_{kl}\mathbf{s}_n, \tag{2.9}$$

where $\mathbf{T}_{kl} \in \mathbb{R}^{3\times 3}$ is the rotation matrix of the k-th UE, relative to the l-th subarray.[3]

The unitary vector pointing from the k-th UE's reference antenna to the l-th subarray's reference antenna, in the UE's local coordinate system, is $\dot{\mathbf{k}}_{kl} = \mathbf{k}(\dot{\varphi}_{kl}, \dot{\theta}_{kl})$, being $\dot{\varphi}_{kl}$ and $\dot{\theta}_{kl}$ the azimuth and elevation angles of departure, respectively, *i.e.*, the angles of the signal that comes from the k-th UE's reference antenna to the l-th subarray's reference antenna, with respect to the k-th UE's local coordinate system. The representation of $\dot{\mathbf{k}}_{kl}$

---

[1] Actually, the indices of the subarrays's and UE's reference antennas are m_0 and n_0, respectively. However, for notation simplicity, $\mathbf{r}_0$ and $\mathbf{s}_0$ are used here instead of $\mathbf{r}_{m_0}$ and $\mathbf{s}_{n_0}$.

[2] The azimuth angle is measured from the x axis counterclockwise to the y-axis, while the elevation angle is measured from the xy-plane to the z-axis, counterclockwise.

[3] The rotation matrix is obtained from the UE's rotation angles (e.g., yaw, pitch, and roll), which are periodically estimated by onboard inertial sensors and orientation modules in mobile devices.

in the coordinate system of the l-th subarray is $\mathbf{T}_{kl}\dot{\mathbf{k}}_{kl}$. Since the receiver-to-transmitter direction ($\mathbf{k}_{kl}$) is the opposite of the transmitter-to-receiver direction ($\mathbf{T}_{kl}\dot{\mathbf{k}}_{kl}$), their relationship is

$$\mathbf{k}_{kl} = -\mathbf{T}_{kl}\dot{\mathbf{k}}_{kl}. \tag{2.10}$$

Thus, the angles of departure can be computed based on the angles of arrival and the rotation matrix using the relationship $\mathbf{k}(\dot{\varphi}_{kl}, \dot{\theta}_{kl}) = -\mathbf{T}_{kl}^{\mathrm{T}}\mathbf{k}(\varphi_{kl}, \theta_{kl})$.[4]

Although the channel model derived in this chapter is general and supports UEs with multiple antennas ($N > 1$), the performance analysis, SINR derivations, and resource allocation algorithms presented in subsequent chapters (Chapters 3 and 4) will focus on the specific case of single-antenna UEs ($N = 1$). The extension of these analyses to the multi-antenna case requires additional considerations regarding precoding at the UE side and is outlined as a perspective for future work in Chapter 5.

2.1 Field Regions and Distance Boundaries

Some distinctive channel effects emerge in XL-MIMO that are absent in conventional massive MIMO. As the array aperture grows, the near-field region expands and the standard UPW assumption no longer holds over the entire structure; instead, each UE–BS link must be modeled via spherical wavefronts. A given UE may be close to some antennas and distant from others, inducing substantial variation in propagation distances, signal amplitude, and angular parameters across the array—giving rise to spatial non-stationarities.

Propagation in XL-MIMO systems can be categorized into distinct regions and boundaries based on electromagnetic wave characteristics, which differ significantly from traditional massive MIMO systems due to the substantial array sizes involved.

The complex-valued channel response models how the propagation environment between a transmitter and a receiver affects the signal amplitude and phase. The channel response shows different features in different field regions.

2.1.1 Field Regions Based on Phase Variations

Based on the phase characteristics of the electromagnetic waves, the space around an antenna array is conventionally divided into three distinct regions: the reactive near-field, the radiative near-field (Fresnel region), and the far-field (Fraunhofer region) [1, 13, 23]. As illustrated in Figure 2.3, each region exhibits unique wave behavior. Closest to the array is the reactive near-field, where evanescent fields dominate and energy is primarily stored rather than radiated. Beyond this, the radiative near-field is characterized by

[4] The inverse of a rotation matrix is its transpose, *i.e.*, $\mathbf{T}_{kl}^{-1} = \mathbf{T}_{kl}^{\mathrm{T}}$.

spherical wavefront and significant variation in amplitude across the array. The interface between the reactive and radiative near-fields is defined by the Fresnel distance. In the far-field, wavefronts become approximately planar. The boundary between the near-field and far-field is set by the Rayleigh distance, also known as the Fraunhofer distance.

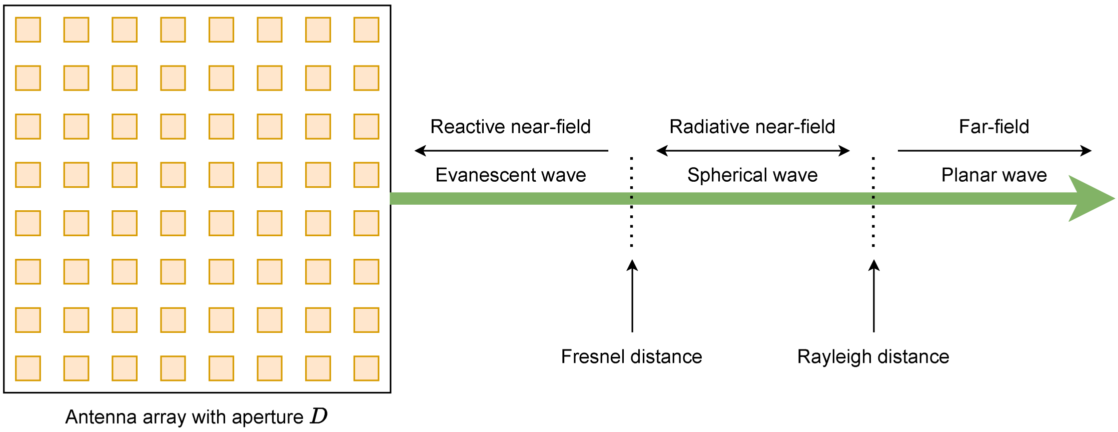


Figure 2.3 – Reactive near-field, radiative near-field, and far-field, separated by the Rayleigh and Fresnel distances.

2.1.1.1 Far-Field Region And Rayleigh Distance

Consider the propagation between two antenna arrays, for instance, an arbitrary subarray at the BS and an arbitrary UE. In the far-field region, electromagnetic waves incident on the array appear nearly planar, resulting in approximately linear phase variations across the array elements. This planar-wave condition is satisfied at distances beyond the Rayleigh distance.

The field regions and distance boundaries are typically discussed from the BS side, in an uplink scenario where UEs transmit towards the BS array. This is because the BS array may range from small to extremely large apertures, whereas UE arrays are usually very small or even contain a single antenna. Assuming that the UE's aperture is considerably smaller than the subarray's aperture, the Rayleigh distance is defined as

$$d_{\text{Rayleigh}} = \frac{2D^2}{\lambda} = \frac{2D^2 f}{c}, \tag{2.11}$$

where $\lambda = c/f$ is the wavelength, f is the carrier frequency, and c is the speed of light in vacuum.

In the particular case of a UPA with N_{y} and N_{z} antennas spaced horizontal and vertically by a multiple of half-wavelengths, *i.e.*, $\Delta_{\text{y}} = I_{\text{y}}\frac{\lambda}{2}$ and $\Delta_{\text{z}} = I_{\text{z}}\frac{\lambda}{2}$, the Rayleigh distance is given by[5]

$$d_{\text{Rayleigh}} = \frac{c[(M_{\text{y}}-1)^2 I_{\text{y}}^2 + (M_{\text{z}}-1)^2 I_{\text{z}}^2)]}{2f}. \tag{2.12}$$

According to (2.11), the Rayleigh distance increases quadratically with the array aperture. According to (2.12), the Rayleigh distance also increases quadratically with antenna number and element spacing. For fixed D, it increases linearly with carrier frequency. On the other hand, for fixed M, it decreases linearly with frequency [13].

In XL-MIMO, large array dimensions can push the Rayleigh distance to hundreds or thousands of meters—comparable to or exceeding typical cell sizes. Thus, near-field effects can dominate even at moderate frequencies, contrary to the common misconception that near-field propagation only matters at high frequencies. In reality, the array's electrical size (physical dimension relative to wavelength) governs the relevance of near-field phenomena, which can be significant in any frequency band, provided the array is sufficiently large [13].

2.1.1.2 Radiative Near-Field Region And Fresnel Distance

In the radiative near-field region, each BS antenna element experiences a distinct propagation distance to the UE, leading to non-linear phase variations across the array. Consequently, the planar wavefront approximation no longer holds, and the propagation must be modeled using spherical waves. In practice, a parabolic wavefront provides a highly accurate approximation of the spherical model. This region is bounded above by the Rayleigh distance and below by the Fresnel distance, which is given by [1]

$$d_{\text{Fresnel}} = 0.62\sqrt{\frac{D^3}{\lambda}}. \tag{2.13}$$

2.1.1.3 Reactive Near-Field Region

In the reactive near-field, the electromagnetic fields store energy rather than radiating effectively. In other words, the waves exhibit strong evanescent behavior and do not effectively propagate. This region is upper-bounded by the Fresnel distance. The Fresnel distance guarantees that the maximum approximation error of the second-order Taylor expansion of the signal phase is $\pi/8$ [1].

2.1.2 Field Regions Based on Amplitude Variations

Two field regions can be defined based on the amplitude variations across the array [1, 13, 23], the uniform region and the non-uniform region.

5 The antenna spacing is typically a multiple of half-wavelength. To avoid mutual coupling among elements, $I_{\text{y}} \geq 1$ and $I_{\text{z}} \geq 1$ is a common practice. Standard element separation is half wavelength ($I_{\text{y}} = I_{\text{z}} = 1$).

2.1.2.1 Uniform Region

The amplitude variations across the antenna array remain below a threshold, typically 10%. Thus, in the uniform region, the amplitudes are approximately constant across the array. This region is lower-bounded by the uniform power distance (UPD). Define the power variation ratio between the strongest and weakest antenna of the array:

$$\Upsilon(d, \varphi, \theta) = \frac{\min_m \beta_m(d, \varphi, \theta)}{\max_m \beta_m(d, \varphi, \theta)} = \frac{\beta_{m'}(d, \varphi, \theta)}{\beta_{m''}(d, \varphi, \theta)}, \tag{2.14}$$

where β_m denotes the LSF channel gain at the m-th antenna, and depends on the propagation distance d, the azimuth angle φ, and the elevation angle θ. The index of the antenna with the lowest and with the highest LSF channel gain are denoted by m' and m'', respectively.

For a given threshold $\Upsilon_{\text{th}} < 1$, the uniform-power distance $d_{\text{UPD}}(\varphi, \theta)$ is the smallest array-to-source distance for which $\Upsilon(d, \varphi, \theta) \geq \Upsilon_{\text{th}}$. In other words, when $d \geq d_{\text{UPD}}(\varphi, \theta)$ the amplitude variation across array elements is within the specified tolerance Υ_{th}, and the uniform-power distance is defined as

$$d_{\text{UPD}}(\varphi, \theta) \triangleq \min \left\{ d : \Upsilon(d, \varphi, \theta) \geq \Upsilon_{\text{th}} \right\}. \tag{2.15}$$

2.1.2.2 Non-Uniform Region

When amplitude variation across elements becomes noticeable, a non-uniform planar wave model is necessary. This typically occurs at the near-field region. The non-uniform region is upper-bounded by the UPD.

2.1.3 Field Regions Based on Amplitude And Phase Variations

Based on the near- and far-field regions, and on the uniform and non-uniform regions, we can define six field regions [1, 13, 23, 62]:

- **UPW ($d \geq d_{\text{Rayleigh}}$ and $d \geq d_{\text{UPD}}$):** In the UPW region, the amplitudes are uniform across the array and the planar wave assumption holds, which greatly simplifies channel modeling, signal processing and performance analysis, since all antennas share nearly identical azimuth and elevation angles and constant amplitude. The UPW region is at the far-field.

- **USW ($d < d_{\text{Rayleigh}}$ and $d \geq d_{\text{UPD}}$):** In the uniform spherical wave (USW) region, which is in the near-field, the amplitudes are uniform, but the waves are spherical. The UPBW model is a common and valid approximation in this region.

- **UPBW ($d_{\text{Fresnel}} \leq d < d_{\text{Rayleigh}}$ and $d \geq d_{\text{UPD}}$):** In the UPBW region, which is in the radiative near-field, the amplitudes are uniform, and parabolic wavefront approximation is accurate.

- **NUPW ($d \geq d_{\text{Rayleigh}}$ and $d < d_{\text{UPD}}$):** In the non-uniform plane wave (NUPW) region, the amplitudes are non-uniform, but the waves are planar. It belongs to the far-field, but occurs at distances just beyond the radiative near-field.
- **NUSW ($d_{\text{Fresnel}} \leq d < d_{\text{Rayleigh}}$ and $d < d_{\text{UPD}}$):** In the NUSW region, which is in the radiative near-field, the amplitudes are non-uniform, and the waves are spherical. Thus, the exact propagation distance should be considered in the modeling of the channel in this region.
- **NUEW ($d < d_{\text{Fresnel}}$ and $d < d_{\text{UPD}}$):** In the non-uniform evanescent wave (NUEW) region, also known as reactive near-field, the amplitudes are non-uniform, and the waves are evanescent.

Table 2.1 summarizes the field regions in XL-MIMO based on amplitude and phase variations, propagation distances, and wavefront models.

Table 2.1 – Summary of field regions in XL-MIMO

Region	Distance	Amplitude	Wavefront	Phase variations
UPW	$d \geq d_{\text{Rayleigh}}$ and $d \geq d_{\text{UPD}}$	Uniform	Planar	Linear
UPBW	$d_{\text{Fresnel}} \leq d < d_{\text{Rayleigh}}$ and $d \geq d_{\text{UPD}}$	Uniform	Parabolic	Non-linear
USW	$d < d_{\text{Rayleigh}}$ and $d \geq d_{\text{UPD}}$	Uniform	Spherical	Non-linear
NUPW	$d \geq d_{\text{Rayleigh}}$ and $d < d_{\text{UPD}}$	Non-uniform	Planar	Linear
NUSW	$d_{\text{Fresnel}} \leq d < d_{\text{Rayleigh}}$ and $d < d_{\text{UPD}}$	Non-uniform	Spherical	Non-linear
NUEW	$d < d_{\text{Fresnel}}$ and $d < d_{\text{UPD}}$	Non-uniform	Evanescent	Non-linear

To reconcile modeling accuracy and tractability, the array may be partitioned into subarrays, each sufficiently compact for every UE–subarray link to remain in the subarray's far-field, locally restoring the UPW approximation, as illustrated in Figure 2.4. Within a subarray, signal phase varies linearly and large-scale gains and angles can be considered approximately constant, as summarized in Table 2.2.

Table 2.2 – Near-field vs. far-field

Difference in	**Antennas belonging to**	
	the same subarray	**different subarrays**
propagation distance	irrelevant	big
azimuth and elevation angles	irrelevant	big
signal amplitude	irrelevant	big
signal phase	probably linear or parabolic	non-linear

For any UE, each subarray is assumed fully LoS-visible or purely NLoS, making the channel stationary within each subarray but probably non-stationary across subarrays.

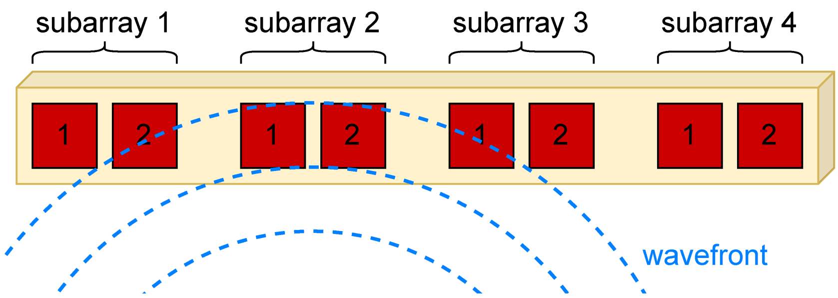


Figure 2.4 – The entire array experiences a spherical wavefront, while each individual subarray observes an approximately planar wavefront.

The resulting model captures inter-subarray amplitude differences and phase curvature, while retaining much of the tractability of conventional massive MIMO.

2.2 Hybrid LoS/NLoS Channel Model

This work adopts the standard block-fading model [8]: the channel remains constant for τ_c symbols and takes an independent realization in each coherence block. Details on coherence block structure are in Appendix I. Within an arbitrary block, the channel between the N antennas of the k-th UE and the M antennas of the l-th subarray is denoted $\mathbf{H}_{kl} \in \mathbb{C}^{M \times N}$, comprising both LoS and NLoS components ($\overline{\mathbf{H}}_{kl}$ and $\check{\mathbf{H}}_{kl}$, respectively). The NLoS matrix captures the practical scenario of spatially correlated multipath propagation, where received signals result from the superposition of multiple paths due to reflections, diffractions, and scattering occurring in the environment [63]. These scattered components are typically grouped into clusters. Thus, the channel between the k-th UE and the l-th subarray is expressed as [13]

$$\mathbf{H}_{kl} = \omega_{kl}\overline{\mathbf{H}}_{kl} + \underbrace{\sum_{i=1}^{C} \omega_{kl,i}\check{\mathbf{H}}_{kl,i}}_{\check{\mathbf{H}}_{kl}}, \tag{2.16}$$

where $\check{\mathbf{H}}_{kl,i}$ denotes the contribution of the i-th cluster, C is the number of clusters, $\omega_{kl} \in \{0,1\}$ is a Bernoulli random variable indicating the presence (1) or absence (0) of a LoS component between UE k and subarray l [64], and $\omega_{kl,i} \in \{0,1\}$ indicates whether the i-th cluster is jointly visible to the k-th UE and the l-th subarray (1 if so, 0 otherwise). The l-th subarray is considered to lie within the k-th UE's VR if a LoS path exists between them; otherwise, the received power is markedly weaker and originates solely from NLoS contributions. The LoS probability is inversely related to the distance between the UE and the subarray, as modeled by (2.21) in Section 2.3.1.

The NLoS term is zero-mean, implying that the LoS component represents the channel mean. Subsequent sections provide a detailed description of the proposed XL-MIMO channel model.

LSF manifests as slow mean power variations over tens to hundreds of meters due to path loss and shadowing, while small-scale fading (SSF) describes rapid fluctuations in amplitude and phase over distances on the order of a wavelength, caused by multipath propagation and Doppler shifts. Variability in the NLoS term generates SSF, while fluctuations in the covariance matrix and LoS term reflect LSF.

Figure 2.5 illustrates an exemplary XL-MIMO deployment where a BS equipped with $L = 5$ subarrays serves $U = 4$ UEs. This scenario highlights the concept of spatial non-stationarity, wherein each UE is characterized by a distinct VR comprising only those subarrays with which it maintains a LoS link. For instance, the VR of UE 1 spans subarrays 1–3, whereas physical obstructions (e.g., a house and a tree) restrict the VR of UE 2 to subarrays 2 and 3. Notably, UE 3 is heavily obstructed and lacks a LoS path to any subarray.

To illustrate resource allocation, consider a subarray selection strategy that assigns the two subarrays with the strongest channel gains to each user. Under this policy, subarrays 1 and 2 are selected for UE 1, subarrays 2 and 3 for UE 2, and subarrays 4 and 5 for UE 4. Due to the absence of LoS links and the consequent severe signal attenuation, UE 3 is not scheduled for service at this time. This spatial configuration also underscores the need for intelligent pilot assignment. For example, assigning the same pilot to UE 1 and UE 2 would induce pilot contamination at subarrays 2 and 3, where their VRs overlap, thereby degrading channel estimation quality.

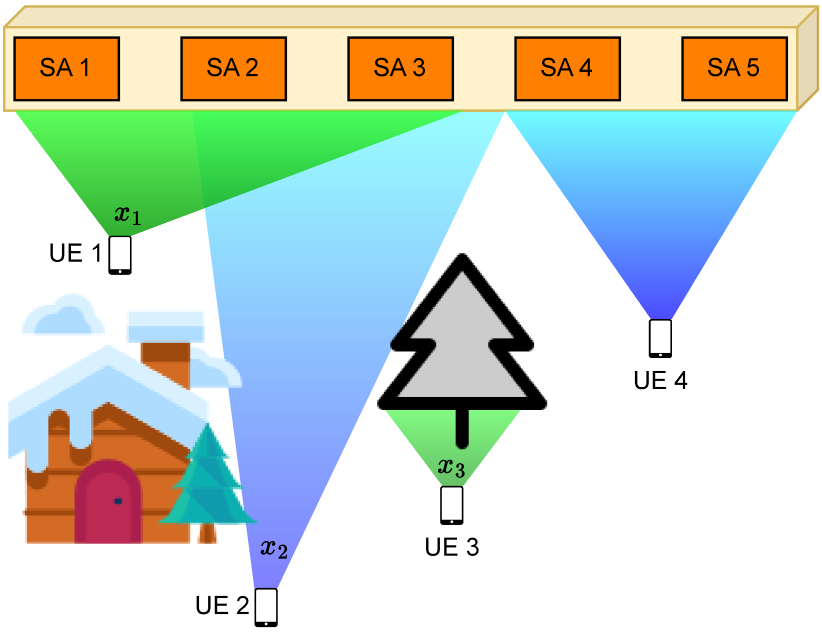


Figure 2.5 – Illustrative XL-MIMO scenario with $L = 5$ subarrays and $U = 4$ UEs exhibiting distinct VRs due to spatial non-stationarities.

2.3 LoS Channel Model

The LoS channel gain and phase shift between the reference antennas of the k-th UE and the l-th subarray are given, respectively, by $\alpha_{kl} = \frac{\lambda}{4\pi d_{kl}}$ and $\phi_{kl} = -\frac{2\pi}{\lambda} d_{kl}$.[6] The resulting complex-valued LoS channel coefficient between the reference antennas is the (m_0, n_0)-th entry of the matrix $\overline{\mathbf{H}}_{kl}$. It is expressed as

$$\overline{h}_{kl} = [\overline{\mathbf{H}}_{kl}]_{m_0,n_0} = \omega_{kl}\alpha_{kl}\mathrm{e}^{\mathrm{j}\phi_{kl}} = \omega_{kl}\frac{\lambda}{4\pi d_{kl}}\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}d_{kl}}, \tag{2.17}$$

and the corresponding channel power gain is [65]

$$\bar{\beta}_{kl} = |\overline{h}_{kl}|^2 = \omega_{kl}\alpha_{kl}^2 = \omega_{kl}\frac{\beta_0}{d_{kl}^2}, \tag{2.18}$$

where $\beta_0 = \frac{\lambda^2}{(4\pi)^2}$ is the average power gain for a distance of 1 m.

The LoS channel between the n-th antenna of the k-th UE and the m-th antenna of the l-th subarray is denoted by $\overline{h}_{klmn}$, and corresponds to the (m,n)-th entry of $\overline{\mathbf{H}}_{kl}$, *i.e.*, $\overline{h}_{klmn} = [\overline{\mathbf{H}}_{kl}]_{m,n}$. Similarly to (2.17), $\overline{h}_{klmn}$ is expressed as

$$\overline{h}_{klmn} = \underbrace{\omega_{kl}\frac{\lambda}{4\pi d_{kl}}\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}d_{kl}}}_{\overline{h}_{kl}}\underbrace{\frac{d_{kl}}{d_{klmn}}\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}(d_{klmn}-d_{kl})}}_{a_{klmn}}, \tag{2.19}$$

where a_{klmn} accounts for the difference in amplitude and phase of the channel between antennas n and m, relative to the channel between the reference antennas, and $d_{klmn} = \|\mathbf{s}_{kln} - \mathbf{r}_m\|$ is the distance between these antennas. Using (2.9), this can be written as

$$d_{klmn} = \sqrt{d_{kl}^2 + 2d_{kl}\mathbf{k}_{kl}^{\mathrm{T}}(\mathbf{T}_{kl}\mathbf{s}_n - \mathbf{r}_m) + \|\mathbf{T}_{kl}\mathbf{s}_n - \mathbf{r}_m\|^2}. \tag{2.20}$$

2.3.1 LoS Probability

For the micro-urban scenario, the probability of existing LoS between UE k and subarray l depends on the propagation distance d_{kl}, and is modeled according to [66] as:

$$\Pr(\overline{\mathbf{H}}_{kl} \neq \mathbf{0}_{M\times N}) = \min\left(\frac{18 \text{ m}}{d_{kl}}, 1\right)\left[1 - \mathrm{e}^{-\frac{d_{kl}}{36 \text{ m}}}\right] + \mathrm{e}^{-\frac{d_{kl}}{36 \text{ m}}}. \tag{2.21}$$

According to (2.21), the propagation is always in LoS conditions for $d_{kl} \leq 18$ m, while the LoS probability decays exponentially when $d_{kl} > 18$ m. Using this model and the simulation parameters in Table 3.2, the average LoS probability between a UE-subarray pair is 60.76%. It represents an expected condition of the XL-MIMO system operating in micro-urban channel scenarios. Thus, from the trustworthy numerical simulations perspective, we cannot disregard the LoS component in this scenario.

[6] Let P_{t} and P_{r} denote the transmitted and the received powers, respectively. The free-space channel gain is $\beta = \sqrt{\frac{P_{\mathrm{r}}}{P_{\mathrm{t}}}} = \sqrt{\frac{A_{\mathrm{e}}}{A_{\mathrm{s}}}}$, where $A_{\mathrm{e}} = \frac{\lambda^2}{4\pi}$ is the effective antenna area, and $A_{\mathrm{s}} = 4\pi d^2$ is the surface area of a sphere of radius d. Consequently, $\alpha_{kl} = \frac{\lambda}{4\pi d_{kl}}$ [65].

2.3.2 NUSW Model

The NUSW model is the more generic model, since it uses the exact propagation distance d_{klmn}, leading to the most general form for the relative channel term:

$$a_{klmn} = \frac{d_{kl}}{d_{klmn}} \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}(d_{klmn} - d_{kl})}. \tag{2.22}$$

See from (2.22) and (2.20) that we do not need to estimate $\overline{h}_{klmn}$, since we can compute it based on the channel between the reference antennas ($\overline{h}_{kl}$), the UE and subarray geometries, the direction of arrival (DoA), and the rotation matrix.

However, to further reduce the complexity of computing $\overline{h}_{klmn}$, we can approximate a_{klmn} using the first or the second-order Taylor expansions of d_{klmn} around $\|\mathbf{T}_{kl}\mathbf{s}_n - \mathbf{r}_m\| = 0$, which are respectively given by

$$d_{klmn} \approx d_{kl} - \mathbf{k}_{kl}^{\mathrm{T}}\mathbf{r}_m + \mathbf{k}_{kl}^{\mathrm{T}}\mathbf{T}_{kl}\mathbf{s}_n \tag{2.23}$$

and

$$d_{klmn} \approx d_{kl} - \mathbf{k}_{kl}^{\mathrm{T}}\mathbf{r}_m + \mathbf{k}_{kl}^{\mathrm{T}}\mathbf{T}_{kl}\mathbf{s}_n + \frac{\|\mathbf{T}_{kl}\mathbf{s}_n - \mathbf{r}_m\|^2 - [\mathbf{k}_{kl}^{\mathrm{T}}(\mathbf{T}_{kl}\mathbf{s}_n - \mathbf{r}_m)]^2}{2d_{kl}}. \tag{2.24}$$

It will then be necessary only to compute inner products between 3-dimensional vectors, instead of computing square roots.

2.3.3 UPBW Model

The UPBW model simplifies the channel by assuming a uniform amplitude across all possible pairs of transmit and receive antennas, which corresponds to the approximation $d_{klmn}/d_{kl} \approx 1$. For the phase term, it maintains higher fidelity by approximating the spherical wavefronts as parabolic. This is achieved by using the second-order Taylor expansion for the distance difference $d_{klmn} - d_{kl}$, given in (2.24). Applying these approximations in (2.22) results in

$$a_{klmn} = \exp\left[\mathrm{j}\frac{2\pi}{\lambda}\left(\mathbf{k}_{kl}^{\mathrm{T}}\mathbf{r}_m - \mathbf{k}_{kl}^{\mathrm{T}}\mathbf{T}_{kl}\mathbf{s}_n - \frac{\|\mathbf{T}_{kl}\mathbf{s}_n - \mathbf{r}_m\|^2 - [\mathbf{k}_{kl}^{\mathrm{T}}(\mathbf{T}_{kl}\mathbf{s}_n - \mathbf{r}_m)]^2}{2d_{kl}}\right)\right]. \tag{2.25}$$

2.3.4 UPW Model

In the far-field, the UPW model is applicable. This model assumes a uniform channel gain across all transmit-receive antenna pairs and utilizes a first-order Taylor expansion to approximate the distance d_{mn}, as previously shown in (2.23). This yields[7]

$$a_{klmn} = \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}\left(\mathbf{k}_{kl}^{\mathrm{T}}\mathbf{r}_m + \dot{\mathbf{k}}_{kl}^{\mathrm{T}}\mathbf{s}_n\right)}. \tag{2.26}$$

[7] Approximating the phase under the plane wave assumption is advantageous because it greatly simplifies the mathematical analysis of the channel model, since it allows the phase to be modeled as a linear term, facilitating the use of tools such as Fourier analysis and angular domain representations, and avoids nonlinear or quadratic terms associated with wavefront curvature, resulting in more tractable expressions for the channel and SINR.

In the far-field, a_{klmn} can be decoupled into separate transmit and receive components, and can be written explicitly as a function of the angles of arrival and departure:

$$a_{klmn} = a_{mn}(\varphi_{kl}, \theta_{kl}, \dot{\varphi}_{kl}, \dot{\theta}_{kl}) = [\mathbf{a}(\varphi_{kl}, \theta_{kl})]_m [\dot{\mathbf{a}}(\dot{\varphi}_{kl}, \dot{\theta}_{kl})]_n, \tag{2.27}$$

where $\mathbf{a}(\varphi, \theta)$ is the subarray's response vector for a signal coming from the direction (φ, θ). Its m-th entry is given by

$$[\mathbf{a}(\varphi, \theta)]_m = \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}\mathbf{k}^{\mathrm{T}}(\varphi,\theta)\mathbf{r}_m}. \tag{2.28}$$

Similarly, $\dot{\mathbf{a}}(\varphi, \theta)$ is the UE's response vector, whose n-th entry is given by

$$[\dot{\mathbf{a}}(\varphi, \theta)]_n = \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}\mathbf{k}^{\mathrm{T}}(\varphi,\theta)\mathbf{s}_n}. \tag{2.29}$$

Therefore, the far-field LoS channel matrix can be expressed as the outer product of these vectors:

$$\overline{\mathbf{H}}_{kl} = \frac{\lambda}{4\pi d_{kl}} \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}d_{kl}} \mathbf{a}(\varphi_{kl}, \theta_{kl})[\dot{\mathbf{a}}(\dot{\varphi}_{kl}, \dot{\theta}_{kl})]^{\mathrm{T}}, \tag{2.30}$$

Assuming planar antenna array, the array response vector can be factored into the Kronecker product of two vectors corresponding to the responses along the y and z axes. For the subarray and the UE, this decomposition is respectively

$$\mathbf{a}(\varphi, \theta) = \mathbf{a}_{\mathrm{z}}(\theta) \otimes \mathbf{a}_{\mathrm{y}}(\varphi, \theta), \tag{2.31}$$

$$\dot{\mathbf{a}}(\varphi, \theta) = \dot{\mathbf{a}}_{\mathrm{z}}(\theta) \otimes \dot{\mathbf{a}}_{\mathrm{y}}(\varphi, \theta), \tag{2.32}$$

where the elements of the constituent vectors for the subarray are

$$[\mathbf{a}_{\mathrm{y}}(\varphi, \theta)]_{m_{\mathrm{y}}} = \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}(m_{\mathrm{y}} - m_{\mathrm{y},0})\Delta_{\mathrm{y}} \sin(\varphi)\cos(\theta)}, \tag{2.33}$$

$$[\mathbf{a}_{\mathrm{z}}(\theta)]_{m_{\mathrm{z}}} = \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}(m_{\mathrm{z}} - m_{\mathrm{z},0})\Delta_{\mathrm{z}} \sin(\theta)}, \tag{2.34}$$

and for the UE are

$$[\dot{\mathbf{a}}_{\mathrm{y}}(\varphi, \theta)]_{n_{\mathrm{y}}} = \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}(n_{\mathrm{y}} - n_{\mathrm{y},0})\dot{\Delta}_{\mathrm{y}} \sin(\varphi)\cos(\theta)}, \tag{2.35}$$

$$[\dot{\mathbf{a}}_{\mathrm{z}}(\theta)]_{n_{\mathrm{z}}} = \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}(n_{\mathrm{z}} - n_{\mathrm{z},0})\dot{\Delta}_{\mathrm{z}} \sin(\theta)}. \tag{2.36}$$

2.4 NLoS Channel Component

Let $\mathbf{p}_{l,iq}$ denote the position of the q-th scatterer in the i-th cluster, relative to the reference antenna of the l-th subarray. The signal from this scatterer arrives at the subarray with specific azimuth and elevation angles, $(\varphi_{l,iq}, \theta_{l,iq})$, which define the DoA unit vector $\mathbf{k}_{l,iq} = \mathbf{k}(\varphi_{l,iq}, \theta_{l,iq})$.[8] Therefore, the scatterer's position can be written as

$$\mathbf{p}_{l,iq} = d_{l,iq}\mathbf{k}_{l,iq}. \tag{2.37}$$

[8] In other words, $\mathbf{k}_{l,iq}$ is the unit vector pointing from the subarray's reference antenna to scatterer q of cluster i.

Similarly, the signal departs from the reference antenna of the k-th UE towards the scatterer with angles $(\dot{\varphi}_{k,iq}, \dot{\theta}_{k,iq})$, defining the direction of departure (DoD) unit vector $\dot{\mathbf{k}}_{k,iq} = \mathbf{k}(\dot{\varphi}_{k,iq}, \dot{\theta}_{k,iq})$ in the UE's local coordinate system. The position of the scatterer in this local frame is

$$\dot{\mathbf{p}}_{k,iq} = \dot{d}_{k,iq}\dot{\mathbf{k}}_{k,iq}, \tag{2.38}$$

where $\dot{d}_{k,iq} = \|\dot{\mathbf{p}}_{k,iq}\|$ is the distance between this scatterer and the UE's reference antenna. The two representations of the scatterer's position are related by the transformation

$$\mathbf{p}_{l,iq} = \mathbf{s}_{kl} + \mathbf{T}_{kl}\dot{\mathbf{p}}_{k,iq}. \tag{2.39}$$

Using (2.7) and (2.37), we can express the distance $\dot{d}_{k,iq}$ as

$$\dot{d}_{k,iq} = \|\mathbf{p}_{l,iq} - \mathbf{s}_{kl}\| = \sqrt{d_{l,iq}^2 - 2d_{kl}\mathbf{k}_{kl}^{\mathrm{T}}\mathbf{p}_{l,iq} + d_{kl}^2}. \tag{2.40}$$

This geometric relationship also implies that the DoD is dependent on the DoA and the positions of the subarray and the UE. Thus, there is no need to independently estimate the angles of departure. By rearranging (2.39), we can solve for the DoD unit vector:[9]

$$\dot{\mathbf{k}}_{k,iq} = \frac{1}{\dot{d}_{k,iq}}\mathbf{T}_{kl}^{\mathrm{T}}(d_{l,iq}\mathbf{k}_{l,iq} - d_{kl}\mathbf{k}_{kl}). \tag{2.41}$$

We now generalize the distances for paths involving arbitrary antenna elements. The distance between the m-th antenna of the l-th subarray and the q-th scatterer of the i-th cluster is

$$d_{lm,iq} = \|\mathbf{p}_{l,iq} - \mathbf{r}_m\| = \sqrt{d_{l,iq}^2 + \|\mathbf{r}_m\|^2 - 2d_{l,iq}\mathbf{k}_{l,iq}^{\mathrm{T}}\mathbf{r}_m}. \tag{2.42}$$

The distance between this scatterer and the n-th antenna of the k-th UE is most compactly expressed in the UE's coordinate system:

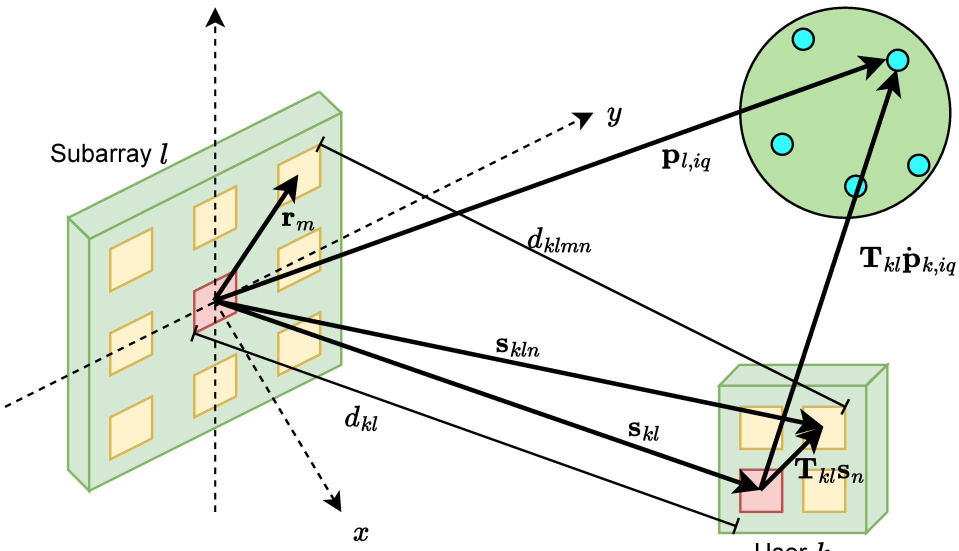
$$\dot{d}_{kn,iq} = \|\dot{\mathbf{p}}_{k,iq} - \mathbf{s}_n\| = \sqrt{\dot{d}_{k,iq}^2 + \|\mathbf{s}_n\|^2 - 2\dot{d}_{k,iq}\dot{\mathbf{k}}_{k,iq}^{\mathrm{T}}\mathbf{s}_n}. \tag{2.43}$$

It can also be expressed as

$$\dot{d}_{kn,iq} = \|\mathbf{p}_{l,iq} - \mathbf{s}_{kln}\| = \sqrt{d_{l,iq}^2 + d_{kl}^2 + \|\mathbf{s}_n\|^2 + 2d_{kl}\mathbf{k}_{kl}^{\mathrm{T}}\mathbf{T}_{kl}\mathbf{s}_n - 2d_{l,iq}\mathbf{k}_{l,iq}^{\mathrm{T}}(d_{kl}\mathbf{k}_{kl} + \mathbf{T}_{kl}\mathbf{s}_n)}. \tag{2.44}$$

Figure 2.6 illustrates a subarray of $M = 9$ antennas, a UE with $N = 4$ antennas, and a cluster with 5 scatterers. The antennas in red are their respective reference antennas. The figure shows the main vectors and distances related to the reference antennas, an arbitrary subarray antenna m, an arbitrary UE antenna n, and an arbitrary scatterer i.

[9] This relation is valid assuming two-hop scattering, *i.e.*, assuming that the signal travels from the transmitter to a scatterer and then to receiver, instead of being reflected by other scatterers before going to the receiver.



- $d_{kl} = \|\mathbf{s}_{kl}\|$
- $d_{klmn} = \|\mathbf{s}_{kln} - \mathbf{r}_m\|$
- $d_{l,iq} = \|\mathbf{p}_{l,iq}\|$
- $\dot{d}_{k,iq} = \|\dot{\mathbf{p}}_{k,iq}\| = \|\mathbf{p}_{l,iq} - \mathbf{T}_{kl}\mathbf{s}_n\|$
- $d_{lm,iq} = \|\mathbf{p}_{l,iq} - \mathbf{r}_m\|$
- $\dot{d}_{kn,iq} = \|\dot{\mathbf{p}}_{k,iq} - \mathbf{s}_n\| = \|\mathbf{p}_{l,iq} - \mathbf{s}_{kln}\|$

Figure 2.6 – Main vectors and distances related to the hybrid LoS/NLoS channel between a subarray and a UE.

Figure 2.7 illustrates the impingent wave at an arbitrary antenna of subarray l, coming from an arbitrary antenna of UE k. The signal is traveling from the UE to the subarray through the LoS path, and through three scatterers.

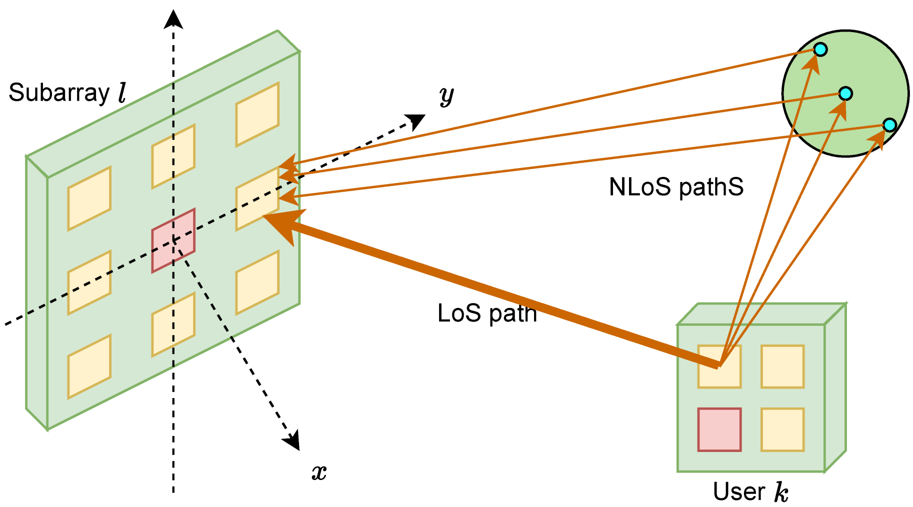


Figure 2.7 – LoS and NLoS propagation between a subarray and a UE.

2.4.1 Discrete Scatterer Model

The i-th cluster is composed of Q_i discrete scatterers. The path through the q-th scatterer in this cluster can be described as a two-hop link. The first hop, from the k-th UE to the scatterer, is characterized by the channel vector $\dot{\mathbf{h}}_{k,iq} \in \mathbb{C}^{N\times 1}$. The second hop, from the scatterer to the l-th subarray, is described by $\mathbf{h}_{l,iq} \in \mathbb{C}^{M\times 1}$. These are given by

$$\dot{\mathbf{h}}_{k,iq} = \frac{\lambda}{4\pi \dot{d}_{k,iq}} \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}\dot{d}_{k,iq}} \dot{\mathbf{a}}_{k,iq}, \tag{2.45}$$

$$\mathbf{h}_{l,iq} = \frac{\lambda}{4\pi d_{l,iq}} \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}d_{l,iq}} \mathbf{a}_{l,iq}, \tag{2.46}$$

where $\dot{\mathbf{a}}_{k,iq}$ and $\mathbf{a}_{l,iq}$ are the respective array response vectors for this scatterer. The n-th and m-th entries of these vectors are given by

$$[\dot{\mathbf{a}}_{k,iq}]_n = \frac{\dot{d}_{k,iq}}{\dot{d}_{kn,iq}} \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}(\dot{d}_{kn,iq}-\dot{d}_{k,iq})}, \tag{2.47}$$

$$[\mathbf{a}_{l,iq}]_m = \frac{d_{l,iq}}{d_{lm,iq}} \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}(d_{lm,iq}-d_{l,iq})}. \tag{2.48}$$

After arriving at the scattering cluster and before traveling to the receiver, the signal amplitude and phase are modified by the scatterers. The complex random variable $c_{iq}\mathrm{e}^{\mathrm{j}\psi_{iq}}$ accounts for this effect. The contribution of a single scatterer to the channel matrix is a rank-one matrix. Summing over all scatterers gives

$$\check{\mathbf{H}}_{kl} = \sum_{i=1}^{C} \omega_{kl,i} \sum_{q=1}^{Q_i} c_{iq}\mathrm{e}^{\mathrm{j}\psi_{iq}} \mathbf{h}_{l,iq}\dot{\mathbf{h}}_{k,iq}^{\mathrm{T}}. \tag{2.49}$$

By substituting (2.45) and (2.46) into (2.49), we can rewrite $\check{\mathbf{H}}_{kl}$ as

$$\check{\mathbf{H}}_{kl} = \sum_{i=1}^{C} \omega_{kl,i} \sum_{q=1}^{Q_i} \zeta_{kliq}\mathrm{e}^{\mathrm{j}\phi_{kliq}} \mathbf{a}_{l,iq}\dot{\mathbf{a}}_{k,iq}^{\mathrm{T}}, \tag{2.50}$$

where $\zeta_{kliq} = \frac{c_{iq}\lambda^2}{(4\pi)^2 \dot{d}_{k,iq} d_{l,iq}}$, and $\phi_{kliq} = \psi_{iq} - \frac{2\pi}{\lambda}\dot{d}_{k,iq} - \frac{2\pi}{\lambda}d_{l,iq}$ represent respectively the random gain and phase shift for the path through the q-th scatterer of the i-th cluster, relative to the reference antennas.

Since the total phase shift is uniformly distributed, *i.e.*, $\phi_{kliq} \sim \mathcal{U}[-\pi,\pi)$, its complex exponential has zero mean: $\mathbb{E}\{\mathrm{e}^{\mathrm{j}\phi_{kliq}}\} = 0$. If ζ_{kliq} and ϕ_{kliq} are independent, the expectation of each term in the sum is zero, leading to a zero-mean channel matrix for the cluster:

$$\mathbb{E}\{\check{\mathbf{H}}_{kl}\} = \sum_{i=1}^{C}\sum_{q=1}^{Q_i} \omega_{kl,i}\mathbb{E}\{\zeta_{kliq}\}\mathbb{E}\{\mathrm{e}^{\mathrm{j}\phi_{kliq}}\}\mathbf{a}_{l,iq}\dot{\mathbf{a}}_{k,iq}^{\mathrm{T}} = \mathbf{0}_{M\times N}. \tag{2.51}$$

Consequently, the total NLoS channel matrix $\check{\mathbf{H}}_{kl}$ also has a zero mean.

2.4.2 Continuous Angular Spread Model

When each scattering cluster contains a large number of scatterers, the discrete summation of spherical waves in (2.50) can be accurately approximated by a continuous integral of plane waves over the angular domain [67]. This leads to a double-directional channel representation given by

$$\check{\mathbf{H}}_{kl} = \sum_{i=1}^{C} \omega_{kl,i} \iiiint g_{kl,i}(\varphi, \theta, \dot{\varphi}, \dot{\theta}) \mathbf{a}(\varphi, \theta) [\dot{\mathbf{a}}(\dot{\varphi}, \dot{\theta})]^{\mathrm{T}} \mathrm{d}\varphi \mathrm{d}\theta \mathrm{d}\dot{\varphi} \mathrm{d}\dot{\theta}, \tag{2.52}$$

where $\mathbf{a}(\varphi, \theta)$ and $\dot{\mathbf{a}}(\varphi, \theta)$ are the far-field array response vectors defined in (2.28) and (2.29). The function $g_{kl,i}(\varphi, \theta, \dot{\varphi}, \dot{\theta})$ represents the complex angular spreading function of the i-th cluster, describing the distribution of path gains over the joint angular domain. The integration intervals are $[-\pi, \pi)$ for azimuth and $[-\pi/2, \pi/2]$ for elevation angles.

From this representation, the n-th column of $\check{\mathbf{H}}_{kl}$ corresponds to the NLoS channel vector between the n-th antenna of the k-th UE and subarray l, expressed as

$$\check{\mathbf{h}}_{kln}^{\mathrm{SA}} = \sum_{i=1}^{C} \omega_{kl,i} \underbrace{\iiiint g_{kl,i}(\varphi, \theta, \dot{\varphi}, \dot{\theta}) \mathbf{a}(\varphi, \theta) [\dot{\mathbf{a}}(\dot{\varphi}, \dot{\theta})]_n \, \mathrm{d}\varphi \mathrm{d}\theta \mathrm{d}\dot{\varphi} \mathrm{d}\dot{\theta}}_{\check{\mathbf{h}}_{kln,i}^{\mathrm{SA}}}, \tag{2.53}$$

where $\check{\mathbf{h}}_{kln,i}^{\mathrm{SA}}$ represents the contribution from the i-th scattering cluster.

Similarly, the channel vector from the k-th UE and the m-th antenna of subarray l, which corresponds to the transpose of the m-th row of $\check{\mathbf{H}}_{kl}$, is:

$$\check{\mathbf{h}}_{klm}^{\mathrm{UE}} = \sum_{i=1}^{C} \omega_{kl,i} \underbrace{\iiiint g_{kl,i}(\varphi, \theta, \dot{\varphi}, \dot{\theta}) [\mathbf{a}(\varphi, \theta)]_m \dot{\mathbf{a}}(\dot{\varphi}, \dot{\theta}) \mathrm{d}\varphi \mathrm{d}\theta \mathrm{d}\dot{\varphi} \mathrm{d}\dot{\theta}}_{\check{\mathbf{h}}_{klm,i}^{\mathrm{UE}}}, \tag{2.54}$$

where $\check{\mathbf{h}}_{klm,i}^{\mathrm{UE}}$ denotes the contribution of the i-th cluster.

The NLoS channel coefficient between the reference antennas is denoted by the complex scalar $\check{h}_{kl}$, corresponding to the (m_0, n_0)-th entry of $\check{\mathbf{H}}_{kl}$, given by[10]

$$\check{h}_{kl} = [\check{\mathbf{H}}_{kl}]_{m_0,n_0} = \sum_{i=1}^{C} \omega_{kl,i} \underbrace{\iiiint g_{kl,i}(\varphi, \theta, \dot{\varphi}, \dot{\theta}) [\mathbf{a}(\varphi, \theta)]_{m_0} [\dot{\mathbf{a}}(\dot{\varphi}, \dot{\theta})]_{n_0}^{*} \mathrm{d}\varphi \mathrm{d}\theta \mathrm{d}\dot{\varphi} \mathrm{d}\dot{\theta}}_{\check{h}_{kl,i}}, \tag{2.55}$$

where $\check{h}_{kl,i}$ is the contribution of the i-th cluster.

Assuming that paths originating from different clusters are uncorrelated, the total channel covariance is the sum of the covariances from each cluster. Furthermore, within a cluster, assume that paths with different angles are uncorrelated. This implies that the

[10] For notation simplicity, the channel between the reference antenna of UE k and the reference antenna of subarray l is denoted by $\check{h}_{kl}$, instead of $\check{h}_{klm_0n_0}$.

expectation of the product of two angular spreading functions is zero unless the angles are identical:[11]

$$\mathbb{E}\{g_{kl,i}(\varphi,\theta,\dot{\varphi},\dot{\theta})g^*_{klj}(\varphi',\theta',\dot{\varphi}',\dot{\theta}')\} = \mathbb{E}\{|g_{kl,i}(\varphi,\theta,\dot{\varphi},\dot{\theta})|^2\} \times \delta_{ij}\delta(\varphi-\varphi')\delta(\theta-\theta')\delta(\dot{\varphi}-\dot{\varphi}')\delta(\dot{\theta}-\dot{\theta}'), \tag{2.56}$$

where δ_{ij} is the Kronecker delta. Then, the average power gain between the reference antennas is

$$\check{\beta}_{kl} = \mathbb{E}\{|\check{h}_{kl}|^2\} = \sum_{i=1}^{C}\omega_{kl,i}\underbrace{\iiiint \mathbb{E}\{|g_{kl,i}(\varphi,\theta,\dot{\varphi},\dot{\theta})|^2\}\mathrm{d}\varphi\mathrm{d}\theta\mathrm{d}\dot{\varphi}\mathrm{d}\dot{\theta}}_{\check{\beta}_{kl,i}}, \tag{2.57}$$

where $\check{\beta}_{kl,i} = \mathbb{E}\{|\check{h}_{kl,i}|^2\}$ denotes the average power gain contribution of the i-th cluster.

2.4.3 Statistical NLoS Channel Model

The NLoS channel is characterized by the second-order statistics of the paths through the scattering clusters. These statistics are modeled based on the angular distribution of the scattered signal. According to [65, page 314], the NLoS component can be reasonably approximated as Rayleigh fading with as few as five propagation paths.

2.4.3.1 Angular Power Distribution

Each cluster, composed of many scatterers, creates a spread of propagation paths. Denote the mean DoD of a signal departing from the reference antenna of the k-th UE to the i-th cluster as $(\dot{\varphi}_{k,i},\dot{\theta}_{k,i})$, and the mean DoA of a signal arriving at the reference antenna of the l-th subarray from the i-th cluster as $(\varphi_{l,i},\theta_{l,i})$. The actual angles for paths within the cluster are random variables centered on these means.

Assume the DoD and the angle of departure (AoD) are independent,[12] and that their angular spreads follow a Gaussian distribution.[13] The joint probability density function (PDF) for the DoD is given by $\dot{f}_{k,i}(\varphi,\theta) = f(\varphi-\dot{\varphi}_{k,i},\theta-\dot{\theta}_{k,i})$, and for the DoA by $f_{l,i}(\varphi,\theta) = f(\varphi-\varphi_{l,i},\theta-\theta_{l,i})$, where

$$f(\varphi,\theta) = \frac{1}{2\pi\sigma_\varphi\sigma_\theta}\exp\left(-\frac{\varphi^2}{2\sigma_\varphi^2}-\frac{\theta^2}{2\sigma_\theta^2}\right) \tag{2.58}$$

[11] This relies on the uncorrelated scattering assumption, where the random phases of any two distinct paths are independent and uniformly distributed in $[0,2\pi)$, causing their cross-correlation to be zero. Suppose $x_m = \alpha_m e^{j\phi_m}$ and $x_n = \alpha_n e^{j\phi_n}$ are random variables with random amplitudes and phases. Assuming independent amplitudes and phases, $\mathbb{E}\{x_m x_n^*\} = \mathbb{E}\{\alpha_m\alpha_n\}\mathbb{E}\{e^{j(\phi_m-\phi_n)}\}$. Since both ϕ_m and ϕ_n are uniformly distributed over $[0,2\pi)$, their difference also does, *i.e.*, $\phi_m-\phi_n \sim \mathcal{U}(0,2\pi)$, resulting in $\mathbb{E}\{e^{j(\phi_m-\phi_n)}\}=0$. Consequently, $\mathbb{E}\{x_m x_n^*\}=0$ if $n \neq m$.

[12] This is a reasonable assumption, particularly for multi-hop scattering.

[13] The sum of a large number of zero-mean random NLoS paths approaches a zero-mean normal distribution, as stated by the law of large numbers [63].

is a zero-mean Gaussian PDF with standard deviations $\sigma_\varphi \geq 0$ and $\sigma_\theta \geq 0$ in azimuth and elevation, respectively [9].

The average power gain of the angular spreading function is modeled as the product of the cluster's total average power, $\check{\beta}_{kl,i}$, and the angular PDFs:

$$\mathbb{E}\{|g_{kl,i}(\varphi,\theta,\dot{\varphi},\dot{\theta})|^2\} = \check{\beta}_{kl,i} f_{l,i}(\varphi,\theta)\dot{f}_{k,i}(\dot{\varphi},\dot{\theta}). \tag{2.59}$$

Since $\dot{f}_{k,i}(\cdot)$ and $f_{l,i}(\cdot)$ are PDFs, which integrate to one, (2.59) is consistent with the definition of $\check{\beta}_{kl,i}$ as $\iiiint \mathbb{E}\{|g_{kl,i}(\varphi,\theta,\dot{\varphi},\dot{\theta})|^2\}\mathrm{d}\varphi\mathrm{d}\theta\mathrm{d}\dot{\varphi}\mathrm{d}\dot{\theta} = \check{\beta}_{kl,i}$.

2.4.3.2 Channel Covariance Matrices

The channel covariance matrices capture the spatial correlation at the receiver and transmitter arrays. Under the uniform waves assumption, $|[\dot{\mathbf{a}}(\cdot)]_n| = |[\mathbf{a}(\cdot)]_m| = 1$ for any m and n. Then, the covariance matrix for the subarray's antennas is the same for any UE antenna n, and the covariance matrix for the UE's antennas is the same for any subarray antenna m, *i.e.*, $\mathbf{R}_{kl}^{\mathrm{SA}} = \mathbb{E}\{\check{\mathbf{h}}_{kln}^{\mathrm{SA}}(\check{\mathbf{h}}_{kln}^{\mathrm{SA}})^{\mathrm{H}}\}$ for any n and $\mathbf{R}_{kl}^{\mathrm{UE}} = \mathbb{E}\{\check{\mathbf{h}}_{klm}^{\mathrm{UE}}(\check{\mathbf{h}}_{klm}^{\mathrm{UE}})^{\mathrm{H}}\}$ for any m. Then,

$$\mathbf{R}_{kl}^{\mathrm{SA}} = \sum_{i=1}^{C}\omega_{kl,i}\mathbf{R}_{kl,i}^{\mathrm{SA}}, \tag{2.60}$$

$$\mathbf{R}_{kl}^{\mathrm{UE}} = \sum_{i=1}^{C}\omega_{kl,i}\mathbf{R}_{kl,i}^{\mathrm{UE}}, \tag{2.61}$$

where $\mathbf{R}_{kl,i}^{\mathrm{SA}} = \mathbb{E}\{\check{\mathbf{h}}_{kln,i}^{\mathrm{SA}}(\check{\mathbf{h}}_{kln,i}^{\mathrm{SA}})^{\mathrm{H}}\}$ and $\mathbf{R}_{kl,i}^{\mathrm{UE}} = \mathbb{E}\{\check{\mathbf{h}}_{klm,i}^{\mathrm{UE}}(\check{\mathbf{h}}_{klm,i}^{\mathrm{UE}})^{\mathrm{H}}\}$ are the contribution from the i-th cluster. Using (2.59), the subarray-side covariance matrix is given by[14]

$$\begin{aligned}\mathbf{R}_{kl}^{\mathrm{SA}} &= \sum_{i=1}^{C}\omega_{kl,i}\iiiint \mathbb{E}\{|g_{kl,i}(\varphi,\theta,\dot{\varphi},\dot{\theta})|^2\}\mathbf{a}(\varphi,\theta)[\mathbf{a}(\varphi,\theta)]^{\mathrm{H}}\mathrm{d}\varphi\mathrm{d}\theta\mathrm{d}\dot{\varphi}\mathrm{d}\dot{\theta} \\ &= \sum_{i=1}^{C}\omega_{kl,i}\underbrace{\check{\beta}_{kl,i}\iint f(\varphi-\varphi_{l,i},\theta-\theta_{l,i})\mathbf{a}(\varphi,\theta)[\mathbf{a}(\varphi,\theta)]^{\mathrm{H}}\mathrm{d}\varphi\mathrm{d}\theta}_{\mathbf{R}_{kl,i}^{\mathrm{SA}}}.\end{aligned} \tag{2.62}$$

By symmetry, the transmitter-side covariance matrix is given by

$$\mathbf{R}_{kl}^{\mathrm{UE}} = \sum_{i=1}^{C}\omega_{kl,i}\underbrace{\check{\beta}_{kl,i}\iint f(\varphi-\dot{\varphi}_{k,i},\theta-\dot{\theta}_{k,i})\dot{\mathbf{a}}(\varphi,\theta)[\dot{\mathbf{a}}(\varphi,\theta)]^{\mathrm{H}}\,\mathrm{d}\varphi\mathrm{d}\theta}_{\mathbf{R}_{kl,i}^{\mathrm{UE}}}. \tag{2.63}$$

Thus, the i-th scattering cluster's contribution on the receiver's and the transmitter's channel vectors follow zero-mean complex Gaussian distributions:

$$\check{\mathbf{h}}_{kln,i}^{\mathrm{SA}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathbf{R}_{kl,i}^{\mathrm{SA}}), \tag{2.64}$$

$$\check{\mathbf{h}}_{klm,i}^{\mathrm{UE}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathbf{R}_{kl,i}^{\mathrm{UE}}). \tag{2.65}$$

[14] This is similar to the covariance matrices derived in [68] for ULAs.

The hybrid LoS and NLoS channel vector between the k-th subarray and the m-th antenna of the l-th subarray is given by the n-th column of $\mathbf{H}_{kl}$ and is distributed as

$$\mathbf{h}_{kln}^{\mathrm{SA}} \sim \mathcal{N}_{\mathbb{C}}(\bar{\mathbf{h}}_{kln}^{\mathrm{SA}}, \mathbf{R}_{kl}^{\mathrm{SA}}), \tag{2.66}$$

where $\bar{\mathbf{h}}_{kln}^{\mathrm{SA}}$ is the LoS contribution and corresponds to the n-th column of $\overline{\mathbf{H}}_{kl}$.

Similarly, the channel vector between the l-th subarray and the n-th antenna of the k-th subarray is given by the transpose of the m-th row of $\mathbf{H}_{kl}$ and is distributed as

$$\mathbf{h}_{klm}^{\mathrm{UE}} \sim \mathcal{N}_{\mathbb{C}}(\bar{\mathbf{h}}_{klm}^{\mathrm{UE}}, \mathbf{R}_{kl}^{\mathrm{UE}}), \tag{2.67}$$

where $\bar{\mathbf{h}}_{klm}^{\mathrm{UE}}$ corresponds to the transpose of the m-th row of $\overline{\mathbf{H}}_{kl}$.

By substituting (2.28) into (2.62), we obtain the (m, m')-th entry of $\mathbf{R}_{kl}^{\mathrm{SA}}$ as

$$\left[\mathbf{R}_{kl}^{\mathrm{SA}}\right]_{m,m'} = \sum_{i=1}^{C} \omega_{kl,i}\check{\beta}_{kl,i} \iint f(\varphi - \varphi_{l,i}, \theta - \theta_{l,i}) \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}[\mathbf{k}(\varphi,\theta)]^{\mathrm{T}}(\mathbf{r}_m - \mathbf{r}_{m'})} \mathrm{d}\varphi \mathrm{d}\theta. \tag{2.68}$$

Similarly, by substituting (2.29) into (2.63), we obtain

$$\left[\mathbf{R}_{kl}^{\mathrm{UE}}\right]_{n,n'} = \sum_{i=1}^{C} \omega_{kl,i}\check{\beta}_{kl,i} \iint f(\varphi - \dot{\varphi}_{k,i}, \theta - \dot{\theta}_{k,i}) \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}[\mathbf{k}(\varphi,\theta)]^{\mathrm{T}}(\mathbf{s}_n - \mathbf{s}_{n'})} \mathrm{d}\varphi \mathrm{d}\theta. \tag{2.69}$$

Considering planar array, the spatial covariance between the antennas $(m_{\mathrm{y}}, m_{\mathrm{z}})$ and $(m'_{\mathrm{y}}, m'_{\mathrm{z}})$ of the l-th subarray is given by

$$\begin{aligned}\left[\mathbf{R}_{kl}^{\mathrm{SA}}\right]_{(m_{\mathrm{y}},m_{\mathrm{z}}),(m'_{\mathrm{y}},m'_{\mathrm{z}})} =& \sum_{i=1}^{C} \omega_{kl,i}\check{\beta}_{kl,i} \iint f(\varphi - \varphi_{l,i}, \theta - \theta_{l,i}) \\ &\times \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}\left[(m_{\mathrm{y}}-m'_{\mathrm{y}})\Delta_{\mathrm{y}}\sin(\varphi)\cos(\theta)+(m_{\mathrm{z}}-m'_{\mathrm{z}})\Delta_{\mathrm{z}}\sin(\theta)\right]} \mathrm{d}\varphi \mathrm{d}\theta,\end{aligned} \tag{2.70}$$

while the spatial covariance between the antennas $(n_{\mathrm{y}}, n_{\mathrm{z}})$ and $(n'_{\mathrm{y}}, n'_{\mathrm{z}})$ of the k-th UE is given by

$$\begin{aligned}\left[\mathbf{R}_{kl}^{\mathrm{UE}}\right]_{(n_{\mathrm{y}},n_{\mathrm{z}}),(n'_{\mathrm{y}},n'_{\mathrm{z}})} =& \sum_{i=1}^{C} \omega_{kl,i}\check{\beta}_{kl,i} \iint f(\varphi - \dot{\varphi}_{k,i}, \theta - \dot{\theta}_{k,i}) \\ &\times \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}\left[(n_{\mathrm{y}}-n'_{\mathrm{y}})\dot{\Delta}_{\mathrm{y}}\sin(\varphi)\cos(\theta)+(n_{\mathrm{z}}-n'_{\mathrm{z}})\dot{\Delta}_{\mathrm{z}}\sin(\theta)\right]} \mathrm{d}\varphi \mathrm{d}\theta.\end{aligned} \tag{2.71}$$

The total power gain of the channel between the reference antennas of the k-th UE and the the l-th subarray is given by

$$\beta_{kl} = \mathbb{E}\{|h_{kl}|^2\} = \bar{\beta}_{kl} + \check{\beta}_{kl}. \tag{2.72}$$

In practice, the power gain associated to the NLoS channel component is usually computed as

$$\check{\beta}_{kl} = \frac{\beta_0}{d_{kl}^{\gamma}}, \tag{2.73}$$

where γ is known as the path-loss exponent of the NLoS component.

2.4.3.3 The Kronecker Model

The assumption that the DoA and DoD are independent leads to a separable covariance structure. Assuming that the covariance at the receiver is independent of the covariance at the transmitter:

$$\check{\mathbf{H}}_{kl} = \sum_{i=1}^{C} \omega_{kl,i} (\mathbf{R}_{kl,i}^{\mathrm{SA}})^{1/2} \dot{\mathbf{H}}_{kl,i} (\mathbf{R}_{kl,i}^{\mathrm{UE}})^{\mathrm{T}/2}, \tag{2.74}$$

where $\dot{\mathbf{H}}_{kl,i} \in \mathbb{C}^{M \times N}$ has independent and identically distributed (i.i.d.) entries with zero mean and unit variance.

The covariance of the full channel matrix is defined as $\mathbf{R}_{kl} = \mathbb{E}\{\mathrm{vec}(\check{\mathbf{H}}_{kl})\mathrm{vec}(\check{\mathbf{H}}_{kl})^{\mathrm{H}}\}$, whose elements are $\mathbf{R}_{kl}(m, m'; n, n') = \mathbb{E}\{\check{h}_{klmn} \check{h}^{*}_{klm'n'}\}$, where $\check{h}_{klmn}$ is the NLoS channel between the m-th antenna of the l-th subarray and the n-th antenna of the k-th UE, *i.e.*, the (m, n)-th entry of $\check{\mathbf{H}}_{kl}$. The covariance matrix can then be factored as:

$$\begin{aligned}
\mathbf{R}_{kl}(m, m'; n, n') &= \sum_{i=1}^{C} \omega_{kl,i} \iiiint \mathbb{E}\{|g_{kl,i}(\varphi, \theta, \dot{\varphi}, \dot{\theta})|^2\} [\mathbf{a}(\varphi, \theta)]_m [\mathbf{a}(\varphi, \theta)]^{*}_{m'} \\
&\qquad \times [\dot{\mathbf{a}}(\dot{\varphi}, \dot{\theta})]_n [\dot{\mathbf{a}}(\dot{\varphi}, \dot{\theta})]^{*}_{n'} \, \mathrm{d}\varphi \mathrm{d}\theta \mathrm{d}\dot{\varphi} \mathrm{d}\dot{\theta} \\
&= \sum_{i=1}^{C} \omega_{kl,i} \check{\beta}_{kl,i} \iint f(\varphi - \varphi_{l,i}, \theta - \theta_{l,i}) [\mathbf{a}(\varphi, \theta)]_m [\mathbf{a}(\varphi, \theta)]^{*}_{m'} \mathrm{d}\varphi \mathrm{d}\theta \\
&\qquad \times \iint f(\varphi - \dot{\varphi}_{k,i}, \theta - \dot{\theta}_{k,i}) [\dot{\mathbf{a}}(\varphi, \theta)]_n [\dot{\mathbf{a}}(\varphi, \theta)]^{*}_{n'} \mathrm{d}\varphi \mathrm{d}\theta \\
&= \sum_{i=1}^{C} \omega_{kl,i} \frac{1}{\check{\beta}_{kl,i}} [\mathbf{R}_{kl,i}^{\mathrm{SA}}]_{m,m'} [\mathbf{R}_{kl,i}^{\mathrm{UE}}]_{n,n'}.
\end{aligned} \tag{2.75}$$

2.4.3.4 Rich Scattering Environment

If the scatterers are well distributed in the propagation environment, then we can use the same channel model described here. The only change is to consider the special case where there is only one cluster, *i.e.*, $C = 1$, and the azimuth and elevation angles follow uniform distributions within the intervals $[-\pi, \pi)$ and $[\pi/2, \pi/2)$, respectively. Then, in this case, the PDF of the angles is

$$f(\varphi, \theta) = \frac{1}{2\pi^2}, \tag{2.76}$$

and the correlation matrix at the subarray and at the UE are respectively given by

$$\mathbf{R}_{kl}^{\mathrm{SA}} = \frac{\check{\beta}_{kl}}{2\pi^2} \iint \mathbf{a}(\varphi, \theta) [\mathbf{a}(\varphi, \theta)]^{\mathrm{H}} \mathrm{d}\varphi \mathrm{d}\theta \tag{2.77}$$

and

$$\mathbf{R}_{kl}^{\mathrm{UE}} = \frac{\check{\beta}_{kl}}{2\pi^2} \iint \dot{\mathbf{a}}(\varphi, \theta) [\dot{\mathbf{a}}(\varphi, \theta)]^{\mathrm{H}} \mathrm{d}\varphi \mathrm{d}\theta. \tag{2.78}$$

As a result, the entries of the joint correlation matrix simplify to

$$\begin{aligned}\mathbf{R}_{kl}(m,m';n,n') &= \frac{1}{\check{\beta}_{kl}}[\mathbf{R}_{kl}^{\mathrm{SA}}]_{m,m'}[\mathbf{R}_{kl}^{\mathrm{UE}}]_{n,n'} \\ &= \frac{\check{\beta}_{kl}}{4\pi^4}\iint [\mathbf{a}(\varphi,\theta)]_m[\mathbf{a}(\varphi,\theta)]_{m'}^{*}\mathrm{d}\varphi\mathrm{d}\theta \iint [\dot{\mathbf{a}}(\varphi,\theta)]_n[\dot{\mathbf{a}}(\varphi,\theta)]_{n'}^{*}\mathrm{d}\varphi\mathrm{d}\theta. \quad (2.79)\end{aligned}$$

2.5 Accuracy of The Approximated Channel Models

In the previous sections, this chapter introduced the main XL-MIMO channel models. The NUSW model is the exact channel model as it accounts for the precise propagation distance between any pair of antennas. Unlike the approximations, it incorporates non-uniform channel gain across the array and describes a spherical wavefront. In contrast, the approximated models (USW, UPBW, and UPW) invoke simplifications: the USW model imposes a uniform channel gain assumption, while the UPBW and UPW models further approximate the wavefront curvature via second-(parabolic) and first-order (planar) Taylor expansions of the distance, respectively. The closer the transmitter and the receiver, the less accurate these approximations become, as the effects of wavefront curvature and propagation distance variations across the array become more significant.

This section systematically evaluates the accuracy of these three approximated models. Consider a square array with M antennas in the y-z plane and aperture D, such that its horizontal and vertical sides are both $D_{\mathrm{y}} = D_{\mathrm{z}} = D/\sqrt{2}$. The reference (central) antenna is located at the point $[0,0,0]^{\mathrm{T}}$, and a single-antenna UE is positioned at $d\mathbf{k}(\varphi,\theta)$, *i.e.*, at a distance d from the subarray's reference antenna. The NUSW channel response at the m-th antenna is:

$$h_m^{\mathrm{NUSW}}(d,\varphi,\theta) = \frac{\alpha_0}{d_m(d,\varphi,\theta)}\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}d_m(d,\varphi,\theta)}, \quad (2.80)$$

where

$$d_m(d,\varphi,\theta) = \sqrt{d^2 - 2d\,\mathbf{k}^{\mathrm{T}}(\varphi,\theta)\,\mathbf{r}_m + \|\mathbf{r}_m\|^2} \quad (2.81)$$

is the precise distance from the UE to the m-th antenna, and $\alpha_0 = \sqrt{\beta_0} = \frac{\lambda}{4\pi}$ is the free-space channel gain at 1 m.

The USW, UPBW and UPW models all assume uniform channel gain, effectively replacing the individual antenna distance with that to the reference element. The USW model uses the exact phase from NUSW, but applies the uniform gain:

$$h_m^{\mathrm{USW}}(d,\varphi,\theta) = \frac{\alpha_0}{d}\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}d_m(d,\varphi,\theta)}. \quad (2.82)$$

The UPBW model applies the uniform gain and substitutes the exact propagation distance by its second-order Taylor approximation, when computing the phase, thus modeling a parabolic wavefront:

$$h_m^{\mathrm{UPBW}}(d,\varphi,\theta) = \frac{\alpha_0}{d}\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}d_m^{\mathrm{parab}}(d,\varphi,\theta)}, \quad (2.83)$$

where

$$d_m^{\mathrm{parab}}(d,\varphi,\theta) = d - \mathbf{k}^{\mathrm{T}}(\varphi,\theta)\mathbf{r}_m + \frac{\|\mathbf{r}_m\|^2 - [\mathbf{k}^{\mathrm{T}}(\varphi,\theta)\mathbf{r}_m]^2}{2d}. \tag{2.84}$$

The UPW model is even more simplified, substituting the distance by its first-order expansion, thus assuming a planar wavefront:

$$h_m^{\mathrm{UPW}}(d,\varphi,\theta) = \frac{\alpha_0}{d}\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}d_m^{\mathrm{plane}}(d,\varphi,\theta)}, \tag{2.85}$$

where

$$d_m^{\mathrm{plane}}(d,\varphi,\theta) = d - \mathbf{k}^{\mathrm{T}}(\varphi,\theta)\mathbf{r}_m. \tag{2.86}$$

The next subsection quantifies the accuracy of the uniform channel gain assumption inherent in the USW, UPBW, and UPW models, while the subsequent subsection analyzes the phase approximation accuracy of the UPBW and UPW models. Additionally, these analyses illuminate the dependence of approximation accuracy on propagation distance, array aperture, and antenna number.

2.5.1 Uniform Channel Gain Assumption

The accuracy of the uniform channel gain assumption can be quantified by the ratio of the exact channel power to the approximated one. Values closer to 1 indicate a more accurate approximation. For the m-th antenna, this ratio is given by

$$\tilde{\alpha}_m(d,\varphi,\theta) = \frac{|h_m^{\mathrm{NUSW}}(d,\varphi,\theta)|^2}{|h_m^{\mathrm{USW}}(d,\varphi,\theta)|^2} = \frac{d^2}{d_m^2(d,\varphi,\theta)}. \tag{2.87}$$

Figure 2.8 displays a heatmap of this accuracy metric for a setup with a UE at 10 m from a square array comprising $M = 625$ antennas ($M_{\mathrm{y}} = \mathrm{z} = 25$) and an antenna spacing of $\Delta_{\mathrm{y}} = \Delta_{\mathrm{z}} = 0.1$ m.[15] The heatmap represents the average accuracy over all M antennas as a function of the azimuth and elevation angles, *i.e.*, $\frac{1}{M}\sum_{m=1}^{M}\tilde{\alpha}_m(10\ \mathrm{m},\varphi,\theta)$. Notably, one of the directions where the USW model exhibits its poorest performances in approximating the channel gain is the broadside direction, *i.e.*, $(\varphi,\theta) = (0,0)$.

To determine the minimum link distance for which the uniform gain approximation is sufficiently accurate, it is necessary to consider the worst-case error. This occurs at the antenna farthest from the array center (i.e., the M-th antenna) with the UE placed at broadside ($\varphi = 0$, $\theta = 0$):

$$\tilde{\alpha}^*(d) = \tilde{\alpha}_M(d,0,0) = \frac{d^2}{d_M^2(d,0,0)} = \frac{d^2}{d^2 + \|\mathbf{r}_M\|^2}. \tag{2.88}$$

Assuming M_y and M_z are both even and that $M_y = M_z = \sqrt{M}$, with half-wavelength spacing in each dimension ($\Delta_y = \Delta_z = \lambda/2$), the position vector of the most

[15] This configuration yields an array aperture of approximately $D = 3.5$ m.

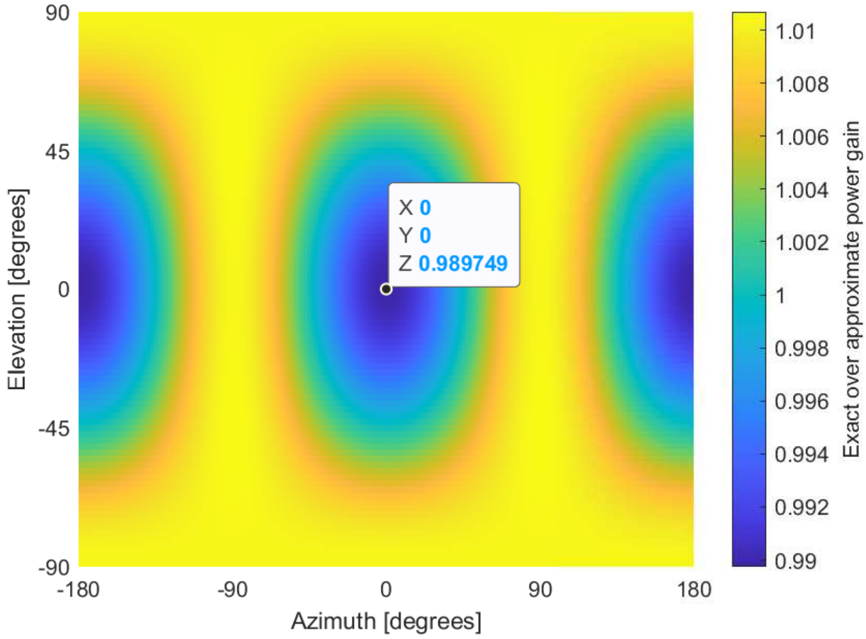


Figure 2.8 – Average ratio of exact to approximated channel power gain as a function of azimuth and elevation angles. Each value represents an average across all the array antennas. One of the directions where the USW model exhibits its poorest performances in approximating the channel gain is the broadside direction, *i.e.*, $(\varphi, \theta) = (0, 0)$.

distant antenna can be written as $\mathbf{r}_M = \left[0, \frac{\lambda}{2}\frac{M_\mathrm{y}}{2}, \frac{\lambda}{2}\frac{M_\mathrm{z}}{2}\right]^\mathrm{T} = \frac{\lambda\sqrt{M}}{4}\,[0,1,1]^\mathrm{T}$, based on (2.3). Therefore,

$$\tilde{\alpha}^*(d) = \frac{8d^2}{8d^2 + \lambda^2 M}. \tag{2.89}$$

Alternatively, the position of the M-th antenna can be expressed in terms of array aperture as $\mathbf{r}_M \approx \frac{1}{2}[0, D_\mathrm{y}, D_\mathrm{z}]^\mathrm{T} = \frac{D}{2\sqrt{2}}[0,1,1]^\mathrm{T}$. This results in

$$\tilde{\alpha}^*(d) \approx \frac{4d^2}{4d^2 + D^2}. \tag{2.90}$$

Using (2.90), Figure 2.9 presents the worst-case channel power gain accuracy as a function of array aperture and distance: the uniform gain approximation holds as the distance greatly exceeds the array aperture, or for sufficiently compact arrays.[16] For extra-large arrays (e.g., $D = 10$ m), achieving 95% accuracy requires distances above roughly 40 m, indicating that the USW channel model becomes inaccurate for nearby users, which is often the case in XL-MIMO deployments.

Next, Figure 2.10 demonstrates how the number of antennas, with half-wavelength inter-element spacing, affects the UPD—which is defined in this section as the minimum

[16] The USW model is accurate when the distance between the UE and the M-th antenna approximately the same as the distance between the UE and the reference antenna. As shown in Figure 2.9, it happens when the UE is far enough from the array, or when the array aperture is small enough.

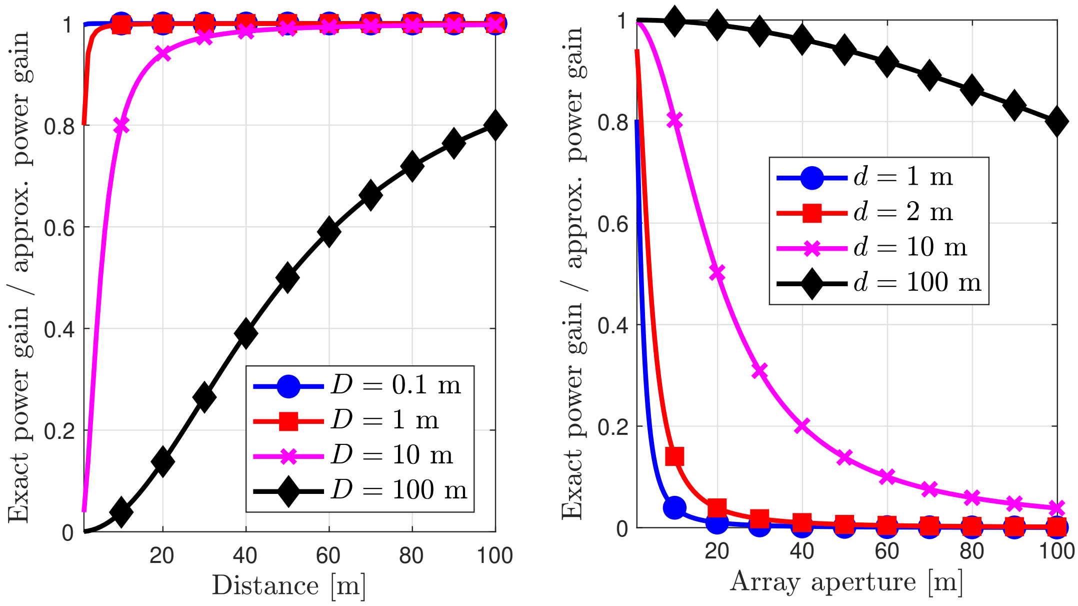


Figure 2.9 – Accuracy of the uniform channel gain assumption as a function of distance and array aperture.

distance at which the uniform gain approximation is sufficiently accurate. When the antenna spacing remains fixed, increasing M expands the array aperture, thereby increasing the required UPD for the uniform gain assumption to hold. For example, the UPD of an extremely large array with 1024 antennas reaches approximately 10 m for 99% accuracy. However, partitioning the XL-MIMO array into smaller subarrays allows the USW model to hold in each subarray if modeled independently. For example, using $M_y = M_z = 8$ ($M = 64$) reduces the UPD to approximately 2 m.

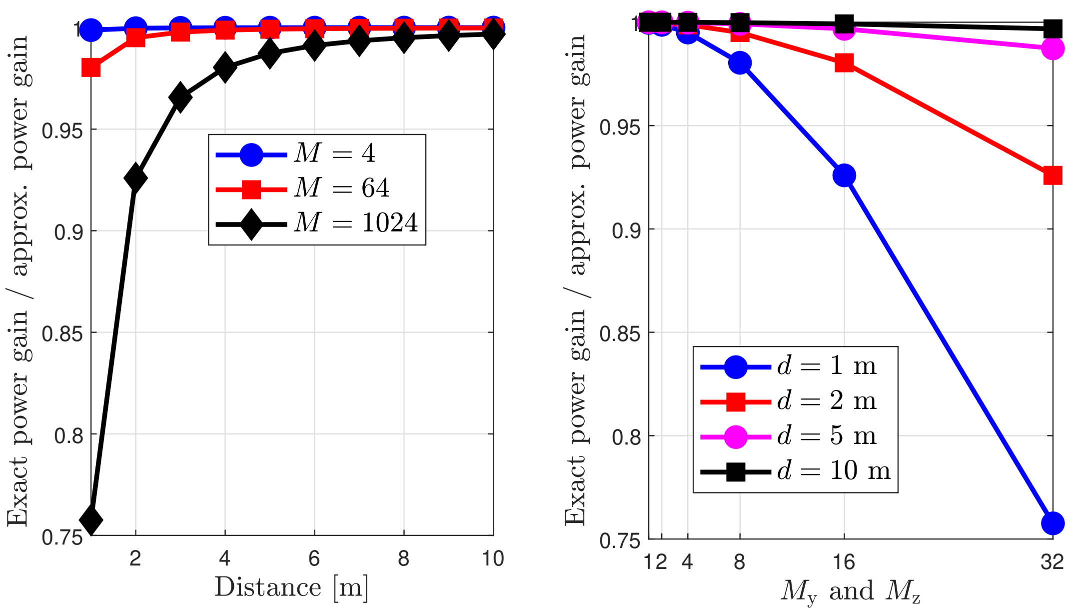


Figure 2.10 – Accuracy of the uniform channel gain assumption as a function of distance and number of antennas.

2.5.2 Parabolic and Plane Wavefront Assumptions

The accuracy of approximating the wavefront as parabolic or planar can be measured by the phase error. At the m-th antenna, it is given by

$$\tilde{\phi}_m^{\mathrm{UPBW}}(d,\varphi,\theta) = \angle h_m^{\mathrm{UPBW}}(d,\varphi,\theta) - \angle h_m^{\mathrm{NUSW}}(d,\varphi,\theta) \tag{2.91}$$

$$= \frac{2\pi}{\lambda}\left(d_m(d,\varphi,\theta) - d_m^{\mathrm{parab}}(d,\varphi,\theta)\right) \tag{2.92}$$

for the UPBW channel model. Similarly, the phase error of the UPW model is

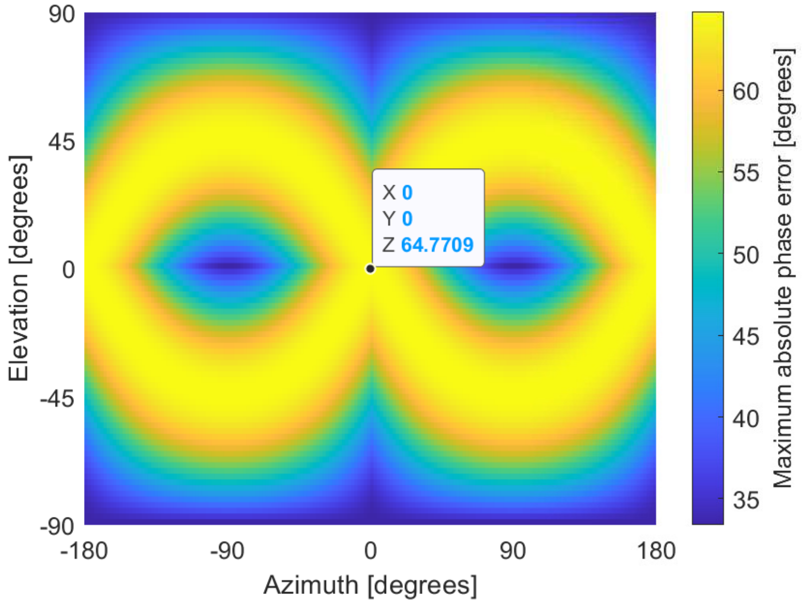
$$\tilde{\phi}_m^{\mathrm{UPW}}(d,\varphi,\theta) = \frac{2\pi}{\lambda}\left(d_m(d,\varphi,\theta) - d_m^{\mathrm{plane}}(d,\varphi,\theta)\right). \tag{2.93}$$

Figure 2.11 presents heatmaps of the maximum absolute phase error, evaluated over all M antennas, as a function of azimuth and elevation, for the both the UPW and UPBW channel models. The simulation setup considers a UE positioned 10 m from the reference antenna of a square array comprising $M = 625$ antennas ($M_{\mathrm{y}} = M_{\mathrm{z}} = 25$) with an inter-element spacing of 0.025 m, operating at a 6 GHz carrier frequency. For each angular direction, the plotted error value corresponds to the largest absolute phase error among all antennas. According to Figure 2.11a, the UPW model exhibits one of its greatest errors at the broadside direction, $(\varphi,\theta) = (0,0)$. In contrast, Figure 2.11b demonstrates that the UPBW model attains one of its maximum errors around $(\varphi,\theta) = (\pi/2,0)$.

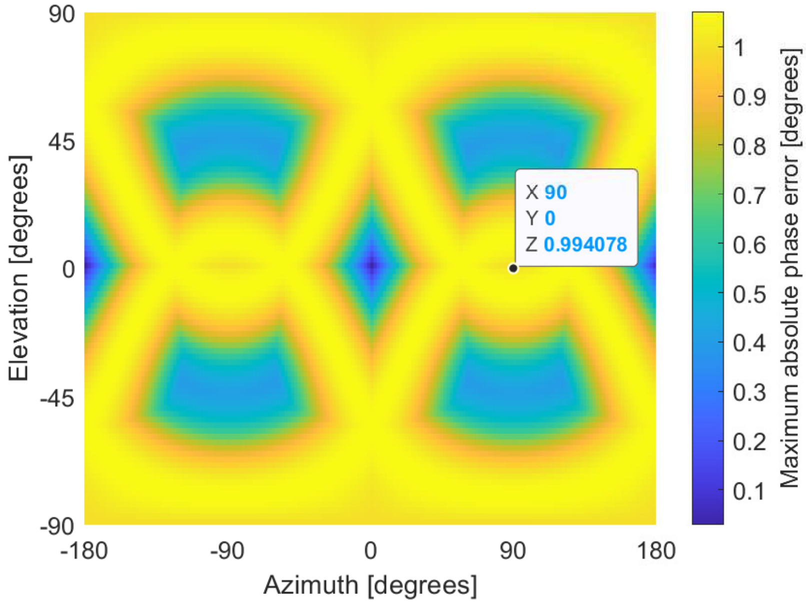
With the same simulation setup, Figure 2.12 illustrates how the absolute phase approximation error varies with both the distance d between the UE and the array's reference antenna (left figure) and the number of antennas (right figure). Each subplot presents the maximum error at the M-th antenna—the element farthest from the reference antenna, where the approximation is typically least accurate. Results are shown for the worst-case angular direction, such that the plotted curves correspond to $\tilde{\phi}_M^{\mathrm{UPBW}}(d,\pi/2,0)$ and $\tilde{\phi}_M^{\mathrm{UPW}}(d,0,0)$. As expected, both models exhibit improved accuracy as the UE moves farther from the array, while accuracy degrades with increasing antenna number (and, consequently, growing array aperture at fixed half-wavelength spacing). Crucially, the UPBW model consistently outperforms the UPW, maintaining much lower phase errors even for moderate array apertures and small propagation distances. Partitioning the XL-MIMO array into smaller subarrays enables highly accurate modeling with the UPBW model (and often also the UPW), provided the channels are modeled independently in each subarray. For instance, using $M_{\mathrm{y}} = M_{\mathrm{z}} = 8$ ($M = 64$) reduces the maximum (at the M-th antenna) phase error to only 0.036° for the UPBW model and 7.2° for the UPW model when the UE is as close as 10 m to the subarray.

2.6 Summary

This chapter provided a detailed investigation of propagation modeling for XL-MIMO systems in micro-urban scenarios, focusing on the case where both the transmitter



(a) UPW channel model



(b) UPBW channel model

Figure 2.11 – Maximum absolute phase error across the $M = 625$ antennas as a function of azimuth and elevation angles. The evaluation considers a UE located at $d = 10$ m from the array operating at 6 GHz. (a) The UPW model exhibits significant phase errors, notably at the broadside direction $(0, 0)$. (b) The UPBW model substantially reduces the error magnitude across the angular domain.

and receiver are equipped with uniform planar arrays. It began with a rigorous examination of electromagnetic field regions, highlighting the limitations of the classical far-field (Fraunhofer) and radiative near-field (Fresnel) approximations for large-aperture arrays.

A hybrid channel model was then formulated, incorporating both line-of-sight

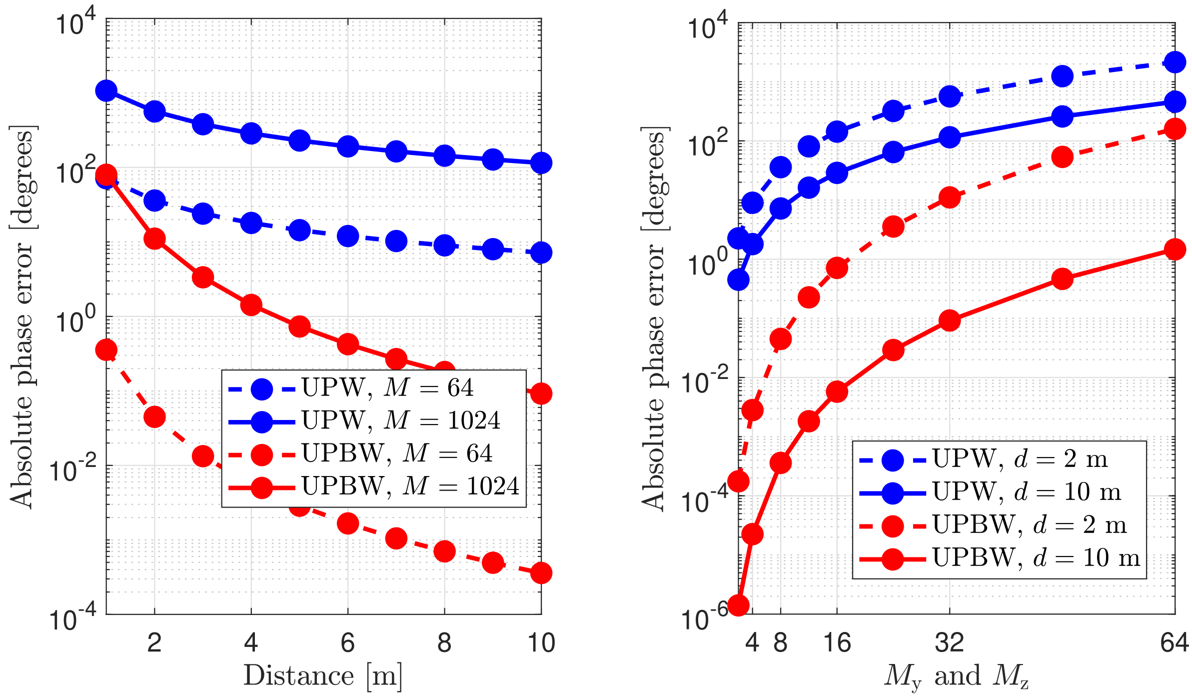


Figure 2.12 – Variation of the absolute phase error on the UE-to-array distance (left figure) and on the number of antennas (right figure).

(LoS) and non-line-of-sight (NLoS) components. Three distinct wave models were considered in detail: non-uniform spherical wave (NUSW), uniform parabolic wave (UPBW), and uniform plane wave (UPW). The NUSW model captures full near-field effects and spatial non-stationarity by accounting for both channel gain variations and non-linear phase variations across the antennas due to spherical wavefront. In contrast, the UPBW and UPW models approximate the path-loss as uniform across the array and approximate the phase by modeling the wavefront as parabolic and planar, respectively. The parabolic approximation remains accurate in the radiative near-field, whereas the UPW model assumes infinitely distant users and neglects curvature, thus being accurate only in the far-field.

For the NLoS component, a clustered scattering model was adopted, and both discrete and continuous scattering models were derived. The discrete model represents the channel as a superposition of waves from individual scatterers, while the continuous model integrates over angular spectrum, capturing richer scattering environments.

Quantitative analyses compared the amplitude and phase errors of the UPW and UPBW approximations against the reference NUSW model as functions of user distance, array aperture, and angular direction. Results show that, for typical urban deployments and moderate aperture sizes, UPBW maintains near-exact phase and amplitude accuracy well into the radiative near-field, while UPW incurs significant phase errors, especially for users close to the array.

Building on the propagation framework developed here, Chapter 3 derives deterministic SINR approximations for both centralized and distributed processing.

3 DETERMINISTIC SINR EXPRESSIONS FOR XL-MIMO

Deterministic SINR approximations are powerful tools for system analysis and design because they depend only on channel statistics, which vary far more slowly than the instantaneous channel. Since LSF changes primarily with user location, these formulas need re-evaluation only when user positions change, not in every channel coherence block. This approach significantly reduces computational load and eliminates the need for frequent instantaneous CSI updates from the subarrays to the CPU, allowing resource allocation algorithms to operate over extended periods with a single SINR calculation. This statistical method is particularly effective in XL-MIMO, where the channel hardening effect makes detection based solely on channel statistics nearly optimal.

Consequently, these closed-form expressions provide a tractable way to address real-world optimization problems. They enable the streamlined design of resource allocation algorithms by avoiding computationally expensive Monte Carlo simulations and offering valuable analytical insights into system behavior. For instance, they can be used to quantify the impact of key system parameters—such as the XL-MIMO architecture, power control strategies, estimation errors, and pilot contamination—on array gains and interference terms [20].

Most studies that derive deterministic SINR or rate expressions for XL-MIMO systems rely on the Rayleigh fading model. While this assumption is analytically convenient, it fails to capture channel characteristics when a LoS component is present. As ultra-dense deployments in 6G networks lead to shorter propagation distances and more prevalent LoS links, Rician fading becomes a more appropriate and realistic model for performance analysis. This motivates the derivation of novel deterministic SINR expressions that are valid for XL-MIMO channels incorporating LoS and spatially correlated NLoS components, as well as VRs. Furthermore, this thesis advances beyond the common assumptions of perfect CSI and simpler MR or ZF combiners. The expressions derived herein uniquely account for channel estimation errors and the use of MMSE receive combining, which distinguishes this work from most existing studies.

This chapter derives two types of deterministic SINR expressions—asymptotic and ergodic—for both centralized and distributed XL-MIMO uplink operations. The **asymptotic expressions** approximate the *instantaneous* SINR as the number of subarray's antennas (M) grows large. For the centralized architecture, the approximation applies to the *global* SINR as $ML_k \to \infty$, where L_k is the number of subarrays serving UE k, and is highly accurate when the number of scheduled UEs, denoted by U, is no greater than the number of serving antennas ($U \leq ML_k$). Similarly, for the distributed architecture, it approximates the *local* SINR as $M \to \infty$, being highly accurate when $U \leq M$. In contrast, the **ergodic expressions** approximate the *average instantaneous* SINR and are highly accurate in crowded scenarios where the number of UEs is much larger than the number of antennas. This condition is $U \gg ML_k$ for centralized operation and $U \gg M$ for distributed operation. Table 3.1 consolidates this information for easy reference.

Table 3.1 – Characterization of the deterministic uplink SINR expressions for centralized and distributed operation.

Uplink Operation	Asymptotic SINR	Ergodic SINR
Centralized	Approximates the **instantaneous global SINR** when $ML_k \to \infty$ and $U \leq ML_k$.	Approximates the **mean global SINR** when $U \gg ML_k$.
Distributed	Approximates the **instantaneous local SINR** when $M \to \infty$ and $U \leq M$.	Approximates the **mean local SINR** when $U \gg M$.

3.1 XL-MIMO System Model

Consider the particular case of K single-antenna UEs trying to communicate to a BS equipped with an extra-large aperture antenna UPA. Let $\mathcal{D} = \{1, 2, \ldots, L\}$ and $\mathcal{K} = \{1, 2, \ldots, K\}$ be the subarray and UE index sets, respectively. At a given time-frequency block, $U = |\mathcal{U}|$ UEs are scheduled, *i.e.*, served by at least one subarray, where the set $\mathcal{U} \subseteq \mathcal{K}$ contains the scheduled UEs. UE k is served by $L_k = |\mathcal{D}_k|$ subarrays, where the set $\mathcal{D}_k \subseteq \mathcal{D}$ contains the indices of these subarrays.

Conversely, subarray l serves $U_l = |\mathcal{U}_l|$ UEs, and $\mathcal{U}_l \subseteq \mathcal{U}$ contains the indices of these UEs. This means that each subarray receives a list of UEs that it should serve. For notational convenience, define the matrix $\mathbf{D}_{kl} \in \mathbb{C}^{M \times M}$ as the identity matrix $\mathbf{I}_M$ if subarray l is allowed to decode signals from UE k and $\mathbf{0}_M$ otherwise:

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_M & \text{if } l \in \mathcal{D}_k, \\ \mathbf{0}_M & \text{otherwise.} \end{cases} \tag{3.1}$$

Define also the collective matrix $\mathbf{D}_k \in \mathbb{C}^{ML \times ML}$ indicating which of the M antennas take part in decoding signals from UE k:

$$\mathbf{D}_k = \text{blkdiag}\left(\mathbf{D}_{k1}, \mathbf{D}_{k2}, \ldots, \mathbf{D}_{kL}\right). \tag{3.2}$$

Because each UE is only near a limited number of subarrays, it is unnecessary to serve all UEs with all subarrays. If, for example, $\mathcal{D}_k$ contains the closest subarrays to UE k, performance loss may be negligible, while fronthaul signaling and computational load are both substantially reduced, and network scalability is ensured. The sets $\mathcal{D}_k$ are dynamic, which means they can be adapted to time-variant characteristics, such as UE locations, service requirements, interference situation, etc [9]. Moreover, these sets are assumed to be known everywhere needed, since dynamic variations occur over much larger time intervals than SSF variations.

The position of the m-th antenna of any subarray relative to its reference antenna is denoted by $\mathbf{r}_m$ and given by (2.3).[1] The position of UE k relative to the l-th subarray's reference antenna is $\mathbf{s}_{kl} = d_{kl}\mathbf{k}_{kl}$, where $d_{kl} = \|\mathbf{s}_{kl}\|$ is the distance between them, and $\mathbf{k}_{kl} = \mathbf{k}(\varphi_{kl}, \theta_{kl})$ is a unitary vector pointing from the l-th subarray's reference antenna to the k-th UE. This vector is given by (2.8), and φ_{kl} and θ_{kl} are respectively the azimuth and elevation angles between the k-th UE and the l-th subarray's reference antenna.

Based on the channel model described in Chapter 2, the LoS channel vector between the k-th UE and the l-th subarray is given by

$$\bar{\mathbf{h}}_{kl} = \omega_{kl}\sqrt{\bar{\beta}_{kl}}\mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}d_{kl}}\mathbf{a}_{kl}, \tag{3.3}$$

where ω_{kl} is a Bernoulli random variable indicating whether there is LoS between the k-th UE and the l-th subarray (1) or not (0), and $\mathbf{a}_{kl}$ is the array response vector. The channel power gain associated to the LoS path is given by[2,3]

$$\bar{\beta}_{kl} = \frac{1}{M}\|\bar{\mathbf{h}}_{kl}\|^2 = \omega_{kl}\frac{\lambda^2}{(4\pi)^2 d_{kl}^2}. \tag{3.4}$$

Considering that the UPW channel model is accurate from the subarray's point of view, the m-th element of the array response vector $\mathbf{a}_{kl} = \mathbf{a}(\varphi_{kl}, \theta_{kl})$ is given by

$$[\mathbf{a}(\varphi, \theta)]_m = \mathrm{e}^{\mathrm{j}\frac{2\pi}{\lambda}[\mathbf{k}(\varphi,\theta)]^{\mathrm{T}}\mathbf{r}_m}. \tag{3.5}$$

Assuming the scatterers are grouped into C clusters, the stochastic NLoS component is given by

$$\check{\mathbf{h}}_{kl} = \sum_{i=1}^{C}\omega_{kl,i}\check{\mathbf{h}}_{kl,i}, \tag{3.6}$$

where $\omega_{kl,i}$ indicates whether the scatterers in the i-th cluster are simultaneously visible to UE k and subarray l or not, and $\check{\mathbf{h}}_{kl,i}$ is the contribution from the i-th cluster. The NLoS component follows a zero-mean complex Gaussian distribution:

$$\check{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_{kl}), \tag{3.7}$$

1 For an arbitrary subarray l, with $l \in \{1, \ldots, L\}$, its reference antenna is located at $[0, 0, 0]^{\mathrm{T}}$ in this subarray's local coordinate system.

2 Some channel models consider the shadow fading, which can be represented by a log-normal variation around the LSF channel gain. The shadowing is described in Appendix G.

3 Considering small to moderate subarray aperture, the variations in propagation distance among its M antennas are negligible for LSF, allowing us to treat the LSF gain as constant within the subarray. However, even these slight distance differences produce distinct phase shifts at each antenna, significantly impacting the received signal phase across the subarray.

such that the overall channel vector, given by

$$\mathbf{h}_{kl} = \overline{\mathbf{h}}_{kl} + \check{\mathbf{h}}_{kl}, \tag{3.8}$$

also follows a complex Gaussian distribution, whose mean value is the LoS channel component, as described by

$$\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\overline{\mathbf{h}}_{kl}, \mathbf{R}_{kl}). \tag{3.9}$$

The channel covariance matrix $\mathbf{R}_{kl} = \mathbb{E}\{\check{\mathbf{h}}_{kl}\check{\mathbf{h}}_{kl}^{\mathrm{H}}\}$ is given by

$$\mathbf{R}_{kl} = \sum_{i=1}^{C} \omega_{kl,i}\check{\beta}_{kl,i} \iint f(\varphi - \varphi_{l,i}, \theta - \theta_{l,i})\mathbf{a}(\varphi, \theta)[\mathbf{a}(\varphi, \theta)]^{\mathrm{H}} \mathrm{d}\varphi \mathrm{d}\theta, \tag{3.10}$$

where $\varphi_{l,i}$ and $\theta_{l,i}$ are the azimuth and elevation angles from the center of this cluster to the l-th subarray, respectively, $f(\varphi - \varphi_{l,i}, \theta - \theta_{l,i})$ is the joint PDF for the DoA from the i-th cluster, given by (2.58), with standard deviations denoted by σ_φ and σ_θ, and $\check{\beta}_{kl,i}$ is the power gain associated to the NLoS paths that passes through the scatterers in cluster i, which is basically composed by the channel gain attenuation along the distance $\dot{d}_{k,i}$ between UE k and cluster i, a random attenuation c_i due to refraction and energy absorption at this cluster, and the channel gain attenuation along the distance $d_{l,i}$ between this cluster and subarray l. Thus,

$$\check{\beta}_{kl,i} = \frac{\omega_{kl,i} c_i^2 \lambda^4}{(4\pi)^4 \dot{d}_{k,i}^2 d_{l,i}^2}. \tag{3.11}$$

The LSF channel power gain associated to all the NLoS paths is given by[4]

$$\check{\beta}_{kl} = \frac{1}{M}\mathbb{E}\{\|\check{\mathbf{h}}_{kl}\|^2\} = \frac{1}{M}\,\mathrm{tr}(\mathbf{R}_{kl}) = \sum_{i=1}^{C} \omega_{kl,i}\check{\beta}_{kl,i}. \tag{3.12}$$

Finally, the overall LSF channel power gain between UE k and subarray l is

$$\beta_{kl} = \frac{1}{M}\mathbb{E}\{\|\mathbf{h}_{kl}\|^2\} = \frac{1}{M}\,\mathrm{tr}(\mathbf{Q}_{kl}) = \bar{\beta}_{kl} + \check{\beta}_{kl}, \tag{3.13}$$

where $\mathbf{Q}_{kl} = \mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kl}^{\mathrm{H}}\}$ is the channel correlation matrix, which can be computed as

$$\mathbf{Q}_{kl} = \mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kl}^{\mathrm{H}}\} = \overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{kl}^{\mathrm{H}} + \mathbf{R}_{kl}. \tag{3.14}$$

The average LSF channel gain for user k across the L subarrays is obtained as

$$\beta_k = \frac{1}{L}\sum_{l=1}^{L} \beta_{kl}. \tag{3.15}$$

[4] The expectation operator is only in (3.12) and not in (3.4) because the LoS component is assumed to be deterministic, since it varies much slower than the NLoS component.

In modular XL-MIMO, the large inter–subarray spacing justifies modeling the NLoS channels of different subarrays as statistically independent. Hence,

$$\mathbb{E}\{\check{\mathbf{h}}_{kl}\check{\mathbf{h}}_{k'l'}^{\mathrm{H}}\} = \begin{cases} \mathbf{R}_{kl} & \text{if } (k', l') = (k, l), \\ \mathbf{0}_{M\times M} & \text{otherwise.} \end{cases} \tag{3.16}$$

As a result, the overall channel covariance matrix for user k, defined as $\mathbf{R}_k = \mathbb{E}\{\check{\mathbf{h}}_k\check{\mathbf{h}}_k^{\mathrm{H}}\}$, is assumed as a block-diagonal matrix:

$$\mathbf{R}_k = \mathrm{blkdiag}\left(\mathbf{R}_{k1}, \mathbf{R}_{k2}, \ldots, \mathbf{R}_{kL}\right), \tag{3.17}$$

and the collective channel vector $\mathbf{h}_k = [\mathbf{h}_{k1}^{\mathrm{T}},\ \ldots,\ \mathbf{h}_{kL}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{ML}$ follows the distribution

$$\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\overline{\mathbf{h}}_k, \mathbf{R}_k), \tag{3.18}$$

where $\overline{\mathbf{h}}_k = [\overline{\mathbf{h}}_{k1}^{\mathrm{T}},\ \ldots,\ \overline{\mathbf{h}}_{kL}^{\mathrm{T}}]^{\mathrm{T}}$ and $\check{\mathbf{h}}_k = [\check{\mathbf{h}}_{k1}^{\mathrm{T}},\ \ldots,\ \check{\mathbf{h}}_{kL}^{\mathrm{T}}]^{\mathrm{T}}$.

3.1.1 Uplink Pilot Transmission

The statistical properties — namely, the mean vector and the covariance matrix — are assumed to be perfectly known at the CPU, as they remain constant over hundreds of channel coherence blocks. In practice, these parameters can be estimated accurately via pilot-based training during connection establishment. The subarrays and the CPU can use sample means and sample covariance matrices to obtain this information, which, given its slow variation, remains unchanged over multiple coherence intervals [64]. UEs transmit dedicated pilots so that the BS estimates the channel statistics with affordable resource consumption.

On the other hand, the instantaneous channels can be estimated using uplink pilot signaling and channel reciprocity [8]. The estimates can be either computed at the CPU or at the subarray. Both options give the same estimation quality. The CSI is used for coherent uplink receive combining and downlink transmit precoding [69].

During one channel coherence time, τ_{c} symbols can be transmitted. The first τ_{p} symbols are used for pilot transmission, for the BS to acquire CSI, while τ_{u} symbols are used for data transmission in the uplink and τ_{d} symbols are used for data transmission in the downlink. Here, $\tau_{\mathrm{c}} = \tau_{\mathrm{p}} + \tau_{\mathrm{u}}$, since only uplink is considered.

The system operates in time-division-duplexing (TDD) mode.[5] Within each channel coherence block, we can transmit τ_{c} symbols. The channels must be estimated once per coherence block, requiring τ_{p} symbols to be reserved for uplink pilot signaling, where τ_{p} is the length of the pilot sequences. The remaining $\tau_{\mathrm{c}} - \tau_{\mathrm{p}}$ symbols are allocated for data transmission. This work focuses exclusively on uplink data transmission.

[5] For more details about the TDD mode, see Section I.1 in Appendix I.

The pilot assigned to UE k is denoted as t_k, and the corresponding sequence is denoted as $\boldsymbol{\phi}_{t_k} \in \mathbb{C}^{\tau_\mathrm{p}}$. To obtain a constant power level, the sequences are assumed to have unit-magnitude elements [8], and they form an orthogonal set, such that[6]

$$\boldsymbol{\phi}_{t_k}^\mathrm{H} \boldsymbol{\phi}_{t_i} = \begin{cases} \tau_\mathrm{p}, & \text{if } t_i = t_k, \\ 0, & \text{otherwise.} \end{cases} \tag{3.19}$$

Accurate channel estimation at the BS requires knowledge of each UE's pilot sequence. Let $\mathcal{P}_t \in \mathcal{K}$ denote the set of UEs using pilot t:

$$\mathcal{P}_t = \{ k \in \mathcal{K} \mid t_k = t \} . \tag{3.20}$$

The elements of $\sqrt{p_k}\boldsymbol{\phi}_{t_k}$ are transmitted by UE k over τ_p consecutive samples at the beginning of the coherence block, with transmit power p_k. Thus, during the entire pilot transmission, the received signal $\mathbf{Y}_l^\mathrm{pilot} \in \mathbb{C}^{M \times \tau_\mathrm{p}}$ at subarray l is

$$\mathbf{Y}_l^\mathrm{pilot} = \sum_{i \in \mathcal{U}} \sqrt{p_i} \mathbf{h}_{il} \boldsymbol{\phi}_{t_i}^\mathrm{T} + \mathbf{N}_l, \tag{3.21}$$

where $\mathbf{N}_l \in \mathbb{C}^{ML \times \tau_\mathrm{p}}$ is the receiver noise with i.i.d. elements following a complex Gaussian distribution $\mathcal{N}_\mathbb{C}(0, \sigma_\mathrm{n}^2)$.

The channel estimation may be performed locally at each subarray or centrally at the CPU, yet the uplink pilot transmission model remains identical in both scenarios.

Consider that the pilot assigned to UE k is $t_k = t$. The first step in estimating the channel $\mathbf{h}_{kl}$ based on the received signal $\mathbf{Y}_l^\mathrm{pilot}$ is to remove the interference from UEs that use other pilots, which is done by multiplying $\mathbf{Y}_l^\mathrm{pilot}$ with the conjugate of $\boldsymbol{\phi}_t$ [9]:

$$\mathbf{y}_{tl}^\mathrm{pilot} = \mathbf{Y}_l^\mathrm{pilot} \boldsymbol{\phi}_t^*, \tag{3.22}$$

resulting in a superposition of the channels from all UEs that use pilot t (including UE k), plus a noise term $\mathbf{n}_{tl} = \mathbf{N}_l \boldsymbol{\phi}_t^* \sim \mathcal{N}_\mathbb{C}(\mathbf{0}, \tau_\mathrm{p} \sigma_\mathrm{n}^2 \mathbf{I}_M)$:

$$\mathbf{y}_{tl}^\mathrm{pilot} = \sum_{i \in \mathcal{P}_t} \sqrt{p_i} \tau_\mathrm{p} \mathbf{h}_{il} + \mathbf{n}_{tl}. \tag{3.23}$$

Two widely used channel estimators are the LS and MMSE estimators. The LS estimator is simple but cannot distinguish channels of UEs sharing the same pilot sequence. In other words, the estimates for UEs k and i are identical if $t_k = t_i$. Under the assumption that the channel means are known, the performance is improved. In contrast, the MMSE estimator exploits statistical channel knowledge to separate pilot-sharing UEs and, given the observation $\mathbf{y}_{t_k l}^\mathrm{pilot}$ and the channel statistics, it minimizes the estimation mean square error (MSE). This improved performance, however, implies additional complexity due to matrix inversions and multiplications.

6 A set of τ_p orthogonal sequences of length τ_p, for any τ_p, is obtained using Fourier complex matrices, for example.

3.1.2 MMSE Channel Estimation

The MMSE estimate of $\mathbf{h}_{kl}$ is [69]:

$$\widehat{\mathbf{h}}_{kl} = \overline{\mathbf{h}}_{kl} + \sqrt{p_k}\mathbf{R}_{kl}\mathbf{\Psi}_{t_k l}^{-1}(\mathbf{y}_{t_k l}^{\text{pilot}} - \overline{\mathbf{y}}_{t_k l}^{\text{pilot}}), \tag{3.24}$$

where

$$\mathbf{\Psi}_{tl} = \frac{1}{\tau_{\text{p}}}\text{Cov}(\mathbf{y}_{tl}^{\text{pilot}}) = \sum_{i \in \mathcal{P}_t} p_i \tau_{\text{p}} \mathbf{R}_{il} + \sigma_{\text{n}}^2 \mathbf{I}_M \tag{3.25}$$

is the normalized covariance matrix of $\mathbf{y}_{tl}^{\text{pilot}}$, and

$$\overline{\mathbf{y}}_{tl}^{\text{pilot}} = \sum_{i \in \mathcal{P}_t} \sqrt{p_i}\tau_{\text{p}} \overline{\mathbf{h}}_{il}. \tag{3.26}$$

The MMSE estimate $\widehat{\mathbf{h}}_{kl}$ and the estimation error $\widetilde{\mathbf{h}}_{kl} = \mathbf{h}_{kl} - \widehat{\mathbf{h}}_{kl}$ are independent random variables distributed as [69]

$$\widehat{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}\left(\overline{\mathbf{h}}_{kl}, \mathbf{R}_{kl} - \mathbf{C}_{kl}\right), \tag{3.27}$$

$$\widetilde{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_M, \mathbf{C}_{kl}\right), \tag{3.28}$$

where $\mathbf{C}_{kl}$ is the correlation matrix of the channel estimation error, given by

$$\mathbf{C}_{kl} = \mathbf{R}_{kl} - p_k \tau_{\text{p}} \mathbf{R}_{kl} \mathbf{\Psi}_{t_k l}^{-1} \mathbf{R}_{kl}. \tag{3.29}$$

Similarly to (3.24), the collective channel estimation vector for UE k is given by $\widehat{\mathbf{h}}_k = [\widehat{\mathbf{h}}_{k1}^{\text{T}} \cdots \widehat{\mathbf{h}}_{kL}^{\text{T}}]^{\text{T}}$ and can be computed as

$$\widehat{\mathbf{h}}_k = \overline{\mathbf{h}}_k + \sqrt{p_k}\mathbf{R}_k \mathbf{\Psi}_{t_k}^{-1}(\mathbf{y}_{t_k}^{\text{pilot}} - \overline{\mathbf{y}}_{t_k}^{\text{pilot}}), \tag{3.30}$$

where

$$\mathbf{\Psi}_t = \sum_{i \in \mathcal{P}_t} p_i \tau_{\text{p}} \mathbf{R}_i + \sigma_{\text{n}}^2 \mathbf{I}_{ML} \tag{3.31}$$

and

$$\overline{\mathbf{y}}_t^{\text{pilot}} = \sum_{i \in \mathcal{P}_t} \sqrt{p_i}\tau_{\text{p}} \overline{\mathbf{h}}_i. \tag{3.32}$$

The estimate $\widehat{\mathbf{h}}_k$ and the error $\widetilde{\mathbf{h}}_k = \mathbf{h}_k - \widehat{\mathbf{h}}_k$ are independent and distributed as

$$\widehat{\mathbf{h}}_k \sim \mathcal{N}_{\mathbb{C}}\left(\overline{\mathbf{h}}_k, \mathbf{R}_k - \mathbf{C}_k\right), \tag{3.33}$$

$$\widetilde{\mathbf{h}}_k \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_{ML}, \mathbf{C}_k\right), \tag{3.34}$$

where $\mathbf{C}\mathbf{C}_k = \text{blkdiag}\left(\mathbf{C}_{k1}, \mathbf{C}_{k2}, \ldots, \mathbf{C}_{kL}\right)$ is a block-diagonal matrix.

Lemma 3.1. *The cross-correlation between the MMSE estimates of* $\mathbf{h}_{kl}$ *and* $\mathbf{h}_{il}$ *is:*

$$\mathbb{E}\{\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\} = \begin{cases} \mathbf{Q}_{kl} - \mathbf{C}_{kl}, & \text{if } i = k, \\ \overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{il}^{\mathrm{H}} + \sqrt{p_k p_i}\tau_{\mathrm{p}}\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_k l}^{-1}\mathbf{R}_{il}, & \text{if } t_i = t_k, \\ \overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{il}^{\mathrm{H}}, & \text{otherwise.} \end{cases} \tag{3.35}$$

Note that pilot sharing not only reduces the estimation quality but also makes the channel estimates more statistically dependent [9, page 107].

Lemma 3.2. *The cross-correlation between the MMSE estimates of* $\mathbf{h}_{k}$ *and* $\mathbf{h}_{i}$ *is:*

$$\mathbb{E}\{\widehat{\mathbf{h}}_{k}\widehat{\mathbf{h}}_{i}^{\mathrm{H}}\} = \begin{cases} \mathbf{Q}_{k} - \mathbf{C}_{k}, & \text{if } i = k, \\ \overline{\mathbf{h}}_{k}\overline{\mathbf{h}}_{i}^{\mathrm{H}} + \sqrt{p_k p_i}\tau_{\mathrm{p}}\mathbf{R}_{k}\boldsymbol{\Psi}_{t_k}^{-1}\mathbf{R}_{i}, & \text{if } t_i = t_k, \\ \overline{\mathbf{h}}_{k}\overline{\mathbf{h}}_{i}^{\mathrm{H}}, & \text{otherwise.} \end{cases} \tag{3.36}$$

Lemma 3.3. *Suppose the* $M \times M$ *complex matrices* $\mathbf{A}$ *and* $\mathbf{B}$*, independent of the MMSE channel estimates* $\widehat{\mathbf{h}}_{il}$ *and* $\widehat{\mathbf{h}}_{kl}$*. From the law of large numbers [57, Lemma 4],*

$$\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\mathbf{B}\mathbf{A}\widehat{\mathbf{h}}_{kl} \xrightarrow{M\to\infty} \mathbb{E}\{\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\mathbf{B}\mathbf{A}\widehat{\mathbf{h}}_{kl}\} = \mathrm{tr}\left(\mathbf{A}\mathbb{E}\{\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\}\mathbf{B}\right), \tag{3.37}$$

and applying (3.35)*:*

$$\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\mathbf{B}\mathbf{A}\widehat{\mathbf{h}}_{kl} \xrightarrow{M\to\infty} \begin{cases} \mathrm{tr}\left[\mathbf{A}(\mathbf{Q}_{il} - \mathbf{C}_{il})\mathbf{B}\right], & i = k, \\ \overline{\mathbf{h}}_{il}^{\mathrm{H}}\mathbf{B}\mathbf{A}\overline{\mathbf{h}}_{kl} + \sqrt{p_i p_k}\tau_{\mathrm{p}}\,\mathrm{tr}\left(\mathbf{A}\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_i l}^{-1}\mathbf{R}_{il}\mathbf{B}\right), & t_i = t_k, \\ \overline{\mathbf{h}}_{il}^{\mathrm{H}}\mathbf{B}\mathbf{A}\overline{\mathbf{h}}_{kl}, & t_i \neq t_k. \end{cases} \tag{3.38}$$

Lemma 3.4. *Suppose the* $ML \times ML$ *complex matrices* $\mathbf{A}$ *and* $\mathbf{B}$*, independent of the MMSE channel estimates* $\widehat{\mathbf{h}}_{i}$ *and* $\widehat{\mathbf{h}}_{k}$*. From the law of large numbers [57, Lemma 4],*

$$\widehat{\mathbf{h}}_{i}^{\mathrm{H}}\mathbf{B}\mathbf{A}\widehat{\mathbf{h}}_{k} \xrightarrow{ML\to\infty} \mathbb{E}\{\widehat{\mathbf{h}}_{i}^{\mathrm{H}}\mathbf{B}\mathbf{A}\widehat{\mathbf{h}}_{k}\} = \mathrm{tr}\left(\mathbf{A}\mathbb{E}\{\widehat{\mathbf{h}}_{k}\widehat{\mathbf{h}}_{i}^{\mathrm{H}}\}\mathbf{B}\right), \tag{3.39}$$

and applying (3.36)*:*

$$\widehat{\mathbf{h}}_{i}^{\mathrm{H}}\mathbf{B}\mathbf{A}\widehat{\mathbf{h}}_{k} \xrightarrow{ML\to\infty} \begin{cases} \mathrm{tr}\left[\mathbf{A}(\mathbf{Q}_{i} - \mathbf{C}_{i})\mathbf{B}\right], & i = k, \\ \overline{\mathbf{h}}_{i}^{\mathrm{H}}\mathbf{B}\mathbf{A}\overline{\mathbf{h}}_{k} + \sqrt{p_i p_k}\tau_{\mathrm{p}}\,\mathrm{tr}\left(\mathbf{A}\mathbf{R}_{k}\boldsymbol{\Psi}_{t_i}^{-1}\mathbf{R}_{i}\mathbf{B}\right), & t_i = t_k, \\ \overline{\mathbf{h}}_{i}^{\mathrm{H}}\mathbf{B}\mathbf{A}\overline{\mathbf{h}}_{k}, & t_i \neq t_k. \end{cases} \tag{3.40}$$

The estimation accuracy can be measured by the NMSE, evaluated at subarray l and at the entire array by [9]

$$\gamma_{kl} = \frac{\mathbb{E}\{\|\widetilde{\mathbf{h}}_{kl}\|^2\}}{\mathbb{E}\{\|\mathbf{h}_{kl}\|^2\}} = \frac{\mathrm{tr}(\mathbf{C}_{kl})}{\mathrm{tr}(\mathbf{Q}_{kl})} \tag{3.41}$$

and

$$\gamma_{kl} = \frac{\mathbb{E}\{\|\widetilde{\mathbf{h}}_{k}\|^2\}}{\mathbb{E}\{\|\mathbf{h}_{k}\|^2\}} = \frac{\mathrm{tr}(\mathbf{C}_{k})}{\mathrm{tr}(\mathbf{Q}_{k})}, \tag{3.42}$$

respectively, where $\mathbf{Q}_k = \mathbb{E}\{\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\} = \overline{\mathbf{h}}_k\overline{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{R}_k$.

3.1.2.1 Pilot Contamination

Suppose that UEs k and i use the same pilot sequence. Intuitively, the closer UE i is to the antenna array, the higher its channel gain, and consequently, the more it interferes with the estimation of $\mathbf{h}_{kl}$. This effect is reflected in (3.41) and (3.42), where the NMSE increases with $\check{\beta}_{il}$, which appears implicitly in $\mathbf{\Psi}_{t_k l}^{-1}$.

Furthermore, when UE k is close to the uniform linear array (ULA), especially in the presence of a LoS component, the channel estimate tends to be more accurate, since the LoS component varies slowly and is often assumed to be perfectly known. On the other hand, according to (3.41), the NMSE of $\overline{\mathbf{h}}_{kl}$ is inversely proportional to β_{kl}.

If $\mathbf{R}_{kl}$ is invertible, the following relation holds [9]:

$$\widehat{\mathbf{h}}_{il} = \overline{\mathbf{h}}_{il} + \sqrt{\frac{p_i}{p_k}}\mathbf{R}_{il}\mathbf{R}_{kl}^{-1}(\widehat{\mathbf{h}}_{kl} - \overline{\mathbf{h}}_{kl}). \tag{3.43}$$

This expression reveals that the more similar $\mathbf{R}_{kl}$ and $\mathbf{R}_{il}$ are, the more correlated the estimates $\widehat{\mathbf{h}}_{kl}$ and $\widehat{\mathbf{h}}_{il}$ become. Although the true channels of the pilot-sharing UEs k and i are statistically independent if at most one of them has LoS to subarray l, the NMSE estimates will be fully correlated in spatially uncorrelated NLoS environments, *i*.e., if $\mathbf{R}_{kl} = \check{\beta}_{kl}\mathbf{I}_M$ and $\mathbf{R}_{il} = \check{\beta}_{il}\mathbf{I}_M$, as the estimates will differ only by a scaling factor [9].

Pilot contamination is also problematic when the pilot-sharing UEs k and i are geographically close, as they tend to exhibit similar azimuth and elevation angles and, consequently, similar spatial correlation matrices, *i.e.*, $\mathbf{R}_{kl} \approx \mathbf{R}_{il}$. This is demonstrated both analytically and numerically in [9] to cause strong pilot contamination.

3.1.3 Uplink Data Transmission

During the uplink data transmission, all subarrays receives a superposition of the signals sent from the scheduled UEs. The signal $\mathbf{y}_l^{\mathrm{ul}} \in \mathbb{C}^M$ received by subarray l is

$$\mathbf{y}_l^{\mathrm{ul}} = \sum_{i\in\mathcal{U}} \mathbf{h}_{il}x_i + \mathbf{n}_l^{\mathrm{ul}}, \tag{3.44}$$

where $\mathbf{h}_{il} \in \mathbb{C}^M$ is the channel vector between UE i and subarray l, $x_i \in \mathbb{C}$ is the zero-mean signal transmitted from i-th UE during the uplink data transmission with transmit power $p_i = \mathbb{E}\{|x_i|^2\}$, and $\mathbf{n}_l^{\mathrm{ul}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{\mathrm{n}}^2\mathbf{I}_M)$ is the received additive white Gaussian noise (AWGN) vector. Although the channels are constant over a coherence block, both the transmitted signals and the noise take new independent realizations at every transmission symbol [9]. The collective received signal during the uplink data transmission is given by

$$\mathbf{y}^{\mathrm{ul}} = \begin{bmatrix} \mathbf{y}_1^{\mathrm{ul}} \\ \vdots \\ \mathbf{y}_L^{\mathrm{ul}} \end{bmatrix} = \sum_{i\in\mathcal{U}} \mathbf{h}_i x_i + \mathbf{n}^{\mathrm{ul}}, \tag{3.45}$$

where $\mathbf{n}^{\text{ul}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{\text{n}}^2 \mathbf{I}_{ML})$. The signals transmitted by the scheduled users are estimated using a receive combining vector, which is computed based on the channel estimates.

3.2 Centralized Uplink Operation

In a fully centralized XL-MIMO uplink operation, each subarray forwards its received pilot and data signals through fronthaul links to a CPU. The CPU then performs all channel estimation and data detection based on the aggregated signals from all ML antennas. This architecture achieves the highest spectral efficiency (SE) because it enables joint interference suppression across all subarrays. However, the practical deployment of this approach is often limited by its demanding fronthaul bandwidth and processing requirements, which motivates the use of a distributed architecture with local processing at the subarrays, which is described in Section 3.3.

3.2.1 Instantaneous Uplink SINR

In a centralized operation, the collective receive combiner $\mathbf{v}_k = [\mathbf{v}_{k1}^{\text{T}}, \ \ldots \ , \mathbf{v}_{kL}^{\text{T}}]^{\text{T}} \in \mathbb{C}^{ML}$ is computed at the CPU, in a centralized way, and the uplink data signals received by the subarrays are jointly used at the CPU to estimate x_k by computing the inner product between the combining vector and the received signal:

$$\widehat{x}_k = \sum_{l \in \mathcal{D}_k} \mathbf{v}_{kl}^{\text{H}} \mathbf{y}_l^{\text{ul}} = \mathbf{v}_k^{\text{H}} \mathbf{D}_k \mathbf{y}^{\text{ul}}, \tag{3.46}$$

Note that x_k is jointly estimated by the subarrays in the set $\mathcal{D}_k$, while the other subarrays do not take part in estimating x_k. Substituting (3.45) into (3.46) results in:

$$\begin{aligned} \widehat{x}_k &= \sum_{i \in \mathcal{U}} \mathbf{v}_k^{\text{H}} \mathbf{D}_k \mathbf{h}_i x_i + \mathbf{v}_k^{\text{H}} \mathbf{D}_k \mathbf{n}^{\text{ul}} \\ &= \underbrace{\mathbf{v}_k^{\text{H}} \mathbf{D}_k \mathbf{h}_k x_k}_{\text{desired signal}} + \underbrace{\sum_{i \in \mathcal{U} \setminus \{k\}} \mathbf{v}_k^{\text{H}} \mathbf{D}_k \mathbf{h}_i x_i}_{\text{inter-user interference}} + \underbrace{\mathbf{v}_k^{\text{H}} \mathbf{D}_k \mathbf{n}^{\text{ul}}}_{\text{noise}}. \end{aligned} \tag{3.47}$$

Figure 3.1 illustrates the signal estimation process within a centralized uplink operation. In this scenario, two scheduled UEs are transmitting distinct signals, denoted as x_1 and x_2. Suppose UE 1 is close to subarrays 1 and 2, but far from subarray 3, while UE 2 is close to subarrays 2 and 3, but far from subarray 1. Based on this proximity, the subsets of subarrays there are selected to serve UEs 1 and 2 are $\mathcal{D}_1 = \{1, 2\}$ and $\mathcal{D}_2 = \{2, 3\}$, respectively. Note that subarray 2 is part of the selection for both UEs, indicating an overlap in coverage. The CPU then estimates x_1 utilizing the received signals from the selected subarrays in $\mathcal{D}_1$, specifically $\mathbf{y}_1^{\text{ul}}$ and $\mathbf{y}_2^{\text{ul}}$, to produced the estimate $\hat{x}_1$. Similarly, the CPU uses the received signals from the selected subarrays in $\mathcal{D}_2$, namely $\mathbf{y}_2^{\text{ul}}$ and $\mathbf{y}_3^{\text{ul}}$, to generate the estimate $\hat{x}_2$.

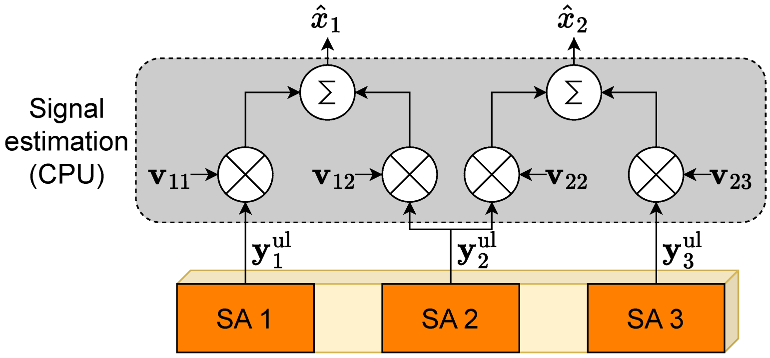


Figure 3.1 – Centralized uplink operation.

The first term in (3.47) carries the intended symbol, the second aggregates multi-user interference, and the third is the thermal noise. Decomposing the desired term into its projection over the estimated channel and over the channel estimation error yields

$$\widehat{x}_k = \underbrace{\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k x_k}_{\substack{\text{desired signal}\\ \text{(over estimated channel)}}} + \underbrace{\sum_{i\in\mathcal{U}\setminus\{k\}} \mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_i x_i}_{\substack{\text{inter-user interference}\\ \text{(over estimated channel)}}} + \underbrace{\sum_{i\in\mathcal{U}} \mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widetilde{\mathbf{h}}_i x_i}_{\substack{\text{interference due}\\ \text{to estimation error}}} + \underbrace{\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\mathbf{n}^{\mathrm{ul}}}_{\text{noise}}. \tag{3.48}$$

Only the desired signal over the estimated channel is useful for signal detection [9]. The desired signal over the channel estimation error is not useful since the estimation error is unknown, and thus it is treated as interference or noise when computing the SINR.

The received power associated to the desired signal of the k-th UE's over the known channel is computed as:

$$S_k = \mathbb{E}\{|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k x_k|^2\} = p_k|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k|^2. \tag{3.49}$$

The received power of the interference provoked by the i-th UE on the k-th UE, over the known channel, is given by:

$$\widehat{I}_{ik} = \mathbb{E}\{|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_i x_i|^2\} = p_i|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_i|^2. \tag{3.50}$$

On the other hand, the received power of the interference provoked by the i-th UE on the k-th UE, over the unknown channel, is given by:

$$\widetilde{I}_{ik} = \mathbb{E}\{|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widetilde{\mathbf{h}}_i x_i|^2\} = p_i\mathbb{E}\{|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widetilde{\mathbf{h}}_i|^2\} = p_i\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\mathbf{C}_i\mathbf{D}_k\mathbf{v}_k^{\mathrm{H}}. \tag{3.51}$$

Finally, the noise power that affects the estimation of x_k is:

$$N_k = \mathbb{E}\{|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\mathbf{n}^{\mathrm{ul}}|^2\} = \sigma_{\mathrm{n}}^2\|\mathbf{D}_k\mathbf{v}_k\|^2. \tag{3.52}$$

During the uplink data transmission with centralized operation, the SINR of UE k can be computed as

$$\Gamma_k^{\mathrm{cent}} = \frac{S_k}{\sum_{i\in\mathcal{U}\setminus\{k\}} \widehat{I}_{ik} + \sum_{i\in\mathcal{U}} \widetilde{I}_{ik} + N_k}. \tag{3.53}$$

Replacing (3.49), (3.50), (3.51) and (3.52) into (3.53) results in the uplink instantaneous SINR given by

$$\Gamma_k^{\text{cent}} = \frac{p_k|\mathbf{v}_k^{\text{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k|^2}{\sum\limits_{i\in\mathcal{U}\setminus\{k\}} p_i|\mathbf{v}_k^{\text{H}}\mathbf{D}_k\widehat{\mathbf{h}}_i|^2 + \sum\limits_{i\in\mathcal{U}} p_i\mathbf{v}_k^{\text{H}}\mathbf{D}_k\mathbf{C}_i\mathbf{D}_k\mathbf{v}_k + \sigma_{\text{n}}^2\left\|\mathbf{D}_k\mathbf{v}_k\right\|^2} \tag{3.54}$$

$$= \frac{p_k|\mathbf{v}_k^{\text{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k|^2}{\mathbf{v}_k^{\text{H}}\mathbf{D}_k\left(\sum\limits_{i\in\mathcal{U}\setminus\{k\}} p_i\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\text{H}} + \sum\limits_{i\in\mathcal{U}} p_i\mathbf{C}_i + \sigma_{\text{n}}^2\mathbf{I}_{ML}\right)\mathbf{D}_k\mathbf{v}_k}. \tag{3.55}$$

Corollary 3.1. *The instantaneous SINR in* (3.55) *is maximized by the MMSE combiner*[7]

$$\mathbf{v}_k = p_k\left(\sum_{i\in\mathcal{U}} p_i\mathbf{D}_k(\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\text{H}} + \mathbf{C}_i)\mathbf{D}_k + \sigma_{\text{n}}^2\mathbf{I}_{ML}\right)^{-1}\mathbf{D}_k\widehat{\mathbf{h}}_k, \tag{3.56}$$

which leads to the maximum SINR value

$$\Gamma_k^{\text{cent}} = p_k\widehat{\mathbf{h}}_k^{\text{H}}\mathbf{D}_k\mathbf{Z}_k^{-1}\mathbf{D}_k\widehat{\mathbf{h}}_k = p_k\,\text{tr}(\mathbf{D}_k\mathbf{Z}_k^{-1}\mathbf{D}_k\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^{\text{H}}\mathbf{D}_k), \tag{3.57}$$

where

$$\mathbf{Z}_k = \sum_{i\in\mathcal{U}\setminus\{k\}} p_i\mathbf{D}_k\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\text{H}}\mathbf{D}_k + \sum_{i\in\mathcal{U}} p_i\mathbf{D}_k\mathbf{C}_i\mathbf{D}_k + \sigma_{\text{n}}^2\mathbf{I}_{ML}. \tag{3.58}$$

The proof is provided in [9, page 135].

The combining vector that maximizes the instantaneous SINR also minimizes the conditional MSE between the true data signal x_k and the estimate $\hat{x}_k$, defined as

$$\text{MSE}_k = \mathbb{E}\{|x_k - \hat{x}_k|^2 \mid \{\widehat{\mathbf{h}}_i : i \in \mathcal{U}\}\}. \tag{3.59}$$

Corollary 3.2. *The instantaneous SE of UE k is defined as*[8]

$$\eta_k^{\text{cent}} = \left(1 - \frac{\tau_{\text{p}}}{\tau_{\text{c}}}\right)\log_2(1 + \Gamma_k^{\text{cent}}). \tag{3.60}$$

3.2.2 Ergodic Uplink SINR

Theorem 3.1 (Ergodic Uplink SINR in Centralized Operation)**.** *The ergodic global uplink SINR for centralized operation with MMSE combiner is:*

$$\mathbb{E}\{\Gamma_k^{\text{cent}}\} \approx \Gamma_k^{\text{cent,erg}} = p_k\,\text{tr}\left[\mathbf{D}_k\overline{\mathbf{Z}}_k^{-1}\mathbf{D}_k\left(\mathbf{Q}_k - \mathbf{C}_k\right)\mathbf{D}_k\right], \tag{3.61}$$

where

$$\overline{\mathbf{Z}}_k = \mathbb{E}\{\mathbf{Z}_k\} = \sum_{i\in\mathcal{U}\setminus\{k\}} p_i\mathbf{D}_k\mathbf{Q}_i\mathbf{D}_k + p_k\mathbf{D}_k\mathbf{C}_k\mathbf{D}_k + \sigma_{\text{n}}^2\mathbf{I}_{ML}. \tag{3.62}$$

This approximations holds if the selected subarrays yield strong LoS components or under high user load conditions, i.e., $U \gg ML_k$.

[7] MMSE combiner is scalable as long as the number of scheduled UEs, U, stays finite as $K \to \infty$.

[8] The SE expressions in this thesis assume that each channel coherence block is split into uplink pilot transmission and uplink data transmission, with no downlink data transmission. Alternatively, one can assume that the downlink SINR equals the uplink SINR.

Proof. The average of the SINR in (3.57) can be expressed as

$$\mathbb{E}\{\Gamma_k^{\text{cent}}\} = p_k \operatorname{tr}\left[\mathbb{E}\{\mathbf{D}_k\mathbf{Z}_k^{-1}\mathbf{D}_k\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^{\text{H}}\mathbf{D}_k\}\right]. \tag{3.63}$$

Under typical pilot-reuse patterns, where few or no UEs share the pilot of UE k, and considering limited LoS overlap across subarrays, $\mathbf{D}_k\mathbf{Z}_k\mathbf{D}_k$ and $\mathbf{D}_k\widehat{\mathbf{h}}_k\mathbf{D}_k$ become approximately uncorrelated.[9] Hence,

$$\mathbb{E}\{\mathbf{D}_k\mathbf{Z}_k^{-1}\mathbf{D}_k\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^{\text{H}}\mathbf{D}_k\} \approx \mathbf{D}_k\mathbb{E}\{\mathbf{Z}_k^{-1}\}\mathbf{D}_k\mathbb{E}\{\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_k^{\text{H}}\}\mathbf{D}_k = \mathbf{D}_k\mathbb{E}\{\mathbf{Z}_k^{-1}\}\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)\mathbf{D}_k. \tag{3.64}$$

When $U \gg ML_k$ or when subarray selection favors subarrays with a strong LoS component to UE k, the matrix $\mathbf{Z}_k$ concentrates around its mean $\overline{\mathbf{Z}}_k = \mathbb{E}\{\mathbf{Z}_k\}$. In this regime, $\mathbf{Z}_k$ exhibits small fluctuations, and F.1 in Appendix F implies

$$\mathbb{E}\{\mathbf{Z}_k^{-1}\} \approx (\mathbb{E}\{\mathbf{Z}_k\})^{-1} = \overline{\mathbf{Z}}_k^{-1}. \tag{3.65}$$

Finally, substituting (3.65) into (3.64) results in

$$\mathbb{E}\{\mathbf{D}_k\mathbf{Z}_k^{-1}\mathbf{D}_k\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^{\text{H}}\mathbf{D}_k\} \approx \mathbf{D}_k\overline{\mathbf{Z}}_k^{-1}\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)\mathbf{D}_k, \tag{3.66}$$

and inserting (3.64) into (3.63) completes the proof of (3.61). □

Corollary 3.3. *The SE estimate obtained with the ergodic SINR approximation is*

$$\eta_k^{\text{cent,erg}} = \left(1 - \frac{\tau_\text{p}}{\tau_\text{c}}\right)\log_2\left(1 + \Gamma_k^{\text{cent,erg}}\right). \tag{3.67}$$

Corollary 3.4. *Evaluating* (3.61) *entails a computational complexity equivalent to*

$$\mathcal{C}_k^{\text{cent,erg}} = 2M^3L_k^3 + \frac{1}{2}M^2L_k^2 - \frac{1}{2}ML_k \tag{3.68}$$

real multiplications. The proof is provided in Appendix J.

3.2.3 Asymptotic Uplink SINR

Theorem 3.2 (Asymptotic Uplink SINR in Centralized Operation)**.** *Assume centralized operation with MMSE channel estimation, MMSE receive combining, user loads satisfying $U \leq ML_k$, and a) no pilot reuse or b) relatively accurate channel estimates and spatially uncorrelated NLoS channel component. Then, as the number of antennas grows large, the*

[9] Although $\mathbf{Z}_k$, given by (3.58), only contains the estimated channels of UEs other than k, it may still be correlated with $\widehat{\mathbf{h}}_k$ especially if UE k shares its pilot with other UEs. According to (3.36), the estimates $\widehat{\mathbf{h}}_k$ and $\widehat{\mathbf{h}}_i$ are correlated if both UEs k and i have LoS to the same subarrays. Furthermore, the correlation between these estimates are more significant if UEs k and i share the same pilot sequence, as demonstrated in the first case in (3.36). However, this correlation will be neglected when deriving the SINR expression, since the overlap of LoS regions of different UEs is usually small and UE k will normally share its pilot with only a few UEs, or even with no other user.

instantaneous SINR converges deterministically to an asymptotic limit, i.e., $\Gamma_k^{\mathrm{cent}} \xrightarrow{ML_k \to \infty} \Gamma_k^{\mathrm{cent,asy}}$. *This asymptotic SINR is given by*

$$\Gamma_k^{\mathrm{cent,asy}} = p_k \frac{\operatorname{tr}\left[\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)\mathbf{D}_k\right] - \sum\limits_{i \in \mathcal{U} \setminus \{k\}} \frac{\operatorname{tr}\left[\mathbf{D}_k(\mathbf{Q}_i - \mathbf{C}_i)\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)\mathbf{D}_k\right]}{\operatorname{tr}\left[\mathbf{D}_k(\mathbf{Q}_i - \mathbf{C}_i)\mathbf{D}_k\right]}}{\sum\limits_{i \in \mathcal{U}} \frac{p_i}{ML_k} \operatorname{tr}(\mathbf{D}_k \mathbf{C}_i \mathbf{D}_k) + \sigma_{\mathrm{n}}^2}. \tag{3.69}$$

Proof. When $ML_k \geq U$, the ZF combiner satisfies

$$\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_i \approx \begin{cases} 1, & \text{if } i = k, \\ 0, & \text{if } i \neq k. \end{cases} \tag{3.70}$$

Under high SNR and small channel estimation errors, the MMSE combiner converges to the ZF. Thus, the SINR in (3.55) can be approximated as

$$\Gamma_k^{\mathrm{cent}} \approx \frac{p_k}{\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \left(\sum\limits_{i \in \mathcal{U}} p_i \mathbf{C}_i + \sigma_{\mathrm{n}}^2 \mathbf{I}_{ML} \right) \mathbf{D}_k \mathbf{v}_k} \overset{(a)}{=} \frac{p_k}{\sum\limits_{l \in \mathcal{D}_k} \mathbf{v}_{kl}^{\mathrm{H}} \left(\sum\limits_{i \in \mathcal{U}} p_i \mathbf{C}_{il} + \sigma_{\mathrm{n}}^2 \mathbf{I}_M \right) \mathbf{v}_{kl}}, \tag{3.71}$$

where (a) holds since $\mathbf{C}_i$ is block-diagonal for any i. When the NLoS channel component is spatially uncorrelated, we can replace $\mathbf{C}_{il}$ by $\frac{1}{ML_k} \sum\limits_{l \in \mathcal{D}_k} \operatorname{tr}(\mathbf{C}_{il}) \mathbf{I}_M$, for any $l \in \mathcal{D}_k$, yielding[10]

$$\Gamma_k^{\mathrm{cent}} \approx \frac{p_k}{\left(\sum\limits_{i \in \mathcal{U}} \sum\limits_{l \in \mathcal{D}_k} \frac{p_i}{ML_k} \operatorname{tr}(\mathbf{C}_{il}) + \sigma_{\mathrm{n}}^2 \right) \sum\limits_{l \in \mathcal{D}_k} \|\mathbf{v}_{kl}\|^2} \overset{(b)}{=} \frac{p_k}{\left(\sum\limits_{i \in \mathcal{U}} \frac{p_i}{ML_k} \operatorname{tr}(\mathbf{D}_k \mathbf{C}_i \mathbf{D}_k) + \sigma_{\mathrm{n}}^2 \right) \|\mathbf{v}_k\|^2}, \tag{3.72}$$

where (b) comes from $\sum_{l \in \mathcal{D}_k} \|\mathbf{v}_{kl}\|^2 = \|\mathbf{D}_k \mathbf{v}_k\|^2$. The squared norm $\|\mathbf{v}_k\|^2$ is approximated using the ZF combiner, which behaves similarly to the MMSE combiner when the channel estimates at the selected subarrays are accurate and the SNR is high. For notational convenience, assume for a moment that all UEs are scheduled, *i.e.*, let $\mathcal{U} = \mathcal{K}$. Then, $\mathbf{v}_k$ is approximated as the k-th column of $\mathbf{V} = \mathbf{D}_k \widehat{\mathbf{H}} (\widehat{\mathbf{H}}^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{H}})^{-1}$, where $\widehat{\mathbf{H}} = [\widehat{\mathbf{h}}_1, \ldots, \widehat{\mathbf{h}}_K]$, such that $\|\mathbf{v}_k\|^2 = [\mathbf{V}^{\mathrm{H}} \mathbf{V}]_{k,k} = [(\mathbf{H}^{\mathrm{H}} \mathbf{D}_k \mathbf{H})^{-1}]_{k,k}$. The block matrix inversion lemma yields [50, Appendix A]

$$\|\mathbf{v}_k\|^2 = [\widehat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k - \widehat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{D}_k \ddot{\mathbf{H}}_k (\ddot{\mathbf{H}}_k^{\mathrm{H}} \mathbf{D}_k \ddot{\mathbf{H}}_k)^{-1} \ddot{\mathbf{H}}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k]^{-1}, \tag{3.73}$$

where $\ddot{\mathbf{H}}_k \in \mathbb{C}^{ML \times (K-1)}$ denotes the estimated channel matrix excluding UE k, i.e., $\ddot{\mathbf{H}}_k = [\widehat{\mathbf{h}}_1, \ldots, \widehat{\mathbf{h}}_{k-1}, \widehat{\mathbf{h}}_{k+1}, \ldots, \widehat{\mathbf{h}}_K]$. Substituting (3.73) into (3.72) results in

$$\Gamma_k^{\mathrm{cent}} \approx p_k \frac{\widehat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k - \widehat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{D}_k \ddot{\mathbf{H}}_k (\ddot{\mathbf{H}}_k^{\mathrm{H}} \mathbf{D}_k \ddot{\mathbf{H}}_k)^{-1} \ddot{\mathbf{H}}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k}{\sum\limits_{i \in \mathcal{U}} \frac{p_i}{ML_k} \operatorname{tr}(\mathbf{D}_k \mathbf{C}_i \mathbf{D}_k) + \sigma_{\mathrm{n}}^2}. \tag{3.74}$$

[10] The approximation in (3.72) remains accurate for correlated NLoS components, as long as there is no pilot reuse. This is because the eigenvalues of the matrix within parentheses in (3.55) are closely clustered in the absence of pilot reuse. Consequently, replacing $\mathbf{C}_{il}$ by $\frac{1}{ML_k} \sum_{l \in \mathcal{D}_k} \operatorname{tr}(\mathbf{C}_{il}) \mathbf{I}_M$ introduces negligible variations in the SINR denominator.

The diagonal elements of $\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k$ consists of the squared norms $\|\mathbf{D}_k\widehat{\mathbf{h}}_i\|^2$ for $i \in \mathcal{K} \setminus \{k\}$, which converge to $\mathrm{tr}[\mathbf{D}_k(\mathbf{Q}_i - \mathbf{C}_i)\mathbf{D}_k]$ as $ML_k \to \infty$, according to the first case of (3.40). In XL-MIMO, the off-diagonal terms are negligible since the channels are spatially non-stationary.[11] We can thus approximate (3.74) by

$$\Gamma_k^{\mathrm{cent}} \approx p_k \frac{\widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k - \widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k\mathbf{T}_k^{-1}\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k}{\sum\limits_{i\in\mathcal{U}} \frac{p_i}{ML_k}\,\mathrm{tr}(\mathbf{D}_k\mathbf{C}_i\mathbf{D}_k) + \sigma_{\mathrm{n}}^2}, \tag{3.75}$$

where $\mathbf{T}_k$ is obtained by retaining only the diagonal entries of $\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k$:

$$\begin{aligned}(\mathbf{T}_k)_{i,i'} &= \begin{cases} (\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k)_{i,i} & \text{if } i = i', \\ 0 & \text{otherwise,} \end{cases} \\ &= \mathrm{diag}\left(\|\mathbf{D}_k\widehat{\mathbf{h}}_1\|^2, \ldots, \|\mathbf{D}_k\widehat{\mathbf{h}}_{k-1}\|^2, \|\mathbf{D}_k\widehat{\mathbf{h}}_{k+1}\|^2, \ldots, \|\mathbf{D}_k\widehat{\mathbf{h}}_K\|^2\right). \end{aligned} \tag{3.76}$$

Applying the first case of (3.40) results in

$$\widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k \xrightarrow{ML_k\to\infty} \mathrm{tr}\left[\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)\mathbf{D}_k\right] \tag{3.77}$$

and

$$\begin{aligned}\widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k\mathbf{T}_k^{-1}\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k &\overset{(c)}{=} \sum_{i\in\mathcal{U}\setminus\{k\}}^{K} \frac{\widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k}{\widehat{\mathbf{h}}_i^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_i} \\ &\xrightarrow{ML_k\to\infty} \frac{\mathrm{tr}\left[\mathbf{D}_k(\mathbf{Q}_i - \mathbf{C}_i)\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)\mathbf{D}_k\right]}{\mathrm{tr}\left[\mathbf{D}_k(\mathbf{Q}_i - \mathbf{C}_i)\mathbf{D}_k\right]}, \end{aligned} \tag{3.78}$$

where (c) comes from simple algebraic manipulation. The proof of Theorem 3.2 is completed by substituting (3.77) and (3.78) into (3.75). □

Corollary 3.5. *The SE estimate obtained with the asymptotic SINR approximation is*

$$\eta_k^{\mathrm{cent,asy}} = \left(1 - \frac{\tau_{\mathrm{p}}}{\tau_{\mathrm{c}}}\right)\log_2(1 + \Gamma_k^{\mathrm{cent,asy}}). \tag{3.79}$$

Corollary 3.6. *Evaluating* (3.69) *entails a computational complexity equivalent to*

$$\mathcal{C}_k^{\mathrm{cent,asy}} = 3(U-1)M^2L_k^2 \tag{3.80}$$

real multiplications.

[11] The off-diagonal elements of $\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k$ are $\widehat{\mathbf{h}}_i^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_{i'}$, for any $i, i' \in \mathcal{K} \setminus \{k\}$ with $i \neq i'$. According to the second or third cases in (3.40), these terms converge to $\overline{\mathbf{h}}_i^{\mathrm{H}}\mathbf{D}_k\overline{\mathbf{h}}_{i'} + \sqrt{p_ip_{i'}}\tau_{\mathrm{p}}\,\mathrm{tr}\left(\mathbf{D}_k\mathbf{R}_i\boldsymbol{\Psi}_{t_i}^{-1}\mathbf{R}_{i'}\mathbf{D}_k\right)$ as $ML_k \to \infty$, if $t_i = t_{i'}$, and to $\overline{\mathbf{h}}_i^{\mathrm{H}}\mathbf{D}_k\overline{\mathbf{h}}_{i'}$ otherwise. In practice, different UEs are typically assigned distinct pilot sequences. Furthermore, the large dimension of the XL-MIMO antenna array ensures that the LoS regions of different UEs overlap on only a few subarrays or may not overlap at all. As a result, the off-diagonal terms of $\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k$ are generally much smaller than the diagonal terms, especially for a large number of antennas. On the other hand, these terms are normally non-negligible in conventional massive MIMO, where the antennas are co-located such that the LoS region of a given UE encompasses all the BS antennas or none of them.

Proof. The computational cost primarily arises from evaluating the matrix-product traces $\mathrm{tr}[\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)\mathbf{D}_k(\mathbf{Q}_i - \mathbf{C}_i)\mathbf{D}_k]$. Since only the diagonal entries of the resulting matrix (restricted to the rows and columns associated with the L_k subarrays serving UEk) are requires to compute the trace, the cost is $(ML_k)^2$ complex multiplications per UE $i \in \mathcal{U} \setminus \{k\}$. Accounting for all $U - 1$ interfering UEs, and assuming that each complex multiplication is implemented using three real multiplications [8, page 378], the total cost is given by (3.80). □

3.3 Distributed Uplink Operation

Rather than centralizing both channel estimation and data detection at the CPU, an XL-MIMO network can operate in a distributed manner, where signal processing is performed locally at the subarrays. Since each subarray can be equipped with a baseband processor that is at least as powerful as those used in the UEs, additional subarrays can be integrated into the network without requiring any upgrade to the CPUs.

In the distributed scheme, uplink data estimation is executed in two stages: local estimation and final estimation. During the local estimation, each subarray $l \in \mathcal{D}_k$ produces a local estimate of the transmitted signal x_k by applying a receive combining vector $\mathbf{v}_{kl}$ to its received signal $\mathbf{y}_l^{\mathrm{ul}}$, as expressed by[12]

$$\widehat{x}_{kl} = \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{y}_l^{\mathrm{ul}}. \tag{3.81}$$

Then, the CPU combines these estimates using a weighted sum to obtain the final estimate of x_k:

$$\hat{x}_k = \sum_{l \in \mathcal{D}_k} \mu_{kl}\, \hat{x}_{kl}, \tag{3.82}$$

where the nonnegative weights μ_{kl} satisfy

$$\mu_{kl} \begin{cases} > 0, & \text{if } l \in \mathcal{D}_k, \\ = 0, & \text{otherwise.} \end{cases} \tag{3.83}$$

The local estimate $\hat{x}_{kl}$ produced by subarray l is therefore assigned zero weight if that subarray is not selected to serve UE k, *i.e.*, if $l \notin \mathcal{D}_k$.[13]

Figure 3.2 illustrates the estimation process in a distributed uplink operation. As in the centralized scenario shown in Figure 3.1, two UEs are scheduled, with subarray selection sets $\mathcal{D}_1 = \{1, 2\}$ and $\mathcal{D}_2 = \{2, 3\}$. The crucial difference lies in where the estimation takes place. In centralized processing, signal estimation is performed in a single step at the CPU, whereas in the distributed approach, local estimation occurs at the

[12] The expression for $\mathbf{y}_l^{\mathrm{ul}}$ is given in (3.44).

[13] In practice, when $l \notin \mathcal{D}_k$, the estimate $\hat{x}_{kl}$ is not computed.

subarrays and the final estimation at the CPU. Initially, subarrays 1 and 2 independently estimate the transmitted signal x_1, yielding the local estimates $\hat{x}_{11}$ and $\hat{x}_{12}$. Similarly, subarrays 2 and 3 estimate x_2, producing $\hat{x}_{22}$ and $\hat{x}_{23}$. These local estimates are then forwarded to the CPU, which computes the final estimates $\hat{x}_1$ and $\hat{x}_2$.

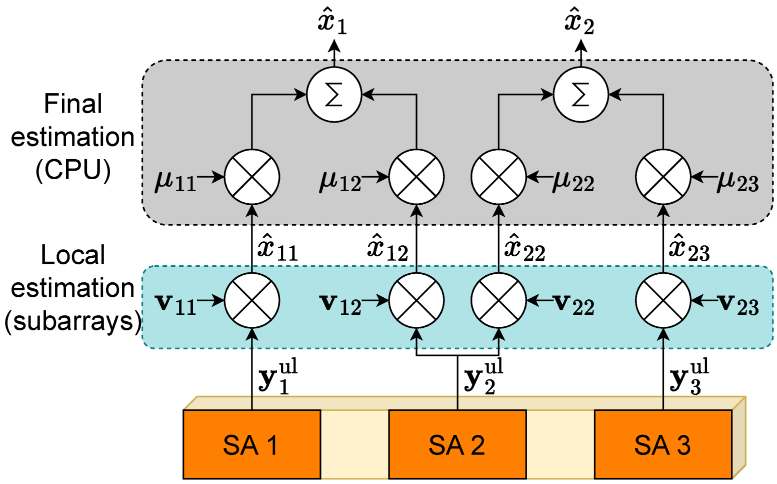


Figure 3.2 – Distributed uplink operation.

Figure 3.3 illustrates a simplified XL-MIMO uplink scenario involving a single UE k and four two-antenna subarrays mounted along the rooftop of a building. The transmitted symbol x_k is predominantly received by the two nearest subarrays (1 and 2), which experience notably higher received power. As a result, only these subarrays are selected to serve the considered UE. Each selected subarray performs L-MMSE receive combining to locally estimate x_k. Finally, the CPU obtains the final signal estimate.

3.3.1 Instantaneous Local Uplink SINR

Similarly to (3.54) and (3.55), UE k's local uplink SINR evaluated at subarray l is

$$\Gamma_{kl}^{\mathrm{dist}} = \frac{p_k |\mathbf{v}_{kl}^{\mathrm{H}} \widehat{\mathbf{h}}_{kl}|^2}{\sum\limits_{\substack{i \in \mathcal{U} \\ i \neq k}} p_i |\mathbf{v}_{kl}^{\mathrm{H}} \widehat{\mathbf{h}}_{il}|^2 + \sum\limits_{i \in \mathcal{U}} p_i \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{C}_{il} \mathbf{v}_{kl} + \sigma_{\mathrm{n}}^2 \|\mathbf{v}_{kl}\|^2}. \tag{3.84}$$

Corollary 3.7. *The conditional MSE associated to the estimation $\hat{x}_{kl}$ is defined as*

$$\mathrm{MSE}_{kl} = \mathbb{E}\{|x_k - \hat{x}_{kl}|^2 \mid \{\widehat{\mathbf{h}}_{il} : i \in \mathcal{U}\}\}. \tag{3.85}$$

The combining vector $\mathbf{v}_{kl}$ that minimizes (3.85) *and maximizes the local SINR is called L-MMSE combiner and is computed as [9, page 149]*

$$\mathbf{v}_{kl} = p_k \mathbf{W}_l^{-1} \widehat{\mathbf{h}}_{kl}, \tag{3.86}$$

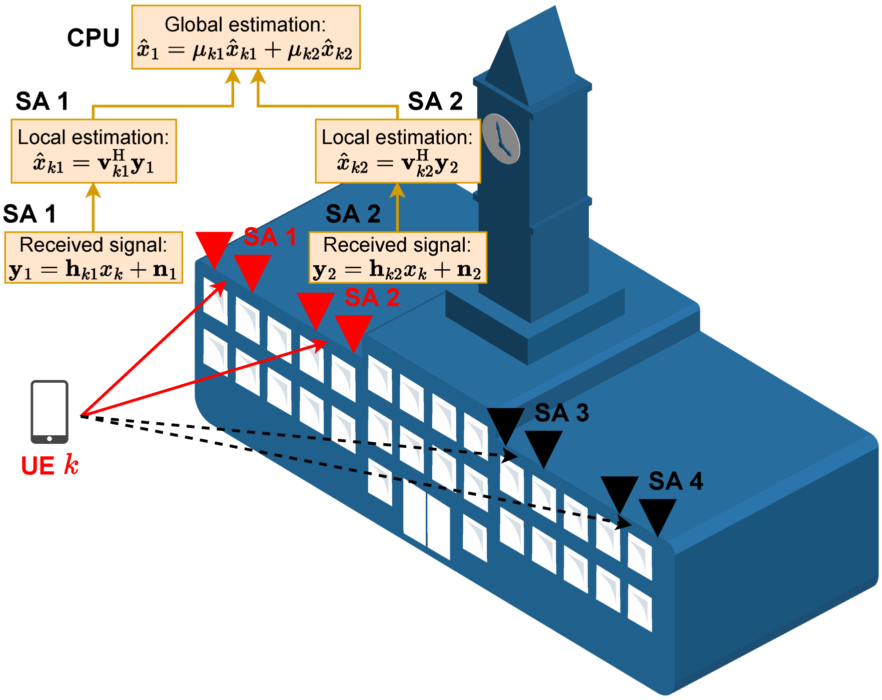


Figure 3.3 – Simplified XL-MIMO system illustrating subarray selection for signal detection. The red subarrays locally estimate x_k. The CPU aggregates the local estimates into a final estimate.

where[14]

$$\mathbf{W}_l = \sum_{i \in \mathcal{U}} p_i(\widehat{\mathbf{h}}_{il}\widehat{\mathbf{h}}_{il}^{\mathrm{H}} + \mathbf{C}_{il}) + \sigma_{\mathrm{n}}^2 \mathbf{I}_M. \tag{3.87}$$

The L-MMSE combiner leads to the maximum local SINR

$$\Gamma_{kl}^{\mathrm{dist}} = p_k \widehat{\mathbf{h}}_{kl}^{\mathrm{H}} \mathbf{Z}_{kl}^{-1} \widehat{\mathbf{h}}_{kl} = p_k \operatorname{tr}(\mathbf{Z}_{kl}^{-1}\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{kl}^{\mathrm{H}}), \tag{3.88}$$

where

$$\mathbf{Z}_{kl} = \mathbf{W}_l - p_k \widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{kl}^{\mathrm{H}} = \sum_{i \in \mathcal{U} \setminus \{k\}} p_i \widehat{\mathbf{h}}_{il}\widehat{\mathbf{h}}_{il}^{\mathrm{H}} + \sum_{i \in \mathcal{U}} p_i \mathbf{C}_{il} + \sigma_{\mathrm{n}}^2 \mathbf{I}_M. \tag{3.89}$$

Corollary 3.8. *Assuming $M \geq U$, high SNR, and relatively small channel estimation errors, the L-MMSE combiner satisfies the approximation*

$$\mathbf{v}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{il} \approx \begin{cases} 1, & \text{if } i = k, \\ 0, & \text{otherwise.} \end{cases} \tag{3.90}$$

Proof. When $M \geq U$, the local ZF combiner satisfies $\mathbf{v}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{il} = \delta_{ik}$, where δ_{il} is the Kronecker delta. Since the L-MMSE combiner converges to the local ZF combiner as the SNR increases and channel estimation errors vanish, the approximation in (3.90) holds. $\square$

[14] Although subarray l only estimates the data signals of the U_l UEs it serves, computing the L-MMSE combiner for each of those UEs requires this subarray to estimate the channels of all U scheduled UEs. Thus, the L-MMSE combiner is scalable only if U stays finite as $K \to \infty$.

Corollary 3.9. *Assuming $M \geq U$, L-MMSE combiner, high SNR, and relatively low channel estimation errors, the local SINR defined in* (3.84) *admits the approximation*

$$\Gamma_{kl}^{\text{dist}} \approx \frac{p_k}{\mathbf{v}_{kl}^{\mathrm{H}} \left(\sum_{i \in \mathcal{U}} p_i \mathbf{C}_{il} + \sigma_{\mathrm{n}}^2 \mathbf{I}_M \right) \mathbf{v}_{kl}}. \tag{3.91}$$

This follows directly from Corollary 3.8 and holds true even under pilot reuse, provided that a robust subarray selection strategy is in place. Specifically, by restricting computation to the set $\mathcal{D}_k$ comprising the subarrays with the strongest channel gains, the system operates in a regime of good channel estimation quality. Conversely, selecting all subarrays, including those dominated by noise or severe interference, would degrade estimation accuracy, rendering the approximation in (3.91) *inaccurate.*

3.3.2 Instantaneous Global Uplink SINR

Substituting (3.44) into (3.81), and then substituting (3.81) into (3.82), yields

$$\hat{x}_k = \sum_{l \in \mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\mathrm{H}} \widehat{\mathbf{h}}_{kl} x_k + \sum_{\substack{i \in \mathcal{U} \\ i \neq k}} \sum_{l \in \mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\mathrm{H}} \widehat{\mathbf{h}}_{il} x_i + \sum_{i \in \mathcal{U}} \sum_{l \in \mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\mathrm{H}} \widetilde{\mathbf{h}}_{il} x_i + \sum_{l \in \mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{n}_l. \tag{3.92}$$

After combining, the average received power of the k-th UE's desired signal, over the known channel, is

$$S_k = \mathbb{E} \left\{ \left| \sum_{l \in \mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\mathrm{H}} \widehat{\mathbf{h}}_{kl} x_k \right|^2 \right\} = p_k \left| \sum_{l \in \mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\mathrm{H}} \widehat{\mathbf{h}}_{kl} \right|^2. \tag{3.93}$$

Applying Corollary 3.8 in (3.93), we observe that maintaining the received desired signal power equal to the transmitted power (*i.e.*, preserving $S_k = p_k$) imposes a unit-sum constraint on the weights:

$$\sum_{l=1}^{L} \mu_{kl} = 1. \tag{3.94}$$

The average power of the interference from UE i onto UE k splits into two components. The first one is over the known channel:

$$\widehat{I}_{ik} = \mathbb{E} \left\{ \left| \sum_{l \in \mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\mathrm{H}} \widehat{\mathbf{h}}_{il} x_i \right|^2 \right\} = p_i \left| \sum_{l \in \mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\mathrm{H}} \widehat{\mathbf{h}}_{il} \right|^2, \tag{3.95}$$

and the second one is over the unknown channel:

$$\widetilde{I}_{ik} = \mathbb{E} \left\{ \left| \sum_{l \in \mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\mathrm{H}} \widetilde{\mathbf{h}}_{il} x_i \right|^2 \right\} = p_i \, \mathbb{E} \left\{ \left| \sum_{l \in \mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\mathrm{H}} \widetilde{\mathbf{h}}_{il} \right|^2 \right\} = p_i \sum_{l \in \mathcal{D}_k} \mu_{kl}^2 \mathbf{v}_{kl}^{\mathrm{H}} \mathbf{C}_{il} \mathbf{v}_{kl}. \tag{3.96}$$

Finally, the noise power after combining is

$$N_k = \mathbb{E}\left\{\left|\sum_{l=1}^{L} \mu_{kl}\mathbf{v}_{kl}^{\mathrm{H}}\mathbf{n}_l\right|^2\right\} = \sigma_{\mathrm{n}}^2 \sum_{l\in\mathcal{D}_k} \mu_{kl}^2\|\mathbf{v}_{kl}\|^2. \tag{3.97}$$

The global SINR for UE k is obtained by substituting (3.93), (3.95), (3.96) and (3.97) into (3.53):

$$\Gamma_k^{\mathrm{dist}} = \frac{p_k\left|\sum\limits_{l\in\mathcal{D}_k}\mu_{kl}\mathbf{v}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{kl}\right|^2}{\sum\limits_{i\in\mathcal{U}\setminus\{k\}} p_i\left|\sum\limits_{l\in\mathcal{D}_k}\mu_{kl}\mathbf{v}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{il}\right|^2 + \sum\limits_{i\in\mathcal{U}} p_i \sum\limits_{l\in\mathcal{D}_k}\mu_{kl}^2\mathbf{v}_{kl}^{\mathrm{H}}\mathbf{C}_{il}\mathbf{v}_{kl} + \sigma_{\mathrm{n}}^2\sum\limits_{l\in\mathcal{D}_k}\mu_{kl}^2\|\mathbf{v}_{kl}\|^2}. \tag{3.98}$$

Corollary 3.10. *The instantaneous SE of UE k is given by*

$$\eta_k^{\mathrm{dist}} = \left(1 - \frac{\tau_{\mathrm{p}}}{\tau_{\mathrm{c}}}\right)\log_2\left(1 + \Gamma_k^{\mathrm{dist}}\right). \tag{3.99}$$

Corollary 3.11. *Assuming $M \geq U$, L-MMSE combiner, high SNR, and relatively small channel estimation errors, the global SINR in* (3.98) *can be approximated as*

$$\Gamma_k^{\mathrm{dist}} \approx \frac{p_k}{\sum\limits_{l\in\mathcal{D}_k}\mu_{kl}^2\mathbf{v}_{kl}^{\mathrm{H}}\left(\sum\limits_{i\in\mathcal{U}} p_i\mathbf{C}_{il} + \sigma_{\mathrm{n}}^2\mathbf{I}_M\right)\mathbf{v}_{kl}}. \tag{3.100}$$

This follows directly from Corollary 3.8 and the unit-sum constraint on the weights.

Corollary 3.12 (Global SINR Approximation)**.** *Assuming $M \geq U$, L-MMSE combiner, high SNR, and relatively small channel estimation errors, the global SINR can be approximated from the local SINRs and the weights as*

$$\widehat{\Gamma}_k^{\mathrm{dist}} = \left(\sum_{l\in\mathcal{D}_k}\mu_{kl}^2(\Gamma_{kl}^{\mathrm{dist}})^{-1}\right)^{-1}. \tag{3.101}$$

Proof. Corollary 3.12 follows directly by rewriting (3.100) and comparing it with the approximation of $(\Gamma_{kl}^{\mathrm{dist}})^{-1}$ in (3.91):

$$(\Gamma_k^{\mathrm{dist}})^{-1} \approx \sum_{l\in\mathcal{D}_k}\mu_{kl}^2\underbrace{\frac{1}{p_k}\mathbf{v}_{kl}^{\mathrm{H}}\left(\sum_{i\in\mathcal{U}} p_i\mathbf{C}_{il} + \sigma_{\mathrm{n}}^2\mathbf{I}_M\right)\mathbf{v}_{kl}}_{(\Gamma_{kl}^{\mathrm{dist}})^{-1}}.$$

□

Corollary 3.13. *The instantaneous SE can also be computed approximately, based on the instantaneous local SINRs, as*

$$\hat{\eta}_k^{\mathrm{dist}} = \left(1 - \frac{\tau_{\mathrm{p}}}{\tau_{\mathrm{c}}}\right)\log_2(1 + \widehat{\Gamma}_k^{\mathrm{dist}}). \tag{3.102}$$

3.3.3 Weighting

Corollary 3.14 (SINR-Based (Optimal) Weights)**.** *Assume $M \geq U$ and L-MMSE combiner. The weights μ_{kl} that maximize the global SINR, given the selected subarrays, are*

$$\mu_{kl}^{*} = \begin{cases} \dfrac{\Gamma_{kl}^{\text{dist}}}{\sum\limits_{l' \in \mathcal{D}_k} \Gamma_{kl'}^{\text{dist}}}, & \text{if } l \in \mathcal{D}_k, \\ 0, & \text{otherwise.} \end{cases} \tag{3.103}$$

and the resulting maximum global SINR is

$$\widehat{\Gamma}_{k}^{\text{dist}} = \sum_{l \in \mathcal{D}_k} \Gamma_{kl}^{\text{dist}}. \tag{3.104}$$

Note that each optimal weight is proportional to the local SINR at the corresponding subarray, and the maximum global SINR equals the sum of the local SINRs. Thus, we also refer to the optimal weighting strategy as SINR-based weighting.

Proof. The optimal weights for distributed uplink operation are obtained by maximizing the approximate global SINR in (3.101), given the local SINRs. Since $\mu_{kl} \geq 0$ and $\Gamma_{kl}^{\text{dist}} \geq 0$ for all $l \in \mathcal{D}_k$, we have $\sum_{l \in \mathcal{D}_k} \mu_{kl}^2 (\Gamma_{kl}^{\text{dist}})^{-1} > 0$. Thus, maximizing $\widehat{\Gamma}_{k}^{\text{dist}}$ is equivalent to solving the following minimization problem:

$$\min_{\{\mu_{kl}\}_{l \in \mathcal{D}_k}} \quad \sum_{l \in \mathcal{D}_k} \mu_{kl}^2 (\Gamma_{kl}^{\text{dist}})^{-1} \tag{3.105a}$$

$$\text{s.t.} \quad \sum_{l \in \mathcal{D}_k} \mu_{kl} = 1, \tag{3.105b}$$

$$\mu_{kl} \geq 0, \quad l \in \mathcal{D}_k, \tag{3.105c}$$

where constraints (3.105b) and (3.105c) enforce nonnegative, unit-sum weights. Neglecting (3.105c), which will be satisfied a posteriori, the Lagrangian function is defined as

$$\mathcal{L}_k = \sum_{l \in \mathcal{D}_k} \mu_{kl}^2 (\Gamma_{kl}^{\text{dist}})^{-1} + \lambda_k \left(\sum_{l \in \mathcal{D}_k} \mu_{kl} - 1 \right), \tag{3.106}$$

where $\lambda_k \neq 0$ is the Lagrangian multiplier associated to the constraint.

To find the optimal weights, we set each partial derivative of $\mathcal{L}_k$ (with respect to the corresponding weight) to zero. This procedure is an application of the first-order Karush–Kuhn–Tucker (KKT) optimality condition [70], which enforces that the gradients of the objective function (3.105a) and the constraint (3.105b) are parallel at the solution:

$$\frac{\partial \mathcal{L}_k}{\partial \mu_{kl}} = 2(\Gamma_{kl}^{\text{dist}})^{-1} \mu_{kl} + \lambda_k = 0, \quad l \in \mathcal{D}_k, \tag{3.107}$$

which results in

$$\mu_{kl} = -\frac{1}{2} \Gamma_{kl}^{\text{dist}} \lambda_k, \quad l \in \mathcal{D}_k. \tag{3.108}$$

Since $\Gamma_{kl}^{\text{dist}}$ is positive, λ_k must be negative to attain constraint (3.105c), *i.e.*, to guarantee that the weights are positive. Constraint (3.105b) is itself another KKT condition. Replacing μ_{kl} in it gives the optimal Lagrangian multiplier:

$$\lambda_k = -\frac{2}{\sum\limits_{l \in \mathcal{D}_k} \Gamma_{kl}^{\text{dist}}}. \tag{3.109}$$

The optimal weights for the distributed uplink operation are found by replacing λ_k in μ_{kl}, which completes the proof of (3.103). Finally, by replacing (3.103) into (3.101), we obtain the maximum global SINR based on the local SINRs, resulting in (3.104). □

A practical alternative to SINR-based weighting is LSF-based weighting, which avoids local SINR estimation by using weights proportional to the LSF chbannel gains:

$$\mu_{kl} = \begin{cases} \dfrac{\beta_{kl}}{\sum\limits_{l' \in \mathcal{D}_k} \beta_{kl'}}, & \text{if } l \in \mathcal{D}_k, \\ 0, & \text{otherwise.} \end{cases} \tag{3.110}$$

While (3.110) is suboptimal as it neglects interference and channel estimation errors, it generally outperforms the equal weighting strategy, defined as:

$$\mu_{kl} = \begin{cases} 1/L_k, & \text{if } l \in \mathcal{D}_k, \\ 0, & \text{otherwise.} \end{cases} \tag{3.111}$$

In all proposed weighting schemes, the weights are normalized to sum to unity.

3.3.4 Ergodic Local Uplink SINR

This section derives an ergodic expression for the local SINR, denoted by $\Gamma_{kl}^{\text{dist,erg}}$, which serves as a deterministic counterpart to the exact ergodic value:

$$\Gamma_{kl}^{\text{dist,erg}} \approx \mathbb{E}\{\Gamma_{kl}^{\text{dist}}\}. \tag{3.112}$$

In other words, $\Gamma_{kl}^{\text{dist,erg}}$ provides a closed-form expression that approximates the expected instantaneous local SINR.

Theorem 3.3 (Ergodic Uplink SINR in Distributed Operation)**.** *The ergodic local uplink SINR for distributed operation with L-MMSE combiner is:*

$$\mathbb{E}\{\Gamma_{kl}^{\text{dist}}\} \approx \Gamma_{kl}^{\text{dist,erg}} = p_k \operatorname{tr}\left[\overline{\mathbf{Z}}_{kl}^{-1} \left(\mathbf{Q}_{kl} - \mathbf{C}_{kl}\right)\right], \tag{3.113}$$

where

$$\overline{\mathbf{Z}}_{kl} = \mathbb{E}\{\mathbf{Z}_{kl}\} = \sum_{i \in \mathcal{U} \setminus \{k\}} p_i \mathbf{Q}_{il} + p_k \mathbf{C}_{kl} + \sigma_n^2 \mathbf{I}_M. \tag{3.114}$$

This approximations holds if the selected subarrays yield strong LoS components or under high user load conditions, i.e., $U \gg M$.

Proof. The average of the local SINR in (3.88) can be expressed as

$$\mathbb{E}\{\Gamma_{kl}^{\text{dist}}\} = p_k \operatorname{tr}(\mathbb{E}\{\mathbf{Z}_{kl}^{-1}\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{kl}^{\text{H}}\}). \tag{3.115}$$

Under typical pilot-reuse patterns, where few or no UEs share the pilot of UE k, and considering limited LoS overlap across subarrays, $\mathbf{Z}_{kl}$ and $\widehat{\mathbf{h}}_{kl}$ become approximately uncorrelated.[15] Hence,

$$\mathbb{E}\{\mathbf{Z}_{kl}^{-1}\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{kl}^{\text{H}}\} \approx \mathbb{E}\{\mathbf{Z}_{kl}^{-1}\}\mathbb{E}\{\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{kl}^{\text{H}}\} = \mathbb{E}\{\mathbf{Z}_{kl}^{-1}\}(\mathbf{Q}_{kl} - \mathbf{C}_{kl}). \tag{3.116}$$

When $U \gg M$ or when subarray selection favors subarrays with a strong LoS component to UE k, the matrix $\mathbf{Z}_{kl}$ concentrates around its mean $\overline{\mathbf{Z}}_{kl} = \mathbb{E}\{\mathbf{Z}_{kl}\}$. In this regime, $\mathbf{Z}_{kl}$ exhibits small fluctuations, and F.1 in Appendix F implies

$$\mathbb{E}\{\mathbf{Z}_{kl}^{-1}\} \approx (\mathbb{E}\{\mathbf{Z}_{kl}\})^{-1} = \overline{\mathbf{Z}}_{kl}^{-1}. \tag{3.117}$$

Finally, substituting (3.117) into (3.116) results in

$$\mathbb{E}\{\mathbf{Z}_{kl}^{-1}\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{kl}^{\text{H}}\} \approx \overline{\mathbf{Z}}_{kl}^{-1}(\mathbf{Q}_{kl} - \mathbf{C}_{kl}), \tag{3.118}$$

and inserting (3.118) into (3.115) completes the proof of (3.113). □

Corollary 3.15. *Using the ergodic approximation for the local SINR, the corresponding SE estimate is given by*

$$\hat{\eta}_k^{\text{dist,erg}} = \left(1 - \frac{\tau_{\text{p}}}{\tau_{\text{c}}}\right) \log_2 \left[1 + \left(\sum_{l\in\mathcal{D}_k} \mu_{kl}^2 \left(\Gamma_{kl}^{\text{dist,erg}}\right)^{-1}\right)^{-1}\right]. \tag{3.119}$$

Corollary 3.16. *Evaluating* (3.113) *across the L_k serving subarrays entails a computational complexity equivalent to*

$$\mathcal{C}_k^{\text{dist,erg}} = 2M^3 L_k + \frac{1}{2}M^2 L_k - \frac{1}{2}M L_k \tag{3.120}$$

real multiplications. The proof is provided in Appendix K.

3.3.5 Asymptotic Local Uplink SINR

Theorem 3.4 (Asymptotic Local Uplink SINR in Distributed Operation)**.** *Assume distributed operation with MMSE channel estimation, L-MMSE receive combining, user loads satisfying $U \leq M$, and a) no pilot reuse or b) relatively accurate channel estimates and*

[15] Although $\mathbf{Z}_{kl}$, given by (3.89), only contains the estimated channels of UEs other than k, it may still be correlated with $\widehat{\mathbf{h}}_{kl}$ especially if UE k shares its pilot with other UEs. According to (3.35), the estimates $\widehat{\mathbf{h}}_{kl}$ and $\widehat{\mathbf{h}}_{il}$ are correlated if both UEs k and i have a LoS component to subarray l. Furthermore, the correlation between these estimates are more significant if UEs k and i share the same pilot sequence, as demonstrated in the first case in (3.35). However, this correlation will be neglected when deriving the SINR expression, since the overlap of LoS regions of different UEs is usually small and user k will normally share its pilot with only a few UEs, or even with no other user.

spatially uncorrelated NLoS channel component. Then, as the number of antennas grows large, the instantaneous local SINR converges deterministically to an asymptotic limit, i.e., $\Gamma_{kl}^{\text{dist}} \xrightarrow{M\to\infty} \Gamma_{kl}^{\text{dist,asy}}$. *This asymptotic SINR is given by*

$$\Gamma_{kl}^{\text{dist,asy}} = p_k \frac{\operatorname{tr}(\mathbf{Q}_{kl} - \mathbf{C}_{kl}) - \sum\limits_{i\in\mathcal{U}\setminus\{k\}} \frac{\operatorname{tr}\left[(\mathbf{Q}_{kl} - \mathbf{C}_{kl})(\mathbf{Q}_{il} - \mathbf{C}_{il})\right]}{\operatorname{tr}(\mathbf{Q}_{il} - \mathbf{C}_{il})}}{\sum\limits_{i\in\mathcal{U}} \frac{p_i}{M}\operatorname{tr}(\mathbf{C}_{il}) + \sigma_{\text{n}}^2}. \tag{3.121}$$

Proof. Under the assumption of spatially uncorrelated NLoS components, where $\mathbf{C}_{il} = \frac{1}{M}\operatorname{tr}(\mathbf{C}_{il})\mathbf{I}_M$, the instantaneous SINR approximation in (3.91) reduces to[16]

$$\Gamma_{kl}^{\text{dist}} \approx \frac{p_k}{\left(\sum\limits_{i\in\mathcal{U}} \frac{p_i}{M}\operatorname{tr}(\mathbf{C}_{il}) + \sigma_{\text{n}}^2\right)\|\mathbf{v}_{kl}\|^2}. \tag{3.122}$$

The squared norm $\|\mathbf{v}_{kl}\|^2$ is approximated using the local-ZF combiner, which behaves similarly to the L-MMSE combiner when the channel estimates at the selected subarrays are accurate and the SINR is high. For notational convenience, assume for a moment that all UEs are scheduled, *i.e.*, let $\mathcal{U} = \mathcal{K}$. Then, $\mathbf{v}_{kl}$ is approximated as the k-th column of $\mathbf{V}_l = \widehat{\mathbf{H}}_l(\widehat{\mathbf{H}}_l^{\text{H}}\widehat{\mathbf{H}}_l)^{-1}$, where $\widehat{\mathbf{H}}_l = [\widehat{\mathbf{h}}_{1l}, \ldots, \widehat{\mathbf{h}}_{Kl}]$, such that

$$\|\mathbf{v}_{kl}\|^2 = [\mathbf{V}_l^{\text{H}}\mathbf{V}_l]_{k,k} = [(\widehat{\mathbf{H}}_l^{\text{H}}\widehat{\mathbf{H}}_l)^{-1}]_{k,k} \overset{(a)}{=} [\widehat{\mathbf{h}}_{kl}^{\text{H}}\widehat{\mathbf{h}}_{kl} - \widehat{\mathbf{h}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl}(\ddot{\mathbf{H}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl})^{-1}\ddot{\mathbf{H}}_{kl}^{\text{H}}\widehat{\mathbf{h}}_{kl}]^{-1}, \tag{3.123}$$

where (a) comes from the block matrix inversion lemma yields [50, Appendix A], and $\ddot{\mathbf{H}}_{kl} \in \mathbb{C}^{M\times(K-1)}$ denotes the estimated channel matrix at subarray l excluding UE k, i.e., $\ddot{\mathbf{H}}_{kl} = [\widehat{\mathbf{h}}_{1l}, \ldots, \widehat{\mathbf{h}}_{(k-1)l}, \widehat{\mathbf{h}}_{(k+1)l}, \ldots, \widehat{\mathbf{h}}_{Kl}]$. Substituting (3.123) into (3.122) results in

$$\Gamma_{kl}^{\text{dist}} \approx p_k \frac{\widehat{\mathbf{h}}_{kl}^{\text{H}}\widehat{\mathbf{h}}_{kl} - \widehat{\mathbf{h}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl}(\ddot{\mathbf{H}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl})^{-1}\ddot{\mathbf{H}}_{kl}^{\text{H}}\widehat{\mathbf{h}}_{kl}}{\sum\limits_{i\in\mathcal{U}} \frac{p_i}{M}\operatorname{tr}(\mathbf{C}_{il}) + \sigma_{\text{n}}^2}. \tag{3.124}$$

The diagonal elements of $\ddot{\mathbf{H}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl}$ consists of the squared norms $\|\widehat{\mathbf{h}}_{il}\|^2$ for $i \in \mathcal{K}\setminus\{k\}$, which converge to $\operatorname{tr}(\mathbf{Q}_{il} - \mathbf{C}_{il})$ as $M \to \infty$, according to the first case of (3.38). In XL-MIMO, the off-diagonal terms are negligible since the channels are spatially non-stationary.[17] We

[16] The approximation in (3.122) remains accurate for correlated NLoS components, as long as there is no pilot reuse. This is because the eigenvalues of the matrix within parentheses in (3.91) are closely clustered in the absence of pilot reuse. Consequently, replacing $\mathbf{C}_{il}$ by $\frac{1}{M}\operatorname{tr}(\mathbf{C}_{il})\mathbf{I}_M$ introduces negligible variations in the SINR denominator.

[17] The off-diagonal elements of $\ddot{\mathbf{H}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl}$ are $\widehat{\mathbf{h}}_{il}^{\text{H}}\widehat{\mathbf{h}}_{i'l}$, for any $i, i' \in \mathcal{K}\setminus\{k\}$ with $i \neq i'$. According to the second or third cases in (3.38), these terms converge to $\overline{\mathbf{h}}_{il}^{\text{H}}\overline{\mathbf{h}}_{i'l} + \sqrt{p_i p_{i'}}\tau_{\text{p}}\operatorname{tr}\left(\mathbf{R}_{il}\boldsymbol{\Psi}_{t_il}^{-1}\mathbf{R}_{i'l}\right)$ as $M \to \infty$, if $t_i = t_{i'}$, and to $\overline{\mathbf{h}}_{il}^{\text{H}}\overline{\mathbf{h}}_{i'l}$ otherwise. In practice, different UEs are typically assigned distinct pilot sequences. Furthermore, the large dimension of the XL-MIMO antenna array ensures that the LoS regions of different UEs overlap on only a few subarrays or may not overlap at all. As a result, the off-diagonal terms of $\ddot{\mathbf{H}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl}$ are generally much smaller than the diagonal terms, especially for a large number of antennas. On the other hand, they are normally non-negligible in conventional massive MIMO, where the antennas are co-located such that the LoS region of a given UE encompasses all the BS antennas or none of them.

can thus approximate (3.124) by

$$\Gamma_{kl}^{\text{dist}} \approx p_k \frac{\widehat{\mathbf{h}}_{kl}^{\text{H}}\widehat{\mathbf{h}}_{kl} - \widehat{\mathbf{h}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl}\mathbf{T}_{kl}^{-1}\ddot{\mathbf{H}}_{kl}^{\text{H}}\widehat{\mathbf{h}}_{kl}}{\sum_{i\in\mathcal{U}} \frac{p_i}{M}\operatorname{tr}(\mathbf{C}_{il}) + \sigma_{\text{n}}^2}. \tag{3.125}$$

where $\mathbf{T}_{kl}$ is obtained by retaining only the diagonal entries of $\ddot{\mathbf{H}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl}$:

$$\begin{aligned}(\mathbf{T}_{kl})_{i,i'} &= \begin{cases}(\ddot{\mathbf{H}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl})_{i,i} & \text{if } i' = i,\\ 0 & \text{otherwise,}\end{cases}\\ &= \operatorname{diag}\left(\|\widehat{\mathbf{h}}_{1l}\|^2, \dots, \|\widehat{\mathbf{h}}_{(k-1)l}\|^2, \|\widehat{\mathbf{h}}_{(k+1)l}\|^2, \dots, \|\widehat{\mathbf{h}}_{Kl}\|^2\right).\end{aligned} \tag{3.126}$$

Applying the first case of (3.38) results in

$$\widehat{\mathbf{h}}_{kl}^{\text{H}}\widehat{\mathbf{h}}_{kl} \xrightarrow{M\to\infty} \operatorname{tr}(\mathbf{Q}_{kl} - \mathbf{C}_{kl}) \tag{3.127}$$

and

$$\widehat{\mathbf{h}}_{kl}^{\text{H}}\ddot{\mathbf{H}}_{kl}\mathbf{T}_{kl}^{-1}\ddot{\mathbf{H}}_{kl}^{\text{H}}\widehat{\mathbf{h}}_{kl} \overset{(a)}{=} \sum_{i\in\mathcal{U}\setminus\{k\}}^{K} \frac{\widehat{\mathbf{h}}_{kl}^{\text{H}}\widehat{\mathbf{h}}_{il}\widehat{\mathbf{h}}_{il}^{\text{H}}\widehat{\mathbf{h}}_{kl}}{\widehat{\mathbf{h}}_{il}^{\text{H}}\widehat{\mathbf{h}}_{il}} \xrightarrow{M\to\infty} \frac{\operatorname{tr}\left[(\mathbf{Q}_{kl} - \mathbf{C}_{kl})(\mathbf{Q}_{il} - \mathbf{C}_{il})\right]}{\operatorname{tr}\left(\mathbf{Q}_{il} - \mathbf{C}_{il}\right)}, \tag{3.128}$$

where (a) comes from simple algebraic manipulation. The proof of Theorem 3.4 is completed by substituting (3.127) and (3.128) into (3.125). □

Corollary 3.17. *Using the asymptotic approximation for the local SINR, the corresponding SE estimate is given by*

$$\hat{\eta}_k^{\text{dist,asy}} = \left(1 - \frac{\tau_{\text{p}}}{\tau_{\text{c}}}\right)\log_2\left[1 + \left(\sum_{l\in\mathcal{D}_k} \mu_{kl}^2 \left(\Gamma_{kl}^{\text{dist,asy}}\right)^{-1}\right)^{-1}\right]. \tag{3.129}$$

Corollary 3.18. *The computational complexity (in real multiplications) to evaluate* (3.121) *at the L_k serving subarrays is*

$$\mathcal{C}_k^{\text{dist,asy}} = 3(U-1)M^2L_k. \tag{3.130}$$

Proof. The computational cost primarily arises from evaluating the matrix product trace $\operatorname{tr}[(\mathbf{Q}_{kl} - \mathbf{C}_{kl})(\mathbf{Q}_{il} - \mathbf{C}_{il})]$ for each $i \in \mathcal{U}\setminus\{k\}$. Since only the diagonal elements of the matrix product are needed to compute the trace, the cost is M^2 complex multiplications per UE i. Considering all $U-1$ interfering UEs and all L_k subarrays that serve UE k, and assuming that each complex multiplication requires three real multiplications [8, page 378], the computational cost is given by (3.130). □

3.4 Performance Comparison

This section evaluates the uplink performance of an XL-MIMO system, contrasting centralized and distributed processing architectures under various subarray selection and weighting strategies. The BS comprises $L = 16$ subarrays arranged in a linear horizontal layout ($L_\mathrm{y} = 16$, $L_\mathrm{z} = 1$), each equipped with a square planar array of $M = 16$ antennas ($M_\mathrm{y} = M_\mathrm{z} = 4$). The system serves $U = 16$ single-antenna UEs, randomly distributed within a 100×50 m coverage area located in front of the array.

The coherence block length is set to $\tau_\mathrm{c} = 16$ samples. The UEs are randomly assigned to one of $\tau_\mathrm{p} = 8$ orthogonal pilot sequences. Since the number of scheduled UEs exceeds the number of pilots ($U > \tau_\mathrm{p}$), pilot reuse is mandatory, rendering pilot contamination unavoidable. Table 3.2 summarizes the remaining simulation parameters, including the geometric configuration and channel characteristics.

Table 3.2 – Simulation parameters

Parameter	**Value**
Number of subarrays (L)	16
Antennas per subarray (M)	$4 \times 4 = 16$
Average height of the antenna array	3.0 m
Height of the UEs	1.5 m
Horizontal distance between consecutive subarrays	6.0 m
Vertical distance between consecutive subarrays	2.0 m
Azimuth angle standard deviation (σ_φ)	10°
Elevation angle standard deviation (σ_θ)	10°
Carrier frequency (f)	6 GHz
Intra-subarray antenna separation (d)	$0.5\lambda = 2.5$ cm
UE transmit power (p_k)	20 dBm
Noise power (σ_n^2)	-96 dBm

Table 3.3 summarizes the twelve scenarios defined to evaluate the system performance under different configurations. In all cases, each UE k is served by a subset of L_k subarrays. The scenarios are categorized by operation mode, subarray selection criterion, and weighting strategy. Scenarios 1 through 3 address the centralized operation, while scenarios 4 through 12 explore the distributed operation. Three subarray selection criteria are considered: (i) random; (ii) LSF-based, which prioritizes signal strength by selecting subarrays with the highest LSF channel gains (β_{kl}); and (iii) SINR-based, which optimizes the trade-off between signal power and interference by selecting subarrays with the highest local SINRs. For the distributed operation, the three weighting schemes defined in Section 3.3.3 are compared: (i) equal, (ii) LSF-based, and (iii) SINR-based.[18]

[18] Note that weighting is not listed for the centralized mode, as the CPU performs joint interference suppression and signal processing across all selected subarrays.

Table 3.3 – Comparison of scenarios for centralized and distributed operations.

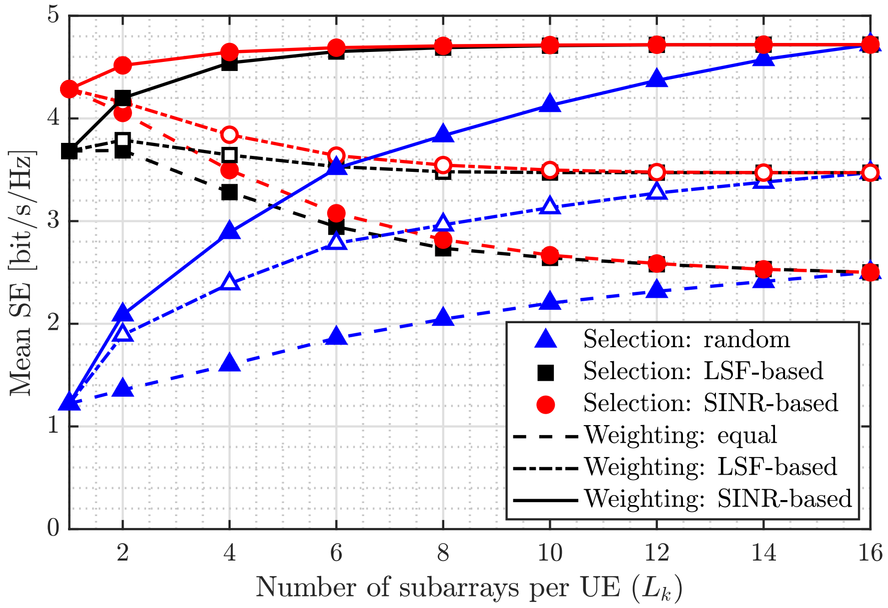
Scenario	**Operation**	**Subarray selection**	**Weighting**
1	Centralized	Random	
2	Centralized	LSF-based	
3	Centralized	SINR-based	
4	Distributed	Random	Equal
5	Distributed	LSF-based	Equal
6	Distributed	SINR-based	Equal
7	Distributed	Random	LSF-based
8	Distributed	LSF-based	LSF-based
9	Distributed	SINR-based	LSF-based
10	Distributed	Random	SINR-based
11	Distributed	LSF-based	SINR-based
12	Distributed	SINR-based	SINR-based

Figure 3.4 depicts the mean per-user SE as a function of the number of selected subarrays, L_k, for the distinct scenarios detailed in Table 3.3. Specifically, Figure 3.4a evaluates the distributed architecture (scenarios 4–12), assessing the impact of different subarray selection and weighting strategies. Subsequently, Figure 3.4b benchmarks the distributed operation with SINR-based weighting against the centralized architecture (scenarios 1–3 and 10–12), while varying the subarray selection method.

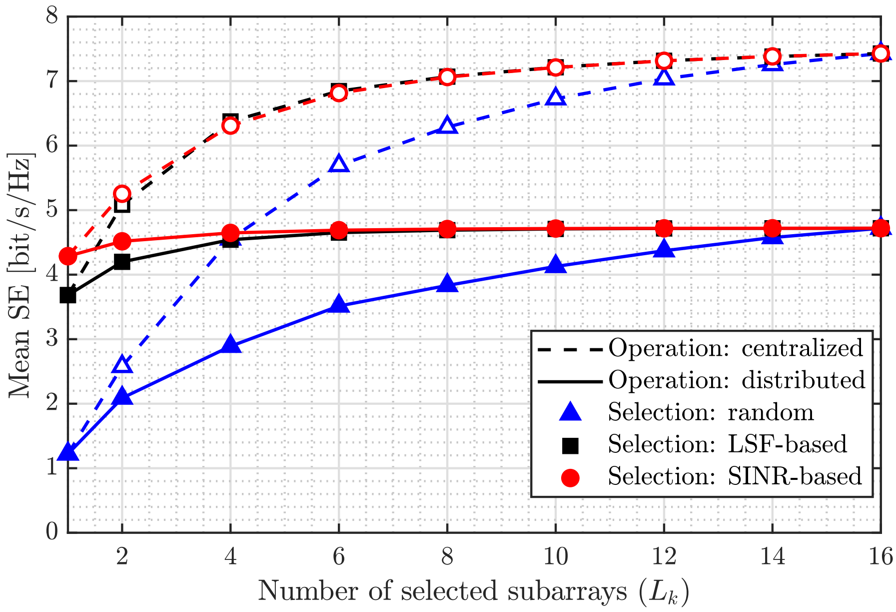
Figure 3.4a reveals distinct convergence behaviors at the boundaries of the subarray selection range. At $L_k = 1$, performance is dictated solely by the *subarray selection* strategy, causing the curves to cluster into three distinct groups. In this regime, the choice of weighting strategy is irrelevant, as the final signal estimate is simply the local estimate from the single selected subarray. Conversely, at $L_k = L = 16$, the selection strategy is irrelevant since the full set of subarrays is utilized regardless of the selection criterion ($\mathcal{D}_k = \mathcal{D}$). Consequently, the curves cluster into three groups determined exclusively by the *weighting* scheme. Further insights drawn from Figure 3.4 are detailed below.

In the regime of limited subarray selection (e.g., $L_k = 2$), LSF- and SINR-based methods demonstrate a substantial performance advantage over random selection. Specifically, the superiority of the SINR-based scheme confirms that accounting for interference is essential when the serving set is small. As L_k grows, the LSF-based selection aligns with the SINR-based optimum. This occurs because the subsets of subarrays selected by both strategies tend to coincide; the subarrays exhibiting the strongest LSF gains are typically the main contributors to useful signal power, dominating the numerator of the SINR expression. Thus, maximizing the channel gain serves as an accurate proxy for SINR maximization when L_k is sufficiently large.

The SINR-based weighting strategy establishes a performance upper bound, exhibiting a monotonic increase in mean per-user SE with L_k that converges to a max-



(a) Distributed operation: impact of subarray selection and weights.



(b) Architecture comparison: centralized vs. distributed with SINR-based weights.

Figure 3.4 – Mean per-user SE versus number of selected subarrays L_k.

imum of approximately 4.7 bit/s/Hz at $L_k = 16$. In contrast, performance degrades as L_k increases when equal weighting is employed with LSF- or SINR-based selection. This indicates that indiscriminately adding subarrays with poor channel conditions or high interference levels, without attenuating their contributions via appropriate weighting, compromises the quality of the combined signal, thereby reducing the global SINR. Consequently, expanding the set of serving subarrays in distributed operation proves beneficial only when SINR-based weights are applied. Finally, LSF-based weighting offers an intermediate solution between the optimal SINR-based and the equal weighting strategies.

Figure 3.4b demonstrates that centralized processing consistently outperforms the distributed architecture, due to its capability to jointly suppress inter-user interference across all active antennas. While SINR-based selection yields the highest performance for both architectures, the SE in centralized operation saturates at $\approx$ 7.5 bit/s/Hz, whereas the distributed approach hits a saturation plateau at significantly lower values ($\approx$ 4.7 bit/s/Hz). Notably, the LSF-based selection scheme closely tracks the SINR-based trend, confirming that LSF channel gain is a robust metric for subarray selection.

The results highlight the trade-off between SE and complexity. While centralized operation yields superior SE, distributed operation offers a viable low-complexity alternative. To maximize its efficacy, two design guidelines emerge: i) LSF-based subarray selection is recommended to achieve near-optimal performance without the overhead of local SINR estimation, and b) SINR-based weighting is essential to prevent performance degradation when scaling L_k.

Finally, the analysis demonstrates that selecting only a small subset of subarrays to serve each UE is sufficient to achieve high SE, offering significant reductions in computational complexity and fronthaul load compared to the full activation baseline ($L_k = L$).

3.5 Accuracy of The Deterministic SINR Expressions

This section presents numerical results validating the accuracy of the derived SINR approximations. Unless stated otherwise, simulations assume $L = 16$, $L_k = 4$, LSF-based subarray selection, SINR-based weighting, random pilot assignment, and other parameters listed in Table 3.2. The performance-complexity trade-off observed in Figure 3.4b justifies the choice of $L_k = 4$. Pilot contamination is enforced by setting $\tau_{\mathrm{p}} = U/2$, while $\tau_{\mathrm{c}} = U$.

The ergodic SE approximations are expected to closely match the exact average SE in the regime where $U \gg ML_k$ (centralized operation) or $U \gg M$ (distributed operation). Accordingly, the normalized absolute error (NAE) between the closed-form approximation and the exact average SE for centralized and distributed operations is defined, respectively, as[19]

$$\tilde{\eta}_k^{\mathrm{cent,erg}} = \mathrm{NAE}(\mathbb{E}\{\eta_k^{\mathrm{cent}}\}, \eta_k^{\mathrm{cent,erg}}) = \frac{|\eta_k^{\mathrm{cent,erg}} - \mathbb{E}\{\eta_k^{\mathrm{cent}}\}|}{\mathbb{E}\{\eta_k^{\mathrm{cent}}\}}, \tag{3.131}$$

$$\tilde{\eta}_k^{\mathrm{dist,erg}} = \mathrm{NAE}(\mathbb{E}\{\eta_k^{\mathrm{dist}}\}, \hat{\eta}_k^{\mathrm{dist,erg}}) = \frac{|\hat{\eta}_k^{\mathrm{dist,erg}} - \mathbb{E}\{\eta_k^{\mathrm{dist}}\}|}{\mathbb{E}\{\eta_k^{\mathrm{dist}}\}}, \tag{3.132}$$

Conversely, the asymptotic SE approximation relies on channel hardening, converging to the instantaneous SE as the number of effective antennas increases ($ML_k \to \infty$ in centralized operation and $M \to \infty$ in distributed). Its accuracy is quantified by the mean

[19] Throughout this thesis, $\mathbb{E}\{\cdot\}$ denotes the expectation over SSF realizations.

NAE (MNAE) relative to the instantaneous SE:

$$\tilde{\eta}_k^{\text{cent,asy}} = \text{MNAE}(\mathbb{E}\{\eta_k^{\text{cent}}\}, \eta_k^{\text{cent,asy}}) = \mathbb{E}\left\{\frac{|\eta_k^{\text{cent,asy}} - \eta_k^{\text{cent}}|}{\eta_k^{\text{cent}}}\right\}, \tag{3.133}$$

$$\tilde{\eta}_k^{\text{dist,asy}} = \text{MNAE}(\mathbb{E}\{\eta_k^{\text{dist}}\}, \hat{\eta}_k^{\text{dist,asy}}) = \mathbb{E}\left\{\frac{|\hat{\eta}_k^{\text{dist,asy}} - \eta_k^{\text{dist}}|}{\eta_k^{\text{dist}}}\right\}, \tag{3.134}$$

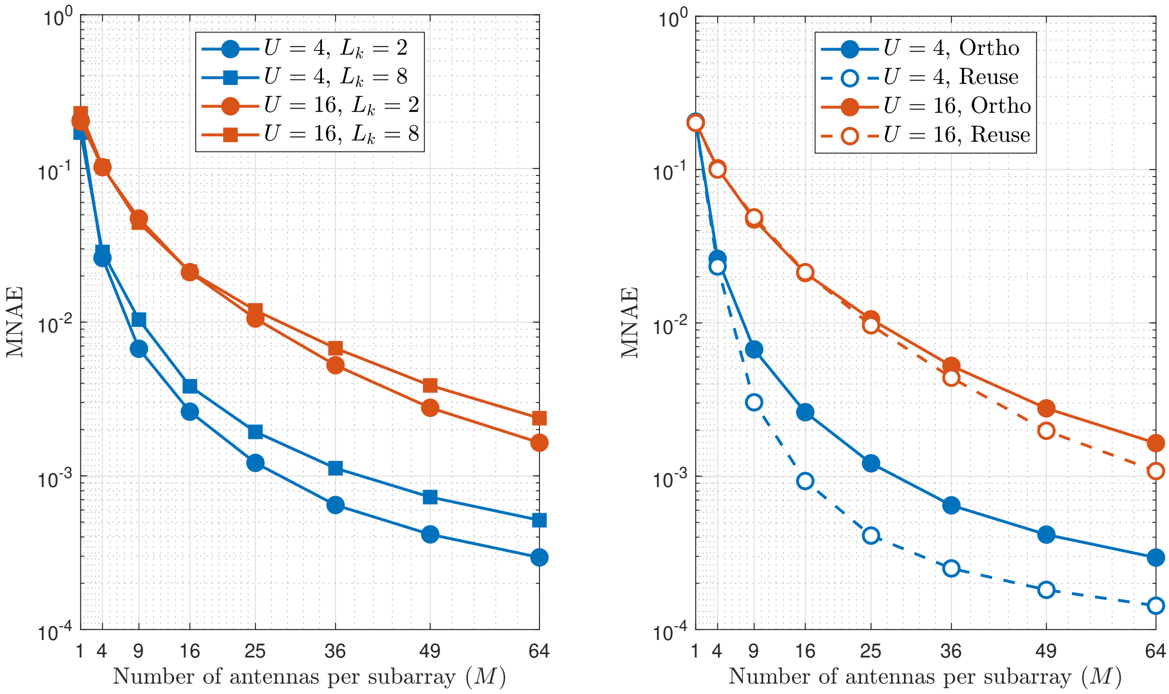
Finally, the MNAE associated with the instantaneous SE approximation in distributed operation is given by

$$\tilde{\eta}_k^{\text{dist}} = \text{MNAE}\left(\eta_k^{\text{dist}}, \hat{\eta}_k^{\text{dist}}\right) = \mathbb{E}\left\{\frac{|\hat{\eta}_k^{\text{dist}} - \eta_k^{\text{dist}}|}{\eta_k^{\text{dist}}}\right\}. \tag{3.135}$$

Each data point in the following figures represents the mean accuracy computed by averaging over the scheduled UEs and LSF realizations.

3.5.1 Accuracy of The Global SINR Approximation (Distributed Operation)

Figure 3.5 illustrates the MNAE of the instantaneous SE approximation ($\tilde{\eta}_k^{\text{dist}}$) as a function of the number of antennas per subarray (M).



(a) Ideal scenario: orthogonal pilots. (b) Impact of pilot reuse (fixed $L_k = 2$).

Figure 3.5 – MNAE of the instantaneous global SE approximation as a function of the number of subarray antennas (M) in distributed uplink operation. The results highlight the impact of the number of UEs (U), the number of selected subarrays (L_k), and pilot reuse on the approximation accuracy.

Figure 3.5a assesses the approximation accuracy when each UE is assigned a different pilot from a set of mutually orthogonal pilots, taking $\tau_{\text{p}} = U$ and $\tau_{\text{c}} = 2U$. The

MNAE exhibits a monotonic decline with M and an increase with U, as a larger ratio M/U enhances interference suppression, satisfying the derivation assumption of negligible coherent interference. Notably, selecting fewer subarrays (e.g., $L_k = 2$) yields lower MNAE compared to broader connectivity (e.g., $L_k = 8$). This behavior arises because the LSF-based selection restricts connectivity to the strongest subarrays, where $\mathbf{v}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{kl} \approx 1$, thus better fulfilling the conditions underlying Corollaries 3.12 and 3.13. These factors validate the robustness of the approximation, even in the overloaded regime ($M < U$).

Figure 3.5b isolates the impact of pilot contamination by comparing orthogonal assignment (solid lines) with random pilot reuse (dashed lines) for fixed $L_k = 2$. Interestingly, pilot reuse leads to a consistent reduction in the approximation error, with the MNAE for the reuse cases being strictly lower than their orthogonal counterparts. This occurs because pilot reuse significantly increases the covariance matrices of the channel estimation errors, rendering the residual interference terms $\mathbf{v}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{il}$ ($i \neq k$)—which are approximated as zero when deriving the global SINR approximation—negligible relative to the dominant interference power. Consequently, the approximation assumptions hold more robustly in the reuse regime.

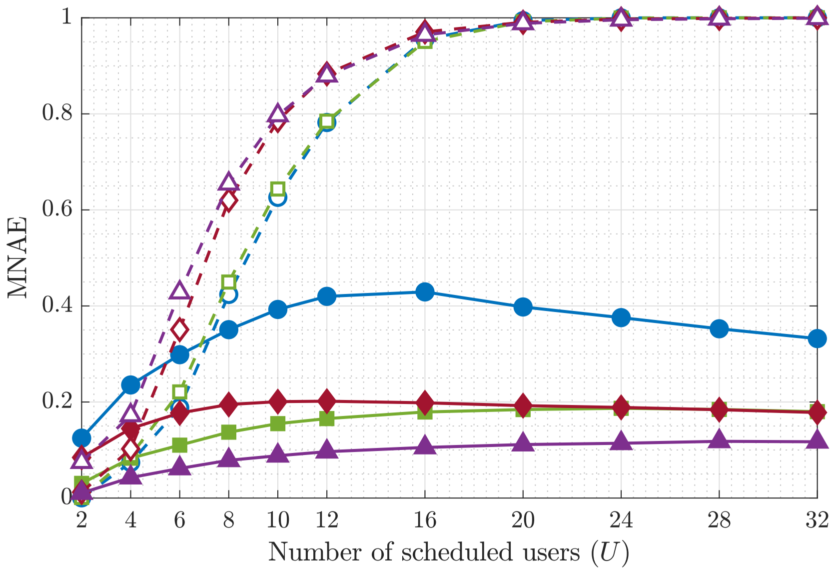
3.5.2 Accuracy of The Ergodic & Asymptotic SINR Approximations: Impact of System Load

Figure 3.6 evaluates the accuracy of the ergodic and asymptotic approximations as a function of the number of scheduled UEs, U. The analysis contrasts distributed (Fig. 3.6a) and centralized (Fig. 3.6b) operations, assessing the impact of spatial correlation and pilot contamination.

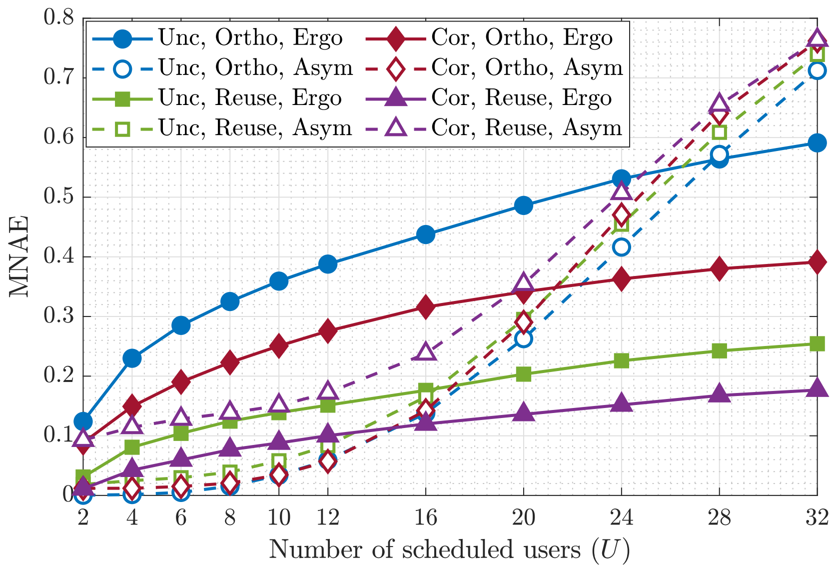
The MNAE of the asymptotic expression increases with U. As the user load grows, the rising variance of the aggregate interference causes the instantaneous SE to deviate from the deterministic limit. This divergence is significantly more pronounced in distributed operation. Centralized operation benefits from a larger number of antennas used by the receive combiner (ML_k versus M), which enhances the channel hardening effect and accelerates convergence to the deterministic mean, effectively delaying the error growth.

Furthermore, spatial correlation and pilot reuse degrade the accuracy of the asymptotic approximations by violating the tractability assumptions behind (3.69) and (3.121), which premise on uncorrelated NLoS components and small channel estimation errors. Moreover, correlation reduces the effective degrees of freedom, hindering channel hardening and delaying SINR convergence to the deterministic regime, while pilot contamination introduces coherent interference that scales with U.

Counter-intuitively, spatial correlation and pilot reuse reduce the MNAE of the ergodic approximations. This occurs because the increased deterministic covariance, $\mathbf{C}_{kl}$, stabilizes $\mathbf{Z}_{kl}$ around its mean, better satisfying the Neumann series convergence condition



(a) Distributed uplink operation.



(b) Centralized uplink operation.

Figure 3.6 – MNAE of the deterministic SE approximations versus the number of scheduled users (U). The results evaluate the accuracy of ergodic and asymptotic approximations under spatially correlated and uncorrelated fading, comparing orthogonal pilot assignment against pilot reuse.

required for these approximations. The error evolution is governed by competing effects: it initially rises with U as added interferers increase the variance of $\mathbf{Z}_{kl}$,[20] but peaks near the system's spatial degrees of freedom ($U \approx M$ for distributed and $U \approx ML_k$ for centralized operation). Beyond this peak, $\mathbf{Z}_{kl}$ concentrates around its expectation $\overline{\mathbf{Z}}_{kl}$, diminishing the gap between the exact mean SINR and the ergodic approximation. This explains

[20] In the low-load regime, each new interferer increases the aggregate interference variance. At this stage, the "interference hardening" effect is not yet dominant, making the statistical mean a less stable predictor of instantaneous realizations.

why the ergodic approximation becomes accurate at lower U in distributed operation: it saturates the available degrees of freedom (M) earlier than the centralized case (ML_k).[21]

In the distributed scenario (Figure 3.6a), the peak and subsequent decay of the error are discernible since the saturation point $U \approx M = 16$ falls within the simulated range. Conversely, in the centralized scenario (Figure 3.6b), the MNAE exhibits a monotonic increase because the saturation point $U \approx ML_k = 64$ lies beyond the maximum simulated load of $U = 32$.

Given these contrasting behaviors, a switching point, U_{switch}, is defined to select the optimal expression. The asymptotic formula is prioritized for $U \leq U_{\text{switch}}$ due to its superior accuracy in low-load regimes and lower complexity, while the ergodic formula is adopted for $U > U_{\text{switch}}$. The optimal U_{switch} shifts toward lower values in the presence of spatial correlation or pilot reuse. Table 3.4 summarizes the thresholds empirically determined from Figure 3.6.

Table 3.4 – Switching threshold U_{switch} values for the deterministic SE approximations.

Operation Mode	**Uncorrelated NLoS**		**Correlated NLoS**	
	Orthogonal	Pilot Reuse	Orthogonal	Pilot Reuse
Centralized	$ML_k/2$	$ML_k/4$	$ML_k/4$	0
Distributed	$M/2$	$M/4$	$M/4$	0

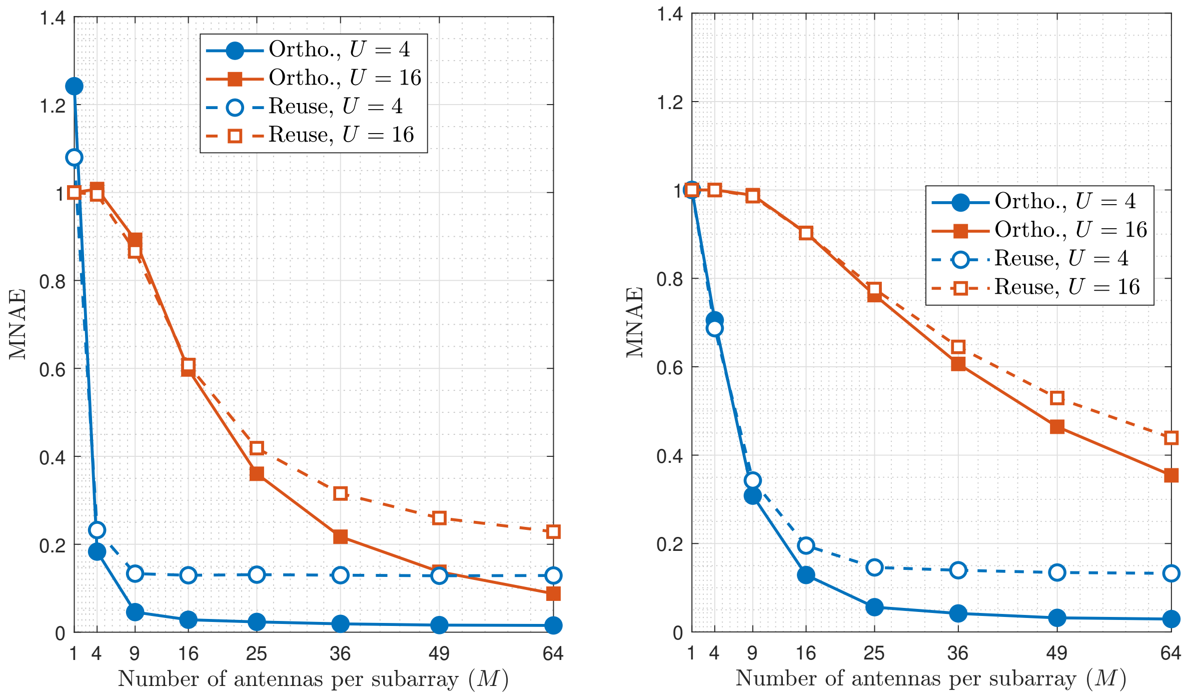
Notably, the thresholds for centralized operation scale by L_k relative to the distributed case. For instance, with uncorrelated NLoS and orthogonal pilots, the threshold sits at half the total combining antennas: $U_{\text{switch}} = ML_k/2$ (centralized) versus $U_{\text{switch}} = M/2$ (distributed). Furthermore, in the most severe scenario (correlated NLoS with pilot reuse), the threshold drops to $U_{\text{switch}} \approx 0$, indicating that the asymptotic approximation becomes unreliable under strong coherent interference and spatial correlation, mandating the exclusive use of the ergodic approximation. These results confirm the decision rules implemented in Algorithms 4.3 and 4.4, supporting the use of the ergodic expression for heavy loading or correlated channels, and the asymptotic expression as a deterministic proxy for instantaneous performance in the large-system limit.

3.5.3 Asymptotic Approximation Accuracy

Figure 3.7 assesses the accuracy of the asymptotic SE approximation as a function of the subarray size (M) for both centralized (Figure 3.7a) and distributed (Figure 3.7b) architectures. In all scenarios, the MNAE exhibits a monotonic decline as M increases, confirming the instantaneous SE converges toward the deterministic asymptotic limit. However, the convergence rate differs markedly between the architectures. Centralized

[21] In distributed operation, receive combining is local (M antennas), whereas centralized operation combines jointly over ML_k antennas.

operation yields significantly lower errors than distributed operation for the same M, owing to the effective array aperture of ML_k antennas discussed in Section 3.5.2. Consequently, the centralized architecture renders the asymptotic approximation accurate even for moderate values of M.



(a) Centralized Uplink Operation (b) Distributed Uplink Operation

Figure 3.7 – MNAE of the asymptotic SE approximation as a function of the number of subarray antennas (M), comparing orthogonal pilot assignment (solid lines) versus pilot reuse (dashed lines), under spatially correlated NLoS channels.

The results also underscore the sensitivity of the asymptotic approximation to interference. Pilot reuse curves consistently manifest higher errors than the orthogonal baseline. Unlike the instantaneous approximation in Figure 3.5, where pilot reuse reduced the error, the coherent interference hinders the convergence to the asymptotic limit, requiring larger arrays to achieve the same accuracy as the orthogonal case. Finally, higher user loads ($U = 16$) exacerbate the error by increasing the aggregate interference variance, particularly in the distributed case where spatial degrees of freedom are more constrained.

3.6 Summary

This chapter derived novel closed-form deterministic SINR approximations for both centralized and distributed uplink operations. These expressions were obtained assuming Rician fading, MMSE channel estimation (imperfect CSI), and MMSE receive combining—a specific combination of factors that distinguishes this work from the existing literature. These deterministic approximations serve as powerful tools for system analysis and design, as they depend solely on slowly varying channel statistics rather than instantaneous channel.

The derived expressions are categorized into two groups: ergodic expressions, which approximate the average SINR, and asymptotic expressions, which approximate the instantaneous SINR when the number of subarray antennas is sufficiently large.

Building on the foundations laid in Chapter 2, Section 3.1 modeled the hybrid LoS+NLoS channels between a modular BS array and multiple single-antenna UEs, each served by a dynamic subset of subarrays.

Sections 3.2 and 3.3 addressed the centralized and distributed operations, respectively. For both architectures, closed-form expressions for the instantaneous, ergodic, and asymptotic uplink SINR were provided, along with an analysis of their computational complexity. In the context of distributed operation, Section 3.3 also derived the global uplink SINR based on local estimates and formulated the optimal combining weights to maximize it.

The numerical results in Section 3.4 compared the spectral efficiency of both architectures. While centralized processing delivers superior performance due to joint interference suppression, distributed operation offers a competitive alternative with lower complexity. Notably, the results demonstrated that LSF-based subarray selection achieves performance similar to SINR-based selection without requiring local SINR estimation. Furthermore, serving each UE with a small subset of subarrays yields SE close to the full-array baseline, but with significantly reduced computational load.

Section 3.5 validated the accuracy of the proposed approximations under various conditions, including spatially correlated NLoS propagation and distinct pilot assignment strategies (orthogonal vs. reuse). The analysis identified the specific regimes—defined by user load and antenna number—in which the ergodic and asymptotic expressions provide accurate performance estimates.

These findings justify the use of the proposed approximations as reliable performance predictors and as practical tools for resource allocation. Building on the deterministic SINR framework and the numerical insights developed here, Chapter 4 proposes algorithms for joint subarray selection, user scheduling, and pilot assignment.

4 APPLYING USER SCHEDULING, PILOT ASSIGNMENT & SUBARRAY SELECTION IN XL-MIMO

XL-MIMO systems deliver substantial advances in spectral efficiency, spatial resolution, and link reliability compared to conventional massive MIMO. Thanks to an order-of-magnitude increase in antenna elements and unique near-field propagation effects, XL-MIMO enables unprecedented spatial degrees of freedom and precise user separation. However, tapping into these capabilities in practical deployments introduces new challenges in scalability, interference management, and channel estimation overhead.

This chapter focuses on the joint design of subarray selection, user scheduling, and pilot assignment, which are key enablers for resource-efficient operation of XL-MIMO under realistic system constraints, and exploits their interdependence to maximize overall system performance under practical constraints.

Although it may seem optimal to serve each UE with all available antennas, doing so is neither necessary nor scalable in practical XL-MIMO deployments. In most scenarios, each UE is close to only a small subset of subarrays. Through subarray selection, each UE is dynamically served only by those subarrays that significantly contribute to its received signal, maintaining service quality while reducing fronthaul signaling and computational complexity [9, 36, 37].[1] Constraining the number of UEs served per subarray ensures network scalability, especially as user density grows [9].

User scheduling plays a crucial role in environments where the number of UEs exceeds the system's spatial degrees of freedom. Serving too many UEs increases interference and degrades throughput [38]. Therefore, selecting a subset of UEs that can be served concurrently while meeting rate requirements is essential to reduce interference and improve fairness, throughput and computational efficiency [39].

Accurate channel estimation is fundamental for coherent processing in XL-MIMO. This depends on pilot assignment: ideally, orthogonal pilots are used to avoid pilot contamination, but the limited coherence block length in densely deployed or high-mobility networks limits the number of orthogonal pilots that can be used [9], and pilot reuse becomes necessary, which introduces pilot contamination, channel estimation errors and performance loss [40]. Smart pilot assignment, aligned to subarray selection and user scheduling, are thus crucial to control interference and estimation overhead.

Importantly, in XL-MIMO, subarrays serving different UEs typically do not overlap, which enables more aggressive pilot reuse strategies than in co-located massive MIMO

[1] Subarray selection reduces fronthaul signaling since only the selected subarrays need to exchange uplink and downlink data with the CPU, and reduces computational complexity since the subarrays process signals for fewer users.

arrays. The spatial sparsity of VRs means that pilot sharing-induced interference is naturally mitigated, setting XL-MIMO apart from conventional massive MIMO wherein every scheduled user is typically 'seen' by all antennas. Together, joint strategies for subarray selection, user scheduling, and pilot assignment form the backbone of scalable, high-performance XL-MIMO operation.

A widely adopted guideline for pilot assignment in XL-MIMO is that each subarray should serve at most one UE per pilot, typically selecting the UE with the strongest channel gain for that subarray. This ensures high effective SNR with only one UE using that pilot [9], and limits pilot contamination, as sharing a pilot among UEs with overlapping VRs would cause the strongest UE to dominate the channel estimation process and degrade the estimation quality for others. Another common approach is to allocate pilots such that the covariance matrices of pilot-sharing UEs are sufficiently distinct [9].

Designing resource allocation algorithms based on instantaneous CSI is computationally intractable in XL-MIMO due to the massive number of antennas and users [41]. A more practical and scalable solution relies on statistical CSI, which vary slowly over time and allows resource allocation decisions to remain valid over many coherence blocks. This reduces both computational and signaling overhead.

To enable scalable and practical resource allocation in XL-MIMO, this work leverages the deterministic SINR approximations derived in Chapter 3. These closed-form expressions capture the effects of system geometry, channel estimation errors, and pilot contamination based solely on large-scale channel statistics, thus sidestepping the need for costly Monte Carlo simulations or instantaneous channel realizations. As these deterministic SINR formulas depend only on slowly varying statistical CSI, resource allocation decisions can remain valid over many coherence blocks, substantially reducing computational and signaling overhead. Channel hardening in XL-MIMO further justifies statistical-based decisions, since the variance of the effective channel decreases with larger number of antennas. Consequently, the proposed methods, built on statistical CSI and deterministic SINR formulas, achieve scalable, robust, and near-optimal performance, with solutions that need not be recomputed every coherence block but remain effective across extended periods.

Optimal algorithms for user scheduling, pilot assignment or subarray selection are combinatorial optimization problems with prohibitively high computational costs, since the complexity of evaluating all possibilities grows exponentially with the number of UEs. Thus, only suboptimal methods are feasible in practice [9, 41].

The simplest strategy is random pilot assignment combined with random subarray selection for each UE. While this method is computationally trivial, it offers no control over pilot contamination. The following sections present more sophisticated algorithms based on statistical CSI, aiming to maximize the minimum per-user SE with much lower computational complexity burden than algorithms based on instantaneous CSI.

For instance, the algorithm in Section 4.2 bases pilot assignment on minimizing the worst-case channel estimation NMSE among scheduled users, thereby avoiding assigning the same pilot to spatially close users or users with highly similar spatial correlation. This enhances the minimum achievable per-user SE. By contrast, the methods in Sections 4.3 and 4.4 use deterministic SINR expressions, aligning resource allocation directly with SE maximization objectives, as validated by the numerical results in Section 4.5.

4.1 Greedy Pilot Assignment & Subarray Selection

A simple and practical strategy for joint pilot assignment and subarray selection aimed at mitigating pilot contamination is the greedy algorithm described in [9, Sec. 4.4]. Although it does not strictly guarantee that all UEs are scheduled, in practice it almost always ensures that each user is served by at least one subarray. The algorithm operates in two stages, as outlined in Algorithm 4.1. Initally (lines 2-4), a set of τ_p orthogonal pilot sequences is randomly assigned to τ_p randomly selected UEs. Subsequently (lines 5-8), the remaining UEs are processed sequentially: each UE k is assigned the pilot t that minimizes the total NLoS channel gain of the UEs already using that pilot.[2] The contamination metric relies on the NLoS channel gains to the UE's strongest subarray, thereby encouraging nearby UEs (which tend to exhibit similar spatial correlation characteristics) to use distinct pilots. This approach effectively reduces pilot contamination while maintaining low algorithmic complexity [9].

After pilot assignment, the CPU communicates the pilot-UE maping to all subarrays. This step introduces negligible fronthaul overhead since pilot allocation depends only on large-scale channel statistics, which vary slowly over time. Subarray selection is then performed (lines 9-14) once all UEs have been assigned pilots. Each subarray is allowed to serve at most one UE per pilot—specifically, the UE with the largest LSF coefficient among those sharing that pilot. This policy is both simple and practical: it allows each subarray to make autonomous decisions based on slowly varying statistical information and, when combined with a pilot assignment that separates spatially proximate UEs onto different pilots, typically guarantees that every UE is served by at least one subarray.

4.2 NMSE-Based User Scheduling & Pilot Assignment

Algorithm 4.2 proposes a joint user scheduling and pilot assignment strategy designed to maximize the number of scheduled UEs while ensuring that each of them meets

[2] Only the NLoS component is considered because the LoS channel is deterministic and does not contribute to the channel estimation error.

Algorithm 4.1 Joint Greedy Pilot Assignment and Subarray Selection

Require:
LSF coefficients: $\{\beta_{kl}, \check{\beta}_{kl}\}$
Number of pilot sequences: τ_p
Ensure:
Assigned pilot sequences: $\{t_k\}_{k\in\mathcal{K}}$
Subarray sets: $\{\mathcal{D}_k\}_{k\in\mathcal{K}}$
1: Initialize: $\mathcal{D}_k \leftarrow \emptyset, \quad k \in \mathcal{K}$
2: **for** $k = 1, \dots, \tau_\text{p}$ **do**
3: $\quad t_k \leftarrow k$ ▷ Assign orthogonal pilots to the first τ_p UEs
4: **end for**
5: **for** $k = \tau_\text{p} + 1, \dots, K$ **do**
6: $\quad l_k \leftarrow \arg\max\limits_{l\in\mathcal{D}} \beta_{kl}$ ▷ Select the best subarray for UE k
7: $\quad t_k \leftarrow \arg\min\limits_{t\in\{1,\dots,\tau_\text{p}\}} \sum\limits_{\substack{i=1\\ t_i=t}}^{k-1} \check{\beta}_{il_k}$ ▷ Choose the best pilot for user k
8: **end for**
9: **for** $l = 1, \dots, L$ **do**
10: $\quad$ **for** $t = 1, \dots, \tau_\text{p}$ **do**
11: $\quad\quad k \leftarrow \arg\max\limits_{i\in\mathcal{K}|t_i=t} \beta_{il}$ ▷ Choose the strongest UE using pilot t
12: $\quad\quad \mathcal{D}_k \leftarrow \mathcal{D}_k \cup \{l\}$
13: $\quad$ **end for**
14: **end for**

a predefined channel estimation NMSE threshold, γ_th. This constraint is essential for mitigating excessive interference, especially in dense deployment scenarios.

Line 1 sets the scheduled UE set to empty and initializes both the number of pilot sequences and the maximum NMSE to zero. It then proceeds to schedule users iteratively in descending order of their LSF gains, terminating when either all UEs have been scheduled or the maximum NMSE exceeds the predefined threshold (line 2).

In each iteration, a candidate user k is selected (line 3), and a tentative set $\mathcal{U}'$ is formed (line 4). The algorithm evaluates all possible pilot assignments for user k, including reuse of the existing τ_p pilots and the introduction of a new pilot $t_k = \tau_\text{p} + 1$ (line 6). Assigning a new pilot increases both the number of pilots and their length (line 7).

For each candidate pilot, the algorithm computes the maximum NMSEs among the UEs in $\mathcal{U}'$, using the expression in (3.41) (lines 8-10). Subsequently, the pilot that minimizes the maximum NMSE is assigned to user k (line 12). If the resulting maximum NMSE (line 13) remains below the threshold, UE k is officially scheduled and the number of pilot sequences is updated accordingly if a new pilot has been created (lines 15-16).

In line 8, the algorithm only needs to update $\mathbf{\Psi}_{t'l}$ for $t' = t$, rather than recomputing it for all $t' \in 1, \dots, \tau'_\text{p}$, except in the case where τ'_p increases during the current

Algorithm 4.2 Deterministic-NMSE-Based Joint User Scheduling & Pilot Assignment

Require:
NMSE threshold: γ_{th}
Subarray selection sets: $\{\mathcal{D}_k\}_{k\in\mathcal{K}}$
Channel statistics: $\{\mathbf{Q}_{kl}, \mathbf{R}_{kl}\}_{k\in\mathcal{K}, l\in\mathcal{D}}$, $\{\beta_k\}_{k\in\mathcal{K}}$
Transmit powers: $\{p_k\}_{k\in\mathcal{K}}$
Ensure:
Set of scheduled users: $\mathcal{U}$
Pilot indices: $\{t_k\}_{k\in\mathcal{U}}$

```
Initialize: U ← ∅, τ_p ← 0, γ_max ← 0
while |U| < K and γ_max ≤ γ_th do
    k ← argmax_{i∈K\U} β_i                                          ▷ Select next user
    U' ← U ∪ {k}
    for t = 1 to τ_p + 1 do
        t_k ← t                                      ▷ Try a candidate pilot for the new user
        τ'_p ← max(τ_p, t_k)                        ▷ Update the number of pilot sequences
        Ψ_{t'l} ← Σ_{i∈U', t_i=t'} p_i τ'_p R_il + σ_n^2 I_M,   t' ∈ {1, ..., τ'_p},   l ∈ D
        C_il ← R_il − p_i τ'_p R_il Ψ_{t_i l}^{-1} R_il,   i ∈ U',   l ∈ D_i
        γ_max^(t) ← max_{i∈U'} (Σ_{l∈D_i} tr(C_il)) / (Σ_{l∈D_i} tr(Q_il))   ▷ Find maximum NMSE for pilot t
    end for
    t_k ← argmin_{t∈{1,...,τ_p+1}} γ_max^(t)                  ▷ Assign the best pilot to user k
    γ_max ← γ_max^(t_k)                                        ▷ Update maximum NMSE
    if γ_max ≤ γ_th then
        U ← U'                                            ▷ Update set of scheduled users
        τ_p ← max(τ_p, t_k)                          ▷ Update number of pilot sequences
    end if
end while
```

Lines 8–10 in math notation:

$$\boldsymbol{\Psi}_{t'l} \leftarrow \sum_{i\in\mathcal{U}',\, t_i=t'} p_i\tau'_{\text{p}}\mathbf{R}_{il} + \sigma_{\text{n}}^2\mathbf{I}_M, \quad t'\in\{1,\ldots,\tau'_{\text{p}}\}, \quad l\in\mathcal{D}$$

$$\mathbf{C}_{il} \leftarrow \mathbf{R}_{il} - p_i\tau'_{\text{p}}\mathbf{R}_{il}\boldsymbol{\Psi}_{t_il}^{-1}\mathbf{R}_{il}, \quad i\in\mathcal{U}', \quad l\in\mathcal{D}_i$$

$$\gamma_{\max}^{(t)} \leftarrow \max_{i\in\mathcal{U}'} \frac{\sum_{l\in\mathcal{D}_i}\operatorname{tr}(\mathbf{C}_{il})}{\sum_{l\in\mathcal{D}_i}\operatorname{tr}(\mathbf{Q}_{il})}$$

iteration, *i.e.*, when $t = \tau_{\text{p}} + 1$. Similarly, the algorithm only needs to recompute $\mathbf{C}_{il}$ and NMSE_i for the UEs assigned to pilot t, rather than for all $i \in \mathcal{U}'$, unless a new pilot has been added in this iteration. These selective updates ensure computational efficiency and are only valid for the current trial assignment. Only the updates corresponding to the pilot ultimately selected for user k will be retained and carried over to the next iteration of the while loop.

4.3 SINR-Based User Scheduling & Pilot Assignment For Centralized Operation

Algorithm 4.3 introduces a joint user scheduling and pilot assignment strategy for XL-MIMO systems operating under a centralized architecture. It is designed to maximize the number of scheduled UEs while guaranteeing that each of them achieves a minimum SE, denoted by η_{th}.

Algorithm 4.3 Deterministic-SINR-Based User Scheduling & Pilot Assignment For Centralized Operation

Require:
Switching point: U_{switch}
SE threshold: η_{th}
Subarray selection matrices: $\{\mathbf{D}_k\}_{k\in\mathcal{K}}$
Number of subarrays serving each user: $\{L_k\}_{k\in\mathcal{K}}$
Channel statistics: $\{\mathbf{Q}_k, \mathbf{R}_k, \beta_k\}_{k\in\mathcal{K}}$
Transmit powers: $\{p_k\}_{k\in\mathcal{K}}$

Ensure:
Set of scheduled users: $\mathcal{U}$
Pilot indices: $\{t_k\}_{k\in\mathcal{U}}$

Initialize: $\mathcal{U} \leftarrow \emptyset,\ \ \tau_{\mathrm{p}} \leftarrow 0,\ \ \eta_{\min} \leftarrow 2\eta_{\mathrm{th}}$.
while $|\mathcal{U}| < K$ **and** $\eta_{\min} > \eta_{\mathrm{th}}$ **do**
  $k \leftarrow \underset{i\in\mathcal{K}\setminus\mathcal{U}}{\arg\max}\, \beta_i$ ▷ Select next user
  $\mathcal{U}' \leftarrow \mathcal{U} \cup \{k\}$
  for $t = 1$ **to** $\tau_{\mathrm{p}} + 1$ **do**
    $t_k \leftarrow t$ ▷ Try a candidate pilot for the new user
    $\tau'_{\mathrm{p}} \leftarrow \max(\tau_{\mathrm{p}},\, t_k)$ ▷ Update the number of pilot sequences
    $\boldsymbol{\Psi}_{t'} \leftarrow \sum_{i\in\mathcal{U}',\, t_i = t'} p_i \tau'_{\mathrm{p}} \mathbf{R}_i + \sigma_{\mathrm{n}}^2 \mathbf{I}_{ML}, \quad t' \in \{1, \ldots, \tau'_{\mathrm{p}}\}$
    $\mathbf{C}_i \leftarrow \mathbf{R}_i - p_i \tau'_{\mathrm{p}} \mathbf{R}_i \boldsymbol{\Psi}_{t_i}^{-1} \mathbf{R}_i, \quad i \in \mathcal{U}'$
    if $|\mathcal{U}'| > U_{\mathrm{switch}}$ **then**: for $i \in \mathcal{U}'$: ▷ Use ergodic SINR expression
$$\Gamma_i \leftarrow p_i \operatorname{tr}\left\{ \mathbf{D}_i \left[\mathbf{D}_i \left(\sum_{\substack{i'\in\mathcal{U}' \\ i'\neq i}} p_{i'} \mathbf{Q}_{i'} + p_i \mathbf{C}_i \right) \mathbf{D}_i + \sigma_{\mathrm{n}}^2 \mathbf{I}_{ML} \right]^{-1} \mathbf{D}_i \left(\mathbf{Q}_i - \mathbf{C}_i \right) \mathbf{D}_i \right\}$$
    else: for $i \in \mathcal{U}'$: ▷ Use asymptotic SINR expression
$$\Gamma_i \leftarrow p_i \frac{\operatorname{tr}[\mathbf{D}_i(\mathbf{Q}_i - \mathbf{C}_i)\mathbf{D}_i] - \sum_{\substack{i'\in\mathcal{U}' \\ i'\neq i}} \frac{\operatorname{tr}[\mathbf{D}_i(\mathbf{Q}_i - \mathbf{C}_i)\mathbf{D}_i(\mathbf{Q}_{i'} - \mathbf{C}_{i'})\mathbf{D}_i]}{\operatorname{tr}[\mathbf{D}_i(\mathbf{Q}_{i'} - \mathbf{C}_{i'})\mathbf{D}_i]}}{\sum_{i'\in\mathcal{U}'} \frac{p_{i'}}{ML_i} \operatorname{tr}(\mathbf{D}_i \mathbf{C}_{i'} \mathbf{D}_i) + \sigma_{\mathrm{n}}^2}$$
    end if
    $\eta_{\min}^{(t)} \leftarrow \min_{i\in\mathcal{U}'} \left(1 - \frac{\tau'_{\mathrm{p}}}{\tau_{\mathrm{c}}}\right) \log_2(1 + \Gamma_i)$ ▷ Find minimum SE for pilot t
  end for
  $t_k \leftarrow \underset{t\in\{1,\ldots,\tau_{\mathrm{p}}+1\}}{\arg\max}\ \eta_{\min}^{(t)}$ ▷ Assign the best pilot to user k
  $\eta_{\min} \leftarrow \eta_{\min}^{(t_k)}$ ▷ Update minimum SE
  if $\eta_{\min} \geq \eta_{\mathrm{th}}$ **then**
    $\mathcal{U} \leftarrow \mathcal{U}'$ ▷ Update set of scheduled users
    $\tau_{\mathrm{p}} \leftarrow \max(\tau_{\mathrm{p}}, t_k)$ ▷ Update number of pilot sequences
  end if
end while

Initially (line 1), the algorithm sets the scheduled UE set to empty, initializes the number of pilot sequences to zero, and sets the minimum per-user SE to a value above the threshold (e.g., twice $\mathrm{SE}_{\mathrm{th}}$). Then, it iteratively schedules UEs in descending order of their LSF channel gains, terminating when either all UEs are scheduled or the minimum per-user SE falls below the threshold (line 2).

In each iteration, a candidate user k is selected (line 3), and a tentative set $\mathcal{U}'$ is formed (line 4). The algorithm evaluates all possible pilot assignments for user k, including reuse of the existing τ_{p} pilots and the introduction of a new pilot $t_k = \tau_{\mathrm{p}} + 1$ (line 6).

For each candidate pilot, the algorithm evaluates the resulting SINRs for all UEs in $\mathcal{U}'$ by employing either the ergodic expression in (3.61), if $|\mathcal{U}'| > U_{\mathrm{switch}}$, or the asymptotic expression in (3.69), otherwise (lines 8–14). This switching rule enhances the accuracy of the SINR estimates by selecting the analytical approximation best suited for each loading regime. The parameter U_{switch} defines the user-loading threshold above which the ergodic approximation is preferred, ensuring that the most reliable expression is utilized based on the numerical characterization provided in Section 3.5.

After evaluating all candidates, the pilot that maximizes the minimum per-user SE is selected for UE k (lines 15–17). If the resulting minimum SE (line 18) satisfies the predefined threshold (line 19), UE k is added to the set of scheduled UEs (line 20), and the number of pilots is incremented if a new sequence was utilized (line 21).

4.4 SINR-Based User Scheduling & Pilot Assignment For Distributed Operation

Following the logic of Algorithm 4.3, Algorithm 4.4 introduces joint scheduling and pilot assignment for distributed operation, ensuring each UE meets the minimum SE requirement, η_{th}. Additionally, the algorithm computes the weights used for signal estimation (lines 15 and 23), which are necessary to combine the local estimates generated by each serving subarray.

When trying to schedule a new user k (line 3), the algorithm forms a candidate user set $\mathcal{U}'$ and evaluates all pilot options for user k. For each candidate pilot, the local SINRs at the selected subarrays are estimated for all UEs in $\mathcal{U}'$, using either the ergodic expression from (3.61), if $|\mathcal{U}'| > U_{\mathrm{switch}}$, or the asymptotic expression from (3.69), otherwise (lines 8-14). Line 15 then computes the optimal weights for each selected subarray, and line 16 computes the resulting minimum per-user SE for the current candidate pilot. Recall that when optimal weighting is employed, the global SINR can be estimated by summing the local SINRs at the selected subarrays, following the expression in (3.104). The pilot sequence that yields the highest minimum SE is selected (line 18), and if the corresponding SE is above the threshold, the UE is scheduled, the number of pilot sequences is updated

Algorithm 4.4 Deterministic-SINR-Based User Scheduling & Pilot Assignment For Distributed Operation

Require:
Switching point: U_{switch}
SE threshold: η_{th}
Subarray selection sets: $\{\mathcal{D}_k\}_{k\in\mathcal{K}}$
Channel statistics: $\{\mathbf{Q}_{kl}, \mathbf{R}_{kl}\}_{k\in\mathcal{K},\, l\in\mathcal{D}}$, $\{\beta_k\}_{k\in\mathcal{K}}$
Transmit powers: $\{p_k\}_{k\in\mathcal{K}}$

Ensure:
Set of scheduled users: $\mathcal{U}$
Pilot indices: $\{t_k\}_{k\in\mathcal{U}}$
Weights: $\{\mu_{kl}\}_{k\in\mathcal{U},\, l\in\mathcal{D}_k}$

Initialize: $\mathcal{U} \leftarrow \emptyset,\ \tau_{\text{p}} \leftarrow 0,\ \eta_{\min} \leftarrow 2\eta_{\text{th}}$
while $|\mathcal{U}| < K$ **and** $\eta_{\min} > \eta_{\text{th}}$ **do**
  $k \leftarrow \arg\max_{i\in\mathcal{K}\setminus\mathcal{U}} \beta_i$ ▷ Select next user
  $\mathcal{U}' \leftarrow \mathcal{U} \cup \{k\}$
  for $t = 1$ **to** $\tau_{\text{p}} + 1$ **do**
    $t_k \leftarrow t$ ▷ Try a candidate pilot for the new user
    $\tau'_{\text{p}} \leftarrow \max(\tau_{\text{p}},\, t_k)$ ▷ Update the number of pilot sequences
    $$\boldsymbol{\Psi}_{t'l} \leftarrow \sum_{i\in\mathcal{U}',\, t_i = t'} p_i \tau'_{\text{p}} \mathbf{R}_{il} + \sigma_{\text{n}}^2 \mathbf{I}_M, \quad t' \in \{1,\ldots,\tau'_{\text{p}}\}, \quad l \in \mathcal{D}$$
    $$\mathbf{C}_{il} \leftarrow \mathbf{R}_{il} - p_i \tau'_{\text{p}} \mathbf{R}_{il} \boldsymbol{\Psi}_{t_i l}^{-1} \mathbf{R}_{il}, \quad i \in \mathcal{U}', \quad l \in \mathcal{D}_i$$
    if $|\mathcal{U}'| > U_{\text{switch}}$ **then**: for $i \in \mathcal{U}'$ and $l \in \mathcal{D}_i$: ▷ Use ergodic SINR expression
      $$\Gamma_{il} \leftarrow p_i \operatorname{tr}\left[\left(\sum_{i'\in\mathcal{U}',\, i'\neq i} p_{i'} \mathbf{Q}_{i'l} + p_i \mathbf{C}_{il} + \sigma_{\text{n}}^2 \mathbf{I}_M\right)^{-1} (\mathbf{Q}_{il} - \mathbf{C}_{il})\right]$$
    else: for $i \in \mathcal{U}'$ and $l \in \mathcal{D}_i$: ▷ Use asymptotic SINR expression
      $$\Gamma_{il} \leftarrow p_i \frac{\operatorname{tr}(\mathbf{Q}_{il} - \mathbf{C}_{il}) - \displaystyle\sum_{i'\in\mathcal{U}',\, i'\neq i} \frac{\operatorname{tr}[(\mathbf{Q}_{il} - \mathbf{C}_{il})(\mathbf{Q}_{i'l} - \mathbf{C}_{i'l})]}{\operatorname{tr}(\mathbf{Q}_{i'l} - \mathbf{C}_{i'l})}}{\displaystyle\sum_{i'\in\mathcal{U}'} \frac{p_{i'}}{M} \operatorname{tr}(\mathbf{C}_{i'l}) + \sigma_{\text{n}}^2}$$
    end if
    $$\mu_{il}^{(t)} \leftarrow \Gamma_{il} \Big/ \sum_{l'\in\mathcal{D}_i} \Gamma_{il'}, \quad i \in \mathcal{U}', \quad l \in \mathcal{D}_i$$ ▷ Compute weights
    $$\eta_{\min}^{(t)} \leftarrow \min_{i\in\mathcal{U}'} \left(1 - \frac{\tau'_{\text{p}}}{\tau_{\text{c}}}\right) \log_2\left(1 + \sum_{l\in\mathcal{D}_i} \Gamma_{il}\right)$$ ▷ Find minimum SE for pilot t
  end for
  $t_k \leftarrow \arg\max_{t\in\{1,\ldots,\tau_{\text{p}}+1\}} \eta_{\min}^{(t)}$ ▷ Assign the best pilot to user k
  $\eta_{\min} \leftarrow \eta_{\min}^{(t_k)}$ ▷ Update minimum SE
  if $\eta_{\min} \geq \eta_{\text{th}}$ **then**
    $\mathcal{U} \leftarrow \mathcal{U}'$ ▷ Update set of scheduled users
    $\tau_{\text{p}} \leftarrow \max(\tau_{\text{p}}, t_k)$ ▷ Update number of pilot sequences
    $\mu_{il} \leftarrow \mu_{il}^{(t_k)}, \quad i \in \mathcal{U}, \quad l \in \mathcal{D}_i$ ▷ Update weights
  end if
end while

if a new pilot has been created, and the combining weights are stored (lines 21–23).

4.5 Numerical Results

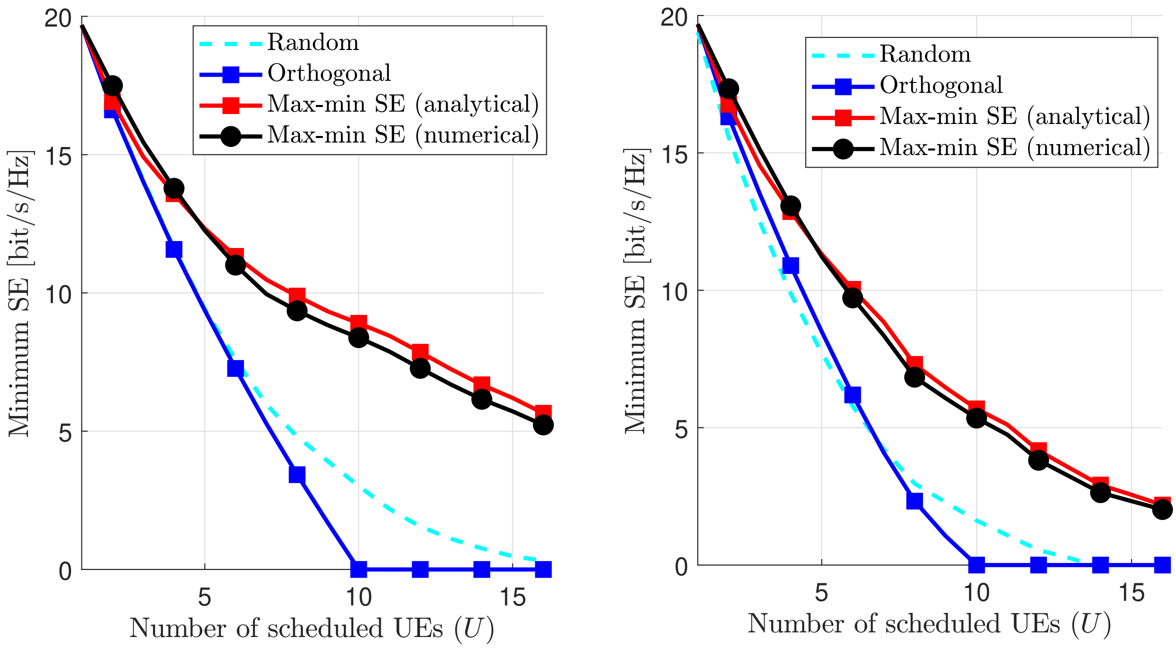
This section presents numerical results that validate the proposed strategies for joint subarray selection, user scheduling, and pilot assignment in XL-MIMO systems operating under either centralized or distributed architectures. The objective is to assess the effectiveness of the proposed algorithms in maximizing the minimum per-user SE and to demonstrate that approaches based on deterministic SINR expressions achieve performance comparable to those relying on instantaneous SINR, while incurring significantly lower computational complexity. The simulation setup adopts $L = 16$, $L_k = 4$, $M = 16$, and the parameters established in Section 3.4.

Given that LSF-based subarray selection achieves performance comparable to the SINR-based strategy, as demonstrated in Section 3.4, it is adopted hereafter due to its lower computational complexity. This approach eliminates the need to estimate the local SINRs of all K UEs across all L subarrays. For distributed operation, optimal weighting is employed, as it remains computationally efficient by requiring local SINR estimation only at the selected subarrays.

Figure 4.1 evaluates the system performance in terms of minimum per-user SE as a function of the number of scheduled users (U). Algorithms 4.3 and 4.4, labeled *Max-Min SE (Ana)*, are compared against a numerical benchmark denoted as *Max-Min SE (Num)*. This benchmark follows the same logic as the proposed algorithms but relies on numerically averaged SINR values for pilot assignment instead of the derived deterministic approximations. Two baseline schemes are included for reference: Algorithm 4.2, labeled *Orthogonal* since it assigns mutually orthogonal pilots by minimizing the maximum channel estimation NMSE, and a *Random* pilot assignment.

All schemes perform user scheduling based on descending LSF gains and employ LSF-based subarray selection. The scenario considers $K = 16$ UEs and a short coherence block ($\tau_c = 10$), rendering pilot contamination unavoidable. Results are shown up to the full load $U = K$ for illustrative purposes. In a practical implementation, the algorithm would terminate once the SE threshold is violated, potentially leaving some UEs unscheduled.

Figures 4.1a and 4.1b, corresponding to centralized and distributed architectures, respectively, show that the centralized approach consistently achieves a higher minimum SE. This is expected, as the CPU in a centralized architecture jointly processes signals from all L_k selected subarrays, enabling global interference suppression. In contrast, the distributed architecture relies on local combining, limiting its ability to mitigate inter-user interference. Notably, if each UE were served by a single subarray ($L_k = 1$), both



(a) Centralized Operation.

(b) Distributed Operation.

Figure 4.1 – Minimum per-user SE versus number of scheduled UEs.

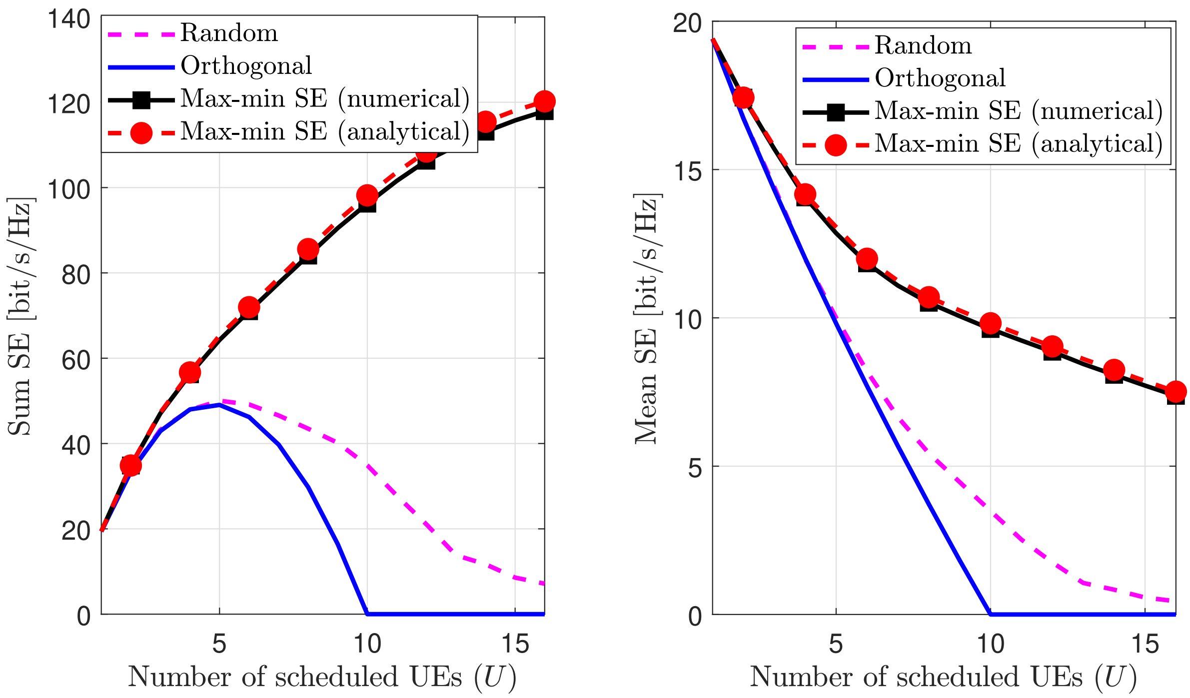
architectures would converge to identical performance.

In both architectures, the *Max-Min SE (Num)* strategy effectively mitigates interference by exploiting the true mean SINR; however, it entails prohibitive computational costs due to its reliance on extensive instantaneous CSI processing. In contrast, the *Max-Min SE (Ana)* strategy achieves nearly identical performance using only deterministic SINR approximations, thereby providing a tractable alternative. This result is particularly significant given that the analytical approach relies solely on statistical CSI. It effectively captures the dominant structure of pilot contamination—namely, the spatial separation of users sharing the same pilots—enabling near-optimal interference mitigation without requiring instantaneous CSI or computationally intensive search procedures. The negligible performance gap indicates that the dominant gains in pilot assignment stem from resolving large-scale spatial conflicts, which are efficiently captured by the derived deterministic expressions without the need for Monte Carlo averaging.

Furthermore, the proposed strategy rivals the numerical benchmark in maximizing the number of scheduled UEs that satisfy a per-user SE requirement. For instance, given a 5 bit/s/Hz threshold in Fig. 4.1a, the *Random* and *Orthogonal* baselines schedule only 7 UEs, whereas both *max-min* algorithms successfully schedule all 16 UEs. These results highlight the effectiveness of deterministic SINR-based resource allocation in enhancing fairness within crowded XL-MIMO networks.

While the intelligent strategies maintain robust performance as U increases, the *Random* and *Orthogonal* assignments suffer severe degradation due to uncoordinated pilot reuse and reduced data transmission intervals (caused by increasing pilot sequence lengths), respectively. Specifically, the *Orthogonal* assignment strategy incurs zero SE for all scheduled UEs when $U \geq \tau_c$, as no time remains for data transmission within the coherence block.

Figures 4.2a and 4.2b, which depict centralized operation, demonstrate that the proposed algorithm also exhibits sum SE and mean per-user SE remarkably close to the numerical benchmark. A cross-comparison between Figures 4.2b and 4.1a reveals that the minimum SE is not significantly lower than the mean SE. For instance, when $U = 16$ UEs are scheduled, the proposed *Max-Min SE (Ana)* strategy achieves a mean SE of approximately 7.5 bit/s/Hz and minimum SE of around 5.8 bit/s/Hz. This narrow gap indicates that the proposed resource allocation achieves fairness, characterized by low variations in user rates. Additionally, the sum SE scales with the number of scheduled UEs, a trend not observed in the *Random* assignment (due to severe pilot contamination) nor in the *Orthogonal* assignment (where increasing pilot overhead reduces the effective data transmission time).



(a) Sum SE vs. U

(b) Mean SE vs. U

Figure 4.2 – Joint subarray selection, user scheduling, and pilot assignment algorithms for centralized uplink operation.

Finally, Figure 4.3 illustrates the minimum per-user SE as a function of the coherence block length for $K = 6$ UEs. A *Max-Min SE (Num - Exhaustive Search)* benchmark is introduced to establish the theoretical performance upper bound. Unlike the sequential approach of the *Max-Min SE (Num)* strategy—where pilots are assigned one-by-one to newly scheduled UEs without re-evaluating previous assignments—the exhaustive search evaluates all possible pilot combinations to identify the global optimum. Notably, the sequential mechanism employed by the proposed Algorithms 4.3 and 4.4 proves highly effective, achieving performance nearly identical to the optimal exhaustive search.

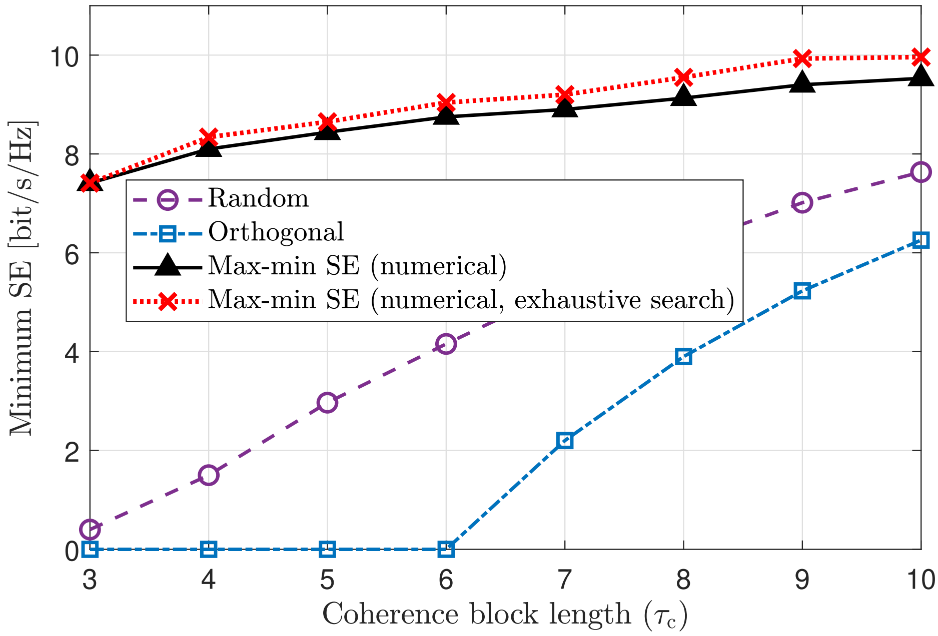


Figure 4.3 – Minimum per-user SE versus coherence block length.

Moreover, observations from Figure 4.1 suggest that the *Max-Min SE (Ana)* approach may approximate the optimal pilot assignment even more closely than its numerical counterpart. Notably, the analytical strategy occasionally outperforms the *Max-Min SE (Num)* benchmark. This behavior is likely attributed to the deterministic expressions effectively capturing long-term spatial conflicts, free from the estimation variance inherent in computing the mean SINR from a limited set of instantaneous samples.

4.6 Summary

This chapter addressed the critical challenges of scalability and interference management in XL-MIMO systems by developing joint strategies for subarray selection, user scheduling, and pilot assignment. Recognizing that resource allocation based on instantaneous CSI is computationally prohibitive in systems with massive antenna arrays, the proposed frameworks relied exclusively on long-term statistical CSI and the closed-form deterministic SINR approximations derived in Chapter 3.

The core contribution of this chapter focused on SINR-based algorithms designed to maximize the minimum per-user SE for both centralized and distributed architectures. These algorithms leverage a novel switching mechanism that selects between ergodic and asymptotic SINR expressions based on the system load, ensuring high accuracy across different operating regimes.

The numerical results validated the efficacy of the proposed strategies, leading to several key conclusions that highlight the practical relevance of the proposed resource allocation framework for future XL-MIMO networks:

- **Accuracy and tractability:** The *Max-Min SE (Ana)* strategy, which utilize deterministic SINR approximations, achieve performance nearly identical to benchmarks that rely on true mean SINRs computed via extensive Monte Carlo simulations. This confirms that the derived expressions effectively capture the system performance with significantly lower computational complexity.
- **User coverage:** The proposed methods successfully schedule the entire user population in scenarios where baseline strategies (random or orthogonal) fail to serve a large portion of users due to interference or pilot overhead limitations.
- **Near-optimality:** Comparisons with an exhaustive search benchmark revealed that the proposed sequential algorithms achieve near-optimal performance, maximizing the number of scheduled users while ensuring fairness.
- **Robustness in high-mobility scenarios:** Unlike orthogonal assignment strategies, which are fundamentally limited by the coherence block length, and random assignment, which suffers from uncontrolled interference, the proposed joint design maintains robust performance even in overloaded scenarios where pilot reuse is unavoidable.

Chapter 5 summarizes the main findings of the thesis and outlines promising directions for future research.

5 CONCLUSIONS AND RESEARCH PERSPECTIVES

This thesis investigates the modeling, analysis, and optimization of XL-MIMO systems, focusing on developing deterministic SINR expressions and efficient resource allocation strategies. XL-MIMO is a key technology for meeting the stringent requirements of future wireless networks, particularly in terms of throughput, reliability, and improved connectivity in densely populated urban environments.

The first major contribution was the development of a realistic channel model tailored to modular XL-MIMO architectures. This model incorporates hybrid propagation characteristics, including distance-dependent LoS components and spatially correlated NLoS components, thereby representing correlated Rician fading channels with spatial non-stationarities and VRs, allowing for practical and accurate simulation and analysis of XL-MIMO deployments in micro-urban scenarios.

Using this model as a foundation, closed-form deterministic expressions for the uplink SINR were derived under both centralized and distributed processing architectures. These expressions assume imperfect CSI, acquired via MMSE channel estimation, and MMSE receive combining. Unlike existing works, which usually consider uncorrelated Rayleigh fading and MR combining, the expressions presented here are novel in that they support spatially correlated Rician fading and MMSE combining, in the XL-MIMO context. Importantly, the proposed deterministic expressions depend only on channel statistics, enabling resource allocation and scheduling decisions that are both tractable and robust across multiple channel coherence blocks. Numerical simulations demonstrated the high accuracy of the deterministic SINR expressions.

Based on the derived SINR approximations, the thesis proposes efficient algorithms for joint user scheduling, pilot assignment, and subarray selection. These algorithms aim to optimize either the sum SE or the minimum per-user SE, with a particular focus on enhancing fairness in crowded scenarios with pilot reuse. The use of statistical CSI significantly reduces computational complexity and fronthaul signaling, making the proposed strategies well-suited to practical large-scale deployments. Simulation results confirmed the efficacy of the proposed algorithms in achieving fairness and throughput gains under realistic constraints.

5.1 Research Perspectives

The work presented in this thesis opens several promising directions for future research:

1. **Extension to multi-antenna UEs:** the current analysis assumes single-antenna UEs. Extending the models and algorithms to support multi-antenna UEs would allow a richer spatial multiplexing structure and enhanced diversity gains.

2. **Downlink analysis:** while this thesis is focused on the uplink, deriving deterministic SINR expressions for the downlink would complement the presented framework and enable full-duplex optimization.

3. **Machine learning-based scheduling and assignment:** statistical CSI-driven algorithms are scalable but may still require tuning or reconfiguration in dynamic scenarios. Machine learning models, particularly those based on graph learning or reinforcement learning, can be trained to generalize across environments and automate decision-making.

4. **Joint power control and SINR optimization:** integrating power allocation into the proposed deterministic SINR framework could further improve system performance, especially in energy-constrained or heterogeneous networks.

5. **Hardware impairments and practical constraints:** realistic deployment studies should consider non-idealities such as quantization noise, hardware nonlinearities, and limited fronthaul capacity. Incorporating these into the analysis would increase the practical relevance of the results.

6. **Integration with cell-free and ISAC architectures:** the proposed methods are readily extensible to cell-free massive MIMO and ISAC systems, which share architectural similarities with distributed XL-MIMO.

BIBLIOGRAPHY

[1] WANG, Z. et al. A Tutorial on Extremely Large-Scale MIMO for 6G: Fundamentals, Signal Processing, and Applications. *IEEE Communications Surveys & Tutorials*, v. 26, n. 3, p. 1560–1605, 2024.

[2] ANDREWS, J. G. et al. What Will 5G Be? *IEEE Journal on Selected Areas in Communications*, v. 32, n. 6, p. 1065–1082, 2014.

[3] SHAFI, M. et al. 5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice. *IEEE Journal on Selected Areas in Communications*, v. 35, n. 6, p. 1201–1221, 2017.

[4] AGIWAL, M.; ROY, A.; SAXENA, N. Next Generation 5G Wireless Networks: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials*, v. 18, n. 3, p. 1617–1655, 2016.

[5] MARZETTA, T. L. Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Transactions on Wireless Communications*, v. 9, n. 11, p. 3590–3600, 2010.

[6] NGO, H. Q.; LARSSON, E. G.; MARZETTA, T. L. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems. *IEEE Transactions on Communications*, v. 61, n. 4, p. 1436–1449, 2013.

[7] LARSSON, E. G. et al. Massive MIMO for Next Generation Wireless Systems. *IEEE Communications Magazine*, v. 52, n. 2, p. 186–195, 2014.

[8] 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. ISSN 1932-8346.

[9] DEMIR, U. T.; BJÖRNSON, E.; SANGUINETTI, L. Foundations of User-Centric Cell-Free Massive MIMO. *Foundations and Trends® in Signal Processing*, v. 14, n. 3-4, p. 162–472, 2021. ISSN 1932-8346.

[10] TARIQ, F. et al. A Speculative Study on 6G. *IEEE Wireless Communications*, v. 27, n. 4, p. 118–125, 2020.

[11] SAAD, W.; BENNIS, M.; CHEN, M. A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems. *IEEE Network*, v. 34, n. 3, p. 134–142, 2020.

[12] GIORDANI, M. et al. Toward 6G Networks: Use Cases and Technologies. *IEEE Communications Magazine*, v. 58, n. 3, p. 55–61, 2020.

[13] LU, H. et al. A Tutorial on Near-Field XL-MIMO Communications Toward 6G. *IEEE Communications Surveys & Tutorials*, v. 26, n. 4, p. 2213–2257, 2024.

[14] WANG, Z. et al. Extremely Large-Scale MIMO: Fundamentals, Challenges, Solutions, and Future Directions. *IEEE Wireless Communications*, v. 31, n. 3, p. 117–124, 2024.

[15] BJÖRNSON, E. et al. Massive MIMO Is A Reality—What Is Next?: Five Promising Research Directions for Antenna Arrays. *Digital Signal Processing*, Elsevier, v. 94, p. 3–20, 2019.

[16] RONG, B. 6G: The Next Horizon: From Connected People and Things to Connected Intelligence. *IEEE Wireless Communications*, v. 28, n. 5, p. 8–8, 2021.

[17] WANG, C.-X. et al. On the Road to 6G: Visions, Requirements, Key Technologies, and Testbeds. *IEEE Communications Surveys & Tutorials*, v. 25, n. 2, p. 905–974, 2023.

[18] AMIRI, A. et al. Extremely Large Aperture Massive MIMO: Low Complexity Receiver Architectures. In: *2018 IEEE Globecom Workshops (GC Wkshps)*. Abu Dhabi, UAE: [s.n.], 2018. p. 1–6.

[19] AMIRI, A.; MANCHóN, C. N.; CARVALHO, E. de. A Message Passing Based Receiver for Extra-Large Scale MIMO. In: *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. Le Gosier, Guadeloupe: IEEE, 2019. p. 564–568.

[20] ZHANG, J. et al. Local Partial Zero-Forcing Combining for Cell-Free Massive MIMO Systems. *IEEE Transactions on Communications*, v. 69, n. 12, p. 8459–8473, 2021.

[21] HAN, Y. et al. Channel Estimation for Extremely Large-Scale Massive MIMO Systems. *IEEE Wireless Communications Letters*, v. 9, n. 5, p. 633–637, 2020.

[22] CUI, M. et al. Near-Field MIMO Communications for 6G: Fundamentals, Challenges, Potentials, and Future Directions. *IEEE Communications Magazine*, v. 61, n. 1, p. 40–46, 2023.

[23] LU, H.; ZENG, Y. Communicating With Extremely Large-Scale Array/Surface: Unified Modeling and Performance Analysis. *IEEE Transactions on Wireless Communications*, v. 21, n. 6, p. 4039–4053, 2022.

[24] ZHANG, H. et al. 6G Wireless Communications: From Far-Field Beam Steering to Near-Field Beam Focusing. *IEEE Communications Magazine*, v. 61, n. 4, p. 72–77, 2023.

[25] LU, Y.; DAI, L. Near-Field Channel Estimation in Mixed LoS/NLoS Environments for Extremely Large-Scale MIMO Systems. *IEEE Transactions on Communications*, v. 71, n. 6, p. 3694–3707, 2023.

[26] SHI, X. et al. Sparse Estimation for XL-MIMO with Unified LoS/NLoS Representation. In: *ICC 2024 - IEEE International Conference on Communications*. Denver, CO, USA: IEEE, 2024. p. 5027–5032.

[27] CARVALHO, E. D. et al. Non-Stationarities in Extra-Large-Scale Massive MIMO. *IEEE Wireless Communications*, v. 27, n. 4, p. 74–80, 2020.

[28] IIMORI, H. et al. Joint Activity and Channel Estimation for Extra-Large MIMO Systems. *IEEE Transactions on Wireless Communications*, v. 21, n. 9, p. 7253–7270, 2022.

[29] AMIRI, A. et al. Distributed Receiver Processing for Extra-Large MIMO Arrays: A Message Passing Approach. *IEEE Transactions on Wireless Communications*, v. 21, n. 4, p. 2654–2667, 2022.

[30] LIU, J. et al. A Non-Stationary Channel Model with Correlated NLoS/LoS States for ELAA-mMIMO. In: *2021 IEEE Global Communications Conference (GLOBECOM)*. Madrid, Spain: IEEE, 2021. p. 1–6.

[31] GUERRA, D. W. M.; ABRÃO, T. Clustered Double-Scattering Channel Modeling for XL-MIMO With Uniform Arrays. *IEEE Access*, v. 10, p. 20173–20186, 2022.

[32] LIU, L. et al. The COST 2100 MIMO channel model. *IEEE Wireless Communications*, v. 19, n. 6, p. 92–99, 2012.

[33] YUAN, Z. et al. Spatial Non-Stationary Near-Field Channel Modeling and Validation for Massive MIMO Systems. *IEEE Transactions on Antennas and Propagation*, v. 71, n. 1, p. 921–933, 2023.

[34] WANG, Z. et al. Uplink Performance of Cell-Free Massive MIMO Over Spatially Correlated Rician Fading Channels. *IEEE Communications Letters*, v. 25, n. 4, p. 1348–1352, 2021.

[35] BJÖRNSON, E.; SANGUINETTI, L. Making Cell-Free Massive MIMO Competitive With MMSE Processing and Centralized Implementation. *IEEE Transactions on Wireless Communications*, v. 19, n. 1, p. 77–90, 2020.

[36] Marinello, J. C. et al. Antenna Selection for Improving Energy Efficiency in XL-MIMO Systems. *IEEE Transactions on Vehicular Technology*, v. 69, n. 11, p. 13305–13318, 2020.

[37] LIU, D. et al. Location-Based Visible Region Recognition in Extra-Large Massive MIMO Systems. *IEEE Transactions on Vehicular Technology*, v. 72, n. 6, p. 8186–8191, 2023.

[38] SOUZA, J. H. I. de et al. QoS-Aware User Scheduling in Crowded XL-MIMO Systems Under Non-Stationary Multi-State LoS/NLoS Channels. *IEEE Transactions on Vehicular Technology*, v. 72, n. 6, p. 7639–7652, 2023.

[39] MASHDOUR, S. et al. Robust Resource Allocation in Cell-Free Massive MIMO Systems. *IEEE Transactions on Communications*, v. 73, n. 8, p. 5745–5759, 2025.

[40] BUZZI, S. et al. Pilot Assignment in Cell-Free Massive MIMO Based on the Hungarian Algorithm. *IEEE Wireless Communications Letters*, v. 10, n. 1, p. 34–37, 2021.

[41] YANG, X. et al. On the Uplink Transmission of Extra-Large Scale Massive MIMO Systems. *IEEE Transactions on Vehicular Technology*, v. 69, n. 12, p. 15229–15243, 2020.

[42] ZHI, K. et al. Ergodic Rate Analysis of Reconfigurable Intelligent Surface-Aided Massive MIMO Systems With ZF Detectors. *IEEE Communications Letters*, v. 26, n. 2, p. 264–268, 2022.

[43] D'ANDREA, C. et al. Analysis of UAV Communications in Cell-Free Massive MIMO Systems. *IEEE Open Journal of the Communications Society*, v. 1, p. 133–147, 2020.

[44] LEI, H. et al. Uplink Performance of Cell-Free Extremely Large-Scale MIMO Systems. *IEEE Transactions on Vehicular Technology*, v. 74, n. 5, p. 8339–8344, 2025.

[45] ZHANG, Y. et al. Enhancing Uplink Performance for Cell-Free Massive MIMO With Low-Resolution ADCs by RSMA. *IEEE Journal on Selected Areas in Communications*, v. 43, n. 3, p. 720–735, 2025.

[46] HOSANY, M. A. On the Performance Analysis of Multi-User Massive MIMO Systems with Error Vector Signals for 5G Cellular Networks. *Wireless Personal Communications*, Springer, v. 123, n. 2, p. 917–934, 2022. ISSN 1572-834X.

[47] LI, J. et al. Low Altitude 3-D Coverage Performance Analysis of Cell-Free RAN for 6G Systems. *IEEE Transactions on Vehicular Technology*, v. 72, n. 12, p. 16163–16176, 2023.

[48] JIN, S.-N.; YUE, D.-W.; NGUYEN, H. H. Spectral and Energy Efficiency in Cell-Free Massive MIMO Systems Over Correlated Rician Fading. *IEEE Systems Journal*, v. 15, n. 2, p. 2822–2833, 2021.

[49] FAN, Y. et al. Power Allocation for Cell-Free Massive MIMO ISAC Systems With OTFS Signal. *IEEE Internet of Things Journal*, v. 12, n. 7, p. 9314–9331, 2025.

[50] ALI, A.; CARVALHO, E. D.; HEATH, R. W. Linear Receivers in Non-Stationary Massive MIMO Channels With Visibility Regions. *IEEE Wireless Communications Letters*, v. 8, n. 3, p. 885–888, 2019.

[51] CHEN, S. et al. Structured Massive Access for Scalable Cell-Free Massive MIMO Systems. *IEEE Journal on Selected Areas in Communications*, v. 39, n. 4, p. 1086–1100, 2021.

[52] QIU, J. et al. Downlink Power Optimization for Cell-Free Massive MIMO Over Spatially Correlated Rayleigh Fading Channels. *IEEE Access*, v. 8, p. 56214–56227, 2020.

[53] NGUYEN, T. H. et al. Pilot Assignment for Joint Uplink-Downlink Spectral Efficiency Enhancement in Massive MIMO Systems With Spatial Correlation. *IEEE Transactions on Vehicular Technology*, v. 70, n. 8, p. 8292–8297, 2021.

[54] ZHOU, M. et al. Spatially Correlated Rayleigh Fading for Cell-Free Massive MIMO Systems. *IEEE Access*, v. 8, p. 42154–42168, 2020.

[55] SUN, Q. et al. Uplink Performance of Hardware-Impaired Cell-Free Massive MIMO With Multi-Antenna Users and Superimposed Pilots. *IEEE Transactions on Communications*, v. 71, n. 11, p. 6711–6726, 2023.

[56] LI, N.; FAN, P. Joint Data Power Control and LSFD Design in Distributed Cell-Free Massive MIMO under Non-Ideal UE Hardware. *EURASIP Journal on Wireless Communications and Networking*, SpringerOpen, v. 2024, n. 9, p. 1–17, 2024.

[57] WAGNER, S. et al. Large System Analysis of Linear Precoding in Correlated MISO broadcast Channels Under Limited Feedback. *IEEE Transactions on Information Theory*, v. 58, n. 7, p. 4509–4537, 2012.

[58] HARITHA, H. et al. Superimposed Versus Regular Pilots for Hardware Impaired Rician-Faded Cell-Free Massive MIMO Systems. *IEEE Transactions on Communications*, v. 72, n. 11, p. 6688–6706, 2024.

[59] TENTU, V. et al. UAV-Enabled Hardware-Impaired Spatially Correlated Cell-Free Massive MIMO Systems: Analysis and Energy Efficiency Optimization. *IEEE Transactions on Communications*, v. 70, n. 4, p. 2722–2741, 2022.

[60] MA, X. et al. Scalable Cell-Free Massive MIMO Systems With Finite Resolution ADCs/DACs Over Spatially Correlated Rician Fading Channels. *IEEE Transactions on Vehicular Technology*, v. 72, n. 6, p. 7699–7716, 2023.

[61] KAUR, H.; KANSAL, A. Closed-form Analysis of RZF in Multicell Massive MIMO Over Correlated Rician Channel. *Wireless Personal Communications*, Springer, v. 131, n. 3, p. 1685–1719, 2023. ISSN 1572-834X.

[62] LIU, Y. et al. Near-Field Communications: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials*, v. 27, n. 3, p. 1687–1728, 2025.

[63] SANGUINETTI, L.; BJÖRNSON, E.; HOYDIS, J. Toward Massive MIMO 2.0: Understanding Spatial Correlation, Interference Suppression, and Pilot Contamination. *IEEE Transactions on Communications*, v. 68, n. 1, p. 232–257, 2020.

[64] MUKHERJEE, S.; CHOPRA, R. Performance Analysis of Cell-Free Massive MIMO Systems in LoS/ NLoS Channels. *IEEE Transactions on Vehicular Technology*, v. 71, n. 6, p. 6410–6423, 2022.

[65] BJÖRNSON, E.; DEMIR, O. T. *Introduction to Multiple Antenna Communications and Reconfigurable Surfaces*. Norwell, MA: Now Publishers, 2024. 688 p. ISBN 978-1-63828-314-0.

[66] 3rd Generation Partnership Project (3GPP). *Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects*. Sophia Antipolis, France, 2017. (Release 9, TR 36.814 v9.2.0).

[67] BJÖRNSON, E. et al. *Towards 6G MIMO: Massive Spatial Multiplexing, Dense Arrays, and Interplay Between Electromagnetics and Processing*. 2024.

[68] DONG, Z.; LI, X.; ZENG, Y. Characterizing and Utilizing Near-Field Spatial Correlation for XL-MIMO Communication. *IEEE Transactions on Communications*, v. 72, n. 12, p. 7922–7937, 2024.

[69] OZDOGAN, O.; BJÖRNSON, E.; LARSSON, E. G. Massive MIMO With Spatially Correlated Rician Fading Channels. *IEEE Transactions on Communications*, v. 67, n. 5, p. 3234–3250, 2019.

[70] BOYD, S.; VANDENBERGHE, L. *Convex Optimization*. Cambridge: Cambridge University Press, 2004.

[71] WANG, H.; ZENG, Y. Can Sparse Arrays Outperform Collocated Arrays for Future Wireless Communications? In: *2023 IEEE Globecom Workshops (GC Wkshps)*. Kuala Lumpur, Malaysia: IEEE, 2023. p. 667–672.

[72] JEON, J. et al. MIMO Evolution toward 6G: Modular Massive MIMO in Low-Frequency Bands. *IEEE Communications Magazine*, v. 59, n. 11, p. 52–58, 2021.

[73] LI, X. et al. Near-Field Modeling and Performance Analysis of Modular Extremely Large-Scale Array Communications. *IEEE Communications Letters*, v. 26, n. 7, p. 1529–1533, 2022.

[74] IRMER, R. et al. Coordinated Multipoint: Concepts, Performance, and Field Trial Results. *IEEE Communications Magazine*, v. 49, n. 2, p. 102–111, 2011.

[75] PENG, M. et al. Recent Advances in Cloud Radio Access Networks: System Architectures, Key Techniques, and Open Issues. *IEEE Communications Surveys & Tutorials*, v. 18, n. 3, p. 2282–2308, 2016.

[76] HOSSEINI, K.; YU, W.; ADVE, R. S. Large-Scale MIMO Versus Network MIMO for Multicell Interference Mitigation. *IEEE Journal of Selected Topics in Signal Processing*, v. 8, n. 5, p. 930–941, 2014.

[77] NGO, H. Q. et al. Cell-Free Massive MIMO Versus Small Cells. *IEEE Transactions on Wireless Communications*, v. 16, n. 3, p. 1834–1850, 2017.

[78] WANG, Z. et al. Uplink Precoding Design for Cell-Free Massive MIMO With Iteratively Weighted MMSE. *IEEE Transactions on Communications*, v. 71, n. 3, p. 1646–1664, 2023.

[79] HORN, R. A.; JOHNSON, C. R. *Matrix Analysis*. 2. ed. Cambridge, UK: Cambridge University Press, 2012.

[80] WU, M. et al. Approximate Matrix Inversion for High-Throughput Data Detection in the Large-Scale MIMO Uplink. In: *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*. Beijing, China: IEEE, 2013. p. 2155–2158.

[81] LU, S.; MAY, J.; HAINES, R. J. Effects of Correlated Shadowing Modeling on Performance Evaluation of Wireless Sensor Networks. In: *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*. Boston, MA, USA: IEEE, 2015. p. 1–5.

[82] LU, S.; MAY, J.; HAINES, R. J. Efficient Modeling of Correlated Shadow Fading in Dense Wireless Multi-Hop Networks. In: *2014 IEEE Wireless Communications and Networking Conference (WCNC)*. Istanbul, Turkey: IEEE, 2014. p. 311–316.

[83] ABBAS, T. et al. A Measurement Based Shadow Fading Model for Vehicle-to-Vehicle Network Simulations. *International Journal of Antennas and Propagation*, Wiley Online Library, v. 2015, n. 1, p. 190607, 2015.

[84] TSE, D.; VISWANATH, P. *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge University Press, 2005.

[85] MARZETTA, T. et al. *Fundamentals of Massive MIMO*. Cambridge, UK: Cambridge University Press, 2016. ISBN 9781107175570.

[86] RAPPAPORT, T. S. *Wireless Communications: Principles and Practice.* Upper Saddle River, NJ: Prentice Hall, 1996.

[87] GOLDSMITH, A. *Wireless Communications.* Cambridge, UK: Cambridge University Press, 2005.

[88] 3rd Generation Partnership Project (3GPP). *Technical Specification Group Radio Access Network; Spatial channel model for Multiple Input Multiple Output (MIMO) simulations.* Sophia Antipolis, France, 2020.

Appendix

APPENDIX A – PAPER A

- **Title:** Improving spectral efficiency via pilot assignment and subarray selection under realistic XL-MIMO channels
- **Authors:** Gabriel Avanzi Ubiali, Taufik Abrão and José Carlos Marinello Filho
- **Journal:** Physical Communication (IF: 2.0), vol. 58, 102062, June 2023
- **DOI:** 10.1016/j.phycom.2023.102062
- **Citation:** Ubiali, G. A., Abrão T., Marinello, J. C., "Improving spectral efficiency via pilot assignment and subarray selection under realistic XL-MIMO channels," *Physical Communication*, vol. 58, 102062, June 2023.

Physical Communication 58 (2023) 102062

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

Improving spectral efficiency via pilot assignment and subarray selection under realistic XL-MIMO channels[☆]

Gabriel Avanzi Ubiali[a], Taufik Abrão[a,*], José Carlos Marinello[b]

[a] Electrical Engineering Department, State University of Londrina, PR, Brazil. Rod. Celso Garcia Cid - PR445, s/n, Campus Universitário, 10.011, CEP: 86057-970, Brazil

[b] Electrical Engineering Department, the Federal University of Technology PR, Cornélio Procópio, PR, CEP: 86300-000, Brazil

ARTICLE INFO

Article history:
Received 11 December 2022
Received in revised form 9 February 2023
Accepted 23 March 2023
Available online 31 March 2023

Keywords:
Extra-large-scale antenna array
Non-stationary channels
Channel model
Spectral efficiency
Pilot assignment
Subarray selection

ABSTRACT

The main requirements for 5G and beyond connectivity include a uniform high quality of service, which can be attained in crowded scenarios by extra-large MIMO (XL-MIMO) systems. Another requirement is to support an increasing number of connected users in (over)crowded machine-type communication (mMTC). In such scenarios, pilot assignment (PA) becomes paramount to reduce pilot contamination and consequently improve spectral efficiency (SE). We propose a novel quasi-optimal low-complexity iterative pilot assignment strategy for XL-MIMO systems, based on a genetic algorithm (GA). The proposed GA-based PA procedure turns the quality of service more uniform, taking into account the normalized mean-square error (NMSE) of channel estimation from each candidate of the population. The simulations reveal that the proposed iterative procedure minimizes the channel estimation NMSE averaged over the UEs. The second procedure is the subarray (SA) selection. In XL-MIMO systems, commonly a UE is close to an SA antenna subset such that a sufficient data rate can be achieved if only a specific SA serves that UE. Thus, an SA selection procedure is investigated to make the system scalable by defining the maximum number of UEs each SA can help. Hence, the SA clusters are formatted based on the PA decision. Furthermore, we introduce an appropriate channel model for XL-MIMO, which considers a deterministic LoS component with a distance-dependent probability of existence combined with a stochastic spatially correlated Rayleigh NLoS fading component. The developed simulations and analyses rely on this suitable channel model under realistic assumptions of pilot contamination and correlated channels.

1. Introduction

One of the main enabling technologies for achieving high data rates in wireless networks is *Massive multiple-input multiple-output* (mMIMO). The massive number of channel observations, at the access point (AP) antennas, increases the array gain and the ability to use the spatial domain to discriminate the desired signal from the interfering signals, in the *uplink* (UL), and to transmit the signals very directly, in the *downlink* (DL). Indeed, the UEs that are close to one of the APs, *i.e.*, located at a *cell-center area*, will experience high data rates; however, the UEs that are relatively distant from any AP, *i.e.*, located at a *cell-edge area*, may have bad quality of service (QoS) due to low *signal-to-noise-ratio* (SNR) and to the inter-cell interference, which comes from the UEs in the neighboring cells, due to the reuse of pilot sequences and to the inability of an AP to cancel the interference from the other cells' UEs, since it only estimates the channel of the UEs it is serving.

Thus, despite the current cellular networks can achieve high peak data rates, the QoS could result unreliable due to significant data rate variations within the cell. However, wireless access is supposed to be ubiquitous since payments, navigation, and entertainment rely on wireless connectivity. It is more important for future mobile networks to provide more uniform high rates over the coverage area than to increase the peak rates. More uniform high rates can be achieved under large antenna arrays dimensionality, the so-called *extra-large* MIMO (XL-MIMO) networks, which can serve relatively large areas such as stadiums and shopping malls densely populated. XL-MIMO networks can turn the QoS more uniform across the entire cell area by achieving moderate to high macro-diversity and reducing the average distance from a UE to the closest AP antenna.

An XL-MIMO antenna array is split into *subarray* (SA) arrangements. As each UE is physically close to a subset of SAs only, the

☆ This work was partly supported by the National Council for Scientific and Technological Development (CNPq) of Brazil under Grants 310681/2019-7 and 163568/2021-9 (scholarship), and in part by State University of Londrina (UEL), PR, Brazil.

* Corresponding author.
E-mail addresses: g.avanziubiali@gmail.com (G.A. Ubiali), taufik@uel.br (T. Abrão), jcmarinello@utfpr.edu.br (J.C. Marinello).

https://doi.org/10.1016/j.phycom.2023.102062
1874-4907/

configuration where all the SAs serve all the UEs is impractical and unnecessary. Limiting the set of SAs that serve each UE can negatively affect the QoS, but if the SA selection is made judiciously, the system can still provide suitable SE, while also reducing the computational complexity and the amount of signaling and control data that needs to be transferred from the SAs to the CPU and vice-versa. Furthermore, SA selection procedure is essential in XL-MIMO systems to guarantee scalability, *i.e.*, to assure that the complexity will not go to infinity as the number of UEs grows unbounded.

Knowledge of the UE–SA channel responses is paramount to coherent processing. The most efficient way to estimate the channels is to consider a TDD scheme where we only need to transmit pilots in the uplink. Depending on the users' mobility, the size of the normalized coherence block may be smaller than the number of UEs. The mutually orthogonal pilot sequences must be reused among the UEs in this case. Most of XL-MIMO papers consider perfect channel estimation [1–6], which is unrealistic. During the UL pilot transmission, the estimation error is generated by two sources: the background noise and the signals coming from the pilot-sharing UEs. The last one results in so-called coherent interference. In this paper, we develop our analysis assuming imperfect channel estimations to obtain realistic numerical simulation results.

The XL-MIMO channel can be described as the composition of a deterministic *line-of-sight* (LoS) path and a stochastic *non*-LoS (NLoS) component. Due to the low probability of LoS occurrence, generally, the LoS component modeling is not distinguished between suburban and urban macro scenarios; indeed, it is selectable for the urban micro scenario only [7]. Usually, research papers in mMIMO only consider the NLoS channel component. On the other hand, papers in cell-free mMIMO (CF-mMIMO) usually consider either LoS or NLoS component. However, due to the uncertainties in the propagation environment in urban areas, it is impossible to accurately determine the presence of an LoS link [8], and a probabilistic LoS model is necessary. Furthermore, the LoS component contributes considerably to the total channel power [9], since it is associated to a shorter propagation distance and to a slower signal attenuation with the distance. Particularly, in XL-MIMO systems, the distance between a UE and the BS antenna array may be small relative to the XL-array length, so the LoS probability should not be neglected. In this paper, we introduce a realistic channel model that incorporates a deterministic LoS component, with a distance-dependent probability of existence, and a spatially correlated NLoS component. Shadow fading correlation between two different links is also considered.

Depending on the mobility, the size of the coherence block may be smaller than the number of UEs. In this case, the mutually orthogonal pilots must be reused among the UEs and can be appropriately assigned to them to limit the pilot contamination [10]. The most simple and naive approach is to give pilots randomly. However, it cannot prevent closely located UEs from being assigned the same pilot, leading to high interference and system performance degradation. On the other extreme, the exhaustive search, which tries all possible assignments and chooses the best one, leads to a computational complexity that presents a combinatorial growth with the number of active UEs. Thus, only suboptimal methods are feasible in practice. One approach is to optimize a utility function concerning one UE at a time, considering each UE once or iterating until convergence [10]. Several suboptimal algorithms consider a golden rule that closely located UEs should not be assigned the same pilot.

Works on cellular mMIMO usually consider that two UEs within the same cell are assigned mutually orthogonal pilot sequences. If the pilot reuse factor is one, the performance of cell-edge UEs may be degraded by pilot-sharing UEs of adjacent cells. In this scenario, the pilot contamination problem can be reasonably relieved greedily by simply assigning the best pilots, *i.e.*, the ones that result in the most negligible pilot contamination to the cell-edge UEs [11]. In [12], the problem of pilot assignment in multi-cell mMIMO systems is tackled by using a *deep reinforcement learning* (DRL) scheme. The distance and the angle-of-arrival (AoA) information of the users are deployed to define the cost function, representing the pilot contamination effects. The DRL algorithm is applied to find an effective pilot assignment strategy by tracking the changes in the environment, learning the near-optimal pilot assignment, and achieving comparable performance to that of the optimum (genie) PA performed by exhaustive search.

However, due to the increased density of UEs in typical applications of XL-MIMO systems and the scarcity of pilots, crucial tasks, such as pilot assignment, network access, and user scheduling, become challenging [13]. In our work, we consider a crowded XL-MIMO scenario where the number of UEs in the coverage area is more significant than the number of orthogonal pilot sequences, representing a scenario with pilot scarcity. A greedy algorithm would assign the orthogonal pilots randomly to some of the UEs, and then assign pilots to the remaining UEs to minimize the pilot contamination, each UE at a time. Works in [2, 13–15] address the problem of random access and pilot assignment in XL-MIMO, considering uncorrelated Rayleigh fading and no subarray selection. Specifically, a joint random access and scheduling protocol was proposed in [13], taking advantage of the different visibility regions (VRs) that arise in the UE and XL-MIMO link, besides seeking UEs with non-overlapping VRs to be scheduled in the same payload data pilot resource. The difference to our work is that we focus on the pilot assignment problem and develop an efficient PA algorithm for a probabilistic LoS/NLoS XL-MIMO channel model based on a heuristic evolutionary genetic algorithm (GA) approach and build up the analysis considering the employment of SA selection to make the receive combining scalable.

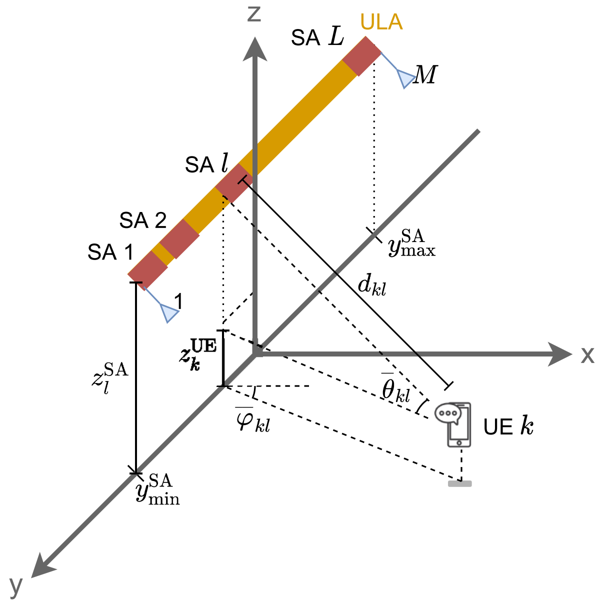
The contribution of this work is threefold.

(*i*) We introduce a realistic XL-MIMO spatially *correlated channel model*, since uncorrelated channels appear under extremely strict physical requirements. Since the XL-MIMO antennas are not co-located, sometimes the UEs will be very close to some SAs, such that the channel modeling must be a composition of a sum of NLoS links and an LoS link, with a distance-dependent probability of occurrence. To the best of our knowledge, all the works on XL-MIMO have considered only one of these two components.

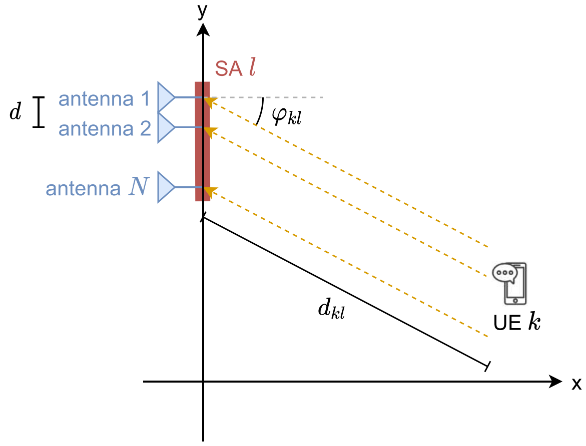
(*ii*) We propose a GA-based PA strategy for crowded XL-MIMO systems, mitigating pilot contamination by approaching the optimal PA solution w.r.t. minimizing the average *normalized mean-square error* (NMSE) of the channel estimates among all the active UEs.

(*iii*) We provide extensive numerical results corroborating the effectiveness and efficiency of the proposed SA selection and PA methodology for XL-MIMO systems, exploring crowded scenarios by considering both the LoS and NLoS channel components. All the analysis and obtained results of this paper are based on the realistic assumption of imperfect channel estimation, the employment of a scalable receive combiner, and a proper SA selection, which is also essential to make the XL subarrays scalable.

The remainder of this paper is organized as follows. Section 2 describes the proposed channel model for XL-MIMO. Section 3 deals with the XL-MIMO system model, describing the signal model for the UL pilot and data transmissions and the adopted channel estimator and receive combiner. Section 4 presents the



(a) 3-D view of the XL-MIMO system



(b) The l-th SA and k-th UE projected in plane $z = 0$

Fig. 1. 3-D XL-MIMO representation, showing (a) the ULA and the azimuth and elevation angles corresponding to the LoS path between UE k and SA l and (b) the antenna spacing d (inside SA l) and the distance between UE k and SA l.

considered strategies for PA in XL-MIMO and the adopted SA selection algorithm. Finally, numerical results corroborating our findings are provided in Section 5, and final remarks and conclusions are presented in Section 6.

2. XL-MIMO channel model

We consider an XL-MIMO antenna array with a M-antenna ULA, divided into L SAs containing N antennas each, such that $M = LN$, serving K UEs. The antennas in each SA are separated by a distance d. If N is small, then each SA will experience spatial stationarity, *i.e.*, all its N antennas will experience the same condition of existence or absence of the LoS path component. Each SA is connected to a central processing unit (CPU), which is responsible for the SA cooperation. The height of the lth SA and the kth UE are denoted by z_l^{SA} and z_k^{UE}, respectively, and d_{kl} denotes the distance from the kth UE to the first antenna of the lth SA. Fig. 1 illustrates the XL-MIMO antenna array (yellow bar) highlighting the azimuth and elevation angles associated to the LoS path between a given UE k and a given SA l, represented by the red bar.

The associated channel model with LoS and the NLoS components can be expressed as:

$$\mathbf{h}_{kl} = \alpha_{kl}\mathbf{h}_{kl}^{\mathrm{LoS}} + \mathbf{h}_{kl}^{\mathrm{NLoS}}, \tag{1}$$

where $\mathbf{h}_{kl}^{\mathrm{LoS}}, \mathbf{h}_{kl}^{\mathrm{NLoS}}, \mathbf{h}_{kl} \in \mathbb{C}^N$ denote, respectively, the LoS, the NLoS and the probabilistic LoS/NLoS channel vectors between UE k and SA l, while α_{kl} is a Bernoulli random variable indicating the presence ($\alpha_{kl} = 1$) or absence ($\alpha_{kl} = 0$) of an LoS component between the kth UE and the lth SA.

The LoS channel vector between the kth UE and the lth SA is given by [8]

$$\mathbf{h}_{kl}^{\mathrm{LoS}} = \sqrt{G_k G_l \frac{z_k^{\mathrm{UE}} z_l^{\mathrm{SA}}}{4\pi d_{kl}}}\sqrt{\mathcal{X}_{kl}^{\mathrm{LoS}}}\, e^{-j2\pi\frac{d_{kl}}{\lambda}}\mathbf{a}(\overline{\varphi}_{kl}, \overline{\theta}_{kl}), \tag{2}$$

where G_k and G_l denote the antenna gain at the kth UE and the lth SA, respectively; $\lambda = c/f$ is the wavelength, being c the speed of light and f the carrier frequency; $\mathcal{X}_{kl}^{\mathrm{LoS}}$ is the shadow fading of the LoS component; and $\mathbf{a}(\overline{\varphi}_{kl}, \overline{\theta}_{kl})$ is the array response vector of the ULA, regarding to the LoS ray. The azimuth and elevation angles of the LoS path between the kth UE and the lth SA are denoted by $\overline{\varphi}_{kl}$ and $\overline{\theta}_{kl}$, respectively. The *array response vector*, as a function of the LoS azimuth and elevation angles between UE k and SA l, respectively $\overline{\varphi}_{kl}$ and $\overline{\theta}_{kl}$, is given by

$$\mathbf{a}(\overline{\varphi}_{kl}, \overline{\theta}_{kl}) = \left[1, e^{-j2\pi\frac{d}{\lambda}\frac{\sin(\overline{\varphi}_{kl})}{\cos(\overline{\theta}_{kl})}}, \dots, e^{-j2(N-1)\pi\frac{d}{\lambda}\frac{\sin(\overline{\varphi}_{kl})}{\cos(\overline{\theta}_{kl})}}\right]^{\mathrm{T}}, \tag{3}$$

where d is the antenna spacing inside the lth SA, Fig. 1(b). Indeed, the array response vector $\mathbf{a}(\cdot,\cdot)$ in Eq. (3) accounts for the phase differences among the signal that arrives in different antennas of the same *uniform linear SA* (ULSA). These variations appear when the azimuth angle is not zero, *i.e.*, when the wavefront is not parallel to the ULA.

Near-field vs. Far-field. An important observation on XL-MIMO arrays is that the wavefront propagation could result in spherical, since some users may be in the near-field distance w.r.t. the entire XL-MIMO array, *i.e.*, $d_{kl} \ll d_{\mathrm{Rayl}}$, being d_{Rayl} the Rayleigh distance [16]. Also, such conditions will result in different propagation distances for different SAs. On the other hand, since the SA aperture is small compared to the UE-SA propagation distances, one can assume that all users are in the far-field propagation from each SA perspective, such that we can assume that the wavefront propagation is *planar* to the SA point of view. The implication is that the difference between the propagation distance for two antennas belonging to the same SA is negligible, such that the corresponding large-scale fading (LSF) coefficients are nearly the same for all antenna elements of the same SA. However, even this negligible difference is sufficient to produce phase differences among these antennas. Such phase differences are accounted by the array response vector in (3).

The LSF channel gain for the LoS component in each XL-MIMO ULSA is determined by the normalized trace as

$$\beta_{kl}^{\mathrm{LoS}} = \frac{1}{N}\|\mathbf{h}_{kl}^{\mathrm{LoS}}\|^2 = \underbrace{\frac{G_k G_l (z_k^{\mathrm{UE}} z_l^{\mathrm{SA}})^2}{(4\pi)^2}}_{\beta_0}\frac{\mathcal{X}_{kl}^{\mathrm{LoS}}}{d_{kl}^2} = \frac{\beta_0}{d_{kl}^2}\mathcal{X}_{kl}^{\mathrm{LoS}}, \tag{4}$$

where β_0 is the average channel gain at the reference distance of 1 m, and β_0/d_{kl}^2 is the path-loss of the LoS component. Replacing (4) in (2) gives

$$\mathbf{h}_{kl}^{\text{LoS}} = \sqrt{\beta_{kl}^{\text{LoS}}}e^{-j2\pi\frac{d_{kl}}{\lambda}}\mathbf{a}(\overline{\varphi}_{kl}, \overline{\theta}_{kl}). \tag{5}$$

On the other hand, the NLoS component describes a practical *spatially correlated multi path* environment. The NLoS channel vector can be obtained as the superposition of B physical signal paths, each one reaching the ULSA as a plane wave from particular azimuth and elevation angles $\varphi_{kl}^{(b)}$ and $\theta_{kl}^{(b)}$ [17]:

$$\mathbf{h}_{kl}^{\text{NLoS}} = \sum_{b=1}^{B} g_{kl}^{(b)}\mathbf{a}(\varphi_{kl}^{(b)}, \theta_{kl}^{(b)}), \tag{6}$$

where $g_{kl}^{(b)}$ is an i.i.d. complex random variable with zero mean and variance $\mathbb{E}\{|g_{kl}^{(b)}|^2\}$, accounting for the gain and phase rotation of the bth physical NLoS path, while $\varphi_{kl}^{(b)}$ and $\theta_{kl}^{(b)}$ are, respectively, the SA azimuth and elevation angles formed by the bth NLoS path between UE k and SA l.

The NLoS LSF coefficient can be calculated as

$$\beta_{kl}^{\text{NLoS}} = \frac{1}{N}\mathbb{E}\{\|\mathbf{h}_{kl}^{\text{NLoS}}\|^2\} = \sum_{b=1}^{B}\mathbb{E}\{|g_{kl}^{(b)}|^2\}, \tag{7}$$

and is given by

$$\beta_{kl}^{\text{NLoS}} = \frac{\beta_0}{d_{kl}^{\gamma}}\mathcal{X}_{kl}^{\text{NLoS}}, \tag{8}$$

where γ is the path-loss exponent of the NLoS component, and $\mathcal{X}_{kl}^{\text{NLoS}}$ is the shadow fading of the NLoS component. Considering the law of large numbers, the sum of the B random NLoS paths approaches a normal distribution [17]:

$$\mathbf{h}_{kl}^{\text{NLoS}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_{kl}), \tag{9}$$

with the positive semi-definite *spatial correlation matrix* $\mathbf{R}_{kl}$ composed by the elements:

$$[\mathbf{R}_{kl}]_{m,n} = \beta_{kl}^{\text{NLoS}} \iint e^{j2(m-n)\pi\frac{d}{\lambda}\frac{\sin(\varphi)}{\cos(\theta)}} f_{kl}(\varphi, \theta)d\varphi d\theta, \tag{10}$$

where $f_{kl}(\varphi, \theta)$ is the joint probability density function (jPDF) of φ and θ, which depends on the azimuth and elevation angles of the LoS link (φ_{kl} and θ_{kl}) and on the angular standard deviations ($\sigma_\varphi \geq 0$ and $\sigma_\theta \geq 0$). The integration interval for φ and θ is $[-180°, 180°]$ and $(-90°, 90°)$, respectively. As in [10], we also consider the joint Gaussian distribution:

$$f_{kl}(\varphi, \theta) = \frac{1}{2\pi\sigma_\varphi\sigma_\theta}\exp\left\{-\frac{(\varphi - \overline{\varphi}_{kl})^2}{2\sigma_\varphi^2} - \frac{(\theta - \overline{\theta}_{kl})^2}{2\sigma_\theta^2}\right\}. \tag{11}$$

The proof of (9) and (10) can be found in Appendix. From (10), notice that $\text{tr}(\mathbf{R}_{kl}) = N\beta_{kl}^{\text{NLoS}}$, which matches with the definition in (7).

Separating the large-scale and the small-scale fading, the NLoS channel vector can be described as

$$\mathbf{h}_{kl}^{\text{NLoS}} = \sqrt{\beta_{kl}^{\text{NLoS}}} \cdot \ddot{\mathbf{h}}_{kl}, \tag{12}$$

where $\ddot{\mathbf{h}}_{kl}$ represents the *spatially correlated small-scale* Rayleigh *fading* following a complex normal distribution [10,17,18]. Note that the double-integral in (10) computes $\mathbb{E}\left\{e^{j2(m-n)\pi\frac{d}{\lambda}\frac{\sin(\varphi)}{\cos(\theta)}}\right\}$ w.r.t. randomly located multipath components distributed according to the joint PDF in (11).

In practical XL-MIMO panel scenarios, as the antennas of different SAs are sufficiently distant, we assume the channel vectors of other SAs are independently distributed, which means that $\mathbb{E}\{\ddot{\mathbf{h}}_{kl}\ddot{\mathbf{h}}_{kj}^{\text{H}}\} = \mathbf{0}_{N\times N}$ for $l \neq j$, where $(.)^{\text{H}}$ denotes the Hermitian operator.

Correlation between F_{kl} ***and*** F_{ij} ***(Shadowing)***. The *shadow fading* can be modeled as a log-normal r.v., in dB scale as $F_{kl} = 10\log_{10}(\mathcal{X}_{kl})$, following the distribution [19]

$$F_{kl} \sim \mathcal{N}\left[0, \sigma_{\text{SF}}^2(1 - e^{-d_{kl}/\delta})^2\right]. \tag{13}$$

Note that the variance of the r.v. F_{kl} in (13) increases with the distance and converges to σ_{SF}^2 when $d_{kl} \gg \delta$, where δ denotes the decorrelation distance. The shadow fading is typically modeled as independent links, although in reality proximate links present spatially correlated shadowing, since the signal is impacted by almost the same set of objects in the propagation environment. Ignoring such correlations results in over-estimating the diversity in XL-MIMO systems and may produce significant differences in network performance [19]. The cross-correlation between F_{kl} and F_{ij} is given by:

$$\mathbb{E}\{F_{kl}F_{ij}\} = \frac{\sigma_{\text{SF}}^2}{2}\frac{(1 - e^{-d_{kl}/\delta})(1 - e^{-d_{ij}/\delta})}{\sqrt{(1 + e^{-d_{kl}/\delta})(1 + e^{-d_{ij}/\delta})}} \times (e^{-d_{kj}/\delta} + e^{-d_{il}/\delta} + e^{-D_{ki}^{\text{UE}}/\delta} + e^{-D_{lj}^{\text{SA}}/\delta}). \tag{14}$$

where d_{kl} is the distance between the kth UE and the lth SA, D_{ki}^{UE} is the distance between UE k and UE i, and D_{lj}^{SA} is the distance between SA l and SA j [19,20]. Note that the closer the kth UE is from the ith UE, the more correlated their shadow fading, F_{kl} and F_{ij}, will be. Also, the closer the lth SA is from the jth SA, the more correlated F_{kl} and F_{ij} will be.

Finally, from (1) and (9), the channel vector follows a *complex normal distribution* with mean $\overline{\mathbf{h}}_{kl} = \alpha_{kl}\mathbf{h}_{kl}^{\text{LoS}}$ and *covariance matrix*

$$\begin{aligned}\mathbb{C}\text{ov}(\mathbf{h}_{kl}) &= \mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kl}^{\text{H}}\} - \overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{kl}^{\text{H}} \\ &= \mathbb{E}\{(\overline{\mathbf{h}}_{kl} + \mathbf{h}_{kl}^{\text{NLoS}})(\overline{\mathbf{h}}_{kl} + \mathbf{h}_{kl}^{\text{NLoS}})^{\text{H}}\} - \overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{kl}^{\text{H}} \\ &= \mathbb{E}\{\mathbf{h}_{kl}^{\text{NLoS}}(\mathbf{h}_{kl}^{\text{NLoS}})^{\text{H}}\} \\ &= \mathbf{R}_{kl},\end{aligned} \tag{15}$$

such that

$$\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\overline{\mathbf{h}}_{kl}, \mathbf{R}_{kl}). \tag{16}$$

Notice that the correlation matrix of $\mathbf{h}_{kl}^{\text{NLoS}}$ is also $\mathbf{R}_{kl}$, since $\mathbf{h}_{kl}^{\text{NLoS}}$ is a zero mean vector. Indeed, as demonstrated in (15), the covariance matrix of $\mathbf{h}_{kl}$ is also $\mathbf{R}_{kl}$; however, its correlation matrix is $\mathbf{R}_{kl} + \overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{kl}^{\text{H}}$, since it is a non-zero mean vector.

Both $\overline{\mathbf{h}}_{kl}$ and $\mathbf{R}_{kl}$ describe the macroscopic propagation effects, *i.e.*, path-loss and shadowing, whereas the distribution in (16) describes the small-scale fading. If there is an LoS path between UE k and SA l, then we say that SA l is inside the visibility region (VR) of UE k. Otherwise, the received power is much weaker as the signal is coming only from NLoS paths, and then we say that SA l is not in the VR of UE k.

Finally, the probabilistic LoS/NLoS LSF channel gain is defined as

$$\beta_{kl} = \frac{1}{N}\mathbb{E}\{\|\mathbf{h}_{kl}\|^2\}, \tag{17}$$

and, by replacing (1) into (17), and recalling that $\|\mathbf{h}_{kl}^{\text{LoS}}\|^2 = N\beta_{kl}^{\text{LoS}}$ and $\mathbb{E}\{\|\mathbf{h}_{kl}^{\text{NLoS}}\|^2\} = \text{tr}(\mathbf{R}_{kl}) = N\beta_{kl}^{\text{NLoS}}$, we come to

$$\beta_{kl} = \alpha_{kl}\beta_{kl}^{\text{LoS}} + \beta_{kl}^{\text{NLoS}}. \tag{18}$$

In the following sections, we describe the XL-MIMO system model and the procedures for both PA and SA selection. An overview of the operations involved in such processes is shown

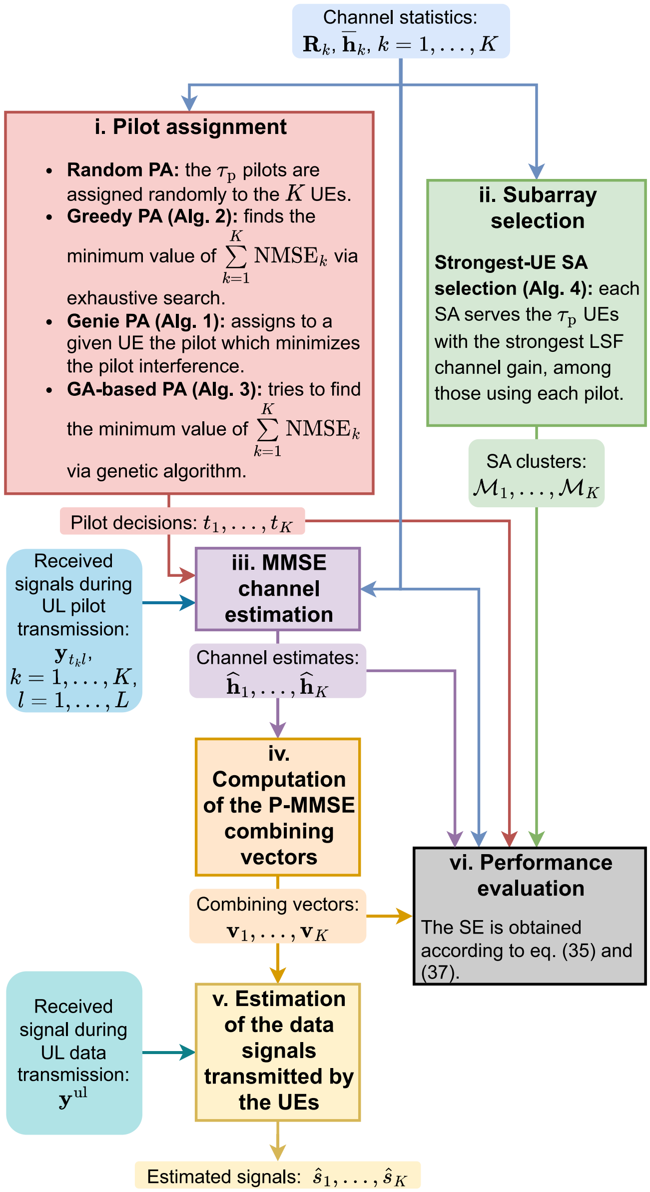


Fig. 2. System operation phases: (*i*) pilot assignment (four methods were compared); (*ii*) subarray selection; (*iii*) channel estimation; (*iv*) computation of the combining vectors; (*v*) signal estimation; (*vi*) SE calculation.

in Fig. 2: first, the pilots are assigned to the K UEs based on the channel statistics (channel average and covariance matrices). Second, the SA selection is performed, also based on the channel statistics. Third, the instantaneous channels are estimated in each coherence block, based on the pilot decisions and the received signals during the UL pilot transmission. Fourth, the combining vectors are computed, based on the channel statistics and the pilot decisions (which are necessary to obtain the channel estimation covariance matrices) and on the channel estimates. Fifth, the signals transmitted by the UEs are estimated, based on the combining vectors and on the received signal at the ULA during the UL data transmission. Finally, when evaluating the system performance, we can compute the SINR based on Eq. (35), based on the channel estimates, channel statistics, pilot decisions, SA clusters and combining vectors.

3. XL-MIMO system model

3.1. UL pilot transmission

Pilots are τ_p-dimensional vectors, and we can only find at most τ_p mutually orthogonal sequences. The set $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{\tau_p} \in \mathbb{C}^{\tau_p}$, containing τ_p orthogonal pilot sequences, is defined such that

$$\boldsymbol{\phi}_i^{\mathrm{H}} \boldsymbol{\phi}_j = \begin{cases} \tau_{\mathrm{p}}, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{19}$$

The channels need to be estimated once per coherence block. Therefore, τ_p samples must be reserved for uplink pilot signaling in each coherence block, but the finite length of the coherence blocks imposes the constraint $\tau_p \le \tau_c$, where τ_c is the number of symbols (samples) that can be transmitted over one coherence block. Frequently, $K \gg \tau_c$ makes it impossible to assign mutually orthogonal pilots to all UEs. Therefore, two or more UEs might be assigned the same pilot. We denote the index of the pilot assigned to UE k as $t_k \in \{1, \ldots, \tau_p\}$ and define the set of UEs using the same pilot as UE k, including itself:

$$\mathcal{P}_k = \{i : t_i = t_k, i = 1, \ldots, K\} \subset \{1, \ldots, K\}. \tag{20}$$

The elements of $\sqrt{p_k}\boldsymbol{\phi}_{t_k}$ are transmitted by UE k over τ_p consecutive samples at the beginning of the coherence block, where p_k is the uplink transmit power of UE k. Then, during the entire pilot transmission, the received signal $\mathbf{Y}_l^{\mathrm{p}} \in \mathbb{C}^{N \times \tau_{\mathrm{p}}}$ at SA l is

$$\mathbf{Y}_l^{\mathrm{p}} = \sum_{i=1}^{K} \sqrt{p_i} \mathbf{h}_{il} \boldsymbol{\phi}_{t_i}^{\mathrm{T}} + \mathbf{N}_l^{\mathrm{p}}, \tag{21}$$

where $\mathbf{N}_l^{\mathrm{p}} \in \mathbb{C}^{N \times \tau_{\mathrm{p}}}$ is the receiver noise with i.i.d. elements distributed as $\mathcal{N}_{\mathbb{C}}(0, \sigma_{\mathrm{n}}^2)$.

Besides, the interference coming from non-pilot-sharing UEs can be eliminated, by multiplying the received signal $\mathbf{Y}_l^{\mathrm{p}}$ with the normalized conjugate of the pilot $\boldsymbol{\phi}_{t_k}$, resulting in the processed received pilot signal $\mathbf{y}_{t_k l}^{\mathrm{p}} \in \mathbb{C}^N$, which is a sufficient statistic for estimating $\mathbf{h}_{kl}$:

$$\mathbf{y}_{t_k l}^{\mathrm{p}} = \mathbf{Y}_l^{\mathrm{p}} \boldsymbol{\phi}_{t_k}^* = \sqrt{p_k}\tau_{\mathrm{p}} \mathbf{h}_{kl} + \sum_{i \in \mathcal{P}_k \setminus \{k\}} \sqrt{p_i}\tau_{\mathrm{p}} \mathbf{h}_{il} + \mathbf{n}_{t_k l}^{\mathrm{p}}, \tag{22}$$

where $\mathbf{n}_{t_k l}^{\mathrm{p}} = \mathbf{N}_l^{\mathrm{p}} \boldsymbol{\phi}_{t_k}^* \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tau_{\mathrm{p}}\sigma_{\mathrm{n}}^2 \mathbf{I}_N)$.

3.2. Channel estimation

Channel estimates can be obtained based on the observation $\mathbf{y}_{t_k l}^{\mathrm{p}}$. The MMSE estimator uses statistical information, *i.e.*, the channel means, and the covariance matrices to partially remove the interference from pilot-sharing UEs. Based on the received signal $\mathbf{y}_{t_k l}^{\mathrm{p}}$, the MMSE estimate of $\mathbf{h}_{kl}$ is defined as [21]:

$$\widehat{\mathbf{h}}_{kl} = \bar{\mathbf{h}}_{kl} + \sqrt{p_k} \mathbf{R}_{kl} \boldsymbol{\Psi}_{t_k l}^{-1} \left(\mathbf{y}_{t_k l}^{\mathrm{p}} - \bar{\mathbf{y}}_{t_k l}^{\mathrm{p}} \right), \tag{23}$$

where

$$\boldsymbol{\Psi}_{t_k l} = \frac{1}{\tau_{\mathrm{p}}} \mathrm{Cov}(\mathbf{y}_{t_k l}^{\mathrm{p}}) = \sum_{i \in \mathcal{P}_k} p_i \tau_{\mathrm{p}} \mathbf{R}_{il} + \sigma_{\mathrm{n}}^2 \mathbf{I}_N, \tag{24}$$

and

$$\bar{\mathbf{y}}_{t_k l}^{\mathrm{p}} = \sqrt{p_k}\tau_{\mathrm{p}} \bar{\mathbf{h}}_{kl} + \sum_{i \in \mathcal{P}_k \setminus \{k\}} \sqrt{p_i}\tau_{\mathrm{p}} \bar{\mathbf{h}}_{il}. \tag{25}$$

The statistical distributions (the mean vector and covariance matrices) are assumed perfectly known, since they are constant during hundreds of channel coherence blocks. In practice, they can be estimated using the sample mean and sample covariance matrices.

The resulting estimation error $\boldsymbol{\varepsilon}_{\mathbf{h}_{kl}} = \mathbf{h}_{kl} - \widehat{\mathbf{h}}_{kl}$ can be modeled as a zero-mean random variable with covariance matrix [21] given by

$$\mathbf{C}_{kl} = \mathbf{R}_{kl} - p_k \tau_{\mathrm{p}} \mathbf{R}_{kl} \boldsymbol{\Psi}_{t_{kl}}^{-1} \mathbf{R}_{kl}. \tag{26}$$

The estimate $\widehat{\mathbf{h}}_{kl}$ and the estimation error $\boldsymbol{\varepsilon}_{\mathbf{h}_{kl}}$ are independent random variables distributed as [21]

$$\widehat{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}\left(\overline{\mathbf{h}}_{kl}, \mathbf{R}_{kl} - \mathbf{C}_{kl}\right), \tag{27}$$

$$\boldsymbol{\varepsilon}_{\mathbf{h}_{kl}} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_N, \mathbf{C}_{kl}\right). \tag{28}$$

The estimation accuracy is defined by the NMSE of the channel estimate [10]. The NMSE of $\widehat{\mathbf{h}}_k$ can be computed as

$$\mathrm{NMSE}_k = \frac{\mathbb{E}\{\|\boldsymbol{\varepsilon}_{\mathbf{h}_k}\|^2\}}{\mathbb{E}\{\|\mathbf{h}_k\|^2\}} = \frac{\mathbb{E}\{\boldsymbol{\varepsilon}_{\mathbf{h}_k}^{\mathrm{H}}\}\mathbb{E}\{\boldsymbol{\varepsilon}_{\mathbf{h}_k}\} + \sum_{l=1}^{L} \mathrm{tr}(\mathbf{C}_{kl})}{N \sum_{l=1}^{L} \beta_{kl}}, \tag{29}$$

where $\boldsymbol{\varepsilon}_{\mathbf{h}_k} = [\boldsymbol{\varepsilon}_{\mathbf{h}_{k1}}{}^{\mathrm{T}}, \ldots, \boldsymbol{\varepsilon}_{\mathbf{h}_{kL}}{}^{\mathrm{T}}]^{\mathrm{T}}$ is the collective channel estimation error vector of UE k. The term $\mathbb{E}\{\|\boldsymbol{\varepsilon}_{\mathbf{h}_k}\|^2\}$ is the mean square error (MSE) of the channel estimation $\widehat{\mathbf{h}}_k$.

3.3. UL data transmission

During the UL data transmission, the ith UE transmits the signal $s_i \in \mathbb{C}$, with power $p_i = \mathbb{E}\{|s_i|^2\}$, and the SAs receive a superposition of the signals sent from the K UEs. The received signal $\mathbf{y}_l^{\mathrm{ul}} \in \mathbb{C}^N$ at the lth SA is then given by

$$\mathbf{y}_l^{\mathrm{ul}} = \sum_{i=1}^{K} \mathbf{h}_{il} s_i + \mathbf{n}_l^{\mathrm{u}}, \tag{30}$$

where $\mathbf{h}_{il} \in \mathbb{C}^N$ denotes the channel vector between the kth UE and the lth SA, and $\mathbf{n}_l^{\mathrm{u}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{\mathrm{n}}^2 \mathbf{I}_N)$ is the independent additive noise during UL data transmission.

We define the set $\mathcal{M}_k$, containing the SAs that serve the kth UE, and the diagonal matrix $\mathbf{D}_{kl} \in \mathbb{C}^{N\times N}$, which indicates whether the lth SA serves the kth UE or not:

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_N, & \text{if } l \in \mathcal{M}_k, \\ \mathbf{0}_N, & \text{if } l \notin \mathcal{M}_k. \end{cases} \tag{31}$$

If $\alpha_{kl} = 1$ and $\alpha_{kj} = 0$, *i.e.*, if UE k has LoS to SA l but not to SA j, a proper SA selection scheme will very likely give preference to SA l over j in serving UE k.

Also, let us define a receive combining vector $\mathbf{v}_{kl} \in \mathbb{C}^N$ that SA l could use if it serves UE k, and the collective combining vector $\mathbf{v}_k = [\mathbf{v}_{k1}^{\mathrm{T}} \ldots \mathbf{v}_{kL}^{\mathrm{T}}]^{\mathrm{T}}$. The collective UL data signal $\mathbf{y}^{\mathrm{ul}} \in \mathbb{C}^{LN}$ is given by

$$\mathbf{y}^{\mathrm{ul}} = \begin{bmatrix} \mathbf{y}_1^{\mathrm{ul}} \\ \vdots \\ \mathbf{y}_L^{\mathrm{ul}} \end{bmatrix} = \sum_{i=1}^{K} \mathbf{h}_i s_i + \mathbf{n}, \tag{32}$$

where $\mathbf{h}_i = [\mathbf{h}_{i1}^{\mathrm{T}} \ldots \mathbf{h}_{1L}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{LN}$ is the ith UE collective channel, and $\mathbf{n} = [\mathbf{n}_1^{\mathrm{T}} \ldots \mathbf{n}_l^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{LN}$ is the collective noise vector along the antennas of all SAs.

In order to obtain an estimate of s_k, the CPU computes the inner product between the effective receive combining vector $\mathbf{D}_k \mathbf{v}_k$, where $\mathbf{D}_k = \mathrm{diag}(\mathbf{D}_{k1}, \ldots, \mathbf{D}_{kL})$, and the received UL data signals at the serving SAs:

$$\widehat{s}_k = \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{y}^{\mathrm{ul}} \tag{33}$$

$$= \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k s_k + \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \boldsymbol{\varepsilon}_{\mathbf{h}_k} s_k + \sum_{\substack{i=1 \\ i\neq k}}^{K} \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{h}_i s_i + \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{n}. \tag{34}$$

Notice that the desired signal $\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{h}_k s_k$ was divided into two parts in Eq. (34): the desired signal over the estimated channel, $\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k s_k$, and the desired signal over the unknown channel, $\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \boldsymbol{\varepsilon}_{\mathbf{h}_k} s_k$. Only the first one can be utilized for signal detection. The second one is not useful since only the distribution of the estimation error is known, and thus it is treated as interference or noise when computing the instantaneous effective uplink SINR of UE k, given by

$$\mathrm{SINR}_k^{\mathrm{ul,c}} = \frac{p_k |\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k|^2}{\sum_{\substack{i=1 \\ i\neq k}}^{K} p_i |\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_i|^2 + \mathbf{v}_k^{\mathrm{H}} \mathbf{Z}_k \mathbf{v}_k + \sigma_{\mathrm{n}}^2 \|\mathbf{D}_k \mathbf{v}_k\|^2}, \tag{35}$$

where

$$\mathbf{Z}_k = \sum_{i=1}^{K} p_i \mathbf{D}_k \mathbf{C}_i \mathbf{D}_k. \tag{36}$$

Finally, being $\tau_{\mathrm{u}}/\tau_{\mathrm{c}}$ the fraction of each coherence block that is used for uplink data transmission, the *achievable* SE of UE k in the centralized operation can be expressed as

$$\mathrm{SE}_k^{\mathrm{ul,c}} = \frac{\tau_{\mathrm{u}}}{\tau_{\mathrm{c}}} \mathbb{E}\left\{\log_2\left(1 + \mathrm{SINR}_k^{\mathrm{ul,c}}\right)\right\}. \tag{37}$$

3.4. Receive combining

A critical implication of deploying an XL antenna array is that the interference that affects UE k is mainly generated by the UEs that are located at its neighborhood [10]. Thus, we will utilize the *partial* MMSE (P-MMSE) for receive combining, defined as:

$$\mathbf{v}_k = p_k \left(\sum_{i\in\mathcal{S}_k} p_i \mathbf{D}_k (\widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}} + \mathbf{C}_i) \mathbf{D}_k + \sigma_{\mathrm{n}}^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k, \tag{38}$$

where only the UEs that are served by at least one SA that also serves UE k are considered in the inverse matrix, *i.e.*, the UEs that have indices in the set

$$\mathcal{S}_k = \{i : \mathbf{D}_k \mathbf{D}_i \neq \mathbf{0}_{LN}\}. \tag{39}$$

In the special case where $\mathcal{S}_k$ contains the K UEs, the vector $\mathbf{v}_k$ becomes the MMSE receive combiner, which is the optimal combiner w.r.t. minimizing the MSE of the estimates $\widehat{s}_k$, and also maximizes the SINR of the kth UE, as defined in (35). The P-MMSE combiner does not attain optimal SE, but it is scalable, since both the computation of the combining vector $\mathbf{v}_k$ and the estimation of the channel vectors $\mathbf{h}_i$, $i \in \mathcal{S}_k$, do not result in a computational complexity that grows without limit when $K \to \infty$.

4. Pilot assignment and subarray selection

Depending on the mobility, the size of the normalized coherence block may be smaller than the number of UEs, *i.e.*, $\tau_{\mathrm{c}} < K$. The finite length of the coherence block imposes that $\tau_{\mathrm{p}} < \tau_{\mathrm{c}}$, such that it is impossible to assign mutually orthogonal pilots to all UEs in overcrowded networks where $K > \tau_{\mathrm{c}}$, and consequently, $K > \tau_{\mathrm{p}}$. Furthermore, it may happen that $\tau_{\mathrm{p}} < K$. Thus, in crowded scenarios, we may need to reuse the τ_{p} mutually orthogonal pilots among the XL-MIMO UEs.

If UEs k and i share the same pilot, *i.e.*, $t_k = t_i$, UE i will probably cause much more interference on the estimate $\widehat{s}_k$ than the UEs that use other pilots than t_k. If the LSF channel gains β_{kl} is stronger than β_{il}, it may be fortunate to assign to this SA the task of estimating s_k but not s_i. Thus, each SA shall estimate the signal of one UE only per pilot sequence, namely the one with the strongest channel, such that each UE will be served by a different subset of SAs. It is a necessary and sufficient condition for scalability since the computational complexity will remain finite as $K \to \infty$ [10].

However, in some cases, it may happen that the kth UE channel is not the strongest among the UEs that share the pilot t_k in any SA, such that the CPU would not even process its signal. Thus, the SA selection is insufficient to ensure a uniform quality of service among the XL-MIMO UEs, and a pilot assignment strategy will be required. A possible way to avoid the impairment mentioned above is to assign exclusive orthogonal pilots to the worst channel UEs. Section 4.2 discusses the proposed pilot assignment strategy.

Both pilot assignment (PA) and subarray selection (SAS) strategies are based only on the channel statistics, not on the instantaneous channel realizations; hence, we do not need to run both algorithms in every coherence block. A basic algorithm for SA selection is to first assign pilots to all UEs and then let each SA serve only one UE per pilot, namely the one with the strongest LSF channel gain[1] among the subset of UEs that have been assigned that pilot. Note that each SA can make these decisions locally [10].

4.1. Pilot contamination effect

Suppose that UEs k and i use the same pilot. Intuitively, the closer the UE i is to the antenna array, the higher its NLoS LSF channel gain ($\beta_{il}^{\mathrm{NLoS}}$) will be and, therefore, the more it will interfere with the estimation of $\mathbf{h}_{kl}$. It can be verified in (26) and (29) that the NMSE increases with $\beta_{il}^{\mathrm{NLoS}}$, which is implicitly included in $\boldsymbol{\Psi}_{t_k l}^{-1}$. Furthermore, if UE k is close to the ULA, and especially if there is an LoS path, the associated channel will be stronger, and the relative estimation error will be smaller since the LoS component changes slowly; therefore, it is assumed to be perfectly known. On the other hand, yet according to (29), the NMSE of estimating $\mathbf{h}_k$ is inversely proportional to $\sum_{l=1}^{L} \beta_{kl}$.

Moreover, if $\mathbf{R}_{kl}$ is invertible, one can write the relation [10]

$$\widehat{\mathbf{h}}_{il} = \bar{\mathbf{h}}_{il} + \sqrt{\frac{p_i}{p_k}} \mathbf{R}_{il} \mathbf{R}_{kl}^{-1} (\widehat{\mathbf{h}}_{kl} - \bar{\mathbf{h}}_{kl}). \tag{40}$$

Notice that the more similar the matrices $\mathbf{R}_{kl}$ and $\mathbf{R}_{il}$ are, the more correlated the estimates $\widehat{\mathbf{h}}_{kl}$ and $\widehat{\mathbf{h}}_{il}$ will be. Although the channels $\mathbf{h}_{kl}$ and $\mathbf{h}_{il}$ are statistically independent, the MMSE estimates will be fully correlated if the NLoS channels are spatially uncorrelated, *i.e.*, if $\mathbf{R}_{kl} = \beta_{kl}^{\mathrm{NLoS}} \mathbf{I}_N$ and $\mathbf{R}_{il} = \beta_{il}^{\mathrm{NLoS}} \mathbf{I}_N$, since they will be the same except for a scaling factor [10].

Since two closely located UEs, k and i, have approximately the same azimuth and elevation angles, they have similar spatial correlation properties, *i.e.*, the matrices $\mathbf{R}_{kl}$ and $\mathbf{R}_{il}$ can result quite similar. As a result, if they use the same pilot, the MMSE estimates of $\mathbf{h}_{kl}$ and $\mathbf{h}_{il}$ will be correlated, which was demonstrated analytically and numerically in [10] to cause strong pilot contamination.

4.2. Pilot assignment (PA) in XL-MIMO

The pilot assignment (PA) is a combinatorial problem with τ_{p}^{K} possible assignments and it is unfeasible to evaluate all of them in realistic 5G mMIMO network scenarios [10]. Hence, a practical, suboptimal PA algorithm typically assigns pilots to the UEs iteratively. The following PA techniques have been considered in this work.

Random PA: is the most naive strategy, consisting of assigning pilots randomly to all the K UEs. It does not guarantee any pilot contamination control but has zero computational complexity.

Genie PA: an exhaustive search can be used to obtain the best PA solution concerning an optimization criterion, for instance, the average channel estimation NMSE among the K UEs. The closed-form expression given in (29) can be used as a cost function for evaluating each candidate solution. The NMSE can be evaluated through the closed-form expression in (29), such that the cost function is given by

$$\sum_{k=1}^{K} \frac{\sum_{l=1}^{L} \mathrm{tr}(\mathbf{C}_{kl})}{N \sum_{l=1}^{L} \beta_{kl}} \tag{41}$$

where the term $\boldsymbol{\varepsilon}_{\mathbf{h}_{kl}}$ in (29) was removed since we consider MMSE channel estimator, which is unbiased. The genie PA procedure is shown in Algorithm 1. However, this method is feasible only if the number of possible assignments is small. It guarantees optimality, but it is unfeasible since there are $(\tau_{\mathrm{p}})^K$ possible assignments [10].

Algorithm 1 Genie Pilot Assignment

Input: $\mathbf{R}_{kl}$, $\forall kl$.

1: Initialize the population $\boldsymbol{\Theta}_{\mathrm{G}}$ containing in its columns all the A_{G} possible PA candidates.
2: **for** $i = 1, \ldots,$ **K do**
3: Using equation (29), compute $\frac{1}{N} \sum_{k=1}^{K} \mathrm{NMSE}_k$ as if the pilots were assigned according to the ith column of the matrix $\boldsymbol{\Theta}_{\mathrm{G}}$, and assign it to the ith column of the cost vector $\mathbf{c}_{\mathrm{G}}$.
4: **end for**
5: Find the pilots $t_1, \ldots, t_K$ corresponding to the column of $\boldsymbol{\Theta}_{\mathrm{G}}$ associated to the smallest element of $\mathbf{c}_{\mathrm{G}}$.

Output: pilot assignment $t_1, \ldots, t_K$

Greedy PA: is a simple PA strategy consisting in assigning the τ_{p} orthogonal pilots to τ_{p} random UEs, in the first step, and then assigning pilots to the remaining UEs, one after the other, as to minimize pilot contamination, in the second step. The pilot contamination faced by a UE k when assigning to it a pilot t can be quantified as the sum of the NLoS LSF channel gains of the UEs that were already assigned pilot t. It can be considered only the LSF channel gains to SA l_k, which is defined as the SA to which the UE k has the strongest LSF channel gain, as in Algorithm 2, or the sum LSF channel gains among the L SAs. The greedy algorithm comes with no performance guarantee but effectively avoids closely-located UEs from being assigned the same pilots [10].

Algorithm 2 Greedy Pilot Assignment

Input: LSF coefficients

1: **for** $k = 1, \ldots, \tau_{\mathrm{p}}$ **do**
2: $t_k \leftarrow k$
3: **end for**
4: **for** $k = \tau_{\mathrm{p}}, \ldots, K$ **do**
5: $l_k \leftarrow \underset{l \in \{1,\ldots,L\}}{\arg\max} \beta_{kl}$
6: $\tau \leftarrow \underset{t \in \{1,\ldots,\tau_{\mathrm{p}}\}}{\arg\min} \sum_{\substack{i=1 \\ t_i = t}}^{k-1} \beta_{il_k}^{\mathrm{NLoS}}$
7: $t_k \leftarrow \tau$
8: **end for**

Output: pilot assignment $t_1, \ldots, t_K$

Genetic Algorithm (GA) PA. Herein we propose to solve the PA problem in XL-MIMO networks by applying heuristic genetic algorithm (GA). It uses A random PA candidates per iteration. The

[1] Recall that this is a channel statistical information, which has been assumed known. Indeed, it is assumed that the algorithms can count on the perfect knowledge of the channel statistics (channel covariance matrix and the channel mean) since they change much slower than the channel realizations.

candidates are K-length vectors, where the kth entry is the index of the pilot assigned to the kth UE. Each candidate is evaluated in terms of a cost function. Here, for a fair comparison, we will take the same cost function used by the genie algorithm (exhaustive search), given by (41), *i.e.*, the average channel estimation NMSE among the K UEs.

The ϕ best candidates in the population are used to generate the A members of a new population. From the list of the best candidates, A pairs of candidates are picked at random to generate the A descendants (elements of the population in the next iteration of the algorithm). The descendants are generated by applying the crossover operator between their correspondent parents, with a random crossover point that is comprised in the interval $[2, \tau_p]$. Then, the mutation operator is applied to modify the pilot of each UE. The probability of modification is p_{mut}, which means that the probability of UE k pilot changing from t_k to another pilot is p_{mut}. Specifically, the probability of changing from t_k to each of the other τ_p is the same. At the end of each iteration, the best candidate is updated if it is better than the previous solution (at the previous iteration). The predefined number of iterations is denoted by N_{it}. The pseudo-code for the heuristic GA-based PA method is presented in Algorithm 3.

Algorithm 3 Pilot Assignment via Genetic Algorithm

Input: A, ϕ, p_{mut}, N_{it}, $\mathbf{R}_{kl}$, $\forall kl$.

Initialize the population $\boldsymbol{\Theta}$ with A random candidates.
Evaluate the cost of each candidate in $\boldsymbol{\Theta}$, according to (29), forming the vector $\mathbf{c}$.
Sort $\mathbf{c}$ in ascending order, reorganizing the columns of $\boldsymbol{\Theta}$ accordingly.
Initialize $c_{best} = c_1$, and $\boldsymbol{\theta}_{best} = \boldsymbol{\theta}_1$.
for $i = 2, \ldots, N_{it}$ **do**
  for $a = 1, \ldots, A$ **do**
    Select randomly two of the ϕ best candidates in the population $\boldsymbol{\Theta}$ to be the parents of $\boldsymbol{\theta}_a$;
    Apply the crossover operator in a random crossover point $\in \{2, K\}$.
    Apply the mutation operator in $\boldsymbol{\theta}_a$ with probability p_{mut}.
    Evaluate the cost function of $\boldsymbol{\theta}_a$, according to (29), and assign it to c_a.
  end for
  Sort $\mathbf{c}$ in descending order, reorganizing the columns of $\boldsymbol{\Theta}$ accordingly.
  if $c_1 > c_{best}$ **then**
    Update $c_{best} = c_1$ and $\boldsymbol{\theta}_{GA} = \boldsymbol{\theta}_1$.
  end if
end for

Output: $\boldsymbol{\theta}_{GA}$

After assigning the pilots, the CPU must inform each SA of the designated pilot for each UE covered by the specific SA. This is not significant in terms of fronthaul traffic since the pilot decisions change according to the channel statistics, not on the instantaneous channels basis.

The proposed GA-based PA for XL-MIMO systems is a sub-optimal algorithm since it is an iterative method searching for solutions in a very reduced search space, capable to approach the *genie* solution, which uses the entire search space. Moreover, the heuristic evolutionary GA-based pilot assignment method has the potential to be much more successful than *greedy* PA – and mainly random-based searching methods – in reducing pilot contamination since the former aggregates some intelligence in selecting potential candidates/solution beyond the simple cost function evaluation.

Table 1
Adopted simulation parameters values.

Parameter	Value
Number of subarrays: L	{25,50}
Number of antennas per subarray: N	4
Height of the antenna array: z_l^{SA}	10 m
Height of the UEs: z_l^{UE}	1.5 m
NLoS path-loss attenuation exponent: γ	4
Shadowing standard deviation: σ_s^2	3 dB (LoS) 4 dB (NLoS)
Azimuth angle standard deviation: σ_φ	10°
Elevation angle standard deviation: σ_θ	10°
Median channel gain at a distance of 1 m: β_0	$8.9125 \cdot 10^{-4}$
Array length	100 m
Wavelength: λ	12.5 cm
Separation distance between consecutive antennas in the same SA: d	$\lambda/2$
Number of antennas in each SA: N	4
LoS Probability between UE$_k$–SA$_l$: $\Pr(\alpha_{kl} = 1)$	Eq. (42)
UE transmit power: $p_1, \ldots, p_K$	10 dBm
Noise power: σ_n^2	−96 dBm
Pilot sequences' length: τ_p	4
Number of Monte-Carlo realizations	1000

4.3. Subarray selection in XL-MIMO

The SA selection is tightly connected to the pilot assignment procedure and will be done after all the UEs have been assigned to pilots. For that, in Algorithm 4, as in [10, Algorithm 4.1], each SA serves one and only one UE per pilot, namely the one with the largest LSF coefficient among the UEs using that pilot. It makes good practical sense since each SA can decide locally which UEs to serve. If the pilot assignment is performed properly, it is highly probable that every UE will be served by at least one SA.

Algorithm 4 Strongest-UE SA Selection

Input: LSF coefficients and pilot assignment $t_1, \ldots, t_K$

Initialization: Set $\mathcal{M}_1 = \cdots = \mathcal{M}_K = \emptyset$.
for $l = 1, \ldots, L$ **do**
  for $t = 1, \ldots, \tau_p$ **do**
    $k \leftarrow \underset{i \in \{1,\ldots,K\}; t_i = t}{\arg\max}\ \beta_{il}$
    $\mathcal{M}_k \leftarrow \mathcal{M}_k \cup \{l\}$
  end for
end for

Output: $\mathcal{M}_1 = \cdots = \mathcal{M}_K$

5. Numerical results

In this section, we present numerical results based on Monte Carlo simulations (MCS) to show the influence of the different proposed pilot assignment strategies and channel estimators on the estimation accuracy and the SE in XL-MIMO systems. The antenna array containing L uniformly spaced SAs is centered in the origin of the Cartesian plane and has a length of 100 m, extending along the y-axis, as depicted in Fig. 1. Each SA contains $N = 4$ antennas, spaced by a distance of half a wavelength. In each MCS realization, the K UEs are randomly dropped on a rectangular area, extending in the range [±100] m, both on the x-axis and on the y-axis. Table 1 contains the adopted values for the simulation parameters.

The GA-based PA algorithm was also configured to approach the optimal solution in minimizing the average NMSE. The stopping criterion is based on the number of iterations N_{it} and the number of candidates in the population is set as twice the number of UEs; the probability of mutation is adjusted to $p_{mut} = 0.02$

Table 2
Adopted parameters values for simulating the GA-based PA algorithm.

Parameter	Value
Number of iterations: $N_{\rm it}$	15
Number of candidates in the population: A	2 K
Probability of mutation: $p_{\rm mut}$	2%
Number of best candidates: ϕ	$\lceil A/2 \rceil$

and the number of best candidates is set to be constrained by $A/2$. The adopted parameter values in the GA-based PA algorithm are listed in Table 2. The proposed GA-based pilot allocation method is compared with three different procedures: (a) genie-PA, (b) greedy-PA, and (c) random-PA methods. SA selection step is applied equally for the four PA methods.

For the micro-urban scenario, the probability of existing LoS between UE k and SA l is a function of the distance between them and is modeled according to [22] as:

$$\Pr(\alpha_{kl} = 1) = \min\left(\frac{18\text{ m}}{d_{kl}}, 1\right) \cdot \left[1 - e^{-\frac{d_{kl}}{36\text{ m}}}\right] + e^{-\frac{d_{kl}}{36\text{ m}}}. \tag{42}$$

According to (42), the propagation is always in LoS conditions for $d_{kl} \leq 18$ m, *i.e.*, $q = 1$, while, for $d_{kl} > 18$ m, the LoS probability decays exponentially. Considering the LoS probability given by Eq. (42) and the adopted simulation parameters for array length and cell dimensions, we obtained an average LoS probability of 36%. It represents an expected condition of the XL-MIMO system operating in micro-urban channel scenarios. From the trustworthy numerical simulations perspective, we cannot disregard the LoS component in urban crowded channel scenarios since most UEs are close to the BS array, which is much longer than co-located mMIMO arrays.

5.1. Achievable spectral efficiency

We compare the channel estimation NMSE and the per-user SE obtained with the proposed GA-based PA strategy and the other three strategies mentioned in Section 4: random, greedy, and genie PA methods. According to Fig. 2, the system operation is composed by the following steps: pilot assignment, SA selection (Alg. 4), channel estimation and receive combining.

First, recall that the task of both the genie and the GA-based PA algorithms is to find the pilot assignment solution that minimizes the *average* channel estimation NMSE among the K UEs; the first does it via exhaustive search, while the second is a GA approach. Fig. 3 shows the cumulative density function (cdf) of the average channel estimation NMSE, *i.e.*, $\overline{\text{NMSE}} = \frac{1}{K}\sum_{k=1}^{K}\text{NMSE}_k$, where NMSE_k is computed via Eq. (29). The simulation setup assumes $K = 6$, $\tau_{\rm p} = 3$, $L = 25$ and $N = 4$. According to Fig. 3, simulations confirmed that genie and GA strategies actually provide the smallest average NMSE values in every channel realization; one can also verify that the GA-based PA strategy is efficient since it achieves approximately the same performance as the genie PA strategy but requires less computation (reduced search space).

Figs. 4(a) and 4(b) present, respectively, the cdf of the maximum per user channel estimation NMSE and the cdf of the minimum per user SE, among the K UEs that are simultaneously served by the XL-MIMO array, representing the worst case in each channel realization, among the K UEs. According to Fig. 4(b), the genie PA strategy, which is simply an exhaustive search for the PA solution that minimizes the average channel estimation NMSE among the K UEs, achieved the best result regarding the maximization of the minimum per user SE among the K UEs. The GA-based PA algorithm achieved almost the same performance, since this algorithm repeatedly modifies a population, selecting the best ones in terms of the same optimization criterion as

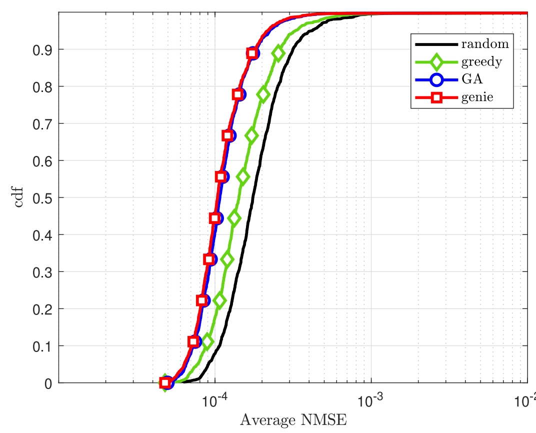


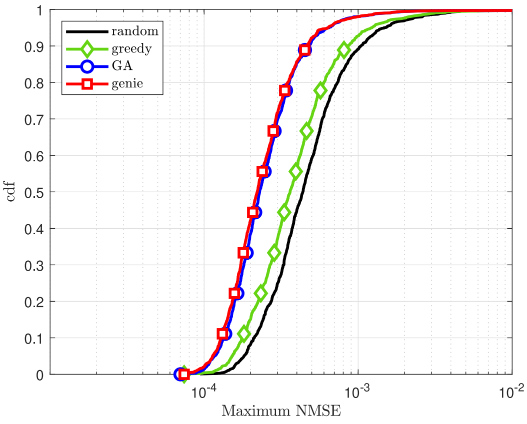
Fig. 3. Cdf of the average channel estimation NMSE, among the K UEs.

genie PA, *i.e.*, the minimization of the average channel estimation NMSE, but using a reduced search space.

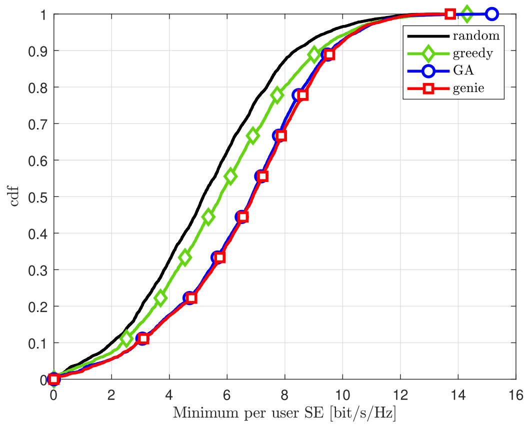
Furthermore, one can see that the proposed GA-based PA beats both random and greedy PA in all channel realizations, which means that it is improving significantly the performance of the UEs with bad channel conditions, compared with these two methods. On the other hand, the proposed method is inferior to the greedy strategy for the UEs with very good channel conditions, according to Figs. 4(c) and 4(d), which present respectively the cdf of the minimum per user channel estimation NMSE and the cdf of the maximum per user SE, among the K UEs. Finally, Figs. 4(e) and 4(f) present, respectively, the cdf of the per-user channel estimation NMSE and the cdf of the per-user SE, considering all the K UEs in each Monte-Carlo channel realization. Again, one can observe that the GA-based PA outperforms the random and greedy strategies, except for the UEs with good channel conditions. The achievements of the proposed PA strategy for XL-MIMO are in line with the requirements of beyond 5G networks, where making the quality of service more uniform is more important than improving the peak data rates. Finally, Fig. 5 shows that the proposed PA strategy also provides higher sum rates than the random and greedy strategies. The simulation setup of Figs. 4 and 5 is $K = 6$, $\tau_{\rm p} = 3$, $L = 25$ and $N = 4$.

Elaborating further on the achieved SE, one can infer that more accurate channel estimates lead to higher SEs. When computing the average of the channel estimation NMSE among a set of UEs, the worst channel UEs have much more impact on the average NMSE than the UEs with good or moderate channel condition. Thus, when pilots are suitably assigned aiming to minimize the average channel NMSE estimation, which is the case of GA-based and genie PA schemes, we are moving towards increasing mainly the SE of the worst channel UEs, which is observed in 4.

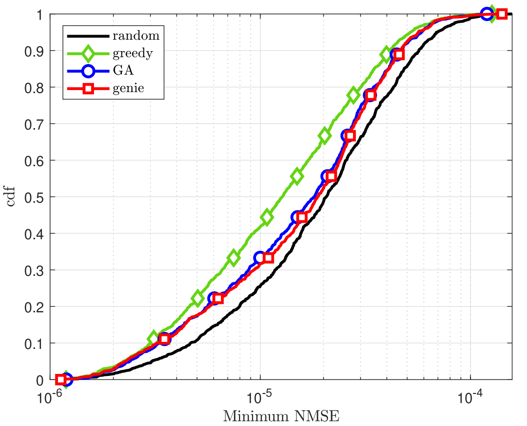
Fig. 6 presents the average values for sum SE, per user SE and minimum per user SE, varying the number of UEs, with $\tau_{\rm p} = 10$, $L = 50$ and $N = 4$. From Fig. 6(c), we confirm the effectiveness of the proposed GA scheme for improving the minimum SE among the K UEs, compared to the greedy and mainly to the random PA strategies, while it did not improve the average per user SE and the average sum SE compared to the greedy PA, only compared to the random approach. Finally, we can observe that the GA-based PA is a suitable way to make the QoS more uniform in XL-MIMO systems and that the channel estimation NMSE is promising in terms of performance-complexity trade-off to evaluate each candidate on the population in each iteration of the Algorithm 3. It reinforces the conjecture that one can make intuitively that the SINR of the active UEs can be improved considerably by reducing the channel estimate error as a consequence of lowering the pilot contamination due to an appropriate pilot assignment and subarray joint selection procedures.



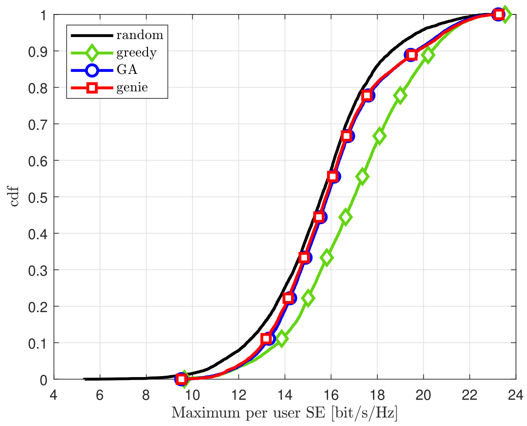
(a) Maximum channel estimation NMSE among the K UEs



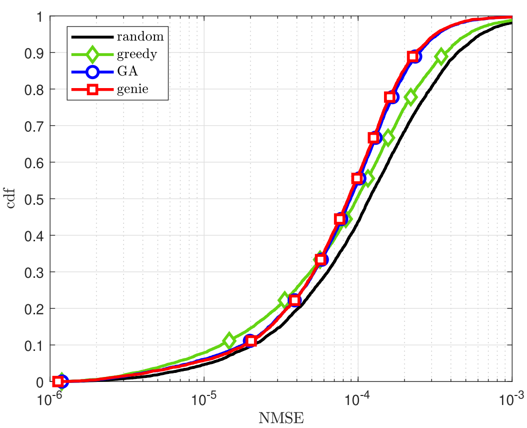
(b) Minimum per-user SE



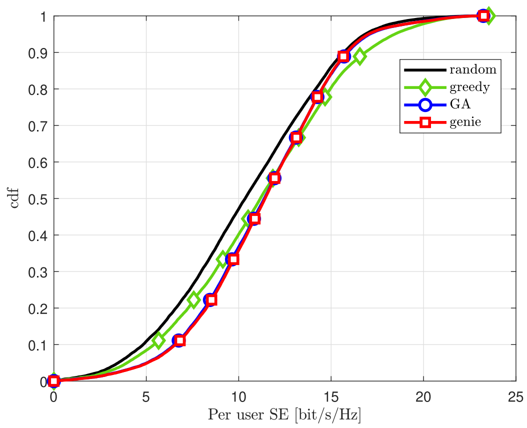
(c) Minimum channel estimation NMSE among the K UEs



(d) Maximum per-user SE



(e) Channel estimation NMSE (NMSE_k)



(f) Per-user SE

Fig. 4. Cdf of the channel estimation NMSE and the per-user SE, considering the worst UE in each Monte-Carlo channel realization (a and b), the best UE (c and d) and all the K UEs (e and f), with $K = 6$, $\tau_p = 3$, $L = 25$ and $N = 4$.

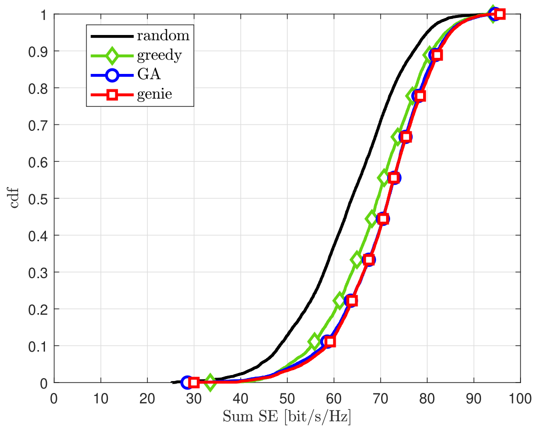
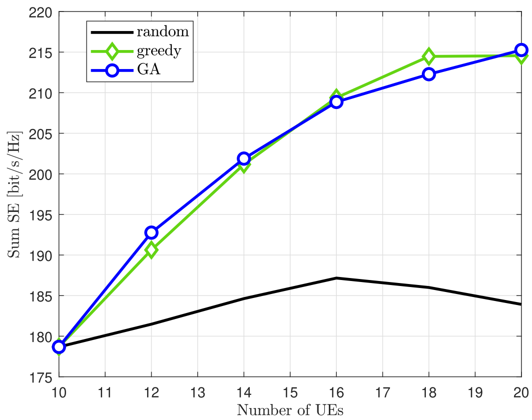


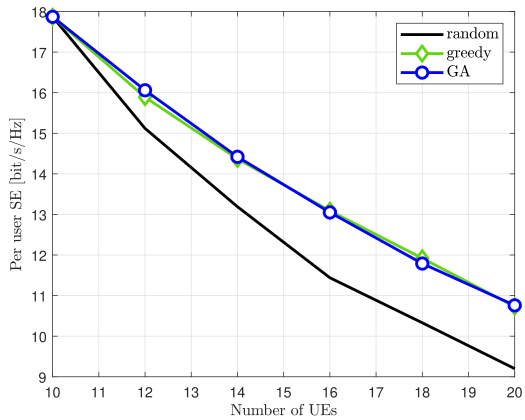
Fig. 5. Cdf of the sum SE.

6. Conclusions

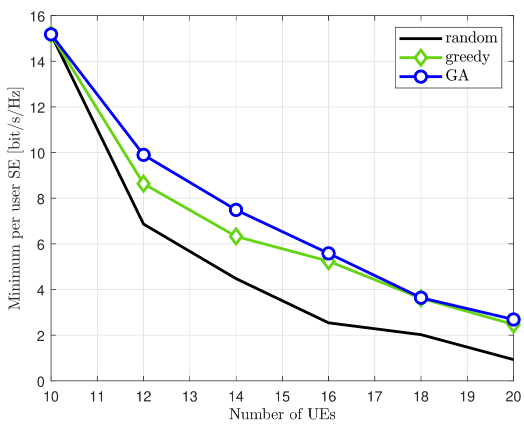
We have investigated the performance of different pilot assignment schemes to improve the QoS uniformity in XL-MIMO systems under both LoS and NLoS channel components. A proper PA strategy is paramount in crowded XL-MIMO systems, where the number of available orthogonal pilots is much smaller than the number of active UEs. Hence, a realistic XL-MIMO spatially correlated channel model has been introduced, considering that the XL-MIMO subarray antennas are not co-located, meaning that a given UE may be close to some SAs and distant from others. Consequently, in most cases, only some of the SAs are relevant to the signal detection of a given UE, such that a simple SA selection strategy based on Strongest-UE was implemented to make the system scalable without compromising the SE. The proposed GA-based PA method for SA-based XL-MIMO systems was compared with three other methods: (a) the genie PA, which is unfeasible in most practical scenarios of applications, as it is an exhaustive search for the best solution in terms of minimizing the channel estimation errors; (b) the greedy PA, which is a low complexity algorithm and was extensively deployed in the literature; (c) and the random PA, the simplest method, treated as the worst SE performance benchmark. The proposed GA-based PA strategy for XL-MIMO systems considerably improved the minimum per-user spectral efficiency compared to the greedy and random PA methods, achieving nearly-optimal results under affordable computational complexity. This is in line with the requirements of beyond 5G networks, where making the quality of service more uniform is more important than improving the peak data rates. Also, the proposed PA strategy provided higher sum rates than the random and greedy strategies. Furthermore, the proposed way for evaluating each candidate in the GA population via the channel estimation NMSE analytical expression has performed remarkably well with respect to improving the SE, especially the minimum per-user SE.



(a) Average sum SE



(b) Average per user SE



(c) Average minimum per user SE

Fig. 6. (a) Average sum SE, (b) average per user SE and (c) average minimum per user SE, as a function of the number of UEs. Parameter values: $\tau_p = 10$, $L = 50$ and $N = 4$.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Taufik Abrao reports financial support was provided by State University of Londrina. Taufik Abrao reports financial support was provided by National Council for Scientific and Technological Development.

Data availability

Data will be made available on request.

Acknowledgment

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

Appendix. Proof of Eqs. (9) and (10)

For notation simplicity, we will use $\mathbf{a}_{kl}^{(b)} = \mathbf{a}(\varphi_{kl}^{(b)}, \theta_{kl}^{(b)})$ and $\mathbf{a}_{kl}^{(0)} = \mathbf{a}(\overline{\varphi}_{kl}, \overline{\theta}_{kl})$. First of all, it is simple to find that $\mathbf{h}_{kl}^{\text{NLoS}}$, as defined in (6), is a zero mean random vector, since $g_{kl}^{(b)}$ and $\mathbf{a}_{kl}^{(b)}$ are independent and therefore uncorrelated:

$$\mathbb{E}\{\mathbf{h}_{kl}^{\text{NLoS}}\} = \sum_{b=1}^{B} \mathbb{E}\{g_{kl}^{(b)}\mathbf{a}_{kl}^{(b)}\} = \sum_{b=1}^{B} \underbrace{\mathbb{E}\{g_{kl}^{(b)}\}}_{0} \mathbb{E}\{\mathbf{a}_{kl}^{(b)}\} = 0.$$

Second, the covariance matrix of $\mathbf{h}_{kl}^{\text{NLoS}}$ is computed as

$$\begin{aligned}
\mathbf{R}_{kl} &= \mathbb{E}\{\mathbf{h}_{kl}^{\text{NLoS}}(\mathbf{h}_{kl}^{\text{NLoS}})^{\text{H}}\} - \mathbb{E}\{\mathbf{h}_{kl}^{\text{NLoS}}\}\mathbb{E}\{(\mathbf{h}_{kl}^{\text{NLoS}})^{\text{H}}\} \\
&= \mathbb{E}\{\left[\sum_{b=1}^{B} g_{kl}^{(b)}\mathbf{a}_{kl}^{(b)}\right]\left[\sum_{b=1}^{B} (g_{kl}^{(b)})^{*}(\mathbf{a}_{kl}^{(b)})^{\text{H}}\right]\} - 0\cdot 0 \\
&= \sum_{b=1}^{B} \mathbb{E}\{|g_{kl}^{(b)}|^2\}\mathbb{E}\{\mathbf{a}_{kl}^{(b)}(\mathbf{a}_{kl}^{(b)})^{\text{H}}\} + \\
&\sum_{b=1}^{B}\sum_{\substack{b'=1\\ b'\neq b}}^{B} \underbrace{\mathbb{E}\{g_{kl}^{(b)}(g_{kl}^{(b')})^{*}\}}_{0}\mathbb{E}\{\mathbf{a}_{kl}^{(b)}(\mathbf{a}_{kl}^{(b')})^{\text{H}}\},
\end{aligned} \tag{43}$$

where $\mathbb{E}\{g_{kl}^{(b)}(g_{kl}^{(b')})^{*}\} = 0$ because $g_{kl}^{(b)}$, $b = 1, \dots, B$, are i.i.d. random variables. Since the array response vector is given by $[\mathbf{a}(\varphi, \theta)]_n = e^{-j2(n-1)\pi\frac{d}{\lambda}\frac{\sin(\varphi)}{\cos(\theta)}}$, the matrix $\mathbf{a}_{kl}^{(b)}(\mathbf{a}_{kl}^{(b)})^{\text{H}}$ can be calculated as

$$[\mathbf{a}_{kl}^{(b)}(\mathbf{a}_{kl}^{(b)})^{\text{H}}]_{m,n} = \exp\left(j2(m-n)\pi\frac{d}{\lambda}\frac{\sin(\varphi_{kl}^{(b)})}{\cos(\theta_{kl}^{(b)})}\right).$$

The expectation $\mathbb{E}\{\mathbf{a}_{kl}^{(b)}(\mathbf{a}_{kl}^{(b')})^{\text{H}}\}$, over the angles $\varphi_{kl}^{(b)}$ and $\theta_{kl}^{(b)}$, is the same for $b = 1, \dots, B$, and by definition is

$$\begin{aligned}
[\mathbb{E}\{\mathbf{a}_{kl}^{(b)}(\mathbf{a}_{kl}^{(b)})^{\text{H}}\}]_{m,n} &= \iint [\mathbf{a}_{kl}^{(b)}(\mathbf{a}_{kl}^{(b)})^{\text{H}}]_{m,n} f_{kl}(\varphi, \theta) d\varphi d\theta \\
&= \iint e^{j2(m-n)\pi\frac{d}{\lambda}\frac{\sin(\varphi)}{\cos(\theta)}} f_{kl}(\varphi, \theta) d\varphi d\theta.
\end{aligned}$$

Thus, this term can go out of the summary in (43). Finally, recalling that $\beta_{kl}^{\text{NLoS}} = \sum_{b=1}^{B} \mathbb{E}\{|g_{kl}^{(b)}|^2\}$:

$$[\mathbf{R}_{kl}]_{m,n} = \beta_{kl}^{\text{NLoS}} \iint e^{j2(m-n)\pi\frac{d}{\lambda}\frac{\sin(\varphi)}{\cos(\theta)}} f_{kl}(\varphi, \theta) d\varphi d\theta.$$

References

[1] A. Amiri, C.N. Manchón, E. de Carvalho, A message passing based receiver for extra-large scale MIMO, in: 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP, 2019, pp. 564–568, http://dx.doi.org/10.1109/CAMSAP45676.2019.9022503.

[2] O.S. Nishimura, J.C. Marinello, T. Abrão, A grant-based random access protocol in extra-large massive mimo system, IEEE Commun. Lett. 24 (11) (2020) 2478–2482, http://dx.doi.org/10.1109/LCOMM.2020.3012586.

[3] J.C. Marinello, T. Abrão, A. Amiri, E. de Carvalho, P. Popovski, Antenna selection for improving energy efficiency in XL-MIMO systems, IEEE Trans. Veh. Technol. 69 (11) (2020) 13305–13318, http://dx.doi.org/10.1109/TVT.2020.3022708.

[4] X. Yang, F. Cao, M. Matthaiou, S. Jin, On the uplink transmission of extra-large scale massive MIMO systems, IEEE Trans. Veh. Technol. 69 (12) (2020) 15229–15243, http://dx.doi.org/10.1109/TVT.2020.3037317.

[5] H. Wang, A. Kosasih, C. Wen, S. Jin, W. Hardjawana, Expectation propagation detector for extra-large scale massive MIMO, IEEE Trans. Wireless Commun. 19 (3) (2020) 2036–2051.

[6] A. Amiri, C.N. Manchón, E. de Carvalho, Uncoordinated and decentralized processing in extra-large MIMO arrays, IEEE Wirel. Commun. Lett. 11 (1) (2022) 81–85, http://dx.doi.org/10.1109/LWC.2021.3120694.

[7] 3GPP, Technical Specification Group Radio Access Network; Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (Release 16), Technical Specification(TS) 25.996, 3rd Generation Partnership Project (3GPP), 2020, version 16.0.0. [Online]. Available: URL https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1382.

[8] S. Mukherjee, R. Chopra, Performance analysis of cell-free massive MIMO systems in los/ nlos channels, IEEE Trans. Veh. Technol. 71 (6) (2022) 6410–6423, http://dx.doi.org/10.1109/TVT.2022.3161702.

[9] S. Elhoushy, M. Ibrahim, W. Hamouda, Cell-free massive MIMO: A survey, IEEE Commun. Surv. Tutor. 24 (1) (2022) 492–523, http://dx.doi.org/10.1109/COMST.2021.3123267.

[10] Ö.T. Demir, E. Björnson, L. Sanguinetti, Foundations of user-centric cell-free massive mimo, 2021.

[11] X. Zhu, Z. Wang, L. Dai, C. Qian, Smart pilot assignment for massive MIMO, IEEE Commun. Lett. 19 (9) (2015) 1644–1647.

[12] Y. Omid, S.M. Hosseini, S.M. Shahabi, M. Shikh-Bahaei, A. Nallanathan, Aoa-based pilot assignment in massive MIMO systems using deep reinforcement learning, IEEE Commun. Lett. 25 (9) (2021) 2948–2952, http://dx.doi.org/10.1109/LCOMM.2021.3089234.

[13] J.C.M. Filho, G. Brante, R.D. Souza, T. Abrão, Exploring the non-Overlapping Visibility Regions in XL-MIMO random access and scheduling, IEEE Trans. Wireless Commun. (2022) 1, http://dx.doi.org/10.1109/TWC.2022.3151329.

[14] M. Cui, L. Dai, Channel estimation for extremely large-scale MIMO: Far-field or near-field? IEEE Trans. Commun. 70 (4) (2022) 2663–2677.

[15] O.S. Nishimura, J.C.M. Filho, T. Abrão, R.D. Souza, Fairness in a class barring power control random access protocol for crowded XL-MIMO systems, IEEE Syst. J. (2022) 1–9, http://dx.doi.org/10.1109/JSYST.2022.3156696.

[16] J.D. Kraus, Antennas for All Applications, McGraw Hill, 2002.

[17] L. Sanguinetti, E. Björnson, J. Hoydis, Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination, IEEE Trans. Commun. 68 (1) (2020) 232–257, http://dx.doi.org/10.1109/TCOMM.2019.2945792.

[18] E. Björnson, J. Hoydis, L. Sanguinetti, Massive MIMO networks: Spectral, energy, and hardware efficiency, Found. Trends® Signal Process. 11 (3–4) (2017) 154–655, http://dx.doi.org/10.1561/2000000093.

[19] S. Lu, J. May, R.J. Haines, Effects of correlated shadowing modeling on performance evaluation of wireless sensor networks, in: 2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall), 2015, pp. 1–5, http://dx.doi.org/10.1109/VTCFall.2015.7390909.

[20] S. Lu, J. May, R.J. Haines, Efficient modeling of correlated shadow fading in dense wireless multi-hop networks, in: 2014 IEEE Wireless Communications and Networking Conference, WCNC, 2014, pp. 311–316, http://dx.doi.org/10.1109/WCNC.2014.6951986.

[21] O. Ozdogan, E. Bjornson, E.G. Larsson, Massive MIMO with spatially correlated rician fading channels, IEEE Trans. Commun. 67 (5) (2019) 3234–3250, http://dx.doi.org/10.1109/TCOMM.2019.2893221.

[22] 3GPP, Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-ULTRA); Further advancements for E-ULTRA physical layer aspects (Release 9), Technical Specification (TS) 36.814, 3rd Generation Partnership Project (3GPP), 2017, version 9.2.0. [Online]. Available: URL https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2493.

Gabriel Avanzi Ubiali received his B.S. and M.S. degree in Electrical Engineering from Londrina State University, PR, Brazil, in Apr. 2018 and Feb. 2021, respectively. His research interests include signal processing and wireless communications, especially massive MIMO, XL-MIMO, and cell-free massive MIMO networks, acquisition of channel-state information, interference management in 5G, massive machine-type communications, RIS-aided communication, resource allocation and random access protocols for crowded networks.

Taufik Abrão (SM'12, SM-SBrT) received the B.S., M.Sc., and Ph.D. degrees in electrical engineering from the Polytechnic School of the 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 Telecomm. & Signal Processing Lab. In 2018, he was with the Connectivity section, at Aalborg University as a Guest Researcher, and with the Southampton Wireless Research Group in 2012 as an Academic Visitor. He has served as Associate Editor for the IEEE Transactions on Vehicular Technology, the IEEE Systems Journal, the IEEE Access, the IEEE Communication Surveys & Tutorials, the AEUe-Elsevier, the IET Signal Processing, and JCIS-SBrT, and as Executive Editor of the ETT-Wiley (2016-2021) journal. His current research interests include communications systems, especially mMIMO, XL-MIMO, RIS-aided communication, optimization methods, machine learning, scheduling, estimation, resource allocation, and random access protocols.

José Carlos Marinello Filho received his B.S. and M.S. degree in Electrical Engineering (the first with Summa Cum Laude) from Londrina State University, PR, Brazil, in Dec. 2012 and Sept. 2014, respectively, and his Ph.D. from Polytechnic School of São Paulo University, São Paulo, Brazil, in Aug. 2018. From 2015 to 2019, he was an Assistant Professor at Londrina State University, and, since 2019, he has been an Adjunct Professor at the Federal University of Technology PR, Cornélio Procópio, Brazil. His research interests include signal processing and wireless communications, especially massive MIMO/XL-MIMO precoding/decoding techniques, acquisition of channel-state information, multicarrier modulation, cross-layer optimization of MIMO/OFDM systems, interference management in 5G, massive machine-type communications, and random access protocols for crowded networks. He has been serving as Associate Editor for the SBrT Journal of Communications and Information Systems since 2019.

APPENDIX B – PAPER B

- **Title:** User Scheduling in Multi-State Los/NLoS Crowded XL-MIMO Channels
- **Authors:** Gabriel Avanzi Ubiali, Taufik Abrão and José Carlos Marinello Filho
- **Conference:** 2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM)
- **DOI:** 10.1109/WINCOM59760.2023.10322996
- **Citation:** G. A. Ubiali, T. Abrão and J. C. Marinello, "User Scheduling in Multi-State Los/NLoS Crowded XL-MIMO Channels," *2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Istanbul, Turkiye, 2023, pp. 1-6, doi: 10.1109/WINCOM59760.2023.10322996.

2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM) | 979-8-3503-2967-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/WINCOM59760.2023.10322996

User Scheduling in Multi-State Los/NLoS Crowded XL-MIMO Channels

Gabriel Avanzi Ubiali
Department of Electrical Engineering
State University of Londrina (UEL).
86057-970, Londrina, PR, Brazil.
gabriel.ubiali@uel.br

Taufik Abrão
Department of Electrical Engineering
State University of Londrina (UEL).
86057-970, Londrina, PR, Brazil.
taufik@uel.br

José Carlos Marinello
Department of Electrical Engineering
Federal University of Technology, PR.
86300-000, Cornélio Procópio, PR, Brazil.
jcmarinello@utfpr.edu.br

***Abstract*—In this paper, we have derived a closed-form expression for the signal-to-interference-and-noise ratio (SINR) in extra-large multiple-input multiple-output (XL-MIMO) systems with spatially correlated multi-state line-of-sight (LoS) / non-LoS (NLoS) channels, which is valid for a sufficiently large number of base station (BS) antennas, when least-squares (LS) channel estimation is used and under the absence of pilot reuse among the scheduled user equipments (UEs). The derived SINR expression can be computed based on the channel statistics instead of the instantaneous channel realizations, which makes it valid for hundreds of channel coherence blocks. We also have proposed an expeditious user scheduling (US) algorithm for crowded XL-MIMO systems, which aims to maximize the sum spectral efficiency (SE) with the constraint of assigning mutually orthogonal pilots to all the scheduled UEs, such that their channel estimation accuracy is very high. The numerical results reveal that the proposed SINR expression is very accurate for a relatively large number of BS antennas, and that the proposed US algorithm improves considerably the sum SE in crowded scenarios.**

I. Introduction

Conventional massive multiple-input multiple-output (M-MIMO) base stations (BSs) are made of colocated antenna arrays. On the other hand, extra-large scale MIMO (XL-MIMO) corresponds to an antenna array deployment with extreme physical dimensions, which can be integrated into walls and columns of buildings [1]. XL-MIMO is a promising technology to address crowded communication scenarios and the requirement of a more uniform quality-of-service (QoS) across the coverage area, since it provides moderate to high spatial macro-diversity and reduces the average distance from a user equipment (UE) to the antenna that is closest to it, compared to conventional M-MIMO [2].

In crowded scenarios, the expenditure of resource blocks must be optimized, to assure that a higher number of UEs are served with a target QoS [3]. This work proposes a procedure to maximize the sum spectral efficiency (SE) while serving as many UEs as possible in a crowded XL-MIMO scenario. This is done by selecting which users will be served by the XL-MIMO BS aiming to maximize the SE, by assigning mutually orthogonal pilot sequences to the scheduled UEs, while achieving almost zero pilot contamination. The US algorithm bases its decision on an expeditious SINR closed-form expression, which is derived in this paper and is valid for XL-MIMO channels. We adopt a robust and realistic channel model for XL-MIMO systems since it considers multi-state line-of-sight (LoS) / non-LoS (NLoS) channels and spatial correlation, as in [4].

User scheduling (US) in XL-MIMO has already been addressed in [3], [5], [6], but the present work is the first one to consider at the same time spatially correlated channels and the unavailability of the channel estimates, such that we propose a US strategy that relies on the channel statistics rather than the instantaneous channels.

In [7]–[10], the ergodic SINR expression under maximum-ratio (MR) processing in cellular massive MIMO or XL-MIMO, with uncorrelated Rayleigh fading, is explored considering the NLoS-only channel model. However, most practical propagation scenarios include both LoS and NLoS channel components, such that we must consider both components for simulation results to be reliable. Furthermore, the performance of MR processing is poor under scenarios with multi-user interference.

The *contribution* of this paper is threefold.

1) We derive an SINR closed-form expression for XL-MIMO networks, which is valid for a sufficiently large number of BS antennas, zero-forcing (ZF) receive combiner, least-squares (LS) channel estimation and under the absence of pilot reuse among the scheduled UEs. Although the expression depends only on the mean channel vectors and covariance matrices, *i.e.*, on the channel statistics, we numerically corroborate that the approximation error is negligible for massive and extra-large MIMO conditions.
2) We propose a US algorithm for maximizing the sum SE. It primarily selects the UEs with the highest large-scale fading (LSF) channel gains, and bases its decision on the derived SINR expression, which depends only on the channel statistics, such that the US decisions remain valid for hundreds of coherence blocks.
3) To corroborate the efficiency and efficacy of the proposed US procedure, we consistently provide and analyze numerical results regarding the system SE and the accuracy of the SINR expression.

The remainder of this work is organized in five sections. Section II deals with the adopted multi-state LoS/NLoS channel, as well as the channel estimation and data transmission steps, for XL-MIMO systems. Section III describes the proposed method for US, while Section IV derives the SINR closed-form expression. Finally, numerical results and conclusions are presented in sections V and VI.

II. XL-MIMO System Model

We consider an over-crowded scenario with time-division-duplexing (TDD) operation where a BS equipped with an XL-MIMO linear antenna array (ULA) serves K single-antenna users. The ULA is equipped with M antennas, which are grouped in L subarrays (SAs), each one containing N antennas separated by a distance $d = \lambda/2$, where λ is the wavelength. The SAs are connected to a central processing unit (CPU), which is responsible for the SA cooperation.

A. XL-MIMO Channel Model

The channel vector between the kth UE and the lth SA can be expressed as

$$\mathbf{h}_{kl} = \alpha_{kl}\bar{\mathbf{h}}_{kl} + \check{\mathbf{h}}_{kl}. \tag{1}$$

where $\bar{\mathbf{h}}_{kl} \in \mathbb{C}^N$ and $\check{\mathbf{h}}_{kl} \in \mathbb{C}^N$ denote respectively the LoS and the NLoS components. The Bernoulli random variable α_{kl} indicates the channel state: it assumes the value 1 or 0, respectively, if there is a LoS component or only NLoS propagation. If the SA has a small aperture, it will experience spatial stationarity, which means that all its N antennas wil have the same channel state. On the other hand, spatial non-stationarity is assumed among different SAs. The LoS channel vector is given by [11]:

$$\bar{\mathbf{h}}_{kl} = \sqrt{\bar{\beta}_{kl}}e^{-j2\pi\frac{d_{kl}}{\lambda}}\mathbf{a}(\bar{\varphi}_{kl}, \bar{\theta}_{kl}), \tag{2}$$

where d_{kl} stands for the distance from the kth UE to the lth SA, and $\bar{\beta}_{kl}$, $\bar{\varphi}_{kl}$ and $\bar{\theta}_{kl}$ are respectively the LSF channel gain and the azimuth and elevation angles w.r.t. the LoS component. Since the NLoS channel component is zero mean, the LoS component corresponds to the channel mean.

The LoS LSF channel gain is determined as

$$\bar{\beta}_{kl} = \frac{1}{N}||\bar{\mathbf{h}}_{kl}||^2 = \alpha_{kl}\frac{\beta_0}{d_{kl}^2}\bar{\mathcal{X}}_{kl}, \tag{3}$$

where β_0 is the average channel gain at the reference distance of 1 m, while β_0/d_{kl}^2 and $\bar{\mathcal{X}}_{kl}$ are respectively the path-loss and shadow fading of the LoS component.

Assuming a small SA aperture, the UEs are very likely to be in the far-field propagation from a SA perspective, such that it is reasonable to model the channel vector in (2) assuming planar wavefront and assuming the same LSF channel gains for the N antennas. On the other hand, the signal arrives at each antenna with different phases. For elevation and azimuth angles θ and φ, these phase differences are accounted by the array response vector $\mathbf{a}(\varphi, \theta)$, whose nth entry is

$$[\mathbf{a}(\varphi, \theta)]_n = e^{-j2(n-1)\pi\frac{d}{\lambda}\sin(\varphi)/\cos(\theta)}. \tag{4}$$

The NLoS channel vector is the superposition of B propagation paths, approaching a Gaussian distribution [12]:

$$\check{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_{kl}). \tag{5}$$

The elements of the *spatial correlation matrix* $\mathbf{R}_{kl}$ are given by:

$$[\mathbf{R}_{kl}]_{m,n} = \check{\beta}_{kl}\int\int e^{j2(m-n)\pi\frac{d}{\lambda}\frac{\sin(\varphi)}{\cos(\theta)}} f_{kl}(\varphi, \theta)d\varphi d\theta. \tag{6}$$

The joint probability density function of φ and θ is given by [2]:

$$f_{kl}(\varphi, \theta) = \frac{1}{2\pi\sigma_\varphi\sigma_\theta}\exp\left\{-\frac{(\varphi - \bar{\varphi}_{kl})^2}{2\sigma_\varphi^2} - \frac{(\theta - \bar{\theta}_{kl})^2}{2\sigma_\theta^2}\right\}. \tag{7}$$

being $\sigma_\varphi \geq 0$ and $\sigma_\theta \geq 0$ the angular standard deviations. The angles φ and θ are integrated in the interval $[-180°, 180°]$ and $(-90°, 90°)$, respectively. The proof of (5) and (6) can be found in [4, Appendix A].

Fig. 1 illustrates the LoS + NLoS propagation in XL-MIMO systems. The NLoS components are generated by a scattering cluster. The distance d_{kl} and the LoS azimuth angle $\bar{\varphi}_{kl}$ are assumed the same for all the N antennas. Notice that the UE is in the near-field (spherical wavefront) with respect to the whole antenna array and in the far-field (planar wavefront) w.r.t. each SA. For simplicity, only the azimuth plane and a pair of multipath components are shown in the figure.

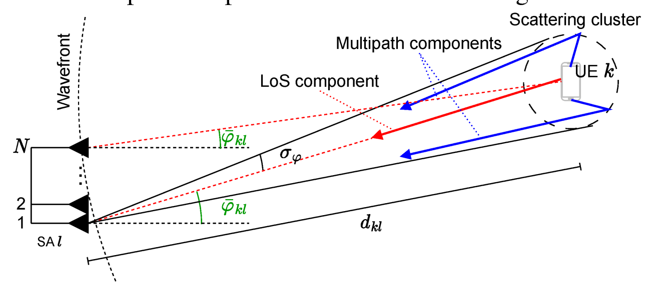


Figure 1: LoS + NLoS propagation.

We can obtain the NLoS LSF coefficient as

$$\check{\beta}_{kl} = \frac{1}{N}\mathbb{E}\{||\check{\mathbf{h}}_{kl}||^2\} = \frac{1}{N}\text{tr}(\mathbf{R}_{kl}) = \frac{\beta_0}{d_{kl}^\gamma}\check{\mathcal{X}}_{kl}, \tag{8}$$

where γ and $\check{\mathcal{X}}_{kl}$ are the path-loss exponent and the shadow fading, both regarding to the NLoS component. We model the LoS and NLoS *shadow fading* as log-normal random variables, such that $\bar{F}_{kl}, \check{F}_{kl} \sim \mathcal{N}(0, \sigma_{\text{SF}}^2)$, where $\bar{F}_{kl} = 10\ \log_{10}(\bar{\mathcal{X}}_{kl})$ and $\check{F}_{kl} = 10\ \log_{10}(\check{\mathcal{X}}_{kl})$. For simplicity, the shadow fading related to different UEs and/or different SAs are assumed independently distributed. Cross-correlated shadowing is considered in [4].

Finally, the kth UE channel vector is distributed as

$$\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\bar{\mathbf{h}}_k, \mathbf{R}_k), \tag{9}$$

and the LSF channel gain of the kth UE is given by

$$\beta_{kl} = \frac{1}{N}\mathbb{E}\{||\mathbf{h}_{kl}||^2\} = \frac{1}{N}\text{tr}(\mathbf{Q}_{kl}) = \bar{\beta}_{kl} + \check{\beta}_{kl}, \tag{10}$$

where $\mathbf{Q}_{kl} = \bar{\mathbf{h}}_{kl}\bar{\mathbf{h}}_{kl}^{\mathrm{H}} + \mathbf{R}_{kl}$, $\mathbf{R}_k = \mathrm{diag}(\mathbf{R}_{k1}, \dots, \mathbf{R}_{kL})$, $\mathbf{h}_k = [\mathbf{h}_{k1}^{\mathsf{T}}, \dots, \mathbf{h}_{kL}^{\mathsf{T}}]^{\mathsf{T}}$ and $\bar{\mathbf{h}}_k = [\bar{\mathbf{h}}_{k1}^{\mathsf{T}}, \dots, \bar{\mathbf{h}}_{kL}^{\mathsf{T}}]^{\mathsf{T}}$.

Lastly, defining $\mathbf{Q}_k = \bar{\mathbf{h}}_k\bar{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{R}_k$, the average LSF channel gain between the ULA and the kth UE is

$$\beta_k = \frac{1}{NL}\mathrm{tr}(\mathbf{Q}_k) = \frac{1}{NL}\sum_{l=1}^{L}\mathrm{tr}(\mathbf{Q}_{kl}) = \frac{1}{L}\sum_{l=1}^{L}\beta_{kl}. \tag{11}$$

B. Channel Estimation and Uplink Data Transmission

The channels must be estimated once per coherence block. There are τ_{c} symbols in each coherence block, but τ_{p} out of them are reserved for uplink (UL) pilot signaling, such that $\tau_{\mathrm{c}} - \tau_{\mathrm{p}}$ symbols are used for UL data transmission, where τ_{p} is the length of the pilot sequences.

The US method we propose assigns mutually orthogonal pilots to all the scheduled UEs, such that $||\boldsymbol{\phi}_i||^2 = \tau_{\mathrm{p}}$ and $\boldsymbol{\phi}_i^{\mathrm{H}}\boldsymbol{\phi}_j = 0$ for $i \neq j$, where $\boldsymbol{\phi}_i \in \mathbb{C}^{\tau_{\mathrm{p}}}$ stands for the pilot sequence assigned to the ith UE. At the beginning of the coherence block, UE k transmits the elements of $\sqrt{p_k}\boldsymbol{\phi}_k$ over τ_{p} consecutive samples, with power p_k. Thus, during the pilot transmission step, the BS receives the signal

$$\mathbf{Y}^{\mathrm{p}} = \sum_{i\in\mathcal{S}}\sqrt{p_i}\mathbf{h}_i\boldsymbol{\phi}_i^{\mathsf{T}} + \mathbf{N}^{\mathrm{p}}, \tag{12}$$

where $\mathcal{S} \subseteq \{1, \dots, K\}$ is the set of scheduled UEs, $\mathbf{N}^{\mathrm{p}} \in \mathbb{C}^{M\times\tau_{\mathrm{p}}}$ is the noise matrix with i.i.d. elements distributed as $\mathcal{N}_{\mathbb{C}}(0, \sigma_{\mathrm{n}}^2)$ and σ_{n}^2 stands for the noise power.

The channel vector is then estimated by multiplying $\mathbf{Y}^{\mathrm{p}}$ with the normalized conjugate $\boldsymbol{\phi}_k$:

$$\widehat{\mathbf{h}}_k = \frac{1}{\sqrt{p_k}\tau_{\mathrm{p}}}\mathbf{Y}^{\mathrm{p}}\boldsymbol{\phi}_k^* = \mathbf{h}_k + \frac{1}{\sqrt{p_k}\tau_{\mathrm{p}}}\mathbf{n}_k^{\mathrm{p}}, \tag{13}$$

where $\mathbf{n}_k^{\mathrm{p}} = \mathbf{N}^{\mathrm{p}}\boldsymbol{\phi}_k^* \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tau_{\mathrm{p}}\sigma_{\mathrm{n}}^2\mathbf{I}_M)$. The covariance matrix of the (zero mean) channel estimation error, $\tilde{\mathbf{h}}_k = \mathbf{h}_k - \widehat{\mathbf{h}}_k$, is:

$$\mathbf{C}_{kl} = \frac{\sigma_{\mathrm{n}}^2}{p_k\tau_{\mathrm{p}}}\mathbf{I}_M. \tag{14}$$

Suppose that UE k is scheduled and transmits the signal $x_k \sim \mathcal{N}_{\mathbb{C}}(0, p_k)$, during the uplink data transmission stage, with transmit power p_k. Then, the BS receives the signal

$$\mathbf{y}^{\mathrm{ul}} = \sum_{i\in\mathcal{S}}\mathbf{h}_i x_i + \mathbf{n}, \tag{15}$$

where $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{\mathrm{n}}^2\mathbf{I}_{LN})$ is the collective noise vector. If UE k is scheduled, the CPU will estimate x_k by computing the inner product between its receive combining vector, denoted by $\mathbf{v}_k$, and the received UL signal, $\mathbf{y}^{\mathrm{ul}}$, as follows:

$$\begin{aligned}\widehat{x}_k &= \mathbf{v}_k^{\mathrm{H}}\mathbf{y}^{\mathrm{ul}} \\ &= \mathbf{v}_k^{\mathrm{H}}\widehat{\mathbf{h}}_k x_k + \sum_{i\in\mathcal{S}\setminus\{k\}}\mathbf{v}_k^{\mathrm{H}}\widehat{\mathbf{h}}_i x_i + \sum_{i\in\mathcal{S}}\mathbf{v}_k^{\mathrm{H}}\tilde{\mathbf{h}}_i x_i + \mathbf{v}_k^{\mathrm{H}}\mathbf{n}. \end{aligned} \tag{16}$$

When applying ZF combiner, the first and second terms in (16) will be x_k and 0, respectively. Thus, if UE k is scheduled, its instantaneous UL SINR is:

$$\Gamma_{k,\mathcal{S}} = \frac{p_k}{\sum\limits_{i\in\mathcal{S}} p_i|\mathbf{v}_k^{\mathrm{H}}\tilde{\mathbf{h}}_i|^2 + \sigma_{\mathrm{n}}^2||\mathbf{v}_k||^2}, \tag{17}$$

which depends on the set $\mathcal{S}$. The sum SE over the scheduled users is computed as

$$\mathrm{SE}_{\mathcal{S}} = \left(1 - \frac{|\mathcal{S}|}{\tau_{\mathrm{c}}}\right)\sum_{k\in\mathcal{S}}\log_2(1 + \Gamma_{k,\mathcal{S}}). \tag{18}$$

III. User scheduling for XL-MIMO

We want to find the set $\mathcal{S}$ of scheduled UEs that maximizes the sum SE in an XL-MIMO system under the condition that all UEs in $\mathcal{S}$ are assigned mutually orthogonal pilot sequences. It corresponds to solving the optimization problem below:

$$\begin{aligned} \max_{\mathcal{S}} \quad & \mathrm{SE}_{\mathcal{S}} && \text{(19a)} \\ \text{s.t.} \quad & \tau_{\mathrm{p}} = |\mathcal{S}|, && \text{(19b)} \\ & ||\boldsymbol{\phi}_i||^2 = \tau_{\mathrm{p}}, \quad \forall i \in \mathcal{S}, && \text{(19c)} \\ & \boldsymbol{\phi}_i^{\mathrm{H}}\boldsymbol{\phi}_j = 0, \quad \forall i, j \in \mathcal{S}, && \text{(19d)} \\ & \mathcal{S} \subseteq \{1, \dots, K\}. && \text{(19e)} \end{aligned}$$

By solving (19), we implicitly optimize the length of the pilot sequences since it is equal to the number of scheduled UEs. Therefore, by increasing $|\mathcal{S}|$, we are also increasing τ_{p}, and the summation in (18) increases, whereas the number of samples available for data transmission in each coherence block decreases. In other words, there is a value for the cardinality $|\mathcal{S}| = \tau_{\mathrm{p}}$ which maximizes $\mathrm{SE}_{\mathcal{S}}$.

The proposed heuristic algorithm tries to solve (19): it maximizes the sum SE by admitting the UEs with the strongest LSF channel gains, one by one, and checking the SE after each admission. For that, it relies on the availability of a closed-form, expeditious equivalent expression to compute the sum SE repeatedly. The heuristic procedure stops admitting more UEs before it starts to decrease. Algorithm 1 describes the operations of the proposed US method.

Algorithm 1 User scheduling for XL-MIMO

Input: $\beta_k, \mathbf{Q}_k$, $k = 1, \dots, K$.

Initialize the set of scheduled UEs: $\mathcal{S} \leftarrow \emptyset$.
Initialize the approximated system SE: $\mathrm{SE}^{(0)} \leftarrow 0$.
Initialize: $\tau_{\mathrm{p}} \leftarrow 0$.
while $\tau_{\mathrm{p}} \leq K$ **do**
  Select a UE to be scheduled:

$$\begin{aligned} k \leftarrow \arg\max_{i} \quad & \beta_i && \text{(20a)} \\ \text{s.t.} \quad & i \in \{1, \dots, K\}\setminus\mathcal{S} && \text{(20b)} \end{aligned}$$

  Update: $\mathcal{S} \leftarrow \mathcal{S} \cup \{k\}$ and $\tau_{\mathrm{p}} \leftarrow \tau_{\mathrm{p}} + 1$.
  Compute the current approximated SINR for the scheduled UEs, using (34), and $\mathrm{SE}^{(\tau_{\mathrm{p}})}$, using (35).
  if $\mathrm{SE}^{(\tau_{\mathrm{p}})} < \mathrm{SE}^{(\tau_{\mathrm{p}}-1)}$ **then**
    Undo the last update: $\mathcal{S} \leftarrow \mathcal{S}\setminus\{k\}$ and $\tau_{\mathrm{p}} \leftarrow \tau_{\mathrm{p}} - 1$.
    Exit loop.
  end if
end while

Output: scheduling set: $\mathcal{S}$

Instead of being based on the instantaneous channels, the proposed closed-form expressions for the SINR and the SE are based on the channel statistics (mean vector and covariance matrix), which we assume to be perfectly known since they only change considerably after hundreds of channel coherence blocks. Thus, each US solution remains valid for hundreds of coherence blocks, requiring a small amount of computational resources.

An extension of Alg. 1 aiming to increase the number of served UEs is devised as follows. Maintaining the number of orthogonal pilot sequences obtained with Alg. 1, the UEs that were not scheduled can be admitted by being assigned a pilot that is already used. A possible criterion to minimize pilot contamination is to assign to a UE $k \notin \mathcal{S}$ the pilot that least compromises the sum SE. Here, we omit this step due to lack of space.

IV. Low-complexity SINR Computation

Considering the XL-MIMO channel model described in Section II, we derive herein a closed-form expression for the SINR, which is consistent for a sufficiently large number of BS antennas, under the absence of pilot reuse among the UEs in $\mathcal{S}$, and when LS channel estimation is employed.

Let us define the matrix $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_{|\mathcal{S}|}] = [\mathbf{h}_{s_1}, \ldots, \mathbf{h}_{s_{|\mathcal{S}|}}]$ containing only the columns of $\mathbf{H}$ that corresponds to the channel vectors of the UEs in $\mathcal{S}$, in descending order of LSF channel gain, such that UE s_m has the m-th highest LSF channel gain. Then, the m-th column of $\mathbf{G}$ contains the channel vector of the s_m-th UE: $\mathbf{g}_m = \mathbf{h}_{s_m}$. Similarly, the matrix $\widehat{\mathbf{G}} = [\widehat{\mathbf{g}}_1, \ldots, \widehat{\mathbf{g}}_{|\mathcal{S}|}] = [\widehat{\mathbf{h}}_{s_1}, \ldots, \widehat{\mathbf{h}}_{s_{|\mathcal{S}|}}]$ contains the estimated channels of the UEs in $\mathcal{S}$, and the ZF combining matrix, $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_{|\mathcal{S}|}] = [\mathbf{v}_{s_1}, \ldots, \mathbf{v}_{s_{|\mathcal{S}|}}]$, can be computed as $\mathbf{W} = \widehat{\mathbf{G}}(\widehat{\mathbf{G}}^{\mathrm{H}}\widehat{\mathbf{G}})^{-1}$. In the opposite way, we say that d_k is the index of the column of $\mathbf{G}$ that corresponds to $\mathbf{h}_k$, *i.e.*, $\mathbf{h}_k = \mathbf{g}_{d_k}$. Consequently, $\mathbf{v}_k = \mathbf{w}_{d_k}$.

The first term in the denominator of (17) approximates to

$$|\mathbf{v}_k^{\mathrm{H}}\tilde{\mathbf{h}}_i|^2 = \tilde{\mathbf{h}}_i\mathbf{v}_k\mathbf{v}_k^{\mathrm{H}}\tilde{\mathbf{h}}_i^{\mathrm{H}} \overset{M\to\infty}{\longrightarrow} \mathrm{tr}(\mathbf{v}_k\mathbf{v}_k^{\mathrm{H}}\mathbf{C}_i) = \frac{\sigma_{\mathrm{n}}^2}{p_i\tau_{\mathrm{p}}}||\mathbf{v}_k||^2, \quad (21)$$

when the number of antennas is large, due to [13, Lemma 4], and defining $\mathbf{C}_i = \mathrm{diag}(\mathbf{C}_{i1}, \ldots, \mathbf{C}_{iL})$. Then,

$$\Gamma_{k,\mathcal{S}} \overset{M\to\infty}{\longrightarrow} \frac{p_k}{2\sigma_{\mathrm{n}}^2||\mathbf{v}_k||^2} = \frac{p_k}{2\sigma_{\mathrm{n}}^2||\mathbf{w}_{d_k}||^2} = \frac{p_k}{2\sigma_{\mathrm{n}}^2(\mathbf{W}^{\mathrm{H}}\mathbf{W})_{d_k,d_k}} \quad (22)$$

Besides, since ZF combiner is given by $\mathbf{W} = \widehat{\mathbf{G}}(\widehat{\mathbf{G}}^{\mathrm{H}}\widehat{\mathbf{G}})^{-1}$ and $\widehat{\mathbf{G}} \approx \mathbf{G}$ (absence of pilot reuse), the asymptotic SINR is:

$$\Gamma_{k,\mathcal{S}} \overset{M\to\infty}{\longrightarrow} \frac{p_k}{2\sigma_{\mathrm{n}}^2(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}_{d_k,d_k}}, \quad (23)$$

The entry $(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}_{d_k,d_k}$ can be re-written as [14]

$$\begin{aligned}(\mathbf{G}^{\mathrm{H}}\mathbf{G})^{-1}_{d_k,d_k} &= (\mathbf{g}_{d_k}^{\mathrm{H}}\mathbf{g}_{d_k} - \mathbf{g}_{d_k}^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}(\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k})^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{g}_{d_k})^{-1} \\ &= (\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_k - \mathbf{h}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}(\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k})^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{h}_k)^{-1}, \quad (24)\end{aligned}$$

where $\ddot{\mathbf{G}}_{d_k} = [\mathbf{g}_1, \ldots, \mathbf{g}_{d_k-1}, \mathbf{g}_{d_k+1}, \cdots, \mathbf{g}_{|\mathcal{S}|}]$.

Based on [13, Lemma 4], $\check{\mathbf{h}}_k^{\mathrm{H}}\check{\mathbf{h}}_k \overset{M\to\infty}{\longrightarrow} \mathrm{tr}(\mathbf{R}_k)$. Since $\check{\mathbf{h}}_k$ has zero-mean entries, $\bar{\mathbf{h}}_k^{\mathrm{H}}\check{\mathbf{h}}_k \overset{M\to\infty}{\longrightarrow} 0$. Thus, we can approximate the first term in the right-hand side of (24) as

$$\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_k \overset{M\to\infty}{\longrightarrow} \bar{\mathbf{h}}_k^{\mathrm{H}}\bar{\mathbf{h}}_k + \mathrm{tr}(\mathbf{R}_k) = \mathrm{tr}(\mathbf{Q}_k). \quad (25)$$

Due to [13, Lemma 5], $\check{\mathbf{g}}_i^{\mathrm{H}}\check{\mathbf{g}}_j \overset{M\to\infty}{\longrightarrow} 0$, for $i \neq j$. Therefore,

$$\mathbf{g}_i^{\mathrm{H}}\mathbf{g}_j \overset{M\to\infty}{\longrightarrow} \bar{\mathbf{g}}_i^{\mathrm{H}}\bar{\mathbf{g}}_j. \quad (26)$$

In general, since XL-MIMO antenna arrays have large dimensions, the LoS regions of different UEs overlap in few antennas or even do not overlap, such that approximating the off-diagonal terms of $\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}$ to 0 is reasonable since they are highly probable to be smaller than the diagonal terms. On the other hand, this approximation is not reasonable for conventional M-MIMO BSs, whose antennas have the same physical location, such that the LoS region of a given UE encompasses all the M BS antennas or none of them. Thus, in XL-MIMO systems, the term $\mathbf{h}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}(\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k})^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{h}_k$ can be approximated as $\mathbf{h}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{h}_k$, where the matrix $\mathbf{T}_{d_k}$ is obtained by retaining the diagonal entries of $\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}$:

$$(\mathbf{T}_{d_k})_{i,j} = \begin{cases} (\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k})_{i,i} & \text{for } i = j, \\ 0 & \text{otherwise.} \end{cases} \quad (27)$$

Now, let us expand $\mathbf{h}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{h}_k$:

$$\begin{aligned}\mathbf{h}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{h}_k = \bar{\mathbf{h}}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\bar{\mathbf{h}}_k + \bar{\mathbf{h}}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\check{\mathbf{h}}_k + \\ \check{\mathbf{h}}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\bar{\mathbf{h}}_k + \check{\mathbf{h}}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\check{\mathbf{h}}_k. \quad (28)\end{aligned}$$

When M is very large, both second and third terms of (28) can be neglected. Moreover, due to [13, Lemma 4]:

$$\check{\mathbf{h}}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\check{\mathbf{h}}_k \overset{M\to\infty}{\longrightarrow} \mathrm{tr}(\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{R}_k). \quad (29)$$

Consequently,

$$\mathbf{h}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{h}_k \overset{M\to\infty}{\longrightarrow} \mathrm{tr}(\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{Q}_k). \quad (30)$$

By simple algebraic manipulation [14]:

$$\mathrm{tr}(\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{Q}_k) = \sum_{\substack{i\in\mathcal{S} \\ i\neq k}} \frac{\mathbf{g}_{d_i}^{\mathrm{H}}\mathbf{Q}_k\mathbf{g}_{d_i}}{\mathbf{g}_{d_i}^{\mathrm{H}}\mathbf{g}_{d_i}} = \sum_{\substack{i\in\mathcal{S} \\ i\neq k}} \frac{\mathbf{h}_i^{\mathrm{H}}\mathbf{Q}_k\mathbf{h}_i}{\mathbf{h}_i^{\mathrm{H}}\mathbf{h}_i}. \quad (31)$$

Using [13, Lemma 4] one more time:

$$\begin{aligned}\mathbf{h}_i^{\mathrm{H}}\mathbf{Q}_k\mathbf{h}_i \overset{M\to\infty}{\longrightarrow} \bar{\mathbf{h}}_i^{\mathrm{H}}\mathbf{Q}_k\bar{\mathbf{h}}_i + \mathrm{tr}(\mathbf{Q}_k\mathbf{R}_i) &= \mathrm{tr}(\mathbf{Q}_k(\bar{\mathbf{h}}_i\bar{\mathbf{h}}_i^{\mathrm{H}} + \mathbf{R}_i)) \\ &= \mathrm{tr}(\mathbf{Q}_k\mathbf{Q}_i). \quad (32)\end{aligned}$$

As in (25), the denominator of (31) approximates to $\mathrm{tr}(\mathbf{Q}_i)$ when M is very large. Thus,

$$\mathbf{h}_k^{\mathrm{H}}\ddot{\mathbf{G}}_{d_k}\mathbf{T}_{d_k}^{-1}\ddot{\mathbf{G}}_{d_k}^{\mathrm{H}}\mathbf{h}_k \overset{M\to\infty}{\longrightarrow} \sum_{\substack{i\in\mathcal{S} \\ i\neq k}} \frac{\mathrm{tr}(\mathbf{Q}_k\mathbf{Q}_i)}{\mathrm{tr}(\mathbf{Q}_i)}. \quad (33)$$

By replacing (25) and (33) into (24), and (24) into (23), we obtain the closed-form SINR expression below:

$$\Gamma_{k,\mathcal{S}}^{\text{approx}} = \frac{p_k}{2\sigma_{\text{n}}^2}\left(\text{tr}(\mathbf{Q}_k) - \sum_{\substack{i\in\mathcal{S}\\ i\neq k}} \frac{\text{tr}(\mathbf{Q}_k\mathbf{Q}_i)}{\text{tr}(\mathbf{Q}_i)}\right), \quad (34)$$

and the following closed-form expression for the sum SE:

$$\text{SE}_{\mathcal{S}}^{\text{approx}} = \left(1 - \frac{|\mathcal{S}|}{\tau_{\text{c}}}\right) \sum_{k\in\mathcal{S}}^{K} \log_2(1 + \Gamma_{k,\mathcal{S}}^{\text{approx}}). \quad (35)$$

V. Numerical Results

Monte Carlo numerical simulations reveal the accuracy of the low-complexity SINR expression in (34) and the performance of Alg. 1. The positions of the K UEs are randomly generated, and the antenna array contains L uniformly spaced SAs. Table I displays the adopted simulation parameters.

Table I: Adopted simulation parameters values

Parameter	**Value**
Length of the channel coherence block: τ_{c}	50
Number of antennas in each SA: N	4
Exponent of the NLoS path-loss: γ	4
Shadowing standard deviation: σ_{s}^2	3 dB (LoS) 4 dB (NLoS)
Standard deviation of the azimuth angle: σ_φ	10°
Standard deviation of the elevation angle: σ_θ	10°
Average channel gain at a distance of 1 m: β_0	$8.9125 \cdot 10^{-4}$
Array length	100 m
Wavelength: λ	12.5 cm
Intra-SA antenna separation: d	$\lambda/2$
UE transmit power: $p_1, \ldots, p_K$	10 dBm
Noise power: σ_{n}^2	−96 dBm
Number of Monte-Carlo channel realizations	1000

The LoS probability between UE k and SA l is distance-dependent and, for the micro-urban scenario, is modeled according to [15] as:

$$\Pr(\alpha_{kl} = 1) = \min\left(\frac{18 \text{ m}}{d_{kl}}, 1\right) \cdot \left[1 - e^{-\frac{d_{kl}}{36 \text{ m}}}\right] + e^{-\frac{d_{kl}}{36 \text{ m}}} \quad (36)$$

A. Accuracy for the SINR Approximation

We evaluate the accuracy of the low-complexity SINR expression in (34) by the probability of returning a negative value, *i.e.*, $\Pr(\Gamma_{k,\mathcal{S}}^{\text{approx}} < 0)$, and by the normalized absolute error (NAE):

$$\epsilon = \frac{|\text{SE}_{\mathcal{S}}^{\text{ZF}} - \text{SE}_{\mathcal{S}}^{\text{approx}}|}{\text{SE}_{\mathcal{S}}^{\text{ZF}}}. \quad (37)$$

Fig. 2 demonstrates that, for a relatively large number of antennas, the proposed SINR approximation given by (34) is almost surely non-negative and the NAE of the approximated SE is sufficiently small. The greater the number of antennas (M), the more evident channel hardening and favorable propagation will be, resulting in smaller variations in the instantaneous SINR around its large-scale average. Therefore, the approximated SINR expression, which depends on the channel statistics, will be more accurate. Increasing the number of UEs (K) causes the opposite effect.

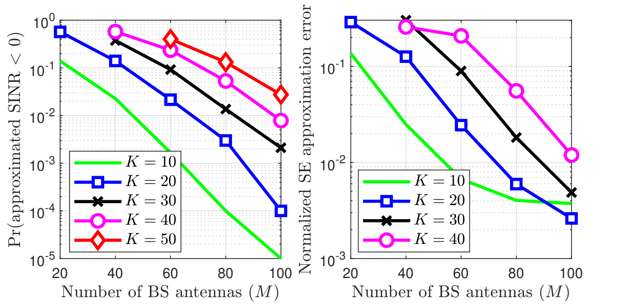


Figure 2: Accuracy of the approximated SINR expression in (34), when all UEs are scheduled, regarding $\Pr(\Gamma_{k,\mathcal{S}}^{\text{approx}} < 0)$ and the NAE in (37). Points where $M < K$ were not simulated.

B. User Scheduling Analysis

Fig. 3 plots the sum SE versus the number of UEs that are scheduled with orthogonal pilots, $|\mathcal{S}|$, which herein is also the pilot sequence length, τ_{p}. We compare two cases: **a)** the UEs are scheduled by prioritizing the ones with the strongest LSF channel gain, as in Alg. 1, and **b)** the UEs are scheduled randomly. The first one performed considerably better than the latter. Furthermore, there is a value of τ_{p} that maximizes the sum SE, since as more UEs are scheduled, the summation in (18) increases, but a smaller portion of the channel coherence block is available for data transmission. Moreover, increasing M improves the SE. Increasing K also improves the SE, since there are more possibilities for choosing the τ_{p} strongest UEs.

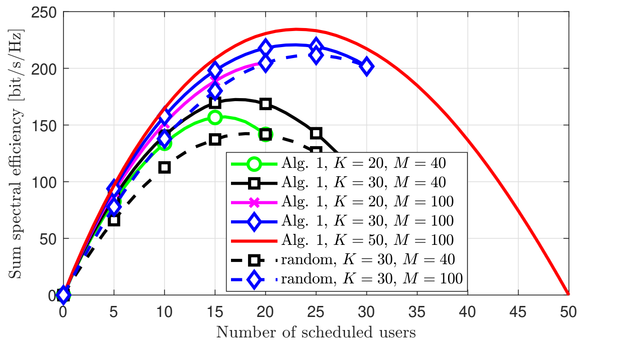


Figure 3: SE dependency on $|\mathcal{S}|$, varying M and K. Two US criteria are compared: random, and highest LSF gain (Alg. 1).

Fig. 4 plots the number of scheduled UEs and the sum SE versus the total number of UEs (K), under three US strategies: **a)** all the K UEs are scheduled, **b)** 50% of the K UEs are randomly scheduled, **c)** the UEs are scheduled by Alg. 1. Results show that the proposed US strategy performs better, in terms of SE, than the random US strategy, independent of the number of UEs. Furthermore, the performance of all scheduling strategy degrades when the number of UEs approaches τ_{c}, which does not happen with the other strategies. Also, increasing the number of SAs at the BS (L) results in a higher sum-SE.

Fig. 5 depicts how the XL-MIMO channel conditions impacts on the performance of the proposed US strategy. Since Alg. 1 gives priority to the strongest UEs, the probability of a UE being scheduled increases with the number of SAs

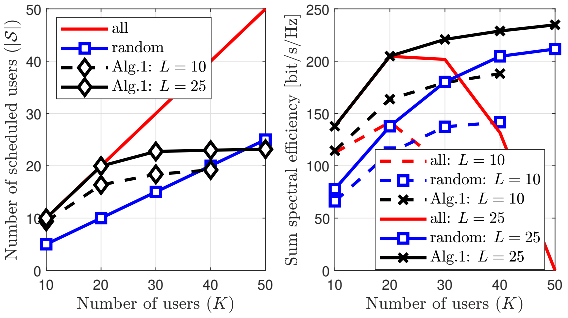


Figure 4: Number of scheduled UEs ($|\mathcal{S}|$) and sum SE, depending on K; all-scheduled, random scheduling and Alg. 1 strategies are compared, for different number of SAs (L); $N = 4$.

with LoS. For instance, considering the case of $K = 20$ and $L = 10$, a UE will surely be scheduled if it has LoS to at least 50% of the SAs, *i.e.*, 5 out of 10 SAs. This percentage increases to 70%, *i.e.*, 7 out of 10 SAs, when K goes from 20 to 40, since there is more concurrence. On the other hand, for $L = 25$, the scenario is more comfortable and a UE needs to have LoS to about 12% and 50% of the SAs in order to be almost surely scheduled, for $K = 20$ and $K = 40$, respectively.

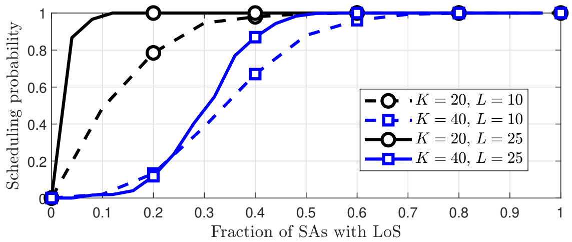


Figure 5: Probability of a UE being scheduled by Alg. 1, depending on the fraction of SAs with LoS; $N = 4$.

VI. Conclusions

For a sufficiently large number of antennas, numerical results corroborate that the proposed low-complexity SINR closed-form expression is very accurate, although being based only on the channel statistics, instead of the instantaneous channels. Results also confirm the proposed US algorithm is effective for crowded XL-MIMO systems, regarding the sum SE improvement, and demonstrate that it really prioritizes the strongest UEs, since the scheduling probability increases with the number of SAs to which a UE has LoS.

Future research for this work includes an extension of the proposed US algorithm aiming to increase the number of served UEs is devised as follows in XL-MIMO systems. In this scenario, since distant UEs are likely to see non-overlapping parts of the ULA, the UEs that were not scheduled might be admitted by being assigned a pilot that is already used.

References

[1] A. Amiri, C. N. Manchón, and E. de Carvalho, "A message passing based receiver for extra-large scale mimo," in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2019, pp. 564–568.

[2] Özlem Tugfe Demir, E. Björnson, and L. Sanguinetti, *Foundations of User-Centric Cell-Free Massive MIMO*, 2021, vol. 14, no. 3-4.

[3] J. H. I. de Souza, J. C. Marinello, A. Amiri, and T. Abrão, "Qos-aware user scheduling in crowded xl-mimo systems under non-stationary multi-state los/nlos channels," *IEEE Transactions on Vehicular Technology*, pp. 1–13, 2023.

[4] G. A. Ubiali, T. Abrão, and J. C. Marinello, "Improving spectral efficiency via pilot assignment and subarray selection under realistic xl-mimo channels," *Physical Communication*, vol. 58, p. 102062, 2023.

[5] 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.

[6] J. P. González-Coma, F. J. López-Martínez, and L. Castedo, "Low-complexity distance-based scheduling for multi-user xl-mimo systems," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2407–2411, 2021.

[7] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[8] J. C. Marinello, T. Abrão, A. Amiri, E. de Carvalho, and P. Popovski, "Antenna selection for improving energy efficiency in xl-mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13 305–13 318, 2020.

[9] F. E. Kadan and O. Haliloglu, "A performance bound for maximal ratio transmission in distributed mimo," *IEEE Wireless Communications Letters*, vol. 12, no. 4, pp. 585–589, 2023.

[10] J. Zhang, J. Zhang, Y. Han, J. Wang, and S. Jin, "Average spectral efficiency for tdd-based non-stationary xl-mimo with vr estimation," in *2022 14th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2022, pp. 973–977.

[11] S. Mukherjee and R. Chopra, "Performance analysis of cell-free massive mimo systems in los/ nlos channels," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6410–6423, 2022.

[12] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive mimo 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 232–257, 2020.

[13] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in correlated miso broadcast channels under limited feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.

[14] A. Ali, E. D. Carvalho, and R. W. Heath, "Linear receivers in non-stationary massive mimo channels with visibility regions," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 885–888, June 2019.

[15] 3GPP, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-ULTRA); Further advancements for E-ULTRA physical layer aspects (Release 9)," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.814, 03 2017, version 9.2.0.

APPENDIX C – PAPER C

- **Title:** User Scheduling in Crowded XL-MIMO Under Multi-state LoS/NLoS Channels
- **Authors:** Gabriel Avanzi Ubiali, Wilson Souza Jr., José Carlos Marinello Filho and Taufik Abrão
- **Journal:** Journal of Network and Systems Management, vol. 33, July 2025, IF = 4.1
- **DOI:** 10.1007/s10922-025-09961-w
- **Citation:** Ubiali, G.A., Souza, W., Marinello, J.C. et al., User Scheduling in Crowded XL-MIMO Under Multi-state LoS/NLoS Channels. *J Netw Syst Manage* **33**, 93 (2025).

Journal of Network and Systems Management (2025) 33:93
https://doi.org/10.1007/s10922-025-09961-w

User Scheduling in Crowded XL-MIMO Under Multi-state LoS/NLoS Channels

Gabriel Avanzi Ubiali[1] · Wilson Souza Jr.[1] · José Carlos Marinello[2] · Taufik Abrão[1]

Received: 7 October 2024 / Revised: 20 June 2025 / Accepted: 24 June 2025

Abstract
User scheduling methods in 5G and B5G networks are evolving from purely radio-centric optimizations to integral components of comprehensive network and system management strategies. The shift towards more dynamic, intelligent, and automated network operations necessitates user scheduling algorithms that are not only efficient in resource allocation but also adaptable to complex network architectures, capable of managing interference, and aligned with broader goals of automation and energy efficiency. In this paper, we have derived two closed-form expressions for the Signal-to-Interference-Plus-Noise Ratio (SINR) in extra-large scale MIMO (XL-MIMO) systems with spatially correlated multi-state line-of-sight (LoS)/non-LoS (NLoS) channels. One expression is asymptotic, valid for a sufficiently large number of antennas and in the absence of pilot reuse among scheduled users, while the other is ergodic, approximating the average SINR when the number of scheduled users exceeds the number of antennas serving each user. Numerical results validated the accuracy of both expressions, and we also provided simplified versions that demand significantly less computational effort. Furthermore, we proposed an efficient user scheduling (US) algorithm for crowded XL-MIMO systems that maximizes the sum spectral efficiency (SE) leveraging the derived analytical expressions instead of relying on instantaneous SE computations. This approach enables the US to remain valid over hundreds of channel coherence blocks, significantly reducing computational complexity. Our numerical results demonstrate that the proposed US algorithm substantially improves the sum-SE in crowded scenarios. By properly admitting pilot reuse, the algorithm effectively minimizes inter-user interference in XL-MIMO systems, mainly as different users are most likely served by a different set of antennas and thus interfere minimally with one another.

Keywords Extra-large-scale antenna array · User scheduling · Subarray clustering · Spectral efficiency

Extended author information available on the last page of the article

Published online: 15 July 2025

1 Introduction

A conventional massive multiple-input multiple-output (M-MIMO) base station (BS) comprises collocated antenna arrays. However, widely spreading a significant number of antennas, which has been called XL-MIMO, is a promising research direction to address crowded communication scenarios and the requirement of a more uniform quality of service (QoS) throughout the cell area by achieving moderate to high macro-diversity and reducing the average distance from user equipment (UE) to the closest antenna [1–3]. XL-MIMO antenna elements can be integrated into walls and columns of a stadium or shopping center, for example [4].

Different regions of a large aperture array receive different powers because they can see different scatterers [3]. Furthermore, different UEs may see different array parts via LoS propagation. At the same time, the other antennas may receive their signal due to the multipath components only, i.e., NLoS propagation. Since the transmission complexity for many antennas can be prohibitive, exploiting the visibility region (VR) information is essential to reduce the complexity in XL-MIMO [5].

The channels can be suitably estimated if each UE is assigned a different pilot sequence from an orthogonal set. However, in crowded massive multiple-input multiple-output (MIMO) scenarios, especially under high mobility open area conditions, the overhead caused by the channel estimation step over each channel coherence block can negatively impact the system spectral efficiency (SE). Thus, in these scenarios, assigning pilots judiciously becomes paramount to achieving high SE and compatible QoS for all UEs.

The optimal approach for massive MIMO channel estimation involves assigning each UE a distinct pilot sequence from an orthogonal set. However, in densely populated settings, especially under conditions of rapid mobility, the overhead associated with channel estimation within each channel coherence block can detrimentally impact the SE. In such challenging scenarios, the finite length of the coherence blocks makes it impossible to assign mutually orthogonal pilots to all UEs [1], which makes it harder to acquire channel state information (CSI) with no interference [6]. Thus, it is crucial to assign pilots wisely to reduce coherent interference between pilot-sharing UEs and enhance the overall performance of the communication system.

With this motivation, we propose a pilot assignment (PA) strategy for improving the SE in XL-MIMO systems under realistic multi-state LoS/NLoS channels with spatial correlation. Since the pilots are assigned based on channel statistics rather than on instantaneous channel realizations, the assignment decisions may remain valid for hundreds of coherence blocks, depending on the user's mobility, which alleviates the computational complexity.

Closed-form SINR expressions for centralized and distributed operation (UL). The PA algorithm bases its decision on the derived SINR expressions, which are valid for XL-MIMO channels. Therefore, these expressions and PA strategy are based on channel statistics, including VR information, which is paramount in

XL-MIMO since most of the energy received from a UE is focused on VR. For instance, one can assign pilots so that two UEs whose VRs overlap are assigned different pilots, while two UEs with nonoverlapping VRs could be assigned the same pilot. Thus, we take into account such a particular characteristic of XL-MIMO channels (VRs) to assign a limited number of pilot sequences for the UEs in a crowded massive machine-type communication (mMTC) scenario, therefore reducing the overhead due to channel estimation while reducing the pilot contamination. The optimal strategy to mitigate pilot contamination is to mutually assign the best configuration to the available set of pilots. Also, we adopt a robust and realistic channel model for XL-MIMO systems since it considers multi-state LoS/NLoS channels and spatial correlation, as in our previous contribution [7].

A deterministic SINR approximation depends solely on channel statistics, such as the channel mean vectors and covariance matrices, which vary much more slowly than the instantaneous realizations of the wireless channel. Deterministic expressions for the achievable SINR have been extensively studied in multi-user massive MIMO, XL-MIMO, and cell-free massive MIMO systems since 2019. However, most existing works are limited to uncorrelated Rayleigh fading channels or are only valid for maximum ratio (MR) receiver or precoder. Closed-form deterministic SINR expressions under spatially correlated Rayleigh fading have been derived in [8–12], but their validity is restricted to the use of MR combining or precoding. For spatially uncorrelated Rician fading channels, deterministic SINR expressions have been derived in [13–17]. In particular, Zhi et al. [14] provides a deterministic uplink expression under the assumption of perfect CSI and zero-forcing (ZF) combining. In contrast, [13, 15, 16] derive uplink deterministic expressions for MR combining under imperfect CSI. Furthermore, Li et al. [17] presents deterministic uplink SINR expressions for MR, partial ZF (PZF) and partial MMSE (PMMSE) combiners, also accounting for imperfect CSI. Nonetheless, all three works assume spatially uncorrelated channel models. Deterministic SINR expressions for cell-free massive MIMO with correlated Rician fading channels are derived in [18–22], accounting for channel estimation errors, i.e., imperfect CSI assumption. However, all these works, except [22], are valid only for MR combining or precoding. Jin et al.[18] also derives a closed-form deterministic SINR expression for ZF precoding, but this one is only valid for uncorrelated Rician fading.

In [23–26], the ergodic SINR expression under maximum-ratio (MR) processing in cellular massive MIMO or XL-MIMO with uncorrelated Rayleigh fading is explored considering the NLoS-only channel model. However, the expected condition of the XL-MIMO system operating in micro-urban channel scenarios should include both LoS and NLoS channel components. From the trustworthy numerical simulation perspective, we cannot disregard the LoS channel component in urban crowded channel scenarios since most UEs are close to the BS array. Furthermore, the performance of MR processing is poor under scenarios with strong multi-user interference.

1.1 State-of-the-Art in User Scheduling Methods for 5G and B5G Networks

Several innovative methods have been developed to meet the unique demands of scheduling in crowded XL-MIMO systems. These approaches tackle issues like interference management, computational scalability, and adaptation to dynamic conditions.

Related works. User scheduling (US) in XL-MIMO has already been addressed in [27–29], but the present work is the first one to consider at the same time spatially correlated channels and the unavailability of the channel estimates, such that we propose a US strategy that is based on the channel statistics rather than the instantaneous channel information. Five key solutions for user scheduling in 5G and B5G Networks are discussed in the sequel.

i) *Statistical-Based Scheduling* [30]: Statistical-based scheduling (SbS) leverages long-term channel statistics, such as mean gains or covariance matrices, rather than instantaneous channel state information (CSI). This method reduces the overhead associated with frequent CSI updates, which is particularly burdensome in XL-MIMO systems with many antennas and users. Scheduling decisions rely on these statistical properties, often integrating pilot assignment techniques to optimize resource allocation. It is especially suitable for fast-changing environments where acquiring real-time CSI is impractical.

   SbS uses channel statistics to make decisions, avoiding the need for frequent CSI updates and reducing overhead in crowded, dynamic scenarios. It often pairs with pilot assignment to enhance resource use. However, it misses instantaneous channel variations, potentially leading to suboptimal performance, and relies on accurate statistical models, which can be challenging to maintain. Its complexity is lower than CSI-based methods and varies depending on the sophistication of the statistical model employed.

ii) *Machine Learning-Based Scheduling* [31]: Machine learning (ML)-based scheduling (MLS) employs techniques like reinforcement or supervised learning to develop adaptive policies by analyzing historical data or real-time feedback. ML models optimize scheduling based on patterns in user behavior, channel conditions, and system performance, making this approach highly adaptable. It excels in complex, variable settings where traditional methods may struggle to adjust effectively, integrating diverse inputs such as quality of service requirements or user priorities.

   ML-based scheduling optimizes decisions through data-driven insights, adapting to changing conditions and user needs while handling diverse inputs like QoS or priorities. It requires large training datasets, which may be costly to acquire, and incurs high computational costs during training. Additionally, its decisions may lack transparency, complicating system analysis. Complexity is high during training but can be low during deployment if the model is efficiently implemented.

iii) *Heuristic Algorithms* [32]: Heuristic algorithms, such as greedy or genetic algorithms, swarm optimization, offer practical solutions by prioritizing simplicity and efficiency over exhaustive optimization. These methods iteratively select users based on predefined criteria, like maximizing throughput or minimizing interference, without ensuring global optimality. They are particularly useful when computational resources are limited and near-optimal solutions suffice, providing a customizable and straightforward implementation.

iv) *Graph-Based Scheduling* [33]: Graph-based scheduling models interference relationships among users as a graph, with nodes representing users and edges indicating interference levels. Using graph theory algorithms, such as maximum independent set or graph coloring, it selects non-interfering user sets to minimize conflicts. This approach is highly effective in dense environments where interference is a primary concern, leveraging established mathematical tools for robust solutions.

GbS directly addresses interference using graph models and robust graph theory tools, excelling in dense, interference-heavy settings. Its complexity increases with the number of users, as larger graphs are more complex to process, and it requires precise interference data, which may be inaccurate in practice. Computational complexity is generally high, often polynomial, e.g., $\mathcal{O}(n^2)$ or exponential, depending on the algorithm.

v) *Cluster-Based Scheduling* [34]: Cluster-based scheduling (CbS) groups users into clusters based on spatial locations or channel characteristics, simplifying the scheduling problem by managing smaller subsets. The system first forms clusters, then performs intra-cluster scheduling, often exploiting spatial separation to reduce interference. This method supports parallel processing and is well-suited to distributed XL-MIMO setups where users naturally form groups.

CbS simplifies the problem by dividing users into clusters, reducing interference via spatial grouping and enabling scalable, parallel processing. The clustering process adds overhead, and suboptimal cluster design may limit performance. Complexity varies based on the clustering algorithm (e.g., k-means) and the intra-cluster scheduling method applied.

Table 1 summarizes features, drawbacks and complexities for the five user scheduling approaches. User scheduling is fundamental to optimizing crowded XL-MIMO systems. The methods discussed, namely statistical-based, machine

Table 1 Comparison of user scheduling methods and their relevance to network and system management

Method	Features	Drawbacks	Complexity	Relevance to network and system management
SbS	Uses statistics, low CSI overhead	Misses real-time data, needs accurate models	Lower than CSI-based	Reduces management overhead by minimizing frequent CSI updates; relies on robust statistical models for resource planning
MLS	Data-driven, highly adaptive	Data-intensive, high training cost	High training, low use	Enables intelligent automation and self-optimization; requires significant data management and computational resources for training and deployment
Hbs	Cost-effective, customizable	Suboptimal, heuristic-dependent	Low to moderate	Provides practical, low-overhead solutions for resource allocation; useful for rapid deployment in resource-constrained management environments
GbS	Interference focus, graph theory tools	Complex with scale, needs interference data	High (e.g., $O(n^2)$)	Directly addresses interference management, a critical network management concern; requires accurate real-time interference data for effective operation
CbS	Simplifies via clusters, scalable	Clustering overhead, suboptimal clusters	Varies by method	Enhances scalability and distributed management by breaking down complex problems; optimal clustering is key for efficient resource utilization and interference mitigation

learning-based, heuristic, graph-based, and cluster-based, each address specific challenges. Statistical approaches reduce overhead, ML offers adaptability, heuristics provide practicality, graph-based methods excel in interference management, and cluster-based strategies enhance scalability. The optimal choice depends on factors like user density, resources, and performance goals, enabling tailored solutions for peak efficiency in crowded XL-MIMO environments

Notice that the proposed two-step user scheduling XL-MIMO method, Fig. 2, most closely aligns with Statistical-Based scheduling approach. Its defining feature is the use of channel statistics for both SINR derivation and scheduling decisions, minimizing reliance on instantaneous CSI and reducing overhead – core traits of this category. The integration of pilot assignment (orthogonal and reuse phases) further strengthens this classification, as statistical-based methods often pair with such strategies to optimize resource use. Moreover, while it exhibits heuristic elements (greedy selection in user prioritization), these are secondary to its statistical foundation. The method's analytical SINR expressions and long-term applicability distinguish it from purely heuristic approaches. Similarly, although it uses subarray clusters for signal processing, it does not schedule within user-defined clusters, ruling out a primary cluster-based classification.

1.2 User Scheduling Methods—A Network and System Management Perspective

User scheduling in 5G and B5G networks is not merely about allocating resources to maximize throughput; it is intricately linked with the broader objectives of network and system management, including resource orchestration, interference management, energy efficiency, and automation. The following discussion elaborates on how various user scheduling methods contribute to or are influenced by these management aspects.

Resource Management and Orchestration. Effective user scheduling is a cornerstone of efficient radio resource management (RRM) in 5G and B5G networks. RRM encompasses the strategies and mechanisms for allocating and managing radio resources (e.g., time, frequency, power, and spatial resources) to meet diverse service requirements and optimize network performance [35]. User scheduling algorithms play a pivotal role in this by determining which users are served, when, and with what resources.

In the context of network and system management, resource orchestration becomes paramount. This involves the automated arrangement, coordination, and management of network resources to deliver services. User scheduling, therefore, must integrate seamlessly with orchestration frameworks to dynamically adapt to changing network conditions, traffic demands, and service level agreements (SLAs). For instance, in network slicing, different slices may have distinct QoS requirements (e.g., URLLC for critical communications vs. enhanced Mobile Broadband (eMBB) for high-throughput applications). User schedulers must be capable of prioritizing and allocating resources according to these slice-specific policies, often requiring intelligent decision-making at various network layers [36].

Interference Management. Interference is a pervasive challenge in dense 5G and B5G deployments, particularly in heterogeneous networks (HetNets) where small cells are overlaid with macro cells. Efficient user scheduling is a primary tool for mitigating inter-cell and intra-cell interference. From a network management perspective, interference management strategies often involve coordinated scheduling among base stations, dynamic power control, and advanced antenna techniques. User scheduling algorithms can be designed to minimize interference by grouping users with favorable channel conditions or by employing techniques like coordinated multi-point (CoMP) transmission/reception [37].

The management of interference extends beyond simple resource allocation. It requires real-time monitoring of channel conditions, prediction of interference patterns, and adaptive adjustment of scheduling decisions. This often leverages machine learning (ML) techniques to learn complex interference dynamics and optimize scheduling policies. The goal from a system management standpoint is to achieve a balance between maximizing spectral efficiency and ensuring a certain level of fairness and QoS across all users, even in highly interfered environments.

Automation and Artificial Intelligence. The complexity of wireless systems in 5G and B5G, coupled with the dynamic nature of traffic and channel conditions, makes manual configuration and optimization of user scheduling impractical. This has led to a strong emphasis on automation and the application of Artificial Intelligence (AI) and Machine Learning (ML) in network and system management. AI/ML-driven user scheduling can learn from historical data and real-time feedback to make intelligent decisions, adapting to unforeseen circumstances and optimizing performance beyond what traditional rule-based algorithms can achieve [38].

From a management perspective, AI/ML-assisted user scheduling contributes to self-organizing networks (SONs) and zero-touch network management. These approaches aim to reduce operational expenditure (OPEX) and improve network agility by automating tasks such as resource allocation, fault management, and performance optimization. Deep Reinforcement Learning (DRL), for example, has shown promise in optimizing user scheduling in complex scenarios, such as High Altitude Platform Station (HAPS) networks, where traditional methods struggle due to the dynamic environment and large state spaces [39].

Energy Efficiency. As network density increases and more devices connect to 5G and B5G networks, energy consumption becomes a significant concern for network operators. User scheduling methods can contribute to energy efficiency by optimizing resource utilization and minimizing unnecessary power consumption. This can involve scheduling users during periods of low traffic to allow base stations to enter sleep modes, or by employing energy-aware scheduling algorithms that prioritize users with better channel conditions to reduce transmit power [40].

Network and system management frameworks are increasingly incorporating energy efficiency as a key performance indicator. This requires schedulers to consider not only throughput and QoS but also the energy footprint of their decisions. The integration of energy management policies with user scheduling algorithms is crucial for achieving sustainable and cost-effective network operations in the long term.

1.3 Contribution

The *Contribution* of this paper is threefold:

1. We derive two SINR closed-form expressions for XL-MIMO networks with centralized uplink (UL) operation, which are valid for a sufficiently large number of BS antennas. One of the expressions is valid for general conditions, while the other is valid in the absence of pilot reuse. Furthermore, we derive a simplified version of both expressions, simplifying the channel correlation matrix as a block diagonal matrix. Although the derived SINR expressions depend only on the channel statistics, i.e., on the mean channel vectors and covariance matrices, we numerically corroborate that the approximation error is negligible for massive and extra-large MIMO conditions.
2. We propose a PA algorithm for sum-SE maximization purposes. It primarily selects the UEs with the highest large-scale fading (LSF) channel gains and establishes its decision on the derived SINR expressions, which depend only on the channel statistics, such that the US decisions remain valid for hundreds of coherence blocks.
3. To corroborate the efficiency and efficacy of the proposed US procedure, we consistently provide and analyze numerical results w.r.t. the SE system and the precision of the SINR expression.

The remainder of this work is organized into five sections. Section 2 describes the adopted multi-state LoS/NLoS channel, as well as the channel estimation and data transmission steps, for XL-MIMO systems. Section 3 derives the SINR closed-form expressions, while Sect. 4 deals with the proposed method for US. Sections 5 and 6 present the numerical results and the main conclusions, respectively.

2 XL-MIMO System Model

We consider a crowded scenario with time-division-duplexing (TDD) operation where a BS equipped with an XL-MIMO system serves K single-antenna users. The system is implemented using a distributed network of subarray (SA)s. Each of the L SAs contains a set of antennas (in this work, a uniform linear array (ULA) with N elements, such that $M = LN$ is the total number of antennas. The SAs can be strategically deployed to ensure efficient coverage of the target area (e.g., on building facades, within stadiums, or in shopping centers) [4]. The following practical aspects are considered:

- **Subarray distribution:** each SA is placed to maximize coverage. Due to the inherent non-stationarity of XL-MIMO channels, a given UE is typically visible only to a subset of the SAs. This distribution placement enables spatial diversity and helps exploit favorable propagation conditions [41].

- **Local processing and central coordination:** each SA performs local signal processing tasks such as channel estimation. The estimates are then forwarded via a high-speed front-haul network to a central processing unit (CPU). The CPU coordinates joint processing across SAs, enabling advanced user scheduling strategies based on the aggregated channel statistics.
- **Interconnection and synchronization:** the architecture relies on dedicated interconnections between the SAs and the CPU. These links are designed to support low-latency and high-bandwidth communications, which are critical for exchanging CSI and coordinating timing synchronization accross the distributed antennas. Proper synchronization is essential to ensure that the combined signals are coherent and that the user scheduling decisions are based on accurate channel conditions.
- **Integration with existing infrastructure:** the proposed architecture is flexible and can be integrated with existing network infrastructures. The number and placement of SAs can be adapted to the specific environmental and coverage requirements, making it possible to enhance current systems with XL-MIMO capabilities.

We define $\mathcal{U} \subseteq \{1, \dots, K\}$ as the set of scheduled UEs such that $U = |\mathcal{U}|$ is the number of scheduled UEs. A CPU is connected to the SAs and is responsible for the SA cooperation.

Given the extra-large size of the antenna array, each UE is close to only a subset of SAs, making it both impractical and unnecessary for all SAs to handle all UEs. Restricting the number of SAs allowed to serve the UE can lead to some degradation in service quality. However, the performance loss should be minimal if all the SAs that can reach the UE with a signal power non-negligible compared to the termal noise power actively contributes to serving that UE. Hence, a user-centered approach is adopted to form the SA clusters that serve each UE. Practical benefits include reduced front-haul signaling and lower computational demands [1].

The serving SA cluster formation is dynamic which may depend on user locations, channel conditions, service requirements, etc. We denote the set of SAs that serve UE k as $\mathcal{L}_k$. The number of SAs serving UE k is $L_k = |\mathcal{L}_k|$, and the number of antennas serving UE k is $M_k = NL_k$. For notation convenience, we also define the diagonal matrix $\mathbf{D}_{kl} \in \mathbb{C}^{N \times N}$, which is an identity matrix if SA l serves UE k and a zero matrix otherwise:

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_N & \text{if } l \in \mathcal{L}_k \\ \mathbf{0}_N & \text{if } l \notin \mathcal{L}_k. \end{cases} \tag{1}$$

2.1 XL-MIMO Channel Model

The channel vector between the kth UE and the lth SA can be expressed as

$$\mathbf{h}_{kl} = \overline{\mathbf{h}}_{kl} + \check{\mathbf{h}}_{kl}, \tag{2}$$

where $\overline{\mathbf{h}}_{kl} \in \mathbb{C}^N$ and $\check{\mathbf{h}}_{kl} \in \mathbb{C}^N$ denote, respectively, the LoS and the NLoS components. We assume that the channel is constant during a coherence block and varies independently from one block to another. If N is small, then the N antennas inside a given SA are highly likely to experience the same condition of existence or absence of the LoS path component, which is called spatial stationarity. On the other hand, spatial non-stationarity is assumed among different SAs.

The statistical distributions, i.e., mean vector, and covariance matrix, are assumed perfectly known and available at the BS, since they are constant during hundreds of channel coherence blocks.[1] Since the NLoS channel component has zero mean, the LoS channel vector corresponds to the channel mean. It is given by [42]

$$\overline{\mathbf{h}}_{kl} = \alpha_{kl}\sqrt{\overline{\beta}_{kl}}e^{-j2\pi\frac{d_{kl}}{\lambda}}\mathbf{a}(\overline{\varphi}_{kl}, \overline{\theta}_{kl}), \tag{3}$$

where d_{kl} denotes the distance from the kth UE to the first antenna of the lth SA, λ is the wavelength and $\overline{\beta}_{kl}$ is the LSF channel gain related to the LoS path, which is determined by the normalized trace as

$$\overline{\beta}_{kl} = \frac{1}{N}||\overline{\mathbf{h}}_{kl}||^2 = \alpha_{kl}\frac{\beta_0}{d_{kl}^2}\overline{\mathcal{X}}_{kl}, \tag{4}$$

where α_{kl} is a Bernoulli random variable indicating the presence ($\alpha_{kl} = 1$) or absence ($\alpha_{kl} = 0$) of a LoS component between the kth UE and the lth SA, β_0 is the average channel gain at the reference distance of 1 m, β_0/d_{kl}^2 is the path-loss of the LoS component, and $\overline{\mathcal{X}}_{kl}$ is the shadow fading of the LoS component. If there is a LoS path between UE k and SA l, then we say that SA l is within VR of UE k. Otherwise, the received power is much weaker as the signal is coming only from NLoS paths, and then we say that SA l is not in the VR of UE k.

We model the LoS and NLoS *shadow fading*, which are respectively denoted by $\overline{\mathcal{X}}_{kl}$ and $\check{\mathcal{X}}_{kl}$, as log-normal random variables, such that $\ln(\overline{X}_{kl}), \ln(\check{X}_{kl}) \sim \mathcal{N}(0, \sigma_{\mathrm{SF}}^2)$. For simplicity, the shadow fading related to different UEs and/or different SAs are assumed independently distributed. Cross-correlated shadowing is considered in [7].

Finally, the array response vector regarding to the LoS ray, denoted by $\mathbf{a}(\overline{\varphi}_{kl}, \overline{\theta}_{kl})$ is a function of the LoS azimuth and elevation angles between UE k and SA l, respectively $\overline{\varphi}_{kl}$ and $\overline{\theta}_{kl}$, and its nth entry is given by

$$[\mathbf{a}(\overline{\varphi}_{kl}, \overline{\theta}_{kl})]_n = e^{-j2(n-1)\pi\frac{d}{\lambda}\sin(\overline{\varphi}_{kl})/\cos(\overline{\theta}_{kl})}, \tag{5}$$

where d is the antenna spacing inside the l-th SA. It accounts for the phase differences among the signal that arrives in different antennas of the same SA. The phase variations appear when the planar wavefront is not parallel to the ULA, i.e., when $\varphi \neq 0$.

The NLoS channel vector is the superposition of multiple propagation paths, approaching a Gaussian distribution [43]:

[1] In practice, they can be estimated using the sample mean and sample covariance matrices.

$$\check{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_{kl}), \tag{6}$$

with a positive semi-definite spatial correlation matrix $\mathbf{R}_{kl}$, whose (m,n)-th entry is given by:

$$[\mathbf{R}_{kl}]_{m,n} = \check{\beta}_{kl} \int\int e^{j2(m-n)\pi \frac{d}{\lambda} \frac{\sin(\varphi)}{\cos(\theta)}} f_{kl}(\varphi, \theta) d\varphi d\theta, \tag{7}$$

The joint probability density function of $\overline{\varphi}$ and $\overline{\theta}$ is given by [1]:

$$f_{kl}(\varphi, \theta) = \frac{1}{2\pi\sigma_\varphi\sigma_\theta} \exp\left\{ -\frac{(\varphi - \overline{\varphi}_{kl})^2}{2\sigma_\varphi^2} - \frac{(\theta - \overline{\theta}_{kl})^2}{2\sigma_\theta^2} \right\}, \tag{8}$$

where $\sigma_\varphi \geq 0$ and $\sigma_\theta \geq 0$ are the angular standard deviations. The integration interval for φ and θ is $[-180°, 180°]$ and $(-90°, 90°)$, respectively. The proof of (6) and (7) can be found in [7, Appendix A].

The NLoS LSF coefficient can be calculated as

$$\check{\beta}_{kl} = \frac{1}{N}\mathbb{E}\{||\check{\mathbf{h}}_{kl}||^2\} = \frac{1}{N}\mathrm{tr}(\mathbf{R}_{kl}) = \frac{\beta_0}{d_{kl}^{\gamma}}\check{\mathcal{X}}_{kl}, \tag{9}$$

where γ is the path-loss exponent of the NLoS component.

Finally, the channel vector is distributed as

$$\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\overline{\mathbf{h}}_{kl}, \mathbf{R}_{kl}). \tag{10}$$

Similarly, the collective channel of UE k, $\mathbf{h}_k = [\mathbf{h}_{k1}^{\mathrm{T}}, \dots, \mathbf{h}_{kL}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^M$, follows the distribution

$$\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\overline{\mathbf{h}}_k, \mathbf{R}_k), \tag{11}$$

where $\mathbf{h}_k = [\mathbf{h}_{k1}^{\mathrm{T}}, \dots, \mathbf{h}_{kL}^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{R}_k = \mathrm{diag}(\mathbf{R}_{k1}, \dots, \mathbf{R}_{kL})$. In practical XL-MIMO panel scenarios, as the antennas of different SAs are sufficiently distant, we assume that the channel vectors of other SAs are independently distributed, which means that $\mathbb{E}\{\check{\mathbf{h}}_{kl}\check{\mathbf{h}}_{kj}^{\mathrm{H}}\} = \mathbf{0}_{N\times N}$ for $l \neq j$, where $(.)^{\mathrm{H}}$ denotes the Hermitian operator. That is why we defined $\mathbf{R}_k$ as a block-diagonal matrix.

Figure 1 illustrates the LoS + NLoS propagation in XL-MIMO systems, between a UE k and a SA l. The NLoS components are generated by a scattering cluster. The distance d_{kl} and the LoS azimuth angle $\overline{\varphi}_{kl}$ are almost the same for all the N antennas. Notice that the UE is in the near-field (spherical wavefront) with respect to the whole antenna array and in the far-field (planar wavefront) w.r.t. each SA. For simplicity, the figure only shows the azimuth plane and two multipath components.

Recalling that $||\overline{\mathbf{h}}_{kl}||^2 = N\overline{\beta}_{kl}$ and $\mathbb{E}\{||\check{\mathbf{h}}_{kl}||^2\} = \mathrm{tr}(\mathbf{R}_{kl}) = N\check{\beta}_{kl}$, the multi-state probabilistic LoS/NLoS LSF channel gain is given by

$$\beta_{kl} = \frac{1}{N}\mathbb{E}\{||\mathbf{h}_{kl}||^2\} = \frac{1}{N}\mathrm{tr}(\mathbf{Q}_{kl}) = \overline{\beta}_{kl} + \check{\beta}_{kl}, \tag{12}$$

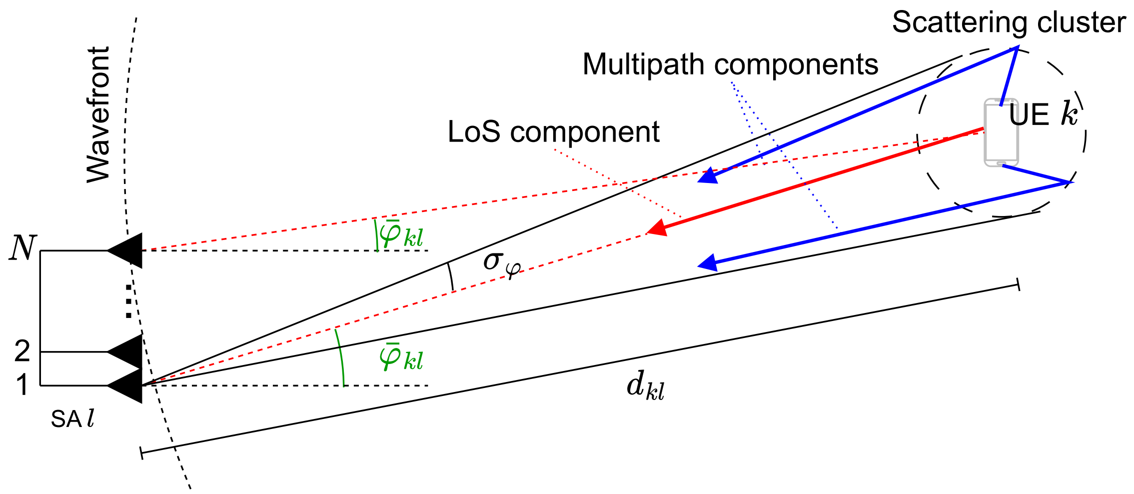


Fig. 1 LoS + NLoS propagation between UE k and SA l

where $\mathbf{Q}_{kl} = \bar{\mathbf{h}}_{kl}\bar{\mathbf{h}}_{kl}^{\mathrm{H}} + \mathbf{R}_{kl}$. Lastly, defining $\mathbf{Q}_{k} = \bar{\mathbf{h}}_{k}\bar{\mathbf{h}}_{k}^{\mathrm{H}} + \mathbf{R}_{k}$, the average LSF channel gain between UE k and the ULA is

$$\beta_k = \frac{1}{NL}\mathbb{E}\{||\mathbf{h}_k||^2\} = \frac{1}{NL}\sum_{l=1}^{L}\mathrm{tr}(\mathbf{Q}_{kl}) = \frac{1}{L}\sum_{l=1}^{L}\beta_{kl}. \tag{13}$$

2.2 Uplink Pilot Transmission

In each channel coherence block, we can transmit τ_{c} symbols. The channels must be estimated once per coherence block, such that τ_{p} symbols must be reserved for uplink pilot signaling in each coherence block, where τ_{p} is the length of the pilot sequences. The remaining $\tau_{\mathrm{c}} - \tau_{\mathrm{p}}$ symbols are used for data transmission. In this work, we only consider uplink data transmission.

The pilot sequence assigned to the kth UE is denoted as $\boldsymbol{\phi}_{t_k} \in \mathbb{C}^{\tau_{\mathrm{p}}}$, and $\boldsymbol{\phi}_{t_i}^{\mathrm{H}}\boldsymbol{\phi}_{t_j}$ is τ_{p} if UEs i and j are assigned the same pilot sequence, i.e., if $t_i = t_j$, and zero otherwise, which means that the pilot sequences form an orthogonal set. Since we can only find τ_{p} mutually orthogonal sequences, the finite length of the coherence blocks imposes the constraint $\tau_{\mathrm{p}} \leq \tau_{\mathrm{c}}$. In crowded scenarios, frequently $K \gg \tau_{\mathrm{p}}$, and therefore there will be pilot sharing UEs. We define the set of UEs using the same pilot as UE k, including itself, as

$$\mathcal{P}_k = \{i : t_i = t_k, i = 1, \ldots, K\} \subset \{1, \ldots, K\}. \tag{14}$$

The elements of $\sqrt{p_k}\boldsymbol{\phi}_{t_k}$ are transmitted by the kth UE over τ_{p} consecutive samples at the beginning of the coherence block, with transmit power p_k. Thus, during the entire pilot transmission, the received signal $\mathbf{Y}^{\mathrm{pilot}} \in \mathbb{C}^{M\times\tau_{\mathrm{p}}}$ at the BS is

$$\mathbf{Y}^{\text{pilot}} = \sum_{i \in \mathcal{U}} \sqrt{p_i} \mathbf{h}_i \boldsymbol{\phi}_{t_i}^{\text{T}} + \mathbf{N}, \tag{15}$$

where $\mathbf{N} \in \mathbb{C}^{M \times \tau_{\text{p}}}$ is the receiver noise with i.i.d. elements distributed as $\mathcal{N}_{\mathbb{C}}(0, \sigma_{\text{n}}^2)$.

2.3 MMSE Channel Estimation

For estimating the channel of the kth UE, we must at first separate the channels of non-pilot-sharing UEs, which is done by multiplying $\mathbf{Y}^{\text{pilot}}$ with the normalized conjugate of $\boldsymbol{\phi}_{t_k}$, resulting in

$$\mathbf{y}_{t_k}^{\text{pilot}} = \mathbf{Y}^{\text{pilot}} \boldsymbol{\phi}_{t_k}^* = \sum_{i \in \mathcal{P}_k} \sqrt{p_i} \tau_{\text{p}} \mathbf{h}_i + \mathbf{n}_{t_k}, \tag{16}$$

where $\mathbf{n}_{t_k} = \mathbf{N}\boldsymbol{\phi}_{t_k}^* \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \tau_{\text{p}} \sigma_{\text{n}}^2 \mathbf{I}_M)$. The MMSE estimate of $\mathbf{h}_k$ is [44]:

$$\widehat{\mathbf{h}}_k = \overline{\mathbf{h}}_k + \sqrt{p_k} \mathbf{R}_k \boldsymbol{\Psi}_{t_k}^{-1} \left(\mathbf{y}_{t_k}^{\text{pilot}} - \overline{\mathbf{y}}_{t_k}^{\text{pilot}} \right), \tag{17}$$

where $\boldsymbol{\Psi}_{t_k}$ is the covariance matrix of $\mathbf{y}_{t_k}^{\text{p}}$ and is given by

$$\boldsymbol{\Psi}_{t_k} = \frac{1}{\tau_{\text{p}}} \text{Cov}(\mathbf{y}_{t_k}^{\text{p}}) = \sum_{i \in \mathcal{P}_k} p_i \tau_{\text{p}} \mathbf{R}_i + \sigma_{\text{n}}^2 \mathbf{I}_M, \tag{18}$$

and

$$\overline{\mathbf{y}}_{t_k}^{\text{pilot}} = \mathbb{E}\{\mathbf{y}_{t_k}^{\text{pilot}}\} = \sum_{i \in \mathcal{P}_k} \sqrt{p_i} \tau_{\text{p}} \overline{\mathbf{h}}_i. \tag{19}$$

Similarly, we can also define

$$\boldsymbol{\Psi}_{t_k} = \sum_{i \in \mathcal{P}_k} p_i \tau_{\text{p}} \mathbf{R}_i + \sigma_{\text{n}}^2 \mathbf{I}_M. \tag{20}$$

By adopting the MMSE channel estimation, the estimation error $\widetilde{\mathbf{h}}_k = \mathbf{h}_k - \widehat{\mathbf{h}}_k$ can be modeled as a zero-mean random variable with covariance matrix [44] given by

$$\mathbf{C}_k = \mathbf{R}_k - p_k \tau_{\text{p}} \mathbf{R}_k \boldsymbol{\Psi}_{t_k}^{-1} \mathbf{R}_k. \tag{21}$$

The estimate $\widehat{\mathbf{h}}_k$ and the estimation error $\widetilde{\mathbf{h}}_k$ are independent random variables distributed as [44]

$$\widehat{\mathbf{h}}_k \sim \mathcal{N}_{\mathbb{C}}\left(\overline{\mathbf{h}}_k, \mathbf{R}_k - \mathbf{C}_k \right), \tag{22}$$

$$\widetilde{\mathbf{h}}_k \& \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_M, \mathbf{C}_k \right). \tag{23}$$

2.4 Uplink Data Transmission

During the UL data transmission, the BS receives the signal:

$$\mathbf{y}^{\mathrm{ul}} = \sum_{i \in \mathcal{U}} \mathbf{h}_i x_i + \mathbf{n}^{\mathrm{ul}}, \tag{24}$$

where $x_i \sim \mathcal{N}_{\mathbb{C}}(0, p_i)$ is the signal transmitted by UE k, with transmit power p_i, and $\mathbf{n}^{\mathrm{ul}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{\mathrm{n}}^2 \mathbf{I}_M)$ is the collective noise vector.

If UE k is scheduled, the CPU will estimate x_k by computing the inner product between its **receive combining vector**, denoted by $\mathbf{v}_k$, and the received signal $\mathbf{y}^{\mathrm{ul}}$, and considering only the SAs that serve UE k:

$$\hat{x}_k = \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{y}^{\mathrm{ul}} = \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k x_k + \sum_{i \in \mathcal{U} \setminus \{k\}} \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_i x_i + \sum_{i \in \mathcal{U}} \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widetilde{\mathbf{h}}_i x_i + \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{n}^{\mathrm{ul}}, \tag{25}$$

where $\mathbf{D}_k = \mathrm{diag}(\mathbf{D}_{k1}, \ldots, \mathbf{D}_{kL})$. Notice that x_k is jointly estimated by the SAs that serve UE k. Notice also that the desired signal, $\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{h}_k x_k$, was divided into two parts in (25): the desired signal on the estimated channel, $\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k x_k$, and the desired signal over the unknown channel, $\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widetilde{\mathbf{h}}_k x_k$. Only the first can be used for signal detection [1]. The second one is not useful since only the distribution of the estimation error is known, and thus it is treated as interference or noise when computing the instantaneous effective uplink SINR of a scheduled UE k, which is given by:

$$\Gamma_k = \frac{p_k |\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k|^2}{\sum\limits_{i \in \mathcal{U} \setminus \{k\}} p_i |\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_i|^2 + \sum\limits_{i \in \mathcal{U}} p_i |\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widetilde{\mathbf{h}}_i|^2 + \sigma_{\mathrm{n}}^2 ||\mathbf{D}_k \mathbf{v}_k||^2}. \tag{26}$$

The sum spectral efficiency (SE) is computed as [1]:

$$\mathrm{SE} = \left(1 - \frac{\tau_{\mathrm{p}}}{\tau_{\mathrm{c}}}\right) \sum_{k \in \mathcal{U}} \log_2(1 + \Gamma_k). \tag{27}$$

Since only the statistics (covariance matrices) of the estimation errors are known, not the instantaneous values, we can only maximize the SINR if we replace $|\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widetilde{\mathbf{h}}_i|^2$, with $\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{C}_i \mathbf{D}_k \mathbf{v}_k$. Thereby, the SINR in (26) is maximized by the MMSE receive combining vector [1], given by

$$\mathbf{v}_k = p_k \left(\sum_{i \in \mathcal{U}} p_i \mathbf{D}_k (\widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}} + \mathbf{C}_i) \mathbf{D}_k + \sigma_{\mathrm{n}}^2 \mathbf{I}_M \right)^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k, \tag{28}$$

which leads to the maximum SINR value for the scheduled k-th user on the set $\mathcal{U}$ as [1]:

$$\Gamma_k = p_k \widehat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{D}_k \mathbf{Z}_k^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k = p_k \cdot \mathrm{tr}\left(\mathbf{Z}_k^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k \widehat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{D}_k \right), \tag{29}$$

where

$$\mathbf{Z}_k = \sum_{i \in \mathcal{U} \setminus \{k\}} p_i \mathbf{D}_k \widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}} \mathbf{D}_k + \sum_{i \in \mathcal{U}} p_i \mathbf{D}_k \mathbf{C}_i \mathbf{D}_k + \sigma_{\mathrm{n}}^2 \mathbf{I}_M. \tag{30}$$

Finally, the instantaneous SE is given by:

$$\eta_k = \left(1 - \frac{\tau_{\mathrm{p}}}{\tau_{\mathrm{c}}} \right) \log_2(1 + \Gamma_k). \tag{31}$$

3 Analytical SINR Expressions

We aim to derive an SINR expression that remains valid across hundreds of coherence blocks, relying solely on channel statistics. Specifically, we seek an analytical expression for the expected value of the SINR in (29), averaged over the small-scale fading (SSF) channel realizations.

First of all, as shown in [1, page 107], the correlation between the MMSE estimates of $\mathbf{h}_k$ and $\mathbf{h}_i$ is:

$$\mathbb{E}\{\widehat{\mathbf{h}}_k \widehat{\mathbf{h}}_i^{\mathrm{H}}\} = \begin{cases} \bar{\mathbf{h}}_k \bar{\mathbf{h}}_i^{\mathrm{H}} + \mathbf{B}_{ki}, & \text{if } t_i = t_k, \\ \bar{\mathbf{h}}_k \bar{\mathbf{h}}_i^{\mathrm{H}}, & \text{if } t_i \neq t_k. \end{cases} \tag{32}$$

where $\mathbf{B}_{ki} = \sqrt{p_k p_i} \tau_{\mathrm{p}} \mathbf{R}_k \mathbf{\Psi}_{t_k}^{-1} \mathbf{R}_i$.

One can see that, although $\mathbf{Z}_k$ only contains the estimated channels of users other than k, it may still be correlated with $\widehat{\mathbf{h}}_k$, especially if user k shares its pilot with other users. However, this correlation will be neglected when deriving the SINR expression, since the overlap of LoS regions of different UEs is usually small and user k will normally share its pilot with only a few users, or even with no other user. Thus, since $\mathbb{E}\{\mathbf{Z}_k^{-1}\}$ can be approximated as $\mathbb{E}\{\mathbf{Z}_k\}^{-1}$ when $M_k \ll U$, and assuming no correlation between $\mathbf{Z}_k$ and $\widehat{\mathbf{h}}_k$, we have:

$$\begin{aligned} \mathbb{E}\{\Gamma_k\} &= p_k \cdot \mathrm{tr}\left(\mathbb{E}\{\mathbf{Z}_k^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k \widehat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{D}_k\} \right), \\ &\approx p_k \cdot \mathrm{tr}\left(\mathbb{E}\{\mathbf{Z}_k^{-1}\} \mathbf{D}_k \mathbb{E}\{\widehat{\mathbf{h}}_k \widehat{\mathbf{h}}_k^{\mathrm{H}}\} \mathbf{D}_k \right), \\ &\approx p_k \cdot \mathrm{tr}\left(\mathbb{E}\{\mathbf{Z}_k\}^{-1} \mathbf{D}_k \mathbb{E}\{\widehat{\mathbf{h}}_k \widehat{\mathbf{h}}_k^{\mathrm{H}}\} \mathbf{D}_k \right). \end{aligned}$$

Since $\mathbb{E}\{\widehat{\mathbf{h}}_k \widehat{\mathbf{h}}_k^{\mathrm{H}}\} = \mathbf{Q}_k - \mathbf{C}_k$, according to the distribution in (22), we finally obtain the ergodic SINR expression, which approximates $\mathbb{E}\{\Gamma_k\}$:

$$\Gamma_k^{\mathrm{erg}} = p_k \cdot \mathrm{tr}\left[\overline{\mathbf{Z}}_k^{-1}\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)\mathbf{D}_k\right] \approx \mathbb{E}\{\Gamma_k\}, \tag{33}$$

where

$$\overline{\mathbf{Z}}_k = \mathbb{E}\{\mathbf{Z}_k\} = \sum_{i \in \mathcal{U}\setminus\{k\}} p_i\mathbf{D}_k\mathbf{Q}_i\mathbf{D}_k + p_k\mathbf{D}_k\mathbf{C}_k\mathbf{D}_k + \sigma_{\mathrm{n}}^2\mathbf{I}_M. \tag{34}$$

Finally, the ergodic approximation for the average SE is:

$$\eta_k^{\mathrm{erg}} = \left(1 - \frac{\tau_{\mathrm{p}}}{\tau_{\mathrm{c}}}\right)\log_2(1 + \Gamma_k^{\mathrm{erg}}). \tag{35}$$

From (34), we observe that the SINR of UE k is affected by its channel estimation errors, the correlation matrix of the other scheduled UEs and the AWGN. Furthermore, recall that the approximation $\mathbb{E}\{\Gamma_k\} \approx \Gamma_k^{\mathrm{erg}}$ is only valid when $M_k \ll U$.

Computing the ergodic SINR requires inverting an $M_k \times M_k$ matrix. To reduce computational complexity, we can disregard channel correlations between antennas in different subarrays and approximate $\mathbf{Q}_k$ with a block-diagonal matrix $\mathbf{Q}_k^{(\mathrm{bd})} = \mathrm{diag}(\mathbf{Q}_{i1}, \dots, \mathbf{Q}_{iL})$. This approximation allows us to obtain the SINR for UE k by inverting only the smaller L_k matrices $N \times N$, thereby significantly reducing the computational burden. The resulting SINR is

$$\Gamma_k^{\mathrm{erg\text{-}bd}} = p_k \cdot \sum_{l \in \mathcal{L}_k} \mathrm{tr}\left[\overline{\mathbf{Z}}_{kl}^{-1}(\mathbf{Q}_{kl} - \mathbf{C}_{kl})\right], \tag{36}$$

where

$$\overline{\mathbf{Z}}_{kl} = \sum_{i \in \mathcal{U}\setminus\{k\}} p_i\mathbf{Q}_{il} + p_k\mathbf{C}_{kl} + \sigma_{\mathrm{n}}^2\mathbf{I}_N. \tag{37}$$

3.1 No Pilot Reuse

Equation (33) accommodates both scenarios—with and without pilot reuse—but remains valid only when $M_k \ll U$. By enforcing mutually orthogonal pilots, we can derive an analytical expression for the SINR that is applicable regardless of the number of scheduled UEs. In the absence of pilot reuse, the LS and MMSE channel estimators exhibit similar performance, while the ZF and MMSE receive combiners achieve approximately the same SINR. Consequently, the SINR expression presented in this section is derived under the assumptions of LS channel estimation, ZF receive combiner, and no pilot reuse.

Suppose $\mathbf{A}$ is a random $M \times M$ matrix independent of $\widehat{\mathbf{h}}_i$ and $\widehat{\mathbf{h}}_j$. Considering the law of large numbers [45, Lemma 4]:

$$\widehat{\mathbf{h}}_i^{\mathrm{H}}\mathbf{A}\widehat{\mathbf{h}}_j \xrightarrow{M_k\to\infty} \mathbb{E}\{\widehat{\mathbf{h}}_i^{\mathrm{H}}\mathbf{A}\widehat{\mathbf{h}}_j\} = \mathrm{tr}(\mathbf{A}\mathbb{E}\{\widehat{\mathbf{h}}_j\widehat{\mathbf{h}}_i^{\mathrm{H}}\}), \tag{38}$$

and applying (22) and (32):

$$\widehat{\mathbf{h}}_j^{\mathrm{H}} \mathbf{A} \widehat{\mathbf{h}}_i \xrightarrow{M_k \to \infty} \begin{cases} \mathrm{tr}[\mathbf{A}(\mathbf{Q}_i - \mathbf{C}_i)] & i = j, \\ \mathrm{tr}[\mathbf{A}(\overline{\mathbf{h}}_i \overline{\mathbf{h}}_j^{\mathrm{H}} + \mathbf{B}_{ij})] & t_i = t_j, \\ \mathrm{tr}[\mathbf{A} \overline{\mathbf{h}}_i \overline{\mathbf{h}}_j^{\mathrm{H}}] & t_i \neq t_j. \end{cases} \tag{39}$$

The second term in the denominator of (26) is often neglected in the literature, as many studies assume perfect CSI. For systems with a large number of antennas, this term can be approximated using (38) as follows:

$$|\mathbf{v}_k^{\mathrm{H}} \widetilde{\mathbf{h}}_i|^2 = \widetilde{\mathbf{h}}_i^{\mathrm{H}} \mathbf{v}_k \mathbf{v}_k^{\mathrm{H}} \widetilde{\mathbf{h}}_i \xrightarrow{M_k \to \infty} \mathrm{tr}(\mathbf{v}_k \mathbf{v}_k^{\mathrm{H}} \mathbf{C}_i) = \frac{\sigma_{\mathrm{n}}^2}{p_i \tau_{\mathrm{p}}} ||\mathbf{v}_k||^2, \tag{40}$$

considering that the covariance matrix of the channel estimation error for LS channel estimation is given by

$$\mathbf{C}_k = \frac{\sigma_{\mathrm{n}}^2}{p_k \tau_{\mathrm{p}}} \mathbf{I}_M. \tag{41}$$

The equality $\mathbf{v}_k^{\mathrm{H}} \widetilde{\mathbf{h}}_i = \mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widetilde{\mathbf{h}}_i$ holds because $\mathbf{D}_k \mathbf{v}_k = \mathbf{v}_k$, since the entries of $\mathbf{v}_k$ corresponding to SAs not serving UE k are zero. The above approximation is valid since $\mathbf{v}_k$ and $\widetilde{\mathbf{h}}_i$ are not correlated, since the combining vectors depend solely on the channel estimates rather than on the channel estimation errors. Furthermore, because the ZF combiner ensures

$$|\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_i|^2 = \begin{cases} 1, & \text{if } i = k, \\ 0, & \text{if } i \neq k, \end{cases}$$

and $|\mathcal{U}| = \tau_{\mathrm{p}}$ when all users are assigned orthogonal pilots, replacing $|\mathbf{v}_k^{\mathrm{H}} \mathbf{D}_k \widetilde{\mathbf{h}}_i|^2$ in (26) by $\frac{\sigma_{\mathrm{n}}^2}{p_i \tau_{\mathrm{p}}} ||\mathbf{v}_k||^2$ gives

$$\Gamma_k^{(\mathrm{ort})} \xrightarrow{M_k \to \infty} \frac{p_k}{2\sigma_{\mathrm{n}}^2 ||\mathbf{v}_k||^2} = \frac{p_k}{2\sigma_{\mathrm{n}}^2 \left[(\widehat{\mathbf{H}}^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{H}})^{-1} \right]_{k,k}}, \tag{42}$$

since the ZF combining vector of UE k is given by

$$\mathbf{v}_k = [\mathbf{D}_k \widehat{\mathbf{H}} (\widehat{\mathbf{H}}^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{H}})^{-1}]_{:,k}, \tag{43}$$

being $\widehat{\mathbf{H}} = [\widehat{\mathbf{h}}_1, \dots, \widehat{\mathbf{h}}_K]$ the estimated channel matrix of all UEs. The (k, k)-th entry of $(\widehat{\mathbf{H}}^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{H}})^{-1}$ can be re-written as [46]

$$\left[(\widehat{\mathbf{H}}^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{H}})^{-1} \right]_{k,k} = (\widehat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{D}_k \widehat{\mathbf{h}}_k - \widehat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{F}_k \widehat{\mathbf{h}}_k)^{-1}, \tag{44}$$

where $\mathbf{F}_k = \mathbf{D}_k \ddot{\mathbf{H}}_k (\ddot{\mathbf{H}}_k^{\mathrm{H}} \mathbf{D}_k \ddot{\mathbf{H}}_k)^{-1} \ddot{\mathbf{H}}_k^{\mathrm{H}} \mathbf{D}_k$ and $\ddot{\mathbf{H}}_k \in \mathbb{C}^{M_k \times (K-1)}$ can be obtained directly from $\widehat{\mathbf{H}}$, by removing its k-th column: $\ddot{\mathbf{H}}_k = [\widehat{\mathbf{h}}_1, \dots, \widehat{\mathbf{h}}_{k-1}, \widehat{\mathbf{h}}_{k+1}, \cdots, \widehat{\mathbf{h}}_K]$.

The off-diagonal terms of $\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k$ represent the trace of the cross-correlation matrix of the estimated channels of different users, say UEs i and j, corresponding to the second or third situation described in (39), depending on whether UEs i and j share the same pilot sequence. However, in practice, different UEs are typically assigned distinct pilots. Furthermore, the large dimensions of the XL-MIMO antenna array mean that the LoS regions of different UEs overlap on only a few antennas or may not overlap at all. As a result, the off-diagonal terms of $\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k$ are generally much smaller than the diagonal terms, especially as the number of antennas increases. Therefore, it is reasonable to approximate these off-diagonal terms as zero, and the following approximation holds:

$$\widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{F}_k\widehat{\mathbf{h}}_k \xrightarrow{M_k\to\infty} \widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{G}_k\widehat{\mathbf{h}}_k, \tag{45}$$

where $\mathbf{G}_k = \mathbf{D}_k\ddot{\mathbf{H}}_k\mathbf{T}_k^{-1}\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k$ and $\mathbf{T}_k$ is obtained by retaining the diagonal entries of $\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k$:

$$(\mathbf{T}_k)_{i,j} = \begin{cases} (\ddot{\mathbf{H}}_k^{\mathrm{H}}\mathbf{D}_k\ddot{\mathbf{H}}_k)_{i,i} & \text{for } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{46}$$

On the other hand, the approximation in (45) is not reasonable for conventional M-MIMO systems, where the antennas are co-located such that the LoS region of a given UE encompasses all the M BS antennas or none of them.

Replacing (44) into (42), and using the approximation in (45) results in

$$\Gamma_k^{(\mathrm{ort})} \xrightarrow{M_k\to\infty} \frac{p_k}{2\sigma_{\mathrm{n}}^2}(\widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k - \widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{G}_k\widehat{\mathbf{h}}_k). \tag{47}$$

Applying (39) on the terms of (47) gives the approximations

$$\widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k \xrightarrow{M_k\to\infty} \mathrm{tr}[\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)], \tag{48}$$

and

$$\widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{G}_k\widehat{\mathbf{h}}_k \xrightarrow{M_k\to\infty} \mathrm{tr}[\mathbf{G}_k(\mathbf{Q}_k - \mathbf{C}_k)]. \tag{49}$$

By simple algebraic manipulation over (49) [46]:

$$\mathrm{tr}[\mathbf{G}_k(\mathbf{Q}_k - \mathbf{C}_k)] = \sum_{i\neq k} \frac{\widehat{\mathbf{h}}_i^{\mathrm{H}}\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)\mathbf{D}_k\widehat{\mathbf{h}}_i}{\widehat{\mathbf{h}}_i^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_i}. \tag{50}$$

Applying (39) over the numerator and the denominator of (50) results in

$$\widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{G}_k\widehat{\mathbf{h}}_k \xrightarrow{M_k\to\infty} \sum_{i\neq k} \frac{\mathrm{tr}[\mathbf{D}_k(\mathbf{Q}_k - \mathbf{C}_k)\mathbf{D}_k(\mathbf{Q}_i - \mathbf{C}_i)]}{\mathrm{tr}[\mathbf{D}_k(\mathbf{Q}_i - \mathbf{C}_i)]}. \tag{51}$$

Finally, by substituting (48) and (51) into (47), we derive the asymptotic SINR expression presented in (52), valid under the condition of no pilot reuse.

$$\Gamma_k^{\text{asymp}} \xrightarrow{M_k \to \infty} \frac{p_k}{2\sigma_n^2}\left[\operatorname{tr}\left(\mathbf{D}_k\left(\mathbf{Q}_k - \mathbf{C}_k\right)\right) - \sum_{\substack{i \in \mathcal{U} \\ i \neq k}} \frac{\operatorname{tr}\left(\mathbf{D}_k\left(\mathbf{Q}_k - \mathbf{C}_k\right)\mathbf{D}_k\left(\mathbf{Q}_i - \mathbf{C}_i\right)\right)}{\operatorname{tr}\left(\mathbf{D}_k\left(\mathbf{Q}_i - \mathbf{C}_i\right)\right)}\right]. \tag{52}$$

Furthermore, by replacing $\mathbf{C}_k$ as defined in (41), the SINR expression is transformed into (53).

$$\Gamma_k^{\text{asymp}} \xrightarrow{M_k \to \infty} \frac{1}{2\tau_{\text{p}}\sigma_{\text{n}}^2}\left[\operatorname{tr}\left(\mathbf{D}_k\left(p_k\tau_{\text{p}}\mathbf{Q}_k - \sigma_{\text{n}}^2\mathbf{I}_M\right)\right) - \sum_{\substack{i \in \mathcal{U} \\ i \neq k}} \frac{\operatorname{tr}\left(\mathbf{D}_k\left(p_k\tau_{\text{p}}\mathbf{Q}_k - \sigma_{\text{n}}^2\mathbf{I}_M\right)\mathbf{D}_k\left(p_i\tau_{\text{p}}\mathbf{Q}_i - \sigma_{\text{n}}^2\mathbf{I}_M\right)\right)}{\operatorname{tr}\left(\mathbf{D}_k\left(p_i\tau_{\text{p}}\mathbf{Q}_i - \sigma_{\text{n}}^2\mathbf{I}_M\right)\right)}\right]. \tag{53}$$

Finally, the asymptotic approximation for the instantaneous SE is:

$$\eta_k^{\text{asymp}} = \left(1 - \frac{\tau_{\text{p}}}{\tau_{\text{c}}}\right)\log_2(1 + \Gamma_k^{\text{asymp}}). \tag{54}$$

Each element of the matrix product $\mathbf{D}_k\mathbf{Q}_k\mathbf{Q}_i\mathbf{D}_k$ requires NL_k complex multiplications. Consequently, computing the trace $\operatorname{tr}(\mathbf{D}_k\mathbf{Q}_k\mathbf{Q}_i\mathbf{D}_k)$, which involves only the diagonal elements of $\mathbf{D}_k\mathbf{Q}_k\mathbf{Q}_i\mathbf{D}_k$, demands $(NL_k)^2$ complex multiplications. To approximate the SINR as in (52), this trace must be calculated $U-1$ times - once for each UE i in the scheduling set $\mathcal{U}$, excluding UE k. This results in a total computational complexity of $(U-1)(NL_k)^2$.

To reduce this complexity, we approximate the correlation matrices $\mathbf{Q}_k$ and $\mathbf{Q}_i$ by $\ddot{\mathbf{Q}}_k$ and $\ddot{\mathbf{Q}}_i$, respectively, retaining only their diagonal entries and setting all off-diagonal elements to zero. With this approximation, the SINR expression in (52) simplifies to (55), which eliminates the need for matrix multiplication and thereby reduces the computational complexity to near zero. Furthermore, the asymptotic SINR in (55) yields the same result as (52) under conditions of uncorrelated fading and zero LoS probability.

$$\Gamma_k^{\text{asymp-bd}} \xrightarrow{M_k \to \infty} \frac{1}{2\tau_{\text{p}}\sigma_{\text{n}}^2}\left[\sum_{l \in \mathcal{L}_k}\left(p_k\tau_{\text{p}}\beta_{kl} - \sigma_{\text{n}}^2\right) - \sum_{\substack{i \in \mathcal{U} \\ i \neq k}} \frac{\sum_{l \in \mathcal{L}_k}\left(p_k\tau_{\text{p}}\beta_{kl} - \sigma_{\text{n}}^2\right)\left(p_i\tau_{\text{p}}\beta_{il} - \sigma_{\text{n}}^2\right)}{\sum_{l \in \mathcal{L}_k}\left(p_i\tau_{\text{p}}\beta_{il} - \sigma_{\text{n}}^2\right)}\right]. \tag{55}$$

4 User Scheduling

The US procedure is executed in two steps. In the first step, τ_p out of the K UEs are assigned orthogonal pilot sequences. The objective is to maximize the sum SE by scheduling as many UEs as possible with mutually orthogonal pilots. Given that XL-MIMO systems often involve UEs with significantly different VRs, we can further enhance SE by allowing pilot reuse, in the second step. This approach enables the scheduling of additional UEs that were not accommodated in the first step, resulting in the set $\mathcal{U}$. Figure 2 illustrates the proposed US procedure for XL-MIMO systems.

4.1 User Scheduling: Step 1 (Orthogonal Pilots)

The first step of the US algorithm corresponds to solving the following optimization problem:

$$\max_{\mathcal{U}} \quad \left(1 - \frac{|\mathcal{U}|}{\tau_\mathrm{c}}\right) \sum_{k \in \mathcal{U}} \log_2(1 + \Gamma_k) \tag{56a}$$

$$\text{s.t} \quad ||\boldsymbol{\phi}_{t_i}||^2 = |\mathcal{U}|, \quad \forall i \in \mathcal{U}, \tag{56b}$$

$$\boldsymbol{\phi}_{t_i}^{\mathrm{H}} \boldsymbol{\phi}_{t_j} = 0, \qquad \forall i, j \in \mathcal{U}, t_i \neq t_j, \tag{56c}$$

$$\mathcal{U} \subseteq \{1, \dots, K\}. \tag{56d}$$

Constraints (56b) and (56c) ensure that all UEs in $\mathcal{U}$ are assigned mutually orthogonal pilot sequences, such that the pilot sequence length is equal to the number of scheduled UEs, *i*.e., $\tau_\mathrm{p} = |\mathcal{U}|$. By solving (56), we implicitly optimize the pilot sequence length, balancing the trade-off between increasing the number of scheduled users (which enhances the summation in the objective function) and decreasing

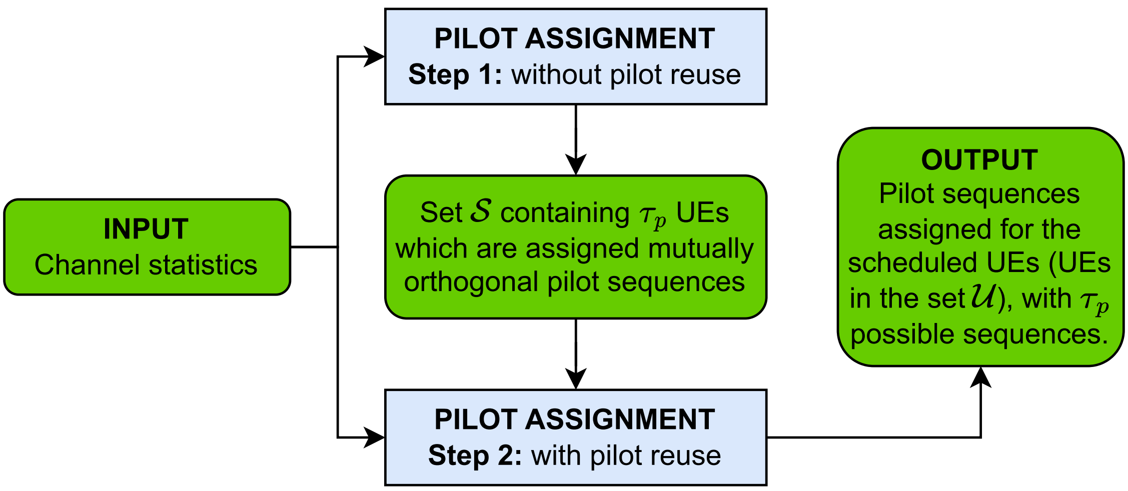


Fig. 2 Two-step user scheduling for XL-MIMO systems

the fraction of the coherence block available for data transmission (as the pre-log factor decreases). Consequently, an optimal τ_p exists that maximizes the sum SE.

Algorithm 1 describes the operation of the first step of the proposed US algorithm to solve (56). It consists of admitting the UEs with the strongest average LSF channel gain, one by one, and checking the sum SE after each new UE based on the SINR expression in (52). The algorithm stops admitting more UEs before the sum SE starts to decrease, ensuring sub-optimal SE performance.

Instead of relying on instantaneous channels, the deterministic SINR expressions that were derived in Sect. 3 are based on channel statistics, specifically the channel mean vectors and covariance matrices, which are assumed to be perfectly known due to their slow variation. Consequently, it is unnecessary to execute Algorithm 1 within each channel coherence block, significantly reducing computational resources as each US solution remains valid across hundreds of coherence blocks.

4.2 User Scheduling: Step 2 (Non-orthogonal Pilots)

In the second step, the US algorithm schedules UEs that were not scheduled in Step 1 by allowing pilot reuse. The number and length of pilot sequences are maintained as determined by Algorithm 1. The operation of the second step of the proposed US method is detailed in Algorithm 2. The final set of UEs scheduled after both steps is denoted as $\mathcal{U}$.

Similar to Algorithm 1, UEs are scheduled in descending order of their LSF channel gains. For each UE $k \notin \mathcal{U}$, the algorithm assigns the pilot sequence t_k that maximizes the ergodic SINR, as defined in (33), considering that other UEs have already been assigned each of the τ_p pilot sequences in the first step. The sum SE is evaluated based on the computed SINR for the scheduled UEs. The algorithm terminates before the sum SE starts to decrease, which can occur due to increased interference, particularly in densely populated XL-MIMO scenarios.

Algorithm 1 User Scheduling for XL-MIMO (Step 1)

Input:

- $\mathbf{Q}_i$ for each $i \in \{1, \ldots, K\}$: channel parameters
- $\mathbf{D}_i$ for each $i \in \{1, \ldots, K\}$: matrices indicating which SAs serve each UE
- p_i for each $i \in \{1, \ldots, K\}$: uplink (UL) transmit power
- σ_{n}^2: noise power
- τ_{c}: number of samples per coherence block

Output:

- $\mathcal{U}$: set of scheduled UEs
- t_i for each $i \in \mathcal{U}$: pilot sequences
- $\mathbf{C}_i$ for each $i \in \mathcal{U}$: channel estimation error matrices
- $\mathrm{SE}^{(\tau_{\mathrm{p}})}$: sum SE after running step 1

Initialize the set of scheduled UEs: $\mathcal{U} \leftarrow \emptyset$
Initialize the sum SE: $\mathrm{SE}^{(0)} \leftarrow 0$
Initialize the number of pilot sequences: $\tau_{\mathrm{p}} \leftarrow 0$
Compute $\mathbf{C}_i$, for each $i \in \{1, \ldots, K\}$, as in (41), using τ_{p}, p_i and σ_{n}^2
while $\tau_{\mathrm{p}} \leq K$ **do**
    Select UE k with the highest LSF coefficient among unscheduled UEs:

$$k \leftarrow \arg \max_{i \in \{1,\ldots,K\} \setminus \mathcal{U}} \mathrm{tr}(\mathbf{Q}_i)$$

    Update the set of scheduled UEs: $\mathcal{U} \leftarrow \mathcal{U} \cup \{k\}$
    Update the number of pilot sequences: $\tau_{\mathrm{p}} \leftarrow \tau_{\mathrm{p}} + 1$
    Compute the asymptotic SINR of the UEs in $\mathcal{U}$, eq. (52), using $\mathbf{Q}_i$, $\mathbf{C}_i$, $\mathbf{D}_i$, τ_{p}, p_i and σ_{n}^2, with $i \in \mathcal{U}$
    Calculate the sum SE, eq. (27), using τ_{p}, τ_{c} and the SINRs, and assign to $\mathrm{SE}^{(\tau_{\mathrm{p}})}$
    if $\mathrm{SE}^{(\tau_{\mathrm{p}})} < \mathrm{SE}^{(\tau_{\mathrm{p}}-1)}$ **then**
        Reverse last update: $\mathcal{U} \leftarrow \mathcal{U} \setminus \{k\}$ and $\tau_{\mathrm{p}} \leftarrow \tau_{\mathrm{p}} - 1$
        Exit the loop
    else
        Assign a pilot to UE k: $t_k \leftarrow \tau_{\mathrm{p}}$
    end if
end while

Algorithm 2 User Scheduling (Step 2)

Input:

- $\mathcal{U}$: set of scheduled UEs
- t_i for each $i \in \mathcal{U}$: pilot sequences
- $\mathbf{C}_i$ for each $i \in \mathcal{U}$: channel estimation error matrices
- $\mathrm{SE}^{(\tau_\mathrm{p})}$: sum SE after running step 1
- $\mathbf{Q}_i$ for each $i \in \{1, \ldots, K\}$: channel parameters
- $\mathbf{D}_i$ for each $i \in \{1, \ldots, K\}$: matrices indicating which SAs serve each UE
- p_i for each $i \in \{1, \ldots, K\}$: UL transmit power
- σ_n^2: noise power
- τ_p: pilot sequences length
- τ_c: number of samples per coherence block

Output:

- $\mathcal{U}$: scheduled UEs
- t_i for each $i \in \mathcal{U}$: pilot sequences

Initialize the number of scheduled UEs: $U \leftarrow |\mathcal{U}|$.
while $U \leq K$ **do**
  Select UE k with the highest LSF coefficient among unscheduled UEs:
$$k \leftarrow \arg \max_{i \in \{1,\ldots,K\} \setminus \mathcal{U}} \beta_i$$
  Update the set of scheduled UEs: $\mathcal{U} \leftarrow \mathcal{U} \cup \{k\}$
  Update the number of scheduled UEs: $U \leftarrow U + 1$
  for $t \leftarrow 1$ **to** τ_p **do**
    Assign pilot t for UE k: $t_k \leftarrow t$
    Compute the channel estimation error matrices for each UE that use pilot t, and UE k, using (21), and assign the results to $\mathbf{C}_i^{(t)}$
    Compute the SINRs, applying (33), and utilizing $\mathbf{C}_i^{(t)}$ for each UE in $\mathcal{U}$, and assign the results to $\overline{\Gamma}_k(t)$
    Compute the sum SE among the UEs in $\mathcal{U}$, using (27), and assign the result to $\mathrm{SE}^{(U,t)}$
  end for
  Assign a pilot to UE k:
$$t_k \leftarrow \arg \max_{t \in \{1,\ldots,\tau_\mathrm{p}\}} \mathrm{SE}^{(U,t)}$$
  Update the sum SE: $\mathrm{SE}^{(U)} \leftarrow \mathrm{SE}^{(U,t_k)}$
  Update the covariance matrices of the UEs that utilize pilot t_k, according to (21): $\mathbf{C}_i \leftarrow \mathbf{C}_i^{(t_k)}$, for each $i \in \mathcal{P}_k$
  if $\mathrm{SE}^{(U)} < \mathrm{SE}^{(U-1)}$ **then**
    Undo the last update: $\mathcal{U} \leftarrow \mathcal{U} \setminus \{k\}$ and $U \leftarrow U - 1$
    exit the loop
  end if
end while

5 Numerical Results

Monte Carlo numerical simulations reveal the accuracy of the low-complexity SINR expressions and the performance of Algorithms 1 and 2. The massive antenna array containing L uniformly spaced SAs centered on the origin of the Cartesian plane serving K UEs randomly dropped on a rectangular area, extending in the range [±100] m on both horizontal dimensions. Table 2 displays the adopted simulation parameters.

Table 2 Adopted simulation parameters values

Parameter	Value
Length of the channel coherence block: τ_c	50
Number of antennas per subarray: N	4
NLoS path-loss attenuation exponent: γ	4
Shadowing standard deviation: σ_s^2	3 dB (LoS)
	4 dB (NLoS)
Azimuth angle standard deviation: σ_φ	10°
Elevation angle standard deviation: σ_θ	10°
Average channel gain at a distance of 1 m: β_0	$8.9125 \cdot 10^{-4}$
Array length	100 m
Wavelength: λ	12.5 cm
Intra-SA antenna separation: d	$\lambda/2$
UE transmit power: $p_1, \ldots, p_K$	10 dBm
Noise power: σ_n^2	– 96 dBm
Number of Monte-Carlo channel realizations	1000

For the micro-urban scenario, the probability of existing LoS between UE k and SA l, i.e., $\alpha_{kl} = 1$ depends on the distance d_{kl}, and is modeled according to [47] as:

$$\Pr(\alpha_{kl} = 1) = \min\left(\frac{18\text{ m}}{d_{kl}}, 1\right) \cdot \left[1 - e^{-\frac{d_{kl}}{36\text{ m}}}\right] + e^{-\frac{d_{kl}}{36\text{ m}}} \tag{57}$$

According to (57), the propagation is always in LoS conditions for $d_{kl} \leq 18$ m, i.e., $q = 1$, while, for $d_{kl} > 18$ m, the LoS probability decays exponentially. Considering the LoS probability given by Eq. (57) and the adopted simulation parameters for array length and cell dimensions, we obtained an average LoS probability of 36%. It represents an expected condition of the XL-MIMO system operating in micro-urban channel scenarios. From the trustworthy numerical simulations perspective, we cannot disregard the LoS component in urban crowded channel scenarios since most UEs are close to the BS array, which is much longer than co-located mMIMO arrays.

5.1 Accuracy of the Closed-Form SINR Approximations

Since we aim for η_k^{erg} to approximate $\mathbb{E}\{\eta_k\}$ when $U \gg M_k$, we quantify the accuracy of estimating the mean SE using (35) by the normalized absolute error (NAE), defined as[2]

[2] Here and everywhere in this paper, the expectation operator is taken over the SSF variations.

$$\mathrm{NAE}(\mathbb{E}\{\eta_k\}, \eta_k^{\mathrm{erg}}) = \frac{|\eta_k^{\mathrm{erg}} - \mathbb{E}\{\eta_k\}|}{\mathbb{E}\{\eta_k\}}. \tag{58}$$

On the other hand, the accuracy associated to approximating η_k by $\hat{\eta}_k^{\mathrm{asymp}}$ is measured by the mean normalized absolute error (MNAE), which is given by

$$\mathrm{MNAE}(\eta_k, \eta_k^{\mathrm{asymp}}) = \mathbb{E}\left\{ \frac{|\eta_k^{\mathrm{asymp}} - \eta_k|}{\eta_k} \right\}. \tag{59}$$

Figure 3a analyzes the probability of the asymptotic SINR expression returning a negative value, i.e., $\Pr(\Gamma_k^{\mathrm{asymp}} < 0)$, while Fig. 3b plots the MNAE for different number of UEs. Figure 3 demonstrates that, for a relatively large number of antennas, the proposed asymptotic SINR approximation is almost surely non-negative and the MNAE of the asymptotic SE is sufficiently small. The greater the number of antennas (M), the more evident favorable propagation and channel hardening will be, resulting in smaller variations in the instantaneous SINR around its large-scale average. Therefore, the approximated SINR expression, which depends on the channel statistics, will be more accurate. On the other hand, increasing the number of UEs (K) causes the opposite effect.

Figure 4 illustrates the accuracy of the analytical SINR expressions. Both original and block-diagonal versions of the analytical expressions are considered. We compare the scenarios with and without pilot reuse, with $\tau_{\mathrm{p}} = U/2$ and $\tau_{\mathrm{p}} = U$, respectively. In this simulation, we consider $L = 32$ subarrays, with each user being served by $L_k = 4$ subarrays serving each user (specifically, those with the best LSF channel gains), and a channel coherence block length of $\tau_{\mathrm{c}} = 100$ symbols. The results are average over the scheduled users' SEs and 50 LSF channel realizations. For both the

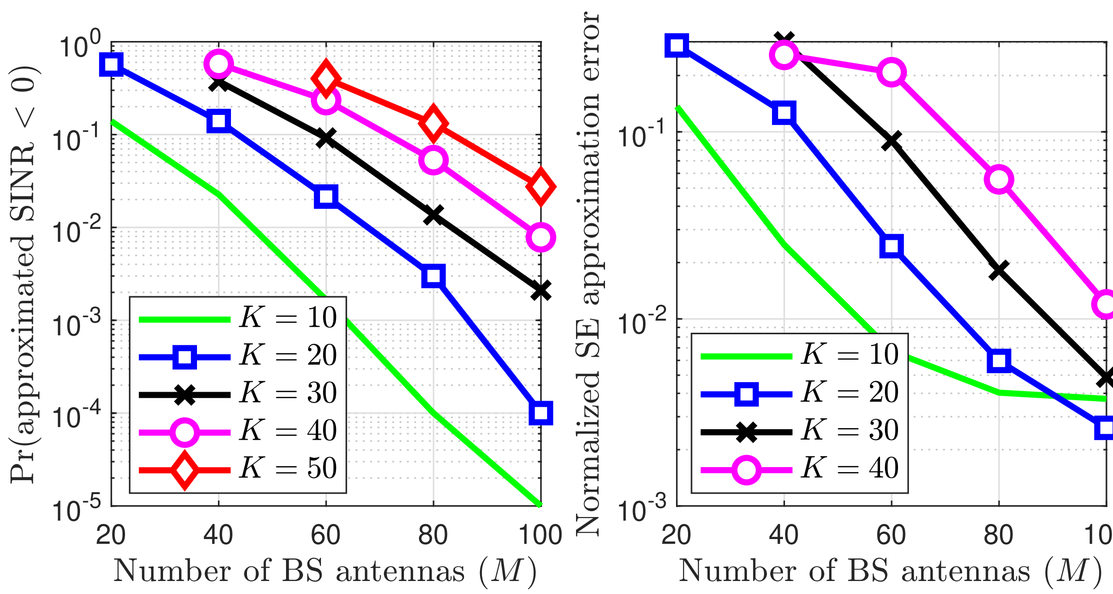
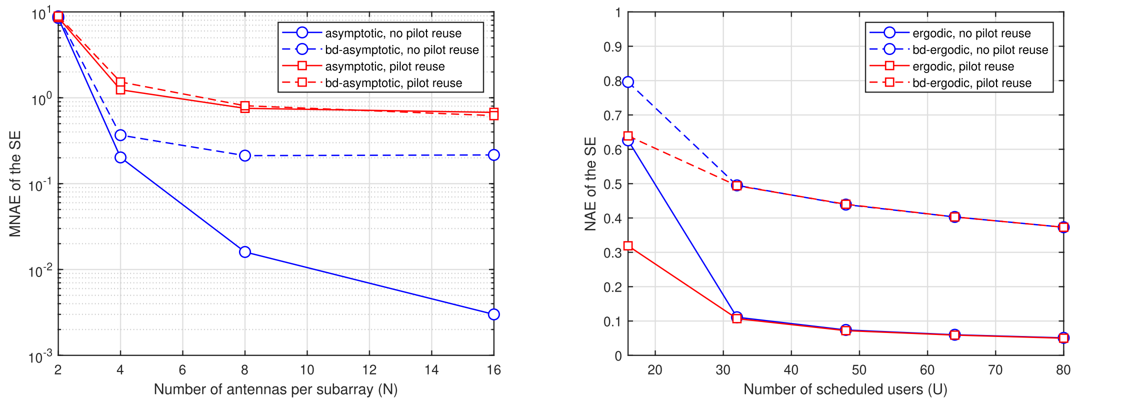


Fig. 3 Accuracy of the approximated SINR expression in (52), when all UEs are scheduled, regarding $\Pr(\Gamma_{k,\mathcal{S}}^{\approx} < 0)$ and the MNAE over the SE based on the asymptotic SINR. Points where $M < K$ were not simulated



(a) Asymptotic SE. Number of scheduled users: $U = 8$. (b) Ergodic SE. Number of antennas per subarray: $N = 4$.

Fig. 4 NAE of the ergodic SE and MNAE of the asymptotic SE. $L = 32$, $L_k = 4$, $\tau_c = 100$

NAE and MNAE, the expectations are computed as the average over 50 SSF realizations per LSF realization, resulting in a total of 1,000 channel realizations.

Figure 4a presents the accuracy of the asymptotic SE. In this scenario, $U = 8$ users are scheduled. Since each user is served by $L_k = 4$ SAs, the number of antennas serving each user, $M_k = NL_k$, ranges from 8 to 64. In the absence of pilot reuse, the asymptotic SE becomes highly accurate as the number of antennas increases, particularly when $U \leq M_k$. Although the block-diagonal asymptotic expression is notably less precise, it remains a viable option when computational resources are very limited. Conversely, as expected, the asymptotic expressions lose accuracy when pilot reuse is employed.

Figure 4b shows the accuracy of the ergodic SE. Here, each SA comprises $N = 4$ antennas, so that a total of $M_k = 16$ antennas serve each UE. It is observed that, regardless of pilot reuse, the ergodic SE is an accurate approximation of the average instantaneous SE when $U > M_k$. The block-diagonal ergodic expression is significantly less accurate, though it can be utilized under conditions of severely constrained computational resources.

Next, we analyze the effect of **Step 1** and by aggregating the **Step 2** in the user scheduling XL-MIMO systems. Our numerical results are organized into two parts:

Part 1: accuracy of the four SINR expressions in two scenarios: (a) all K users are scheduled with mutually orthogonal pilots; (b) $K/2$ orthogonal pilots are assigned randomly. Such an accuracy is measured as:

- Fig. 5: ΣSE$\times$ Scheduled UEs
- Fig. 6a: Scheduled UEs $\times K$
- Fig. 6b: ΣSE $\times$ K
- Fig. 7: Probability of a UE being scheduled $\times$ Fraction of SA in LoS.

Part 2: performance of pilot assignment algorithm with pilot reuse combining Algorithm 1 and .

- Fig. 8: Σ SE $\times$ # scheduled users with pilot reuse.

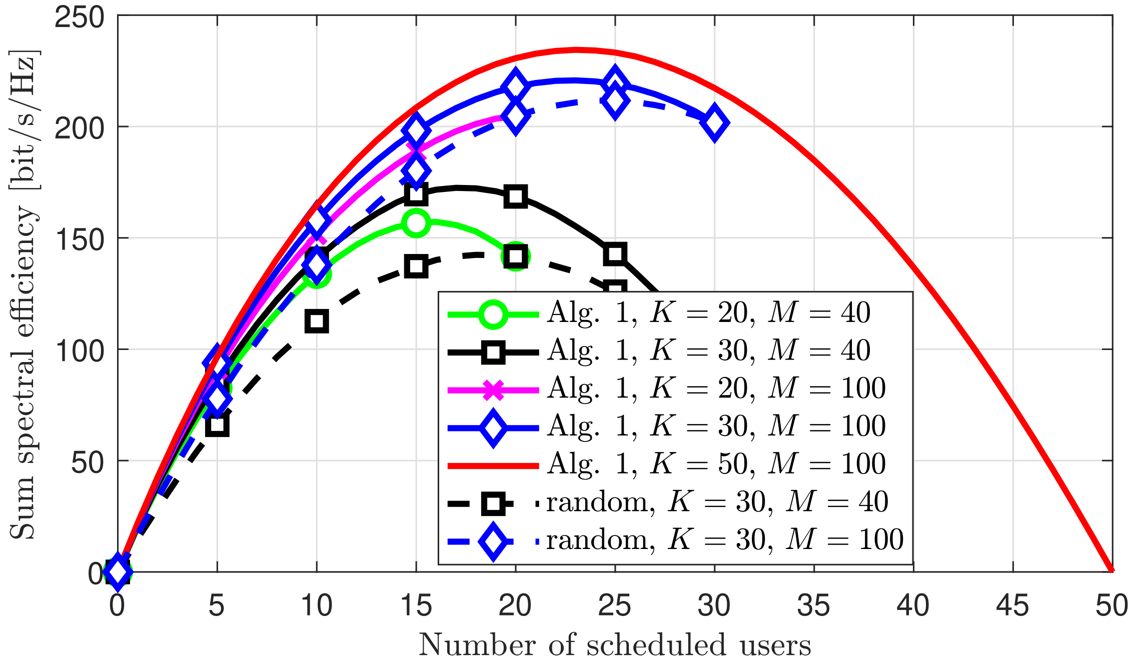


Fig. 5 SE dependency on $|\mathcal{S}|$, varying M and K. Two US criteria are compared: random, and highest LSF gain (Algorithm 1)

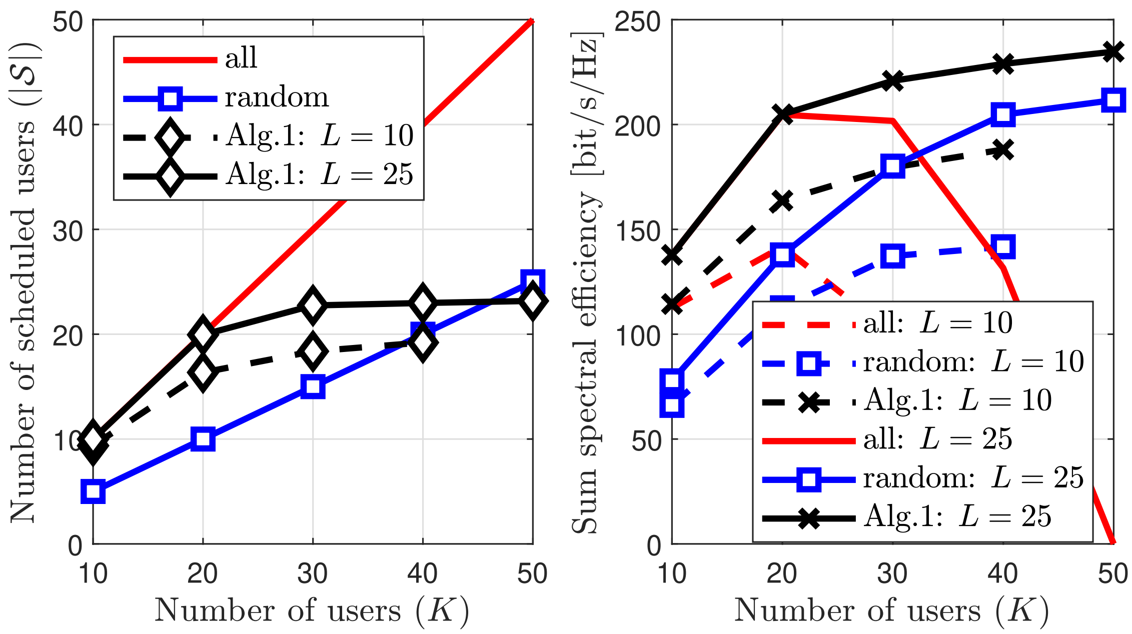


Fig. 6 Number of scheduled UEs ($|\mathcal{S}|$) and sum SE, depending on K; all-scheduled, random scheduling and Algorithm 1 strategies are compared, for different number of SAs (L); $N = 4$

5.2 User Scheduling Analysis: Step 1

Figure 5 reveals the dependency of the sum SE on the length of the pilot sequences, τ_p, which herein is also the number of UEs that are scheduled with orthogonal pilots, $|\mathcal{S}|$. We compare two cases: (a) the UEs are scheduled by prioritizing the ones with the strongest LSF channel gain, as in Algorithm 1, and (b)

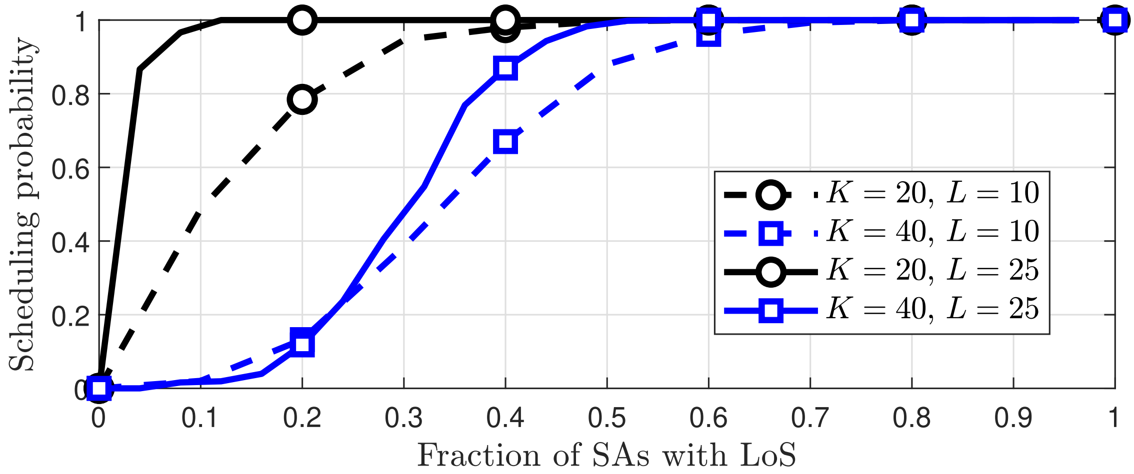
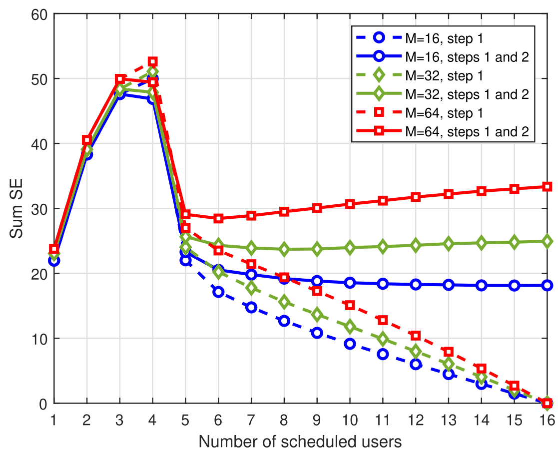
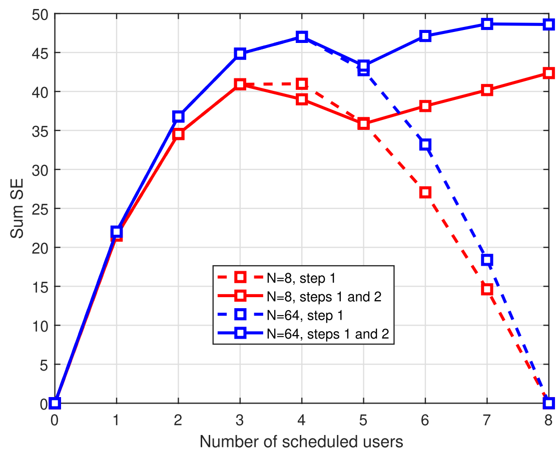


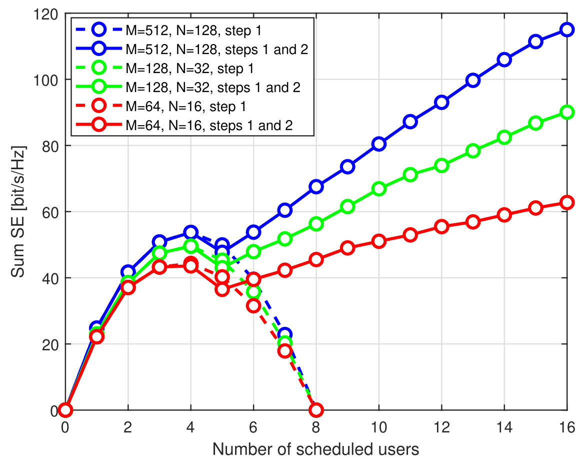
Fig. 7 Probability of a UE being scheduled by Algorithm 1 (Step 1), depending on the fraction of SAs with LoS; $N = 4$



(a) $\tau_c = 16,\ K = 16,\ N = 4,\ M \in [16; 32; 64]$.



(b) $\tau_c = 8,\ K = 8,\ M = 64;\ N \in [8; 64]$.



(c) $\tau_c = 8,\ K = 16,\ M \in [64; 128; 512]$ and $N \in [16; 32; 128]$.

Fig. 8 User schedules with pilot reuse (Steps 1 and 2). Sum-SE ×# scheduled users plots are depicted as a function of (parameterized): **a** increasing number of BS antennas, M; **b** different number of SA, N; **c** combining increasing number of BS antennas and subarrays M and N

the UEs are scheduled randomly with 50% scheduling probability. The first one performed considerably better than the latter. Furthermore, there is a value of τ_p that maximizes the sum SE, since as more UEs are scheduled, the summation in (27) increases, but a smaller portion of the channel coherence block is available for data transmission. Moreover, increasing M improves the SE. Increasing K also improves the SE, since there are more possibilities for choosing the τ_p strongest UEs.

Figure 6 depicts the behavior of the number of scheduled UEs and the sum SE depending on the number of UEs (K), under three US strategies: (a) all the K UEs are scheduled, (b) 50% of the K UEs are randomly scheduled, (c) the UEs are scheduled by Algorithm 1. Results show that the proposed US strategy performs better, in terms of SE, than the random US strategy, independent of the number of UEs. Furthermore, the performance of all-scheduling strategy degrades when the number of UEs approaches τ_c, which does not happen with the other strategies. Also, increasing the number of SAs at the BS (L) results in a higher sum-SE.

Figure 7 shows how the conditions of the XL-MIMO channel impact the performance of the proposed US strategy. Since Algorithm 1 gives priority to scheduling UEs with substantial channel gains, the probability of a UE being scheduled increases with the number of SAs with LoS. For example, considering the case of $K = 20$ and $L = 10$, a UE will surely be scheduled if it has LoS to 50% of the SAs, i.e., 5 out of 10 SAs. This percentage increases to 70%, (7 out of 10 SAs) when K increases from 20 to 40, since there is more concurrence. On the other hand, for $L = 25$ the scenario is more comfortable, and a UE needs to have LoS to about 12% and 50% of the SAs to be almost surely scheduled for $K = 20$ and $K = 40$, respectively.

5.3 User Scheduling Analysis: Steps 1 and 2

User scheduling admitting pilot reuse is implemented by combining Steps 1 and 2. Figure 8 explores the Σ SE $\times$ the number of scheduled users with and without pilot reuse. A drop in sum-SE occurs from $K = 4$ to $K = 5$ in the three plots, Fig. 8a–c This abrupt drop in terms of Σ SE is due to the distributed processing (combined reception done separately in each subarray), and occurs even under different number of antennas per subarray $N \in [4;16;32;64;128]$. While in Fig. 8b there is no longer an abrupt drop in Σ SE, since K is never greater than N. Figure 8c reveals the substantial increase in the Σ SE in terms of the number of scheduled users with pilot reuse (Step 1 and 2) and increasing number of BS antennas (M). As expected, increasing the number of antennas per subarray (N), substantially increases the sum-SE. However, the "valley" formed at $K = 5$ UEs always remains as N increases. Finally, Fig. 8c reveals the huge gain in terms of Σ SE when combining reuse pilot (Step 2) in user scheduling in XL-MIMO systems with increasing number of BS antennas M and number of antenna per subarray N.

6 Conclusions

The work presented in the paper focuses on user scheduling in crowded extra-large multiple-input multiple-output (XL-MIMO) systems under multi-state line-of-sight (LoS) and non-line-of-sight (NLoS) channels.Future research will likely continue to explore the synergy between advanced user scheduling techniques and AI-driven network management frameworks to unlock the full potential of 5G and B5G technologies. We successfully derived closed-form expressions for the Signal-to-Interference-Plus-Noise Ratio (SINR) in XL-MIMO systems. These expressions are based on channel statistics rather than instantaneous channel realizations, making them applicable across hundreds of coherence blocks. This approach significantly reduces computational complexity while maintaining accuracy. An efficient user scheduling (US) algorithm was proposed, which maximizes the sum spectral efficiency (SE) by assigning mutually orthogonal pilots to scheduled user equipment (UE). The algorithm effectively prioritizes UEs with higher large-scale fading (LSF) channel gains, demonstrating improved performance in crowded scenarios. The paper introduces a pilot assignment (PA) strategy that leverages visibility region (VR) information to mitigate pilot contamination. This strategy allows pilots reuse among UEs with non-overlapping VRs, thus reducing overhead and enhancing channel estimation accuracy.

The numerical results validate the proposed SINR expressions' accuracy and the US algorithm's effectiveness. The findings indicate that the proposed methods significantly enhance the sum SE in crowded XL-MIMO environments, particularly when the number of antennas is large. The study emphasizes the importance of considering both LoS and NLoS components in channel modeling, especially in urban environments where UEs are typically close to the BS subarrays of the XL-MIMO. This consideration is crucial for realistic performance assessments. Results also confirm the effectiveness of the proposed US algorithm for crowded XL-MIMO systems, in terms of improving the sum SE, and demonstrate that it really prioritizes the strongest UEs, since the scheduling probability increases with the number of SAs to which a UE has LoS.

6.1 Future Research

Addressing the following future research directions will allow the field of XL-MIMO systems to continue evolving, leading to more robust, efficient, and adaptable communication networks. Future research could explore (a) dynamic user scheduling algorithms that adapt to changing channel conditions and user mobility; (b) it could also involve real-time adjustments to pilot assignments and scheduling decisions based on instantaneous channel feedback; (c) Further research could focus on developing more sophisticated channel models that account for additional factors such as user mobility patterns, varying environmental conditions, and the impact of different types of interference; (d) user scheduling and pilot assignment aided by machine learning techniques; (e) explore hybrid architectures.

Integrating machine learning techniques for user scheduling and pilot assignment could lead to more intelligent and adaptive systems. Machine learning algorithms could analyze historical channel data to predict optimal scheduling strategies. Future studies could also address energy efficiency in XL-MIMO systems, particularly in the context of massive machine-type communications (mMTC). This could involve optimizing power allocation strategies alongside scheduling.

Conducting experimental validations in real-world scenarios would be beneficial to confirm the theoretical findings and numerical simulations. This could help understand the practical challenges and limitations of deploying XL-MIMO systems.

Research could explore hybrid architectures by combining XL-MIMO with other technologies, such as reconfigurable intelligent surfaces (RIS) or cell-free massive MIMO, aiming to enhance performance in complex environments.

Author Contributions G.A.U. developed the system model, performed the simulations, and wrote the initial draft of the manuscript. W.S.J. contributed to the problem formulation and assisted with the literature review. J.C.M. provided guidance on the XL-MIMO concepts and helped refine the methodology. T.A. supervised the project, contributed to the algorithm development, and critically revised the manuscript. All authors participated in the analysis and interpretation of the results, reviewed the manuscript, and approved the final version for submission.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Demir, T., Björnson, E., Sanguinetti, L.: Foundations of user-centric cell-free massive MIMO. Found. Trends Signal Process. **14**, 162–472 (2021). https://doi.org/10.1561/2000000109
2. Björnson, E., Sanguinetti, L., Wymeersch, H., Hoydis, J., Marzetta, T.L.: Massive MIMO is a reality—what is next? Five promising research directions for antenna arrays (2019)
3. Han, Y., Jin, S., Wen, C.-K., Ma, X.: Channel estimation for extremely large-scale massive MIMO systems. IEEE Wirel. Commun. Lett. **9**(5), 633–637 (2020). https://doi.org/10.1109/LWC.2019.2963877
4. Amiri, A., Manchón, C.N., Carvalho, E.: A message passing based receiver for extra-large scale MIMO. In: 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), pp. 564–568 (2019). https://doi.org/10.1109/CAMSAP45676.2019.9022503
5. Liu, D., Wang, J., Li, Y., Han, Y., Ding, R., Zhang, J., Jin, S., Quek, T.: Location-based visible region recognition in extra-large massive MIMO systems. IEEE Trans. Veh. Technol. **72**(6), 8186–8191 (2023). https://doi.org/10.1109/TVT.2023.3242615
6. Buzzi, S., D'Andrea, C., Fresia, M., Zhang, Y.-P., Feng, S.: Pilot assignment in cell-free massive MIMO based on the Hungarian algorithm. IEEE Wirel. Commun. Lett. **10**(1), 34–37 (2021). https://doi.org/10.1109/LWC.2020.3020003
7. Ubiali, G.A., Abrão, T., Marinello, J.C.: Improving spectral efficiency via pilot assignment and subarray selection under realistic XL-MIMO channels. Phys. Commun. **58**, 102062 (2023). https://doi.org/10.1016/j.phycom.2023.102062

8. Qiu, J., Xu, K., Xia, X., Shen, Z., Xie, W.: Downlink power optimization for cell-free massive MIMO over spatially correlated Rayleigh fading channels. IEEE Access **8**, 56214–56227 (2020). https://doi.org/10.1109/ACCESS.2020.2981967
9. Nguyen, T.H., Chien, T.V., Ngo, H.Q., Tran, X.N., Björnson, E.: Pilot assignment for joint uplink-downlink spectral efficiency enhancement in massive MIMO systems with spatial correlation. IEEE Trans. Veh. Technol. **70**(8), 8292–8297 (2021). https://doi.org/10.1109/TVT.2021.3091020
10. Fan, Y., Wu, S., Bi, X., Li, G.: Power allocation for cell-free massive MIMO ISAC systems with OTFS signal. IEEE Internet Things J. **12**(7), 9314–9331 (2025). https://doi.org/10.1109/JIOT.2024.3507778
11. Zhou, M., Zhang, Y., Qiao, X., Yang, L.: Spatially correlated Rayleigh fading for cell-free massive MIMO systems. IEEE Access **8**, 42154–42168 (2020). https://doi.org/10.1109/ACCESS.2020.2976672
12. Sun, Q., Ji, X., Wang, Z., Chen, X., Yang, Y., Zhang, J., Wong, K.-K.: Uplink performance of hardware-impaired cell-free massive MIMO with multi-antenna users and superimposed pilots. IEEE Trans. Commun. **71**(11), 6711–6726 (2023). https://doi.org/10.1109/TCOMM.2023.3301071
13. D'Andrea, C., Garcia-Rodriguez, A., Geraci, G., Giordano, L.G., Buzzi, S.: Analysis of UAV communications in cell-free massive MIMO systems. IEEE Open J. Commun. Soc. **1**, 133–147 (2020). https://doi.org/10.1109/OJCOMS.2020.2964983
14. Zhi, K., Pan, C., Ren, H., Wang, K.: Ergodic rate analysis of reconfigurable intelligent surface-aided massive MIMO systems with ZF detectors. IEEE Commun. Lett. **26**(2), 264–268 (2022). https://doi.org/10.1109/LCOMM.2021.3128904
15. Lei, H., Zhang, J., Wang, Z., Xiao, H., Ai, B.: Uplink performance of cell-free extremely large-scale MIMO systems. IEEE Trans. Veh. Technol. **74**(5), 8339–8344 (2025). https://doi.org/10.1109/TVT.2024.3523373
16. Zhang, Y., Xia, W., Zhao, H., Mao, Y., Zhang, J., Zheng, G.: Enhancing uplink performance for cell-free massive MIMO with low-resolution ADCs by RSMA. IEEE J. Sel. Areas Commun. **43**(3), 720–735 (2025). https://doi.org/10.1109/JSAC.2025.3536544
17. Li, J., Pan, Q., Wan, Z., Zhu, P., Wang, D., Lou, M., Jin, J., Liu, F., You, X.: Low altitude 3-D coverage performance analysis of cell-free RAN for 6G systems. IEEE Trans. Veh. Technol. **72**(12), 16163–16176 (2023). https://doi.org/10.1109/TVT.2023.3296794
18. Jin, S.-N., Yue, D.-W., Nguyen, H.H.: Spectral and energy efficiency in cell-free massive MIMO systems over correlated Rician fading. IEEE Syst. J. **15**(2), 2822–2833 (2021). https://doi.org/10.1109/JSYST.2020.2993048
19. Haritha, H., Amudala, D.N., Budhiraja, R., Chaturvedi, A.K.: Superimposed versus regular pilots for hardware impaired Rician-faded cell-free massive MIMO systems. IEEE Trans. Commun. **72**(11), 6688–6706 (2024). https://doi.org/10.1109/TCOMM.2024.3392780
20. Tentu, V., Sharma, E., Amudala, D.N., Budhiraja, R.: UAV-enabled hardware-impaired spatially correlated cell-free massive MIMO systems: analysis and energy efficiency optimization. IEEE Trans. Commun. **70**(4), 2722–2741 (2022). https://doi.org/10.1109/TCOMM.2022.3144470
21. Wang, Z., Zhang, J., Björnson, E., Ai, B.: Uplink performance of cell-free massive MIMO over spatially correlated Rician fading channels. IEEE Commun. Lett. **25**(4), 1348–1352 (2021). https://doi.org/10.1109/LCOMM.2020.3041899
22. Ma, X., Lei, X., Mathiopoulos, P.T., Yu, K., Tang, X.: Scalable cell-free massive MIMO systems with finite resolution ADCs/DACs over spatially correlated Rician fading channels. IEEE Trans. Veh. Technol. **72**(6), 7699–7716 (2023). https://doi.org/10.1109/TVT.2023.3243921
23. Marzetta, T.L.: Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wireless Commun. **9**(11), 3590–3600 (2010). https://doi.org/10.1109/TWC.2010.092810.091092
24. Marinello, J.C., Abrão, T., Amiri, A., de Carvalho, E., Popovski, P.: Antenna selection for improving energy efficiency in XL-MIMO systems. IEEE Trans. Veh. Technol. **69**(11), 13305–13318 (2020). https://doi.org/10.1109/TVT.2020.3022708
25. Kadan, F.E., Haliloglu, O.: A performance bound for maximal ratio transmission in distributed MIMO. IEEE Wirel. Commun. Lett. **12**(4), 585–589 (2023). https://doi.org/10.1109/LWC.2023.3234326
26. Zhang, J., Zhang, J., Han, Y., Wang, J., Jin, S.: Average spectral efficiency for TDD-based non-stationary XL-MIMO with VR estimation. In: 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 973–977 (2022). https://doi.org/10.1109/WCSP55476.2022.10039284

27. Filho, J.C.M., Brante, G., Souza, R.D., Abrão, T.: Exploring the non-overlapping visibility regions in XL-MIMO random access and scheduling. IEEE Trans. Wirel. Commun. (2022). https://doi.org/10.1109/TWC.2022.3151329
28. Souza, J.H.I., Marinello, J.C., Amiri, A., Abrão, T.: Qos-aware user scheduling in crowded XL-MIMO systems under non-stationary multi-state LoS/NLoS channels. IEEE Trans. Veh. Technol. (2023). https://doi.org/10.1109/TVT.2023.3243488
29. González-Coma, J.P., López-Martínez, F.J., Castedo, L.: Low-complexity distance-based scheduling for multi-user XL-MIMO systems. IEEE Wirel. Commun. Lett. **10**(11), 2407–2411 (2021). https://doi.org/10.1109/LWC.2021.3101940
30. Björnson, E., Hoydis, J., Sanguinetti, L.: Massive MIMO networks: spectral, energy, and hardware efficiency. IEEE Trans. Wireless Commun. **16**(5), 2989–3002 (2017). https://doi.org/10.1109/TWC.2017.2676161
31. Yang, H., Zhong, W.-D., Chen, C., Du, P.: Machine learning-based scheduling in mmwave MIMO networks. IEEE Trans. Wirel. Commun. **20**(11), 7229–7243 (2021). https://doi.org/10.1109/TWC.2021.3082582
32. Björnson, E., Sanguinetti, L., Debbah, M.: Massive MIMO networks: spectral, energy, and hardware efficiency. Found. Trends Signal Process. **11**(3–4), 154–655 (2017). https://doi.org/10.1561/2000000093
33. Li, X., He, C., Dai, H.: Graph-based user scheduling algorithms for multiuser MIMO systems. IEEE Trans. Veh. Technol. **67**(9), 8728–8741 (2018). https://doi.org/10.1109/TVT.2018.2848485
34. Demir, O.T., Bjornson, E., Sanguinetti, L.: Foundations of user-centric cell-free massive MIMO. Found. Trends Signal Process. **14**(3–4), 162–472 (2021). https://doi.org/10.1561/2000000109
35. Agarwal, B., Togou, M.A., Marco, M., Muntean, G.-M.: A comprehensive survey on radio resource management in 5G HetNets: current solutions, future trends and open issues. IEEE Commun. Surv. Tutor. **24**(4), 2495–2534 (2022). https://doi.org/10.1109/COMST.2022.3204993
36. Li, Y.: Towards secure and intelligent network slicing for 5G networks. IEEE Access **8**, 174138–174152 (2020). https://doi.org/10.1109/ACCESS.2020.3025952
37. Yoon, P., Hong, J., Ahn, S., Cho, Y., Na, J., Kwak, J.: Ultima: ultimate balance of centralized and distributed benefits for interference management in 5G cellular networks. IEEE Access **11**, 84080–84096 (2023). https://doi.org/10.1109/ACCESS.2023.3303462
38. Han, M.: Ai-assisted network-slicing based next-generation wireless networks: a survey. IEEE Access **7**, 165885–165902 (2019). https://doi.org/10.1109/ACCESS.2019.2953199
39. Sharma, S.: Deep reinforcement learning approach for haps user scheduling in 5G and beyond networks. IEEE Access **11**, 1130634–1130648 (2023). https://doi.org/10.1109/ACCESS.2023.3327652
40. Abbas, W.B., Ahmed, Q.Z., Khan, F.A., Mian, N.S., Lazaridis, P.I., Sureephong, P.: Designing future wireless networks (FWN)s with net zero (NZ) and zero touch (ZT) perspective. IEEE Access **11**, 83078–83095 (2023). https://doi.org/10.1109/ACCESS.2023.3301909
41. Ngo, H.Q., Ashikhmin, A., Yang, H., Larsson, E.G., Marzetta, T.L.: Cell-free massive MIMO versus small cells. IEEE Trans. Wirel. Commun. **16**(3), 1834–1850 (2017). https://doi.org/10.1109/TWC.2017.2655515
42. Mukherjee, S., Chopra, R.: Performance analysis of cell-free massive MIMO systems in LoS/ NLoS channels. IEEE Trans. Veh. Technol. **71**(6), 6410–6423 (2022). https://doi.org/10.1109/TVT.2022.3161702
43. Sanguinetti, L., Björnson, E., Hoydis, J.: Toward massive MIMO 2.0: understanding spatial correlation, interference suppression, and pilot contamination. IEEE Trans. Commun. **68**(1), 232–257 (2020). https://doi.org/10.1109/TCOMM.2019.2945792
44. Ozdogan, O., Bjornson, E., Larsson, E.G.: Massive MIMO with spatially correlated Rician fading channels. IEEE Trans. Commun. **67**(5), 3234–3250 (2019). https://doi.org/10.1109/TCOMM.2019.2893221
45. Wagner, S., Couillet, R., Debbah, M., Slock, D.: Large system analysis of linear precoding in correlated miso broadcast channels under limited feedback. IEEE Trans. Inf. Theory **58**(7), 4509–4537 (2012). https://doi.org/10.1109/TIT.2012.2191700
46. Ali, A., Carvalho, E.D., Heath, R.W.: Linear receivers in non-stationary massive MIMO channels with visibility regions. IEEE Wirel. Commun. Lett. **8**(3), 885–888 (2019). https://doi.org/10.1109/LWC.2019.2898572

47. 3GPP: Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9). Technical Specification (TS) 36.814, 3rd Generation Partnership Project (3GPP). Version 9.2.0 (2017)

Authors and Affiliations

Gabriel Avanzi Ubiali[1] · Wilson Souza Jr.[1] · José Carlos Marinello[2] · Taufik Abrão[1]

✉ Taufik Abrão
taufik@uel.br

Gabriel Avanzi Ubiali
gabriel.ubiali@uel.com.br

Wilson Souza Jr.
wilsoonjr98@gmail.com

José Carlos Marinello
jcmarinello@utfpr.edu.br

1 Department of Electrical Engineering, State University of Londrina (UEL), Celso Garcia Cid, Londrina, Paraná 86057-970, Brazil

2 Electrical Engineering Department, Federal University of Technology - Paraná (UTFPR), Cornélio Procópio, Paraná 86300-000, Brazil

APPENDIX D – PAPER D

- **Title:** Joint Subarray Selection, User Scheduling, and Pilot Assignment for XL-MIMO
- **Authors:** Gabriel Avanzi Ubiali, José Carlos Marinello Filho and Taufik Abrão
- **Journal:** Physical Communication

Joint Subarray Selection, User Scheduling, and Pilot Assignment for XL-MIMO

Gabriel Avanzi Ubiali*, José Carlos Marinello Filho†, Taufik Abrão*
*Department of Electrical Engineering, State University of Londrina (UEL), Londrina, Brazil
†Department of Electrical Engineering, Federal Technological University of Paraná (UTFPR), Cornélio Procópio, Brazil
E-mail: gabriel.ubiali@uel.br, jcmarinello@utfpr.edu.br, taufik@uel.br

***Abstract*—Extra-large scale MIMO (XL-MIMO) is a key technology for meeting sixth-generation (6G) requirements for high-rate connectivity and uniform quality of service (QoS); however, its deployment is challenged by the prohibitive complexity of resource management based on instantaneous channel state information (CSI). To address this intractability, this work derives novel closed-form deterministic signal-to-interference-plus-noise ratio (SINR) expressions for both centralized and distributed uplink operations. Valid for Rician fading channels with minimum mean square error (MMSE) receive combining and MMSE channel estimation, these expressions depend exclusively on long-term channel statistics, providing a tractable alternative to computationally expensive instantaneous CSI-driven optimization. Building on these results, we develop statistical-CSI-based algorithms for joint subarray selection, users scheduling, and pilot assignment, leveraging the derived SINR approximations to maximize the minimum spectral efficiency (SE) among scheduled users while preserving computational tractability. The proposed framework exploits the spatial sparsity of user equipment (UE) visibility regions (VRs) to enable more aggressive pilot reuse than is possible in conventional massive MIMO. Numerical results validate the high accuracy of the derived SINR approximations and demonstrate that the proposed algorithms significantly enhance fairness and throughput in crowded scenarios.**

***Index Terms*—XL-MIMO, deterministic SINR, large-scale fading (LSF) subarray selection, user scheduling, pilot assignment, fairness, spectral efficiency.**

I. Introduction

The evolution toward 6G networks demands unprecedented SE and link reliability [1]–[4]. XL-MIMO, characterized by antenna deployments with hundreds or thousands of elements, have emerged as a key enabling technology to meet these requirements [5]–[8]. Unlike co-located massive MIMO, XL-MIMO features antenna arrays distributed over large areas (typically organized into discrete subarrays), introducing near-field effects and spatial non-stationarity that fundamentally alter system design principles [1], [9], [10].

In XL-MIMO deployments, each UE is typically close to only a few subarrays. Consequently, serving each UE with all antennas is neither necessary nor scalable. Intelligent subarray selection ensures UEs are served only by subarrays contributing significantly to their signal, preserving service quality while reducing computational complexity [5], [11], [12]. Furthermore, in dense scenarios where UEs outnumber spatial degrees of freedom, user scheduling is essential to manage interference and satisfy QoS constraints [13], [14].

A fully centralized uplink architecture, where a central processing unit (CPU) performs joint detection, achieves the highest SE by enabling global interference suppression [15]. However, its deployment is often limited by fronthaul constraints. A more scalable alternative is distributed processing, where channel estimation and data detection are performed locally at the subarrays. This enhances implementation flexibility and scalability, as new subarrays can be integrated without upgrading the CPU [5, Sec. 5.2].

Coherent processing relies on accurate CSI. While ideally UEs would use orthogonal pilot sequences, pilot reuse is often needed due to limited channel coherence blocks [5], [16]. Pilot sharing reduces estimation accuracy and induces statistical dependence between channel estimates. Fortunately, the spatial sparsity of VRs in XL-MIMO permits more aggressive pilot reuse than in conventional massive MIMO, where every scheduled UE is "seen" by all antennas.

We develop statistical-CSI-based algorithms for joint subarray selection, user scheduling, and pilot assignment. Our key innovation lies in the use of deterministic SINR approximations derived from channel statistics. Because these expressions depend on slowly varying statistical CSI rather than instantaneous CSI, they remain valid over multiple coherence blocks. This significantly reduces computational and signaling overhead, providing a highly efficient framework for resource allocation. Furthermore, channel hardening in XL-MIMO ensures that statistical-CSI-driven algorithm achieves SE comparable to schemes relying on instantaneous CSI.

A. Related Works

Deterministic SINR expressions have been extensively investigated in multi-user massive MIMO, XL-MIMO, and cell-free architectures. However, most existing expressions are derived under spatially uncorrelated Rayleigh fading with simple maximum-ratio (MR) processing, which performs poorly in interference-limited regimes. In realistic 6G micro-urban environments, channels are characterized by a combination of line-of-sight (LoS) and non-LoS (NLoS) components, best captured by the Rician fading model. Furthermore, while MR is analytically convenient, MMSE combining is essential for interference suppression in dense networks [15].

The literature regarding deterministic SINR can be categorized by channel models and combining schemes. Expressions for spatially uncorrelated Rician fading are available in [17]–[23], but assume MR processing or perfect CSI. Conversely, works focusing on spatially correlated Rayleigh fading, such as [24]–[26] for perfect CSI and [27]–[33] for imperfect CSI, neglect the LoS component and generally restrict analysis to MR combiners, thereby limiting their applicability in dense deployments. Regarding other combining schemes, [25] and [33] investigate zero-forcing (ZF) and regularized ZF (RZF) variants, under perfect and imperfect CSI, respectively.

More recent efforts have addressed spatially correlated Rician fading with imperfect CSI and MR processing [23], [34]–[38]. A notable exception is [38], which analyzes RZF precoding. However, the resulting expressions are not fully closed-form, requiring iterative solutions of coupled fixed-point equations.

To the best of our knowledge, no existing work provides closed-form deterministic SINR expressions that jointly account for: (i) spatially correlated Rician fading, (ii) under imperfect CSI, where channel estimates are obtained via MMSE estimation, and (iii) MMSE receive combining. This paper fills this gap, providing a robust analytical framework for both centralized and distributed operations.

B. Contributions

This paper's contributions are summarized as:

- **Novel closed-form deterministic SINR expressions:** We derive original, closed-form deterministic expressions for the uplink SINR under both centralized and distributed XL-MIMO operations. These expressions are unique in that they jointly account for spatially correlated Rician fading, MMSE receive combining, and imperfect CSI obtained through MMSE channel estimation. By depending exclusively on statistical CSI, they enable rapid performance evaluation and provide a tractable analytical foundation for optimization. Numerical results validate that these approximations closely track the true SINR.
- **Statistical-CSI-driven resource allocation:** We develop algorithms for joint subarray selection, user scheduling, and pilot assignment that leverage the derived SINR approximations to maximize the minimum SE among the scheduled UEs. By utilizing only slowly-varying channel statistics, the proposed methods drastically reduce computational complexity and signaling overhead compared to instantaneous-CSI benchmarks while maintaining near-optimal performance. Numerical results demonstrate that our statistical approach closely matches the performance of instantaneous-CSI-based strategies.

II. XL-MIMO Channel and System Models

Consider a system with L subarrays, coordinated by a CPU, serving K single-antenna UEs. Each subarray utilizes a uniform planar array (UPA) with M antennas. Defining the sets of all subarrays and UEs as $\mathcal{D} = \{1, \ldots, L\}$ and $\mathcal{K} = \{1, \ldots, K\}$, respectively, we denote $\mathcal{U} \subseteq \mathcal{K}$ as the set of scheduled UEs.

Each UE $k \in \mathcal{U}$ is served by a subarray subset $\mathcal{D}_k \subseteq \mathcal{D}$ of cardinality $L_k = |\mathcal{D}_k|$, while subarray l serves a UE subset $\mathcal{U}_l \subseteq \mathcal{U}$ of size $U_l = |\mathcal{U}_l|$. For convenience, define the selection matrix $\mathbf{D}_k = \mathrm{blkdiag}(\mathbf{D}_{k1}, \ldots, \mathbf{D}_{kL})$, where

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_M & \text{if } l \in \mathcal{D}_k, \\ \mathbf{0}_M & \text{otherwise.} \end{cases} \tag{1}$$

We define three subarray selection criteria: *random*; *LSF-based*, which prioritizes signal strength by selecting subarrays with the highest large-scale fading (LSF) channel gains; and *SINR-based*, which optimizes the trade-off between signal power and interference by selecting subarrays with the highest local SINRs.

During uplink data transmission, subarray l receives the signal

$$\mathbf{y}_l^{\mathrm{ul}} = \sum_{i\in\mathcal{U}} \mathbf{h}_{il} x_i + \mathbf{n}_l^{\mathrm{ul}}, \tag{2}$$

where $\mathbf{h}_{il}$ is the channel vector between UE i and subarray l, $x_i \in \mathbb{C}$ is the zero-mean symbol transmitted by UE i, with transmit power $p_i = \mathbb{E}\{|x_i|^2\}$, and $\mathbf{n}_l^{\mathrm{ul}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{\mathrm{n}}^2 \mathbf{I}_M)$ denotes the additive white Gaussian noise (AWGN) vector.

The channel vector between UE k and subarray l comprises a deterministic LoS component, $\overline{\mathbf{h}}_{kl}$, and a stochastic NLoS component, $\check{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_{kl})$, where $\mathbf{R}_{kl} \in \mathbb{C}^{M\times M}$ is the spatial covariance matrix. It follows the distribution $\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\omega_{kl}\overline{\mathbf{h}}_{kl}, \mathbf{R}_{kl})$, where $\omega_{kl} \in \{0,1\}$ is a Bernoulli random variable indicating the presence of a LoS path. The average channel gain between UE k and the antennas of subarray l is

$$\beta_{kl} = \frac{1}{M}\mathbb{E}\{\|\mathbf{h}_{kl}\|^2\} = \frac{1}{M}\,\mathrm{tr}(\mathbf{Q}_{kl}), \tag{3}$$

where $\mathbf{Q}_{kl} = \mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kl}^{\mathrm{H}}\} = \overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{kl}^{\mathrm{H}} + \mathbf{R}_{kl}$ is the channel correlation matrix. Similarly, the channel gain associated only to the NLoS component is $\check{\beta}_{kl} = \frac{1}{M}\,\mathrm{tr}(\mathbf{R}_{kl})$. The average LSF channel gain across all L subarrays for UE k is $\beta_k = \frac{1}{L}\sum_{l=1}^{L}\beta_{kl}$.

In modular XL-MIMO, the large inter–subarray spacing justifies modeling the NLoS channels of different subarrays as statistically independent. Hence, $\mathbb{E}\{\check{\mathbf{h}}_{kl}\check{\mathbf{h}}_{kl'}^{\mathrm{H}}\} = \mathbf{R}_{kl}\delta_{ll'}$, where $\delta_{ll'}$ is the Kronecker delta. As a result, the overall channel covariance matrix for user k is taken as a block-diagonal matrix, $\mathbf{R}_k = \mathrm{blkdiag}\left(\mathbf{R}_{k1}, \mathbf{R}_{k2}, \ldots, \mathbf{R}_{kL}\right)$, and the collective channel vector $\mathbf{h}_k = [\mathbf{h}_{k1}^{\mathrm{T}}, \ldots, \mathbf{h}_{kL}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{ML}$ is distributed as $\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\overline{\mathbf{h}}_k, \mathbf{R}_k)$, where $\overline{\mathbf{h}}_k = [\overline{\mathbf{h}}_{k1}^{\mathrm{T}}, \ldots, \overline{\mathbf{h}}_{kL}^{\mathrm{T}}]^{\mathrm{T}}$.

A. MMSE Channel Estimation

The CPU is assumed to have perfect knowledge of the channel statistics (mean vectors and covariance matrices), as these remain approximately constant over numerous coherence blocks.[1] Under time-division-duplexing (TDD) operation, instantaneous channels are estimated via uplink pilot signaling. Each coherence block of length τ_{c} consists of τ_{p} symbols reserved for pilots and $\tau_{\mathrm{c}} - \tau_{\mathrm{p}}$ symbols dedicated to data transmission. This study focuses exclusively on the uplink data transmission. The pilot assigned to UE k is denoted by t_k, and its corresponding sequence is $\boldsymbol{\phi}_{t_k} \in \mathbb{C}^{\tau_{\mathrm{p}}}$. The sequences form an orthogonal set:

$$\boldsymbol{\phi}_{t_k}^{\mathrm{H}}\boldsymbol{\phi}_{t_i} = \begin{cases} \tau_{\mathrm{p}}, & t_i = t_k, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Let $\mathcal{P}_t = \{k \in \mathcal{K} \mid t_k = t\}$ denote the set of UEs using pilot t. The MMSE estimator exploits the channel statistics to separate pilot-sharing UEs and minimize the estimation mean square error (MSE). Denote by $\widehat{\mathbf{h}}_{kl}$ the MMSE estimate of $\mathbf{h}_{kl}$ and by $\widetilde{\mathbf{h}}_{kl} = \mathbf{h}_{kl} - \widehat{\mathbf{h}}_{kl}$ the corresponding estimation error. They are independent and distributed as [40]:

$$\widehat{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}\left(\overline{\mathbf{h}}_{kl}, \mathbf{R}_{kl} - \mathbf{C}_{kl}\right), \tag{5}$$

$$\widetilde{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_M, \mathbf{C}_{kl}\right), \tag{6}$$

where the error covariance matrix is

$$\mathbf{C}_{kl} = \mathbf{R}_{kl} - p_k\tau_{\mathrm{p}}\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_k l}^{-1}\mathbf{R}_{kl}, \tag{7}$$

and $\boldsymbol{\Psi}_{tl} = \sum_{i\in\mathcal{P}_t} p_i\tau_{\mathrm{p}}\mathbf{R}_{il} + \sigma_{\mathrm{n}}^2\mathbf{I}_M$.

The estimation accuracy can be measured by the normalized mean square error (NMSE), evaluated in subarray l and in the entire array by, respectively [5]:

$$\gamma_{kl} = \frac{\mathbb{E}\{\|\widetilde{\mathbf{h}}_{kl}\|^2\}}{\mathbb{E}\{\|\mathbf{h}_{kl}\|^2\}} = \frac{\mathrm{tr}(\mathbf{C}_{kl})}{\mathrm{tr}(\mathbf{Q}_{kl})}, \tag{8}$$

$$\gamma_k = \frac{\mathbb{E}\{\|\widetilde{\mathbf{h}}_k\|^2\}}{\mathbb{E}\{\|\mathbf{h}_k\|^2\}} = \frac{\mathrm{tr}(\mathbf{C}_k)}{\mathrm{tr}(\mathbf{Q}_k)}, \tag{9}$$

where $\mathbf{C}_k = \mathrm{blkdiag}\left(\mathbf{C}_{k1}, \ldots, \mathbf{C}_{kL}\right)$, $\widetilde{\mathbf{h}}_k = [\widetilde{\mathbf{h}}_{k1}^{\mathrm{T}}, \ldots, \widetilde{\mathbf{h}}_{kL}^{\mathrm{T}}]^{\mathrm{T}}$, and $\mathbf{Q}_k = \mathbb{E}\{\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\} = \overline{\mathbf{h}}_k\overline{\mathbf{h}}_k^{\mathrm{H}} + \mathbf{R}_k$.

B. Centralized Uplink Operation

In centralized operation, the CPU computes the collective receive combiner $\mathbf{v}_k = [\mathbf{v}_{k1}^{\mathrm{T}}, \ldots, \mathbf{v}_{kL}^{\mathrm{T}}]^{\mathrm{T}}$ and jointly processes the uplink signals from the selected subarrays to estimate x_k, resulting in $\widehat{x}_k = \mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\mathbf{y}^{\mathrm{ul}}$, where $\mathbf{y}^{\mathrm{ul}} = [(\mathbf{y}_1^{\mathrm{ul}})^{\mathrm{T}}, \ldots, (\mathbf{y}_L^{\mathrm{ul}})^{\mathrm{T}}]^{\mathrm{T}}$. The instantaneous, centralized uplink SINR for UE k is given by [5, Ch. 5]

$$\Gamma_k^{\mathrm{cent}} = \frac{p_k|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k|^2}{\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\mathbf{Z}_k\mathbf{D}_k\mathbf{v}_k}, \tag{10}$$

where $\widehat{\mathbf{h}}_k = [\widehat{\mathbf{h}}_{k1}^{\mathrm{T}}, \ldots, \widehat{\mathbf{h}}_{kL}^{\mathrm{T}}]^{\mathrm{T}}$,

$$\mathbf{Z}_k = \mathbf{W} - p_k\mathbf{D}_k\widehat{\mathbf{h}}_k\widehat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{D}_k, \tag{11}$$

and

$$\mathbf{W} = \sum_{i\in\mathcal{U}} p_i\mathbf{D}_k(\widehat{\mathbf{h}}_i\widehat{\mathbf{h}}_i^{\mathrm{H}} + \mathbf{C}_i)\mathbf{D}_k + \sigma_{\mathrm{n}}^2\mathbf{I}_{ML}. \tag{12}$$

This SINR is maximized by the MMSE combining vector, which is given by [5, Ch. 5]

$$\mathbf{v}_k = p_k\mathbf{W}^{-1}\mathbf{D}_k\widehat{\mathbf{h}}_k, \tag{13}$$

[1]Such statistics can be acquired during connection setup [39].

and yields the maximum SINR value

$$\Gamma_k^{\text{cent}} = p_k \widehat{\mathbf{h}}_k^{\text{H}} \mathbf{D}_k \mathbf{Z}_k^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k. \quad (14)$$

The resulting SE is given by $\eta_k^{\text{cent}} = g(\Gamma_k^{\text{cent}})$, where

$$g(x) = \left(1 - \frac{\tau_{\text{p}}}{\tau_{\text{c}}}\right) \log_2(1 + x). \quad (15)$$

As we focus exclusively on the uplink, the SE expression assumes that each channel coherence block is divided into uplink pilot transmission and uplink data transmission, ignoring downlink slots.

C. Distributed Uplink Operation

In the distributed scheme, each serving subarray $l \in \mathcal{D}_k$ first estimates the uplink symbol of UE k using the local combining vector $\mathbf{v}_{kl}$ as:

$$\hat{x}_{kl} = \mathbf{v}_{kl}^{\text{H}} \mathbf{y}_l^{\text{ul}}, \qquad l \in \mathcal{D}_k. \quad (16)$$

Subsequently, the CPU forms the final estimate $\hat{x}_k$ by aggregating these local estimates:

$$\hat{x}_k = \sum_{l \in \mathcal{D}_k} \mu_{kl}\, \hat{x}_{kl}. \quad (17)$$

The weights satisfy $\sum_{l \in \mathcal{D}_k} \mu_{kl} = 1$ for all $k \in \mathcal{U}$, where $\mu_{kl} > 0$ if $l \in \mathcal{D}_k$, and $\mu_{kl} = 0$ otherwise.

The instantaneous local SINR for UE k evaluated at a subarray l that serves this UE is given by [5, Ch. 5]

$$\Gamma_{kl}^{\text{dist}} = \frac{p_k \left| \mathbf{v}_{kl}^{\text{H}} \widehat{\mathbf{h}}_{kl} \right|^2}{\sum\limits_{\substack{i \in \mathcal{U} \\ i \neq k}} p_i \left| \mathbf{v}_{kl}^{\text{H}} \widehat{\mathbf{h}}_{il} \right|^2 + \sum\limits_{i \in \mathcal{U}} p_i \mathbf{v}_{kl}^{\text{H}} \mathbf{C}_{il} \mathbf{v}_{kl} + \sigma_{\text{n}}^2 \left\| \mathbf{v}_{kl} \right\|^2}. \quad (18)$$

This SINR is maximized by the local MMSE (L-MMSE) combiner, given by [5, page 149]

$$\mathbf{v}_{kl} = p_k \mathbf{W}_l^{-1} \widehat{\mathbf{h}}_{kl}, \quad (19)$$

which yields the maximum value

$$\Gamma_{kl}^{\text{dist}} = p_k \widehat{\mathbf{h}}_{kl}^{\text{H}} \mathbf{Z}_{kl}^{-1} \widehat{\mathbf{h}}_{kl} = p_k \operatorname{tr}\left(\mathbf{Z}_{kl}^{-1} \widehat{\mathbf{h}}_{kl} \widehat{\mathbf{h}}_{kl}^{\text{H}} \right), \quad (20)$$

where

$$\mathbf{W}_l = \sum_{i \in \mathcal{U}} p_i (\widehat{\mathbf{h}}_{il} \widehat{\mathbf{h}}_{il}^{\text{H}} + \mathbf{C}_{il}) + \sigma_{\text{n}}^2 \mathbf{I}_M \quad (21)$$

and

$$\mathbf{Z}_{kl} = \mathbf{W}_l - p_k \widehat{\mathbf{h}}_{kl} \widehat{\mathbf{h}}_{kl}^{\text{H}}. \quad (22)$$

The global SINR for UE k under distributed operation, denoted as Γ_k^{dist}, is given by (23), and the resulting SE is given by $\eta_k^{\text{dist}} = g(\Gamma_k^{\text{dist}})$.

Proposition 1 (Global SINR Approximation and Optimal Weights)**.** *Assuming $M \geq U$ and L-MMSE combining, the global SINR for UE k can be approximated by:*

$$\widehat{\Gamma}_k^{\text{dist}} = \left(\sum_{l \in \mathcal{D}_k} \mu_{kl}^2 (\Gamma_{kl}^{\text{dist}})^{-1} \right)^{-1}. \quad (24)$$

The weights that maximize (24) *subject to* $\sum\limits_{l \in \mathcal{D}_k} \mu_{kl} = 1$ *are:*

$$\mu_{kl} = \begin{cases} \dfrac{\Gamma_{kl}^{\text{dist}}}{\sum_{l' \in \mathcal{D}_k} \Gamma_{kl'}^{\text{dist}}}, & \textit{if } l \in \mathcal{D}_k, \\ 0, & \textit{otherwise}, \end{cases} \quad (25)$$

yielding the maximum approximate global SINR

$$\widehat{\Gamma}_k^{\text{dist}} = \sum_{l \in \mathcal{D}_k} \Gamma_{kl}^{\text{dist}}. \quad (26)$$

Proof. See Appendix A. □

We also refer to the optimal weighting in (25) as SINR-based weighting. A practical alternative is LSF-based weighting, which avoids local SINR estimation by using weights proportional to the LSF gains:

$$\mu_{kl} = \begin{cases} \dfrac{\beta_{kl}}{\sum_{l' \in \mathcal{D}_k} \beta_{kl'}}, & \text{if } l \in \mathcal{D}_k, \\ 0, & \text{otherwise}. \end{cases} \quad (27)$$

While (27) is suboptimal as it neglects interference and channel estimation errors, it generally outperforms the equal weighting strategy, defined as:

$$\mu_{kl} = \begin{cases} 1/L_k, & \text{if } l \in \mathcal{D}_k, \\ 0, & \text{otherwise}. \end{cases} \quad (28)$$

The approximate SE can be expressed via the global mapping $\hat{\eta}_k^{\text{dist}} = g(\widehat{\Gamma}_k^{\text{dist}})$ or, equivalently, via the component-wise mapping $\hat{\eta}_k^{\text{dist}} = \hat{g}(\{\Gamma_{kl}^{\text{dist}}\}_{l \in \mathcal{D}_k})$, defined as:

$$\hat{g}\left(\{\Gamma_{kl}\}_{l \in \mathcal{D}_k}\right) = \left(1 - \frac{\tau_{\text{p}}}{\tau_{\text{c}}}\right) \log_2 \left(1 + \frac{1}{\sum_{l \in \mathcal{D}_k} \mu_{kl}^2 \Gamma_{kl}^{-1}}\right). \quad (29)$$

D. Performance Comparison

This subsection evaluates the uplink performance of centralized and distributed XL-MIMO architectures as a function of the number of serving subarrays, for different subarray selection and weighting strategies.

The base station (BS) comprises $L = 16$ linearly arranged subarrays, each equipped with a 4×4 UPA ($M = 16$), serving $U = 16$ UEs in a $100 \times 50\,\text{m}$ coverage area. The coherence block length is $\tau_{\text{c}} = 16$ and $\tau_{\text{p}} = 8$ orthogonal pilots are randomly assigned to the UEs, which implies pilot contamination since

$$\Gamma_k^{\text{dist}} = \frac{p_k \left| \sum_{l\in\mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\text{H}} \widehat{\mathbf{h}}_{kl} \right|^2}{\sum_{i\in\mathcal{U}\setminus\{k\}} p_i \left| \sum_{l\in\mathcal{D}_k} \mu_{kl} \mathbf{v}_{kl}^{\text{H}} \widehat{\mathbf{h}}_{il} \right|^2 + \sum_{i\in\mathcal{U}} p_i \sum_{l\in\mathcal{D}_k} \mu_{kl}^2 \mathbf{v}_{kl}^{\text{H}} \mathbf{C}_{il} \mathbf{v}_{kl} + \sigma_{\text{n}}^2 \sum_{l\in\mathcal{D}_k} \mu_{kl}^2 \|\mathbf{v}_{kl}\|^2}. \tag{23}$$

Table I
SIMULATION PARAMETERS

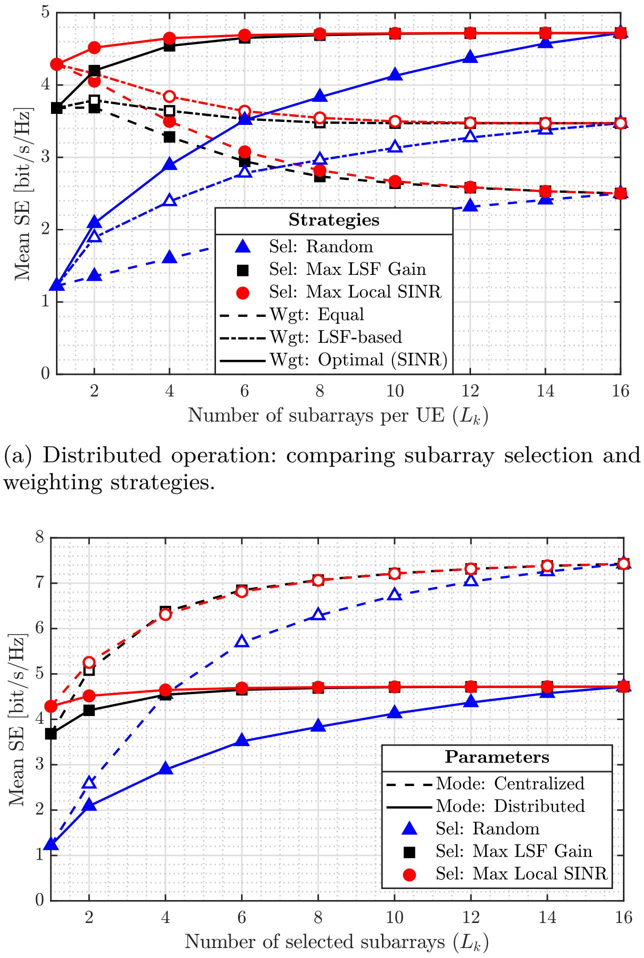
Parameter	Value
Number of subarrays (L)	16
UE transmit power (p_k)	20 dBm
Noise power (σ_{n}^2)	-96 dBm
Channel coherence block length (τ_{c})	16
Carrier frequency	6 GHz
Antenna spacing (half-wavelength)	2.5 cm
Subarray spacing	6.0 m
Array height	3 m
UE height	1.5 m

$U > \tau_{\text{p}}$. Table I summarizes the remaining parameter values adopted in this validation.

Figure 1a illustrates the impact of subarray selection and weighting on the mean SE in distributed operation. Two distinct convergence regimes are observed at the boundaries of the L_k range. At $L_k = 1$, performance is dictated solely by the *selection* strategy, clustering the curves into three groups; weighting is irrelevant here, as the final signal estimate coincides with the local estimate from the selected subarray. Conversely, at $L_k = L = 16$, the selection strategy is irrelevant, and the curves cluster into three groups determined exclusively by the *weighting* scheme.

For small numbers of selected subarrays (e.g., $L_k = 2$), intelligent selection (LSF- and SINR-based) significantly outperforms random selection. The SINR-based scheme yields the highest performance, confirming that interference awareness is especially critical when L_k is low. As L_k increases, the LSF-based selection converges to the SINR-based scheme in performance.

The optimal (SINR-based) weighting provides a clear upper bound, increasing monotonically with L_k up to about 4.7 bit/s/Hz. In contrast, *equal* weighting degrades performance for intelligent selection schemes as L_k grows, since it does not attenuate the contribution of the subarrays with weak channels or strong interference. Consequently, increasing L_k in distributed operation is only beneficial if optimal



(a) Distributed operation: comparing subarray selection and weighting strategies.

(b) Architecture comparison: centralized vs. distributed with optimal weighting, varying subarray selection strategy.

Figure 1. Mean SE versus number of selected subarrays L_k.

weights are applied. The LSF-based weighting offers an intermediate solution, alleviating performance loss but still falling short of the optimum as it fails to penalize interference.

Figure 1b compares the distributed architecture (with optimal weighting) against the fully centralized implementation. As expected, centralized processing consistently outperforms the distributed approach due to its ability to jointly suppress inter-user interference. The *SINR-based* selection yields the highest performance for both architectures. The *LSF-based* scheme closely follows the SINR-based trend, reinforcing that LSF is a robust selection metric.

The results highlight the trade-off between complexity and SE. While centralized processing yields superior SE, the distributed architecture offers a viable low-complexity alternative. To maximize its efficacy, two design choices are critical: *a)* LSF-based subarray selection is recommended for achieving near-optimal performance without requiring local SINR estimation; and *b)* optimal weighting is mandatory to prevent performance degradation when scaling L_k.

Finally, results demonstrated that selecting only a small subset of subarrays to serve each UE is sufficient to achieve high mean SE while significantly reducing computational complexity and fronthaul load compared to utilizing the full antenna array ($L_k = L$).

III. Uplink SINR Approximations

This section derives deterministic asymptotic and ergodic SINR approximations for centralized and distributed uplink operation. The asymptotic expressions describe the deterministic convergence of the instantaneous SINR as the number of antennas grows large. Conversely, the ergodic expressions approximate the average SINR and are particularly accurate in crowded scenarios with high user loads.

Theorem 1 (Ergodic SINR)**.** *The ergodic global uplink SINR for centralized operation with MMSE combiner and the ergodic local uplink SINR for distributed operation with L-MMSE combiner are:*

$$\Gamma_k^{\text{cent,erg}} \approx p_k \operatorname{tr}\left[\mathbf{D}_k \overline{\mathbf{Z}}_k^{-1} \mathbf{D}_k \left(\mathbf{Q}_k - \mathbf{C}_k\right) \mathbf{D}_k\right], \tag{30a}$$

$$\Gamma_{kl}^{\text{dist,erg}} \approx p_k \operatorname{tr}\left[\overline{\mathbf{Z}}_{kl}^{-1} \left(\mathbf{Q}_{kl} - \mathbf{C}_{kl}\right)\right], \tag{30b}$$

where

$$\overline{\mathbf{Z}}_k = \mathbf{D}_k \left(\sum_{i \in \mathcal{U}\setminus\{k\}} p_i \mathbf{Q}_i + p_k \mathbf{C}_k \right) \mathbf{D}_k + \sigma_{\text{n}}^2 \mathbf{I}_{ML}, \tag{31a}$$

$$\overline{\mathbf{Z}}_{kl} = \sum_{i \in \mathcal{U}\setminus\{k\}} p_i \mathbf{Q}_{il} + p_k \mathbf{C}_{kl} + \sigma_{\text{n}}^2 \mathbf{I}_M. \tag{31b}$$

These approximations hold if the selected subarrays yield strong LoS components or under high user load conditions (i.e., $U \gg ML_k$ for centralized or $U \gg M$ for distributed operation).

Proof. See Appendix B. □

Corollary 1. *The computational complexity (in real multiplications) to evaluate* (30a) *and* (30b) *across the L_k serving subarrays is, respectively:*

$$\mathcal{C}_k^{\text{cent,erg}} = 3M^3 L_k^3 + \frac{1}{2} M^2 L_k^2 - \frac{3}{2} M L_k, \tag{32a}$$

$$\mathcal{C}_k^{\text{dist,erg}} = 3M^3 L_k + \frac{1}{2} M^2 L_k - \frac{3}{2} M L_k. \tag{32b}$$

Proof. See Appendix C. □

Theorem 2 (Asymptotic SINR)**.** *Assume MMSE channel estimation and user loads satisfying $U \leq ML_k$ for centralized or $U \leq M$ for distributed operation. Furthermore, assume MMSE (centralized) or L-MMSE (distributed) combining is used, and that a) there is no pilot reuse or b) the channel estimates are relatively accurate and the NLoS channel component is spatially uncorrelated. As the number of antennas grows large ($ML_k \to \infty$ for centralized or $M \to \infty$ for distributed operation), the instantaneous SINR converges deterministically to the asymptotic limits given by:*

$$\Gamma_k^{\text{cent,asy}} = \frac{\operatorname{tr}(\mathbf{D}_k \mathbf{X}_k \mathbf{D}_k) - \displaystyle\sum_{i \in \mathcal{U}\setminus\{k\}} \frac{\operatorname{tr}(\mathbf{D}_k \mathbf{X}_i \mathbf{D}_k \mathbf{X}_k \mathbf{D}_k)}{\operatorname{tr}(\mathbf{D}_k \mathbf{X}_i \mathbf{D}_k)}}{\displaystyle\sum_{i \in \mathcal{U}} \frac{p_i}{ML_k p_k} \operatorname{tr}(\mathbf{D}_k \mathbf{C}_i \mathbf{D}_k) + \sigma_{\text{n}}^2}, \tag{33a}$$

$$\Gamma_{kl}^{\text{dist,asy}} = \frac{\operatorname{tr}\left(\mathbf{X}_{kl}\right) - \displaystyle\sum_{i \in \mathcal{U}\setminus\{k\}} \frac{\operatorname{tr}(\mathbf{X}_{kl} \mathbf{X}_{il})}{\operatorname{tr}(\mathbf{X}_{il})}}{\displaystyle\sum_{i \in \mathcal{U}} \frac{p_i}{M p_k} \operatorname{tr}(\mathbf{C}_{il}) + \sigma_{\text{n}}^2}, \tag{33b}$$

where $\mathbf{X}_k = \mathbf{Q}_k - \mathbf{C}_k$ and $\mathbf{X}_{kl} = \mathbf{Q}_{kl} - \mathbf{C}_{kl}$.

Proof. See Appendix D. □

Corollary 2. *The computational complexity (in real multiplications) to evaluate* (33a) *and* (33b) *across the L_k serving subarrays is, respectively:*

$$\mathcal{C}_k^{\text{cent,asy}} = 3(U-1) M^2 L_k^2, \tag{34a}$$

$$\mathcal{C}_k^{\text{dist,asy}} = 3(U-1) M^2 L_k. \tag{34b}$$

Proof. Complexity is dominated by evaluating trace terms $\operatorname{tr}(\mathbf{AB})$ for the $U-1$ interfering UEs ($i \neq k$). Computing the trace of a product of two square matrices of size N requires calculating only the diagonal elements of the product, costing N^2 complex multiplications. For distributed operation (33b), this involves $M \times M$ matrices ($N = M$) for each of the L_k subarrays. For centralized operation (33a), the matrices have effective size $ML_k \times ML_k$ ($N = ML_k$). Assuming three real multiplications per complex multiplication [41, p. 378], we arrive at (34a) and (34b). □

The approximate achievable SE is obtained via the mapping $g(\cdot)$ in (15). For centralized operation, we evaluate $\eta_k^{\text{cent,asy}} = g(\Gamma_k^{\text{cent,asy}})$ and $\eta_k^{\text{cent,erg}} = g(\Gamma_k^{\text{cent,erg}})$. For distributed operation, it is reconstructed from the local SINR approximations as $\hat{\eta}_k^{\text{dist,asy}} = \hat{g}(\{\Gamma_{kl}^{\text{dist,asy}}\}_{l\in\mathcal{D}_k})$ and $\hat{\eta}_k^{\text{dist,erg}} = \hat{g}(\{\Gamma_{kl}^{\text{dist,erg}}\}_{l\in\mathcal{D}_k})$.

A. Approximation Accuracy

This section validates the derived SINR approximations. Simulations assume $L = 16$ and $L_k = 4$, employing LSF-based subarray selection and optimal weighting.[2] We enforce pilot contamination by setting $\tau_p = U/2$ with random assignment, while $\tau_c = U$. Remaining parameters are listed in Table I.

We assess the accuracy of the derived approximations using the mean normalized absolute error (MNAE) metric. For a generic variable x and its estimate $\hat{x}$, the MNAE is defined as:[3]

$$\nu(x, \hat{x}) = \mathbb{E}\left\{\frac{|\hat{x} - x|}{x}\right\}. \tag{35}$$

To validate the global SINR approximation for distributed operation derived in Proposition 1, we evaluate the error between the true distributed SE and its estimate based on local SINRs:

$$\tilde{\eta}_k^{\text{dist}} = \nu(\eta_k^{\text{dist}}, \hat{\eta}_k^{\text{dist}}). \tag{36}$$

Since the asymptotic expressions target the *instantaneous* performance, their accuracy is measured relative to the true instantaneous SE:

$$\tilde{\eta}_k^{\text{op,asy}} = \nu(\eta_k^{\text{op}}, \hat{\eta}_k^{\text{op,asy}}), \quad \text{for op} \in \{\text{cent}, \text{dist}\}. \tag{37}$$

Conversely, the ergodic expressions target the *average* performance. Thus, we compare them against the true mean SE:

$$\tilde{\eta}_k^{\text{op,erg}} = \nu(\mathbb{E}\{\eta_k^{\text{op}}\}, \hat{\eta}_k^{\text{op,erg}}), \quad \text{for op} \in \{\text{cent}, \text{dist}\}. \tag{38}$$

We first assess the accuracy of the global SINR approximation for distributed operation in (24) using Fig. 2, which plots the MNAE $\tilde{\eta}_k^{\text{dist}}$ versus M.[4] The approximation error decreases steadily with M and increases with U, since a larger M/U ratio provides stronger interference suppression, satisfying the derivation's assumption of negligible coherent interference. On the other hand, selecting fewer subarrays (e.g., $L_k = 2$), yields lower MNAE than larger subsets (e.g., $L_k = 8$), as the LSF-based strategy can restrict selection to the strongest subarrays where $\mathbf{v}_{kl}^{\text{H}}\widehat{\mathbf{h}}_{kl} \approx 1$.[5] These factors render the approximation robust even in the overloaded regime ($M < U$).

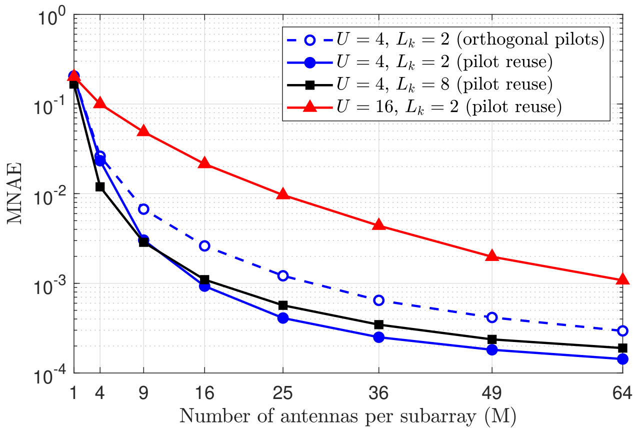


Figure 2. MNAE of the global SE approximation for distributed operation versus subarray antennas.

Figure 3 evaluates the ergodic and asymptotic approximations as a function of U, with fixed $M = 16$ and $L_k = 4$. The subfigures depict the MNAE for distributed (Fig. 3a) and centralized (Fig. 3b) operations, assessing the impact of spatial correlation and pilot contamination. To this end, we compare spatially correlated vs. uncorrelated fading, and contrast orthogonal pilot assignment ($\tau_\text{p} = U$) with pilot reuse ($\tau_\text{p} = U/2$, with random assignment).

The MNAE of the asymptotic expression increases with U. As the system load grows, the variance of the aggregate interference becomes more significant, causing the instantaneous realizations to deviate further from the deterministic limit. This divergence is more pronounced in compared to the centralized case. The latter benefits from a higher diversity order (ML_k antennas involved in combining), which accelerates convergence to the deterministic mean.

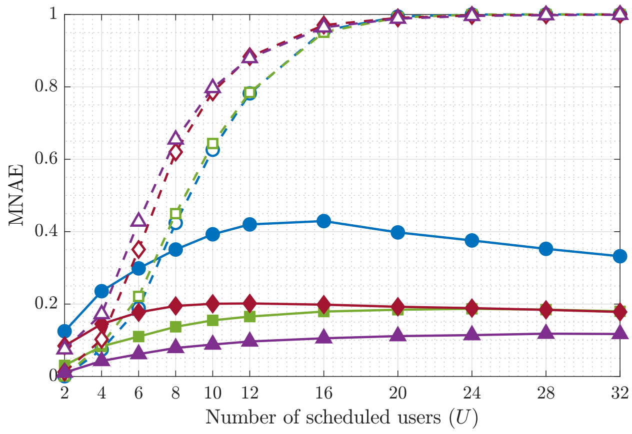
According to Fig 3, spatial correlation and pilot reuse degrade the accuracy of the asymptotic approximations by violating the tractability assumptions underlying (33a) and (33b), which rely on uncorrelated NLoS components and small channel estimation errors. Furthermore, correlation reduces spatial diversity, which hinders channel hardening and thereby delays the SINR convergence to the deterministic regime, while pilot contamination scales with U. These factors introduce direct modeling discrepancies, caus-

[2] The choice of $L_k = 4$ is justified by the performance-complexity trade-off observed in Fig. 1b.

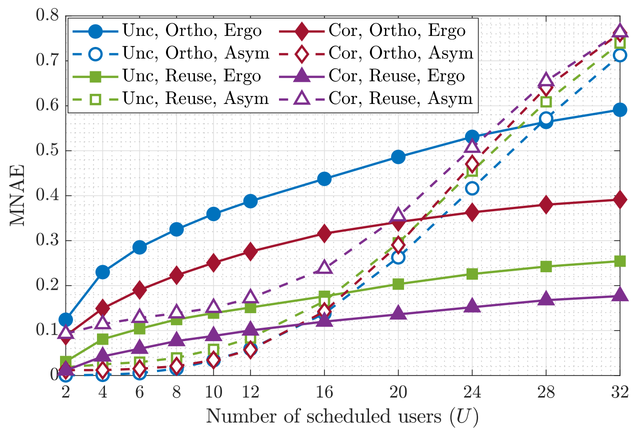
[3] Throughout this work, $\mathbb{E}\{\cdot\}$ denotes the expectation over small-scale fading (SSF) realizations.

[4] We actually plot the accuracy averaged across both scheduled UEs and channel realizations.

[5] See Appendix A for more details on the validity conditions for the approximation.



(a) Distributed operation.



(b) Centralized operation.

Figure 3. MNAE of the deterministic SE approximations vs. number of UEs, contrasting ergodic and asymptotic approximations under varying spatial correlation and pilot reuse scenarios.

ing the instantaneous SE to diverge from the asymptotic approximation.

Counter-intuitively, spatial correlation and pilot reuse reduce the MNAE of the ergodic approximations by increasing the deterministic covariance $\mathbf{C}_{kl}$, which stabilizes $\mathbf{Z}_{kl}$ around its mean, better satisfying the Neumann series condition required for these approximations. The error evolution is driven by competing effects: it initially rises with U as additional interferers increase the variance of $\mathbf{Z}_{kl}$, and then it peaks near the spatial degrees of freedom ($U \approx M$ for distributed operation and $U \approx ML_k$ for centralized). As we increase U beyond this peak, $\mathbf{Z}_{kl}$ concentrates around its expected value. This explains why, for the same M, the ergodic approximation is more accurate in distributed operation than in centralized: it reaches saturation significantly earlier than the centralized.

Given the contrasting error behaviors of the ergodic and asymptotic approximations, we define a switching point, U_{switch}, to select the best approximation. The asymptotic expression is prioritized for $U \leq U_{\text{switch}}$ due to its superior accuracy and lower complexity, while the ergodic expression is adopted for $U > U_{\text{switch}}$. The optimal U_{switch} depends on channel conditions and the pilot assignment strategy, shifting toward lower values of U when correlation or pilot reuse is present. Table II summarizes the empirically determined switching thresholds derived from the intersection points of the asymptotic and ergodic error curves in Fig. 3. Note that the thresholds for centralized operation are consistently scaled by L_k compared to the distributed case. Note also that, in the most severe scenario (correlated NLoS with pilot reuse), the threshold drops to $U_{\text{switch}} \approx 0$. This indicates that the asymptotic approximation is unreliable for any practical user load under these conditions, and the ergodic approximation should be employed exclusively for all U.

Table II
SWITCHING THRESHOLD U_{SWITCH} VALUES.

Scenario		Operation	
NLoS	**Pilots**	**Centralized**	**Distributed**
Uncorrelated	Orthogonal	$ML_k/2$	$M/2$
	Reuse	$ML_k/4$	$M/4$
Correlated	Orthogonal	$ML_k/4$	$M/4$
	Reuse	0	0

The validity of the asymptotic approximations relies on channel hardening, confirmed by the monotonic MNAE decrease when increasing M in Fig. 4. Centralized operation converges significantly faster than distributed operation, achieving accuracy even at moderate M due to the larger effective array size (ML_k vs. M) involved in joint combining. Conversely, accuracy degrades with pilot reuse and higher user loads (U). The simulation setup includes spatially correlated NLoS components and $L_k = 2$.

IV. Applying User Scheduling, Pilot Assignment and Subarray Selection in XL-MIMO

A widely used benchmark for joint pilot assignment and subarray selection is the greedy algorithm in [5, Sec. 4.4], which follows a three-stage procedure: (i) assign τ_{p} orthogonal pilots to τ_{p} random UEs; (ii) sequentially assign pilots to the remaining UEs so as to

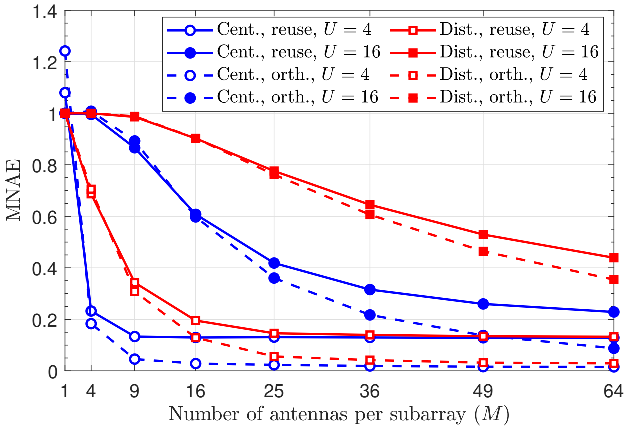


Figure 4. MNAE of the asymptotic SE approximation versus subarray antennas (M).

minimize pilot contamination[6]; and (iii) perform subarray selection, where each subarray locally chooses which τ_p UEs to serve, specifically one UE per pilot (namely, the strongest UE using that pilot). While this approach ensures minimal computational complexity, it does not strictly guarantee full user scheduling, as certain UEs may remain unselected by any subarray.

This section proposes joint user scheduling and pilot assignment algorithms to enhance the minimum per-user SE. Algorithm 1 maximizes the number of scheduled UEs subject to a maximum channel estimation error threshold, γ_th (line 2). It iteratively schedules UEs in descending order of LSF gain (line 3). In each step, the candidate UE is assigned the pilot (either an existing or a new one) that minimizes the maximum channel estimation NMSE across the candidate user set $\tilde{\mathcal{U}}$ (lines 4-9). If the resulting maximum NMSE satisfies γ_th, the UE is scheduled (lines 10-12).

Designed for centralized and distributed operation, respectively, Algorithms 2 and 3 adopt the same framework of Algorithm 1. However, rather than minimizing the maximum channel estimation NMSE, their pilot assignment criteria maximizes the minimum per-user SE, and their user scheduling strategy maximizes the number of scheduled UEs subject to a minimum per-user SE, η_th. More than scheduling UEs and assigning pilots, Algorithm 3 computes the optimal weights used to obtain the final signal estimates from the local estimates (lines 15 and 21).

[6]When assigning a candidate pilot t to a UE k, [5, Sec. 4.4] measures pilot contamination as the sum of the NLoS channel gains of the UEs already using pilot t, evaluated at UE k's strongest subarray. Only the NLoS component is considered because the deterministic LoS component does not contribute to the channel estimation error.

Algorithm 1 NMSE-Based User Scheduling & Pilot Assignment

Require:
NMSE threshold: γ_th
Subarray selection sets: $\{\mathcal{D}_i\}_{i\in\mathcal{K}}$
Channel statistics: $\{\mathbf{Q}_{kl}, \mathbf{R}_{kl}\}_{k\in\mathcal{K}, l\in\mathcal{D}}$, $\{\beta_k\}_{k\in\mathcal{K}}$
Transmit powers: $\{p_k\}_{k\in\mathcal{K}}$
Ensure:
Set of scheduled users: $\mathcal{U}$
Pilot indices: $\{t_k\}_{k\in\mathcal{U}}$
1: Initialize: $\mathcal{U} \leftarrow \emptyset,\ \tau_\mathrm{p} \leftarrow 0,\ \gamma_\mathrm{max} \leftarrow 0$
2: **while** $|\mathcal{U}| < K$ **and** $\gamma_\mathrm{max} \leq \gamma_\mathrm{th}$ **do**
3: $k \leftarrow \arg\max_{i\in\mathcal{K}\setminus\mathcal{U}} \beta_i$ and $\tilde{\mathcal{U}} \leftarrow \mathcal{U} \cup \{k\}$
4: **for** $t = 1$ **to** $\tau_\mathrm{p} + 1$ **do**
5: $t_k \leftarrow t,\ \ \tau \leftarrow \max(\tau_\mathrm{p}, t_k),\ \ \zeta \leftarrow \{1, \ldots, \tau\}$
6: $\boldsymbol{\Psi}_{\tilde{t}l} \leftarrow \sum_{\substack{i\in\tilde{\mathcal{U}} \\ t_i=\tilde{t}}} p_i \tau \mathbf{R}_{il} + \sigma_\mathrm{n}^2 \mathbf{I}_M, \quad l \in \mathcal{D},\ \tilde{t} \in \zeta$
7: $\gamma_\mathrm{max}^{(t)} \leftarrow \max_{i\in\tilde{\mathcal{U}}} \frac{\sum_{l\in\mathcal{D}_i} \mathrm{tr}(\mathbf{R}_{il} - p_i\tau\mathbf{R}_{il}\boldsymbol{\Psi}_{t_il}^{-1}\mathbf{R}_{il})}{\sum_{l\in\mathcal{D}_i} \mathrm{tr}(\mathbf{Q}_{il})}$
8: **end for**
9: $t_k \leftarrow \underset{t\in\{1,\ldots,\tau_\mathrm{p}+1\}}{\arg\min}\ \gamma_\mathrm{max}^{(t)}$ and $\gamma_\mathrm{max} \leftarrow \gamma_\mathrm{max}^{(t_k)}$
10: **if** $\gamma_\mathrm{max} \leq \gamma_\mathrm{th}$ **then**
11: $\mathcal{U} \leftarrow \tilde{\mathcal{U}}$ and $\tau_\mathrm{p} \leftarrow \max(\tau_\mathrm{p}, t_k)$
12: **end if**
13: **end while**

For each candidate pilot, Algorithm 2 evaluates the SINRs of the UEs $\tilde{\mathcal{U}}$ using the ergodic approximation in (30a) when $|\tilde{\mathcal{U}}| > U_\mathrm{switch}$ and the asymptotic approximation (33a) otherwise (lines 6–14), and then it estimates the minimum SE based on these SINRs (line 15). This conditional switching rule improves the accuracy of the SINR estimates by selecting the analytical expression that is known to be more reliable for each loading regime. Its value is chosen based on the numerical analysis in Section III-A.

Similarly, Algorithm 3 evaluates the resulting SINRs for each candidate pilot using (30b) when $|\tilde{\mathcal{U}}| > U_\mathrm{switch}$ and (33b) otherwise (lines 6–14). Since optimal weighting is employed, line 16 approximates the global SINR as the sum of the local SINRs at the selected subarrays, following Proposition 1.

A. Performance Evaluation

This subsection assesses the efficacy of the proposed algorithms in maximizing the minimum per-user SE, demonstrating that strategies based on deterministic SINR expressions achieve performance comparable to those based on instantaneous SINR, but with significantly reduced computational complexity. We assume $L = 16$, $L_k = 4$, and $M = 16$, with the simulation

Algorithm 2 SINR-Based Joint User Scheduling & Pilot Assignment For Centralized Operation

Require:

Switching point: U_{switch}
SE threshold: η_{th}
Subarray selection matrices: $\{\mathbf{D}_k\}_{k\in\mathcal{K}}$
Number of subarrays serving each user: $\{L_k\}_{k\in\mathcal{K}}$
Channel statistics: $\{\mathbf{Q}_k, \mathbf{R}_k, \beta_k\}_{k\in\mathcal{K}}$
Transmit powers: $\{p_k\}_{k\in\mathcal{K}}$

Ensure:

Set of scheduled users: $\mathcal{U}$
Pilot indices: $\{t_k\}_{k\in\mathcal{U}}$

Initialize: $\mathcal{U} \leftarrow \emptyset,\ \tau_{\text{p}} \leftarrow 0,\ \eta_{\min} \leftarrow 2\eta_{\text{th}}$
while $|\mathcal{U}| < K$ **and** $\eta_{\min} > \eta_{\text{th}}$ **do**
  $k \leftarrow \arg\max_{i\in\mathcal{K}\setminus\mathcal{U}} \beta_i$ and $\tilde{\mathcal{U}} \leftarrow \mathcal{U}\cup\{k\}$
  for $t = 1$ **to** $\tau_{\text{p}}+1$ **do**
    $t_k \leftarrow t,\ \tau \leftarrow \max(\tau_{\text{p}}, t_k),\ \zeta \leftarrow \{1,\ldots,\tau\}$
    $\boldsymbol{\Psi}_{\tilde{t}} \leftarrow \sum_{\substack{i\in\tilde{\mathcal{U}}\\ t_i=\tilde{t}}} p_i\tau\mathbf{R}_i + \sigma_{\text{n}}^2\mathbf{I}_{ML}, \quad \tilde{t}\in\zeta$
    $\mathbf{C}_i \leftarrow \mathbf{R}_i - p_i\tau\mathbf{R}_i\boldsymbol{\Psi}_{t_i}^{-1}\mathbf{R}_i, \quad i\in\tilde{\mathcal{U}}$
    if $|\tilde{\mathcal{U}}| > U_{\text{switch}}$ **then**: for $i\in\tilde{\mathcal{U}}$:
      $\overline{\mathbf{Z}}_i \leftarrow \mathbf{D}_i\Big(\sum_{\substack{i'\in\tilde{\mathcal{U}}\\ i'\neq i}} p_{i'}\mathbf{Q}_{i'} + p_i\mathbf{C}_i\Big)\mathbf{D}_i + \sigma_{\text{n}}^2\mathbf{I}_{ML}$
      $\Gamma_i \leftarrow p_i\,\mathrm{tr}\big[\mathbf{D}_i\overline{\mathbf{Z}}_i^{-1}\mathbf{D}_i\left(\mathbf{Q}_i - \mathbf{C}_i\right)\mathbf{D}_i\big]$
    else: for $i\in\tilde{\mathcal{U}}$:
      $\mathbf{S}_{ii'} \leftarrow \mathbf{D}_i(\mathbf{Q}_{i'} - \mathbf{C}_{i'})\mathbf{D}_i,\ i'\in\tilde{\mathcal{U}}$
      $\Gamma_i \leftarrow \dfrac{\mathrm{tr}(\mathbf{S}_{ii}) - \sum\limits_{i'\in\tilde{\mathcal{U}}\setminus\{i\}} \dfrac{\mathrm{tr}(\mathbf{S}_{ii}\mathbf{S}_{ii'})}{\mathrm{tr}(\mathbf{S}_{ii'})}}{\sum\limits_{i'\in\tilde{\mathcal{U}}} \dfrac{p_{i'}}{ML_ip_i}\mathrm{tr}(\mathbf{D}_i\mathbf{C}_{i'}\mathbf{D}_i) + \sigma_{\text{n}}^2}$
    end if
    $\eta_{\min}^{(t)} \leftarrow \min\limits_{i\in\tilde{\mathcal{U}}}\left(1 - \dfrac{\tau}{\tau_{\text{c}}}\right)\log_2(1+\Gamma_i)$
  end for
  $t_k \leftarrow \arg\max\limits_{t\in\{1,\ldots,\tau_{\text{p}}+1\}} \eta_{\min}^{(t)}$ and $\eta_{\min} \leftarrow \eta_{\min}^{(t_k)}$
  if $\eta_{\min} \geq \eta_{\text{th}}$ **then**
    $\mathcal{U} \leftarrow \tilde{\mathcal{U}}$ and $\tau_{\text{p}} \leftarrow \max(\tau_{\text{p}}, t_k)$
  end if
end while

Algorithm 3 SINR-Based Joint User Scheduling & Pilot Assignment For Distributed Operation

Require:

Switching point: U_{switch}
SE threshold: η_{th}
Subarray selection sets: $\{\mathcal{D}_i\}_{i\in\mathcal{K}}$
Channel statistics: $\{\mathbf{Q}_{kl}, \mathbf{R}_{kl}\}_{k\in\mathcal{K},\, l\in\mathcal{D}},\ \{\beta_k\}_{k\in\mathcal{K}}$
Transmit powers: $\{p_k\}_{k\in\mathcal{K}}$

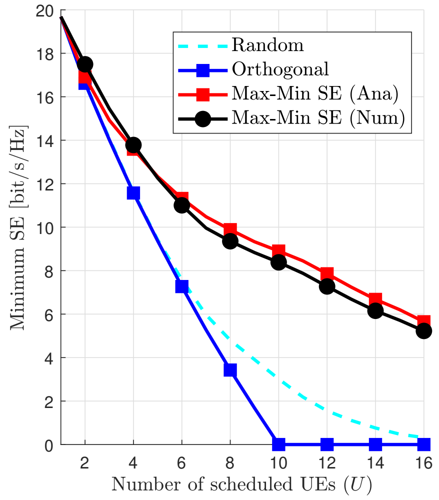
Ensure:

Set of scheduled users: $\mathcal{U}$
Pilot indices: $\{t_k\}_{k\in\mathcal{U}}$
Weights: $\{\mu_{kl}\}_{k\in\mathcal{U},\, l\in\mathcal{D}_k}$

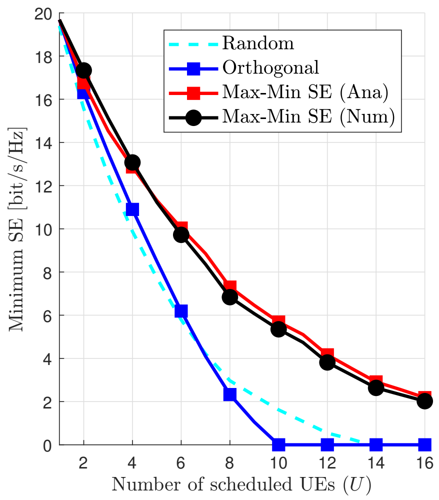
Initialize: $\mathcal{U} \leftarrow \emptyset,\ \tau_{\text{p}} \leftarrow 0,\ \eta_{\min} \leftarrow 2\eta_{\text{th}}$
while $|\mathcal{U}| < K$ **and** $\eta_{\min} > \eta_{\text{th}}$ **do**
  $k \leftarrow \arg\max_{i\in\mathcal{K}\setminus\mathcal{U}} \beta_i$ and $\tilde{\mathcal{U}} \leftarrow \mathcal{U}\cup\{k\}$
  for $t = 1$ **to** $\tau_{\text{p}}+1$ **do**
    $t_k \leftarrow t,\ \tau \leftarrow \max(\tau_{\text{p}}, t_k),\ \zeta \leftarrow \{1,\ldots,\tau\}$
    $\boldsymbol{\Psi}_{\tilde{t}l} \leftarrow \sum_{\substack{i\in\tilde{\mathcal{U}}\\ t_i=\tilde{t}}} p_i\tau\mathbf{R}_{il} + \sigma_{\text{n}}^2\mathbf{I}_{M}, \quad l\in\mathcal{D},\ \tilde{t}\in\zeta$
    $\mathbf{C}_{il} \leftarrow \mathbf{R}_{il} - p_i\tau\mathbf{R}_{il}\boldsymbol{\Psi}_{t_il}^{-1}\mathbf{R}_{il}, \quad l\in\mathcal{D},\ i\in\tilde{\mathcal{U}}$
    $\mathbf{S}_{il} = \mathbf{Q}_{il} - \mathbf{C}_{il}, \quad l\in\mathcal{D},\ i\in\tilde{\mathcal{U}}$
    if $|\tilde{\mathcal{U}}| > U_{\text{switch}}$ **then**: for $i\in\tilde{\mathcal{U}}$ and $l\in\mathcal{D}_i$:
      $\overline{\mathbf{Z}}_{il} \leftarrow \sum\limits_{i'\in\tilde{\mathcal{U}}\setminus\{i\}} p_{i'}\mathbf{Q}_{i'l} + p_i\mathbf{C}_{il} + \sigma_{\text{n}}^2\mathbf{I}_M$
      $\Gamma_{il} \leftarrow p_i\,\mathrm{tr}\big[\overline{\mathbf{Z}}_{il}^{-1}(\mathbf{Q}_{il} - \mathbf{C}_{il})\big]$
    else: for $i\in\tilde{\mathcal{U}}$ and $l\in\mathcal{D}_i$:
      $\Gamma_{il} \leftarrow \dfrac{\mathrm{tr}(\mathbf{S}_{il}) - \sum\limits_{i'\in\tilde{\mathcal{U}}\setminus\{i\}} \dfrac{\mathrm{tr}(\mathbf{S}_{il}\mathbf{S}_{i'l})}{\mathrm{tr}(\mathbf{S}_{i'l})}}{\sum\limits_{i'\in\tilde{\mathcal{U}}} \dfrac{p_{i'}}{Mp_i}\mathrm{tr}(\mathbf{C}_{i'l}) + \sigma_{\text{n}}^2}$
    end if
    $\mu_{il}^{(t)} \leftarrow \Gamma_{il}\Big/\sum\limits_{l'\in\mathcal{D}_i}\Gamma_{il'},\ i\in\tilde{\mathcal{U}}$ and $l\in\mathcal{D}_i$
    $\eta_{\min}^{(t)} \leftarrow \min\limits_{i\in\tilde{\mathcal{U}}}\left(1 - \dfrac{\tau}{\tau_{\text{c}}}\right)\log_2\left(1 + \sum\limits_{l\in\mathcal{D}_i}\Gamma_{il}\right)$
  end for
  $t_k \leftarrow \arg\max\limits_{t\in\{1,\ldots,\tau_{\text{p}}+1\}} \eta_{\min}^{(t)}$ and $\eta_{\min} \leftarrow \eta_{\min}^{(t_k)}$
  if $\eta_{\min} \geq \eta_{\text{th}}$ **then**
    $\mathcal{U} \leftarrow \tilde{\mathcal{U}}$ and $\tau_{\text{p}} \leftarrow \max(\tau_{\text{p}}, t_k)$
    $\mu_{il} \leftarrow \mu_{il}^{(t_k)},\ i\in\mathcal{U}$ and $l\in\mathcal{D}_i$
  end if
end while

setup following the parameters established in Section III-A. Since LSF-based subarray selection yields performance comparable to the SINR-based strategy (as shown in Section III-A), it is adopted hereafter for its lower complexity.

Figure 5 depicts the minimum per-user SE as a function of the number of scheduled users (U). We compare Algorithms 2 and 3 (labeled *Max-Min SE (Ana)*) against a numerical benchmark, *Max-Min SE (Num)*. This benchmark follows the same logic as the proposed algorithms but utilizes the numerical mean SINR for pilot assignment instead of the derived deterministic approximations. Two baselines serve as references: Algorithm 1, which is labeled *Orthogonal* since it assigns mutually orthogonal pilots by minimizing the maximum channel estimation NMSE, and a *Random* pilot assignment. All schemes employ user scheduling based on descending LSF gains and LSF-based subarray selection. The scenario features $K = 16$ UEs and a short coherence block ($\tau_c = 10$), rendering pilot contamination unavoidable.[7]



(a) Centralized Operation.



(b) Distributed Operation.

Figure 5. Minimum per-user SE versus number of scheduled UEs.

Figures 5a and 5b, corresponding to centralized and distributed architectures respectively, show that the centralized approach consistently yields a higher minimum SE. This is expected, as the CPU in a centralized architecture jointly processes signals from all L_k selected subarrays, enabling global interference suppression. Conversely, the distributed architecture relies on local combining, limiting interference mitigation capabilities. Notably, if each UE were served by a single subarray ($L_k = 1$), both architectures would converge to identical performance.

In both architectures, the *Max-Min SE (Num)* strategy effectively mitigates interference by leveraging the true mean SINR; however, it incurs prohibitive computational costs by relying on extensive instantaneous CSI processing. In contrast, the *Max-Min SE (Ana)* strategy achieves remarkably similar performance using only deterministic SINR approximations, offering a tractable solution. The negligible performance gap suggests that the primary gains in pilot assignment stem from resolving large-scale spatial conflicts, which the derived deterministic expressions capture efficiently without requiring Monte Carlo sampling or instantaneous fading optimization.

Furthermore, the proposed strategy rivals the numerical benchmark in maximizing the number of scheduled UEs that satisfy a per-user SE requirement. For instance, given a 5 bit/s/Hz threshold in Fig. 5a, the *Random* and *Orthogonal* baselines schedule only 7 UEs, whereas both *max-min* algorithms successfully schedule all 16 UEs. These results highlight the effectiveness of deterministic SINR-based resource allocation in enhancing fairness within crowded XL-MIMO networks.

While the intelligent strategies maintain robust performance as U increases, the *Random* and *Orthogonal* assignments suffer severe degradation due to uncoordinated pilot reuse and reduced data transmission intervals (caused by increasing pilot sequence lengths), respectively.

Finally, Figure 6 illustrates the minimum per-user SE as a function of the coherence block length (τ_c) for $K = 6$ UEs. A *Max-Min SE (Num - Exhaustive Search)* benchmark is introduced to establish the theoretical performance upper bound. Unlike the sequential approach of the *Max-Min SE (Num)* strategy—where pilots are assigned one-by-one to newly scheduled UEs without re-evaluating previous assignments—the exhaustive search evaluates all possible pilot combinations to identify the global optimum.

Notably, the sequential mechanism employed by the proposed Algorithms 2 and 3 proves highly effective, achieving performance nearly identical to the optimal exhaustive search. Moreover, results suggest that the *Max-Min SE (Ana)* approach may approximate the optimal pilot assignment even more closely than the numerical version. This aligns with observations in Figure 5, where the analytical strategy occasionally outperforms the *Max-Min SE (Num)* benchmark,

[7]Results are plotted up to the full load $U = K$ for illustrative purposes. In practice, the algorithm terminates when the SE threshold is violated, potentially leaving some UEs unscheduled.

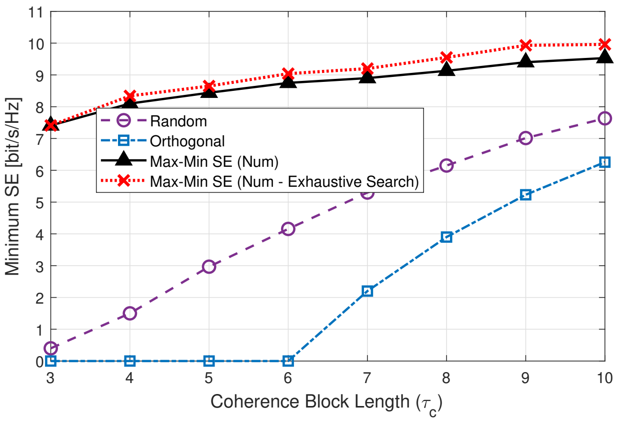


Figure 6. Minimum per-user SE versus coherence block length.

likely because the deterministic expressions effectively capture long-term spatial conflicts without the variance associated with limited instantaneous SINR samples.

V. Conclusion

This paper addressed the critical challenge of resource allocation in XL-MIMO systems, where the reliance on instantaneous CSI for optimization is computationally prohibitive. We introduced novel, closed-form deterministic SINR expressions for both centralized and distributed uplink operations. These expressions, which are valid for Rician fading channels and account for MMSE receive combining and channel estimation, depend exclusively on long-term channel statistics. This key innovation provides a tractable analytical foundation for XL-MIMO system design and optimization, eliminating the need for computationally expensive instantaneous CSI.

Building on these theoretical results, we developed a suite of low-complexity algorithms for joint subarray selection, user scheduling, and pilot assignment. By leveraging the derived statistical SINR approximations, these algorithms are designed to maximize the minimum spectral efficiency among scheduled users, thereby enhancing fairness. Our numerical results validate the high accuracy of the SINR approximations and demonstrate that the proposed statistical-CSI-driven framework achieves performance remarkably close to that of instantaneous-CSI-based benchmarks. The proposed methods effectively exploit the spatial sparsity of user visibility regions to enable more aggressive pilot reuse, significantly improving both fairness and throughput in crowded network scenarios. This work demonstrates that by relying on slowly-varying channel statistics, it is possible to design resource allocation schemes that are both highly efficient and practical for real-world XL-MIMO deployments.

A. Future Research Directions

The findings of this work open up several promising avenues for future research, including:

- *Hardware Impairments and Channel Aging*: Future studies could extend the deterministic SINR framework to incorporate the effects of practical hardware impairments, such as phase noise and power amplifier nonlinearity, as well as channel aging in high-mobility scenarios. Developing robust resource allocation algorithms would be a significant step toward practical implementation.
- *Downlink Transmission and Precoding Design*: This work focused exclusively on the uplink. A natural extension is to develop a similar statistical-CSI-based framework for the downlink, including the design of low-complexity precoding schemes that can achieve near-optimal performance without requiring instantaneous CSI at the transmitter.
- *Energy Efficiency Optimization*: While the proposed algorithms enhance spectral efficiency and fairness, they do not explicitly consider energy consumption. Future research could focus on developing resource allocation strategies that jointly optimize for spectral and energy efficiency, a critical requirement for sustainable 6G networks.
- *Machine Learning-based Approaches*: The statistical nature of the proposed framework makes it well-suited for machine learning-based enhancements. Deep learning models could be trained to learn the complex relationships between long-term channel statistics and optimal resource allocation decisions, potentially leading to even lower-complexity and more adaptive solutions.
- *Integration with Reconfigurable Intelligent Surfaces* (RIS): The synergy between XL-MIMO and RIS is a promising area of investigation. Future work could explore the joint design of subarray selection in XL-MIMO and passive beamforming at the RIS, leveraging statistical CSI to create a programmable and highly efficient wireless environment.
- *Near-Field Communications*: As XL-MIMO systems operate in the near-field, future research should more deeply investigate the implications of spherical wave propagation on channel modeling and resource allocation. Extending the deterministic SINR expressions to explicitly account for

near-field effects would provide a more accurate analytical foundation for such systems.

Appendix A
Proof of Proposition 1

When $M \geq U$, the local ZF combiner satisfies $\mathbf{v}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{il} = \delta_{ik}$. Under high signal-to-noise ratio (SNR) and small channel estimation errors, the L-MMSE combiner converges to the local ZF, and thus obeys the approximation $\mathbf{v}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{il} \approx \delta_{ik}$, yielding the local and global SINR approximations given by

$$\Gamma_{kl} \approx \frac{p_k}{\mathbf{v}_{kl}^{\mathrm{H}}\left(\sum_{i\in\mathcal{U}} p_i\mathbf{C}_{il} + \sigma_{\mathrm{n}}^2\mathbf{I}_M\right)\mathbf{v}_{kl}}, \tag{39}$$

$$\Gamma_k \approx \frac{p_k}{\sum_{l\in\mathcal{D}_k}\mu_{kl}^2\mathbf{v}_{kl}^{\mathrm{H}}\left(\sum_{i\in\mathcal{U}} p_i\mathbf{C}_{il} + \sigma_{\mathrm{n}}^2\mathbf{I}_M\right)\mathbf{v}_{kl}}. \tag{40}$$

Corollary 1 follows directly by rewriting (40) and comparing it with the approximation of Γ_{kl}^{-1} in (39):

$$\Gamma_k^{-1} \approx \sum_{l\in\mathcal{D}_k}\mu_{kl}^2 \underbrace{\frac{1}{p_k}\mathbf{v}_{kl}^{\mathrm{H}}\left(\sum_{i\in\mathcal{U}} p_i\mathbf{C}_{il} + \sigma_{\mathrm{n}}^2\mathbf{I}_M\right)\mathbf{v}_{kl}}_{\Gamma_{kl}^{-1}}.$$

Although L-MMSE and local ZF combiners perform similarly only under low channel estimation errors, the accuracy of the approximation (26) may improve as these errors grow. The resulting increase in $\mathbf{C}_{il}$ dominates the SINR denominator, rendering the neglected interference terms $\mathbf{v}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{il}$ negligible compared to the $\mathbf{v}_{kl}^{\mathrm{H}}\mathbf{C}_{il}\mathbf{v}_{kl}$ components.

The optimal weights for distributed uplink operation are obtained by maximizing the approximate global SINR in (24), given the local SINRs. As $\mu_{kl} \geq 0$ and $\Gamma_{kl} \geq 0$ for all $l \in \mathcal{D}_k$, we have $\sum_{l\in\mathcal{D}_k}\mu_{kl}^2\Gamma_{kl}^{-1} > 0$. Thus, maximizing $\widehat{\Gamma}_k$ is equivalent to solving

$$\min_{\{\mu_{kl}\}_{l\in\mathcal{D}_k}} \quad \sum_{l\in\mathcal{D}_k}\mu_{kl}^2\Gamma_{kl}^{-1} \tag{41a}$$

$$\text{s.t.} \quad \sum_{l\in\mathcal{D}_k}\mu_{kl} = 1, \tag{41b}$$

$$\mu_{kl} \geq 0, \quad \forall l \in \mathcal{D}_k, \tag{41c}$$

where constraints (41b) and (41c) enforce nonnegative, unit-sum weights. Neglecting (41c), which will be satisfied a posteriori, the Lagrangian is

$$\mathcal{L}_k = \sum_{l\in\mathcal{D}_k}\mu_{kl}^2\Gamma_{kl}^{-1} + \lambda_k\left(\sum_{l\in\mathcal{D}_k}\mu_{kl} - 1\right), \tag{42}$$

where λ_k is the lagrangian multiplier. The first-order optimality condition gives [42]

$$\frac{\partial\mathcal{L}_k}{\partial\mu_{kl}} = 2\Gamma_{kl}^{-1}\mu_{kl} + \lambda_k = 0, \quad l \in \mathcal{D}_k, \tag{43}$$

which yields

$$\mu_{kl} = -\frac{1}{2}\Gamma_{kl}\lambda_k, \quad l \in \mathcal{D}_k. \tag{44}$$

Substituting (44) into (41b) results in $\lambda_k = -\frac{2}{\sum_{l\in\mathcal{D}_k}\Gamma_{kl}}$. Substituting this into (44) completes the proof of (25), and inserting (25) into (24) yields (26).

Appendix B
Proof of Theorem 1

We first establish the following lemma regarding the expectation of the inverse of a random matrix.

Lemma 1. *Let $\mathbf{X}$ be an invertible random matrix with mean $\mathbf{A} = \mathbb{E}\{\mathbf{X}\}$. If the fluctuation $\Delta = \mathbf{X} - \mathbf{A}$ is sufficiently small to satisfy $\|\mathbf{A}^{-1}\Delta\| < 1$, then the approximation $\mathbb{E}\{\mathbf{X}^{-1}\} \approx (\mathbb{E}\{\mathbf{X}\})^{-1}$ holds.*

Proof. Using the Neumann series expansion [43], we have $\mathbf{X}^{-1} = (\mathbf{A} + \Delta)^{-1} = \mathbf{A}^{-1}\sum_{n=0}^{\infty}(-1)^n(\Delta\mathbf{A}^{-1})^n$. For small Δ, truncating the series at $n = 0$ and taking the expectation (noting that $\mathbb{E}\{\Delta\} = \mathbf{0}$) yields $\mathbb{E}\{\mathbf{X}^{-1}\} = \mathbf{A}^{-1} + \mathcal{O}(\mathbb{E}\{\|\Delta\|^2\})$, which justifies the approximation. □

The average local SINR in (20) can be expressed as

$$\mathbb{E}\{\Gamma_{kl}\} = p_k\,\mathrm{tr}(\mathbb{E}\{\mathbf{Z}_{kl}^{-1}\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{kl}^{\mathrm{H}}\}). \tag{45}$$

Under typical pilot-reuse patterns (where few or no UEs share the pilot of UE k) and limited LoS overlap across subarrays, $\mathbf{Z}_{kl}$ and $\widehat{\mathbf{h}}_{kl}$ are approximately uncorrelated. Thus,

$$\mathbb{E}\{\mathbf{Z}_{kl}^{-1}\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{kl}^{\mathrm{H}}\} \approx \mathbb{E}\{\mathbf{Z}_{kl}^{-1}\}(\mathbf{Q}_{kl} - \mathbf{C}_{kl}). \tag{46}$$

Furthermore, when $U \gg M$ or when subarray selection favors subarrays with a strong LoS component to UE k, the matrix $\mathbf{Z}_{kl}$ concentrates around its mean $\overline{\mathbf{Z}}_{kl} = \mathbb{E}\{\mathbf{Z}_{kl}\}$. In this regime, $\mathbf{Z}_{kl}$ exhibits small fluctuations, and Lemma 1 implies

$$\mathbb{E}\{\mathbf{Z}_{kl}^{-1}\} \approx (\mathbb{E}\{\mathbf{Z}_{kl}\})^{-1} = \overline{\mathbf{Z}}_{kl}^{-1}. \tag{47}$$

Substituting (47) into (46) yields

$$\mathbb{E}\{\mathbf{Z}_{kl}^{-1}\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{kl}^{\mathrm{H}}\} \approx \overline{\mathbf{Z}}_{kl}^{-1}(\mathbf{Q}_{kl} - \mathbf{C}_{kl}). \tag{48}$$

Inserting (48) into (45) completes the proof of (30b). The proof of (30a) follows similar steps.

Appendix C
Proof of Corollary 1

Evaluating (30b) primarily involves computing the trace $\mathrm{tr}(\mathbf{A}^{-1}\mathbf{B})$, where $\mathbf{A} = \overline{\mathbf{Z}}_{kl}$ and $\mathbf{B} = \mathbf{Q}_{kl} - \mathbf{C}_{kl}$. Let $\mathbf{C} = \mathbf{A}^{-1}\mathbf{B}$. We efficiently compute the diagonal elements of $\mathbf{C}$ by solving the linear systems $\mathbf{A}\mathbf{c}_m = \mathbf{b}_m$ for $m = 1, \ldots, M$, where $\mathbf{c}_m$ and $\mathbf{b}_m$ are the m-th columns of $\mathbf{C}$ and $\mathbf{B}$, respectively.

First, the $\mathbf{LDL}^{\mathrm{H}}$ decomposition of the Hermitian matrix $\mathbf{A}$ requires $\frac{1}{3}(M^3 - M)$ complex multiplications [41, App. B.1.1]. The system $\mathbf{LDL}^{\mathrm{H}}\mathbf{c}_m = \mathbf{b}_m$ is then solved for the m-th row of each column $\mathbf{c}_m$ in three steps: (i) solving $\mathbf{L}\mathbf{z}_m = \mathbf{b}_m$ for each m requires a total of $M \times \frac{1}{2}(M^2 - M)$ complex multiplications; (ii) solving $\mathbf{D}\mathbf{y}_m = \mathbf{z}_m$ involves M complex-by-real divisions per column, *i.e.*, $M \times 2M = 2M^2$ real divisions; and (iii) to obtain the diagonal element $[\mathbf{c}_m]_m$, the system $\mathbf{L}^{\mathrm{H}}\mathbf{c}_m = \mathbf{y}_m$ is solved from the M-th row up to the m-th row, which requires $\sum_{n=0}^{M-m} n = \frac{(M-m)(M-m+1)}{2}$ complex multiplications. Hence, the total cost for all $m = 1, \ldots, M$ is $\sum_{m=1}^{M} \frac{(M-m)(M-m+1)}{2} = \frac{M^3 - M}{6}$ complex multiplications.

Assuming each complex multiplication is implemented via three real multiplications and each real division is computationally equivalent to a real multiplication [41, App. B.1.1], the total complexity of computing $\Gamma_{kl}^{\mathrm{erg}}$ across the L_k subarrays that serve UE k is equivalent to $\mathcal{C}_k^{\mathrm{erg}} = L_k[M^3 - M + \frac{3}{2}(M^3 - M^2) + 2M^2 + \frac{1}{2}(M^3 - M)]$ real multiplications, which simplifies to (32b). The proof of (32a) follows similar steps, by substituting the subarray antenna dimension M with the total number of serving antennas ML_k.

Appendix D
Proof of Theorem 2

Under the assumption of spatially uncorrelated NLoS components, where $\mathbf{C}_{il} = \frac{1}{M}\,\mathrm{tr}(\mathbf{C}_{il})\mathbf{I}_M$, the instantaneous SINR approximation in (39) reduces to[8]

$$\Gamma_{kl} \approx \frac{p_k}{\left(\displaystyle\sum_{i\in\mathcal{U}} \frac{p_i}{M}\,\mathrm{tr}(\mathbf{C}_{il}) + \sigma_{\mathrm{n}}^2\right)\|\mathbf{v}_{kl}\|^2}. \tag{49}$$

The term $\|\mathbf{v}_{kl}\|^2$ is approximated using the local-ZF combiner, which behaves similarly to the L-MMSE combiner at high SNR and sufficiently accurate channel estimates. Letting $\mathcal{U} = \mathcal{K}$, the local ZF matrix at subarray l is $\mathbf{V}_l = \widehat{\mathbf{H}}_l(\widehat{\mathbf{H}}_l^{\mathrm{H}}\widehat{\mathbf{H}}_l)^{-1}$, where $\widehat{\mathbf{H}}_l = [\widehat{\mathbf{h}}_{1l}, \ldots, \widehat{\mathbf{h}}_{Kl}]$ and $\mathbf{v}_{kl}$ is the k-th column of $\mathbf{V}_l$. The block matrix inversion lemma yields [26, Appendix A]

$$\begin{aligned}\|\mathbf{v}_{kl}\|^2 &= [\mathbf{V}_l^{\mathrm{H}}\mathbf{V}_l]_{k,k} = [(\widehat{\mathbf{H}}_l^{\mathrm{H}}\widehat{\mathbf{H}}_l)^{-1}]_{k,k} \\ &= [\widehat{\mathbf{h}}_{kl}^{\mathrm{H}}(\mathbf{I}_M - \mathbf{P}_{kl})\widehat{\mathbf{h}}_{kl}]^{-1},\end{aligned} \tag{50}$$

where $\mathbf{P}_{kl} = \ddot{\mathbf{H}}_{kl}(\ddot{\mathbf{H}}_{kl}^{\mathrm{H}}\ddot{\mathbf{H}}_{kl})^{-1}\ddot{\mathbf{H}}_{kl}^{\mathrm{H}}$ and $\ddot{\mathbf{H}}_{kl} \in \mathbb{C}^{M\times(K-1)}$ denotes the estimated channel matrix at subarray l excluding UE k. Substituting (50) into (49) results in

$$\Gamma_{kl} \approx p_k \frac{\widehat{\mathbf{h}}_{kl}^{\mathrm{H}}(\mathbf{I}_M - \mathbf{P}_{kl})\widehat{\mathbf{h}}_{kl}}{\displaystyle\sum_{i\in\mathcal{U}} \frac{p_i}{M}\,\mathrm{tr}(\mathbf{C}_{il}) + \sigma_{\mathrm{n}}^2}. \tag{51}$$

Lemma 2. *The cross-correlation between the MMSE estimates $\widehat{\mathbf{h}}_{kl}$ and $\widehat{\mathbf{h}}_{il}$ is given by [5, page 107]:*

$$\mathbb{E}\{\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\} = \begin{cases} \mathbf{Q}_{kl} - \mathbf{C}_{kl}, & i = k, \\ \overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{il}^{\mathrm{H}} + \mathbf{A}_{kil}, & t_i = t_k, \\ \overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{il}^{\mathrm{H}}, & \textit{otherwise}, \end{cases} \tag{52}$$

where $\mathbf{A}_{kil} = \sqrt{p_k p_i}\tau_{\mathrm{p}}\mathbf{R}_{kl}\mathbf{\Psi}_{t_k l}^{-1}\mathbf{R}_{il}$. Furthermore, for any matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{M\times M}$ independent of the channel estimates, the quadratic form $\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\mathbf{B}\mathbf{A}\widehat{\mathbf{h}}_{kl}$ converges to its expectation as $M \to \infty$ [33, Lemma 4]:

$$\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\mathbf{B}\mathbf{A}\widehat{\mathbf{h}}_{kl} \xrightarrow[M\to\infty]{a.s.} \mathbb{E}\{\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\mathbf{B}\mathbf{A}\widehat{\mathbf{h}}_{kl}\} = \mathrm{tr}(\mathbf{A}\mathbb{E}\{\widehat{\mathbf{h}}_{kl}\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\}\mathbf{B}). \tag{53}$$

Applying (52)*, this yields*

$$\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\mathbf{B}\mathbf{A}\widehat{\mathbf{h}}_{kl} \xrightarrow[M\to\infty]{a.s.} \begin{cases} \mathrm{tr}[\mathbf{A}(\mathbf{Q}_{kl} - \mathbf{C}_{kl})\mathbf{B}], & i = k, \\ \mathrm{tr}[\mathbf{A}(\overline{\mathbf{h}}_{kl}\overline{\mathbf{h}}_{il}^{\mathrm{H}} + \mathbf{A}_{kil})\mathbf{B}], & t_i = t_k, \\ \overline{\mathbf{h}}_{il}^{\mathrm{H}}\mathbf{B}\mathbf{A}\overline{\mathbf{h}}_{kl}, & t_i \neq t_k. \end{cases} \tag{54}$$

The diagonal elements of $\ddot{\mathbf{H}}_{kl}^{\mathrm{H}}\ddot{\mathbf{H}}_{kl}$ consist of the squared norms $\|\widehat{\mathbf{h}}_{il}\|^2$ for $i \in \mathcal{K} \setminus \{k\}$, which converge to $\mathrm{tr}\,(\mathbf{Q}_{il} - \mathbf{C}_{il})$ as $M \to \infty$, according to (54). In XL-MIMO, the off-diagonal terms are negligible due to the channel spatial non-stationarity.[9] We can thus approximate (51) by

$$\Gamma_{kl} \approx p_k \frac{\widehat{\mathbf{h}}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{kl} - \widehat{\mathbf{h}}_{kl}^{\mathrm{H}}\ddot{\mathbf{H}}_{kl}\mathbf{T}_{kl}^{-1}\ddot{\mathbf{H}}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{kl}}{\displaystyle\sum_{i\in\mathcal{U}} \frac{p_i}{M}\,\mathrm{tr}(\mathbf{C}_{il}) + \sigma_{\mathrm{n}}^2}. \tag{55}$$

[8]This remains accurate for correlated NLoS components in the absence of pilot reuse. This is because the eigenvalues of the matrix term within parentheses in (39) are closely clustered without pilot reuse. Thus, the SINR denominator presents negligible variations when replacing $\mathbf{C}_{il}$ with $\frac{1}{M}\,\mathrm{tr}(\mathbf{C}_{il})\mathbf{I}_M$.

[9]From (54), the off-diagonal elements $\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\widehat{\mathbf{h}}_{i'l}$ $(i \neq i')$ converge to $\overline{\mathbf{h}}_{il}^{\mathrm{H}}\overline{\mathbf{h}}_{i'l}$ as $M \to \infty$, since UEs are typically assigned distinct pilot sequences. In XL-MIMO, however, these off-diagonal terms are negligible compared to the diagonal components, since the large array dimensions ensure that LoS footprints of different UEs have minimal overlap across subarrays. This contrasts with conventional massive MIMO, where VRs typically encompass the entire co-located array.

where $\mathbf{T}_{kl} = \mathrm{diag}(\{\|\widehat{\mathbf{h}}_{il}\|^2\}_{i\in\mathcal{U}\setminus\{k\}})$ contains only the diagonal terms of $\ddot{\mathbf{H}}_{kl}^{\mathrm{H}}\ddot{\mathbf{H}}_{kl}$.

Applying the asymptotic results from (54), the terms in the numerator converge as follows:

$$\widehat{\mathbf{h}}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{kl} \xrightarrow{M\to\infty} \mathrm{tr}(\mathbf{Q}_{kl} - \mathbf{C}_{kl}) \tag{56}$$

and

$$\begin{aligned}\widehat{\mathbf{h}}_{kl}^{\mathrm{H}}\ddot{\mathbf{H}}_{kl}\mathbf{T}_{kl}^{-1}\ddot{\mathbf{H}}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{kl} &= \sum_{i\in\mathcal{U}\setminus\{k\}} \frac{\widehat{\mathbf{h}}_{kl}^{\mathrm{H}}\widehat{\mathbf{h}}_{il}\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\widehat{\mathbf{h}}_{kl}}{\widehat{\mathbf{h}}_{il}^{\mathrm{H}}\widehat{\mathbf{h}}_{il}} \\ &\xrightarrow{M\to\infty} \frac{\mathrm{tr}(\mathbf{X}_{kl}\mathbf{X}_{il})}{\mathrm{tr}(\mathbf{X}_{il})}.\end{aligned} \tag{57}$$

Substituting these into (55) concludes the proof of (33b). The proof of (33a) folllows similar steps.

References

[1] Z. Wang, J. Zhang, H. Du, D. Niyato, S. Cui, B. Ai, M. Debbah, K. B. Letaief, and H. V. Poor, "A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 3, pp. 1560–1605, 2024.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[3] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[4] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[5] Özlem Tugfe Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Foundations and Trends® in Signal Processing*, vol. 14, no. 3-4, pp. 162–472, 2021. [Online]. Available: http://dx.doi.org/10.1561/2000000109

[6] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—what is next?: Five promising research directions for antenna arrays," *Digital Signal Processing*, vol. 94, pp. 3–20, 2019.

[7] Y. Han, S. Jin, C.-K. Wen, and X. Ma, "Channel estimation for extremely large-scale massive MIMO systems," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 633–637, 2020.

[8] J. Zhang, J. Zhang, E. Björnson, and B. Ai, "Local partial zero-forcing combining for cell-free massive MIMO systems," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8459–8473, 2021.

[9] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions," *IEEE Communications Magazine*, vol. 61, no. 1, pp. 40–46, 2023.

[10] 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.

[11] J. C. Marinello, T. Abrão, A. Amiri, E. de Carvalho, and P. Popovski, "Antenna selection for improving energy efficiency in XL-MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13 305–13 318, 2020.

[12] D. Liu, J. Wang, Y. Li, Y. Han, R. Ding, J. Zhang, S. Jin, and T. Q. S. Quek, "Location-based visible region recognition in extra-large massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 8186–8191, 2023.

[13] J. H. I. de Souza, J. C. Marinello, A. Amiri, and T. Abrão, "Qos-aware user scheduling in crowded XL-MIMO systems under non-stationary multi-state LoS/NLoS channels," *IEEE Transactions on Vehicular Technology*, pp. 1–13, 2023.

[14] S. Mashdour, A. R. Flores, S. Salehi, R. C. De Lamare, A. Schmeink, and P. B. Da Silva, "Robust resource allocation in cell-free massive MIMO systems," *IEEE Transactions on Communications*, pp. 1–1, 2025.

[15] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2020.

[16] S. Buzzi, C. D'Andrea, M. Fresia, Y.-P. Zhang, and S. Feng, "Pilot assignment in cell-free massive MIMO based on the hungarian algorithm," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 34–37, 2021.

[17] K. Zhi, C. Pan, H. Ren, and K. Wang, "Ergodic rate analysis of reconfigurable intelligent surface-aided massive MIMO systems with ZF detectors," *IEEE Communications Letters*, vol. 26, no. 2, pp. 264–268, 2022.

[18] C. D'Andrea, A. Garcia-Rodriguez, G. Geraci, L. G. Giordano, and S. Buzzi, "Analysis of UAV communications in cell-free massive MIMO systems," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 133–147, 2020.

[19] H. Lei, J. Zhang, Z. Wang, H. Xiao, and B. Ai, "Uplink performance of cell-free extremely large-scale MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 74, no. 5, pp. 8339–8344, 2025.

[20] Y. Zhang, W. Xia, H. Zhao, Y. Mao, J. Zhang, and G. Zheng, "Enhancing uplink performance for cell-free massive MIMO with low-resolution ADCs by RSMA," *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 3, pp. 720–735, 2025.

[21] M. A. Hosany, "On the performance analysis of multi-user massive MIMO systems with error vector signals for 5G cellular networks," *Wireless Personal Communications*, vol. 123, no. 2, pp. 917–934, 2022. [Online]. Available: https://doi.org/10.1007/s11277-021-09162-z

[22] J. Li, Q. Pan, Z. Wan, P. Zhu, D. Wang, M. Lou, J. Jin, F. Liu, and X. You, "Low altitude 3-D coverage performance analysis of cell-free RAN for 6G systems," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 12, pp. 16 163–16 176, 2023.

[23] S.-N. Jin, D.-W. Yue, and H. H. Nguyen, "Spectral and energy efficiency in cell-free massive MIMO systems over correlated rician fading," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2822–2833, 2021.

[24] Y. Fan, S. Wu, X. Bi, and G. Li, "Power allocation for cell-free massive MIMO ISAC systems with otfs signal," *IEEE Internet of Things Journal*, vol. 12, no. 7, pp. 9314–9331, 2025.

[25] X. Yang, F. Cao, M. Matthaiou, and S. Jin, "On the uplink transmission of extra-large scale massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 229–15 243, 2020.

[26] A. Ali, E. D. Carvalho, and R. W. Heath, “Linear receivers in non-stationary massive MIMO channels with visibility regions,” *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 885–888, June 2019.

[27] S. Chen, J. Zhang, E. Björnson, J. Zhang, and B. Ai, “Structured massive access for scalable cell-free massive MIMO systems,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 4, pp. 1086–1100, 2021.

[28] J. Qiu, K. Xu, X. Xia, Z. Shen, and W. Xie, “Downlink power optimization for cell-free massive MIMO over spatially correlated rayleigh fading channels,” *IEEE Access*, vol. 8, pp. 56 214–56 227, 2020.

[29] T. H. Nguyen, T. V. Chien, H. Q. Ngo, X. N. Tran, and E. Björnson, “Pilot assignment for joint uplink-downlink spectral efficiency enhancement in massive MIMO systems with spatial correlation,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 8292–8297, 2021.

[30] M. Zhou, Y. Zhang, X. Qiao, and L. Yang, “Spatially correlated rayleigh fading for cell-free massive MIMO systems,” *IEEE Access*, vol. 8, pp. 42 154–42 168, 2020.

[31] Q. Sun, X. Ji, Z. Wang, X. Chen, Y. Yang, J. Zhang, and K.-K. Wong, “Uplink performance of hardware-impaired cell-free massive MIMO with multi-antenna users and superimposed pilots,” *IEEE Transactions on Communications*, vol. 71, no. 11, pp. 6711–6726, 2023.

[32] N. Li and P. Fan, “Joint data power control and LSFD design in distributed cell-free massive MIMO under non-ideal ue hardware,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2024, no. 9, pp. 1–17, 2024. [Online]. Available: https://doi.org/10.1186/s13638-024-02334-y

[33] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, “Large system analysis of linear precoding in correlated miso broADCast channels under limited feedback,” *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.

[34] H. Haritha, D. N. Amudala, R. Budhiraja, and A. K. Chaturvedi, “Superimposed versus regular pilots for hardware impaired rician-faded cell-free massive MIMO systems,” *IEEE Transactions on Communications*, vol. 72, no. 11, pp. 6688–6706, 2024.

[35] V. Tentu, E. Sharma, D. N. Amudala, and R. Budhiraja, “UAV-enabled hardware-impaired spatially correlated cell-free massive MIMO systems: Analysis and energy efficiency optimization,” *IEEE Transactions on Communications*, vol. 70, no. 4, pp. 2722–2741, 2022.

[36] Z. Wang, J. Zhang, E. Björnson, and B. Ai, “Uplink performance of cell-free massive MIMO over spatially correlated rician fading channels,” *IEEE Communications Letters*, vol. 25, no. 4, pp. 1348–1352, 2021.

[37] X. Ma, X. Lei, P. T. Mathiopoulos, K. Yu, and X. Tang, “Scalable cell-free massive MIMO systems with finite resolution ADCs/DACs over spatially correlated rician fading channels,” *IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 7699–7716, 2023.

[38] H. Kaur and A. Kansal, “Closed-form analysis of RZF in multicell massive MIMO over correlated rician channel,” *Wireless Personal Communications*, vol. 131, no. 3, pp. 1685–1719, 2023. [Online]. Available: https://doi.org/10.1007/s11277-023-10518-w

[39] S. Mukherjee and R. Chopra, “Performance analysis of cell-free massive MIMO systems in LoS/ NLoS channels,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6410–6423, 2022.

[40] O. Ozdogan, E. Bjornson, and E. G. Larsson, “Massive MIMO with spatially correlated rician fading channels,” *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3234–3250, 2019.

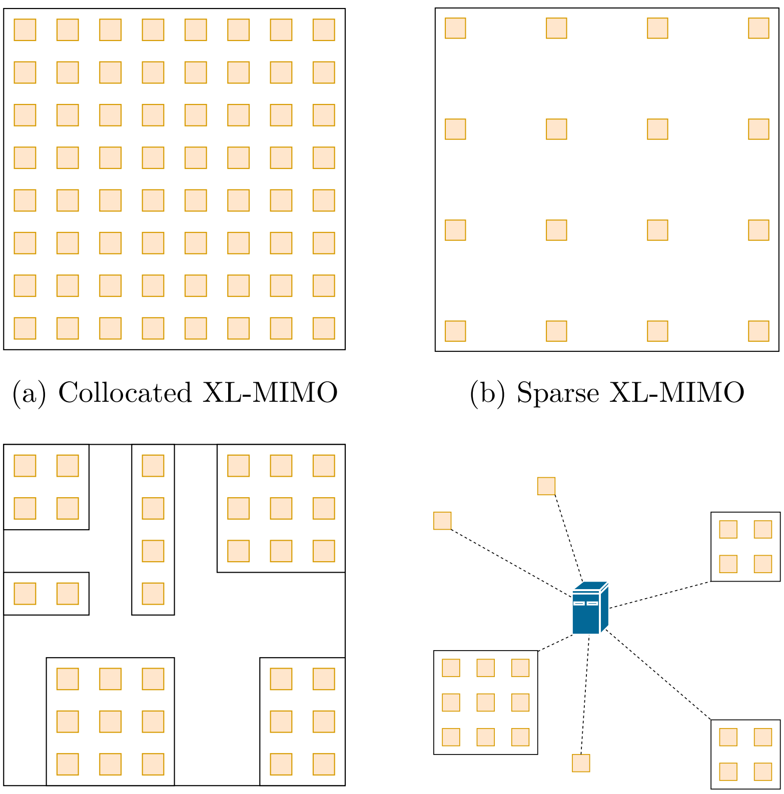
[41] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO networks: Spectral, energy, and hardware efficiency,” *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

[42] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[43] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2012.

APPENDIX E – XL-MIMO ARCHITECTURES

XL-MIMO can be deployed in four main architectures: collocated, sparse, modular, and distributed [13], which are illustrated in Figure E.1. The standard array architecture is the collocated one, where adjacent antennas are typically separated by half wavelength. Since a large continuous platform may not always be available, it may face practical deployment difficulty as the number of antennas drastically increases. The sparse XL-MIMO architecture is characterized by antenna spacings larger than half wavelength [71], and enables a larger aperture without extra antennas. It also faces practical deployment difficulty. In a modular XL-MIMO architecture, the antennas are arranged in compact modules [72,73]. Within each module, elements remain half-wavelength–spaced. The modules can be spaced farther apart to match deployment constraints. For example, when the modular XL-MIMO is mounted to the facades of buildings, the neighbouring modules can be separated by windows [13]. Finally, in distributed XL-MIMO, the antennas are distributed in multiple access points (APs) over a large geographical region, instead of being accommodated on a common platform [13]. Typical distributed XL-MIMO systems include coordinated multipoint (CoMP) [74], cloud radio access network (C-RAN) [75], network MIMO [76], and cell-free massive MIMO [44, 77, 78].



(a) Collocated XL-MIMO (b) Sparse XL-MIMO

(c) Modular XL-MIMO (d) Distributed XL-MIMO

Figure E.1 – Different architectures of XL-MIMO.

APPENDIX F – APPROXIMATION OF THE INVERSE EXPECTATION

Theorem F.1. *Let* $\mathbf{X}$ *be an invertible random matrix with mean* $\mathbf{A} = \mathbb{E}\{\mathbf{X}\}$*. The approximation* $\mathbb{E}\{\mathbf{X}^{-1}\} \approx (\mathbb{E}\{\mathbf{X}\})^{-1}$ *is valid as long as the random fluctuation* $\Delta = \mathbf{X} - \mathbf{A}$ *is sufficiently small to satisfy* $\|\mathbf{A}^{-1}\Delta\| < 1$*.*

Proof. Since

$$\mathbf{X} = \mathbf{A} + \Delta = \mathbf{A}(\mathbf{I} + \mathbf{A}^{-1}\Delta), \tag{F.1}$$

we can express the inverse as

$$\mathbf{X}^{-1} = (\mathbf{I} + \mathbf{A}^{-1}\Delta)^{-1}\mathbf{A}^{-1}. \tag{F.2}$$

Given the condition $\|\mathbf{A}^{-1}\Delta\| < 1$, the inverse can be expanded using a convergent Neumann series [79, 80]:

$$\mathbf{X}^{-1} = \left(\sum_{n=0}^{\infty}(-1)^n(\mathbf{A}^{-1}\Delta)^n\right)\mathbf{A}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\Delta\mathbf{A}^{-1} + (\mathbf{A}^{-1}\Delta)^2\mathbf{A}^{-1} - \cdots . \tag{F.3}$$

Taking the expectation of the series term-by-term, and noting that $\mathbb{E}\{\Delta\} = \mathbf{0}$ by definition, the first-order term vanishes:

$$\mathbb{E}\{\mathbf{X}^{-1}\} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\underbrace{\mathbb{E}\{\Delta\}}_{\mathbf{0}}\mathbf{A}^{-1} + \mathbb{E}\{(\mathbf{A}^{-1}\Delta)^2\mathbf{A}^{-1}\} - \cdots \tag{F.4}$$

Consequently, the expectation is dominated by the zero-order term, with the remaining error scaling quadratically with the fluctuation:

$$\mathbb{E}\{\mathbf{X}^{-1}\} = \mathbf{A}^{-1} + \mathcal{O}(\mathbb{E}\{\|\Delta\|^2\}) \approx \mathbf{A}^{-1} = \Big(\mathbb{E}\{\mathbf{X}\}\Big)^{-1}. \tag{F.5}$$

This completes the proof. $\square$

APPENDIX G – SHADOWING

Both LoS and NLoS shadow fading, are denoted by $\overline{\mathcal{X}}_{kl}$ and $\check{\mathcal{X}}_{kl}$, and modeled as independent log-normal random variables: $\ln(\overline{X}_{kl}), \ln(\check{X}_{kl}) \sim \mathcal{N}(0, \sigma_{\mathrm{SF}}^2)$. The random variables $\overline{\mathcal{X}}_{kl}$ and $\check{\mathcal{X}}_{kl}$ behave as multiplicative terms in the LSF channel power gain, given by (3.4) and (3.12).

The shadow fading ($\mathcal{X}_{kl}$) is modeled as a log-normal random variable; hence, the corresponding variable in dB scale, defined as

$$F_{kl} = 10 \log_{10}(\mathcal{X}_{kl}), \tag{G.1}$$

follows a normal distribution according to [81]

$$F_{kl} \sim \mathcal{N}\left[0, \sigma_{\mathrm{SF}}^2 (1 - \mathrm{e}^{-d_{kl}/\delta})^2\right], \tag{G.2}$$

where σ_{SF} is the shadow fading standard deviation and δ denotes the decorrelation distance. Notably, the variance of F_{kl} increases with distance and converges to σ_{SF}^2 when $d_{kl} \gg \delta$.

While shadow fading is often modeled as independent across links, in reality, adjacent links exhibit spatially correlated shadowing because they are influenced by a largely overlapping set of environmental obstructions. Neglecting these correlations may lead to an overestimation of diversity gains in crowded XL-MIMO systems, where users are typically in close proximity, thereby affecting the accuracy of network performance simulations [81]. According to [81, 82], the cross-correlation between F_{kl} and $F_{k'l'}$ is given by:

$$\mathbb{E}\{F_{kl}F_{k'l'}\} = \frac{\sigma_{\mathrm{SF}}^2}{2} \frac{(1 - \mathrm{e}^{-d_{kl}/\delta})(1 - \mathrm{e}^{-d_{k'l'}/\delta})}{\sqrt{(1 + \mathrm{e}^{-d_{kl}/\delta})(1 + \mathrm{e}^{-d_{k'l'}/\delta})}} (\mathrm{e}^{-d_{kl'}/\delta} + \mathrm{e}^{-d_{k'l}/\delta} + \mathrm{e}^{-D_{kk'}^{\mathrm{UE}}/\delta} + \mathrm{e}^{-D_{ll'}^{\mathrm{SA}}/\delta}), \tag{G.3}$$

where $D_{kk'}^{\mathrm{UE}}$ is the distance between users k and k', and $D_{ll'}^{\mathrm{SA}}$ is the distance between subarrays l and l'. As expected, the closer the users k and k' (or subarrays l and l') are to each other, the stronger the correlation between F_{kl} and $F_{k'l'}$.

The direct signal path, *i.e.*, the LoS component, is generally less influenced by obstacles, resulting in reduced variability in shadow fading. Conversely, in NLoS scenarios, obstacles such as buildings, trees, or vehicles block the direct path, causing the signal to propagate through reflections, diffraction, and scattering. This leads to greater variability and higher standard deviation in the shadowing effect [83]. Due to the shadow fading in LoS and NLoS conditions having different standard deviations, it is desirable to adopt different notations for them: in this work, $\overline{\mathcal{X}}_{kl}$ and $\check{\mathcal{X}}_{kl}$.

APPENDIX H – NETWORK SCALABILITY

The network must be designed to be scalable in the sense that one can add more subarrays and/or UEs without having to increase the capabilities of the existing ones. An XL-MIMO network is scalable if all the following tasks remain practically implementable at each subarray when $K \to \infty$ [9]:

1. Signal processing for channel estimation;
2. Signal processing for data reception and transmission;
3. Fronthaul signaling for data and CSI sharing;
4. Power allocation optimization.

Tasks 1–3 remain bounded (and therefore practically implementable) if each subarray serves only a finite number of UEs. If the number of users served by each subarray remains finite as $K \to \infty$, then both computational complexity and fronthaul load remain finite, since each subarray only needs to compute channel estimates and precoding/combining vectors for the UEs that it is serving, and it only needs to send/receive data related to these UEs over the fronthaul links. This fulfills the first three scalability requirements. Ideally, we should assign subarrays so that each UE is served by all the subarrays that influence its performance noticeably [9].

The fourth scalability condition is not guaranteed by the finite number of served UEs. If power allocation is done in a centralized way that depends on all K UEs, complexity explodes as $K \to \infty$. Even if each subarray only cares about its own K_l UEs, we still need to be sure that no subarray ever needs to solve a problem whose size grows with K. Hence, to satisfy the fourth scalability condition, *i.e.*, to assure that the complexity of the power allocation is finite as $K \to \infty$, each subarray must choose its powers locally, and based only on the UEs it serves. Furthermore, each UE's transmit power must be chosen by exactly one of its serving subarrays, and using only that subarray's local information. An example of scalable pilot allocation scheme is every UE simply transmitting at its full power. However, it is much harder to create scalable algorithms that operate close to the network-wide power allocation benchmarks [9].

APPENDIX I – CHANNEL COHERENCE BLOCK

Wireless channels are linear, due to the superposition principle prescribed by Maxwell's equations [9], and time-varying, since small movements on the order of a wavelength—made by the transmitter, receiver or objects in the wireless propagation environment—can substantially change the channel [9]. However, over a sufficiently short interval, called *channel coherence time* and denoted by T_{c}, the channel may be treated as constant and, therefore, the communication system is time-invariant. Likewise, over a sufficiently narrow frequency band—the *channel coherence bandwidth* B_{c}—the channel's frequency response is approximately flat.

A *channel coherence block* is a time-frequency block whose time duration is T_{c} and frequency width is B_{c}. According to the Nyquist-Shannon sampling theorem, the number of symbols per block is $\tau_{\mathrm{c}} = T_{\mathrm{c}}B_{\mathrm{c}}$. Within a coherence block, the channel between two antennas is constant and frequency-flat, and hence it can described by a complex-valued scalar coefficient. The communication system needs to estimate the channel properties once per channel coherence interval. Under the block-fading model—which assumes that the channel fading takes one independent random realization in each coherence block—one can study one block at a time without loss of generality [9].

The coherence time is typically a few milliseconds in mobile networks [9], but it depends on the mobility and the carrier frequency. It is common to approximate the coherence time as the time the UE takes to move a quarter of a wavelength, such that $T_{\mathrm{c}} = \lambda/(4v)$, where v is the UE velocity [84, Sec.2.3]. Hence, the higher the mobility, the faster the channel changes. Note also that if the carrier frequency is high, the channel changes more rapidly for a given mobility.

The coherence bandwidth can be approximated by $B_{\mathrm{c}} = 1/(2T_{\mathrm{d}})$, where T_{d} is the channel delay spread, i.e., the time difference between the earliest and last propagation path.

The coherence block size ranges from hundreds (when mobility and channel dispersion are high) to hundred of thousands of samples (when mobility and channel dispersion are low). A typical value for the outdoor scenario is $\tau_{\mathrm{c}} = 200$ samples [8, 85].

I.1 Time-Division-Duplex Protocol

TDD is a communication technique that enables a single frequency channel to be shared for both uplink and downlink transmissions by dividing the transmission time into distinct time slots. This temporal separation permits both directions to share the same spectrum without simultaneous transmission, thereby minimizing the risk of interference

[66, 84]. One of the primary benefits of TDD is its flexibility, since the ratio of uplink to downlink time slots can be dynamically adapted to suit varying traffic loads, making it particularly effective in environments where data demand is asymmetric [66,86]. However, the reuse of the same frequency for both transmission directions may require strategies to reduce self-interference, and demands timing and synchronization to prevent transmission collisions.

In TDD operation mode, the τ_c transmission symbols of a coherence block are used for three purposes: τ_p symbols for uplink pilots, τ_u symbols for uplink data, and τ_d symbols for downlink data. Note that the pilots and data are transmitted at different times and $\tau_\mathrm{c} = \tau_\mathrm{p} + \tau_\mathrm{u} + \tau_\mathrm{d}$. The uplink pilots must be transmitted first so that the channels are updated before performing uplink receive combining and downlink transmit precoding.

TDD is extensively used in various wireless systems, including long term evolution (LTE), Wi-Fi and mid-band 5G (notably in the 3.6 GHz range). In the context of 5G, TDD is especially advantageous because it permits a more flexible allocation of resources, thereby accommodating diverse and dynamic service requirements, such as enhanced mobile broadband and ultra-reliable low-latency communications [66, 86].

TDD is a half-duplex (HD) system, as uplink and downlink transmissions do not occur within the same time slot. In contrast, frequency-division-duplexing (FDD) utilizes separate frequency bands for uplink and downlink, enabling simultaneous operation in both directions without interference between the two streams [84, 87].

TDD offers several advantages over FDD, making it particularly attractive in modern wireless communication systems. Some of the key benefits include:

- **Dynamic traffic adaptability:** TDD allows for flexible allocation of uplink and downlink time slots based on current traffic conditions. This means that if one direction (typically downlink) experiences heavier data transmission demands, additional time slots can be allocated accordingly. In contrast, FDD requires fixed frequency bands, which can lead to inefficient spectrum usage when traffic is asymmetrical [66, 84, 88].

- **Spectrum efficiency:** since TDD uses a single frequency band for both uplink and downlink, it maximizes the efficient utilization of available spectrum. This is especially valuable in frequency-constrained environments, as it avoids the need for separate, and potentially underutilized, frequency allocations for each transmission direction [84].

- **Simplified radio frequency (RF) hardware:** by using a single frequency band, TDD eliminates the need for duplexers to separate uplink and downlink transmissions, resulting in simpler, less costly RF designs [87].

APPENDIX J – COMPLEXITY OF THE ERGODIC APPROXIMATION FOR THE GLOBAL UPLINK SINR IN CENTRALIZED OPERATION

Evaluating (3.61) involves computing $\mathrm{tr}(\mathbf{C})$, where $\mathbf{C} = \mathbf{A}^{-1}\mathbf{B}$. Here, $\mathbf{A}$ and $\mathbf{B}$ represent the effective $ML_k \times ML_k$ submatrices containing only the rows and columns selected by $\mathbf{D}_k$ from $\overline{\mathbf{Z}}_k$ and $(\mathbf{Q}_k - \mathbf{C}_k)$, respectively. We efficiently compute the diagonal elements of $\mathbf{C}$ by solving $\mathbf{A}\mathbf{c}_n = \mathbf{b}_n$ for $n = 1, \ldots, ML_k$, where $\mathbf{c}_n$ and $\mathbf{b}_n$ are the n-th columns of $\mathbf{C}$ and $\mathbf{B}$.

The $\mathbf{LDL}^{\mathrm{H}}$ decomposition of the Hermitian matrix $\mathbf{A}$ facilitates a computationally and memory-efficient hardware implementation, particularly when computing the product of a matrix inverse and another matrix. This approach avoids explicit matrix inversion and requires $\frac{1}{3}((ML_k)^3 - ML_k)$ complex multiplications [8, App. B.1.1].

The resulting linear system, $\mathbf{LDL}^{\mathrm{H}}\mathbf{c}_n = \mathbf{b}_n$, is solved for the n-th row of each column $\mathbf{c}_n$ in three steps:[1] (i) solving $\mathbf{L}\mathbf{z}_n = \mathbf{b}_n$ for each n requires $\sum_{i=0}^{ML_k-1} i = \frac{ML_k(ML_k-1)}{2}$ complex multiplications; (ii) solving $\mathbf{D}\mathbf{y}_n = \mathbf{z}_n$ involves ML_k complex-by-real divisions, *i.e.*, $2ML_k$ real divisions; and (iii) solving $\mathbf{L}^{\mathrm{H}}\mathbf{c}_n = \mathbf{y}_n$ until obtaining the n-th diagonal element of $\mathbf{C}$, *i.e.*, $[\mathbf{c}_n]_n$, requires $\sum_{i=0}^{ML_k-n} i = \frac{(ML_k-n)(ML_k-n+1)}{2}$ complex multiplications.

By adopting the convention that a single complex multiplication can be implemented using three real multiplications, and assuming the computational cost of a real division is equivalent to that of a real multiplication, the total complexity for computing Γ_k^{erg} for all $n = 1, \ldots, ML_k$ is equivalent to

$$\begin{aligned}
\mathcal{C}_k^{\mathrm{cent,erg}} &= \sum_{n=1}^{ML_k} \left(3\frac{ML_k(ML_k-1)}{2} + 2ML_k + 3\frac{(ML_k-n)(ML_k-n+1)}{2} \right) \\
&= 3\frac{(ML_k)^2(ML_k-1)}{2} + 2(ML_k)^2 + \frac{3}{2}\sum_{n=1}^{ML_k}(ML_k-n)(ML_k-n+1) \\
&\overset{(a)}{=} \frac{3}{2}(ML_k)^3 + \frac{1}{2}(ML_k)^2 + \frac{3}{2}\sum_{n'=0}^{ML_k-1} n'(n'+1) \\
&= \frac{3}{2}(ML_k)^3 + \frac{1}{2}(ML_k)^2 + \frac{3}{2}\frac{(ML_k)^3 - ML_k}{3} \\
&= 2(ML_k)^3 + \frac{1}{2}(ML_k)^2 - \frac{1}{2}ML_k
\end{aligned}$$

real multiplications, where (a) comes from defining $n' = ML_k - n$.

[1] $\mathbf{D}$ and $\mathbf{L}$ are diagonal and lower triangular matrices of dimension $ML_k \times ML_k$, respectively.

APPENDIX K – COMPLEXITY OF THE ERGODIC APPROXIMATION FOR THE LOCAL UPLINK SINR IN DISTRIBUTED OPERATION

Evaluating (3.113) involves computing $\mathrm{tr}(\mathbf{C})$, where $\mathbf{C} = \mathbf{A}^{-1}\mathbf{B}$, $\mathbf{A} = \overline{\mathbf{Z}}_{kl}$, and $\mathbf{B} = \mathbf{Q}_{kl} - \mathbf{C}_{kl}$. We efficiently compute the diagonal elements of $\mathbf{C}$ by solving $\mathbf{A}\mathbf{c}_m = \mathbf{b}_m$ for $m = 1, \ldots, M$, where $\mathbf{c}_m$ and $\mathbf{b}_m$ are the m-th columns of $\mathbf{C}$ and $\mathbf{B}$, respectively.

The $\mathbf{LDL}^{\mathrm{H}}$ decomposition of the Hermitian matrix $\mathbf{A}$ facilitates a computationally and memory-efficient hardware implementation, particularly when computing the product of a matrix inverse and another matrix. This approach avoids explicit matrix inversion and requires $\frac{1}{3}(M^3 - M)$ complex multiplications [8, App. B.1.1].

The resulting linear system, $\mathbf{LDL}^{\mathrm{H}}\mathbf{c}_m = \mathbf{b}_m$, is solved for the m-th row of each column $\mathbf{c}_m$ in three steps:[1] (i) solving $\mathbf{L}\mathbf{z}_m = \mathbf{b}_m$ for each m requires $\sum_{n=0}^{M-1} n = \dfrac{M(M-1)}{2}$ complex multiplications; (ii) solving $\mathbf{D}\mathbf{y}_m = \mathbf{z}_m$ involves M complex-by-real divisions, *i.e.*, $2M$ real divisions; and (iii) solving $\mathbf{L}^{\mathrm{H}}\mathbf{c}_m = \mathbf{y}_m$ until obtaining the m-th diagonal element of $\mathbf{C}$, *i.e.*, $[\mathbf{c}_m]_m$, requires $\sum_{n=0}^{M-m} n = \dfrac{(M-m)(M-m+1)}{2}$ complex multiplications.

By adopting the convention that a single complex multiplication can be implemented using three real multiplications, and assuming the computational cost of a real division is equivalent to that of a real multiplication, the total complexity for computing $\Gamma_{kl}^{\mathrm{erg}}$ for all $m = 1, \ldots, M$ at the L_k serving subarrays is equivalent to

$$
\begin{aligned}
\mathcal{C}_k^{\mathrm{dist,erg}} &= L_k \sum_{m=1}^{M} \left(3\frac{M(M-1)}{2} + 2M + 3\frac{(M-m)(M-m+1)}{2} \right) \\
&= L_k \left[3\frac{M^2(M-1)}{2} + 2M^2 + \frac{3}{2}\sum_{m=1}^{M} (M-m)(M-m+1) \right] \\
&\overset{(a)}{=} L_k \left[\frac{3}{2}M^3 + \frac{1}{2}M^2 + \frac{3}{2}\sum_{m'=0}^{M-1} m'(m'+1) \right] \\
&= L_k \left[\frac{3}{2}M^3 + \frac{1}{2}M^2 + \frac{3}{2}\frac{M^3 - M}{3} \right] \\
&= 2M^3 L_k + \frac{1}{2}M^2 L_k - \frac{1}{2}M L_k
\end{aligned}
$$

real multiplications, where (a) comes from defining $m' = M - m$.

[1] $\mathbf{D}$ and $\mathbf{L}$ are diagonal and lower triangular matrices, respectively.