Massive MIMO in Cellular Networks

Vom Promotionsausschuss der
Technischen Universität Hamburg

zur Erlangung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing.)

genehmigte Dissertation

von
Günter Martin Kurras

aus
Halle (Saale), Deutschland

2020

Vorsitzender des Prüfungsausschusses:
Prof. Dr.-Ing. Hoc Khiem Trieu

Gutachter:
Prof. Dr.-Ing. Gerhard Bauch
Prof. Dr.-Ing. habil. Tobias Weber

Tag der mündlichen Prüfung:
12.02.2020

Acknowledgement

I would like to thank Professor Bauch for being my supervisor and his team at Hamburg University of Technology (TUHH). Beside the scientific research, Professor Bauch is a great person and mentor who values excellent teaching. You can't wish more from a supervisor. Furthermore, being at TUHH a couple of times each year always felt a bit like family, especially at the Christmas events.

While the thesis was done at TUHH the daily work was carried out at the Fraunhofer Heinrich Hertz Institute (HHI) in Berlin. Therefore, I would like to thank Lars Thiele, who happens to be not only my friend but also my manager at HHI, for the support and most important the alignment of my project work with the thesis topic. Besides, I would like to thank the System Level Innovations Group and other colleagues who supported me during the writing of the thesis. Also the three months offered by HHI to focus on thesis writing have been very helpful to push through the final stage of the thesis and my gratitude goes to the responsible persons at HHI.

I am also very grateful to Professor Weber from University Rostock for examining my thesis.

Last but not least I would like to thank my best friend, partner, wife and love of my live, Barbara for her seemingly endless patience and support. I'm not sure who is more happy that "our" long Phd-journey has finally come to a successful end, she or I. Finally, I dedicate this thesis to my first-born child, my daughter Helen Greta, who enriched my life more than I can possibly express in words.

Berlin, February 2020

Summary

This thesis studies the application of centralized large antenna arrays in cellular networks, widely called massive multiple-input multiple-output (MIMO). The first part of the thesis is focused on spectral efficiency improvements in downlink transmission. Namely, in frequency division duplex downlink, two factors limit the number of antennas at the base station, the pilot overhead in the downlink, and second, the amount of feedback that can be sent in the uplink. One solution that balances between loss due to pilot overhead and gain from spatial multiplexing under a feedback constraint, is hybrid precoding. Here, hybrid precoding refers to the combination of codebook-based first-stage precoding and effective channel-aware second-stage precoding. The thesis, in Chapter 2, addresses this challenge by proposing a novel codebook design called "sub-codebook splitting" - developed for the first-stage precoding. With the proposed solution, any number of base station antennas can be used while providing a design parameter that adjusts the pilot overhead. Furthermore, the sub-codebook splitting technique provides the same degree of channel quantization as a large codebook, but with lower pilot overhead. Numerical simulations demonstrate, that the proposed hybrid precoding technique achieves a sum spectral efficiency close to full digital minimum mean square error precoding in the regime of spatial multiplexing with much fewer users than antennas. There, it is also shown that explicit channel state information (CSI) feedback is essential for multi-user downlink transmission, where the phase and amplitude of complex channel coefficients is reported. This is in contrast to implicit CSI in state-of-the-art systems, where only codebook indices are reported. Existing feedback reduction techniques, also in Chapter 2, demonstrate, that for a feedback constraint as in new radio (NR) Release-15, explicit CSI in combination with the proposed scheme, achieves a larger sum spectral efficiency than the codebook-based precoding, even without the feedback constraint for the latter. Furthermore, the thesis in Section 4.1, demonstrates that the lack of multi-cell interference knowledge at the user results in a loss of sum spectral efficiency. Thus, a low overhead feedback is proposed, consisting of the wideband interference power measured at the mobile users. With the proposed feedback up to 100 % sum spectral efficiency gain with scheduling technique and up to 400 % gain for random user selection are achieved.

In the second part of this thesis, direction of arrival (DoA) estimation in the uplink for search-based techniques is studied. Search based DoA estimation schemes require a power spectrum computation for each searched angular direction. Consequently, the challenge to achieve a low DoA estimation error, which requires a high quantization of the angular search space, is a high computational complexity. To address this challenge, in Chapter 3, a new scheme on adaptive search space quantization (ASSQ) is proposed. The principle of ASSQ is that the search space is quantized with a low resolution in a first step and based on the maximum DoA found in the first step, a smaller search space is quantized with a higher resolution in the next step. This can be repeated several times and the thesis specifically focuses on the challenge of the derivation of ASSQ parameters. The ASSQ scheme significantly reduces the required complexity compared to a brute-force search with hardly any performance loss, e.g. by a factor of 9300 for a scan of 120° in the azimuth- and elevation-plane. Furthermore, user grouping for the purpose of joint multi-user DoA estimation is investigated in order to reduce the amount of resources required for the uplink positioning pilots. Namely, it is shown that the difference in received signal power between users is the major source of DoA estimation errors in joint detection. Therefore, a novel two-step algorithm is developed, where the first step divides users in power-groups with equivalent received power. Subsequently, the second step selects users based on minimum angular distance criteria. The input parameters of the grouping algorithm allow either scaling the DoA estimation error or the number of jointly estimated users. There, it is shown that the DoA estimation error distribution can be approximated by a t-location-scale probability distribution. This simplifies future algorithm development by the direct emulation of estimation errors. Finally, the calculation of 3D user positions, requires DoA estimates of the same user from two different base stations and their corresponding positions. The thesis addresses this by deriving the details of the minimum-distance-between-skew-lines technique in Section 4.2. The evaluation shows that the geometry dilution of precision (GDoP) effect causes large position errors, even for low DoA estimation errors. Hence, two compensation techniques that perform averaging over narrowband subarriers are proposed - one is averaging of DoAs, and the other over positions. Both methods are compatible with the current NR Release-15 and are capable to reduce most of the GDoP-based position errors.

Contents

Notations xi

Symbols xiii

1. Introduction 1
1.1. Massive MIMO in Cellular Communications . . . 3
1.2. Localization in Cellular Systems . . . 4
1.2.1. Geometrical Approaches . . . 5
1.2.2. Non-Geometrical Approaches . . . 10
1.2.3. Summary . . . 10
1.3. Thesis Contributions and Organization . . . 11

2. Downlink Data Transmission 15
2.1. Radio Channel - A Geometry-Based Stochastic Model . . . 15
2.2. Multiple-User Multiple-Cell System Model . . . 18
2.3. Full Channel State Information in Time Division Duplex (TDD) . . . 22
2.3.1. Linear Precoding Schemes . . . 23
2.3.2. Power Constraints . . . 25
2.3.3. Numerical Evaluation of Massive MIMO Effects . . . 26
2.3.4. User Grouping . . . 33
2.3.5. Channel Hardening . . . 38
2.3.6. Conclusions . . . 39
2.4. Limited Channel State Information in Frequency Division Duplex (FDD) . . . 40
2.4.1. Codebook Design . . . 47
2.4.2. Sub-Codebook Precoding - Single User . . . 56
2.4.3. Sub-Codebook Precoding - Multiple Users . . . 62
2.4.4. Hybrid Precoding . . . 64
2.4.5. Feedback - Sum Spectral Efficiency - Trade-Off . . . 65
2.4.6. Conclusions . . . 74

3. Uplink Search-Based Direction-of-Arrival (DoA) Estimation 77
3.1. Multiple-Source System Model . . . 77
3.1.1. Receive Signal . . . 77
3.1.2. DoA Estimation Techniques . . . 83
3.1.3. Performance Metrics . . . 86
3.2. Parameter Analysis and Complexity Reduction . . . 88
3.2.1. Impact of Parameters . . . 88
3.2.2. Complexity Reduction . . . 98
3.2.3. Conclusions . . . 111
3.3. Multiple-Source Estimation . . . 112
3.3.1. Challenges . . . 113
3.3.2. Grouping Algorithm . . . 119
3.3.3. Conclusions . . . 132

4. Challenges in Cellular Deployments **135**
4.1. Downlink Inter-Cell Interference in Multiple-User MIMO . . . 136
4.1.1. Multiple-Sector Scenarios . . . 136
4.1.2. Interference Compensation . . . 138
4.1.3. Conclusions . . . 142
4.2. Uplink Positioning based on Two DoA Estimates . . . 143
4.2.1. Skew Lines in Euclidean Space . . . 144
4.2.2. Compensation of Dilution of Precision (DoP) Errors . . . 150
4.2.3. Conclusions . . . 158

5. Conclusions **159**

Appendix A. Rearrangements of Equations **165**

Appendix B. Additional Material: Algorithms, Figures, Tables **174**

Bibliography **193**

Acronyms

2D	two dimensional.
2G	second generation.
3D	three dimensional.
3G	third generation.
3GPP	3rd Generation Partnership Project.
4G	fourth generation.
5G	fifth generation.
AoA	angle of arrival.
ASSQ	adaptive search space quantization.
AWGN	additive white Gaussian noise.
BD	block diagonalization.
BS	base station.
CDF	cumulative distribution function.
CoMP	coordinated multi-point.
CQI	channel quality indicator.
CRB	Cramer Rao bound.
CRLB	Cramer Rao lower bound.
CSI	channel state information.
DFT	discrete Fourier transform.
DoA	direction of arrival.
DoF	degree of freedom.
DoP	dilution of precision.
FDD	frequency division duplex.
GDoP	geometry dilution of precision.
GNSS	global navigation satellite system.
GPS	global positioning system.
GSM	global system for mobile communications.
HPBW	half-power beam-width.
i.i.d.	independent identically distributed.
IEEE	institute of electrical and electronics engineers.
ISD	inter-site distance.
KPI	key performance indicator.
LoS	line of sight.
LTE	Long Term Evolution.
LTE-A	Long Term Evolution (LTE)-Advanced.
MCS	modulation and coding scheme.
MIMO	multiple-input multiple-output.
MISO	multiple-input single-output.
MMSE	minimum mean square error.
MRT	maximum ratio transmission.
MS	mobile station.
MSLR	minimum side lobe ratio.
MUSIC	multiple signal classification.
NLoS	non line of sight.

NR	new radio.
OFDM	orthogonal frequency division multiplexing.
PAPC	per antenna power constraint.
PAPR	peak-to-average power ratio.
PBZF	projection based zero forcing.
PHY	physical.
PMI	precoding matrix indicator.
QAM	quadrature amplitude modulation.
QuaDRiGa	QUAsi Deterministic RadIo channel GenerAtor.
RADAR	radio detection and ranging.
RB	resource block.
RE	resource element.
RSRP	reference signal receive power.
RSS	receive signal strength.
SINR	signal to interference and noise ratio.
SLNR	signal to leakage and noise ratio.
SNR	signal to noise ratio.
SUS	semi-orthogonal user selection.
SVD	singular value decomposition.
TDD	time division duplex.
TDoA	time difference of arrival.
ToA	time of arrival.
ToF	time of flight.
UE	user equipment.
ULA	uniform linear array.
UMTS	universal mobile telecommunications system.
UPA	uniform planar array.
WLAN	wireless local area network.
ZF	zero forcing.

Notations

Notation	Description
$\Re(z)$	Real part of complex number z.
$\Im(z)$	Imaginary part of complex number z.
$\arg(z)$	Angle between real and imaginary part of complex number z, also called argument of z.
a, A	Italic lower and upper case letters denote scalars.
$\mathbf{a}$	Boldface lower case letters denote vectors.
$\mathbf{A}$	Boldface upper case letters denote matrices.
$\mathcal{A}$	Calligraphic upper case letters denote sets.
$\mathbf{a}_i$	Italic lower case letters as subscripts indicate dependency on the subscript variable.
$[a_1 \ \dots \ a_i]$	Vector, of size $1 \times i$, where $i \in \mathbb{N}_+$ and $i > 1$.
$\begin{bmatrix} a_1 \\ \vdots \\ a_i \end{bmatrix}$	Vector, of size $i \times 1$, where $i \in \mathbb{N}_+$ and $i > 1$.
$\{1, \dots, i\}$	Set that contains the elements in the interval $[1, i] = \{x \in \mathbb{N} \mid 1 \leq x \leq i\}$.
$[\mathbf{A}\ \mathbf{B}]$	Column concatenation of matrix $\mathbf{A}$ of size $a \times b$ and matrix $\mathbf{B}$ of size $a \times c$, where $a, b, c \in \mathbb{N}_+$, result in a matrix of size $a \times b + c$.
$[a, b]$	Interval from a to b, including a and b.
$\mathbf{I}_N$	Identity matrix of size $N \times N$.
$\mathbf{A}^{(i)}$	Italic lower case letters as superscripts in () brackets indicate dependency on the superscript variable. This is for clear differentiation from x^2 which means x to the power of 2.
$\mathbf{A}^{(\mathrm{ZF})}, A_{(\mathrm{ZF})}$	Upright letter subscripts or superscripts in () brackets are used to distinguish different variables.
$[\mathbf{a}]_i$	The i-th element of vector $\mathbf{a}$.
$[\mathbf{A}]_{i,j}$	The element of row i and column j of matrix $\mathbf{A}$.
$[\mathbf{A}]_{:,j}$	All elements of column j of matrix $\mathbf{A}$.
$[\mathbf{A}]_{i,:}$	All elements of row i of matrix $\mathbf{A}$.
$\mathbf{A} \odot \mathbf{B}$	Hadamard product of matrix $\mathbf{A}$ and $\mathbf{B}$. Also called Schur product or entry-wise product.
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of matrix $\mathbf{A}$ and $\mathbf{B}$.
$\mathbf{a} \times \mathbf{b}$	Cross or vector product of vector $\mathbf{a}$ and $\mathbf{b}$.
$\mathbf{a} \cdot \mathbf{b}$	Scalar product of vector $\mathbf{a}$ and $\mathbf{b}$.
$(\cdot)^{\mathrm{T}}$	Transpose of vector or matrix.
$(\cdot)^*$	Conjugate complex of vector or matrix.
$(\cdot)^{\mathrm{H}}$	Conjugate complex and transpose of vector or matrix.

Notation	Description
$(\cdot)^{-1}$	Inverse of a matrix.
$(\cdot)^{\dagger}$	Pseudo inverse of matrix.
$\left\|\mathbf{a}\right\|, \left\|\mathbf{A}\right\|, \left\|\mathcal{A}\right\|$	Number of elements in $\mathbf{a}$, $\mathbf{A}$, or set $\mathcal{A}$, respectively.
$\left\|\cdot\right\|$	Euclidean norm.
$\left\|\mathbf{A}\right\|_{\mathrm{F}}$	Frobenius norm of matrix $\mathbf{A}$.
$\mathbb{E}\left(\cdot\right)$	Expectation value.
$\mathbb{N}, \mathbb{R}, \mathbb{C}$	Set of natural, real, and complex numbers, respectively.
$\mathbb{R}^3$	Three dimensional Euclidean space.
$\mathbf{e}^{(\mathrm{x})}, \mathbf{e}^{(\mathrm{y})}, \mathbf{e}^{(\mathrm{z})}$	Orthonormal unit vectors that span the three dimensional Euclidean space.
$\mathbf{x} \sim \mathcal{CN}(a, \mathbf{B})$	$\mathbf{x}$ is a circularly symmetric complex Gaussian vector with mean a and covariance matrix $\mathbf{B}$.
dy/dx	The derivative of y with respect to x.
$\partial y/\partial x$	Partial derivative of y with respect to x.
$\max\left(\cdot\right)$	Maximum value of the elements in set, vector or matrix.
$\min\left(\cdot\right)$	Minimum value of the elements in set, vector or matrix.
$\mathrm{trace}\left(\mathbf{A}\right)$	Trace of matrix $\mathbf{A}$, $\mathrm{trace}\left(\mathbf{A}\right) = \sum_{a=1}^{\left\|\mathbf{A}\right\|} \left[\mathbf{A}\right]_{a,a}$.
$\mathrm{vec}(\mathbf{A})$	Vectorization of matrix $\mathbf{A}$ of size $M \times N$ into vector $\mathbf{a}$ of size $MN \times 1$, $\mathrm{vec}(\mathbf{A}) = \mathbf{a}$ such that elements $\left[\mathbf{A}\right]_{:,n}$ of $\mathbf{A}$ correspond to elements $\left[\left(n-1\right)M+1, \ldots, nM\right]$ of $\mathbf{a}$.
$\overset{A}{=}$	Rearrangement of equation with several steps in between, where details are provided in "Appendix A: Rearrangements of Equations" with headline "Equation N", where N corresponds to the equation number.
$f\left(a_1, \ldots, a_i\right)$	Function depending on the $i \in \mathbb{N}_+$ variables $a_1, \ldots, a_i$.
$\mathbf{J}_i(\mathbf{a}(i))$	Jacobian matrix of vector $\mathbf{a}(i)$ with respect to i.
$\mathrm{floor}\left(\mathbf{A}\right)$	Rounds down each element of $\mathbf{A}$ to the nearest integers towards minus infinity.
$\mathrm{mod}\left(a, b\right)$	Modulo operator, where $\mathrm{mod}\left(a, b\right)$ returns the remainder after division of a by b, where a is the dividend and b is the divisor and $a, b \in \mathbb{R}$. $\mathrm{mod}\left(a, b\right) = a - b \cdot n$ where $n = \mathrm{floor}\left(\frac{a}{b}\right)$ if $b \neq 0$. If $b = 0$ the modulo operator is defined as $\mathrm{mod}\left(a, 0\right) = a$.
$\Gamma\left(\cdot\right)$	Gamma function.

Symbols

Symbol	Description
α	True horizontal angle of direction of arrival ϕ.
α_k	True horizontal angle of source or beam k.
$\boldsymbol{\alpha}^{(\mathrm{f})}$	Horizontal search range for search based direction of arrival methods.
$\boldsymbol{\alpha}_s^{(\mathrm{f})}$	Horizontal search range for search based direction of arrival methods in adaptive search space quantization step s.
$\alpha_{i,j}^{(\mathrm{joint})}$	Horizontal angle between sources i and j.
$\alpha_{(\mathrm{min})}^{(\mathrm{f})}$	Absolute minimum of horizontal search range over all adaptive search space quantization steps.
$\alpha_{(\mathrm{max})}^{(\mathrm{f})}$	Absolute maximum of horizontal search range over all adaptive search space quantization steps.
$\hat{\alpha}$	Estimated horizontal angle.
$\hat{\alpha}_i$	Estimated horizontal angle of the i-th source.
$\hat{\alpha}_{i,j}$	Estimated horizontal angle of the i-th source on the j-th subcarrier.
$\hat{\boldsymbol{\alpha}}_i^{(N_{(\mathrm{s})})}$	Vector notation of estimated horizontal angles of source i of $N_{(\mathrm{s})}$ subcarriers.
$\hat{\alpha}_{(\mathrm{HPBW})}$	Half-power beam-width of estimated α in power spectrum.
$\hat{\alpha}_{(\mathrm{right})}$	Right hand side (larger) angle of estimated azimuth half-power beam-width.
$\hat{\alpha}_{(\mathrm{left})}$	Left hand side (smaller) angle of estimated azimuth half-power beam-width.
$\bar{\alpha}_k$	Estimated horizontal angle of source k transformed to reference coordinate system.
$\breve{\alpha}_i^{(N_{(\mathrm{s})})}$	Horizontal angle averaged over $N_{(\mathrm{s})}$ estimated angles of source i.
$\tilde{\alpha}_i$	Horizontal orientation angle of the i-th base station antenna array.
$\alpha_{i,j}^{(\mathrm{ULA})}$	Horizontal angle of a steered beam by phase controlled antenna elements i and j, where $i \neq j$.
$\alpha_{i,j}^{(\mathrm{ULA,DFT})}$	Horizontal angle of a steered beam by phase controlled antenna elements i and j with phase difference obtained from elements i and j in a discrete Fourier transform codeword.
$\tilde{\alpha}_{i,j}^{(\mathrm{ULA,DFT})}$	Horizontal angle of a steered beam of an uniform linear array with a discrete Fourier transform codeword and inputs to the inverse sine function mapped to the interval $[-1, 1]$ of stream i with j codewords.
$\boldsymbol{\alpha}^{(\mathrm{ULA})}$	Horizontal beam directions of a discrete Fourier transform codebook in vector notation.
$\boldsymbol{\alpha}^{(\mathrm{CW})}$	Vector containing lower and upper horizontal angle for codeword directions in discrete Fourier transform codebook.
$\alpha^{(\mathrm{CW,min})}$	Lower horizontal angle for codeword directions in discrete Fourier transform codebook.
$\alpha^{(\mathrm{CW,max})}$	Upper horizontal angle for codeword directions in discrete Fourier transform codebook.

Symbol	Description
β	True vertical angle of direction of arrival ϕ.
β_k	True vertical angle of source or beam k.
$\boldsymbol{\beta}^{(\mathrm{f})}$	Vertical search range for search based direction of arrival methods.
$\boldsymbol{\beta}_s^{(\mathrm{f})}$	Vertical search range for search based direction of arrival methods in adaptive search space quantization step s.
$\beta_{i,j}^{(\mathrm{joint})}$	Vertical angle between sources i and j.
$\beta_{(\mathrm{min})}^{(\mathrm{f})}$	Absolute minimum of vertical search range over all adaptive search space quantization steps.
$\beta_{(\mathrm{max})}^{(\mathrm{f})}$	Absolute maximum of vertical search range over all adaptive search space quantization steps.
$\hat{\beta}$	Estimated vertical angle of direction of arrival ϕ.
$\hat{\beta}_i$	Estimated vertical angle of the i-th source.
$\hat{\beta}_{i,j}$	Estimated vertical angle of the i-th source on the j-th subcarrier.
$\hat{\boldsymbol{\beta}}_i^{(N_{(\mathrm{s})})}$	Vector notation of estimated vertical angles of source i of $N_{(\mathrm{s})}$ subcarriers.
$\hat{\beta}_{(\mathrm{HPBW})}$	Half-power beam-width of estimated β in power spectrum.
$\hat{\beta}_{(\mathrm{right})}$	Right hand side (larger) angle of estimated elevation half-power beam-width.
$\hat{\beta}_{(\mathrm{left})}$	Left hand side (smaller) angle of estimated elevation half-power beam-width.
$\bar{\beta}_k$	Estimated vertical angle of source k transformed to reference coordinate system.
$\check{\beta}_i^{(N_{(\mathrm{s})})}$	Vertical angle averaged over $N_{(\mathrm{s})}$ estimated angles of source i.
$\tilde{\beta}_i$	Vertical orientation angle of the i-th base station antenna array.
$\boldsymbol{\beta}^{(\mathrm{ULA})}$	Vertical beam directions of a discrete Fourier transform codebook in vector notation.
$\boldsymbol{\beta}^{(\mathrm{CW})}$	Vector containing lower and upper vertical angle for codeword directions in discrete Fourier transform codebook.
$\beta^{(\mathrm{CW,min})}$	Lower vertical angle for codeword directions in discrete Fourier transform codebook.
$\beta^{(\mathrm{CW,max})}$	Upper vertical angle for codeword directions in discrete Fourier transform codebook.
$\gamma_{k,t}$	Signal to interference and noise ratio of stream t at receiver k.
$\tilde{\gamma}$	Signal to noise ratio.
$\tilde{\gamma}_{k,l}$	Uplink receive signal to noise ratio at base station l from source k.
$\gamma^{(\mathrm{E})}$	Receive signal energy-to-noise ratio.
$\hat{\gamma}_{\tilde{t}}$	Estimated signal to noise ratio of stream $\tilde{t}$.
$\gamma^{(\mathrm{SLNR})}$	Signal to leakage and noise ratio.
$\gamma^{(\mathrm{min})}$	Minimum signal to interference and noise ratio. Values below $\gamma^{(\mathrm{min})}$ are assumed with $\gamma \to 0$.
$\gamma^{(\mathrm{max})}$	Maximum signal to interference and noise ratio. Values above $\gamma^{(\mathrm{max})}$ are set to $\gamma^{(\mathrm{max})}$.
$\gamma^{(\mathrm{CPR})}$	Compression ratio due to multiple-source grouping.
$\delta_{l_{(\mathrm{MPC})}}$	Dirac impulse corresponding to the $l_{(\mathrm{MPC})}$-th multi-path component.
$\boldsymbol{\Delta}_k$	Diagonal matrix with Dirac impulses of the $L_{(\mathrm{MPC})}$ multi-path component of source k.
$\Delta_{(\mathrm{xyz})}$	Position error in three dimensional space.
$\Delta_{(\mathrm{xyz})}^{(i)}$	Position error for mobile user i.
$\Delta_{(\mathrm{x})}, \Delta_{(\mathrm{y})}, \Delta_{(\mathrm{z})}$	$x^{(\mathrm{C})}$, $y^{(\mathrm{C})}$, and $z^{(\mathrm{C})}$ component of position error.

Symbol	Description
$\Delta_{(\mathrm{xy})}$	$x^{(\mathrm{C})} - y^{(\mathrm{C})}$ component of position error.
$\Delta_{(\mathrm{xz})}$	$x^{(\mathrm{C})} - z^{(\mathrm{C})}$component of position error.
$\Delta_{(\mathrm{yz})}$	$y^{(\mathrm{C})} - z^{(\mathrm{C})}$ component of position error.
$\Delta_{(\mathrm{x})}^{(\mathrm{opt})}$	$x^{(\mathrm{C})}$ component of position error with optimal weighting factor $g_{(\mathrm{p})}^{(0)}$.
$\Delta_{(\mathrm{y})}^{(\mathrm{opt})}$	$y^{(\mathrm{C})}$ component of position error with optimal weighting factor $g_{(\mathrm{p})}^{(0)}$.
$\Delta_{(\mathrm{z})}^{(\mathrm{opt})}$	$z^{(\mathrm{C})}$ component of position error with optimal weighting factor $g_{(\mathrm{p})}^{(0)}$.
$\Delta_{(\alpha)}$	Horizontal direction of arrival estimation error.
$\Delta_{(\alpha)}^{(i,j)}$	Horizontal direction of arrival estimation error at base station i for mobile user j.
$\Delta_{(\beta)}$	Vertical direction of arrival estimation error.
$\Delta_{(\mathrm{ToF})}$	Time of flight measurement.
$\Delta_{(\mathrm{RTT})}$	Round trip time of flight measurement.
$\Delta\omega_{i,j}$	Phase difference between antenna elements i and j in a phased antenna array, where $i \neq j$.
$\mathbf{\Delta}_{(i)}^{(\Phi)}$	Direction of arrival estimation error in Cartesian coordinates at base station i.
$\tilde{\mathbf{\Delta}}$	Distance vector in Cartesian coordinates from base station 1 to base station 2.
$\tilde{\Delta}_{(\mathrm{x})}, \tilde{\Delta}_{(\mathrm{y})}, \tilde{\Delta}_{(\mathrm{z})}$	$x^{(\mathrm{C})}$, $y^{(\mathrm{C})}$, and $z^{(\mathrm{C})}$ component of distance vector $\tilde{\mathbf{\Delta}}$, respectively.
ϵ	Arbitrary threshold defined at each occasion.
$\epsilon_{(\mathrm{SUS})}$	Orthogonality threshold for semi-orthogonal user selection.
$\epsilon_{(\mathrm{SUS})}^{(\max)}$	Orthogonality threshold that achieves the maximum sum spectral efficiency with the semi-orthogonal user selection algorithm for a given number of available streams.
$\epsilon_{(\alpha)}$	Threshold for mapping estimated to true horizontal angles.
$\epsilon_{(\beta)}$	Threshold for mapping estimated to true vertical angles.
η	Index of sets in the same power group.
$\vartheta_{k,t}$	Intra-sector interference at user k from the serving base station of user k caused by other streams than stream t.
λ, λ_n	Solution parameters in multiple signal classification method.
$\lambda_{(\mathrm{min})}$	Minimum solution parameter in multiple signal classification method.
$\lambda_{(\mathrm{c})}$	Signal wavelength according to frequency $f_{(\mathrm{c})}$.
$\mu_{k,n}$	Phase difference between antenna element n and a reference element of source k.
$\mu^{(\mathrm{a})}, \mu^{(\mathrm{b})}, \mu^{(\mathrm{c})}, \mu^{(\mathrm{tmp})}$	Variables aggregating part of an equation for the sake of readability used to obtain the joint angle between two sources.
$\mu^{(\mathrm{t})}$	Location parameter of t-location-scale distribution.
$\nu^{(\mathrm{t})}$	Shape parameter of t-location-scale distribution.
ξ	Index of source set for joint direction of arrival estimation.
π	Mathematical constant pi, $\pi \approx 3.14159265358979323846264 33$.
$\Pi(I)$	Stream I that maximizes the orthogonality of selected streams in semi-orthogonal user selection algorithm.
σ_k	Noise variance at user k.
$\sigma_l^{(\mathrm{BS})}$	Noise variance at base station l.
$\sigma^{(\mathrm{RSS})}$	Offset parameter in [dB] for mapping distance to receive power with receive signal strength method.

Symbol	Description
$\bar{\boldsymbol{\Sigma}}_{k,l}$	Rectangular diagonal matrix of singular values from singular value decomposition of channel matrix between base station l and user k.
$\sigma_{(\phi)}$	Variance of normal distribution for approximation of direction of arrival estimation error.
$\sigma^{(\mathrm{CRB})}$	Cramer Rao lower bound on the variance of the unbiased direction of arrival estimator.
$\sigma^{(\mathrm{t})}$	Scale parameter of t-location-scale distribution.
τ_l	Delay or time of flight of the l-th path.
$\tau^{(\mathrm{MS})}$	Processing delay at the mobile station.
$\tau^{(\mathrm{C})}$	Number of symbols in a coherence resource.
$\tau_{i,j}^{(\mathrm{N})}$	Time difference of arrival between antenna element i and j.
ϕ	True direction of arrival consisting of horizontal angle α and vertical angle β.
ϕ'	Direction searched with direction of arrival estimation techniques.
$\boldsymbol{\phi}_{(\mathrm{ULA})}^{(\mathrm{DFT})}$	Direction vector of discrete Fourier transform beams from an uniform linear array.
$\boldsymbol{\phi}_{(\mathrm{UPA})}^{(\mathrm{DFT})}$	Direction matrix of discrete Fourier transform beams from an uniform planar array.
$\tilde{\boldsymbol{\phi}}_{(\mathrm{UPA})}^{(\mathrm{DFT})}$	Direction matrix of discrete Fourier transform beams from an uniform planar array with notation such that a row corresponds to the same vertical/elevation angle and a column to the same horizontal/azimuth angle.
$\boldsymbol{\Phi}_{(i)}$	True direction of arrival in Cartesian coordinate parametrized by horizontal angle α and vertical angle β from the i-th base station to the mobile device.
$\Phi_{(i)}^{(\mathrm{x})}, \Phi_{(i)}^{(\mathrm{y})}, \Phi_{(i)}^{(\mathrm{z})}$	x, y, z component of true direction of arrival vector $\boldsymbol{\Phi}_{(i)}$, respectively.
$\hat{\phi}$	Estimated direction of arrival consisting of horizontal angle α and vertical angle β.
$\hat{\phi}^{(2)}$	Estimated direction of arrival of largest side lobe or second largest peak in power spectrum.
$\phi_{i,j}^{(\mathrm{joint})}$	Angle spanned by sources i and j in three dimensional Euclidean space from base station perspective.
$\boldsymbol{\Phi}_{\mathcal{K}}^{(\mathrm{joint})}$	Joint angles of all sources in $\mathcal{K}$.
$\phi_{(\mathrm{min})}$	Minimum angular distance between sources in the same group.
$\hat{\phi}_{(\mathrm{HPBW})}$	Joint estimated half-power beam-width angle.
$\phi_k^{(\mathrm{false})}$	Boolean value per source if detected sources in multiple source direction of arrival estimation is outside of a threshold around the input direction of arrival.
$\boldsymbol{\Psi}_{\tilde{t}}$	Subspace projector of stream $\tilde{t}$ for projection based zero forcing in Algorithm 1.
$\boldsymbol{\Psi}_{\mathcal{T}}$	Subspace projector of stream set $\mathcal{T}$ for projection based zero forcing in Algorithm 1.
$\boldsymbol{\omega}_i^{(\mathrm{DFT})}$	i-th codeword from discrete Fourier transform matrix of size $N \times 1$.
$\boldsymbol{\Omega}$	Codebook of size $N \times N^{(\boldsymbol{\Omega})}$.
$\boldsymbol{\Omega}_{(\mathrm{ULA})}^{(\mathrm{DFT})}$	Discrete Fourier transform codebook for uniform linear array.
$\boldsymbol{\Omega}_{(\mathrm{UPA})}^{(\mathrm{DFT})}$	Kronecker discrete Fourier transform codebook for uniform planar array.
$\tilde{\boldsymbol{\Omega}}_{n_{(\mathrm{SCB})}}$	$n_{(\mathrm{SCB})}$-th sub codebook of size $N \times N^{(\tilde{\boldsymbol{\Omega}})}$.
$\omega_{i,j}^{(\mathrm{N})}$	Phase difference of incoming signal between antenna elements i and j.
$\mathbf{a}_k, l_{(\mathrm{MPC})}$	Uplink direction of arrival steering vector of source k of the $l_{(\mathrm{MPC})}$-th multi-path component.

Symbol	Description
$\mathbf{A}_k$	Uplink direction of arrival steering matrix of source k of the $L_{(\mathrm{MPC})}$ multi-path components.
$\mathbf{A}_{(\mathrm{f})}$	Horizontal search space of the S adaptive search space quantization steps, where $\mathbf{A}_{\mathrm{f}} \in \mathbb{R}_{>0}^{2\times S}$.
$a_{(\mathrm{h})}$	Auxiliary variable to improve readability introduced in Section 4.2.1 in Eq. (4.17).
$\tilde{a}_{(\mathrm{h})}$	Auxiliary variable to improve readability introduced in Eq. (A.25).
B	Used signal bandwidth.
$B^{(\mathrm{C})}$	Coherence bandwidth of channel.
b_l	Number of streams in first stage precoder at base station l.
$\mathbf{B}_l$	Downlink first stage precoding matrix of base station l.
$\mathbf{b}_{(\mathrm{BS})}(\phi)$	Uplink receive beamformer at base station.
$\mathbf{b}_{(\mathrm{BS})}(\alpha)$	Uplink receive beamformer of horizontal uniform linear array.
$\mathbf{b}_{(\mathrm{BS})}(\beta)$	Uplink receive beamformer of vertical uniform linear array.
$\mathbf{b}_{(\mathrm{BS})}^{(\mathrm{Bartlett})}$	Uplink receive beamformer at base station with Bartlett method.
$\mathbf{b}_{(\mathrm{BS})}^{(\mathrm{Capon})}$	Uplink receive beamformer at base station with Capon method.
$\mathbf{b}_{(\mathrm{BS})}^{(\mathrm{MUSIC})}$	Uplink receive beamformer at base station with multiple signal classification method.
$\mathbf{B}_{(\mathrm{f})}$	Vertical search space of the S adaptive search space quantization steps, where $\mathbf{B}_{\mathrm{f}} \in \mathbb{R}_{>0}^{2\times S}$.
$\mathbf{B}_{(\mathrm{CRB})}$	Array geometry dependent part of Cramer Rao lower bound.
$b_{(\mathrm{h})}$	Auxiliary variable to improve readability introduced in Section 4.2.1 in Eq. (4.17).
$C_{k,t}$	Spectral efficiency in [bit/s/Hz] at user k on stream t.
$C_l^{(\mathrm{sum})}$	Sum spectral efficiency in [bit/s/Hz] of base station l.
$\tilde{C}_{k,t}$	Bounded spectral efficiency in [bit/s/Hz] at user k on stream t.
$\tilde{C}^{(\mathrm{sum})}$	Bounded sum spectral efficiency in [bit/s/Hz] in case of single base station scenario.
$\tilde{C}_l^{(\mathrm{sum})}$	Bounded sum spectral efficiency in [bit/s/Hz] of base station l.
$\tilde{C}_{k,t}^{(\mathrm{O})}$	Effective bounded spectral efficiency in [bit/s/Hz] considering pilot signaling overhead at user k on stream t.
$\tilde{C}_l^{(\mathrm{sum,O})}$	Effective bounded sum spectral efficiency in [bit/s/Hz] considering pilot signaling overhead of base station l.
$\hat{C}_t$	Estimated stream spectral efficiency based on subspace projection of stream t in Algorithm 1.
$\hat{C}_t^{(\mathrm{sum})}$	Estimated sum spectral efficiency based on subspace projection of stream t in Algorithm 1.
$\hat{C}_t^{(\mathrm{sum,old})}$	Estimated sum spectral efficiency reference in Algorithm 1 without stream t.
$\hat{C}_t^{(\mathrm{sum,new})}$	Estimated sum spectral efficiency in Algorithm 1 with new stream t.
$c_{(\mathrm{h})}$	Auxiliary variable to improve readability introduced in Section 4.2.1 in Eq. (4.17).
d_i	Distance from coordinate system origin to the i-th mobile user or source.
$d^{(\mathrm{BS})}, d_k^{(\mathrm{BS})}$	Distance between a base station and a source or the k-th source if subscript is used.
$d^{(\mathrm{RSS})}$	Reference distance for receive signal strength method.
$d_{i,j}^{(\mathrm{N})}$	Distance between antenna elements i and j where $i \neq j$.
$d_{(\mathrm{spread})}^{(\mathrm{row-norm})}$	Spread of row norms of a matrix.
$d_{(\mathrm{h})}$	Auxiliary variable to improve readability introduced in Section 4.2.1 in Eq. (4.19).

Symbol	Description
$\mathbf{e}^{(\mathbf{R}_{(\mathbf{yy})})}h_i$	i-th eigenvector of $\mathbf{R}_{(\mathbf{yy})}$.
$e_{(\mathrm{h})}$	Auxiliary variable to improve readability introduced in Section 4.2.1 in Eq. (4.19).
$\mathbf{E}_{(\mathrm{n})}$	Noise subspace matrix in multiple signal classification direction of arrival estimation.
$\mathbf{F}_l$	Downlink diagonal power allocation matrix at base station l.
$f_{(\phi)}$	Scaling factor for minimum angular distance $\phi_{\min}$.
$f_{(\mathrm{c})}$	Center frequency of transmitted signal.
$f^{(\mathrm{s})}$	Rounding function as part of counting the number of power spectrum computations required for direction of arrival estimation.
$f_{(\mathrm{h})}$	Auxiliary variable to improve readability introduced in Section 4.2.1 introduced in Eq. (4.19).
$f_{(1)}$	Auxiliary variable to improve readability introduced in Section 4.2.1 introduced in Eq. (A.25).
$f_{(2)}$	Auxiliary variable to improve readability introduced in Section 4.2.1 introduced in Eq. (A.25).
$G^{(\mathrm{r})}$	Antenna gain at receiver.
$G^{(\mathrm{t})}$	Antenna gain at transmitter.
$G^{(\mathrm{Path})}$	Path gain that includes all transmitter and receiver gains, transmit power and system loss.
$L^{(\mathrm{sys})}$	System loss.
$\mathbf{g}_{\tilde{t}}^{(\mathrm{SUS})}$	Orthogonality metric between available stream $\tilde{t}$ and already scheduled streams in semi-orthogonal user selection algorithm.
$g_{(\mathrm{p})}$	Weighting factor that determines the position obtained with the "shortest distance between skew lines" method.
$g_{(\mathrm{p})}^{(0)}$	The optimal weighting factor $g_{(\mathrm{p})}$ that minimizes the position error with the "shortest distance between skew lines" method.
$g^{(\mathrm{norm})}$	Binary value that indicates channel normalization. 1 means normalized and 0 means not normalized.
$\mathbf{G}_{(\mathrm{CRB})}$	Geometry dependent part of Cramer Rao bound.
$G_{(\mathrm{DoP})}$	Value that gives a measure about the impact from dilution of precision.
$\mathbf{H}_{k,l}$	Frequency domain downlink channel matrix, between base station l and user k.
$\underline{\mathbf{H}}_{\mathcal{T}_l}$	Compound channel matrix that consists of streams $\mathcal{T}_l$.
$\mathbf{H}_{\mathcal{K}_l \backslash k}^{(\mathrm{SLNR})}$	Signal to leakage and noise ratio compound channel matrix of set $\mathcal{K}_l$ without k.
$\tilde{\mathbf{h}}_{k,t}$	Effective receive channel of stream t at user k without post-processing.
$\hat{\mathbf{H}}_{k,l}$	Effective channel at user k from base station l obtained with first-stage precoder by $\hat{\mathbf{H}}_{k,l} = \mathbf{H}_{k,l}\mathbf{B}_l$.
$\mathbf{h}_{m,k,l}^{(\mathrm{MISO})}$	m-th equivalent multiple-input single-output channel of the channel between base station l and user k.
$\mathbf{h}_{k,l}^{(\mathrm{T})}$	Time domain downlink channel vector between base station l and source k, matrix, respectively.
$\mathbf{H}^{(\mathrm{T})}$	Time domain downlink channel matrix.
$\underline{\mathbf{H}}_{\mathcal{K}_l}^{(\mathrm{T})}$	Time domain downlink compound channel matrix of source set $\mathcal{K}_l$ at base station l.

Symbol	Description
$\bar{\mathbf{H}}_k^{(\mathrm{T})}$	Time domain downlink multi-path channel matrix of a single antenna source k, where the first dimension is the number of multi-path components and the second dimension is the number of receive antennas.
$\hat{\mathbf{H}}_k^{(\mathrm{T})}$	Time domain downlink multi-path channel gain of a single antenna user, where the first dimension is the number of multi-path components and the second dimension is the number of receive antennas.
$\tilde{\mathbf{H}}_{\mathcal{K}_l}^{(\mathrm{T})}$	Time domain normalized channel matrix of source set $\mathcal{K}_l$ at base station l.
$\hat{\mathbf{h}}_{1,k}^{(\mathrm{TFS})}$	Deterministic channel vector assuming free space propagation of source k.
i, I, i'	General iterators or counting variables used multiple times with value ranges defined per appearance.
$i^{(\max)}$	Codeword index that maximizes the receive signal power.
j	Index if used as subscript, otherwise imaginary unit.
$\mathbf{J}_{(\mathrm{CRB})}$	Jacobian matrix that computes the direction of arrival dependent part of the Cramer Rao bound.
$\mathcal{K}$	Set of users in the system.
$\mathcal{K}_l$	Set of users spatially multiplexed for multi-user transmission on the same time-frequency resource at base station l.
$\tilde{\mathcal{K}}_l$	Set of users connected to base station l that requests downlink transmission, also called available users.
$K_{(\mathrm{TDD})}^{(\max)}$	Maximum number of users that can be served in time division duplex for a given number of coherent resources with overhead below 50 %.
$k, \tilde{k}$	Mobile user index.
$K, K_l, \tilde{K}_l$	Number of users in sets $\mathcal{K}, \mathcal{K}_l$, and $\tilde{\mathcal{K}}_l$, respectively.
$K^{(\xi)}$	Number of groups for multiple source direction of arrival estimation as output from angular grouping without power grouping.
$\mathcal{K}_\xi^{(\xi)}$	ξ-th source subset for joint direction of arrival estimation without power grouping.
$K_\xi^{(\xi)}$	Number of devices in $\mathcal{K}_\xi^{(\xi)}$.
$K^{(\eta)}$	Number of power groups for multiple source direction of arrival estimation as output from power grouping.
$\mathcal{K}_\eta^{(\eta)}$	η-th subset of sources as output from power grouping.
$K^{(\xi\eta)}$	Number of groups for joint direction of arrival estimation with power grouping.
$\mathcal{K}_{\xi,\eta}^{(\xi\eta)}$	ξ-th subset of sources for joint direction of arrival estimation from the η-th power subset $\mathcal{K}_\eta^{(\eta)}$.
$K_\eta^{(\xi)}$	Number of subsets as output of the angular grouping for the η-th power group $\mathcal{K}_\eta^{(\eta)}$.
$\mathcal{K}^{(\mathrm{RxP})}$	Set of receive power values of K sources at the base station.
$K^{(\xi,\mathrm{random})}$	Number of random selected devices for joint direction of arrival estimation.
$K_{k,n}^{(\mathrm{R})}$	Rician K-factor of source k at base station antenna element n.
$k_{(\mathrm{x})}, k_{(\mathrm{y})}, k_{(\mathrm{z})}$	Auxiliary parameters for readability used to derived the optimal weighting factor $g_{(\mathrm{p})}^{(0)}$ introduced in Eq. (4.33).
l	Base station index.
$l_{(\mathrm{MPC})}$	multi-path component index.
L	Number of base stations in systems.
$\mathcal{L}$	Set of base stations indices.

Symbol	Description
$L_{(\mathrm{MPC})}$	Number of multi-path components.
$l_{(\mathrm{x})}, l_{(\mathrm{y})}, l_{(\mathrm{z})}$	Auxiliary parameters for readability used to derived the optimal weighting factor $g_{(\mathrm{p})}^{(0)}$ introduced in Eq. (4.33).
$L^{(\mathrm{sys})}$	System loss factor.
m	Interfering base station index.
M	Number of elements at mobile station antenna array.
$M_k^{(\mathrm{MISO})}$	Number reported equivalent multiple-input single-output channels of user k.
$\mathbf{m}_{(\mathrm{DoA})}$	Estimated direction of arrival vector $\mathbf{m}_{(\mathrm{BS})}$ of base station 1.
$m_{(\mathrm{x})}, m_{(\mathrm{y})}, m_{(\mathrm{z})}$	$x^{(\mathrm{C})}$, $y^{(\mathrm{C})}$, and $z^{(\mathrm{C})}$ component of estimated direction of arrival vector $\mathbf{m}_{(\mathrm{DoA})}$ of base station 1.
$n^{(\mathrm{RSS})}$	Scaling factor for distance based path gain computation with receive signal strength method.
N	Number of elements in base station array antenna.
n	Antenna index.
$\mathbf{n}_k$	Additive white Gaussian noise at mobile station k.
$\mathbf{n}_l^{(\mathrm{BS})}$	Additive white Gaussian noise at base station l.
$n_{(\mathrm{r,max})}$	Maximum row norm of precoding matrix $\mathbf{V}$.
$n_{(\mathrm{r,min})}$	Minimum row norm of precoding matrix $\mathbf{V}$.
$n_{(\mathrm{s})}$	Subcarrier index.
$N_{(\mathrm{s})}$	Number of orthogonal frequency division multiplexing subcarriers.
$N^{(\mathrm{RB})}$	Number of resource blocks in frequency domain.
$N_{(\mathrm{subband})}^{(\mathrm{PBZF})}$	Subband size given in resource blocks where the same set of selected users is assigned.
$\tilde{N}^{(\mathrm{RB})}$	Number of reported resource blocks in frequency domain.
$N^{(\mathrm{bit})}$	Number of bits used to quantize the channel, such that a codebook can contain up to $2^{N^{(\mathrm{bit})}}$ codewords.
$N^{(*)} = N - K$	Number of antenna minus number of sources in uplink direction of arrival estimation.
$\tilde{N}$	Number of elements in the non-massive MIMO base station array antenna in the heterogeneous multiple sector scenario.
$N_{(\mathrm{FDD})}^{(\mathrm{max})}$	Maximum number of base station antennas in a frequency division duplex system with uplink feedback considerations.
$\tilde{N}_{(\mathrm{FDD})}^{(\mathrm{max})}$	Maximum number of base station antennas in a frequency division duplex system without uplink feedback considerations.
$N^{(\boldsymbol{\Omega})}$	Number of codewords in codebook $\boldsymbol{\Omega}$.
$N_{(\alpha)}^{(\boldsymbol{\Omega})}$	Number of horizontal codewords in Kronecker codebook.
$N_{(\beta)}^{(\boldsymbol{\Omega})}$	Number of vertical codewords in Kronecker codebook.
$N_{(\boldsymbol{\alpha}^{(\mathrm{CW})})}^{(\boldsymbol{\Omega})}$	Number of codewords in angular range $\boldsymbol{\alpha}^{(\mathrm{CW})}$.
$N^{(\tilde{\boldsymbol{\Omega}})}$	Number of codewords in sub codebooks.
$\tilde{N}^{(\tilde{\boldsymbol{\Omega}})}$	Number of reported streams/codewords from a sub codebook.
$N_{(\alpha)}^{(\boldsymbol{\Omega},\mathrm{SCB})}$	Number of horizontal codewords in sub codebooks.
$N_{(\beta)}^{(\boldsymbol{\Omega},\mathrm{SCB})}$	Number of vertical codewords in sub codebooks.
$N^{(\mathrm{SCB})}$	Number of sub codebooks.

Symbol	Description
$n_{(\mathrm{SCB})}$	Counting index of sub codebooks.
$\mathbf{n}^{(\mathrm{SCB})}_{n_{(\mathrm{SCB})}}$	Vector of indices that selects the codeword entries of sub codebook $n_{(\mathrm{SCB})}$ from the large codebook.
$N^{(\mathrm{CW})}_{(\mathrm{FB})}$	Number of codewords fed back from the mobile user to the base station.
$N_{(\alpha)}, N_{(\alpha)_l}$	Number of horizontal antennas or columns in the uniform planar array of base station l or without index there is a single base station.
$\mathbf{N}_{(\alpha)}$	Vector containing the number of horizontal antennas selected in each step of adaptive search space quantization.
$\tilde{N}_{(\alpha)}$	Number of columns of the non-massive MIMO uniform planar array in the heterogeneous multiple sector scenario.
$N_{(\beta)}, N_{(\beta)_l}$	Number of vertical antennas or rows in the uniform planar array of base station l or without index there is a single base station.
$\mathbf{N}_{(\beta)}$	Vector containing the number of vertical antennas selected in each step of adaptive search space quantization.
$\tilde{N}_{(\beta)}$	Number of rows of the non-massive MIMO uniform planar array in the heterogeneous multiple sector scenario.
$N^{(\phi)}$	Number of angles searched with search based direction of arrival methods.
$\mathbf{n}^{(\phi,\mathrm{S})}$	Vector that contains the number of angles searched with search based direction of arrival methods at each step.
$N^{(\phi)}_s$	Number of angles searched in adaptive search space quantization step s.
$N^{(\phi)}_{(\alpha)}$	Number of horizontal angles for search based direction of arrival methods.
$N^{(\phi)}_{(\beta)}$	Number of vertical angles for search based direction of arrival methods.
$N^{(\mathrm{seq})}$	Number of discrete time samples or sequence length.
$N^{(\mathrm{false})}$	Number of "false" detected sources in multiple source direction of arrival estimation, meaning that estimated direction of arrival is outside of a threshold around the input direction of arrival.
$\mathbf{n}_{(\mathrm{DoA})}$	Estimated direction of arrival vector $\mathbf{n}_{(\mathrm{BS})}$ of base station 2.
$n_{(\mathrm{x})}, n_{(\mathrm{y})}, n_{(\mathrm{z})}$	$x^{(\mathrm{C})}$, $y^{(\mathrm{C})}$, and $z^{(\mathrm{C})}$ component of estimated direction of arrival vector $\mathbf{n}_{(\mathrm{DoA})}$ of base station 2.
$\mathcal{O}$	Number of combinations for stream selection.
$O^{(\mathrm{CB})}$	Overhead scaling factor due to pilots in the data channel per coherence block.
$\mathbf{p}_{t,l}$	Second stage precoding vector of stream t at base station l.
$\mathbf{P}_l$	Second stage precoding matrix of base station l.
$P^{(\mathrm{RB})}$	Total transmit power per resource block.
$P_{(\mathrm{PAPC})}$	Per antenna power constraint, that is the maximum transmit power per antenna.
$\mathbf{P}_{\mathcal{T}_l}$	Second stage precoder for stream set $\mathcal{T}_l$.
$\mathbf{P}^{(\mathrm{MMSE})}_{\mathcal{T}_l}$	Minimum mean square error second stage precoder for stream set $\mathcal{T}_l$.
$\mathbf{p}^{(\mathrm{MRT})}_t, \mathbf{P}^{(\mathrm{MRT})}_{\mathcal{T}_l}$	Maximum ratio transmission second stage precoder for stream t and stream set $\mathcal{T}_l$, respectively.
$\tilde{\mathbf{p}}^{(\mathrm{SLNR})}_t$	Signal to leakage and noise ratio second stage precoder for stream t.
$\mathbf{P}^{(\mathrm{SLNR})}_{\mathcal{T}_l}$	Signal to leakage and noise ratio second stage precoder for stream set $\mathcal{T}_l$.
$\mathbf{P}^{(\mathrm{ZF})}_{\mathcal{T}_l}$	Zero forcing second stage precoder for stream set $\mathcal{T}_l$.

Symbol	Description
$\mathbf{P}_{k}^{(\mathrm{MPC})}$	Path gain matrix of multi-path components of source k.
$\mathbf{P}_{(\mathrm{FB})}^{(\mathrm{eff})}$	Vector of descend ordered power values of the effective channel.
$p^{(\mathrm{O})}$	Projection order, parameter in projection based zero forcing user grouping algorithm.
$P_{k,n}^{(\mathrm{MPC,NLoS})}$	Sum power of non line of sight paths between the transmit antenna of user k and the n-th receive antenna at the base station.
$P_{k,n}^{(\mathrm{MPC,LoS})}$	Power of line of sight path between the transmit antenna of user k and the n-th receive antenna at the base station.
$P_{l,k}^{(\mathrm{r,s})}$	Receive signal power at receiver l from transmitter k.
$P_{l}^{(\mathrm{t})}$	Transmit power of transmitter l.
$P_{k,l}^{(\mathrm{s})}$	Receive signal power at base station l from user k.
$P_{l}^{(\mathrm{n})}$	Receiver noise power at base station l.
$\tilde{P}^{(\mathrm{Rx})}$	Instantaneous receive power per symbol.
$P^{(\mathrm{Rx})}, P_{i}^{(\mathrm{Rx})}$	Uplink receiver power spectrum at base station. The method used to obtain the power spectrum is indicated by subscript $i \in \{(\mathrm{Bartlett}), (\mathrm{Capon}), (\mathrm{MUSIC})\}$.
$P_{(\mathrm{Bartlett})}^{(\mathrm{Rx})}$	Uplink receive power spectrum at base station with Bartlett method.
$P_{(\mathrm{Capon})}^{(\mathrm{Rx})}$	Uplink receive power spectrum at base station with Capon method.
$P_{(\mathrm{MUSIC})}^{(\mathrm{Rx})}$	Uplink receive power spectrum at base station with multiple signal classification method.
$P^{(\mathrm{MSLR})}$	Power ratio between main lobe and largest side lobe.
$P^{(\epsilon)}$	Input parameter that defines the power range of groups in Algorithm 3 for power grouping.
$\tilde{\mathbf{p}}^{(\mathrm{BS})}, \tilde{\mathbf{p}}_{i}^{(\mathrm{BS})}$	Position of the base station in case of a single base station scenario (subscript omitted), and position of the i-th base station, respectively.
$\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}$	Position of virtual point required to obtain the joint angle between device i and j.
$\tilde{\mathbf{p}}_{i}^{(\mathrm{N})}$	Position of i-th antenna element in $\mathbb{R}^3$ in Cartesian coordinates.
$P_{(\mathrm{CRB})}$	Parametric dependent part of Cramer Rao bound.
$\mathbf{p}^{(\mathrm{array})}$	Geometric center of the base station antenna array in Cartesian coordinates.
$P^{(\mathrm{min})}$	Minimum receive power value of $\mathcal{K}^{(\mathrm{RxP})}$.
$P^{(\mathrm{max})}$	Maximum receive power value of $\mathcal{K}^{(\mathrm{RxP})}$.
$\mathbf{p}^{(\mathrm{thr})}$	Power threshold values for power grouping.
$\mathcal{P}^{(\mathrm{t})}$	Probability distribution function of t-location scale distribution.
$P_{k,l}^{(\mathrm{RSRP})}$	Unprecoded reference signal receive power at user k from base station l.
$q^{(\phi)}$	Quantization of search space in case of linear array and without adaptive search space quantization.
$q_{s}^{(\phi)}$	Quantization of search space in case of linear array in adaptive search space quantization step s.
$q_{s}^{(\alpha)}$	Horizontal quantization of search space in adaptive search space quantization step s.
$q_{s}^{(\beta)}$	Vertical quantization of search space in adaptive search space quantization step s.
$\mathbf{q}^{(\alpha)}$	Vector of horizontal quantization values in the $\mathcal{S}$ steps.
$\mathbf{q}^{(\beta)}$	Vector of vertical quantization values in the $\mathcal{S}$ steps.

Symbol	Description
$q^{(\mathrm{A})}$	Number of bits for amplitude quantization of channel state information resulting to $2^{q^{(\mathrm{A})}}$ quantization levels.
$q^{(\mathrm{A,R})}$	Amplitude quantization range.
$q^{(\mathrm{P})}$	Number of bits for phase quantization resulting to $2^{q^{(\mathrm{P})}}$ quantization levels.
$q^{(\mathrm{CSI})}$	Number of bits for channel state information quantization as sum of $q^{(\mathrm{A})}$ and $q^{(\mathrm{P})}$.
$q^{(\mathrm{CW})}$	Number of bits required for a codeword/stream index.
$\mathbf{R}_{k,l}$	Downlink transmit covariance matrix between base station l and user k.
$R_{k,t}$	Link capacity or rate in [bit/s] at user k on stream t.
$R_l^{(\mathrm{sum})}$	Sum capacity or sum rate in [bit/s] of base station l.
$\tilde{R}_{k,t}$	Bounded link rate in [bit/s] at user k on stream t.
$\tilde{R}_l^{(\mathrm{sum})}$	Bounded sum spectral efficiency in [bit/s] of base station l.
$\tilde{R}_{k,t}^{(\mathrm{O})}$	Effective link capacity in [bit/s] considering pilot signaling overhead at user k on stream t.
$\tilde{R}_l^{(\mathrm{sum,O})}$	Effective sum capacity in [bit/s] considering pilot signaling overhead of base station l.
$\mathbf{R}^{(\mathrm{MMSE})}$	Regularization matrix for minimum mean square error precoder.
$R_{k,t}^{(\mathrm{FB})}$	Feedback in bits per feedback interval of user k for codeword t.
$R_{\tilde{\mathcal{K}}_l}^{(\mathrm{FB,sum})}$	Sum feedback rate per feedback interval of all users connected to base station l.
$R^{(\mathrm{FB,HP,1})}$	Per user feedback in bits for hybrid precoding assuming full reporting.
$R^{(\mathrm{FB,HP,2})}$	Per user feedback in bits for hybrid precoding reporting a subset of codewords/stream.
$R^{(\mathrm{FB,HP,RB})}$	Per user feedback in bits for hybrid precoding reporting a subset of resource blocks.
$R^{(\mathrm{false})}$	False detection ratio with joint direction of arrival estimation of multiple sources.
$R_{(\mathrm{random})}^{(\mathrm{false})}$	False detection ratio with joint direction of arrival estimation of multiple sources indicating "random" source grouping.
$R_{(\mathrm{angular})}^{(\mathrm{false})}$	False detection ratio with joint direction of arrival estimation of multiple sources indicating "angular" source grouping.
$\mathbf{R}_{(\mathbf{yy})}$	Uplink receive covariance matrix at base station.
$\mathbf{R}_{(\mathbf{yy})}^{(0)}$	Null space of $\mathbf{R}_{(\mathbf{yy})}$ assuming that the number of users is less than the number of receive antennas at the base station.
$\mathbf{R}_{(\mathbf{xx})}$	Positive definite matrix that contains the pair-wise correlation between receive signals used in multiple signal classification method.
s	Current adaptive search space quantization step.
S	Number of adaptive search space quantization steps.
$\mathcal{S}$	Set of adaptive search space quantization steps.
$\mathbf{s}^{(\mathrm{BS})}$	Vector that aggregates the position and estimated direction of arrival of base station 1 in Cartesian coordinates.
$s^{(\mathrm{BS})}$	Auxiliary parameter that scales the distance along the direction of arrival estimate of base station 1.
$\tilde{\mathbf{s}}^{(\mathrm{BS})}$	Point on direction of arrival of base station 1 that is the shortest distance to direction of arrival of base station 2.
$s_{(\mathrm{PAPC})}$	Scaling factor for per antenna power constraint.
t	Stream index in Chapter 2 and time sample index in Chapter 3.

Symbol	Description
t'	Stream index with the condition that $t' \neq t$.
T_l	Number of multiplexed downlink spatial layers/streams on the same time-frequency resource at base station l.
$\mathcal{T}_l$	Set of multiplexed spatial layers/streams at base station l.
$\tilde{t}$	Available multiple-input single-output channel/stream index.
$\tilde{T}_l$	Number of available spatial layers/streams at base station l.
$\tilde{\mathcal{T}}_l$	Set of available spatial layers/streams at base station l.
$T^{(\max)}$	Maximum number of streams that can be selected with linear precoding.
$\tilde{\mathcal{T}}^{(I)}$	Set of streams that fulfills the orthogonality condition in semi-orthogonal user selection algorithm.
$\check{t}$	Stream index with minimum orthogonality to already selected streams for projection based zero forcing in Algorithm 1.
t_i	Time at discrete index i.
$t^{(\mathrm{obs})}$	Observation time corresponding to sequence of length $N^{(\mathrm{seq})}$.
$T^{(\mathrm{C})}$	Coherence time of the channel.
$T_{(\mathrm{s})}$	Signal duration.
$T^{(\mathrm{FI})}$	Feedback interval from users to base station in frequency division duplex system.
$t^{(\mathrm{sym})}$	Symbol duration.
$\mathbf{t}^{(\mathrm{BS})}$	Vector that aggregates the position and estimated direction of arrival of base station 2 in Cartesian coordinates.
$t^{(\mathrm{BS})}$	Auxiliary parameter that scales the distance along the direction of arrival estimate of base station 2.
$\tilde{\mathbf{t}}^{(\mathrm{BS})}$	Point on direction of arrival of base station 2 that is the shortest distance to direction of arrival of base station 1.
$\bar{\mathbf{U}}_{k,l}$	Left hand side eigenvectors of singular value decomposition of channel matrix $\mathbf{H}_{k,l}$.
$\mathbf{u}_{(\phi)}$	Direction in $\mathbb{R}^3$ as unit vector in Cartesian coordinates parametrized by spherical coordinates.
$\mathbf{u}, \mathbf{u}_i$	True source or mobile device position general or referring to the i-th mobile device with subscript i.
$u_{(\mathrm{x})}, u_{(\mathrm{y})}, u_{(\mathrm{z})}$	$x^{(\mathrm{C})}$, $y^{(\mathrm{C})}$, and $z^{(\mathrm{C})}$ component of true source position.
$\hat{\mathbf{u}}, \hat{\mathbf{u}}_i$	Estimated position of mobile device in general or referring to the i-th mobile device with subscript i.
$\hat{u}_{(\mathrm{x})}, \hat{u}_{(\mathrm{y})}, \hat{u}_{(\mathrm{z})}$	$x^{(\mathrm{C})}$, $y^{(\mathrm{C})}$, and $z^{(\mathrm{C})}$ component of estimated source position.
$\hat{\mathbf{u}}_{i,j}$	Estimated position of mobile device i on subcarrier or resource block j.
$\hat{\mathbf{U}}_i^{(N_{(\mathrm{s})})}$	Estimated position of source i averaged over $N_{(\mathrm{s})}$ subcarrier measurements.
$\check{U}_k^{(N_{(\mathrm{s})})}$	Averaged estimated position of source k over $N_{(\mathrm{s})}$ subcarriers.
$\mathbf{v}_t$	Downlink precoding vector of stream t.
$\mathbf{V}_l$	Downlink precoding matrix of base station l.
$\mathbf{V}_l^{(\mathrm{EPS})}$	Equal power per stream normalized downlink precoding matrix of base station l.
$\bar{\mathbf{V}}_{k,l}$	Right hand side eigenvectors of singular value decomposition of channel matrix $\mathbf{H}_{k,l}$.
$v_{(\mathrm{c})}$	Speed of light.
$\mathbf{v}^{(\mathrm{st})}$	Vector in Cartesian coordinates from direction of arrival $\mathbf{s}^{(\mathrm{BS})}$ of base station 1 to direction of arrival $\mathbf{t}^{(\mathrm{BS})}$ of base station 2.

Symbol	Description
$\mathbf{v}_{(\mathrm{n})}^{(\mathrm{st})}$	Normalized direction vector in Cartesian coordinates from direction of arrival $\mathbf{s}^{(\mathrm{BS})}$ of base station 1 to direction of arrival $\mathbf{t}^{(\mathrm{BS})}$ of base station 2.
$\tilde{\mathbf{v}}^{(\mathrm{st})}$	Segment in Cartesian coordinates that is the shortest distance between the direction of arrivals of base station 1 and two.
$v_{(\mathrm{h})}$	Auxiliary parameter used to derive the optimal weighting factor introduced in Eq. (A.13).
$w_{i',N_{(\alpha)}^{(\mathbf{\Omega})}}^{(-1,1)}$	Values of the inverse sine function mapped to the the interval $[-1,1]$ by Eq. (2.66) for codeword i' and $N_{(\alpha)}^{(\mathbf{\Omega})}$ codewords.
$\mathbf{w}_{k,t}$	Receive beamformer or post-coder at user k for stream t.
$w_{(\mathrm{h})}$	Auxiliary parameter used to derive the optimal weighting factor introduced in Eq. (A.15).
$\mathbf{x}_l$	Downlink vector of transmit symbols from base station l.
$\mathbf{x}_{\mathcal{K}_l}$	Uplink vector of transmit symbols of mobile users in set $\mathcal{K}_l$.
$\tilde{\mathbf{x}}_{\mathcal{K}_l}$	Effective uplink vector of transmit symbols of mobile users in set $\mathcal{K}_l$.
$x^{(\mathrm{C})}$	$x^{(\mathrm{C})}$ component of Cartesian coordinates.
$x_i^{(\mathrm{C})}$	$x^{(\mathrm{C})}$ component of Cartesian coordinates of base station i.
$x^{(\mathrm{t})}$	Input to t-location scale probability distribution function.
$y_k, \mathbf{y}_k$	Downlink receive signal scalar or vector at user k.
$\mathbf{y}_l$	Uplink receive signal vector at base station l.
$\tilde{y}_k, \tilde{\mathbf{y}}_k$	Downlink effective receive signal scalar or vector including receive beamforming at user k.
$\tilde{y}_l, \tilde{\mathbf{y}}_l$	Uplink effective receive signal scalar or vector including receive beamforming at base station l.
$y^{(\mathrm{C})}$	$y^{(\mathrm{C})}$ component of Cartesian coordinates.
$y_i^{(\mathrm{C})}$	$y^{(\mathrm{C})}$ component of Cartesian coordinates of base station i.
$\mathbf{z}_k$	Interference vector received at user k from other than the serving base station of user k.
$\tilde{\mathbf{Z}}_{k,t}$	Interference covariance plus noise matrix at user k with respect to stream t.
$z_k^{(\mathrm{IF},i)}$	Interference level of user k used in regularization of minimum mean square error precoder.
$z_k^{(\mathrm{IF,est})}$	Estimated inter-cell interference receive power at user k.
$z^{(\mathrm{C})}$	$z^{(\mathrm{C})}$ component of Cartesian coordinates.
$z_i^{(\mathrm{C})}$	$z^{(\mathrm{C})}$ component of Cartesian coordinates of base station i.

1. Introduction

The utilization of multiple antennas in wireless communications is subject to research and development since the 1920s [Jen16]. This covers, inter alia, multiple antennas at either the receiver or transmitter side, both sides deployed as localized or distributed arrays. Thereby, insights in the multiplexing-diversity trade-off are of major interest [TVZ04]. That means, to find out operation points where systems should switch from diversity to multiplexing, or vice versa, in order to optimize certain key performance indicators (KPIs). This switching is a soft transition and depends on the current transmission properties, e.g. distribution of the antennas, signal to noise ratio (SNR) at users and so on. With the development of military and commercial radio transmission systems, spectrum becomes regulated and fragmented for specific usage, inter alia television, satellite navigation systems, wireless local area network (WLAN) or cellular mobile communications [TS12]. With the exponential wireless traffic increase [Cis12, Cis17] the value and cost of spectrum also increases. This motivates the need towards more efficient spectrum usage. In the past, the main driver for this has been an increased spectrum reuse per area, called cell densification [BLM$^+$14]. The costs of further densification are the limiting factor, especially in large cities, where space for new base stations (BSs) is expensive [AZDG16]. Therefore, more and more effort is spend to increase the spectral efficiency by increasing the spatial reuse of used spectrum in the same area [JBR$^+$11]. The most promising candidate to achieve significant spatial multiplexing gains is multiple-user massive multiple-input multiple-output (MIMO)[LETM14].

Marzetta's paper titled "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas" published in 2010 [Mar10] is considered by many as the origin of massive MIMO. On the other hand, there are publication from before that also provide insights and results about the scaling of antennas [CRT01, ZT03, SH05, TJSZ07]. Therefore, it is safe to state, that massive MIMO, or the idea to use large scale antenna systems at the base station, is a logical and consequent continuation in the development of MIMO research. In [Mar10] the following assumptions are considered:

1. An unlimited number of BS antennas.
2. An orthogonal frequency division multiplexing (OFDM) time division duplex (TDD) system using channel reciprocity for channel state information (CSI) estimation on BS side (Channel sounding).
3. Time required for pilots is proportional to the number of terminals served and is independent of the number of base station antennas.
4. The number of mobile stations (MSs) is limited by the coherence time depending on their mobility.

With these assumptions the following advantages based on mathematically-exact conclusions are obtained:

1. Spectral efficiency per user is independent of the bandwidth.
2. Required transmitted energy per bit vanishes.
3. Effects from uncorrelated noise and fast fading vanish.
4. The number of MSs per cell and the throughput per MS are independent of the absolute size of the cell.

Note, that the above conclusions only hold under the impossible assumption of an unlimited number of antennas at the BS, however, they indicate advantages that can be obtained from large antenna arrays. For

example point 1 and 3 kick in already at realistic numbers as shown in this thesis. Point 2 is only true if power consumption of hardware is ignored. In point 4, the beamforming gain compensates the path-loss, but with every 3 dB additional path-loss a doubling of the antenna number is required so this statement is purely of academic nature. Furthermore, approximate conclusions claim that:

1. A larger coherence time would not affect user capacity, but increases proportionally the number of MSs that can be served.

2. The uplink performance is nearly identical to the downlink performance.

As a drawback, the paper mentioned that MS receive signal to interference and noise ratio (SINR) in massive MIMO is fundamentally limited by pilot contamination due to inter-cell interference. Though it is later shown in [AM12] that pilot contamination can be mitigated due to pilot precoding, this is based on knowledge of large-scale fading coefficients and BS cooperation. [BHS18] and [BHS17] confirm that pilot contamination is not a fundamental asymptotic limitation in massive MIMO, but they assume perfect knowledge of channel covariance matrices. Due to this ideal assumptions pilot contamination remains up-to-date a major SINR limiting factor in practical massive MIMO systems [YCQL18]. Nevertheless, these findings attracted a lot of attention on massive MIMO. After the initial research hype more and more practical limitations have been taken into account [Ngo15, BLM16] such that massive MIMO is now mature and considered a key component for fifth generation (5G) systems [Hua13, Doc14, Eri15].

Due to the vast usage of the term massive MIMO in literature the exact meaning of it is fuzzy and far from being a clearly defined terminology. In part of the literature massive MIMO is based on the large system assumptions that the number of BS antennas $N \to \infty$ [HCPR11, HtBD13, BSHD15, BHS17]. Other literature considers massive MIMO whenever the number of antennas is much larger than the number of spatial multiplexed streams [NAAC12, LHYS13, MAA$^+$13, KL14]. Furthermore, other literature uses the term massive MIMO more loosely and considers any number of antennas $N \geq 4$ as massive, i.e. $N = 4$ in [DKTH14], $N = 8$ in [KSHT15], $N = 32$ in [LSZ$^+$13, WTK$^+$16], or $N = 50$ in [YCG14]. Independent of the number of antennas is the assumption of a centralized or distributed antenna system [QWC14]. While parts of the literature also label distributed antenna systems as massive MIMO [KBT$^+$12, RL14, YGC14, HKT15] this thesis focuses on centralized arrays for practical reasons. To achieve beamforming gains over multiple antennas phase coherent joint transmission is required. While this is already a challenge in centralized arrays [GSPE$^+$14] it becomes harder to realize in distributed arrays. Distributed massive MIMO corresponds to network MIMO [KJTB15] or joint transmission coordinated multi-point (CoMP) [MZ10]. In this work, the term massive MIMO means a large number of antennas in a centralized array, that means 100 antennas in Chapter 2 and 64 antennas in Chapter 3 are assumed. Furthermore, the application of massive MIMO in mobile cellular systems from a system level perspective is considered focusing on the frequency range below 6 GHz. This limitation of the frequency range follows a decisions of the recent World Radio-communication Conferences (WRC), where no spectrum above 6 GHz was allocated for mobile broadband service [ITU15].

A more detailed introduction to massive MIMO in cellular communications is given in Section 1.1. While the term massive MIMO is mostly used in the context of data transmission and its improvement, multiple antennas are also used in localization systems for direction of arrival (DoA) estimation. It is known that more antennas improve DoA estimation accuracy [TF09]. Therefore, the second aspect covered by this thesis is the utilization of massive MIMO for DoA estimation and localization in cellular networks. Hence, Section 1.2 provides a general overview on localization techniques and why DoA estimation is the preferred technique for localization in the context of massive MIMO in cellular systems. Finally, Section 1.3 of this introduction gives details about the remaining organization and an overview on the contributions of this thesis.

1.1. Massive MIMO in Cellular Communications

5G is the next cellular mobile communications network and is called new radio (NR) by 3rd Generation Partnership Project (3GPP) standardization. The first technical specification of 5G is published as Release 15 with documents containing a "38" at the beginning, e.g. [3GP17b] for the general description of the physical (PHY) layer. Therein, OFDM is defined as the waveform for 5G and is considered throughout this thesis. In contrast to Long Term Evolution (LTE) as the fourth generation (4G) network, a flexible numerology is defined supporting the various use cases in 5G, e.g. resource blocks (RBs) can be shorter than 14 OFDM symbols and subcarrier spacing can be a multiple of 15 kHz. In 5G massive MIMO is considered as a key technology, using a finite but larger number of antennas at the base station compared to 4G [HH15, MMB+15, NGM15]. The usage of ever more antennas in cellular networks is a continuation that has already started in second generation (2G) or global system for mobile communications (GSM) networks. At that time the goal was to increase reliability by the selection of transmit antennas at the BS [Ake03] utilizing transmit diversity, e.g. from multiple directive antennas pointing in different directions. In the following third generation (3G) or universal mobile telecommunications system (UMTS) cellular network, single data stream beamforming over multiple antennas at the BS has been introduced to increase spatial diversity [YYU01]. In the evolution of 3G and 4G, multiple antennas at the BS are used for spatial multiplexing, e.g. transmitting multiple data streams over multiple beams at the same time-frequency resource [LCG+12]. Thereby, the number of antennas has been increased up to 64 in Release 14 [3GP17c]. Multiple antennas are not only exploited by BSs, also MSs utilize multiple antennas. By that, the use of spatial multiplexing has been extended, for example from open loop to closed loop beamforming [SM07] or to cooperative multi-point transmission [Zir15]. Furthermore, the use of multiple antennas is not limited to spatial multiplexing or beam steering. Also polarization multiplexing [JBTJ12] or combinations with non-orthogonal multiplexing are being researched [NYB+14].

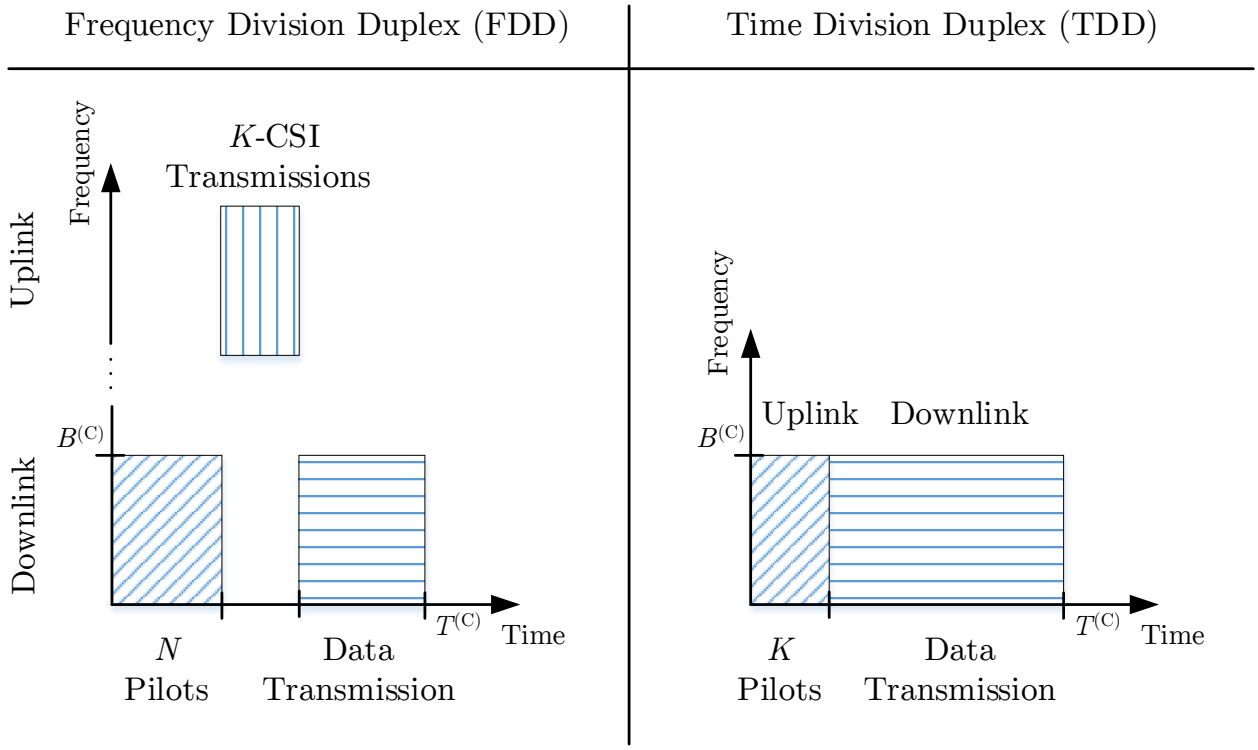


Figure 1.1.: Comparison of downlink CSI acquisition in a FDD and TDD system.

With the use of massive MIMO, the fundamental differences between frequency division duplex (FDD) and TDD become a key factor for deployment decisions [BLM16]. In FDD, the downlink and uplink are separated by operating at different frequencies; while in TDD downlink and uplink are separated by different time slots, see Fig. 1.1. For efficient transmit precoding at the BS, downlink channel knowledge is required at the BS, meaning the amplitude and phase of the complex channel coefficients[1], henceforth called CSI. In TDD the downlink CSI can be obtained from channel estimation in the uplink utilizing channel reciprocity [JAMV11], see Fig. 1.1. The advantage of massive MIMO in TDD is that, based on a single pilot transmission in the uplink, the channel to all transmit antennas can be estimated at once, thus only a single time-slot in the uplink is required per user for channel estimation at the BS. The channel knowledge at the BS can be used for spatial multiplexing either by receive post-coding in the uplink or transmit precoding in downlink. This is not possible in FDD systems due to the frequency separation of downlink and uplink resulting in non-reciprocity [HKL02]. Therefore, in FDD the downlink channel has to be estimated at the mobile users and then send back via a feedback channel. The first challenge of massive MIMO in FDD is the pilot overhead. In LTE reference signals also called "pilots", are integrated into the downlink data channel with one pilot sequence distributed in the time-frequency grid for each antenna port [3GP14]. Thus, scaling the number of pilots with the number of antennas will increase the pilot overhead and decrease the available resources for data downlink transmission. The second main challenge is the feedback that has to be sent in the capacity limited uplink channel from the mobile users to the BS. If the amount of feedback is scaling with the number of antennas at the BS, the resulting feedback overhead will become too large. The trade-off between the amount of feedback and system performance KPIs depends on specific system parameters. In [TKBH12, TKOB13] CSI feedback reduction for joint transmission CoMP is studied. [HKT15] shows that in case of correlated channels time-domain feedback can be reduced with compressed sensing. In [KJTB15] the CSI feedback is reduced by reporting only relevant multiple-path components depending on the SNR and channel estimation quality. In Section 2.4 both challenges, the pilot and feedback overhead, are addressed adapting hybrid precoding as the state of the art assumption in NR Release 15 [3GP17b]. Thereby, the focus is on a scalable solution that provide operators with the ability to adapt the cellular network according to varying mobile user demands and other KPIs.

1.2. Localization in Cellular Systems

In the recent century, knowledge of the position of moving objects henceforth called mobile users or short devices, became of great interest. A prominent example is the detection of airplanes and ships by radio detection and ranging (RADAR) during World War II. Also prominent is the navigation of cars with global navigation satellite systems (GNSSs). In cellular mobile communication systems, for example GSM, UMTS and LTE, various principles are used, but accuracy achieved is much lower in comparison with GNSS [Fis14].

Therefore, a huge interest exists to achieve higher accuracies within 5G mobile networks. In this thesis the aim is an accuracy below $\mathbf{1\,m}$ according to the requirement for next generation network (see Section 9.2 "Location/Positioning Service" in [3GP16d]), which can be used e.g. for guidance of devices.

The task of localization is to obtain a unique position of an object with respect to a reference system, e.g. global positioning system (GPS) coordinates. This unique position is given in three dimensional (3D) Euclidean space by Cartesian coordinates $\left[x^{(\mathrm{c})}\ y^{(\mathrm{c})}\ z^{(\mathrm{c})}\right]^{\mathrm{T}}$ according to Fig. 1.2, where an example for $[3\ 4\ 5]^{\mathrm{T}}$ m is given. The 3D Euclidean space is denoted as $\mathbb{R}^3$ and spanned by $\mathbf{e}^{(\mathrm{x})} = [1\ 0\ 0]^{\mathrm{T}}$, $\mathbf{e}^{(\mathrm{y})} = [0\ 1\ 0]^{\mathrm{T}}$, and $\mathbf{e}^{(\mathrm{z})} = [1\ 0\ 0]^{\mathrm{T}}$. The $x^{(\mathrm{c})}$-$y^{(\mathrm{c})}$ plane given in the legend of Fig. 1.2 is the plane spanned by $\mathbf{e}^{(\mathrm{x})}$ and $\mathbf{e}^{(\mathrm{y})}$ with $z^{(\mathrm{c})} = 0$. Accordingly, the $x^{(\mathrm{c})}$-$z^{(\mathrm{c})}$ plane is spanned by $\mathbf{e}^{(\mathrm{x})}$ and $\mathbf{e}^{(\mathrm{z})}$ with $y^{(\mathrm{c})} = 0$, and the $y^{(\mathrm{c})}$-$z^{(\mathrm{c})}$ plane is spanned by $\mathbf{e}^{(\mathrm{y})}$ and $\mathbf{e}^{(\mathrm{z})}$ with $x^{(\mathrm{c})} = 0$. The estimated position of a mobile device or source is

[1] There are also other channel representations that contain the required information. It is assumed that other channel representations can be transformed into complex coefficients in frequency domain, e.g. via discrete Fourier transform (DFT) from time domain representation, see Section 2.1.

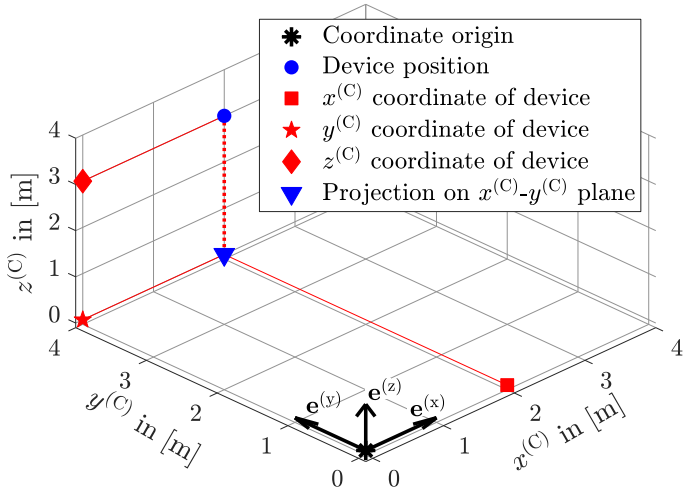


Figure 1.2.: Coordinate system in three dimensional Euclidean space $\mathbb{R}^3$ spanned by $\mathbf{e}^{(x)}, \mathbf{e}^{(y)}$, and $\mathbf{e}^{(z)}$.

based on some kind of measurements. In radio systems these measurements can be inter alia time of flight (ToF), time difference of arrival (TDoA), signal power, angle of arrival (AoA) or power delay profile. The measurements are obtained by sensors at a receiver which can be a BS or movile device.

Chapter 2 shows that large antenna systems at the BS increase the sum or system throughput due to spatial multiplexing, a solution also in frequencies below 6 GHz where spectrum is expensive. Nevertheless, once large antenna systems are deployed, they can also be used for DoA estimation of uplink signals from the user, in this chapter referred to as "sources". It is shown in Chapter 3, that the large number of antennas enables accurate DoA estimation meeting 5G requirements of 1 m position accuracy [3GP15b]. Previous cellular communications systems, as well as WiFi, only used one to four antennas, thus limiting the effectiveness of DoA estimation. This is also the reason, why in LTE the focus was on ToF based positioning approaches for localization. However, with large antenna arrays in 5G systems the focus is again on the well-studied field of DoA estimation. Therefore, one part of this thesis focuses on how to exploit the large number of antennas at the BS for radio based localization. Thereby, the challenging environment of a cellular mobile communication system is taken into account and existing algorithms are investigated with respect to robustness against impairments.

To complete this introduction the main localization techniques in wireless systems are briefly discussed. Thereby, localization techniques are divided into geometrical and non-geometrical approaches described in Section 1.2.1 and Section 1.2.2, respectively [VGL+15]. Note that although these techniques are described separately they can also be combined with each other for positioning[2].

1.2.1. Geometrical Approaches

For geometrical techniques, the position is obtained from one or more channel characteristics transformed into a geometric output. The unknown position of the mobile device is derived from equations taking into account the known position of the BSs.

Range measurements that suffer only from white noise are limited in accuracy by receive signal energy-to-noise ratio, $\gamma^{(E)}$ and the used bandwidth B. The Cramer Rao bound (CRB) or Cramer Rao lower bound (CRLB) provides a lower bound for the variance of range estimates with white noise [Cra45]. For example in an one-way ranging system that uses institute of electrical and electronics engineers (IEEE) 802.15.4

[2]The combination of different measurement methods is often referred to as "sensor fusion".

modulation [IEE11], the CRB is given as [MF09]

$$\sigma^{(\mathrm{CRB})} \geq \frac{v_{(\mathrm{c})}^2}{4\pi^2 B^2 \gamma^{(\mathrm{E})}}, \tag{1.1}$$

where the $\sigma^{(\mathrm{CRB})}$ denotes the noise variance, $v_{(\mathrm{c})}$ is the speed of light, B is the signal bandwidth given in [Hz], and $\gamma^{(\mathrm{E})}$ is the energy-to-noise ratio [Tre04]. The energy-to-noise ratio $\gamma^{(\mathrm{E})}$ is proportional to the SNR denoted as $\tilde{\gamma}$ such that

$$\gamma^{(\mathrm{E})} = T_{(\mathrm{s})} B \tilde{\gamma}, \tag{1.2}$$

where $T_{(\mathrm{s})}$ is the symbol duration. Most signals in commercial systems, e.g. LTE, 5G or WiFi, have a bandwidth symbol-time product of one such that $T_{(\mathrm{s})}B = 1$ and the energy-to-noise ratio becomes $\gamma^{(\mathrm{E})} \approx \tilde{\gamma}$. Since signal design is not in the focus of this work, signals with $T_{(\mathrm{s})}B \lesssim 1$ are not considered. From Eq. (1.1) it is seen that, given a constant signal energy and noise density, an increase in bandwidth reduces the estimation variance. This is one reason why wide signal bandwidths are used in ranging systems, e.g. in ultra-wideband [YS15]. If $\tilde{\gamma} \gg 1$ the CRB can be closely approached. Both, bandwidth and $\tilde{\gamma}$ play significant roles in determining noise-limited performance [Tre04]. In the following, the main principals of geometry based localization techniques are briefly discussed.

ToF Method

ToF or time of arrival (ToA) estimation is a basic principle of RADAR and uses the known speed of a radio wave by measurement of the the propagation time between a source and a destination. Without limitation to the generality, in Fig. 1.3a the source is the BS and the destination is the mobile device. Start time of the radio signal at the source is denoted by t_1 and the time of reception at the destination by t_2. Consequently, the ToF is given by

$$\Delta_{(\mathrm{ToF})} = t_2 - t_1. \tag{1.3}$$

From the ToF measurement, the distance can be obtained using knowledge of the velocity of radio waves. The velocity of radio waves is equal to the speed of light $v_{(\mathrm{c})} = 299\,792\,458\,\mathrm{m/s}$ and the distance $d^{(\mathrm{BS})}$ between a source and a destination (receive BS) can be obtained by

$$d^{(\mathrm{BS})} = v_{(\mathrm{c})} \Delta_{(\mathrm{ToF})}. \tag{1.4}$$

From Eq. (1.4) it becomes clear that uncertainties in the ToF measurement are multiplied by $v_{(\mathrm{c})}$ and result in large distance uncertainties. Uncertainties are caused, among other things, by the following effects.

1. There is limited resolution or precision of the local clock, as well as frequency errors caused by local oscillators.
2. Synchronization mismatch between source and destination, meaning that the exact t_1 has to be known at the destination or vice versa for t_2 at the source.
3. Other uncertainties are for example thermal noise or non line of sight propagation.

There are many techniques known in literature to reduce or mitigate these error sources, but they are out of the scope of this thesis and therefore omitted here[3].

TDoA Technique

One option to overcome the drawback of the tight synchronization requirement of ToF measurement is to measure the round trip time. In Fig. 1.3b a signal is sent from the BS to the device at t_1 and then another

[3] The interested reader finds further details in [Ben07].

(a) Direct ToF measurement

(b) Round trip ToF measurement

Figure 1.3.: ToF measurement principles.

signal is transmitted back, which is later received at t_2 at the BS. With this method the ToF is obtained by

$$\Delta_{(\mathrm{RTT})} = \frac{t_2 - t_1}{2} - \tau^{(\mathrm{MS})}, \tag{1.5}$$

where $\tau^{(\mathrm{MS})}$ is processing delay at the device. The subscript "RTT" stands for "round trip time". This processing delay is device specific and how to obtain it is an open challenge.

Another option to circumvent the requirement of the synchronization between transmitter and receiver is TDoA measurement, where the time difference of arrival of signals at BS pairs used. Thus, only BSs have to be synchronized to each other. This method utilizes trilateration to obtain the position and details can be found in [SD10]. The basic idea is captured by Fig. 1.4a, where $\tilde{\mathbf{p}}_1^{(\mathrm{BS})}$ and $\tilde{\mathbf{p}}_2^{(\mathrm{BS})}$ are the positions of the two BSs. Without loss of generality BS 1 and BS 2 are positioned on the $x^{(\mathrm{C})}$ axis by $\tilde{\mathbf{p}}_1^{(\mathrm{BS})} = \begin{bmatrix} -x_1^{(\mathrm{C})} & 0 & 0 \end{bmatrix}^{\mathrm{T}}$ and $\tilde{\mathbf{p}}_2^{(\mathrm{BS})} = \begin{bmatrix} x_1^{(\mathrm{C})} & 0 & 0 \end{bmatrix}^{\mathrm{T}}$. Thus, $2x_1^{(\mathrm{C})}$ is the distance between the two BSs. In Fig. 1.4a the device is represented by a square and the distance to the BSs are denoted by $d_1^{(\mathrm{BS})}$ and $d_2^{(\mathrm{BS})}$. The device can only detect the time difference between the arriving radio waves. Thus, the device only knows the distance difference given as a hyperbolic line [Ben07] by

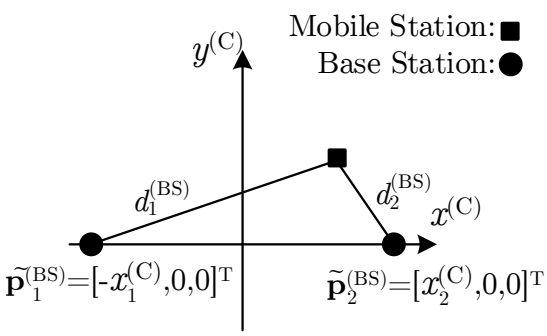
$$\left| d_1^{(\mathrm{BS})} - d_2^{(\mathrm{BS})} \right| = \left| \sqrt{\left(u_{(\mathrm{y})}\right)^2 + \left(u_{(\mathrm{x})} + x_1^{(\mathrm{C})}\right)^2} - \sqrt{\left(u_{(\mathrm{y})}\right)^2 + \left(u_{(\mathrm{x})} - x_1^{(\mathrm{C})}\right)^2} \right|, \tag{1.6}$$

where $u_{(\mathrm{x})}$ and $u_{(\mathrm{y})}$ denote the x and y coordinate of the device position, respectively. Considering a two dimensional (2D) space such that BSs and the devices are in the same plane, than three BSs are required to obtain a unique position with the TDoA method. Including also the uncertainty of the time measurement, the estimated device position is the intersection area of two hyperbolas as shown in Fig. 1.4b [Fis14].

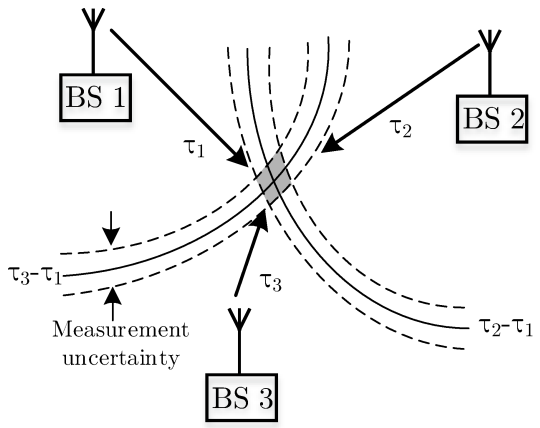
The same uncertainties mentioned in Section 1.2.1 for the ToA method remain and limit the achievable localization accuracy. In [Fis14] the observed TDoA method used in LTE is evaluated in detail, and therein even with optimistic assumptions only a standard deviation in the range of 50 m is achieved, failing the 5G target requirement of 1 m.

RSS Method

The receive signal strength (RSS) method utilizes the property of electromagnetic waves that they propagate with their power distributed on the surface of a sphere and the distance of a source can be estimated from



(a) TDoA principle with two BSs in 2D space.



(b) TDoA with three BSs in 2D space. The figure is adapted from Fig. 4.1 in [Fis14].

Figure 1.4.: Principle of TDoA based positioning.

the receive signal power. Assuming a free space propagation model, the receive signal power $P^{(\mathrm{r,s})}\left(d^{(\mathrm{BS})}\right)$ is a function of the distance $d^{(\mathrm{BS})}$ and can be represented by [Bal16]

$$P^{(\mathrm{r,s})}\left(d^{(\mathrm{BS})}\right) = \frac{P^{(\mathrm{t})} G^{(\mathrm{r})} G^{(\mathrm{t})} \lambda_{(\mathrm{c})}^2}{(4\pi)^2 \left(d^{(\mathrm{BS})}\right)^2 L^{(\mathrm{sys})}}, \tag{1.7}$$

where $P^{(\mathrm{t})}$ is the transmitted signal power, $G^{(\mathrm{r})}$ and $G^{(\mathrm{t})}$ are the antenna gains at the receiver and transmitter, respectively. $L^{(\mathrm{sys})}$ is the system loss, which equals 1 in free space, and $\lambda_{(\mathrm{c})}$ is the wavelength of the signal. Rearranging Eq. (1.7) and aggregating all gains and powers to the path gain $G^{(\mathrm{Path})}$ given in [dB] yields

$$G^{(\mathrm{Path})} = 10\log_{10}\left(\frac{P^{(\mathrm{r,s})} L^{(\mathrm{sys})}}{P^{(\mathrm{t})} G^{(\mathrm{r})} G^{(\mathrm{t})}}\right) = 10\log_{10}\left(\frac{\lambda_{(\mathrm{c})}^2}{(4\pi)^2 \left(d^{(\mathrm{BS})}\right)^2}\right) = 20\log_{10}\left(\frac{\lambda_{(\mathrm{c})}}{4\pi d^{(\mathrm{BS})}}\right). \tag{1.8}$$

Solving Eq. (1.8) the distance is obtained as a function of the path gain $G^{(\mathrm{Path})}$ according to

$$d^{(\mathrm{BS})} = \frac{\lambda_{(\mathrm{c})}}{4\pi} 10^{\left(-\frac{G^{(\mathrm{Path})}}{20}\right)}. \tag{1.9}$$

The accuracy of the RSS depends on the path gain $G^{(\mathrm{Path})}$. The path gain factor highly depends on the environment by the system loss and has to be approximated based on measurements, e.g. in [Ben07] by

$$G^{(\mathrm{Path})} = \begin{cases} -20\log_{10}\frac{4\pi d^{(\mathrm{RSS})}}{\lambda_{(\mathrm{c})}} - 20\log_{10}\frac{d^{(\mathrm{BS})}}{d^{(\mathrm{RSS})}} & \text{if } d^{(\mathrm{BS})} \leq d^{(\mathrm{RSS})} \\ -20\log_{10}\frac{4\pi d^{(\mathrm{RSS})}}{\lambda_{(\mathrm{c})}} - 10n^{(\mathrm{RSS})}\log_{10}\frac{d^{(\mathrm{BS})}}{d^{(\mathrm{RSS})}} + \sigma^{(\mathrm{RSS})} & \text{if } d^{(\mathrm{BS})} > d^{(\mathrm{RSS})} \end{cases} \tag{1.10}$$

where the reference distance $d^{(\mathrm{RSS})}$, $n^{(\mathrm{RSS})}$ and $\sigma^{(\mathrm{RSS})}$ are empirical obtained parameters given in Table 1.1.

The main advantage of RSS is that this measurement is available in most of the existing radio networks and therefore only little or no hardware modifications of existing devices are required. While multipath reflections are accounted as "noise" in other geometrical methods, they can improve the results for RSS when taken into account in the empirical obtained parameters. On the other hand RSS is generally less accurate than ToA or AoA [VGL+15].

Table 1.1.: Environment specific parameters required for path gain computation $G^{(\text{Path})}$ in Eq. (1.10), according to [Ben07].

Environment	Frequency $f_{(\text{c})}$ in [MHz]	$n^{(\text{RSS})}$	$\sigma^{(\text{RSS})}$ in [dB]
Retail store	914	2.2	8.7
Office, hard partition	1500	3.0	7.0
Office, soft partition	900	2.4	9.6
Factory, line of sight	1900	2.6	14.1

Direction of Arrival Technique

Direction of arrival (DoA), in literature often referred to as angle of arrival (AoA), is a widely studied method [Cap69, BK83, Sch86, KV96] and still in the focus of recent research [CPT+13, TI13, ASP14, BLG+15, KTH+16b]. DoA estimation is a method for determining the direction of propagation of a radio-frequency wave incident on an antenna array. DoA estimation determines the direction by utilizing the time difference of arrival at individual elements of the array by impinging wave fronts. These time differences correspond to phase differences, from which the the DoA can be calculated [Bal16].

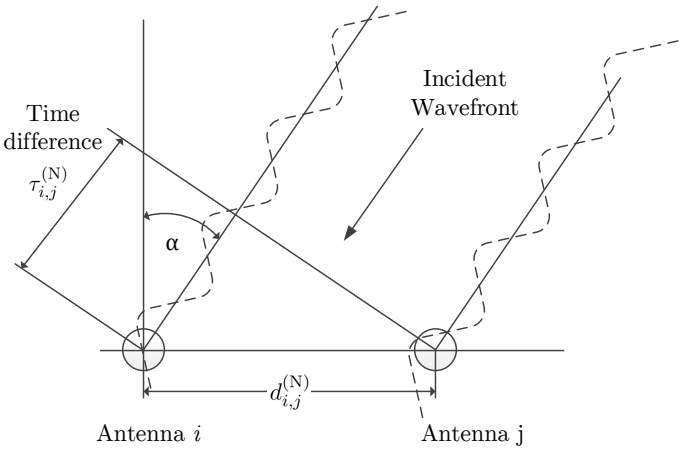


Figure 1.5.: Angle or direction of arrival measurement based on phase difference at multiple antennas.

In Fig. 1.5 the principle of DoA measurement is depicted. There are multiple antennas at the receiver, which are assumed to be in the far field of the transmitter. Far field means, that a part of a radiating spherical wave can be approximated by a planar wave. Based on this relaxation, the time difference $\tau_{i,j}^{(\text{N})}$ between antenna elements i and j can be determined by

$$\tau_{i,j}^{(\text{N})} = \frac{d_{i,j}^{(\text{N})} \sin \alpha}{v_{(\text{c})}}, \tag{1.11}$$

where $v_{(\text{c})}$ is the speed of light and α the azimuth AoA according to Fig. 1.5. In practice, instead of the time difference $\tau_{i,j}^{(\text{N})}$, the corresponding phase difference $\omega_{i,j}^{(\text{N})}$ is measured according to

$$\omega_{i,j}^{(\text{N})} = 2\pi \frac{d_{i,j}^{(\text{N})} \sin \alpha}{\lambda_{(\text{c})}}, \tag{1.12}$$

where $\lambda_{(\text{c})}$ is the length of the radio-wave. Usually the distance between antenna elements is known at the receiver. With a given antenna element spacing, e.g. $d_{i,j}^{(\text{N})} = \lambda_{(\text{c})}/2$ Eq. (1.12), the angle α of the incident

wave front depends only on the phase difference $\omega_{i,j}^{(\mathrm{N})}$ by

$$\alpha = \arcsin\left(\frac{\omega_{i,j}^{(\mathrm{N})}}{\pi}\right). \tag{1.13}$$

In principle, the DoA estimation method requires no coordination or synchronization with other BSs or devices, it is independent of the modulation characteristics of the incoming signal and accuracy is not limited by the signal bandwidth. General drawbacks are, that either the position or the phase difference between the antenna elements has to be known and antenna patterns are required. On the other hand, antennas positions are known from the manufacturing process and antenna patterns, if not already provided by the manufacturer, can be obtained from simulations or measurements in an anechoic chamber.

1.2.2. Non-Geometrical Approaches

In contrast to geometrical techniques, which are using trilateration, triangulation or other geometrical characteristics, non-geometrical techniques don't require a line of sight (LoS) path. A well-known non-geometrical techniques is fingerprinting as used in WiFi systems[1] for positioning [Ben07]. Fingerprinting is also called "scene analysis", because measurements are correlated with an a-priori obtained database to determine the position estimate. For example in WiFi a set of receive signal power measurements from multiple access points is taken for comparison with the data base. The principle is the following.

In a preliminary or offline step, the area of interest is quantized into segments and a database is built based on certain metrics such as channel or other signal signatures, like receive power. Each measured metric is mapped to a segment, access node and other system parameters, e.g. receive structure. The required values of the database can be assembled either by measurements or from simulation data. Once this is done, in the online phase the estimated values are compared or correlated with the values in the database and the estimated position is obtained by minimizing or maximizing certain metrics. Examples for values in the data base are receive signal power to multiple access points in [WLL+14], magmatic field measurements as in [AGSR14], or a combination of ToF and AoA in [VGL+15].

Beside other parameters, the accuracy of fingerprinting depends foremost on the quality of the database. This means that once the database is created, it has to be updated continuously, especially after significant changes in the environment. In addition a special location server is required for hosting the database. Another drawback of fingerprinting is, that measurements may be different from devices with different hardware. On the other hand no synchronization or LoS is required and methods to circumvent these drawbacks are of interest, e.g. [LDBL07, ZZL09, DAKN11]. As mentioned before, the offline step to assemble the database can be time consuming and expensive, especially when measurements are used with a high resolution. Therefore, interpolation methods are considered in [CWK+12] or blind crowd sourcing to generate the database [LJLH12].

1.2.3. Summary

In Sections 1.2.1 and 1.2.2 fundamental estimation techniques used in radio based localization are discussed. In Table 1.2 the pros and cons with respect to large antenna systems are provided. From the techniques compared, only in AoA estimation the accuracy scales with the number of antennas and is independent of the bandwidth. Furthermore, AoA is independent of required synchronization and clock accuracy limiting the performance of ToA and TDoA technique. Hence, AoA is the candidate technique to be used for localization with large antenna systems and in the focus of this thesis.

[1] WiFi here means a WLAN using the IEEE 802.11 standards.

Table 1.2.: Overview of pros and cons of localization estimation techniques.

	ToA	TDoA	AoA	RSS	Fingerprinting
Synchronization between BS and mobile device required	Yes	No	No	No	No
Synchronization between BSs required	No	Yes	No	No	No
Clock dependent	Yes	Yes	No	No	No
LoS required	Yes	Yes	Yes	Yes	No
Multiple antennas and calibration required	No	No	Yes	No	No
Accuracy scales with number of antennas	No	No	Yes	No	No
Database required	No	No	No	No	Yes

1.3. Thesis Contributions and Organization

This section describes the organization of the remaining thesis, whereby novel contributions in each chapter are highlighted by ⇒ signs. In general, this thesis contributes to and further develops the application of massive MIMO in cellular deployments. In Chapter 2 downlink data transmission with a single BS is treated. First, Section 2.1 introduces the underlying radio channel model, that is used throughout this thesis, followed by Section 2.2 providing the system model for downlink data transmission. It is worthwhile to mention that in contrast to the majority of massive MIMO literature, in this work multiple receive antennas at the user side are considered and taken into account for downlink data transmission, by the concept of equivalent multiple-input single-output (MISO) channels. In Section 2.3 TDD is assumed and downlink precoders based on perfect CSI under power constraints are investigated. In Section 2.3.4 the relevance of user grouping also for massive MIMO systems is shown, with the following contribution.

⇒ In Section 2.3.4 the user grouping algorithm "projection based zero forcing (PBZF)" is presented in detail together with an evaluation and comparison with state of the art user grouping. The PBZF has been used in previous publications [TKBH12, KRTT13, KTH14b], but only insufficient text descriptions have been provided, making reproducibility nearly impossible.

In the following Section 2.4 the "One Grand Question" from [BLM16] is investigated: "Can Massive MIMO work in FDD operation?". First codebook designs for massive MIMO antenna systems in FDD are studied in Section 2.4.1.

⇒ In Section 2.4.2, a novel codebook scheme, that splits a large codebook into smaller sub-codebook, is proposed. It is shown, that the proposed "sub-codebook splitting" is scalable with respect to the number of antennas and pilot overhead. The scheme quantizes the channel and keeps close to maximum ratio transmission (MRT) spectral efficiency, while providing coverage to all mobile users in the cellular coverage area.

In Section 2.4.3, the "sub-codebook splitting" scheme is investigated for multiple users showing that codebook based multiple-user transmission is limited by spatial inter-stream interference. Therefore, in Section 2.4.4 a two-stage hybrid precoding approach from 3GPP is adapted and combined with the proposed "sub-codebook splitting" scheme. Using the CSI of the codebook precoded channels, multiple user spatial multiplexing gains in FDD are investigated, considering a 100 antenna massive MIMO array and pilot overhead. A major advantage compared to other schemes in literature is, that no second order channel statistic is required at the BS.

⇒ In Section 2.4.5 CSI feedback reduction is investigated and it is shown, that also with a feedback constraint similar as in 5G NR Release 15, sum spectral efficiency gains with MIMO in FDD systems can be realized. Thereby, a novel scalable feedback mechanism that combines CSI quantization, number of reported RBs and number of streams in the first stage precoder is proposed. The scalable number of streams in the first stage precoding is based on the previous proposed "sub-codebook splitting" scheme.

Next, in Chapter 3 uplink DoA estimation with a massive MIMO array is investigated. While sum-throughput gains have been the original motivation for massive MIMO, multiple antennas can also be used for DoA estimation. The uplink DoA estimation model is given in Section 3.1. Section 3.2 provides a thorough parameter study for a massive MIMO uniform planar array (UPA) in a cellular outdoor deployment. Thereby, the focus is on array geometry independent search based DoA estimation techniques. A major drawback of these techniques is the large number of power spectrum computations required to realize the high accuracy offered by massive MIMO arrays. Especially, with 2D antenna structures a quantization in horizontal and vertical direction results in not feasible computational costs.

⇒ Section 3.2 proposes a novel complexity reduction method, called adaptive search space quantization (ASSQ), that is applicable to search based DoA estimation methods. The ASSQ scheme is based on adaptive quantization of the search space, going from a large search space with coarse quantization to small search spaces with fine quantization. A huge complexity reduction is achieved with approximately no performance loss in DoA estimation accuracy.

A further complexity reduction results in large DoA estimation errors due to coarse quantization in the first step the of ASSQ algorithm.

⇒ In Section 3.2 the proposed ASSQ scheme is extended to adaptive antenna selection per step. The adaptive antenna selection reduces most of the DoA estimation errors caused by large quantization steps. These large quantization steps, especially in the first steps, enable further complexity reduction.

Next, in Section 3.3 multiple source estimation is investigated to reduce the number of time-frequency resources that are required for single source DoA estimation in uplink. Such resources are valuable in commercial mobile cellular systems. In contrast to other DoA estimation literature, a uniform random user distribution in a cellular coverage area is considered and the goal is to "select" mobile users in order to minimize the DoA errors of multiple-source estimation. To the best knowledge of the author, this is the first approach that explicitly targets on user-grouping for multiple source DoA estimation with a large antenna array for cellular 5G systems.

⇒ In Section 3.3, a new two-step user grouping algorithm is proposed, that selects users with the goal of minimizing DoA errors for multiple-source estimation. In the first step, users with similar uplink receive power are grouped together. In the second step, per power group users are selected based on angular separation. The algorithm is discussed in detail, and simulations show that the input design parameters enable a scaling between estimation accuracy and number of jointly estimated mobile users.

Finally, in Chapter 4 challenges that arise in cellular systems are treated. In the first part, in Section 4.1, downlink data transmission in the presence of uncoordinated multiple-sector interference is investigated, extending the work from Chapter 2. Thereby, a homogeneous scenario with only massive MIMO BSs and a heterogeneous scenario where non-massive MIMO BSs surround a massive MIMO BS is taken into account.

⇒ Section 4.1.2 shows, that interference knowledge at the mobile users is required to utilize spatial multiplexing gains of massive MIMO. A low-feedback overhead metric is proposed, promising significant gains in the homogeneous and heterogeneous scenario.

In Section 4.2, the challenge of localization in 3D space based on DoA estimated from Chapter 3 is investigated. The minimum distance of skew lines method is presented in Section 4.2.1, requiring two BSs together with their position and corresponding DoA estimates of the same source.

⇒ In Section 4.2.2, methods to compensate dilution of precision (DoP) based position errors are proposed. While these methods themselves are not new, their application and shown compatibility to the 5G NR Release 15 frame and data structure is novel. The section shows, that with "simple" averaging in the frequency domain most of the DoP based position errors can be compensated.

Finally, Chapter 5 concludes this thesis, followed by Appendices A and B for rearrangement of equations and additional material, respectively. At the end of this thesis the bibliography and the authors curriculum vitae are given.

2. Downlink Data Transmission

This chapter studies multiple-user downlink transmission with a centralized massive multiple-input multiple-output (MIMO) array. While the deployment of the base station (BS) and mobile users follows cellular deployment assumptions, a single BS is assumed in order to avoid the impact of uncoordinated interference. In Section 2.1 the underlying radio channel model is introduced and used throughout the thesis. The subsequent Section 2.2 describes the multiple-user multiple-cell system model. For downlink multiple-user spatial multiplexing the amount and quality of channel state information (CSI) at the transmitter is crucial. Accordingly, Section 2.3 assumes perfect and complete CSI at the transmitter, while Section 2.4 considers limited and imperfect CSI.

2.1. Radio Channel - A Geometry-Based Stochastic Model

All scientific models and simulations are based on certain assumptions and simplifications compared to "reality". For wireless communications that focuses on the physical layer, the underlying radio channel is an essential part, in particular as innovations in the field are mainly validated and justified by simulations. Therefore, a valid and often asked question is, how realistic is the underlying channel model? In general, the more realistic a channel model is, the more complex it becomes. In wireless communications, the channel models can be classified into two groups [MJY10]:

1. Empirical Models
2. Deterministic Models

The empirical models can be further divided into two approaches, a) stored channel impulse responses and b) stochastic channel modeling methods. For stored channel impulse responses expensive channel sounders are required. However, with respect to research papers and standardization stochastic channel models are used most often, above all the Rayleigh fading model [SFGK00]. Stochastic channel models are mainly used to design and evaluate system performance. The motivation is that e.g. with the "urban macro" scenario, not a particular city but the general propagation environment of cities is modeled in order to obtain generally valid conclusions.

On the other hand, if the focus is on the optimization of a particular environment, e.g. the cabin of an air plane, then deterministic channel models are used based on Maxwell's equations using the characteristics and geometry of the environment. For example channel modeling of "millimeter-wave" or even higher frequencies with highly-directive antennas uses deterministic channel models, e.g. using ray-tracing tools, because only few dominant paths are visible. However, the selection of the channel model for a specific communications system also depends on further parameters inter alia the signal bandwidth and the scattering environment.

In this work, a hybrid channel model is used combining features of both the deterministic and stochastic approach, called QUAsi Deterministic RadIo channel GenerAtor (QuaDRiGa) [JRBT14]. Especially, the recent massive MIMO extensions and validations by measurements clinched the decision for QuaDRiGa with details in the next paragraphs followed by the system model description in Section 2.2.

QuaDRiGa - QUAsi Deterministic RadIo channel GenerAtor

The QuaDRiGa is a geometry-based stochastic channel model that is open source and public available under the Lesser General Public License (LGPL) at [HHI17]. It follows the modeling principles of geometry-based stochastic channel models such as the 3rd Generation Partnership Project (3GPP) "spatial channel model extended" [BHS05], or the wireless world initiative new radio (WINNER) channel model [WIN10] or the COST2100 [LOP+12]. In contrast to "spatial channel model extended", a full three dimensional (3D) modeling approach is used in QuaDRiGa differentiating the horizontal and vertical plane. This 3D modeling includes

- path-loss distance,
- angle of arrival/departure due to 3D scatterer placement, and
- geometric polarization propagation to enable co-/cross-polarized antenna arrays.

Especially, a 3D modeling approach is essential for massive MIMO evaluations, because two dimensional (2D), e.g. rectangular arrays, or even 3D, e.g. cylindrical array structures are required in order to keep antenna apertures to reasonable limits. For example with $f_{(\mathrm{c})} = 4\,\mathrm{GHz}$ and 32×32 antenna elements in a uniform planar array (UPA) with $\frac{\lambda_{(\mathrm{c})}}{2}$ horizontal/vertical element spacing, the total number of antennas is $N = 1024$ with an aperture in horizontal/vertical of only $\approx 1.17\,\mathrm{m}$. This also enables so called "3D beamforming" one of the extensions of Long Term Evolution (LTE) in Release 13, see [3GP15a]. 3D beamforming provides performance gains in particular for indoor user equipments (UEs) in upper floors of high-rise buildings.

Other QuaDRiGa extensions with respect to the geometry-based stochastic channel model that are utilized in this work are:

- Time evolution - Short-term time evolution of the channel coefficients is realized by updating the delays, the departure- and arrival angels, the polarization, the shadow fading and the K-Factor based on the position of the UE.
- Improved method for calculating correlated large scale parameter - The wireless world initiative new radio (WINNER) model calculates maps of correlated parameter values using filtered random numbers. QuaDRiGa uses the same method but extends the map generation algorithm considering also diagonal movement directions to create smooth outputs.
- New functions for modifying antenna patterns - Antenna patterns can now be freely rotated in 3D-coordinates while maintaining the polarization properties.

Time evolution is required for feedback delay, e.g. in frequency division duplex (FDD) when UEs report their CSI to the BS. The correlation of large scale parameters has an inherent impact on the multiple-user performance and feature three is used for the generation of antenna arrays in the cellular deployment. Furthermore, a wide range of propagation scenarios is supported, such as

- macro, micro, pico, femto, relay and satellite [BDJT12],
- 100 % compatible with measurements from wireless world initiative new radio (WINNER), and
- own parametrization for urban-macro with high altitude base stations.

Each propagation scenario defines parameters that can be directly applied to the QuaDRiGa channel model derived from extensive measurement campaigns, e.g in wireless world initiative new radio (WINNER) [WIN10] or 3GPP. This work focuses on the urban macro/micro line of sight (LoS) or non line of sight (NLoS) scenarios. QuaDRiGa is in constant developments and the following recent extensions from version 2.0 are considered in this work:

- Spatially consistent small-scale-fading
- Support for large array antennas - ”massive MIMO”

Spatially consistent small-scale-fading has a large impact on multiple-user performance, e.g. if UEs are geographically close to each other they observe similar scatterers and thus correlated small-scale fading. This correlation results in rank reduction and loss due to precoder normalization for MIMO precoders. Support for large array antennas means that small scale parameters change over the elements of an array and each antenna element observes different but correlated multiple-path delays and angles.

Calibration with 3GPP

To the best of the author's knowledge, QuaDRiGa is the only available open source implementation that is constantly calibrated against 3GPP recommendations, the main standardization body for cellular mobile communications network. The recent QuaDRiGa update v2.0 is calibrated against 3GPP fifth generation (5G) channel model requirements [3GP17f] and LTE extensions [3GP17e].

Relevant Modeling Aspects

In this paragraph, further relevant modeling aspect are discussed as well as their impact on system level simulations.

- In contrast to the one-ring model, QuaDRiGa can model radiation patterns of antenna elements. Therefore, multi-path components from certain directions are attenuated or amplified with respect to an ideal isotropic radiator, see Fig. 3.6a. The radiation pattern in Fig. 3.6a is according to recommendations for 3GPP system level simulations [3GP17d].
- Due to explicit 3D modeling of array patterns with main- and side-lobes, precoding can result in unintended interferences along multi-path components, e.g. a beamformer designed to point in a certain direction also radiates power into other directions, see Fig. 3.6b.
- Channels between closely spaced antenna elements, e.g. $\frac{\lambda_{(c)}}{2}$, are correlated resulting in rank deficient channels or in other words the multiplexing gain is less than the number of antennas.
- Large scale parameters can change in large antenna arrays and therefore QuaDRiGa provides the option that delays of multi-path components are modeled for all antenna elements instead of the array center only.
- Scatterer positions and large-scale parameters are modeled by random distributions. These random distributions are spatially correlated, meaning for example that users close to each other observe similar path-loss.
- In order to obtain a sufficiently large statistic, Monte-Carlo simulations are performed for each set of simulation parameters. Due to the stochastic generation of large and small-scale parameters in each channel realization, the random seed is changed in each realization. Due to the deterministic modeling of path-loss, users are placed randomly and independently in each channel realization in areas called sectors. Often, independent channel realizations are referred to as “drops” because the user placement is similar to “rain drops”.

2.2. Multiple-User Multiple-Cell System Model

According to Section 1.1, this thesis focuses on system level perspective meaning that the abstraction level of the physical (PHY) layer is the signal to interference and noise ratio (SINR). The downlink SINR describes the quality of a receive stream at the mobile user. In contrast to link-level simulations, where bits are transmitted and received to obtain bit error rates, system level simulations use mapping tables to obtain throughput values from SINRs [3GP18b]. These so called link-to-system mapping tables need to be obtained for each parameter configuration, each modulation and coding scheme (MCS), and each possible bandwidth (number of resource blocks (RBs) a user can get) to calibrate the mapping interfaces such as mutual information effective SINR mapping (MIESM) [YXXW09] or mean mutual information per bit (MMIB) [BWK08]. It is computationally too complex to obtain these link-to-system tables only for this thesis, because each changed parameter in the simulation assumption would require a new link-to-system mapping curve and a corresponding link-level simulation. Instead, the capacity of the links is used as key performance indicator (KPI) with a detailed description in the paragraphs below.

In this work a multiple-user multiple-sector downlink orthogonal frequency division multiplexing (OFDM) system with K users and L BSs is considered, where $\mathcal{L}$ is the set of BSs IDs $\{1,\dots,L\}$ and $\mathcal{K}$ the set of user IDs $\{1,\dots,K\}$. Each user is equipped with M receive antennas and each BS with N transmit antennas. With this the receive signal $\mathbf{y}_k \in \mathbb{C}^M$ at mobile station (MS) $k \in \mathcal{K}$ on a coherence block[1] is given by

$$\mathbf{y}_k = \sum_{l=1}^{l\in\mathcal{L}} \mathbf{H}_{k,l}\mathbf{V}_l\sqrt{\mathbf{F}_l}\mathbf{x}_l + \mathbf{n}_k, \tag{2.1}$$

where $\mathbf{H}_{k,l} \in \mathbb{C}^{M\times N}$ is the downlink channel matrix from BS l to MS k, $\mathbf{V}_l \in \mathbb{C}^{N\times T_l}$ is the precoding matrix applied by BS l normalized such that trace $\left(\mathbf{V}_l\mathbf{V}_l^{\mathrm{H}}\right) = 1$, where T_l is the cardinality of set $\mathcal{T}_l = \{1,\dots,T_l\}$ of spatial layer IDs transmitted by BS l, $\mathbf{F}_l \in \mathbb{R}_+^{T_l\times T_l}$ is the diagonal power allocation matrix at BS l with the power constraint $P^{(\mathrm{RB})} \geq \mathrm{trace}(\mathbf{F}_l)$, $\mathbf{x}_l \in \mathbb{C}^{T_l}$ is the vector of transmit data symbols at BS l to be send to the MSs, and $\mathbf{n}_k \in \mathbb{C}^M$ denotes the additive white Gaussian noise (AWGN) vector with covariance $\mathbb{E}\left(\mathbf{n}_k\mathbf{n}_k^{\mathrm{H}}\right) = \mathbf{I}_M\sigma_k^2$. $P^{(\mathrm{RB})}$ is the maximum transmit power per subcarrier. $\mathbb{E}(\cdot)$ is the expectation value and σ_k^2 is the noise power comprising the receiver and thermal noise at user k. $\mathcal{K}_l$ is the set of served users by BS l and consequently the number of multiplexed layers $T_l \leq K_l$, where equality is achieved if each users is served with one layer. $\mathcal{K}_l$ is a subset of the available users connected to BS l denoted by $\tilde{\mathcal{K}}_l$ such that $\mathcal{K}_l \subseteq \tilde{\mathcal{K}}_l$. The selection of $\mathcal{K}_l$ from $\tilde{\mathcal{K}}_l$ is done by a scheduling entity, discussed in Section 2.3.4. Each column of the precoding matrix $\mathbf{V}_l = [\mathbf{v}_1 \ \dots \ \mathbf{v}_T]$ is a spatial multiplexed stream or layer and assigned to a certain MS. It is shown in [BLM16] that linear precoding achieves capacity close to optimal non-linear precoding schemes such as dirty paper coding [Cos83] or successive interference cancellation [Fos96] in the large antenna regime, e.g. already at 100 antennas in [BLM16]. Therefore, this work focuses on linear precoding such that the number of spatial layers is constraint by the number of BS antennas as $T_l \leq N$. Assuming that MS k is connected to and served by BS l on stream $t \in \mathcal{T}_l$ the receive signal from Eq. (2.1) can be divided into three parts according to

$$\mathbf{y}_{k,t} = \underbrace{\mathbf{H}_{k,l}\mathbf{v}_{t,l}\sqrt{[\mathbf{F}_l]_{t,t}}}_{\tilde{\mathbf{h}}_{k,t}}[\mathbf{x}_l]_t + \underbrace{\sum_{j\neq t}^{j\in\mathcal{T}_l} \mathbf{H}_{k,l}\mathbf{v}_{j,l}\sqrt{[\mathbf{F}_l]_{j,j}}}_{\boldsymbol{\vartheta}_{k,t}}[\mathbf{x}_l]_j + \underbrace{\sum_{m\neq l}^{m\in\mathcal{L}} \mathbf{H}_{k,m}\mathbf{V}_m\sqrt{\mathbf{F}_m}\mathbf{x}_m + \mathbf{n}_k}_{\mathbf{z}_k}, \tag{2.2}$$

where the intra-sector interference caused by streams $j \neq t$ is aggregated in $\boldsymbol{\vartheta}_{k,t} \in \mathbb{C}^M$, inter-sector interference from other BSs is denoted as $\mathbf{z}_k \in \mathbb{C}^M$, and $\mathcal{T}_l$ is the set of spatial layers from BS l. The subscript l of stream t is omitted for notational brevity and not required, since t corresponds to the stream with symbols

[1] Coherence block means a resource with time and frequency dimension equal or less than the coherence time and coherence frequency, respectively, such that the channel in the coherence block can be considered as constant over time and frequency. In literature this is also called block fading channel model [HZT+16].

for MS k, and BS l is the serving BS of MS k. The linear receive filter at MS k for stream t is denoted by $\mathbf{w}_{k,t} \in \mathbb{C}^M$. Assuming Gaussian distributed normalized receive symbols the SINR of stream t at MS k including post-processing at the receiver is obtained by

$$\gamma_{k,t} = \frac{\mathbf{w}_{k,t}^{\mathrm{H}} \tilde{\mathbf{h}}_{k,t} \tilde{\mathbf{h}}_{k,t}^{H} \mathbf{w}_{k,t}}{\mathbf{w}_{k,t}^{\mathrm{H}} \tilde{\mathbf{Z}}_{k,t} \mathbf{w}_{k,t}}, \tag{2.3}$$

where $\tilde{\mathbf{Z}}_{k,t} \in \mathbb{C}^{M \times M}$ is the interference covariance plus noise matrix obtained by

$$\tilde{\mathbf{Z}}_{k,t} = \boldsymbol{\vartheta}_{k,t} \boldsymbol{\vartheta}_{k,t}^{\mathrm{H}} + \mathbf{z}_k \mathbf{z}_k^{\mathrm{H}} + \mathbf{n}_k \mathbf{n}_k^{\mathrm{H}}. \tag{2.4}$$

Finally, the capacity of the effective channel-link between BS l serving user k on stream t in the presence of AWGN is characterized by [TV05]

$$R_{k,t} = B \log_2 \left(1 + \gamma_{k,t}\right), \tag{2.5}$$

where B is the bandwidth of the link and $R_{k,t}$ is given in [bit/s]. Note, that the capacity expression in Eq. (2.5) is often referred to as "Shannon" capacity in literature, named after the developer of the underlying information theory Claude E. Shannon [Sha98]. The rate is often normalized by the bandwidth to the spectral efficiency, providing an metric independent of the bandwidth such that

$$C_{k,t} = \frac{R_{k,t}}{B} = \log_2 \left(1 + \gamma_{k,t}\right), \tag{2.6}$$

where $C_{k,t}$ is given in [bit/s/Hz]. The "Shannon" capacity is an upper bound for the performance enabling comparison and analysis of physical layer techniques without computational expensive link-level simulations. Furthermore, the sum over all layers T_l of BS l given by $R_l^{(\mathrm{sum})}$ as

$$R_l^{(\mathrm{sum})} = \sum_{t=1}^{T_l} R_{k,t}, \tag{2.7}$$

and the corresponding sum spectral efficiency as

$$C_l^{(\mathrm{sum})} = \sum_{t=1}^{T_l} C_{k,t}. \tag{2.8}$$

In real systems the SINR that can be utilized is upper and lower bounded due to MCSs, e.g. in the LTE link level simulator [RST16] the signal to noise ratio (SNR) ranges from $-5\,\mathrm{dB}$ to $19.83\,\mathrm{dB}$. Therein the highest modulation is a 64 quadrature amplitude modulation (QAM) which corresponds to $C_{k,t} < 6\,\mathrm{bit/s/Hz}$ spectral efficiency per link considering channel coding. 3GPP standardization investigated up to 1024 QAM for the LTE downlink in order to utilize higher SINRs [3GP17a]. Therefore, the Shannon rate in Eq. (2.5) is bounded by

$$\tilde{R}_{k,t} = \begin{cases} 0\,\mathrm{bit/s} & ,\text{if } \gamma_{k,t} < \gamma^{(\min)}, \\ R_{k,t} & ,\text{if } \gamma^{(\min)} \leq \gamma_{k,t} \leq \gamma^{(\max)}, \\ B \log_2 \left(1 + \gamma^{(\max)}\right)\,\mathrm{bit/s} & ,\text{if } \gamma_{k,t} > \gamma^{(\max)}. \end{cases} \tag{2.9}$$

As a consequence, streams with SINR below the lower bound $\gamma^{(\min)}$ do not contribute to the sum rate $R_l^{(\mathrm{sum})}$ and streams with SINR above the upper bound $\gamma^{(\max)}$ do not increase the sum rate. The corresponding bounded spectral efficiency $\tilde{C}_{k,t}$, bounded sum rate $\tilde{R}_l^{(\mathrm{sum})}$, and bounded sum spectral efficiency $\tilde{C}_l^{(\mathrm{sum})}$ are obtained according to Eq. (2.6), Eq. (2.7), and Eq. (2.8), respectively, by replacing $R_{k,t}$ with $\tilde{R}_{k,t}$.

In the time before massive MIMO the dominant system assumption was FDD due to stable interference situations in downlink direction [CHE13]. Nevertheless, time division duplex (TDD) is considered in the

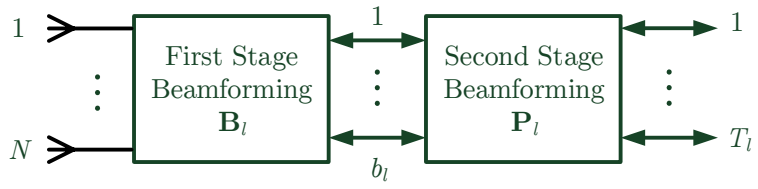


Figure 2.1.: Two stage precoding at BS l with first-stage precoder $\mathbf{B}_l$, second stage precoder $\mathbf{P}_l$, and T_l data streams.

introduction of massive MIMO [Mar10] because the explicit downlink CSI is required for the design of the downlink precoder $\mathbf{V}$. The explicit downlink CSI can be obtained by channel estimation in the uplink from a single shot transmission per MS at all BS antennas. In contrast to this, in FDD a pilot per antenna has to be transmitted in downlink for channel estimation at the MS side. Additionally, the estimated CSI has to be fed back to the BS in the uplink. The time for pilot transmission, channel estimation and feedback transmission is limited by the coherence time of the channel. Therefore, N is limited by the coherence time of the channel. In TDD, the coherence time limits the number of MSs that can be estimated in the uplink slots but not the number of antennas at the BS. However, most of the world wide spectrum is allocated to FDD mode and solutions to minimize the performance loss of massive MIMO in FDD compared to TDD are in the focus [BLM16]. These solutions can be classified into

- hybrid-beamforming [KKS13, HJL$^+$14],
- precoded pilots [AM12, KTH$^+$16a], and
- two-stage precoding [KTC15a, KTC15b].

They have all in common that the precoder $\mathbf{V}_l$ is divided into two parts such that

$$\mathbf{V}_l = \mathbf{B}_l \mathbf{P}_l, \tag{2.10}$$

where $\mathbf{B}_l \in \mathbb{C}^{N \times b_l}$ is the first stage precoding matrix, $\mathbf{P}_l \in \mathbb{C}^{b_l \times T_l}$ is the second stage or multi-user precoding matrix, and $T_l \leq b_l \leq N$ is a integer design parameter, see Fig. 2.1. The precoders are normalized such that $\|\mathbf{B}_l\| = 1, \forall l \in \mathcal{L}$ and $\|\mathbf{P}_l\| = 1, \ \forall l \in \mathcal{L}$. Similar precoder splitting is also considered in 3GPP standardization for full dimension MIMO [NRL$^+$15]. Furthermore, in [TK14] and [KSHT15] the same two-stage precoder structure is applied for spatial coordination between macro and small cells.

This work assumes that the first stage precoder $\mathbf{B}_l$ is either obtained from codebooks or a function of the second order statistics such that

$$\mathbf{B}_l = f\left(\mathbf{R}_{1,l}, \ldots, \mathbf{R}_{\tilde{K}_l,l}\right), \tag{2.11}$$

where $\mathbf{R}_{k,l} \in \mathbb{C}^{N \times N}$ is the downlink transmit covariance matrix between BS l and user k obtained by

$$\mathbf{R}_{k,l} = \mathbb{E}\left(\mathbf{H}_{k,l}^{\mathrm{H}} \mathbf{H}_{k,l}\right). \tag{2.12}$$

These second order statistics change with the same coherence time as large scale parameters and can be assumed constant for a longer time than the small-scale dependent fading. With the first stage precoder $\mathbf{B}_l$ an effective channel $\hat{\mathbf{H}}_{k,l} \in \mathbb{C}^{M \times b_l}$ is generated by

$$\hat{\mathbf{H}}_{k,l} = \mathbf{H}_{k,l} \mathbf{B}_l. \tag{2.13}$$

This effective channel is the input to the second stage precoder $\mathbf{P}_l$ such that

$$\mathbf{P}_l = f\left(\hat{\mathbf{H}}_{1,l}, \ldots, \hat{\mathbf{H}}_{K,l}\right). \tag{2.14}$$

Note that the case of explicit and perfect channel information can be achieved by $\mathbf{B}_l = \mathbf{I}_N$. Throughout this work equal power distribution is assumed such that

$$[\mathbf{F}_l]_{1,1} = [\mathbf{F}_l]_{2,2} = \ldots = [\mathbf{F}_l]_{T_l,T_l}\,. \tag{2.15}$$

Thus, the transmit power per stream t is inherently given by the norm of column t of the precoding matrix $\mathbf{V}_l$.

The above system model adopts assumptions in 3GPP standardization [NRL+15], such that the first stage precoder $\mathbf{B}_l$ corresponds to a wideband or analog beamformer mapping physical antenna elements to antenna ports. Each antenna port is connected to a transceiver chain.

A common assumption in massive MIMO literature is that users are equipped with a single antenna [Mar10, HtBD11, HCPR12, ANAC13, BHKD14, LETM14, SMD15, LP16, FRZ+17]. One reasons for this is simplicity to derive analytical expressions. On the other hand, LTE mobile phones use already two antennas for reception and more are expected in future 5G systems [WTL+16]. Therefore, the extension of massive MIMO schemes to $M > 1$ receive antennas is required. Throughout this thesis, multiple antennas at the user side are explicitly considered and taken into account for downlink transmissions or feedback. One approach to achieve this was proposed in [Thi13] for joint transmission in coordinated multi-point systems in FDD. The principle idea is that the $M \times N$ MIMO channel is decomposed into M channels of size $1 \times N$ at the user side, so called multiple-input single-output (MISO) channels.

In this thesis, the approach from [Thi13] is adopted to ensure that the massive MIMO transmission scheme proposed in this thesis are compatible with 5G. Therein, a linear receive filters, in this work defined as $\mathbf{w}_{k,t}$ in (2.3), is multiplied to the channel matrix resulting in an effective channel. The advantage of this MISO feedback is that each MISO channel can be treated as an independent single antenna users and massive MIMO schemes based on single antennas users can be used without adaptation.

In essence, the MIMO channel $\mathbf{H}_{k,l}$ from BS l to user k is decomposed by singular value decomposition (SVD) to

$$\mathbf{H}_{k,l} = \bar{\mathbf{U}}_{k,l}\bar{\mathbf{\Sigma}}_{k,l}\bar{\mathbf{V}}_{k,l}^{\mathrm{H}}, \tag{2.16}$$

where $\bar{\mathbf{U}}_{k,l} \in \mathbb{C}^{M\times M}$ is a unitary matrix, $\bar{\mathbf{\Sigma}}_{k,l} \in \mathbb{R}_{+}^{M\times N}$ is a rectangular diagonal matrix, and $\bar{\mathbf{V}}_{k,l} \in \mathbb{C}^{N\times N}$ is also a unitary matrix. The columns of $\bar{\mathbf{U}}_{k,l}$ and $\bar{\mathbf{V}}_{k,l}$ are the left and right singular vectors, respectively. These singular vectors correspond to the ordered non negative singular values such that

$$\left[\bar{\mathbf{\Sigma}}_{k,l}\right]_{1,1} \geq \left[\bar{\mathbf{\Sigma}}_{k,l}\right]_{2,2} \geq ... \geq \left[\bar{\mathbf{\Sigma}}_{k,l}\right]_{M,M}\,. \tag{2.17}$$

Indeed, the number of singular values is limited by $\min(M, N)$, rather than M. However, according to the massive MIMO assumption that the number of BS antennas N is larger than the number of user antennas M such that $N > M$. Then, the m-th equivalent MISO channel is obtained by

$$\begin{aligned}\mathbf{h}_{m,k,l}^{(\mathrm{MISO})} &= \left[\bar{\mathbf{U}}_{k,l}\right]_{:,m}^{\mathrm{H}} \mathbf{H}_{k,l} \\ &= \left[\bar{\mathbf{U}}_{k,l}\right]_{:,m}^{\mathrm{H}} \bar{\mathbf{U}}_{k,l}\bar{\mathbf{\Sigma}}_{k,l}\bar{\mathbf{V}}_{k,l}^{\mathrm{H}} = \left[\bar{\mathbf{\Sigma}}_{k,l}\right]_{m,m}\left[\bar{\mathbf{V}}_{k,l}\right]_{:,m}^{\mathrm{H}},\end{aligned} \tag{2.18}$$

where the colon operator notation $[\mathbf{A}]_{:,j}$ represents all elements of column j of matrix $\mathbf{A}$ and $m \in \{1, \ldots, M\}$. Let $M_k^{(\mathrm{MISO})} \leq M$ be the number of equivalent MISO channels reported by user k, then the amount of feedback is scalable and $M_k^{(\mathrm{MISO})}$ can be a function of the uplink capacity, user traffic requirements, load in the network, or any other metric. For clarification, in case of $M_k^{(\mathrm{MISO})} < M$ it is assumed that the MISO channels $\mathbf{h}_{1,k,l}^{(\mathrm{MISO})}, \ldots, \mathbf{h}_{M_k^{(\mathrm{MISO})},k,l}^{(\mathrm{MISO})}$, that correspond to the first $M_k^{(\mathrm{MISO})}$ eigenvalues of $\bar{\mathbf{\Sigma}}_{k,l}$, are reported. In case all M MISO channels are reported the notation is

$$\mathbf{H}_{k,l}^{(\mathrm{MISO})} = \left[\mathbf{h}_{1,k,l}^{(\mathrm{MISO})} \;\; \ldots \;\; \mathbf{h}_{M_k,k,l}^{(\mathrm{MISO})}\right], \tag{2.19}$$

where the antenna subscript is omitted. In case of first-stage precoded channels $\hat{\mathbf{H}}_{k,l} = \mathbf{H}_{k,l}\mathbf{B}_l$ according to (2.13), the same notation for equivalent MISO channels as above is used in combination with the ˆ accent. Consequently, the equivalent MISO channel of the first-stage precoded effective channel $\hat{\mathbf{H}}_{k,l}$ from Eq. (2.13) is given as

$$\hat{\mathbf{H}}_{k,l}^{(\mathrm{MISO})} = \left[\hat{\mathbf{h}}_{1,k,l}^{(\mathrm{MISO})} \ \ldots \ \hat{\mathbf{h}}_{M_k,k,l}^{(\mathrm{MISO})}\right], \tag{2.20}$$

where $\hat{\mathbf{h}}_{m,k,l}^{(\mathrm{MISO})}$ is the m-th equivalent MISO channel obtained with Eq. (2.16) - Eq. (2.18) by replacing $\mathbf{H}_{k,l}$ with the effective channel $\hat{\mathbf{H}}_{k,l}$.

Note, that zero forcing (ZF) precoding in combination with the above described MISO channels is equivalent to block diagonalization (BD) precoding with MIMO channels, ensuring zero inter-user and inter-stream interference in case multiple users are served by multiple streams each. ZF precoding is described in detail in the next section.

Remark. *The above described MISO channel decomposition is assuming a FDD system where the estimated downlink channel is known at the user and reported as feedback to the BS. In a TDD system each user has to transmit a pilot per antenna and the same signal processing as described in Eq.* (2.16)*-Eq.* (2.20) *can be done at the BS on the estimated downlink channels prior to other signal processing. This adaptation is straightforward and omitted here for the sake of brevity.*

2.3. Full Channel State Information in Time Division Duplex (TDD)

In [Mar10] the term massive MIMO has been introduced with the assumption that the system operates in TDD, where the downlink channel can be obtained from the uplink channel estimated at the BS. Thus, most massive MIMO related literature assumes single antenna MS such that $M = 1$. Then, the channel can be estimated at the BS at once in a single time slot. Thus, in this section the full channel is assumed to be known at the BS for the design of the downlink precoder such that

$$\mathbf{B}_l = \mathbf{I}_N, \tag{2.21}$$

resulting in $\hat{\mathbf{H}}_{k,l} = \mathbf{H}_{k,l}$ according to Eq. (2.13), also referred to as "full digital" precoding in literature [SJJS00].

Massive MIMO means that asymptotic arguments based on random matrix theory [Mar10] are used for performance evaluation. The main effects are:

1. The impact from uncorrelated noise and small-scale fading vanishes due to averaging.
2. Spectral efficiency of a single user scales logarithmically with the number of antennas that corresponds to a linear scaling with SINR.
3. The required transmitted energy per bit vanishes as the number of antennas in a MIMO cell grows to infinity.
4. The sum spectral efficiency scales linearly with the number of served users assuming single antenna MSs, which corresponds to the number of streams T_l in this work.

One advantage of massive MIMO is that linear precoding schemes are approximately optimal [Mar10], compensating the increase in complexity. Therefore, the focus is on linear precoding schemes for the numerical evaluation of massive MIMO effects, described in the next section.

Note that the author of this thesis has already published parts of this section in [KRTT13, TKBH13, KTH14b, KJTB15]. However, due to changes in simulation assumptions in this thesis and evolution of the underlying channel model compared to the publications, numerical simulation results presented in this work have not been published before and differ from results in the publications.

2.3.1. Linear Precoding Schemes

In order to increase the SINR in Eq. (2.3), the numerator can be increased, the denominator decreased, or the ratio of both increased. Due to the multiple-user interference, inter-cell interference part, and other side-constraints such as quantization or delay, the design of precoders is still subject to a lot of research. The following set of well-known linear precoding schemes is considered: maximum ratio transmission (MRT) [Lo99], ZF [YG06], signal to leakage and noise ratio (SLNR) [STS07a], and minimum mean square error (MMSE) or regularized ZF [PHS05], each optimizing a different objective. In order to reflect the fact that the set of users connected to a BS can be larger than the set of served users, $\mathcal{K}_l \subset \tilde{\mathcal{K}}_l$ denotes the set of users selected for downlink transmission on the same time-frequency resource by a scheduling entity. Furthermore, if users have more than one receive antenna, they can be served with more than one spatial layer, thus the number of spatial layers can be larger than the number of served users such that $T_l \geq |\mathcal{K}_l|$. With the equivalent MISO feedback according to (2.18) each entry in $\mathcal{T}_l$ corresponds to a user k and the m-th MISO channel. It is intuitive that if a user is selected only once, the first MISO channel $m = 1$ is used, however it may happen that a scheduling entity selects the $m = 2$ due to orthogonality to other selected users. So in the following the equivalent single antenna channel $\mathbf{h}_{t,l}^{(\mathrm{MISO})}$ corresponds to $\mathbf{h}_{m,k,l}^{(\mathrm{MISO})}$. Note that the index l indicating the service BS l is omitted if not required in the following for better readability.

Maximum Ration Transmission (MRT)

The objective of MRT is to increase the SINR from Eq. (2.3) due to maximizing the numerator by obtaining

$$\mathbf{p}_t^{(\mathrm{MRT})} = \arg\max_{\mathbf{p}_t} \mathbf{H}_k \mathbf{p}_t, \tag{2.22}$$

where subscript t indicates stream t assigned to user k. From Eq. (2.22), it can be seen that the receive signal is optimized for each MS independently and multiple-user or intra-cell interference caused towards others is ignored. Representing $\mathbf{H}_k$ of MS k by SVD as in Eq. (2.16) the precoding vector $\mathbf{p}_t$ of stream t intended for MS k is formed by

$$\mathbf{p}_t^{(\mathrm{MRT})} = \mathbf{v}_{k,1}, \tag{2.23}$$

where $\mathbf{v}_{k,1}$ is the right eigenvector corresponding to the largest singular value $[\mathbf{\Sigma}_k]_{1,1}$. Consequently, if user k is assigned to a second stream $t' \neq t$ the MRT precoding vector is $\mathbf{p}_{t'}^{(\mathrm{MRT})} = \mathbf{v}_{k,2}$. The extension to more than two streams is straightforward. Thus, the precoder $\mathbf{P}^{(\mathrm{MRT})}$ of set $\mathcal{T}_l$ is obtained by

$$\mathbf{P}_{\mathcal{T}_l}^{(\mathrm{MRT})} = \left[\mathbf{p}_1^{(\mathrm{MRT})} \ \ldots \ \mathbf{p}_T^{(\mathrm{MRT})}\right]. \tag{2.24}$$

With MISO channels according to Eq. (2.18) the precoder $\mathbf{p}_t^{(\mathrm{MRT})}$ is obtained by

$$\begin{aligned} \mathbf{p}_t^{(\mathrm{MRT})} &= \frac{1}{\left|\mathbf{h}_{m,k}^{(\mathrm{MISO})}\right|} \left(\mathbf{h}_{m,k}^{(\mathrm{MISO})}\right)^{\mathrm{H}} = \frac{1}{\left|[\mathbf{\Sigma}_k]_{m,m} [\mathbf{V}_k]_{:,m}^{\mathrm{H}}\right|} \left([\mathbf{\Sigma}_k]_{m,m} [\mathbf{V}_k]_{:,m}^{\mathrm{H}}\right)^{\mathrm{H}} \\ &= \frac{1}{[\mathbf{\Sigma}_k]_{m,m}} [\mathbf{V}_k]_{:,m} [\mathbf{\Sigma}_k]_{m,m} = \mathbf{v}_{k,m}, \end{aligned} \tag{2.25}$$

where subscript m refers to the m-th MISO channel. Eq. (2.25) shows that from MISO channel $m = 1$ the same MRT precoder as in Eq. (2.23) can be obtained. If user k is assigned to a second stream t' the second MISO channel $m = 2$ is used and so on.

Zero Forcing (ZF)

The objective of ZF is to increase the SINR due to minimizing the denominator by "forcing" multiple-user interference to "zero". This is achieved for MS k served on stream t by the condition

$$\mathbf{H}_k \mathbf{p}_{t'} = 0, \quad \text{for } t' \neq t. \tag{2.26}$$

In order to ensure Eq. (2.26) for all $t \in \mathcal{T}_l$, first the compound channel matrix $\underline{\mathbf{H}}_{\mathcal{T}_l}$ is build such that

$$\underline{\mathbf{H}}_{\mathcal{T}_l} = \left[\mathbf{h}_1^{(\mathrm{MISO})} \ \dots \ \mathbf{h}_{T_l}^{(\mathrm{MISO})}\right]. \tag{2.27}$$

According to [YG06], the $\mathbf{P}_{\mathcal{T}_l}$ that yields zero multiple-user interference is the pseudoinverse of $\underline{\mathbf{H}}_{\mathcal{T}_l}$ obtained by

$$\mathbf{P}_{\mathcal{T}_l}^{(\mathrm{ZF})} = \underline{\mathbf{H}}_{\mathcal{T}_l}^{\dagger} = \underline{\mathbf{H}}_{\mathcal{T}_l}^{\mathrm{H}} \left(\underline{\mathbf{H}}_{\mathcal{T}_l} \underline{\mathbf{H}}_{\mathcal{T}_l}^{\mathrm{H}}\right)^{-1}. \tag{2.28}$$

Without MISO channels and multiple antennas at the users the construction of the precoding matrix is the same as the BD precoder in [SSH04] in order to keep the zero interference condition in Eq. (2.28). The BD precoder is equivalent to the ZF precoder with MISO channels, see Section 8 in [YG06], and therefore omitted.

Signal to Leakage and Noise Ratio (SLNR)

The objective of SLNR precoding is to maximize the ratio of signal power to user k over interference generated towards other users plus noise. In other words, the SLNR precoder can be seen as a weighted combination of the MRT and ZF approach. In [STS07a] the SLNR of MS k is defined as

$$\gamma_k^{(\mathrm{SLNR})} = \frac{\left\|\mathbf{H}_k \mathbf{p}_t\right\|^2}{\sigma_k^2 + \left\|\underline{\mathbf{H}}_{\mathcal{K}_l \backslash k} \mathbf{p}_t\right\|^2}, \tag{2.29}$$

where $\underline{\mathbf{H}}_{\mathcal{K}_l \backslash k} \in \mathbb{C}^{(K_l - 1)M \times N}$ is the virtual channel of the other users except k such that $\mathcal{K}_l \backslash k$ and is obtained by

$$\underline{\mathbf{H}}_{\mathcal{K}_l \backslash k} = \left[\mathbf{H}_1^{\mathrm{T}} \ \dots \ \mathbf{H}_{k-1}^{\mathrm{T}} \ \mathbf{H}_{k+1}^{\mathrm{T}} \ \dots \ \mathbf{H}_{\mathcal{K}_l}^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{2.30}$$

$\|\underline{\mathbf{H}}_{\mathcal{K}_l \backslash k} \mathbf{p}_t\|^2$ is the sum power of the multiple-user interferences caused towards other MSs by the precoder $\mathbf{p}_t$, the so called leakage. The optimal $\mathbf{p}_t$ to maximize the SLNR in Eq. (2.29) corresponds to [STS07b]

$$\tilde{\mathbf{p}}_t^{(\mathrm{SLNR})} \propto \max \text{ eigenvector}(\mathbf{H}_{\mathcal{K}_l \backslash k}^{(\mathrm{SLNR})}), \tag{2.31}$$

where "max eigenvector" is the eigenvector corresponding to the largest eigenvalue and

$$\mathbf{H}_{\mathcal{K}_l \backslash k}^{(\mathrm{SLNR})} = \left(\mathbf{I}_N \sigma_k^2 + \underline{\mathbf{H}}_{\mathcal{K}_l \backslash k}^H \underline{\mathbf{H}}_{\mathcal{K}_l \backslash k}\right)^{-1} \mathbf{H}_k^H \mathbf{H}_k. \tag{2.32}$$

With MISO channels according to Eq. (2.18) the virtual channel in Eq. (2.30) changes to

$$\underline{\mathbf{H}}_{\mathcal{T}_l \backslash t} = \left[\mathbf{h}_1^{(\mathrm{MISO})} \ \dots, \mathbf{h}_{t-1}^{(\mathrm{MISO})} \ \mathbf{h}_{t+1}^{(\mathrm{MISO})} \ \dots \ \mathbf{h}_{T_l}^{(\mathrm{MISO})}\right] \tag{2.33}$$

of size $[T_l \times N]$ and $\mathbf{H}_k$ is replaced with $\mathbf{h}_t^{(\mathrm{MISO})}$ in (2.32) such that

$$\mathbf{p}_t^{(\mathrm{SLNR})} \propto \max \text{ eigenvector} \left(\left(\mathbf{I}_N \sigma_k^2 + \underline{\mathbf{H}}_{\mathcal{T}_l \backslash t}^H \underline{\mathbf{H}}_{\mathcal{T}_l \backslash t}\right)^{-1} \left(\mathbf{h}_t^{(\mathrm{MISO})}\right)^{\mathrm{H}} \mathbf{h}_t^{(\mathrm{MISO})}\right). \tag{2.34}$$

Consequently, the multiple-user SLNR precoder is obtained by

$$\mathbf{P}_{\mathcal{T}_l}^{(\mathrm{SLNR})} = \left[\mathbf{p}_1^{(\mathrm{SLNR})} \ \dots \ \mathbf{p}_{\mathcal{T}_l}^{(\mathrm{SLNR})}\right]. \tag{2.35}$$

Minimum Mean Square Error (MMSE)

The minimum mean square error (MMSE) precoder [PHS05], also referred to as "regularized" ZF, can be interpreted as an extension of the ZF precoder by an additional regularization summand in the inversion term of Eq. (2.28). The MMSE precoder is given by

$$\mathbf{P}_{\mathcal{T}_l}^{(\mathrm{MMSE})} = \underline{\mathbf{H}}_{\mathcal{T}_l}^{\mathrm{H}} \left(\underline{\mathbf{H}}_{\mathcal{T}_l} \underline{\mathbf{H}}_{\mathcal{T}_l}^{\mathrm{H}} + \mathbf{R}^{(\mathrm{MMSE})} \right)^{-1}, \tag{2.36}$$

where $\mathbf{R}^{(\mathrm{MMSE})} \in \mathbb{C}^{\mathcal{T}_l \times \mathcal{T}_l}$ is a diagonal matrix with elements $\left[\mathbf{R}^{(\mathrm{MMSE})}\right]_{t,t} = z_k^{(\mathrm{IF},i)}$, where $z_k^{(\mathrm{IF},i)}$ is the estimated interference power term at MS k. The knowledge about the interference situation at MSs k selected on stream t depends on side constraints such as the availability of signals to measure the interference or enough uplink resources for the corresponding feedback. Therefore, different levels of information about the interference situation at the MSs are considered due to index i by

$$z_k^{(\mathrm{IF},i)} = \begin{cases} \sigma_k^2 & , \text{if } i = 0 \\ \frac{\left\| \mathbf{z}_k^{\mathrm{H}} \mathbf{z}_k \right\|^2}{M} + \sigma_k^2 & , \text{if } i = 1 \\ \mathbb{E}\left(\frac{\left\| \mathbf{z}_k^{\mathrm{H}} \mathbf{z}_k \right\|^2}{M} \right) + \sigma_k^2 & , \text{if } i = 2. \end{cases} \tag{2.37}$$

In Eq. (2.37) $i = 0$ corresponds to no interference knowledge, $i = 1$ means perfect and complete inter-cell interference knowledge by taking into account $\mathbf{z}_k$ from (2.2) that comprises the interference from other BSs, and $i = 3$ corresponds to statistical knowledge about the interference by the expectation of the interference covariance $\mathbf{z}_k^{\mathrm{H}} \mathbf{z}_k$, e.g. obtained by averaging over RBs in frequency or time. Note that $\mathbf{z}_k$ is defined in (2.2) and represents interference from other BSs than the serving BS l at user k served by stream t, so the index l in $z_{k,t}^{(\mathrm{IF},i)}$ is not necessary and omitted intentionally. The abstraction levels in Eq. (2.37) reflect the trend in 3GPP standardization to design reference signals or muting patterns for interference measurement [3GP17g]. Interference knowledge is essential for accurate MCS adaptation, multiple-user scheduling, and precoder design. By multiplexing another device to already selected devices in the spatial domain, it has to be ensured that the sum spectral efficiency increases. For example adding a device experiencing high interference can result in zero rate for this device and in parallel it will lower the SINR of all the other multiplexed devices by splitting the transmit power to more devices. Independent of TDD or FDD downlink interference can only be measured at the mobile user. Therefore, it has to be transmitted in the uplink channel from MSs to the BS and is essential for precoder design of multiple-user systems.

Remark. *The default value for $z_k^{(\mathrm{IF},i)}$ is $i = 0$ if not explicitly mentioned otherwise.*

2.3.2. Power Constraints

Beside the already mentioned sum power constraint $P^{(\mathrm{RB})} \geq \mathrm{trace}(\mathbf{F}_l)$ per OFDM block, introduced in Section 2.2 below (2.1), additional power constraints are taken into account and presented in this section. The "problem" of optimal power allocation in wireless multiple-user communication systems is a cross-layer problem coupled with user (channel) selection and precoder design [LSS06]. In cellular communications systems practical constraints can "overrule" optimal solutions achieved under "ideal/perfect" system assumptions, e.g. in the LTE-Advanced (LTE-A) standard [3GP16c] each data stream is transmitted with identical power. The main reason for this equal power per stream assumption is peak-to-average power ratio (PAPR) reduction, because the maximum PAPR defines a requirement for radio frequency amplifiers which are the main driver of operational expenditure cost at the BSs [LGP12]. Another reason for equal power per stream is to ensure a "predictable" interference situation for receivers for reasons given in Section 2.3.1. In order to achieve equal transmit power per stream, first the transmit power $P^{(\mathrm{RB})}$ is equally divided by

the number of streams T_l

$$[\mathbf{F}_l]_{t,t} = \frac{P^{(\mathrm{RB})}}{T_l} \tag{2.38}$$

and second each column of $\mathbf{V}_l$ is scaled with the corresponding column norm such that

$$\left[\mathbf{V}_l^{(\mathrm{EPS})}\right]_{:,t} = [\mathbf{V}_l]_{:,t} \frac{1}{\sqrt{\sum_{n=1}^{N} \left|[\mathbf{V}_l]_{n,t}\right|^2 T_l}}, \tag{2.39}$$

where $\mathbf{V}_l^{(\mathrm{EPS})}$ denotes the precoding matrix with equal power per stream. If equal power per beam is considered, then $\mathbf{V}_l$ in Eq. (2.1) and following equations is replaced by $\mathbf{V}_l^{(\mathrm{EPS})}$. Additionally, a per antenna power constraint (PAPC) is considered, which means that the total power transmitted by each antenna is constraint by a maximum, again for cost reasons due to radio frequency amplifiers. This is incorporated by the scaling factor $s_{(\mathrm{PAPC})}$ such that

$$s_{(\mathrm{PAPC})} = \begin{cases} 1 & ,\text{if } n_{(\mathrm{r,max})}^2 \leq P_{(\mathrm{PAPC})} \\ \frac{P_{(\mathrm{PAPC})}}{n_{(\mathrm{r,max})}^2} & ,\text{if } n_{(\mathrm{r,max})}^2 > P_{(\mathrm{PAPC})}, \end{cases} \tag{2.40}$$

where $n_{(\mathrm{r,max})}$ is the largest row norm according to

$$n_{(\mathrm{r,max})} = \sqrt{\left([\mathbf{F}_l]_{t,t}\right)} \max_{n\in[1,\ldots,N]} \left(\sum_{t=1}^{T_l} \left|[\mathbf{V}]_{t,n}\right|^2 \right). \tag{2.41}$$

If the PAPC assumption is considered then $\mathbf{F}_l$ in Eq. (2.1) and following equations is replaced by $s_{(\mathrm{PAPC})}\mathbf{F}_l$. A row in the precoder matrix $\mathbf{V}_l$ corresponds to an antenna while columns correspond to spatial streams, therefore the PAPC constraint results in only one BS antenna transmitting with maximum power since the complete power matrix $\mathbf{F}_l$ is scaled with $s_{(\mathrm{PAPC})}$. Hence, the remaining $N-1$ antennas transmit with less power than $\frac{P^{(\mathrm{RB})}}{N}$.

In order to measure the power distribution over the antennas the row-norm distance with respect to the precoder is introduced according to

$$d_{(\mathrm{spread})}^{(\mathrm{row-norm})} = n_{(\mathrm{r,max})} - n_{(\mathrm{r,min})}, \tag{2.42}$$

where $n_{(\mathrm{r,max})}$ is defined in Eq. (2.41) and $n_{(\mathrm{r,min})}$ is the minimum row norm given by

$$n_{(\mathrm{r,min})} = \sqrt{[\mathbf{F}_l]_{t,t}} \min_{t\in\mathcal{T}_l} \left(\sum_{n=1}^{N} \left|[\mathbf{V}]_{t,n}\right|^2 \right). \tag{2.43}$$

In one extreme case all the power is transmitted by a single antennas resulting in $d_{(\mathrm{spread})}^{(\mathrm{row-norm})} = n_{(\mathrm{r,max})} = 1$ and $n_{(\mathrm{r,min})} = 0$ or in the other all antennas transmit with the same power such that $n_{(\mathrm{r,min})} = n_{(\mathrm{r,max})}$ resulting in $d_{(\mathrm{spread})}^{(\mathrm{row-norm})} = 0$. A spread close to zero is desired in practical system because it corresponds to a small dynamic range of the amplifiers making them cheap.

2.3.3. Numerical Evaluation of Massive MIMO Effects

In this work, no approximation or large system assumption is considered due to the practical number of antennas considered and to show that even then a part of the massive MIMO gains can be utilized. Especially network vendors and operators are interested in cost-efficient operating points, meaning the trade-off between more performance on one side and more capital expenditure and operational expenditure

costs on the other side. This so-called "sweet spot" may not be in the large system regime. Another reason is that in realistic wireless systems, MSs do not have the same SNR as considered in a lot of massive MIMO literature [HtBD13, RPL+13, ANAC13], instead the opposite is the case. The SNR of MSs depends on the effective path loss[2] and other factors resulting in a large variance in SNR, especially in cellular outdoor systems with cell radii in the range of several hundred meters. Additional different multiple-user and inter-cell interference received at each MS further increases the variance in SINR. Also, the definition of SNR is not clear in a multiple-user system, e.g. by adding an additional user to a set of already scheduled users, assuming constant transmit power, the SNR of all the users decreases. Due to the availability of a high performance computation cluster and "massive storage", the numerical result in this work are obtained by processing of all channel coefficients without any relaxation.

In system level simulations the abstraction level is SINR and the task is to evaluate the performance of new techniques under various conditions, e.g. to ensure that algorithms work not only under favorable channel conditions. Therefore, the placement of MSs is subject to a random process over the Monte-Carlo simulations runs, also referred to as "drops" due to the MSs "dropping" in a certain region. In the example of Fig. 2.2a $K = 500$ MSs are "dropped" in the "virtual sector" region obtained from the hexagonal deployment with 300 m inter-site distance (ISD) [3GP17e]. In order to allow comparison of results obtained with system-level simulation between different entities most of the parameters are taken from system level simulation assumptions in 3GPP standardization.

In Fig. 2.2b the spectral efficiency for $T_l = 2$ layers over the number of antennas N is given and the expected logarithmic increase is observed for MRT over the whole antenna range and for the other precoders until 50 antennas. The saturation of the other precoders is caused by the bounded SINR according to (2.9) reflecting limitations of realistic systems due to modulation and coding. Thus, also the difference between MRT and ZF precoder decreases, however in the practical regime a significant gap remains. The performance of ZF, SLNR, and MMSE precoder is similar and the curves overlap in Fig. 2.2b. The maximum power per antenna in the precoder is limited by the PAPC $P_{(\mathrm{PAPC})} = \frac{P^{(\mathrm{RB})}}{N}$ resulting in a power normalization loss shown in Fig. 2.3a for $N = 2$ with approximately the same behaviour for all precoders. The power normalization loss is caused by the maximum allowed transmit power for each antenna as explained in Eq. (2.39) and the subsequent paragraph, see also Section 3.2 in [MGZF13]. Thus, the [5, 50, 95] %-ile of the MMSE precoder as a representative is given in Fig. 2.3b. The power normalization loss due to the PAPC is increasing up to more than 70 % for $N \geq 100$ with a variance < 0.02.

The comparison between PAPC and without PAPC in terms of median sum spectral efficiency $\tilde{C}^{(\mathrm{sum})}$ is given in Table 2.2 where the corresponding result without PAPC is shown in Fig. 2.4a and loss due to the PAPC constraint is ≤ 0.5 bit/s/Hz. Note that according to Eq. (2.9) for $T_l = 2$ the theoretical maximum of $\tilde{C}^{(\mathrm{sum})} \approx 26.6$ bit/s/Hz, but in Table 2.2 the sum spectral efficiency saturates at 24.4 bit/s/Hz. This is caused by the cyclic prefix overhead of 7 % taken into account as in LTE OFDM systems.

The corresponding cumulative distribution function (CDF) of the row-norm spread according to Eq. (2.42) that can be interpreted as the power transmitted by the antennas assuming a power budget of one is shown in Fig. 2.4b for $N = 2$ and $N = 300$ antennas. With $N = 2$ the maximum row-norm spread $d_{(\mathrm{spread})}^{(\mathrm{row-norm})} \approx 1$, which means that in some cases all transmit power is radiated by one of the antennas, so both amplifiers have to be designed for the maximum power budget, whereas with $N = 300$ the maximum row-norm spread $d_{(\mathrm{spread})}^{(\mathrm{row-norm})} = 0.018$. This means, that the larger the number of antennas the higher the probability that antennas transmit approximately with equal power and the PAPC as required in low-antenna systems can be dropped. In both cases, the four investigated precoder schemes yield approximately the same performance, therefore the results over the number of antennas N are shown for MMSE precoder only in Fig. 2.5.

Up to $N = 300$ antennas are considered in the above investigations which results in approximately 22.43 m aperture size assuming $\frac{\lambda_{(\mathrm{c})}}{2}$ spacing between elements. It is obvious that such antenna arrays can hardly be deployed in dense urban environments where space is precious and expensive. Another disadvantage of

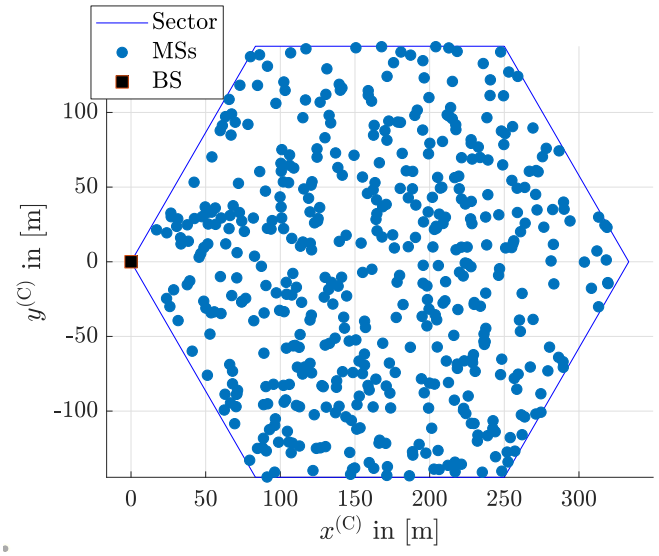
[2] Effective path loss means distance depended path loss plus effects from antenna patterns and multipath propagation.

Table 2.1.: Simulation assumptions for numerical evaluation of downlink massive MIMO in TDD.

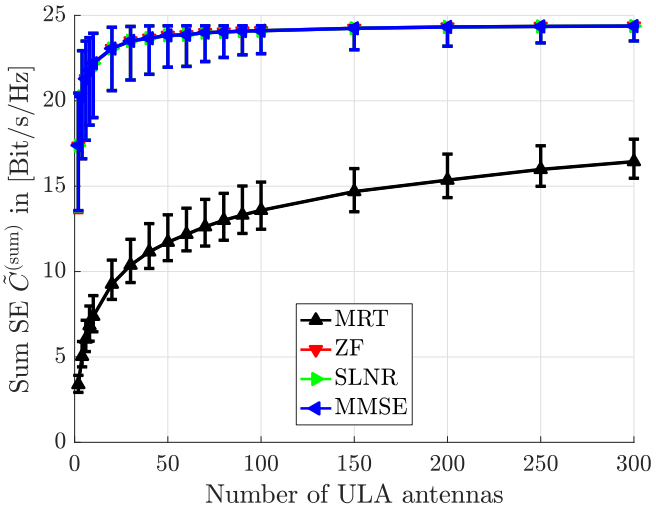
Parameter	**Value**
Simulation type	Monte Carlo 500 realizations
Channel model	QuaDRiGa version 2.0 [JRB+17]
Scenario	3GPP 3D Urban Macro NLoS, [3GP17e]
Center frequency	4 GHz
Number of multi-path components $L_{(\mathrm{MPC})}$	21
OFDM RB Bandwidth	180 kHz
Number of RBs	50
Utilized bandwidth	9 MHz
Cyclic prefix overhead	7 %
BS antenna distribution	Horizontal ULA or UPA
Number of BS antenna elements N	$[2, 300]$
Antenna element spacing	$\lambda_{(\mathrm{c})}/2$
Antenna element type	Patch
Horizontal, vertical HPBW	$65°, 65°$
Element directive gain	≈ 9.4 dBi
BS height	25 m
Total transmit power	40 dBm
SINR bounds, $\gamma^{(\min)}, \gamma^{(\max)}$	$\gamma^{(\min)} = -5$ dB , $\gamma^{(\max)} = 40$ dB
$x^{(\mathrm{C})} - y^{(\mathrm{C})}$ MS distribution	Random uniform i.i.d. in hexagonal layout [3GP17e]
Number of users per hexagonal sector	20
$z^{(\mathrm{C})}$ coordinate of MSs	1.5 m
Minimum distance MS-BS	25 m

Table 2.2.: Median sum spectral efficiency comparison with and without PAPC. The numbers correspond to Fig. 2.2b and Fig. 2.4a. MMSE precoder.

	N	2	10	50	100	300
with PAPC	$\tilde{C}^{(\mathrm{sum})}$ in [bit/s/Hz]	17.4	22.2	23.8	24.1	24.4
without PAPC		17.9	23.2	24.3	24.4	24.4

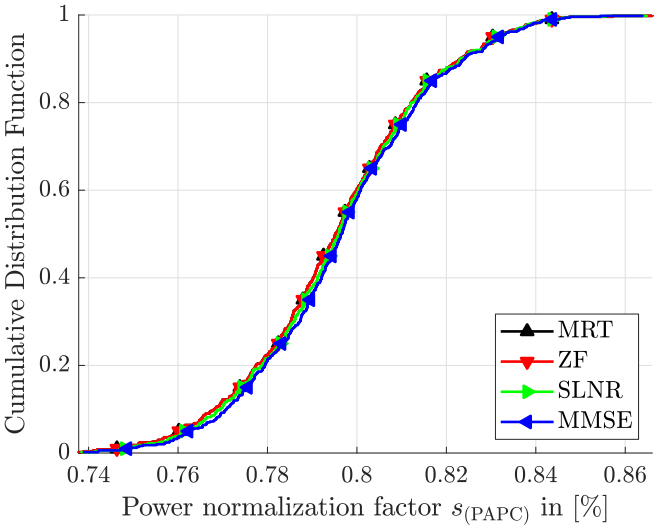


(a) Distribution of MSs in $x^{(C)}$-$y^{(C)}$ plane with the virtual sector boarder following 3GPP assumptions. Example with 500 MSs.

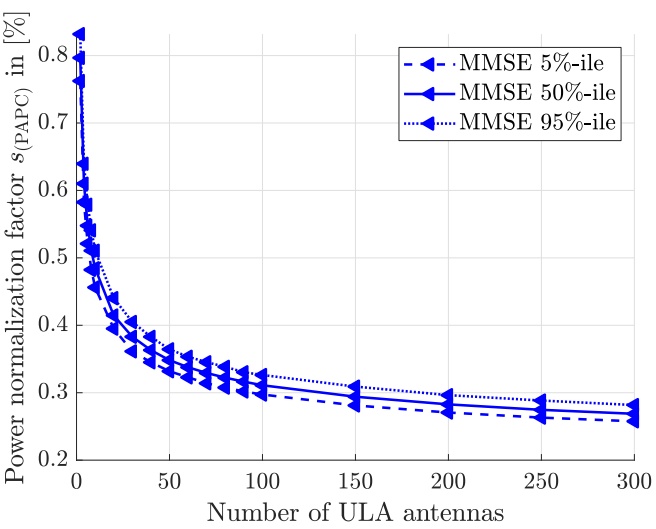


(b) Sum spectral efficiency (SE) for $T_l = 2$ over number of antennas N for horizontal uniform linear array (ULA) with PAPC constraint $P_{(\mathrm{PAPC})} = \frac{P^{(\mathrm{RB})}}{N}$. The curves for MMSE, SLNR, and ZF overlap.

Figure 2.2.: Massive MIMO downlink deployment and performance evaluation.

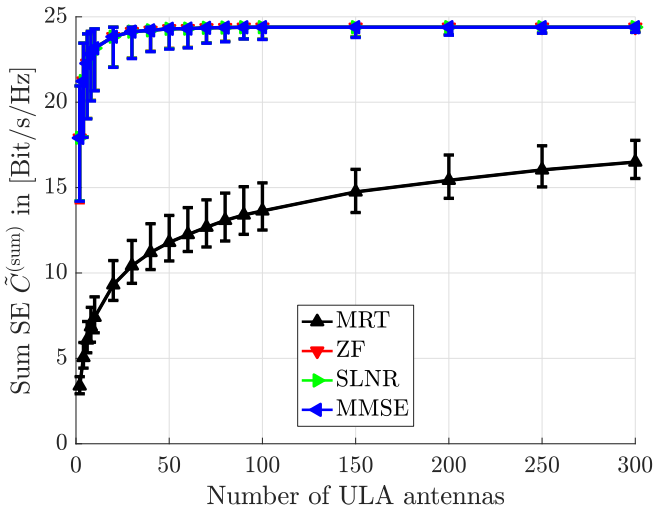


(a) Distribution of power normalization loss $s_{(\mathrm{PAPC})}$ in [%] according to (2.40) in the non-massive MIMO regime for $N = 2$ antennas. The curves for MMSE, MRT, SLNR, and ZF overlap.

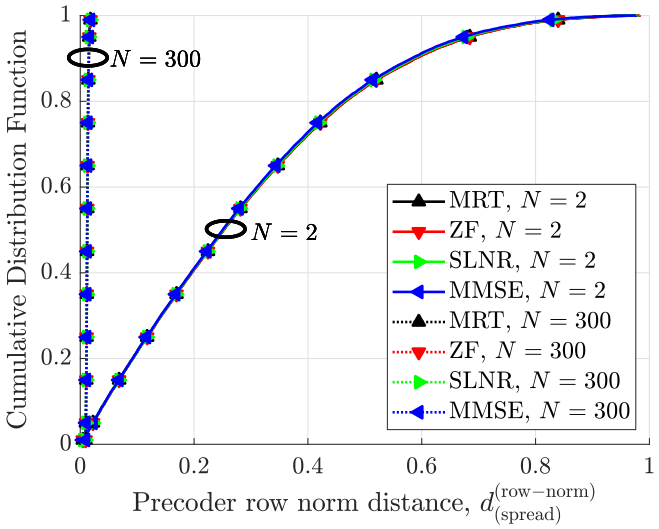


(b) Power normalization factor $s_{(\mathrm{PAPC})}$ in [%] according to (2.40) over antennas for the MMSE precoder.

Figure 2.3.: Power normalization factor due to PAPC constraint with $P_{(\mathrm{PAPC})} = \frac{P^{(\mathrm{RB})}}{N}$.



(a) Sum spectral efficiency (SE) for $T_l = 2$ over number of antennas N for horizontal ULA without PAPC. The curves for MMSE, SLNR, and ZF overlap.



(b) Cumulative distribution of the row-norm spread according to Eq. (2.42) for $N = 2$ and $N = 300$ antennas. The curves for MMSE, SLNR, and ZF overlap.

Figure 2.4.: Performance without PAPC constraint.

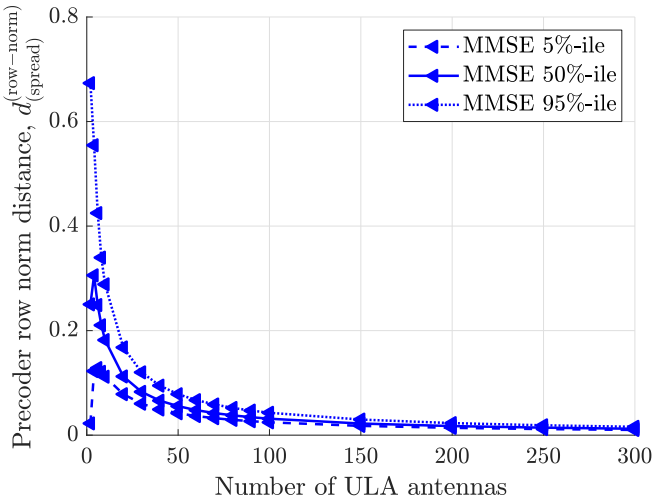
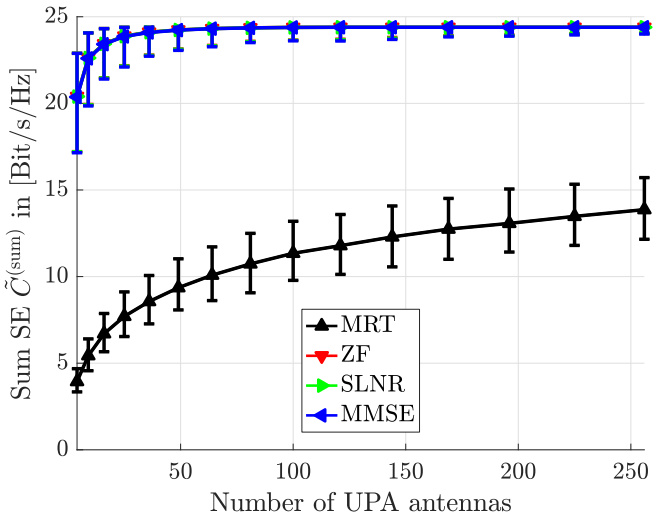
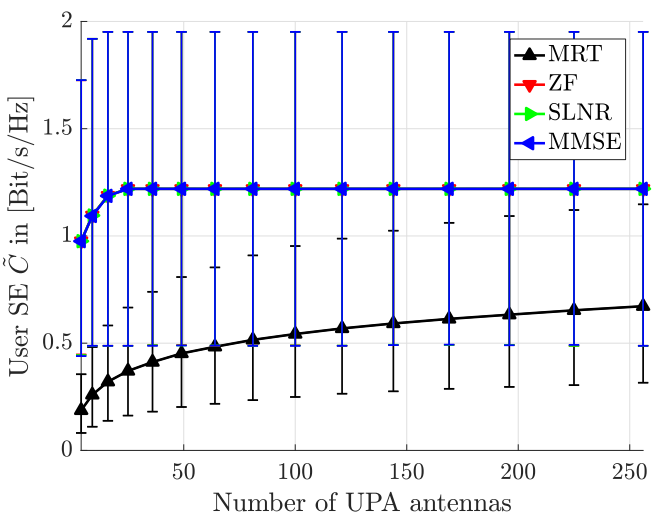


Figure 2.5.: Precoder row-power spread according to Eq. (2.42) over the number of antennas N for horizontal ULA and $T_l = 2$.



(a) Sum spectral efficiency (SE).

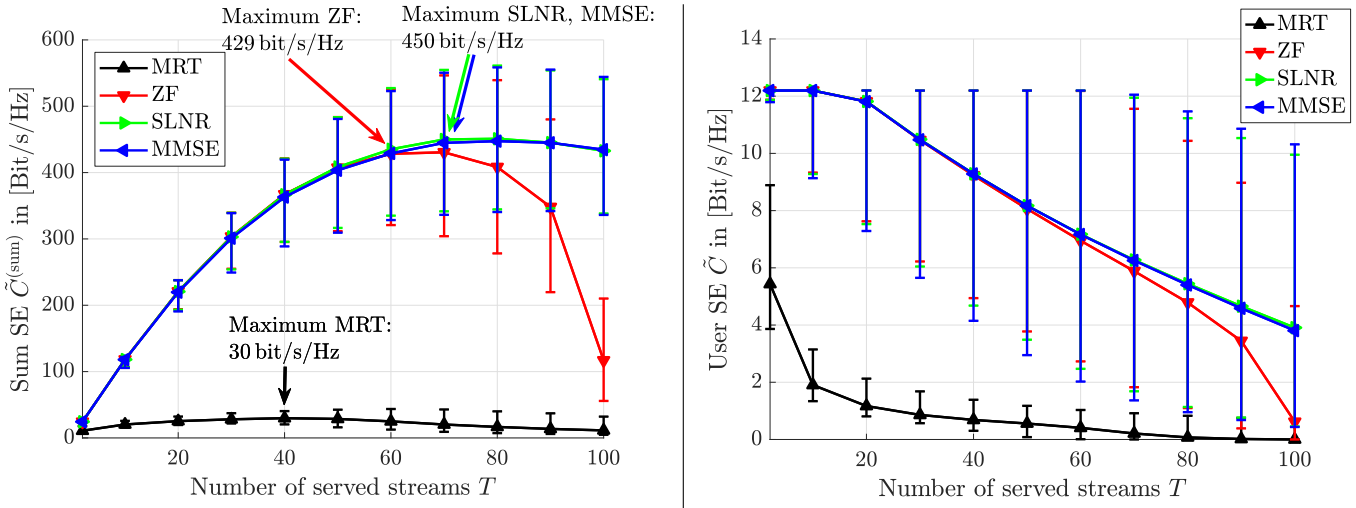


(b) User spectral efficiency (SE). The large variance is caused by the equal power per stream assumption in combination with a large variance in path loss among users due to random dropping within the sector.

Figure 2.6.: Performance of downlink precoding for $T_l = 2$ over the number of antennas $N = N_{(\alpha)}N_{(\beta)}$ with $N_{(\alpha)} = N_{(\beta)}$ for a UPA without PAPC constraint. The curves of MMSE, SLNR, and ZF overlap.

ULAs is that they have only one degree of freedom (DoF) for beam steering, however users in cities are not only distributed in the $x^{(\mathrm{C})}$-$y^{(\mathrm{C})}$ plane also in elevation e.g. due large buildings or non-flat surface. Therefore, standardization adopted 2D uniform planar arrays in [3GP15a] to allow for 3D also called full-dimension beamforming. An example of a UPA is shown in Fig. 3.3, where the number of elements in horizontal/azimuth domain is defined by $N_{(\alpha)}$, the number of elements in vertical/elevation domain as $N_{(\beta)}$, such that $N = N_{(\alpha)}N_{(\beta)}$. The sum spectral efficiency over the number of antennas is given in Fig. 2.6a and the trends are the same as for the ULA above. In the corresponding user spectral efficiency in Fig. 2.6b a large variance can be observed due to the equal power per stream but large variance in user path losses. This is caused by the random dropping of users within the sector according to Fig. 2.2a. However, it can be observed that sum spectral efficiency gain of the UPA is similar compared to the ULA, see Fig. 2.4a and Fig. 2.6. In combination with the advantage that the maximum aperture size of the UPA is reduced by a factor of $\sqrt{N}$ compared to the ULA the focus in the remaining part of the chapter is on UPA. For example with $N = 100$ the edge length of the UPA is $\approx 0.68\,\mathrm{m}$ compared to $7.5\,\mathrm{m}$ of the ULA. The UPA assumption also follows current assumption in 5G 3GPP standardization, see Section A.2 in [3GP17g].

The main motivation to use massive MIMO is the achievable spatial multiplexing gain, e.g. in areas with a high user density. By utilizing the DoFs and with a full rank channel the sum spectral efficiency scales linearly with the number of spatial multiplexed users or corresponding in this work to the number of effective MISO channels T_l [RPL+13]. With an unlimited number of antennas $\lim N \to \infty$ the channel vectors of users become orthogonal, then MRT precoding also achieves maximum capacity. However, due to the limited number of antennas in practical systems, this property is not fulfilled and the sum spectral efficiency of MRT precoding is significantly less than the other precoders, see Fig. 2.7a. Therein, MRT achieves a maximum of $30\,\mathrm{bit/s/Hz}$ at $T = 40$ which corresponds to a multiplexing gain of ≈ 3 compared to $T = 2$, where $11\,\mathrm{bit/s/Hz}$ are achieved. The number of devices that can be spatially multiplexed depends on the spatial correlation and user SNRs. Spatial correlation means that due to the non-orthogonality of channels inter-user interference is generated which limits the spectral efficiency, see Eq. (2.2) and Eq. (2.3). Regarding the SNR, independent of channel orthogonality among users, the more users are served the lower the power per beam, see Eq. (2.15). Even in the massive MIMO regime at $N \gg T$, splitting the power to two users results in the optimal case to a SINR reduction of $3\,\mathrm{dB}$ compared to the case that one user is served. Gain in sum rate, e.g. for the 2 users case, is only achieved on condition $\log_2(1+\gamma_1) > \log_2(1+\gamma_1/2)+\log_2(1+\gamma_2)$, where



(a) Sum spectral efficiency (SE).

(b) User spectral efficiency (SE).

Figure 2.7.: Performance of spatial multiplexing with a 10×10 UPA over the number of served streams T_l. The curves of MMSE, SLNR, and ZF partly overlap.

γ_1 and γ_2 are the SINRs of user 1 and user 2, respectively, according to Eq. (2.3). In contrast to this the sum spectral efficiency of ZF has a maximum at $T = 60$ with a gain of ≈ 18 compared to $T = 2$. A further increase of T results in a significant sum spectral efficiency loss due to noise enhancement and normalization loss caused by the zero interference constraint in Eq. (2.26), see [KRTT13]. While the sum spectral efficiency of SLNR and MMSE overlap with ZF up to $T = 60$ in Fig. 2.6a, their maximum is achieved at $T = 70$ and remains approximately constant up to $T = 100$. The reason for that is, that in both precoders the noise variance is taken into account as a regularization factor, see Eq. (2.32) and Eq. (2.36). Note that in the single BS the interference term for MMSE in Eq. (2.37) is zero. The corresponding user spectral efficiency is given in Fig. 2.7 and it can be observed that due to the non-orthogonal user channels SLNR and MMSE scale with $\propto \frac{1}{K}$. For the study in Fig. 2.7 the assumption was that $\tilde{K}_l = K_l$ meaning that all users connected to the BS that request downlink data are served and due to the random user placement this corresponds to round robin scheduling. All the precoders in Fig. 2.7 have in common that adding more users to the system decreases the sum spectral efficiency at some point. This result also shows that in practical systems user scheduling is still required for massive MIMO. The expected complexity increase of user grouping due to large number of antennas and users in the system can be compensated according to [BLM16] by the frequency flat effective channels of massive MIMO also called channel hardening. This means that the scheduling has to be performed only once for a given time slot and can be applied to all RBs in the frequency domain. In contrast to this in LTE systems with a lower number of antennas channels are highly frequency selective and thus scheduling has to be performed on a RBs or group of RBs level [STW+09]. Therefore, the impact by user grouping and the assumption of frequency flat channels is investigated in Section 2.3.4.

From the results in this section, it can be concluded that requirements on power amplifiers relax with an increasing number of antennas due to decreased power variance over antennas. It is also shown that with a low number of served streams, e.g. $T \ll N$ most of the spectral efficiency is already achieved with a moderate number of antennas, e.g. in Fig. 2.4, 90 % of the performance with $N = 300$ is already achieved at $N = 10$. Thus, the deployment of massive MIMO has to be considered carefully. From a sum spectral efficiency perspective the best operating point of massive MIMO is with a high number of users to select users with orthogonal channels [BLD16], where "high" means more than half the antenna number. As a drawback, user selection is required in order to achieve the maximum sum throughput.

2.3.4. User Grouping

It is shown in the previous paragraph that with the number of users or MISO channels being in the regime $T \geq \frac{N}{2}$ downlink performance estimation for user scheduling is required. The main challenge is to ensure that additional users increase the objective KPI. Thereby, the interference situation at the user, consisting of multiple users and multiple-cell interference, has to be taken into account by the scheduling entity. In this paragraph, the focus is on a single BS so only inter-user interference from the same BS is taken into account. The impact from inter-cell interference is discussed in Section 4.1. In the system model described in Section 2.2, the MISO channels are introduced for users with $M > 1$ antennas. The resulting $M\tilde{K}$ MISO channels can be used as $M\tilde{K}$ single antenna users. Thus the set of "available" MISO channels is denoted as $\tilde{\mathcal{T}}$. The set $\tilde{\mathcal{T}}$ of size $M\tilde{K}$ is considered for user selection.

Remark. *Note that each MISO channel* $\mathbf{h}_{\tilde{t},l}^{(\mathrm{MISO})}$*, where* $\tilde{t} \in \tilde{\mathcal{T}}$*, corresponds to a certain* $\mathbf{h}_{m,k,l}^{(\mathrm{MISO})}$*. Therefore, in the following it is assumed that any stream* $\tilde{t}$ *corresponds one-to-one to a specific user* k *and his* m*-th MISO feedback.*

With linear precoding the maximum possible number of selected streams $T^{(\max)}$ is the minimum of the number of antennas N or available streams $\tilde{T}$ such that

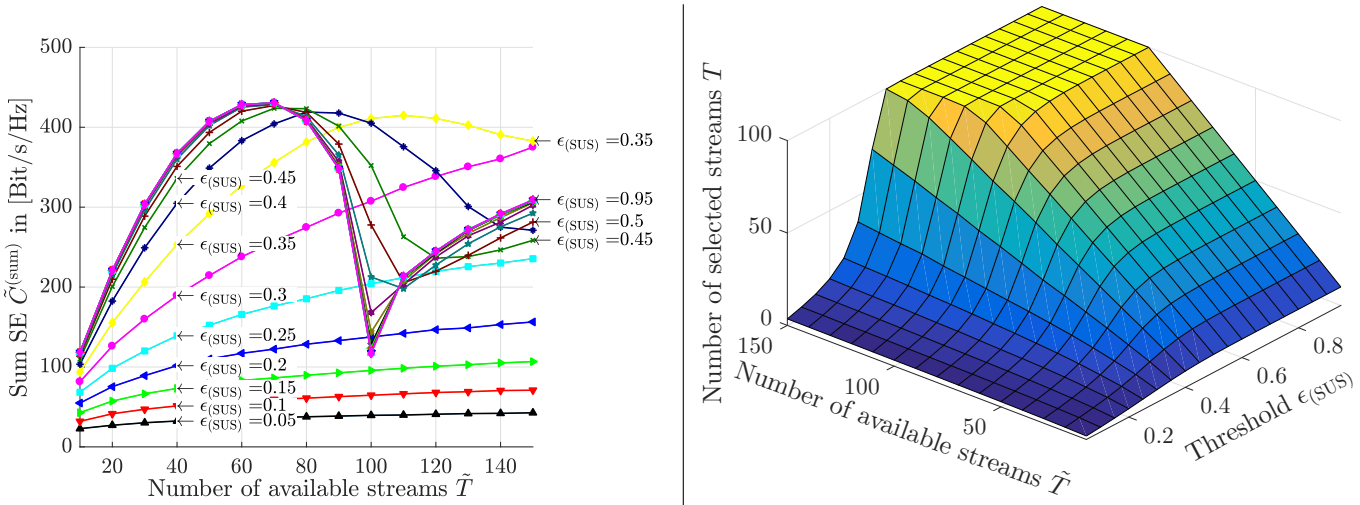
$$T^{(\max)} = \min\left(N, \tilde{T}\right). \tag{2.44}$$

In order to obtain the optimal stream selection for a given number of selected streams T, all possible combinations of length T out of the set $\tilde{\mathcal{T}}$ have to be tried. Therefore, the optimization is a combinatorial problem with complexity

$$\mathcal{O} = \sum_{T=1}^{T^{(\max)}} \binom{\tilde{T}}{T} = \sum_{T=1}^{T^{(\max)}} \frac{\tilde{T}!}{T!\left(\tilde{T}-T\right)}. \tag{2.45}$$

Considering the example of $N = 64$ antennas and $\tilde{T} = 20$ MISO channels this results in $\mathcal{O} = 1.048.575$. Additionally, the complexity of the multiple-user precoder $\mathbf{P}$ computation scales with N and is required for each combination. Thus, the optimum solution of user grouping is not derived in this thesis. However, as a baseline for user selection the well-known semi-orthogonal user selection (SUS) is used. Semi-orthogonal user selection (SUS) was proposed in [YG06] which shows that SUS combined with ZF precoding achieves the same asymptotic sum rate as that of optimal dirty paper coding if the number of streams $\tilde{T}$ goes to infinity. The SUS algorithm works iteratively in four steps as follows [YG06]:

- Step 1: Initialize the set of available streams as $\tilde{\mathcal{T}}^{(I)} = \tilde{\mathcal{T}}$, the set of selected streams as $\mathcal{T} = \emptyset$, and the counting variable $I = 1$.
- Step 2: Compute for each $\tilde{t} \in \tilde{\mathcal{T}}^{(I)}$ the variable $\mathbf{g}_{\tilde{t}}^{(\mathrm{SUS})} = \left(\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}\right)^{\mathrm{T}} \left(\mathbf{I}_N - \sum_{i=1}^{I-1} \frac{\left(\mathbf{g}_i^{(\mathrm{SUS})}\right)^{\mathrm{H}} \mathbf{g}_i^{(\mathrm{SUS})}}{\left\|\mathbf{g}_i^{(\mathrm{SUS})}\right\|^2}\right)$, that is the channel $\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}$ projected to the null subspace of already scheduled users. Note, that in the first step with $I = 1$ the sum term vanishes such that $\mathbf{g}_{\tilde{t}}^{(\mathrm{SUS})} = \left(\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}\right)^{\mathrm{T}}$.
- Step 3: Select stream $t^{(\mathrm{SUS})}$ such that $\mathbf{g}_{t^{(\mathrm{SUS})}} = \underset{\tilde{t} \in \tilde{\mathcal{T}}^{(I)}}{\arg\max} \left\|\mathbf{g}_{\tilde{t}}^{(\mathrm{SUS})}\right\|$, $\mathcal{T} \leftarrow \mathcal{T} \cup t^{(\mathrm{SUS})}$, and $\mathbf{g}_I^{(\mathrm{SUS})} = \mathbf{g}_{t^{(\mathrm{SUS})}}^{(\mathrm{SUS})}$.
- Step 4: If $|\mathcal{T}| < N$, compute the set of streams $\tilde{\mathcal{T}}^{(I+1)}$ that fulfils the orthogonality condition to already scheduled users by $\tilde{\mathcal{T}}^{(I+1)} = \left\{ \tilde{t} \in \tilde{\mathcal{T}}^{(I)}, \tilde{t} \neq t^{(\mathrm{SUS})}, \forall \frac{\left\|\left(\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}\right)^{\mathrm{T}} \left(\mathbf{g}_I^{(\mathrm{SUS})}\right)^{\mathrm{H}}\right\|}{\left\|\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}\right\| \left\|\mathbf{g}_I^{(\mathrm{SUS})}\right\|} \leq \epsilon_{(\mathrm{SUS})} \right\}$, otherwise $\tilde{\mathcal{T}}^{(I+1)} = \emptyset$. Increase the iterator: $I \leftarrow I + 1$. Go to step 2 if $\tilde{\mathcal{T}}^{(I+1)} \neq \emptyset$.



(a) 50 %-ile sum spectral efficiency (SE) of SUS algorithm over number of connected streams $\tilde{T}$ and different threshold value $\epsilon_{(\text{SUS})}$.

(b) 50 %-ile number of selected streams T by SUS over the two-dimensional parameter space of available streams $\tilde{K}$ and threshold value $\epsilon_{(\text{SUS})}$.

Figure 2.8.: Performance evaluation of SUS. For visibility the confidence interval is omitted.

Table 2.3.: List of $\epsilon_{(\text{SUS})}$ that achieves the maximum spectral efficiency in Fig. 2.8a for a given number of available users $\tilde{K}_l$.

$\tilde{T}$	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
$\epsilon_{(\text{SUS})}^{(\max)}$	0.9	0.8	0.85	0.75	0.8	0.8	0.8	0.45	0.4	0.35	0.35	0.35	0.35	0.35	0.35

The SUS was designed for the case $\tilde{T} > N$, where the number of streams for downlink transmission is larger than the number of antennas at the BS and it is stated in [YG06] that the threshold value $\epsilon_{(\text{SUS})}$ has to be obtained for each $\mathcal{T}$ to achieve the maximum sum throughput. This is shown in Fig. 2.8a where the 50 %-ile sum throughput is given over the number of available streams $\tilde{T}$ for various $\epsilon_{(\text{SUS})}$. A significant observation is, that the maximum sum spectral efficiency is achieved at $\tilde{T} = 70$ for $\epsilon_{(\text{SUS})} = 0.8$ and an increase of $\tilde{T}$ does not increase the sum spectral efficiency despite the additional stream selection diversity. One reason for this could be the finite set of simulated $\epsilon_{(\text{SUS})}$ indicating the computational complexity required to find optimal $\epsilon_{(\text{SUS})}$. Another and probably the main reason is, that the objective of the SUS is not the optimization of sum spectral efficiency, meaning that the streams with the most orthogonal channel does not necessarily provide a sum spectral efficiency gain, e.g. due to low SNR and splitting of the transmit power to all the scheduled streams. A third reasons is the realistic number of $N = 100$ antennas which prevents the law of large numbers to kick in for channels to become orthogonal, see Marzettas paper [Mar10] discussed in Section 1.1. However, for further reference in this work, the sum spectral efficiency maximizing $\epsilon_{(\text{SUS})}$ according to Fig. 2.8a is used for each $\tilde{T}$, see Table 2.3.

The corresponding number of selected streams T is shown in Fig. 2.8b. It can be observed that the lower $\epsilon_{(\text{SUS})}$ the less streams are selected and in the range of $\epsilon_{(\text{SUS})} > 0.5$ and $\tilde{T} \leq N$ the number of selected streams $T \approx \tilde{T}$. Note that the maximum T is bounded by $T \leq \tilde{T}$.

Additionally, the optimal $\epsilon_{(\text{SUS})}$ may also vary from one environment to another. In practical systems, this would require a lot of training overhead to obtain the optimal $\epsilon_{(\text{SUS})}$. Furthermore, it is stated in [YG06] that in order to achieve optimal performance equal SNR is assumed $\forall \tilde{T}$, which corresponds to an equal path-loss for all users. This is not a valid assumption in real world scenarios where users have different distances to the same BS. Thus, a sum rate aware scheduling algorithm that does not require a training phase to find optimal configuration parameters, and at the same time takes into account stream SNRs is preferred. Note, that SUS doesn't consider fairness among users. In a practical system this can be achieved

by the selection of the right users for input to the SUS algorithm, e.g. omitting users that have been selected in the previous time slot, or in frequency domain on a previous subcarrier. This means fairness would be considered by an over the top algorithm rather than in the SUS itself.

The author of this thesis and others published in [TKBH12, TKOB13] a suitable greedy sum throughput aware user/stream selection scheme called projection based zero forcing (PBZF), similar to SUS, for joint transmission coordinated multi-point (CoMP) systems. It was originally proposed in [FGH06, FGH07] and later used for massive MIMO performance evaluation in interference limited scenarios in [KRTT13, KTH14b]. However, no comparison with a reference user/stream selection is presented to verify the effectiveness of the projection based zero forcing (PBZF) stream selection. Therefore, a modified version of the PBZF algorithm according to [TKOB13] is discussed next and compared with the SUS as reference. Before the PBZF algorithm, given in Algorithm 1 is described in detail the basic principle is explained.

The greedy PBZF algorithm calculates for each user in the set of available users an rate estimates of the sum throughput, assuming the user is added to an set of already scheduled users. The calculation of the rate estimate is achieved with a projection matrix where the null spaces of the scheduled users are multiplied with each other. Then, the user from the set of available users that maximizes and increases the sum throughput is selected and added to the set of already scheduled users. In the next paragraph Algorithm 1 is explained in detail.

As input, the projection order denoted as $p^{(\mathrm{O})}$ and the equivalent MISO channels $\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}, \forall \tilde{t} \in \tilde{\mathcal{T}}$ according to Eq. (2.19) are required. The projection order $p^{(\mathrm{O})}$ is explained later and assumed constant for a given hardware. If the user channel $\mathbf{H}_k$ is given, the equivalent MISO channels can be obtained by a SVDs operation according to Eq. (2.18). With this, the null space of each MISO channel $\tilde{t}$ is computed according to

$$\mathbf{\Psi}_{\tilde{t}} = \mathbf{I}_N - \left(\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}\right)^{\mathrm{H}} \left(\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})} \left(\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}\right)^{\mathrm{H}}\right) \mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}, \tag{2.46}$$

where matrix $\mathbf{\Psi}_{\tilde{t}}$ of size $N \times N$ can be interpreted as the null space projector of the respective MISO channel. Then, a first stream t is selected randomly in Algorithm 1 for initialization, however any other KPI can be used for the initial selection, e.g. the the user with the highest traffic demand. After that, the set of available streams $\tilde{\mathcal{T}}$ and the set of selected streams $\mathcal{T}$ are updated such that $\tilde{\mathcal{T}} = \tilde{\mathcal{T}} \setminus t$ and $\mathcal{T} = t$. Then the sum spectral efficiency $\hat{C}_{\mathcal{T}}^{(\mathrm{sum})}$ of the set of selected streams is set as the reference value denoted by $\hat{C}^{(\mathrm{sum,old})}$. With this, the for loop from 2 to the maximum number of spatial layers $\tilde{T}^{(\max)}$ starts. Remember, that the maximum number of spatial layers is given by $\tilde{T}^{(\max)} = \min(N, \tilde{T})$. As the fist step in the for loop the joint null subspace of already scheduled MISO channels in set $\mathcal{T}$, denoted by $\tilde{\mathbf{\Psi}}_{\mathcal{T}}$, is obtained as

$$\tilde{\mathbf{\Psi}}_{\mathcal{T}} = \left(\mathbf{\Psi}_1 \cdot \mathbf{\Psi}_2 \cdot \ldots \cdot \mathbf{\Psi}_T\right)^{p^{(\mathrm{O})}}, \tag{2.47}$$

by multiplying the null spaces of the respective users obtained in Eq. (2.46) with each other. This product is then taken to the power of $p^{(\mathrm{O})}$ and therefore denoted as the projection order, one of the inputs to the algorithm. Note that $\tilde{\mathbf{\Psi}}_{\mathcal{T}}$ is an approximation of the joint null-subspace of scheduled users in $\mathcal{T}$ and equality is only achieved for $\lim p^{(\mathrm{O})} \to \infty$ [FGH06], but authors in [FGH07] showed that an projection order of $1 \leq p^{(\mathrm{O})} \leq 3$ is already sufficient. These findings are based on numerical simulations and can be different in correlated MIMO channels. Thus, a parameter study on $p^{(\mathrm{O})}$ for the given scenario is performed and shown in Fig. 2.9. Note that analytically the larger $p^{(\mathrm{O})}$ the better the approximation of the null space and the better the sum spectral efficiency [FGH07]. Yet, numerical computation programs[3] are limited in accuracy by the number of bits used for data types. Therefore, the sum spectral efficiency decreases for $p^{(\mathrm{O})} > 1$ contrary to the analytical theorem. The advantage compared to the orthogonality parameter in SUS is that the projection order has to be searched only once for the given computation system where the algorithm is operated, but not a parameter search for each $\tilde{T}$ and scenario is required as for SUS. Therefore,

[3] MATLAB 2016b on a Linux 64 bit system was used in this thesis. The MATLAB command "computer" shows "GLNXA64" as output. MATLAB uses 64 bit for a double precision according to institute of electrical and electronics engineers (IEEE) Standard 754.

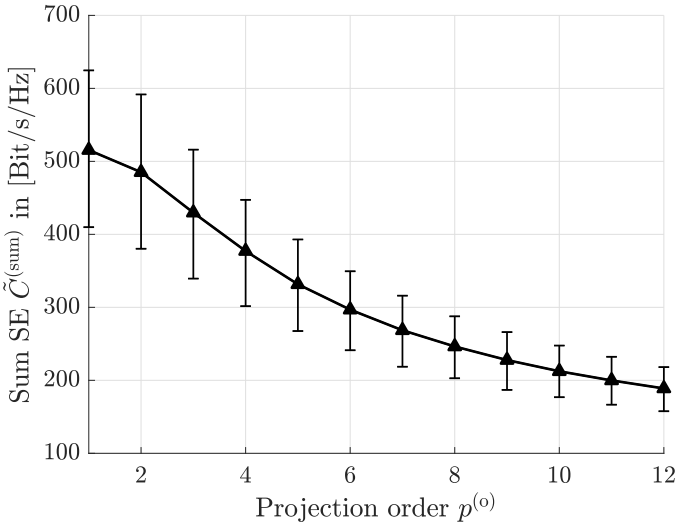


Figure 2.9.: Sum spectral efficiency (SE) over projection order $p^{(\mathrm{O})}$ of the PBZF stream grouping Algorithm 1.

$p^{(\mathrm{O})} = 1$ is selected for the following simulations in this section. In the next step of Algorithm 1, the stream $\check{t}$ that minimizes the norm $\left\|\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}\tilde{\mathbf{\Psi}}_{\mathcal{T}}\right\|$such that $\check{t} \leftarrow \arg\max_{\tilde{t}\in\tilde{\mathcal{T}}}(\left\|\mathbf{h}_{\tilde{t}}^{(\mathrm{MISO})}\tilde{\mathbf{\Psi}}_{\mathcal{T}}\right\|)$ is obtained. This step can be interpreted as selecting the MISO channel $\mathbf{h}_{\check{t}}^{(\mathrm{MISO})}$ with highest orthogonality to the null space of already scheduled users $\tilde{\mathbf{\Psi}}_{\mathcal{T}}$ obtained in Eq. (2.47). Then the sets of selected and available MISO channels are updated by $\mathcal{T} = \mathcal{T} \cup \check{t}$ and $\tilde{\mathcal{T}} = \tilde{\mathcal{T}} \setminus \check{t}$, respectively. A modification compared to [TKOB13] is that a sum spectral efficiency constraint is added in order to ensure that a new stream increases the sum spectral efficiency compared to the previous iteration. Therefore, first the link spectral efficiency of stream $\tilde{t}$ is estimated according to [FGH07] by

$$\hat{C}_{\tilde{t}} = \begin{cases} 0 & ,\text{if } \hat{\gamma}_{\tilde{t}} < \gamma^{(\min)}, \\ \log_2\left(1+\hat{\gamma}_{\tilde{t}}\right) & ,\text{if } \gamma^{(\min)} \leq \hat{\gamma}_{\tilde{t}} \leq \gamma^{(\max)}, \\ \log_2\left(1+\gamma^{(\max)}\right) & ,\text{if } \hat{\gamma}_{\tilde{t}} > \gamma^{(\max)}. \end{cases} \tag{2.48}$$

where $\hat{\gamma}_{\tilde{t}}$ is given by

$$\hat{\gamma}_{\tilde{t}} = \frac{P^{(\mathrm{RB})}}{T\sigma_{\tilde{k}}^2}\|\mathbf{h}_{\tilde{t},l}^{(\mathrm{MISO})}\tilde{\mathbf{\Psi}}_{\mathcal{T}}\|_{\mathrm{F}}^2. \tag{2.49}$$

Note that $\hat{\gamma}_{\tilde{t}}$ can be interpreted as the receive SNR at stream $\tilde{t}$ taking into account the equal power splitting by the current number of selected spatial layers t in the denominator. In Eq. (2.49) the interference due to non-perfect orthogonality by the approximation in Eq. (2.47) with $p^{(\mathrm{O})} < \infty$ is not considered. Consequently, the estimated sum spectral efficiency with SINR bounds is given by

$$\hat{C}_{\mathcal{T}}^{(\mathrm{sum})} = \sum_{t=1}^{T}\hat{C}_t. \tag{2.50}$$

Finally, if the sum rate of set $\mathcal{T}$ including $\check{t}$ in step t' denoted as $\hat{C}^{(\mathrm{sum,new})}$ is larger than the sum rate of step $t'-1$ denoted as $\hat{C}^{(\mathrm{sum,old})}$ such that $\hat{C}^{(\mathrm{sum,new})} > \hat{C}^{(\mathrm{sum,old})}$, the MISO channel $\check{t}$ is accepted. Then the sum rate is updated by $\hat{C}^{(\mathrm{sum,old})} = \hat{C}^{(\mathrm{sum,new})}$ and the for loop proceeds with the next step. On the other hand, if $\hat{C}^{(\mathrm{sum,new})} \leq \hat{C}^{(\mathrm{sum,old})}$ the for loop stops and $\check{t}$ is removed from the set of selected streams by $\mathcal{T} = \mathcal{T} \setminus \check{t}$. The for loop also stops if $\tilde{T}^{(\max)}$ according to Eq. (2.44) is reached.

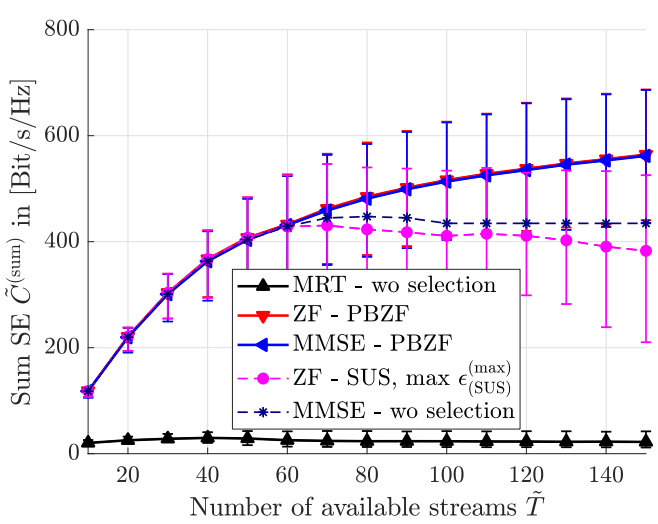
Algorithm 1 Modified projection based zero forcing (PBZF) stream selection algorithm based on [TKOB13].

Require: Equivalent MISO channel $\mathbf{h}_{\tilde{t},l}^{(\mathrm{MISO})}, \forall \tilde{t} \in \tilde{\mathcal{T}}$, projection order $p^{(\mathrm{O})}$, $T^{(\max)}$

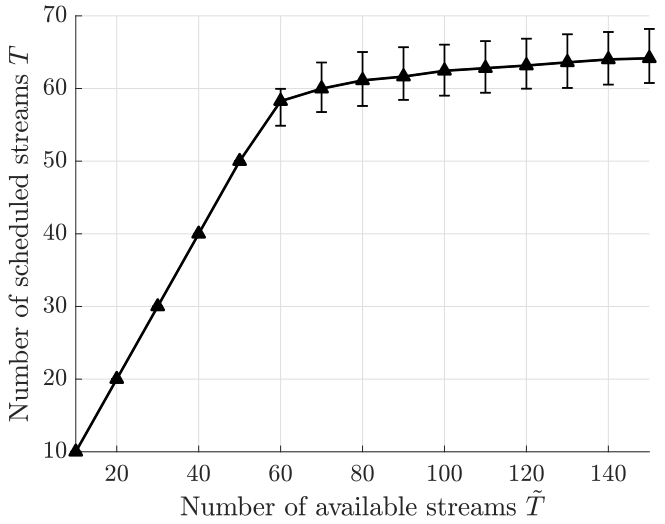
```
Select a start stream: t ∈ T̃
Update available streams: T̃ = T̃ \ t
Update selected streams: T = t
Compute projectors Ψ_t̃, ∀t̃ ∈ T̃ according to Eq. (2.46)
Compute Ĉ^(sum,old) = Ĉ_t̃ according to Eq. (2.48)
for t' = 2 : T̃^(max) do
    Compute the projection matrix Ψ̃_T according to Eq. (2.47)
    Get ť ← arg min_{t̃∈T̃} (|h_{t̃,l}^(MISO) Ψ̃_T|)
    Update the set of selected streams: T = T ∪ ť
    Update the set of available streams: T̃ = T̃ \ ť
    Compute Ĉ^(sum,new) = Ĉ_T^(sum) according to Eq. (2.50)
    if Ĉ^(sum,new) > Ĉ^(sum,old) then
        Update sum spectral efficiency: Ĉ^(sum,old) = Ĉ^(sum,new)
    else
        Update selected streams: T = T \ ť
        Stop for loop
    end if
end for
```

In Fig. 2.10a, PBZF stream selection is compared with SUS where $\epsilon_{(\mathrm{SUS})}$ is set according to Table 2.3, and the case that all streams are selected if $\tilde{T} \leq N$ or randomly if $\tilde{T} > N$. Fist of all, it can be observed that there is no difference until $\tilde{T} = 60$ which corresponds to $\frac{T}{N} \approx 60\,\%$. In this range all streams in the system are selected such that $T \approx \tilde{T}$ shown in Fig. 2.10b where the number of selected streams T is given over the number of available streams $\tilde{T}$. In other words no grouping is required and all streams can be scheduled without performance loss with respect to the investigated greedy scheduling algorithms. In the range $60 < \tilde{T}$ the SUS with ZF precoder achieves a lower sum spectral efficiency compared to MMSE precoder with random selection. The PBZF stream selection yields the highest sum spectral efficiency of the three schemes utilizing diversity gains from increased $\tilde{T}$. The number of selected streams saturates to ≈ 65. Furthermore, the performance degradation of the ZF precoder vanishes with PBZF stream selection and the curve overlaps with MMSE. The reason for that is the assumption in PBZF grouping that user experience no inter-stream interference, the same as in ZF precoding, but considering noise in Eq. (2.49) for the rate SINR calculation. Therefore, PBZF stops adding streams before the sum spectral efficiency of ZF precoding drops from serving to many streams.

In short summary, in order to utilize multiplexing gains in practical massive MIMO systems inter user interference has to be reduced by precoder design, ideally only down to the noise level of the users. Therefore, the noise variance estimates of the users are required. With user or stream grouping further sum spectral efficiency gains can be achieved in the regime of $T > \frac{N}{2}$ available MISO channels compared to "without scheduling". For stream grouping with the objective to increase sum spectral efficiency a sum spectral efficiency objective should be considered rather than only channel orthogonality. As a drawback, stream grouping requires additional complexity compared to "without scheduling", on the other hand, with scheduling less hardware is required to achieve similar sum spectral efficiency. For example, each transmit-stream requires a transmit chain including inter alia a modulator, a digital to analogue converter, a power amplifier, filters etc.. Note that for each spatial layer a base band unit is required. Thus, the additional complexity of scheduling can be justified with less required hardware and thus less energy consumption. According to Fig. 2.10b approximately $\frac{2}{3}$ of the maximum possible number of streams is selected by the PBZF even for $\tilde{T} > \frac{2}{3}N$. Thereby, a saturation of sum spectral efficiency gains is observed. Another possibility to re-



(a) Sum spectral efficiency (SE) of PBZF stream selection and optimal threshold SUS according to Table 2.3.



(b) Number of served streams T of PBZF stream selection.

Figure 2.10.: Performance of PBZF stream selection.

duce computational complexity of stream selection in massive MIMO systems is the utilization of "channel hardening".

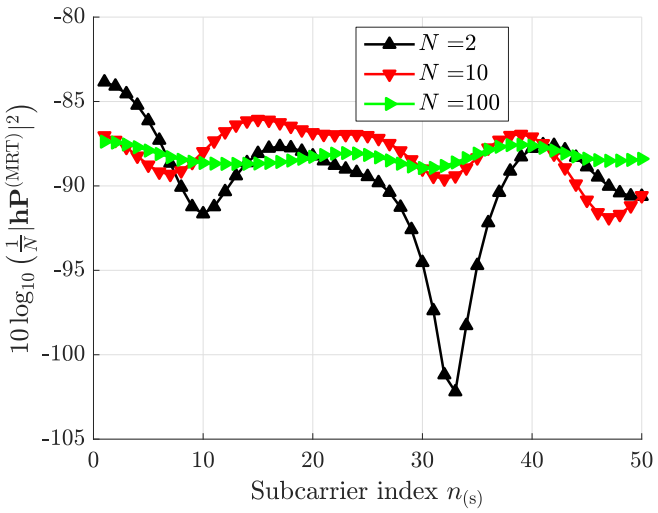
2.3.5. Channel Hardening

[BLM16] states that with increasing number of antennas channels become frequency flat, also called "channel hardening". This is because small-scale fading is averaged out over the antennas as a consequence of the law of large numbers. In Fig. 2.11a the normalized power of the MRT precoded channel in logarithmic scale is given over the frequency domain with 180 kHz subcarriers comparing 2, 10 and 100 antennas confirming the statement in [BLM16]. More precisely, in Fig. 2.11b the channel power variance of $N^{(\mathrm{RB})} = 50$ subcarriers is given over the number of antennas and it can be observed that the variance decreases monotonic to < 0.2 dB at $N = 100$ antennas.

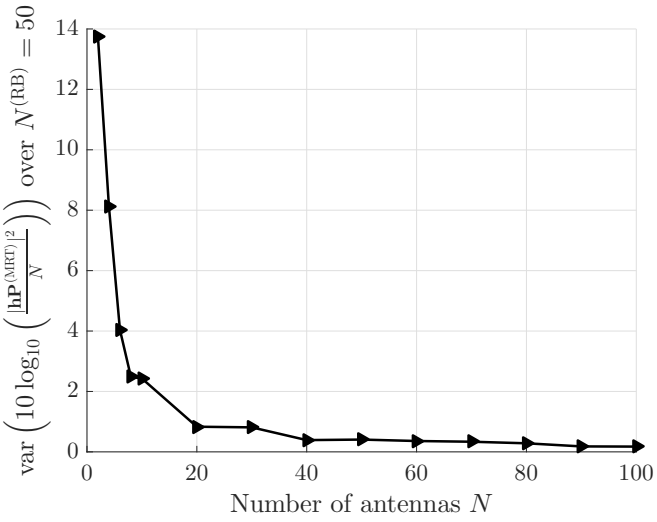
With frequency flat channels, the need for frequency adaptive scheduling vanishes, an essential part in LTE systems. Thus, the idea is that the increased signal processing complexity caused by a large number of antennas can be compensated by decreased complexity for stream selection. In the extreme case stream selection can be applied to the complete system bandwidth. This effect is studied in Fig. 2.12 by the following two assumptions:

1. The selection is performed on every $N_{(\mathrm{subband})}^{(\mathrm{PBZF})}$-th subcarrier according to $1 : N_{(\mathrm{subband})}^{(\mathrm{PBZF})} : N^{(\mathrm{RB})}$, where $N_{(\mathrm{subband})}^{(\mathrm{PBZF})}$ is increased in the X-axis in Fig. 2.12. In one extreme case $N_{(\mathrm{subband})}^{(\mathrm{PBZF})} = 1$ PBZF is performed on every subcarrier and is maximum in Fig. 2.12. In the other extreme case $N_{(\mathrm{subband})}^{(\mathrm{PBZF})} = 50$ PBZF is performed only on the first subcarrier.

2. The previous stream selection of the $N_{(\mathrm{subband})}^{(\mathrm{PBZF})}$-th subcarriers $i \in \left\{1 : N_{(\mathrm{subband})}^{(\mathrm{PBZF})} : N^{(\mathrm{RB})}\right\}$ is applied to subcarriers $i : i + N_{(\mathrm{subband})}^{(\mathrm{PBZF})}$. For example if PBZF stream selection is applied on subcarriers $1, 11, 21, 31, 41$ then these decisions are applied on subcarriers $1-10, 11-20, 21-30, 31-40, 41-50$, respectively, such that PBZF of subcarrier 1 is applied to subcarriers $1-10$ and so on. This corresponds to the LTE mode, where scheduling decisions are applied to a group of RBs called subbands [3GP16c].



(a) Normalized effective channel power over frequency domain with 180 MHz subcarriers comparing 2, 10, 100 antennas.



(b) Y-axis shows the variance of normalized power of the precoded channel over $N^{(\mathrm{RB})} = 50$ RBs as $\mathrm{var}\left(10\log_{10}\left(\frac{|\mathbf{hP}^{(\mathrm{MRT})}|^2}{N}\right)\right)$. X-axis is the number of antennas N.

Figure 2.11.: Channel hardening effect of massive MIMO.

Note, that the subcarrier size in this work is 180 kHz corresponding to the bandwidth of a LTE RB assuming that a RB is frequency flat. The subband size $N^{(\mathrm{PBZF})}_{(\mathrm{subband})}$ is defined as $N^{(\mathrm{PBZF})}_{(\mathrm{subband})} \in \mathbb{N}_+$ and bounded by $N^{(\mathrm{PBZF})}_{(\mathrm{RB})} \leq N^{(\mathrm{RB})}$.

The main observation is, that the sum spectral efficiency is approximately constant up to $N^{(\mathrm{subband})}_{(\mathrm{RB})} = 50$. This means that a single scheduling process, e.g. based on the first subcarrier, can be applied to the complete bandwidth without loss in sum spectral efficiency. Thus, the signal processing overhead due to stream/user scheduling does not scale with the bandwidth.

2.3.6. Conclusions

In this section, fundamentals of massive MIMO are studied and confirmed in a realistic environment with a "moderate" number of antennas. New relevant aspects, regarding the row norm and power distribution over the transmit antennas are investigated. In the single cell scenario, where users usually have a high SNR the MRT precoder is not a good choice in terms of sum spectral efficiency of the system. Furthermore, it is shown that the utilization of multiplexing gains in terms of sum spectral efficiency requires stream-selection, if the noise variance of users is unknown at the BS. Then, it is critical to take not only orthogonality, but also sum rate constraints into account. This is shown by the proposed and modified PBZF that outperforms the SUS reference scheme. Not only sum rate constraints, but also practical SINR bounds are considered by the algorithm, as well as power splitting with each additional stream. Another advantage of the PBZF stream-grouping algorithm is that no scenario (number of users) dependent orthogonality parameter is required. Finally, in order to decrease processing complexity for massive MIMO scheduling, the channel-hardening feature of massive MIMO is studied and exploited showing, that stream selection based on narrow band feedback can be applied to the overall system bandwidth without performance degradation.

On the other hand, if the noise variances of the users are known at the BS, then random stream selection up to the number of antennas also provides performance close to the sum spectral efficiency aware algorithm. However, the advantage of stream selection is that the same spectral efficiency can be achieved with less hardware (less transceiver chains), compared to the case without stream selection, e.g. in the studied

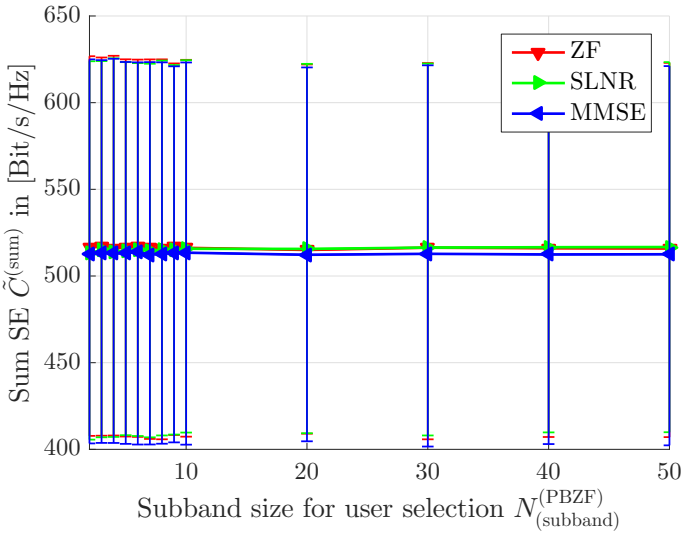


Figure 2.12.: Sum spectral efficiency (SE) over subband size for PBZF stream selection with $\tilde{K} = 100$ available streams.

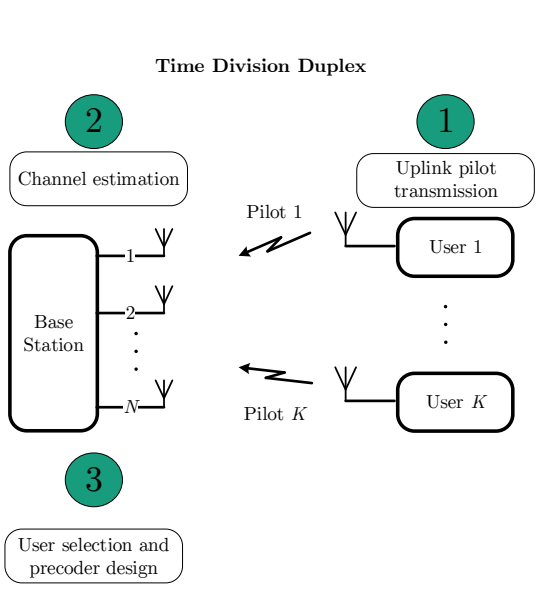
scenario with $N = 100$ antennas 65 streams are selected in average. In the greedy and sum spectral efficiency maximizing scheduling algorithms user fairness and quality of service is ignored, but can be included by over the top algorithms. For example an algorithm can select appropriate MISO channels or users to determine the set of available streams $\tilde{T}$, e.g. if a users was selected in a previous time slot, he can be left out in the next slot according to some fairness criteria.

2.4. Limited Channel State Information in Frequency Division Duplex (FDD)

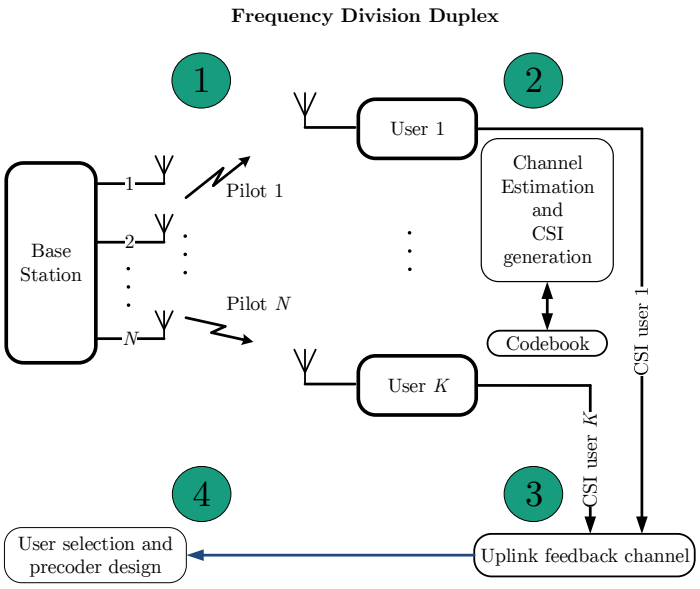
In the previous sections performance analysis assumes TDD mode and uplink-downlink reciprocity. This means, that the channel is estimated at the BS in the uplink transmission of the mobile user. Based on this the downlink channel is obtained by the transpose conjugate of the estimated uplink channel, assuming uplink-downlink reciprocity of the radio channel, see Fig. 2.13a. This is an ideal assumption in real systems, however recent work shows that the principle findings of massive MIMO still hold in the presence of CSI errors [Luo16]. The true research challenge of massive MIMO is to utilize the same gains from TDD systems also in FDD systems [BLM16]. The following two bottlenecks in FDD mode limit massive MIMO gains:

1. Downlink pilot signaling overhead
2. Uplink feedback capacity

The transmitter requires channel state information (CSI) for downlink precoding. In FDD systems, the downlink and uplink channels are located at different frequencies, therefore, in LTE the downlink channel is estimated at the mobile user and then CSI is fed back in the uplink to the transmit BS, see Fig. 2.13b. With delays caused by processing, higher layer protocols, and the electromagnetic wave propagation, the challenge is to obtain the downlink CSI at the transmitter within the coherence time of the channel. Coherence time means the time duration in which the channel is approximately constant with respect to the time of the channel estimation [RABT02]. This first bottleneck, the downlink signaling overhead is due to the need of "pilots" that has to be sent in the downlink for channel estimation at the user. To clarify this, a time-frequency block with a certain coherence time and coherence bandwidth is considered. It is further assumed, that the coherence time is divided into a number of time slots and the coherence bandwidth is divided into a number of narrowband sub channels resulting in $\tau^{(\mathrm{C})} \in \mathbb{N}_+$ resource elements within the coherence block. As shown in Fig. 2.14, in LTE a physical RB is considered a coherence block and consists of $\tau^{(\mathrm{C})} = 12 \cdot 14 = 168$



(a) Three steps in the TDD system consisting of 1) uplink pilot transmission by the users, 2) channel estimation at the BS, and 3) subsequent user selection and precoder design.



(b) Four steps in the FDD system consisting of 1) downlink pilot transmission from the BS, 2) channel estimation and CSI generation at the users, 3) uplink feedback transmission, and 4) subsequent user selection and precoder design.

Figure 2.13.: Downlink CSI acquisition at the BS, equipped with multiple antennas, for multiple-user downlink transmission.

resource elements, where 12 is the number of subcarriers and 14 is the number of symbols[4]. With one pilot per antenna and a maximum allowed pilot overhead of 50 %, the maximum number of antennas $\tilde{N}_{(\mathrm{FDD})}^{(\max)}$ is given by

$$\tilde{N}_{(\mathrm{FDD})}^{(\max)} = \frac{\tau^{(\mathrm{C})}}{2}. \tag{2.51}$$

Eq. (2.51) would result in a maximum of $\tilde{N}_{(\mathrm{FDD})}^{(\max)} = 84$ antennas for LTE. Note that Eq. (2.51) just considers the pilot overhead and does not include the number of uplink resources required to feed back the CSI within the coherence time. Taking this into account, it is assumed that each user requires a single time slot to transmit his feedback in the uplink. Therefore, with K spatially multiplexed users for downlink transmissions the maximum number of antennas is less than in Eq. (2.51) and given by

$$N_{(\mathrm{FDD})}^{(\max)} = \frac{\tau^{(\mathrm{C})} - K}{2} < \tilde{N}_{(\mathrm{FDD})}^{(\max)}. \tag{2.52}$$

Fig. 2.15 shows on the left hand the feasibility region of a FDD system according to Eq. (2.52). Therein, the overhead is plotted colour coded over the number of antennas and users. This example illustrates the fundamental limitation on the number of antennas and number of spatially multiplexed users for downlink in an FDD system.

In contrast to this, in a TDD system the maximum number of antennas $N_{(\mathrm{TDD})}^{(\max)}$ is not fundamentally limited by the system design, see Fig. 2.13a. In TDD only the number of served users K is limited by

$$K_{(\mathrm{TDD})}^{(\max)} \leq \frac{\tau^{(\mathrm{C})}}{2}, \tag{2.53}$$

[4]Note, that the coherence time and bandwidth of the channel depend on the speed of the mobile user and environment and can be smaller or larger than the LTE RB.

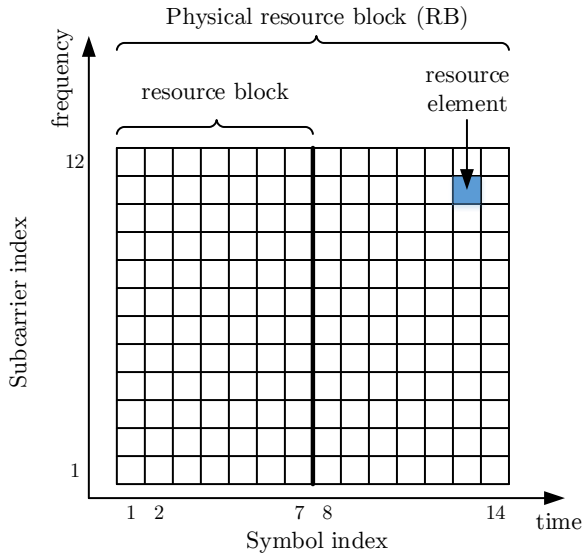


Figure 2.14.: Time-frequency structure of a physical RB in LTE. A physical RB is the smallest unit for resource assignment.

assuming again a maximum allowed overhead of 50 % as for FDD. This is shown on the right hand side of Fig. 2.15. Note that Eq. (2.53) does not consider time slots for uplink data transmission explicitly, so the maximum number of served users $K_{(\mathrm{TDD})}^{(\max)}$ can be less than $\frac{\tau^{(\mathrm{C})}}{2}$ indicated by the $\leq$ sign.

Remark. *In the subsequent thesis, a time-frequency resource is referred to as RB following LTE terminology for readability and comprehension. In LTE, a physical RB is the smallest unit for resource scheduling. In release 15 the first "5G" standard the concept of flexible numerology is introduced and RBs can be of various size in time and frequency domain [3GP17b]. However, if not stated otherwise a RB in this thesis is considered with* 180 kHz *bandwidth and* 1 ms *duration taking into account cyclic prefix overhead. Such a RB may or may not be coherent in time or frequency, depending on other factors such as user velocity.*

In an "ideal" system the complete estimated channel would be fed back to the BS without delay, however the second bottleneck is caused by several limitations in the uplink channel that prevent this "ideal" case. First, the uplink channel is capacity limited and mobile users usually share the uplink channel with each other. Therefore, a balance between CSI feedback and data-traffic has to be found. Another more practical limitation is the power budget of the mobile users caused by batteries. Uplink transmission requires transmit power and thus empties the batteries of mobile users. Therefore, CSI compression or reduction to decrease the required uplink feedback without losing relevant information are of high interest in current massive MIMO research [KKT12, NAAC14, GDWC15, JMCN15, KJTB15, SPCH16, JKS17] including compressed sensing, principal component analysis, feedback in various domains, quantization, joint feedback over multiple users, and so on. Another impairment is additional delay caused by processing and transmission at the UE. In this work, the CSI feedback is distinguished into "explicit" and "implicit" CSI.

- Explicit CSI refers to compression schemes that allow a reconstruction of the downlink channel at the BS for phase adapted precoding, e.g. the techniques presented in Section 2.3.1. Explicit CSI is usually without any assumptions about the transmission scheme.
- Implicit CSI refers to codebook-based feedback schemes. Implicit CSI usually consists of indices pointing to entries in a codebook known at the receiver and transmitter side, e.g. in LTE the CSI consists among other things of a precoding matrix indicator (PMI). Implicit CSI usually includes one or more assumptions on the transmission and the precoder is not phase adapted with respect to the

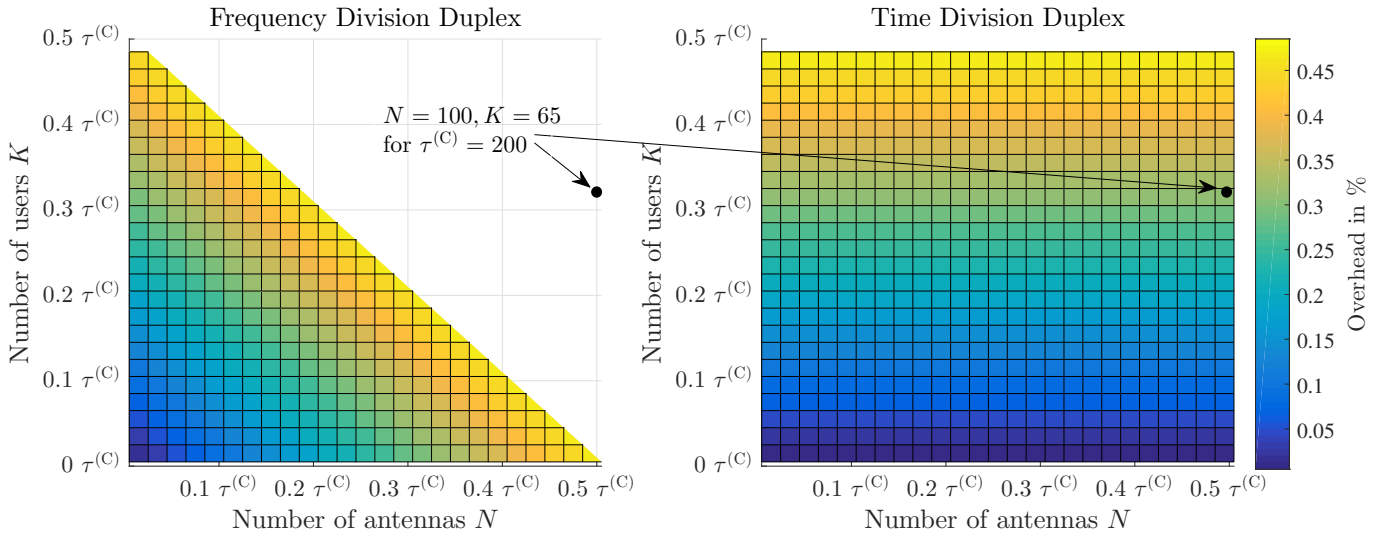


Figure 2.15.: Illustration of the overhead signaling trade-off in Massive MIMO FDD (left hand side) and TDD (right hand side). The main difference is that FDD limits the number of antennas, while TDD can have any number of antennas. As an example, a coherence resource with $\tau^{(C)} = 200$, $N = 100$ antennas, and $K = 65$ users is supported in TDD systems but not in FDD.

multiple paths of the downlink channel. This means that with a codebook the signal transmitted into a certain direction, however, signals arriving over multiple paths at the user are not phase adapted to superimpose constructively.

A brief discussion on CSI reduction principles is given in the next paragraph. Note that part of these reduction principles can be applied to both implicit and explicit CSI feedback while others are limited to implicit CSI only.

Overview on CSI Feedback Reduction Principles

1. A straightforward way of feedback reduction is an optimization of the feedback interval. In [SM11] the feedback interval is optimized with respect to the coherence time which itself depends on the velocity of the MSs and the CSI format. This means for example that the feedback interval of CSI which contains information about large-scale parameters (angular spread or direction of departure), e.g. the channel covariance in (2.12), the feedback interval can be larger compared to CSI on small-scale parameters such as shadow fading.

2. The second option is about compression in the time-delay domain from [LKCSM09, SMJ13], where the idea is that non-significant taps in the $\mathbf{H}^{(\mathrm{T})}(\tau, t)$ representation of the channel can be removed. This is motivated by the fact that the power of taps decreases exponentially with delay τ and values below certain thresholds can be neglected. In [SMJ13] it turns out that with respect to interference an compression rate of more than 90 % can be achieved without significant performance loss compared to the channel feedback $\mathbf{H}$ in the frequency domain.

3. The third point is about compressed sensing. In [KKT12] compressed sensing is used to reduce the spatial dimension exploiting correlation between transmit and receive antennas. Compressed sensing can also be used in other domains such as frequency and time domain. On top of compressed sensing in the spatial domain the distributed approach from [DSB+05] can be used where the spatial correlation between closely located MSs is utilized. The idea is to split the feedback of each MS into a common and a individual part and compress the common parts with a transformation matrix known at the UEs.

4. A similar approach compared to the third option is presented in [ANAC13] where the idea is to use only the dominant eigenspace of the channel covariance matrices in Eq. (2.12) for precoding resulting in feedback a with a much smaller dimension than the number of BS antennas.

5. In [TKBH12], CSI feedback to multiple BSs for joint transmission CoMP is reduced, exploiting a power threshold selection of a subset of the BSs. This can interpreted as an antenna selection, exploiting the different path-loss from the user to multiple not co-located BSs. In principle this could be adapted to the large antenna system such that antenna entries below a certain threshold are set to zero. However, with a practical number of antennas in a centralized deployment all antennas have a similar path loss to the user and reduction of the precoder size would be limited.

6. Another simple approach is to apply fractional frequency reuse among UEs presented in [XYYS10]. Therein each MS reports only CSI for a certain part of the available bandwidth and a CSI reduction by a factor of four with only 4 % performance loss in spectral efficiency and 5 %-ile user throughput is achieved. With this, the amount of feedback could be coupled to the rate requirements of the users. This indicates a great potential to reduce the feedback overhead but it requires many users in the system for diversity, e.g. in [XYYS10] 20 active users have been considered per sector. Furthermore, this approach does not reduce the amount of feedback with respect to the antenna dimension so it can be considered as complementary to other feedback reduction approaches.

7. In digital communications systems information is always quantized, therefore quantization has to be applied on top of any compression scheme. Quantizing continues values, e.g. the amplitude and phase of the complex channel coefficient, always results to a quantization-loss. On the other hand, the quantization loss scales with the number of bits used for quantization such that the more bits are used the lower the quantization-loss. Furthermore, with digital signal processing, e.g. for channel estimation, a finite number of bits is used for processing and storing the channel coefficients already induces a quantization loss. However, before transmitting such information over the air, another quantization with less bits is applied. This quantization can be straightforward by independent quantization of entries in $\mathbf{H}$ independently, jointly, or by approximation from predefined matrices [LHL+08]. The difference compared to codebook based precoding matrix indicator (PMI) feedback is that each MS designs its own precoder instead of selecting from a common codebook.

Note that combinations of the above listed CSI feedback reduction schemes are possible, for example, a reduction with scheme 1 can be combined with all the others, or a combination of scheme 2 with scheme 3. The lossy quantization of the channel in scheme 7 with a certain bit constraint $N^{(\mathrm{bit})}$ is similar to the codebook approach, where an index of the same bit size points to a codebook of size $2^{N^{(\mathrm{bit})}}$. A brief overview of codebook based CSI feedback is discussed in the next paragraph and summarized in Table 2.5.

Overview on Codebook Based CSI Feedback

As mentioned in the section before the goal is to design CSI feedback with a bit constraint given a certain channel correlation profile. In LTE, this is realized by pointing to entries in predefined codebook matrices. These pointers are called precoding matrix indicators (PMI). A more detailed overview of limited feedback in wireless communication systems is provided in [LHL+08]. In the following overview the focus is on correlated MIMO channels for "large" antenna arrays at the transmitter based on [SZL+13].

1. The so called "Kerdock" codebook was first proposed for limited feedback in MIMO systems in [IH08, IH09]. The general idea is that Kerdock codebooks are systematically generated with elements drawn from a quaternary alphabet[5] requiring low complexity and memory (or storage). On the other hand, according to [SZL+13] and verified with simulations based on a modified wireless world initiative new radio (WINNER) II channel model for 3-dimensional beamforming, a Kerdock codebook

[5] An alphabet that consists of four entries, e.g. the digits [0 1 2 3].

Table 2.4.: Overview of CSI feedback reduction principles distinguished by the reduction domain.

Number	CSI Reduction Domain	Reference	Brief description
1	Time	[SM11]	Optimization of the feedback interval depending on the coherence time $\tau^{(C)}$
2	Time-Delay	[LKCSM09, SMJ13]	Compression in the time-delay domain, remove non-significant taps
3	Spatial, Frequency, Time	[KKT12, DSB+05]	Compressed sensing to reduce spatial correlation between antennas. Distributed compressed sensing to reduce common parts of feedback among users
4	Spatial	[ANAC13]	Report only significant dominant eigenvalues
5	Spatial	[TKBH12]	Report only for significant antennas
6	Frequency	[XYYS10]	Report only parts of the bandwidth, depending on coherence bandwidth $B^{(C)}$ and user grouping strategies
7	Quantization	[LHL+08]	Report a quantized precoder approximating the CSI with respect to the precoding strategy

is not optimized for correlated channels which is the case when $\lambda_{(c)}/2$ spaced antennas are assumed. Furthermore, a Kerdock codebook can only be constructed for a number of antennas that is a power of two, which is a "strong" constraint on practical deployments.

2. [LH05] shows that the problem of finding the optimal codebook/precoder is equivalent to packing subspaces in the Grassmannian manifold, also called Grassmannian line packaging in literature. Briefly summarized, Grassmannian codebooks are fixed codebooks designed by maximizing the minimum distance between beamforming vectors. In [YVT+14] Grassmannian line packaging is applied for a 2D antenna array in a 3D ray traced based channel model and they provide a close to optimal solution for uncorrelated MIMO channels. As a drawback, Grassmannian codebooks are complex to construct with respect to the number of antennas and [LH05] shows that they are not optimal for correlated MIMO channels, e.g. for UPAs with $\lambda_{(c)}/2$ element spacing.

3. To overcome the drawbacks of the Grassmannian codebooks, Roh and Rao proposed vector quantization in [RR06] based on the Lloyd algorithm from Gray in [Gra84]. As stated in [RV13], vector quantization codebooks meet the necessary conditions for an optimal codebook construction, but not the sufficient condition. Nevertheless, it is widely believed that vector quantization constructions are optimal. Besides the complexity it is hard to obtain insights on the structure of the optimal codebook as a function of the channel correlation structure. Therefore, random vector quantization codebooks/precoders as a subset of vector quantization have been proposed in [SH03] with the advantage of low-complexity since they are instantiations of a random construction and the beamforming vectors are isotropic uniformly distributed and i.i.d. over the ambient space. In [RHV09] and subsequent [RV13] random vector quantization codebooks are studied for limited-feedback beamforming but only considering Rayleigh fading for the channel coefficients and the Kronecker-product correlation model for the second-order statistics, e. g. the transmit and receive covariance matrices [RKS10]. To the authors knowledge there are no publications on random vector quantization for limited feedback with 2D large antenna arrays. One reason for this is that the codebook size increases exponentially with the number of transmit antennas for a given distortion loss constraint with respect to the not quantized CSI. Thus, memory requirements would be infeasible for large antenna arrays. Another drawback of vector quantization codebook is, that they have no structure and exhaustive search in the

Table 2.5.: Summary of literature overview on PMI feedback with respect to 3D beamforming.

Number	Precoder type	Reference	Brief description
1	Kerdock	[IH09]	Systematically generated with elements from a quaternary alphabet, reduced storage and complexity but not optimized for correlated channels.
2	Grassmannian	[LH05, YVT$^+$14]	Codebooks generated with with Grassmannian line packaging based on the Kroenecker correlation model by separation of horizontal and vertical spatial correlation.
3	Vector quantization	[RHV09, RV13]	Optimal codebook design but not practical for large antenna arrays because the codebook size increases exponentially with the number of transmit antennas.
4	QAM	[RCC$^+$07]	Each vector in the codebook is a sequence of i-ary QAM symbols scaled to have unit norm. Not suitable for large number of antennas because i^N sequences are required for the codebook construction.
5	DFT	[LSZ$^+$13, YWZP13, SZL$^+$13]	DFT rotation matrices for horizontal and vertical beamforming combined by a Kroenecker product. Assumes independent antenna correlation in horizontal and vertical direction.

codebook is required at each UE, e.g. as in LTE.

4. Overcoming the disadvantages of random vector quantization codebooks, QAM codebooks were proposed [RCC$^+$07]. QAM codebooks are structured and each vector in the codebook is a sequence of QAM symbols scaled to have unit norm. The codebook is constructed by generating all possible sequences of a i-QAM, e.g. ± 1 for $i = 2$ corresponding to binary phase-shift keying. Since QAM is a "simple" bit-to-symbol mapping algorithm no storage is required at either transmitter or receiver and [RCC$^+$07] shows that the QAM-based codebooks can achieve the full diversity order. However, the drawback of the QAM codebooks is that the set of all possible codewords of size i^N is required, where i stands here for the QAM constellation number, e.g. $i = 2$ corresponds to binary phase-shift keying. Thus, the required memory is infeasible even for a moderate number of $N = 64$ antennas and $i = 2$ which results in $2^{64} \approx 1.810^{19}$ codewords.

5. In [LH03b, LH03a] discrete Fourier transform (DFT) matrices are proposed for equal gain transmission to achieve the full diversity order in correlated MIMO channels and for closely spaced antenna arrays in [SWK11]. The modification of the precoder design from a ULA towards 2D UPAs is given in [YWZP13] where separate precoders for horizontal and vertical beamforming are proposed. The extension to large antenna arrays in [SZL$^+$13] also assumes independent correlation in vertical and horizontal dimension and codebooks are obtained by a Kronecker-product of the horizontal and vertical precoder. In [LSZ$^+$13] the same is shown also for dual polarized antennas but despite the "massive" MIMO claim only results for a 4×4 antenna array with 32 elements (2 polarizations) in total are presented.

Conclusion

In the first 5G new radio (NR) specification from 3GPP standardization, short release 15 [3GP17c], implicit CSI feedback is specified for FDD systems. This implicit CSI feedback is based on DFT codebooks and is

a straightforward extension from the LTE specification, so one can see it as an "evolution" rather than an "revolution". The main reasons for DFT codebooks are:

1. They are robust against CSI estimation errors and well suited for spatially correlated channels.
2. An extension for any N is straightforward and the construction is of low complexity.
3. They are structured which provides the potential for low complexity selection algorithms at the UE.
4. They have an orthonormal basis that means any two columns are orthogonal to each other and have the same norm.
5. The codebook size can be used as a design parameter, e.g. for keeping a given feedback constraint.
6. Inter-stream interference can be estimated and taken into account at the UE side, e.g. for beam and channel quality indicator (CQI) selection.

The effectiveness of implicit CSI for FDD massive MIMO schemes is studied next in Section 2.4.1 - Section 2.4.4 with focus on the pilot overhead rather than feedback. The amount of CSI feedback that can be sent in the uplink depends on several other parameters, e.g. the uplink capacity depends on the "path-loss", interference from other UEs, receive algorithms at the BS, the uplink bandwidth, and others. The feedback versus spectral efficiency trade-off for the investigated techniques is studied separately in Section 2.4.5. codebooks.

Remark. *Due to the single BS scenario assumption SINRs* $> 40\,\mathrm{dB}$ *can be achieved in the single user case. Therefore, in Section 2.4.1, Section 2.4.2 and Section 2.4.3 the maximum SINR is set from* $40\,\mathrm{dB}$ *to* ∞ *for evaluation of codebook channel quantization.*

2.4.1. Codebook Design

In release 15, the 3GPP's first specification of 5G NR, the maximum number of downlink spatial layers is set to $T_l = 32$, see Table 5.2.2.2.1-2 in [3GP17c]. This value includes explicitly two polarizations, so maximum 16 different spatial directions are considered for various antenna configurations and oversampling factors. The oversampling factors are required for the DFT based codebook construction, explained in detail later in this section.

Codebook construction is far from simple and a lot of research was done in this field, e.g. the overview article on limited feedback in wireless communication systems from 2008 [LHL+08] includes 373 references. Also research has not stopped since then, e.g. codebooks for 3D beamforming with planar arras have been studied in standardization [3GP15a] and adopted in release 15 the first specification of 5G NR, [3GP17c]. There, the focus is on DFT based codebooks only due to the limitation on uniform planar array deployments. The reason for this is that the channel space given by the eigenvectors of the channel covariance matrix can be approximated by the columns of an DFT matrix for a ULA, see [ANAC13] for the proof. For reasons of practical relevance, the focus in this thesis is also on UPAs and thus DFT based codebooks, however the author explicitly points out that the presented idea in this section is by design not limited to DFT codebooks and can be applied to other codebooks.

An uniform DFT codebook $\mathbf{\Omega}_{(\mathrm{ULA})}^{(\mathrm{DFT})}$ for a ULA with N antennas and $N^{(\mathbf{\Omega})}$ codewords is described by

$$\mathbf{\Omega}_{(\mathrm{ULA})}^{(\mathrm{DFT})}\left(N, N^{(\mathbf{\Omega})}\right) = \left[\boldsymbol{\omega}_1^{(\mathrm{DFT})} \ \ldots \ \boldsymbol{\omega}_{N^{(\mathbf{\Omega})}}^{(\mathrm{DFT})}\right], \tag{2.54}$$

where the i-th codeword $\boldsymbol{\omega}_i^{(\mathrm{DFT})}$ is constructed as [XJW+13]

$$\boldsymbol{\omega}_i^{(\mathrm{DFT})} = \frac{1}{\sqrt{N}}\left[1 \ e^{j\frac{2\pi}{N^{(\mathbf{\Omega})}}i} \ e^{j\frac{2\pi}{N^{(\mathbf{\Omega})}}2i} \ \ldots \ e^{j\frac{2\pi}{N^{(\mathbf{\Omega})}}(N-1)i}\right]^{\mathrm{T}}. \tag{2.55}$$

Thus, the unitary DFT codebook can be constructed straightforwardly for any number of antennas and codewords. The extension to a UPA by a Kronecker product was proposed in [LSZ+13] and is obtained by

$$\mathbf{\Omega}_{(\mathrm{UPA})}^{(\mathrm{DFT})} = \mathbf{\Omega}_{(\mathrm{ULA})}^{(\mathrm{DFT})}\left(N_{(\alpha)}, N_{(\alpha)}^{(\mathbf{\Omega})}\right) \otimes \mathbf{\Omega}_{(\mathrm{ULA})}^{(\mathrm{DFT})}\left(N_{(\beta)}, N_{(\beta)}^{(\mathbf{\Omega})}\right), \tag{2.56}$$

where $N_{(\alpha)}$ and $N_{(\beta)}$ are the numbers of horizontal and vertical antennas, respectively, $N_{(\alpha)}^{(\mathbf{\Omega})}$ and $N_{(\beta)}^{(\mathbf{\Omega})}$ are the number of horizontal and vertical codewords, respectively. $\mathbf{\Omega}_{(\mathrm{ULA})}^{(\mathrm{DFT})}\left(N_{(\alpha)}\right)$ and $\mathbf{\Omega}_{(\mathrm{ULA})}^{(\mathrm{DFT})}\left(N_{(\beta)}\right)$ are obtained according to Eq. (2.55). The codebook construction in Eq. (2.56) is following the assumption that the channel correlation in horizontal and vertical dimension is independent according to [YVT+14]. The resulting codebook $\mathbf{\Omega}_{(\mathrm{UPA})}^{(\mathrm{DFT})}$ is of dimensions $N_{(\alpha)}N_{(\beta)} \times N_{(\alpha)}^{(\mathbf{\Omega})}N_{(\beta)}^{(\mathbf{\Omega})}$. A numerical performance evaluation for the UPA is given in Fig. 2.16a, where other assumptions than the antenna array are following Table 2.1. The $x^{(\mathrm{C})}$- and $y^{(\mathrm{C})}$-axis are presented in logarithmic scale such that the numbers $[0, \ldots, 5]$ can be interpreted as the number of bits required to index the best codeword in the respective codebook. On the $z^{(\mathrm{C})}$-axis the gain with respect to the achievable MRT rate is shown in order to quantify the gap to the upper bound. Note that the results in Fig. 2.16a are obtained for a single user served with a single stream. Each user k selects the codeword that maximizes the receive signal power according to

$$i^{(\max)} \leftarrow \underset{i \in \left[1,\ldots,N^{(\mathbf{\Omega})}\right]}{\arg\max} \|\mathbf{h}_{1,k,l}^{(\mathrm{MISO})}\boldsymbol{\omega}_i^{(\mathrm{DFT})}\|. \tag{2.57}$$

It can be observed that the rate is strictly monotonous increasing, however a saturation can be observed for large codebook sizes in both horizontal and vertical dimension with approximately 92 % at $N_{(\alpha)} = N_{(\beta)} = 2^5 = 32$. The remaining gap with respect to MRT precoding cannot be closed if the signal power is received over multiple paths at the user. Each codeword in the codebook corresponds to beam steering in a certain direction, see Fig. 3.14b for an example. The beam is selected to maximize the receive power which corresponds to matching the direction of the strongest multi-path component. The main message from Fig. 2.16a is the motivation for a large codebook size to increase the user rate.

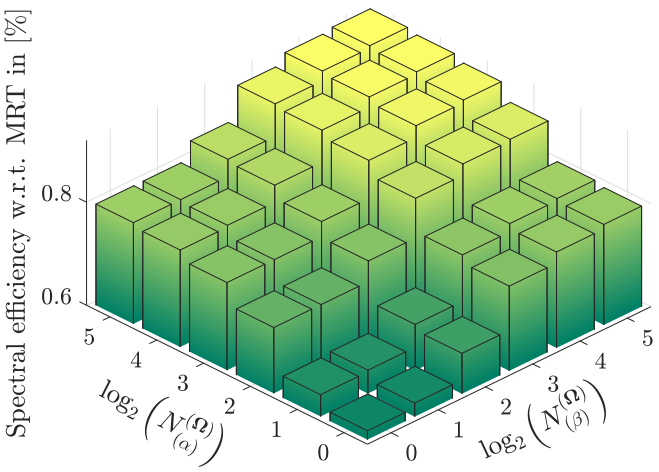
On the other hand, the larger the codebook the larger the required amount of feedback. Fig. 2.15 shows that the overhead from pilots in an FDD system increases linearly with the number of antennas. This is true for non-precoded pilots. However, with correlation in the channel, a lower number of pilots than the number of antennas is required, e.g. see [ANAC13]. Therein, the concept of precoded pilots is utilized in order to reduce signaling overhead. Precoded pilots are also used in LTE [3GP17c]. That means, if a user has to select from a codebook of size $N^{(\mathbf{\Omega})} < N$ it requires less signaling overhead to send $N^{(\mathbf{\Omega})}$ precoded pilots. On the other hand, if $N^{(\mathbf{\Omega})} \geq N$ it is better in terms of signaling overhead to send N non-precoded pilots. So an overhead scaling factor according to

$$O^{(\mathrm{CB})} = \begin{cases} \frac{\tau^{(\mathrm{C})} - N^{(\mathbf{\Omega})}}{\tau^{(\mathrm{C})}} & \text{, if } N^{(\mathbf{\Omega})} < N < \tau^{(\mathrm{C})}, \\ \frac{\tau^{(\mathrm{C})} - N}{\tau^{(\mathrm{C})}} & \text{, if } N \leq N^{(\mathbf{\Omega})} \wedge N < \tau^{(\mathrm{C})}, \\ 0 & \text{, if } N^{(\mathbf{\Omega})} \geq \tau^{(\mathrm{C})} \wedge N \geq \tau^{(\mathrm{C})} \end{cases} \tag{2.58}$$

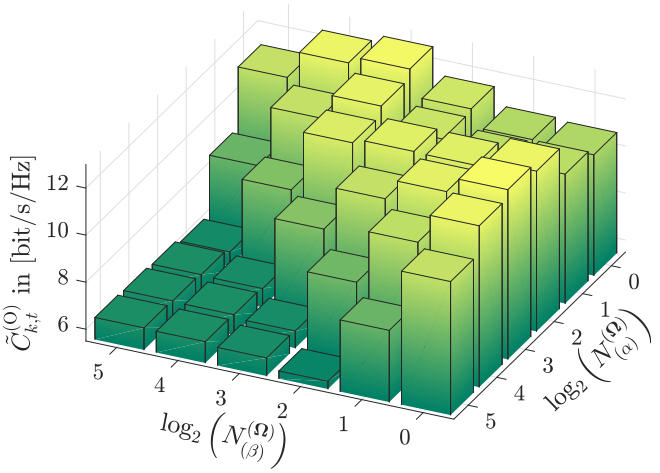
is considered for an effective downlink rate. Note that $O^{(\mathrm{CB})}$ is in the interval $0 \leq O^{(\mathrm{CB})} < 1$. Consequently, the effective downlink rate is obtained by combining the overhead and the rate according to

$$\tilde{R}_{k,t}^{(\mathrm{O})} = \tilde{R}_{k,t} O^{(\mathrm{CB})}. \tag{2.59}$$

Accordingly, $\tilde{R}_l^{(\mathrm{sum,O})}, \tilde{C}_{k,t}^{(\mathrm{O})}, \tilde{C}_l^{(\mathrm{sum,O})}$ denote the sum rate, stream spectral efficiency, and sum-spectral efficiency obtained by multiplying $O^{(\mathrm{CB})}$ with Eq. (2.6), Eq. (2.7), and Eq. (2.8), respectively. The same holds also for the bounded sum rate and spectral efficiency expressions with respect to Eq. (2.9). Fig. 2.16b shows $\tilde{C}_{k,t}^{(\mathrm{B,O})}$ where the other parameters are the same as in Fig. 2.16a and the $z^{(\mathrm{C})}$-axis is in [bit/s/Hz] instead of the ratio with respect to MRT precoding. However, Fig. 2.16b shows the trade-off between the



(a) Single user spectral efficiency with respect to (w.r.t.) optimal MRT precoding over horizontal and vertical codebook size $N_{(\alpha)}$ and $N_{(\beta)}$, respectively. A value of 0.8 means that codebook precoding achieves 80 % of the spectral efficiency of optimal MRT precoding.



(b) Single user effective spectral efficiency according to Eq. (2.59) taking into account pilot signaling overhead over horizontal and vertical codebook size $N_{(\alpha)}$ and $N_{(\beta)}$, respectively. The number of resource elements is $\tau^{(\mathrm{C})} = 168$ and each stream requires one resource element overhead. The maximum of 12.7 bit/s/Hz is achieved for $\log 2\left(N_{(\alpha)}^{(\mathbf{\Omega})} = 3\right)$ and $\log 2\left(N_{(\beta)}^{(\mathbf{\Omega})} = 0\right)$.

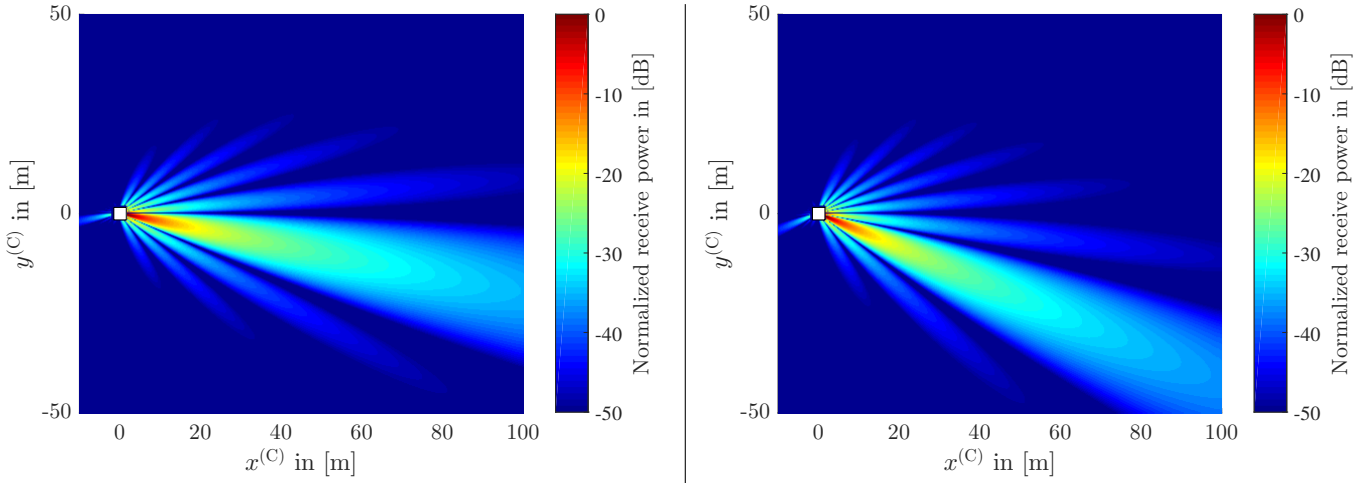
Figure 2.16.: Performance evaluation of the codebook size and signaling overhead for codebook based precoding in FDD. The UPA dimension is 10×10.

codebook size and the required pilot signaling overhead on the effective downlink rate in a FDD system. Note that in Fig. 2.16b the number of coherence elements is assumed with $\tau^{(\mathrm{C})} = 168$ according to a LTE resource block. The achieved maximum depends on the user positions, e.g. if the elevation spread of the user positions increases, a larger quantization of the vertical codebook would be more effective than the horizontal codebook in terms of spectral efficiency. Note, that the spread of user positions is with respect to the Monte-Carlo simulations drops where in each drop the user is randomly positioned within the sector. Consequently, the density distribution of the users has an impact and a non-unitary DFT codeword distribution would be better. Therefore, the spatial structure of DFT Kronecker codebooks is discussed in detail in the next paragraph.

Spatial Structure of DFT Kronecker Codebooks

In order to understand the structure of the UPA Kronecker codebook $\mathbf{\Omega}_{(\mathrm{UPA})}^{(\mathrm{DFT})}$ a visualization of the codewords from the DFT codebook is discussed.

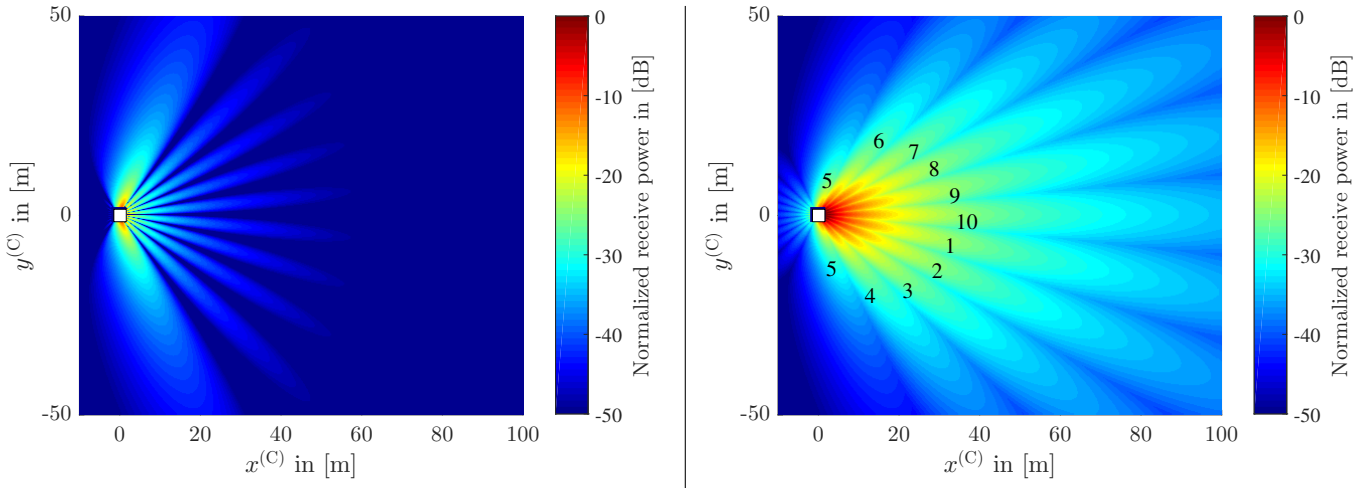
Fig. 2.17 shows the normalized receive power of codewords $\boldsymbol{\omega}_1^{(\mathrm{DFT})}$ and $\boldsymbol{\omega}_2^{(\mathrm{DFT})}$ in the $x^{(\mathrm{C})}$-$y^{(\mathrm{C})}$ plane. The corresponding codebook $\mathbf{\Omega}$ of size $N \times 10$ is obtained by Eq. (2.54). For simplicity, a ULA with $N = N_{(\alpha)}$, orientation $[1\ 0\ 0]^{\mathrm{T}}$, and positioned at $[0\ 0\ 0]^{\mathrm{T}}$ marked by white squares is assumed. It can be observed that the direction of the main lobe shifts from codeword $\boldsymbol{\omega}_1^{(\mathrm{DFT})}$ to $\boldsymbol{\omega}_2^{(\mathrm{DFT})}$. One property of UPAs is, that beamforming gains decrease with increased angular distance with respect to the array orientation, discussed in more detail in Section 3.2. Fig. 2.18a shows the beam of codeword $\boldsymbol{\omega}_5^{(\mathrm{DFT})}$ and it can be observed that two lobes of similar power exist pointing to the array sides. The largest receive power over all codewords of the codebook is given in Fig. 2.18b with the corresponding codeword indices. Each codeword corresponds to a certain direction with respect to the main lobe and the direction of the steered beam can be obtained analytically directly from the phases of the codewords for a ULA. The phase of the complex coefficients of



(a) Codeword $\boldsymbol{\omega}_1^{(\mathrm{DFT})}$.

(b) Codeword $\boldsymbol{\omega}_2^{(\mathrm{DFT})}$.

Figure 2.17.: Normalized receive power of different codewords plotted over the area at $z^{(\mathrm{C})} = 0$. The $N = N_{(\alpha)}$ ULA has orientation $[1\ 0\ 0]^{\mathrm{T}}$ positioned at $[0\ 0\ 0]^{\mathrm{T}}$ marked by the white square. The codebook $\boldsymbol{\Omega}$ of size $N \times N^{(\boldsymbol{\Omega})} = 10 \times 10$ is obtained by Eq. (2.54).



(a) Codeword $\boldsymbol{\omega}_5^{(\mathrm{DFT})}$.

(b) Largest receive power over all codewords at each positions.

Figure 2.18.: Normalized receive power of different codewords plotted over the area at $z^{(\mathrm{C})} = 0$. The $N = N_{(\alpha)}$ ULA has orientation $[1\ 0\ 0]^{\mathrm{T}}$ positioned at $[0\ 0\ 0]^{\mathrm{T}}$ marked by the white square. The codebook $\boldsymbol{\Omega}$ of size $N \times 10$ is obtained by Eq. (2.54).

codeword $\boldsymbol{\omega}$ can interpreted as phase shifters, see Fig. 2.19. According to [RP08] the phase difference $\Delta\omega_{i,j}$ between two antennas elements i and j, where $i \neq j$ and $i, j \in \{1, \dots N\}$, is proportional to the distance of the incoming wave to the antenna elements denoted by $x_{i,j}^{(\mathrm{C})}$, see Fig. 2.19, according to

$$\frac{360^\circ}{\Delta\omega_{i,j}} = \frac{\lambda_{(\mathrm{c})}}{x_{i,j}^{(\mathrm{C})}}. \tag{2.60}$$

The distance of the incoming wave from the antennas depends on the antenna spacing $d_{i,j}^{(\mathrm{N})}$ and the angle of the incoming wave by $x_{i,j}^{(\mathrm{C})} = d_{i,j}^{(\mathrm{N})} \sin(\alpha_{i,j})$. By substituting $x_{i,j}^{(\mathrm{C})}$ in Eq. (2.60) the angular direction $\alpha_{i,j}$ of the steered beam is given by

$$\alpha_{i,j} = \arcsin\left(\frac{\Delta\omega_{i,j}\lambda_{(\mathrm{c})}}{360^\circ d_{i,j}^{(\mathrm{N})}}\right). \tag{2.61}$$

In a ULA the distance between adjacent elements is $d_{1,2}^{(\mathrm{N})} = d_{2,3}^{(\mathrm{N})} \dots = d_{N-1,N}^{(\mathrm{N})} = d^{(\mathrm{N})}$ and the distance between any two elements i, j is given by

$$d_{i,j}^{(\mathrm{N})} = |i-j|\, d^{(\mathrm{N})}. \tag{2.62}$$

Thus, in order to steer the beam over all elements the phase difference $\Delta\omega_{i,j}$ must fulfill $\Delta\omega_{i,j} = (i-j)\,\Delta\omega_{1,2}$. In other words, the phase difference between non-adjacent antenna elements must be an integer multiple of the phase difference between adjacent antenna elements, where the multiplication factor is the difference of the antenna element indices. Note that due to property of the sinus function a negative argument due to a negative phase difference results in a negative angle. From the DFT codeword construction in Eq. (2.55) the phase difference between antenna element i and j in codeword i' of codebook $\boldsymbol{\Omega}_{(\mathrm{ULA})}^{(\mathrm{DFT})}\left(N, N_{(\alpha)}^{(\boldsymbol{\Omega})}\right)$ is given by

$$\Delta\omega_{i,j}\left(i', N_{(\alpha)}^{(\boldsymbol{\Omega})}\right) = \arg\left(\left[\boldsymbol{\omega}_{i'}^{(\mathrm{DFT})}\right]_i - \left[\boldsymbol{\omega}_{i'}^{(\mathrm{DFT})}\right]_j\right) = \frac{360^\circ i' i}{N_{(\alpha)}^{(\boldsymbol{\Omega})}} - \frac{360^\circ i' j}{N_{(\alpha)}^{(\boldsymbol{\Omega})}} = \frac{360^\circ i'}{N_{(\alpha)}^{(\boldsymbol{\Omega})}}(i-j). \tag{2.63}$$

It can be observed that the phase difference of two elements in a codeword is a multiple of the antenna index difference $i-j$. Substituting $\Delta\omega_{i,j}$ in Eq. (2.61) with Eq. (2.63) yields

$$\alpha_{i,j}^{(\mathrm{ULA})}\left(i', N_{(\alpha)}^{(\boldsymbol{\Omega})}\right) = \arcsin\left(\frac{\frac{360^\circ i'}{N_{(\alpha)}^{(\boldsymbol{\Omega})}}(i-j)\,\lambda_{(\mathrm{c})}}{360^\circ d_{i,j}^{(\mathrm{N})}}\right) = \arcsin\left(\frac{i'}{N_{(\alpha)}^{(\boldsymbol{\Omega})}}\frac{\lambda_{(\mathrm{c})}}{d_{i,j}^{(\mathrm{N})}}(i-j)\right). \tag{2.64}$$

Applying the uniform antenna distribution property from Eq. (2.62) on Eq. (2.64) results in

$$\alpha_{i,j}^{(\mathrm{ULA,DFT})}\left(i', N_{(\alpha)}^{(\boldsymbol{\Omega})}\right) = \arcsin\left(\frac{i'}{N_{(\alpha)}^{(\boldsymbol{\Omega})}}\frac{\lambda_{(\mathrm{c})}}{d^{(\mathrm{N})}}\frac{i-j}{|i-j|}\right) = \begin{cases} \arcsin\left(\frac{i'}{N_{(\alpha)}^{(\boldsymbol{\Omega})}}\frac{\lambda_{(\mathrm{c})}}{d^{(\mathrm{N})}}\right) & , \text{if } i-j \geq 0 \\ -\arcsin\left(\frac{i'}{N_{(\alpha)}^{(\boldsymbol{\Omega})}}\frac{\lambda_{(\mathrm{c})}}{d^{(\mathrm{N})}}\right) & , \text{if } i-j < 0 \end{cases} \tag{2.65}$$

Eq. (2.65) shows that the beam direction of a ULA with a codeword from the DFT codebook in Eq. (2.54) depends only on the codeword index i' and the number of codewords $N_{(\alpha)}^{(\boldsymbol{\Omega})}$, because the factor $\frac{\lambda_{(\mathrm{c})}}{d^{(\mathrm{N})}}$ is constant for a given center frequency and antenna spacing. Note that the codeword index is constrained by $i' \leq N_{(\alpha)}^{(\boldsymbol{\Omega})}$ such that the argument of the inverse sinus function in Eq. (2.65) can be > 1 if the spacing between antenna elements $d^{(\mathrm{N})}$ is less than the wavelength $\lambda_{(\mathrm{c})}$ such that $\frac{\lambda_{(\mathrm{c})}}{d^{(\mathrm{N})}} > 1$. However, the argument of the inverse sinus function is only defined in the interval $[-1, 1]$ corresponding to an angular range of $[-90^\circ, 90^\circ]$. Therefore, a mapping of the argument $\frac{i'}{N^{(\boldsymbol{\Omega})}}\frac{\lambda_{(\mathrm{c})}}{d^{(\mathrm{N})}}$ from $\mathbb{R}_+$ to the interval $[-1, 1]$ is required to take into account the periodicity of the sine function by

$$w_{i', N_{(\alpha)}^{(\boldsymbol{\Omega})}}^{(-1,1)} = \mathrm{mod}\left(\frac{i'}{N_{(\alpha)}^{(\boldsymbol{\Omega})}}\frac{\lambda_{(\mathrm{c})}}{d^{(\mathrm{N})}} + 1, 2\right) - 1, \tag{2.66}$$

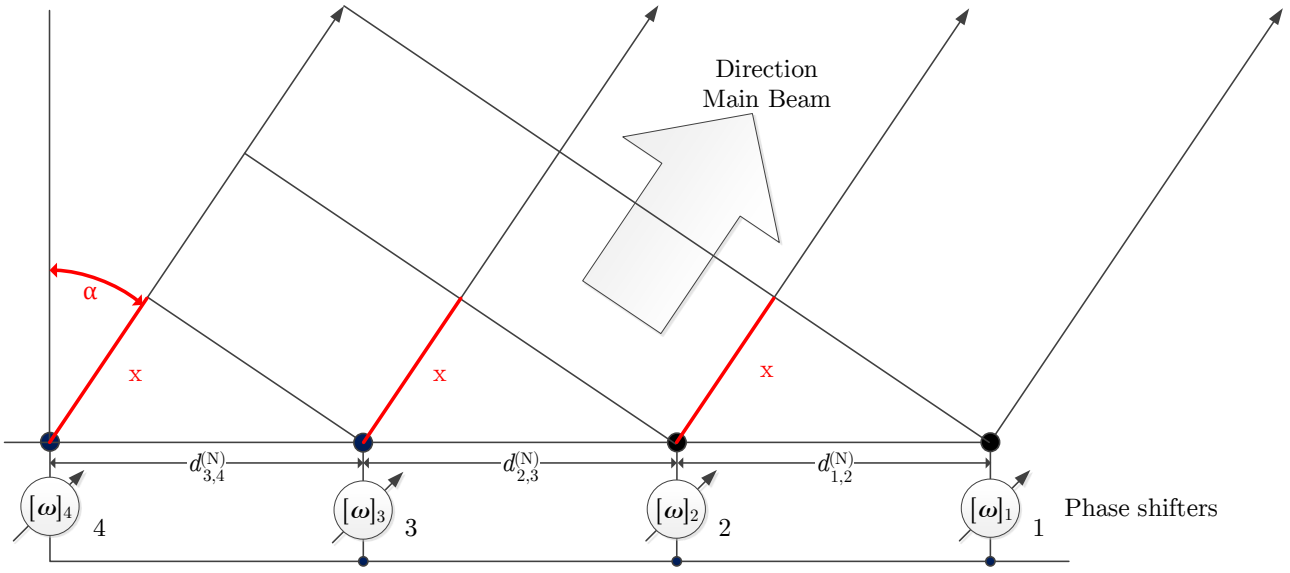


Figure 2.19.: Beam steering with phase shifters in a ULA.

such that Eq. (2.65) can be rewritten as

$$\tilde{\alpha}^{(\mathrm{ULA,DFT})}_{i',N^{(\mathbf{\Omega})}_{(\alpha)}} = \begin{cases} \arcsin\left(w^{(-1,1)}_{i',N^{(\mathbf{\Omega})}_{(\alpha)}}\right) & \text{, if } w^{(-1,1)}_{i',N^{(\mathbf{\Omega})}_{(\alpha)}} \geq 0 \\ -\arcsin\left(w^{(-1,1)}_{i',N^{(\mathbf{\Omega})}_{(\alpha)}}\right) & \text{, if } w^{(-1,1)}_{i',N^{(\mathbf{\Omega})}_{(\alpha)}} < 0, \end{cases} \tag{2.67}$$

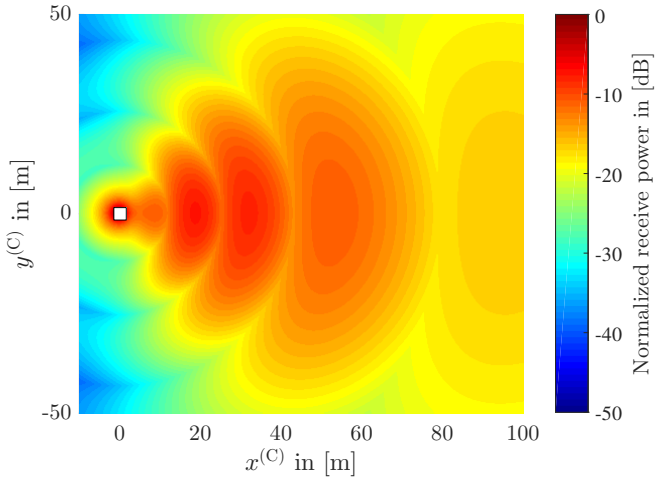
where the antenna indices i and j are omitted assuming that the mapping of codewords to antenna elements is constant without limit to generality. Thus the factor $\frac{i'}{N^{(\mathbf{\Omega})}}$ results in an uniform angular quantization of the interval $[-90°, 90°]$ by $N^{(\mathbf{\Omega})}$ steps. Note that the beam $\frac{i'=N^{(\mathbf{\Omega})}}{N^{(\mathbf{\Omega})}}$ corresponds to $\alpha = 0°$ and is included in all possible DFT codebook with respect to Eq. (2.54).

The same beam steering as for the horizontal ULA holds also for the vertical ULA, however due to the projection of beams to the plane spanned by $[1, 1, 0]^{\mathrm{T}}$ the steered beams result in unequal coverage areas, see Fig. 2.20b and Fig. 2.20a. Consequently, the beam direction of the resulting Kronecker codebook Eq. (2.56) can be obtained from the linear combination of the beam directions of the horizontal and vertical DFT codebook. Therefore, the direction vector $\boldsymbol{\phi}^{(\mathrm{DFT})}_{(\mathrm{UPA})}$ of size $N^{(\mathbf{\Omega})}_{(\alpha)} N^{(\mathbf{\Omega})}_{(\beta)} \times 1$ is defined where entries correspond to the $N^{(\mathbf{\Omega})}_{(\alpha)} N^{(\mathbf{\Omega})}_{(\beta)}$ directions of the codewords in the Kronecker codebook $\mathbf{\Omega}^{(\mathrm{DFT})}_{(\mathrm{UPA})}$. Each entry of $\boldsymbol{\phi}^{(\mathrm{DFT})}_{(\mathrm{UPA})}$ is a direction and represented by an azimuth and elevation angle according to the definition in Fig. 3.1. Note that the azimuth angles stem from the horizontal codebook $\mathbf{\Omega}^{(\mathrm{DFT})}_{(\mathrm{ULA})}\left(N, N^{(\mathbf{\Omega})}_{\alpha}\right)$ and the vertical angles stem from the vertical codebook $\mathbf{\Omega}^{(\mathrm{DFT})}_{(\mathrm{ULA})}\left(N, N^{(\mathbf{\Omega})}_{\beta}\right)$. For notational convenience the horizontal and vertical directions are defined as

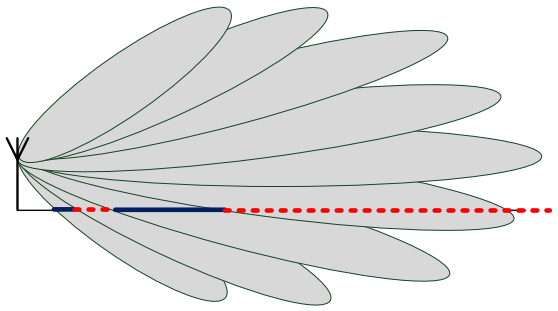
$$\begin{aligned} \boldsymbol{\alpha}^{(\mathrm{ULA})} &= \left[\alpha_1 \ \ldots \ \alpha_{i'} \ \ldots \ \alpha_{N^{(\mathbf{\Omega})}_{\alpha}}\right]^{\mathrm{T}} \\ \boldsymbol{\beta}^{(\mathrm{ULA})} &= \left[\beta_1 \ \ldots \ \beta_{j'} \ \ldots \ \beta_{N^{(\mathbf{\Omega})}_{\beta}}\right]^{\mathrm{T}}, \end{aligned} \tag{2.68}$$

where i' and j' correspond to the codewords in the respective codebooks. Following the UPA codebook construction in (2.56) this results in codeword directions according to

$$\boldsymbol{\phi}^{(\mathrm{DFT})}_{(\mathrm{UPA})} = \begin{bmatrix} \alpha_1 & \alpha_1 & \ldots & \alpha_1 & \alpha_2 & \ldots & \alpha_{N^{(\mathbf{\Omega})}_{\alpha}} & \ldots & \alpha_{N^{(\mathbf{\Omega})}_{\alpha}} \\ \beta_1 & \beta_2 & \ldots & \beta_{N^{(\mathbf{\Omega})}_{\beta}} & \beta_1 & \ldots & \beta_1 & \ldots & \beta_{N^{(\mathbf{\Omega})}_{\beta}} \end{bmatrix}^{\mathrm{T}}. \tag{2.69}$$



(a) $N = N_{(\beta)}$ ULA with orientation $\mathbf{e}^{(x)}$ positioned at $[0,0,0]^T$ marked by the white square. The codebook $\mathbf{\Omega}$ of size $N \times 10$ is obtained by Eq. (2.54). The maximum normalized receive power over all codewords is plotted in the $x^{(C)}$-$y^{(C)}$ plane.



(b) Beam projection on the $x^{(C)}$-axis from a BS located at $\mathbf{e}^{(z)}$. The alternating dashed and solid lines mark the projected beam width in the $x^{(C)}$-dimension of the $x^{(C)}$-$y^{(C)}$ plane.

Figure 2.20.: Projection of DFT beams from a vertical ULA at height 25 m.

$\boldsymbol{\phi}_{(\mathrm{UPA})}^{(\mathrm{DFT})}$ in Eq. (2.69) can be transformed to a matrix representation such that rows correspond to vertical angles and columns to horizontal angles according to

$$\tilde{\boldsymbol{\phi}}_{(\mathrm{UPA})}^{(\mathrm{DFT})} = \begin{bmatrix} \beta_1, \alpha_1 & \beta_1, \alpha_2 & \dots & \beta_1, \alpha_{N_\alpha^{(\mathbf{\Omega})}} \\ \beta_2, \alpha_1 & \beta_2, \alpha_2 & \dots & \beta_2, \alpha_{N_\alpha^{(\mathbf{\Omega})}} \\ \vdots & \vdots & & \vdots \\ \beta_{N_\beta^{(\mathbf{\Omega})}}, \alpha_1 & \beta_{N_\beta^{(\mathbf{\Omega})}}, \alpha_2 & \dots & \beta_{N_\beta^{(\mathbf{\Omega})}}, \alpha_{N_\alpha^{(\mathbf{\Omega})}} \end{bmatrix}. \tag{2.70}$$

An example of the corresponding receive power plot of the Kronecker DFT codebook is given in Fig. 2.21. Due to the normalization to the maximum receive power, the color-code in Fig. 2.17, Fig. 2.18, Fig. 2.20, and Fig. 2.21 are not directly comparable, e.g. by distribution of transmit power to a different number of streams, however they visualize the general distribution of the steered beams.

From the distribution of the steered beams it becomes clear that the statistic on selected codewords depends on the user distribution, e.g. if users are located in upper floors of skyscraper buildings elevation beams of $\beta < 0°$ are selected. The underlying user distribution of Fig. 2.16 is given in Table 2.1 and users are distributed in the $x^{(C)}$-$y^{(C)}$ plane at $z^{(C)} = 1.5\,\mathrm{m}$. The corresponding distribution of selected codewords for $N_\alpha^{(\mathbf{\Omega})} = N_\beta^{(\mathbf{\Omega})} = 2^4$ is given in Fig. 2.22a and a sparse and non-uniform distribution can be observed. The reason for this is the non-uniform user distribution in the angular domain from the perspective of the UPA. Fig. 2.22 shows the angular directions that correspond to the selected codewords according to Eq. (2.70). The numbers in the squares are the selection probabilities in % which is also color coded from green to yellow, where yellow corresponds to the highest probability. Due to the given geometry, the elevation angles of selected codewords range from 7° to −39° and the azimuth range is from −49° to 49°, ignoring codeword directions with probabilities ≤ 0.2. This angular range can change by changing the deployment geometry[6]. Therefore, the general angular codeword range is introduced by $\alpha^{(\mathrm{CW,min})}, \alpha^{(\mathrm{CW,max})}$ and $\beta^{(\mathrm{CW,min})}, \beta^{(\mathrm{CW,max})}$ for the azimuth and elevation range, respectively. These angular ranges can be obtained and updated by the BS due to "probing" over the complete angular range on a much larger timescale than required for CSI feedback. Thus, the additional overhead for such probing is not taken into account

[6] Geometry here includes the deployment of the antenna array, the users with respect to the orientation of the array, and also the propagation conditions, e.g. the angular range can also be different for street canyons and open squares.

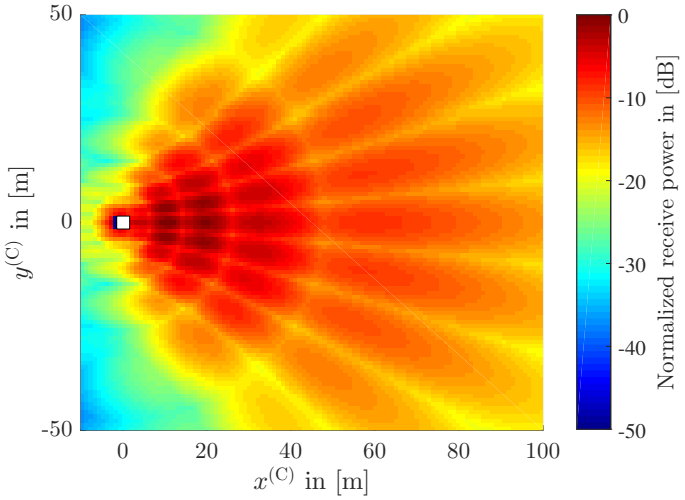


Figure 2.21.: Normalized receive power of the strongest beam in the $x^{(C)}$-$y^{(C)}$ plane for the DFT Kronecker codebook.

for overhead and feedback calculation and the angular ranges $\alpha^{(\mathrm{CW,min})}, \alpha^{(\mathrm{CW,max})}$, $\beta^{(\mathrm{CW,min})}, \beta^{(\mathrm{CW,max})}$ are assumed to be known at the BS.

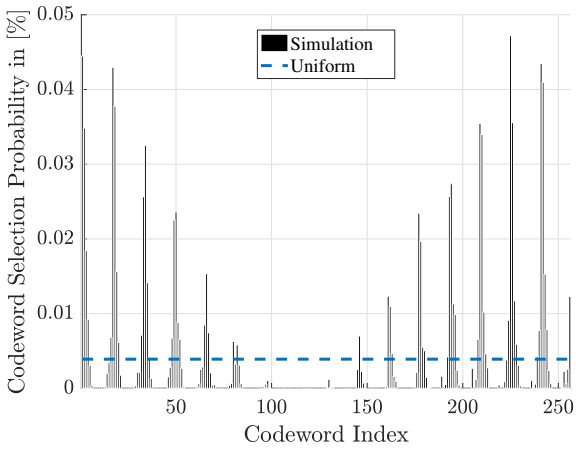
For ease of notation the angular ranges are denoted in a vector notation by

$$\begin{aligned} \boldsymbol{\alpha}^{(\mathrm{CW})} &= \left[\alpha^{(\mathrm{CW,min})} \;\; \alpha^{(\mathrm{CW,max})}\right]^{\mathrm{T}} \\ \boldsymbol{\beta}^{(\mathrm{CW})} &= \left[\beta^{(\mathrm{CW,min})} \;\; \beta^{(\mathrm{CW,max})}\right]^{\mathrm{T}} \end{aligned} . \tag{2.71}$$

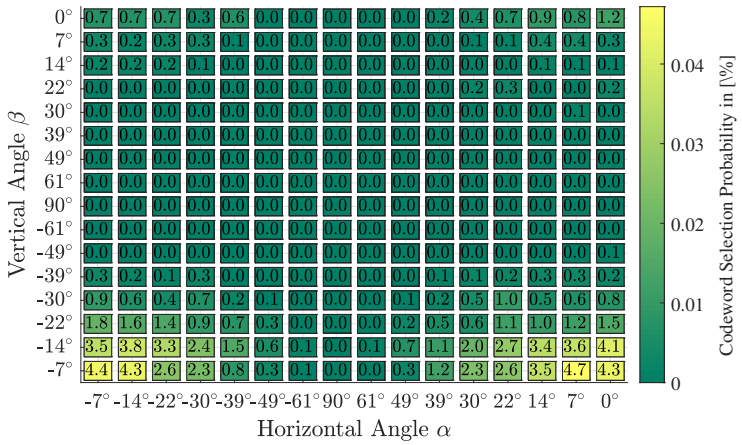
Note, that by limiting the angular ranges the number of codewords can be reduced keeping the beam resolution constant or by keeping the number of codewords constant the resolution of the reduced angular range is increased. Assuming a target angular resolution the amount of CSI feedback can be reduced by constructing codewords only within the limited angular ranges compared to construction in the complete angular range. On the other hand, the resolution and thus spectral efficiency can be increased for a given amount of CSI feedback, see Fig. 2.16. In order to obtain $N^{(\boldsymbol{\Omega})}_{(\boldsymbol{\alpha}^{(\mathrm{CW})})}$ codewords within the angular range $\boldsymbol{\alpha}^{(\mathrm{CW})}$ the codebook size $N^{(\boldsymbol{\Omega})}_{(\alpha)}$ in Eq. (2.54) over the entire angular range is increased until $N^{(\boldsymbol{\Omega})}_{(\boldsymbol{\alpha}^{(\mathrm{CW})})}$ is satisfied. The respective effective spectral efficiency according to Eq. (2.59) is given in Fig. 2.23a. In Table B.1 the numerical values of Fig. 2.23a and Fig. 2.16a with and without pilot overhead are compared. Therein, it can be observed that with the adapted angular range $N^{(\boldsymbol{\Omega})} = 16$ achieves the highest effective spectral efficiency for $N^{(\boldsymbol{\Omega})}_{(\alpha)} = N^{(\boldsymbol{\Omega})}_{(\beta)} = 4$ providing a gain of 0.8 bit/s/Hz compared to the full angular range. $N^{(\boldsymbol{\Omega})} = 16$ is also the maximum possible number of CSI reference signals antenna ports per polarization in 5G NR Release 15, see Table 5.2.2.2.1-2 in [3GP17c][7]. Without pilot overhead 99.9 % of the maximum from $N^{(\boldsymbol{\Omega})} = 1024$ is already achieved with $N^{(\boldsymbol{\Omega})} = 256$ for $N^{(\boldsymbol{\Omega})}_{(\alpha)} = N^{(\boldsymbol{\Omega})}_{(\beta)} = 16$.

Remark. *The codebook design for a certain angular range can be straightforwardly extended to discontinuous angular ranges by applying the method multiple times. The unequal distribution of selected codewords indicates that a non-uniform angular quantization can achieve an even higher throughput compared to the investigated uniform distribution, however, this would favor users within the orientation of the array and is omitted due to fairness and coverage reasons.*

[7] In the first column of Table 5.2.2.2.1-2 in [3GP17c] the maximum number of CSI-RS antenna ports is given with 32, however, this always assumes two polarizations, which can be seen by the antenna configurations given by N_1 and N_2 such that $32 = 2N_1N_2$.

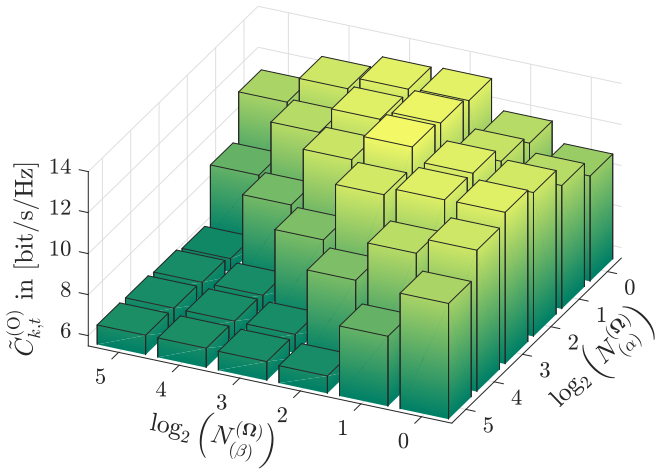


(a) Probability distribution of selected codewords indices for $N_{\alpha}^{(\mathbf{\Omega})} = N_{\beta}^{(\mathbf{\Omega})} = 2^4$ with respect to Eq. (2.56). The dashed line corresponds to the uniform distribution.

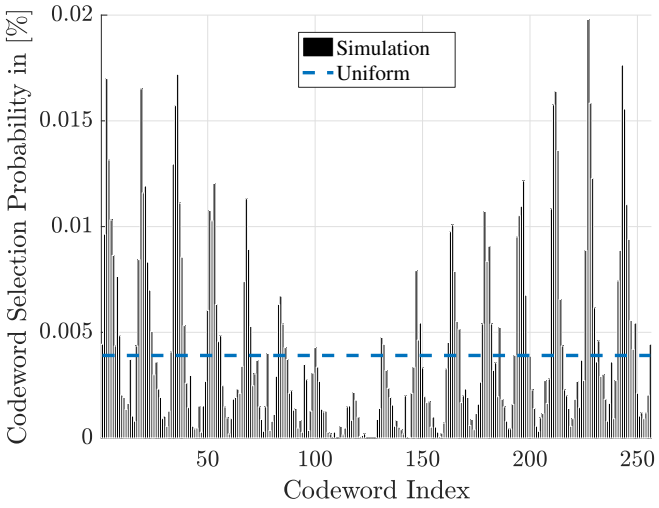


(b) Probability distribution of selected codewords depending on directions. The selection probabilities are given as values in the rectangles and also by the color according to the color-code on the right hand side of the figure.

Figure 2.22.: Probability distribution of codeword selection for a uniform DFT based Kronecker codebook with non-uniform user distribution.



(a) Single user effective spectral efficiency according to Eq. (2.59) taking into account pilot signaling overhead over horizontal and vertical codebook size $N_{(\alpha)}$ and $N_{(\beta)}$, respectively. The UPA dimension is $[10 \times 10]$ and $\tau^{(\mathrm{C})} = 168$.



(b) Probability distribution of selected codewords indices for $N_{(\alpha)}^{(\mathbf{\Omega})} = N_{(\beta)}^{(\mathbf{\Omega})} = 2^4$ with respect to Eq. (2.56). The dashed line corresponds to the uniform distribution.

Figure 2.23.: Performance of DFT codebook with adapted angular range, where $\boldsymbol{\alpha}^{(\mathrm{CW})} = [-49\ 49]^\circ$ and $\boldsymbol{\beta}^{(\mathrm{CW})} = [-40\ 7]^\circ$.

However, the principle trade-off between codebook size and pilot overhead shown in Fig. 2.23a remains. The required per user feedback can be directly obtained from the axes in Fig. 2.16 by the sum of horizontal and vertical codeword index size. More precisely, the feedback rate $R_{k,t}^{(\mathrm{FB})}$ of user k for codeword t is

$$R_{k,t}^{(\mathrm{FB})} = \frac{\log_2\left(N_{(\alpha)}^{(\mathbf{\Omega})}\right) + \log_2\left(N_{(\beta)}^{(\mathbf{\Omega})}\right)}{T^{(\mathrm{FI})}} = \frac{\log_2\left(N_{(\alpha)}^{(\mathbf{\Omega})} N_{(\beta)}^{(\mathbf{\Omega})}\right)}{T^{(\mathrm{FI})}}, \tag{2.72}$$

where $T^{(\mathrm{FI})}$ is the feedback interval and $R_{k,t}^{(\mathrm{FB})}$ is given in bits per feedback interval. Assuming typical LTE parameters, e.g. a $T^{(\mathrm{FI})} = 10\,\mathrm{ms}$, a subband size of 6 RBs [3GP17h][8] and $\log_2\left(N_{(\alpha)}^{(\mathbf{\Omega})} N_{(\beta)}^{(\mathbf{\Omega})}\right) = 16\,\mathrm{bit}$ that results in a feedback rate of $3.33\,\mathrm{kbit/s}$[9]. The sum feedback rate $R_{\tilde{\mathcal{K}}_l}^{(\mathrm{FB,sum})}$ of BS l for the connected set of users $\tilde{\mathcal{K}}_l$ is scaling linearly with $\tilde{K}_l$, where the gradient of the scaling depends on the number of codeword indices reported. Note that in real systems more feedback is required, e.g. link quality indicators such as CQI such that the BS can select a "suitable" modulation and coding scheme. This additional overhead is not considered in this thesis, because the focus is on massive MIMO and implications from pilot signaling and feedback overhead caused by the large number of antennas.

One idea to decrease both the pilot signaling and feedback overhead is to split and distribute large codebooks in time and/or frequency dimension to combine the advantage of fine quantization while keeping the feedback overhead low. This idea is studied in detail in the next paragraphs.

2.4.2. Sub-Codebook Precoding - Single User

The basic principle of the idea on codebook splitting is published by the author of this thesis et. al. in [WKT+15] as a patent and in [KTH+16a] as a conference paper. In this section, further novel aspects and more details are provided.

As before, a codebook $\mathbf{\Omega}$ of size $N \times N^{(\mathbf{\Omega})}$ is considered and divided along the second dimension, the number of codewords $N^{(\mathbf{\Omega})}$, into $N^{(\mathrm{SCB})}$ sub-codebooks of size $N \times N^{(\tilde{\mathbf{\Omega}})}$. Thus, $N^{(\tilde{\mathbf{\Omega}})} \in \mathbb{N}_+$ is considered as a system design parameter that controls the number of precoded pilots per RB, assuming that $N^{(\tilde{\mathbf{\Omega}})} < N$ and $N^{(\tilde{\mathbf{\Omega}})} < N^{(\mathbf{\Omega})}$. It follows that the number of sub-codebooks $N^{(\mathrm{SCB})}$ is obtained as

$$N^{(\mathrm{SCB})} = \left\lceil \frac{N^{(\mathbf{\Omega})}}{N^{(\tilde{\mathbf{\Omega}})}} \right\rceil. \tag{2.73}$$

The $N^{(\mathrm{SCB})}$ sub-codebooks are assigned to disjunctive RBs. Possible mappings between sub-codebooks and RBs are discussed later. The $n_{(\mathrm{SCB})}$-th sub-codebook, where $n_{(\mathrm{SCB})} \in \left\{1, \ldots, N^{(\mathrm{SCB})}\right\}$, is denoted as $\tilde{\mathbf{\Omega}}_{n_{(\mathrm{SCB})}}$. Using these sub-codebooks for precoded pilots such that $\mathbf{B} = \tilde{\mathbf{\Omega}}_{n_{(\mathrm{SCB})}}$ yields the following system relevant aspects:

1. The effective channel $\hat{\mathbf{H}}_{k,l} = \mathbf{H}_{k,l}\tilde{\mathbf{\Omega}}_{n_{(\mathrm{SCB})}}$ in Eq. (2.13) is of dimension $M \times N^{(\tilde{\mathbf{\Omega}})}$ and can be directly estimated by the user k which means that the sub-codebook is transparent to the users and can be changed by the network or BS without knowledge at user k. Consequently, no additional control information has to be signaled from the BS to the user. The complexity of the effective channel estimation is reduced by $\frac{\min\left(N, N^{(\mathbf{\Omega})}\right)}{N^{(\tilde{\mathbf{\Omega}})}}$.
2. With un-precoded pilots and shared codebooks, e.g. as in LTE, user k estimates the channel $\mathbf{H}_{k,l}$ and requires $N^{(\mathbf{\Omega})}$ matrix multiplications of the channel and the codewords to find the best one. With the

[8] The data channel bandwidth is assumed to be 9 MHz that corresponds to 50 RBs. Note that the feedback rate is independent of the coherence time $T^{(\mathrm{C})}$ and if the feedback interval is larger than $T^{(\mathrm{C})}$ this results in loss in spectral efficiency.

[9] The feedback rate can be further decreased by mechanisms such as differential feedback, where a codeword index is selected over the entire bandwidth and per subband a difference value with a smaller value-range is reported.

precoded pilots the best codeword can be directly obtained from estimated effective channel $\hat{\mathbf{H}}_{k,l}$, e.g. by selecting the column with the largest norm.

3. The precoded pilots can also be used for demodulation and no additional demodulation reference signals, as in current LTE, are required when transmitting data[10].

4. The feedback bits to indicate the best codeword of user k reduce from $\log_2\left(N^{(\mathbf{\Omega})}\right)$ to $\log_2\left(N^{(\tilde{\mathbf{\Omega}})}\right)$.

The mapping from the large codebook $\mathbf{\Omega}$ to sub-codebook $\tilde{\mathbf{\Omega}}_{n_{(\mathrm{SCB})}}$ can be described by an vector of indices acording to

$$\tilde{\mathbf{\Omega}}_{n_{(\mathrm{SCB})}} = [\mathbf{\Omega}]_{:,\mathbf{n}^{(\mathrm{SCB})}_{n_{(\mathrm{SCB})}}}, \tag{2.74}$$

where $\mathbf{n}^{(\mathrm{SCB})}_{n_{(\mathrm{SCB})}}$ is of size $N^{(\tilde{\mathbf{\Omega}})} \times 1$ and the elements are from the set of natural numbers in the interval $[1, ..., N^{(\mathbf{\Omega})}]$. In other words, the index vector $\mathbf{n}_{(\mathrm{SCB})}$ selects $N^{(\tilde{\mathbf{\Omega}})}$ codewords from the full codebook $\mathbf{\Omega}$ for sub-codebook $\tilde{\mathbf{\Omega}}_{n_{(\mathrm{SCB})}}$. The sub-codebook splitting exploits the spatial beam structure discussed in the previous paragraph.

The optimum design of $\mathbf{n}^{(\mathrm{SCB})}_{n_{(\mathrm{SCB})}}$ with respect to the spectral efficiency is a combinatorial problem with complexity $\binom{N^{(\mathbf{\Omega})}}{N^{(\tilde{\mathbf{\Omega}})}}$ for a single sub-codebook designed independently from the others. However, the $N^{(\mathrm{SCB})}$ codebooks have to be designed jointly such that $\mathbf{n}^{(\mathrm{SCB})}_1, \ldots \mathbf{n}^{(\mathrm{SCB})}_{N^{(\mathrm{SCB})}}$ are disjunctive sets. The disjunctive property ensures that all codewords and therefore implicitly beam directions have the same probability for the user to select from. Without prior knowledge about the user position distribution preferring certain directions is not necessarily beneficial in terms of spectral efficiency. Furthermore, a uniform and equally spaced codeword selection from the full codebook is assumed, such that sub-codebook $\tilde{\mathbf{\Omega}}_{N^{(\mathrm{SCB})}}$ equals the "small" full codebook $\mathbf{\Omega}\left(N^{(\mathbf{\Omega},\mathrm{SCB})}_{(\alpha)}, N^{(\mathbf{\Omega},\mathrm{SCB})}_{(\beta)}\right)$ according to Eq. (2.56), where $N^{(\mathbf{\Omega},\mathrm{SCB})}_{(\alpha)}$ and $N^{(\mathbf{\Omega},\mathrm{SCB})}_{(\beta)}$ denote the number of horizontal and vertical directions such that $N^{(\mathbf{\Omega},\mathrm{SCB})}_{(\alpha)} N^{(\mathbf{\Omega},\mathrm{SCB})}_{(\beta)} = N^{(\tilde{\mathbf{\Omega}})}$. The other sub-codebooks $\tilde{\mathbf{\Omega}}_1, \ldots, \tilde{\mathbf{\Omega}}_{N^{(\mathrm{SCB})}-1}$ can be interpreted as angular shifted versions of $\tilde{\mathbf{\Omega}}_{N^{(\mathrm{SCB})}}$, with the same angular distance between the codewords.

The algorithm of the sub-codebook construction is straightforward and given as Matlab code in the appendix in Algorithm 4. The example in Fig. 2.24 with $N^{(\mathbf{\Omega})}_{(\alpha)} = N^{(\mathbf{\Omega})}_{(\beta)} = 16$ and $N^{(\mathbf{\Omega},\mathrm{SCB})}_{(\alpha)} = N^{(\mathbf{\Omega},\mathrm{SCB})}_{(\beta)} = 4$ illustrates how the sub-codebooks are constructed. The selection of $N^{(\mathbf{\Omega},\mathrm{SCB})}_{(\alpha)}$ and $N^{(\mathbf{\Omega},\mathrm{SCB})}_{(\beta)}$ in the example in Fig. 2.24 is based on the numerical results in Fig. 2.23 where maximum effective spectral efficiency is achieved with $N^{(\mathbf{\Omega})}_{(\alpha)} = N^{(\mathbf{\Omega})}_{(\beta)} = 4$. Note that without limitation to generality codeword indices according to Fig. 2.25 are assumed for the sub-codebook construction in Algorithm 4.

Next, the sub codewords $\tilde{\mathbf{\Omega}}_{n_{(\mathrm{SCB})}}$ are frequency multiplexed as shown in Fig. 2.26. In principle the sub-codebooks can also be distributed into the time dimension within the coherence time of the channel, however here only frequency multiplex is considered due to simulation complexity[11]. Note that the multiplexing of sub-codebooks to orthogonal time-frequency resources can be subject to further optimization depending on the user density and traffic distribution, however, it is not in the focus of this thesis and therefore only the distributed mode as shown in Fig. 2.26 is used. The number of orthogonal frequency resource is denoted by $N^{(\mathrm{RB})}$, e.g. in LTE the smallest unit for user selection in the frequency dimension is a physical RB with 180 kHz bandwidth. The same resolution is assumed in this thesis for sub-codebook resolution in the frequency domain. If the number of sub-codebooks $N^{(\mathrm{SCB})}$ is smaller than the number of physical RBs $N^{(\mathrm{RB})}$ such that $N^{(\mathrm{SCB})} < N^{(\mathrm{RB})}$, several or all sub-codebooks appear multiple times. For example in Fig. 2.25 sub-codebook $\tilde{\mathbf{\Omega}}_1$ is assigned to RB 1 and again on RB 17 assuming $N^{(\mathrm{SCB})} = 16$ sub-codebooks. With this the term sub-codebook block is introduced. A sub-codebook block is a complete set of all sub-codebooks

[10] Demodulation reference signals are precoded reference signals with the precoder used for downlink data transmission that is determined by the BS after scheduling.

[11] Time evolving channels would linearly increase the simulation complexity and storage.

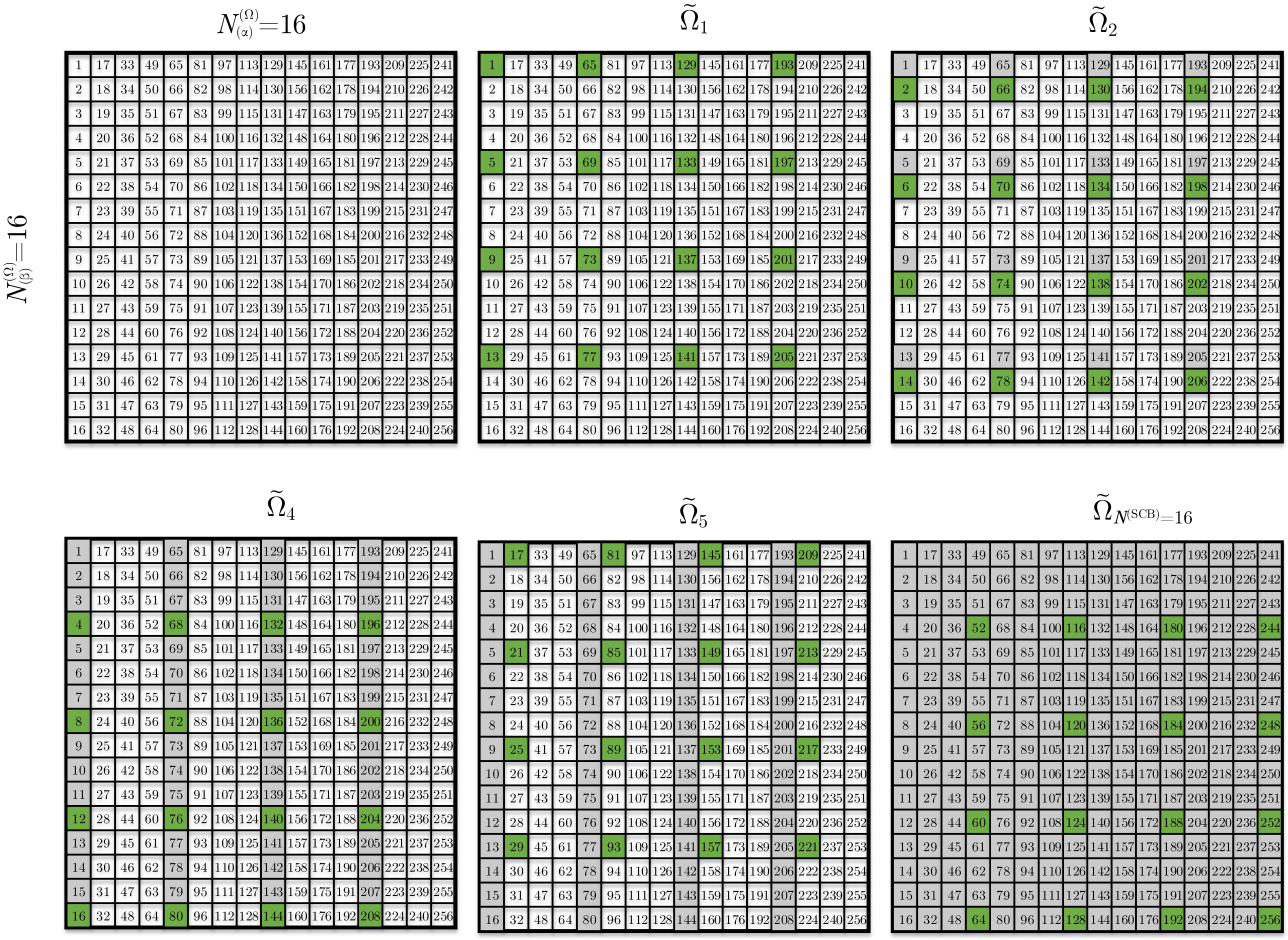


Figure 2.24.: Example of sub-codebook construction $\tilde{\mathbf{\Omega}}_{n_{(\mathrm{SCB})}}$ of size $N_{(\alpha)}^{(\mathbf{\Omega},\mathrm{SCB})} = N_{(\beta)}^{(\mathbf{\Omega},\mathrm{SCB})} = 4$ from a "large" codebook of size $N_{(\alpha)}^{(\mathbf{\Omega})} = N_{(\beta)}^{(\mathbf{\Omega})} = 16$. The green filled squares are codewords of sub-codebook $n_{(\mathrm{SCB})}$, while gray filled squares represent codewords from sub-codebooks with a lower index $n_{(\mathrm{SCB})}$. Codewords have maximum average angular distance from each other to keep inter-stream interference low.

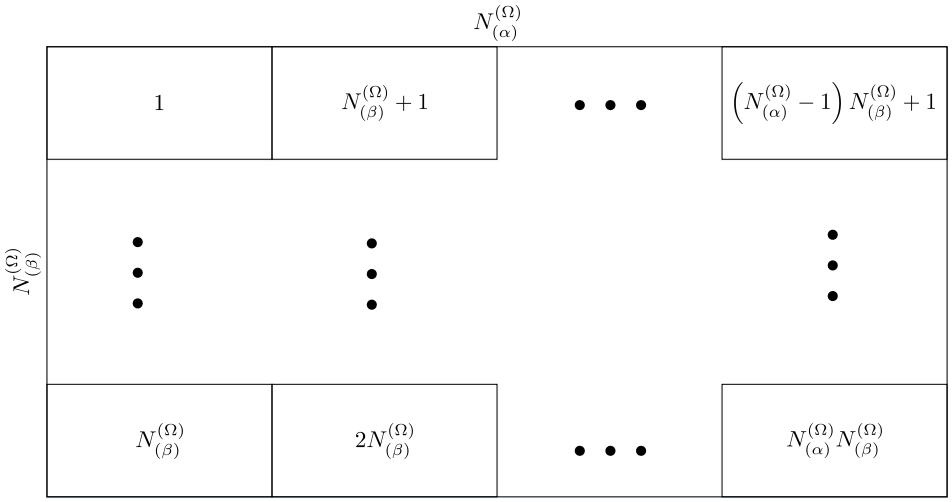


Figure 2.25.: Codeword indexing in the full codebook with respect to the number of horizontal and vertical codewords $N_{(\alpha)}^{(\mathbf{\Omega})}$ and $N_{(\beta)}^{(\mathbf{\Omega})}$, respectively.

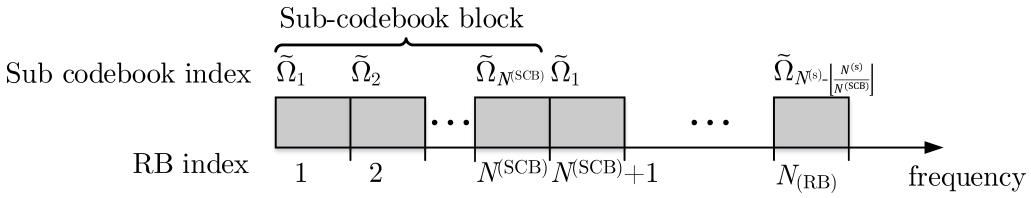


Figure 2.26.: Sub-codebook multiplexing in frequency domain to provide user with selection diversity.

Table 2.6.: Codebook specific simulation assumptions for performance evaluation of sub-codebook splitting.

Parameter	**Value**
SINR bounds, $\gamma^{(\min)}, \gamma^{(\max)}$	$\gamma^{(\min)} = -5\,\mathrm{dB}$, $\gamma^{(\max)} = 100\,\mathrm{dB}$
Number of available users	1, or given
Number of selected users	1, or according to Algorithm 1
$N_{(\alpha)}, N_{(\beta)}, N$	$10, 10, 100$
Full "large" codebook: $N_{(\alpha)}^{(\mathbf{\Omega})}, N_{(\beta)}^{(\mathbf{\Omega})}, N^{(\mathbf{\Omega})}$	$16, 16, 256$
Full "small" codebook: $N_{(\alpha)}^{(\mathbf{\Omega})}, N_{(\beta)}^{(\mathbf{\Omega})}, N^{(\mathbf{\Omega})}$	$4, 4, 16$
	$16, 16, 256$
$\alpha^{(\mathrm{CW,min})}, \alpha^{(\mathrm{CW,max})}$	$-50°, 50°$
$\beta^{(\mathrm{CW,min})}, \beta^{(\mathrm{CW,max})}$	$-40°, 8°$
Sub-codebooks: $N_{(\alpha)}^{(\mathbf{\Omega},\mathrm{SCB})}, N_{(\beta)}^{(\mathbf{\Omega},\mathrm{SCB})}, N^{(\tilde{\mathbf{\Omega}})}$	$4, 4, 16$
$N^{(\mathrm{SCB})}$	According to Algorithm 4

from $1, \ldots, N^{(\mathrm{SCB})}$, e.g. with the settings in Table 2.6 with $N^{(\mathrm{SCB})} = 16$ sub-codebooks and $N^{(\mathrm{RB})} = 50$ physical RBs, there are 3 sub-codebook blocks.

A summary of the codebook specific parameters and their setting for the following performance evaluation is given in Table 2.6. The parameter values are derived based on numerical simulations as explained above. Fig. 2.27a compares the following four cases:

1. All sub-codebooks: This means that the 16 sub-codebooks are distributed in the frequency domain according to Fig. 2.26. Each user selects on each RB his "best" codeword of the sub-codebook according to Eq. (2.57). This BS than applies the selected codeword for transmission to the user.

2. Max. of sub-codebook block: Again the 16 sub-codebooks are distributed in the frequency domain according to Fig. 2.26. In contrast to case 1, each user selects his best codeword according to Eq. (2.57) per sub-codebook block and not per RB. With the parameters in Table 2.6 the user selects his first codeword from RBs 1-16, the second codeword from RBs 17-32, the third codeword from RB 33-48 and the forth from RBs 49-50. While the last two RBs are not a complete sub-codebook block the user can report a codeword in order not to waste these resources for downlink transmission.

3. Full codebook 16×16: This case is the reference and the spectral efficiency of the full codebook $N_{(\alpha)}^{(\mathbf{\Omega})} = N_{(\beta)}^{(\mathbf{\Omega})} = 16, N^{(\mathbf{\Omega})} = 256$ that is used for the sub-codebook splitting in the previous two cases is given. This codebook is used on all RBs.

4. Full codebook 4×4 - This case is for comparison with the sub-codebook splitting and the spectral efficiency of the "small" codebook $N_{(\alpha)}^{(\mathbf{\Omega})} = N_{(\beta)}^{(\mathbf{\Omega})} = 4, N^{(\mathbf{\Omega})} = 16$ having the same size as the sub-

codebook is given. The the same "small" codebook is applied on all RBs.

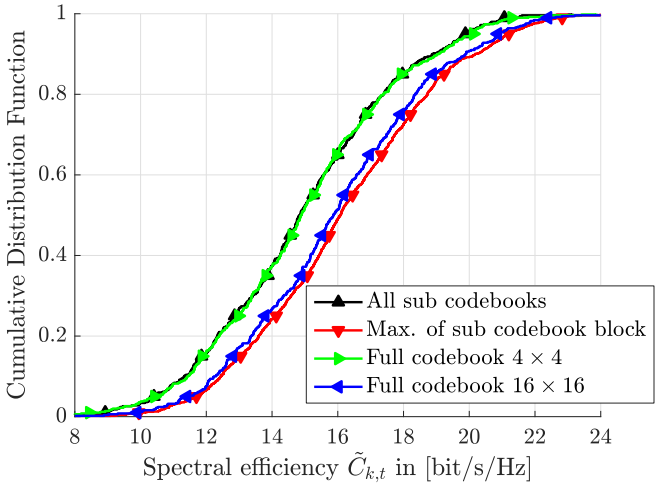
The first observation in Fig. 2.27a is that the spectral efficiency of the sub-codebooks (case 1) overlaps with case 4 the "small" full codebook of size 4×4, that is of the same size as the sub-codebooks. This verifies that the proposed sub-codebook splitting strategy achieves similar performance with each sub-codebook. Second, the spectral efficiency of the "large" full codebook of size 16×16 achieves a higher spectral efficiency than case 1 and case 4, e.g. 1 bit/s/Hz more at the 50 %-ile. However, the main observation is that the sub-codebook splitting with codeword selection out of a sub-codebook block achieves a higher spectral efficiency than the "large" full codebook over the complete range of the CDF, e.g. 0.2 bit/s/Hz at the 50 %-ile. Note, that the original motivation for sub-codebook splitting was the reduction of pilot overhead, that can be achieved without performance loss compared to the full codebook. However, some discussion is required to interpret this result correctly. In the full codebook of size 16×16 a single user is randomly assigned to each RB and receives data on his selected codeword. In contrast to this, in case 2 where each user selects the best codeword out of the sub-codebook block, there are now 16 users that can select their best codeword over 16 RBs. Thus, the gain for case two is actually a frequency diversity gain inherently utilized by the codeword selection over the sub-codebook block in case 2. Therefore, the frequency diversity of the simulated scenario is investigated next for clarification.

Fig. 2.27b shows the frequency selectivity of the effective channel for codeword 1 of the full 16×16 codebook in comparison to MRT precoding for the $[10, 10]$ UPA. This frequency selectivity of the codebook contradicts the findings of "channel hardening" in Section 2.3. Channel hardening with a practical number of antennas requires phase adapted precoding, however with DFT based codebooks only "directions" are approximated. Thus a part of the frequency selectivity of the channel remains, e.g. the variance with respect to the median value over 50 subcarriers is 19 dB and 1 dB for codebook and MRT precoding, respectively. A ULA with 10 elements has a half-power beam-width in the main radiation direction of approximately 10°. With 80° azimuth spread and 52° elevation spread at the BS in the simulated urban macro NLoS scenario [3GP17e] a lot of the multiple path components remain with codebook precoding. Increasing the number of antennas the beam-width becomes smaller and the effective channel with codebook precoding becomes less frequency selective. In the extreme case the beam-width is so small that only a single path will remain visible above the noise threshold. One finding of this is, that frequency selective scheduling is still beneficial w.r.t. spectral efficiency in massive MIMO FDD systems using codebook based precoding.

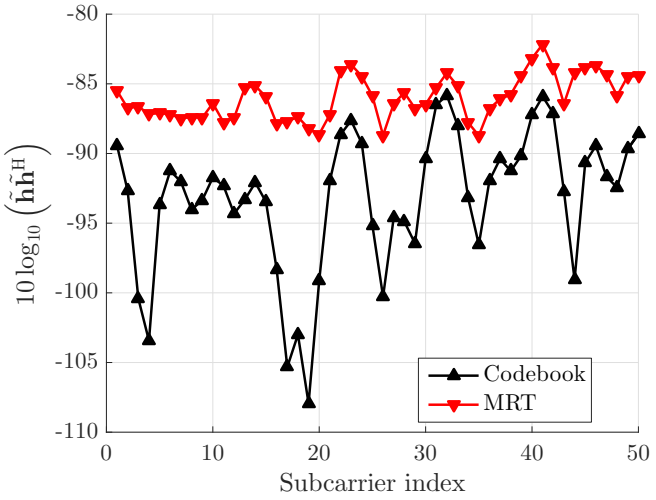
Therefore, the sub-codebook precoding is next combined with frequency selective opportunistic scheduling to further increase spectral efficiency. In Fig. 2.28 the number of available users is increased and on each sub-codebook the user with largest effective channel power according to Eq. (2.57) over all codewords is selected. Note, that the assignment of sub-codebooks to RBs is still the same as in Fig. 2.26. Only the assignment of users to the sub-codebooks and codewords is done by the BS. The assumption here is that each users reports per RB for the respective sub-codebook the power values of the best codeword. That is similar to PMI and CQI reporting in LTE. This means, that per RB the BS searches the maximum value out the $\tilde{K}$ feedback values. It can be observed in Fig. 2.28 that the sub-codebook splitting utilizes the same user diversity gains as MRT with an approximately constant gap. Overall, the sub-codebook splitting can achieve the same beamforming gain as MRT due to the high channel quantization but with less than 10 % pilot overhead, assuming one pilot per codeword on a RB resolution in LTE.

Remark. *It is shown that with the proposed sub-codebook splitting scheme beamforming gains similar to MRT can be achieved in a FDD system while keeping the pilot overhead low, e.g. in the shown example by less than* 10 %*.*

Remark. *With codebook based precoding, frequency aware user scheduling is still required in massive MIMO FDD systems, because the channel hardening effect does not kick in without channel aware phase adaptation in the precoder.*



(a) Spectral efficiency performance of codebook splitting, where the reference cases 4×4 full codebook and 16×16 full codebook consider no sub-codebook splitting. 4×4 corresponds to the same size as the sub-codebooks and 16×16 to the full codebook before splitting. In case of maximum of sub-codebook block, the best codebook out of 16 sub-codebooks in the block is selected.



(b) Channel hardening comparing codebook and MRT precoder which correspond to channel phase adaptive and non-phase adaptive precoders, respectively.

Figure 2.27.: Sub-codebook splitting evaluation.

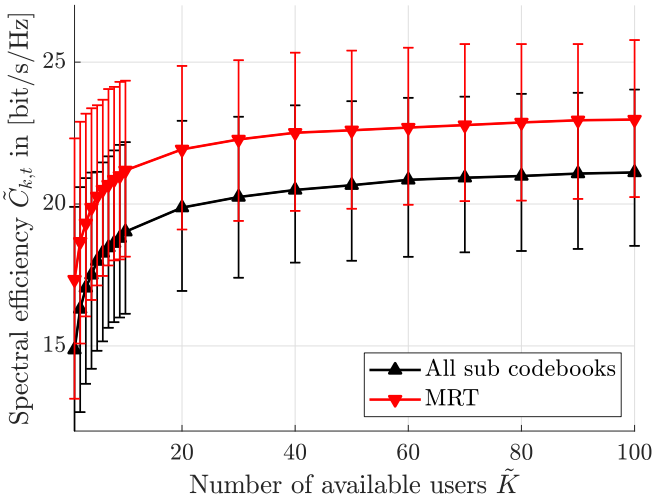


Figure 2.28.: Spectral efficiency over the number of available users in the cell.

2.4.3. Sub-Codebook Precoding - Multiple Users

In the above investigation on sub-codebook splitting, single user transmission ($T_l = 1$) was assumed to derive the trade-offs. However, the main benefit of massive MIMO is the offered spatial multiplexing gain. Therefore, the extension to multiple user transmission is in the focus of this Section.

Multiple-user selection in codebook based FDD systems is a typical chicken egg problem, meaning that codeword selection at the BS generates unknown interference towards other users. Thus multiplexing of additional users can result in sum spectral efficiency reduction. The same effect was shown in Section 2.3.4 for SUS. On the other hand, without beam selection at the BS users do not know which beams are interference while reporting CSI in the uplink. In contrast to this, the proposed sub-codebook splitting has the advantage that the set of available codewords is already known at the user, because the set of codewords for data transmission is a subset of the codewords used for precoded pilots and channel estimation. Thus, interference can be estimated at the user side and taken into account for CSI feedback. Keeping the same assumptions as in Section 2.3 it is assumed that the transmit power is equally divided to the T_l streams. The number of fed back codewords per time-frequency block per user is denoted by $N_{(\mathrm{FB})}^{(\mathrm{CW})} \in \mathbb{N}_+$ bounded by $N_{(\mathrm{FB})}^{(\mathrm{CW})} \leq N^{(\tilde{\mathbf{\Omega}})}$ and refers to the $N_{(\mathrm{FB})}^{(\mathrm{CW})}$ first elements of the descend ordered list of power values $\mathbf{P}_{(\mathrm{FB})}^{(\mathrm{eff})} \in \mathbb{R}_+^{N^{(\tilde{\mathbf{\Omega}})}}$ of the effective channel such that

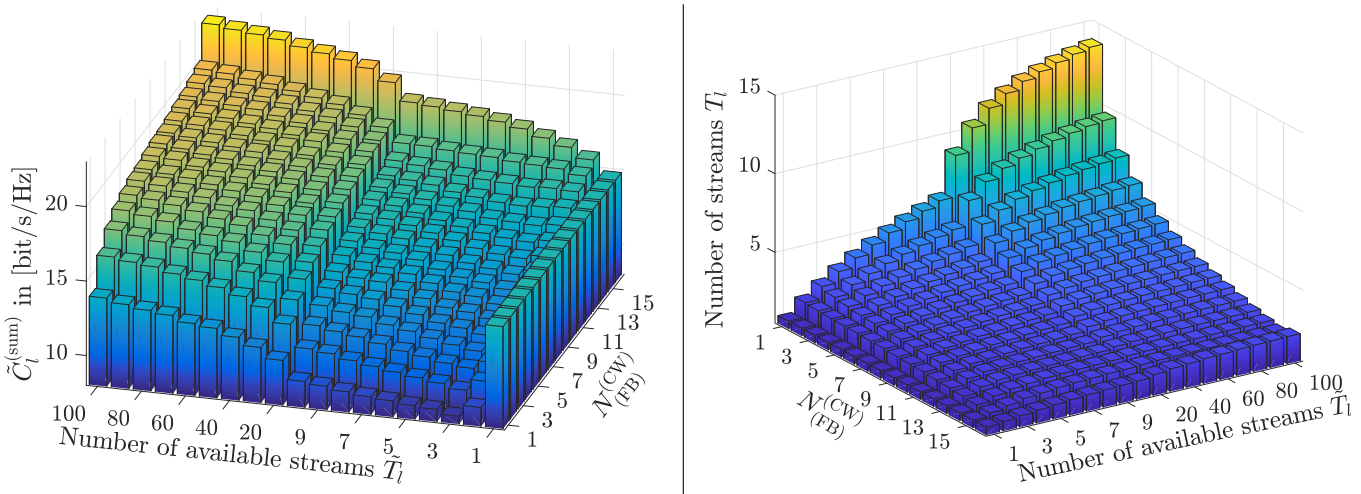
$$\left[\mathbf{P}_{(\mathrm{FB})}^{(\mathrm{eff})}\right]_1 \geq \left[\mathbf{P}_{(\mathrm{FB})}^{(\mathrm{eff})}\right]_2 \ldots \geq \left[\mathbf{P}_{(\mathrm{FB})}^{(\mathrm{eff})}\right]_{N_{(\mathrm{FB})}^{(\mathrm{CW})}}, \tag{2.75}$$

where $\left[\mathbf{P}_{(\mathrm{FB})}^{(\mathrm{eff})}\right]_1 = \arg\max_i \left|\left[\hat{\mathbf{H}}\right]_{1,i}\right|^2$ and $\hat{\mathbf{H}}$ is assumed of size $1 \times N^{(\tilde{\mathbf{\Omega}})}$[12]. A heuristic scheduler that aims to maximize the sum spectral efficiency given by pseudo-code in Algorithm 5 is used for performance evaluation. An optimal approach is not available and to try all possible combinations is too computational expensive for the same reasons as given at the beginning of Section 2.3.4. Briefly summarized, the algorithm selects at each step the stream that maximizes the estimated sum rate. If the sum rate is not increasing the algorithm stops. It is worthwhile to mention, that by using power values for feedback in Eq. (2.75) the worst case interference scenario is considered, e.g. the power of interfering codewords is summed up, which corresponds to coherent superposition of interference in the same subspace as the signal. The true interference is equal or lower and therefore the algorithm has the tendency to stop adding streams before the true optimum is achieved.

Fig. 2.29a shows the sum spectral efficiency $\tilde{C}_l^{(\mathrm{sum})}$ over the two dimensional parameter space, the number of available streams and the number of reported codewords, with corresponding numerical values given in Table B.2. Leaving out the single user case $\tilde{T}_l = 1$ where no inter-stream interference is generated, the sum spectral efficiency is monotonically increasing with the number of available streams and with the number of reported streams. However, there is a wide range of low $N_{(\mathrm{FB})}^{(\mathrm{CW})}$ where the sum spectral efficiency is lower than the single user performance. This demonstrates that without the respective interference information, a sum spectral efficiency gain cannot be ensured in massive MIMO FDD systems. The maximum gain with $N_{(\mathrm{FB})}^{(\mathrm{CW})} = 16$ compared to single user is less than two times which is much smaller than the five times gains with full CSI knowledge in Section 2.3.4. Accompanying Fig. 2.29 shows the number of selected streams T_l corresponding to Fig. 2.29a, numerical values are listed in Table B.3. With $N_{(\mathrm{FB})}^{(\mathrm{CW})} = 16$ and $\tilde{T}_l = 100$, that corresponds to the largest achieved sum spectral efficiency, on average $T_l = 2.2$ streams are selected indicating large inter-stream interference. This is contradicting the intention of massive spatial multiplexing in massive MIMO and emphasizes that without multiple-user user interference suppression the multiplexing gain can't be realized. This indicates that pure codebook based precoder design for multiple-user transmission is not a good choice. The next paragraphs clarify this further.

Fig. 2.30a compares the sub-codebook splitting with the full codebook and MRT precoding. The sub-codebook splitting and full codebook assume feedback of all available codewords on each RB such that

[12] If the number of antennas at the user is $M > 1$ the effective channel is assumed as MISO feedback according to Eq. (2.18).



(a) Sum spectral efficiency $\tilde{C}_l^{(\mathrm{sum})}$.

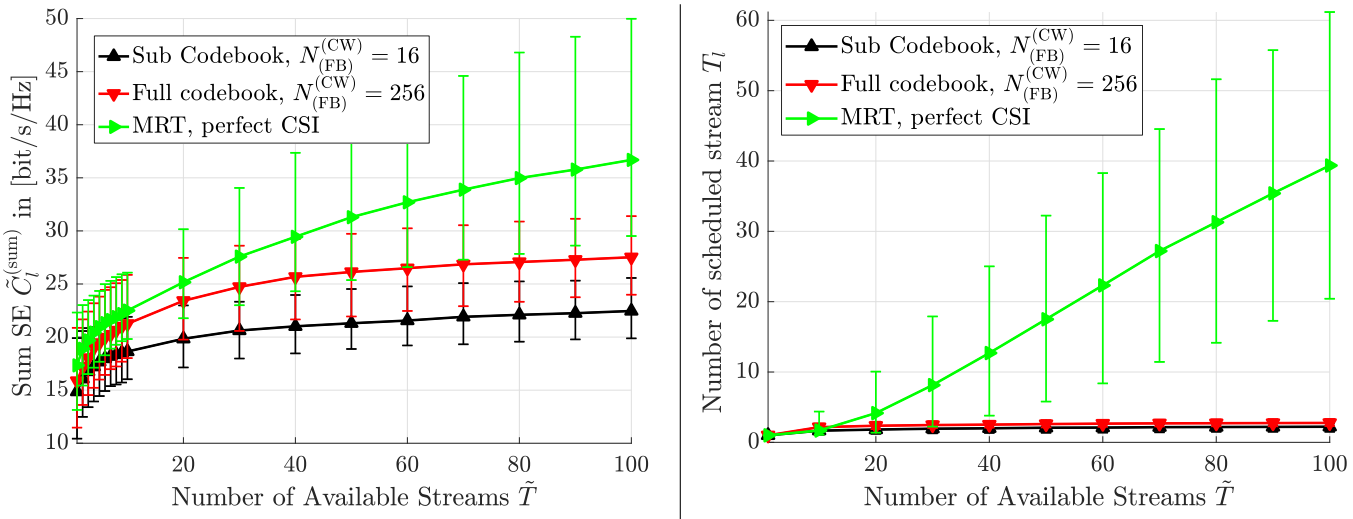
(b) Number of selected streams T_l.

Figure 2.29.: Performance evaluation of multiple-stream transmission with sub-codebook splitting according to Table 2.6 over the number of available streams $\tilde{T}_l$ and the number of fed back streams $N_{(\mathrm{FB})}^{(\mathrm{CW})}$.

$N_{(\mathrm{FB})}^{(\mathrm{CW})} = 16$ and $N_{(\mathrm{FB})}^{(\mathrm{CW})} = 256$, respectively. At a low number of available users codebook based precoding can utilize the same multiplexing gains as MRT. Increasing the number of available users increases the gap to MRT, while the gap between full and sub-codebook splitting remains approximately constant with 5 bit/s/Hz. Fig. 2.30b shows the number of scheduled streams. With MRT the number of selected streams scales approximately linearly while for codebook based precoding it remains constant. That means the sum spectral efficiency increase in Fig. 2.29a is achieved by increased SINR of the same number of scheduled users with the codebook based approaches. With perfect CSI and MMSE precoding ≈ 500 bit/s/Hz are achieved in Fig. 2.10a, thus less than 5% of the TDD sum spectral efficiency with MMSE is achieved in the codebook based FDD system.

One advantage that is not investigated in this work is that with precoded pilots and the a-priori defined sub-codebook, user k is able to estimate multiple-user interference from other streams unequal to the serving stream t. This estimated interference can be taken into account for SINR estimation to improve link adaptation. Link adaptation means that according to the estimated SINR user k selects CQI corresponding to a desired MCS. This selection usually aims for a certain block error rate. The smaller the difference between the estimated and true SINR, the smaller the link-adaptation loss. There are two types of link-adaptation loss. First, if the estimated SINR is larger than the true SINR the block error rate increases and reduces the realized data-rate. Second, if the estimated SINR is lower, then a lower data-rate is realized than possible by the channel link capacity. In LTE users report their CQI based on the estimated SNR resulting in average realized data rates much lower than theoretically possible [DNO+16]. Consequently, one trade-off of any codebook based precoding approach is to reduce the inter-stream interference while providing enough coverage that users find a stream close to the optimum. One solution for the optimum beam selection is offered in this work by the sub-codebook approach with users selecting their codeword over the complete sub-codebook block in Fig. 2.27. The inter-stream interference is taken into account by the design of the sub-codebooks with the maximum angular distance criteria between streams, see Fig. 2.24.

Remark. *Massive MIMO multiplexing gains cannot be utilized with codebook based precoding if inter-stream interference is the limiting factor.*



(a) Sum spectral efficiency (SE) $\tilde{C}_l^{(\mathrm{sum})}$.

(b) Number of selected streams T_l.

Figure 2.30.: Performance evaluation of multiple-stream transmission comparing sub-codebook splitting with the full codebook, according to Table 2.6, and MRT precoder assuming perfect channel knowledge.

2.4.4. Hybrid Precoding

The previous section shows that with user selection based on receive power information of the codewords and codebook based beamforming less than 5% compared to full CSI knowledge is achieved. Therefore, this section takes into account the complete effective user channels $\hat{\mathbf{H}}_{\tilde{k},l} = \mathbf{H}_{\tilde{k},l}\mathbf{B}_l$ from Eq. (2.13) for feedback, instead of only power information as in the previous section. Thus, multiple-user precoding techniques from Section 2.3.1 can be used for inter-stream interference reduction for the second stage precoder $\mathbf{P}_l$ according to Eq. (2.14), following the system model description in Section 2.2. In case of multiple receive antennas the equivalent MISO channels according to Eq. (2.18) are considered. This requires, not only power values but also phase information about the effective channel $\hat{\mathbf{H}}$. The trade-offs between feedback and sum spectral efficiency is studied in the next section and therefore not taken into account in this section. By using the sub-codebook splitting from Section 2.4.2 as first stage precoders $\mathbf{B}_l$, each user observes a precoded effective channel of the same size and the same pilot overhead as in Section 2.4.2. Fig. 2.31 compares the sum spectral efficiency of the following cases:

1. Sub-codebook based beamforming according to Table 2.6 with power values and beam index feedback using the sum spectral efficiency maximizing heuristic according to Algorithm 5.

2. MRT precoding with perfect CSI with sum spectral efficiency maximizing heuristic according to Algorithm 5.

3. Hybrid precoding with sub-codebooks as first stage precoder, MMSE as second stage precoder, and perfect CSI of the effective channel using PBZF scheduling according to Algorithm 1.

4. MMSE with perfect CSI using PBZF scheduling according to Algorithm 1.

In Fig. 2.31, it can be observed that the hybrid precoding achieves 4-5 times of the sum spectral efficiency compared to sub-codebook based or MRT precoding, due to the multiple-user interference suppression. Furthermore, up to $\tilde{T} = 10$ the hybrid precoding scales similarly to MMSE precoding, that is considered here as an upper performance bound. At $\tilde{T} = 20$ the gap is significant because the number of multiplexed streams T is bounded by the number of codewords in the sub-codebooks, $T \leq N^{(\tilde{\mathbf{\Omega}})}$ that is sixteen in the setting of Fig. 2.31. Thus the hybrid precoding cannot follow the linear scaling of the MMSE for $\tilde{T} > N^{(\tilde{\mathbf{\Omega}})}$. Note that for the hybrid precoding simulation in Fig. 2.31 the upper SINR was set again to $\gamma^{(\max)} = 10^{\frac{40}{10}}$,

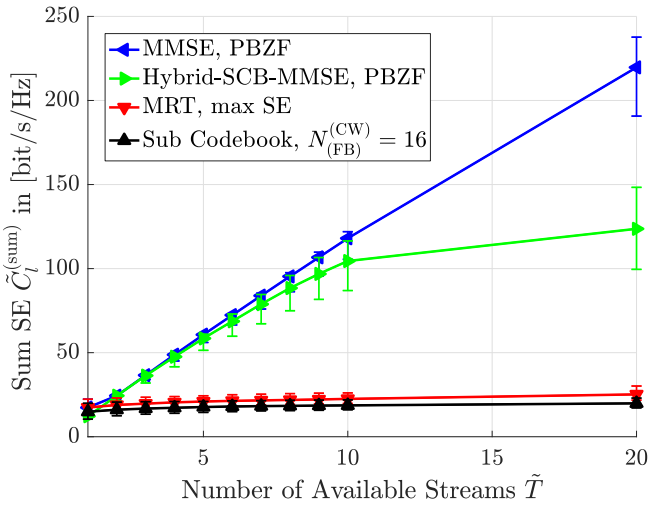


Figure 2.31.: Sum spectral efficiency (SE) comparison of a) Sub-codebook based beamforming according to Table 2.6 with feedback of power values and beam indices, b) MRT precoding with perfect CSI, c) Hybrid precoding using sub-codebooks as first stage precoder and MMSE as second stage precoder with perfect CSI knowledge of the effective channel, and d) full digital MMSE precoding with perfect CSI knowledge.

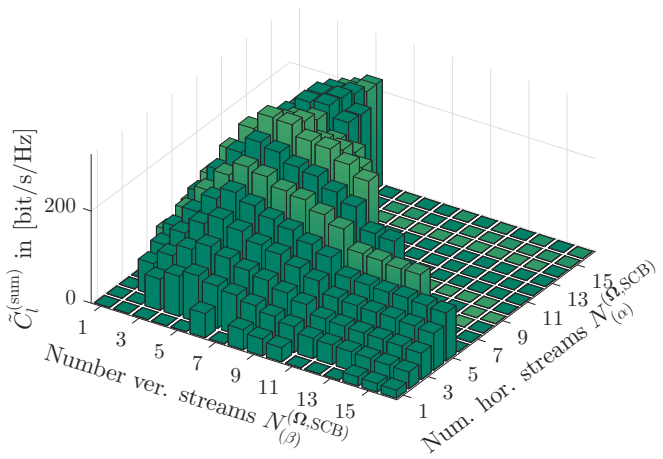
the same value as for MMSE.

It is intuitive that by adding more codewords to each sub-codebook the sum spectral efficiency can be increased, but similar as in Fig. 2.23a the question is whether the effective sum spectral efficiency, taking into account pilot overhead, is also increased. Thus, Fig. 2.32b shows the effective sum spectral efficiency according to Eq. (2.59) over $N^{(\tilde{\mathbf{\Omega}})}$ the number of codewords in the sub-codebooks with sufficient available streams such that $\tilde{T} > N^{(\tilde{\mathbf{\Omega}})}$. For comparison the sum spectral efficiency without the pilot overhead is given next to it in Fig. 2.32a. First, it can be observed that the sum spectral efficiency increases with the number of codewords in both dimension, horizontal and vertical. However, taking the pilot overhead into account with one resource element (RE) per codeword per LTE RB with 168 REs, the maximum is achieved for $N_{(\alpha)}^{(\mathbf{\Omega},\mathrm{SCB})} = 8$ and $N_{(\beta)}^{(\mathbf{\Omega},\mathrm{SCB})} = 6$ with an effective spectral efficiency of $\approx 247\,\mathrm{bit/s/Hz}$. The corresponding data values to Fig. 2.32a and Fig. 2.32b are given in the annex in Table B.4 and Table B.5, respectively. Simulations, where values are zero have been omitted, either because the number of sub-codebook would be larger than the number RBs or the pilot overhead would be larger than 50 %.

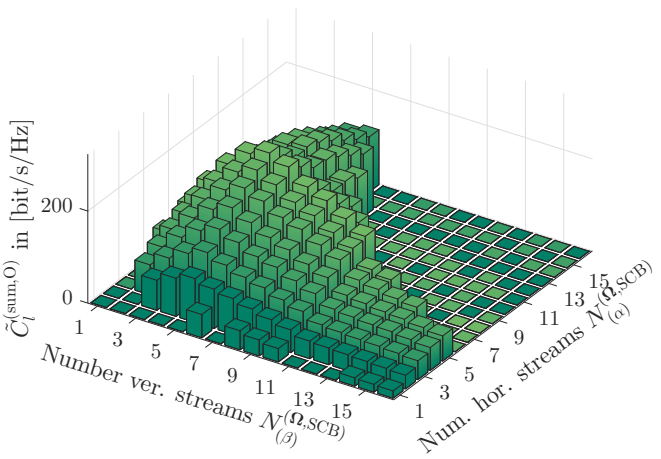
From a solely sum spectral efficiency perspective hybrid precoding outperforms codebook based precoding, due to the additional interference mitigation by the second stage precoder. In combination with the proposed sub-codebook splitting, a tunable solution to adjust for the pilot overhead multiplexing trade-off is presented that scales up to the number of codewords $N^{(\tilde{\mathbf{\Omega}})}$ similar to TDD performance. On the other hand, in the current NR Release 15, still codebook based precoding is considered, due to restrictions in the uplink feedback channel. Therefore, a critical question that has to be answered by all proposed massive MIMO schemes for FDD systems is the following: How much of the performance with perfect CSI assumption can be achieved with a limited feedback channel?

2.4.5. Feedback - Sum Spectral Efficiency - Trade-Off

In this section, the trade-off between feedback and sum spectral efficiency is investigated. The target is to realize as much of the sum spectral efficiency performance of perfect CSI while assuming similar feedback as in Release 15 systems [AVW18]. Fist the performance degradation due to quantization is studied. Based on these findings further feedback reduction is investigated with special focus on the total feedback.



(a) Sum spectral efficiency (SE) without pilot overhead.



(b) Effective sum spectral efficiency (SE) with pilot overhead assuming one RE per pilot per LTE RB consisting of 168 REs.

Figure 2.32.: Performance evaluation of hybrid precoding with a sum spectral efficiency maximizing scheduling heuristic over sub-codebook sizes used as first stage precoder. The second stage precoder is MMSE and the number of available streams is $\tilde{T}_l = 100$.

For feedback of CSI values, an amplitude and phase quantization is required. While the range of the phase values is constant in $[0, 360]^\circ$ the range of the amplitude values is variable, e.g. due to dependency on path-loss among others. Therefore, also the range of amplitude quantization is of relevance in this investigation. Therefore, the following notation is introduced:

- $q^{(\mathrm{A})}$ ... Number of bits for amplitude quantization resulting in $2^{q^{(\mathrm{A})}}$ quantization levels.
- $q^{(\mathrm{A,R})}$ ... Amplitude quantization range, where the maximum of the CSI values to be quantized corresponds to the upper quantization level and values below the quantization range are set to zero.
- $q^{(\mathrm{P})}$ ... Number of bits for phase quantization resulting in $2^{q^{(\mathrm{P})}}$ quantization levels.

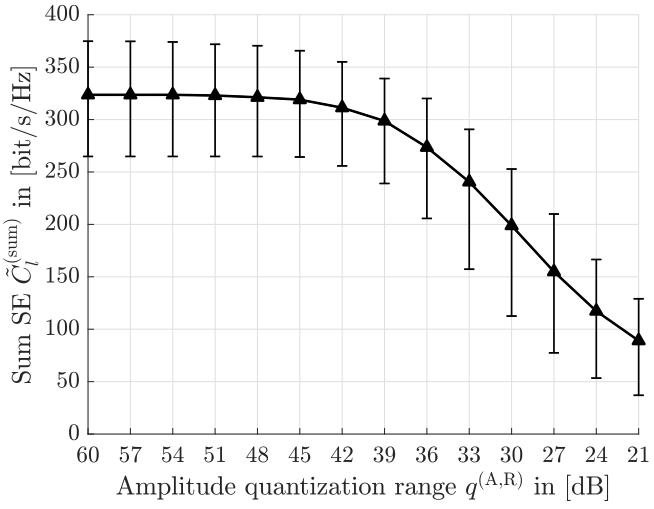
Due to the assumed uniform distribution of the phase and an unknown distribution of amplitude values, a linear uniform quantization is used for both amplitude and phase. The amplitude quantization range $q^{(\mathrm{A,R})}$ refers to the maximum amplitude of CSI values to be quantized and can be different for each user, whereas the phase range is the same for all users. The maximum amplitude of the CSI as a reference value is not taken into account for the short-term feedback calculation. Such a reference value can be sent per user with a different feedback interval, e.g. depending on the velocity of users. Since the amplitude depends on the path-loss and changes slowly compared to the phase, a larger feedback interval is foreseen and the overhead can be neglected compared to the CSI feedback. Alternatively, other existing feedback can be used, e.g. in LTE the reference signal receive power (RSRP) [3GP14]. Furthermore, the amplitude range $q^{(\mathrm{A,R})}$ effectively introduces a power based threshold where CSI values below are set to zero. Table 2.7 lists the simulation parameters for the investigation on the quantization values.

Fig. 2.33a shows the sum spectral efficiency over the quantization range $21\,\mathrm{dB} \leq q^{(\mathrm{A,R})} \leq 60\,\mathrm{dB}$ for "high" amplitude and phase quantization, $q^{(\mathrm{A})} = 16\,\mathrm{bit}$ and $q^{(\mathrm{P})} = 16$ bit, respectively. The high amplitude and phase quantization is selected in order to isolate the spectral efficiency impact caused by the change of the amplitude range. Reducing the amplitude range in Fig. 2.33a down to 54 dB achieves approximately the same sum spectral efficiency as the un-quantized CSI[13], see Table B.4. At a amplitude quantization range of 45 dB the sum spectral efficiency loss is less than 2 %. This loss is considered acceptable and therefore 45 dB amplitude quantization range is used for further study on phase and frequency quantization.

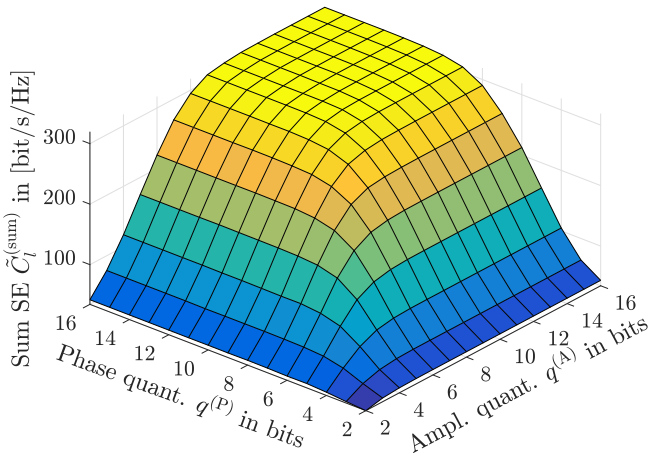
[13] Unquantized here means using Matlab double precision of complex numbers that is using 128 bit in total, 64 bit for real and complex values each, according to IEEE Standard 754.

Table 2.7.: Simulation assumptions for hybrid precoding for feedback investigation.

Parameter	Value
SINR bounds, $\gamma^{(\min)}, \gamma^{(\max)}$	$\gamma^{(\min)} = -5\,\text{dB}$, $\gamma^{(\max)} = 40\,\text{dB}$
Number of available users	100
Number of selected users	≤ 40
$N_{(\alpha)}, N_{(\beta)}, N$	$10, 10, 100$
First stage precoder $\mathbf{B}_l$	Sub-codebook splitting with values below
$N_{(\alpha)}^{(\mathbf{\Omega})}, N_{(\beta)}^{(\mathbf{\Omega})}, N^{(\mathbf{\Omega})}$	$16, 16, 256$
$\alpha^{(\text{CW,min})}, \alpha^{(\text{CW,max})}$	$-50°, 50°$
$\beta^{(\text{CW,min})}, \beta^{(\text{CW,max})}$	$-40°, 8°$
$N_{(\alpha)}^{(\mathbf{\Omega},\text{SCB})}, N_{(\beta)}^{(\mathbf{\Omega},\text{SCB})}, N^{(\tilde{\mathbf{\Omega}})}$	$8, 5, 40$
$N^{(\text{SCB})}$	8, according to Algorithm 4
Second stage precoder $\mathbf{P}$	MMSE based on first-stage precoded channels with MISO feedback



(a) Sum spectral efficiency (SE) over the amplitude quantization range $q^{(\text{A,R})}$ with $q^{(\text{A})} = 16\,\text{bit}$ amplitude quantization and $q^{(\text{P})} = 16\,\text{bit}$ phase quantization. The high amplitude and phase quantization is selected in order to isolate the spectral efficiency impact caused by the change of the amplitude range.



(b) Sum spectral efficiency (SE) over amplitude and phase quantization, $q^{(\text{A})}$ and $q^{(\text{P})}$, respectively.

Figure 2.33.: CSI quantization and effective sum spectral efficiency trade-off. Other hybrid precoding simulation parameters are given in Table 2.7.

Fig. 2.33b shows the sum spectral efficiency over amplitude and phase quantization, $q^{(\mathrm{A})}$ and $q^{(\mathrm{P})}$, respectively. It can be observed that the sum spectral efficiency is decreasing with changing slope similar for decreasing amplitude or phase. However, since multiple configurations require the same number of bits, $q^{(\mathrm{CSI})} = q^{(\mathrm{A})} + q^{(\mathrm{P})}$ Table 2.8 lists the configuration for each $q^{(\mathrm{CSI})}$ that achieves the largest sum spectral efficiency. Furthermore, it is observed that in case of even $q^{(\mathrm{CSI})}$ values, the number of bits for amplitude and phase quantization is equal, and for odd $q^{(\mathrm{CSI})}$ values, the difference is always one bit. With $q^{(\mathrm{CSI})} = 20\,\mathrm{bit}$ the spectral efficiency loss compared to the achieved maximum is $< 2\,\%$ and therefore considered as "perfect" quantization. One recommendation derived from these observations is that the same number of bits should be used for phase and amplitude quantization of CSI feedback. From quantization theory perspective, a scalar quantizer as considered in this work, is optimal with respect to the required average bits per quantized value for high resolution quantizers if quantization levels are uniform. High resolution quantizer means that the probability density function of the source is approximately constant on each quantization bin. Here the source that is to be quantized corresponds to the first-stage precoded channel coefficients. Therefore, the uniform quantizer considered in this work can be considered as suitable for amplitude and phase quantization of channel coefficients.

Table 2.8.: Quantization bits $q^{(\mathrm{CSI})}$ as the sum of amplitude and phase quantization bits, $q^{(\mathrm{A})}$ and $q^{(\mathrm{P})}$, respectively. The configuration that achieves the maximum sum spectral efficiency $\tilde{C}^{(\mathrm{sum})}$ from Fig. 2.33b is listed.

$q^{(\mathrm{CSI})}$ in [bit]	$q^{(\mathrm{A})}$ in [bit]	$q^{(\mathrm{P})}$ in [bit]	$\tilde{C}^{(\mathrm{sum})}$ in [bit/s/Hz]	$q^{(\mathrm{CSI})}$ in [bit]	$q^{(\mathrm{A})}$ in [bit]	$q^{(\mathrm{P})}$ in [bit]	$\tilde{C}^{(\mathrm{sum})}$ in [bit/s/Hz]	$q^{(\mathrm{CSI})}$ in [bit]	$q^{(\mathrm{A})}$ in [bit]	$q^{(\mathrm{P})}$ in [bit]	$\tilde{C}^{(\mathrm{sum})}$ in [bit/s/Hz]
4	2	2	34.4	14	7	7	238.8	24	12	12	318.6
5	3	2	40.5	15	7	8	255.9	25	13	12	318.6
6	3	3	51.9	16	8	8	277.8	26	13	13	318.9
7	3	4	64.8	17	8	9	289.6	27	14	13	319.0
8	4	4	85.8	18	9	9	302.2	28	14	14	319.0
9	4	5	106.8	19	9	10	308.7	29	15	14	319.0
10	5	5	136.0	20	10	10	313.5	30	16	14	319.0
11	5	6	159.1	21	10	11	315.0	31	16	15	319.0
12	6	6	189.8	22	11	11	317.1	32	16	16	319.0
13	6	7	210.5	23	12	11	317.5				

With a given quantization, the actual required feedback rate can be calculated. The feedback rate per user for hybrid precoding is denoted as

$$R^{(\mathrm{FB,HP,1})} = \frac{1}{T^{(\mathrm{FI})}} q^{(\mathrm{CSI})} N^{(\tilde{\mathbf{\Omega}})} N^{(\mathrm{RB})}, \tag{2.76}$$

where $q^{(\mathrm{CSI})} = q^{(\mathrm{A})} + q^{(\mathrm{P})}$ denotes the bits required for quantization for the complex effective channel values. With $q^{(\mathrm{CSI})} = 20\,\mathrm{bit}$ this results in 40 kbit per feedback interval. Assuming a $T^{(\mathrm{FI})} = 10\,\mathrm{ms}$ feedback interval this results in 4 Mbit feedback rate per user. As a reference for realistic amount of feedback the recently standardized NR phase I Type II CSI feedback [3GP17g] is considered based on [AVW18]. Therein, a short-term feedback overhead of 264 bit is determined. Therefore, the 40 kbit per user feedback that are required to achieve transmission close to the maximum sum spectral efficiency is ≈ 150 times more than the 264 bit reference and is considered as unrealistic in current cellular systems.

According to Eq. (2.76) one option of feedback reduction is to reduce the size of the effective MISO channel that is fed back in the uplink channel. There are two options:

1. Reduction of the sub-codebook size $N^{(\tilde{\mathbf{\Omega}})}$.

2. Reporting of only a subset of codewords per sub-codebook.

Note that option 2, the reporting of only a subset of codewords per sub-codebook, requires additional feedback in the form of codeword indices so that the BS knows which codewords have been reported. Therefore, the region where option two reduces feedback is obtained next. The user feedback for option two is given by

$$R^{(\mathrm{FB,HP,2})} = \frac{1}{T^{(\mathrm{FI})}} \left(q^{(\mathrm{CSI})} + q^{(\mathrm{CW})} \right) \tilde{N}^{(\tilde{\mathbf{\Omega}})} N^{(\mathrm{RB})}, \tag{2.77}$$

where $\tilde{N}^{(\tilde{\mathbf{\Omega}})} \leq N^{(\tilde{\mathbf{\Omega}})}$ is the number of reported codeword entries and $q^{(\mathrm{CW})}$ is the number of bits required for codeword (stream) indices given by

$$q^{(\mathrm{CW})} = \left\lceil \log_2 N^{(\tilde{\mathbf{\Omega}})} \right\rceil. \tag{2.78}$$

Feedback is only reduced if $R^{(\mathrm{FB,HP,2})} < R^{(\mathrm{FB,HP,1})}$. By substitution of $q^{(\mathrm{CW})}$ with Eq. (2.78) in Eq. (2.77) this yields

$$\begin{aligned} \frac{\left(q^{(\mathrm{CSI})} + \left\lceil \log_2 N^{(\tilde{\mathbf{\Omega}})} \right\rceil \right) \tilde{N}^{(\tilde{\mathbf{\Omega}})} N^{(\mathrm{RB})}}{T^{(\mathrm{FI})}} &< \frac{q^{(\mathrm{CSI})} N^{(\tilde{\mathbf{\Omega}})} N^{(\mathrm{RB})}}{T^{(\mathrm{FI})}} \\ \tilde{N}^{(\tilde{\mathbf{\Omega}})} &\leq \left\lfloor \frac{N^{(\tilde{\mathbf{\Omega}})} q^{(\mathrm{CSI})}}{q^{(\mathrm{CSI})} + \left\lceil \log_2 N^{(\tilde{\mathbf{\Omega}})} \right\rceil} \right\rfloor, \end{aligned} \tag{2.79}$$

as a condition for feedback reduction, where the equal condition is ensured by the floor function to the next integer. Table B.5 lists $\tilde{N}^{(\tilde{\mathbf{\Omega}})}$ for equality according to Eq. (2.79) over $q^{(\mathrm{CSI})}$ and $N^{(\tilde{\mathbf{\Omega}})}$. Under the assumption that $N_2^{(\tilde{\mathbf{\Omega}})} < N_1^{(\tilde{\mathbf{\Omega}})}$ and $\tilde{N}^{(\tilde{\mathbf{\Omega}})}$ fulfills condition Eq. (2.79) the feedback reduction for option one and two are given as

$$R^{(\mathrm{FB,HP,1})}(N_1^{(\tilde{\mathbf{\Omega}})}) - R^{(\mathrm{FB,HP,1})}(N_2^{(\tilde{\mathbf{\Omega}})}) = \left(N_1^{(\tilde{\mathbf{\Omega}})} - N_2^{(\tilde{\mathbf{\Omega}})} \right) q^{(\mathrm{CSI})} \frac{N^{(\mathrm{RB})}}{T^{(\mathrm{FI})}}, \tag{2.80}$$

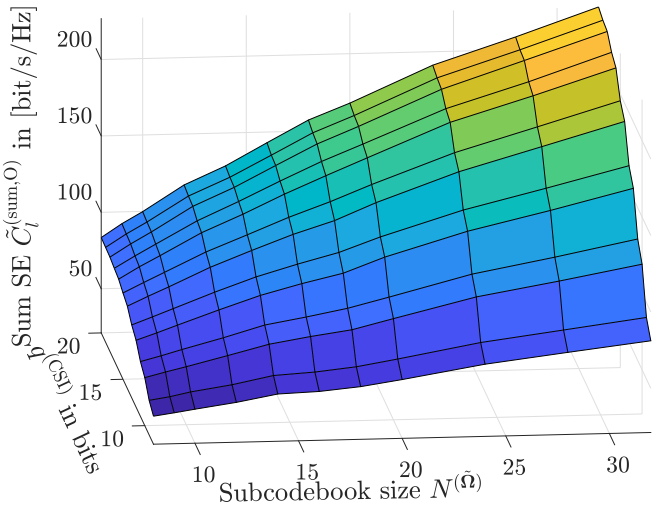
$$R^{(\mathrm{FB,HP,1})}(N_1^{(\tilde{\mathbf{\Omega}})}) - R^{(\mathrm{FB,HP,2})}(\tilde{N}^{(\tilde{\mathbf{\Omega}})}) = \left(\left(N_1^{(\tilde{\mathbf{\Omega}})} - \tilde{N}^{(\tilde{\mathbf{\Omega}})} \right) q^{(\mathrm{CSI})} + \left\lceil \log_2 N_1^{(\tilde{\mathbf{\Omega}})} \right\rceil \tilde{N}^{(\tilde{\mathbf{\Omega}})} \right) \frac{N^{(\mathrm{RB})}}{T^{(\mathrm{FI})}}, \tag{2.81}$$

respectively. From a feedback perspective the feedback reduction is larger with option one. However, the trade-off with the sum spectral efficiency cannot be captured analytically and is obtained numerically.

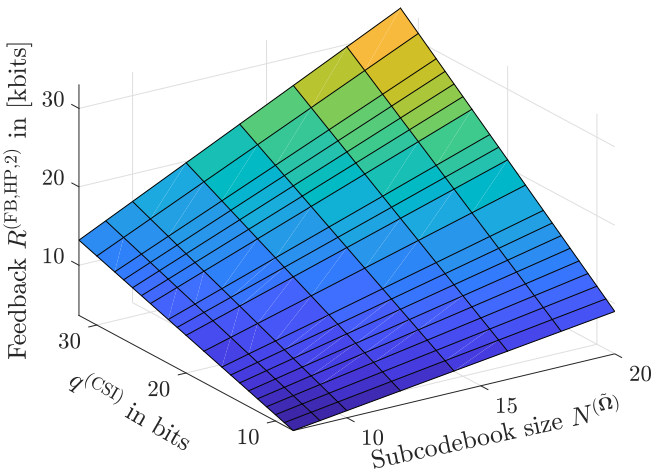
Fig. 2.34 shows the effective sum spectral efficiency versus feedback trade-off over $q^{(\mathrm{CSI})}$ and $N^{(\tilde{\mathbf{\Omega}})}$. The best configuration in terms of effective sum spectral efficiency is selected for each value of $q^{(\mathrm{CSI})}$ and $N^{(\tilde{\mathbf{\Omega}})}$ according to Table 2.8 and Table B.5, respectively. The effective sum spectral efficiency is calculated according to Eq. (2.59) and the feedback according to Eq. (2.79). The corresponding numerical values are listed in Table B.6 and Table B.7. It can be observed that the sum spectral efficiency is not decreasing monotonically over the sub-codebook size, because some configurations correspond to suboptimal beam distributions, e.g. configuration $N^{(\tilde{\mathbf{\Omega}})} = 26$ corresponds to $N_{(\alpha)}^{(\mathbf{\Omega},\mathrm{SCB})} = 13$ and $N_{(\beta)}^{(\mathbf{\Omega},\mathrm{SCB})} = 2$, thus the vertical dimension is quantized only coarse.

Fig. 2.35 shows reduction of reported streams by $\tilde{N}^{(\tilde{\mathbf{\Omega}})}$ for sub-codebook size $N^{(\tilde{\mathbf{\Omega}})} = 32$. This value is selected because the loss in effective sum spectral efficiency is less than 7 % by a feedback reduction of 20 % compared to $N^{(\tilde{\mathbf{\Omega}})} = 40$, see Table B.7. This also reduces the required bits for the stream index $q^{(\mathrm{CW})}$ from six to five bits.

Table 2.9 summarizes and compares the three investigated feedback reduction techniques “codeword subset reporting”, “sub-codebook size”, and “RB subset reporting” for certain $R^{(\mathrm{FB,Thr})}$ values showing only the

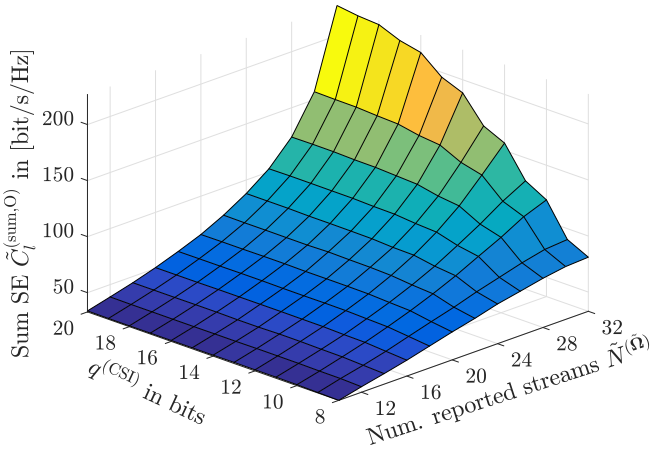


(a) Effective sum spectral efficiency (SE) $\tilde{C}_l^{(\mathrm{sum,O})}$ according to Eq. (2.59).

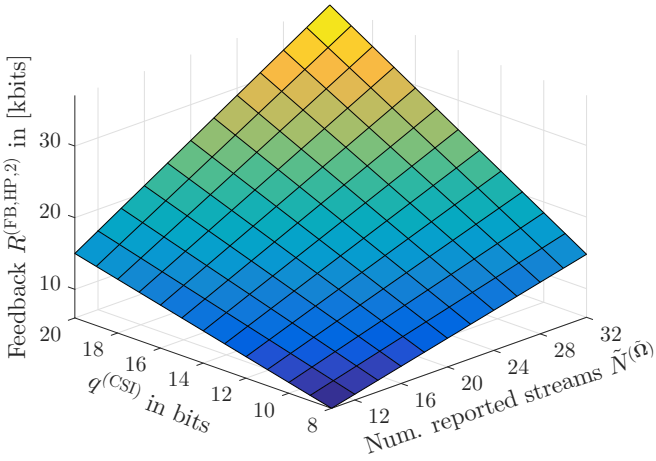


(b) Feedback $R^{(\mathrm{FB,HP},2)}$ according to Eq. (2.79).

Figure 2.34.: Effective sum spectral efficiency (SE) and feedback trade-off over quantization $q^{(\mathrm{CSI})}$ and subcodebook size $N^{(\tilde{\Omega})}$.



(a) Effective sum spectral efficiency (SE) $\tilde{C}_l^{(\mathrm{sum,O})}$ according to Eq. (2.59).

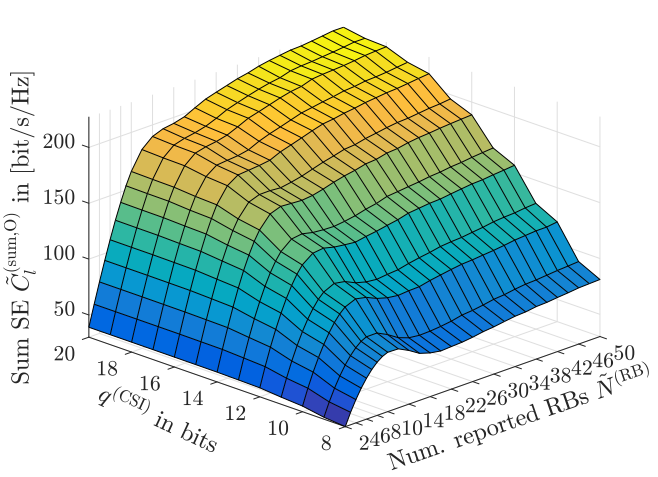


(b) Feedback $R^{(\mathrm{FB,HP},2)}$ according to Eq. (2.79).

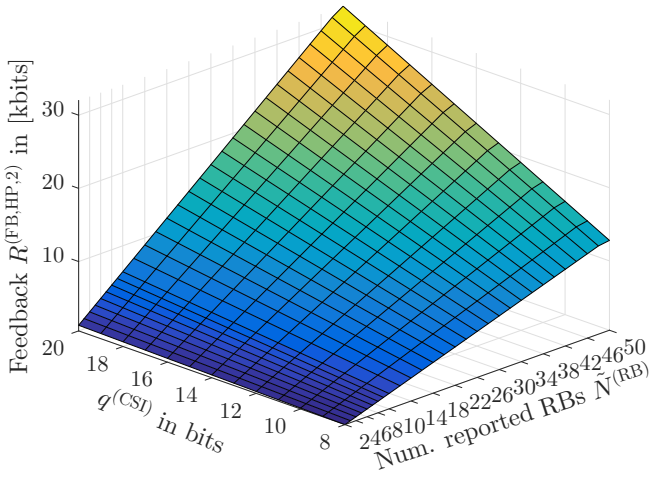
Figure 2.35.: Effective sum spectral efficiency (SE) and feedback trade-off over quantization $q^{(\mathrm{CSI})}$ and subcodebook subset reporting $\tilde{N}^{(\tilde{\Omega})}$ for $N^{(\tilde{\Omega})} = 32$.

Table 2.9.: Comparison of feedback schemes. Given a feedback constraint $R^{(\mathrm{FB,Thr})}$ the configuration that achieves the maximum effective sum spectral efficiency is selected based on Fig. 2.34 and Fig. 2.35.

$R^{(\mathrm{FB,Thr})}$ in [kbit]	Codeword subset reporting			Sub-codebook size			RB subset reporting		
	$\tilde{C}_l^{(\mathrm{sum,O})}$ in [bit/s/Hz]	$q^{(\mathrm{CSI})}$ in [bit]	$\tilde{N}^{(\tilde{\Omega})}$	$\tilde{C}_l^{(\mathrm{sum,O})}$ in [bit/s/Hz]	$q^{(\mathrm{CSI})}$ in [bit]	$N^{(\tilde{\Omega})}$	$\tilde{C}_l^{(\mathrm{sum,O})}$ in [bit/s/Hz]	$q^{(\mathrm{CSI})}$ in [bit]	$\tilde{N}^{(\mathrm{RB})}$
5	-	-	-	57.0	11	8	156.0	15	10
10	57.8	8	20	103.1	15	12	191.3	19	16
15	84.0	9	28	138.9	14	20	203.5	19	24
20	119.1	10	32	178.1	16	24	214.4	20	30



(a) Effective sum spectral efficiency (SE) $\tilde{C}_l^{(\mathrm{sum,O})}$ according to Eq. (2.59).



(b) Feedback $R^{(\mathrm{FB,HP,RB})}$ according to Eq. (2.82).

Figure 2.36.: Effective sum spectral efficiency (SE) versus feedback trade-off over quantization $q^{(\mathrm{CSI})}$ and RB subset reporting $\tilde{N}^{(\mathrm{RB})}$ for $N^{(\tilde{\mathbf{\Omega}})} = 32$.

configurations that achieve the largest $\tilde{C}_l^{(\mathrm{sum,O})}$. From Table 2.9 it can be concluded that in terms of effective sum spectral efficiency it is better to use smaller sub-codebook for feedback reduction instead of reporting a subset of streams from a larger sub-codebook. However, even the lowest feedback rate of 3.7 kbit in Table B.7 is still ≈ 14 times the reference feedback of 264 bit. This is clearly to much feedback for a practical 5G system in the near future and further reduction is required. Therefore, the remaining DoF for feedback reduction, the frequency domain, is investigated next.

Similar to the reporting of the best codewords each user reports a subset of RBs. The number of reported RBs is denoted by $\tilde{N}^{(\mathrm{RB})} \leq N^{(\mathrm{RB})}$. For the selection of the "best" RBs each user k computes on each RB the receive signal assuming a MRT receive beamformer according to (2.3). Based on the descended ordered list of receive signal values per subband the first $\tilde{N}^{(\mathrm{RB})}$ values that correspond to the largest values are reported. Thus, $q^{(\mathrm{RB})}$ additional bits are required for RB indices in the frequency domain, e.g. in the assumed 9 MHz band with 50 RBs this results in $q^{(\mathrm{RB})} = \left\lceil \log_2 N^{(\mathrm{RB})} \right\rceil$ bit $= \lceil \log_2 50 \rceil$ bit $= 6$ bit. Accordingly the feedback bits for RB subset reporting is given by

$$R^{(\mathrm{FB,HP,RB})} = \frac{1}{T^{(\mathrm{FI})}} \left(q^{(\mathrm{CSI})} N^{(\tilde{\mathbf{\Omega}})} + q^{(\mathrm{RB})} \right) \tilde{N}^{(\mathrm{RB})}. \tag{2.82}$$

Similar as in Eq. (2.79) the additional feedback of the index reduces feedback only below a certain threshold that is given as

$$\tilde{N}^{(\mathrm{RB})} \leq \left\lfloor \frac{N^{(\mathrm{RB})} q^{(\mathrm{CSI})} N^{(\tilde{\mathbf{\Omega}})}}{N^{(\tilde{\mathbf{\Omega}})} q^{(\mathrm{CSI})} + \lceil \log_2 N^{(\mathrm{RB})} \rceil} \right\rfloor. \tag{2.83}$$

With $N^{(\tilde{\mathbf{\Omega}})}$ and $N^{(\mathrm{RB})}$ the maximum of transmit streams over all RBs is $N^{(\tilde{\mathbf{\Omega}})} N^{(\mathrm{RB})} = 1600$ for the given assumptions. In order to achieve theoretically the full multiplexing gain, at least $\tilde{K}$ users have to report $\tilde{N}^{(\mathrm{RB})} \geq \frac{N^{(\tilde{\mathbf{\Omega}})} N^{(\mathrm{RB})}}{\tilde{K}}$ RBs. However, the probability that users report the same RB index is high and results in a reduced number of scheduled streams. Additionally, the probability is low that $N^{(\tilde{\mathbf{\Omega}})}$ suitable users can be found by the scheduling heuristic at each RB. This available stream reduction can be seen in Fig. 2.36a where the slope of the effective sum spectral efficiency decreases significantly at $\tilde{N}^{(\mathrm{RB})} = 12$ for $\tilde{K} = 100$. The corresponding values are listed in Table B.11, while the required feedback $R^{(\mathrm{FB,HP,RB})}$ is shown in Fig. 2.36b and Table B.12.

Table 2.10.: Achieved performance for selected configurations with feedback close to reference feedback of 264 bit. The blue colored fields indicate the largest achieved median sum and 5 %-ile user spectral efficiency.

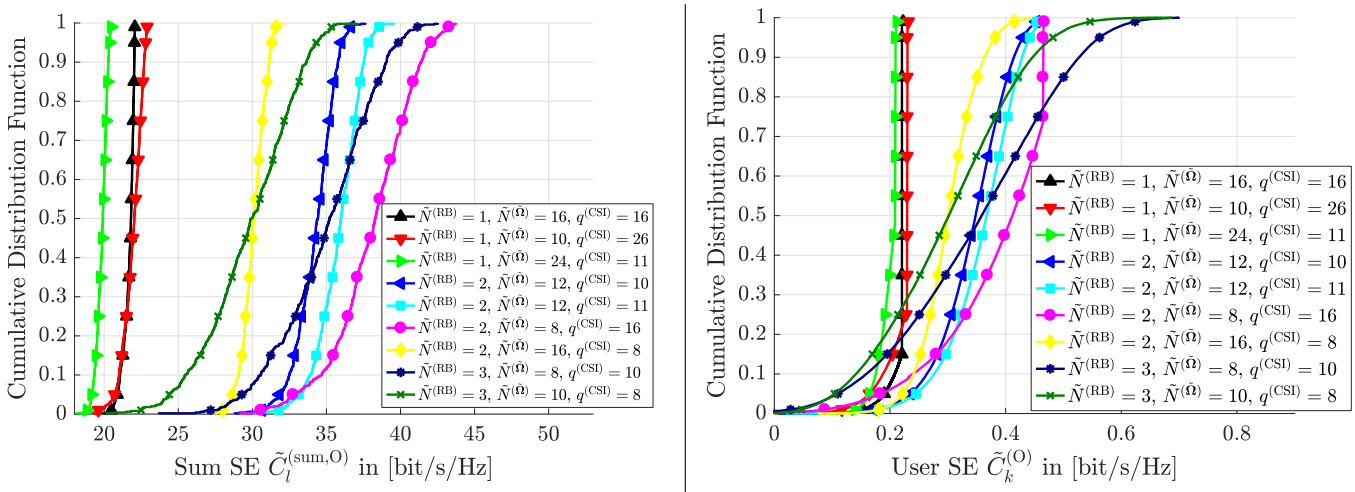
$R^{(\mathrm{FB,HP,RB})}$	$q^{(\mathrm{CSI})}$ in [bits]	$N^{(\tilde{\mathbf{\Omega}})}$	$\tilde{N}^{(\mathrm{RB})}$	Median sum spectral efficiency $\tilde{C}_l^{(\mathrm{sum,O})}$ in [bit/s/Hz]	5 %-ile user spectral efficiency $\tilde{C}_k^{(\mathrm{O})}$ in [bit/s/Hz]
262	16	16	1	22	0.19
266	26	10	1	22	0.16
270	11	24	1	20	0.16
252	10	12	2	34	0.24
276	11	12	2	36	0.25
268	16	8	2	38	0.18
268	8	16	2	30	0.22
258	10	8	3	35	0.11
258	8	10	3	30	0.10

Table 2.9 compares the three feedback reduction schemes, it can be observed that for a "low" feedback threshold a combination of less quantization bits and smaller sub-codebook size achieves the largest effective sum spectral efficiency. In contrast to this, for lower feedback thresholds reporting a subset of RBs is the best option. However, in order to achieve a feedback rate that is close to the reference of 264 bit a combination of RB subset and lower sub-codebook size is required. Due to the parameter space consisting of $q^{(\mathrm{CSI})}$ the quantization bits, $N^{(\tilde{\mathbf{\Omega}})}$ the sub-codebook size, and $\tilde{N}^{(\mathrm{RB})}$ the number of reported RBs the parameter space is too large for numerical optimization.

Some selected configurations that achieve "similar" feedback as the reference feedback of 264 bit are computed. Table 2.10 lists the selected configuration together with their required feedback, according to Eq. (2.82), the achieved median effective sum spectral efficiency, and the 5 %-ile user effective spectral efficiency. The corresponding result are shown in Fig. 2.37. The large variance in the user spectral efficiency in Fig. 2.37b is a typical characteristic of the maximum throughput heuristic from Section 2.3.4. A larger $\tilde{N}^{(\mathrm{RB})}$ and therefore a greater number of DoFs at the scheduler results in a larger user spectral efficiency variance and a lower 5 %-ile user spectral efficiency. This inherent fairness is provided by the sub-codebook construction and distribution in frequency domain.

The largest achieved sum spectral efficiency of 38 bit/s/Hz is provided by configuration $q^{(\mathrm{CSI})} = 16$ bit, $N^{(\tilde{\mathbf{\Omega}})} = 8$, and $\tilde{N}^{(\mathrm{RB})} = 2$. This is a "large" performance loss compared to sum spectral efficiency in Fig. 2.32, but in comparison with codebook based precoding shown in Fig. 2.30 in Section 2.4.3 hybrid precoding achieves a larger sum spectral efficiency. Even with the full knowledge of 256 streams and without considering pilot overhead, codebook based precoding achieves only 27.5 bit/s/Hz. On top of this, the codebook based precoding requires also a lot of feedback, e.g. the sub-codebook splitting with 16 beams, 9 bit amplitude quantization on two RBs results in 288 bit. Fig. 2.29 shows that reporting only a subset of the codewords results in large performance loss. This means that for similar feedback amount the gain of the hybrid precoding would even be larger. Finally, a discussion on the pros and cons of TDD versus FDD for massive MIMO is provided.

First of all, a fair comparison between FDD and TDD is difficult to obtain due to a lot of different system assumptions. For example, should the bandwidth in TDD be equal to the sum of uplink and downlink in FDD? How much of the uplink capacity in FDD can be used for feedback? Therefore, in order to provide a final assessment between FDD with the proposed solution and TDD the following assumptions are made:



(a) Effective sum spectral efficiency (SE).

(b) Effective user spectral efficiency (SE).

Figure 2.37.: Performance for configurations according to Table 2.10 requiring (262, 266, 270, 252, 276, 268, 268, 258, 258) bits. The order from left to right corresponds to the order in the legend from top to bottom.

- The coherence time is set to three LTE physical RBs, because in FDD one slot is required to transmit pilots, the next slot is required to send back CSI, and the third slot is required for the data transmission. This can be seen as the minimum required coherence time because no signal processing delay is considered. With 168 resource elements per physical RB this sums up to 504 resource elements or coherence resources. The effect of a larger coherence block in time or frequency is discussed afterwards.
- For FDD the pilot overhead, required for the sub-codebook precoded channel, is taken into account with the assumption that each pilot requires one coherence resource. The same is considered in Fig. 2.15.
- For TDD, the pilot overhead corresponds to the uplink resources required for channel estimation, again one resource element per coherence block per user is considered, the same assumption as for FDD.
- In FDD it is assumed that all users can transmit their feedback error free in the uplink channel to the BS. Similar, in TDD no channel estimation error is considered.
- Spectral efficiency is evaluated instead of the throughput to avoid the impact from different bandwidth in uplink and downlink in both duplex modes, see Fig. 1.1.

The following observations summarize result from previous sections as well as Fig. 2.38. With pilot overhead and infinite uplink capacity for feedback, the FDD system achieves 247 bit/s/Hz in Fig. 2.32b. As an equivalent for the infinite uplink capacity assumption for FDD, in TDD perfect channel knowledge and no uplink traffic, besides the resource elements occupied by uplink pilots, is considered. With this and other parameters being the same as for FDD, the TDD system achieves 412 bit/s/Hz, see Fig. 2.38.

The next point that is discussed is how both duplex modes would benefit from a larger coherence block size. Note, that a larger coherence block size can be achieved due to a larger coherence time or a larger coherence frequency. With a larger coherence block in FDD, the codebook size could be increased yielding a larger multiplexing gain, while keeping the same pilot overhead ratio. Due to the sub-codebook size of 16, the sum spectral efficiency is already saturating with 20 available mobile users in the system, see Fig. 2.31. Therefore, increasing the codebook size together with the coherence block size significant sum-spectral efficiency gains can be expected up to 60 users with the assumed 100 antennas similar as in the TDD system. In contrast to FDD, the TDD system shows a saturation in sum spectral efficiency due to the limited channel rank. The 100 antennas assumed in this thesis serve ≈ 60 users with the PBZF scheduling algorithm and an increase of available users hardly increases the sum spectral efficiency, see Fig. 2.10. Therefore, a larger coherence time

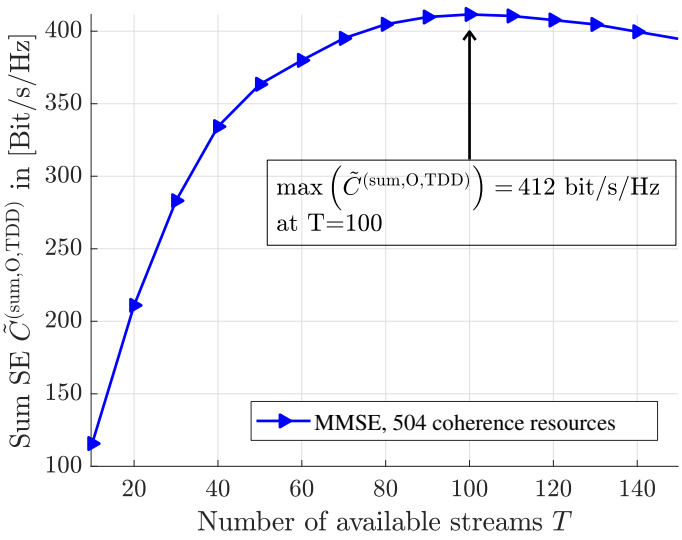


Figure 2.38.: Sum spectral efficiency of TDD with MMSE precoder and PBZF scheduling. Overhead for CSI acquisition due to channel estimation in uplink considers one resource element out of the coherence block of three physical RBs that is $3 \cdot 12 \cdot 14 = 504$ resource elements in total.

and reporting of additional users increases the uplink channel estimation overhead but hardly the spectral efficiency. For the TDD an increase of the coherence block has to be accompanied with an increase in the number of antennas in order to increase sum spectral efficiency. Hence, the FDD system would benefit most in terms of sum spectral efficiency from a larger coherence block size compared to TDD.

Finally, both the impact of limited uplink capacity in FDD and equivalently the impact of uplink traffic in TDD is discussed in the following paragraph. With limited uplink capacity, the performance of the FDD system strongly depends on the available uplink capacity that can be used for feedback as discussed in this chapter. E.g. if only few users are in the system, each could send much more feedback than the 264 bit. On the other hand, if many users are in the system each of them may send only the 264 bit feedback but user diversity gain can compensate a part of the spectral efficiency degradation due to the limited CSI feedback. With the assumptions in this thesis a sum spectral efficiency of approximately 200 bit/s/Hz is achieved with an uplink feedback of 40 kbit per feedback interval per user, see Fig. 2.35. On the other hand, the limited uplink feedback of 264 bit per feedback interval per user, degrades the sum spectral efficiency down to 40 bit/s/Hz, see Fig. 2.37. In TDD, the equivalent of the limited uplink capacity of FDD are time slots reserved for uplink data transmission. Thus the effective downlink spectral efficiency reduces linear with the number of time slots reserved for uplink traffic. Note that time duplex means that all resource elements in frequency of the same symbol are blocked for downlink if this slot is assigned to uplink data transmission. E.g. with $\frac{1}{3}$ of the time slots reserved for uplink data transmission, the 412 bit/s/Hz in Fig. 2.38 reduce to 275 bit/s/Hz. Still, the TDD achieves a larger sum spectral efficiency than the FDD.

2.4.6. Conclusions

This section shows that massive MIMO can be efficiently integrated into FDD systems by using two-stage hybrid precoding considering pilot and feedback constraints. One outcome of this section is that purely codebook precoding cannot utilize spatial multiplexing gains of massive MIMO arrays. In systems with a "realistic" number of antennas interference is still limiting the sum spectral efficiency and phase information of the channels for interference cancellation is required. The scalability of pilot and feedback overhead according to user demands or system KPIs is an important requirement of 5G systems due to the range of use cases. A combination of codebook based first stage precoding and effective CSI aware second stage precoding is applied in this section. In order to provide a sufficient large channel quantization by the codebook based first-stage precoding while keeping at the same time the pilot overhead limited, the sub-codebook

splitting scheme is proposed in this thesis. It is shown that with sub-codebook splitting, beamforming gains and spatial multiplexing gains of massive MIMO can be utilized to increase the sum spectral efficiency compared to NR Release 15 systems. This result is achieved assuming the same amount of feedback as in NR Release 15 by a joint combination of feedback reduction in frequency, space and quantization. The proposed sub-codebook splitting scheme is scalable with respect to the number of BS antennas and number of streams/pilots given by the first stage precoder. Furthermore, sub-codebook splitting provides inherent user fairness due to the distribution of the sub-codebooks in frequency domain where each sub-codebook covers different directions. In combination with feedback reduction in frequency domain, where each user reports only his best subset of RBs, sub-codebook splitting is an efficient solution that provides cell-coverage, sum spectral efficiency gains, and fairness for multiple users at the same time.

Furthermore, the discussion on FDD versus TDD has revealed, that for massive MIMO TDD mode should be selected for short coherence blocks, e.g. in high mobility scenarios, whereas in static scenarios, e.g. for fixed wireless access, FDD is an option that can achieve similar spectral efficiency as TDD. However, the more antennas are used the more beneficial TDD becomes in terms of sum spectral efficiency under the condition that sufficient active users are in the system. On the other hand, there is a lot of spectrum dedicated to FDD and operators don't have a choice on the duplex mode. In such cases the two-stage precoding and feedback compression schemes presented in this thesis can be applied to improve the sum-spectral efficiency with massive MIMO compared to non massive MIMO.

3. Uplink Search-Based Direction-of-Arrival (DoA) Estimation

In this chapter, direction of arrival (DoA) estimation in the uplink at the base station (BS) is studied. Research on DoA estimation with multiple antennas is an ongoing topic of huge interest since the 1920s (see introduction in [TF09]). Due to the focus of this thesis on search-based techniques, grouping for multiple-source estimation and challenges in cellular deployment, the interested reader is referred to the book "Classical and Modern Direction-of-Arrival Estimation" from Tuncer and Friedlander [TF09] as a comprehensive overview on DoA estimation in general. One approach to classify DoA estimation methods is to divid these into search-based and search-free algorithms. Prominent search-free DoA estimation techniques are Root-multiple signal classification (MUSIC) [Bar83][1] and ESPRIT, which stands for "Estimation of Signal Parameter via Rotational Invariance Technique" [PRK85]. Search-free methods are less complex than search-based methods, however they usually are restricted to certain array geometries and are less robust against hardware impairments, such as disposition of antenna elements, calibration errors etc.. Therefore, the focus in this work is on search-based techniques, which can be applied to arbitrary antenna structures. The most prominent approaches of search-based DoA estimation are given in Section 3.1. However, due to the huge research interest in improving localization systems based on DoA estimation there are a lot of advances and not all of them can be captured in this thesis, e.g. compressed sensing-based DoA estimation [MCW05]. Nevertheless, the Cramer Rao lower bound for arbitrary arrays from [BM03] is included for comparison.

In the remainder of this chapter, the uplink signal model is introduced in Section 3.1. Afterward, in Section 3.2.2 a parameter analysis is done and a novel complexity reduction scheme is derived and evaluated. Further on, in Section 3.3 multiple-source estimation is taken into account and the focus is on grouping to reduce network resources for positioning, a typical task in wireless communications systems where network resources are valuable and limited. Finally, Section 3.3.3 summarizes and concludes this chapter.

3.1. Multiple-Source System Model

In contrast to previous Chapter 2, where downlink direction is considered in the system model description in Section 2.2, this chapter is on DoA estimation at the massive multiple-input multiple-output (MIMO) array and therefore on uplink direction. This requires a new system model given in the following. Nevertheless, wherever suitable the same notations as before are used but for the sake of readability in this chapter, some of them are introduced again in this section.

3.1.1. Receive Signal

In the ideal case of synchronized reception the samples of the uplink baseband receive signal at BS l can be written as

$$\mathbf{y}_l(t) = \left(\underline{\mathbf{H}}_{\mathcal{K},l}^{(\mathrm{T})}(t)\right)^{\mathrm{H}} \mathbf{x}_{\mathcal{K}}(t) + \mathbf{n}_l^{(\mathrm{BS})}(t), \tag{3.1}$$

[1] Note while the original MUSIC algorithm is a search-based method [Sch86], the root MUSIC in [Bar83] is a search-free method and works only for uniform linear arrays.

where $\mathbf{y}_l(t) \in \mathbb{C}^N$ denotes the receive signal at the N antenna elements of BS l, $\underline{\mathbf{H}}_{\mathcal{K}}^{(\mathrm{T})} \in \mathbb{C}^{K \times N}$ denotes the time domain channel matrix[2] from the K single antenna devices (sources) in the set $\mathcal{K} = \{1, \ldots, K\}$ such that

$$\underline{\mathbf{H}}_{\mathcal{K},l}^{(\mathrm{T})}(t) = \left[\left(\mathbf{h}_{1,l}^{(\mathrm{T})}(t) \right)^{\mathrm{T}} \ \ldots \ \left(\mathbf{h}_{K_l,l}^{(\mathrm{T})}(t) \right)^{\mathrm{T}} \right]^{\mathrm{T}}, \tag{3.2}$$

where $\mathbf{h}_{k,l}^{(\mathrm{T})}(t) \in \mathbb{C}^N$ is the channel of source k to BS l, $\mathbf{x}_{\mathcal{K}}(t) \in \mathbb{C}^K$ denotes the transmitted signals of the K_l sources, $\mathbf{n}_l^{(\mathrm{BS})}(t) \sim \mathcal{CN}\left(0, \left(\sigma_l^{(\mathrm{BS})}\right)^2 \mathbf{I}_N\right)$ constitutes the Gaussian distributed uncorrelated noise at BS l with the noise variance $\left(\sigma_l^{(\mathrm{BS})}\right)^2$, and $t \in \mathbb{N}_+$ represents the discrete time index of the samples. Note that for notational simplicity and better readability the time dependency (t) is omitted whenever operations (e.g. matrix multiplications) are based on the same time index, as well as the BS index l if only a single BS is considered. Furthermore, the signal power from user k and noise power of $\mathbf{y}$ in Eq. (3.1) are given by

$$P_{k,l}^{(\mathrm{s})} = \mathrm{trace}\left(\left(\mathbf{h}_{k,l}^{(\mathrm{T})} \right)^{\mathrm{H}} x_k x_k^* \mathbf{h}_{k,l}^{(\mathrm{T})} \right), \tag{3.3}$$

and

$$P_l^{(\mathrm{n})} = \mathrm{trace}\left(\mathbf{n}_l^{(\mathrm{BS})} \left(\mathbf{n}_l^{(\mathrm{BS})} \right)^{\mathrm{H}} \right), \tag{3.4}$$

respectively. Consequently the signal to noise ratio (SNR) of the receive signal from user k in Eq. (3.1) is

$$\tilde{\gamma}_{k,l} = \frac{P_{k,l}^{(\mathrm{s})}}{P_l^{(\mathrm{n})}}. \tag{3.5}$$

In case devices are equipped with multiple antennas but transmit a single beamformed layer, the same representation as in Eq. (3.1) can be used by replacing elements in $\underline{\mathbf{H}}_{\mathcal{K}}^{(\mathrm{T})}$ with an effective channel. Multipath propagation is an essential part in wireless communications systems and explicitly modeled in cellular communications channels, see Section 2.1. Therefore, the channel impulse response of a single device k is represented by matrix $\bar{\mathbf{H}}_k^{(\mathrm{T})} \in \mathbb{C}^{L_{(\mathrm{MPC})} \times N}$ and structured as

$$\bar{\mathbf{H}}_k^{(\mathrm{T})} = \boldsymbol{\Delta}_k \hat{\mathbf{H}}_k^{(\mathrm{T})}, \tag{3.6}$$

where $\hat{\mathbf{H}}_k^{(\mathrm{T})} \in \mathbb{C}^{L_{(\mathrm{MPC})} \times N}$ are the channel coefficients for the $L_{(\mathrm{MPC})}$ multi-path components and $\boldsymbol{\Delta}_k \in \mathbb{R}^{L_{(\mathrm{MPC})} \times L_{(\mathrm{MPC})}}$ is a diagonal matrix with elements $[\boldsymbol{\Delta}]_{l_{(\mathrm{MPC})}, l_{(\mathrm{MPC})}} = \delta_{l_{(\mathrm{MPC})}}(t - \tau_{l_{(\mathrm{MPC})}})$. Therein $\delta_{l_{(\mathrm{MPC})}}(t - \tau_{l_{(\mathrm{MPC})}})$ denotes the Dirac impulse of path delay $\tau_{l_{(\mathrm{MPC})}}$, and subscript $l_{(\mathrm{MPC})}$ indicates the l-th multi-path component.

Remark. *The BS index l is omitted in Eq.* (3.6) *and following equations for better readability because a single BS is considered in this chapter. Furthermore, the subscript* $_{(\mathrm{MPC})}$ *is used to clearly distinguish the multi-path index $l_{(\mathrm{MPC})}$ from the BS index l.*

In contrast to Eq. (3.1) the first dimension of $\hat{\mathbf{H}}^{(\mathrm{T})}$ is of size $L_{(\mathrm{MPC})}$ instead of K. However, based on the assumption that the delay spread of the $L_{(\mathrm{MPC})}$ multi-path components is smaller than a symbol length, e.g. in Long Term Evolution (LTE) they are within the cyclic prefix, individual multi-path components can only be resolved with sufficient bandwidth. Therefore, the channel vector of device k is assumed as

$$\mathbf{h}_k^{(\mathrm{T})} = \sum_{l_{(\mathrm{MPC})}=1}^{L_{(\mathrm{MPC})}} \left[\hat{\mathbf{H}}_k^{(\mathrm{T})} \right]_{l_{(\mathrm{MPC})},:}, \tag{3.7}$$

[2] $\mathbf{H}$ indicates the frequency domain downlink channel matrix for a single frequency chunk, while $\mathbf{H}^{(\mathrm{T})}$ indicates the time domain downlink channel matrix.

where notation $[\mathbf{A}]_{i,:}$ denotes all column elements of row i, such that $\mathbf{h}_k^{(\mathrm{T})} \in \mathbb{C}^{1\times N}$ is the complex sum or superposition of all multi-path components per antenna. The following numerical example justified the above assumption. With a typical LTE resource block (RB) bandwidth of 180 kHz a time resolution of $\approx 5.6\,\mu\mathrm{s}$ can be achieved which is a multiple of the channel impulse response time according to measurements in [EKPP10] ($< 0.8\,\mu\mathrm{s}$) and [WWH$^+$11] ($< 0.6\,\mu\mathrm{s}$).

One property of the QUAsi Deterministic RadIo channel GenerAtor (QuaDRiGa) model is that the channel coefficients include path-loss and antenna pattern gains such that they are not-normalized to power one. In contrast to this, in most DoA estimation literature normalized channels are assumed, [SS15]. In order to evaluate the effect of not-normalized channels and to be comparable with other literature, the case of normalized channels is introduced and denoted as $\underline{\tilde{\mathbf{H}}}_{\mathcal{K}}^{(\mathrm{T})}$ and derived by

$$\underline{\tilde{\mathbf{H}}}_{\mathcal{K}}^{(\mathrm{T})} = \left[\frac{\left(\mathbf{h}_1^{(\mathrm{T})}\right)^{\mathrm{T}}}{\left\|\mathbf{h}_1^{(\mathrm{T})}\right\|} \ \cdots \ \frac{\left(\mathbf{h}_K^{(\mathrm{T})}\right)^{\mathrm{T}}}{\left\|\mathbf{h}_K^{(\mathrm{T})}\right\|} \right]^{\mathrm{T}}. \tag{3.8}$$

Accordingly a Boolean parameter $g^{(\mathrm{norm})} \in \{0, 1\}$ is introduced which indicates if channel normalization is applied such that

$$\underline{\mathbf{H}}_{\mathcal{K}}^{(\mathrm{T})} = \begin{cases} \underline{\tilde{\mathbf{H}}}_{\mathcal{K}}^{(\mathrm{T})} & \text{if } g^{(\mathrm{norm})} = 1 \\ \underline{\mathbf{H}}_{\mathcal{K}}^{(\mathrm{T})} & \text{if } g^{(\mathrm{norm})} = 0. \end{cases} \tag{3.9}$$

For reasons of consistency with DoA literature the matrix of channel coefficients $\hat{\mathbf{H}}_k^{(\mathrm{T})}$ from Eq. (3.6) is rewritten as

$$\hat{\mathbf{H}}_k^{(\mathrm{T})} = \mathbf{P}_k^{(\mathrm{MPC})} \odot \mathbf{A}_k, \tag{3.10}$$

where $\mathbf{P}_k^{(\mathrm{MPC})} \in \mathbb{R}^{L_{(\mathrm{MPC})}\times N}$ denotes the matrix of path gains between user k and BS l and $\mathbf{A}_k \in \mathbb{C}^{L_{(\mathrm{MPC})}\times N}$ is the respective steering matrix[3]. $\mathbf{A}_k = \left[\mathbf{a}_{k,1}^{\mathrm{T}} \ \cdots \ \mathbf{a}_{k,L_{(\mathrm{MPC})}}^{\mathrm{T}}\right]^T$ contains the steering vectors $\mathbf{a}_{k,l_{(\mathrm{MPC})}}$ of each multi-path $l_{(\mathrm{MPC})}$, where $\left\|\mathbf{a}_{k,l_{(\mathrm{MPC})}}\right\| = 1$. In general the steering vector represents the angular dependence of the array response in different directions, which depends on the geometry and the antenna pattern of each element in the array antenna. In this work all antenna elements are assumed as identical. As already introduced in Section 1.2.1 the phase of the n-th element can be expressed by a phase difference to a reference element. Without limiting the generality the reference element is always the first element $n = 0$. In literature, the steering vector is also referred to as the space vector, array response, or array manifold [TF09]. The steering vector of user k of path $l_{(\mathrm{MPC})}$ is

$$\mathbf{a}_{k,l_{(\mathrm{MPC})}} = \left[1 \ e^{j\mu_{k,1}} \ \cdots \ e^{j\mu_{k,N-1}}\right]^{\mathrm{T}}, \tag{3.11}$$

where $\mu_{k,n}$ is the phase difference of the n-th antenna element with respect to the first element as reference. In Eq. (1.12) the phase difference is determined under the assumption that antenna elements and the incident wave front are in the same plane thus depending only on the azimuth angle. However, the phase difference $\mu_{k,n}$ in Eq. (3.11) the incident wave and the antenna elements can be in different planes thus the phase difference depends on the azimuth and elevation angle.

In three dimensional (3D) scenario the DoA ϕ can be defined by the horizontal angle α and vertical angle β, see Fig. 3.1. Consequently, the DoA ϕ in Cartesian coordinates parametrized by spherical coordinates is given as

$$\mathbf{u}_{(\phi)}(\alpha, \beta) = -\begin{bmatrix} \cos(\beta)\cos(\alpha) \\ \cos(\beta)\sin(\alpha) \\ \sin(\beta) \end{bmatrix}, \tag{3.12}$$

[3] In most of the DoA literature only the steering matrix is considered and the path gains are ignored, e.g. [Cap69, BK83, PRK85, Sch86, SN89, Dho10, BLG$^+$15].

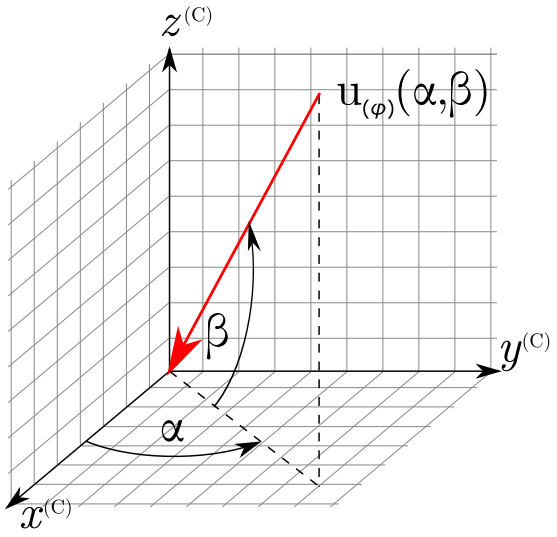


Figure 3.1.: Direction of arrival in spherical coordinates, where $-180° \leq \alpha \leq +180°$ denotes the horizontal/azimuth angle and $-90° \leq \beta \leq +90°$ denotes the vertical/elevation angle.

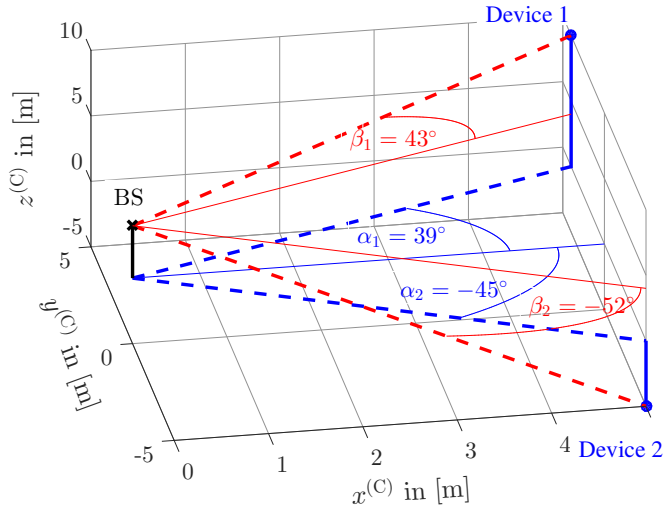


Figure 3.2.: Example of BS, mobile devices, and corresponding DoA angles. The orientation of the BS antenna array is $[1, 0, 0]^{\mathrm{T}}$.

where $-180° \leq \alpha \leq +180°$ and $-90° \leq \beta \leq +90°$ are the horizontal and vertical angles, respectively, according to Fig. 3.2. From Eq. (1.13) it is clear that the phase difference $\mu_{k,n}$ depends on the DoA of the incoming signal and the array geometry, e.g. the distance to the reference element.

One interpretation from Eq. (3.10) is that multi-path components in the propagation channel are the same as multiple-source DoA estimation [ASP14, BLG+15]. In order to quantify the quality of the fading channel the Rician K-factor denoted by $K^{(\mathrm{R})}$ is introduced and defined according to [Par92] for a link between the antenna of user k and BS antenna element n by

$$K_{k,n}^{(\mathrm{R})} = \frac{\left[\mathbf{P}_k^{(\mathrm{MPC})}\right]_{1,n}}{\sum_{l_{(\mathrm{MPC})}=2}^{L_{(\mathrm{MPC})}} \left[\mathbf{P}_k^{(\mathrm{MPC})}\right]_{l_{(\mathrm{MPC})},n}} = \frac{P_{k,n}^{(\mathrm{MPC,LoS})}}{P_{k,n}^{(\mathrm{MPC,NLoS})}}, \tag{3.13}$$

where $\left[\mathbf{P}_k^{(\mathrm{MPC})}\right]_{l_{(\mathrm{MPC})},n}$ denotes the path-loss of the $l_{(\mathrm{MPC})}$-th path between the antenna of user k and BS antenna n. It is assumed that paths are ordered by their delay such that $l_{(\mathrm{MPC})} = 1$ is the line of sight (LoS) path and remaining components $l_{(\mathrm{MPC})} = \left[2 \ \ldots \ L_{(\mathrm{MPC})}\right]$ are non line of sight (NLoS) paths. It can

be interpreted as the ratio of the power of the LoS component to the sum power of the NLoS components. One extreme of the K-factor is without fading, often called free space, and no NLoS components exist, and the other extreme is that only NLoS multi-path components exist, often referred to as Rayleigh fading or Rayleigh channel. Both extremes are expressed by

$$K_{k,n}^{(\mathrm{R})} = \begin{cases} \text{if free space :} & \lim\limits_{P_{k,n}^{(\mathrm{MPC,NLoS})} \to 0} K_{k,n}^{(\mathrm{R})} \to \infty \\ \text{if Rayleigh fading :} & \lim\limits_{P_{k,n}^{(\mathrm{MPC,LoS})} \to 0} K_{k,n}^{(\mathrm{R})} \to 0. \end{cases} \tag{3.14}$$

In practical systems, the number of detectable multi-path components is limited inter alia by thermal noise and receiver sensitivity [GME99]. This work follows the QuaDRiGa based channel (see Section 2.1) that considers $L_{(\mathrm{MPC})} = 8$ multi-path components in the dense-urban outdoor LoS scenario. Furthermore, it is assumed that the K-factor is approximately constant over co-located antennas of the same array.

Including receive beamforming at the BS, Eq. (3.1) can be rewritten as

$$\tilde{y}(t) = \mathbf{b}_{(\mathrm{BS})}^{\mathrm{H}}(\phi)\mathbf{y}(t) = \mathbf{b}_{(\mathrm{BS})}^{\mathrm{H}}(\phi)\left(\underline{\mathbf{H}}_{\mathcal{K}}^{(\mathrm{T})}(t)\right)^{\mathrm{H}}\mathbf{x}(t) + \mathbf{n}^{(\mathrm{BS})}(t), \tag{3.15}$$

where $\mathbf{b}_{(\mathrm{BS})}(\phi) \in \mathbb{C}^N$ is the steering matrix at a particular direction ϕ. The beamforming vector of a linear array as depicted in Fig. 1.5 at direction $\phi(\beta = 0)$ is obtained according to [Bal16] by

$$\mathbf{b}_{(\mathrm{BS})}(\alpha) = \begin{bmatrix} 1 \\ \exp\left(-j\frac{2\pi}{\lambda_{(\mathrm{c})}} d_{1,2}^{(\mathrm{N})} \cos(\alpha)\right) \\ \exp\left(-j\frac{2\pi}{\lambda_{(\mathrm{c})}} d_{1,3}^{(\mathrm{N})} \cos(\alpha)\right) \\ \vdots \\ \exp\left(-j\frac{2\pi}{\lambda_{(\mathrm{c})}} d_{1,N}^{(\mathrm{N})} \cos(\alpha)\right) \end{bmatrix}, \tag{3.16}$$

where according to Fig. 1.5 $d_{i,j}^{(\mathrm{N})}$ is the distance between the i-th and j-th antenna element. The design of $\mathbf{b}_{(\mathrm{BS})}$ becomes more complex for other geometries, e.g. in [KTH+16b] the beamforming vector is obtained for a uniform planar array (UPA) by the multiplication of two beamforming vectors for a linear array according to

$$\mathbf{b}_{(\mathrm{BS})} = \mathrm{vec}\left[\mathbf{b}_{(\mathrm{BS})}(\beta)\left(\mathbf{b}_{(\mathrm{BS})}(\alpha)\right)^{\mathrm{H}}\right], \tag{3.17}$$

where $\mathbf{b}_{(\mathrm{BS})}(\alpha) \in \mathbb{C}^{N_{(\alpha)}}$ and $\mathbf{b}_{(\mathrm{BS})}(\beta) \in \mathbb{C}^{N_{(\beta)}}$. $N_{(\alpha)}$ and $N_{(\beta)}$ denote the respective number of antennas in horizontal and vertical direction such that $N = N_{(\alpha)}N_{(\beta)}$, see Fig. 3.3. However, in order to be independent of a specific array configuration a different approach to calculate $\mathbf{b}_{(\mathrm{BS})}$ is used in this thesis. With the QuaDRiGa channel model [JRB+17] the free space coefficients of Eq. (3.6) of a mobile device positioned at direction ϕ_k are computed for $\lim K^{(\mathrm{R})} \to \infty$ such that $L_{(\mathrm{MPC})} = 1$ in Eq. (3.14). Note that the free space channel coefficients denoted by $\hat{\mathbf{h}}_k^{(\mathrm{TFS})}$ are deterministic and depend only on the geometry. These coefficients have to be computed only once and can be stored in a look-up table. Thus, the final beamformer is computed as

$$\mathbf{b}_{(\mathrm{BS})}(\phi_k) = \frac{\left(\hat{\mathbf{h}}_k^{(\mathrm{TFS})}(\phi_k)\right)^{\mathrm{H}}}{||\hat{\mathbf{h}}_k^{(\mathrm{TFS})}||}, \tag{3.18}$$

which inherently includes the antenna structure[4]. In Eq. (3.1) and following equations, the dependency on baseband time samples is introduced by index t. Without limitation to any specific waveform or modulation the time duration of a transmitted symbol is denoted by $t^{(\mathrm{sym})}$. In practical systems pilot or reference

[4] In QuaDRiGa version 1.4 function "qd_channel_builder.get_los_channels" can be used to directly calculate LoS only channel coefficients [HHI17].

signals are transmitted using sequences with suitable properties for a specific purpose, e.g. Zadoff-Chu (or Frank-Zadoff-Chu) sequences in LTE. Given a sequence length $N^{(\mathrm{seq})}$ this results in an observation time of

$$t^{(\mathrm{obs})} = N^{(\mathrm{seq})} t^{(\mathrm{sym})}. \tag{3.19}$$

The impact of channel ageing on parameter estimation is well studied in literature and not specific for massive MIMO. Therefore, in this work a constant and time-invariant channel is assumed over $t^{(\mathrm{obs})}$. The receive covariance matrix of the receive signal in Eq. (3.1) for an infinite small symbol length is given by the expectation value over the observation time according to

$$\mathbf{R}_{(\mathbf{yy})} = \mathbb{E}\left(\mathbf{y}(t)\mathbf{y}^{\mathrm{H}}(t)\right), \tag{3.20}$$

and in real systems with finite symbol length over sequence length $N^{(\mathrm{seq})}$ as

$$\mathbf{R}_{(\mathbf{yy})} = \frac{1}{N^{(\mathrm{seq})}} \sum_{t=1}^{N^{(\mathrm{seq})}} \mathbf{y}(t)\mathbf{y}^{\mathrm{H}}(t). \tag{3.21}$$

Including receive beamforming according to Eq. (3.15) the receive power $\tilde{P}^{(\mathrm{Rx})}(t)$ is given as

$$\tilde{P}_l^{(\mathrm{Rx})}(t) = \tilde{y}(t)\tilde{y}^*(t) \tag{3.22}$$

and the receive power $P^{(\mathrm{Rx})}(\phi)$ averaged over $N^{(\mathrm{seq})}$ samples by

$$\begin{aligned} P^{(\mathrm{Rx})}(\phi) &= \frac{1}{N^{(\mathrm{seq})}} \sum_{t=1}^{N^{(\mathrm{seq})}} \tilde{y}(t)\tilde{y}^{\mathrm{H}}(t) \\ &= \frac{1}{N^{(\mathrm{seq})}} \sum_{t=1}^{N^{(\mathrm{seq})}} \mathbf{b}_{(\mathrm{BS})}^{\mathrm{H}}(\phi)\mathbf{y}(t)\left(\mathbf{b}_{(\mathrm{BS})}^{\mathrm{H}}(\phi)\mathbf{y}(t)\right)^{\mathrm{H}} = \mathbf{b}_{(\mathrm{BS})}^{\mathrm{H}}(\phi)\mathbf{R}_{(\mathbf{yy})}\mathbf{b}_{(\mathrm{BS})}(\phi) \end{aligned} \tag{3.23}$$

denoted as the power spectrum due to the dependency on ϕ. In Eq. (3.6) the time domain representation of the channel is given where each multipath corresponds to a received delay. From previous equations, one can conclude that DoA estimation is independent of the bandwidth. That is not true, because usually the physical position of the antennas elements is fixed, e.g. in this work the distance between antenna elements is half the wavelength of the center frequency. Thus, if the frequency of the receive signal is higher or lower than the center frequency the antenna spacing in multiple of wavelengths is larger or smaller, respectively. Therefore, the bandwidth has an impact on beamforming and DoA estimation. However, the signal bandwidth can also be utilized to improve the DoA estimation or the positioning accuracy, e.g. in Chapter 4.

Furthermore, in current LTE and fifth generation (5G) Release 15 the channel estimation is done in the "frequency" domain [3GP17b] and the time-domain channel from above equations has to be obtained from it. Therefore, the orthogonal frequency division multiplexing (OFDM)-based frequency domain representation of the channel vector and matrix is introduced and denoted as $\mathbf{h}$ and $\mathbf{H}$, respectively. Given the signal bandwidth B and the number of subcarriers $N_{(\mathrm{s})}$, the channel coefficient of the n-th antenna element $[\mathbf{h}(t, n_s)]_n$ and the $n_{(\mathrm{s})}$-th subcarrier is obtained from Eq. (3.6) by a discrete Fourier transform (DFT) as

$$[\mathbf{h}(t, n_{(\mathrm{s})})]_n = \sum_{l_{(\mathrm{MPC})}=1}^{L_{(\mathrm{MPC})}} \left[\bar{\mathbf{h}}_{l_{(\mathrm{MPC})}}^{(\mathrm{T})}\right]_n \exp\left(-j2\pi n_{(\mathrm{s})} \tau_{l_{(\mathrm{MPC})}} B\right), \tag{3.24}$$

where $n_{(\mathrm{s})} \in [1, \ldots, N_{(\mathrm{s})}]$ is the position of the $n_{(\mathrm{s})}$-th OFDM subcarrier relative to the beginning of the frequency response.

3.1.2. DoA Estimation Techniques

It is already stated in the introduction and at the beginning of this chapter, the focus in this thesis is on searched-based techniques. Therefore, the well-known and often used search-based methods Bartlett, Capon and MUSIC are described in the next paragraphs as well as the Cramer Rao lower bound (CRLB) as reference.

Bartlett Method The Bartlett method [Bar48], also known as "conventional" beamformer, is a Fourier-based spectral analysis, or delay and sum method, where the receive power $P^{(\mathrm{Rx})}_{(\mathrm{Bartlett})}$ at a certain direction ϕ' is subject to the following optimization

$$\underset{\mathbf{b}_{(\mathrm{BS})}(\phi')}{\arg\max}\, \mathbb{E}\left(\tilde{P}^{(\mathrm{Rx})}(t)\right) = \underset{\mathbf{b}_{(\mathrm{BS})}(\phi')}{\arg\max}\, \mathbf{b}^{\mathrm{H}}_{(\mathrm{BS})}(\phi')\mathbb{E}\left(\mathbf{y}(t)\mathbf{y}^{\mathrm{H}}(t)\right)\mathbf{b}_{(\mathrm{BS})}(\phi'). \tag{3.25}$$

According to [KV96] the resulting solution is

$$\mathbf{b}^{(\mathrm{Bartlett})}_{(\mathrm{BS})}(\phi) = \frac{\mathbf{b}_{(\mathrm{BS})}(\phi)}{\sqrt{\mathbf{b}^{\mathrm{H}}_{(\mathrm{BS})}(\phi)\mathbf{b}_{(\mathrm{BS})}(\phi)}} = \frac{\mathbf{a}_{k,1}(\phi)}{\sqrt{\mathbf{a}^{\mathrm{H}}_{k,1}(\phi)\,\mathbf{a}_{k,1}}}(\phi), \tag{3.26}$$

assuming that the first multi-path component is the LoS path to be detected corresponding to the source direction ϕ, where $\mathbf{a}_{k,1}$ is given in Eq. (3.11). Inserting Eq. (3.26) into Eq. (3.23), the "Bartlett" power spectrum is obtained as

$$P^{(\mathrm{Rx})}_{(\mathrm{Bartlett})}(\phi) = \frac{\mathbf{b}^{\mathrm{H}}_{(\mathrm{BS})}(\phi)\mathbf{R}_{(\mathbf{yy})}\mathbf{b}_{(\mathrm{BS})}(\phi)}{\mathbf{b}^{\mathrm{H}}_{(\mathrm{BS})}(\phi)\mathbf{b}_{(\mathrm{BS})}(\phi)}. \tag{3.27}$$

This method can also be interpreted as the normalized covariance of the effective receive signal from Eq. (3.15). The Bartlett method is categorized as a search-based technique because DoAs can only be found in the applied search space that is here the power values in directions ϕ. Consequently, the complexity of the Bartlett method scales linearly with the number of searched directions $N^{(\phi)}$.

Capon Method J. Capon published in 1969 an algorithm [Cap69] that became later known as the "Capon" method. The intention of this approach was to overcome the limited resolution of the Bartlett method. It proposes to minimize the power from noise and any signals coming from other directions than the source DoA ϕ_k, while keeping a constant gain in the desired direction ϕ_k. The optimization problem is stated as [KV96]

$$\begin{aligned} &\min P^{(\mathrm{Rx})}\left(\mathbf{b}_{(\mathrm{BS})}\right) \\ &\text{subject to } \mathbf{b}^{\mathrm{H}}_{(\mathrm{BS})}(\phi')\mathbf{h}_k(\phi) = 1 \end{aligned} \tag{3.28}$$

resulting in the beamformer

$$\mathbf{b}^{(\mathrm{Capon})}_{(\mathrm{BS})}(\phi) = \frac{\mathbf{R}^{-1}_{(\mathbf{yy})}\mathbf{b}_{(\mathrm{BS})}(\phi)}{\mathbf{b}^{\mathrm{H}}_{(\mathrm{BS})}(\phi)\mathbf{R}^{-1}_{(\mathbf{yy})}\mathbf{b}_{(\mathrm{BS})}(\phi)}. \tag{3.29}$$

Inserting Eq. (3.29) into Eq. (3.23) leads to the following "Capon" power spectrum

$$P^{(\mathrm{Rx})}_{(\mathrm{Capon})}(\phi) = \frac{1}{\mathbf{b}^{\mathrm{H}}_{(\mathrm{BS})}(\phi)\mathbf{R}^{-1}_{(\mathbf{yy})}\mathbf{b}_{(\mathrm{BS})}(\phi)}. \tag{3.30}$$

From the above equation it is clear that the Capon method is also a search-based approach where the complexity scales with the number of searched directions $N^{(\phi)}$.

MUSIC Method MUSIC stands for multiple signal classification and was first proposed by Schmidt [Sch86] and [BK83]. In contrast to the previous described Bartlett and Capon method which can be seen as correlation-based approaches, the MUSIC approach utilizes the subspaces of the covariance matrix $\mathbf{R}_{(\mathbf{yy})}$. Therefore, Eq. (3.20) is rewritten as

$$\mathbf{R}_{(\mathbf{yy})} = \mathbb{E}\left(\mathbf{y}(t)\mathbf{y}^{\mathrm{H}}(t)\right) = \mathbb{E}\left(\underline{\mathbf{H}}^{(\mathrm{T})}(t)\mathbf{x}_{\mathcal{K}}(t)\left(\underline{\mathbf{H}}^{(\mathrm{T})}(t)\mathbf{x}_{\mathcal{K}}(t)\right)^{\mathrm{H}}\right) + \mathbb{E}\left(\mathbf{n}^{(\mathrm{BS})}(t)\left(\mathbf{n}^{(\mathrm{BS})}(t)\right)^{\mathrm{H}}\right) \tag{3.31}$$

under the assumption that the incident signals and noise are uncorrelated. Note, that the BS index l from Eq. (3.20) is omitted in Eq. (3.31) for ease of notation. With the assumption of time invariant channel over the observation time, Eq. (3.31) can be relaxed to

$$\mathbf{R}_{(\mathbf{yy})} = \left(\underline{\mathbf{H}}^{(\mathrm{T})}\right)^{\mathrm{H}}\mathbf{R}_{(\mathbf{xx})}\underline{\mathbf{H}}^{(\mathrm{T})} + \lambda\mathbf{R}^{(0)}_{(\mathbf{yy})}, \tag{3.32}$$

where $\mathbf{R}_{(\mathbf{xx})} \in \mathbb{C}^{K\times K}$ is a positive definite matrix that contains the pair-wise correlation between the transmit signals, $\mathbf{R}^{(0)}_{(\mathbf{yy})} \in \mathbb{C}^{N\times N}$ is the noise covariance matrix or noise space, and $\lambda \in \mathbb{R}_+$ is a scalar multiplier that corresponds to the noise variance $\left(\sigma^{(\mathrm{BS})}\right)^2$ under certain conditions which are clarified below. Notation λ is used to follow the general case and notation in [Sch86]. If the number of sources is less than the number of receive antennas such that $K < N$, then $\left(\underline{\mathbf{H}}^{(\mathrm{T})}\right)^{\mathrm{H}}\mathbf{R}_{(\mathbf{xx})}\underline{\mathbf{H}}^{(\mathrm{T})}$ is singular with rank $< N$ Eq. (3.32) can be reformulated according to [Sch86] as

$$\left|\left(\underline{\mathbf{H}}^{(\mathrm{T})}\right)^{\mathrm{H}}\mathbf{R}_{(\mathbf{xx})}\underline{\mathbf{H}}^{(\mathrm{T})}\right| = \left|\mathbf{R}_{(\mathbf{yy})} - \lambda\mathbf{R}^{(0)}_{(\mathbf{yy})}\right| = 0. \tag{3.33}$$

With $\mathbf{H}^{(\mathrm{T})}$ being a full rank matrix and $\mathbf{R}_{(\mathbf{xx})}$ being a positive definite matrix, the left hand side of Eq. (3.33) is positive semi-definite. Therefore, Eq. (3.33) can only be satisfied when λ is an eigenvalue of $\mathbf{R}_{(\mathbf{yy})}$ in the space of $\mathbf{R}^{(0)}_{(\mathbf{yy})}$. It follows that the smallest solution to $\left|\mathbf{R}_{(\mathbf{yy})} - \lambda\mathbf{R}^{(0)}_{(\mathbf{yy})}\right|$ is $\lambda_{(\min)} \geq 0$. If $\mathbf{n}^{(\mathrm{BS})}$ has zero mean and variance $\left(\sigma^{(\mathrm{BS})}\right)^2$ the special case $\lambda\mathbf{R}^{(0)}_{(\mathbf{yy})} = \left(\sigma^{(\mathrm{BS})}\right)^2\mathbf{I}_N$ holds. Skipping the details of the solution which can be found in [Sch86] it follows that the N eigenvectors $\mathbf{e}_1^{(\mathbf{R}_{(\mathbf{yy})})}, \ldots, \mathbf{e}_N^{(\mathbf{R}_{(\mathbf{yy})})}$ of $\mathbf{R}_{(\mathbf{yy})}$ must fulfill

$$\mathbf{R}_{(\mathbf{yy})}\mathbf{e}_n^{(\mathbf{R}_{(\mathbf{yy})})} = \lambda_n\mathbf{R}^{(0)}_{(\mathbf{yy})}. \tag{3.34}$$

According to Eq. (3.32) and $\lambda_{(\min)}$ the following condition is obtained,

$$\left(\mathbf{H}^{(\mathrm{T})}\right)^{\mathrm{H}}\mathbf{R}_{(\mathbf{xx})}\mathbf{H}^{(\mathrm{T})}\mathbf{e}_n^{(\mathbf{R}_{(\mathbf{yy})})} = (\lambda_n - \lambda_{(\min)})\mathbf{R}^{(0)}_{(\mathbf{yy})}\mathbf{e}_n^{(\mathbf{R}_{(\mathbf{yy})})}. \tag{3.35}$$

If λ_n is equal to $\lambda_{(\min)}$ there must be an eigenvector $\mathbf{e}_n^{(\mathbf{R}_{(\mathbf{yy})})}$ that satisfies $\left(\mathbf{H}^{(\mathrm{T})}\right)^{\mathrm{H}}\mathbf{R}_{(\mathbf{xx})}\mathbf{H}\mathbf{e}_n^{(\mathbf{R}_{(\mathbf{yy})})} = 0$ or more specifically $\mathbf{H}^{(\mathrm{T})}\mathbf{e}_n^{(\mathbf{R}_{(\mathbf{yy})})} = 0$. In other words the eigenvectors that correspond to the noise space are orthogonal to the space spanned by $\mathbf{H}^{(\mathrm{T})}$, which contains according to Eq. (3.6) the DoA vectors of the incoming signals. In [Sch86] it is shown that there are $N^{(*)} = N - K$ such eigenvalues[5]. Therefore, $\mathbf{R}_{(\mathbf{yy})}$ can be divided into a $N^{(*)}$ dimensional noise subspace and a disjoint K dimensional "signal" space. Consequently, the eigenvalues of $\mathbf{R}_{(\mathbf{yy})}$, represented by $\lambda_1, \lambda_2, \ldots, \lambda_N$ can also be divided into K "signal" eigenvalues and $N^{(*)}$ noise eigenvalues, such that

$$\lambda_n = \begin{cases} \text{signal eigenvalue if} & : \quad \lambda_n > \left(\sigma^{(\mathrm{BS})}\right)^2 \\ \text{noise eigenvalue if} & : \quad \lambda_n = \left(\sigma^{(\mathrm{BS})}\right)^2. \end{cases} \tag{3.36}$$

[5] This is only true for $L_{(\mathrm{MPC})} = 1$ because in [Sch86] only the LoS path is considered. However, without limit to generality each multipath for $L_{(\mathrm{MPC})} > 1$ can be interpreted as another independent source.

As mentioned before, the steering vectors are orthogonal to the noise subspace in the direction of incident signals such that

$$\mathbf{b}_{(\mathrm{BS})}^{\mathrm{H}}(\phi_k)\mathbf{E}_{(\mathrm{n})}\mathbf{E}_{(\mathrm{n})}^{\mathrm{H}}\mathbf{b}_{(\mathrm{BS})}(\phi_k) = 0, \tag{3.37}$$

where ϕ_k is the direction to a device (source) k and $\mathbf{E}_{(\mathrm{n})} \in \mathbb{C}^{N \times N^{(*)}}$ is the noise subspace matrix spanned by the noise eigenvectors corresponding to the noise eigenvalues from Eq. (3.36). Consequently, the power spectrum of the MUSIC approach is given by

$$P_{(\mathrm{MUSIC})}^{(\mathrm{Rx})}(\phi) = \frac{1}{\mathbf{b}_{(\mathrm{BS})}^{\mathrm{H}}(\phi)\mathbf{E}_{(\mathrm{n})}\mathbf{E}_{(\mathrm{n})}^{\mathrm{H}}\mathbf{b}_{(\mathrm{BS})}(\phi)}. \tag{3.38}$$

While using a different approach than the Bartlett and Capon method, the MUSIC is also a search-based approach, where the pseudo-power spectrum in each direction has to be computed and complexity scales with $N^{(\phi)}$.

For the sake of completion and clarification the MUSIC algorithm requires certain conditions to work properly [SN89]. As shown below these conditions are fulfilled in this thesis.

1. As weak assumption that guarantees uniqueness of the estimator, the number of antennas must be larger than the number of signals $N > K$. With large N in massive MIMO arrays this condition is fulfilled in this thesis and with high probability in real systems.

2. Spatially and temporally uncorrelated noise, such that $\mathbb{E}\left(\mathbf{n}^{(\mathrm{BS})}(t)\left(\mathbf{n}^{(\mathrm{BS})}(t)\right)^{\mathrm{H}}\right) = \sigma^{(\mathrm{BS})}\mathbf{I}$ and $\mathbb{E}\left(\mathbf{n}^{(\mathrm{BS})}(t)\left(\mathbf{n}^{(\mathrm{BS})}(t)\right)^{\mathrm{T}}\right) = 0$. Furthermore, $\mathbb{E}\left(\mathbf{x}_{\mathcal{K}}(t)\left(\mathbf{x}_{\mathcal{K}}(t)\right)^{\mathrm{H}}\right)$ must be non-singular (positive definite). These assumptions coincide with the assumptions in Section 3.1.1.

Estimated Direction of Arrival According to Eq. (3.12) the DoA ϕ consists of α and β indicating the true DoA. The estimated DoA is indicated by the accent $\hat{}$, such that $\hat{\phi}$, $\hat{\alpha}$, and $\hat{\beta}$ denote the estimated DoA, estimated horizontal angle, and estimated vertical angle, respectively. For the search-based techniques described above, these estimates are obtained according to

$$\hat{\phi}\left(\hat{\alpha}, \hat{\beta}\right) = \arg\max_{\phi}\left(P_i^{(\mathrm{Rx})}\left(\phi\right)\right), \tag{3.39}$$

where $i \in \{(\mathrm{Bartlett}), (\mathrm{Capon}), (\mathrm{MUSIC})\}$ and $P_i^{(\mathrm{Rx})}(\phi)$ is the respective power spectrum obtained by Eq. (3.27), Eq. (3.30), or Eq. (3.38).

Cramer-Rao Lower Bound In statistics and estimation theory the Cramer Rao bound (CRB) (also referred to as CRLB), named after Harald Cramer [Cra45] and Calyampudi Radhakrishna Rao [Rao92], is a lower bound on the variance of an estimator of a deterministic parameter. This lower bound is also referred to as the Cramer-Rao inequality or the information inequality. The CRB is the lower performance bound for any unbiased estimator. Especially when the exact minimum mean square error (MMSE) is intractable to derive, the CRB can be used to compare the relative performance of the estimator. In [SN89] the CRB is derived for DoA estimation of linear arrays. The CRB of a signal from direction ϕ is generalized to arbitrary array geometries for a DoA estimate in the $\mathbb{R}^3$ and given as [BM03]

$$\sigma^{(\mathrm{CRB})}\left(\phi\right) = \left[\mathbf{G}_{(\mathrm{CRB})}(\phi)P_{(\mathrm{CRB})}\right]^{-1}, \tag{3.40}$$

where

$$\mathbf{G}_{(\mathrm{CRB})}(\phi) = \mathbf{J}_{(\mathrm{CRB})}^{\mathrm{T}}(\mathbf{u}_{(\phi)})\mathbf{B}_{(\mathrm{CRB})}\mathbf{J}_{(\mathrm{CRB})}(\mathbf{u}_{(\phi)}) \tag{3.41}$$

and

$$P_{(\mathrm{CRB})} = 2\left(\frac{f_{(\mathrm{c})}}{v_{(\mathrm{c})}}\right)^2 \frac{N^{(\mathrm{seq})} N P^{(\mathrm{s})}}{P^{(\mathrm{n})}} \left(1 - \frac{P^{(\mathrm{n})}}{P^{(\mathrm{s})} N + P^{(\mathrm{n})}}\right) \overset{A}{=} 2\left(\frac{2\pi}{\lambda_{(\mathrm{c})}}\right)^2 \frac{N^{(\mathrm{seq})} N^2 \tilde{\gamma}^2}{N\tilde{\gamma} + 1}, \tag{3.42}$$

where $N^{(\mathrm{seq})}$ is introduced above as the sequence length , $P^{(\mathrm{s})}$ denotes the source power according to Eq. (3.3), and $P^{(\mathrm{n})}$ the noise power following Eq. (3.4). The details of the reformulation of Eq. (3.42) from left to right are given in Chapter A below the respective headline[6] indicated by the notation $\overset{A}{=}$. In Eq. (3.41) $\mathbf{J}_{(\mathrm{CRB})}(\mathbf{u}_{(\phi)}(\alpha, \beta))$ denotes the Jacobian matrix[7] given by

$$\mathbf{J}_{(\mathrm{CRB})}(\mathbf{u}_{(\phi)}(\alpha, \beta)) = \left[\frac{\partial \mathbf{u}_{(\phi)}(\alpha, \beta)}{\partial \alpha} \; \frac{\partial \mathbf{u}_{(\phi)}(\alpha, \beta)}{\partial \beta}\right] = \begin{bmatrix} -\cos(\beta)\sin(\alpha) & -\sin(\beta)\cos(\alpha) \\ \cos(\beta)\cos(\alpha) & -\sin(\beta)\sin(\alpha) \\ 0 & \cos(\beta) \end{bmatrix}, \tag{3.43}$$

where $\mathbf{u}_{(\phi)}(\alpha, \beta)$ is defined in Eq. (3.12), and $\mathbf{B}_{(\mathrm{CRB})} \in \mathbb{R}^{3\times 3}$ given by

$$\mathbf{B}_{(\mathrm{CRB})} = \frac{1}{N} \sum_{n=1}^{N} (\tilde{\mathbf{p}}_n^{(\mathrm{N})} - \mathbf{p}^{(\mathrm{array})})(\tilde{\mathbf{p}}_n^{(\mathrm{N})} - \mathbf{p}^{(\mathrm{array})})^{\mathrm{T}}, \tag{3.44}$$

where $\tilde{\mathbf{p}}_n^{(\mathrm{N})} \in \mathbb{R}^3$ is the position of the n-th antenna element in Cartesian coordinates. Eq. (3.44) can be interpreted as an array geometry dependent term depending on the number and position of the antenna elements in the array where $\mathbf{p}^{(\mathrm{array})}$ is the centroid of the array, i.e.,

$$\mathbf{p}^{(\mathrm{array})} = \frac{1}{N} \sum_{n=1}^{N} \tilde{\mathbf{p}}_n^{(\mathrm{N})}. \tag{3.45}$$

3.1.3. Performance Metrics

In this section the performance metrics used for evaluation are described. The DoA estimation errors are defined as

$$\Delta_{(\alpha)} = \alpha - \hat{\alpha} \tag{3.46}$$

and

$$\Delta_{(\beta)} = \beta - \hat{\beta}, \tag{3.47}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are obtained by Eq. (3.39). From the definition in this thesis, the angles are constraint by

$$\alpha \in \mathbb{R}, \alpha \in [-180, 180)^\circ \tag{3.48}$$

and

$$\beta \in \mathbb{R}, \beta \in [-90, 90)^\circ. \tag{3.49}$$

In contrast to this, $\hat{\alpha}$ and $\hat{\beta}$ are constraint by the number of searched directions previously introduced by $N^{(\phi)}$ and the quantization of the search space. The search space is split into horizontal and vertical search space, denoted by $\boldsymbol{\alpha}^{(\mathrm{f})}$ and $\boldsymbol{\beta}^{(\mathrm{f})}$, respectively. The angular quantization step size is denoted by $q^{(\phi)}$ and can be interpreted similar to the quantization noise from analog to digital converters. Under the assumption of uniformly distributed α and β, a uniform linear quantization minimizes the estimation error that is uniformly distributed in

$$0 \leq \left|\Delta_{(\alpha)}\right| \leq \frac{q^{(\phi)}}{2} \tag{3.50}$$

[6] Headlines in Appendix A are given for example as "Equation (3.42)" which indicates further equations on rearrangements for Eq. (3.42). The headlines are ordered numerically ascending.

[7] The Jacobian matrix is the matrix of all first-order partial derivatives of a vector-valued function, see nomenclature in the front matter.

and

$$0 \leq \left|\Delta_{(\beta)}\right| \leq \frac{q^{(\phi)}}{2}. \tag{3.51}$$

Thus for a certain target quantization $q^{(\phi)}$ of the search space $\boldsymbol{\alpha}^{(\mathrm{f})}$ this requires

$$N_{(\alpha)}^{(\phi)} = \left\lceil \frac{[\boldsymbol{\alpha}^{(\mathrm{f})}]_2 - [\boldsymbol{\alpha}^{(\mathrm{f})}]_1}{q^{(\phi)}} \right\rceil + f^{(\mathrm{s})}(\boldsymbol{\alpha}^{(\mathrm{f})}, q^{(\phi)}) \tag{3.52}$$

computations of the power spectrum, where function $f^{(\mathrm{s})}(\boldsymbol{\alpha}^{(\mathrm{f})}, q^{(\phi)})$ is defined by

$$f^{(\mathrm{s})}(\boldsymbol{\alpha}^{(\mathrm{f})}, q^{(\phi)}) = \begin{cases} 1, & \text{if mod}\left(\frac{[\boldsymbol{\alpha}^{(\mathrm{f})}]_2 - [\boldsymbol{\alpha}^{(\mathrm{f})}]_1}{q^{(\phi)}}, 1\right) = 0 \\ 0, & \text{else.} \end{cases} \tag{3.53}$$

From Eq. (3.12) it is clear that $\hat{\alpha}$ and $\hat{\beta}$ cannot be estimated independent from each other. Therefore, each $\hat{\beta} \in \boldsymbol{\beta}^{(\mathrm{f})}$ requires $N_{(\alpha)}^{(\phi)}$ computations resulting in

$$N^{(\phi)} = N_{(\alpha)}^{(\phi)} N_{(\beta)}^{(\phi)} = \left(\left\lceil \frac{[\boldsymbol{\alpha}^{(\mathrm{f})}]_2 - [\boldsymbol{\alpha}^{(\mathrm{f})}]_1}{q^{(\phi)}} \right\rceil + f^{(\mathrm{s})}(\boldsymbol{\alpha}^{(\mathrm{f})}, q^{(\phi)}) \right) \left(\left\lceil \frac{[\boldsymbol{\beta}^{(\mathrm{f})}]_2 - [\boldsymbol{\beta}^{(\mathrm{f})}]_1}{q^{(\phi)}} \right\rceil + f^{(\mathrm{s})}(\boldsymbol{\beta}^{(\mathrm{f})}, q^{(\phi)}) \right), \tag{3.54}$$

where $N_{(\beta)}^{(\phi)}$ is the number of computations for $\boldsymbol{\beta}^{(\mathrm{f})}$ equivalent to Eq. (3.52) for $\boldsymbol{\alpha}^{(\mathrm{f})}$. Thus $N^{(\phi)}$ is considered as another metric.

The half-power beam-width (HPBW), often called 3 dB beam-width, of the detected peak $\hat{\phi}$ and the power ratio between the peak and the strongest side-lobe are used to measure the quality of the power spectrum. An illustration of the metrics is shown in Section 3.2.1 in Fig. 3.8. The HPBW of $\hat{\alpha}$ and $\hat{\beta}$ denoted by $\hat{\alpha}_{(\mathrm{HPBW})}$ and $\hat{\beta}_{(\mathrm{HPBW})}$, respectively, is obtained by

$$\begin{aligned} \hat{\alpha}_{(\mathrm{HPBW})} &= \hat{\alpha}_{(\mathrm{right})} - \hat{\alpha}_{(\mathrm{left})} \\ \hat{\beta}_{(\mathrm{HPBW})} &= \hat{\beta}_{(\mathrm{right})} - \hat{\beta}_{(\mathrm{left})}, \end{aligned} \tag{3.55}$$

where

$$\begin{aligned} \hat{\alpha}_{(\mathrm{right})} &= \arg\min_{\alpha} \left| \frac{P_i^{(\mathrm{Rx})}\left(\hat{\alpha}, \hat{\beta}\right)}{2} - P_i^{(\mathrm{Rx})}\left(\alpha, \hat{\beta}\right) \right|, \quad \alpha > \hat{\alpha} \\ \hat{\alpha}_{(\mathrm{left})} &= \arg\min_{\alpha} \left| \frac{P_i^{(\mathrm{Rx})}\left(\hat{\alpha}, \hat{\beta}\right)}{2} - P_i^{(\mathrm{Rx})}\left(\alpha, \hat{\beta}\right) \right|, \quad \alpha < \hat{\alpha} \\ \hat{\beta}_{(\mathrm{right})} &= \arg\min_{\beta} \left| \frac{P_i^{(\mathrm{Rx})}\left(\hat{\alpha}, \hat{\beta}\right)}{2} - P_i^{(\mathrm{Rx})}\left(\hat{\alpha}, \beta\right) \right|, \quad \beta > \hat{\beta} \\ \hat{\beta}_{(\mathrm{left})} &= \arg\min_{\beta} \left| \frac{P_i^{(\mathrm{Rx})}\left(\hat{\alpha}, \hat{\beta}\right)}{2} - P_i^{(\mathrm{Rx})}\left(\hat{\alpha}, \beta\right) \right|, \quad \beta < \hat{\beta}. \end{aligned} \tag{3.56}$$

The ratio of $P_i^{(\mathrm{Rx})}\left(\hat{\alpha}, \hat{\beta}\right)$ and the strongest side-lobe denoted by $P_i^{(\mathrm{Rx})}\left(\hat{\alpha}^{(2)}, \hat{\beta}^{(2)}\right)$ is computed according to

$$P^{(\mathrm{MSLR})} = \frac{P_i^{(\mathrm{Rx})}\left(\hat{\alpha}, \hat{\beta}\right)}{P_i^{(\mathrm{Rx})}\left(\hat{\alpha}^{(2)}, \hat{\beta}^{(2)}\right)}, \tag{3.57}$$

where the superscript MSLR stands for minimum side lobe ratio (MSLR). Note, that also in the MUSIC pseudo-spectrum a side-lobe will be detected due to unavoidable numerical inaccuracies with digital signal processing.

Table 3.1.: Summary of DoA estimation performance metrics.

Metric	Description
$\Delta_{(\alpha)}, \Delta_{(\beta)}$	Horizontal and vertical angle estimation error, see Eq. (3.46) and Eq. (3.47)
$\mathbf{n}^{(\phi,\mathrm{S})}$	Vector that contains the number of computations of power spectrum at each step of adaptive search space quantization (ASSQ)
$N^{(\phi)} = \sum_{s=1}^{S} \left[\mathbf{n}^{(\phi,\mathrm{S})}\right]_s$	Total number of computations of power spectrum for search-based techniques, see Eq. (3.54)
$\hat{\alpha}_{(\mathrm{HPBW})}, \hat{\beta}_{(\mathrm{HPBW})}$	HPBW of estimated DoA $\hat{\phi}$, see Eq. (3.55).
$P^{(\mathrm{MSLR})}$	Power ratio of peak to second largest peak in the power spectrum, see Eq. (3.57).

Finding $P_i^{(\mathrm{Rx})}\left(\hat{\alpha}^{(2)}, \hat{\beta}^{(2)}\right) \in P_i^{(\mathrm{Rx})}(\alpha, \beta)$ over the two angular directions corresponds to the problem of multi-peak detection in a two dimensional (2D) data-array $\mathbb{R}^2$. This is a classical problem in image processing [VP11] and there exist several algorithms to solve it, e.g. [KB75, Lem06, VH92]. In this work, an approach similar to Lemires algorithm [Lem06] is used.

For numerical evaluation, a statistic over multiple Monte-Carlo runs is generated and the cumulative distribution functions (CDFs) of the above mentioned metrics are obtained. From these CDFs the $(5, 50, 95)$ %-ile values are taken. The median value is used to visualize trends for parameter studies, and the $(5, 95)$ %-ile values provide a measure on the variance of the CDF.

The main performance metrics discussed in the section are summarized in Table 3.1.

3.2. Parameter Analysis and Complexity Reduction

3.2.1. Impact of Parameters

For the study and evaluation of the DoA estimation techniques described in Section 3.1 and the impact of the relevant parameters an uniform planar antenna array is considered. The main reasons for this are that UPAs are assumed in current standardization [3GP17e] and often used in literature for DoA estimation performance evaluation [Sch86, Dho10, KKA13, TI13, BLG+15, SS15]. A major drawback of UPAs is the ambiguity of the beam pattern formed by the receive beamformer $\mathbf{b}_{(\mathrm{BS})}$ from Eq. (3.15). An example for an 8×8 UPA is shown in Fig. 3.6a for $\phi(\alpha = 30°, \beta = 0°)$. Without limiting the generality, the N antenna elements are positioned on the $y^{(\mathrm{C})}$-$z^{(\mathrm{C})}$ plane[8]. Further a distance of $\lambda_{(\mathrm{c})}/2$ between two neighbor elements in $y^{(\mathrm{C})}$ or $z^{(\mathrm{C})}$ direction is assumed. With N antenna elements at the BS this results in $N = N_{(\alpha)}N_{(\beta)}$, where $N_{(\alpha)}$ and $N_{(\beta)}$ are the number of elements parallel to the $y^{(\mathrm{C})}$-axis and $z^{(\mathrm{C})}$-axis, respectively.

Computing the spectrum, e.g. according to Eq. (3.27), for an angular range $\alpha \in [-180, 180]°$ results in two maxima, visualized in Fig. 3.6b. It becomes clear that for each maximum in the intervall $[-180\ 0)°$ there is a second maximum in the interval $[180\ 0)°$. Such ambiguities can be avoided with circular arrays or directional antenna elements. Consequently, the angular search range in case of a UPA is set to $\boldsymbol{\alpha}^{(\mathrm{f})} = [-90\ 90]$ to avoid ambiguities in the results.

According to [3GP16d] there is a 1 m accuracy target for positioning in the 5G mobile communications system. This requirement can be transformed to a DoA estimation accuracy for a certain distance between BS and mobile device. In Fig. 3.5 the position error caused by DoA estimates is denoted by $\Delta_{(\mathrm{xyz})}$. In

[8] The notations $x^{(\mathrm{C})}$, $y^{(\mathrm{C})}$ and $z^{(\mathrm{C})}$ are used for directions in the 3D Cartesian coordinate system to clearly distinguish them from transmit signal x and receive signal y. See also symbol definition in the front matter.

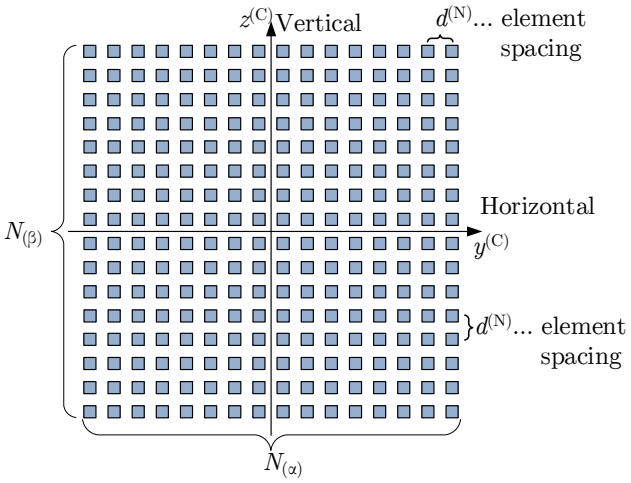
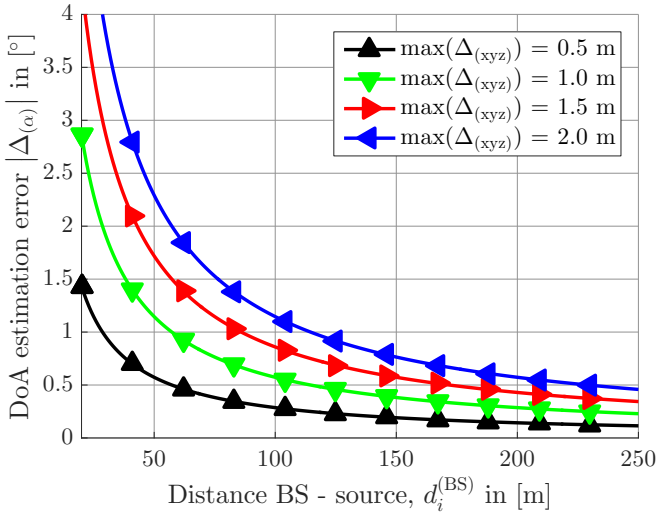
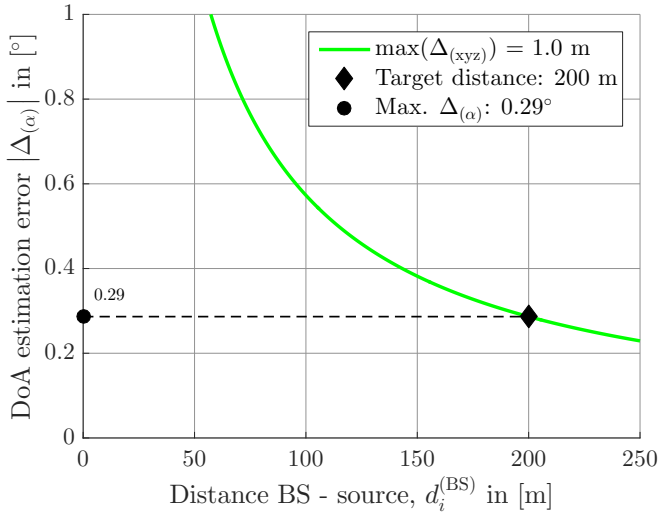


Figure 3.3.: Exemplary UPA with $N_{(\alpha)} = N_{(\beta)} = 16$ antenna elements in horizontal and vertical dimension, respectively.



(a) Required angle estimation accuracy $|\Delta_{(\alpha)}|$ to ensure $\Delta_{(xyz)} \leq \epsilon$, where ϵ is a threshold set to $(0.5, 1, 1.5, 2)$ m.



(b) Required angle estimation accuracy $|\Delta_{(\alpha)}|$ to ensure $\Delta_{(xyz)} \leq 1$ m at $d_i^{(BS)} = 200$ m.

Figure 3.4.: Maximum DoA estimation error to guarantee a maximum position error over distance.

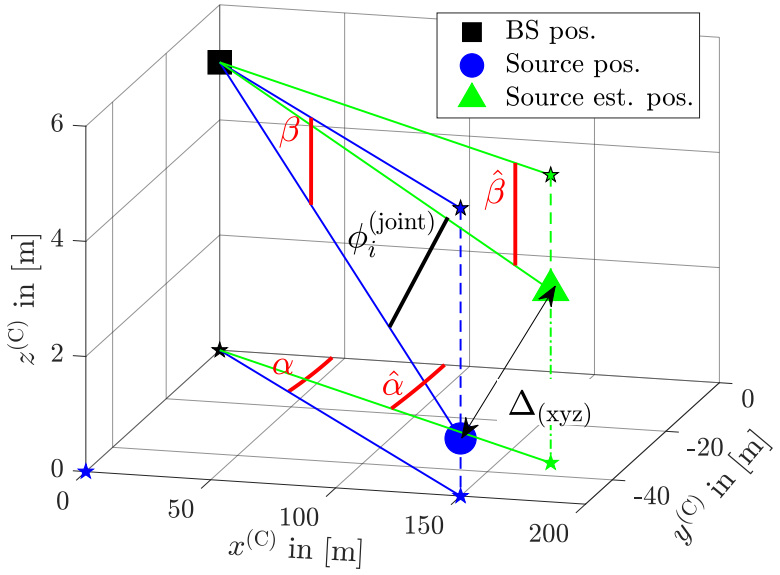


Figure 3.5.: Estimation errors in three-dimensional coordinates.

order to analyze the relation between DoA estimation error and position error the distance $d_i^{(\text{BS})}$ between mobile device i and the BS is assumed to be known error free, e.g. from time of flight (ToF) measurement in Section 1.2.1. A geometrical interpretation of this is that the mobile device is located on a sphere with radius $d_i^{(\text{BS})}$. By DoA estimation, a unique position estimate is derived by the intersection of the DoA vector and the sphere. Under these assumptions, an upper bound on the DoA estimation error over distance for an upper bound position error can be analytically derived by basic geometry. The position error $\Delta_{(\text{xyz})}$ is obtained by Lambert's cosine law [MMWW04] such that

$$\Delta_{(\text{xyz})} = d_i^{(\text{BS})} \sqrt{2\left(1 - \cos \phi_i^{(\text{joint})}\right)}, \tag{3.58}$$

where $\phi_i^{(\text{joint})}$ is the angle spanned by the true direction ϕ and the estimated direction $\hat{\phi}$ of device i according to Fig. 3.5. In Eq. (3.98) $\phi_i^{(\text{joint})}$ is obtained as $\cos \phi_i^{(\text{joint})} = \cos \Delta_{(\alpha)} \cos \Delta_{(\beta)}$, where $\Delta_{(\alpha)}$ and $\Delta_{(\beta)}$ are the horizontal and vertical angle difference between the true and estimated source position, respectively. The position error in Eq. (3.58) can be rewritten as

$$\Delta_{(\text{xyz})} = d_i^{(\text{BS})} \sqrt{2\left(1 - \cos \Delta_{(\alpha)} \cos \Delta_{(\beta)}\right)}. \tag{3.59}$$

Rearranging Eq. (3.59) to $\Delta_{(\alpha)}$ this yields

$$\Delta_{(\alpha)} = \arccos\left(\frac{1 - \frac{1}{2}\left(\frac{\Delta_{(\text{xyz})}}{d_i^{(\text{BS})}}\right)^2}{\cos \Delta_{(\beta)}}\right). \tag{3.60}$$

An example for $\Delta_{(\beta)} = 0°$ is shown in Fig. 3.4a which can be used to design systems, e.g. for a maximum position error of $\Delta_{(\text{xyz})} \leq 1\,\text{m}$ at a distance of $d_i^{(\text{BS})} = 200\,\text{m}$ the maximum acceptable DoA estimation error is $|\Delta_\alpha| \leq 0.29°$. According to Eq. (3.60), $\Delta_{(\text{xyz})}$ scales linearly with $d_i^{(\text{BS})}$ and by $\propto \sqrt{2\left(1 - \cos \Delta_{(\alpha)} \cos \Delta_{(\beta)}\right)}$. With angle estimation errors $\left|\Delta_{(\alpha)}\right| < 10°$ the small-angle approximation [Boa66]

$$\cos \Delta_{(\alpha)} \approx 1 - \frac{\Delta_{(\alpha)}^2}{2} \tag{3.61}$$

Table 3.2.: Simulation parameters for DoA estimation.

Parameter	**Value**
Simulation type	Monte Carlo 2500 realizations
Channel model	QuaDRiGa version 1.4 [JRBT14]
Scenario	Urban macro LoS [JRBT14]
Center frequency	3.75 GHz
Number of multi-path components $L_{(\mathrm{MPC})}$	1 if single path, 8 if multipath
If $L_{(\mathrm{MPC})} > 1$, Rician K-Factor $K^{(\mathrm{R})}$	10 dB
BS antenna distribution	UPA
Number of antenna elements N at BS	64
Number of elements in $y^{(\mathrm{C})}$-direction $N_{(\alpha)}$	8
Number of elements in $z^{(\mathrm{C})}$-direction $N_{(\beta)}$	8
Antenna element spacing in $y^{(\mathrm{C})}$- and $z^{(\mathrm{C})}$-dimension	$\lambda_{(\mathrm{c})}/2$
Antenna element type	Isotrop
BS height $\left[\tilde{\mathbf{p}}^{(\mathrm{BS})}\right]_3$	0 m
Sequence length, number of samples $N^{(\mathrm{seq})}$	10
Angular quantization $q^{(\phi)}$	0.1°
Horizontal coverage range $\boldsymbol{\alpha}^{(\mathrm{f})}$	$[-90\ 90]^\circ$
Horizontal angle of device α	random i.i.d. in $[25, 35]^\circ \in \mathbb{R}$
Elevation angle of device β	0°
Normalization value $g^{(\mathrm{norm})}$ Eq. (3.9)	1
Distance BS-Device d_{xyz}	50 m

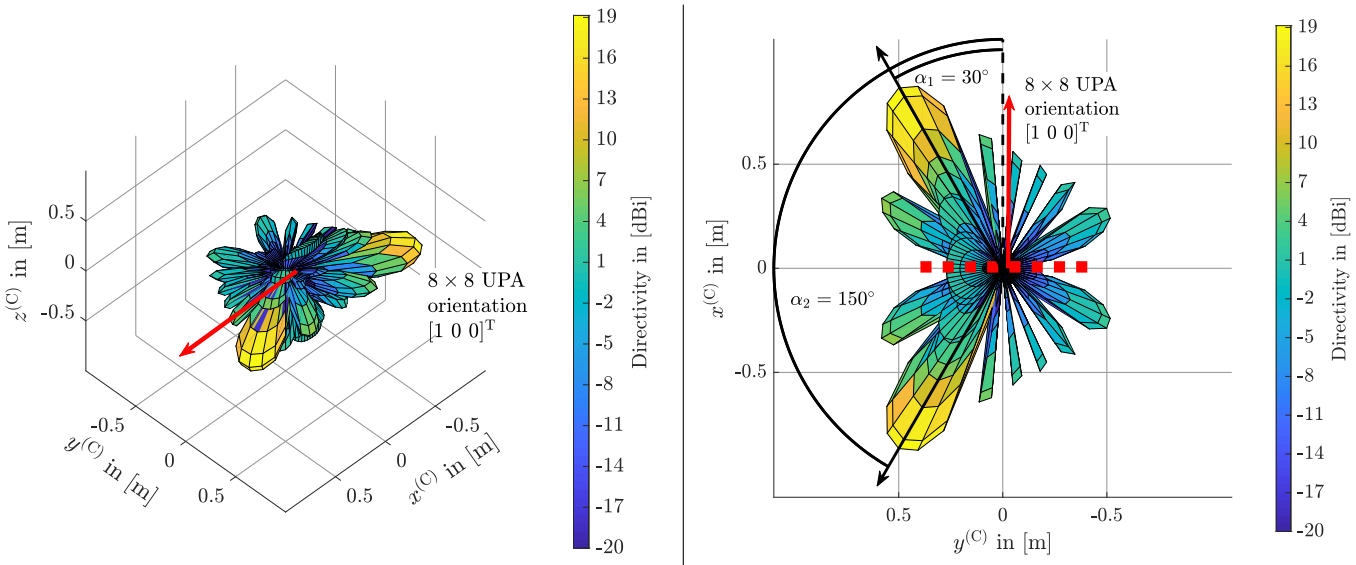
can be applied to Eq. (3.59) with a relative error $< 0.0039\,\%$ such that

$$\Delta_{(\mathrm{xyz})} \approx d^{(\mathrm{BS})} \sqrt{2\left(1-\left(1-\frac{\Delta_{(\alpha)}^2}{2}\right)\left(1-\frac{\Delta_{(\beta)}^2}{2}\right)\right)} = d^{(\mathrm{BS})} \sqrt{\Delta_{(\alpha)}^2 + \Delta_{(\beta)}^2 - \frac{\Delta_{(\alpha)}^2 \Delta_{(\beta)}^2}{2}}. \tag{3.62}$$

It is clear from Eq. (3.62) that $\Delta_{(\mathrm{xyz})}$ scales linearly with $\max(\Delta_{(\alpha)}, \Delta_{(\beta)})$. Therefore, the quantization of the spectrum (angular search space) is set to $q^{(\phi)} = 0.1^\circ$ which results in $N^{(\phi)} = 1801$ computations of the power spectrum, according to Eq. (3.54). Other simulation parameters as default setting for the numerical analysis of the various parameters impacting the DoA estimation techniques are given in Table 3.2.

Fig. 3.7a shows an example for the power spectrum of Eq. (3.27), Eq. (3.30), and Eq. (3.38), according to Eq. (3.27), Eq. (3.30), Eq. (3.38), respectively. As a reference, the solid line without marker represents the true angle. The Bartlett technique generates significant side-lobes over the observed angular range, whereas the Capon method has approximately the same power level at -19 dB besides the peak, e.g. the difference between side-lobe peaks and valleys is < 1 dB. The largest distance of side-lobes to the peak is provided by MUSIC, which is 29 dB. The corresponding metric, the minimum side lobe ratio according to Eq. (3.57) is illustrated on the left-hand side in Fig. 3.8. In order to emphasize the different behavior around the peak, Fig. 3.7b shows on the right-hand side a smaller angular range of $\boldsymbol{\alpha}_{(\mathrm{f})} = [15\ 45]^\circ$. According to [BLG+15] the gradient around the peak is increasing from Bartlett, to Capon, to MUSIC. Consequently, the HPBW according to Eq. (3.55) is largest for Bartlett and smallest for MUSIC, given on the left-hand side of Fig. 3.8.

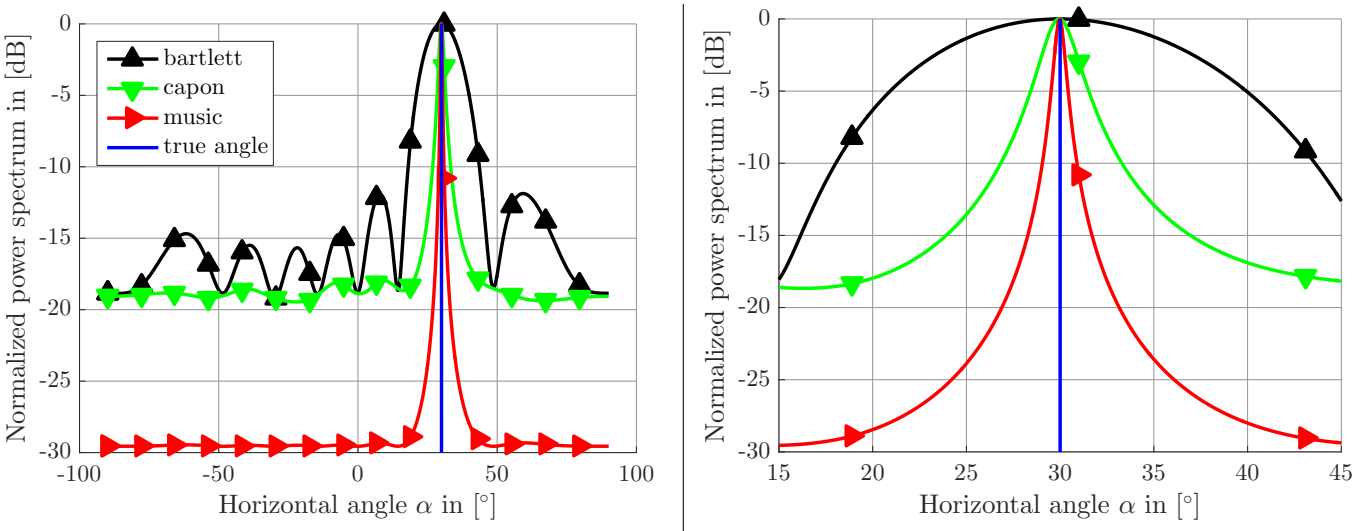
Note that Fig. 3.7 and Fig. 3.8 are generated assuming a single path such that $L_{(\mathrm{MPC})} = 1$ in order to illustrate the principle differences of the DoA estimation methods. However, in ground-based scenarios



(a) Diagonal view.

(b) Pattern ambiguity in the $x^{(C)}$-$y^{(C)}$ plane.

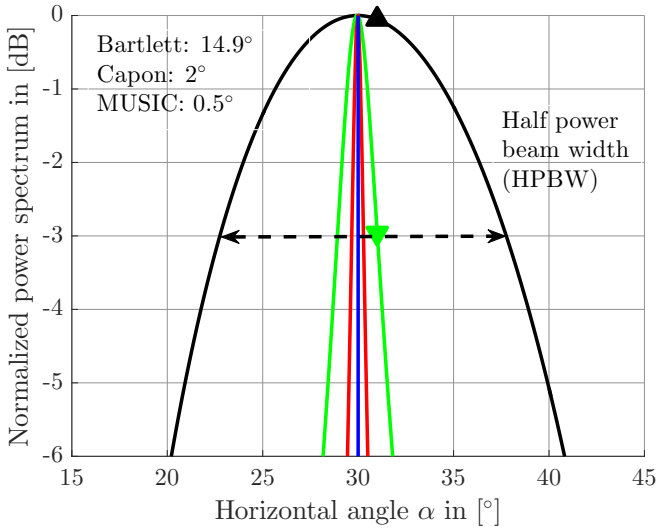
Figure 3.6.: 3D beam pattern of 8×8 UPA with a beamformer towards 30° in horizontal and 0° in vertical direction. The red arrow is the orientation vector of the array. These beam patterns are generated with the open source QuaDRiGa software [JRBT14, HHI17].



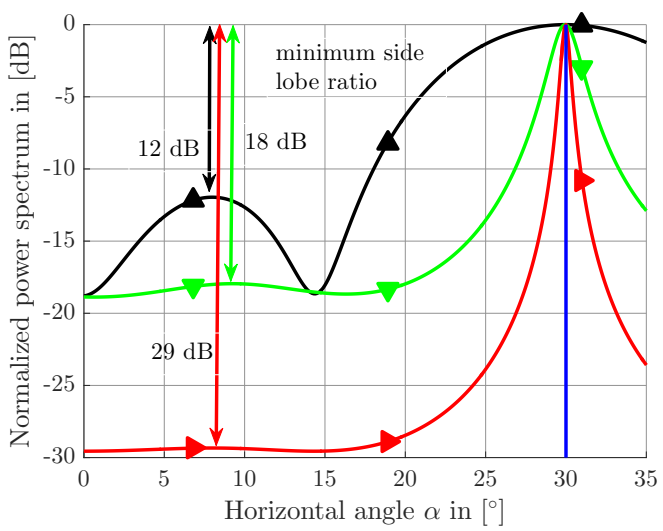
(a) Horizontal angular range $\boldsymbol{\alpha}_{(\mathrm{f})} = [-90\ 90]°$.

(b) Zoom to angular range $\boldsymbol{\alpha}_{(\mathrm{f})} = [15\ 45]°$.

Figure 3.7.: Power spectrum of search-based DoA techniques for a single path such that $L_{(\mathrm{MPC})} = 1$. Other parameters are set according to Table 3.2: Bartlett, Capon and MUSIC.



(a) half-power beam-width of detected peak according to Eq. (3.55).



(b) Minimum side lobe ratio according to Eq. (3.57).

Figure 3.8.: Illustration of performance metrics in power spectrum for search-based DoA techniques for a single path such that $L_{(\mathrm{MPC})} = 1$. Other parameters are set according to Table 3.2.

such as the "urban macro" assumed in this chapter, see Table 3.2, multi-path components are also present. Therefore, Fig. 3.9 shows the power spectrum and HPBW with multi-path components such that $L_{(\mathrm{MPC})} = 8$. In Fig. 3.9a it is observed that the minimum side lobe ratio of the MUSIC and Capon method is worse than that of the Bartlett method. Accordingly, also the HPBW of the MUSIC and Capon method shown in Fig. 3.9b are larger compared to the single path assumption in Fig. 3.8a. As explained in Section 3.1.1, the multiple paths can be interpreted as other sources. Therefore, assuming them as noise as in the optimizations of the Capon and MUSIC method is not optimal and causes significant performance degradation, even for a high K-factor of $K^{(\mathrm{R})} = 10\,\mathrm{dB}$.

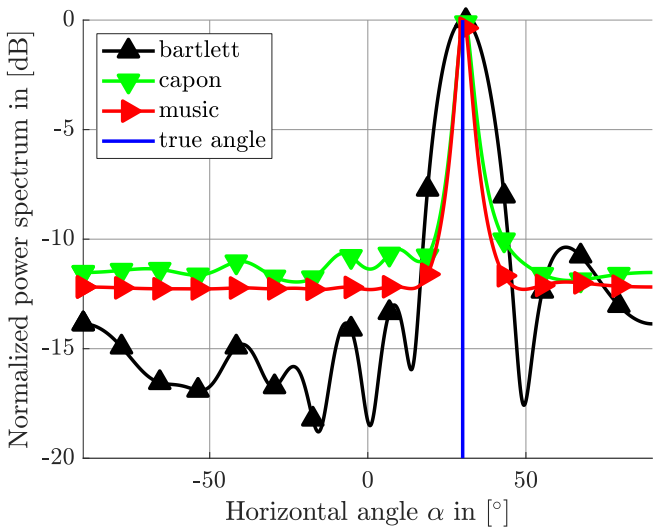
Next, the impact of the various parameters on the DoA estimation performance is conducted. According to Table 3.2 Monte Carlo simulations are performed over a statistic of 2500 trials. In addition to the additive white Gaussian noise (AWGN) (see Eq. (3.1)) also the position of the device is random independent identically distributed (i.i.d.) in the interval $[25, 35]^\circ \in \mathbb{R}$. In order to evaluate the performance of a Monte Carlo statistic the CDF is used. In Fig. 3.10a this is illustrated for increasing the number of columns $N_{(\alpha)}$ in the UPA, where the $(5, 50, 95)$ %-ile values are marked by blue circles on the curve of $N_{(\alpha)} = 8$. Fig. 3.10a already illustrates that the distribution of CDFs can overlap and thus make it difficult to distinguish. Therefore, typically the 50 %-ile (median) of the CDFs is given over the varied parameter. In this case the distribution of the estimation error can be represented by the 90 % variance or confidence interval of $[5, 95]$ %-ile values from the CDF, shown by the error bars in Fig. 3.10b. However, for the sake of presentation in this section only the median values are shown in most of the figures, which is sufficient to show the dependency on the different parameters. Whenever they are of relevance, the corresponding confidence intervals are shown or listed in tables in the annex. Note that according to Fig. 3.10b $N_{(\alpha)} = 8$ antennas are required in order to achieve the angular estimation error target of 0.29° with a probability of 95 %. The value of 0.29° is the lower threshold to achieve the 1 m positioning target at 200 m distance from the BS, see Fig. 3.4.

In Fig. 3.11 horizontal estimation error scales with

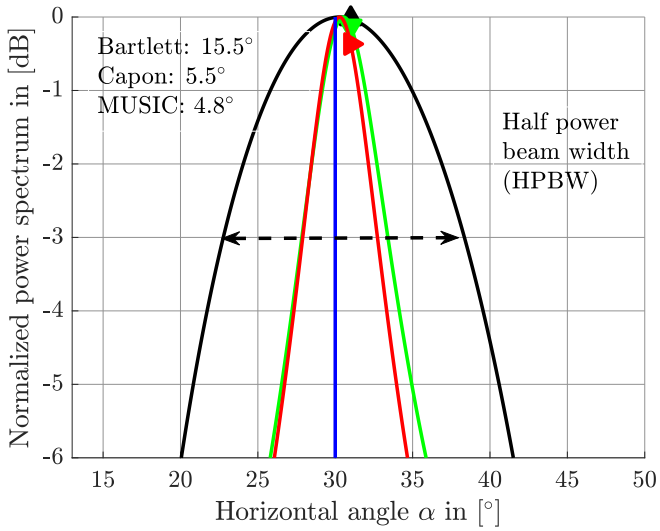
$$\left|\Delta_{(\alpha)}\right| \propto \frac{1}{N_{(\alpha)}} \tag{3.63}$$

over the number of horizontal antennas $N_{(\alpha)}$ and

$$\left|\Delta_{(\alpha)}\right| \propto \frac{1}{\tilde{\gamma}} \tag{3.64}$$

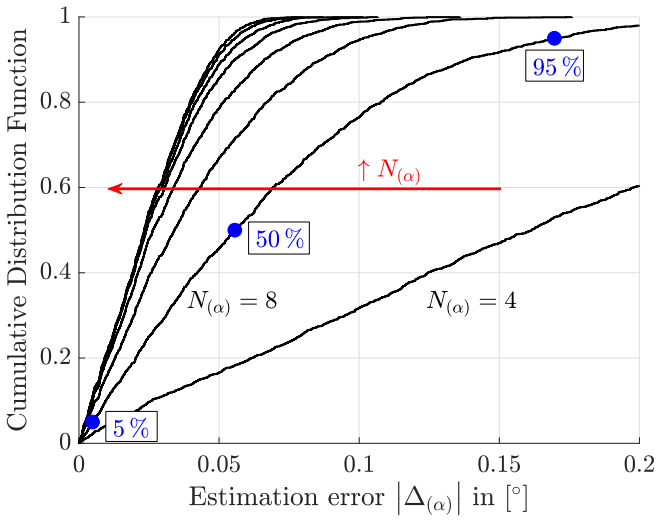


(a) Horizontal angular range $\boldsymbol{\alpha}_{(\mathrm{f})} = [-90\ 90]°$.

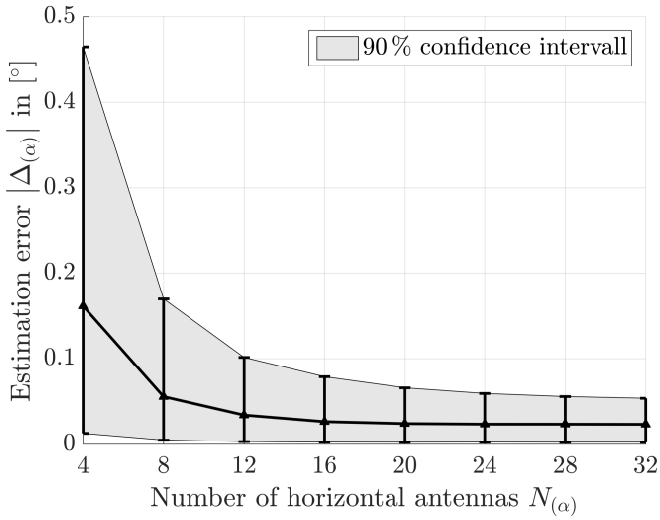


(b) half-power beam-width of detected peak according to Eq. (3.55).

Figure 3.9.: Power spectrum of search-based DoA techniques for multiple paths such that $L_{(\mathrm{MPC})} = 8$. Other parameters are set according to Table 3.2: Bartlett, Capon and MUSIC.

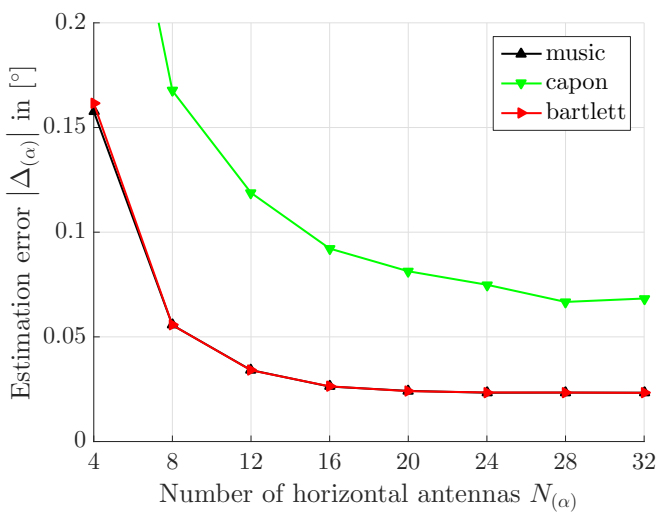


(a) CDF of Monte Carlo simulation over horizontal DoA estimation error according to Eq. (3.46).

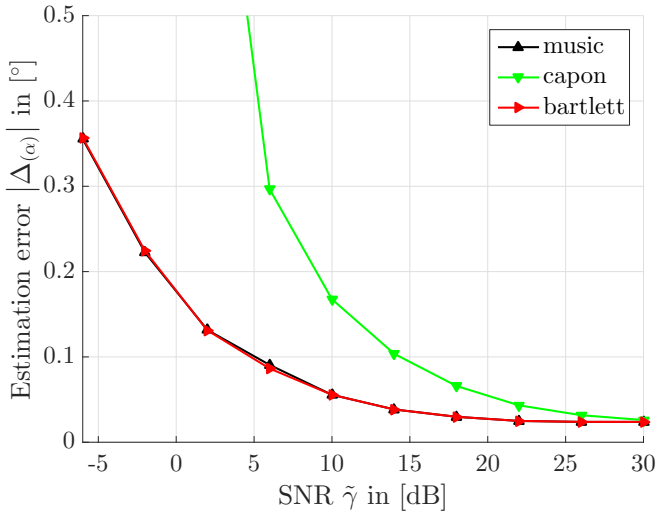


(b) 50 %-ile or median value of horizontal DoA estimation error with 90 % confidence interval.

Figure 3.10.: CDF of estimation error based on the Monte Carlo simulation statistic. The $(5, 50, 95)$ %-ile values are used for confidence interval and further evaluation. Simulation parameters are given in Table 3.2 with Bartlett DoA estimator and K-factor $\lim K^{(\mathrm{R})} \to \infty$.



(a) Variation over number of horizontal antenna elements $N_{(\alpha)}$ at the receive BS.



(b) Variation over SNR $\tilde{\gamma}$ according to Eq. (3.5).

Figure 3.11.: Median horizontal DoA estimation error $\left|\Delta_{(\alpha)}\right|$ over number of antenna elements and SNR. Other parameters are set according to Table 3.2.

over the SNR $\tilde{\gamma}$ until a saturation. A logarithmic scaling of the $y^{(\mathrm{C})}$-axis clarifies this on the left hand side of Fig. B.1a and Fig. B.2a. The saturation is caused by the quantization of the search space $q^{(\phi)}$. In contrast to this, the CRLB in Fig. B.1b and Fig. B.2b scales without saturation, because the quantization of the search range is not considered. Note that the figures use logarithmic scale and $y^{(\mathrm{C})} = \frac{1}{x^{(\mathrm{C})}}$ becomes

$$\log_{10} y^{(\mathrm{C})} = \log_{10} \frac{1}{x^{(\mathrm{C})}} = \log_{10} 1 - \log_{10} x^{(\mathrm{C})} = -\log_{10} x^{(\mathrm{C})} \tag{3.65}$$

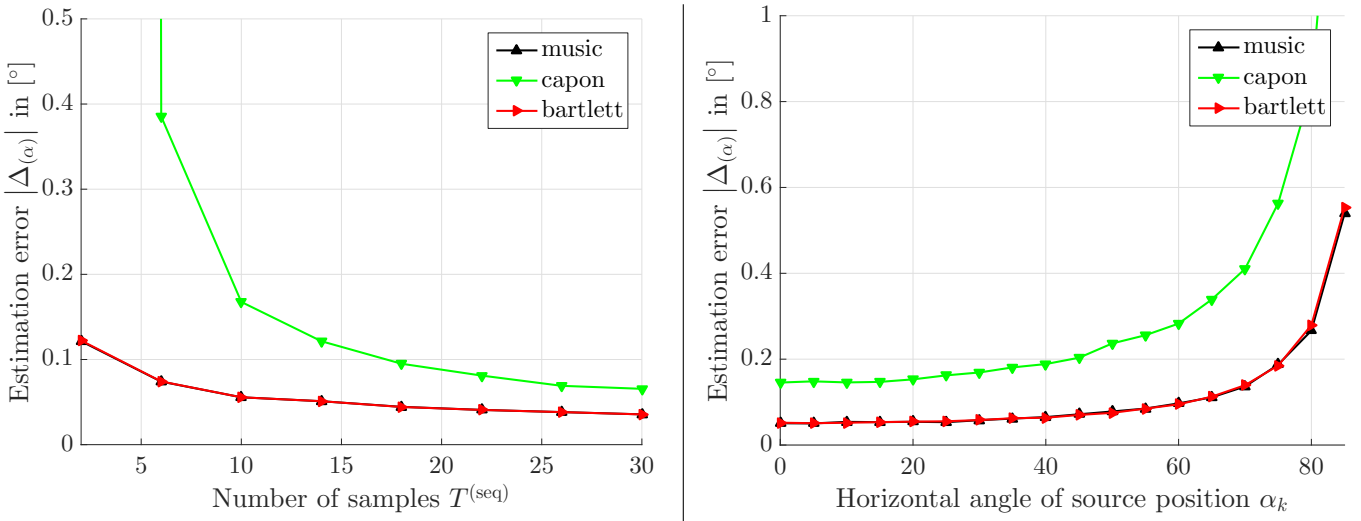
and hence result in a straight line. The MUSIC and Bartlett method achieve approximately the same performance because a single path is assumed and the estimated DoA is obtained from the maximum in the power spectrum. Therefore, the larger HPBW and side-lobes of the Bartlett method compared to the MUSIC method have no impact when the maximum of the peaks of both methods are at the same position. The significant larger estimation error of the Capon method compared to MUSIC and Bartlett methods is caused by the inverse computation of the receive covariance matrix in (3.30). This matrix inversion enhances the noise included in the covariance matrix. This is confirmed by observations in Fig. 3.11b and Fig. 3.12a, where a higher SNR and sequence length closes the estimation error gap between the Capon method and the other methods, respectively. However, the Capon method shows the same scaling of $\left|\Delta_{(\alpha)}\right| \propto \frac{1}{N_{(\alpha)}}$ as the other two methods.

In Fig. 3.12a the variation over the sequence length $N^{(\mathrm{seq})}$ is given. According to Eq. (3.21) the sequence length $N^{(\mathrm{seq})}$ improves the receive covariance matrix by averaging out noise by taking into account more samples of the time-invariant channel. The scaling of the DoA estimation error with $N^{(\mathrm{seq})}$ is

$$\left|\Delta_{(\alpha)}\right| \propto \frac{1}{N^{(\mathrm{seq})}} \tag{3.66}$$

until a saturation for a large sequence length is reached. The reason for the saturation is again the quantization of the search space. The corresponding CRLB is given in the appendix in Fig. B.3. In contrast to the previous parameters $[\tilde{\gamma}, N_{(\alpha)}, N^{(\mathrm{seq})}]$ which are part of $P_{(\mathrm{CRB})}$ in Eq. (3.40), the source position is part of the geometry dependent term $\mathbf{G}_{(\mathrm{CRB})}(\phi)$ in Eq. (3.40). The exponential decrease in estimation performance for $\alpha_k \to 90°$ is due to the Jacobian matrix $\mathbf{J}_{(\mathrm{CRB})}(\mathbf{u}_{(\phi)}(\alpha, \beta))$ from Eq. (3.43) by

$$\lim_{\alpha \to 90°} \mathbf{J}_{(\mathrm{CRB})}(\mathbf{u}_{(\phi)}(\alpha, \beta)) = \lim_{\alpha \to 90°} \begin{bmatrix} -\cos(\beta)\sin(\alpha) & -\sin(\beta)\cos(\alpha) \\ \cos(\beta)\cos(\alpha) & -\sin(\beta)\sin(\alpha) \\ 0 & \cos(\beta) \end{bmatrix} = \begin{bmatrix} 0 & -\sin(\beta) \\ \cos(\beta) & 0 \\ 0 & \cos(\beta) \end{bmatrix}. \tag{3.67}$$



(a) Variation over sequence length $N^{(\mathrm{seq})}$.

(b) Variation of source position (angle).

Figure 3.12.: Median horizontal DoA estimation error $\left|\Delta_{(\alpha)}\right|$ over sequence length and horizontal angle of the source. Other parameters are set according to Table 3.2.

The DoA estimation performance is shown in Fig. 3.12b and the corresponding CRLB in Fig. B.4.

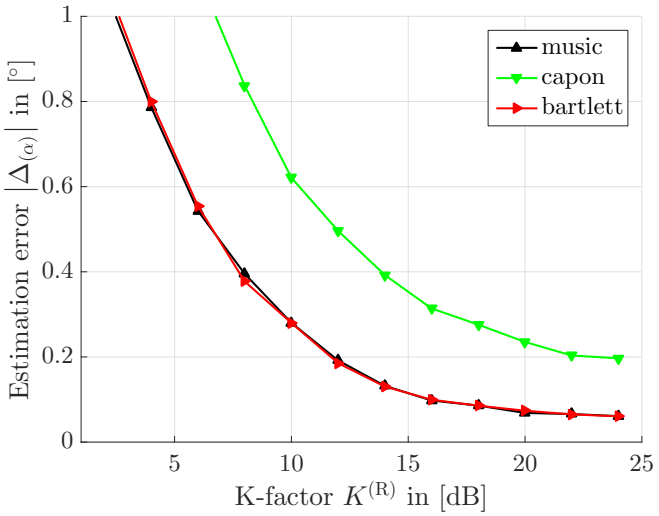
Not taken into account by the CRLB is the Rician K-factor parameter $K^{(\mathrm{R})}$ defined in Eq. (3.13). Until now only the LoS path was considered resulting in $K^{(\mathrm{R})} = \infty$. In Fig. 3.13a the estimation error $\Delta_{(\alpha)}$ scales also with

$$\left|\Delta_{(\alpha)}\right| \propto \frac{1}{K^{(\mathrm{R})}} \tag{3.68}$$

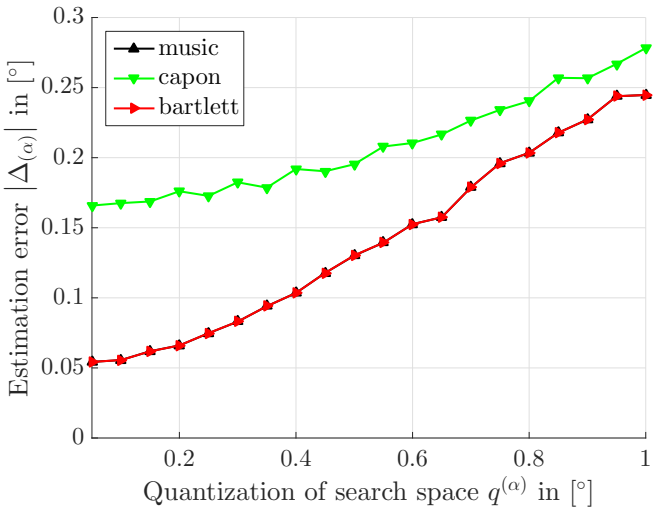
clarified by the straight line in logarithmic scale in Fig. B.5. Similar to $\tilde{\gamma}$, $N^{(\mathrm{seq})}$ and $N_{(\alpha)}$ a saturation at high $K^{(\mathrm{R})}$ is observed caused by the quantization of the search space $q^{(\phi)}$. According to Eq. (3.50) and Eq. (3.51) there are estimation errors caused by the quantization of the search space. The impact of the search space quantization is also not considered in the CRLB and a linear correlation is observed in Fig. 3.13b. At a certain quantization the performance is limited by noise and the array geometry, so saturation is achieved at approximately $q^{(\phi)} = 0.1°$ resulting in $\left|\Delta_{(\alpha)}\right| < 0.055° \approx q^{(\phi)}/2$ which is the quantization noise according to Eq. (3.50).

In the default simulation assumptions an omni-directional idealized beam pattern is assumed for the antenna elements, resulting in the ambiguity as shown in Fig. 3.6. However, in practical antennas systems such UPAs are designed and setup to cover a certain direction. With typical patch antennas [Bie07], e.g. $\alpha_{\mathrm{HPBW}} = \beta_{\mathrm{HPBW}} = 65°$ HPBW and a front-back ratio attenuation of 30 dB as shown in Fig. 3.14a, the resulting pattern no longer contains an ambiguity caused by the symmetry of the uniform linear array (ULA) or UPA. The resulting pattern for the same deployment and beamformer in direction $\alpha = 30°, \beta = 0°$ as in Fig. 3.6 is visualized in Fig. 3.14b. However, due to the same array geometry the form factor is also the same and there is no performance gain for DoA estimation. This ambiguity can also be avoided by using circular arrays, but with the same number of antennas, the resolution in the direction of interest is lower compared to planar arrays [GRK+16].

Finally, the dependency on the distance is evaluated by the position error. Changing the distance is changing the receive power at the BS and therefore the SNR. The impact of this is already shown in Fig. 3.11b and not in the focus, therefore $g^{(\mathrm{norm})} = 1$ and the SNR is kept constant $\tilde{\gamma} = 10\,\mathrm{dB}$ according to Table 3.2. The reason for this assumptions is to isolate the impact of the distance from the SNR. In Fig. 3.15a the median position error scales linearly over the distance, with a steeper gradient for the Capon estimator than Bartlett and MUSIC. These numerical results, confirm the analytical findings from Fig. 3.4, that a constant DoA estimation error, see Fig. 3.15a, results in larger position errors when increasing the distance between

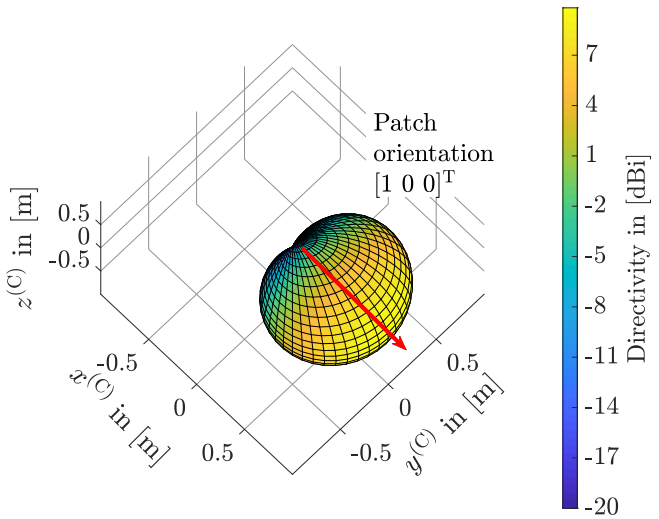


(a) Variation over K-factor $K^{(\mathrm{R})}$.

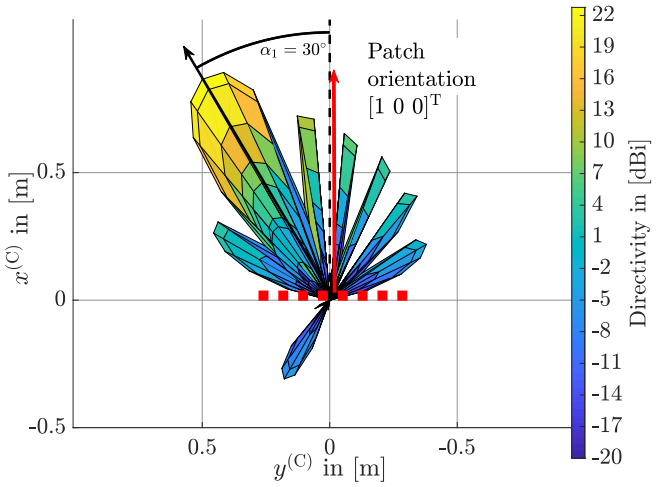


(b) Variation of search space quantization by $\mathbf{q}^{(\alpha)}$ with K-factor $\lim K^{(\mathrm{R})} \to \infty$.

Figure 3.13.: Median horizontal DoA estimation error $\left|\Delta_{(\alpha)}\right|$ over K-factor and search space quantization. Other parameters are set according to Table 3.2.

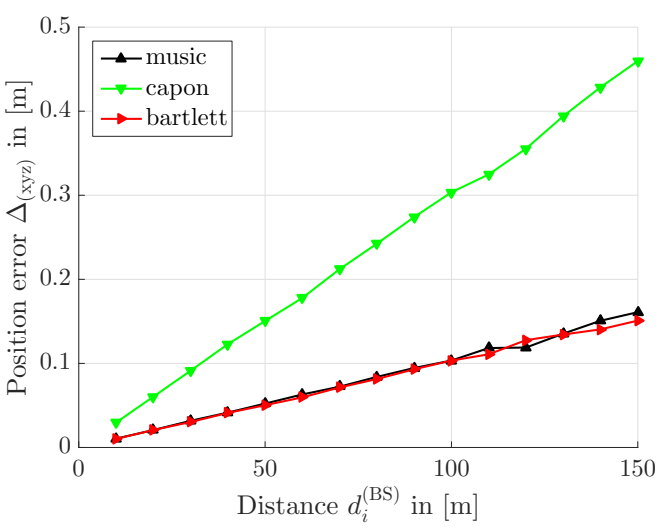


(a) 3D beam pattern of a single patch element with 65° HPBW in azimuth and elevation.

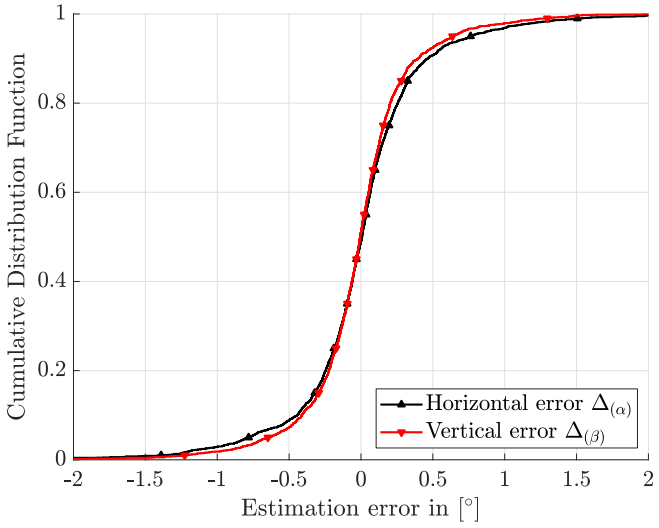


(b) 3D beam pattern of an 8×8 UPA with patch elements.

Figure 3.14.: 3D beam pattern of a single patch element and array of patch elements solving ambiguity of DoA estimation in the $x^{(\mathrm{C})}$-$y^{(\mathrm{C})}$ plane. These beam patterns are generated with the open source QuaDRiGa software [JRBT14, HHI17].



(a) Median position error over the distance between source and BS. Other parameters are set according to Table 3.2.



(b) Distribution of horizontal and vertical DoA estimation errors of MUSIC method.

Figure 3.15.: Performance of DoA estimation methods with parameters from Table 3.2.

the mobile device and BS.

Remark. *In the single source scenario, the HPBW of the peak has approximately no impact. Therefore, Bartlett and MUSIC yield approximately the same performance. Due to the inverse computation of the receive covariance matrix in the Capon beamformer, see Eq.* (3.29)*, noise is enhanced. Therefore, the Capon method shows the same trends but always a larger estimation error than Capon and MUSIC. Therefore, in the following, the focus is on Bartlett and MUSIC.*

The error distribution of the estimated horizontal and vertical error for parameters in Table 3.2 is given in Fig. 3.15b and symmetrical to zero. However, the distribution is not normal due to heavy tails. More details on the approximation by a parametrized probability distribution is shown in Section 3.3.2.

3.2.2. Complexity Reduction

In the previous section, the impact of various parameters was investigated by considering only the horizontal angle component of the DoA. However, the 3D DoA consists of an horizontal and vertical component which have to be estimated jointly, due to coupling according to Eq. (3.40) and Eq. (3.43). In other words, estimating $\hat{\alpha}$ and $\hat{\beta}$ independently results in a systematic unknown error additional to the random error. Therefore, the power spectrum in Eq. (3.23) has to be computed for each DoA in the search range $\boldsymbol{\alpha}^{(\mathrm{f})}$ and $\boldsymbol{\beta}^{(\mathrm{f})}$. With numbers from the default parameter setting in Table 3.2, a search space of $\boldsymbol{\alpha}^{(\mathrm{f})} = \boldsymbol{\beta}^{(\mathrm{f})} = [-60\ 60]^\circ$ and a quantization of the search space with $q^{(\phi)} = 0.1^\circ$ results in $N^{(\phi)} = 1\,442\,401$ computations of the power spectrum according to Eq. (3.54). The scanning range is selected based on insights obtained from Fig. 3.12b, as the typical working point of a UPA. A larger scanning range results in large DoA estimation errors independent of any complexity reduction due to limited resolution of the array geometry, see Fig. 3.12b. Clearly, more than 1.4 million power spectrum computations is a huge obstacle to enable search-based techniques for high accurate DoA estimation with large antenna arrays. The power spectrum computation includes a matrix multiplication for Bartlett, additionally an Eigenvalue decomposition for MUSIC, and a matrix inverse for Capon, This motivates the current section on complexity reduction for search-based DoA estimation techniques without performance loss.

Remark. *The number of computations of the power spectrum, $N^{(\phi)}$, is henceforth considered as a performance measure of the complexity which does not take the complexity of the computation of the power spectrum per DoA estimation into account.*

The general idea is to use different quantization levels $q^{(\phi)}$ in multiple steps, similar to adaptive quantization, e.g. used in video compression [WSBL03]. This approach is called adaptive search space quantization (ASSQ) and was published in [KTH+16b]. A toy example described in Fig. 3.16 compares the conventional approach used in the previous sections on the left-hand side with the proposed ASSQ approach on the right-hand side. In this toy example a step size $S = 2$ is used for the ASSQ approach. What the step size S exactly means is clarified in the following. Assuming a target resolution of 2° for an angular range of 80° the conventional scheme results in $N^{(\phi)} = 41$ power spectrum computations. Using a larger quantization in the first step $s = 1$, denoted as $q_s^{(\phi)} = q_1^{(\phi)} = 8°$, results in $N_s^{(\phi)} = N_1^{(\phi)} = 11$. In the second step $s = 2$, based on the maximum found in step one according to Eq. (3.39), a smaller search range $\boldsymbol{\alpha}_2^{(\mathrm{f})} = [-6, 6]$ is selected and $q_s^{(\phi)} = q_2^{(\phi)} = 2°$, that is the same quantization as on the left-hand side. The second step requires a complexity of $N_2^{(\phi)} = 7$. Summing the complexity over S according to

$$N^{(\phi)} = \sum_{s=1}^{S} N_s^{(\phi)} \tag{3.69}$$

the total complexity in the example of Fig. 3.16 is $N^{(\phi)} = 18$. This is a complexity reduction compared to $S = 1$ by more than 50 % while providing the same target quantization. The ASSQ scheme inherently assumes that the HPBW in the power spectrum is larg enough to detect the peak. This is a valid assumption for the simulation parameters assumed in the previous section if multi-path components are present, see Fig. 3.9 where for the MUSIC method a HPBW of $\approx 5°$ is observed. Note that without limitation to generality the search range $\boldsymbol{\alpha}_s^{(\mathrm{f})}$ is relative to the maximum found in step $s - 1$, e.g. the angle determined in $s = 1$ is $\hat{\alpha}_1 = -25°$ and $\boldsymbol{\alpha}_2^{(\mathrm{f})} = [-5\ 5]°$, then the absolute angle range of $[-30, -20]°$ is searched in the second step.

The introduction of multiple steps by ASSQ requires adaptation on notations that are summarized in Table 3.3. In contrast to the previous section the quantization in horizontal and vertical dimension can be different at step s, but with $q_S^{(\alpha)} = q_S^{(\beta)}$ it can be ensured that the quantization error is the same in both dimensions. In other notations the subscript $s \in \mathcal{S}$ indicates the current step, where $\mathcal{S} = \{1, \ldots, S\}$.

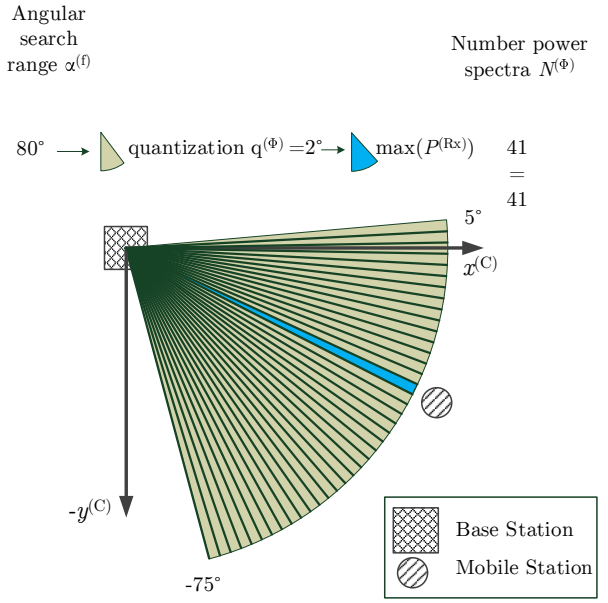
For performance evaluation, the position error is selected containing azimuth and elevation estimation errors. While all mobile devices have the same distance from the BS (one a sphere) it follows that an increase in the position error must be caused by an increase in DoA estimation errors. Larger DoA estimation errors must be then caused by the described ASSQ. As mentioned above, the objective is to minimize the complexity $N^{(\phi)}$ without degradation of the position error. To obtain $N^{(\phi)}$ Eq. (3.54) is put into Eq. (3.69) with notation from Table 3.3 yielding

$$N^{(\phi)} = \sum_{s=1}^{S} \left(\left\lceil \frac{[\boldsymbol{\alpha}_s^{(\mathrm{f})}]_2 - [\boldsymbol{\alpha}_s^{(\mathrm{f})}]_1}{q_s^{(\alpha)}} \right\rceil + f^{(\mathrm{s})}(\boldsymbol{\alpha}_s^{(\mathrm{f})}, q_s^{(\alpha)}) \right) \left(\left\lceil \frac{[\boldsymbol{\beta}_s^{(\mathrm{f})}]_2 - [\boldsymbol{\beta}_2^{(\mathrm{f})}]_1}{q_s^{(\beta)}} \right\rceil + f^{(\mathrm{s})}(\boldsymbol{\beta}_s^{(\mathrm{f})}, q_s^{(\beta)}) \right). \tag{3.70}$$

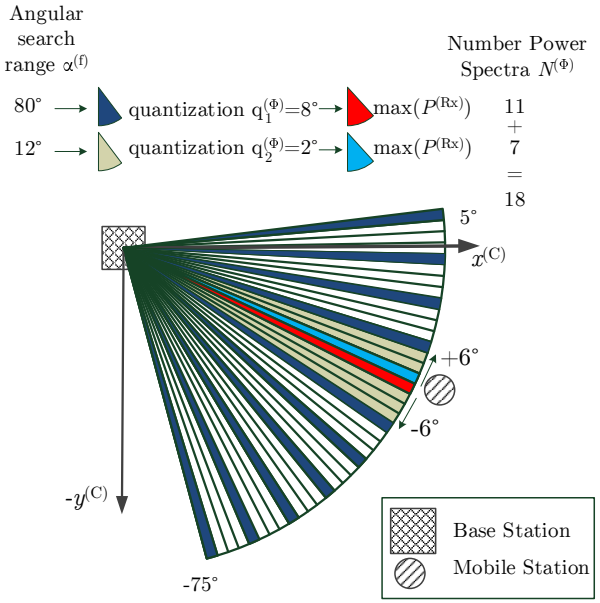
The optimization problem is then given as

$$\underset{S, \mathbf{q}^{(\alpha)}, \mathbf{q}^{(\beta)}, \mathbf{A}_{(\mathrm{f})}, \mathbf{B}_{(\mathrm{f})}}{\text{minimize}} \quad N^{(\phi)} \left(S, \mathbf{q}^{(\alpha)}, \mathbf{q}^{(\beta)}, \mathbf{A}_{(\mathrm{f})}, \mathbf{B}_{(\mathrm{f})} \right), \tag{3.71}$$

where $\mathbf{A}_{(\mathrm{f})}$, $\mathbf{B}_{(\mathrm{f})}$, $\mathbf{q}^{(\alpha)}$, and $\mathbf{q}^{(\beta)}$ are the aggregated search spaces and quantization per step s in matrix and vector notation, respectively, according to Table 3.3. For the minimization of $N^{(\phi)}$ constraints for $S, \mathbf{q}^{(\alpha)}, \mathbf{q}^{(\beta)}, \mathbf{A}_{(\mathrm{f})}, \mathbf{B}_{(\mathrm{f})}$ are required. These constraints are obtained in the following paragraphs.



(a) $S = 1$ steps with angular search range $\mathbf{A}_{(\mathrm{f})} = \boldsymbol{\alpha}_1^{(\mathrm{f})} = [-5° \quad 75°]$ and quantization $q^{(\phi)} = 2°$.



(b) $S = 2$ steps with angular search range per step $\mathbf{A}_{(\mathrm{f})} = \begin{bmatrix} 5° & -75° \\ -6° & 6° \end{bmatrix}$ and quantization per step $\mathbf{q}^{(\alpha)} = [8° \; 2°]$.

Figure 3.16.: Illustration of adaptive search space quantization algorithm with one step on the left hand side and $S = 2$ steps on the right hand side. $S = 2$ requires less power spectrum computations denoted by $N^{(\phi)}$.

Table 3.3.: Summary of notations introduced due to adaptive search space quantization for search-based DoA estimation.

Parameter Description	**Symbol**
Number of ASSQ steps	$S \in \mathbb{N}^+$
Set of ASSQ steps	$\mathcal{S} = \{1, \ldots, S\}$
current ASSQ step	$s \in \mathcal{S}$
Horizontal search space in step s	$\boldsymbol{\alpha}_s^{(\mathrm{f})}$
Vertical search space in step s	$\boldsymbol{\beta}_s^{(\mathrm{f})}$
Horizontal search space of steps $\mathcal{S}$	$\mathbf{A}_{(\mathrm{f})} = \left[\left(\boldsymbol{\alpha}_1^{(\mathrm{f})}\right)^{\mathrm{T}} \; \ldots \; \left(\boldsymbol{\alpha}_S^{(\mathrm{f})}\right)^{\mathrm{T}} \right]^{\mathrm{T}}$
Vertical search space of steps $\mathcal{S}$	$\mathbf{B}_{(\mathrm{f})} = \left[\left(\boldsymbol{\beta}_1^{(\mathrm{f})}\right)^{\mathrm{T}} \; \ldots \; \left(\boldsymbol{\beta}_S^{(\mathrm{f})}\right)^{\mathrm{T}} \right]^{\mathrm{T}}$
Horizontal quantization in step s	$q_s^{(\alpha)}$
Vertical quantization in step s	$q_s^{(\beta)}$
Horizontal quantization of steps $\mathcal{S}$	$\mathbf{q}^{(\alpha)} = [q_1^{(\alpha)} \; \ldots \; q_S^{(\alpha)}]^{\mathrm{T}}$
Vertical quantization of steps $\mathcal{S}$	$\mathbf{q}^{(\beta)} = [q_1^{(\beta)} \; \ldots \; q_S^{(\beta)}]^{\mathrm{T}}$
Complexity in step s	$N_s^{(\phi)}$

From Eq. (3.70) and Fig. 3.16 constraints for the quantization of the search space are obtained:

$$\begin{aligned}\left|\left[\boldsymbol{\alpha}_s^{(\mathrm{f})}\right]_2 - \left[\boldsymbol{\alpha}_s^{(\mathrm{f})}\right]_1\right| &< \left|\left[\boldsymbol{\alpha}_{s-1}^{(\mathrm{f})}\right]_2 - \left[\boldsymbol{\alpha}_{s-1}^{(\mathrm{f})}\right]_1\right| \\ \left|\left[\boldsymbol{\beta}_s^{(\mathrm{f})}\right]_2 - \left[\boldsymbol{\beta}_s^{(\mathrm{f})}\right]_1\right| &< \left|\left[\boldsymbol{\beta}_{s-1}^{(\mathrm{f})}\right]_2 - \left[\boldsymbol{\beta}_{s-1}^{(\mathrm{f})}\right]_1\right|,\end{aligned} \tag{3.72}$$

which means that the search space in step s has to be smaller than the search space in the previous step $s-1$. The same is true for the search space quantization such that

$$\begin{aligned}q_s^{(\alpha)} &< q_{s-1}^{(\alpha)} \\ q_s^{(\beta)} &< q_{s-1}^{(\beta)},\end{aligned} \tag{3.73}$$

which means that the quantization in step s has to be smaller than in step $s-1$. In the toy example of Fig. 3.16 the search space in step 2 is $\left|\boldsymbol{\alpha}_2^{(\mathrm{f})}\right| = q_1^{(\alpha)}$. However, $N^{(\phi)}$ in Eq. (3.70) is minimized for a given $q_{s-1}^{(\alpha)}$ if $\left|\boldsymbol{\alpha}_s^{(\mathrm{f})}\right|$ is minimized. Therefore, the question is: How to select $\left|\boldsymbol{\alpha}_s^{(\mathrm{f})}\right|$ without performance loss in position accuracy?

The search space of step s can be limited by

$$q_{s-1}^{(\alpha)} \leq \left|\left[\boldsymbol{\alpha}_s^{(\mathrm{f})}\right]_2 - \left[\boldsymbol{\alpha}_s^{(\mathrm{f})}\right]_1\right| \leq 2q_{s-1}^{(\alpha)} \tag{3.74}$$

because a larger search space results in overlapping areas with the quantization in step $s-1$ and a lower search space result in gaps also called white spaces of the search space. Fig. 3.17 provides an example with $S=2$. On the left hand side searched angles in step $s=1$ are separated by angular distance $q_1^{(\phi)}$ and the detected maximum is depicted by a red arrow for case a) and a blue arrow for case b). The search space of step two $|\boldsymbol{\alpha}_2^{(\mathrm{f})}|$ is shown on the right hand side and is depicted by striped areas with the same color as the detected maximum in step 1 for case a) and b) accordingly. It can be seen in Fig. 3.17 that a search space of $\left|\left[\boldsymbol{\alpha}_s^{(\mathrm{f})}\right]_2 - \left[\boldsymbol{\alpha}_s^{(\mathrm{f})}\right]_1\right| = q_1^{(\phi)}$ results in adjacent search spaces. This means, that even if the true position is in the middle between searched points in $s=1$ it would be detected by the search in the second step because there are no white spaces. Following this argumentation, the search space with the lowest complexity in step s for a given quantization q_{s-1} in the previous step $s-1$ without performance degradation is given by equality on the left-hand side of Eq. (3.74) such that

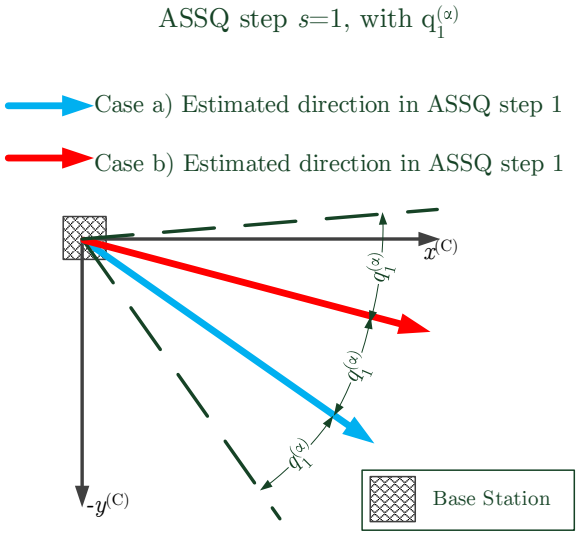
$$q_{s-1}^{(\alpha)} = \left|\left[\boldsymbol{\alpha}_s^{(\mathrm{f})}\right]_2 - \left[\boldsymbol{\alpha}_s^{(\mathrm{f})}\right]_1\right|, \tag{3.75}$$

which is an additional constraint for Eq. (3.71). In other words, the search space in the current step s should be equal to the quantization in the previous step $s-1$. The same conclusion holds for the vertical domain such that

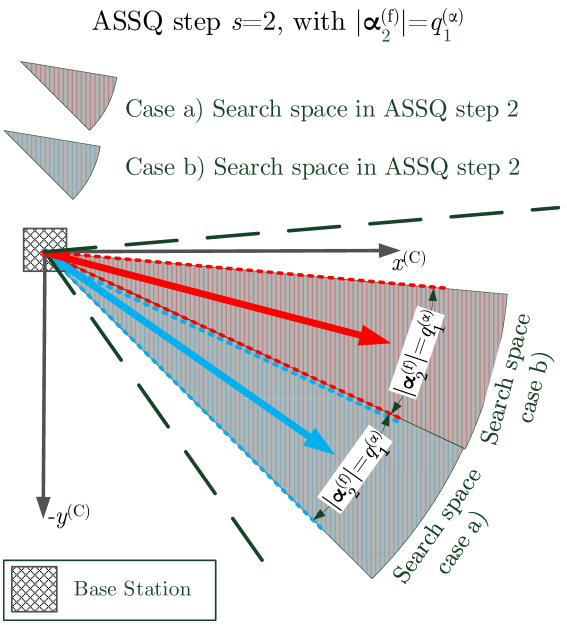
$$q_{s-1}^{(\beta)} = \left|\left[\boldsymbol{\beta}_s^{(\mathrm{f})}\right]_2 - \left[\boldsymbol{\beta}_s^{(\mathrm{f})}\right]_1\right|. \tag{3.76}$$

However, even with the constraints Eq. (3.75) and Eq. (3.76) the optimization in Eq. (3.71) is intractable, due to the degrees of freedom by the parameters, S the number of ASSQ steps, and $\mathbf{q}^{(\alpha)}, \mathbf{q}^{(\beta)}$ the quantization of the horizontal and vertical search spaces for all steps. Furthermore, any analytical solution has to take into account the rounding up operation $\lceil\cdot\rceil$ in Eq. (3.70) which requires a distinction of cases. Therefore, the minimization of $N^{(\phi)}$ was done by exhaustive search for a given S, constraint parameters $\boldsymbol{\alpha}_1^{(\mathrm{f})}, \boldsymbol{\beta}_1^{(\mathrm{f})}, q_S^{(\alpha)}, q_S^{(\beta)}$ and the conditions from Eq. (3.75), Eq. (3.76). This means that out of a set of $q_{\alpha,\tilde{s}}$ and $q_{\beta,\tilde{s}}$ values, the parameters yielding the minimum $N^{(\phi)}$ are selected. From Fig. 3.17 it becomes clear that further constraints to limit this exhaustive search can be included by

$$\begin{aligned}q_s^{(\alpha)} &< q_{s-1}^{(\alpha)}, \\ q_s^{(\beta)} &< q_{s-1}^{(\beta)},\end{aligned} \tag{3.77}$$



(a) ASSQ step $s = 1$ with quantization $q_1^{(\alpha)}$.



(b) ASSQ step $s = 2$ with search space $\left|\alpha_1^{(f)}\right|$ equal to the quantization $q_1^{(\alpha)}$ in the previous step.

Figure 3.17.: ASSQ constraints on search space based on quantization in previous step. Selection of search space $\boldsymbol{\alpha}_2^{(f)}$ in step $s = 2$ based on quantization $q_1^{(\alpha)}$ in step $s = 1$.

meaning that quantization of step s has to be smaller than in the previous step $s - 1$. This limits the exhaustive search significantly.

An example for the exhaustive search for $S = 2$ is given in Fig. 3.18 by the surface plot of $N^{(\phi)}$, over $q_1^{(\alpha)} \in \{2, 2.1, \ldots, 5.9, 6\}^\circ$ and $q_1^{(\beta)} \in \{2, 2.1, \ldots, 5.9, 6\}^\circ$. There are two minima in this non-convex surface with $N^{(\phi)} = 2418$, which is a complexity reduction of ≈ 1000 times compared to $S = 1$ with the same search space quantization. There are two minima because of the symmetrical setup such that $\boldsymbol{\alpha}_1^{(f)} = \boldsymbol{\beta}_1^{(f)}$ and $q_S^{(\alpha)} = q_S^{(\beta)}$. Increasing the limits of $q_1^{(\alpha)}$ or $q_1^{(\beta)}$ as either a small $q_1^{(\alpha)}$ and large $q_1^{(\beta)}$ or vice versa results in a larger $N_{(\alpha)}$ or $N_{(\beta)}$, respectively. In Fig. 3.18 the search for $q_1^{(\alpha)}$ and $q_1^{(\beta)}$ is shown for a resolution of 0.01° and the complexity is reduced to $N^{(\phi)} = 2418$ for the following two parameter settings:

- $\mathbf{q}^{(\alpha)} = [3.88\ 0.1]^\circ$, $\mathbf{q}^{(\beta)} = [3.08\ 0.1]^\circ$
- $\mathbf{q}^{(\alpha)} = [3.08\ 0.1]^\circ$, $\mathbf{q}^{(\beta)} = [3.88\ 0.1]^\circ$

A further increase of the resolution doesn't yield a further minimization of $N^{(\phi)}$. However, the minimization of computational complexity should be achieved optimally without loss in DoA estimation performance. Therefore, using quantization steps that are non-integer multiples, e.g. $f^{(s)} = 1$ in Eq. (3.53), results in white spaces, e.g. areas that are not covered by the search. This effect is emphasized by the following example. Already indicated in [KTH+16b], increasing the ASSQ size S results in large quantization values in the first step $s = 1$. Focusing on the horizontal domain and considering a quantization $q_1^{(\alpha)} = 34.1^\circ$ results in the search of the following horizontal angles $(-60, -35.9, -11.8, 12.3, 36.4)^\circ$ and there is a white space of 23.6°. Moreover, a NLoS path from another direction would be detected instead of the LoS path as the peak, which results in large estimation errors. The same effect occurs at each ASSQ step and it becomes clear that larger quantization values result in larger white space and consequently large estimation errors. The outcome of this thought experiment is the additional constraint that the quantization $q_s^{(\alpha)}$ and $q_s^{(\beta)}$ of each ASSQ step s are limited to integer multiples of the total search range such that $f^{(s)} = 0$ in Eq. (3.53).

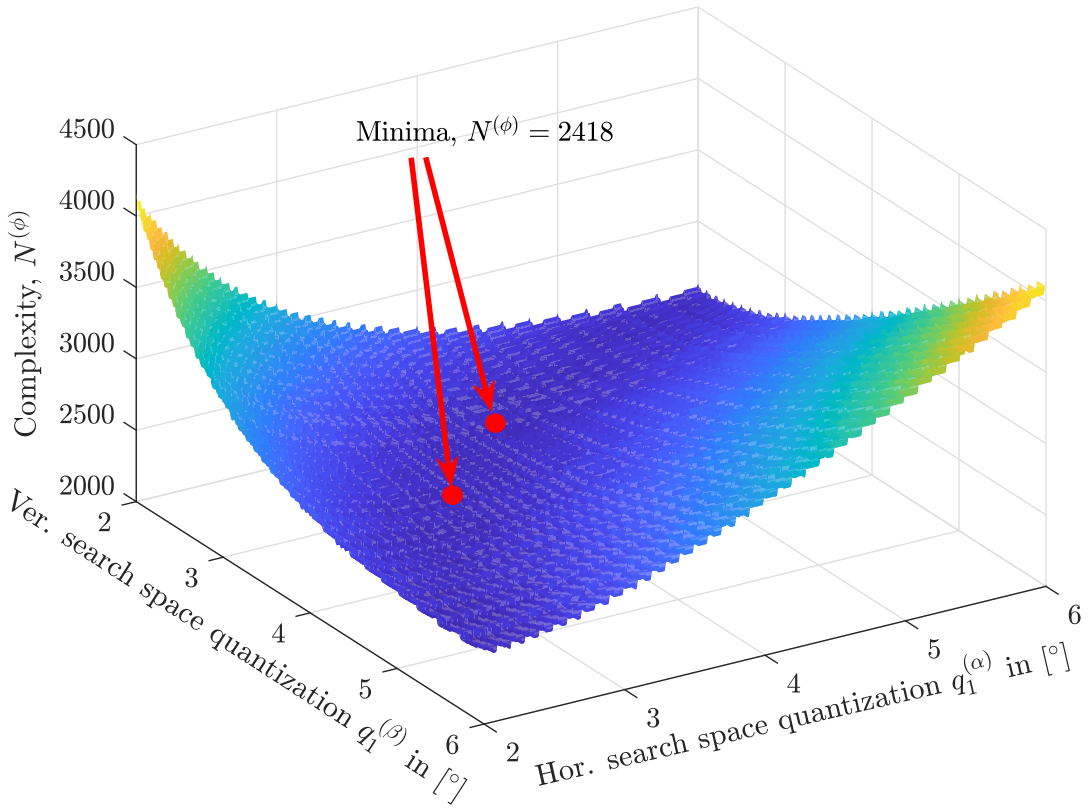


Figure 3.18.: Complexity $N^{(\phi)}$ of search-based DoA estimation with ASSQ over horizontal and vertical quantization in step $s = 1$.

Finally, the ASSQ complexity reduction is evaluated by the median position error for the DoA estimation methods Bartlett and MUSIC. The baseline performance is assumed for $S = 1$ and the parameters are given in Table 3.4. In contrast to the previous Section 3.2.2 the K-factor is set to $K^{(\mathrm{R})}$ =10 dB and users are distributed in the horizontal and vertical domain in the interval $[-60, 60]°$. The search space is accordingly set to $\boldsymbol{\alpha}^{(\mathrm{f})} = \boldsymbol{\beta}^{(\mathrm{f})} = [-60\ 60]°$.

The complexity reduction up to $S = 5$ is given in Fig. 3.20, the median position error on the left-hand side in Fig. 3.20a, and the complexity in Fig. 3.20b on the right-hand side. The corresponding ASSQ parameters obtained by exhaustive search follow the principles explained above are summarized in Table 3.5. In Fig. 3.20a the $y^{(\mathrm{C})}$-axis is in logarithmic scale and decreased by factor ≈ 17 from $S = 2$ to $S = 5$. A further complexity reduction for $S > 5$ is omitted due to the saturation in decreasing $N^{(\phi)}$ and an increase of the position error even with the constraints in Eq. (3.77) and Eq. (3.76). However, the position error is approximately constant from $S = 1$ to $S = 4$, $\Delta_{(\mathrm{xyz})} \approx 0.2\,\mathrm{m}$, but increases to $\Delta_{(\mathrm{xyz})} \approx 1.8\,\mathrm{m}$ for $S = 5$. Both DoA estimation methods, Bartlett and MUSIC, show approximately the same performance which is explained in the following.

Note that according to Table 3.4 for the complexity reduction $L_{(\mathrm{MPC})} = 8$ multi-path components are considered in order to have realistic propagation conditions. While the Bartlett method has a larger HPBW compared to the MUSIC method, 15° versus 5° see Fig. 3.9, the beam-width of the main-lobe which is is useful for ASSQ is approximately the same. What is meant by useful beam-width of the main-lobe? In Fig. 3.19 the useful beam-width of the main-lobe is defined by the peaks of the side-lobes projected on the main-lobe. In other words, whenever a power-spectrum is computed within this effective beam-width it yields to the correct maximum in the first step of the ASSQ algorithm. It turns out in Fig. 3.19 that the MUSIC and Bartlett method have the same effective beam-width, which explains why they yield approximately the same performance when combined with the ASSQ complexity reduction. The useful beam-width of 28.5°

Table 3.4.: Simulation parameters for complexity reduction. Other parameters are the same as in Table 3.2.

Parameter	**Value**
Simulation type	Monte Carlo 2500 realizations
Channel model	QuaDRiGa version 1.4 [JRBT14]
Scenario	Urban macro LoS [3GP17f]
Center frequency	3.75 GHz
Number of multi-path components $L_{(\mathrm{MPC})}$	8
Rician K-Factor $K^{(\mathrm{R})}$	10 dB
BS antenna distribution	UPA
Number of antenna elements N at BS	64
Number of elements in $y^{(\mathrm{C})}$-direction $N_{(\alpha)}$	8
Number of elements in $z^{(\mathrm{C})}$-direction $N_{(\beta)}$	8
Antenna element spacing in $y^{(\mathrm{C})}$- and $z^{(\mathrm{C})}$-direction	$\lambda_{(\mathrm{c})}/2$
Antenna element type	Isotrop
BS height $\left[\tilde{\mathbf{p}}^{(\mathrm{BS})}\right]_3$	0 m
Sequence length, number of samples $N^{(\mathrm{seq})}$	10
DoA estimation method	Bartlett and MUSIC
Target horizontal angular quantization $q_S^{(\alpha)}$	0.1°
Target vertical angular quantization $q_S^{(\beta)}$	0.1°
Horizontal coverage range $\boldsymbol{\alpha}^{(\mathrm{f})}$	$[-60\ 60]^\circ$
Vertical coverage range $\boldsymbol{\beta}^{(\mathrm{f})}$	$[-60\ 60]^\circ$
True horizontal angles of devices α_i	random i.i.d. in $[-60, 60]^\circ \in \mathbb{R}$
True vertical angles of devices β_i	random i.i.d. in $[-60, 60]^\circ \in \mathbb{R}$
Normalization value $g^{(\mathrm{norm})}$, see Eq. (3.9)	1
Distance BS-Sources $d_i^{(\mathrm{BS})}$	50 m

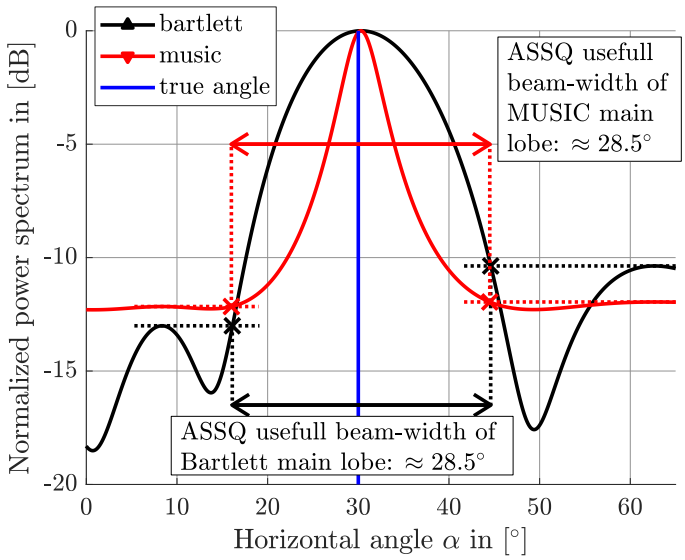


Figure 3.19.: Beam-width of the main-lobe that is useful for ASSQ, comparing Bartlett and MUSIC method, considering $L_{(\mathrm{MPC})} = 8$ multi-path components.

also corresponds to the values in Table 3.5, where for $S = 4$ a quantization of $q_1^{(\alpha)} = q_1^{(\beta)} = 20°$ is selected in the first step without degradation of the position error.

The parameters found in Table 3.5 indicate a monotonic increase in the search range and quantization from $S = 1$ to $S = 5$ for the same step such that

$$\begin{aligned}\left[\mathbf{A}_{(\mathrm{f})}(S = i+1)\right]_{s,1} &\leq \left[\mathbf{A}_{(\mathrm{f})}(S = i)\right]_{s,1} \\ \left[\mathbf{A}_{(\mathrm{f})}(S = i+1)\right]_{s,2} &\geq \left[\mathbf{A}_{(\mathrm{f})}(S = i)\right]_{s,2},\end{aligned} \tag{3.78}$$

where $i \in \mathbb{N}_+$. The same holds for the vertical search range. For the search space quantization, this means that

$$\left[\mathbf{q}^{(\alpha)}(S = i+1)\right]_s \geq \left[\mathbf{q}^{(\alpha)}(S = i)\right]_s \tag{3.79}$$

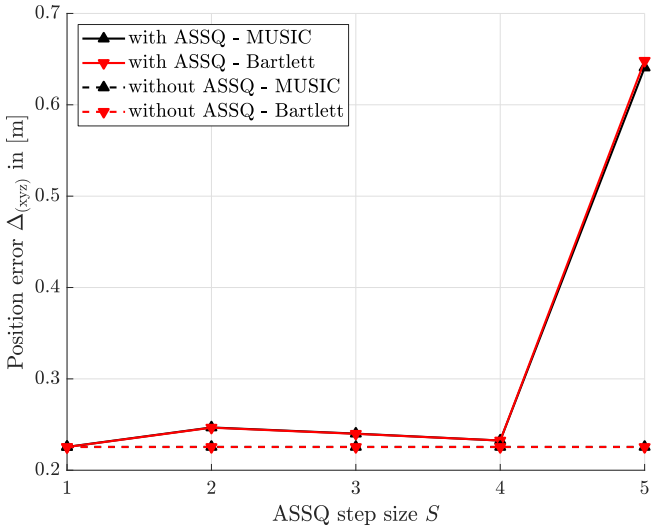
$$\left[\mathbf{q}^{(\beta)}(S = i+1)\right]_s \geq \left[\mathbf{q}^{(\beta)}(S = i)\right]_s. \tag{3.80}$$

This also limits the search on parameter finding for larger S based on parameters obtained from lower S.

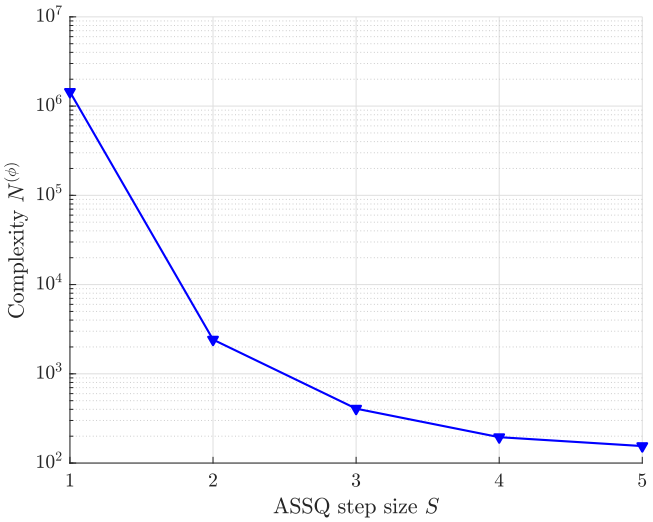
The reason for the performance degradation of $S = 5$ in Fig. 3.20a is caused by the combination of two effects:

1. The effective angular search range is larger than the suitable operational angular range.
2. The combination of multipath propagation with coarse quantization, meaning large values for $q_s^{(\alpha)}$, $q_s^{(\beta)}$.

Effective search range means the effect that based on the maximum in the previous step $s-1$ the search space of step s is relative based on the detected peak, according to Eq. (3.39). This can result in a search space outside of the initial search space in step $s = 1$. Consider the following example with parameters according to $S = 5$ in Table 3.5. The horizontal search angles are $(-60, 0, 60)°$ and $\hat{\alpha}_1 = -60°$ is the maximum, where the subscript indicates the actual step index. The relative horizontal search angles in step $s = 2$ are $(-30, -20, \ldots, 30)°$ which corresponds to the absolute angles of $(-90, -80, \ldots, -30)°$. Based on



(a) Median position error by search-based DoA estimation in horizontal and elevation domain.

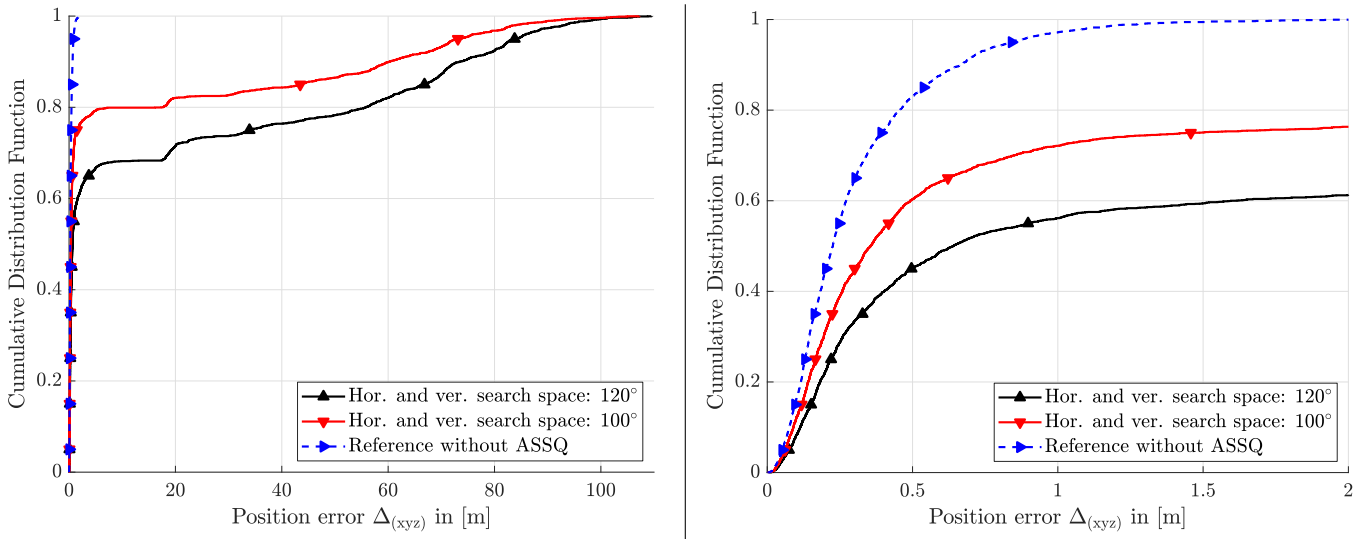


(b) Corresponding number of power spectra $N^{(\phi)}$ required for search-based DoA estimation.

Figure 3.20.: ASSQ performance over varying number of steps. Simulation parameters are set according to Table 3.4 and Table 3.5. ASSQ step size $S = 1$ corresponds to the conventional approach without complexity reduction.

Table 3.5.: ASSQ parameters obtained by (exhaustive) search on minimizing the complexity with corresponding DoA estimation errors of MUSIC and Bartlett.

S	$\left[\mathbf{A}_{(\mathrm{f})}\ \mathbf{B}_{(\mathrm{f})}\right]$ in [°]	$\left[\mathbf{q}^{(\alpha)}\ \mathbf{q}^{(\beta)}\right]$ in [°]	$\mathbf{n}^{(\phi,\mathrm{S})}$	$N^{(\phi)}$	Median $\Delta_{(\mathrm{xyz})}$ in [m]	
					MUSIC	Bartlett
1	$\begin{pmatrix} -60 & 60 & -60 & 60 \end{pmatrix}$	$\begin{pmatrix} 0.1 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 1.442.401 \end{pmatrix}$	1.442.401	0.23	0.23
2	$\begin{pmatrix} -60 & 60 & -60 & 60 \\ -1.94 & 1.94 & -1.54 & 1.54 \end{pmatrix}$	$\begin{pmatrix} 3.88 & 3.08 \\ 0.1 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 1209 \\ 1209 \end{pmatrix}$	2418	0.25	0.25
3	$\begin{pmatrix} -60 & 60 & -60 & 60 \\ -6 & 6 & -6 & 6 \\ -0.5 & 0.5 & -0.6 & 0.6 \end{pmatrix}$	$\begin{pmatrix} 12 & 12 \\ 1 & 1.2 \\ 0.1 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 121 \\ 143 \\ 143 \end{pmatrix}$	407	0.24	0.24
4	$\begin{pmatrix} -60 & 60 & -60 & 60 \\ -10 & 10 & -10 & 10 \\ -2 & 2 & -2 & 2 \\ -0.5 & 0.5 & -0.2 & 0.2 \end{pmatrix}$	$\begin{pmatrix} 20 & 20 \\ 4 & 4 \\ 1 & 0.4 \\ 0.1 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 49 \\ 36 \\ 55 \\ 55 \end{pmatrix}$	195	0.23	0.23
5	$\begin{pmatrix} -60 & 60 & -60 & 60 \\ -30 & 30 & -30 & 30 \\ -5 & 5 & -5 & 5 \\ -1 & 1 & -1 & 1 \\ -0.2 & 0.2 & -0.2 & 0.2 \end{pmatrix}$	$\begin{pmatrix} 60 & 60 \\ 10 & 10 \\ 2 & 2 \\ 0.4 & 0.4 \\ 0.1 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 9 \\ 49 \\ 36 \\ 36 \\ 25 \end{pmatrix}$	155	0.64	0.65



(a) Complete distribution.

(b) Limited distribution range.

Figure 3.21.: Comparison of position error distribution for MUSIC estimator with ASSQ step size $S = 5$ between $\boldsymbol{\alpha}_1^{(\mathrm{f})} = \boldsymbol{\beta}_1^{(\mathrm{f})} = [-60\ 60]^\circ$ and $\boldsymbol{\alpha}_1^{(\mathrm{f})} = \boldsymbol{\beta}_1^{(\mathrm{f})} = [-50\ 50]^\circ$. The dashed reference curve is without ASSQ.

this example the minimum and maximum horizontal search space is given by

$$\begin{aligned} \alpha_{(\mathrm{min})}^{(\mathrm{f})} &= \sum_{s=1}^{S} \left[\mathbf{A}_{(\mathrm{f})}\right]_{1,s}, \\ \alpha_{(\mathrm{max})}^{(\mathrm{f})} &= \sum_{s=1}^{S} \left[\mathbf{A}_{(\mathrm{f})}\right]_{2,s}. \end{aligned} \tag{3.81}$$

Accordingly, the vertical extrema are given as

$$\begin{aligned} \beta_{(\mathrm{min})}^{(\mathrm{f})} &= \sum_{s=1}^{S} \left[\mathbf{B}_{(\mathrm{f})}\right]_{1,s}, \\ \beta_{(\mathrm{max})}^{(\mathrm{f})} &= \sum_{s=1}^{S} \left[\mathbf{B}_{(\mathrm{f})}\right]_{2,s}. \end{aligned} \tag{3.82}$$

From the parameter analysis in Section 3.2.1 the resolution of a UPA decreases towards ±90° which causes ambiguities. This is one of the reasons for the performance degradation of $S = 5$ compared to $S = 1$. As a consequence, this means that a reduction of the search spaces in step $s = 1$ $\boldsymbol{\alpha}_1^{(\mathrm{f})}$ and $\boldsymbol{\beta}_1^{(\mathrm{f})}$ should decrease the position error. This is confirmed in Fig. 3.21[9] where the search space in step $s = 1$ is reduced from $\boldsymbol{\alpha}_1^{(\mathrm{f})} = \boldsymbol{\beta}_1^{(\mathrm{f})} = [-60\ 60]^\circ$ to $\boldsymbol{\alpha}_1^{(\mathrm{f})} = \boldsymbol{\beta}_1^{(\mathrm{f})} = [-50\ 50]^\circ$. The corresponding 50 %-ile position error is approximately halved down to ≈ 0.35 m. The used ASSQ parameters for the reduced search space are listed in Table B.13 in the appendix. Note that also the computation complexity is reduced, from 155 to 142. The maximum/minimum effective search space is changed from ±96.2° to ±81.2°.

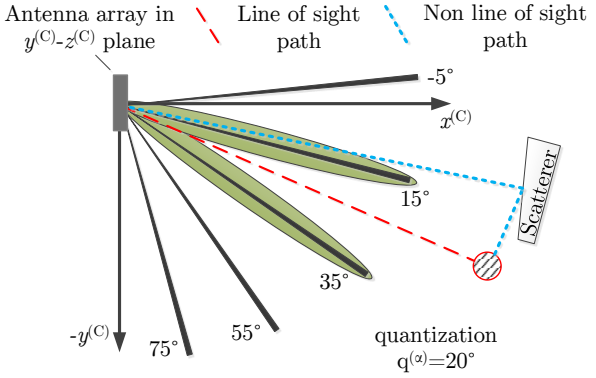
Remark. *The minimum and maximum effective search range with ASSQ search according to Eq.* (3.81)-*Eq.* (3.82) *should be* $> -90^\circ$ *and* $< 90^\circ$*, respectively.*

However, there are still 28 % of the estimation errors $\Delta_{(xyz)}$ >1 m in contrast to 2 % without ASSQ. As mentioned above these errors are caused by the combination of multipath propagation in combination with

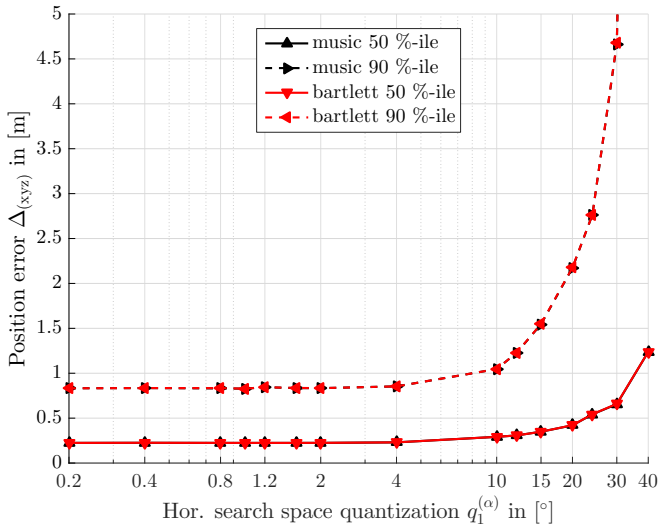
[9] Only MUSIC is shown because the performance of Bartlett is approximately the same.

Table 3.6.: ASSQ parameters for Fig. 3.22b and $S = 2$.

Search Space	Quantization
$\mathbf{A}_{(\mathrm{f})} = \begin{pmatrix} -60° & 60° \\ -\frac{q_1^{(\alpha)}}{2} & \frac{q_1^{(\alpha)}}{2} \end{pmatrix}$, $\mathbf{B}_{(\mathrm{f})} = \begin{pmatrix} -60° & 60° \\ -0.5° & 0.5° \end{pmatrix}$	$\mathbf{q}^{(\alpha)} = \begin{pmatrix} q_1^{(\alpha)} \\ 0.1° \end{pmatrix}$, $\mathbf{q}^{(\beta)} = \begin{pmatrix} 1° \\ 0.1° \end{pmatrix}$



(a) Estimation of a NLoS instead of the LoS path due to a coarse quantization of the search space.



(b) Dependency of the position error on the horizontal quantization $q_1^{(\alpha)}$ in ASSQ step $s = 1$. Parameters are set according to Table 3.4 and Table 3.6.

Figure 3.22.: Performance degradation due to the combination of coarse quantization, multipath propagation and a too narrow beam-width of the main-lobe.

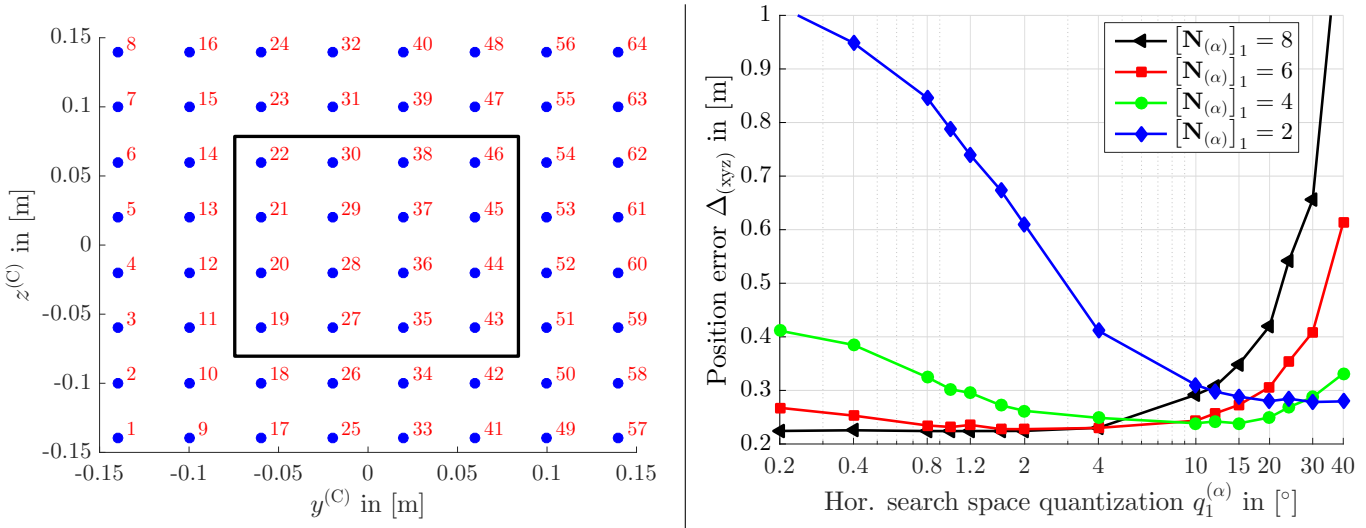
large quantization steps $q_s^{(\alpha)}$, $q_s^{(\beta)}$. A coarse quantization results in a larger probability that NLoS paths are estimated. A toy-example in Fig. 3.22a illustrates this effect. Due to reflection from a scatterer, the superposition of coarse quantization and limited HPBW results in a larger receive power in the power spectrum for the NLoS than the LoS path. The exact probability depends also on the angular spread of the multi-path components at the BS. This effect is shown in Fig. 3.22b for $S = 2$, with parameters in Table 3.6 overwriting those from Table 3.4.

In Fig. 3.22b, both the 50 %-ile and the 90 %-ile position error are approximately constant up to a threshold of $q_1^{(\alpha)} = 4°$. For $q_1^{(\alpha)} > 4°$ a non-linear increase is observed due to the effect described in Fig. 3.22a. Note that due to the dynamic range of the quantization values $q_1^{(\alpha)}$ the $x^{(C)}$-axis is displayed in the logarithmic scale for presentation reasons.

In order to compensate this effect and to "capture" the power of the LoS path for large quantization steps the HPBW has to be increased. According to Eq. (3.84) this is achieved by using less antenna elements which corresponds to the selection of a sub-array out of the large array. This means, that only a subset of receive signal in (3.1) is used for signal processing and no change on the hardware is required. Consequently, a new parameter to indicate the adaptive sub-array antenna selection per ASSQ step is required for both, the horizontal and the vertical domain, given by $\mathbf{N}_{(\alpha)}$ and $\mathbf{N}_{(\beta)}$, respectively, defined in Table 3.7. Note that the sub-array antennas are selected symmetrical with respect to the center of the array. An example for sub-array selection $[\mathbf{N}_{(\alpha)}]_1 = [\mathbf{N}_{(\beta)}]_1 = 4$ out of the 8×8 is given in Fig. 3.23a. This sub-array selection can be combined with all the searched-based DoA estimation methods since only the dimension of the receive signal and beamformer is changed. As before, the focus is on the Bartlett and MUSIC method for performance evaluation.

Table 3.7.: Parameter description for antenna adaptive ASSQ.

$\mathbf{N}_{(\alpha)}$	Vector of size $S \times 1$ where $[\mathbf{N}_{(\alpha)}]_s \in \{1, \ldots, N_{(\alpha)}\}$ is the number of horizontal antennas selected in step s
$\mathbf{N}_{(\beta)}$	Vector $S \times 1$ where $[\mathbf{N}_{(\beta)}]_s \in \{1, \ldots, N_{(\beta)}\}$ is the number of vertical antennas selected in step s.



(a) Positions and indices of the 8×8 UPA. The black box indicates the inner 4×4 sub-array.

(b) Position error over quantization in step $s = 1$ using a sub-array in horizontal domain.

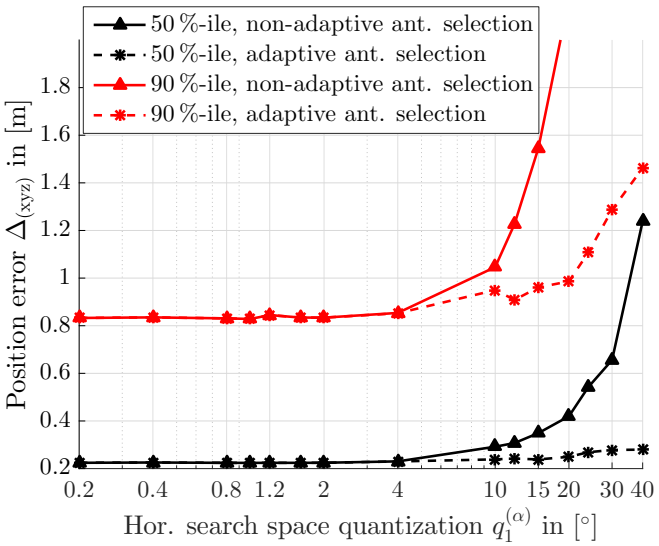
Figure 3.23.: Performance with complete and sub-array antenna selection for quantization in first step of ASSQ algorithm.

In Fig. 3.23b the 50 %-ile position error of the sub-array selection is given for $[\mathbf{N}_{(\alpha)}]_1 = 8$, $[\mathbf{N}_{(\alpha)}]_1 = 6$, $[\mathbf{N}_{(\alpha)}]_1 = 4$, and $[\mathbf{N}_{(\alpha)}]_1 = 2$ antenna elements in the horizontal domain and $[\mathbf{N}_{(\beta)}]_1 = 8$ antenna elements in the vertical domain[10]. The position error for $[\mathbf{N}_{(\alpha)}]_1 = 8$ increases together with the quantization value $q_1^{(\alpha)}$ over the complete range. For $[\mathbf{N}_{(\alpha)}]_1 = 6$ and $[\mathbf{N}_{(\alpha)}]_1 = 4$ the curves have the shape of a bath-tube while the position error for $[\mathbf{N}_{(\alpha)}]_1 = 2$ is decreasing with an increasing $q_1^{(\alpha)}$. These curve shapes are explained as follows. With $q_1^{(\alpha)}$ also the search space in the second step is changed by $\left[\mathbf{A}_{(\mathrm{f})}\right] = \begin{pmatrix} -60 & 60 \\ -\frac{q_1^{(\alpha)}}{2} & +\frac{q_1^{(\alpha)}}{2} \end{pmatrix}$ while the quantization in the second step is constant with $q_1^{(\alpha)} = 0.1°$ and the number of horizontal antennas is $[\mathbf{N}_{(\alpha)}]_2 = 8$. This means, that a low $q_1^{(\alpha)}$ in the first step is limiting the search space in the second step where the complete array with the target quantization of 0.1° is used. Thus, the larger the quantization value $q_1^{(\alpha)}$ in the first step the larger is the search space in the second step, and therefore the lower the position error. This effect is dominant for $[\mathbf{N}_{(\alpha)}]_1 = 2$. For $[\mathbf{N}_{(\alpha)}]_1 = 4$ and $[\mathbf{N}_{(\alpha)}]_1 = 6$ this effect is dominant only in the first part, but for larger $q_1^{(\alpha)}$ the previous explained position error degradation caused by coarse search space quantization illustrated in Fig. 3.22 kicks in and dominates. Thus, curves for $[\mathbf{N}_{(\alpha)}]_1 = 4$ and $[\mathbf{N}_{(\alpha)}]_1 = 6$ have a bath-tube shape.

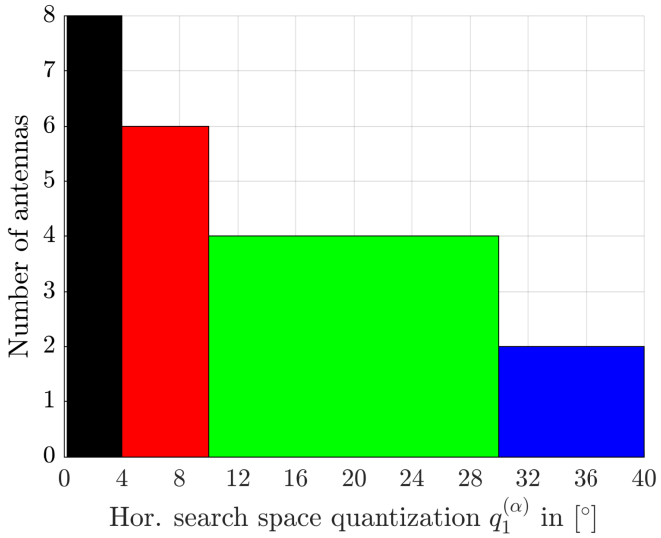
Looking at the minimum envelope of all the curves there are three crossings points between them, which results in four quantization ranges where the corresponding number of antenna elements yield the minimum position error.

The minimum envelope of Fig. 3.23b and the reference of $[\mathbf{N}_{(\alpha)}]_1 = 8$ is given in Fig. 3.24a. With adaptive

[10] The corresponding 90 %-ile result is given in the annex with qualitative the same insights.



(a) Comparison of adaptive antenna selection with non-adaptive antenna selection. In the case of non-adaptive antenna selection $\left[\mathbf{N}_{(\alpha)}\right]_s = 8,\ \forall s \in \mathcal{S}$. In case with adaptive antenna selection $\left[\mathbf{N}_{(\alpha)}\right]_s$ is obtained according to Eq. (3.83).



(b) Number of antenna elements yielding the minimum position error.

Figure 3.24.: Position error for adaptive antenna selection over quantization in step $s = 1$ with ASSQ and limitation to horizontal dimension.

antenna selection the performance of the 50 %-ile is approximately constant, while the 90 %-ile increases up to 1.5 m. The corresponding number of antenna elements selected over the quantization range of $q_1^{(\alpha)}$ is given in Fig. 3.24b which results in

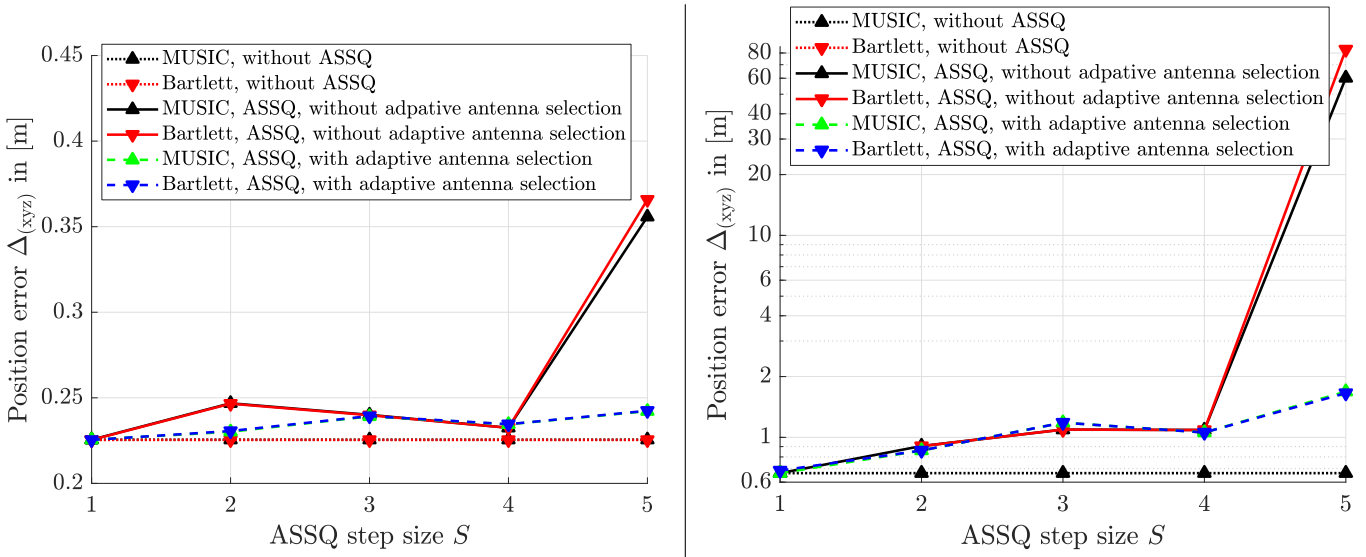
$$\left[\mathbf{N}_{(\alpha)}\right]_s = \begin{cases} 8 & \text{if } q_s^{(\alpha)} \leq 4° \\ 6 & \text{if } 4° < q_s^{(\alpha)} \leq 10° \\ 4 & \text{if } 10° < q_s^{(\alpha)} \leq 30° \\ 2 & \text{if } 30° < q_s^{(\alpha)}. \end{cases} \tag{3.83}$$

The same is applied also for $\left[\mathbf{N}_{(\beta)}\right]_s$. Note that the antenna selection in Eq. (3.83) is derived for the given simulation parameters in Table 3.4, e.g. for a certain number of samples or a certain center frequency. Depending on the system parameters these values may change, however this optimization has to be done only once for a given system.

Remark. *The switching points for ASSQ with adaptive antenna selection in this thesis are numerically derived for only a subset of (non-continuous) quantization values $q_1^{(\alpha)}$ due to the integer quantization condition of the search-space. With other search space values, the exact antenna switching points may differ, but nevertheless the general rule that the coarser the quantization (meaning a large step size) the less antenna elements have to be used, remains.*

Note that the adaptive antenna selection impacts the corresponding steps in the ASSQ, but if in the last step the full antenna array is used, e.g. here 8 antennas for a target quantization of 0.1°, the final resolution is the same as without adaptive antenna selection.

Finally, to verify the effectiveness of the adaptive antenna selection, evaluations for the ASSQ parameters from Table 3.5 are performed with the adaptive antenna selection according to Eq. (3.83). For $S = 5$



(a) 50 %-ile position error from CDF.

(b) 90 %-ile position error from CDF. $y^{(C)}$-axis is in logarithmic scale.

Figure 3.25.: Position error of MUSIC and Barlett method using ASSQ with and without adaptive antenna selection according to Eq. (3.83).

the ASSQ parameter from Table B.13 are used. Especially, the position error for $S = 5$ is reduced from 0.36 m to 0.24 m. The corresponding 90 %-ile for $S = 5$ is reduced from 60 m and 80 m to < 2 m for MUSIC and Bartlett method, respectively. The reason why Bartlett and MUSIC DoA estimation yield approximately the same position errors is because of approximately the same useful beam-width as shown in Fig. 3.19. The reference without ASSQ yields the best performance but requires the largest complexity of $N^{(\phi)} = 1\,442\,401$ power spectrum computations, while in contrast to this ASSQ with $S = 5$ steps requires only 155 computations. Note that the complexity in terms of power spectrum computations is the same with and without adaptive antenna selection. An additional advantage of adaptive antenna selection is that the dimension of the receive signal and beamforming vectors are reduced and thus less complexity is required for matrix operations with the receive signal and beamformer.

3.2.3. Conclusions

A main advantage of search-based algorithms is that generic antenna array structures can be used for DoA estimation. However, search-based DoA estimation algorithms require large computational complexity to achieve high accuracy, e.g. in the sub-degree range to satisfy 3rd Generation Partnership Project (3GPP) new radio (NR) requirements for positioning. This is even more severe in the two dimensional search space, where the cross-dependency of horizontal and vertical estimation prevents straightforward independent searches in each dimension. In this section, a method using adaptive quantization of the search space in multiple steps called ASSQ is proposed to reduce the required complexity without significant performance loss. ASSQ is able to reduce the required computational complexity in the given example from $1.4 \cdot 10^6$ to 155 computations of the power spectrum. While the exact reduction value may change with the search range and the number of antennas, the principle of ASSQ with adaptive antenna selection can be applied to any array shape or planar arrays with any number of antennas to reduce the number of power spectrum computations for search-based DoA estimation methods aiming for low estimation errors. In this thesis it was shown for Bartlett and MUSIC. In 5G a wide range of BS antenna arrays is expected in a similar range as considered in this thesis, e.g. there are already 8×8 antenna arrays for 28 GHz and below 6 GHz on the market [Bjö19]. The following design rules on how to obtain the required ASSQ parameters are summarized as follows.

- The search space in step $s+1$ has to be equal to the quantization range in previous step s, see Fig. 3.17.
- The quantization in step $s+1$ has to be smaller than the quantization in previous step s.
- In each step s the quantization has to be an integer multiple of the search space to avoid white spaces in the search space.
- For planar arrays, the effective search range as the sum of the search range over S steps should be $< 90°$, where necessary reduce search range in step 1.
- If quantization becomes large, adaptive sub-array selection reduces estimation errors.

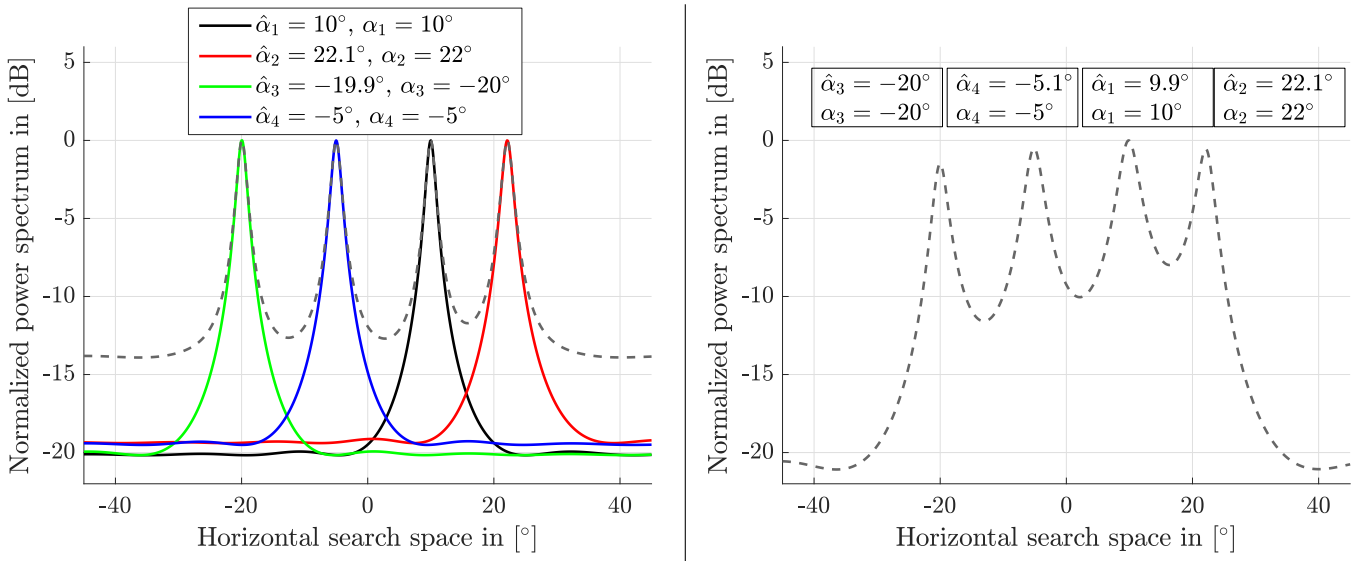
As a drawback, for each scenario, e.g. a given search space to be observed, a parameter search for the ASSQ algorithm is required. While the above design rules help with the parameter search this is not a systematic approach. However, this parameter search has to be done only once for a given setting. The presented complexity reduction by ASSQ is applicable for any search space based DoA estimation method and is shown in this section for Bartlett and MUSIC.

3.3. Multiple-Source Estimation

Multiple-source detection means that multiple directions are estimated in the same power spectrum. It is assumed that these directions correspond to multiple devices. This is considered a likely scenario in cellular systems where multiple devices transmit reference signals in the uplink on the same time-frequency resource synchronized by timing advance [3GP18d, 3GP18a]. This was also partly the motivation for advances on DoA estimation algorithms in the past, e.g. leading to the development of MUSIC, see Section 3.2. From a multiple-user communications point of view, the joint DoA estimation of multiple sources is a classical interference scenario as given in massive MIMO uplink by spatial multiplexing of devices on the same time-frequency resource. The advantage of multiple-spource DoA estimation is that less resources in the network are required for positioning and thus increasing the overall efficiency. Therefore, the focus is on the grouping of devices to multiplex as many as possible while keeping the loss in DoA estimation "affordable" compared to the single source case. For this the same DoA estimators as before are used.

In literature, the topic of multiple-source DoA estimation is widely studied [CPT+13], [ASP14], [ASH14], [KTH14a]. For the basic principles on multiple-source DoA estimation the interested reader is referred to Chapter 1.6 in [TF09]. However, not much attention has been given to the grouping of these devices. While [CPT+13] proposed a novel scheme for joint DoA estimation of multiple sources and studies the trade-off on computation and sampling cost, in simulations at most fours devices are placed on fixed well separated positions. Authors in [ASP14] and [ASH14] compare high-resolution DoA estimation algorithms for non-coherent devices groups, however devices are also located at predefined and favorable angles. Authors in [KTH14a] proposed a robust multiple-source localization method for planar arrays in the area of acoustic source detection, but again, only two and four and well-placed sources are considered. While the separation or minimum angular distance between two sources is studied and limited by the array and algorithm, it is not clear how much sources can be detected in total or what are the trade-offs. Therefore, in this section the focus is on the grouping of sources for multiple-source DoA estimation.

In the previous Section 3.2, the focus was on high accurate low complex DoA estimation for a single source, e.g. for location service providers. In this section the motivation is different. In this section the aim is to reduce the number of resources consumed by positioning signals, due to multiple-source estimation, in order to free these resources for their intended purpose of data-transmission. This motivation is substantiated by the recent 5G frequency auction in Germany [Bun19], where 6 549 651 000 € have been spend for licensed spectrum, 120 MHz at 2 GHz and 300 MHz at 3.6 GHz. This example shows how valuable resources have become for cellular communications.



(a) Single source DoA estimation.

(b) Multiple source DoA estimation.

Figure 3.26.: MUSIC power spectrum with normalized channels and LoS path only, $L_{(\mathrm{MPC})} = 1$.

In the reminder of this section, challenges for multiple-source DoA estimation are discussed in Section 3.3.1, then the grouping algorithm balancing the trade-offs for multiple-source localization is given in Section 3.3.2, and finally Section 3.3.3 provides a summary and conclusion of this section.

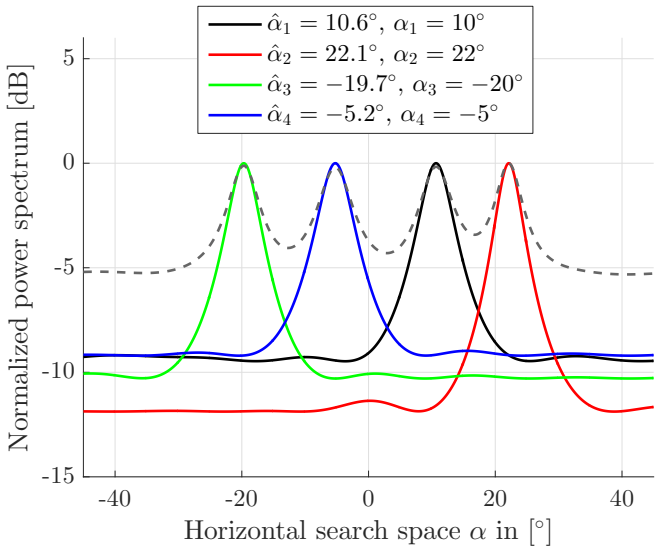
3.3.1. Challenges

The extension from single to multiple-source estimation imposes additional constraints that have to be taken into account. Similar to the minimum side lobe ratio (MSLR) estimation in Section 3.1.3, detection of the K largest peaks is required, where K is the number of sources to be estimated. However, in this section, K is known at the BS because of the grand-based or scheduled uplink transmission controlled by the network in LTE and 5G NR. Due to the overlap of the power spectrum and the HPBW, a minimum angular separation between the sources is required. Otherwise, a side-lobe of the power spectrum is falsely considered as a source DoA. A toy example to illustrate this effect is given in Fig. 3.26 and Fig. 3.27. Firstly, in Fig. 3.26a the power spectra of four separately measured sources are given and the dashed line provides the sum of them. Secondly, the multiple-source joint DoA estimation of the same four sources is given in Fig. 3.26b and the four DoAs are estimated with similar errors compared to single source estimation. The difference between the two figures is that in Fig. 3.26 only the LoS path is considered such that $L_{(\mathrm{MPC})} = 1$, while in Fig. 3.27 multiple paths are assumed such that $L_{(\mathrm{MPC})} = 8$. Again, on the left hand side in Fig. 3.27a the power spectrum of the four sources is measured separately while on the right hand side in Fig. 3.27b the joint power spectrum is given. It can be observed that, in the joint power spectrum, the fourth largest peak at 43.3° no longer corresponds to any of the source DoAs, because the peak of source three is “hidden” within the side-lobes of the other sources. There are two reasons for this false detection.

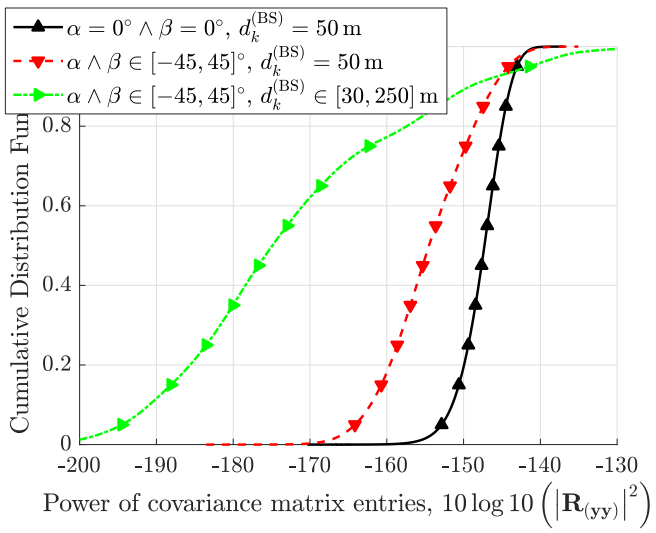
1. The MUSIC power spectrum is a pseudo-spectrum and the joint detection of multiple paths is still limited by the resolution of the antenna array.

2. The power difference of the peaks, due to superposition of signals.

Note that the channels in the above toy example are normalized to one according to Eq. (3.8) and in Chapter B the corresponding figures for Bartlett beamformer are shown. As one of the MUSIC design goals in [Sch86] was to increase resolution of multiple-source estimation compared to previous DoA estimators, it outperforms Bartlett. Therefore, the focus in the the remainder of this section is on MUSIC.



(a) Single source DoA estimation.



(b) Multiple source DoA estimation. False detection for source three. True angle of source one is $\alpha = 10°$. False detected peak is at $\hat{\alpha}_1 = -43.3°$.

Figure 3.27.: MUSIC power spectrum with normalized channels and $L_{(\text{MPC})} = 8$ multi-path components.

Number of antennas $N_{(\alpha)}$ in ULA	4	8	16
HPBW measured from QuaDRiGa in °	25.7	13	6.3
HPBW approximated with Eq. (3.85) in °	25.3	12.6	6.3

Table 3.8.: Comparison of measured and approximated HPBW of a ULA over number of antennas N.

To verify the limitation of the antenna array, Fig. 3.28a shows the normalized receive power spectrum at the BS over the number of antenna elements and their corresponding HPBW $\hat{\alpha}_{(\text{HPBW})}$. In Table 3.8 the HPBW measured from Fig. 3.28a is compared with the approximation for large linear arrays according to [JP06] (page 18 and 19) by

$$\hat{\alpha}_{(\text{HPBW})}(\alpha) = \frac{\hat{\alpha}_{(\text{HPBW})}(0°)}{\cos \alpha}, \tag{3.84}$$

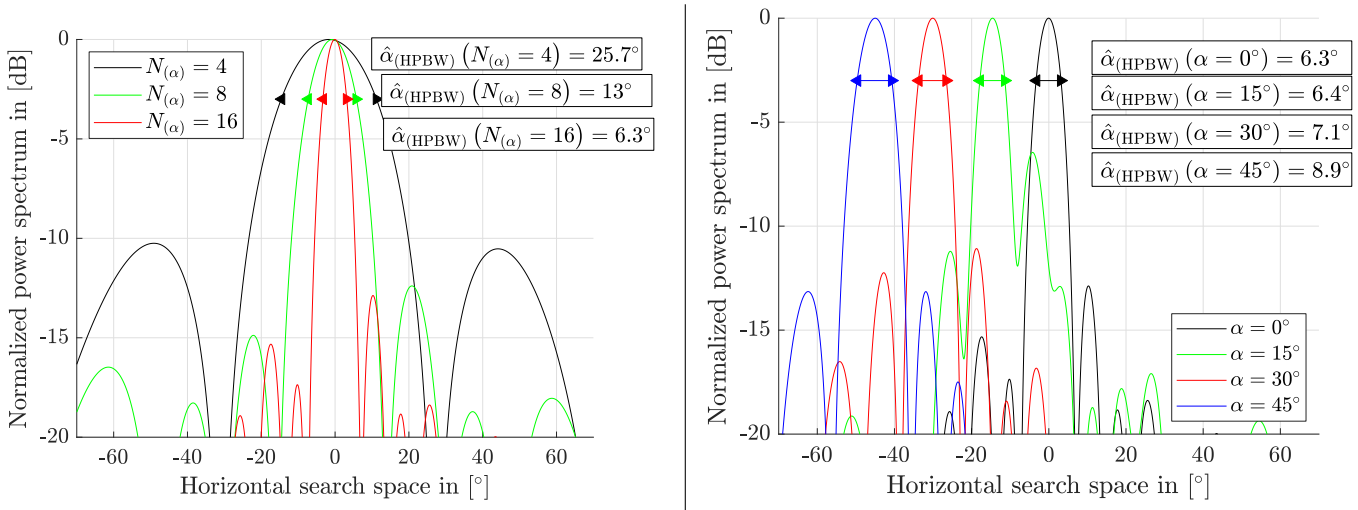
where $\hat{\alpha}_{(\text{HPBW})}(0°)$ is approximated by

$$\hat{\alpha}_{(\text{HPBW})}(0°) = 180° \frac{0.88}{\pi} \frac{\lambda_{(\text{c})}}{N_{(\alpha)} d_{i,i+1}^{(\text{N})}} \approx \frac{101°}{N_{(\alpha)}}, \tag{3.85}$$

where $d_{i,i+1}^{(\text{N})}$ is the distance between neighbor antenna elements, which is assumed with $d_{i,i+1}^{(\text{N})} = \frac{\lambda_{(\text{c})}}{2}$. The maximum error of the HPBW between measured and approximated by Eq. (3.84) in Table 3.8 is $\leq 3\,\%$.

According to Eq. (3.84), the HPBW not only depends on the number of antenna elements, but also on the source angle α itself. In Fig. 3.28b the normalized angular power spectrum is given for changing α from 0° to 45°. The measured HPBW is compared to the approximated HPBW in Table 3.9 and the error between them is $\leq 5\,\%$. Therefore, Eq. (3.84) is used to obtain the HPBW in the remainder of this thesis.

Applying the approximation of the HPBW by Eq. (3.84) to the device at $\alpha_2 = 22°$ in the toy example in Fig. 3.27b $\hat{\alpha}_{(\text{HPBW})}(\alpha_2) \approx 13.75°$. This is more than the angular distance to the neighbor device $\alpha_2 - \alpha_1 = 12°$. Consequently, the true angle of device one is is within an angular range that corresponds to the 3 dB power range around the peak of device two and the peak of the device one at $\alpha_1 = 10°$ disappears in the slope of the larger peak of device two.



(a) Number of antenna elements $N_{(\alpha)}$ changed.

(b) Horizontal angle α of device positions changed.

Figure 3.28.: Normalized receive power spectrum in the horizontal domain.

Horizontal angle of device α in °	0	-15	-30	-45
HPBW measured from QuaDRiGa in °	6.3	6.4	7.1	8.9
HPBW approximated with Eq. (3.84) in °	6.3	6.5	7.3	8.9

Table 3.9.: Comparison of measured and approximated HPBW of a ULA over device angles α.

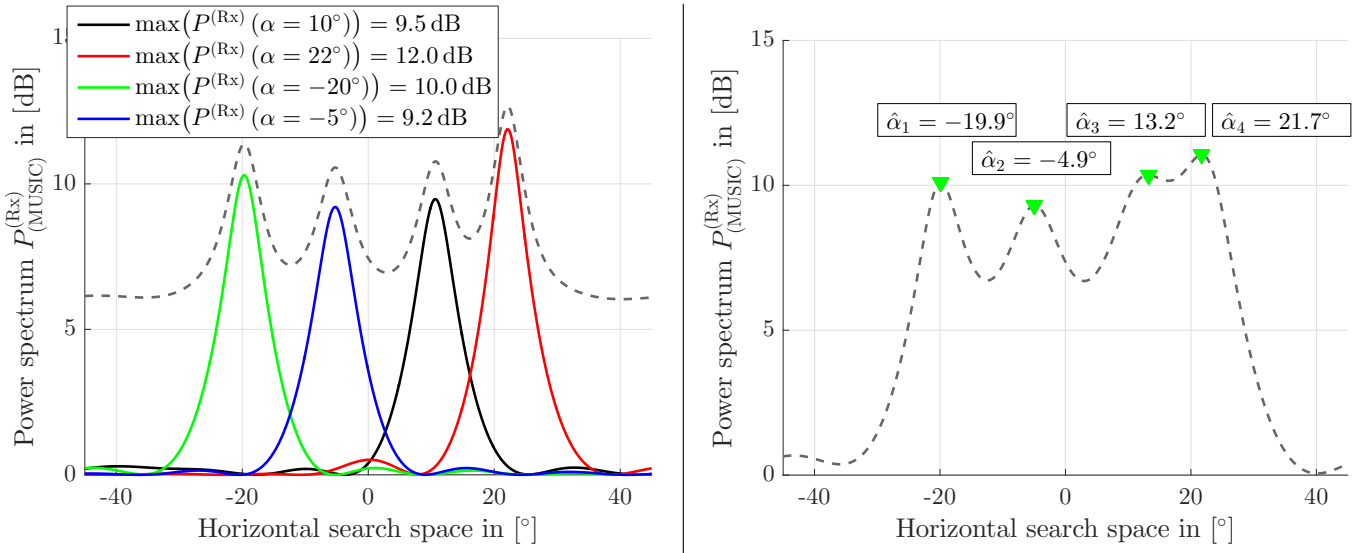
As mentioned before, the channels of all devices are normalized to one by $g^{(\mathrm{norm})} = 1$ according to Eq. (3.8). This is a common but also ideal assumption in most of the literature [CPT+13], [ASP14], [ASH14], [KTH14a], because in reality the receive power of the signals depends on the path loss of the signals. The path-loss itself depends on the distance between BS and device and an environment dependent path-loss factor. These environment path-loss factors are usually empirically obtained values [Ben07]. Note that also the LoS can be obstructed, e.g. by leafs on trees. In order to demonstrate the impact of the path-loss, Fig. 3.29 depicts the power spectrum with not-normalized channels, such that $g^{(\mathrm{norm})} = 0$. All devices have the same distance to the BS and difference of the peaks in the power spectrum depends on the ratio of main to side lobes (K-factor) and the beamforming gain by the different angles of the devices. "Surprisingly", in case of not normalized channels the source at $\alpha = 10°$ from the above toy example can be detected again, demonstrating that dependencies are not straight forward. The variations in the power of the covariance matrices are clarified in Fig. 3.30 for the following scenarios:

a) Black solid line, upward pointing triangles: Devices are located at $\alpha = 0°$, $\beta = 0°$, $d_k^{(\mathrm{BS})} = 50\,\mathrm{m}$.

b) Red dashed line, downward pointing triangles: Devices are located in interval $\alpha \in [-45, 45]°$, $\beta = 0°$ and $d_k^{(\mathrm{BS})} = 50\,\mathrm{m}$.

c) Green, dotted-dashed line, right-hand side pointing triangles: Devices are located in intervals $\alpha \in [-45, 45]°$ and $\beta = 0°$, $d_k^{(\mathrm{BS})} \in [30, 250]\,\mathrm{m}$.

Fig. 3.30 shows the CDF of the power of entries in the covariance matrix in [dB], by $10 \log 10 \left(\left| \mathbf{R}_{(\mathbf{yy})} \right|^2 \right)$. An increase of the variance from scenario a) to c) is observed with a spread between the 5 %-ile and 95 %-ile of 9.8 dB, 20.0 dB, and 53.1 dB, respectively. A consequence of this evaluation is that any user-grouping algorithm for multiple-source estimation has to take into account the receive power of the sources. Due to variation in the range of 10 dB even for users at the same distance and angle, also a power margin has to be considered.



(a) Single source DoA estimation.

(b) Multiple source DoA estimation.

Figure 3.29.: Power spectrum of MUSIC with $L_{(\mathrm{MPC})} = 8$ multi-path components and not-normalized channels.

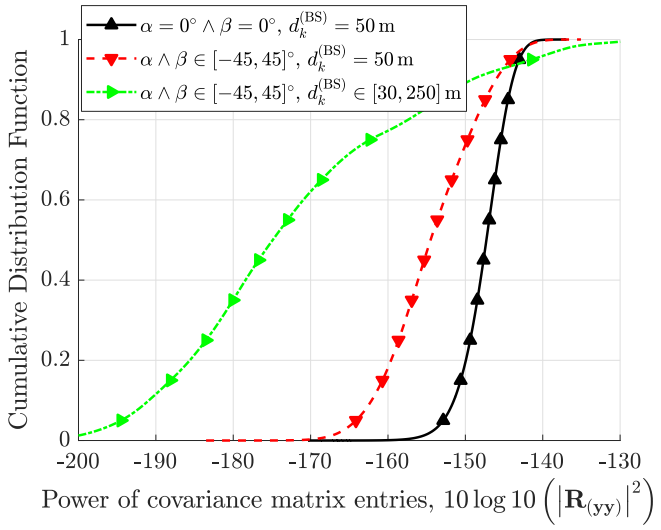


Figure 3.30.: Poewr distribution covariance matrix entries for a) same angle and same distance, b) angle range and same distance, and c) angle range and distance range. The corresponding spread between 5 %-ile and 95 %-ile is $[9.8, 20.0, 53.1]$ dB.

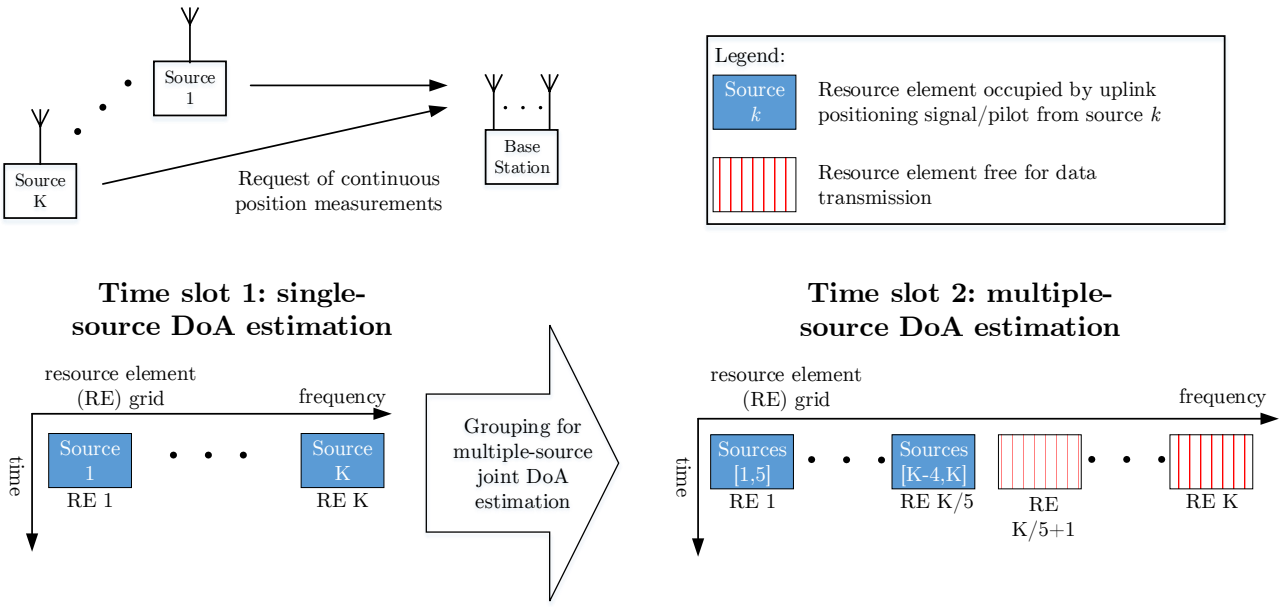


Figure 3.31.: Motivation for multiple-source grouping, where K devices request continuous positioning measurements, e.g. drones requesting navigation assistance. On the left hand side in time slot one, the DoAs of devices 1 to K are estimated on orthogonal time-frequency resources elements, e.g. using the ASSQ method from Section 3.2.2 which works well on small bandwidths, such that one device is estimate per resource element. This occupies K resource elements. Recall that the uplink in cellular network is grand-based, i.e. scheduled by the network. These (single device) estimated DoAs are input to the grouping algorithm in this thesis for multiple-source estimation. On the right hand side, in time slot two, five devices are assigned to each resource element for joint DoA estimation as an exemplary outcome of the grouping algorithm. Multiple-source DoA estimation in time slot two requires $\left\lceil \frac{K}{5} \right\rceil$ resource elements for DoA estimation, thus the remaining $K - \left\lceil \frac{K}{5} \right\rceil$ resource elements are free for data-transmission in contrast to single device DoA estimation.

The second information required for a grouping algorithm are the DoA information of the devices itself, otherwise false detection can occur as demonstrated in the above toy-examples in Fig. 3.26 to Fig. 3.29. This is a typical chicken-egg problem, here the DoAs of the devices are required in order to estimate multiple of them simultaneously. However, in LTE and 5G NR cellular networks the following can be assumed. Any device requiring a position measurement from the network has to send an uplink request. These uplink requests are on scheduled time-frequency resource, where the scheduling is done by the network. These uplink signals on orthogonal time-frequency resources can be used for single-device DoA estimation and the ASSQ algorithm with MUSIC or Bartlett from the previous section can be applied, see Fig. 3.31. Then, all devices that are requesting a position measurement are assigned to a scheduling pool for multiple-source estimation in order to reduce the number of required network resources occupied by positioning signals. The following use case is envisioned. Self-moving vehicles or devices that are assisted or observed by the cellular network, e.g. cars or drones, require a continuous update of the position measurement that is based on estimated uplink DoAs. After a single device DoA estimation based on the uplink request for positioning, the continuous DoA measurement updates can be performed by multiple-source estimation in order to save network resources compared to single-device DoA estimation, where every device is transmitting a positioning signal on an orthogonal time-frequency resource.

Remark. *One can argue that the assumption of a-priori DoA information is a strong constraint and there is no need for multiple-source estimation. However, for mobile devices there are applications that require a*

constant update of the position. In order to reduce the required amount of network resources multiple devices can be orthogosnalized in the spatial (angular) domain by grouping. The a-priori DoA information required for multiple-source grouping can be derived from parallel frequency resources or a previous time slot, e.g. an initial single source DoA estimation using the ASSQ method from Section 3.2.2 in combination with MUSIC or Bartlett. Especially in use cases where a lot of devices require a continuous position update, the multiple-source DoA estimation can reduce the amount of network resources required for positioning signals.

Beside the above discussed challenges, a further "obstacle" of multiple-source estimation is the evaluation of the DoA estimation error itself. The DoA estimation error in the single source case is defined in Eq. (3.46) and Eq. (3.47) where each estimated angle corresponds to a true or input angle one-by-one. However, when estimating multiple sources, there is a set of estimated angles, denoted by $\hat{\boldsymbol{\alpha}}$, and a set of true or input angles, with no unambiguous mapping between both sets. Therefore, "false detection" and the resulting false mapping of device ids to estimated DoAs can happen. The problem of "false detection" is referred to in Chapter 1.6 of [TF09] as a hypothesis testing from the set of estimated DoAs, which is a resolution problem that depends on the number of antennas, correlation of signals, signal to interference and noise ratio (SINR) of signals to each other. Note that the challenge of mapping the estimated DoAs to the device ids based on the input DoAs is effecting the evaluation of the DoA estimation error instead of the DoA estimation itself. The following notation is introduced in order to classify false mapping of estimated DoAs to input DoAs. For the ease of notation, the input DoAs of the multiple-source grouping algorithm are denoted as the true DoAs in the below paragraphs. This avoids introducing a second DoA estimation error, however all the true angles can be replaced by estimated DoAs obtained from single source estimation.

The set of estimated angles is derived from the multiple-peak detection in the power spectrum described in Section 3.1.3. In Fig. 3.26 the mapping of estimated angles to input or true angles is intuitive, as there are four peaks detected such that

$$\alpha_i - \epsilon_{(\alpha)} < \hat{\alpha}_i < \alpha_i + \epsilon_{(\alpha)}, \tag{3.86}$$

where $\epsilon_{(\alpha)} \in \mathbb{R}_+$ denotes a threshold value, e.g. in the discussed example $\epsilon_{(\alpha)} = 0.1°$. In contrast to this, in Fig. 3.27, the estimated angle $\hat{\alpha}_1$ doesn't correspond to one of the true or input angles in $\boldsymbol{\alpha}$. There is no change in the gradient sign from one estimated peak towards the next true DoA. Therefore, no peak is found at the true DoA. Thus, the threshold $\epsilon_{(\alpha)}$ has to be larger than the angular distance between two angles to fulfill Eq. (3.86). Such a large $\epsilon_{(\alpha)}$ can result in multiple solutions for mapping elements in $\hat{\boldsymbol{\alpha}}$ to elements in $\boldsymbol{\alpha}$. Therefore, $\epsilon_{(\alpha)}$ is constraint by

$$\epsilon_{(\alpha)} < \underset{i,j\in[1,\ldots,K],i\neq j}{\arg\min} \frac{|\alpha_i - \alpha_j|}{2}. \tag{3.87}$$

The same holds for vertical estimated angels such that

$$\beta_i - \epsilon_{(\beta)} < \hat{\beta}_i < \beta_i + \epsilon_{(\beta)}, \tag{3.88}$$

where $\epsilon_{(\beta)} \in \mathbb{R}_+$ and

$$\epsilon_{(\beta)} < \underset{i,j\in[1,\ldots,K],i\neq j}{\arg\min} \frac{|\beta_i - \beta_j|}{2}. \tag{3.89}$$

In other words, the threshold in multiple-source estimation has to be smaller than half the minimum angular distance between any two devices. The mapping of estimated to input or true angles with thresholds $\epsilon_{(\alpha)}$ and $\epsilon_{(\beta)}$ corresponds to a classification of the estimated angles such that

$$\begin{aligned} \Delta_{(\alpha)} &= \begin{cases} \hat{\alpha}_j - \alpha_i & \text{if } \alpha_i - \epsilon_{(\alpha)} < \hat{\alpha}_j < \alpha_i + \epsilon_{(\alpha)} \\ \emptyset & \text{if } \hat{\alpha}_j \leq \alpha_i - \epsilon_{(\alpha)} \text{ or } \hat{\alpha}_j \geq \alpha_i + \epsilon_{(\alpha)} \end{cases}, \\ \Delta_{(\beta)} &= \begin{cases} \hat{\beta}_j - \beta_i, & \text{if } \beta_i - \epsilon_{(\beta)} < \hat{\beta}_j < \beta_i + \epsilon_{(\beta)} \\ \emptyset, & \text{if } \hat{\beta}_j \leq \beta_i - \epsilon_{(\beta)} \text{ or } \hat{\beta}_j \geq \beta_i + \epsilon_{(\beta)}, \end{cases} \end{aligned} \tag{3.90}$$

where $i, j \in \{1, \ldots, K\}$. Note that the thresholds $\epsilon_{(\alpha)}$ and $\epsilon_{(\beta)}$ are only used in the evaluation for associating estimated DoAs to input DoAs. This means, that if an estimated angle is above the threshold $\epsilon_{(\alpha)}$ or $\epsilon_{(\beta)}$ it is not mapped to a input or true angle and set to infinity, thus counted as false detection, e.g. in the example in Fig. 3.27b angles $(\hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)$ are mapped to the corresponding true angles while $\hat{\alpha}_1$ is counted as false detection. Consequently, a measure of the number of false detections is required and given as

$$R^{(\text{false})} = \frac{N^{(\text{false})}}{K}, \tag{3.91}$$

where $N^{(\text{false})}$ is the number of false detected sources defined as

$$N^{(\text{false})} = \sum_{k=1}^{K} \phi_k^{(\text{false})}, \tag{3.92}$$

where $\phi_k^{(\text{false})}$ is either *one* or *zero* according to

$$\phi_k^{(\text{false})} = \begin{cases} 1, & \text{if } \Delta_{(\alpha)} = \emptyset \textbf{ or } \Delta_{(\beta)} = \emptyset \\ 0, & \text{if } \Delta_{(\alpha)} \neq \emptyset \textbf{ and } \Delta_{(\beta)} \neq \emptyset. \end{cases} \tag{3.93}$$

The false detection ratio $R^{(\text{false})}$, that is a threshold-based classification, is taken into account as another key performance indicator (KPI) for evaluation of multiple-source DoA estimation performance because false detections are not taken into account in the CDF of the DoA or position estimation error.

3.3.2. Grouping Algorithm

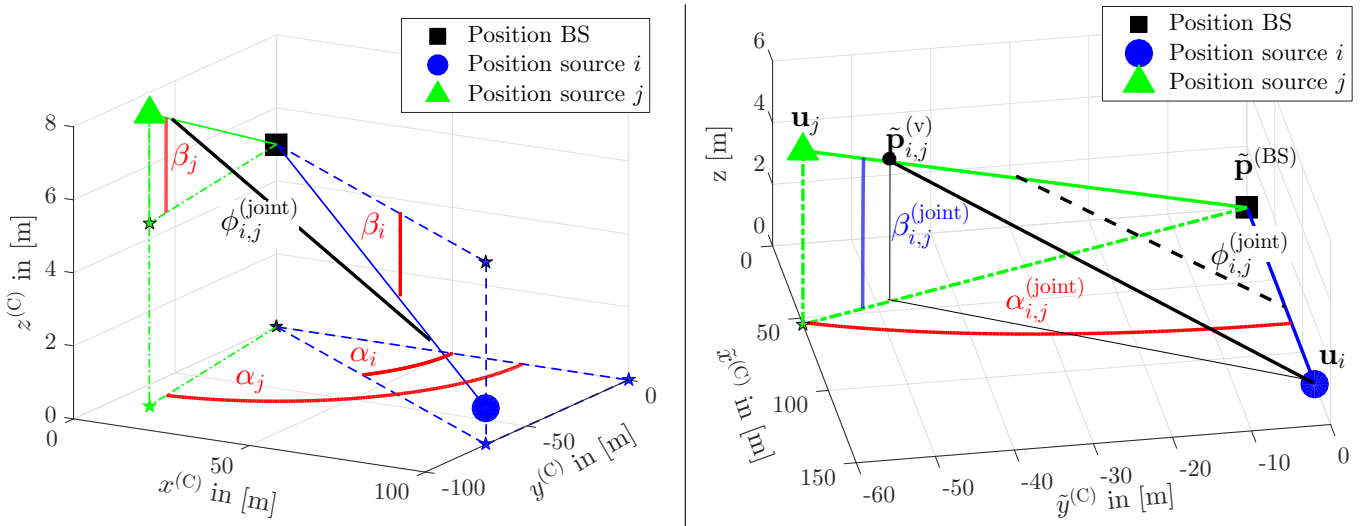
Multiple-source estimation of electromagnetic or acoustic waves has been widely studied and is still in the focus of current research, e.g. with compressed sensing-based approaches [KV96, KKA13, TI13, BLG+15]. However, due to the recent development of planar (large) antenna arrays for 5G mobile communications networks, DoA estimation is enabled and with it the challenge to use network resources as efficiently as possible. With the expected increase in the number of 5G devices compared to LTE [Cis19], these devices can be scheduled for multiple-source DoA estimation, similar to resource scheduling for multiple-user MIMO.

The target and trade-off for multiple-source detection is to estimate as many devices as possible on the same time-frequency resource[11] while keeping increase of the DoA estimation error as low as possible. As pointed out in the previous section, to achieve this, devices have to be

1. separated in the angular domain and
2. the power of their peaks in the power spectrum has to be in a certain range.

In this section the focus is on device selection for multiple-source DoA estimation. Since no literature has been found on this topic, fundamental trade-offs are studied first. The joint scheduling for more than one BS is discussed in Chapter 4 as part of challenges in cellular deployments. Furthermore, the focus is more on practical implications rather then theoretical limits, e.g. the impact from outdated and erroneous DoAs from previous single device DoA estimation. However, perfect a-priori DoA and power information is also considered for comparison. As a lower performance baseline device selection with no DoA knowledge at all is assumed.

[11] Estimation on the same time-frequency resource means that other orthogonalization methods are not considered, e.g. a common approach is orthogonalization in the "code domain" by orthogonal sequences. However, with a finite code length the number of orthogonal sequences is limited and devices with the same non-orthogonal sequences also have to separated by user grouping methods.



(a) Angle $\phi_{i,j}^{(\text{joint})}$ between source i and j.

(b) Transformation of angle $\phi_{i,j}^{(\text{joint})}$ such that $d_i^{(\text{BS})}$ of source i is one the virtual $x^{(\text{C})}$-axis. The BS position is in the origin.

Figure 3.32.: Joint angle between two sources.

Angular Grouping: In angular grouping sources are separated based on their input DoA values from the previous single-source DoA estimation, see Fig. 3.31. According to Section 3.1, $\mathcal{K}$ is the set of available sources in the network considered for DoA estimation. The objective of the angular grouping is to to have $K^{(\xi)} \leq K$ subsets of sources for joint DoA estimation. Thus, a "compression" ratio $\gamma^{(\text{CPR})}$ can be defined by

$$\gamma^{(\text{CPR})} = \frac{K}{K^{(\xi)}}, \tag{3.94}$$

as an indicator on how much resources are saved by the multiple-source estimation. For example, with $K = 64$ sources grouped into $K^{(\xi)} = 8$ subsets, this would result in a compression factor of $\gamma^{(\text{CPR})} = 8$. Note that $\gamma^{(\text{CPR})}$ is bounded by $1 \leq \gamma^{(\text{CPR})} \leq K$ and includes no information about the number of devices per subset. Let $\mathcal{K}_\xi^{(\xi)}$ be the $\xi-$th subset with $K_\xi^{(\xi)} = \left|\mathcal{K}_\xi^{(\xi)}\right|$ sources in it, where the subscript $\xi \in \left\{1, \ldots, K^{(\xi)}\right\}$ and $K = \sum_{\xi=1}^{K^{(\xi)}} K_\xi^{(\xi)}$. If $K^{(\xi)} > 1$, it is considered that each subset of sources $\mathcal{K}_\xi^{(\xi)}$ is estimated on an orthogonal resource, e.g. multiplexed on disjunct time or frequency resources.

It is shown in the previous section that a minimum angular distance between the sources is required. However, the examples therin consider only the horizontal domain and the vertical domain is neglected. Therefore, a joint angular distance metric including horizontal and vertical angular distance is require. In Fig. 3.32a the angle between two sources observed by the BS is denoted by $\phi_{i,j}^{(\text{joint})}$. The objective is that joint angle depends only on the horizontal and elevation angles such that

$$\phi_{i,j}^{(\text{joint})} = \text{f}(\alpha_i, \alpha_j, \beta_i, \beta_j), \tag{3.95}$$

where f is an arbitrary function. In order to derive $\phi_{i,j}^{(\text{joint})}$ the problem is transformed into a coordinate system given in Fig. 3.32b, such that source i is on the $x^{(\text{C})}$-axis with distance $d_i^{(\text{BS})}$ to the BS, such that

$$\mathbf{u}_i = [d_i^{(\text{BS})} \; 0 \; 0]^{\text{T}}, \tag{3.96}$$

where $\mathbf{u}_i$ denotes the position of device i in the new coordinate system. Accordingly, the angles $\alpha_{i,j}^{(\text{joint})} = \alpha_j - \alpha_i$ and $\beta_{i,j}^{(\text{joint})} = \beta_j - \beta_i$ are depicted. The solid black lines in Eq. (3.95) span a right-angled triangle with the LoS connections from the BS to source i and j such that the theorem of Pythagoras and subsequent

relationships can be applied according to

$$\begin{aligned}\left(\mu^{(\mathrm{a})}\right)^2 &= \left([\mathbf{u}_i]_1 - \left[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}\right]_1\right)^2 + \left([\mathbf{u}_i]_2 - \left[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}\right]_2\right)^2 + \left([\mathbf{u}_i]_3 - \left[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}\right]_3\right)^2 \\ \left(\mu^{(\mathrm{b})}\right)^2 &= \left([\mathbf{u}_i]_1 - \left[\tilde{\mathbf{p}}^{(\mathrm{BS})}\right]_1\right)^2 + \left([\mathbf{u}_i]_2 - \left[\tilde{\mathbf{p}}^{(\mathrm{BS})}\right]_2\right)^2 + \left([\mathbf{u}_i]_3 - \left[\tilde{\mathbf{p}}^{(\mathrm{BS})}\right]_3\right)^2 \\ \left(\mu^{(\mathrm{c})}\right)^2 &= \left(\left[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}\right]_1 - \left[\tilde{\mathbf{p}}^{(\mathrm{BS})}\right]_1\right)^2 + \left(\left[\tilde{\mathbf{p}}^{(\mathrm{BS})}\right]_2 - \left[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}\right]_2\right)^2 + \left(\left[\tilde{\mathbf{p}}^{(\mathrm{BS})}\right]_3 - \left[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}\right]_3\right)^2.\end{aligned} \tag{3.97}$$

With this $\phi_{i,j}^{(\mathrm{joint})}$ is obtained by

$$\begin{aligned}\phi_{i,j}^{(\mathrm{joint})} &= \arccos\left(\frac{\mu^{(\mathrm{b})}}{\mu^{(\mathrm{c})}}\right) = \arccos\left(\frac{\mu^{(\mathrm{b})}\sin(\beta_{i,j})}{\left[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}\right]_3}\right) = \arccos\left(\frac{\mu^{(\mathrm{tmp})}\cos(\alpha_{i,j})\sin(\beta_{i,j})}{\left[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}\right]_3}\right) \\ &= \arccos\left(\cos(\alpha_{i,j})\cos(\beta_{i,j})\right),\end{aligned} \tag{3.98}$$

where $\mu^{(\mathrm{tmp})} = \sqrt{\left(\mu^{(\mathrm{a})}\right)^2 + \left[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}\right]_1^2 + \left[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}\right]_2^2}$ and $\frac{\mu^{(\mathrm{tmp})}}{[\tilde{\mathbf{p}}_{i,j}^{(\mathrm{v})}]_3} = \frac{\cos(\beta_{i,j})}{\sin(\beta_{i,j})}$.

Next, Eq. (3.98) is applied to all sources in $\mathcal{K}$ and their joint angular distance to each other is denoted in matrix $\mathbf{\Phi}_{\mathcal{K}}^{(\mathrm{joint})} \in \mathbb{R}^{K\times K}$ defined by

$$\mathbf{\Phi}_{\mathcal{K}}^{(\mathrm{joint})} = \begin{bmatrix} 0 & \phi_{1,2}^{(\mathrm{joint})} & \phi_{1,3}^{(\mathrm{joint})} & \cdots & \phi_{1,K}^{(\mathrm{joint})} \\ \phi_{2,1}^{(\mathrm{joint})} & 0 & \phi_{2,3}^{(\mathrm{joint})} & \cdots & \phi_{2,K}^{(\mathrm{joint})} \\ \phi_{3,1}^{(\mathrm{joint})} & \phi_{3,2}^{(\mathrm{joint})} & 0 & \cdots & \phi_{3,K}^{(\mathrm{joint})} \\ \vdots & \vdots & & \ddots & \vdots \\ \phi_{K,1}^{(\mathrm{joint})} & \phi_{K,2}^{(\mathrm{joint})} & \cdots & \phi_{K,K-1}^{(\mathrm{joint})} & 0 \end{bmatrix}. \tag{3.99}$$

Due to symmetry $\phi_{i,j}^{(\mathrm{joint})} = \phi_{j,i}^{(\mathrm{joint})}$, where $i \in [1,\cdots,K]$ and $j \in [1,\cdots,K]$, all the information are within a strict lower (or upper) triangular matrix. However, for algorithmic design reasons and notational simplicity $\mathbf{\Phi}_{\mathcal{K}}^{(\mathrm{joint})}$ according to Eq. (3.99) is used.

The algorithm for angular grouping is given in Algorithm 2 and requires as input the set of source indices $\mathcal{K}$ and the corresponding angular distance matrix $\mathbf{\Phi}_{\mathcal{K}}^{(\mathrm{joint})}$, which is obtained from the previous single-source DoA estimates, see Fig. 3.31, according to Eq. (3.99). Furthermore, the design parameter $\phi_{(\mathrm{min})}$ is required as an input to Algorithm 2, that is the minimum angular distance between devices in the same group defined as

$$\phi_{(\mathrm{min})} = f_{(\phi)}\hat{\phi}_{(\mathrm{HPBW})}, \tag{3.100}$$

where $f_{(\phi)} \in \mathbb{R}_+$ is a scaling factor and design parameter, and

$$\hat{\phi}_{(\mathrm{HPBW})} = \arccos\left(\cos\left(\hat{\alpha}_{(\mathrm{HPBW})}\right)\cos\left(\hat{\beta}_{(\mathrm{HPBW})}\right)\right) \tag{3.101}$$

is the joint HPBW angle obtained by replacing $\alpha_{i,j}$ and $\beta_{i,j}$ in Eq. (3.98) with $\hat{\alpha}_{(\mathrm{HPBW})}$ and $\hat{\beta}_{(\mathrm{HPBW})}$ according to Eq. (3.85), assuming that the source is located at $\alpha = 0°$ and $\beta = 0°$. Algorithm 2 starts with assigning a random source to a new group $\mathcal{K}_{\xi}^{(\xi)}$. Then the source with the minimum sum-distance to sources already in $\mathcal{K}_{\xi}^{(\xi)}$, out of the subset $\tilde{\mathcal{K}}_{\xi}$ is added. Subset $\tilde{\mathcal{K}}_{\xi}$ contains sources with a distance larger or equal than $\phi_{(\mathrm{min})}$ to sources in $\mathcal{K}_{\xi}^{(\xi)}$.

Algorithm 2 and the impact of design parameter $f_{(\phi)}$ that scales the minimum angular distance is evaluated in Fig. 3.33 and Fig. 3.34 with parameters according to Table 3.10. The design parameter $f_{(\phi)}$ controls the trade-off between the compression ratio $\gamma^{(\mathrm{CPR})}$ and position error, e.g. $f_{(\phi)} \to 0$ result in a single group such that $\gamma^{(\mathrm{CPR})} = K$, and $f_{(\phi)} \to \infty$ results in $\gamma^{(\mathrm{CPR})} = 1$. The position error behaves inverse, the less sources

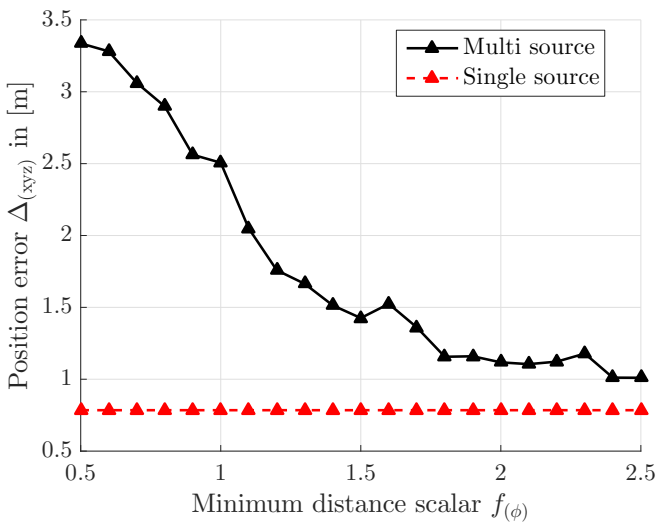
Algorithm 2 Angular source grouping.

Require: $\mathcal{K}, \mathbf{\Phi}_{\mathcal{K}}^{(\text{joint})}, \phi_{(\min)}$
  Initialize $\xi = 1$
  while $\mathcal{K} \neq \emptyset$ **do**
    Get random source $j \in \mathcal{K}$
    Create set $\mathcal{K}_{\xi}^{(\xi)} = \{j\}$
    while $j \neq \emptyset$ **do**
      Find $\tilde{\mathcal{K}}_{\xi}$ sources such that $\left[\mathbf{\Phi}_{\mathcal{K}}^{(\text{joint})}\right]_{\tilde{\mathcal{K}}_{\xi}, \mathcal{K}_{\xi}^{(\xi)}} \geq \phi_{(\min)}$
      Find j corresponding to $\min_{j \in \tilde{\mathcal{K}}_{\xi}} \sum_{k \in \mathcal{K}_{\xi}^{(\xi)}} \left[\mathbf{\Phi}_{\mathcal{K}}^{(\text{joint})}\right]_{j,k}$
      Add j to $\mathcal{K}_{\xi}^{(\xi)}$
    end while
    Remove devices in $\mathcal{K}_{\xi}^{(\xi)}$ from $\mathcal{K}$
    Increase group index: $\xi = \xi + 1$
  end while
  $K^{(\xi)} = \xi$

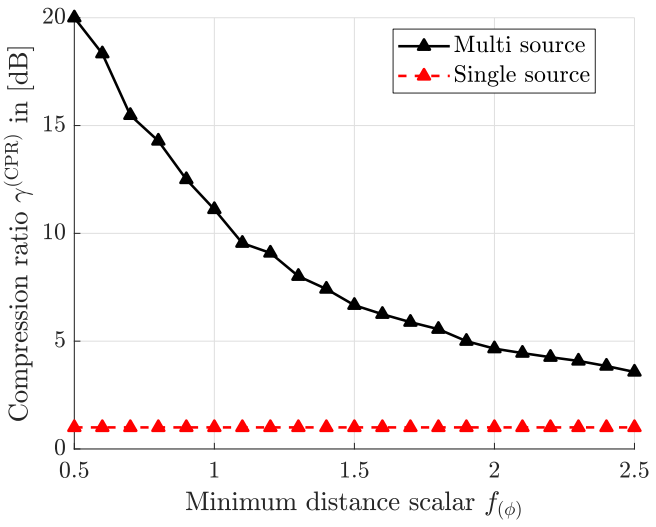
are scheduled to one group the lower the position error. Note that with the parameter setting in Table 3.10, that is a $[8 \times 8]$ UPA, the minimum angular distance between any two sources is $\phi_{(\min)}(f_{(\phi)} = 1) = 17.8°$. For the evaluation of the false detection ratio the horizontal and vertical thresholds are set to $\epsilon_{(\alpha)} = \epsilon_{(\beta)} = 3°$ due to the symmetric antenna structure and to capture large estimation errors. The minimum value of $f_{(\phi)} = 0.5$ corresponds to $\phi_{(\min)} \approx 9°$ satisfying constraints in Eq. (3.87) and Eq. (3.88). The false detection ratio $R^{(\text{false})}$ according to Eq. (3.91) given in Fig. 3.34 is increasing over $f_{(\phi)}$, at first not an intuitive behavior. One can expect the false detection ratio to be decreasing similar to the position error. However, the dominating effect is that with a decreasing compression ratio $\gamma^{(\text{CPR})}$ defined in Eq. (3.94), also the probability decreases that "false" detected peaks are within the threshold of a true angle. Recall that the compression ration $\gamma^{(\text{CPR})}$ is defined by the number of devices divided by the number of groups. This is visualized in Fig. 3.35 where two outcomes of Algorithm 2 are shown, $f_{(\phi)} = 0.5$ on the left-hand side with $K_1 = 48$, and $f_{(\phi)} = 2$ on the right-hand side with $K_2 = 8$. For constant thresholds $\epsilon_{(\alpha)}$ and $\epsilon_{(\beta)}$ the total area around each DoA is larger if $K_1 > K_2$. Consequently, also the probability is larger that an element of the estimated angles is within one of the these areas. Furthermore, the effect of the "noise" floor in the MUSIC power spectrum contributes to the false detection. An example of the MUSIC power spectrum of a group of seven sources, where $N^{(\text{false})} = 3$ of them are false detected, is shown in Fig. 3.36a. The error floor of the MUSIC power spectrum in this scenario is $10\,\text{dB} - 15\,\text{dB}$ and the peaks of the three false estimated sources are below the noise floor. The channel power loss of the sources for $f_{(\phi)} = 0.5$ are given in in Fig. 3.36b comparing true detected and false detections. At the median the difference is ≈ 20 dB which implies that the receive power of the sources has to be taken into account for efficient grouping for multiple-source DoA estimation.

Power Grouping: The false detection rate $R^{(\text{false})}$ in Fig. 3.34 is around 0.3 over the complete range of the minimum distance parameter $f_{(\phi)}$. That is considered as too large for an application and $R^{(\text{false})}$ has to be decreased. Therefore, in the next step the receive power of the sources is taken into account. According to insights from Fig. 3.36 sources within the same group require similar receive power. Therefore, the power-based grouping has to be done before the angular grouping. In order to reuse the angular grouping algorithm and to not further increase implementation complexity a decoupled two-step approach is chosen by:

1. Power grouping
2. Angular grouping per power group



(a) Position error.



(b) Compression ratio $\gamma^{(\mathrm{CPR})}$ according to Eq. (3.94).

Figure 3.33.: Multiple-source estimation performance over design parameter $f_{(\phi)}$ scaling the minimum angular distance between sources according to Eq. (3.100). $\lim f_{(\phi)} \to \infty$ corresponds to single source estimation.

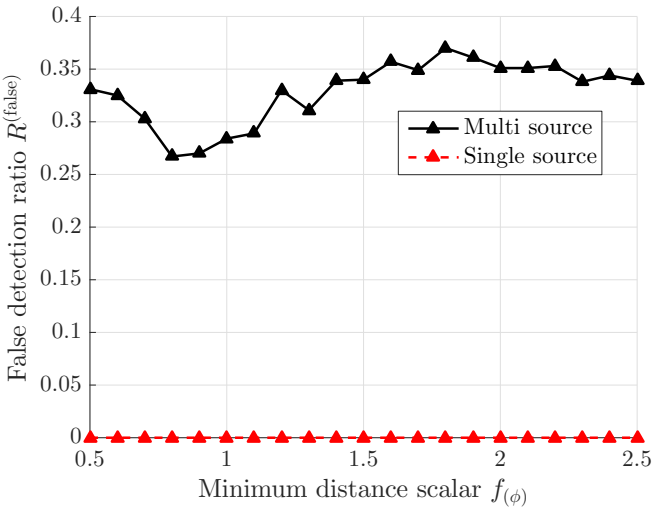
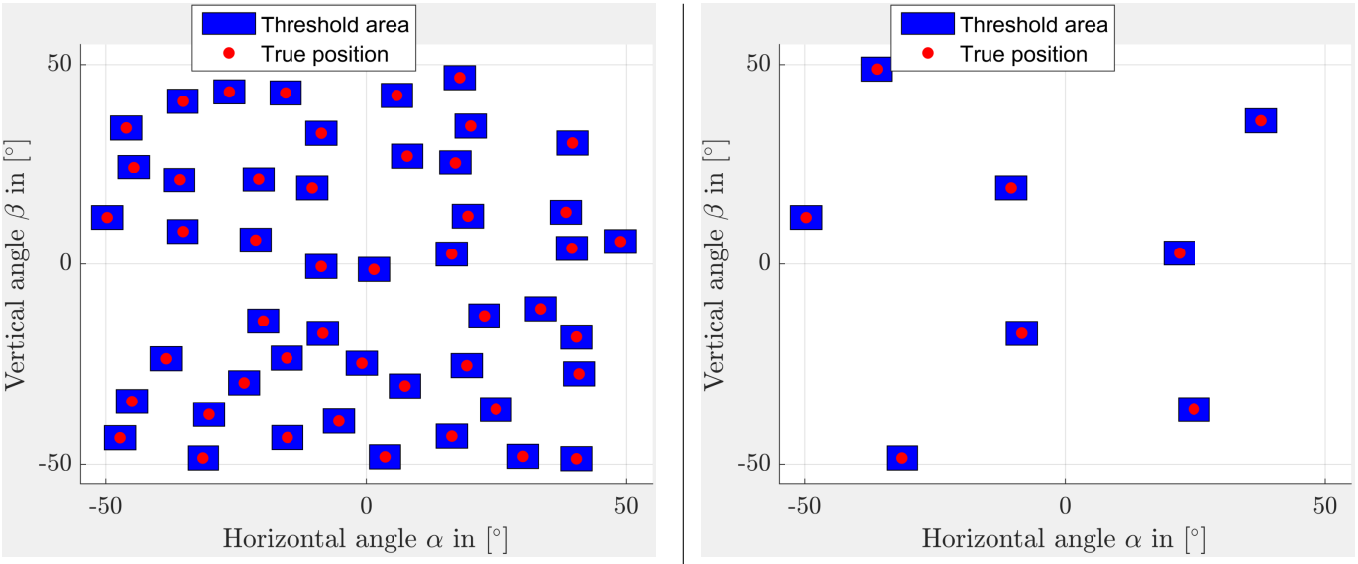


Figure 3.34.: Multiple-source estimation false detection ratio $R^{(\mathrm{false})}$ according to Eq. (3.91) over design parameter $f_{(\phi)}$ scaling the minimum angular distance between sources according to Eq. (3.100). $\lim f_{(\phi)} \to \infty$ corresponds to single source estimation.



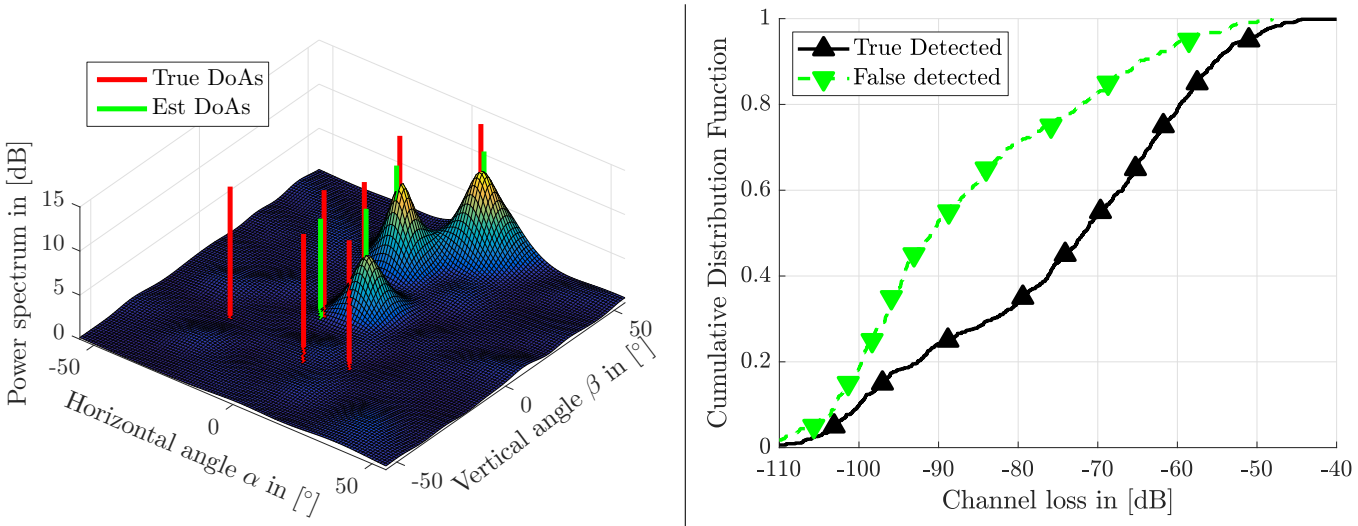
(a) Minimum distance parameter $f_{(\phi)} = 0.5$ resulting in group size $K_1 = 48$.

(b) Minimum distance parameter $f_{(\phi)} = 2$ resulting in group size $K_2 = 8$.

Figure 3.35.: Visualization of selected groups with Algorithm 2 and thresholds $\epsilon_{(\alpha)} = \epsilon_{(\beta)} = 3°$.

Table 3.10.: Simulation parameters for multiple-source DoA estimation.

Parameter	**Value**
Simulation type	Monte Carlo 2500 realizations
Channel model	QuaDRiGa version 1.4 [JRBT14]
Scenario	Urban macro LoS [JRBT14]
Center frequency	3.75 GHz
Number of multi-path components $L_{(\mathrm{MPC})}$	8
Rician K-Factor $K^{(\mathrm{R})}$	10 dB
BS antenna distribution	UPA
Number of antenna elements N at BS	64
Number of elements in $x^{(\mathrm{C})}$-direction $N_{(\alpha)}$	8
Number of elements in $z^{(\mathrm{C})}$-direction $N_{(\beta)}$	8
Antenna element spacing in $x^{(\mathrm{C})}$- and $z^{(\mathrm{C})}$-direction	$\lambda_{(\mathrm{c})}/2$
Antenna element type	Isotrop
BS height z_1	0 m
Sequence length, number of samples $N^{(\mathrm{seq})}$	10
Target horizontal angular quantization $q_S^{(\alpha)}$	0.5°
Target vertical angular quantization $q_S^{(\beta)}$	0.5°
Horizontal search range $\boldsymbol{\alpha}^{(\mathrm{f})}$	$[-50\ 50]°$
Vertical search range $\boldsymbol{\beta}^{(\mathrm{f})}$	$[-50\ 50]°$
Horizontal angles of devices α	random i.i.d. in $[-50, 50]°$
Vertical angles of devices β	random i.i.d. in $[-50, 50]°$
Normalization value $g^{(\mathrm{norm})}$ Eq. (3.9)	1
Distance BS-Device $d_i^{(\mathrm{BS})}$	random i.i.d. in $[30, 250]$ m
Thresholds $\epsilon_{(\alpha)}$ and $\epsilon_{(\beta)}$ for false detection	3° and 3°



(a) MUSIC power spectrum of seven sources with three false detections.

(b) Average power loss comparing true and false detected devices.

Figure 3.36.: Performance evaluation of false detection.

This means that devices require a similar path-loss to be in the same power group, which is a strong constraint. However, on the other hand, ignoring the receive power means that the false detection ratio is around $\frac{1}{3}$ even for large angular distances. Thinking of a service where only a third of the positioning requests are served is hard to sell. Therefore the decoupled two-step approach is necessary. In doing so the set of sources $\mathcal{K}$ is divided into $K^{(\eta)}$ subsets as a first step such that

$$\mathcal{K} = \bigcup_{\eta=1}^{K^{(\eta)}} \mathcal{K}_{\eta}^{(\eta)}, \tag{3.102}$$

where the subscript $\eta \in \left\{1, \ldots, K^{(\eta)}\right\}$ indicates the respective subset. Afterward, in the second step, Algorithm 2 is applied $K^{(\eta)}$ times with input $\mathcal{K}_{\eta}^{(\eta)}$ instead of $\mathcal{K}$. A flow chart of the two-step source grouping approach, along with the required input and system design control parameter, is provided in Fig. 3.44. Consequently, the total number of subsets for joint DoA estimation, taking into account power grouping, is $K^{(\xi\eta)} \leq K^{(\xi)}$ according to

$$K^{(\xi\eta)} = \sum_{\eta=1}^{K^{(\eta)}} \sum_{\xi=1}^{K_{\eta}^{(\xi)}} \left| \mathcal{K}_{\xi,\eta}^{(\xi\eta)} \right|, \tag{3.103}$$

where $K_{\eta}^{(\xi)}$ is the number of angular groups per power subset $\mathcal{K}_{\eta}^{(\eta)}$, an output from Algorithm 2. Due to the approximately linear distribution of receive power values (in dB) in Fig. 3.36b, a linear quantization of the power range (in dB) is selected as straightforward solution. For this purpose, the set of receiver power values of the K sources at the BS, according to Eq. (3.1), is defined as

$$\mathcal{K}^{(\mathrm{RxP})} = \left\{ \mathbf{y}_1^{\mathrm{H}} \mathbf{y}_1, \ldots, \mathbf{y}_K^{\mathrm{H}} \mathbf{y}_K \right\}. \tag{3.104}$$

$\mathcal{K}^{(\mathrm{RxP})}$ and the design parameter $\mathcal{K}_{\xi,\eta}^{(\xi\eta)}$ are inputs to Algorithm 3. The output are the sets of "power groups" denoted by $\mathcal{K}_p$, where the receive power difference of any two sources within a single group is $\leq P^{(\epsilon)}$. Therefore, the design parameter $P^{(\epsilon)}$ can be interpreted as a power range. Furthermore, the minimum and maximum power in Algorithm 3 is given by

$$\begin{aligned} P^{(\min)} &= \min\left(\mathcal{K}^{(\mathrm{RxP})}\right) \\ P^{(\max)} &= \max\left(\mathcal{K}^{(\mathrm{RxP})}\right). \end{aligned} \tag{3.105}$$

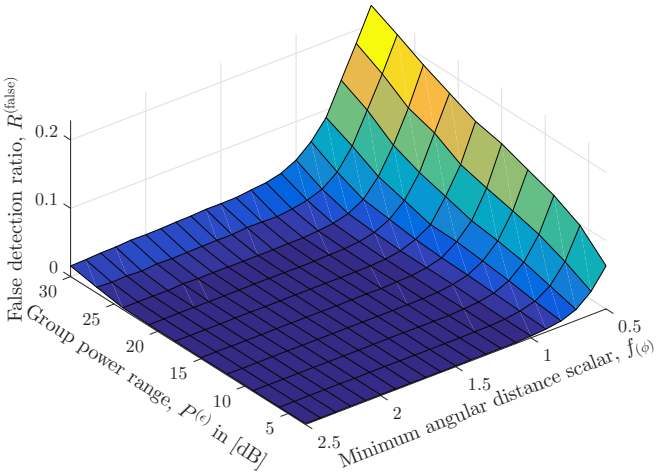


Figure 3.37.: False detection ratio $R^{(\text{false})}$ according to Eq. (3.91) of the two-step decoupled source grouping. Design parameters are $P^{(\epsilon)}$ for first step power grouping, and $f_{(\phi)}$ for second step angular grouping.

One observation while implementing Eq. (3.105) and testing Algorithm 3 was that empty groups can exist where no source has a receive power within two adjacent elements in $\mathbf{p}^{(\text{thr})}$. Therefore, the last statement in Algorithm 3 is added to remove empty group sets.

Algorithm 3 Power grouping of sources.

Require: $\mathcal{K}^{(\text{RxP})}, P^{(\epsilon)}, P^{(\text{min})}, P^{(\text{max})}$

$\mathbf{p}^{(\text{thr})} = \left[P^{(\text{max})} : -P^{(\epsilon)} : \left(P^{(\text{min})} - P^{(\epsilon)}\right)\right]^{\text{T}}$

for $\eta = 2 : \left|\mathbf{p}^{(\text{thr})}\right|$ **do**

Create $\mathcal{K}^{(\eta)}_{\eta-1}$

Find $\tilde{\mathcal{K}} \subseteq \mathcal{K}^{(\text{RxP})}$ that satisfies: $\left[\mathbf{p}^{(\text{thr})}\right]_{\eta} \leq \tilde{\mathcal{K}} < \left[\mathbf{p}^{(\text{thr})}\right]_{\eta-1}$

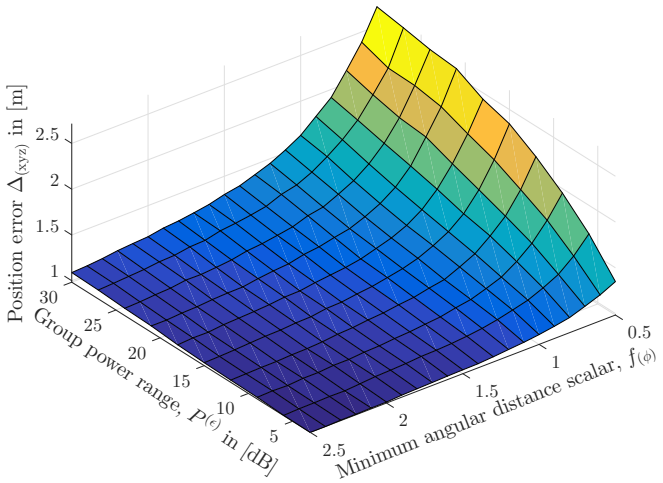
Add $\tilde{\mathcal{K}}$ to $\mathcal{K}_{\eta-1}$

end for

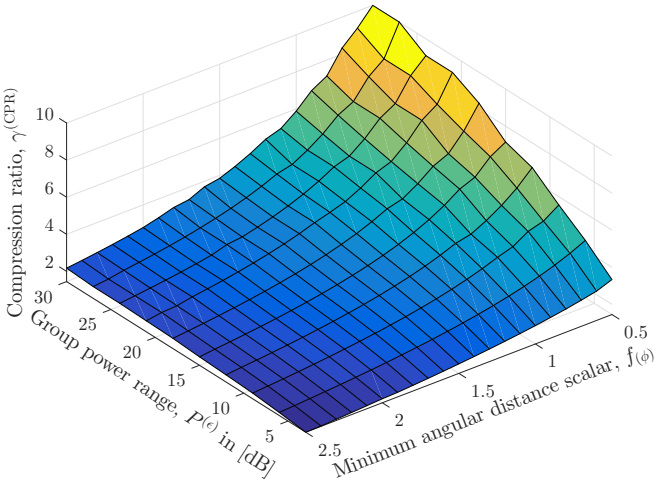
Remove empty group sets

The target of the additional power grouping step is to reduce the false detection ratio of 0.3 achieved by using only angular grouping. Fig. 3.37 shows the $R^{(\text{false})}$ of the two-step source grouping as a 3D surface plot over the parameter space $P^{(\epsilon)}$, the power range parameter for first step power grouping, and $f_{(\phi)}$, the scaling parameter for second step angular grouping. The effectiveness of the power grouping is demonstrated by $R^{(\text{false})} \leq 1\,\%$ for most of the parameter space, e.g. $P^{(\epsilon)} \leq 24\,\text{dB}$ or $f_{(\phi)} \leq 1$. This reduction of false detection is due to additional power grouping of sources before the angular grouping. The price that is payed for the reduces false detection is a reduction of the compression ratio $\gamma^{(\text{CPR})}$, see Fig. 3.38b. In other words, the reduction of $P^{(\epsilon)}$ results in a lower compression ration, e.g. for $R^{(\text{false})} \leq 1\,\%$, the largest achieved compression ratio is $\gamma^{(\text{CPR})} \approx 6$ for $P^{(\epsilon)} = 24\,\text{dB}$ and $f_{(\phi)} = 1$. On the other hand, it is clear from the previous results of only angular grouping, that the position error performance is inverse to the compression ratio $\gamma^{(\text{CPR})}$, see Fig. 3.38a.

Finally, the effectiveness of the developed two-step grouping scheme for multiple-source DoA estimation is evaluated with respect to practical constraints. In the previous results the ideal assumptions of perfect (true) DoAs is used twice, first in the angular grouping Algorithm 2, and second for false detection in Eq. (3.90). In contrast to this, in a real system only estimated DoAs are available. In order to perform a systematic analysis the horizontal and vertical DoA estimation errors are modeled by a parametrized probability distribution. In Fig. 3.39a the CDF of the horizontal DoA error $\Delta_{(\alpha)}$ is given for single source



(a) Position error in, $\Delta_{(\mathrm{xyz})}$ excluding false detected sources.



(b) Compression ration $\gamma^{(\mathrm{CPR})}$ according to Eq. (3.94) with initial number of sources $K = 100$.

Figure 3.38.: Performance evaluation of the two-step decoupled source grouping. Design parameters are $P^{(\epsilon)}$ for first step power grouping, and $f_{(\phi)}$ for second step angular grouping.

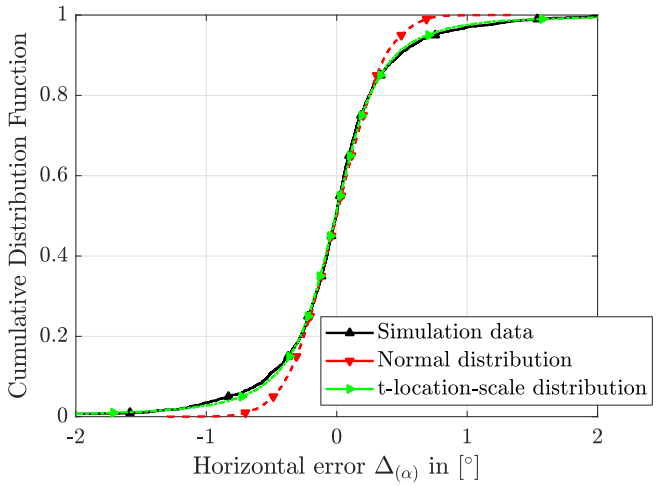
estimation with parameters according to Table 3.10. Additionally to the numerical values obtained by DoA estimation, a normal probability distribution $\mathcal{N}\left(0, \sigma^2_{(\phi)}\right)$ with $\sigma_{(\phi)} = 0.3$ is given as an approximation in Fig. 3.39a. It can be observed that the tails of the data set are not reflected by a normal distribution. Such tails can be approximated with a t-location-scale distribution also referred to as non-standardized Student's t-distribution with the probability distribution function given by

$$\mathcal{P}^{(\mathrm{t})}\left(x^{(\mathrm{t})} \mid \nu^{(\mathrm{t})}, \mu^{(\mathrm{t})}, \sigma^{(\mathrm{t})}\right) = \frac{\Gamma(\frac{\nu^{(\mathrm{t})}+1}{2})}{\Gamma(\frac{\nu^{(\mathrm{t})}}{2})\sqrt{\pi\nu^{(\mathrm{t})}}\sigma^{(\mathrm{t})}} \left(1 + \frac{1}{\nu^{(\mathrm{t})}}\left(\frac{x^{(\mathrm{t})} - \mu^{(\mathrm{t})}}{\sigma^{(\mathrm{t})}}\right)^2\right)^{-\frac{\nu^{(\mathrm{t})}+1}{2}}, \tag{3.106}$$

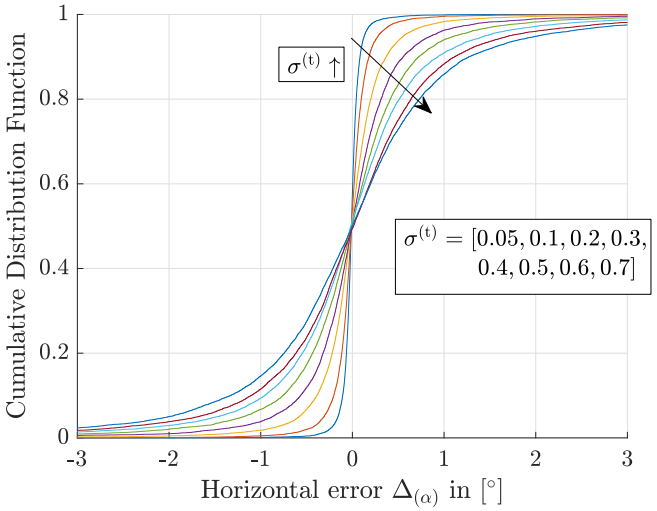
where $\nu^{(\mathrm{t})} \in \mathbb{R}_+$ is a shape parameter, $\sigma^{(\mathrm{t})} \in \mathbb{R}_+$ is a scale parameter, and $\mu^{(\mathrm{t})} \in \mathbb{R}$ in $[^\circ]$ is the location parameter. Note that $\sigma^{(\mathrm{t})}$ sets the overall scaling of the distribution and does "not" correspond to the standard deviation of the scaled "t" distribution, nor is it the standard deviation of the underlying normal distribution. The t-location-scale distribution in Fig. 3.39a is obtained with $\nu^{(\mathrm{t})} = 2,\ \sigma^{(\mathrm{t})} = 0.25,\ \mu^{(\mathrm{t})} = 0^\circ$ and fits very well the data set from the single source DoA estimation. Consequently, the t-location-scale distribution is selected for approximation of $\hat{\alpha}$ and $\hat{\beta}$. The variation of $\sigma^{(\mathrm{t})}$ is shown in Fig. 3.39b and further used as parameter for performance evaluation on the impact of estimated DoA as input for the proposed source grouping. Thus, the estimated DoAs for this evaluation are obtained by

$$\begin{aligned} \hat{\alpha} &= \alpha + \mathcal{P}^{(\mathrm{t})}\left(x^{(\mathrm{t})} \mid \nu^{(\mathrm{t})} = 2, \mu^{(\mathrm{t})} = 0^\circ, \sigma^{(\mathrm{t})}\right) \\ \hat{\beta} &= \beta + \mathcal{P}^{(\mathrm{t})}\left(x^{(\mathrm{t})} \mid \nu^{(\mathrm{t})} = 2, \mu^{(\mathrm{t})} = 0^\circ, \sigma^{(\mathrm{t})}\right). \end{aligned} \tag{3.107}$$

For the performance evaluation of $\sigma^{(\mathrm{t})}$ and comparison with previous results, the working point is set to a compression ratio $\gamma^{(\mathrm{CPR})} \approx 4$, e.g. $P^{(\epsilon)} = 18\,\mathrm{dB}$ and $f_{(\phi)} = 1.2$ with only moderate degratation of the position error compared to single source estimation and a false detection ratio $R^{(\mathrm{false})} \approx 0$. In Fig. 3.40, Fig. 3.41, and Fig. B.9 the position error, the false detection ratio, and the compression ratio are shown, respectively, over the variation of $\sigma^{(\mathrm{t})}$ as a scale parameter of the probability distribution. The robustness of the proposed algorithms against DoA estimation errors for input to the angular grouping algorithm is shown by the approximately constant position error, which increases only in the cm range. The false detection ratio increases approximately linear with larger DoA estimation errors, however for the selected working point it is $< 0.01\,\%$ over the complete range $\sigma^{(\mathrm{t})}$. The compression ration $\gamma^{(\mathrm{CPR})}$ shown in the annex is approximately constant over $\sigma^{(\mathrm{t})}$. Due to an increased $\sigma^{(\mathrm{t})}$ the overall angular range of the input DoAs



(a) CDF of horizontal DoA estimation error $\hat{\alpha}$ and approximations by normal distribution $\mathcal{N}(0, 0.9)$ and t-location-scale distribution according to Eq. (3.106). The t-location-scale parameters are $\nu^{(\mathrm{t})} = 2$, $\sigma^{(\mathrm{t})} = 0.25$, $\mu^{(\mathrm{t})} = 0$. Other parameters according to Table 3.10.



(b) Variation of parameter $\sigma^{(\mathrm{t})}$ in t-location-scale distribution Eq. (3.106).

Figure 3.39.: Parametrization of DoA estimation error by probability functions.

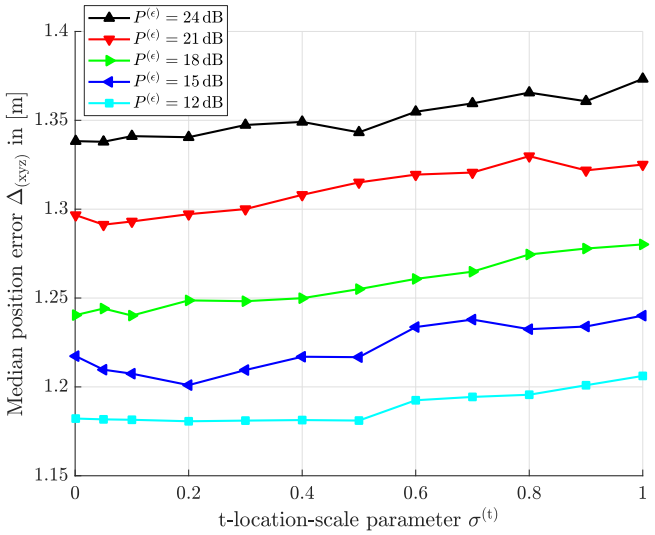
is increased by the tails of the t location-scale distribution. Consequently, in average more sources can be grouped together for a given minimum angular distance.

Finally, the performance of the proposed multiple-source angular grouping is compared with the baseline of random grouping, where the maximum number of sources is given as an input parameter $K^{(\xi,\mathrm{random})}$. Similar to angular grouping this is performed as a two step approach:

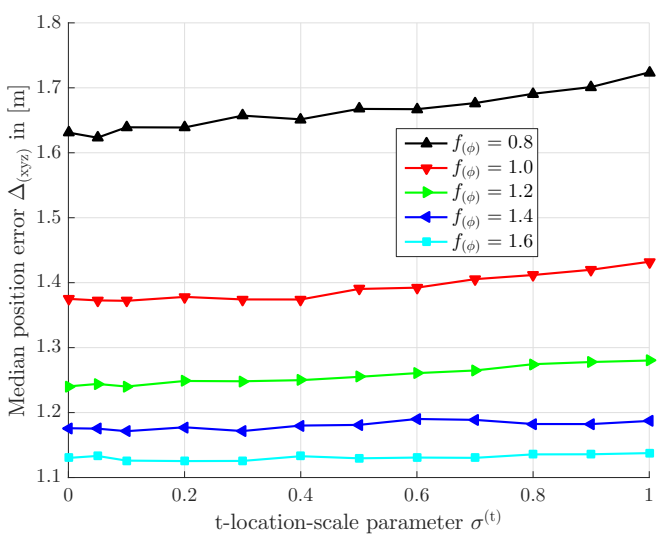
1. Power grouping
2. Per power group: random grouping

Therefore, the compression ratio depends not only on the number of maximum simultaneously randomly grouped sources, but also on how many sources are found per power group. Fig. 3.42a shows the position error as a surface plot over the number of maximum random grouped sources and the power grouping input $P^{(\epsilon)}$. There is an approximately constant position error up to $P^{(\epsilon)} = 60\,\mathrm{dB}$, then the performance decreases. With the number of grouped devices the position error increases approximately linear. The compression ratio $\gamma^{(\mathrm{CPR})}$ in Fig. 3.43 also scales linearly over the number of randomly grouped, but not over $P^{(\epsilon)}$ where $\gamma^{(\mathrm{CPR})}$ is constant for $P^{(\epsilon)} > 15\,\mathrm{dB}$ and decreases non-linear in $P^{(\epsilon)} \leq 15\,\mathrm{dB}$. For random source selection, the minimum false detection ratio in Fig. 3.42b is achieved with the lowest $P^{(\epsilon)}$ where $R^{(\mathrm{false})}_{(\mathrm{random})} > 3\,\% > R^{(\mathrm{false})}_{(\mathrm{angular})}$ with respect to the working point $P^{(\epsilon)} = 18\,\mathrm{dB}$ and $f_{(\phi)} = 1.2$ for angular grouping.

In order to compare the random grouping directly with the angular-based grouping, parameters that result in similar compression ratios are obtained from Fig. 3.43 and Fig. 3.38. Then, the false detection ratio and position error achieved with these parameters are compared in Table 3.11. An outcome of this table is that the position error of random grouping and angular grouping is in the same regime. This is caused by the re-movement of large position errors from the statistic by false detection according to Eq. (3.90). Therefore, the false detection ratio of random grouping is significantly higher compared to angular grouping. Only by selecting a combination of small angular separation and large power range for angular grouping, e.g. $f_{(\phi)} = 1$ $P^{(\epsilon)} = 24\,\mathrm{dB}$, the position error and false detection is larger compared to random grouping.

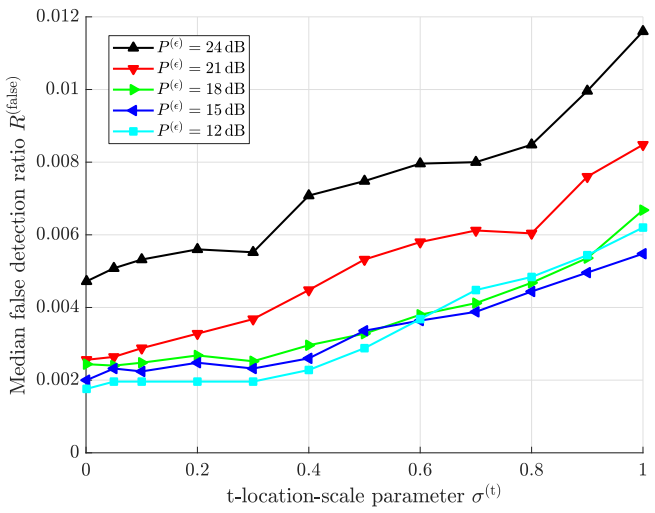


(a) Variation of design parameter $P^{(\epsilon)}$ for first step power grouping algorithm.

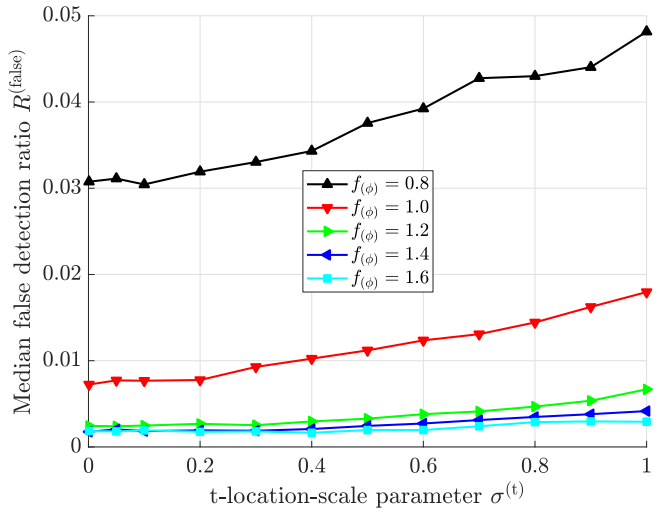


(b) Variation of design parameter $f_{(\phi)}$ for second step angular grouping algorithm.

Figure 3.40.: Position error evaluation for DoA estimation error approximation with t-location-scale probability distribution. Variation of grouping-algorithm input parameters.

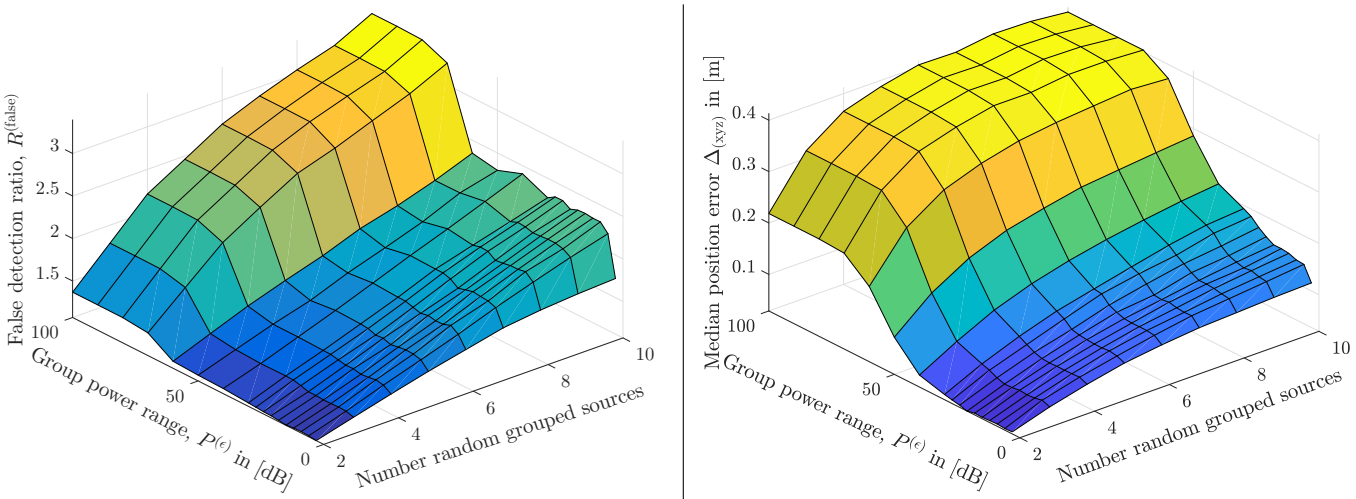


(a) Variation for first step power grouping algorithm design parameter $P^{(\epsilon)}$.



(b) Variation for second step angular grouping algorithm design parameter $f_{(\phi)}$.

Figure 3.41.: False detection evaluation for DoA estimation error approximation with t location-scale probability distribution. Variation of grouping-algorithm input parameters.



(a) Median position error $\Delta_{(\text{xyz})}$.

(b) Median false detection ratio $R^{(\text{false})}$.

Figure 3.42.: Performance of random grouping, surface plots over the number of maximum simultaneously grouped devices and $P^{(\epsilon)}$ as input to power grouping.

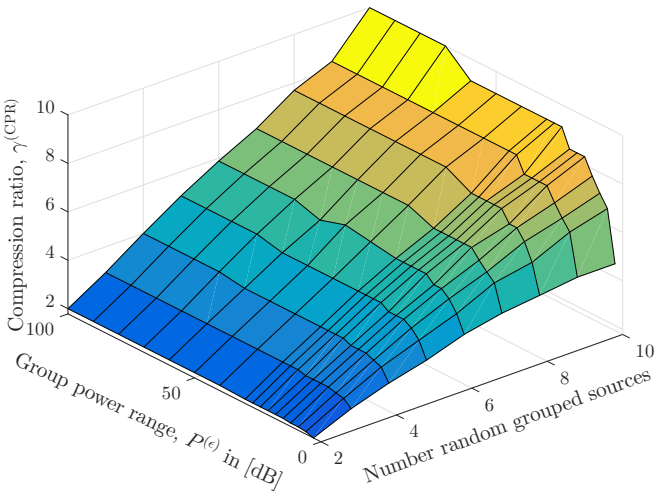


Figure 3.43.: Performance of random grouping on compression ratio, surface plot over the number of maximum simultaneously grouped devices and $P^{(\epsilon)}$ as input to power grouping.

Table 3.11.: Performance comparison of angular and random grouping by compression ratios $\gamma^{(\mathrm{CPR})}$. The input parameters to achieve these compression ratios are jointly listed with the achieved position error and false detection ratio.

Compression Ratio	Parameters Angular Grouping		Parameters Random Grouping		Position Error Angular	Position Error Random	False Detection Ratio Angular	False Detection Ratio Random
$\gamma^{(\mathrm{CPR})}$	$P^{(\epsilon)}$ in [dB]	$f_{(\phi)}$	$P^{(\epsilon)}$in [dB]	$K^{(\xi,\mathrm{random})}$	$\Delta_{(\mathrm{xyz})}$ in [m]	$\Delta_{(\mathrm{xyz})}$ in [m]	$R^{(\mathrm{false})}_{(\mathrm{angular})}$	$R^{(\mathrm{false})}_{(\mathrm{random})}$
2	3	1.4	15	2	1.04	1.07	<1 %	3 %
2	6	1.9	30	2	1.03	1.11	<1 %	4 %
2	12	2.3	40	2	1.02	1.13	<1 %	5 %
2	18	2.5	50	2	1.03	1.14	<1 %	10 %
3	3	0.6	3	4	1.27	1.31	3 %	7 %
3	9	1.3	24	3	1.11	1.25	<1 %	6 %
3	18	1.7	30	3	1.11	1.27	<1 %	7 %
3	27	1.8	50	3	1.18	1.28	<1 %	16 %
4	9	0.9	3	7	1.29	1.61	<1 %	11 %
4	21	1.3	6	5	1.25	1.51	<1 %	10 %
4	27	1.4	70	4	1.29	2.09	1 %	36 %
5	9	0.6	-	-	1.72	-	7 %	
5	15	0.9	9	6	1.42	1.71	1 %	11 %
5	24	1.1	70	5	1.41	2.35	<1 %	38 %
6.25	18	0.7	6	9	1.85	2.09	6 %	14 %
6.25	21	0.8	9	8	1.71	2.03	4 %	13 %
6.25	27	0.9	15	7	1.70	1.86	4 %	13 %
7.7	21	0.6	15	9	2.23	2.16	12 %	14 %
7.7	27	0.7	40	8	2.06	2.07	12 %	20 %
9.1	24	0.5	24	10	2.61	2.33	19 %	17 %
9.1	30	0.6	60	10	2.39	2.35	18 %	37 %
10	27	0.5	70	10	2.66	3.26	21 %	40 %

3.3.3. Conclusions

In this section, grouping for multiple-source estimation is studied with focus on the trade-off to group as many sources as possible while keeping the estimation error from DoA estimation as low as possible. In contrast to state-of-the-art literature, where DoA estimation algorithms are only studied for artificially placed sources, in this thesis a cellular system with many uniform random located sources is assumed. In this section insights on challenges and fundamental trade-offs for source grouping with the aim of multiple-source DoA estimation are obtained with the following new contributions:

- A novel two step grouping algorithm is proposed, taking into account the receive power between sources and the angular distance to each other, where the trade-off between estimation accuracy and the simultaneously estimated source can be controlled by a power and angular system design parameters, see Fig. 3.44.
- An angular distance metric is derived taking into account azimuth and elevation angles and the input scaling factor can be used as system design parameter.
- The false detection ratio, depending on the false detection threshold, as an important KPI for any multiple-source DoA estimation is introduced.
- DoA estimation errors can be approximated by t-location-scale probability distributions for parametrized analysis of multiple-source grouping algorithms.

Furthermore, it is shown that using only angular grouping, ignoring the receive power of sources, results in large probability of false detection, which can interpreted as large estimation errors. Therefore, a power grouping before the angular grouping is recommended and shown in this chapter to decrease the false detection ratio below 1 %. Also the trade-off between position error loss and the number of simultaneously estimated sources is investigated. Finally, it is shown that the proposed angular grouping algorithm is robust with respect to input DoA estimation errors and thus suitable for practical systems.

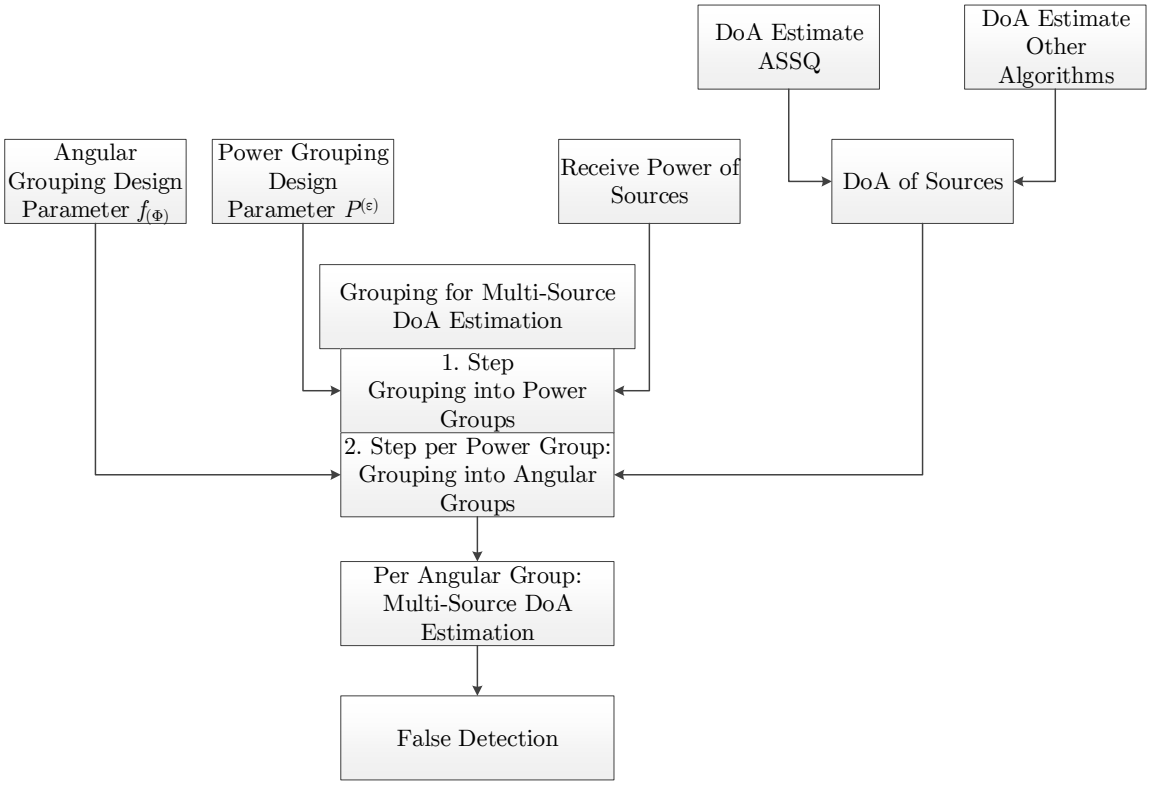


Figure 3.44.: Block diagram of multiple-source grouping algorithms with required input and design parameters.

4. Challenges in Cellular Deployments

While Chapter 2 covers downlink transmission and Chapter 3 uplink direction of arrival (DoA) estimation, both assume a single base station (BS) scenario. Consequently, no uncoordinated inter-sector[1] interference is considered in Chapter 2. However, uncoordinated inter-sector interference is known to limit downlink sum throughput in cellular networks. In contrast to interference limited, the single BS is noise limited. Considering multiple BSs is often referred to as "system level evaluation". These system level simulations are required because algorithms that improve key performance indicators (KPIs) in a noise limited single BS scenario do not necessarily show the same gains in an interference-limited scenario [KRTT13, KTH14b]. Therefore, Section 4.1 investigates the impact from inter-sector interference and implications on findings from Chapter 2 for massive multiple-input multiple-output (MIMO) multiple-user scheduling and precoding are discussed. Furthermore, downlink interference experienced at users cannot be estimated in the uplink, therefore, in time division duplex (TDD) and frequency division duplex (FDD) systems, feedback is required to obtain knowledge about the interference situation of mobile users at the BS. Therefore, Section 4.1.2 proposes an additional feedback based on findings in Section 4.1 to compensate the inter-sector interference loss.

Chapter 3 shows that massive MIMO enables high accurate DoA estimation and position errors below 1 m in a wide parameter range are achieved. However, the assumption of a single BS in Chapter 3 means that in addition to the DoA estimate a different measurement is required to obtain a position estimate of a mobile user. Therefore, in Chapter 3 the distance between the BS and mobile user is assumed to be known, e.g. from time of flight (ToF) measurement as discussed in Section 1.2.1. Yet, the accuracy of distance measurements based on ToF is limited by network synchronization. According to [Fis14] in Section 4.2.2 the smallest time unit in Long Term Evolution (LTE) is ≈ 32 ns which corresponds to a distance of 9.8 m by the time of flight with the speed of light. This means any time of flight measurement is limited in accuracy by this smallest time unit and a corresponding position accuracy below 1 m is not possible in LTE. Hence, the cellular challenge treated in Section 4.2 is to obtain the position estimated only based on DoA estimates. For this task at least two BSs with a DoA estimate from the same mobile user are required. Due to noise these two estimated DoAs may not have an intersection. Thus, Section 4.2.1 presents and evaluates a method to estimate a position based on noisy DoA estimates at two different BSs from the same mobile user. Thereby, a typical cellular deployment for the two BSs is assumed resulting in regions with large position errors caused by dilution of precision (DoP). DoP means that position errors are caused by the geometry of the measurement, here these are the positions of the BSs and the position of the mobile user. Even measurements with high precision, in this thesis low DoA errors, can result to large position errors, hence the term dilution of precision (DoP). The compensation of the DoP effect is studied in Section 4.2.2. The cellular challenges treated in this chapter and their corresponding links to the previous chapters are listed below:

- Section 4.1.1 shows the impact from unknown inter-sector interference on scheduling and precoding schemes from Chapter 2, a challenge for multiple user spatial multiplexing with massive MIMO in cellular deployments.
- Section 4.1.2 presents and evaluates compensation of inter-sector interference.
- Section 4.2.1 presents a positioning algorithm based on two DoA estimates from BSs in a cellular network deployment. Thereby, DoP is identified as a challenge in cellular deployments for DoA based positioning.
- Section 4.2.2 presents and evaluates compensation of DoP.

[1] Sector and cell are used interchangeable in this thesis.

4.1. Downlink Inter-Cell Interference in Multiple-User MIMO

In this section the single BS scenario from Chapter 2 is extended to multiple BSs transmitting on the same time-frequency resource, a typical assumption for evaluation of cellular networks [3GP16d]. Therefore, cellular networks are usually interference limited [ACH07, ZLD+14]. This means that achieved sum throughput per BS in a multiple-cell scenario is lower than the sum throughput of a BS alone in the scenario. System level simulations are used to answer the questions if new techniques, that have been shown gains of some KPIs for a single BS, improve the same KPIs throughout the network at all BSs. One toy example is the following: Increasing the transmit power at BS l, the signal to interference and noise ratio (SINR) of users $\tilde{\mathcal{K}}_l$ connected to BSs l increases according to (2.3). Nevertheless, if other BSs in the network $l' \neq l$ also increase their transmit power by the same amount as BS l, the SINR is approximately the same as without power increase[2]. Thus, increasing transmit power in a cellular network is no solution to increase spectral efficiency.

Each generation of mobile networks from second generation (2G) to fifth generation (5G) networks came with a higher cell densification compared to the predecessor [RF10, ZLD+14]. With higher density of BSs more "interaction" exists between them and system level evaluations taking into account more than a single BS become more relevant. Cell densification can be interpreted as a higher re-use factor of spectrum over space thus increasing the spectral efficiency per area. On the other hand, this also increases interference-limited areas called cell boarders. Research on interference management includes, among other things, interference coordination by time - frequency orthogonalization [ZLD+14], coordinated multi-point (CoMP) transmission [Thi13, KSHT15], or spatial interference coordination [ZM11, DKT13, KTC15a, KTC15b]. Independent of the coordination scheme, in downlink transmission, the unknown interference situation at the user causes loss in spectral efficiency, due to sub-optimal user selection and precoder design, see [KRTT13, KTH14b]. In this section, first the impact from uncoordinated interference is investigated that comes from the coexistence of other massive MIMO BSs or non-massive MIMO BSs in cellular networks. Then, techniques that take into account the interference situation of the mobile users for user selection and precoding are derived and evaluated.

4.1.1. Multiple-Sector Scenarios

For the numerical evaluations in this section two multiple-sector scenarios are considered, labeled as "homogeneous" and "heterogeneous". In the homogeneous scenario all L BSs are equipped with the same antenna arrays such that

$$N_{(\alpha),l} \times N_{(\beta),l} = N_{(\alpha)} \times N_{(\beta)}, \forall l \in \mathcal{L}. \tag{4.1}$$

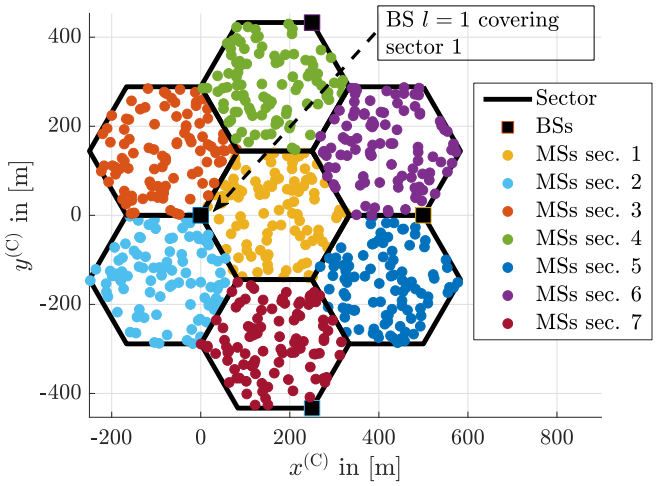
In order words, the homogeneous scenario represent a network, where all BSs are equipped with massive MIMOs arrays. In the heterogeneous scenario, the antenna array of BS $l = 1$ has more antennas than other BSs such that

$$N_{(\alpha),l} \times N_{(\beta),l} = \begin{cases} N_{(\alpha)} \times N_{(\beta)}, & \text{if } l = 1, \\ \tilde{N}_{(\alpha)} \times \tilde{N}_{(\beta)}, & \text{if } l \neq 1, \forall l \in \mathcal{L}, \end{cases} \tag{4.2}$$

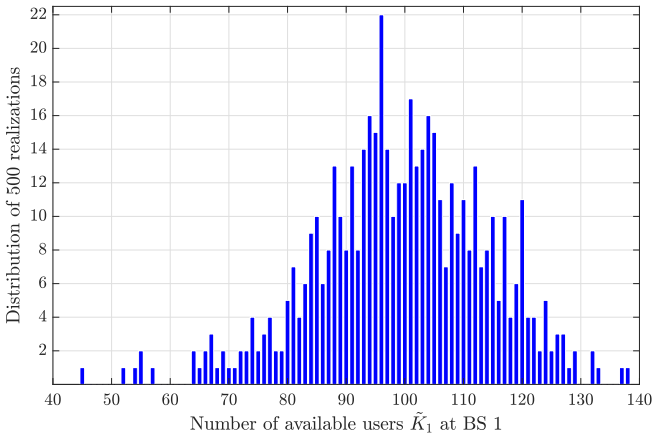
where $\tilde{N}_{(\alpha)} < N_{(\alpha)}$ and $\tilde{N}_{(\beta)} < N_{(\beta)}$ denote the number of rows and columns of the non-massive MIMO uniform planar arrays (UPAs). In other words, the heterogeneous scenario represents a network where the massive MIMO BS is put in to a network of "conventional" or non-massive MIMO BSs. Such a scenario is possible if massive MIMO BSs are deployed only in specific hot-spot sectors with high traffic, e.g. crowded places in cities.

For evaluation of multiple sector or cell scenarios this thesis follows recommendations of 3rd Generation Partnership Project (3GPP) standardization for system level simulations in [3GP17g]. Therein, BSs are placed

[2] In (2.3) the noise variance is constant, therefore the SINR is not exactly the same when increasing transmit power at all BSs, but in practical systems some parts of the noise scale with transmit power as well.



(a) Multiple-sector deployment with virtual sector boarders wherein mobile users are dropped in the $x^{(C)}$-$y^{(C)}$ plane. Acronym MSs means mobile stations and sec. is short for sector.



(b) Number of users connected to BS in sector 1 due to sector assignment of users to the BS with the highest receive power.

Figure 4.1.: Multiple-sector scenario and effect of sector assignment procedure.

in a regular hexagonal grid shown in Fig. 4.1a, where BSs locations are represented by black squares. Furthermore, black solid lines represent virtual sector boarders, wherein the users are uniformly and randomly dropped. Despite the regular deployment of antenna arrays, using a geometry-based stochastic channel model with randomly correlated large and small-scale parameters and a uniform random user placement, sufficient variation in the propagation environment is ensured for algorithm evaluation.

In Fig. 4.1a, 100 mobile users are dropped in each sector, represented as colored circles. Without limitation to generality BS $l = 1$ belongs to the inner sector in Fig. 4.1a (yellow circles).

With the above described multiple-sector scenarios sector assignment of mobile users to BSs becomes relevant. Sector assignment is not required in a single BS scenario and is therefore omitted in the system model description in Chapter 2. Sector assignment means, that due to shadow fading the distance-wise closest BS of a mobile user can be different from the BS with the largest receive power. Therefore, multiple-sector system level simulations include a sector assignment procedure that assign mobile users to BSs based on receive power measurements. Here, a handover procedure recommended by 3GPP standardization is used [HSKS14] and described in more detail in the following.

At the start of the program each mobile user is assigned to the BS that corresponds to the largest un-precoded receive signal power that is given for BS l at user k by

$$P_{k,l}^{(\mathrm{RSRP})} = \frac{1}{N^{(\mathrm{RB})}} \sum_{i=1}^{N^{(\mathrm{RB})}} \sum_{m=1}^{M} \sum_{n=1}^{N_l} |\mathbf{H}_{k,l,i}|_{m,n}^2 \,, \tag{4.3}$$

where "RSRP" is 3GPP terminology for reference signal receive power (RSRP) and i indicates the frequency domain channel of the i-th resource block (RB). This means, that $P_{k,l}^{(\mathrm{RSRP})}$ is a wideband measure averaged over RBs. Then each mobile user k is connected to BS $l^{(\max)}$ that corresponds to the largest RSRP. Consequently, the number of mobile users connected to each BS varies even with the same number of dropped users in each sector. Fig. 4.1b shows the distribution of the number of connected users to BS 1 in Fig. 4.1a from a statistic of 500 realizations.

Fig. 4.2 shows the impact from inter-cell interference on the sum spectral efficiency over the number of interfering BSs with simulation parameters as listed in Table 4.1. The number of dropped users is $\tilde{K} = 20$ per sector and all of them are scheduled. The sum spectral efficiency loss in this setting is more than 50 %

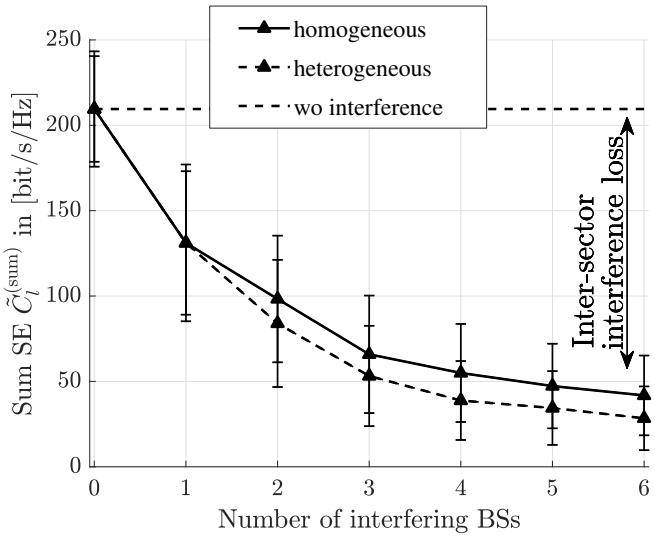


Figure 4.2.: Impact of interfering BSs on sum spectral efficiency for $K = \tilde{K} = 20$ mobile users in each sector. Other simulation assumptions are listed in Table 4.1.

with two interfering BSs and more than 75 % with saturation towards six interfering BSs. Adding more interfering BSs would change results only marginal by the cost of approximately 3 times more computational complexity, as another ring means 18 more interfering BSs. The main observation in Fig. 4.2 is the gap between the homogeneous and heterogeneous scenario caused by the "massive MIMO effect". The massive MIMO effect means that the larger the number of antennas the higher the probability that channels become orthogonal [Mar15]. Briefly summarized, if a user served by BS 1 has orthogonal channels comparing the channel between BS 1 and BS 2, then BS 2 would cause no interference. In the homogeneous scenario the number of antennas $N = 100$ is larger than in the heterogeneous scenario with $\tilde{N} = \tilde{N}_{(\alpha)} \cdot \tilde{N}_{(\beta)} = 2 \cdot 2 = 4$. Therefore, in Fig. 4.2 the sum spectral efficiency of the homogeneous scenario is larger compared to the heterogeneous scenario.

4.1.2. Interference Compensation

In the previous section, sum spectral efficiency degradation due to inter-sector interference is shown. Thereby, $T = 20$ randomly selected users are spatially multiplexed. In this section, the impact of inter-sector interference on user selection is studied. Then, an additional low overhead feedback is proposed to compensate sum spectral efficiency losses from interference.

The impact of unknown interference on the sum spectral efficiency is shown over the number of available users $\tilde{K}_l$ in Fig. 4.3a. Upward and downward pointing triangles represent random user selection in the homogeneous and heterogeneous scenario, respectively, while right- and left-hand pointing triangles represent sum spectral efficiency maximizing projection based zero forcing (PBZF)[3] user selection in the homogeneous and heterogeneous scenario, respectively. In case of the random scheduler the number of selected streams is constraint by the minimum of the number of available users and streams in the first-stage precoder $\mathbf{B}_l$ given by $N^{(\tilde{\mathbf{\Omega}})}$ such that[4]

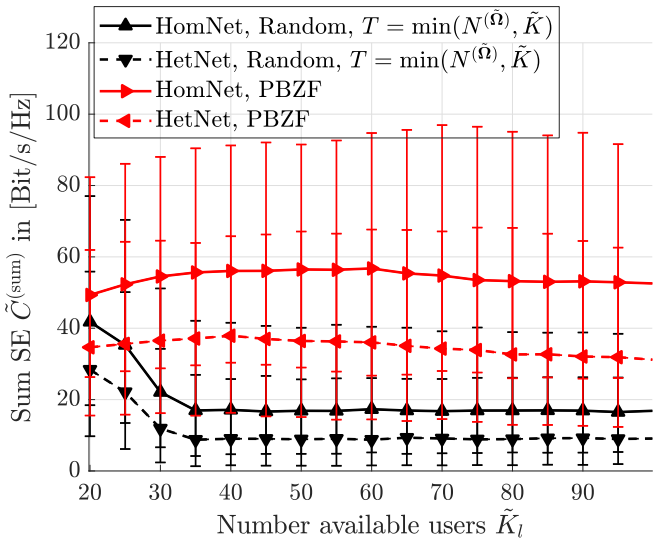
$$T = \min\left(N^{(\tilde{\mathbf{\Omega}})}, \tilde{K}\right). \tag{4.4}$$

With parameters from Table 4.1 maximum $T = 32$ streams can be scheduled. This is shown in Fig. 4.3b, where the number of selected streams T over the number of available users $\tilde{K}$ is given. This constant number of scheduled streams is also the reason for the approximately constant sum spectral efficiency with

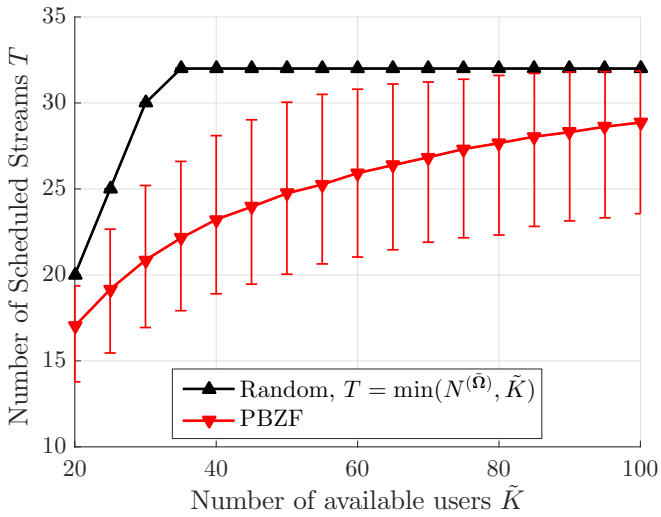
[3] See Section 2.3.4 for details on the PBZF user selection algorithm.
[4] See Section 2.4.4 on hybrid precoding for details.

Table 4.1.: Parameter assumptions for multiple-sector simulations.

Parameter	**Value**
Scenario	3GPP three dimensional (3D) Urban Macro NLoS, [3GP17e]
Center frequency	4 GHz
Number of multi-path components $L_{(\mathrm{MPC})}$	21
OFDM RB Bandwidth	180 kHz
Number of RBs	50
Utilized bandwidth	9 MHz
Cyclic prefix overhead	7 %
SINR bounds, $\gamma^{(\min)}, \gamma^{(\max)}$	$\gamma^{(\min)} = -5\,\mathrm{dB}$, $\gamma^{(\max)} = 40\,\mathrm{dB}$
Number of BSs L	7
Inter-side distance	500 m
Sectorization	Three per BS location
User distribution	Uniform random in hexagons according to Fig. 4.1a
Number of available users $\tilde{K}$	$[20, 100]$
Homogeneous: $N_{(\alpha)}, N_{(\beta)}, N$	$10, 10, 100$
Heterogeneous: $\tilde{N}_{(\alpha)}, \tilde{N}_{(\beta)}, \tilde{N}$	$2, 2, 4$
First stage precoder $\mathbf{B}_l$	Sub codebook splitting with values below
$N_{(\alpha)}^{(\mathbf{\Omega})}, N_{(\beta)}^{(\mathbf{\Omega})}, N^{(\mathbf{\Omega})}$	$16, 16, 256$
$\alpha^{(\mathrm{CW,min})}, \alpha^{(\mathrm{CW,max})}$	$-50°, 50°$
$\beta^{(\mathrm{CW,min})}, \beta^{(\mathrm{CW,max})}$	$-40°, 8°$
$N_{(\alpha)}^{(\mathbf{\Omega},\mathrm{SCB})}, N_{(\beta)}^{(\mathbf{\Omega},\mathrm{SCB})}, N^{(\tilde{\mathbf{\Omega}})}$	$8, 4, 32$
$N^{(\mathrm{SCB})}$	8, according to Algorithm 4
Second stage precoder $\mathbf{P}_l$	Minimum mean square error based on first-stage precoded channels and multiple-input single-output feedback according to Section 2.2
Scheduler	Random or PBZF according to Section 2.3.4



(a) Sum spectral efficiency over the number of available users $\tilde{K}$ comparing random and PBZF scheduling for homogeneous and heterogeneous scenario.



(b) Number of scheduled streams T over the number of available users $\tilde{K}$.

Figure 4.3.: Performance of homogeneous and heterogeneous multiple-sector scenario.

random selection for $\tilde{K}_l > 30$ in Fig. 4.3a. The gap between homogeneous and heterogeneous scenario is approximately constant over $\tilde{K}$.

With the PBZF scheduler the sum spectral efficiency gain is less than 15 % for $\tilde{K} = 60$ with respect to $\tilde{K} = 20$, achieved in the homogeneous scenario. For $\tilde{K} > 60$ the sum spectral efficiency is decreasing due to the unknown inter-sector interference and no further scheduling gains are realized. The gap between homogeneous and heterogeneous scenario increases for the PBZF scheduler because the received interference power in the heterogeneous scenario is larger, see Fig. 4.2.

Fig. 4.3b shows that for PBZF the number of scheduled streams constantly increases along with the number of available users, but without gain in sum spectral efficiency. This is due to the unknown interference situation of the users, which is not taken into account in the scheduling decision and precoder design. Note that this observation is independent of TDD or FDD systems. In [KRTT13] the same observation for second stage precoding using maximum ratio transmission (MRT), signal to leakage and noise ratio (SLNR) and zero forcing (ZF) is described. First stage precoding is not considered in [KRTT13] corresponding to $\mathbf{B}_l = \mathbf{I}_{N_l}$ in this work.

The results in Fig. 4.3 confirm that the interference situation at the mobile users is essential for massive MIMO multiple-user downlink transmission[5], otherwise spatial multiplexing gains are limited. In [KTH14b] the author of this thesis and others proposes to include the interference power received at the mobile users in the regularization matrix of the minimum mean square error (MMSE) precoder. This proposed scheme and corresponding details are presented in the next paragraphs.

In (2.37) the regularization matrix $\mathbf{R}^{(\mathrm{MMSE})} \in \mathbb{C}^{\mathcal{T}_l \times \mathcal{T}_l}$ is introduced as diagonal matrix with elements $\left[\mathbf{R}^{(\mathrm{MMSE})}\right]_{t,t} = z_k^{(\mathrm{IF},i)}$. The term $z_k^{(\mathrm{IF},i)}$ reflects the level of interference at user k that is known at the BS and is defined in Eq. (2.37) for $i \in [0, 1, 2]$. For the sake of readability these three cases are briefly summarized again.

- $i = 0$ corresponds to the case that only noise variance and no interference knowledge is considered.

[5] Actual, the knowledge about interference is essential not only for massive MIMO transmission, however the emphasis here is on massive MIMO because the spatial multiplexing gain is the main advantage of massive MIMO and a large portion of it is rendered useless by the unknown interference.

- $i = 1$ corresponds to the case that interference power is available per user on a per RB granularity, labeled henceforth as "Per RB". This can be seen as the "upper bound" for the proposed wideband feedback in $i = 2$.
- $i = 2$ is the proposed scheme and corresponds to the case that the expectation of the interference power over all RBs is available per user, henceforth labeled as "wideband".

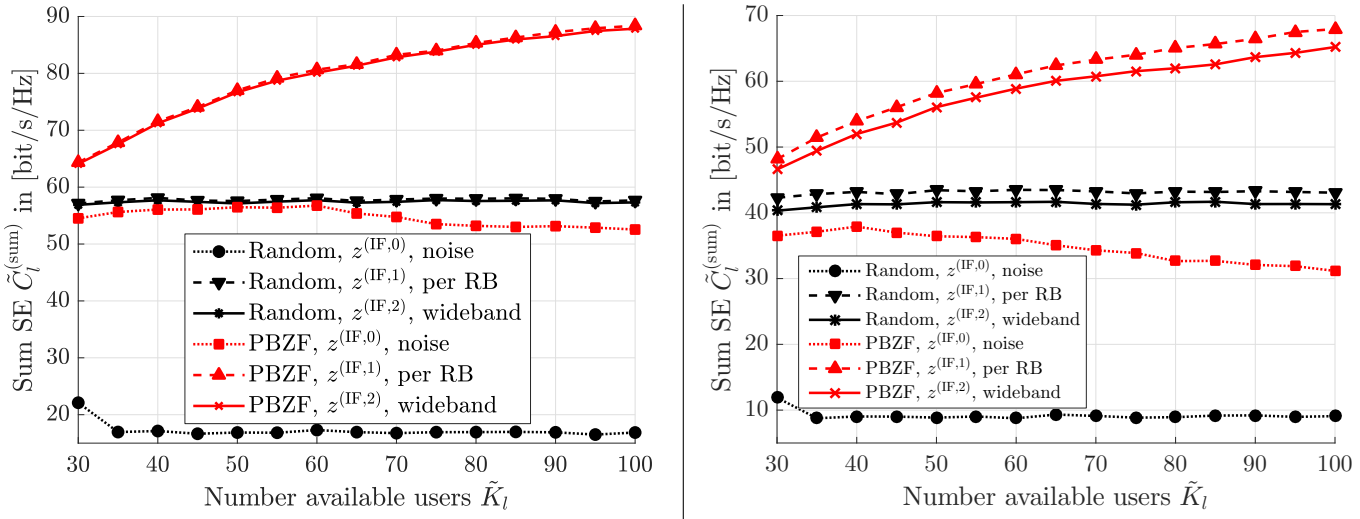
An estimate of the interference power at user k can be obtained by subtracting the power of the transmitted signals of the serving BS l from the power of the complete receive signal such that

$$\begin{aligned} z_k^{(\mathrm{IF,est})} &= \frac{1}{M} \left\| \mathbf{y}_k^{\mathrm{H}} \mathbf{y}_k - \left(\mathbf{H}_{k,l} \mathbf{V}_l \sqrt{\mathbf{F}_l} \mathbf{x}_l \right)^{\mathrm{H}} \mathbf{H}_{k,l} \mathbf{V}_l \sqrt{\mathbf{F}_l} \mathbf{x}_l \right\|^2 \\ &= \frac{1}{M} \left\| \left(\sum_{m \neq l}^{m \in \mathcal{L}} \mathbf{H}_{k,m} \mathbf{V}_m \sqrt{\mathbf{F}_m} \right)^{\mathrm{H}} \sum_{m \neq l}^{m \in \mathcal{L}} \mathbf{H}_{k,m} \mathbf{V}_m \sqrt{\mathbf{F}_m} + \mathbf{n}_k^{\mathrm{H}} \mathbf{n}_k \right\|^2 = \frac{1}{M} \left\| \mathbf{z}_k^{\mathrm{H}} \mathbf{z}_k + \mathbf{n}_k^{\mathrm{H}} \mathbf{n}_k \right\|^2, \end{aligned} \tag{4.5}$$

where M is the number of receive antennas, $\mathbf{V}_l$ is the precoding matrix, and $\mathbf{F}_l$ is the diagonal power allocation matrices following the definitions in Section 2.2 below Eq. (2.1) for the general downlink receive signal description. The separation of the signal from serving BS l and other BSs in Eq. (4.5) assumes that orthogonal pilots are used. Note, different to the assumption in (2.37), where the noise power is considered, in Eq. (4.5) an instantaneous realization of the noise power is included. Yet, in an interference-limited system the assumption is that $\mathbf{z}_k^{\mathrm{H}} \mathbf{z}_k \gg \mathbf{n}_k^{\mathrm{H}} \mathbf{n}_k$ such that the impact from the noise becomes negligible and $z_k^{(\mathrm{IF},1)} \approx z_k^{(\mathrm{IF,est})}$. In this work, $z_k^{(\mathrm{IF},2)}$ is obtained as the average over all RBs in frequency while in real system the expectation value may also be obtained by additional averaging in time. A practical problem is, that additional to the data, a known reference signal has to be transmitted for each stream in order to estimate and subtract signals of serving BS l from the complete receive signal. In LTE and 5G new radio (NR) these signals are available and called demodulation reference signals (DM-RS), see Section 9 in [3GP14] and Section 5 in [3GP18c], respectively. However, the demodulation reference signals from BS l and the other BSs $m \neq l$ may not be completely orthogonal and thus an unknown error is included in $z_k^{(\mathrm{IF,est})}$ in real systems. Thus, the above mentioned averaging in the frequency domain over all RBs reflects to some extend these errors. By using an averaged value the interference power estimate on some RBs is larger or lower than the actual interference.

Fig. 4.4a shows the homogeneous scenario and there is approximately the same sum spectral efficiency for "wideband" and "per RB" interference regularization according to Eq. (2.37). In contrast to Fig. 4.3a the sum spectral efficiency is negligibly increasing with $\tilde{K}_l$ resulting in 60 % gain at $\tilde{K}_l = 100$. The performance of the random scheduler remains approximately constant and the gain compared to no interference knowledge $z_k^{(\mathrm{IF},0)}$ in Fig. 4.3a is more than three times. In Fig. 4.4b the sum spectral efficiency for the heterogeneous scenario is shown. The difference between "per RB" ,$z_k^{(\mathrm{IF},1)}$ and wideband $z_k^{(\mathrm{IF},2)}$ is significantly larger compared to the homogeneous scenario. Especially with the PBZF scheduling heuristic the gap becomes larger with increasing $\tilde{K}_l$. The sum spectral efficiency gains, due to the additional interference regularization in the MMSE precoder, are larger in the heterogeneous scenario compared to the homogeneous scenario. The reasons for this is the larger received inter-cell interference power in heterogeneous scenarios, e.g. the gain for PBZF is more than a factor of two at $\tilde{K}_l = 100$ and more then four for random scheduling.

The additional feedback that is required consist of scalar power values, in case of "Per RB" there are $N^{(\mathrm{RB})}$ values, and in case of "wideband" there is a single value. Considering 10 bit for quantization per power value this results in 500 bit and 10 bit additional feedback for "per RB" and "wideband", respectively. In case of the homogeneous scenario, the sum spectral efficiency degradation from "per RB" to "wideband" is $< 1\,\%$. By sending the interference power feedback on a larger time-scale than the channel state information (CSI) feedback the additional overhead is neglectable compared to the CSI feedback.



(a) Homogeneous scenario.

(b) Heterogeneous scenario.

Figure 4.4.: Effective sum spectral efficiency comparing user scheduler and MMSE precoding with and without interference power regularization according to Eq. (2.37). PBZF is a sum spectral efficiency maximizing scheduler as described in Section 2.3.4. The label "$z^{(\mathrm{IF},0)}$, noise" means that no interference, only noise variance is taken into account. The label "$z^{(\mathrm{IF},1)}$, per RB" means that the interference power per RB is taken into account. The label "$z^{(\mathrm{IF},2)}$, wideband" means that the average interference power over all RBs is taken into account.

4.1.3. Conclusions

It is shown that inter-sector interference is limiting the downlink sum spectral efficiency for massive MIMO multiple-user spatial multiplexing, assuming a practical number of antennas. A homogeneous and heterogeneous scenario is considered in this section and it can be observed that the impact from interference is less in the homogenous scenario. This confirms findings that, according to the law of large numbers, the probability of orthogonal channels increases with the number of antennas. However, even with the $N = 100$ antenna elements at all BSs, considered in the homogeneous scenario, the loss from single BS to six interfering BSs is more than a factor of four.

Independent of TDD or FDD, without knowledge about the interference situation at the users, spatial multiplexing gains from single-cell scenario can vanish or even result in sum spectral efficiency degradation. An interference power feedback metric is proposed, which requires very low additional feedback overhead of 10 bit per feedback interval. Furthermore, it is shown that by including this metric in the MMSE precoder massive MIMO spatial multiplexing gains can also be utilized in interference-limited scenarios. Results in this section indicate that in the homogeneous massive MIMO scenario a wideband interference power feedback is sufficient and additional gains with "per RB" are neglectable. In the heterogeneous scenario a "per RB" feedback provides additional sum spectral efficiency improvement of approximately 3 bit/s/Hz.

4.2. Uplink Positioning based on Two DoA Estimates

In Chapter 3, the position error was evaluated in order to reflect horizontal and elevation estimation errors in one performance indicator, however, the distance between sources and the BS was assumed to be known. In a real cellular system, this assumption is not valid. To obtain the position of a device in the uplink, measurements from multiple BSs, or multiple and different measurements at the same BS are required. For example, to obtain a unique position estimate in 3D coordinates based on ToF measurements, four different BSs are required [Fis14]. As already mentioned in Section 1.2, a major drawback of time based positioning systems is that the minimal time resolution limits the achievable accuracy. A typical example of the error budget for a LTE time-based system is given in Section 8.7 in [Fis14] with approximately 50 m position error.

In order to achieve the position accuracy requirement of 1 m[6], this chapter investigates the utilization of accurate DoA estimates from large antenna arrays in the system, e.g. by low complex search-based DoA estimation given in Chapter 3. There exist at least two straightforward positioning schemes utilizing DoA estimates:

1. Combination of range and DoA estimation from one BS (similar to radio detection and ranging (RADAR)).

2. Combination of at least two DoA estimates from different BSs.

In the first approach, a distance or range estimate based on the ToF, see Section 1.2.1, additional to the DoA estimate is required. The distance estimate corresponds to a sphere around the BS and in combination with the direction estimate a unique 3D position can be obtained. However, without precise synchronization, a range estimate obtained based on time a measurement results in large estimation errors. On the other hand the range estimate can also be obtained by receive signal strength (RSS) methods, see Section 1.2.1. However, the achieved accuracy is based on correction values[7] used for path-loss computation, see Section 1.2.1. These correction values strongly depend on the environment [LZP11, VGL+15]. Therefore, the focus in this chapter is on option two, combining multiple DoA estimates from different BSs. With uncertainty in the DoA estimates this results in skew lines without an intersection. In literature, many algorithms exist to obtain the position based on estimated DoAs, e.g. by calculating centroids [GM13], weighted least square [HN03], robust weighted least square [PP07], or by grid-based search in [GAPM14]. These approaches minimize the estimation error by utilizing more than two DoA estimates. The more DoA estimates the better, however, in cellular deployments there is a distance depending line-of-sight probability, meaning that the larger the distance between a BS and source, the larger the probability of non line of sight (NLoS). The outcome of extensive measurements campaigns is included in the current 3GPP channel model for system level simulations [3GP17f]. Therefore, the focus of this thesis is on the scenario with the minimum requirement of two available DoA estimates from two different BSs with a line of sight (LoS) connection to the same source. A possible scenario is for example position tracking of vehicles to support traffic safety, in the future this can also include unmanned aerial vehicles[8] in flight corridors covered by massive MIMO BS. To minimize capital expenditure and operational expenditure costs, the minimum required number of two BSs is in the focus of this section.

In Fig. 4.5, the principle signal flow to determine the position of a mobile source in the cellular network based on uplink DoA estimation is given and described in the following four steps.

1. The mobile source transmits a sequence in the uplink.

2. The BSs estimate the direction of the source signal.

[6] See Section 9.2 "Location/Positioning Service" in [3GP16d].
[7] These correction values are usually obtained by measurements campaigns.
[8] In 3GPP standardization a study item titled "Study on Enhanced LTE Support for Aerial Vehicles" has been completed in 2017 in technical report 36.777 V15.0.0 and it is most likely that similar support will be provided in 5G NR networks.

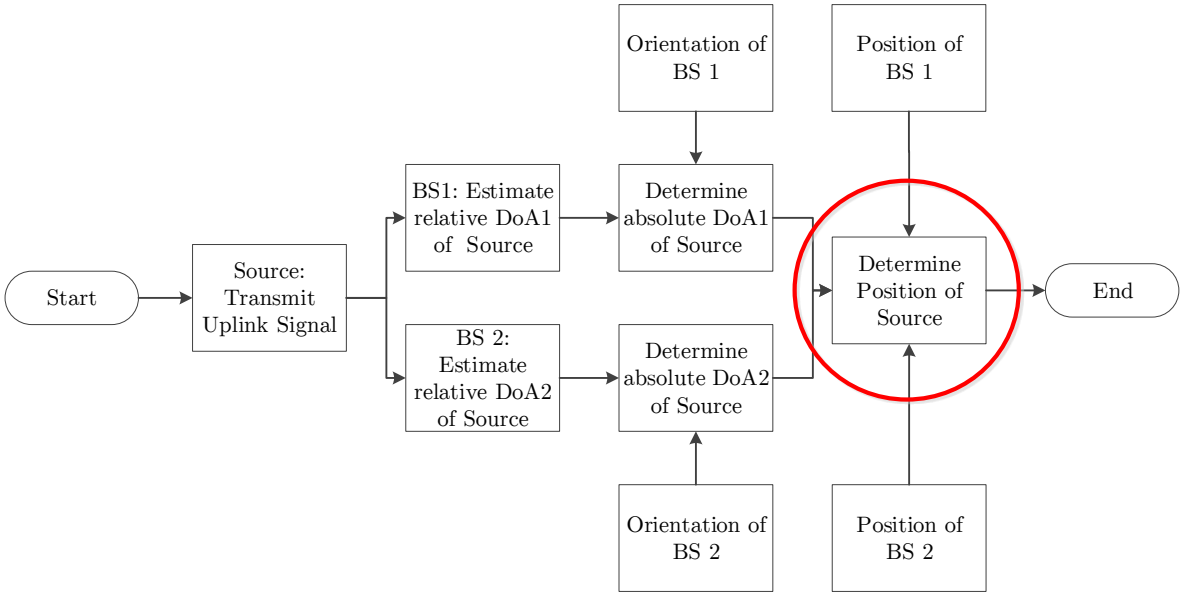


Figure 4.5.: Signal flow for DoA based positioning in a cellular network with the minimum requirement of two BSs.

3. The estimated DoAs are transformed to a reference coordinate system.

4. The position of the source is determined.

Grey blocks indicate informations that are required and can be assumed available in the network because of its static nature, e.g. the orientation and position of the MIMO arrays. Utilizing the BSs positions and estimated DoAs within the reference coordinate system, the position can be determined by a central unit similar to a location server in LTE. Fig. 4.6 shows the above described principle in 3D coordinates. However, for illustrative reasons in Fig. 4.6 "perfect" DoA estimation without error is assumed resulting in an intersection of both DoAs. That such an assumption is almost never true is described Chapter 3 and by the Cramer Rao lower bound (CRLB) in Section 3.1.2. Therefore, this section focuses first on the method to obtain a position estimate in Section 4.2.1 and in Section 4.2.2 on the reduction of the position error.

4.2.1. Skew Lines in Euclidean Space

According to Fig. 4.5, it is assumed that two BSs send a DoA estimated of the same source to a central processing unit, e.g. using the LTE positioning protocol [3GP16a, 3GP16b]. These estimates consist of an azimuth and elevation angle, $\hat{\alpha}$ and $\hat{\beta}$, respectively. However, each BS obtains these estimates relative to its array orientation. Therefore, additionally the absolute position and the orientation of the array are required to use these estimates for positioning. Given, that the orientation and position of the antenna arrays are static we assume both as known in the network. First, the relative DoAs are transformed to a reference coordinate system given by $\bar{\alpha}_i$ and $\bar{\beta}_i$, where the subscript i is the BS ID. The estimated DoA of source k in the reference coordinate system is obtained by

$$\begin{aligned} \bar{\alpha}_k &= \mathrm{mod}(\tilde{\alpha}_i + \hat{\alpha}_k + 180°, 360°) - 180° \\ \bar{\beta}_k &= \mathrm{mod}(\tilde{\beta}_i + \hat{\beta}_k + 90°, 180°) - 90°, \end{aligned} \tag{4.6}$$

where "mod" is the modulo operation and $\tilde{\alpha}, \tilde{\beta}$ are the horizontal/vertical orientation angles of the BS array, respectively. Note that the vertical angle β is counted in this work in the interval $[-90, 90]°$, which is the

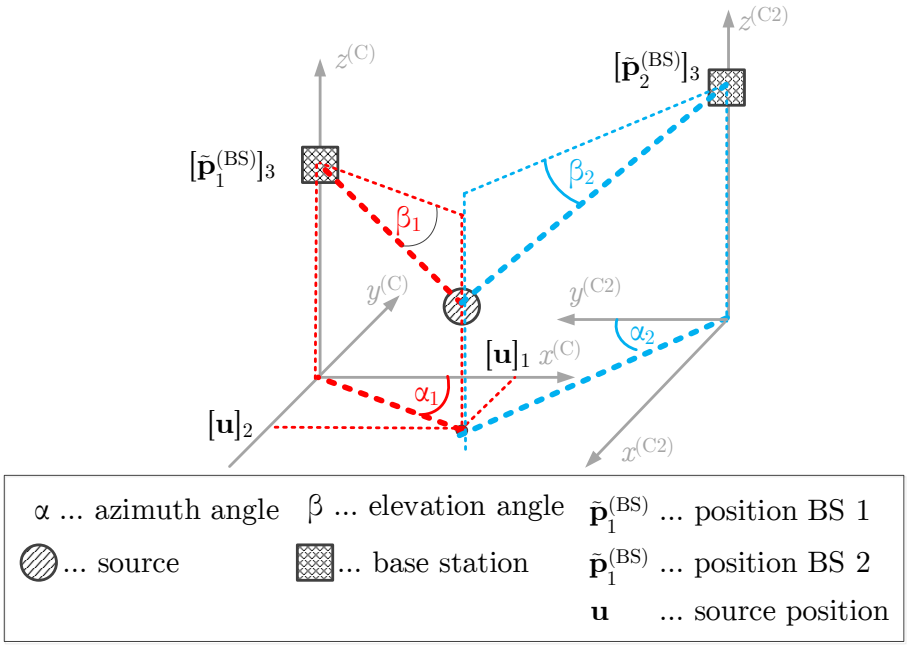


Figure 4.6.: Three dimensional position based on two DoAs from two BS.

reason of the $-90°$ offset after the modulus operation. The next step in Fig. 4.5 is to obtain vector expressions of the DoAs including BS positions. Without limit to generality, the origin of the reference coordinate system is in this work assumed such that BS 1 is located at $\tilde{\mathbf{p}}_1^{(BS)} = \left[x_1^{(C)} = 0\,\text{m}\ y_1^{(C)} = 0\,\text{m}\ z_1^{(C)}\right]^T$ and BS 2 at $\tilde{\mathbf{p}}_2^{(BS)} = \left[x_2^{(C)}\ y_2^{(C)}\ z_2^{(C)}\right]^T$ with orientations of BS 1 as $\tilde{\alpha}_1 = 0°, \tilde{\beta}_1 = 0°$ and BS 2 as $\tilde{\alpha}_2, \tilde{\beta}_2$ according to Fig. 4.6. Combining the position and DoA estimate of BS 1 a linear equation is obtained according to

$$\mathbf{s}^{(BS)} = \begin{pmatrix} x_1^{(C)} + s^{(BS)} \cos\bar{\beta}_1 \cos\bar{\alpha}_1 \\ y_1^{(C)} + s^{(BS)} \cos\bar{\beta}_1 \sin\bar{\alpha}_1 \\ z_1^{(C)} + s^{(BS)} \sin\bar{\beta}_1 \end{pmatrix}, \tag{4.7}$$

where $s^{(BS)} \in \mathbb{R}_+$ is an auxiliary parameter defines the distance from BS 1 along the DoA in [m]. Transforming the DoA estimate of BS 2 into the coordinate system of BS 1 with Eq. (4.6) the linear equation for BS 2 is obtained as

$$\mathbf{t}^{(BS)} = \begin{pmatrix} x_2^{(C)} + t^{(BS)} \cos\bar{\beta}_2 \cos\bar{\alpha}_2 \\ y_2^{(C)} + t^{(BS)} \cos\bar{\beta}_2 \sin\bar{\alpha}_2 \\ z_2^{(C)} + t^{(BS)} \sin\bar{\beta}_2 \end{pmatrix}, \tag{4.8}$$

where the auxiliary parameter $t^{(BS)} \in \mathbb{R}_+$ given in [m] is the corresponding distance along the DoA from BS 2.

As shown in Fig. 4.6, without estimation errors the position can be determined by setting Eq. (4.7) equal to Eq. (4.8). However, the probability that a pair of parameters $s^{(BS)}$ and $t^{(BS)}$ can be found such that $\mathbf{s}^{(BS)} = \mathbf{t}^{(BS)}$ is zero, due to noise and the variance of any DoA estimator. Therefore, $\mathbf{s}^{(BS)}$ and $\mathbf{t}^{(BS)}$ are called skew lines. In [Lev15] it is stated that the source position is determined as the middle point of the smallest segment between these skew lines, but [Lev15] does not describe how the smallest segment and the centroid is obtained. The author of this thesis published this in [KTPI17]. In literature, this method is known as the "*shortest distance between skew lines*" technique [HCV99]. For ease of notation the notations

in Eq. (4.9) and Eq. (4.10) according to Fig. 4.7 are introduced as

$$\mathbf{m}_{(\mathrm{DoA})} = \begin{pmatrix} m_{(\mathrm{x})} \\ m_{(\mathrm{y})} \\ m_{(\mathrm{z})} \end{pmatrix} = \begin{pmatrix} \cos\bar{\beta}_1 \cos\bar{\alpha}_1 \\ \cos\bar{\beta}_1 \sin\bar{\alpha}_1 \\ \sin\bar{\beta}_1 \end{pmatrix} \tag{4.9}$$

and

$$\mathbf{n}_{(\mathrm{DoA})} = \begin{pmatrix} n_{(\mathrm{x})} \\ n_{(\mathrm{y})} \\ n_{(\mathrm{z})} \end{pmatrix} = \begin{pmatrix} \cos\bar{\beta}_2 \cos\bar{\alpha}_2 \\ \cos\bar{\beta}_2 \sin\bar{\alpha}_2 \\ \sin\bar{\beta}_2 \end{pmatrix}. \tag{4.10}$$

$\mathbf{m}_{(\mathrm{DoA})}$ in Eq. (4.9) is the orientation vector of $\mathbf{s}^{(\mathrm{BS})}$ and $\mathbf{n}_{(\mathrm{DoA})}$ in Eq. (4.10) is the orientation vector of $\mathbf{t}^{(\mathrm{BS})}$.

With this, Eq. (4.7) and Eq. (4.8) can be rewritten as

$$\mathbf{s}^{(\mathrm{BS})} = s^{(\mathrm{BS})}\mathbf{m}_{(\mathrm{DoA})} + \tilde{\mathbf{p}}_1^{(\mathrm{BS})} = \begin{pmatrix} s^{(\mathrm{BS})}m_{(\mathrm{x})} + x_1^{(\mathrm{C})} \\ s^{(\mathrm{BS})}m_{(\mathrm{y})} + y_1^{(\mathrm{C})} \\ s^{(\mathrm{BS})}m_{(\mathrm{z})} + z_1^{(\mathrm{C})} \end{pmatrix}, \tag{4.11}$$

and

$$\mathbf{t}^{(\mathrm{BS})} = t^{(\mathrm{BS})}\mathbf{n}_{(\mathrm{DoA})} + \tilde{\mathbf{p}}_2^{(\mathrm{BS})} = \begin{pmatrix} t^{(\mathrm{BS})}n_{(\mathrm{x})} + x_2^{(\mathrm{C})} \\ t^{(\mathrm{BS})}n_{(\mathrm{y})} + y_2^{(\mathrm{C})} \\ t^{(\mathrm{BS})}n_{(\mathrm{z})} + z_2^{(\mathrm{C})} \end{pmatrix}. \tag{4.12}$$

The vector from $\mathbf{s}^{(\mathrm{BS})}$ to $\mathbf{t}^{(\mathrm{BS})}$ is obtained by

$$\mathbf{v}^{(\mathrm{st})} = \mathbf{t}^{(\mathrm{BS})} - \mathbf{s}^{(\mathrm{BS})} \begin{pmatrix} t^{(\mathrm{BS})}n_{(\mathrm{x})} + x_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{x})} - x_1^{(\mathrm{C})} \\ t^{(\mathrm{BS})}n_{(\mathrm{y})} + y_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{y})} - y_1^{(\mathrm{C})} \\ t^{(\mathrm{BS})}n_{(\mathrm{z})} + z_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{z})} - z_1^{(\mathrm{C})} \end{pmatrix}. \tag{4.13}$$

According to [HCV99] the length of vector $\mathbf{v}^{(\mathrm{st})}$ is the shortest distance if and only if $\mathbf{v}^{(\mathrm{st})}$ is perpendicular to the orientation vectors $\mathbf{m}_{(\mathrm{DoA})}$ and $\mathbf{n}_{(\mathrm{DoA})}$ resulting in conditions

$$\begin{aligned} \mathbf{m}_{(\mathrm{DoA})}\mathbf{v}^{(\mathrm{st})} = 0 = & m_{(\mathrm{x})}(t^{(\mathrm{BS})}n_{(\mathrm{x})} + x_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{x})} - x_1^{(\mathrm{C})}) + \\ & m_{(\mathrm{y})}(t^{(\mathrm{BS})}n_{(\mathrm{y})} + y_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{y})} - y_1^{(\mathrm{C})}) + \\ & m_{(\mathrm{z})}(t^{(\mathrm{BS})}n_{(\mathrm{z})} + z_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{z})} - z_1^{(\mathrm{C})}) \end{aligned} \tag{4.14}$$

and

$$\begin{aligned} \mathbf{n}_{(\mathrm{DoA})}\mathbf{v}^{(\mathrm{st})} = 0 = & n_{(\mathrm{x})}(t^{(\mathrm{BS})}n_{(\mathrm{x})} + x_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{x})} - x_1^{(\mathrm{C})}) + \\ & n_{(\mathrm{y})}(t^{(\mathrm{BS})}n_{(\mathrm{y})} + y_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{y})} - y_1^{(\mathrm{C})}) + \\ & n_{(\mathrm{z})}(t^{(\mathrm{BS})}n_{(\mathrm{z})} + z_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{z})} - z_1^{(\mathrm{C})}). \end{aligned} \tag{4.15}$$

The problem of two linear equations for two unknown variables $s^{(\mathrm{BS})}$ and $t^{(\mathrm{BS})}$ has always a unique solution. First, Eq. (4.14) is transformed to

$$\begin{aligned} 0 = & t^{(\mathrm{BS})}m_{(\mathrm{x})}n_{(\mathrm{x})} + m_{(\mathrm{x})}x_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{x})}^2 - m_{(\mathrm{x})}x_1^{(\mathrm{C})} + \\ & t^{(\mathrm{BS})}m_{(\mathrm{y})}n_{(\mathrm{y})} + m_{(\mathrm{y})}y_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{y})}^2 - m_{(\mathrm{y})}y_1^{(\mathrm{C})} + \\ & t^{(\mathrm{BS})}m_{(\mathrm{z})}n_{(\mathrm{z})} + m_{(\mathrm{z})}z_2^{(\mathrm{C})} - s^{(\mathrm{BS})}m_{(\mathrm{z})}^2 - m_{(\mathrm{z})}z_1^{(\mathrm{C})}. \end{aligned} \tag{4.16}$$

Secondly, $t^{(\mathrm{BS})}$ and $s^{(\mathrm{BS})}$ are separated and for the sake of readability other values (constants) are aggregated to auxiliary variables $a_{(\mathrm{h})}, b_{(\mathrm{h})}$, and $c_{(\mathrm{h})}$ below the under-braces in

$$\begin{aligned} 0 =& t^{(\mathrm{BS})} \underbrace{\left(m_{(\mathrm{x})}n_{(\mathrm{x})} + m_{(\mathrm{y})}n_{(\mathrm{y})} + m_{(\mathrm{z})}n_{(\mathrm{z})}\right)}_{=a_{(\mathrm{h})}} - s^{(\mathrm{BS})} \underbrace{\left(m_{(\mathrm{x})}^2 + m_{(\mathrm{y})}^2 + m_{(\mathrm{z})}^2\right)}_{=b_{(\mathrm{h})}} + \\ & \underbrace{\left(m_{(\mathrm{x})}x_2^{(\mathrm{C})} - m_{(\mathrm{x})}x_1^{(\mathrm{C})} + m_{(\mathrm{y})}y_2^{(\mathrm{C})} - m_{(\mathrm{y})}y_1^{(\mathrm{C})} + m_{(\mathrm{z})}z_2^{(\mathrm{C})} - m_{(\mathrm{z})}z_1^{(\mathrm{C})}\right)}_{=c_{(\mathrm{h})}}. \end{aligned} \tag{4.17}$$

Next, Eq. (4.17) is rearranged to $s^{(\mathrm{BS})}$ by

$$s^{(\mathrm{BS})} = \frac{t^{(\mathrm{BS})}a_{(\mathrm{h})} + c_{(\mathrm{h})}}{b_{(\mathrm{h})}}. \tag{4.18}$$

The same steps are performed on Eq. (4.15), where constants are aggregated to $d_{(\mathrm{h})}, e_{(\mathrm{h})}$, and $f_{(\mathrm{h})}$ in

$$\begin{aligned} 0 =& t^{(\mathrm{BS})} \underbrace{\left(n_{(\mathrm{x})}^2 + n_{(\mathrm{y})}^2 + n_{(\mathrm{z})}^2\right)}_{d_{(\mathrm{h})}} - s^{(\mathrm{BS})} \underbrace{\left(n_{(\mathrm{x})}m_{(\mathrm{x})} + n_{(\mathrm{y})}m_{(\mathrm{y})} + n_{(\mathrm{z})}m_{(\mathrm{z})}\right)}_{e_{(\mathrm{h})}} + \\ & \underbrace{\left(n_{(\mathrm{x})}x_2^{(\mathrm{C})} - n_{(\mathrm{x})}x_1^{(\mathrm{C})} + n_{(\mathrm{y})}y_2^{(\mathrm{C})} - n_{(\mathrm{y})}y_1^{(\mathrm{C})} + n_{(\mathrm{z})}z_2^{(\mathrm{C})} - n_{(\mathrm{z})}z_1^{(\mathrm{C})}\right)}_{f_{(\mathrm{h})}}, \end{aligned} \tag{4.19}$$

which is rearranged to $t^{(\mathrm{BS})}$ by

$$t^{(\mathrm{BS})} = \frac{s^{(\mathrm{BS})}e_{(\mathrm{h})} - f_{(\mathrm{h})}}{d_{(\mathrm{h})}}. \tag{4.20}$$

Eq. (4.18) is put into Eq. (4.20) and rearranged to $t^{(\mathrm{BS})}$ by

$$t^{(\mathrm{BS})} = \frac{c_{(\mathrm{h})}e_{(\mathrm{h})} - f_{(\mathrm{h})}b_{(\mathrm{h})}}{d_{(\mathrm{h})}f_{(\mathrm{h})} - e_{(\mathrm{h})}a_{(\mathrm{h})}} = \frac{c_{(\mathrm{h})}a_{(\mathrm{h})} - f_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2}, \tag{4.21}$$

where details can be found in the Appendix in Eq. (A.1). The re-arrangement in Eq. (4.20) utilizes $b_{(\mathrm{h})} = d_{(\mathrm{h})} = 1$ and $a_{(\mathrm{h})} = e_{(\mathrm{h})}$ of which details are given in Table A.1. Finally, $s^{(\mathrm{BS})}$ is obtained by putting Eq. (4.21) in Eq. (4.18) such that

$$s^{(\mathrm{BS})} = \frac{\frac{c_{(\mathrm{h})}a_{(\mathrm{h})} - f_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2}a_{(\mathrm{h})} + c_{(\mathrm{h})}}{b_{(\mathrm{h})}} = \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})}a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2}. \tag{4.22}$$

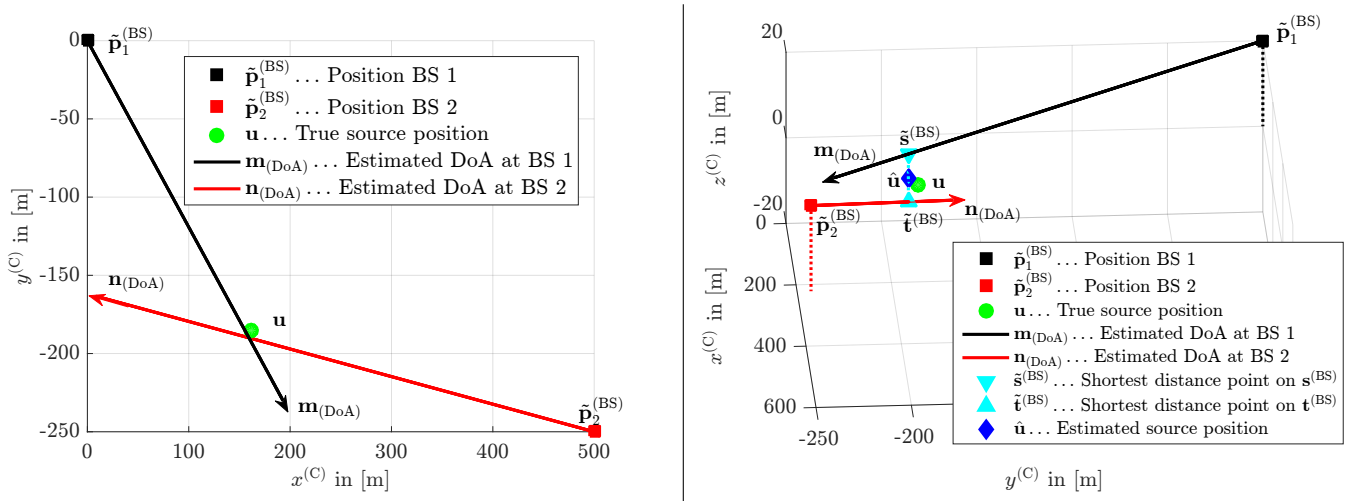
By substituting $s^{(\mathrm{BS})}$ and $t^{(\mathrm{BS})}$ in Eq. (4.7) and Eq. (4.8) with Eq. (4.21) and Eq. (4.22) the points $\tilde{\mathbf{s}}^{(\mathrm{BS})}$ and $\tilde{\mathbf{t}}^{(\mathrm{BS})}$ are obtained, respectively, defining the shortest distance. $\tilde{\mathbf{s}}^{(\mathrm{BS})}$ is marked as a light blue downward pointing triangle in Fig. 4.7b and $\tilde{\mathbf{t}}^{(\mathrm{BS})}$ is marked as a light blue upward pointing triangle in Fig. 4.7b, respectively. Straightforward the segment which is the shortest distance between $\mathbf{s}^{(\mathrm{BS})}$ and $\mathbf{t}^{(\mathrm{BS})}$ is denoted as $\tilde{\mathbf{v}}^{(\mathrm{st})}$ and given by

$$\tilde{\mathbf{v}}^{(\mathrm{st})} = \tilde{\mathbf{s}}^{(\mathrm{BS})} + g_{(\mathrm{p})}\left(\tilde{\mathbf{t}}^{(\mathrm{BS})} - \tilde{\mathbf{s}}^{(\mathrm{BS})}\right), \tag{4.23}$$

where $g_{(\mathrm{p})} \in \mathbb{R}$ is in the interval $[0, 1]$. Finally, the estimated 3D position $\hat{\mathbf{u}}$ is a point on the segment $\tilde{\mathbf{v}}^{(\mathrm{st})}$ for a given $g_{(\mathrm{p})}$, e.g. the middle of $\tilde{\mathbf{v}}^{(\mathrm{st})}$ is obtained by $g_{(\mathrm{p})} = \frac{1}{2}$ as

$$\hat{\mathbf{u}} = \tilde{\mathbf{s}}^{(\mathrm{BS})} + \frac{1}{2}\left(\tilde{\mathbf{t}}^{(\mathrm{BS})} - \tilde{\mathbf{s}}^{(\mathrm{BS})}\right), \tag{4.24}$$

illustrated in Fig. 4.7b by the deep blue diamond.



(a) Top view. (b) Diagonal view.

Figure 4.7.: Illustration for DoA estimation based on "*shortest distance between skew lines*" in 3D space.

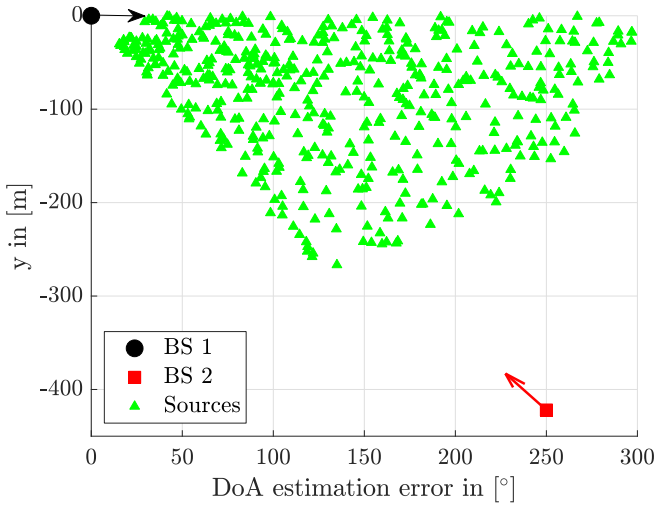


Figure 4.8.: Deployment of BSs and sources. The arrows indicate the orientation of the antenna arrays.

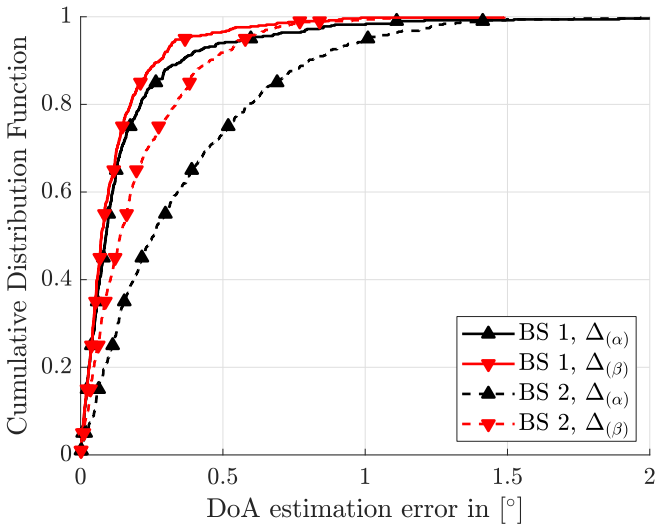
For the numerical evaluation a typical hexagonal deployment as recommended for system level evaluation in 3GPP [3GP17f] is assumed. From this hexagonal deployment two neighbor BSs are selected resulting in BS positions $\tilde{\mathbf{p}}_1^{(\mathrm{BS})} = [0\ 0\ 5]^{\mathrm{T}}$ m and $\tilde{\mathbf{p}}_2^{(\mathrm{BS})} = [250\ \ -422\ 5]^{\mathrm{T}}$ m corresponding to an inter-site distance (ISD) of 500 m.

The corresponding orientation vectors $\tilde{\alpha}_1 = \tilde{\beta}_1 = 0°$ for BS 1 and $\tilde{\alpha}_2 = 120°$, $\tilde{\beta}_2 = 0°$ for BS 2 are illustrated in Fig. 4.8, where arrows represent the orientation vectors. Sources are uniformly randomly distributed in the horizontal and vertical angular range given by $\alpha \in \{0, ..., -65\}°$ and $\beta \in \{-5, ..., 5\}°$, respectively, from BS 1 perspective. The distance between BS 1 and the sources is in the interval $\{30, ..., 300\}$ m. Fig. 4.8 shows and example of 500 source positions in the above defined area. All remaining parameter assumptions are listed in Table 4.2.

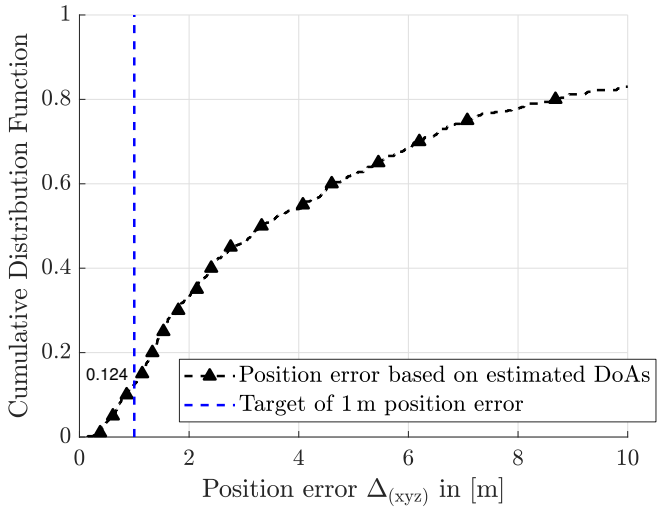
In Fig. 4.9, the DoA estimation and position errors are given. A first observation is that in Fig. 4.9a the cumulative distribution functions (CDFs) of $\Delta_{(\alpha)}$ have a larger variance and larger values compared to the CDFs of $\Delta_{(\beta)}$ for both BSs. This is because the horizontal angular spread of the sources is larger than the vertical angular spread and it is shown in Fig. 3.12 that for angles $\gtrless \pm 60°$ the resolution of the array becomes the limiting factor. A second observation is that the horizontal and vertical estimation errors of BS 2 are larger compared to BS 1, because the average distance of the source to BS 2 is larger than to

Table 4.2.: Parameter assumptions for simulations on positioning based on two DoAs estimates from two BSs.

Parameter	**Value**
Simulation type	Monte Carlo
Number of realization	100
Channel model	QuaDRiGa [JRBT14]
Scenario	Urban macro line of sight [3GP17f]
Center frequency	3.75 GHz
Rician K-Factor	10 dB
BS antenna distribution	UPA
Number of BS antenna elements N	64
Number of elements in $y^{(\mathrm{C})}$-direction $n_{(\mathrm{y})}$	8
Number of elements in $z^{(\mathrm{C})}$-direction $n_{(\mathrm{z})}$	8
Antenna element spacing $d_{i,i+1}^{(\mathrm{N})}$	$\lambda_{(\mathrm{c})}/2$
Antenna element type	Patch
Patch element azimuth HPBW	65°
Patch element elevation HPBW	65°
Element directive gain	9.4 dBi
Bandwidth	180 kHz
BS coordinates $\tilde{\mathbf{p}}_1^{(\mathrm{BS})}, \tilde{\mathbf{p}}_2^{(\mathrm{BS})}$	$[0\ 0\ 5]^{\mathrm{T}}$ m, $[250\ -422\ 5]^{\mathrm{T}}$ m
BS orientations $\tilde{\alpha}_1, \tilde{\beta}_1, \tilde{\alpha}_2, \tilde{\beta}_2$	$0°, 0°, 120°, 0°$
DoA estimator	MUSIC
Target quantization of ASSQ $[q_S^{(\alpha)} q_S^{(\beta)}]$	$[0.1\ 0.1]°$
Position calculation method	Shortest distance of skew lines
Weighting factor $g_{(\mathrm{p})}$	0.5
Source distribution	Random and uniform
Horizontal angular range of sources	$[0, -65]°$
Vertical angular range of sources	$[5, -5]°$
Distance range from BS 1 to sources	$[30, 300]$ m



(a) Distribution of horizontal and vertical DoA estimation errors.



(b) Position error distribution. 0.124 % of the position errors are below the target of 1 m.

Figure 4.9.: Localization performance in cellular deployment of two BSs with 500 m ISD.

BS. Even if both BSs observe the same K-factor[9], the receive signal to noise ratio (SNR) at BS 2 is lower compared to BS 1. In Fig. 4.9b the CDF of the position error $\Delta_{(xyz)}$ is larger compared to $\Delta_{(\alpha)}$ or $\Delta_{(\beta)}$. One reason for this is the distance dependency of the position error additional to the DoA estimation error, see Fig. 3.15a. Another reason for the larger distribution of $\Delta_{(xyz)}$ is caused by the geometry of the deployment, the so-called DoP effect or geometry dilution of precision (GDoP). In Fig. 4.10 the position error is plotted over the positions of the devices in the $x^{(C)}$-$y^{(C)}$ plane. There is a region around the line of sight between BS 1 and BS 2 where position errors $\Delta_{(xyz)} \leq 20\,\text{m}$ are concentrated. This problem is known as dilution of precision and in more detail discussed in the following section as a challenge of DoA based positioning in cellular deployments.

4.2.2. Compensation of Dilution of Precision (DoP) Errors

The impact of DoP on positioning is known in global navigation satellite systems (GNSSs) related literature also as geometry dilution of precision (GDoP) [Zhu92]. Using ToF based positioning methods as described in Section 1.2.1 GDoP has also been observed in cellular networks [CF14]. A visualization of the ToF based GDoP is shown in Fig. 4.11a. Therein the uncertainty region of case C, where the two receivers and the device are on the same line, is larger compared to case A. The same effect happens with DoA based positioning as the following example shows. The same cellular deployment, as used in the previous Section 4.2.1, is given in Fig. 4.11b where the same DoA estimation error at

- BS 1 to device 1 and device 2 is $\Delta_{(\alpha)}^{(1,1)} = \Delta_{(\alpha)}^{(1,2)} = 1°$, and at
- BS 2 to device 1 and device 2 is $\Delta_{(\alpha)}^{(2,1)} = \Delta_{(\alpha)}^{(2,2)} = 0.3°$,

resulting in position errors of $\Delta_{(xyz)}^{(1)} = 2\,\text{m}$ and $\Delta_{(xyz)}^{(2)} = 92\,\text{m}$ for device 1 and 2, respectively. This result explains the error distribution in Fig. 4.9.

In GNSS, for example global positioning system (GPS), uncertainty from dilution of precision is taken into account by a DoP factor $G_{(DoP)}$ calculated based on the geometry of the receiver and four satellites

[9] Power ratio of the LoS and sum of NLoS paths defined in Section 3.1.1 in Eq. (3.13).

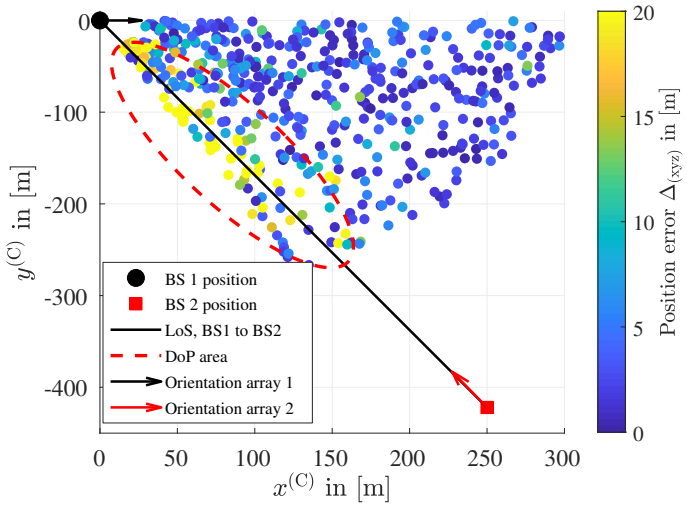


Figure 4.10.: Visualization of the dilution of precision (DoP) problem. Position error $\Delta_{(\mathrm{xyz})}$ plotted in the $x^{(C)}$-$y^{(C)}$ plane at the true source positions. The BSs are separated by 500 m ISD. The red dashed ellipsoid, labeled as the DoP area, contains most of the large position errors, as indicated by the yellow dots of the color coded position error. The DoP area is around the LoS between BS 1 and BS 2, shown as a solid black line. The BS positions are given by a black circle and red square. The orientations are indicated by the corresponding colored arrows at the BS positions.

[KT06, Lan99]. The selection of the satellites for positioning is based on the minimum $G_{(\mathrm{DoP})}$. A further study on the optimal satellite selection is out of scope of this thesis and omitted.

In a cellular deployment, the dilution of precision problem is based on angle of arrival measurements and a similar factor as for GNSSs can be calculated, see [Dem06]. Corresponding results for optical wireless positioning systems are drawn in [BAJ+15]. Assuming that in an ultra-dense network the degree of freedom is large enough to combine the DoA estimates from more than two BSs the pair of estimates with the lowest $G_{(\mathrm{DoP})}$ should be used. However, in this work the focus is on two BSs so there is no degree of freedom on the selection of BSs.

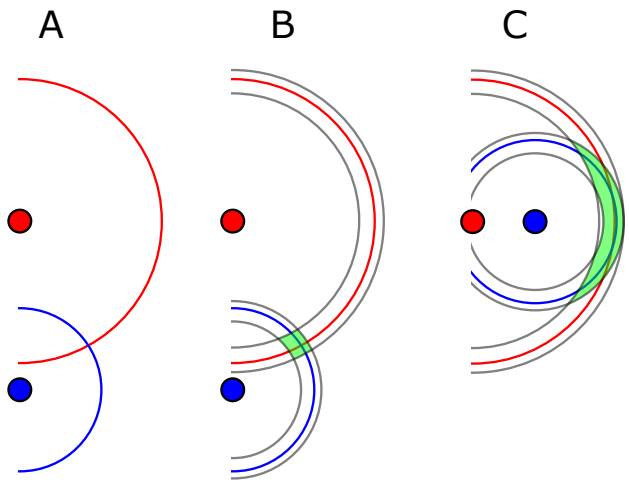
In order to improve the estimation accuracy under the constraint that only two BSs are available one idea is to utilize the frequency domain. Due to frequency-selective small-scale fading, it is expected that DoA estimates vary over subcarriers (equivalent to a LTE RB). However, it is known that one effect of large antenna systems is that small-scale fading is averaged out over the antenna dimension [BLM16]. This means that the frequency selectivity of the channel mainly depends on the frequency-selective large-scale fading. Therefore, gains in localization from utilizing the frequency domain in a LTE typical bandwidth of 20 MMHz are expected to be limited.

In principle, the idea is that the errors are distributed and an averaging should improve estimation accuracy. In the following two options for the averaging are considered:

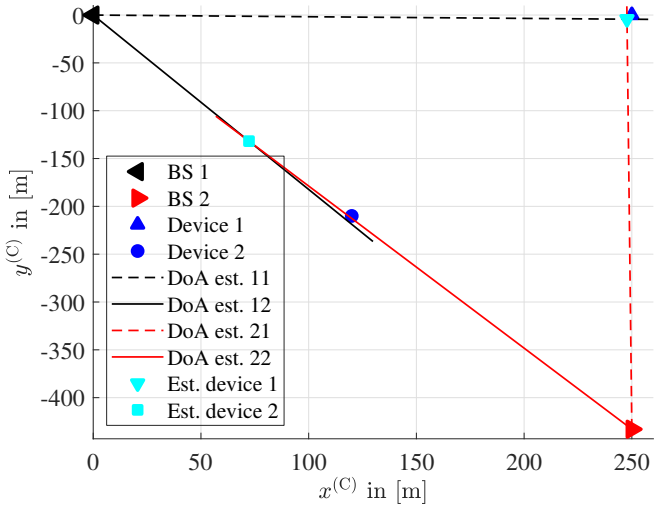
1. DoA estimates such that the single averaged value is used as input for the 3D localization algorithm described in Section 4.2.1.

2. Position estimates obtained on each subcarrier or RB independently are averaged to a final position estimate.

To be more precise, the assumption is that signals are received on $N_{(\mathrm{s})}$ subcarriers in the frequency domain at the BSs. With this definition, the following adaptation in notation is introduced:



(a) TDoA measurement with GNSS (Picture is taken from [Wik19]).



(b) The same DoA estimation errors for device 1 and device 2 of $\Delta_{(\alpha)}^{(1,1)} = \Delta_{(\alpha)}^{(1,2)} = 1°$ and $\Delta_{(\alpha)}^{(2,1)} = \Delta_{(\alpha)}^{(2,2)} = 0.3°$ are assumed at BS 1 and BS 2, respectively. Different geometry results in different position errors in given cellular deployment.

Figure 4.11.: Examples showing the GDoP phenomena, where the same uncertainty in DoA measurements results in different position errors.

Option 1) Average over DoAs: Let $\hat{\alpha}_{k,n_{(\mathrm{s})}}$ and $\hat{\beta}_{k,n_{(\mathrm{s})}}$ be the horizontal and vertical DoA estimate of source k on subcarrier $n_{(\mathrm{s})} \in [1, \ldots, N_{(\mathrm{s})}]$, the vector notation of the DoA estimates on the $N_{(\mathrm{s})}$ subcarriers of the k-th source is given as

$$\hat{\boldsymbol{\alpha}}_k^{(N_{(\mathrm{s})})} = \begin{pmatrix} \hat{\alpha}_{k,1} \\ \vdots \\ \hat{\alpha}_{k,N_{(\mathrm{s})}} \end{pmatrix} \text{ and } \hat{\boldsymbol{\beta}}_k^{(N_{(\mathrm{s})})} = \begin{pmatrix} \hat{\beta}_{k,1} \\ \vdots \\ \hat{\beta}_{k,N_{(\mathrm{s})}} \end{pmatrix}. \tag{4.25}$$

The final estimates after averaging are denoted by $\check{\alpha}^{(N_{(\mathrm{s})})}$ and $\check{\beta}^{(N_{(\mathrm{s})})}$, where the superscript $N_{(\mathrm{s})}$ indicates an averaging over $N_{(\mathrm{s})}$ subcarriers. The DoA estimates $\check{\alpha}^{(N_{(\mathrm{s})})}$ and $\check{\beta}^{(N_{(\mathrm{s})})}$ are obtained as the mean according to

$$\check{\alpha}_k^{(N_{(\mathrm{s})})} = \mathrm{mean}\left(\hat{\boldsymbol{\alpha}}_k^{(N_{(\mathrm{s})})}\right) = \frac{1}{N_{(\mathrm{s})}} \sum_{n_{(\mathrm{s})}=1}^{N_{(\mathrm{s})}} \hat{\alpha}_{k,n_{(\mathrm{s})}}, \tag{4.26}$$

and

$$\check{\beta}_k^{(N_{(\mathrm{s})})} = \mathrm{mean}\left(\hat{\boldsymbol{\beta}}_k^{(N_{(\mathrm{s})})}\right) = \frac{1}{N_{(\mathrm{s})}} \sum_{n_{(\mathrm{s})}=1}^{N_{(\mathrm{s})}} \hat{\beta}_{k,n_{(\mathrm{s})}}. \tag{4.27}$$

Note that $\check{\alpha}_k^{(N_{(\mathrm{s})})}$ and $\check{\beta}_k^{(N_{(\mathrm{s})})}$ are used as input of the k-th source for the 3D localization algorithm if averaging method a) is selected.

Option 2) Average over Positions: In option 2), the position is estimated independently on each subcarrier and according to Fig. 4.7 denoted by $\hat{\mathbf{u}}_{k,n}$ for source k on subcarrier n. Consequently, the estimated positions

over the $N_{(\mathrm{s})}$ subcarriers can be written in vector notation as

$$\hat{\mathbf{U}}_k^{(N_{(\mathrm{s})})} = \begin{pmatrix} \hat{\mathbf{u}}_{k,1} \\ \vdots \\ \hat{\mathbf{u}}_{k,N_{(\mathrm{s})}} \end{pmatrix}. \tag{4.28}$$

With this, the final position estimate is given by $\check{U}^{N_{(\mathrm{s})}}$ and obtained as the mean over $N_{(\mathrm{s})}$ subcarriers by

$$\check{U}_k^{(N_{(\mathrm{s})})} = \mathrm{mean}\left(\hat{\mathbf{U}}_k^{(N_{(\mathrm{s})})}\right) = \frac{1}{N_{(\mathrm{s})}} \sum_{n_{(\mathrm{s})}=1}^{N_{(\mathrm{s})}} \hat{\mathbf{u}}_{k,n_{(\mathrm{s})}}. \tag{4.29}$$

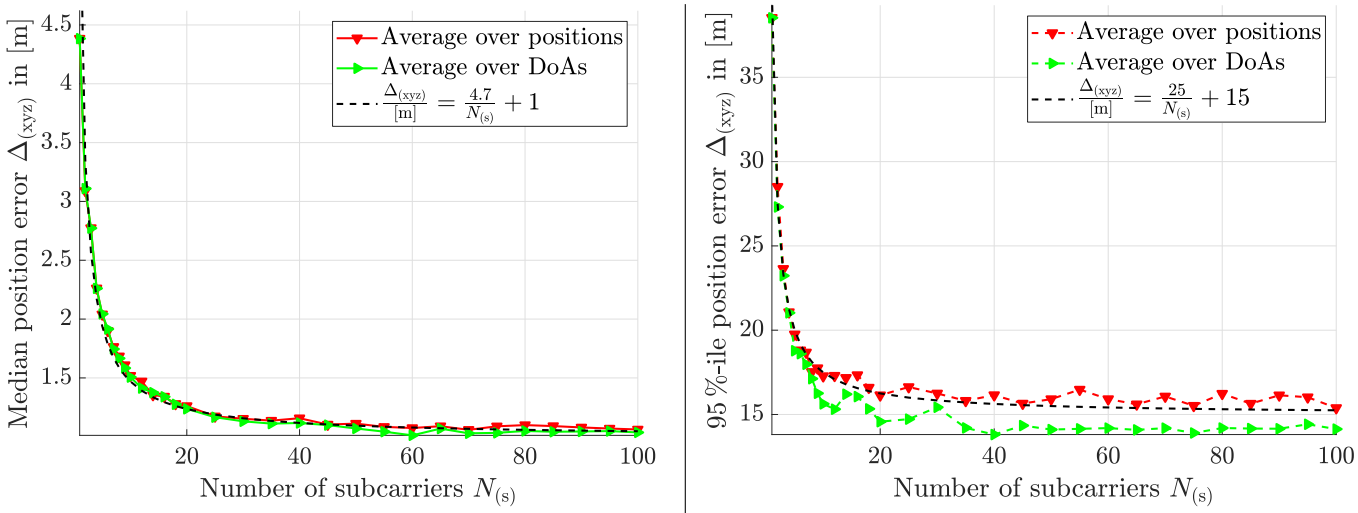
In Fig. 4.12, the performance results of both averaging options are given over the number of subcarriers. Here a subcarrier has the equivalent bandwidth of a LTE RB of 180 kHz and 100 subcarriers correspond to 18 MHz bandwidth. In Fig. 4.12a the 50 %-ile position error decreases approximately by $\propto \frac{1}{N_{(\mathrm{s})}}$ as can be seen by the overlap with the black dashed line which is parametrized as $\frac{4.7}{N_{(\mathrm{s})}} + 1$. This means, that already $N_{(\mathrm{s})} = 2$ improves the average performance significantly, e.g. in Fig. 4.12a by > 1 m. On the other hand at a certain point increasing $N_{(\mathrm{s})}$ does not decrease the position error further. Both averaging options show the same scaling, however, averaging over DoAs with option 1) results in a lower absolute position error, e.g. ≈ 0.3 cm at $N_{(\mathrm{s})} = 100$ in Eq. (4.26).

On the right hand side of Fig. 4.12 the corresponding 95 %-ile of the position error is given and the same $\propto \frac{1}{N_{(\mathrm{s})}}$ scaling over $N_{(\mathrm{s})}$ is observed. The gap between option 1) averaging over DoAs and option 2) averaging over positions is larger compared to the 50 %-ile performance, e.g. ≈ 1.2 m at $N_{(\mathrm{s})} = 100$. The averaging over DoAs is better due to the decrease of the DoA estimation error, which is an input to the position calculation. With a lower DoA estimation error, the distance dependent position error is lower and at the same time the effect from DoP is compensated, while averaging over positions the DoA estimation error is not decreased. Another advantage of option 1) is the lower complexity compared to option 2), e.g. position calculation is require only once, instead $N_{(\mathrm{s})}$ times as in option 1). The saturation at 15 m is given by the selected deployment. In Eq. (3.62) the position error is approximated assuming that the distance to the BS is known. Assuming a distance $d_k^{(\mathrm{BS})} > 250$ m to BS 2 and the 95 %-ile $\Delta_{(\alpha)} \approx 0.34°$ $\Delta_{(\beta)} = 0.19°$ this yields in a position error of $\Delta_{(\mathrm{xyz})} > 1.7$ m. However, due to the position calculation by the shortest distance of skew lines method there exist constellations, where the position error is smaller or larger, by subtractive or additive superposition of DoA errors, respectively. For example if in the given scenario BS 1 has estimated $\hat{\beta}_1 > \beta$ and BS 2 $\hat{\beta}_2 < \beta$, the center of the skew line method decreases the position error.

Finally, the distributions of the DoA estimation errors with DoA averaging over $N_{(\mathrm{s})}$ subcarriers is given in Fig. 4.13a observing a significant improvement compared to Fig. 4.9a[10]. The corresponding position error distribution in the $x^{(\mathrm{C})}$-$y^{(\mathrm{C})}$ plane is given in Fig. 4.13b.

Not discussed in this thesis is the utilization of bandwidth to improve the DoA estimate itself, instead of doing averaging as proposed in this section. There is certainly a trade-off because averaging over more than 20 subcarriers doesn't decrease the position error significantly. On the other, already a small bandwidth increase by averaging over two subcarriers provides significant position error reduction compared to one subcarrier. For such an trade-off investigation, the bandwidth needs to be utilized in the DoA estimation, e.g. by separating the LoS multi-path component from the others, however this is out of scope of this thesis.

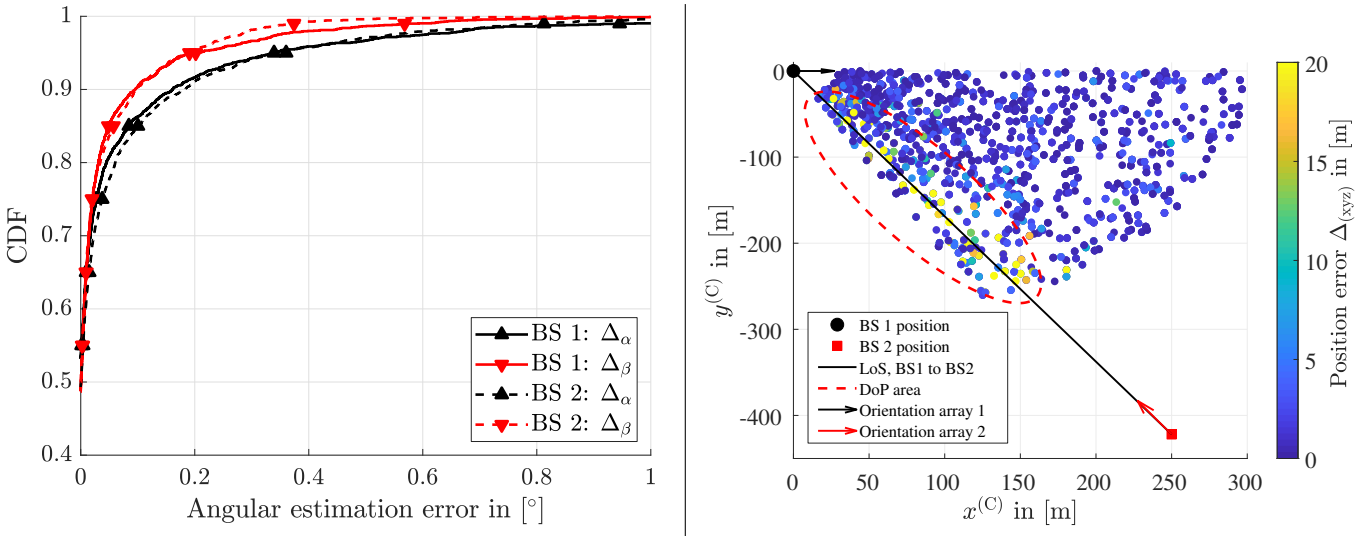
[10] Note that in Fig. 4.9a, without averaging, the error range is from 0° to 2° while in Fig. 4.13a, with averaging, the error range is from 0° to 1°.



(a) 50 %-ile position error from CDF.

(b) 95 %-ile position error from CDF.

Figure 4.12.: Performance of averaging DoA estimates and positions using method a) "Average over DoAs" and method b) "Average over positions", respectively. The averaging is done over $N_{(s)}$ OFDM resources in frequency domain where a subcarrier corresponds to the bandwidth of a LTE RB.



(a) CDF of DoA error.

(b) Position errors in the $x^{(C)}$-$y^{(C)}$ plane at the true source positions.

Figure 4.13.: Impact from GDoP. Performance on angular estimation errors and position error.

Option 3) Optimization of Weighting Factor $g_{(\mathrm{p})}$: Previous results have been obtained assuming $g_{(\mathrm{p})} = 1/2$ in Eq. (4.23) for the position estimate on the shortest distance between the two DoA estimates. The idea is that an optimization of the weighting factor $g_{(\mathrm{p})}$ takes into account estimation uncertainties, e.g. if a user equipment (UE) is much closer to BS one than to BS two $g_{(\mathrm{p})} < \frac{1}{2}$ can be selected such that the position estimate is closer to DoA one. Or if azimuth estimate one is close to zero degree it can be trusted more than azimuth estimate two above 70°, see Fig. 3.12b in the parameter study in Section 3.2.1. However, before deriving such guidelines or an algorithm, the potential improvement is checked. For evaluation on the achievable performance with optimized $g_{(\mathrm{p})}$ the source positions as in Fig. 4.8 and parameters from Table 3.2 are assumed. Fig. 4.14a shows the CDFs for

1. the minimum distance between the DoA estimate from BS 1 denoted by $\mathbf{s}^{(\mathrm{BS})}$ and the true device position $\mathbf{u}$ marked by upward pointing black trangles,
2. the minimum distance between the DoA estimate from BS 2 denoted by $\mathbf{t}^{(\mathrm{BS})}$ and the true device position $\mathbf{u}$ marked by downward pointing red triangles,
3. the minimum distance between the vector $\mathbf{v}^{(\mathrm{st})}$ and the true device position $\mathbf{u}$ marked by light green right pointing triangles, and
4. the position error $\Delta_{(\mathrm{xyz})}$ with $g_{(\mathrm{p})} = \frac{1}{2}$ marked by left pointing deep blue triangles.

Note that the minimum distance between $\mathbf{v}^{(\mathrm{st})}$ and $\mathbf{u}$ is the maximum achievable performance by optimizing $g_{(\mathrm{p})}$. Compared to the minimum distance between $\mathbf{v}^{(\mathrm{st})}$ and $\mathbf{u}$ the minimum distance between the DoA estimates $\mathbf{s}^{(\mathrm{BS})}$, $\mathbf{t}^{(\mathrm{BS})}$, and $\mathbf{u}$ is significant smaller. This confirms again the dilution of precision observation that the "accurate" individual estimates of each BS result in large position errors. In addition, the selected position calculation method by the shortest distance on skew lines is not the optimal solution for sources located around the LoS axis between two BSs.

However, even if not obvious in Fig. 4.14a from the difference between the position error and minimum distance between source and $\mathbf{v}^{(\mathrm{st})}$, the optimization of $g_{(\mathrm{p})}$ can provide significant improvements, especially in the case of near-far situations as given in Fig. 4.15. Under the assumption of similar DoA estimation error at BS 1 and BS 2 the larger distance from BS 2 causes a larger distance between $\tilde{\mathbf{t}}_p$ and the source compared to $\tilde{\mathbf{s}}_p$ and the source. Consequently, a constant weighting factor $g_{(\mathrm{p})} = \frac{1}{2}$ results in a larger estimation compared to $g_{(\mathrm{p})} < \frac{1}{2}$ which moves $\hat{\mathbf{u}}$ closer to $\tilde{\mathbf{s}}_p$ in the example of Fig. 4.15. According to Eq. (4.23), $g_{(\mathrm{p})}$ is in the interval $[0, 1]$, where the extreme cases are

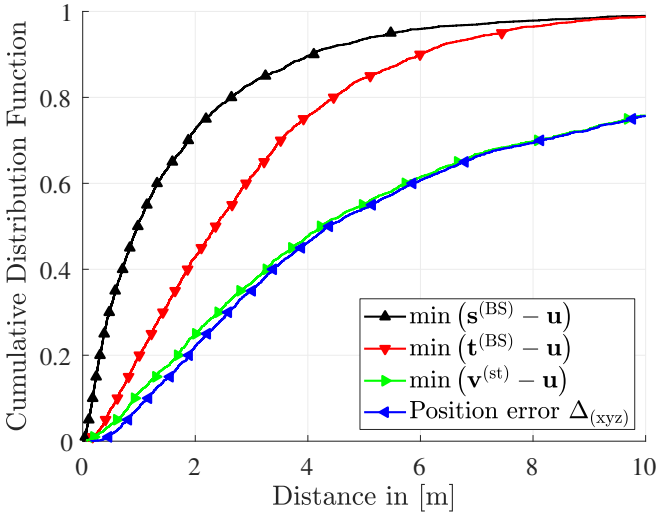
$$\hat{\mathbf{u}} = \begin{cases} \tilde{\mathbf{s}}^{(\mathrm{BS})} & \text{if } g_{(\mathrm{p})} = 0 \\ \tilde{\mathbf{t}}^{(\mathrm{BS})} & \text{if } g_{(\mathrm{p})} = 1. \end{cases} \tag{4.30}$$

The position error for the "shortest distance on skew lines" method is derived from the general position error as a starting point given as Eq. (4.31)

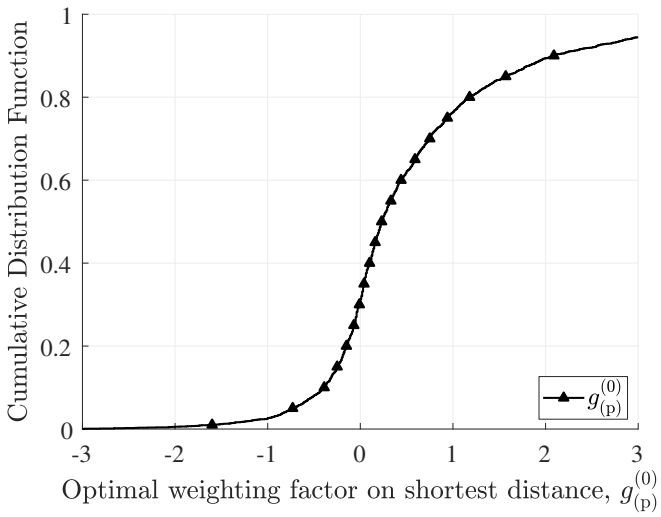
$$\Delta_{(\mathrm{xyz})} = \sqrt{\left(u_{(\mathrm{x})} - \hat{u}_{(\mathrm{x})}\right)^2 + \left(u_{(\mathrm{y})} - \hat{u}_{(\mathrm{y})}\right)^2 + \left(u_{(\mathrm{z})} - \hat{u}_{(\mathrm{z})}\right)^2}, \tag{4.31}$$

where $\mathbf{u}$ is given as

$$\mathbf{u} = \begin{pmatrix} u_{(\mathrm{x})} \\ u_{(\mathrm{y})} \\ u_{(\mathrm{z})} \end{pmatrix} = \begin{pmatrix} x_1^{(\mathrm{C})} + s^{(\mathrm{BS})} \cos\beta_1 \cos\alpha_1 \\ y_1^{(\mathrm{C})} + s^{(\mathrm{BS})} \cos\beta_1 \sin\alpha_1 \\ z_1^{(\mathrm{C})} + s^{(\mathrm{BS})} \sin\beta_1 \end{pmatrix} = \begin{pmatrix} x_2^{(\mathrm{C})} + t^{(\mathrm{BS})} \cos\beta_2 \cos\alpha_2 \\ y_2^{(\mathrm{C})} + t^{(\mathrm{BS})} \cos\beta_2 \sin\alpha_2 \\ z_2^{(\mathrm{C})} + t^{(\mathrm{BS})} \sin\beta_2 \end{pmatrix} \tag{4.32}$$

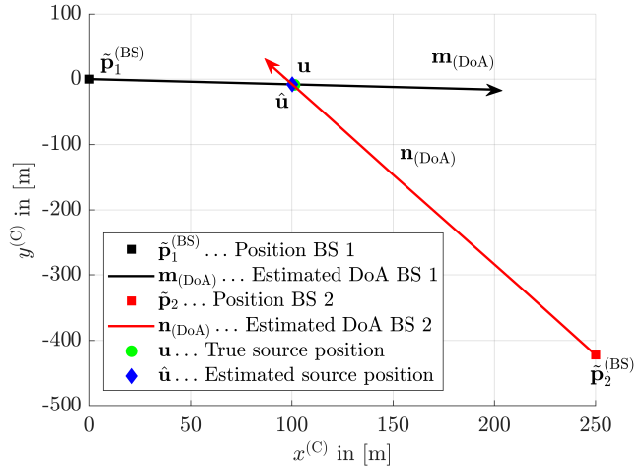


(a) Minimum distance between true source position and virtual points/lines.

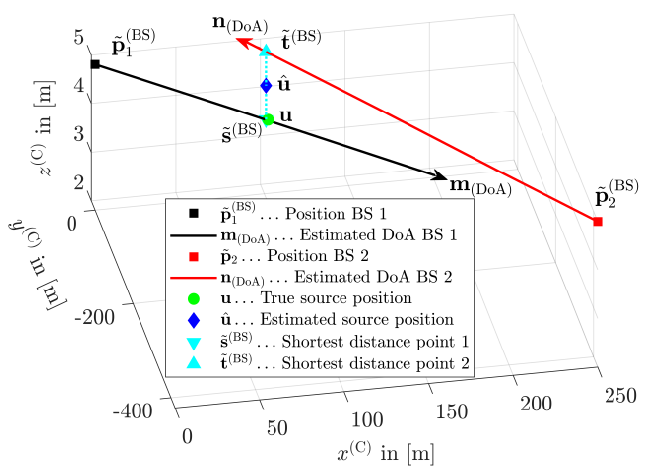


(b) Optimal selection of $g_{(p)}^{(0)}$ to minimize the position error according to Eq. (4.35).

Figure 4.14.: Performance results on position calculation based on DoA estimation and method "*shortest distance between skew lines*".



(a) Top view, BS 1 is closer to the source than BS 2.



(b) 3D view with the estimated source position based on the shortest distance on skew lines method.

Figure 4.15.: Position calculation based on DoAs estimates visualizing the near far effect with similar estimation errors and constant weighting factor $g_{(p)} = \frac{1}{2}$.

and $\hat{\mathbf{u}}$ as

$$\hat{\mathbf{u}} = \begin{pmatrix} \hat{u}_{(\mathrm{x})} \\ \hat{u}_{(\mathrm{y})} \\ \hat{u}_{(\mathrm{z})} \end{pmatrix} = \begin{pmatrix} s^{(\mathrm{BS})} m_{(\mathrm{x})} + x_1^{(\mathrm{C})} + g_{(\mathrm{p})} \frac{1}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \left(t^{(\mathrm{BS})} n_{(\mathrm{x})} + x_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{x})} - x_1^{(\mathrm{C})} \right) \\ s^{(\mathrm{BS})} m_{(\mathrm{y})} + y_1^{(\mathrm{C})} + g_{(\mathrm{p})} \frac{1}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \left(t^{(\mathrm{BS})} n_{(\mathrm{y})} + y_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{y})} - y_1^{(\mathrm{C})} \right) \\ s^{(\mathrm{BS})} m_{(\mathrm{z})} + z_1^{(\mathrm{C})} + g_{(\mathrm{p})} \frac{1}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \left(t^{(\mathrm{BS})} n_{(\mathrm{z})} + z_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{z})} - z_1^{(\mathrm{C})} \right) \end{pmatrix} = \begin{pmatrix} k_{(\mathrm{x})} + g_{(\mathrm{p})} l_{(\mathrm{x})} \\ k_{(\mathrm{y})} + g_{(\mathrm{p})} l_{(\mathrm{y})} \\ k_{(\mathrm{z})} + g_{(\mathrm{p})} l_{(\mathrm{z})} \end{pmatrix}. \tag{4.33}$$

The square root is a concave function, so the minimum of the term inside of the square root is also the minimum of square root corresponding to the position error. The derivative of the square of the position error with respect to $g_{(\mathrm{p})}$ is

$$\frac{\mathrm{d}\Delta_{(\mathrm{xyz})}}{\mathrm{d}g_{(\mathrm{p})}} = \frac{-2 \left(u_{(\mathrm{x})} l_{(\mathrm{x})} - k_{(\mathrm{x})} l_{(\mathrm{x})} - l_{(\mathrm{x})}^2 g_{(\mathrm{p})} + u_{(\mathrm{y})} l_{(\mathrm{y})} - k_{(\mathrm{y})} l_{(\mathrm{y})} - l_{(\mathrm{y})}^2 g_{(\mathrm{p})} + u_{(\mathrm{z})} l_{(\mathrm{z})} - k_{(\mathrm{z})} l_{(\mathrm{z})} - l_{(\mathrm{z})}^2 g_{(\mathrm{p})} \right)}{2 \left(\left(u_{(\mathrm{x})} - \hat{u}_{(\mathrm{x})} \right)^2 + \left(u_{(\mathrm{y})} - \hat{u}_{(\mathrm{y})} \right)^2 + \left(u_{(\mathrm{z})} - \hat{u}_{(\mathrm{z})} \right)^2 \right)}. \tag{4.34}$$

Eq. (4.34) is set to zero and solved to $g_{(\mathrm{p})}$ in

$$\begin{aligned} g_{(\mathrm{p})}^{(0)} =& s^{(\mathrm{BS})} \frac{\left(\left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{n}_{(\mathrm{DoA})} \right) \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \tilde{\mathbf{\Delta}} \right) - \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\mathbf{\Delta}} \right) \left(\mathbf{n}_{(\mathrm{DoA})} \cdot \mathbf{\Phi}_{(1)} \right)}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ & - s^{(\mathrm{BS})} \frac{\left(\left(\mathbf{m}_{(\mathrm{DoA})} \cdot \tilde{\mathbf{\Delta}} \right) - \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{n} \right) \left(\mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\mathbf{\Delta}} \right) \right) \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{\Phi}_{(1)} \right)}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ & + s^{(\mathrm{BS})} \frac{\left(1 - \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{n}_{(\mathrm{DoA})} \right)^2 \right) \left(\tilde{\mathbf{\Delta}} \cdot \mathbf{\Phi}_{(1)} \right)}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} \end{aligned} \tag{4.35}$$

where the details are given in Annex A and $\tilde{a}_{(\mathrm{h})} = 1 - a^2 = 1 - \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{n}_{(\mathrm{DoA})} \right)^2$. Eq. (4.35) is rearranged to

$$g_{(\mathrm{p})}^{(0)} = \frac{f_{(1)} t^{(\mathrm{BS})} \mathbf{\Delta}_{(2)}^{(\Phi)} - f_{(2)} s^{(\mathrm{BS})} \mathbf{\Delta}_{(1)}^{(\Phi)}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + \frac{f_{(1)} \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\mathbf{\Delta}}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + \frac{\left(s^{(\mathrm{BS})} \right)^2 - \left(t^{(\mathrm{BS})} \right)^2 + \tilde{\mathbf{\Delta}}^2}{2 \left| \mathbf{v}^{(\mathrm{st})} \right|}, \tag{4.36}$$

where $f_{(1)}$, $f_{(2)}$ are geometry depended factors which are obtained from the BS positions and DoA estimates, $\mathbf{\Delta}_{(1)}^{(\Phi)} = \mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{\Phi}_{(1)}$ and $\mathbf{\Delta}_{(2)}^{(\Phi)} = \mathbf{n}_{(\mathrm{DoA})} \cdot \mathbf{\Phi}_{(2)}$ are the joint DoA estimation errors from BS 1 and BS 2, and $\tilde{\mathbf{\Delta}} = \tilde{\mathbf{p}}_2^{(\mathrm{BS})} - \tilde{\mathbf{p}}_1^{(\mathrm{BS})}$. $g_{(\mathrm{p})}^{(0)}$ depends on four unknowns, the distances to the BSs $s^{(\mathrm{BS})}$, $t^{(\mathrm{BS})}$, and the DoA estimation errors. The distances can be obtained from the estimated position with a preselected $g_{(\mathrm{p})}$, however, with the still unknown DoA estimation errors a minimization of the position error by $g_{(\mathrm{p})}$ cannot be guaranteed. For the parameters in Table 4.2 the optimal $g_{(\mathrm{p})}^{(0)}$ is shown in Fig. 4.14b. With the given deployment, where sources in average are closer to BS 1, $g_{(\mathrm{p})} = 0.5$ is at 62 %. This means that the position estimate, which minimizes the position error, is in average closer to the DoA estimate of BS 1.

Nevertheless, as can be seen by the gap in Fig. 4.14a, the gain from an optimal selection of $g_{(\mathrm{p})}$ is less than 5 % at the median. This is due to the distribution of the sources and the geometry of the selected deployment. Fig. 4.16 shows the $x^{(\mathrm{c})}$, $y^{(\mathrm{c})}$, and $z^{(\mathrm{c})}$ components of the position error with $g_{(\mathrm{p})} = 0.5$ and the optimal $g_{(\mathrm{p})}^{(0)}$. It can be seen that in the investigated setup, mainly the $z^{(\mathrm{C})}$-component of the position error is optimized. On the other hand, the $z^{(\mathrm{C})}$-component of the position error is also to smallest component so the total position error reduction is less than 5 % at the median.

Even in other scenarios, e.g. where the $x^{(\mathrm{C})}$, $y^{(\mathrm{C})}$, and $z^{(\mathrm{C})}$ component of the position error are equal, the gain from an optimal selection of $g_{(\mathrm{p})}$ even with perfect knowledge is marginal. The degree of freedom for optimization is along the vector $\mathbf{v}^{(\mathrm{st})}$ given by the shortest distance between the DoA estimates.

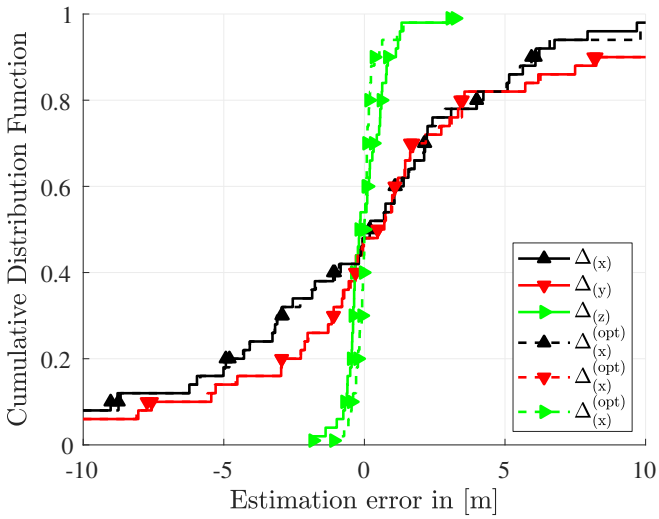


Figure 4.16.: Distribution of $x^{(C)}$, $y^{(C)}$, and $z^{(C)}$ components of position error. Dashed lines represent the errors with optimal $g_{(p)}$ and solid lines with the default value $g_{(p)} = 0.5$.

4.2.3. Conclusions

This section shows that the position of mobile devices can be determined from DoA estimates and the position of two BSs, utilizing the high DoA estimation accuracy of massive MIMO arrays, as shown in previous Chapter 3. Even with a realistic number of antennas, e.g. in this thesis a $[8, 8]$ array is considered, position errors below 1 m can be obtained in large regions of the cellular network. With the minimum of two BSs, the position calculation method "shortest distance on skew lines" is investigated in this section.

With the "shortest distance on skew lines" position method applied in cellular networks, DoP caused by the geometry of the deployment can result in large position errors in certain areas, despite small DoA estimation errors. In this thesis the frequency domain is utilized by averaging over estimated DoAs or positions derived independently on different subcarriers to decrease position errors. It is shown that such averaging is efficient and already 10 subcarriers, where each subcarrier is assumed as a LTE RB of 180 kHz, reduce most of the DoP caused position errors.

Furthermore, the optimization of the position calculation method "shortest distance between skew lines" is studied in this thesis. The optimal parameter $g_{(p)}$ is obtained that minimizes the position error. However, the gain from this optimization is limited and other side-information has to be utilized to further minimize the position error.

One open issue for future work is the extension to more than two DoA estimates to further decrease the impact from geometry caused DoP in cellular networks. In addition, the probability that devices have a LoS connection to several BSs is expected to be low. This also leads to the challenge to detect ambiguities in the network, e.g. the BS does not know if the received signal is a LoS or NLoS component. Therefore, the reliable classification of multi-path components in to LoS or NLoS is of great interest. For example, if a signal is received from an angle of $> 90°$ or $< -90°$ in the horizontal plane, but the search range is from $[-70, 70]°$, a DoA in this interval is estimated which results in large position errors.

5. Conclusions

In this work, the application of centralized massive multiple-input multiple-output (MIMO) antenna arrays in cellular networks is studied[1]. Thereby, the focus has been on two aspects, downlink multiple-user transmission and uplink DoA estimation. While Chapter 2 and Chapter 3 investigate fundamental trade-offs, considering a single base station (BS), in Chapter 4, challenges that arise in cellular networks for both aspects are studied.

Section 2.3.3 shows that with a large number of antennas the distribution of transmit power becomes approximately equal over all antennas, at least for the most commonly used precoding strategies maximum ratio transmission (MRT), signal to leakage and noise ratio (SLNR), zero forcing (ZF) and minimum mean square error (MMSE). Consequently, power ranges of amplifiers in the BS can be significantly reduced, compared to BSs with only few antennas. Furthermore, significant transmit power loss due to per antenna power constraint (PAPC) and required precoder normalization is observed in massive MIMO systems. While PAPC is essential for transmission with a low number of antennas, to keep requirements on the power range and costs for amplifiers reasonable, in massive MIMO, PAPC can be removed without increasing amplifier requirements. Indeed for a constant transmit power, the output-power per antenna decreases proportional with the number of antennas so requirements on the output power of the amplifiers are lower compared to Long Term Evolution (LTE) BSs.

Section 2.3.4 demonstrates that additional users in massive MIMO systems provide significant sum spectral efficiency gains, if the number of spatial multiplexed streams is $T > \frac{N}{2}$. Increasing the number of users further requires user grouping techniques. The investigation with semi-orthogonal user selection (SUS) shows, that sum spectral efficiency gains are not guaranteed by channel orthogonality based user grouping. Therefore, user grouping based on a sum spectral efficiency criteria is necessary and the effectiveness is demonstrated by the projection based zero forcing (PBZF) algorithm. The additional complexity of user grouping can be compensated by applying the scheduling decision from a single resource block (RB) to the complete system bandwidth, because precoded channels become frequency-flat due to the massive MIMO channel hardening effect. However, massive MIMO channel hardening only kicks in with phase adaptive precoders such as MRT, SLNR, ZF or MMSE. In contrast to these precoders, with 3rd Generation Partnership Project (3GPP) Release 15 codebook precoding the channel remains highly frequency-selective.

In [BLM16] a critical questions for massive MIMO is asked: "Can Massive MIMO Work in frequency division duplex (FDD) Operation?". Section 2.4 answers this question with yes. The key solution for massive MIMO in FDD is hybrid precoding. Thereby, the main challenge is to find the right balance between loss due to pilot overhead and gain from spatial multiplexing under a feedback constraint. The combination of codebook based first-stage precoding and effective channel aware second-stage precoding can provide this balance. Using "sub-codebook splitting" for first-stage precoding, proposed in Section 2.4.2, any number of BS antennas can be used while providing a design parameter that adjusts the number of active streams. Controlling the number of streams corresponds to controlling the pilot overhead, e.g. based on the number of active users in the system. The "sub-codebook splitting" technique provides the same high channel quantization of a large codebook, achieving similar spectral efficiency as MRT, while keeping the size of the effective channel small. Numerical simulations in Section 2.4.4 show, up to 10 users, a sum spectral efficiency close to full digital MMSE precoding. Also, if the number of users exceeds the size of the sub-codebooks,

[1] In Chapter 2 the default assumption is a 10×10 uniform planar array (UPA) and in Section 2.3 investigation up to 300 antennas is included. In Chapter 3 for direction of arrival (DoA) estimation the default assumption is an 8×8 UPA. In Chapter 4 both UPA sizes are assumed in the respective sub-sections.

the performance with the proposed hybrid-precoding saturates. This limitation can be overcome by a larger sub-codebook size, which can be applied to scenarios with a larger coherence block size, meaning a larger coherence bandwidth and/or coherence time.

Finally, as an outcome of the discussion on advantages and disadvantages of time division duplex (TDD) versus FDD it turned out that the shorter the coherence block and the larger the number of antennas, the more beneficial TDD is in terms of sum spectral efficiency compared to FDD. However, in scenarios with a larger coherence block, e.g. for fixed wireless access, FDD can achieve similar sum spectral efficiency as TDD with the proposed sub-codebook based hybrid-precoding technique and under the assumptions large uplink capacity for channel state information (CSI) feedback. Additional, with a large coherence time the feedback interval can be reduced accordingly and the uplink capacity demand for CSI feedback scales inverse to the feedback interval.

While in Section 2.4.4 a feedback channel with infinite capacity is considered, in Section 2.4.5 the trade-off between feedback size and sum spectral efficiency is studied. Thereof, it is shown that explicit CSI feedback is essential for multiple user downlink transmission. Results in Section 2.4.5 demonstrate that explicit CSI together with hybrid precoding, achieves a larger sum spectral efficiency than pure codebook based precoding without feedback constraint, even for a feedback constraint as in fifth generation (5G) new radio (NR) Release 15 of 266 bit. Explicit CSI in combination with hybrid precoding means, that the effective channel, after multiplication with the first-stage precoder, is transmitted back to the BS. The study on explicit CSI reduction under feedback constraint takes into account the number of reported RB in frequency domain, the bits used for quantization, and the sub-codebook size. Extensive numerical simulations show a strong coupling among these three parameters. This means, that a small change of one of the parameters has significant impact on sum spectral efficiency. Accordingly, these parameters have to be selected jointly and carefully and further research on the topic is required.

Independent of TDD or FDD, the sum spectral efficiency of downlink spatial multiplexing with massive MIMO suffers from inter-cell interference. On top of this, Section 4.1 points out that scheduling decision without knowledge about the interference situation on the user side, can result to additional sum spectral efficiency loss. Thus, a low overhead feedback, consisting of the wideband interference power received at the mobile users, is proposed and evaluated. The proposed feedback provides up to 100 % sum spectral efficiency gain with PBZF scheduling and up to 400 % for random selection.

An equally important contribution of this thesis is, that massive MIMO arrays can also be used to increase localization accuracy in cellular systems compared to fourth generation (4G). DoA estimation in line of sight (LoS) scenario can achieve accuracy below 1 m position error over a wide range of parameters, see Section 3.2. A major practical challenge for search based DoA estimation schemes is the required complexity in terms of power spectrum computations. A high quantization of the search space is necessary for a low DoA estimation error. The proposed adaptive search space quantization (ASSQ) scheme in Section 3.2.2, significantly reduces the required complexity with approximately no performance loss, e.g. from 1 442 401 to 195. Further complexity reduction increases the DoA estimation errors due to coarse quantization in the first step. This estimation errors can be mostly compensated by the proposed adaptive antenna selection modification of the ASSQ scheme, see Section 3.2.2. The general rule is, the coarser the quantization of the search space, the less antennas should be used in the respective ASSQ step.

Massive MIMO not only offers high accurate DoA estimation, but also the ability to detect multiple users at the same time. Theoretically, this can be done on any kind of transmitted signal, e.g. on a scheduled uplink data transmission of multiple users. Yet, no literature was found on scheduling of users for multiple-source DoA estimation with the aim of reducing DoA estimation errors. For that reasons, a two-step algorithm is developed and evaluated in Section 3.3. A take-away learned in Section 3.3.1 is, that the difference in received signal power between users is the major source of DoA estimation errors in joint detection. Consequently, the first step of the grouping algorithm divides users in “power-groups” with similar receive power. Then, the second step selects users based on minimum angular distance criteria. The derived input parameters of the grouping algorithm allow a scaling of either the DoA estimation error or the number of

jointly detected users. Note, that both scale inverse to each other, so decreasing the DoA estimation errors results in less jointly detected users.

An additional result of the evaluation in Section 3.3.2 is, that the DoA estimation error distribution of searched based methods can be closely approximated by a t-location-scale probability distribution, see Fig. 3.39a. This simplifies future testing and development of robust algorithms based on DoA measurements, because DoA estimation errors can directly be emulated by t-location-scale probability distributions.

In Section 4.2, the position of mobile users in a cellular network is obtained, based only on DoA estimates. The minimum requirements for DoA based positioning are DoA estimates of the same user from two different BSs, their corresponding positions and their array orientations. With this, the details of the "minimum distance between skew lines" technique are derived and evaluated in Section 4.2.1. This study shows that, similar to satellite systems, geometry dilution of precision (GDoP) causes large position errors also in cellular systems, independent of the DoA estimation error. There are two compensation techniques proposed in Section 4.2.2, one is averaging of DoAs, the other is averaging of positions. Both methods reduce a significant part of the GDoP based position errors. Furthermore, both techniques are compatible with the frame structure of current 5G NR Release 15 so they can directly be applied in areas, or for applications, where low position errors are required or needed, e.g. for cellular assisted guidance like automated driving. Finally, the parameter that defines the middle of the minimum distance between the two skew lines is optimized in order to minimize the position error. However, this optimization delivers only marginal reduction of position errors in the investigated cellular network deployment.

Future work

The investigation in Section 2.4 on sum spectral efficiency under feedback constraint for multiple-user hybrid precoding systems has been carried out by numerical simulations. Therefore, only a part of the parameter space could be searched. Optimal parameter sets for a given feedback constraint are still unknown. Thus, further research in this direction is required before massive MIMO can be used in FDD systems. In order to achieve sum spectral efficiency optimal working points, analytical frameworks are needed to avoid numerical simulations over the large parameter spaces. Such numerical simulations of massive MIMO systems are computational expensive and require huge memory.

Another direction of future interest is the investigation of other CSI compression schemes. In this thesis, straightforward scalar linear quantization is applied to the complex channel coefficients, combined with the sub-codebook size and the number of RBs that are reported. This is not the optimal compression scheme and it is of high relevance for massive MIMO in FDD to search for other scalable CSI compression schemes.

Furthermore, in this thesis only a per user feedback constraint is considered. However, in practical systems, all users share the same feedback channel in the uplink. Therefore, a sum-feedback constraint over all users has to be taken into account for future system design. This leads to the following trade-off. Few users can report high quality CSI or many users can report low quality CSI. In terms of sum spectral efficiency it is not clear where the system should operate.

As inter-cell interference is limiting sum spectral efficiency in 5G systems (same as in 3G and 4G) the reduction of inter-cell interference, preferable based on algorithms, will be one of the most important topics for cellular system design in future. In this thesis, existing interference has been taken into account to adapt the downlink precoder, but no interference reduction or cancellation is investigated. One idea for future research to reduce inter-cell interference is the following. The proposed sub-codebooks can be used for spatial inter-cell-interference coordination similar as in interference-floor shaping [TKJ+15]. For example, BSs can exchange angular ranges to constraint the design of codebooks to a certain spatial direction. Using these coordinated codebooks for first-stage precoding, coordination with other BSs can be achieved, while each BS can independently selected users for second-stage precoding.

In the DoA estimation topic the following open issues remain. The parameter search of the ASSQ scheme is done numerically and it is not clear if parameters other than given in Table 3.5, can achieve a lower complexity. It is also possible the $S = 6$ further decreases the complexity. Therefore, an analytical solution of Eq. (3.71) to find ASSQ parameters is preferred and open for future research.

During discussions on the multiple source DoA estimation, it was suggested to use successive cancellation of already detected sources on the power spectrum, before searching the next source. This approach is similar to the well-known successive interference cancellation (SIC) receiver. In this thesis, the multiple source estimation is done by detecting the i strongest peaks.

For positioning and DoA estimation in this thesis, always a LoS channel has been assumed, as an essential requirement. Yet, in real system this is not always fulfilled. Indeed, in urban scenario the majority of connections is non line of sight (NLoS). Note that in case of NLoS, the DoA estimation algorithm is working in the same way and detects the direction of the strongest path. Therefore, the reliable classification of multi-path components into LoS or NLoS is of high interest, in order to decide if an estimated DoA, from a massive MIMO BS, can be used for position calculation or should be dropped.

Appendices

A. Rearrangements of Equations

Equation (3.42)

$$
\begin{aligned}
P_{\mathrm{CRB}} &= 2\left(\frac{f_{(\mathrm{c})}}{v_{(\mathrm{c})}}\right)^2 \frac{T^{(\mathrm{seq})} N s^{(\mathrm{p})}}{n^{(\mathrm{p})}} \left(1 - \frac{n^{(\mathrm{p})}}{s^{(\mathrm{p})} N + n^{(\mathrm{p})}}\right) = 2\left(\frac{2\pi}{\lambda}\right)^2 \frac{T^{(\mathrm{seq})} N s^{(\mathrm{p})}}{n^{(\mathrm{p})}} \left(1 - \frac{n^{(\mathrm{p})}}{s^{(\mathrm{p})} N + n^{(\mathrm{p})}}\right) \\
&= 2\left(\frac{2\pi}{\lambda}\right)^2 \frac{T^{(\mathrm{seq})} N s^{(\mathrm{p})}}{n^{(\mathrm{p})}} \left(\frac{s^{(\mathrm{p})} N}{s^{(\mathrm{p})} N + n^{(\mathrm{p})}}\right) = 2\left(\frac{2\pi}{\lambda}\right)^2 \frac{T^{(\mathrm{seq})} N^2 \left(s^{(\mathrm{p})}\right)^2}{n^{(\mathrm{p})}\left(s^{(\mathrm{p})} N + n^{(\mathrm{p})}\right)} \\
&= 2\left(\frac{2\pi}{\lambda}\right)^2 \frac{T^{(\mathrm{seq})} N^2 \left(s^{(\mathrm{p})}\right)^2}{\left(n^{(\mathrm{p})}\right)^2 \left(\frac{s^{(\mathrm{p})} N}{n^{(\mathrm{p})}} + 1\right)} = 2\left(\frac{2\pi}{\lambda}\right)^2 \frac{T^{(\mathrm{seq})} N^2 \tilde{\gamma}^2}{N\tilde{\gamma} + 1}
\end{aligned}
$$

Equation (4.21)

$$
\begin{aligned}
t^{(\mathrm{BS})} &= \frac{e_{(\mathrm{h})} \frac{t^{(\mathrm{BS})} a_{(\mathrm{h})} + c_{(\mathrm{h})}}{b_{(\mathrm{h})}} - f_{(\mathrm{h})}}{d_{(\mathrm{h})}} \\
t^{(\mathrm{BS})} d_{(\mathrm{h})} &= t^{(\mathrm{BS})} \frac{a_{(\mathrm{h})} e_{(\mathrm{h})}}{b_{(\mathrm{h})}} + \frac{c_{(\mathrm{h})} e_{(\mathrm{h})}}{b_{(\mathrm{h})}} - f_{(\mathrm{h})} \\
t^{(\mathrm{BS})} \frac{d_{(\mathrm{h})} b_{(\mathrm{h})} - a_{(\mathrm{h})} e_{(\mathrm{h})}}{b_{(\mathrm{h})}} &= \frac{c_{(\mathrm{h})} e_{(\mathrm{h})} - f_{(\mathrm{h})} b_{(\mathrm{h})}}{b_{(\mathrm{h})}} \qquad (\mathrm{A.1}) \\
t^{(\mathrm{BS})} &= \frac{c_{(\mathrm{h})} e_{(\mathrm{h})} - f_{(\mathrm{h})} b_{(\mathrm{h})}}{d_{(\mathrm{h})} b_{(\mathrm{h})} - a_{(\mathrm{h})} e_{(\mathrm{h})}} \overset{\substack{a_{(\mathrm{h})}=e_{(\mathrm{h})} \\ b_{(\mathrm{h})}=1_{(\mathrm{h})} \\ d_{(\mathrm{h})}=1}}{=} \frac{c_{(\mathrm{h})} a_{(\mathrm{h})} - f_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} \\
s^{(\mathrm{BS})} &= \frac{t^{(\mathrm{BS})} a_{(\mathrm{h})} + c_{(\mathrm{h})}}{b_{(\mathrm{h})}} = \frac{\frac{c_{(\mathrm{h})} a_{(\mathrm{h})} - f_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} a_{(\mathrm{h})} + c_{(\mathrm{h})}}{b_{(\mathrm{h})}} = \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} \qquad (\mathrm{A.2})
\end{aligned}
$$

Equations on Derivative in Section Section 4.2.2 on Optimization of Weighting factor $g_{(\mathrm{p})}$ for Position Estimation with Shortest Distance between Skew Lines

It is assumed that in the following derivations related to the optimization of the weighting factor $g_{(\mathrm{p})}$ all parameters and variables are given in [m]. The true position $\mathbf{u}$ can be obtained from the position, the error free DoA and distance from both BS 1 and BS 2 given by

$$
\mathbf{u} = \begin{pmatrix} u_{(\mathrm{x})} \\ u_{(\mathrm{y})} \\ u_{(\mathrm{z})} \end{pmatrix} = \begin{pmatrix} x_1^{(\mathrm{C})} + s^{(\mathrm{BS})} \cos\beta_1 \cos\alpha_1 \\ y_1^{(\mathrm{C})} + s^{(\mathrm{BS})} \cos\beta_1 \sin\alpha_1 \\ z_1^{(\mathrm{C})} + s^{(\mathrm{BS})} \sin\beta_1 \end{pmatrix} = \begin{pmatrix} x_1^{(\mathrm{C})} + s^{(\mathrm{BS})} \Phi_{(1)}^{(x)} \\ y_1^{(\mathrm{C})} + s^{(\mathrm{BS})} \Phi_{(1)}^{(y)} \\ z_1^{(\mathrm{C})} + s^{(\mathrm{BS})} \Phi_{(1)}^{(z)} \end{pmatrix} \qquad (\mathrm{A.3})
$$

and

$$\mathbf{u} = \begin{pmatrix} u_{(\mathrm{x})} \\ u_{(\mathrm{y})} \\ u_{(\mathrm{z})} \end{pmatrix} = \begin{pmatrix} x_2^{(\mathrm{C})} + t^{(\mathrm{BS})} \cos\beta_2 \cos\alpha_2 \\ y_2^{(\mathrm{C})} + t^{(\mathrm{BS})} \cos\beta_2 \sin\alpha_2 \\ z_2^{(\mathrm{C})} + t^{(\mathrm{BS})} \sin\beta_2 \end{pmatrix} = \begin{pmatrix} x_2^{(\mathrm{C})} + t^{(\mathrm{BS})} \Phi_{(2)}^{(x)} \\ y_2^{(\mathrm{C})} + t^{(\mathrm{BS})} \Phi_{(2)}^{(y)} \\ z_2^{(\mathrm{C})} + t^{(\mathrm{BS})} \Phi_{(2)}^{(z)} \end{pmatrix}, \tag{A.4}$$

respectively, where

$$\mathbf{\Phi}_{(1)} = \begin{pmatrix} \Phi_{(1)}^{(x)} \\ \Phi_{(1)}^{(y)} \\ \Phi_{(1)}^{(z)} \end{pmatrix} \text{ and } \mathbf{\Phi}_{(2)} = \begin{pmatrix} \Phi_{(2)}^{(x)} \\ \Phi_{(2)}^{(y)} \\ \Phi_{(2)}^{(z)} \end{pmatrix}. \tag{A.5}$$

The position based on the DoA estimated of the two BSs using the "shortest distance between skew lines method" is given by

$$\hat{\mathbf{u}} = \tilde{\mathbf{s}}_{\mathrm{t}} + g_{(\mathrm{p})} \tilde{\mathbf{v}}_{(\mathrm{n})}^{(\mathrm{st})}, \tag{A.6}$$

where

$$\tilde{\mathbf{v}}_{(\mathrm{n})}^{(\mathrm{st})} = \frac{\mathbf{v}^{(\mathrm{st})}}{\left|\mathbf{v}^{(\mathrm{st})}\right|}[\mathrm{m}], \tag{A.7}$$

$$\mathbf{v}^{(\mathrm{st})} = \begin{pmatrix} t^{(\mathrm{BS})} n_{(\mathrm{x})} + x_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{x})} - x_1^{(\mathrm{C})} \\ t^{(\mathrm{BS})} n_{(\mathrm{y})} + y_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{y})} - y_1^{(\mathrm{C})} \\ t^{(\mathrm{BS})} n_{(\mathrm{z})} + z_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{z})} - z_1^{(\mathrm{C})} \end{pmatrix}, \tag{A.8}$$

and

$$\tilde{\mathbf{s}}_{\mathrm{t}} = \begin{pmatrix} s m_{(\mathrm{x})} + x_1^{(\mathrm{C})} \\ s m_{(\mathrm{y})} + y_1^{(\mathrm{C})} \\ s m_{(\mathrm{z})} + z_1^{(\mathrm{C})} \end{pmatrix}. \tag{A.9}$$

By substitution of $\tilde{\mathbf{s}}_{\mathrm{t}}$, $\mathbf{v}_{(\mathrm{n})}^{(\mathrm{st})}$, and $\mathbf{v}^{(\mathrm{st})}$ in Eq. (A.6) with Eq. (A.7), Eq. (A.8), and Eq. (A.9), respectively, yields

$$\hat{\mathbf{u}} = \begin{pmatrix} \hat{u}_{(\mathrm{x})} \\ \hat{u}_{(\mathrm{y})} \\ \hat{u}_{(\mathrm{z})} \end{pmatrix} = \begin{pmatrix} s m_{(\mathrm{x})} + x_1^{(\mathrm{C})} + g_{(\mathrm{p})} \frac{1}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \left(t^{(\mathrm{BS})} n_{(\mathrm{x})} + x_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{x})} - x_1^{(\mathrm{C})} \right) = k_{(\mathrm{x})} + g_{(\mathrm{p})} l_{(\mathrm{x})} \\ s m_{(\mathrm{y})} + y_1^{(\mathrm{C})} + g_{(\mathrm{p})} \frac{1}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \left(t^{(\mathrm{BS})} n_{(\mathrm{y})} + y_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{y})} - y_1^{(\mathrm{C})} \right) = k_{(\mathrm{y})} + g_{(\mathrm{p})} l_{(\mathrm{y})} \\ s m_{(\mathrm{z})} + z_1^{(\mathrm{C})} + g_{(\mathrm{p})} \frac{1}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \left(t^{(\mathrm{BS})} n_{(\mathrm{z})} + z_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{z})} - z_1^{(\mathrm{C})} \right) = k_{(\mathrm{z})} + g_{(\mathrm{p})} l_{(\mathrm{z})} \end{pmatrix}, \tag{A.10}$$

where $k_{(\mathrm{x})}, k_{(\mathrm{y})}, k_{(\mathrm{z})}, l_{(\mathrm{x})}, l_{(\mathrm{y})}$, and $l_{(\mathrm{z})}$ are auxiliary variable aggregating variables that are independent of $g_{(\mathrm{p})}$ by

$$\begin{aligned} k_{(\mathrm{x})} =& s^{(\mathrm{BS})} m_{(\mathrm{x})} + x_1^{(\mathrm{C})} \\ k_{(\mathrm{y})} =& s^{(\mathrm{BS})} m_{(\mathrm{y})} + y_1^{(\mathrm{C})} \\ k_{(\mathrm{z})} =& s^{(\mathrm{BS})} m_{(\mathrm{z})} + z_1^{(\mathrm{C})} \\ l_{(\mathrm{x})} =& \frac{1}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \left(t^{(\mathrm{BS})} n_{(\mathrm{x})} + x_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{x})} - x_1^{(\mathrm{C})} \right) \\ l_{(\mathrm{y})} =& \frac{1}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \left(t^{(\mathrm{BS})} n_{(\mathrm{y})} + y_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{y})} - y_1^{(\mathrm{C})} \right) \\ l_{(\mathrm{z})} =& \frac{1}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \left(t^{(\mathrm{BS})} n_{(\mathrm{z})} + z_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{z})} - z_1^{(\mathrm{C})} \right). \end{aligned} \tag{A.11}$$

The position error in general is the distance between the true position $\mathbf{u}$ and the estimated position $\hat{\mathbf{u}}$ given by

$$\Delta_{(\mathrm{xyz})} = \sqrt{\left(u_{(\mathrm{x})} - \hat{u}_{(\mathrm{x})} \right)^2 + \left(u_{(\mathrm{y})} - \hat{u}_{(\mathrm{y})} \right)^2 + \left(u_{(\mathrm{z})} - \hat{u}_{(\mathrm{z})} \right)^2}. \tag{A.12}$$

In order to minimize the position error by finding the optimal $g_{(\mathrm{p})}$, the extreme point in Eq. (A.12) is searched. The differentiation of Eq. (A.12) with respect to $g_{(\mathrm{p})}$ is obtained by using the chain rule in Leibniz's notation according to

$$\Delta'_{\mathrm{xyz}} = \frac{\mathrm{d}\Delta_{(\mathrm{xyz})}}{\mathrm{d}g_{(\mathrm{p})}} = \frac{\mathrm{d}\sqrt{v_{(\mathrm{h})}}}{\mathrm{d}v_{(\mathrm{h})}} \frac{\mathrm{d}v_{(\mathrm{h})}}{\mathrm{d}g_{(\mathrm{p})}} = \frac{1}{2v_{(\mathrm{h})}} \frac{\mathrm{d}v_{(\mathrm{h})}}{\mathrm{d}g_{(\mathrm{p})}} \tag{A.13}$$

However, since the square root is a concave function and obtaining $g_{(\mathrm{p})}$ in Eq. (A.12) that yields the minimum, it is sufficient to consider the argument under the square root to find $g_{(\mathrm{p})}$ that yields the minimum position error. Therefore, only the second product in Eq. (A.13) is rewritten as

$$\begin{aligned} \frac{\mathrm{d}v_{(\mathrm{h})}}{\mathrm{d}g_{(\mathrm{p})}} &= \frac{\mathrm{d}\left(\left(u_{(\mathrm{x})} - \hat{u}_{(\mathrm{x})}\right)^2 + \left(u_{(\mathrm{y})} - \hat{u}_{(\mathrm{y})}\right)^2 + \left(u_{(\mathrm{z})} - \hat{u}_{(\mathrm{z})}\right)^2\right)}{\mathrm{d}g_{(\mathrm{p})}} \\ &= \frac{\mathrm{d}\left(u_{(\mathrm{x})} - \hat{u}_{(\mathrm{x})}\right)^2}{\mathrm{d}g_{(\mathrm{p})}} + \frac{\mathrm{d}\left(u_{(\mathrm{y})} - \hat{u}_{(\mathrm{y})}\right)^2}{\mathrm{d}g_{(\mathrm{p})}} + \frac{\mathrm{d}\left(u_{(\mathrm{z})} - \hat{u}_{(\mathrm{z})}\right)^2}{\mathrm{d}g_{(\mathrm{p})}}. \end{aligned} \tag{A.14}$$

The three summands in Eq. (A.14) have the same structure and the first summand is differentiated exemplary as

$$\begin{aligned} \frac{\mathrm{d}\left(u_{(\mathrm{x})} - \hat{u}_{(\mathrm{x})}\right)^2}{\mathrm{d}g_{(\mathrm{p})}} &= \frac{\mathrm{d}\left(w_{(\mathrm{h})}\right)^2}{\mathrm{d}w_{(\mathrm{h})}} \frac{\mathrm{d}w_{(\mathrm{h})}}{\mathrm{d}g_{(\mathrm{p})}} = 2w_{(\mathrm{h})} \frac{\mathrm{d}\left(u_{(\mathrm{x})} - \hat{u}_{(\mathrm{x})}\right)}{\mathrm{d}g_{(\mathrm{p})}} = 2w_{(\mathrm{h})} \frac{\mathrm{d}\left(u_{(\mathrm{x})} - k_{(\mathrm{x})} - g_{(\mathrm{p})}l_{(\mathrm{x})}\right)}{\mathrm{d}g_{(\mathrm{p})}} \\ &= -2w_{(\mathrm{h})}l_{(\mathrm{x})} = -2\left(u_{(\mathrm{x})} - k_{(\mathrm{x})} - g_{(\mathrm{p})}l_{(\mathrm{x})}\right) l_{(\mathrm{x})}. \end{aligned} \tag{A.15}$$

According to Eq. (A.15) summand two and three in Eq. (A.14) are differentiated by

$$\frac{\mathrm{d}\left(u_{(\mathrm{y})} - \hat{u}_{(\mathrm{y})}\right)^2}{\mathrm{d}g_{(\mathrm{p})}} = -2\left(u_{(\mathrm{y})} - k_{(\mathrm{y})} - g_{(\mathrm{p})}l_{(\mathrm{y})}\right) l_{(\mathrm{y})} \tag{A.16}$$

and

$$\frac{\mathrm{d}\left(u_{(\mathrm{z})} - \hat{u}_{(\mathrm{z})}\right)^2}{\mathrm{d}g_{(\mathrm{p})}} = -2\left(u_{(\mathrm{z})} - k_{(\mathrm{z})} - g_{(\mathrm{p})}l_{(\mathrm{z})}\right) l_{(\mathrm{z})}. \tag{A.17}$$

Substituting Eq. (A.15), Eq. (A.16) , and Eq. (A.17) in Eq. (A.14) yields

$$\frac{\mathrm{d}v_{(\mathrm{h})}}{\mathrm{d}g_{(\mathrm{p})}} = -2\left(u_{(\mathrm{x})}l_{(\mathrm{x})} - k_{(\mathrm{x})}l_{(\mathrm{x})} - l^2_{(\mathrm{x})}g_{(\mathrm{p})} + u_{(\mathrm{y})}l_{(\mathrm{y})} - k_{(\mathrm{y})}l_{(\mathrm{y})} - l^2_{(\mathrm{y})}g_{(\mathrm{p})} + u_{(\mathrm{z})}l_{(\mathrm{z})} - k_{(\mathrm{z})}l_{(\mathrm{z})} - l^2_{(\mathrm{z})}g_{(\mathrm{p})}\right). \tag{A.18}$$

Substituting Eq. (A.18) in Eq. (A.13) results to

$$\frac{\mathrm{d}\Delta_{(\mathrm{xyz})}}{\mathrm{d}g_{(\mathrm{p})}} = \frac{-2\left(u_{(\mathrm{x})}l_{(\mathrm{x})} - k_{(\mathrm{x})}l_{(\mathrm{x})} - l^2_{(\mathrm{x})}g_{(\mathrm{p})} + u_{(\mathrm{y})}l_{(\mathrm{y})} - k_{(\mathrm{y})}l_{(\mathrm{y})} - l^2_{(\mathrm{y})}g_{(\mathrm{p})} + u_{(\mathrm{z})}l_{(\mathrm{z})} - k_{(\mathrm{z})}l_{(\mathrm{z})} - l^2_{(\mathrm{z})}g_{(\mathrm{p})}\right)}{2\left(\left(u_{(\mathrm{x})} - \hat{u}_{(\mathrm{x})}\right)^2 + \left(u_{(\mathrm{y})} - \hat{u}_{(\mathrm{y})}\right)^2 + \left(u_{(\mathrm{z})} - \hat{u}_{(\mathrm{z})}\right)^2\right)}. \tag{A.19}$$

In order to find $g_{(\mathrm{p})}$ that is the extreme point (maxima), Eq. (A.19) is set to zero and rearranged such that

$$\frac{\mathrm{d}\Delta_{(\mathrm{xyz})}}{\mathrm{d}g_{(\mathrm{p})}} = 0 = u_{(\mathrm{x})}l_{(\mathrm{x})} - k_{(\mathrm{x})}l_{(\mathrm{x})} - l^2_{(\mathrm{x})}g_{(\mathrm{p})} + u_{(\mathrm{y})}l_{(\mathrm{y})} - k_{(\mathrm{y})}l_{(\mathrm{y})} - l^2_{(\mathrm{y})}g_{(\mathrm{p})} + u_{(\mathrm{z})}l_{(\mathrm{z})} - k_{(\mathrm{z})}l_{(\mathrm{z})} - l^2_{(\mathrm{z})}g_{(\mathrm{p})}. \tag{A.20}$$

Solving Eq. (A.20) to $g_{(\mathrm{p})}$ yields

$$g^{(0)}_{(\mathrm{p})} = \frac{u_{(\mathrm{x})}l_{(\mathrm{x})} - k_{(\mathrm{x})}l_{(\mathrm{x})} + u_{(\mathrm{y})}l_{(\mathrm{y})} - k_{(\mathrm{y})}l_{(\mathrm{y})} + u_{(\mathrm{z})}l_{(\mathrm{z})} - k_{(\mathrm{z})}l_{(\mathrm{z})}}{l^2_{(\mathrm{x})} + l^2_{(\mathrm{y})} + l^2_{(\mathrm{z})}}, \tag{A.21}$$

where $g_{(\mathrm{p})}^{(0)}$ indicates the $g_{(\mathrm{p})}$ that minimizes the position error.

The denominator of Eq. (A.21) can be simplified by $l_{(\mathrm{x})}^2 + l_{(\mathrm{y})}^2 + l_{(\mathrm{z})}^2 = \left|\mathbf{v}_{(\mathrm{n})}^{(\mathrm{st})}\right|^2 = 1$ because $\mathbf{v}_{(\mathrm{n})}^{(\mathrm{st})}$ in Eq. (A.7) is a normalized vector resulting to

$$\begin{aligned} g_{(\mathrm{p})}^{(0)} =& u_{(\mathrm{x})} l_{(\mathrm{x})} - k_{(\mathrm{x})} l_{(\mathrm{x})} + u_{(\mathrm{y})} l_{(\mathrm{y})} - k_{(\mathrm{y})} l_{(\mathrm{y})} + u_{(\mathrm{z})} l_{(\mathrm{z})} - k_{(\mathrm{z})} l_{(\mathrm{z})} \\ g_{(\mathrm{p})}^{(0)} =& l_{(\mathrm{x})} \left(u_{(\mathrm{x})} - k_{(\mathrm{x})}\right) + l_{(\mathrm{y})} \left(u_{(\mathrm{y})} - k_{(\mathrm{y})}\right) + l_{(\mathrm{z})} \left(u_{(\mathrm{z})} - k_{(\mathrm{z})}\right). \end{aligned} \tag{A.22}$$

Reformulating the differences in Eq. (A.22) with Eq. (A.3) yields

$$\begin{aligned} u_{(\mathrm{x})} - k_{(\mathrm{x})} &= x_1^{(\mathrm{C})} + s^{(\mathrm{BS})} \cos\beta_1 \cos\alpha_1 - s^{(\mathrm{BS})} m_{(\mathrm{x})} - x_1^{(\mathrm{C})} = s^{(\mathrm{BS})} \cos\beta_1 \cos\alpha_1 - s^{(\mathrm{BS})} m_{(\mathrm{x})}, \\ u_{(\mathrm{x})} - k_{(\mathrm{x})} &= s^{(\mathrm{BS})} \cos\beta_1 \cos\alpha_1 - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} \cos\bar{\beta}_1 \cos\bar{\alpha}_1 = s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} m_{(\mathrm{x})}, \\ u_{(\mathrm{y})} - k_{(\mathrm{y})} &= s^{(\mathrm{BS})} \cos\beta_1 \sin\alpha_1 - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} \cos\bar{\beta}_1 \sin\bar{\alpha}_1 = s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} m_{(\mathrm{y})}, \\ u_{(\mathrm{z})} - k_{(\mathrm{z})} &= s^{(\mathrm{BS})} \sin\beta_1 - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} \sin\bar{\beta}_1 = s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} m_{(\mathrm{z})}, \end{aligned} \tag{A.23}$$

and substituting Eq. (A.23) in Eq. (A.22) yields

$$\begin{aligned} g_{(\mathrm{p})}^{(0)} =& \frac{l_{(\mathrm{x})}}{\left|\mathbf{v}^{(\mathrm{st})}\right|^2} \left(s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} m_{(\mathrm{x})}\right) + \frac{l_{(\mathrm{y})}}{\left|\mathbf{v}^{(\mathrm{st})}\right|^2} \left(s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} m_{(\mathrm{y})}\right) \\ &+ \frac{l_{(\mathrm{z})}}{\left|\mathbf{v}^{(\mathrm{st})}\right|^2} \left(s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} m_{(\mathrm{z})}\right). \end{aligned} \tag{A.24}$$

Substituting $l_{(\mathrm{x})}, l_{(\mathrm{y})}$, and $l_{(\mathrm{z})}$ by

$$\begin{aligned} l_{(\mathrm{x})} &= \frac{1}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \left(t^{(\mathrm{BS})} n_{(\mathrm{x})} + x_2^{(\mathrm{C})} - s^{(\mathrm{BS})} m_{(\mathrm{x})} - x_1^{(\mathrm{C})}\right) = \frac{t^{(\mathrm{BS})} n_{(\mathrm{x})} - s^{(\mathrm{BS})} m_{(\mathrm{x})} + \tilde{\Delta}_{(\mathrm{x})}}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \\ &= \frac{\frac{c_{(\mathrm{h})} a_{(\mathrm{h})} - f_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} n_{(\mathrm{x})} - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} m_{(\mathrm{x})} + \tilde{\Delta}_{(\mathrm{x})}}{\left|\mathbf{v}^{(\mathrm{st})}\right|} \\ l_{(\mathrm{x})} &= \frac{\left(c_{(\mathrm{h})} a_{(\mathrm{h})} - f_{(\mathrm{h})}\right) n_{(\mathrm{x})} - \left(c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}\right) m_{(\mathrm{x})} + \left(1 - a_{(\mathrm{h})}^2\right) \tilde{\Delta}_{(\mathrm{x})}}{\left(1 - a_{(\mathrm{h})}^2\right) \left|\mathbf{v}^{(\mathrm{st})}\right|} = \frac{f_{(1)} n_{(\mathrm{x})} - f_{(2)} m_{(\mathrm{x})} + \tilde{a} \tilde{\Delta}_{(\mathrm{x})}}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|} \\ l_{(\mathrm{y})} &= \frac{\left(c_{(\mathrm{h})} a_{(\mathrm{h})} - f_{(\mathrm{h})}\right) n_{(\mathrm{y})} - \left(c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}\right) m_{(\mathrm{y})} + \left(1 - a_{(\mathrm{h})}^2\right) \tilde{\Delta}_{(\mathrm{y})}}{\left(1 - a_{(\mathrm{h})}^2\right) \left|\mathbf{v}^{(\mathrm{st})}\right|} = \frac{f_{(1)} n_{(\mathrm{y})} - f_{(2)} m_{(\mathrm{y})} + \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{y})}}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|} \\ l_{(\mathrm{z})} &= \frac{\left(c_{(\mathrm{h})} a_{(\mathrm{h})} - f_{(\mathrm{h})}\right) n_{(\mathrm{z})} - \left(c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}\right) m_{(\mathrm{z})} + \left(1 - a_{(\mathrm{h})}^2\right) \tilde{\Delta}_{(\mathrm{z})}}{\left(1 - a_{(\mathrm{h})}^2\right) \left|\mathbf{v}^{(\mathrm{st})}\right|} = \frac{f_{(1)} n_{(\mathrm{z})} - f_{(2)} m_{(\mathrm{z})} + \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{z})}}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|}, \end{aligned} \tag{A.25}$$

where $\tilde{a}_{(\mathrm{h})} = 1 - a_{(\mathrm{h})}^2$ and $\tilde{\Delta}_{(\mathrm{x})} = x_2^{(\mathrm{C})} - x_1^{(\mathrm{C})}$, $\tilde{\Delta}_{(\mathrm{y})} = y_2^{(\mathrm{C})} - y_1^{(\mathrm{C})}$, $\tilde{\Delta}_{(\mathrm{z})} = z_2^{(\mathrm{C})} - z_1^{(\mathrm{C})}$, $f_{(1)} = c_{(\mathrm{h})} a_{(\mathrm{h})} - f_{(\mathrm{h})}$,

and $f_{(2)} = c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}$. Then the products in Eq. (A.22) can be substituted with

$$\begin{aligned} l_{(\mathrm{x})}(u_{(\mathrm{x})} - k_{(\mathrm{x})}) =& \frac{f_{(1)} n_{(\mathrm{x})} - f_{(2)} m_{(\mathrm{x})} + \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{x})}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} \left(s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} - \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} m_{(\mathrm{x})} \right) \\ =& \frac{f_{(1)} n_{(\mathrm{x})} - f_{(2)} m_{(\mathrm{x})} + \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{x})}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} \left(\frac{s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} \tilde{a}_{(\mathrm{h})} - f_{(2)} m_{(\mathrm{x})}}{\tilde{a}_{(\mathrm{h})}} \right) \\ =& \frac{f_{(1)} n_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} \tilde{a}_{(\mathrm{h})} - f_{(1)} f_{(2)} n_{(\mathrm{x})} m_{(\mathrm{x})} - f_{(2)} m_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} \tilde{a}_{(\mathrm{h})} + f_{(2)}^2 m_{(\mathrm{x})}^2}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ &+ \frac{s^{(\mathrm{BS})} \tilde{\Delta}_{(\mathrm{x})} \Phi_{(1)}^{(\mathrm{x})} \tilde{a}_{(\mathrm{h})}^2 - f_{(2)} m_{(\mathrm{x})} \tilde{\Delta}_{(\mathrm{x})} \tilde{a}_{(\mathrm{h})}}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|}, \end{aligned} \tag{A.26}$$

which is exemplarily shown for the x direction in Eq. (A.26). Then $g_{(\mathrm{p})}^{(0)}$ in Eq. (A.24) can be reformulated to

$$\begin{aligned} g_{(\mathrm{p})}^{(0)} =& \frac{f_{(1)} n_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} \tilde{a}_{(\mathrm{h})} - f_{(1)} f_{(2)} n_{(\mathrm{x})} m_{(\mathrm{x})} - f_{(2)} m_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} \tilde{a}_{(\mathrm{h})} + f_{(2)}^2 m_{(\mathrm{x})}^2 + s^{(\mathrm{BS})} \tilde{\Delta}_{(\mathrm{x})} \Phi_{(1)}^{(\mathrm{x})} \tilde{a}_{(\mathrm{h})}^2}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ &- \frac{f_{(2)} m_{(\mathrm{x})} \tilde{\Delta}_{(\mathrm{x})} \tilde{a}_{(\mathrm{h})}}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ &+ \frac{f_{(1)} n_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} \tilde{a}_{(\mathrm{h})} - f_{(1)} f_{(2)} n_{(\mathrm{y})} m_{(\mathrm{y})} - f_{(2)} m_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} \tilde{a}_{(\mathrm{h})} + f_{(2)}^2 m_{(\mathrm{y})}^2 + s^{(\mathrm{BS})} \tilde{\Delta}_{(\mathrm{y})} \Phi_{(1)}^{(\mathrm{y})} \tilde{a}_{(\mathrm{h})}^2}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ &- \frac{f_{(2)} m_{(\mathrm{y})} \tilde{\Delta}_{(\mathrm{y})} \tilde{a}_{(\mathrm{h})}}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ &+ \frac{f_{(1)} n_{(\mathrm{z})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} \tilde{a}_{(\mathrm{h})} - f_{(1)} f_{(2)} n_{(\mathrm{z})} m_{(\mathrm{z})} - f_{(2)} m_{(\mathrm{z})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} \tilde{a}_{(\mathrm{h})} + f_{(2)}^2 m_{(\mathrm{z})}^2 + s^{(\mathrm{BS})} \tilde{\Delta}_{(\mathrm{z})} \Phi_{(1)}^{(\mathrm{z})} \tilde{a}_{(\mathrm{h})}^2}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ &- \frac{f_{(2)} m_{(\mathrm{z})} \tilde{\Delta}_{(\mathrm{z})} \tilde{a}_{(\mathrm{h})}}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|}. \end{aligned} \tag{A.27}$$

The square norm of the normalized orientation vector $\mathbf{m}_{(\mathrm{DoA})}$ is $m_{(\mathrm{x})}^2 + m_{(\mathrm{y})}^2 + m_{(\mathrm{z})}^2 = 1$. Applying this and $n_{(\mathrm{x})} m_{(\mathrm{x})} + n_{(\mathrm{y})} m_{(\mathrm{y})} + n_{(\mathrm{z})} m_{(\mathrm{z})} = a_{(\mathrm{h})}$ from Table A.1 to Eq. (A.27) it simplifies to

$$\begin{aligned} g_{(\mathrm{p})}^{(0)} =& \frac{f_{(2)}^2 - f_{(1)} f_{(2)} a_{(\mathrm{h})}}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ &+ \frac{f_{(1)} n_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} \tilde{a}_{(\mathrm{h})} - f_{(2)} m_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} \tilde{a}_{(\mathrm{h})} + \tilde{a}_{(\mathrm{h})}^2 \tilde{\Delta}_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} - \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{x})} f_{(2)} m_{(\mathrm{x})}}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ &+ \frac{f_{(1)} n_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} \tilde{a}_{(\mathrm{h})} - f_{(2)} m_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} \tilde{a}_{(\mathrm{h})} + \tilde{a}_{(\mathrm{h})}^2 \tilde{\Delta}_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} - \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{y})} f_{(2)} m_{(\mathrm{y})}}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|} \\ &+ \frac{f_{(1)} n_{(\mathrm{z})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} \tilde{a}_{(\mathrm{h})} - f_{(2)} m_{(\mathrm{z})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} \tilde{a}_{(\mathrm{h})} + \tilde{a}_{(\mathrm{h})}^2 \tilde{\Delta}_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} - \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{z})} f_{(2)} m_{(\mathrm{z})}}{\tilde{a}_{(\mathrm{h})}^2 \left| \mathbf{v}^{(\mathrm{st})} \right|}. \end{aligned} \tag{A.28}$$

Reducing the summands two to four in Eq. (A.28) by $\tilde{a}_{(\mathrm{h})}$ yields

$$\begin{aligned} g_{(\mathrm{p})}^{(0)} = \frac{f_{(2)}^2 - f_{(1)} f_{(2)} a_{(\mathrm{h})}}{\tilde{a}_{(\mathrm{h})}^2 \left|\mathbf{v}^{(\mathrm{st})}\right|} &+ \frac{f_{(1)} n_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} - f_{(2)} m_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} + \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} - \tilde{\Delta}_{(\mathrm{x})} f_{(2)} m_{(\mathrm{x})}}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|} \\ &+ \frac{f_{(1)} n_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} - f_{(2)} m_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} + \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} - \tilde{\Delta}_{(\mathrm{y})} f_{(2)} m_{(\mathrm{y})}}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|} \\ &+ \frac{f_{(1)} n_{(\mathrm{z})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} - f_{(2)} m_{(\mathrm{z})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} + \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} - \tilde{\Delta}_{(\mathrm{z})} f_{(2)} m_{(\mathrm{z})}}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|} \end{aligned} \tag{A.29}$$

Moving subtrahends $\tilde{\Delta}_{(\mathrm{x})} f_{(2)} m_{(\mathrm{x})}$, $\tilde{\Delta}_{(\mathrm{y})} f_{(2)} m_{(\mathrm{y})}$, and $\tilde{\Delta}_{(\mathrm{z})} f_{(2)} m_{(\mathrm{z})}$ to the first summand and applying rearrangements from Table A.1 in Eq. (A.29) results to

$$\begin{aligned} g_{(\mathrm{p})}^{(0)} =& \frac{f_{(2)}^2 - f_{(1)} f_{(2)} a_{(\mathrm{h})} - f_{(2)} \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \tilde{\mathbf{\Delta}}\right) \tilde{a}_{(\mathrm{h})}}{\tilde{a}_{(\mathrm{h})}^2 \left|\mathbf{v}^{(\mathrm{st})}\right|} \\ &+ \frac{f_{(1)} n_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} - f_{(2)} m_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})} + \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{x})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{x})}}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|} \\ &+ \frac{f_{(1)} n_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} - f_{(2)} m_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})} + \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{y})}}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|} \\ &+ \frac{f_{(1)} n_{(\mathrm{z})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} - f_{(2)} m_{(\mathrm{z})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})} + \tilde{a}_{(\mathrm{h})} \tilde{\Delta}_{(\mathrm{y})} s^{(\mathrm{BS})} \Phi_{(1)}^{(\mathrm{z})}}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|}. \end{aligned} \tag{A.30}$$

With further compression into vector expression Eq. (A.30) can be rewritten to

$$\begin{aligned} g_{(\mathrm{p})}^{(0)} =& \frac{f_{(2)}^2 - f_{(1)} f_{(2)} a_{(\mathrm{h})} - f_{(2)} \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \tilde{\mathbf{\Delta}}\right) \tilde{a}_{(\mathrm{h})}}{\tilde{a}_{(\mathrm{h})}^2 \left|\mathbf{v}^{(\mathrm{st})}\right|} \\ &+ s^{(\mathrm{BS})} \frac{f_{(1)} \left(\mathbf{n}_{(\mathrm{DoA})} \cdot \mathbf{\Phi}_{(1)}\right) - f_{(2)} \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{\Phi}_{(1)}\right) + \tilde{a}_{(\mathrm{h})} \left(\tilde{\mathbf{\Delta}} \cdot \mathbf{\Phi}_{(1)}\right)}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|} \\ =& \frac{f_{(2)}^2 - f_{(1)} f_{(2)} a - f_{(2)} c \left(1 - a^2\right)}{\tilde{a}_{(\mathrm{h})}^2 \left|\mathbf{v}^{(\mathrm{st})}\right|} + s^{(\mathrm{BS})} \frac{f_{(1)} \left(\mathbf{n}_{(\mathrm{DoA})} \cdot \mathbf{\Phi}_{(1)}\right) - f_{(2)} \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{\Phi}_{(1)}\right) + \tilde{a}_{(\mathrm{h})} \left(\tilde{\mathbf{\Delta}} \cdot \mathbf{\Phi}_{(1)}\right)}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|}. \end{aligned} \tag{A.31}$$

The numerator in the first summand in Eq. (A.31) is equal to zero according to

$$\begin{aligned} & f_{(2)}^2 - f_{(1)} f_{(2)} a_{(\mathrm{h})} - f_{(2)} c_{(\mathrm{h})} \left(1 - a_{(\mathrm{h})}^2\right) = \\ =& \left(c_{(\mathrm{h})} - a_{(\mathrm{h})} f_{(\mathrm{h})}\right) \left(c_{(\mathrm{h})} - a_{(\mathrm{h})} f_{(\mathrm{h})}\right) \\ & - a_{(\mathrm{h})} \left(a_{(\mathrm{h})} c_{(\mathrm{h})} - f_{(\mathrm{h})}\right) \left(c_{(\mathrm{h})} - a_{(\mathrm{h})} f_{(\mathrm{h})}\right) - c_{(\mathrm{h})} \left(c_{(\mathrm{h})} - a_{(\mathrm{h})} f_{(\mathrm{h})}\right) \left(1 - a_{(\mathrm{h})}^2\right) \\ =& c_{(\mathrm{h})}^2 - 2 c_{(\mathrm{h})} a_{(\mathrm{h})} f_{(\mathrm{h})} + a_{(\mathrm{h})}^2 f_{(\mathrm{h})}^2 - a_{(\mathrm{h})} \left(a_{(\mathrm{h})} c_{(\mathrm{h})}^2 - a_{(\mathrm{h})}^2 c_{(\mathrm{h})} f_{(\mathrm{h})} - f_{(\mathrm{h})} c_{(\mathrm{h})} + a_{(\mathrm{h})} f_{(\mathrm{h})}^2\right) \\ & - c_{(\mathrm{h})} \left(c_{(\mathrm{h})} - c_{(\mathrm{h})} a_{(\mathrm{h})}^2 - a_{(\mathrm{h})} f_{(\mathrm{h})} + a_{(\mathrm{h})}^3 f_{(\mathrm{h})}\right) \\ =& c_{(\mathrm{h})}^2 - 2 c_{(\mathrm{h})} a_{(\mathrm{h})} f_{(\mathrm{h})} + a_{(\mathrm{h})}^2 f_{(\mathrm{h})}^2 - a_{(\mathrm{h})}^2 c_{(\mathrm{h})}^2 + a_{(\mathrm{h})}^3 c_{(\mathrm{h})} f_{(\mathrm{h})} + a_{(\mathrm{h})} f_{(\mathrm{h})} c_{(\mathrm{h})} - a_{(\mathrm{h})}^2 f_{(\mathrm{h})}^2 - c_{(\mathrm{h})}^2 + c_{(\mathrm{h})}^2 a_{(\mathrm{h})}^2 \\ & + a_{(\mathrm{h})} c_{(\mathrm{h})} f_{(\mathrm{h})} - c_{(\mathrm{h})} a_{(\mathrm{h})}^3 f_{(\mathrm{h})} \\ =& 0. \end{aligned} \tag{A.32}$$

Applying Eq. (A.32) simplifies Eq. (A.31) to

$$g_{(\mathrm{p})}^{(0)} = s^{(\mathrm{BS})} \frac{f_{(1)} \left(\mathbf{n}_{(\mathrm{DoA})} \cdot \mathbf{\Phi}_{(1)}\right) - f_{(2)} \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{\Phi}_{(1)}\right) + \tilde{a}_{(\mathrm{h})} \left(\tilde{\mathbf{\Delta}} \cdot \mathbf{\Phi}_{(1)}\right)}{\tilde{a}_{(\mathrm{h})} \left|\mathbf{v}^{(\mathrm{st})}\right|}. \tag{A.33}$$

Next, $f_{(1)}$ and $f_{(2)}$ are rearranged according to

$$\begin{aligned}
f_{(1)} =& a_{(\mathrm{h})} c_{(\mathrm{h})} - f_{(\mathrm{h})} \\
=& a_{(\mathrm{h})} \left(\cos\bar{\beta}_1 \cos\bar{\alpha}_1 \left(x_2^{(\mathrm{C})} - x_1^{(\mathrm{C})} \right) + \cos\bar{\beta}_1 \sin\bar{\alpha}_1 \left(y_2^{(\mathrm{C})} - y_1^{(\mathrm{C})} \right) + \sin\bar{\beta}_1 \left(z_2^{(\mathrm{C})} - z_1^{(\mathrm{C})} \right) \right) \\
& - \cos\bar{\beta}_2 \cos\bar{\alpha}_2 \left(x_2^{(\mathrm{C})} - x_1^{(\mathrm{C})} \right) + \cos\bar{\beta}_2 \sin\bar{\alpha}_2 \left(y_2^{(\mathrm{C})} - y_1^{(\mathrm{C})} \right) + \sin\bar{\beta}_2 \left(z_2^{(\mathrm{C})} - z_1^{(\mathrm{C})} \right) \\
=& a_{(\mathrm{h})} \left(m_{(\mathrm{x})} \tilde{\Delta}_{(\mathrm{x})} + m_{(\mathrm{y})} \tilde{\Delta}_{(\mathrm{y})} + m_{(\mathrm{z})} \tilde{\Delta}_{(\mathrm{z})} \right) - n_{(\mathrm{x})} \tilde{\Delta}_{(\mathrm{x})} + n_{(\mathrm{y})} \tilde{\Delta}_{(\mathrm{y})} + n_{(\mathrm{z})} \tilde{\Delta}_{(\mathrm{z})} \\
=& a_{(\mathrm{h})} \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}} \right) - \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}} \\
=& \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{n} \right) \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}} \right) - \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}}
\end{aligned} \tag{A.34}$$

and

$$\begin{aligned}
f_{(2)} =& c_{(\mathrm{h})} - a_{(\mathrm{h})} f_{(\mathrm{h})} \\
=& \cos\bar{\beta}_1 \cos\bar{\alpha}_1 \left(x_2^{(\mathrm{C})} - x_1^{(\mathrm{C})} \right) + \cos\bar{\beta}_1 \sin\bar{\alpha}_1 \left(y_2^{(\mathrm{C})} - y_1^{(\mathrm{C})} \right) + \sin\bar{\beta}_1 \left(z_2^{(\mathrm{C})} - z_1^{(\mathrm{C})} \right) \\
& - a_{(\mathrm{h})} \left(\cos\bar{\beta}_2 \cos\bar{\alpha}_2 \left(x_2^{(\mathrm{C})} - x_1^{(\mathrm{C})} \right) + \cos\bar{\beta}_2 \sin\bar{\alpha}_2 \left(y_2^{(\mathrm{C})} - y_1^{(\mathrm{C})} \right) + \sin\bar{\beta}_2 \left(z_2^{(\mathrm{C})} - z_1^{(\mathrm{C})} \right) \right) \\
=& m_{(\mathrm{x})} \tilde{\Delta}_{(\mathrm{x})} + m_{(\mathrm{y})} \tilde{\Delta}_{(\mathrm{y})} + m_{(\mathrm{z})} \tilde{\Delta}_{(\mathrm{z})} - a_{(\mathrm{h})} \left(n_{(\mathrm{x})} \tilde{\Delta}_{(\mathrm{x})} + n_{(\mathrm{y})} \tilde{\Delta}_{(\mathrm{y})} + n_{(\mathrm{z})} \tilde{\Delta}_{(\mathrm{z})} \right) \\
=& \mathbf{m}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}} - a_{(\mathrm{h})} \left(\mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}} \right) \\
=& \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}} \right) - \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{n}_{(\mathrm{DoA})} \right) \left(\mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}} \right).
\end{aligned} \tag{A.35}$$

According to Eq. (A.34), Eq. (A.35) variables $f_{(1)}$ and $f_{(2)}$ in Eq. (A.33) depends on $\mathbf{m}_{(\mathrm{DoA})}, \mathbf{n}_{(\mathrm{DoA})}$, and $\tilde{\boldsymbol{\Delta}}$ which are the DoA estimates and difference of the BS positions, respectively, which can be considered as known constants.

Substituting $\tilde{\boldsymbol{\Delta}}$ and $\boldsymbol{\Phi}_{(1)}$ in Eq. (A.33) with

$$\begin{aligned}
\tilde{\boldsymbol{\Delta}} &= s^{(\mathrm{BS})} \boldsymbol{\Phi}_{(1)} - t^{(\mathrm{BS})} \boldsymbol{\Phi}_{(2)} \\
\boldsymbol{\Phi}_{(1)} &= \frac{t^{(\mathrm{BS})} \boldsymbol{\Phi}_{(2)} + \tilde{\boldsymbol{\Delta}}}{s^{(\mathrm{BS})}}
\end{aligned} \tag{A.36}$$

$g_{(\mathrm{p})}^{(0)}$ can be reformulated to

$$g_{(\mathrm{p})}^{(0)} = \frac{f_{(1)} \left(t^{(\mathrm{BS})} \mathbf{n}_{(\mathrm{DoA})} \cdot \boldsymbol{\Phi}_{(2)} + \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}} \right)}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} - s^{(\mathrm{BS})} \frac{f_{(2)} \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \boldsymbol{\Phi}_{(1)} \right)}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + s^{(\mathrm{BS})} \frac{s^{(\mathrm{BS})} - t^{(\mathrm{BS})} \boldsymbol{\Phi}_{(2)} \cdot \boldsymbol{\Phi}_{(1)}}{\left| \mathbf{v}^{(\mathrm{st})} \right|}. \tag{A.37}$$

Incorporating DoA estimation errors given by $\mathbf{m}_{(\mathrm{DoA})} \cdot \boldsymbol{\Phi}_{(1)} = \boldsymbol{\Delta}_{(1)}^{(\Phi)}$ and $\mathbf{n}_{(\mathrm{DoA})} \cdot \boldsymbol{\Phi}_{(2)} = \boldsymbol{\Delta}_{(2)}^{(\Phi)}$ into Eq. (A.37) yields

$$\begin{aligned}
g_{(\mathrm{p})}^{(0)} =& \frac{f_{(1)} \left(t^{(\mathrm{BS})} \boldsymbol{\Delta}_{(2)}^{(\Phi)} + \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}} \right)}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} - \frac{s^{(\mathrm{BS})} f_{(2)} \boldsymbol{\Delta}_{(1)}^{(\Phi)}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + s^{(\mathrm{BS})} \frac{s^{(\mathrm{BS})} - t^{(\mathrm{BS})} \boldsymbol{\Phi}_{(2)} \cdot \boldsymbol{\Phi}_{(1)}}{\left| \mathbf{v}^{(\mathrm{st})} \right|} \\
=& \frac{f_{(1)} t^{(\mathrm{BS})} \boldsymbol{\Delta}_{(2)}^{(\Phi)} - f_{(2)} s^{(\mathrm{BS})} \boldsymbol{\Delta}_{(1)}^{(\Phi)}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + \frac{f_{(1)} \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + s^{(\mathrm{BS})} \frac{s^{(\mathrm{BS})} - t^{(\mathrm{BS})} \boldsymbol{\Phi}_{(2)} \cdot \boldsymbol{\Phi}_{(1)}}{\left| \mathbf{v}^{(\mathrm{st})} \right|}.
\end{aligned} \tag{A.38}$$

Now the cosine law is used to replace $\tilde{\boldsymbol{\Delta}}^2$ by $\tilde{\boldsymbol{\Delta}}^2 = \left(s^{(\mathrm{BS})} \right)^2 + \left(t^{(\mathrm{BS})} \right)^2 - 2 s^{(\mathrm{BS})} t^{(\mathrm{BS})} \cos\gamma_{1,2}$, because BS 1, BS 2, and the device form an arbitrary triangle. Additional, $\boldsymbol{\Phi}_{(1)}$ and $\boldsymbol{\Phi}_{(2)}$ are orientation vectors with norm 1 such that

$$\boldsymbol{\Phi}_{(1)} \cdot \boldsymbol{\Phi}_{(2)} = \cos\gamma_{1,2} = \frac{\left(s^{(\mathrm{BS})} \right)^2 + \left(t^{(\mathrm{BS})} \right)^2 - \tilde{\boldsymbol{\Delta}}^2}{2 s^{(\mathrm{BS})} t^{(\mathrm{BS})}} \tag{A.39}$$

Applying the cosine law and Eq. (A.39) to Eq. (A.38) yields

$$\begin{aligned} g_{(\mathrm{p})}^{(0)} =& \frac{f_{(1)} t^{(\mathrm{BS})} \boldsymbol{\Delta}_{(2)}^{(\Phi)} - f_{(2)} s^{(\mathrm{BS})} \boldsymbol{\Delta}_{(1)}^{(\Phi)}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + \frac{f_{(1)} \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + s^{(\mathrm{BS})} \frac{s^{(\mathrm{BS})} - t^{(\mathrm{BS})} \frac{\left(s^{(\mathrm{BS})}\right)^2 + \left(t^{(\mathrm{BS})}\right)^2 - \tilde{\boldsymbol{\Delta}}^2}{2 s^{(\mathrm{BS})} t^{(\mathrm{BS})}}}{\left| \mathbf{v}^{(\mathrm{st})} \right|} \\ =& \frac{f_{(1)} t^{(\mathrm{BS})} \boldsymbol{\Delta}_{(2)}^{(\Phi)} - f_{(2)} s^{(\mathrm{BS})} \boldsymbol{\Delta}_{(1)}^{(\Phi)}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + \frac{f_{(1)} \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + s^{(\mathrm{BS})} \frac{s^{(\mathrm{BS})} - \frac{s^{(\mathrm{BS})}}{2} - \frac{\left(t^{(\mathrm{BS})}\right)^2}{2 s^{(\mathrm{BS})}} + \frac{\tilde{\boldsymbol{\Delta}}^2}{2 s^{(\mathrm{BS})}}}{\left| \mathbf{v}^{(\mathrm{st})} \right|} \\ =& \frac{f_{(1)} t^{(\mathrm{BS})} \boldsymbol{\Delta}_{(2)}^{(\Phi)} - f_{(2)} s^{(\mathrm{BS})} \boldsymbol{\Delta}_{(1)}^{(\Phi)}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + \frac{f_{(1)} \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + \frac{\frac{\left(s^{(\mathrm{BS})}\right)^2}{2} - \frac{\left(t^{(\mathrm{BS})}\right)^2}{2} + \frac{\tilde{\boldsymbol{\Delta}}^2}{2}}{\left| \mathbf{v}^{(\mathrm{st})} \right|} \\ =& \frac{f_{(1)} t^{(\mathrm{BS})} \boldsymbol{\Delta}_{(2)}^{(\Phi)} - f_{(2)} s^{(\mathrm{BS})} \boldsymbol{\Delta}_{(1)}^{(\Phi)}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + \frac{f_{(1)} \mathbf{n}_{(\mathrm{DoA})} \cdot \tilde{\boldsymbol{\Delta}}}{\tilde{a}_{(\mathrm{h})} \left| \mathbf{v}^{(\mathrm{st})} \right|} + \frac{\left(s^{(\mathrm{BS})}\right)^2 - \left(t^{(\mathrm{BS})}\right)^2 + \tilde{\boldsymbol{\Delta}}^2}{2 \left| \mathbf{v}^{(\mathrm{st})} \right|}. \end{aligned} \tag{A.40}$$

The following interpretations can be obtained from Eq. (A.40):

- If the distance from the device to BS 1 and BS 2 is equal such that $s^{(\mathrm{BS})} = t^{(\mathrm{BS})}$, than $g_{(\mathrm{p})}^{(0)}$ depends only on the estimation error weighted with a geometry factor.
- If the estimation errors are equal such that $\boldsymbol{\Delta}_{(1)}^{(\Phi)} = \boldsymbol{\Delta}_{(2)}^{(\Phi)}$, than $g_{(\mathrm{p})}^{(0)}$ depends only on the distance difference weighted with a geometry factor.

Table A.1.: Overview of rearrangements used in deriving the optimal weighting parameter $g_{(\mathrm{p})}^{(0)}$.

$m_{(\mathrm{x})} = \cos\bar{\beta}_1 \cos\bar{\alpha}_1$	(1)
$m_{(\mathrm{y})} = \cos\bar{\beta}_1 \sin\bar{\alpha}_1$	(2)
$m_{(\mathrm{z})} = \sin\bar{\beta}_1$	(3)
$n_{(\mathrm{x})} = \cos\bar{\beta}_2 \cos\bar{\alpha}_2$	(4)
$n_{(\mathrm{y})} = \cos\bar{\beta}_2 \sin\bar{\alpha}_2$	(5)
$n_{(\mathrm{z})} = \sin\bar{\beta}_2$	(6)
$s^{(\mathrm{BS})} = \frac{c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} = \frac{f_{(2)}}{1 - a_{(\mathrm{h})}^2}$	(7)
$t^{(\mathrm{BS})} = \frac{c_{(\mathrm{h})} a_{(\mathrm{h})} - f_{(\mathrm{h})}}{1 - a_{(\mathrm{h})}^2} = \frac{f_{(1)}}{1 - a_{(\mathrm{h})}^2}$	(8)
$a_{(\mathrm{h})} = m_{(\mathrm{x})} n_{(\mathrm{x})} + m_{(\mathrm{y})} n_{(\mathrm{y})} + m_{(\mathrm{z})} n_{(\mathrm{z})}$	(9)
$a_{(\mathrm{h})} = \cos\bar{\beta}_1 \cos\bar{\alpha}_1 \cos\bar{\beta}_2 \cos\bar{\alpha}_2 + \cos\bar{\beta}_1 \sin\bar{\alpha}_1 \cos\bar{\beta}_2 \sin\bar{\alpha}_2 + \sin\bar{\beta}_1 \sin\bar{\beta}_2$	(10)
$a_{(\mathrm{h})} = \cos\bar{\beta}_1 \cos\bar{\beta}_2 \left(\cos\bar{\alpha}_1 \cos\bar{\alpha}_2 + \sin\bar{\alpha}_1 \sin\bar{\alpha}_2\right) = \cos\bar{\beta}_1 \cos\bar{\beta}_2 \cos\left(\bar{\alpha}_1 - \bar{\alpha}_2\right) + \sin\bar{\beta}_1 \sin\bar{\beta}_2$	(11)
$a_{(\mathrm{h})} = \mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{n}_{(\mathrm{DoA})}$	(12)
$b_{(\mathrm{h})} = m_{(\mathrm{x})}^2 + m_{(\mathrm{y})}^2 + m_{(\mathrm{z})}^2$	(13)
$b_{(\mathrm{h})} = \cos^2\bar{\beta}_1 \cos^2\bar{\alpha}_1 + \cos^2\bar{\beta}_1 \sin^2\bar{\alpha}_1 + \sin^2\bar{\beta}_1 = \cos^2\bar{\beta}_1 \left(\cos^2\bar{\alpha}_1 + \sin^2\bar{\alpha}_1\right) + \sin^2\bar{\beta}_1$	(14)
$b_{(\mathrm{h})} = 1$	(15)
$c_{(\mathrm{h})} = m_{(\mathrm{x})} x_2^{(\mathrm{C})} - m_{(\mathrm{x})} x_1^{(\mathrm{C})} + m_{(\mathrm{y})} y_2^{(\mathrm{C})} - m_{(\mathrm{y})} y_1^{(\mathrm{C})} + m_{(\mathrm{z})} z_2^{(\mathrm{C})} - m_{(\mathrm{z})} z_1^{(\mathrm{C})}$	(16)
$c_{(\mathrm{h})} = \cos\bar{\beta}_1 \cos\bar{\alpha}_1 \left(x_2^{(\mathrm{C})} - x_1^{(\mathrm{C})}\right) + \cos\bar{\beta}_1 \sin\bar{\alpha}_1 \left(y_2^{(\mathrm{C})} - y_1^{(\mathrm{C})}\right) + \sin\bar{\beta}_1 \left(z_2^{(\mathrm{C})} - z_1^{(\mathrm{C})}\right)$	(17)
$d_{(\mathrm{h})} = n_{(\mathrm{x})}^2 + n_{(\mathrm{y})}^2 + n_{(\mathrm{z})}^2$	(18)
$d_{(\mathrm{h})} = \cos^2\bar{\beta}_2 \cos^2\bar{\alpha}_2 + \cos^2\bar{\beta}_2 \sin^2\bar{\alpha}_2 + \sin^2\bar{\beta}_2 = \cos^2\bar{\beta}_2 \left(\cos^2\bar{\alpha}_2 + \sin^2\bar{\alpha}_2\right) + \sin^2\bar{\beta}_2$	(19)
$d_{(\mathrm{h})} = 1$	(20)
$e_{(\mathrm{h})} = n_{(\mathrm{x})} m_{(\mathrm{x})} + n_{(\mathrm{y})} m_{(\mathrm{y})} + n_{(\mathrm{z})} m_{(\mathrm{z})} = a_{(\mathrm{h})} = \mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{n}_{(\mathrm{DoA})}$	(21)
$f_{(\mathrm{h})} = n_{(\mathrm{x})} x_2^{(\mathrm{C})} - n_{(\mathrm{x})} x_1^{(\mathrm{C})} + n_{(\mathrm{y})} y_2^{(\mathrm{C})} - n_{(\mathrm{y})} y_1^{(\mathrm{C})} + n_{(\mathrm{z})} z_2^{(\mathrm{C})} - n_{(\mathrm{z})} z_1^{(\mathrm{C})}$	(22)
$f_{(\mathrm{h})} = \cos\bar{\beta}_2 \cos\bar{\alpha}_2 \left(x_2^{(\mathrm{C})} - x_1^{(\mathrm{C})}\right) + \cos\bar{\beta}_2 \sin\bar{\alpha}_2 \left(y_2^{(\mathrm{C})} - y_1^{(\mathrm{C})}\right) + \sin\bar{\beta}_2 \left(z_2^{(\mathrm{C})} - z_1^{(\mathrm{C})}\right)$	(23)
$a_{(\mathrm{h})}^2 = m_{(\mathrm{x})}^2 n_{(\mathrm{x})}^2 + m_{(\mathrm{y})}^2 n_{(\mathrm{y})}^2 + m_{(\mathrm{z})}^2 n_{(\mathrm{z})}^2 + 2(m_{(\mathrm{x})} n_{(\mathrm{x})} m_{(\mathrm{y})} n_{(\mathrm{y})} + m_{(\mathrm{x})} n_{(\mathrm{x})} m_{(\mathrm{z})} n_{(\mathrm{z})} + m_{(\mathrm{y})} n_{(\mathrm{y})} m_{(\mathrm{z})} n_{(\mathrm{z})})$	(24)
$a_{(\mathrm{h})}^2 = \cos^2\bar{\beta}_1 \cos^2\bar{\alpha}_1 \cos^2\bar{\beta}_2 \cos^2\bar{\alpha}_2 + \cos^2\bar{\beta}_1 \sin^2\bar{\alpha}_1 \cos^2\bar{\beta}_2 \sin^2\bar{\alpha}_2 + \sin^2\bar{\beta}_1 \sin^2\bar{\beta}_2$	(25)
$a_{(\mathrm{h})}^2 = \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{n}_{(\mathrm{DoA})}\right) \left(\mathbf{m}_{(\mathrm{DoA})} \cdot \mathbf{n}_{(\mathrm{DoA})}\right)$	(26)
$\tilde{a}_{(\mathrm{h})} = 1 - a_{(\mathrm{h})}^2$	(27)
$\tilde{\Delta}_{(\mathrm{x})} = x_2^{(\mathrm{C})} - x_1^{(\mathrm{C})}$	(28)
$\tilde{\Delta}_{(\mathrm{y})} = y_2^{(\mathrm{C})} - y_1^{(\mathrm{C})}$	(29)
$\tilde{\Delta}_{(\mathrm{z})} = z_2^{(\mathrm{C})} - z_1^{(\mathrm{C})}$	(30)
$f_{(1)} = c_{(\mathrm{h})} a_{(\mathrm{h})} - f_{(\mathrm{h})}$	(31)
$f_{(2)} = c_{(\mathrm{h})} - f_{(\mathrm{h})} a_{(\mathrm{h})}$	(32)

B. Additional Material: Algorithms, Figures, Tables

Section 2.4

Table B.1.: Single user spectral efficiency and effective spectral efficiency for full and adapted angular range of codebooks. The UPA dimension is $[10 \times 10]$ and $\tau^{(\mathrm{C})} = 168$. Red colored fields indicate the largest achieved sum spectral efficiency including pilot overhead for each angular range.

$N_\alpha^{(\boldsymbol{\Omega})}$	$N_\beta^{(\boldsymbol{\Omega})}$	$N^{(\boldsymbol{\Omega})}$	$\boldsymbol{\alpha}^{(\mathrm{CW})} = [-90\ 90]^\circ$ $\boldsymbol{\beta}^{(\mathrm{CW})} = [-90\ 90]^\circ$		$\boldsymbol{\alpha}^{(\mathrm{CW})} = [-50\ 50]^\circ$ $\boldsymbol{\beta}^{(\mathrm{CW})} = [-40\ 7]^\circ$	
			$\tilde{C}_{k,t}^{(\mathrm{O})}$ in [bit/s/Hz]	$\tilde{C}_{k,t}$ in [bit/s/Hz]	$\tilde{C}_{k,t}^{(\mathrm{O})}$ in [bit/s/Hz]	$\tilde{C}_{k,t}$ in [bit/s/Hz]
1	1	1	10.6	10.7	10.6	10.7
2	1	2	11	11.1	11.5	11.7
4	1	4	12.3	12.6	12.5	12.8
8	1	8	12.7	13.3	12.9	13.5
16	1	16	12.4	13.7	12.4	13.7
32	1	32	11.2	13.8	11.1	13.8
1	2	2	10.7	10.9	11.6	11.7
2	2	4	10.9	11.2	12.1	12.4
4	2	8	12	12.6	12.8	13.5
8	2	16	12.1	13.3	12.9	14.2
16	2	32	11.1	13.7	11.6	14.3
32	2	64	8.5	13.8	8.9	14.4
1	4	4	11.5	11.8	13.1	13.4
2	4	8	11.4	12	13.3	14
4	4	16	12	13.3	13.5	14.9
8	4	32	11.2	13.9	12.5	15.4
16	4	64	8.9	14.3	9.7	15.6
32	4	128	5.8	14.4	6.3	15.7
1	8	8	12.6	13.3	13	13.6
2	8	16	12.2	13.5	12.9	14.2
4	8	32	11.9	14.6	12.2	15.1
8	8	64	9.4	15.1	9.7	15.6
16	8	128	6.2	15.4	6.4	15.8
32	8	256	6.3	15.5	6.4	15.9
1	16	16	12.3	13.6	12.4	13.7
2	16	32	11.2	13.8	11.6	14.3
4	16	64	9.3	15	9.4	15.2
8	16	128	6.3	15.5	6.3	15.7
16	16	256	6.4	15.7	6.4	15.9
32	16	512	6.4	15.8	6.4	15.9
1	32	32	11.2	13.8	11.1	13.7
2	32	64	8.6	13.9	8.9	14.3
4	32	128	6.1	15.1	6.2	15.2
8	32	256	6.3	15.6	6.3	15.7
16	32	512	6.4	15.8	6.4	15.9
32	32	1024	6.4	15.9	6.5	15.9

Algorithm 4 Matlab code of algorithm to construct sub codebooks from full codebook. "NumHorCws" and "NumVerCws" correspond to $N_{(\alpha)}^{(\mathbf{\Omega})}$ and $N_{(\beta)}^{(\mathbf{\Omega})}$, respectively. "NumVerCwsSplit" and "NumHorCwsSplit" correspond to $N_{(\alpha)}^{(\mathbf{\Omega},\mathrm{SCB})}$ and $N_{(\beta)}^{(\mathbf{\Omega},\mathrm{SCB})}$, respectively.

```
% Author: Martin Kurras, Matlab Version: 2015b
% Number of vertical and horizontal codewords in the full codebook
NumVerCws = 16;
NumHorCws = 16;
NumCws = NumHorCws*NumVerCws; % Total number of codewords in full codebook
% Number of horizontal and vertical directions for sub codebook splitting
NumVerCwsSplit = 4;
NumHorCwsSplit = 4;
NumCwsSplit = NumVerCwsSplit*NumHorCwsSplit; % Total number of codewords in sub codebooks
% Distance in [codewords] from the full codebook for adjacent codewords in sub codebooks
VerDist = round(NumVerCws./NumVerCwsSplit);
HorDist = round(NumHorCws./NumHorCwsSplit);
% If NumVerCwsSplit and NumHorCwsSplit are not integer multiples of NumVerCws and
%       NumHorCwsSplit, respectively, more codebooks are required to cover all codewords
if rem(NumVerCws,NumVerCwsSplit)
    TmpVec = 1:VerDist:VerDist*NumVerCwsSplit;
    TmpVer = NumVerCws+1-TmpVec(end)-VerDist;
    if TmpVer<0
        TmpVer=0;
    end
else
        TmpVer = 0;
end
if rem(NumHorCws,NumHorCwsSplit)
    TmpVec = 1:HorDist:HorDist*NumHorCwsSplit;
    TmpHor = NumHorCws+1-TmpVec(end)-HorDist;
    if TmpHor<0
        TmpHor=0;
    end
else
    TmpHor=0;
end
% Number of sub-codebooks
NumSubCbs = (VerDist+TmpVer)*(HorDist+TmpHor);
% Selection indices of sub codebooks
SubCodebookIndices = zeros(NumSubCbs,NumCwsSplit);
for ItSubCbs = 1:NumSubCbs
    HorStartIndex = floor((ItSubCbs-1)./(VerDist+TmpVer))+1;
    HorIndices = rem(HorStartIndex:HorDist:HorStartIndex+...
        HorDist*NumHorCwsSplit-1,NumHorCws);
    HorIndices(HorIndices==0)=NumHorCws;
    if sum(HorIndices>NumHorCws)
        error('');
    end
    VerStartIndex = mod(ItSubCbs,VerDist+TmpVer);
    if VerStartIndex==0
        VerStartIndex=VerDist+TmpVer;
    end
    VerIndices = rem(VerStartIndex:VerDist:VerStartIndex+...
        VerDist*NumVerCwsSplit-1,NumVerCws);
    VerIndices(VerIndices==0)=NumVerCws;

    LinearCwIndices = zeros(1,NumCwsSplit);
    for ItVer = 1:NumVerCwsSplit

        VerIndex = VerIndices(ItVer);
        if VerIndex>NumVerCws
            error('');
        end

        LinearCwIndexTmp = sub2ind([NumVerCws,NumHorCws],...
            VerIndex*ones(1,NumHorCwsSplit),HorIndices);
        IndicesAccess = (ItVer-1)*NumHorCwsSplit+1 : ItVer*NumHorCwsSplit;
        LinearCwIndices(IndicesAccess) = LinearCwIndexTmp;
    end
    LinearCwIndices = sort(LinearCwIndices);
    SubCodebookIndices(ItSubCbs,:) = LinearCwIndices;
end
```

Algorithm 5 Pseudo code of codeword selection algorithm based on power values per stream feedback $\mathbf{P}_{(\mathrm{FB})}^{(\mathrm{eff})}$. The algorithm is a heuristic with the objective of maximizing the sum rate.

Input: $\mathbf{P}_{(\mathrm{FB}),k}^{(\mathrm{eff})} \forall k \in \mathcal{K}$, NumberMaxStreams
Initialize SumRateCheck as true
Initialize OldRate as zero
Initialize NumberSelectedStreams as zero
Initialize SelectedCodewords as empty
Initialize SelectedUsers as empty
while SumRateCheck==true and NumberSelectedStreams < NumberMaxStreams **do**
    Computation of user rate for all available codewords with SelectedCodewords as interference
    Select codeword/user with maximum estimated sum-rate
    Update NewRate
    if NewRate > OldRate **then**
        Set OldRate to NewRate
        Update SelectedCodewords
        Update SelectedUsers
        Update NumberSelectedStreams
        Update AvailableCodewords
    else
        Set SumRateCheck to false
    end if
end while

Table B.2.: Sum spectral efficiency values corresponding to Fig. 2.29a.

$\tilde{T}_l$	$N_{(\mathrm{FB})}^{(\mathrm{CW})}$															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Sum spectral efficiency $\tilde{C}_l^{(\mathrm{sum})}$ in [bit/s/Hz]															
1	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9	14.9
2	9.3	10.4	11.0	11.4	11.8	12.2	12.5	12.8	13.1	13.5	13.9	14.3	14.8	15.2	15.7	16.1
3	8.5	9.6	10.4	10.9	11.3	11.7	12.1	12.4	12.8	13.2	13.6	14.1	14.7	15.3	16.0	16.8
4	8.8	9.9	10.6	11.2	11.6	12.0	12.4	12.7	13.0	13.3	13.8	14.2	14.6	15.2	16.0	17.2
5	8.9	10.1	10.9	11.4	11.8	12.2	12.6	12.9	13.2	13.6	13.9	14.2	14.7	15.3	16.1	17.7
6	9.1	10.3	11.1	11.8	12.2	12.7	13.0	13.3	13.6	13.9	14.2	14.6	14.9	15.4	16.3	18.0
7	9.2	10.6	11.4	12.0	12.6	13.1	13.5	13.8	14.0	14.3	14.5	14.8	15.1	15.5	16.3	18.2
8	9.4	10.8	11.7	12.4	13.0	13.5	13.8	14.1	14.4	14.6	14.8	15.1	15.4	15.7	16.4	18.4
9	9.6	11.1	11.9	12.6	13.1	13.6	14.0	14.3	14.6	14.9	15.1	15.3	15.5	15.9	16.5	18.5
10	9.8	11.3	12.2	12.9	13.4	13.8	14.3	14.6	14.8	15.1	15.4	15.5	15.7	16.0	16.5	18.6
20	11.0	12.8	13.8	14.5	15.0	15.4	15.8	16.1	16.3	16.6	16.8	16.9	17.2	17.3	17.5	19.8
30	11.7	13.7	14.8	15.5	16.0	16.3	16.7	16.8	17.0	17.3	17.4	17.6	17.8	18.0	18.1	20.6
40	12.2	14.3	15.3	16.0	16.5	16.8	17.2	17.4	17.6	17.8	17.9	18.2	18.2	18.5	18.7	21.0
50	12.7	14.8	15.9	16.5	16.9	17.3	17.6	17.7	18.0	18.1	18.3	18.4	18.6	18.7	19.0	21.3
60	12.9	15.1	16.2	16.8	17.2	17.5	17.8	18.0	18.2	18.4	18.5	18.7	18.9	19.1	19.3	21.6
70	13.2	15.4	16.4	17.0	17.6	17.9	18.1	18.3	18.5	18.6	18.8	18.9	19.2	19.4	19.6	21.9
80	13.4	15.7	16.7	17.4	17.8	18.1	18.3	18.5	18.6	18.8	19.0	19.1	19.3	19.5	19.7	22.1
90	13.6	15.8	16.9	17.6	18.0	18.3	18.5	18.8	18.8	19.0	19.1	19.3	19.4	19.7	20.0	22.3
100	13.9	16.0	17.2	17.8	18.2	18.5	18.6	18.9	19.0	19.1	19.3	19.4	19.6	19.8	20.1	22.5

Table B.3.: Number of selected streams (average value) corresponding to Fig. 2.29a.

$\tilde{T}_l$	$N_{(\mathrm{FB})}^{(\mathrm{CW})}$															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Number of selected streams T_l															
1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
2	1.9	1.9	1.8	1.8	1.8	1.7	1.7	1.6	1.6	1.6	1.5	1.4	1.4	1.3	1.2	1.1
3	2.7	2.5	2.3	2.2	2.1	2.1	2.0	2.0	1.9	1.9	1.8	1.8	1.7	1.6	1.4	1.2
4	3.4	3.0	2.8	2.6	2.4	2.3	2.2	2.2	2.1	2.1	2.0	2.0	1.9	1.8	1.6	1.3
5	4.0	3.4	3.1	2.9	2.7	2.6	2.4	2.3	2.2	2.2	2.1	2.0	2.0	1.9	1.7	1.4
6	4.6	3.8	3.4	3.1	2.9	2.8	2.6	2.5	2.4	2.3	2.2	2.1	2.0	1.9	1.8	1.5
7	5.1	4.2	3.7	3.3	3.1	2.9	2.7	2.6	2.4	2.3	2.2	2.1	2.0	2.0	1.8	1.5
8	5.6	4.5	3.9	3.5	3.3	3.0	2.9	2.7	2.5	2.4	2.3	2.2	2.1	2.0	1.9	1.6
9	6.0	4.7	4.1	3.7	3.4	3.2	3.0	2.8	2.6	2.5	2.3	2.2	2.1	2.0	1.9	1.6
10	6.4	5.0	4.3	3.8	3.5	3.3	3.0	2.9	2.7	2.5	2.4	2.2	2.1	2.0	1.9	1.7
20	9.0	6.6	5.5	4.8	4.3	3.9	3.6	3.3	3.0	2.8	2.6	2.4	2.2	2.1	2.0	1.8
30	10.5	7.5	6.2	5.3	4.7	4.2	3.8	3.5	3.2	3.0	2.7	2.5	2.3	2.1	2.0	1.9
40	11.5	8.1	6.6	5.6	4.9	4.4	4.0	3.6	3.3	3.0	2.8	2.6	2.3	2.2	2.0	2.0
50	12.2	8.5	6.9	5.8	5.1	4.5	4.1	3.7	3.4	3.1	2.8	2.6	2.4	2.2	2.0	2.1
60	12.7	8.9	7.2	6.0	5.2	4.6	4.2	3.8	3.4	3.1	2.9	2.6	2.4	2.2	2.0	2.1
70	13.1	9.1	7.4	6.2	5.3	4.7	4.2	3.8	3.5	3.2	2.9	2.6	2.4	2.2	2.0	2.2
80	13.4	9.4	7.5	6.3	5.4	4.8	4.3	3.9	3.5	3.2	2.9	2.7	2.4	2.2	2.0	2.2
90	13.7	9.5	7.6	6.4	5.5	4.8	4.3	3.9	3.5	3.2	2.9	2.7	2.4	2.2	2.0	2.2
100	13.9	9.7	7.7	6.4	5.6	4.9	4.3	3.9	3.5	3.2	3.0	2.7	2.4	2.2	2.0	2.2

Table B.4.: Values corresponding to Fig. 2.32a. The red colored field is the achieved maximum over parameters $N_{(\alpha)}^{(\mathbf{\Omega},\mathrm{SCB})}$ and $N_{(\beta)}^{(\mathbf{\Omega},\mathrm{SCB})}$.

$N_{(\alpha)}^{(\mathbf{\Omega},\mathrm{SCB})}$	$N_{(\beta)}^{(\mathbf{\Omega},\mathrm{SCB})}$															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Sum spectral efficiency $\tilde{C}_l^{(\mathrm{sum})}$ in [bit/s/Hz] without pilot overhead															
1	-	-	-	-	-	52.5	-	43.6	36.7	31.0	-	-	-	19.8	18.7	17.7
2	-	-	67.7	87.0	101.2	92.1	85.3	75.9	65.5	56.5	44.1	42.4	41.2	39.9	38.5	36.8
3	-	67.9	98.4	125.0	144.3	131.1	118.5	109.9	96.1	84.2	66.2	63.4	61.9	61.4	59.7	57.4
4	-	88.2	126.9	160.6	183.7	170.4	153.9	146.4	130.0	115.1	90.6	87.9	86.0	86.4	83.3	80.2
5	-	107.2	153.5	194.3	221.3	205.7	189.1	178.5	160.8	145.6	114.2	111.7	109.5	110.5	107.3	104.2
6	65.2	124.3	178.9	225.7	256.5	248.7	230.8	224.6	205.6	187.7	147.2	144.7	140.3	140.9	-	-
7	-	141.5	202.3	254.7	290.0	281.1	262.1	255.1	233.9	215.4	170.5	166.7	-	-	-	-
8	84.4	157.8	226.9	286.3	323.6	321.4	298.5	294.6	273.9	250.8	-	-	-	-	-	-
9	76.3	135.2	192.5	253.0	297.2	308.2	275.9	272.1	246.9	-	-	-	-	-	-	-
10	68.5	124.0	174.2	234.9	279.4	291.2	256.3	246.6	-	-	-	-	-	-	-	-
11	-	141.4	196.4	258.9	296.0	302.4	278.0	-	-	-	-	-	-	-	-	-
12	-	137.8	191.1	251.2	293.7	298.4	274.6	-	-	-	-	-	-	-	-	-
13	-	127.5	175.0	235.9	280.0	286.1	-	-	-	-	-	-	-	-	-	-
14	61.6	114.6	155.4	215.4	261.6	271.6	-	-	-	-	-	-	-	-	-	-
15	50.3	98.6	138.7	196.1	242.4	-	-	-	-	-	-	-	-	-	-	-
16	43.0	87.2	124.7	179.5	228.1	-	-	-	-	-	-	-	-	-	-	-

Table B.5.: Values corresponding to Fig. 2.32b. The red colored field is the achieved maximum over parameters $N_{(\alpha)}^{(\mathbf{\Omega},\mathrm{SCB})}$ and $N_{(\beta)}^{(\mathbf{\Omega},\mathrm{SCB})}$.

$N_{(\alpha)}^{(\mathbf{\Omega},\mathrm{SCB})}$	$N_{(\beta)}^{(\mathbf{\Omega},\mathrm{SCB})}$															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Sum spectral efficiency $\tilde{C}_l^{(\mathrm{sum,O})}$ in [bit/s/Hz] with pilot overhead considering one RE per pilot per LTE RB consisting of 168 REs															
1	-	-	-	-	-	50.6	-	41.6	34.7	29.2	-	-	-	18.1	17.0	16.0
2	-	-	65.3	82.8	95.2	85.5	78.2	68.6	58.5	49.7	38.3	36.3	34.9	33.3	31.6	29.8
3	-	65.4	93.1	116.1	131.4	117.0	103.7	94.2	80.7	69.2	53.2	49.8	47.5	46.0	43.7	41.0
4	-	84.0	117.8	145.3	161.8	146.1	128.2	118.5	102.2	87.7	66.9	62.8	59.4	57.6	53.6	49.6
5	-	100.8	139.8	171.2	188.3	169.0	149.7	136.0	117.8	102.3	76.8	71.8	67.1	64.5	59.4	54.6
6	62.8	115.4	159.8	193.5	210.7	195.4	173.1	160.4	139.5	120.7	89.4	82.7	75.2	70.4	-	-
7	-	129.8	177.0	212.2	229.5	210.8	185.7	170.1	146.2	125.6	92.3	83.4	-	-	-	-
8	80.3	142.8	194.5	231.8	246.6	229.6	199.0	182.3	156.5	131.4	-	-	-	-	-	-
9	72.2	120.7	161.6	198.8	217.6	209.1	172.4	155.5	127.9	-	-	-	-	-	-	-
10	64.4	109.2	143.1	179.0	196.2	187.2	149.5	129.2	-	-	-	-	-	-	-	-
11	-	122.9	157.8	191.1	199.1	183.6	150.6	-	-	-	-	-	-	-	-	-
12	-	118.1	150.1	179.4	188.8	170.5	137.3	-	-	-	-	-	-	-	-	-
13	-	107.8	134.3	162.9	171.7	153.3	-	-	-	-	-	-	-	-	-	-
14	56.5	95.5	116.5	143.6	152.6	135.8	-	-	-	-	-	-	-	-	-	-
15	45.8	81.0	101.6	126.1	134.2	-	-	-	-	-	-	-	-	-	-	-
16	38.9	70.6	89.1	111.1	119.5	-	-	-	-	-	-	-	-	-	-	-

Table B.6.: Values corresponding to Fig. 2.34a.

$N^{(\tilde{\Omega})}$	Quantization bits $q^{(\mathrm{CSI})}$												
	8	9	10	11	12	13	14	15	16	17	18	19	20
	Sum spectral efficiency $\tilde{C}_l^{(\mathrm{sum,O})}$ in [bit/s/Hz] with pilot overhead considering one RE per pilot per LTE RB consisting of 168 REs												
8	39.4	43.7	52.5	57.0	64.8	68.8	75.0	77.6	80.5	81.6	82.8	83.1	83.5
9	41.3	45.8	55.9	60.4	69.3	73.7	81.2	84.1	88.3	89.4	91.1	91.6	92.3
10	43.4	48.6	59.0	64.2	74.1	78.8	87.0	90.4	95.3	96.8	98.7	99.1	99.8
11	21.6	25.0	30.1	34.6	40.4	45.6	51.3	56.5	61.2	65.0	68.3	70.9	72.4
12	47.2	52.6	65.7	71.3	83.4	88.9	99.2	103.1	109.7	111.4	114.8	115.5	116.7
13	21.1	23.8	28.1	31.7	36.3	40.2	45.4	50.0	54.8	58.8	61.9	64.0	65.4
14	52.0	58.1	72.5	79.0	92.3	98.3	109.5	113.7	121.1	123.4	126.6	127.1	128.5
15	52.0	58.3	73.9	80.9	95.9	102.2	114.8	119.6	128.8	131.1	135.5	136.1	138.0
16	53.1	59.7	76.9	83.6	99.4	106.0	119.4	124.5	133.6	136.1	140.2	140.9	143.0
18	55.6	62.5	80.7	88.6	105.5	113.0	127.8	134.1	145.1	148.3	153.7	154.9	157.2
20	59.6	66.9	87.3	95.3	115.4	122.6	138.9	144.7	156.1	159.1	164.6	165.6	167.7
21	62.8	70.7	91.6	99.7	119.6	127.4	144.0	149.8	161.9	164.9	171.1	172.1	174.2
22	30.4	35.0	44.9	51.0	62.5	69.5	81.9	89.4	99.1	104.2	111.0	113.9	117.0
24	68.3	76.5	100.2	109.1	131.1	138.9	157.9	164.3	178.1	180.9	187.6	189.0	191.2
25	48.0	55.6	75.0	84.4	105.5	115.8	134.4	143.5	158.2	163.6	172.9	176.4	180.9
26	29.6	33.4	42.0	47.1	56.5	62.8	73.2	79.0	88.5	92.9	99.2	101.3	103.7
27	64.3	71.5	91.8	98.8	116.4	122.6	137.7	142.2	151.1	153.6	158.3	159.0	159.8
28	74.8	83.4	109.3	118.9	143.4	151.9	173.0	179.9	194.2	197.3	205.1	206.1	208.4
30	54.4	62.9	84.9	95.9	120.0	130.9	152.5	162.1	178.3	184.5	194.6	198.2	202.9
32	81.1	90.3	119.1	129.5	156.6	165.3	188.3	195.7	211.2	215.3	222.9	224.1	226.7

Table B.7.: Values corresponding to Fig. 2.34b.

$N^{(\tilde{\Omega})}$	Quantization bits $q^{(\mathrm{CSI})}$												
	8	9	10	11	12	13	14	15	16	17	18	19	20
	Feedback $R^{(\mathrm{FB,HP},2)}$ in [kbit]												
8	3.6	4.1	4.5	5.0	5.4	5.9	6.3	6.8	7.2	7.7	8.1	8.6	9.0
9	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0
10	4.4	5.0	5.5	6.1	6.6	7.2	7.7	8.3	8.8	9.4	9.9	10.5	11.0
11	4.8	5.4	6.0	6.6	7.2	7.8	8.4	9.0	9.6	10.2	10.8	11.4	12.0
12	5.2	5.9	6.5	7.2	7.8	8.5	9.1	9.8	10.4	11.1	11.7	12.4	13.0
13	5.6	6.3	7.0	7.7	8.4	9.1	9.8	10.5	11.2	11.9	12.6	13.3	14.0
14	6.0	6.8	7.5	8.3	9.0	9.8	10.5	11.3	12.0	12.8	13.5	14.3	15.0
15	6.4	7.2	8.0	8.8	9.6	10.4	11.2	12.0	12.8	13.6	14.4	15.2	16.0
16	6.8	7.7	8.5	9.4	10.2	11.1	11.9	12.8	13.6	14.5	15.3	16.2	17.0
18	7.6	8.6	9.5	10.5	11.4	12.4	13.3	14.3	15.2	16.2	17.1	18.1	19.0
20	8.4	9.5	10.5	11.6	12.6	13.7	14.7	15.8	16.8	17.9	18.9	20.0	21.0
21	8.8	9.9	11.0	12.1	13.2	14.3	15.4	16.5	17.6	18.7	19.8	20.9	22.0
22	9.2	10.4	11.5	12.7	13.8	15.0	16.1	17.3	18.4	19.6	20.7	21.9	23.0
24	10.0	11.3	12.5	13.8	15.0	16.3	17.5	18.8	20.0	21.3	22.5	23.8	25.0
25	10.4	11.7	13.0	14.3	15.6	16.9	18.2	19.5	20.8	22.1	23.4	24.7	26.0
26	10.8	12.2	13.5	14.9	16.2	17.6	18.9	20.3	21.6	23.0	24.3	25.7	27.0
27	11.2	12.6	14.0	15.4	16.8	18.2	19.6	21.0	22.4	23.8	25.2	26.6	28.0
28	11.6	13.1	14.5	16.0	17.4	18.9	20.3	21.8	23.2	24.7	26.1	27.6	29.0
30	12.4	14.0	15.5	17.1	18.6	20.2	21.7	23.3	24.8	26.4	27.9	29.5	31.0
32	13.2	14.9	16.5	18.2	19.8	21.5	23.1	24.8	26.4	28.1	29.7	31.4	33.0

Table B.8.: $\tilde{N}^{(\tilde{\mathbf{\Omega}})}$ that ensures feedback reduction according to Eq. (2.79) over the number of CSI quantization bits $q^{(\mathrm{CSI})}$ and MISO feedback channel size $N^{(\tilde{\mathbf{\Omega}})}$.

$N^{(\tilde{\mathbf{\Omega}})}$	$q^{(\mathrm{CSI})}$ in [bit]																						
	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
	$\tilde{N}^{(\tilde{\mathbf{\Omega}})}$ for equality according to Eq. (2.79)																						
16	10	10	10	10	11	11	11	11	12	12	12	12	12	12	12	12	13	13	13	13	13	13	13
17	10	11	11	11	11	12	12	12	12	12	13	13	13	13	13	13	13	13	14	14	14	14	14
18	11	11	12	12	12	12	13	13	13	13	13	14	14	14	14	14	14	14	14	14	15	15	15
19	11	12	12	13	13	13	13	14	14	14	14	14	14	15	15	15	15	15	15	15	15	15	16
20	12	12	13	13	14	14	14	14	15	15	15	15	15	15	16	16	16	16	16	16	16	16	16
21	13	13	14	14	14	15	15	15	15	15	16	16	16	16	16	16	17	17	17	17	17	17	17
22	13	14	14	15	15	15	16	16	16	16	16	17	17	17	17	17	17	18	18	18	18	18	18
23	14	14	15	15	16	16	16	17	17	17	17	17	18	18	18	18	18	18	18	19	19	19	19
24	15	15	16	16	16	17	17	17	18	18	18	18	18	19	19	19	19	19	19	19	20	20	20
25	15	16	16	17	17	17	18	18	18	19	19	19	19	19	20	20	20	20	20	20	20	20	21
26	16	16	17	17	18	18	18	19	19	19	20	20	20	20	20	20	21	21	21	21	21	21	21
27	16	17	18	18	18	19	19	19	20	20	20	21	21	21	21	21	21	22	22	22	22	22	22
28	17	18	18	19	19	20	20	20	21	21	21	21	22	22	22	22	22	22	23	23	23	23	23
29	18	18	19	19	20	20	21	21	21	22	22	22	22	23	23	23	23	23	23	24	24	24	24
30	18	19	20	20	21	21	21	22	22	22	23	23	23	23	24	24	24	24	24	24	25	25	25
31	19	20	20	21	21	22	22	22	23	23	23	24	24	24	24	25	25	25	25	25	25	25	26
32	20	20	21	21	22	22	23	23	24	24	24	24	25	25	25	25	26	26	26	26	26	26	26
33	20	21	22	22	23	23	24	24	24	25	25	25	25	26	26	26	26	27	27	27	27	27	27
34	21	22	22	23	23	24	24	25	25	25	26	26	26	26	27	27	27	27	28	28	28	28	28
35	21	22	23	23	24	25	25	25	26	26	26	27	27	27	28	28	28	28	28	29	29	29	29
36	22	23	24	24	25	25	26	26	27	27	27	28	28	28	28	29	29	29	29	29	30	30	30
37	23	23	24	25	25	26	26	27	27	28	28	28	29	29	29	29	30	30	30	30	30	31	31
38	23	24	25	26	26	27	27	28	28	28	29	29	29	30	30	30	30	31	31	31	31	31	32
39	24	25	26	26	27	27	28	28	29	29	30	30	30	30	31	31	31	31	32	32	32	32	32
40	25	25	26	27	28	28	29	29	30	30	30	31	31	31	32	32	32	32	32	33	33	33	33

Table B.9.: Values corresponding to Fig. 2.35a.

$\tilde{N}^{(\tilde{\Omega})}$	Quantization bits $q^{(\mathrm{CSI})}$												
	8	9	10	11	12	13	14	15	16	17	18	19	20
	Effective sum spectral efficiency $\tilde{C}_l^{(\mathrm{sum,O})}$ in [bit/s/Hz] with $\approx 8.1\,\%$ pilot overhead												
32	81.1	90.3	119.1	129.5	156.6	165.3	188.3	195.7	211.2	215.3	222.9	224.1	226.7
30	80.0	88.6	111.3	118.6	135.3	140.1	148.9	150.9	154.5	155.3	156.5	156.7	157.1
28	76.7	84.0	100.1	105.4	115.0	117.2	121.1	122.1	123.1	123.3	123.9	124.0	123.9
26	72.4	78.2	90.3	93.6	98.9	100.0	101.7	102.0	102.7	102.8	102.9	103.0	102.8
24	68.3	72.6	80.3	82.3	85.5	85.9	87.0	87.3	87.4	87.4	87.8	87.9	87.9
22	62.6	66.0	71.5	72.8	74.6	75.3	75.5	75.9	75.4	75.5	75.6	75.6	75.8
20	57.8	60.4	63.5	64.4	65.4	65.6	65.7	65.7	66.0	66.0	65.9	65.9	65.8
18	52.4	54.4	56.4	57.0	57.7	57.9	57.8	57.8	58.1	58.1	58.0	58.0	57.9
16	47.0	48.4	49.8	50.1	50.7	50.7	50.9	50.9	50.9	50.9	50.7	50.7	50.7
14	42.2	43.1	44.1	44.3	44.6	44.7	44.3	44.3	44.5	44.5	44.5	44.5	44.4
12	37.8	38.5	38.6	38.9	38.9	39.0	38.8	38.8	38.8	38.8	38.8	38.8	39.0
10	32.8	33.3	33.7	33.8	33.8	33.8	33.7	33.7	33.6	33.6	33.6	33.6	33.7
8	28.3	28.6	29.0	29.0	28.9	28.9	29.0	29.1	28.9	28.9	29.0	29.0	29.0

Table B.10.: Values corresponding to Fig. 2.35b.

$\tilde{N}^{(\tilde{\Omega})}$	Quantization bits $q^{(\mathrm{CSI})}$												
	8	9	10	11	12	13	14	15	16	17	18	19	20
	Feedback $R^{(\mathrm{FB,HP},2)}$ in [kbit]												
32	14.8	16.7	18.5	20.4	22.2	24.1	25.9	27.8	29.6	31.5	33.3	35.2	37.0
30	14.0	15.8	17.5	19.3	21.0	22.8	24.5	26.3	28.0	29.8	31.5	33.3	35.0
28	13.2	14.9	16.5	18.2	19.8	21.5	23.1	24.8	26.4	28.1	29.7	31.4	33.0
26	12.4	14.0	15.5	17.1	18.6	20.2	21.7	23.3	24.8	26.4	27.9	29.5	31.0
24	11.6	13.1	14.5	16.0	17.4	18.9	20.3	21.8	23.2	24.7	26.1	27.6	29.0
22	10.8	12.2	13.5	14.9	16.2	17.6	18.9	20.3	21.6	23.0	24.3	25.7	27.0
20	10.0	11.3	12.5	13.8	15.0	16.3	17.5	18.8	20.0	21.3	22.5	23.8	25.0
18	9.2	10.4	11.5	12.7	13.8	15.0	16.1	17.3	18.4	19.6	20.7	21.9	23.0
16	8.4	9.5	10.5	11.6	12.6	13.7	14.7	15.8	16.8	17.9	18.9	20.0	21.0
14	7.6	8.6	9.5	10.5	11.4	12.4	13.3	14.3	15.2	16.2	17.1	18.1	19.0
12	6.8	7.7	8.5	9.4	10.2	11.1	11.9	12.8	13.6	14.5	15.3	16.2	17.0
10	6.0	6.8	7.5	8.3	9.0	9.8	10.5	11.3	12.0	12.8	13.5	14.3	15.0
8	5.2	5.9	6.5	7.2	7.8	8.5	9.1	9.8	10.4	11.1	11.7	12.4	14.4

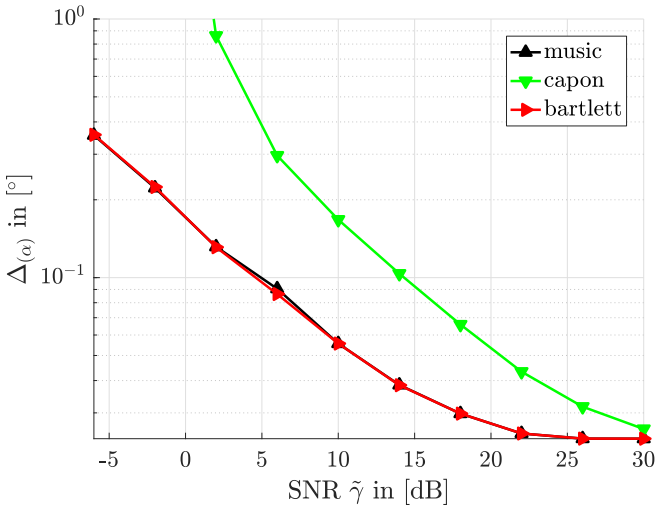
Table B.11.: Values corresponding to Fig. 2.36a.

$\tilde{N}^{(\mathrm{RB})}$	Quantization bits $q^{(\mathrm{CSI})}$												
	8	9	10	11	12	13	14	15	16	17	18	19	20
	Effective sum spectral efficiency $\tilde{C}_l^{(\mathrm{sum,O})}$ in [bit/s/Hz] with $\approx 8.1\,\%$ pilot overhead												
50	81.2	90.6	119.4	129.3	156.7	165.6	188.6	196.1	212.0	215.6	222.8	223.7	226.9
48	80.8	89.9	118.1	128.6	155.8	164.8	188.0	195.0	211.1	214.5	222.3	223.7	227.0
46	80.4	89.3	118.1	127.5	155.0	163.9	187.4	194.4	210.0	213.4	221.7	222.5	225.8
44	79.4	88.4	117.3	127.0	153.9	162.7	186.6	193.4	208.9	212.6	220.4	221.5	224.6
42	78.5	87.6	116.6	126.2	153.5	162.0	185.7	192.3	208.4	212.2	219.5	220.6	223.6
40	77.5	87.1	115.2	124.7	152.5	160.7	185.0	191.8	207.3	210.8	218.8	219.7	223.2
38	77.1	86.0	114.1	123.6	151.3	160.1	183.8	190.6	206.3	210.1	218.3	219.2	222.0
36	75.7	84.5	113.2	122.4	150.0	158.7	181.9	189.0	204.7	208.4	217.4	218.2	220.6
34	75.2	83.7	112.3	121.8	149.6	157.8	181.2	187.6	203.5	207.0	214.2	215.3	218.9
32	73.8	81.9	111.8	120.9	148.6	156.7	179.0	185.4	201.8	205.4	212.9	213.9	216.3
30	72.4	80.5	110.4	119.0	147.5	155.8	178.2	184.6	200.7	203.9	210.6	211.8	214.4
28	71.5	79.6	108.8	117.6	145.0	152.9	175.9	182.1	198.0	201.0	207.9	209.3	211.7
26	69.8	77.5	107.2	115.8	143.3	151.1	174.2	179.8	194.9	198.0	205.1	206.5	209.1
24	68.7	76.2	105.1	113.6	141.4	148.6	171.9	177.0	192.0	195.3	202.5	203.5	206.2
22	68.2	75.3	104.8	112.8	140.1	146.9	169.2	174.1	189.2	192.0	198.2	199.1	201.9
20	67.6	75.0	104.0	111.7	138.3	145.8	166.8	172.0	185.0	187.6	194.1	195.4	197.2
18	69.2	76.1	105.3	112.8	139.4	146.1	165.8	170.4	183.6	185.7	191.8	193.0	194.7
16	72.2	79.6	108.0	115.4	140.6	147.0	166.4	171.2	182.7	184.9	190.4	191.3	193.2
14	78.2	85.7	112.4	119.9	143.1	149.2	167.1	171.0	181.9	184.0	188.1	188.6	189.6
12	83.0	90.5	114.4	121.5	141.9	147.9	163.7	167.6	175.6	177.1	180.1	180.5	181.3
10	83.3	90.4	110.4	117.1	134.8	140.0	152.6	156.0	161.9	162.7	164.3	164.6	165.0
9	81.2	88.1	105.9	112.4	128.2	133.1	144.2	147.2	151.3	152.0	153.4	153.6	153.9
8	77.7	84.0	99.9	105.8	119.7	124.4	133.6	135.9	139.1	139.6	140.4	140.5	140.7
7	72.9	78.6	92.4	97.9	109.8	114.0	121.4	123.3	125.2	125.7	126.2	126.3	126.4
6	66.8	71.9	83.7	88.5	98.4	101.9	107.5	108.7	110.0	110.3	110.5	110.6	110.6
5	59.5	63.8	73.6	77.7	85.5	88.3	92.0	92.8	93.5	93.7	93.8	93.8	93.8
4	51.0	54.5	62.2	65.5	71.2	73.2	75.3	75.7	76.0	76.1	76.2	76.2	76.2
3	41.3	43.9	49.5	51.8	55.5	56.5	57.5	57.7	57.9	57.9	57.9	57.9	57.9
2	30.1	31.9	35.2	36.5	38.2	38.6	38.9	38.9	39.0	39.0	39.0	39.0	39.0

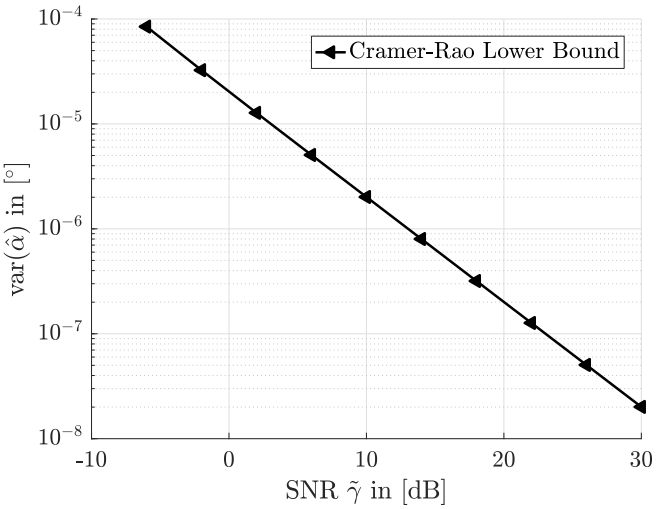
Table B.12.: Values corresponding to Fig. 2.36b.

	Quantization bits $q^{(\mathrm{CSI})}$												
	8	9	10	11	12	13	14	15	16	17	18	19	20
	Feedback $R^{(\mathrm{FB,HP,RB})}$ in [kbit]												
50	12.8	14.4	16.0	17.6	19.2	20.8	22.4	24.0	25.6	27.2	28.8	30.4	32.0
48	12.6	14.1	15.6	17.2	18.7	20.3	21.8	23.3	24.9	26.4	27.9	29.5	31.0
46	12.1	13.5	15.0	16.5	17.9	19.4	20.9	22.4	23.8	25.3	26.8	28.2	29.7
44	11.5	12.9	14.3	15.8	17.2	18.6	20.0	21.4	22.8	24.2	25.6	27.0	28.4
42	11.0	12.3	13.7	15.0	16.4	17.7	19.1	20.4	21.8	23.1	24.4	25.8	27.1
40	10.5	11.8	13.0	14.3	15.6	16.9	18.2	19.4	20.7	22.0	23.3	24.6	25.8
38	10.0	11.2	12.4	13.6	14.8	16.0	17.3	18.5	19.7	20.9	22.1	23.3	24.5
36	9.4	10.6	11.7	12.9	14.0	15.2	16.3	17.5	18.6	19.8	21.0	22.1	23.3
34	8.9	10.0	11.1	12.2	13.3	14.3	15.4	16.5	17.6	18.7	19.8	20.9	22.0
32	8.4	9.4	10.4	11.5	12.5	13.5	14.5	15.6	16.6	17.6	18.6	19.6	20.7
30	7.9	8.8	9.8	10.7	11.7	12.7	13.6	14.6	15.5	16.5	17.5	18.4	19.4
28	7.3	8.2	9.1	10.0	10.9	11.8	12.7	13.6	14.5	15.4	16.3	17.2	18.1
26	6.8	7.6	8.5	9.3	10.1	11.0	11.8	12.6	13.5	14.3	15.1	16.0	16.8
24	6.3	7.1	7.8	8.6	9.4	10.1	10.9	11.7	12.4	13.2	14.0	14.7	15.5
22	5.8	6.5	7.2	7.9	8.6	9.3	10.0	10.7	11.4	12.1	12.8	13.5	14.2
20	5.2	5.9	6.5	7.2	7.8	8.4	9.1	9.7	10.4	11.0	11.6	12.3	12.9
18	4.7	5.3	5.9	6.4	7.0	7.6	8.2	8.7	9.3	9.9	10.5	11.1	11.6
16	4.2	4.7	5.2	5.7	6.2	6.8	7.3	7.8	8.3	8.8	9.3	9.8	10.3
14	3.7	4.1	4.6	5.0	5.5	5.9	6.4	6.8	7.3	7.7	8.1	8.6	9.0
12	3.1	3.5	3.9	4.3	4.7	5.1	5.4	5.8	6.2	6.6	7.0	7.4	7.8
10	2.6	2.9	3.3	3.6	3.9	4.2	4.5	4.9	5.2	5.5	5.8	6.1	6.5
9	2.4	2.6	2.9	3.2	3.5	3.8	4.1	4.4	4.7	5.0	5.2	5.5	5.8
8	2.1	2.4	2.6	2.9	3.1	3.4	3.6	3.9	4.1	4.4	4.7	4.9	5.2
7	1.8	2.1	2.3	2.5	2.7	3.0	3.2	3.4	3.6	3.9	4.1	4.3	4.5
6	1.6	1.8	2.0	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9
5	1.3	1.5	1.6	1.8	2.0	2.1	2.3	2.4	2.6	2.8	2.9	3.1	3.2
4	1.0	1.2	1.3	1.4	1.6	1.7	1.8	1.9	2.1	2.2	2.3	2.5	2.6
3	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.7	1.8	1.9
2	0.5	0.6	0.7	0.7	0.8	0.8	0.9	1.0	1.0	1.1	1.2	1.2	1.3

Section 3.2.1

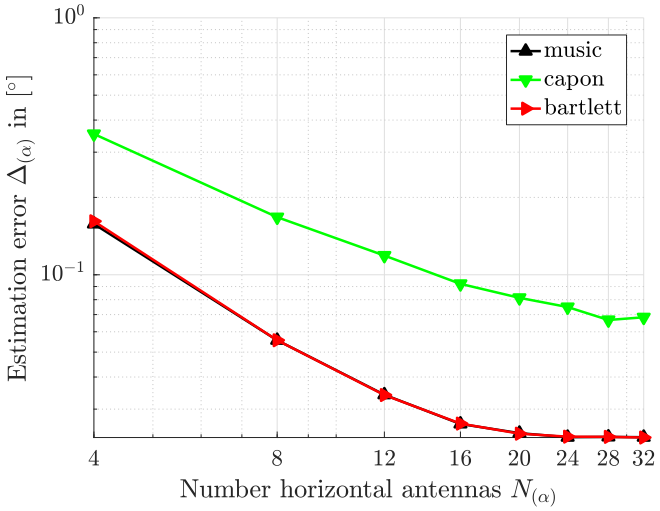


(a) Estimated error of horizontal DoA.

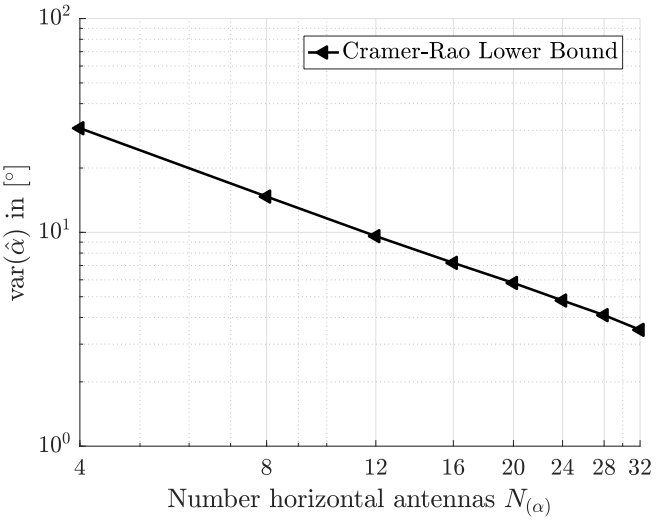


(b) CRLB of the horizontal DoA according to (3.40).

Figure B.1.: Comparison of scaling behavior between CRLB and DoA estimation techniques according over SNR $\tilde{\gamma}$ according to (3.5) in double logarithmic representation.

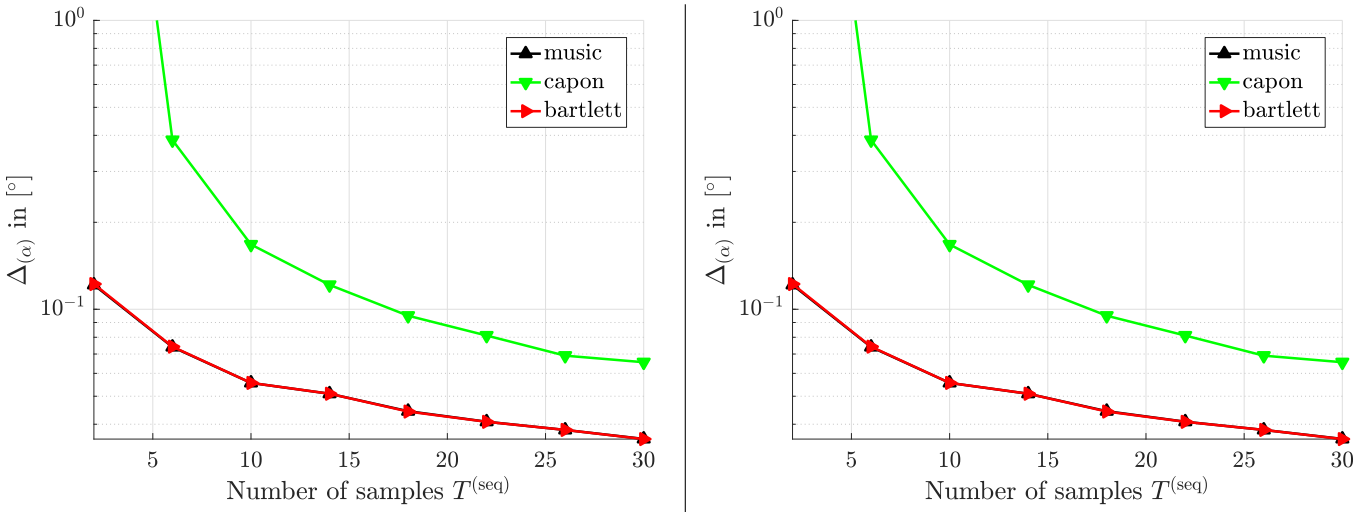


(a) Estimated error of horizontal DoA.



(b) CRLB of the horizontal DoA according to (3.40).

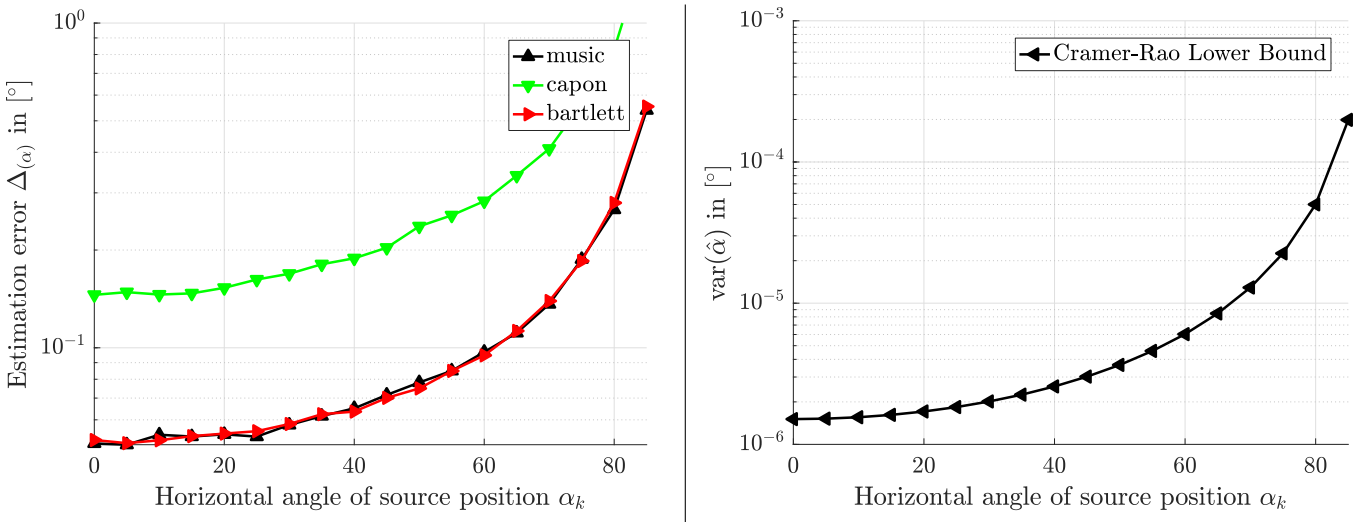
Figure B.2.: Comparison of scaling behavior between CRLB and DoA estimation techniques over number of horizontal antennas N_α in double logarithmic representation.



(a) Estimated error of horizontal DoA.

(b) CRLB of the horizontal DoA according to (3.40).

Figure B.3.: Comparison of scaling behavior between CRLB and DoA estimation techniques over number of samples or sequence length T in double logarithmic representation.



(a) Estimated error of horizontal DoA.

(b) CRLB of the horizontal DoA according to (3.40).

Figure B.4.: Comparison of scaling behavior between CRLB and DoA estimation techniques over horizontal angle of source position α_k in logarithmic representation.

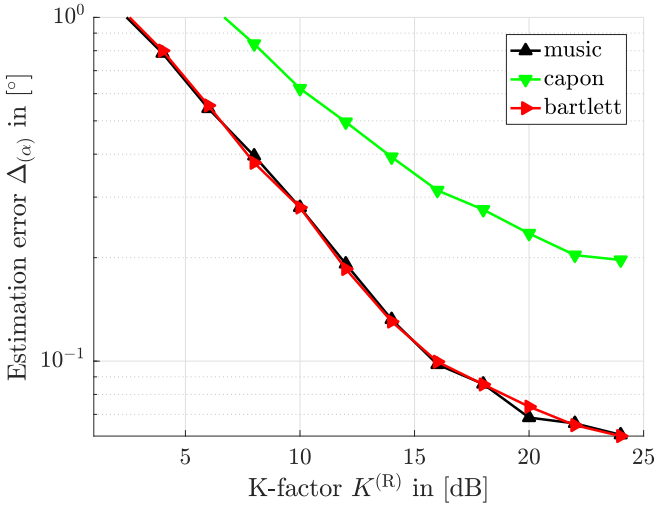


Figure B.5.: Estimated error of horizontal DoA in logarithmic representation.

Section 3.2.2

Table B.13.: ASSQ parameters obtained by exhaustive search for minimized complexity and estimation performance of MUSIC power spectrum.

S	$[\boldsymbol{\alpha}_f\ \boldsymbol{\beta}_f]$ in [°]	$[\mathbf{q}_\alpha\ \mathbf{q}_\beta]$ in [°]	$\mathbf{N}_\phi$	N_ϕ	Median Δ_{xyz} in [m]
5	$\begin{pmatrix} -50 & 50 & -50 & 50 \\ -25 & 25 & -25 & 25 \\ -5 & 5 & -5 & 5 \\ -1 & 1 & -1 & 1 \\ -0.2 & 0.2 & -0.2 & 0.2 \end{pmatrix}$	$\begin{pmatrix} 50 & 50 \\ 10 & 10 \\ 2 & 2 \\ 0.4 & 0.4 \\ 0.1 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 9 \\ 36 \\ 36 \\ 36 \\ 25 \end{pmatrix}$	142	0.36

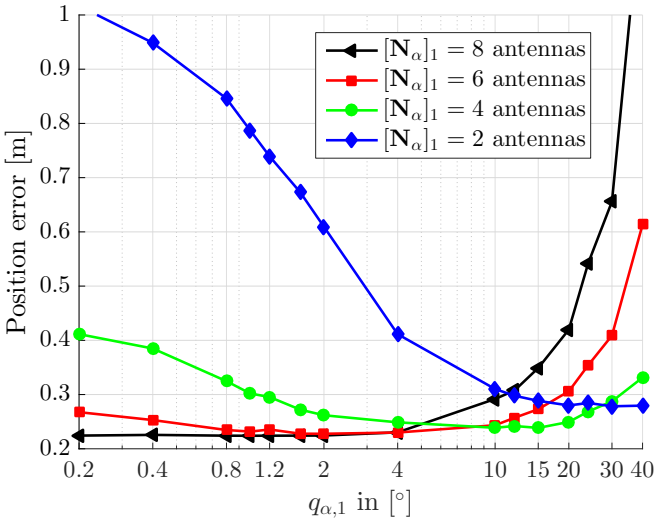
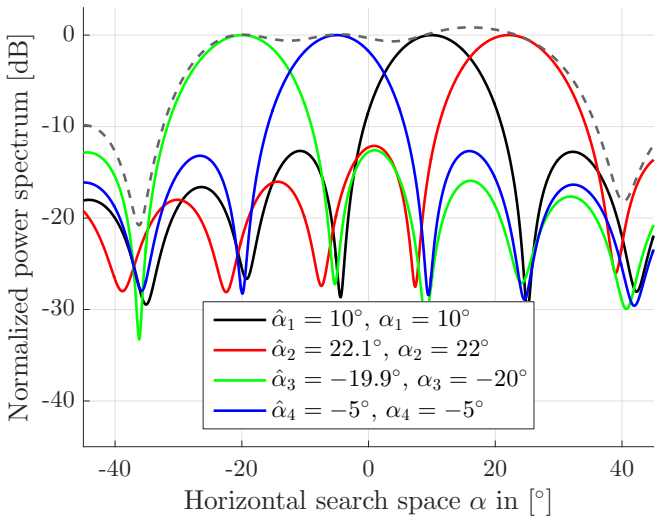
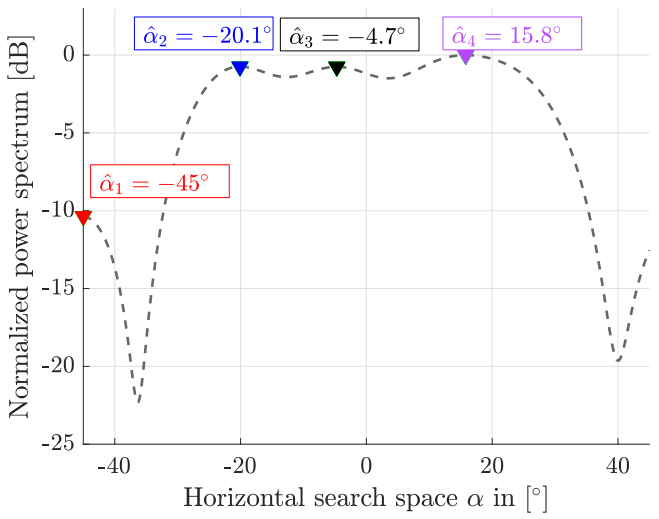


Figure B.6.: 90 %-ile values corresponding to in Fig. 3.23b.

Section 3.3

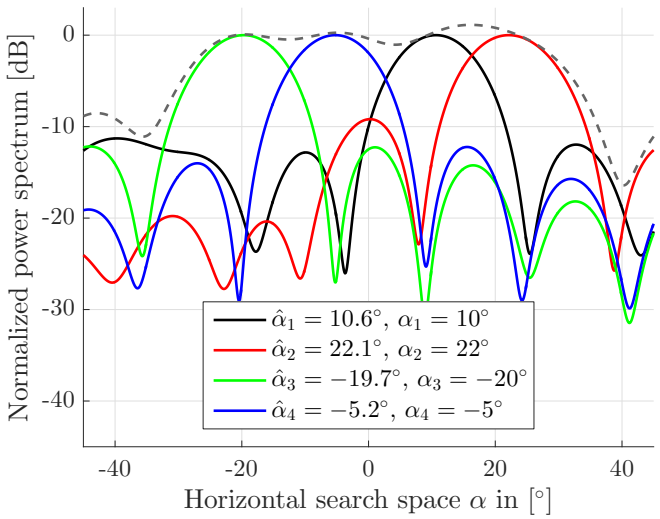


(a) Single source DoA estimation.

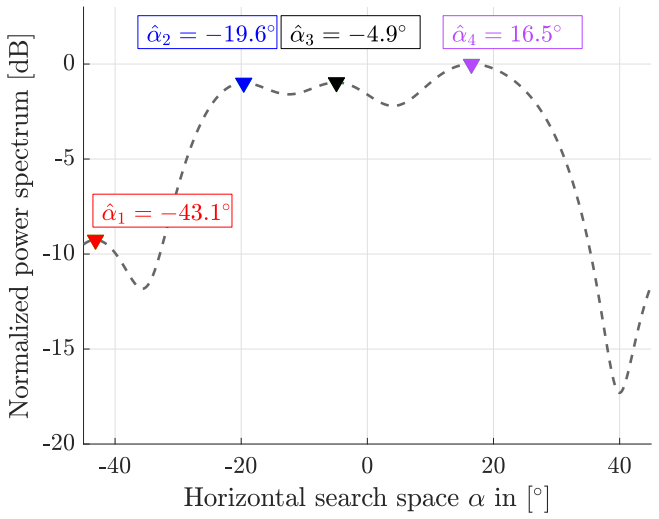


(b) Multi source DoA estimation.

Figure B.7.: Power spectrum of Bartlett and only LoS path, $L = 1$ with normalized channels.

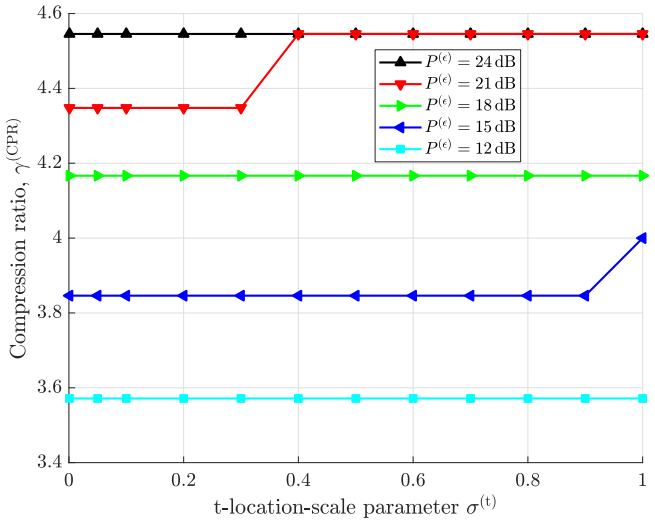


(a) Single source DoA estimation.

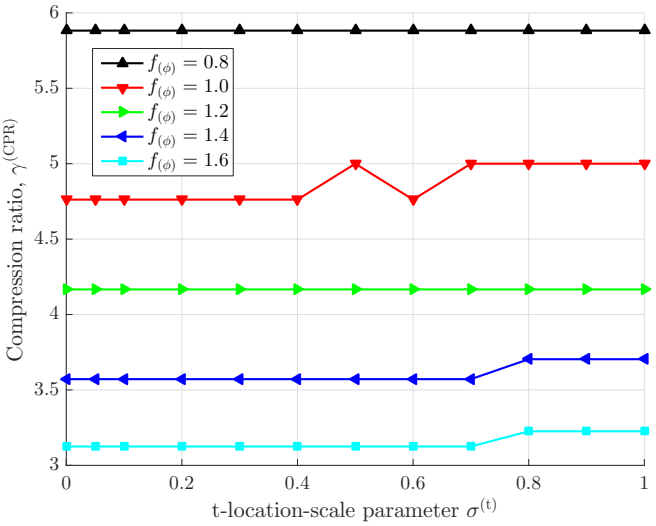


(b) Multi source DoA estimation.

Figure B.8.: Power spectrum of Bartlett and $L = 8$ multi-path components with normalized channels.



(a) Variation of design parameter P_ϵ, the power difference within the same group, for first step power grouping algorithm.



(b) Variation of design parameter f_ϕ, scaling the minimum angular distance between any two users, for second step angular grouping algorithm.

Figure B.9.: Evaluation of compression ratio, where the DoA estimation error is approximated with t location-scale probability distribution. The larger the t-location scale parameter $\sigma^{(\mathrm{t})}$ the larger the DoA estimation error.

Bibliography

[3GP14] 3GPP. Physical layer procedures, Release 11. Specification 36213, 3rd Generation Partnership Project, July 2014.

[3GP15a] 3GPP. Study on elevation beamforming / Full-Dimension (FD) Multiple Input Multiple Output (MIMO) for LTE, Release 13. Technical Report 36897, 3rd Generation Partnership Project, July 2015.

[3GP15b] 3GPP. Study on indoor positioning enhancements for UTRA and LTE, Release 13. Technical Report 37857, 3rd Generation Partnership Project, December 2015.

[3GP16a] 3GPP. LTE Positioning Protocol A (LPPa), Release 13. Specification 36455, 3rd Generation Partnership Project, March 2016.

[3GP16b] 3GPP. LTE Positioning Protocol (LPP), Release 13. Specification 36355, 3rd Generation Partnership Project, March 2016.

[3GP16c] 3GPP. Physical Channels and Modulation, Release 13. Specification 36211, 3rd Generation Partnership Project, January 2016.

[3GP16d] 3GPP. Study on Scenarios and Requirements for Next Generation Access Technologies, Release 14. Technical Report 38913, 3rd Generation Partnership Project, March 2016.

[3GP17a] 3GPP. Introduction of 1024QAM in LTE downlink, Release 15. Technical Report 36783, 3rd Generation Partnership Project, April 2017.

[3GP17b] 3GPP. Physical layer; General description, Release 15. Specification 38201, 3rd Generation Partnership Project, December 2017.

[3GP17c] 3GPP. Physical layer procedures for data, Release 15. Specification 36214, 3rd Generation Partnership Project, December 2017.

[3GP17d] 3GPP. Spatial channel model for Multiple Input Multiple Output (MIMO) simulations, Release 14. Technical Report 25996, 3rd Generation Partnership Project, September 2017.

[3GP17e] 3GPP. Study on 3D channel model for LTE, Release 12. Technical Report 36873, 3rd Generation Partnership Project, June 2017.

[3GP17f] 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz, Release 14. Technical Report 38901, 3rd Generation Partnership Project, March 2017.

[3GP17g] 3GPP. Study on New Radio (NR) Access Technology Physical Layer Aspects, Release 14. Technical Report 38802, 3rd Generation Partnership Project, October 2017.

[3GP17h] 3GPP. User Equipment (UE) radio transmission and reception, Release 14. Specification 36101, 3rd Generation Partnership Project, April 2017.

[3GP18a] 3GPP. Medium Access Control (MAC) protocol specification, Release 15. Specification 36321, 3rd Generation Partnership Project, March 2018.

[3GP18b] 3GPP. Physical channels and modulation, Release 15. Specification 38211, 3rd Generation Partnership Project, March 2018.

[3GP18c] 3GPP. Physical layer procedures for data, Release 15. Specification 38214, 3rd Generation Partnership Project, March 2018.

[3GP18d] 3GPP. Requirements for support of radio resource management, Release 15. Specification 36133, 3rd Generation Partnership Project, March 2018.

[ACH07] J. G. Andrews, W. Choi, and R. W. Heath. Overcoming Interference in Spatial Multiplexing MIMO Cellular Networks. *IEEE Wireless Communications*, 14(6):95–104, December 2007.

[AGSR14] D. D. Arumugam, J. D. Griffin, D. D. Stancil, and D. S. Ricketts. Three-dimensional position and orientation measurements using magneto-quasistatic fields and complex image theory [measurements corner]. *IEEE Antennas and Propagation Magazine*, 56(1):160–173, February 2014.

[Ake03] D. E. Akerberg. Antenna switching diversity. Patent US6553078B1, April 2003.

[AM12] A. Ashikhmin and T. Marzetta. Pilot Contamination Precoding in Multi-Cell Large Scale Antenna Systems. In *Proceedings International Symposium on Information Theory (ISIT)*, pages 1137–1141, July 2012.

[ANAC13] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire. Joint Spatial Division and Multiplexing: The Large-Scale Array Regime. *IEEE Transactions on Information Theory*, 59(10):6441–6463, October 2013.

[ASH14] A. Aminu, M. Secmen, and M. A. Husein. The Effectiveness of JADE Based Minimum Variance Distortionless Response Algorithm For DOA Estimation and Adaptive Beamforming in Multipath Smart Antenna Application. In *IEEE 6th International Conference on Adaptive Science Technology (ICAST)*, October 2014.

[ASP14] A. Aminu, M. Secmen, and S. Poyraz. The Comparison of JADE Based DOA Estimation Methods for Unknown Noncoherent Source Groups Containing Coherent Signals with Frequency Matching. In *20th International Conference on Microwaves, Radar, and Wireless Communication (MIKON)*, pages 1–6, June 2014.

[AVW18] R. Ahmed, E. Visotsky, and T. Wild. Explicit CSI Feedback Design for 5G New Radio phase II. In *22th International ITG Workshop on Smart Antennas WSA*, pages 1–7, March 2018.

[AZDG16] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta. Are we approaching the fundamental limits of wireless network densification? *IEEE Communications Magazine*, 54(10):184–190, October 2016.

[BAJ+15] M. H. Bergen, A. Arafa, X. Jin, R. Klukas, and J. F. Holzman. Characteristics of Angular Precision and Dilution of Precision for Optical Wireless Positioning. *Journal of Lightwave Technology*, 33(20):4253–4260, October 2015.

[Bal16] C. A. Balanis. *Antenna Theory: Analysis and Design*. John Wiley & Sons, 2nd edition, 2016.

[Bar48] M. S. Bartlett. Smoothing Periodograms from Time Series with Continuous Spectra. *Nature*, 161(4096):686–687, May 1948.

[Bar83] A. Barabell. Improving the Resolution Performance of Eigenstructure-Based Direction-Finding Algorithms. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 8, pages 336–339, April 1983.

[BDJT12] K. Börner, J. Dommel, S. Jaeckel, and L. Thiele. On The Requirements for Quasi-Deterministic Radio Channel Models for Heterogeneous Networks. In *International Symposium on Signals, Systems, and Electronics (ISSSE)*, pages 1–5, October 2012.

[Ben07] A. Bensky. *Wireless Positioning Technologies and Applications*. Artech House, 2007.

[BHKD14] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah. Massive MIMO Systems With Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits. *IEEE Transactions on Information Theory*, 60(11):7112–7139, November 2014.

[BHS05] D. S. Baum, J. Hansen, and J. Salo. An Interim Channel Model for Beyond-3G Systems. In *IEEE 61st Vehicular Technology Conference (VTC), Spring*, volume 5, pages 3132–3136, 2005.

[BHS17] E. Björnson, J. Hoydis, and L. Sanguinetti. Pilot Contamination is Not a Fundamental Asymptotic Limitation in Massive MIMO. In *International Conference on Communications (ICC)*, pages 1–6, 2017.

[BHS18] E. Björnson, J. Hoydis, and L. Sanguinetti. Massive MIMO has Unlimited Capacity. *IEEE Transactions on Wireless Communications*, 17(1):574–590, January 2018.

[Bie07] T. Biedermann. A Dual-Polarized Patch Antenna with High Dcecoupling. In *2nd International ITG Conference on Antennas*, pages 166–170, March 2007.

[Bjö19] E. Björnson. Massive MIMO. Online https://ma-mimo.ellintech.se/2019/05/02/when-will-hybrid-beamforming-disappear/, edited on 02 May 2019, accessed 27-August-2019.

[BK83] G. Bienvenu and L. Kopp. Optimality of High Resolution Array Processing Using the Eigensystem Approach. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(5):1235–1248, October 1983.

[BLD16] E. Björnson, E. G. Larsson, and M. Debbah. Massive MIMO for Maximal Spectral Efficiency: How Many Users and Pilots Should Be Allocated? *IEEE Transactions on Wireless Communications*, 15(2):1293–1308, February 2016.

[BLG+15] S. Barua, Sinh Cong Lam, P. Ghosa, Shiqi Xing, and K. Sandrasegaran. A Survey of Direction of Arrival Estimation Techniques and Implementation of Channel Estimation based on SCME. In *12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 1–5, June 2015.

[BLM+14] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer. Network Densification: The Dominant Theme for Wireless Evolution into 5G. *IEEE Communications Magazine*, 52(2):82–89, February 2014.

[BLM16] E. Björnson, E. G. Larsson, and T. L. Marzetta. Massive MIMO: Ten Myths and One Critical Question. *IEEE Communications Magazine*, 54(2):114–123, February 2016.

[BM03] U. Baysal and R. L. Moses. On the Geometry of Isotropic Arrays. *IEEE Transactions on Signal Processing*, 51(6):1469–1478, June 2003.

[Boa66] M. L. Boas. *Mathematical Methods in the Physical Sciences*. Wiley New York, 2nd edition, 1966.

[BSHD15] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah. Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the Answer? *IEEE Transactions on Wireless Communications*, 14(6):3059–3075, June 2015.

[Bun19] Bundesnetzagentur. Electronic Communications Services. Spectrum for wireless access for the

provision of telecommunications services, Mobile broadband - Spectrum for 5G, Frequency auction. Online https://www.bundesnetzagentur.de, edited at least on 12 June 2019, accessed 27-August-2019.

[BWK08] C. Bockelmann, D. Wübben, and K. D. Kammeyer. Mutual Information Based Rate Adaptation for MIMO-OFDM Systems with Finite Signal Alphabets. In *International ITG Workshop on Smart Antennas (WSA)*, pages 60–67, February 2008.

[Cap69] J. Capon. High-Resolution Frequency-Wavenumber Spectrum Analysis. *Proceedings of the IEEE*, 57(8):1408–1418, August 1969.

[CF14] C. H. Chen and K. T. Feng. Geometry-Improved Location Estimation Algorithm for LTE-A Wireless Networks. In *IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pages 2040–2044, September 2014.

[CHE13] CMCC, Huawei, and Ericsson. TDD Spectrum White Paper. White paper, Global TD-LTE Initiative, 2013.

[Cis12] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011-2016. White paper, Cisco, 2012.

[Cis17] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021. White paper, Cisco, 2017.

[Cis19] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017-2022. White paper, Cisco, 2019.

[Cos83] M. H. M. Costa. Writing on dirty paper (Corresp.). *IEEE Transactions on Information Theory*, 29(3):439–441, 5 1983.

[CPT+13] H. Chen, Z. Pan, L. Tian, J. Shi, G. Yang, and M. Suzuki. A Novel AWSF Algorithm for DOA Estimation in Virtual MIMO Systems. *IEEE Journal on Selected Areas in Communications*, 31(10):1994–2003, October 2013.

[Cra45] H. Cramér. *Mathematical Methods of Statistics*, volume 9. Princeton University Press, 1945.

[CRT01] N. Chiurtu, B. Rimoldi, and E. Telatar. On The Capacity Of Multi-Antenna Gaussian Channels. In *Proceedings of IEEE International Symposium on Information Theory*, page 53, June 2001.

[CWK+12] K. Cujia, M. Wirz, M. B. Kjärgaard, D. Roggen, and G. Träster. DactyLoc: A minimally georeferenced WiFi+GSM-fingerprint-based localization method for positioning in urban spaces. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–9, November 2012.

[DAKN11] S. Dayekh, S. Affes, N. Kandil, and C. Nerguizian. Radio-Localization in Underground Narrow-Vein Mines Using Neural Networks with In-built Tracking and Time Diversity. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1788–1793, March 2011.

[Dem06] A. G. Dempster. Dilution of precision in angle-of-arrival positioning systems. *Electronics Letters*, 42(5):291–292, March 2006.

[Dho10] T. S. Dhope. Application of MUSIC, ESPRIT and ROOT MUSIC in DOA Estimation. *Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia*, 2010.

[DKT13] J. Dommel, M. Kurras, and L. Thiele. Inter-Cell Interference-Coordination for Distributed

MIMO-Systems based on Wideband Precoder Selection. In *IEEE AFRICON*, pages 1–5, September 2013.

[DKTH14] J. Dommel, P.-P. Knust, L. Thiele, and T. Haustein. Massive MIMO for Interference Management in Heterogeneous Networks. In *IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 289–292, June 2014.

[DNO+16] M. Dohler, T. Nakamura, A. Osseiran, J. F. Monserrat, O. Queseth, and P. Marsch. *5G Mobile and Wireless Communications Technology.* Cambridge University Press, 2016.

[Doc14] NTT Docomo. 5G Radio Access: Requirements, Concept and Technologies. White paper, NTT Docomo, July 2014.

[DSB+05] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk. Distributed Compressed Sensing of Jointly Sparse Signals. In *Proceedings of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, pages 1537–1541, October 2005.

[EKPP10] H. El-Sallabi, P. Kyritsi, A. Paulraj, and G. Papanicolaou. Experimental Investigation on Time Reversal Precoding for Space-Time Focusing in Wireless Communications. *IEEE Transactions on Instrumentation and Measurement*, 59(6):1537–1543, June 2010.

[Eri15] Ericsson. 5G Radio Access. White paper, Ericsson, February 2015.

[FGH06] M. Fuchs, G. D. Galdo, and M. Haardt. Low complexity spatial scheduling ProSched for MIMO systems with multiple base stations and a central controller. In *Proc. ITG/IEEE Workshop on smart Antennas, Guenzburg, Germany*, March 2006.

[FGH07] M. Fuchs, G. D. Galdo, and M. Haardt. Low-Complexity Space-Time-Frequency Scheduling for MIMO Systems With SDMA. *IEEE Transactions on Vehicular Technology*, 56(5):2775 –2784, September 2007.

[Fis14] S. Fischer. Observed Time Difference Of Arrival (OTDOA) Positioning in 3GPP LTE. *Qualcomm white paper*, 1(1):1–62, June 2014.

[Fos96] G. J. Foschini. Layered Space-Time Architecture for Wireless Communication in a Fading Environment When Using Multi-Element Antennas. *Bell labs technical journal*, 1(2):41–59, 1996.

[FRZ+17] G. Fodor, N. Rajatheva, W. Zirwas, L. Thiele, M. Kurras, K. Guo, A. Tolli, J. H. Sorensen, and E. De Carvalho. An Overview of Massive MIMO Technology Components in METIS. *IEEE Communications Magazine*, 55(6):155–161, 2017.

[GAPM14] A. Griffin, A. Alexandridis, D. Pavlidi, and A. Mouchtaris. Real-Time Localization of Multiple Audio Sources in a Wireless Acoustic Sensor Network. In *Proceedings of 22nd European Signal Processing Conference (EUSIPCO)*, pages 306–310, September 2014.

[GDWC15] Z. Gao, L. Dai, Z. Wang, and S. Chen. Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO. *IEEE Transactions on Signal Processing*, 63(23):6169–6183, December 2015.

[GM13] A. Griffin and A. Mouchtaris. Localizing Multiple Audio Sources from DOA estimates in a Wireless Acoustic Sensor Network. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 1–4, October 2013.

[GME99] L. J. Greenstein, D. G. Michelson, and V. Erceg. Moment-Method Estimation of the Ricean -Factor. *IEEE Communications Letters*, 3(6):175–176, June 1999.

[Gra84] R. Gray. Vector Quantization. *IEEE ASSP Magazine*, 1(2):4–29, April 1984.

[GRK+16] M. Grossmann, V. Ramireddy, J. König, M. Landmann, F. Römer, G. Del Galdo, and R. Perthold. Antenna Array Optimization Strategies for Robust Direction Finding. In *10th European Conference on Antennas and Propagation (EuCAP)*, pages 1–5, April 2016.

[GSPE+14] U. Gustavsson, C. Sanchez-Perez, T. Eriksson, F. Athley, G. Durisi, P. Landin, K. Hausmair, C. Fager, and L. Svensson. On the Impact of Hardware Impairments on Massive MIMO. In *Proceedings of Globecom (Workshop)*, pages 294–300, December 2014.

[HCPR11] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad. Achieving Large Spectral Efficiency with TDD and Not-so-Many Base-Station Antennas. In *IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications*, pages 1346–1349, September 2011.

[HCPR12] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad. Achieving "Massive MIMO" Spectral Efficiency with a Not-so-Large Number of Antennas. *IEEE Transactions on Wireless Communications*, 11(9):3226 –3239, September 2012.

[HCV99] D. Hilbert and S. Cohn-Vossen. *Geometry and the Imagination*, volume 87. American Mathematical Soc., 1999.

[HH15] E. Hossain and M. Hasan. 5G Cellular: Key Enabling Technologies and Research Challenges. *IEEE Instrumentation Measurement Magazine*, 18(3):11–21, June 2015.

[HHI17] Fraunhofer HHI. QuaDRiGa - QUasi Deterministic RadIo channel GenerAtor. Online http://quadriga-channel-model.de/, edited on 27 June 2019, accessed 27-August-2019.

[HJL+14] Y. Han, S. Jin, X. Li, Y. Huang, L. Jiang, and G. Wang. Design of double codebook based on 3D dual-polarized channel for multiuser MIMO system. *EURASIP Journal on Advances in Signal Processing*, 2014(1):1–13, 2014.

[HKL02] K. Hugl, K. Kalliola, and J. Laurila. Spatial Reciprocity of Uplink and Downlink Radio Channels in FDD Systems. *Proc. COST*, 273(2):066, May 2002.

[HKT15] K. S. Hassan, M. Kurras, and L. Thiele. Performance of Distributed Compressive Sensing Channel Feedback in Multi-User Massive MIMO. In *IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 430–436, October 2015.

[HN03] M. Hawkes and A. Nehorai. Wideband Source Localization Using a Distributed Acoustic Vector-Sensor Array. *IEEE Transactions on Signal Processing*, 51(6):1479–1491, June 2003.

[HSKS14] A. Hans, A. Sharma, K. Kumar, and N. Singh. An Overview of Handoff Procedure in LTE Technology. In *International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)*, pages 391–394, November 2014.

[HtBD11] J. Hoydis, S. ten Brink, and M. Debbah. Massive MIMO: How many antennas do we need? In *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 545 –550, September 2011.

[HtBD13] J. Hoydis, S. ten Brink, and M. Debbah. Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need? *IEEE Journal on Selected Areas in Communications*, 31(2):160–171, February 2013.

[Hua13] Huawei. 5G: A Technology Vision. White paper, Huawei, 2013.

[HZT+16] K. Haneda, J. Zhang, L. Tan, G. Liu, Y. Zheng, H. Asplund, J. Li, Y. Wang, D. Steer, C. Li, T. Balercia, S. Lee, Y. Kim, A. Ghosh, T. Thomas, T. Nakamura, Y. Kakishima, T. Imai, H. Papadopoulos, T. S. Rappaport, G. R. MacCartney, M. K. Samimi, S. Sun, O. Koymen, S. Hur, J. Park, C. Zhang, E. Mellios, A. F. Molisch, S. S. Ghassamzadeh, and A. Ghosh. 5G 3GPP-Like Channel Models for Outdoor Urban Microcellular and Macrocellular Environments. In *IEEE 83rd Vehicular Technology Conference (VTC Spring)*, pages 1–7, May 2016.

[IEE11] IEEE. IEEE Standard for Local and metropolitan area networks -Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs). Specification 802.15.4, IEEE Computer Society, September 2011.

[IH08] T. Inoue and R. W. Heath. Kerdock Codes for Limited Feedback MIMO Systems. In *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, pages 3113–3116, March 2008.

[IH09] T. Inoue and R. W. Heath. Kerdock Codes for Limited Feedback Precoded MIMO Systems. *IEEE Transactions on Signal Processing*, 57(9):3711–3716, September 2009.

[ITU15] ITU. World Radiocommunication Conference allocates spectrum for future innovation - Conference outcomes to spur long-term investments in ICT industry. Online https://www.itu.int/net/pressoffice/press_releases/2015/56.aspx, International Telecommunications Union (ITU), edited on 27 November 2015, accessed 27-August-2019.

[JAMV11] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath. Pilot Contamination and Precoding in Multi-Cell TDD Systems. *IEEE Transactions on Wireless Communications*, 10(8):2640–2651, August 2011.

[JBR+11] D. Jonathan, B. Biljana, B. Rajarajan, G. Rizwan, et al. MU-MIMO in LTE Systems. *EURASIP Journal on Wireless Communications and Networking*, 2011:12, 2011.

[JBTJ12] S. Jaeckel, K. Börner, L. Thiele, and V. Jungnickel. A Geometric Polarization Rotation Model for the 3-D Spatial Channel Model. *IEEE Transactions on Antennas and Propagation*, 60(12):5966–5977, 2012.

[Jen16] M. A. Jensen. A History of MIMO Wireless Communications. In *IEEE International Symposium on Antennas and Propagation (APSURSI)*, pages 681–682, June 2016.

[JKS17] J. Joung, E. Kurniawan, and S. Sun. Channel Correlation Modeling and its Application to Massive MIMO Channel Feedback Reduction. *IEEE Transactions on Vehicular Technology*, 66(5):3787–3797, May 2017.

[JMCN15] Z. Jiang, A. Molisch, G. Caire, and Z. Niu. Achievable Rates of FDD Massive MIMO Systems with Spatial Channel Correlation. *IEEE Transactions on Wireless Communications*, 14(5):2868–2882, 2015.

[JP06] L. Josefsson and P. Persson. *Conformal Array Antenna Theory and Design*. Wiley-IEEE Press, 1st edition, February 2006.

[JRB+17] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein. *Quasi Deterministic Radio Channel Generator User Manual and Documentation*, May 2017.

[JRBT14] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele. QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials. *IEEE Transactions on Antennas and Propagation*, 62(6):3242–3256, June 2014.

[KB75] H. P. Kramer and J. B. Bruckner. Iterations of a Non-Linear Transformation for Enhancement of Digital Images. *Pattern Recognition*, 7(1):53 – 58, 1975.

[KBT+12] M. Kurras, K. Börner, L. Thiele, M. Olbrich, and T. Haustein. Achievable System Performance Gains Using Distributed Antenna Deployments. In *IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 143 –148, September 2012.

[KJTB15] M. Kurras, S. Jaeckel, L. Thiele, and V. Braun. CSI Compression and Feedback for Network MIMO. In *IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5, May 2015.

[KKA13] V. Krishnaveni, T. Kesavamurthy, and B. Aparna. Beamforming for Direction-of-Arrival (DOA) Estimation-A Survey. *International Journal of Computer Applications*, 61(11):1–8, January 2013.

[KKS13] C. Kim, T. Kim, and J.-Y. Seol. Multi-Beam Transmission Diversity with Hybrid Beamforming for MIMO-OFDM Systems. In *IEEE Globecom Workshop*, pages 61–65, December 2013.

[KKT12] P.-H. Kuo, H. T. Kungy, and P.-A. Ting. Compressive Sensing Based Channel Feedback Protocols for Spatially-Correlated Massive Antenna Arrays. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 492–497, April 2012.

[KL14] M. Karlsson and E. G. Larsson. On the Operation of Massive MIMO with and without Transmitter CSI. In *IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5, June 2014.

[KRTT13] M. Kurras, L. Raschkowski, M. Talaat, and L. Thiele. Massive SDMA with Large Scale Antenna Systems in a Multi-Cell Environment. In *IEEE Conference, AFRICON*, 2013.

[KSHT15] M. Kurras, M. Shehata, K. Hassan, and L. Thiele. Spatial Interference Management with Hierarchical Precoding in Ultra-Dense Heterogeneous Networks. In *IEEE 11th International Conference onWireless and Mobile Computing, Networking and Communications (WiMob)*, pages 520–526, October 2015.

[KT06] K. Kawamura and T. Tanaka. Study on the improvement of measurement accuracy in GPS. In *SICE-ICASE International Joint Conference*, pages 1372–1375, October 2006.

[KTC15a] M. Kurras, L. Thiele, and G. Caire. Interference Mitigation and Multiuser Multiplexing with Beam-Steering Antennas. In *Proceedings of 19th International ITG Workshop on Smart Antennas (WSA)*, pages 1–5, March 2015.

[KTC15b] M. Kurras, L. Thiele, and G. Caire. Multi-Stage Beamforming for Interference Coordination in Massive MIMO Networks. In *49th Asilomar Conference on Signals, Systems and Computers*, pages 700–703, November 2015.

[KTH14a] L. Kumar, A. Tripathy, and R. M. Hegde. Robust Multi-Source Localization Over Planar Arrays Using MUSIC-Group Delay Spectrum. *IEEE Transactions on Signal Processing*, 62(17):4627–4636, September 2014.

[KTH14b] M. Kurras, L. Thiele, and T. Haustein. Interference Aware Massive SDMA with a Large Uniform Rectangular Antenna Array. In *European Conference on Networks and Communications (EuCNC)*, pages 1–5, June 2014.

[KTH+16a] M. Kurras, L. Thiele, T. Haustein, W. Lei, and C. Yan. Full Dimension MIMO for Frequency Division Duplex under Signaling and Feedback Constraints. In *24th European Signal Processing Conference (EUSIPCO)*, pages 1985–1989, August 2016.

[KTH+16b] M. Kurras, L. Thiele, T. Haustein, X. Peng, and M. Ariyoshi. Complexity Reduction for Direction of Arrival Estimation with Massive MIMO. In *IEEE 84rd Vehicular Technology Conference (VTC Fall)*, September 2016.

[KTPI17] M. Kurras, L. Thiele, X. Peng, and N. Ishii. Direction of Arrival Based Positioning in Three Dimensional Coordinates. In *21th International ITG Workshop on Smart Antennas (WSA)*, pages 1–7, March 2017.

[KV96] H. Krim and M. Viberg. Two Decades of Array Signal Processing Research: The Parametric Approach. *IEEE Signal Processing Magazine*, 13(4):67–94, July 1996.

[Lan99] R. B. Langley. Dilution of Precision. *GPS world*, 10(5):52–59, 1999.

[LCG+12] L. Liu, R. Chen, S. Geirhofer, K. Sayana, Z. Shi, and Y. Zhou. Downlink MIMO in LTE-Advanced: SU-MIMO vs. MU-MIMO. *IEEE Communications Magazine*, 50(2):140–147, February 2012.

[LDBL07] H. Liu, H. Darabi, P. Banerjee, and J. Liu. Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6):1067–1080, November 2007.

[Lem06] D. Lemire. Streaming Maximum-Minimum Filter Using No More Than Three Comparisons Per Element. *Nordic Journal of Computing*, 13(4):328–339, December 2006.

[LETM14] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta. Massive MIMO for Next Generation Wireless Systems. *IEEE Communications Magazine*, 52(2):186–195, February 2014.

[Lev15] R. Levorato. *Robot Audio Localization using Directions Of Arrival*. PhD thesis, University of Padova, July 2015.

[LGP12] J. Lorincz, T. Garma, and G. Petrovic. Measurements and Modelling of Base Station Power Consumption under Real Traffic Loads. *Sensors*, 12(4):4281–4310, 2012.

[LH03a] D. J. Love and R. W. Heath. Corrections to "Equal Gain Transmission in Multiple-Input Multiple-Output Wireless Systems". *IEEE Transactions on Communications*, 51(9):1613 – 1613, September 2003.

[LH03b] D. J. Love and R. W. Heath. Equal Gain Transmission in Multiple-Input Multiple-Output Wireless Systems. *IEEE Transactions on Communications*, 51(7):1102–1110, July 2003.

[LH05] D. J. Love and R. W. Heath. Limited Feedback Unitary Precoding for Spatial Multiplexing Systems. *IEEE Transactions on Information Theory*, 51(8):2967–2976, August 2005.

[LHL+08] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews. An Overview of Limited Feedback in Wireless Communication Systems. *IEEE Journal on Selected Areas in Communications*, 26(8):1341–1365, October 2008.

[LHYS13] W. Liu, S. Han, C. Yang, and C. Sun. Massive MIMO or Small Cell Network: Who is More Energy Efficient? In *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 24–29, April 2013.

[LJLH12] M. Lee, S. H. Jung, S. Lee, and D. Han. Elekspot: A Platform for Urban Place Recognition via Crowdsourcing. In *IEEE/IPSJ 12th International Symposium on Applications and the Internet (SAINT)*, pages 190–195, July 2012.

[LKCSM09] S. J. Lee, D. S. Kwon, G. Caire, and H. Shirani-Mehr. Method for Channel State Feedback by Quantization of Time-Domain Coefficients. Patent US 2009/0185607 A1, July 23 2009.

[Lo99] T. K. Y. Lo. Maximum Ratio Transmission. *IEEE Transactions on Communications*, 47(10):1458 –1461, October 1999.

[LOP+12] L. Lingfeng, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. D. Doncker. The COST 2100 MIMO Channel Model. *IEEE Wireless Communications*, 19(6):92–99, 2012.

[LP16] E. G. Larsson and H. V. Poor. Joint Beamforming and Broadcasting in Massive MIMO. *IEEE Transactions on Wireless Communications*, 15(4):3058–3070, April 2016.

[LSS06] X. Lin, N. B. Shroff, and R. Srikant. A Tutorial on Cross-Layer Optimization in Wireless Networks. *IEEE Journal on Selected Areas in Communications*, 24(8):1452–1463, August 2006.

[LSZ+13] J. Li, X. Su, J. Zeng, Y. Zhao, S. Yu, L. Xiao, and X. Xu. Codebook Design for Uniform Rectangular Arrays of Massive Antennas. In *IEEE 77th Vehicular Technology Conference (VTC Spring)*, pages 1–5, June 2013.

[Luo16] X. Luo. Multiuser Massive MIMO Performance With Calibration Errors. *IEEE Transactions on Wireless Communications*, 15(7):4521–4534, July 2016.

[LZP11] S. Lanzisera, D. Zats, and K. S. J. Pister. Radio Frequency Time-of-Flight Distance Measurement for Low-Cost Wireless Sensor Localization. *Sensors Journal, IEEE*, 11(3):837–845, March 2011.

[MAA+13] Y. Mehmood, W. Afzal, F. Ahmad, U. Younas, I. Rashid, and I. Mehmood. Large Scaled Multi-User MIMO System so called Massive MIMO Systems for Future Wireless Communication Networks. In *19th International Conference on Automation and Computing (ICAC)*, pages 1–4, September 2013.

[Mar10] T. L. Marzetta. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Transactions on Wireless Communications*, 9(11):3590–3600, November 2010.

[Mar15] T. L. Marzetta. Massive MIMO: An Introduction. *Bell Labs Technical Journal*, 20:11–22, 2015.

[MCW05] D. Malioutov, M. Cetin, and A. S. Willsky. Sparse Signal Reconstruction Perspective for Source Localization With Sensor Arrays. *IEEE Transactions on Signal Processing*, 53(8):3010–3022, August 2005.

[MF09] G. Mao and B. Fidan. *Localization Algorithms and Strategies for Wireless Sensor Networks: Monitoring and Surveillance Techniques for Target Tracking*. IGI Global, 2009.

[MGZF13] W. Mennerich, M. Grieger, W. Zirwas, and G. Fettweis. Interference Mitigation Framework for Cellular Mobile Radio Networks. *International Journal of Antennas and Propagation*, 2013:15, 2013.

[MJY10] X. Mao, J. Jin, and J. Yang. Wireless Channel Modeling Methods: Classification, Comparison and Application. In *5th International Conference on Computer Science Education*, pages 1669–1673, August 2010.

[MMB+15] J. F. Monserrat, G. Mange, V. Braun, H. Tullberg, G. Zimmermann, and Ömer Bulakci. METIS Research Advances towards the 5G Mobile and Wireless System Defintion. *EURASIP Journal on Wireless Communications and Networking*, 2015(1):1–16, 2015.

[MMWW04] G. Merziger, G. Mühlbach, D. Wille, and T. Wirth. *Formeln + Hilfen zur Höheren Mathematik*. Binomi Verlag, 2004.

[MZ10] W. Mennerich and W. Zirwas. Implementation Issues of the Partial CoMP Concept. In

IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), pages 1939–1944, September 2010.

[NAAC12] J. Nam, J.-Y. Ahn, A. Adhikary, and G. Caire. Joint Spatial Division and Multiplexing: Realizing Massive MIMO Gains with Limited Channel State Information. In *46th Annual Conference on Information Sciences and Systems (CISS)*, pages 1 –6, March 2012.

[NAAC14] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire. Joint Spatial Division and Multiplexing: Opportunistic Beamforming, User Grouping and Simplified Downlink Scheduling. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):876–890, October 2014.

[NGM15] NGMN. Next Generation Mobile Networks 5G White Paper. White paper, Next Generation Mobile Networks (NGMN) Alliance, 2015.

[Ngo15] H. Q. Ngo. *Massive MIMO: Fundamentals and System Designs*. PhD thesis, Linköping University, 2015.

[NRL+15] Y. H. Nam, M. S. Rahman, Y. Li, G. Xu, E. Onggosanusi, J. Zhang, and J. Y. Seol. Full Dimension MIMO for LTE-Advanced and 5G. In *Information Theory and Applications Workshop (ITA)*, pages 143–148, February 2015.

[NYB+14] H. Nikopour, E. Yi, A. Bayesteh, K. Au, M. Hawryluck, H. Baligh, and M. Jianglei. SCMA for Downlink Multiple Access of 5G Wireless Networks. In *IEEE Global Communications Conference (GLOBECOM)*, pages 3940–3945, December 2014.

[Par92] J. D. Parsons. *The Mobile Radio Propagation Channel*. John Wiley & Sons LTd, 2nd edition, 1992.

[PHS05] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst. A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication - Part I: Channel Inversion and Regularization. *IEEE Transactions on Communications*, 53(1):195–202, 2005.

[PP07] P. Pertila and M. Parviainen. Robust Speaker Localization in Meeting Room Domain. In *IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP*, volume 4, pages IV–497–IV–500, April 2007.

[PRK85] A. Paulraj, R. Roy, and T. Kailath. Estimation Of Signal Parameters Via Rotational Invariance Techniques- ESPRIT. In *Nineteeth Asilomar Conference on Circuits, Systems and Computers*, pages 83–89, November 1985.

[QWC14] D. Qiao, Y. Wu, and Y. Chen. Massive MIMO Architecture for 5G Networks: Co-Located, or Distributed? In *11th International Symposium on Wireless Communications Systems (ISWCS)*, pages 192–197, August 2014.

[RABT02] T. S. Rappaport, A. Annamalai, R. M. Buehrer, and W. H. Tranter. Wireless Communications: Past Events and a Future Perspective. *IEEE Communications Magazine*, 40(5):148–161, 2002.

[Rao92] C. R. Rao. Information and the Accuracy Attainable in the Estimation of Statistical Parameters. In Samuel Kotz and Norman L. Johnson, editors, *Breakthroughs in Statistics: Foundations and Basic Theory*, pages 235–247. Springer New York, 1992.

[RCC+07] D. J. Ryan, I. V. L. Clarkson, I. B. Collings, D. Guo, and M. L. Honig. QAM Codebooks for Low-Complexity Limited Feedback MIMO Beamforming. In *IEEE International Conference on Communications, ICC*, pages 4162–4167, June 2007.

[RF10] F. Richter and G. Fettweis. Cellular Mobile Network Densification Utilizing Micro Base Stations. In *IEEE International Conference on Communications (ICC)*, pages 1–6, May 2010.

[RHV09] V. Raghavan, M. L. Honig, and V. V. Veeravalli. Performance Analysis of RVQ-Based Limited Feedback Beamforming Codebooks. In *IEEE International Symposium on Information Theory ISIT*, pages 2437–2441, June 2009.

[RKS10] V. Raghavan, Jayesh H. Kotecha, and A. M. Sayeed. Why Does the Kronecker Model Result in Misleading Capacity Estimates? *IEEE Transactions on Information Theory*, 56(10):4843–4864, October 2010.

[RL14] X. Rao and V. K. N. Lau. Distributed Compressive CSIT Estimation and Feedback for FDD Multi-User Massive MIMO Systems. *IEEE Transactions on Signal Processing*, 62(12):3261–3271, June 2014.

[RP08] M. A. Richards and R. D. Palmer. Fundamentals of Radar Signal Processing. *Bulletin of the American Meteorological Society*, 89(7):1037, 2008.

[RPL+13] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson. Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays. *IEEE Signal Processing Magazine*, 30(1):40–60, 2013.

[RR06] J. C. Roh and B. D. Rao. Transmit Beamforming in Multiple-Antenna Systems With Finite Rate Feedback: A VQ-Based Approach. *Information Theory, IEEE Transactions on*, 52(3):1101–1112, March 2006.

[RST16] M. Rupp, S. Schwarz, and M. Taranetz. *The Vienna LTE-Advanced Simulators: Up and Downlink, Link and System Level Simulation.* Signals and Communication Technology. Springer Singapore, 1st edition, 2016.

[RV13] V. Raghavan and V. V. Veeravalli. Ensemble Properties of RVQ-Based Limited-Feedback Beamforming Codebooks. *IEEE Transactions on Information Theory*, 59(12):8224–8249, December 2013.

[Sch86] R. Schmidt. Multiple Emitter Location and Signal Parameter Estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, March 1986.

[SD10] A. Schmidt-Dannert. Positioning Technologies and Mechanisms for mobile Devices. In *Seminar Master Module SNET2 (TU-Berlin, ed.)*, 2010.

[SFGK00] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn. Fading Correlation and Its Effect on the Capacity of Multielement Antenna Systems. *IEEE Transactions on Communications*, 48(3):502–513, March 2000.

[SH03] W. Santipach and M. L. Honig. Asymptotic Performance of MIMO Wireless Channels with Limited Feedback. In *IEEE Military Communications Conference (MILCOM)*, volume 1, pages 141–146, October 2003.

[SH05] M. Sharif and B. Hassibi. On the Capacity of MIMO Broadcast Channels With Partial Side Information. *IEEE Transactions on Information Theory*, 51(2):506–522, February 2005.

[Sha98] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IEEE*, 86(2):447–457, 1998.

[SJJS00] Q. H. Spencer, B. D. Jeffs, M. A. Jensen, and A. L. Swindlehurst. Modeling the Statistical Time and Angle of Arrival Characteristics of an Indoor Multipath Channel. *Selected Areas in Communications, IEEE Journal on*, 18(3):347–360, March 2000.

[SM07] L. Sanguinetti and M. Morelli. Non-Linear Pre-Coding for Multiple-Antenna Multi-User Down-

link Transmissions with Different QoS Requirements. *IEEE Transactions on Wireless Communications*, 6(3):852–856, March 2007.

[SM11] W. Santipach and K. Mamat. Optimal Feedback Interval for Temporally-Correlated Multi-antenna Channel. In *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1–5, December 2011.

[SMD15] L. Sanguinetti, A. L. Moustakas, and M. Debbah. Interference Management in 5G Reverse TDD HetNets With Wireless Backhaul: A Large System Analysis. *IEEE Journal on Selected Areas in Communicationss*, 33(6):1187–1200, June 2015.

[SMJ13] L. Schulz, K. Manolakis, and V. Jungnickel. Adaptive Feedback Compression for Joint Transmission Coordinated Multi-Point. In *IEEE 77th Vehicular Technology Conference (VTC Spring)*, pages 1–6, June 2013.

[SN89] P. Stoica and A. Nehorai. MUSIC, Maximum Likelihood, and Cramer-Rao Bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(5):720–741, May 1989.

[SPCH16] M. S. Sim, J. Park, C. B. Chae, and R. W. Heath. Compressed Channel Feedback for Correlated Massive MIMO Systems. *Journal of Communications and Networks*, 18(1):95–104, February 2016.

[SS15] N. Singh and K. M. Singh. A Review of Direction of Arrival (DoA) Estimation for Smart Antenna Structure. *Journal of Information, Knowledge and Research in Electronics and Communication*, 03(02):1141–1146, October 2015.

[SSH04] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt. Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels. *IEEE Transactions on Signal Processing*, 52(2):461–471, February 2004.

[STS07a] M. Sadek, A. Tarighat, and A. H. Sayed. A Leakage-Based Precoding Scheme for Downlink Multi-User MIMO Channels. *IEEE Transactions on Wireless Communications*, 6(5):1711 – 1721, May 2007.

[STS07b] M. Sadek, A. Tarighat, and A. H. Sayed. Active Antenna Selection in Multiuser MIMO Communications. *IEEE Transactions on Signal Processing*, 55(4):1498 –1510, April 2007.

[STW+09] M. Schellmann, L. Thiele, T. Wirth, V. Jungnickel, and T. Haustein. Spatial mode switching in multi-user MIMO-OFDM systems: Performance in a real-world scenario. In *13th International OFDM-Workshop (InOWo)*, Hamburg, Germany, aug 2009.

[SWK11] K. Schober, R. Wichman, and T. Koivisto. MIMO Adaptive Codebook for Closely Spaced Antenna Arrays. In *European Signal Processing Conference (EUSIPCO)*, 2011.

[SZL+13] X. Su, J. Zeng, J. Li, L. Rong, L. Liu, X. Xu, and J. Wang. Limited Feedback Precoding for Massive MIMO. *International Journal of Antennas and Propagation*, 2013:1–9, 2013.

[TF09] T. E. Tuncer and B. Friedlander. *Classical and Modern Direction-of-Arrival Estimation*. Academic Press, 2009.

[Thi13] L. Thiele. *Spatial Interference Management for OFDM-Based Cellular Networks*. PhD thesis, Technical University Munich, 2013.

[TI13] A. M. Tonello and D. Inserra. Radio Positioning Based on DoA Estimation: an Implementation Perspective. In *IEEE International Conference on Communications Workshops (ICC)*, pages 27–31, June 2013.

[TJSZ07] L. Thiele, V. Jungnickel, M. Schellmann, and W. Zirwas. Capacity Scaling of Multi-User MIMO with Limited Feedback in a Multi-Cell Environment. In *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers ACSSC*, pages 93–100, Pacific Grove, CA, November 2007.

[TK14] L. Thiele and M. Kurras. Hierarchical Precoding for Ultra-Dense Heterogeneous Networks. In *Asilomar Conference on Signals, Systems and Computers*, November 2014.

[TKBH12] L. Thiele, M. Kurras, K. Börner, and T. Haustein. User-Aided Sub-Clustering for CoMP Transmission: Feedback Overhead vs. Data Rate Trade-off. In *Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, November 2012.

[TKBH13] L. Thiele, M. Kurras, K. Börner, and T. Haustein. System-Level Study on Multi-User MIMO Transmission for Dense FDD Networks. In *Asilomar Conference on Signals, Systems and Computers*, pages 1885–1889, November 2013.

[TKJ+15] L. Thiele, M. Kurras, S. Jaeckel, S. Fähse, and W. Zirwas. Interference-floor Shaping for Liquid Coverage Zones in Coordinated 5G Networks. In *49th Asilomar Conference on Signals, Systems and Computers*, pages 1102–1106, November 2015.

[TKOB13] L. Thiele, M. Kurras, M. Olbrich, and K. Börner. On Feedback Requirements for CoMP Joint Transmission in the Quasi-Static User Regime. In *IEEE 77th Vehicular Technology Conference (VTC Spring)*, pages 1–5, June 2013.

[Tre04] H. L. Van Trees. *Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 2004.

[TS12] T. Tjelta and R. Struzak. Spectrum Management Overview. *URSI Radio Science Bulletin*, 2012(340):25–28, March 2012.

[TV05] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, July 2005.

[TVZ04] D. N. C. Tse, P. Viswanath, and L. Zheng. Diversity-Multiplexing Tradeoff in Multiple-Access Channels. *IEEE Transactions on Information Theory*, 50(9):1859–1874, September 2004.

[VGL+15] I. Vin, D. P. Gaillot, P. Laly, M. Liénard, and P. Degauque. Overview of mobile localization techniques and performances of a novel fingerprinting-based method. *Comptes Rendus Physique*, 16(9):862–873, 2015.

[VH92] M. Van Herk. A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels. *Pattern Recognition Letters*, 13(7):517–521, 1992.

[VP11] D. Ventzas and N. Petrellis. Peak Searching Algorithms and Applications. In *The IASTED International Conference on Signal and Image Processing and Applications*, 2011.

[Wik19] Wikipedia. Dilution of precision (navigation). Online https://en.wikipedia.org/wiki/Dilution_of_precision_(navigation), edited on 25 July 2019, accessed 27-August-2019.

[WIN10] WINNER+. WINNER+ Final Channel Models. Technical Report D5.3, Wireless World Initiative New Radio, June 2010.

[WKT+15] L. Wang, M. Kurras, L. Thiele, T. Haustein, D. Chen, Y. Wu, and D. Qiao. Precoding Information Obtaining Method, and Device. Patent US20170317795 A1, 2015.

[WLL+14] Y.-H. Wang, H. Li, X.-L. Luo, Q.-M. Sun, and J.-N. Liu. A 3D Fingerprinting Positioning Method Based on Cellular Networks. *International Journal of Distributed Sensor Networks*, 2014:1–9, 2014.

[WSBL03] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the H.264/AVC Video Coding Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):560–576, July 2003.

[WTK+16] T. Wirth, L. Thiele, M. Kurras, M. Mehlhose, and T. Haustein. Massive MIMO Proof-of-Concept: Emulations and Hardware-Field Trials at 3.5 GHz. In *50th Asilomar Conference on Signals, Systems and Computers*, 2016.

[WTL+16] K. L. Wong, C. Y. Tsai, J. Y. Lu, D. M. Chian, and W. Y. Li. Compact Eight MIMO Antennas for 5G Smartphones and Their MIMO Capacity Verification. In *URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)*, pages 1054–1056, August 2016.

[WWH+11] B. Wang, Y. Wu, F. Han, Y. Yang, and K. J. R. Liu. Green wireless communications: A time-reversal paradigm. *IEEE Journal on Selected Areas in Communications*, 29(8):1698–1710, September 2011.

[XJW+13] Y. Xie, S. Jin, J. Wang, Y. Zhu, X. Gao, and Y. Huang. A Limited Feedback Scheme for 3D Multiuser MIMO based on Kronecker Product Codebook. In *IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1130–1135, September 2013.

[XYYS10] Wei Xi, Chang-chuan Yin, Guang-xin Yue, and Hong-mei Sun. Partial channel information feedback mechanism for FFR-based multi-channel systems. *The Journal of China Universities of Posts and Telecommunications*, 17(1):19–43, 2010.

[YCG14] H. Yin, L. Cottatellucci, and D. Gesbert. Enabling massive MIMO systems in the FDD mode thanks to D2D communications. In *48th Asilomar Conference on Signals, Systems, and Computers*, November 2014.

[YCQL18] F. Yang, P. Cai, H. Qian, and X. Luo. Pilot Contamination in Massive MIMO Induced by Timing and Frequency Errors. *IEEE Transactions on Wireless Communications*, 17(7):4477–4492, July 2018.

[YG06] T. Yoo and A. Goldsmith. On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming. *IEEE Journal on Selected Areas in Communications*, 24(3):528 – 541, March 2006.

[YGC14] H. Yin, D. Gesbert, and L. Cottatellucci. Dealing With Interference in Distributed Large-Scale MIMO Systems: A Statistical Approach. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):942–953, October 2014.

[YS15] C. Yang and H.-R. Shao. WiFi-Based Indoor Positioning. *IEEE Communications Magazine*, 53(3):150–157, March 2015.

[YVT+14] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh. Kronecker Product Correlation Model and Limited Feedback Codebook Design in a 3D Channel Model. In *IEEE International Conference on Communications (ICC)*, pages 5865–5870, June 2014.

[YWZP13] Y. Yuan, Y. Wang, W. Zhang, and F. Peng. Separate Horizontal and Vertical Codebook Based 3D MIMO Beamforming Scheme in LTE-A Networks. In *IEEE 78th Vehicular Technology Conference (VTC Fall)*, pages 1–5, September 2013.

[YXXW09] C. Yu, W. Xiangming, L. Xinqi, and Z. Wei. Research on the Modulation and Coding Scheme in LTE TDD Wireless Network. In *International Conference on Industrial Mechatronics and Automation, ICIMA*, pages 468 –471, May 2009.

[YYU01] A. Yener, R. D. Yates, and S. Ulukus. Interference Management for CDMA Systems Through Power Control, Multiuser Detection, and Beamforming. *IEEE Transactions on Communications*, 49(7):1227–1239, July 2001.

[Zhu92] J. Zhu. Calculation of Geometric Dilution of Precision. *IEEE Transactions on Aerospace and Electronic Systems*, 28(3):893–895, July 1992.

[Zir15] W. Zirwas. Opportunistic CoMP for 5G massive MIMO Multilayer Networks. In *Proceedings of 19th International ITG Workshop on Smart Antennas WSA*, pages 1–7, March 2015.

[ZLD+14] Y. Zhou, L. Liu, H. Du, L. Tian, X. Wang, and J. Shi. An Overview on Intercell Interference Management in Mobile Cellular Networks: from 2G to 5G. In *IEEE International Conference on Communication Systems (ICC)*, pages 217–221, November 2014.

[ZM11] W. Zirwas and W. Mennerich. Main Enablers for Advanced Interference Mitigation. In *8th International Workshop on Multi-Carrier Systems Solutions (MC-SS)*, pages 1–5, May 2011.

[ZT03] L. Zheng and D. N. C. Tse. Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels. *IEEE Transactions on Information Theory*, 49(5):1073–1096, May 2003.

[ZZL09] Y. Zhao, H. Zhou, and M. Li. A Novel Overlap Area Matching Algorithm Based on Location Fingerprinting in Wireless Networks. In *5th International Conference on Wireless Communications, Networking and Mobile Computing*, pages 1–5, September 2009.