TECHNISCHE UNIVERSITÄT DRESDEN

Adaptive and Robust Beam Selection in Millimeter-Wave Massive MIMO Systems

M.Sc.

Mostafa Khalili Marandi

der Fakultät Elektrotechnik und Informationstechnik
der Technischen Universität Dresden

zur Erlangung des akademischen Grades

Doktoringenieurs

(Dr.-Ing.)

genehmigte Dissertation

Vorsitzender: Prof. Dr.-Ing. Frank Ellinger
Gutachter: Prof. Dr.-Ing. Dr. h. c. Gerhard Fettweis
Gutachter: Prof. Dr.-Ing. habil. Volker Kühn
Gutachter: Prof. Dr.-Ing. Dirk Plettemeier

Tag der Einreichung: 12.09.2022
Tag der Verteidigung: 20.01.2023

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Professor Gerhard Fettweis, for his guidance, empathy, support, and encouragement throughout my PhD journey. He has been a great role model and because of him, I had priceless opportunities to learn, grow and collaborate with amazing researchers in important 5G and 6G projects.

I am also deeply thankful to my research group leader, Dr. Wolfgang Rave, who has been a great mentor and a kind friend during this challenging period. He has shared his expertise, experience, and profound wisdom with me and helped me overcome many difficulties and doubts along the way. The conduction of this work would have not be imaginable without him. I appreciate his valuable feedback, constructive criticism and insightful suggestions that helped me improve the quality of my research and my personality as a researcher.

Next, I would like to express my sincere gratitude to my colleagues and students at the Vodafone Chair at TU Dresden for their invaluable support and collaboration during my PhD studies. They have contributed to the quality and impact of my research work, and have enriched my academic experience with their insights and feedback. I am fortunate to have worked with such a talented, diverse and kind group of people, and I hope to maintain our professional and personal relationships in the future.

Last but not least, I would also like to acknowledge the love and support of my family and my partner, who have always been there for me in good and bad times. They have given me the motivation, confidence, and strength to pursue my dreams and goals. They have also sacrificed a lot for me and made many compromises to make this possible. I dedicate this thesis to them.

Abstract

Future 6G wireless communications network will increase the data capacity to unprecedented numbers and thus empower the deployment of new real-time applications. Millimeter-Wave (mmWave) band and Massive MIMO are considered as two of the main pillars of 6G to handle the gigantic influx in data traffic and number of mobile users and IoT devices. The small wavelengths at these frequencies mean that more antenna elements can be placed in the same area. Thereby, high spatial processing gains are achievable that can theoretically compensate for the higher isotropic path loss. The propagation characteristics at mmWave band, create sparse channels in typical scenarios, where only few paths convey significant power. Considering this feature, Hybrid (analog-digital) Beamforming introduces a new signal processing framework which enables energy and cost-efficient implementation of massive MIMO with innovative smart arrays. In this setup, the analog beamalignment via beam selection in link access phase, is the critical performance limiting step. Considering the variable operating condition in mmWave channels, a desirable solution should have the following features: efficiency in training (limited coherence time, delay constraints), adaptivity to channel conditions (large SNR range) and robustness to realized channels (LOS, NLOS, Multipath, non-ideal beam patterns). For the link access task, we present a new energy-detection framework based on variable length channel measurements with (orthogonal) beam codebooks. The proposed beam selection technique denoted as composite *M-ary Sequential Competition Test* (SCT) solves the beam selection problem when knowledge about the SNR operating point is *not* available. It adaptively changes the test length when the SNR varies to achieve an essentially constant performance level. In addition, it is robust to non-ideal beam patterns and different types of the realized channel. Compared to the conventional fixed length energy-detection techniques, the SCT can increase the training efficiency up to two times while reducing the delay if the channel condition is good. Having the flexibility to allocate resources for channel measurements through different beams adaptively in time, we improve the SCT to eliminate unpromising beams from the remaining candidate set as soon as possible. In this way, the *Sequential Competition and Elimination Test* (SCET) significantly further reduces training time by increasing the efficiency. The developed ideas can be applied with different codebook types considered for practical applications. The reliable performance of the beam selection technique is evident through experimental evaluation done using the state-of-the-art test-bed developed at the Vodafone Chair that combines a Universal Software Radio Peripheral (USRP) based platform with mmWave frontends.

Publicatios

Parts of this thesis have been published in the following peer reviewed articles.

Journal Publications:

- M. Khalili Marandi, W. Rave, G. Fettweis, **“Beam Selection Based on Sequential Competition”**, IEEE Signal Processing Letters 26 (3), 455-459, March 2019.
- H. Chiang, K. Chen, W. Rave, M. Khalili Marandi, G. Fettweis, **“Machine-Learning Beam Tracking and Weight Optimization for mmWave Multi-UAV Links"**, 2021 IEEE Transactions on Wireless Communications.

Conference Publications:

- M. Khalili Marandi, W. Rave, G. Fettweis, **“Beam Elimination Based on Sequentially Estimated a Posteriori Probabilities of Winning"**, 2020 International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain.
- M. Khalili Marandi, C. Jans, W. Rave, G. Fettweis, **“Evaluation of Detection Accuracy and Efficiency of Considered Beam Alignment Strategies for mmWave Massive MIMO Systems"**, 2021 Asilomar Conference on Signals, Systems and Computers, USA.
- M. Khalili Marandi, S. Wei, B. Khodapanah, W. Rave, G. Fettweis, **“Experimental Evaluation of a Variable Length Beam Selection Framework in a USRP based Testbed with mmWave Frontends and Butler Matrices"**, 2022 IEEE International Symposium on Phased Array Systems and Technology, Boston, USA.
- M. Khalili Marandi, W. Rave, G. Fettweis, **“An Adaptive Sequential Competition Test for Beam Selection in Massive MIMO Systems”**, 2018 IEEE SAM, Sheffield, United Kingdom.
- M. Khalili Marandi, W. Rave, G. Fettweis, **“Multi User Beam Selection Using Sequential Competition Test"**, 2019 IEEE WSA, Vienna, Austria.
- W. Rave, M. Khalili Marandi **“The Elimination Game or: Beam Selection Based on M-ary Sequential Competition & Elimination"**, 2019 IEEE WSA, Vienna, Austria.

- X. Wang, M. Laabs, D. Plettemeier, H. Chiang, M. Khalili Marandi, G. Fettweis, K. Kosaka, Y. Matsunaga, **"28 GHz Multi-Beam Antenna Array Based On A Compact Wideband** 88 **Butler Matrix"**, 2018 IEEE International Symposium on Antennas and Propagation.

- H. Chiang, K. Chen, W. Rave, M. Khalili Marandi, G. Fettweis, **"Multi-UAV mmWave Beam Tracking using Q-Learning and Interference Mitigation"**, 2020 IEEE ICC Workshops.

- K. Dehkordi, O. Schwanitz, M. Khalili Marandi, T. H. Le, M. P. Kaiser, I. Ndip, and G. Caire,**"Sub-THz Layout Optimization under Fabrication Constraints,"** 2022 EUCNC & 6G Summit, Grenoble, France

Contents

List of Figures

List of Acronyms

1/2/3/4/5/6G First/Second/Third/Fourth/Fifth/Sixth Generation

ADC Analog-to-Digital Converter

ANN Artificial Neural Network

AoA Angle of Arrival

AoD Angle of Departure

AR Augmented reality

BA Beam Alignment

BFN Beamforming Network

BM Butler Matrix

BS Base Station

CIR Channel Impulse Response

CSI Channel State Information

DAC Digital-to-Analog Converter

DFT Discrete-Fourier-Transform

EMB Enhanced Mobile Broadband

FC Fully Connected

FDB Frequency Dependent Beamforming

FSPL Free-Space Path Loss

GF Group Factor

GLLR Generalize Log Likelihood Ratio

GLR Generalize Likelihood Ratio

GLRT Generalized Likelihood Ratio Test

HBF Hybrid Beamforming

IA Initial Access

IAB Industry Advisory Board

IoT Internet of Things

LNA Low Noise Amplifier

LO Local Oscillator

LOS Line of Sight

ML Maximum Likelihood

MIMO Multiple Input Multiple Output

MLVL Multi-level

MM Magnitude Maximization

MMSE Minimum Mean Squared Error

mMTC Massive Machine to Machine Type Communications

MU Multi User

MVUE Minimum Variance Unbiased Estimator

mmWave millimeter-wave

NLOS Non Line of Sight

NP Neyman-Pearson

NR New Radio

OFDM Orthogonal Frequency-Division Multiplexing

OMP Orthogonal Matching Pursuit

PA Power Amplifier

PC Partially Connected

PDF Probability Density Function

PL Path Losses

PN Pseudo-random Noise

QPSK Quadrature Phased Shift Keying

RACH Random Access Channel

RAM Radiation Absorbent Material

RF Radio Frequency

SaPP Sequentially Estimated a-Posteriori Probability

SCET Sequential competition and Elimination Test

SCT Sequential competition Test

SINR Signal to Interference plus Noise Ratio

SISO Single Input Single Output

SLMMSE Sequential Linear MMSE

SM SINR Maximization

SNR Signal to Noise Ratio

SPRT Sequential Probability Ratio Test

SVD Singular Value Decomposition

UE User Equipment

ULA Uniform Linear Array

UMP Uniformly Most Powerful

URLLC Ultra-reliable and Low Latency Communications

USRP Universal Software Radio Peripheral

VGA Variable Gain Amplifier

VR Virtual reality

WGN White Gaussian Noise

List of Symbols

M_{TX} Number of antennas at the transmitter

M_{RX} Number of antennas at the receiver

M Codebook size

N_{fix} Fixed number of measurements

N_{tot} Total number of measurements

n Variable number of measurements

$\bar{n}$ Average number of measurements

$\gamma_m[n]$ GLLR for beam m after n observations

$s[n]$ Pseudo-random training sequence

P Number of resolvable plane waves

τ_p Delay of path p

ϕ_p AoA of path p

ρ_p Gain of path p

$\mathbf{w}_m$ Steering vector of beam m

ϕ_m steering direction of beam m

λ_0 Wavelength at the carrier frequency

m Beam index

$\mathbb{E}[.]$ Expected value

$\mathcal{CN}$ Complex Gaussian Distribution

σ^2 Noise variance

$z[n]$ Noise sample

A_{m,τ_p} The combined effective channel and beamforming gain corresponding to beam m and path p

α_p Overall multipath component power

η_p Power ratio between the deterministic LOS and the random scattered NLOS components

ϵ_p Zero-mean unit-variance complex Gaussian random variable

PL Path loss

β Reflection loss

d_p Propagation distance of path p

$y_{m,\tau}[n]$ Correlated observation sequence for each beam m and delay τ

$\mathcal{H}_{m,\tau}$ Hypothesis (m, τ)

$h(t, \phi)$ Channel impulse response

$\bar{l}$ Averaged normalized loss of signal magnitude

d^2 Deflection coefficient

$\bar{a}$ Average captured relative magnitude

$Q(.)$ Right tail probability of a standard Gaussian random variable

$L_{\mathrm{G}}(.)$ Generalized likelihood ratio

$|.|$ Absolute value operation

$\chi_1^2(\lambda_{m,\tau})$ Chi squared distribution with non-centrality parameter $\lambda_{m,\tau}$

P^{FA} Probability of false alarm

P^{MD} Probability of miss detection

γ_{term} Termination threshold

$\mathbf{J}$ Active set of candidate beams

$\hat{\sigma}_A^2[n]$ Variance of the estimated a-posteriori PDF

$K[n]$ Kalman gain

$P_{m,\tau}^{\mathrm{win}}[n]$ Probability of winning the competition

d_{elim}^2 Elimination threshold

$Q_1^{\mathrm{c}}(.)$ Complementary first-order Marcum Q-function

$\beta_{u,m}$ SINR for user u under beam m

$\mathbf{F}_B$ Digital beamforming matrix

$\mathbf{H}_E$ Effective channel matrix

Chapter 1

Introduction

1.1 5G New Radio: Road to 6G

The advent of wireless communications in the 1990s has had an everlasting effect on our everyday life and productivity. It started with voice telephony in the first and second generation and paved its way to data access, applications, and services in third and fourth generations. The demand for data communications is increasing over time, while the transmission networks remain to be the bottleneck. Therefore, it is essential to provide as much capacity as possible, while building an efficient and smart architecture that can accommodate future demands for data communications and groundbreaking innovations like tactile internet [FA14b, RXK+19b].

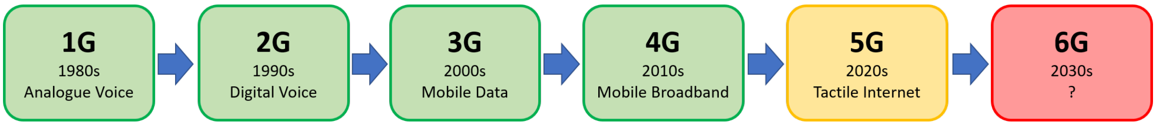


Figure 1.1: Evolution of wireless communications standards over the past decades.

5G New Radio wireless networks start a new era in mobile communications by offering important connectivity advantages such as energy efficiency, higher system capacity, reduced latency and certainly higher data rates. Such high data rates and wide channels are only possible by employing the millimeter-wave (mmWave) carriers [XMH+17], which already overlap with what is considered as the sub-terahertz. However, the increasing number of new applications such as Virtual/Augmented reality (VR/AR), autonomous driving, Internet of Things (IoT), wireless backhaul [BAM+18], smart city/home, industry automation, mission-critical applications like e-health, joint communication and sensing etc. will demand data rates on the order of 100 Gbps and above which can only be achieved at frequencies above 100 GHz, where the available spectrum is abundant [Sie02].

Development of the next generation of transceivers is motivated by future requirements of the 6G network. These are based on desired applications and possible use cases, which

	EMB	mMTC	URLLC	Localization
ERICSSON	●	●	●	
TE	●		●	
AIRBUS	●			●
JOHN DEERE	●	●		●

Figure 1.2: Survey on interesting use cases considered for 6G by the industry advisory board members of the FormiKro-6GKOM project in 2019 at TU Dresden.

can be sorted into the following categories:

- Enhanced Mobile Broadband (EMB)
- Massive Machine to Machine Type Communications (mMTC)
- Ultra-reliable and Low Latency Communications (URLLC)
- Localization and Sensing

Through the joint BMBF project 6GKOM, the next generation of transceivers are being researched at the Vodafone chair for Mobile Communications Systems at TU Dresden in collaboration with TU Berlin, Fraunhofer Institute IZM, Leibnitz Institute iHP, University of Ulm and an Industry Advisory Board (IAB) including major companies from different sectors. We started this project with a workshop at TU Dresden in 2019, where IAB-members presented their perspective of the next generation of wireless communications systems. Many use cases and the corresponding requirements in different scenarios were identified. A summary of this survey including companies like Airbus, TE connectivity, Ericsson and John Deere which are from different industry-segments is presented in Fig. 1.2.

As envisioned and also approved through our investigations, Enhanced Mobile Broadband (EMB) with data rates of 100 Gbs and above is a key desire in the majority of industrial sectors. Therefore, 6G will employ carrier frequencies from higher mmWave band to alleviate the spectrum gridlock while accessing a much larger available bandwidth. This appears vital for thriving of the next generation of wireless communications systems with a wide range of innovative applications.

1.2 Why is Efficient Beam Selection Necessary?

The evolution of mobile communications standards towards higher frequencies creates a need for efficient and robust beamforming as a consequence. While in earlier systems, base stations were sending out their signals with one or two antenna elements in a roughly isotropic manner, trying to provide coverage almost everywhere, at higher frequencies smaller and more antenna elements need to be used with more and more pronounced anisotropy of the antenna patterns. This, inherently, allows reducing the interference on one hand but requires support of controlled energy distribution on the other.

Due to the severe isotropic pathloss and large scattering attenuation incurred at mmWave [V14] frequencies as a consequence of the reduced size of a single antenna element, the application of antenna arrays has become mandatory. This follows from the desire to conserve at least a certain fraction of the received power levels when moving to higher carrier frequencies, so that also a reasonable communication range is conserved. In turn, this means that the area efficiency of the antennas involved has to be maintained at least approximately by using antenna arrays with numerous antennas at one or even both sides of the link.

Under these assumptions, the required antenna gain is obtainable via beamforming. However, to achieve it in a practical multi-antenna system, situation aware beam alignment is still needed, e.g., via digitally steerable beamforming networks (BFN) [But61] [BVBL19] [WLP+18]. Developing fast and accurate BFN at mmWave, however, is not a trivial task. The fully digital transceiver architecture, where one radio frequency (RF) chain is connected to each antenna element, is impractical. Challenges include hardware cost, the power consumption and aso the power dissipation in the integrated arrays. Each RF-chain consists of analog-to-digital/digital-to-analog converters (ADC/DAC), up/down-conversion mixers, filters and power/low-noise amplifiers (PAs/LNAs). The goal for the design of the mmWave transceivers is to reduce the number of RF-chains to a level that is significantly smaller than the number of antenna array elements, without greatly sacrificing the multi-user sum rate. For this reason, the concatenation of digital and analog beamforming, known as *hybrid* beamforming (HBF) architecture, has been widely considered [AELH14, BHR+15, KCW+17, ZHS+17, HM19, SHC18, SHC19a, SKC20].

While for both tasks the same word beamforming is used, only in the case of analog beamforming it has a geometrical meaning in the sense of transmitting or receiving signals using antenna arrays towards specific directions in 3D space. In contrast, digital beamforming operates in a more abstract vector space of arbitrary dimension given by the number of RF-chains at the transmitter and the receiver side, with the goal to optimally combine the analog beams, given some cost criterion.

Therefore, the two steps of hybrid beamforming can be viewed as follows: First, estimate the energy transfer between narrow analog beams to project a Multiple Input Multiple Output (MIMO) channel [Kuh06] of high dimensions given by the product of the numbers of antennas at each side of the link into a subspace of much lower dimensions, so that an 'effective' channel results based on the selected analog beamforming vectors [CRKF18], [KCW+17]. Only for this effective channel, the exact parameters (a matrix-valued impulse response or channel transfer function) need to be estimated to determine the optimal weights for digital beamforming in the second step.

While channel estimation between all antenna elements is infeasible in practice, several earlier works on HBF [SY16, LM17, IE17, DXS+18, COS18], assuming full Channel State Information (CSI), established the important result that the rate loss between fully digital and suboptimal hybrid beamforming is quite moderate in typical scenarios (channels), justifying the application of the latter approach. This is explained by the fact that mmWave channels are sparse in both the angular and delay domains, considering the typical application scenarios. This can be understood intuitively by arguing that only a few paths transports significant levels of reflected or scattered energy from the transmitter to the receiver and need to be addressed by selected beams (sparsifying the channel even further). As a consequence, the number of relevant paths is strongly or even drastically less than the number of antennas [ALS+14], [SR16a], so that after decoupling the analog and digital beamforming problems, performance remains close to the optimum one. The main benefit is that for the lower-dimensional effective channel, its estimation in terms of channel impulse response or transfer function becomes tractable. This means that standard techniques, possibly under different power constraints, for solving the digital problem can be applied, that were developed in the literature for spatial multiplexing (see [TV05, Gol05] and references cited therein) after its initial proposal in [CRT01].

The general Beam Alignment (BA) problem is defined as a selection problem of one or several analog beams from an available discrete set of candidates, often denoted as a 'codebook'. This is the critical performance limiting step of HBF, since the beam pairs with the largest energy coupling result in the best effective channel, which determines the achievable throughput during the data transmission phase [CRKF18]. As described in [CRKF17], a straightforward approach to select suitable analog beams at transmitter and receiver sides is to conduct an exhaustive search, employing estimates of the received power level or energy transfer for each beam pair by correlating the received signal to a pilot sequence. In this approach, the total training cost in terms of required number of measurement symbols while testing the beam pairs via exhaustive search, is given as $N_{\mathrm{tot}} = M_{\mathrm{TX}} \times M_{\mathrm{RX}} \times N_{\mathrm{fix}}$, where M_{TX} and M_{RX} indicate the sizes of the codebooks at TX and RX, while N_{fix} is standing for the length of the training sequence measured for each beam pair. A two-stage pseudo-exhaustive BA technique has been proposed in [PDW17], where in the first stage, the Base Station (BS) omnidirectionally probes the channel, while the User Equipment (UE) sweeps through its codebook to find the beam(s) capturing the largest amount of power. In the second stage, the UE probes the channel along the selected beam in the first stage, while the BS performs beam sweeping to find the best beam in its codebook. This way the cost associated to training becomes $N_{\text{tot}} = M_{\text{RX}} \times N_{\text{fix}}^{\text{1st}} + M_{\text{TX}} \times N_{\text{fix}}^{\text{2nd}}$, where now the quadratic dependence on the codebook sizes is removed, while training sequence length at the first stage is extended.

The problem can also be solved by the orthogonal Matching Pursuit (OMP) [CW11] as well as the Non-Negative Least Squares (NNLS) approach [SHC19b], provided that the analog beamforming vectors are selected from an orthogonal set of beams with corresponding mainlobe directions[1]. These approaches aim to reduce the number of required tests (i.e., $M_{\text{TX}} \times M_{\text{RX}}$) by the exhaustive search during beam alignment phase by employing pseudo-random spreading codes to test an ensemble of beam pairs via each RF chain at each step. However, the price to be paid is that the training sequence length or transmit power should be increased to maintain the effective detection SNR. On the other hand, the simple method of phenomenologically estimating receive power levels does not depend on the correct or even exact knowledge of the array manifold, while OMP like techniques degrade under mismatch between the assumed and actual codebook, as well as in channels with multiple strong paths which hurt the strong sparsity required in these techniques. This makes the strategy using the former idea of estimating the energy transfer more robust.

Further important features of our problem are the following. First, its size scales linearly with the number of antennas (to which the 3 dB mainlobe beam width is inversely related) and/or the codebook size. For orthogonal codebooks that minimize the codebook size by obeying the sampling theorem with the maximum spatial sampling interval and in the regime of narrow pencil beams (large codebook size) it holds that in the majority of channel realizations, any path falls only in the main lobe of a single beam and its energy will be missed, if it is not selected. Therefore, the crucial initial step in HBF during link establishment requires a reliable and time efficient estimate of the energy transfer among a potentially large number of beam pairs with members belonging to the codebooks given at Rx and Tx.

In addition, since the knowledge of exact Probability Density Functions (PDFs) that relates the observations to the parameters to be estimated is not available in practice, but defining parameters of the assumed PDFs such as SNR or noise variance have to be estimated themselves, this amounts to a composite multi-hypothesis detection problem (for a classification of detection problems, see [Kaya]).

[1] Intuitively, such codebooks act as complete dictionaries in terms of compressed sensing techniques [VAGH17]. In case of an over-complete dictionary, the approach corresponds to matching pursuit.

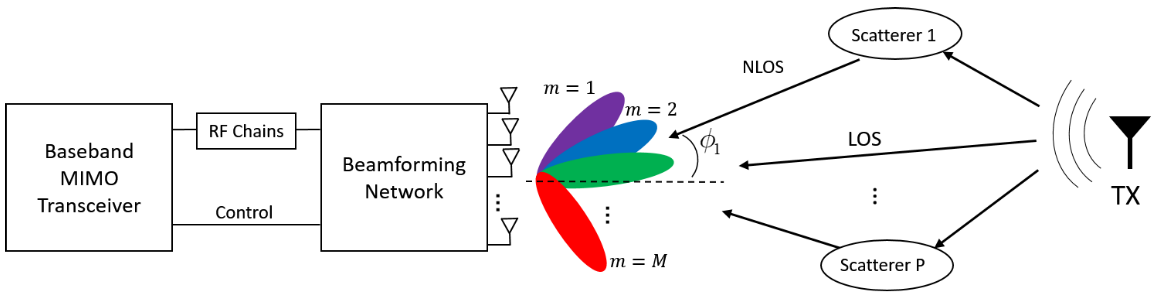


Figure 1.3: Beam selection problem: MIMO transceiver using an electronically controllable beamforming network in a multi-scatterer environment.

1.3 Energy-Detection Frameworks for Beam Selection

Due to the symmetry of the problem, we focus in this work on the one-sided beam selection problem during the Initial Access (IA) procedure [BHM+15] [BHM+16] a.k.a. beam alignment problem, which is shown schematically in Fig 1.3. In [BHR+15] the authors showed that the analog beam selection with *fixed* length probing sequences is simply performed by non-coherently adding the energy from the matched filter outputs and finding the candidate beam with maximum power. State of the art beam-alignment (i.e., two-sided beam selection) techniques based on energy-detection e.g., exhaustive [ALS+14], pseudo-exhaustive [PDW17] or tree search with hierarchical codebooks [NZL17a, KCW+17] as proposed in IEEE 802.11ad, use *fixed* length probing sequences for evaluation of the best beam pairs[2]. However, a *fixed* length probing sequence can only be optimally designed for one particular Signal to Noise Ratio (SNR) which is unknown before beam selection and variable in most practical scenarios. Naively fixing the test length based on a certain assumed operating point can result in a strongly variable performance in practical scenarios. Additionally, if the test length is conservatively set to a high value based on the worst still acceptable operating point, a lot of time and resources spent for detection of the best beam will be wasted, if the channel quality is actually better than expected. This is particularly important when the channel coherence-time is limited so that wasting time for training results in a relatively large loss of throughput and excess delay. Therefore, adaptivity to the channel conditions during the beam selection phase, potentially results in rate gain and delay reduction.

The idea of *variable* length detection was first introduced by A. Wald in his seminal work on the Sequential Probability Ratio Test (SPRT) [Wal44] [Wal45] for the binary hypothesis problem, where the underlying PDFs are assumed fully known. The main goal of the SPRT is to minimize the average sample number while achieving a decision-error probability (e.g., denoted as 'false alarm' or 'missed detection') specified in advance. In

[2] The essential idea presented in this work can be simply extended for the scenario where both user and base station employ a codebook.

case of i.i.d observations, Wald proved the optimality of the SPRT in terms of average decision time while fulfilling the predefined decision reliabilities. There have been extensions to SPRT towards multi-hypothesis tests (MSPRT) [BV94], as well as to composite hypothesis problems (GSPRT) [BT18] for which it was shown that they are asymptotically optimal [DTV99].

Beam selection is a composite multi-hypothesis problem, where each beam stands as a candidate. However, it is distinct from the problems considered under the variants of SPRT due to two main reasons. First, in contrast to classical multi-hypothesis decision problems where only a *single* observation sequence is considered, for beam selection separate *multiple* observation sequences are available corresponding to different candidate beams. Secondly, the price to be paid for not selecting the strongest signal depends on the specific hypothesis that is chosen (e.g., whether a smaller or a larger fraction of the maximum signal level is missed). Therefore, a meaningful cost function for beam selection aims to maximize the average selected signal magnitude, which is a different criterion as in classical SPRT variants trying to achieve specified decision reliabilities. It is related, but does not directly translate into predefined decision accuracies (see Eq. (3.10)). This means that for beam selection, the cost of erroneous decisions depends on the ratio of the selected magnitude over the magnitude of the best candidate, which are both unknown a-priori.

1.4 Contributions

In this work, we investigate the mmWave Massive MIMO systems as a key enabler technology for next generation of wireless systems. In **chapter 2** we introduce two novel practical hardware modules researched and developed at TU Dresden for mmWave massive MIMO systems enabling efficient signal processing based on two different hybrid beamforming structures. These innovative pieces of hardware provide reliable and accurate beamforming capabilities for a variety of possible applications.

In **chapter 3** we formulate the beam selection problem in mmWave systems as a composite multi-hypothesis test while introducing meaningful performance measures. Next, we propose our idea for a *variable*-length detection framework based on the statistics of the Generalized Likelihood Ratio (GLR) and denoted it as *Sequential Competition Test* (SCT) to solve the composite multi-hypothesis beam selection problem [MRF19]. SCT learns on the fly the current statistics of the signals in terms of their amplitudes and noise variances based on the available observations from each beam. The test length adaptively changes with respect to the unknown SNR operating point, while achieving some specific constant level of accuracy. Furthermore, it even speeds up beam selection on average compared to the optimally tuned Fixed-length test (i.e., having genie knowledge about the SNR). This occurs at lower SNR values where it matters the most or most time will be spent for training in a possible application. In contrast to beam selection techniques based on compressed sensing, the SCT is robust to the inaccuracies that are present in the implemented beam patterns when compared to the reference orthogonal beam patterns, as well as not being sensitive to the level of the sparsity in the channel. This variable-length detection framework enables the idea of beam elimination [RM19] to increase the efficiency of SCT by shutting down the unpromising beams during the competition. Our initial idea employs the distance between the GLLR of the leading competitor and all other candidate beams to eliminate the beams that are far behind in terms of their GLLR metric (see the schematic showing the evolution of the metric values $\gamma_m[n]$ towards a termination threshold of four competing beams in Fig. 3.8 with elimination due to a certain metric difference w.r.t. the best current metric). In this algorithm, the elimination criterion was set to a fixed portion of the metric chosen for the termination threshold. The elimination mechanism is then refined based on *a-posteriori* PDFs of the estimated unknown effective channel coupling coefficients (i.e., the estimated signal amplitudes under each beam) [MRF20]. These were acquired via a Sequential Linear MMSE (SLMMSE) estimator. This algorithm evaluates the Bayesian winning probability for each beam at each time step during the competition. The elimination of any beam occurs as soon as its corresponding winning probability falls below a small value, that acts as an optimization or tuning parameter in the algorithm[3]. We denoted the resulting variable-length beam selection technique as Sequential Competition and Elimination Test (SCET).

Chapter 4 discusses the main practical codebook types which are typically used through the searching phase prior to data communication. Furthermore, the effect of the variable vs Fixed-length detection while using different codebook types is evaluated, where a fair comparison is intended throughout this study by applying the same power constraint on all scenarios. Our numerical results [KMJRG21] indicate that the variable-length test used either with the frequency dependent codebook or a Butler codebook are the most promising methods achieving better efficiency, adaptivity, and robustness under varying SNR in practical scenarios.

In **chapter 5** the multi-user beam selection problem in mmWave massive MIMO systems is discussed. We introduce the extension to multi-user sequential competition test [KMRF19] as the variable-length detection technique for this problem. Finally, we discuss the practical solution to the joint analog digital beamforming optimization problem in the multi-user setup to improve the sum rate performance of the MU-MIMO system

[3] In case that the receiver can employ more than one RF-chain during the data communication, it can use the competition ranking after termination to select the best beams and design the optimal digital precoding/combining.

with minimum complexity.

An experimental evaluation of the key ideas presented in this work is done in **chapter 6**. The experimental setup employs a hardware in the loop including the developed Butler Matrix at TU Dresden which provides a codebook of orthogonal beams connected through an electronic switch to a single RF-chain. We investigate multiple measurement scenarios in which beams with different main-lobe directions are automatically probed at the TX and RX through different channel realizations. The acquired measurements are processed in matlab to benchmark the beam selection accuracy and efficiency. The results indicate the resilience of the sequential competition and elimination test in the presence of hardware impairments and non-ideal synchronization.

Finally, in **chapter 7** the main achievements in this work are summarized while providing an outlook on the possible future works.

Chapter 2

mmWave Massive MIMO: A Key Enabler for 6G

mmWave band and Massive MIMO are considered as two of the main pillars of 6G to handle the gigantic influx in data traffic and number of mobile users and IoT devices. The available bandwidth at mmWave frequencies is abundant, while the spectrum gridlock is non-existent. The small wavelengths at these frequencies mean that more antenna elements can be placed in the same area. Thereby, high spatial processing gains are achievable that can theoretically compensate for the higher isotropic path loss. Nevertheless, as these systems are equipped with massive number of antennas, to maintain the anticipated performance gain, several challenges arise including the computation complexity and the hardware energy-efficiency. Toward this end, this chapter discusses the key enabling signal processing framework called hybrid beamforming and its implication in the hardware structure which enables the implementation of massive MIMO with innovative smart arrays. A limited number of RF-chains is used to enable multi-stream processing in baseband, while analog processing in the RF frontend is used to realize the beamforming gain. A primary objective of hybrid beamforming is to maximize the multiuser sum rate, while keeping hardware costs, complexity, and power efficiency, within some desirable targets. In this direction, we present two novel mmWave massive MIMO hardware modules developed at TU-Dresden together with different industry partners to address these challenges.The main ideas in this chapter are partially presented in [WLP+18, DSM+22]

Figure 2.1: mmWave Massive MIMO for mobile communications.

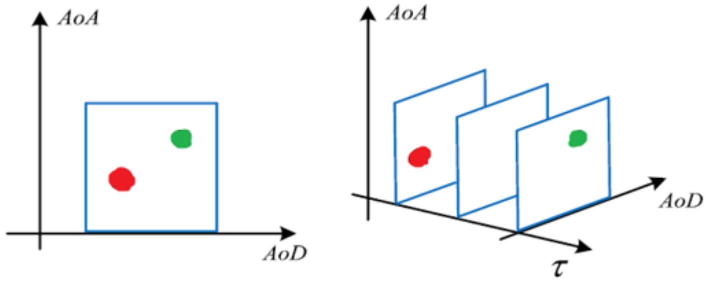


Figure 2.2: Illustration of the sparse mmWave channel in angular and delay domain [SHC19b].

2.1 mmWave-THz Channel Characteristics

mmWave-THz channels are distinguished in terms of the propagation characteristics compared to microwave frequency bands in terms of path loss, diffraction and blockage, rain attenuation, atmospheric absorption, and foliage loss behaviors. In general, the overall loss of mmWave-THz channels is significantly larger than that of microwave channels for a point-to-point link [RMSS15,XR21]. The upper end of the mmWave band a.k.a. the THz band, spanning from 0.1 to 10 THz, offers bandwidth orders huge chunks of Bandwidth. Specifically, the usable bandwidth of the mmWave band in 30 to 100 GHz is usually up to 7 GHz, while that of the terahertz band is at least 10 to 50 GHz [LL16]. Several studies suggest that the channel in the angular and delay domains is sparse at these higher carrier frequencies. This means only a very few paths compared to the number of antennas actually convey a significant amount of power [HWF+17, WHR+11, JWNC19].

Frequency-selective fading due to multipath in mmWave-THz transmission depends on: delay spread vs symbol period, number of dominant paths and finally the spatial filtering, i.e., 3dB beam width vs. angular spread and misalignment. The most important take away is that, in case of a sparse channel and good alignment with increasingly narrower beams, we obtain a flat effective channel with unknown delay parameter, carrier frequency offset, and noise variance [AKM20]. This motivates the use of single carrier modulation schemes compared to Orthogonal Frequency-Division Multiplexing (OFDM) [BDF+18], specially during beam alignment phase where OFDM might suffer from pre-beamforming low SNR and non-ideal synchronization. Moreover, wideband transmission

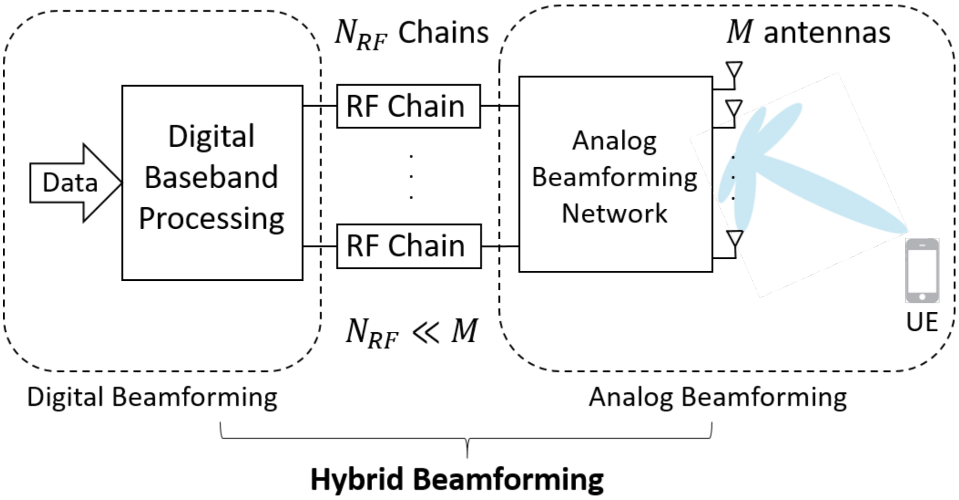


Figure 2.3: Illustration of hybrid beamforming signal processing structure, introducing the concept of analog beamforming networks.

at THz is prone to timing error and inter-symbol interference due to dispersion caused by: frequency dependent propagation, reflection, refraction, and diffraction [HBA15]. To combat these, proper time and frequency domain equalization might be necessary at the receiver [SAAN21].

2.2 Practical Hybrid Beamforming Structures

As explained, mmWave channels are generally sparse and large antenna arrays plus beamforming techniques are required to increase the link budget. Use of high-resolution individual RF-chain per each antenna element is not practical due to large energy consumption by ADCs/DACs and the hardware cost. On the other hand, full channel estimation between all antenna elements is computationally intractable due to the considerable size of the channel Matrix.

However, the sparsity of the channel can be exploited for optimizing channel estimation and beam training. This means the problem can be solved in two consecutive steps, namely the analog and digital beamforming (i.e., combining/precoding) [SY16]. Electronically controllable beamforming networks can be employed to realize the narrow analog beams. In this case, only a subspace projection of the channel can be observed via a limited number of RF-chains, which typically has a much smaller size than the number of antennas. Therefore, the digital beamforming problem should be solved in baseband with the reduced size which makes it tractable. The rate loss between fully digital and suboptimal Hybrid Beamforming (HBF) is quite moderate in realistic scenarios. This practical solution as the signal processing framework for 6G (depicted in Fig. 2.3) offers an energy and cost-efficient hardware structure.

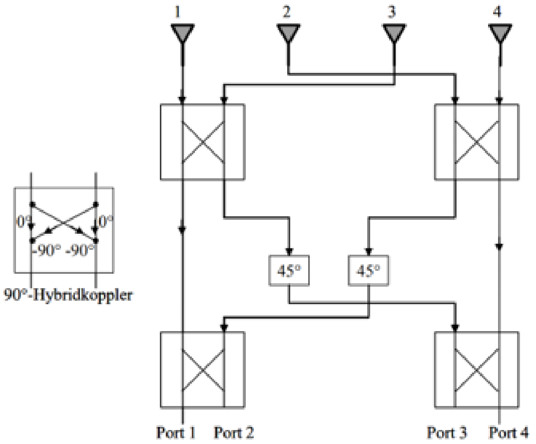
In the following, we introduce two typical HBF architectures, namely, the Fully Connected (FC) and the Partially Connected (PC) architectures used to develop two state-of-the-art pieces of hardware including a beamforming network and an integrated transceiver at TUD in collaboration with other partners. In the FC architecture, each RF antenna port is connected to all antenna elements of the array through individual pathways, while in the PC architecture the RF antenna ports are connected to disjoint subarrays.

2.2.1 Butler Matrix: Analog Beamforming Network

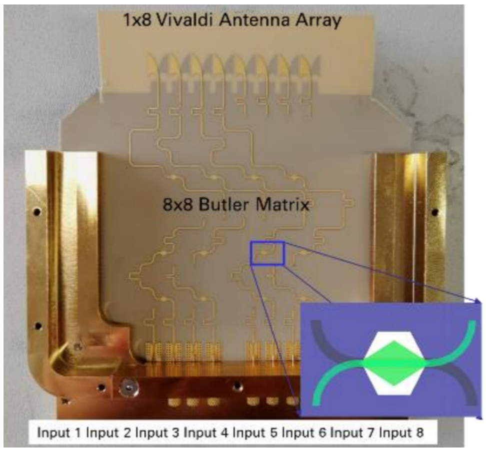
There exists an increasing interest in Butler Matrix (BM) [But61] as the potential beamforming network for the antenna arrays in mm-wave massive MIMO systems. It can simply generate highly directional orthogonal narrow beams, while achieving continuous beam scanning without any mechanical motion. The beam switching can take place via an electronic switch which changes the input port of the BM connected to the RF-chain. Use of multiple orthogonal beams is as easy as connecting multiple RF-chains to different ports of the BM.

Through a collaboration between RF and Vodafone chairs at TU Dresden and NEC Japan in the context of the DFG project mmWave Antenna Array Concept Study, a MIMO analog beamforming network based on the concept of BM at 28 GHz carrier frequency have been developed [WLP+18]. A $M \times M$ BM maps each of the M input ports to all M antenna elements through a fully connected HBF structure, as depicted in Fig. 2.4. A set of M orthogonal beams ('codebook') based on Discrete-Fourier-Transform (DFT) matrix steered into different main lobe directions can be achieved by applying the right relative phase shifts from each port to each antenna via true time delays. The usable bandwidth of a BM design is mainly limited by the hybrid coupler. To overcome the limitation of narrowband branch line couplers with relative bandwidth of around %20, an aperture hybrid coupler [AB07] (relative bandwidth %100) together with Schiffman phase shifters [Sch58] is utilized in this project. The initial 8×8 design in [WLP+18] as shown in Fig. 2.4 is scalable and can be used to build larger BMs to feed larger arrays of 16 and above [WLP+19a, WLP+19b].

As a reference, the ideal beam patterns as Group Factor (GF) using Butler Matrix can be attained theoretically. When the BM feeds the uniform linear antenna array with patch elements, the overall patterns are the result of the multiplication of the GF and the single antenna element pattern. If the single antenna element pattern is ideal, i.e., uniform in all directions, then the group factor and overall beam patterns are identical. Group factor can be calculated as multiplication of DFT codebook by matrix of angel of arrivals (AoA). Fig. 2.5 depicts the magnitude of the group factor regarding the 8×8 BM.



(a) Block diagram of a 4x4 Butler Matrix employing hybrid couplers.



(b) Prototype of a 8×8 Butler Matrix beamforming network including a 1×8 antenna array

Figure 2.4: Implemented hardware for the fully connected beamforming network based on the Butler Matrix and a linear array [WLP+18].

Beam index	1	2	3	4	5	6	7	8
Steering angle (deg)	-61	-39	-22	-7	7	22	39	61

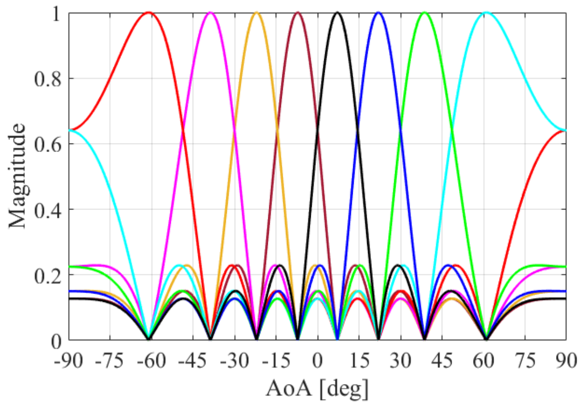


Figure 2.5: Absolute value of the group factor patterns (ideal magnitude patterns) generated by 8×8 BM.

RF chair at TU Dresden has conducted the beam measurements at the frequency of 28 GHz by receiving a directionally transmitted plane wave signal from a distance of 2 meters over angle of arrivals from $[-90, 90]$. At each port of BM, the S-parameters are recorded for each AoA, where S-parameters can be measured as either power values or complex values. Fig. 2.6 shows schematically the measurement procedure. Provided measured S-parameters are then used to plot the magnitude patterns.

To show the effect of the single element antenna pattern on the overall pattern, we have used the single patch antenna pattern provided by RF chair at TU Dresden to attain the plot in Fig. 2.7a. In this figure, we observe that as the angle of arrival shifts away from zero, the magnitude reduces. This is because the single patch antenna used in the design is not ideal and therefore does not have a uniform gain in all directions. Fig. 2.7b

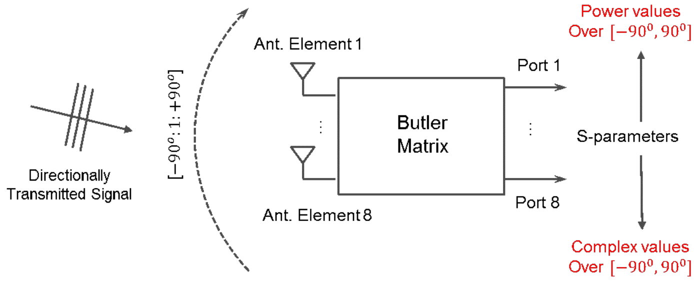
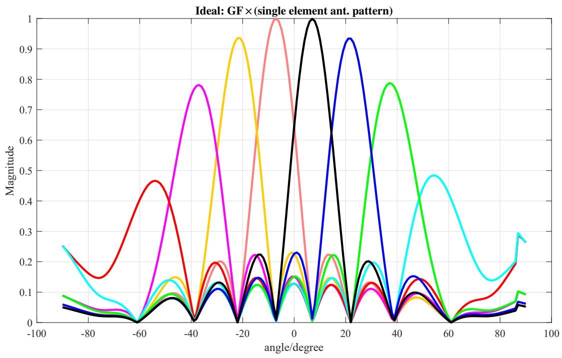
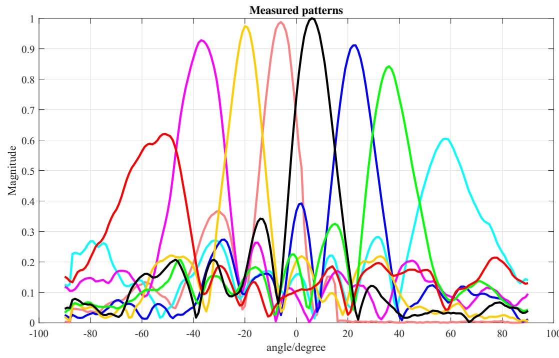


Figure 2.6: Beam measurement procedure for 8 × 8 BM with 1 × 8 uniform linear array.

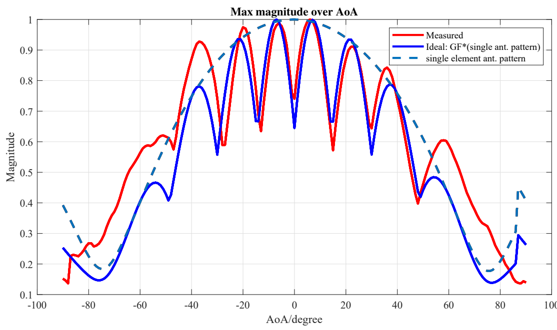
illustrates the magnitude patterns attained by power S-parameter measurements at each port of the BM. Comparing these patterns with Fig. 2.7a, in which we used the same single element antenna pattern, shows that: firstly the steering angles are approximately the same. Secondly, the ratios between maximum value of the main lobes and side lobes are close as well. Fig. 2.7c compares the maximum magnitude patterns of an ideal and measured patterns over different angle of arrivals, which is the interesting part of the patterns in case of single dominant path channel. This qualitatively shows that the performance of the implemented BM is very close to the ideal design in single dominant path channel, specially in the angular range $[-60, 60]$ which is of most interest in practical solutions.



(a) Ideal beam patterns including the element factor.



(b) Measured beam patterns using the S-parameters.



(c) maximum magnitude comparison between measured and ideal patterns

Figure 2.7: Ideal and measured beam patterns of the developed 8 × 8 butler matrix

A variant of this developed BM will be used in chapter 6 to experimentally demonstrate the performance of beam selection in a practical mmWave Massive MIMO system.

2.2.2 6GKOM D-Band Module

The 6GKom research project aims to develop a hardware basis for 6G at an early stage. In contrast to 6G preparatory research activities currently underway worldwide, the main objective of 6GKom is to research and develop an efficient, ultra-wideband and miniaturized massive MIMO D-band module with integrated beamforming capability based on a novel scalable hardware architecture. This integrated system will enable multi-terabit per second (Tbit/s) data rates and very precise localization applications for future 6G mobile communications.

The D-band module is designed based on a partially connected hybrid beamforming structure operating at 140 GHz carrier frequency. The power efficiency of a partially connected architecture compared to a fully connected architecture is higher while suffering from marginal performance loss [SKC20]. Moreover, scalability can be achieved via this HBF structure. This enables us to build any desirable array size to fulfill the required link budget by simply stacking the D-band modules as subarrays. In Fig. 2.8 it is illustrated how an array of subarrays or super arrays can be constructed using the D-band module. This is suitable for a variety of scenarios, such as multi-user massive MIMO in cellular mobile communications and multi-stream LOS MIMO in fixed wireless backhaul [SRB+18].

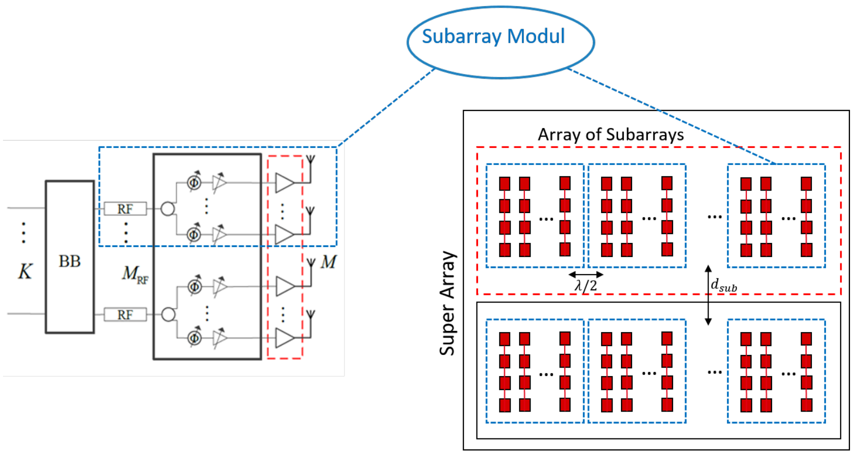
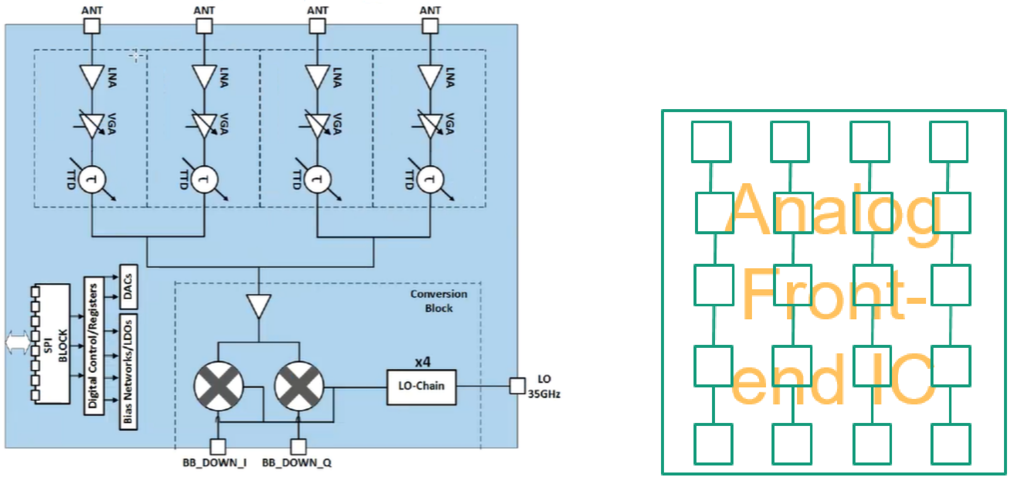


Figure 2.8: Partially connected Hybrid Beamforming structure for D-Band Module on the left and construction arrays of subarrays using the D-band module on the right.

The layout of this module, under the constraints present in the packaging process, has been optimized [DSK+22] to enable the desired scalability feature. This provides a seamless solution for construction of (uniform) arrays of subarrays based on required beamforming gain and number of users. Moreover, these constructed arrays enjoy more flexibility in having individual access to beamforming elements (D-band module). This flexibility can be exploited to enable adaptive and situation aware beamforming resulting in a more energy-efficient network.



(a) Receiver topology (schematic provided by iHP).

(b) Packaged IC with integrated antennas on the top.

Figure 2.9: Developed D-band receiver IC packaged with antennas in the D-band Module.

Fig. 2.9 illustrates the receiver topology and the packaged frontend IC with integrated antennas on the top. The developed D-band module has the following specifications:

- Packaged chip size of 5×5 mm^2.
- Carrier frequency of 140 GHz
- Available bandwidth of 60 GHz (initially usable 20 GHz, bottleneck being the integrated antennas on the package).
- 1 Analog Front End chip per module.
- 4 beamforming channels & one IF channel per module.
- Beamforming in azimuths plane:
  - Antenna gain of 7 dBi (5 elements in elevation).
  - Beamforming gain of 6 dB (4 azimuths beamforming channels).
- Transmit power of 5 dBm (with 7 dB power backoff).
- 4bit phase shifters (true time delays).
- Fast beam switching time of 1 ns.
- Use of Variable Gain Amplifiers (VGAs) per beamforming channel to enable amplitude control in addition to phase control for more precise beamforming.

In the following, we have done a detailed link budget calculation (see Fig. 2.10) for the Single Input Single Output (SISO) LOS scenario where a single D-Band module is employed at the transmitter and the receiver. This provides a reference over achievable spectral efficiency with respect to range. Based on the realistic numbers including detailed

$$\mathrm{PL}(f,d,n,\alpha)[\mathrm{dB}] = \mathrm{FSPL}(f,1\mathrm{m})[\mathrm{dB}] + 10n\log_{10}(d) + \mathrm{AT}[\mathrm{dB}] + \varepsilon_\sigma$$

$$\mathrm{FSPL}(f,1\mathrm{m})[\mathrm{dB}] = 20\log_{10}\left(\frac{4\pi f}{c}\right)$$

$$\mathrm{AT}[\mathrm{dB}] = \alpha[\mathrm{dB/km}] \times d/1000$$

$$\mathrm{TN}(T=290\mathrm{K})[\mathrm{dBm}] = -174\ [\mathrm{dBm/Hz}] + 10\log_{10}(W)$$

$$\mathrm{BG}(M)[\mathrm{dB}] = 10\log_{10}(M)$$

$$\mathrm{LB}[\mathrm{dB}] = \mathrm{TP} + 2\times\mathrm{AG} + 3\times\mathrm{BG} - \mathrm{PL} - \mathrm{TN} - \mathrm{NF} - \mathrm{IL} - \mathrm{FL}$$

TP[dBm]= 5 AG[dBi]= 7 BG[dB]= 6 IL[dB]= 3

$\mathrm{TN}(T=290\mathrm{K}, W=2\mathrm{GHz})[\mathrm{dBm}] = -81$ NF[dB]= 12 FL[dB]= 3

$\mathrm{PL}(f=140\mathrm{GHz}, d=20\mathrm{m}, n=2.5, \alpha=2)[\mathrm{dB}] = 108$

- TP: transmit power
- PL: path loss
- FSPL: free space path loss
- AT: attenuation
- TN: thermal noise
- NF: noise figure
- AG: antenna gain
- BG: beamforming gain
- IL: integration loss
- FL: front-end loss
- LB: link budget
- W: bandwidth
- M: number of antennas
- f: carrier frequency
- d: distance in meters
- n: path loss exponent
- c: speed of light
- α: attenuation per km
- ε_σ: Zero mean Gaussian r.v.

Figure 2.10: D-Band SISO LoS Link Budget for single module per TX/RX. The hardware related numbers are provided by iHP, IZM and Uni Ulm.

hardware specifications and path-loss at 140 GHz, a spectral efficiency of 1 bit/sec/Hz is achievable at 10-20 meters, which translates to 10-20 GB/sec considering the useable bandwidth of 10-20 GHz. This low-level of spectral efficiency per stream is achievable with relatively simple modulation and coding schemes. For instance, as the Signal to Interference plus Noise Ratio (SINR) at the receiver is between 0 and 3 dB, when using binary codes of rate 1/2 mapped onto a Quadrature Phased Shift Keying (QPSK) constellation. The range and/or achieved data rate can be extended by using multiple modules to build a larger array with greater achievable beamforming gain based on any specific application.

2.3 Summary

In this chapter, we have discussed the hybrid digital-analog beamforming as the efficient signal processing framework for mmWave massive MIMO systems. Hybrid beamforming enables novel hardware implementations achieving higher energy efficiency and lower cost. Vodafone chair at TU Dresden has been involved in development of such heardwares from ground up with signal processing aspects related to real life applications in mind. Two of such innovative hardware modules have been mentioned in this work. First, the beamforming network based on the Butler matrix for a uniform linear array. It can simply generate highly directional orthogonal narrow beams, while achieving fast and continuous beam scanning without any mechanical motion. The beam switching can take place via an electronic switch which changes the input port of the BM connected to the RF-chain. Second, a miniaturized massive MIMO D-band module with integrated beamforming capability based on the partially connected hybrid beamforming structure. This enables us to build any desirable array size to fulfill the required link budget by simply stacking the D-band modules as subarrays. The attained scalability and flexibility is desirable for a variety of scenarios, such as multi-user massive MIMO in cellular mobile communications and multi-stream LOS MIMO in fixed wireless.

Chapter 3

Variable-length Energy-Detection Framework for Beam Selection

Beam acquisition is the most vital step in the link-establishment process in mmWave massive MIMO systems using HBF technology. The aim is to maximize the energy transfer through the best possible effective channel. To find the best beams with highest captured signal power, a rather difficult detection problem should be solved. Considering the variable operating conditions in mmWave channels, a desirable solution should have the following desirable features:

- Efficient in training (limited coherence time, delay constraints),
- Adaptive to channel conditions (large SNR range),
- Robust to realized channels (LOS, NLOS, Multipath, non-ideal beam patterns).

For the link access task, we present a novel energy-detection framework based on variable-length channel measurements with orthogonal codebooks. The proposed beam selection technique denoted as composite *M-ary Sequential Competition Test* (SCT) solves the beam selection problem when knowledge about the SNR operating point is *not* available. It adaptively changes its test length when the SNR varies to achieve an essentially constant performance level. In addition, it is robust to deviations of real beam patterns due to fabrication tolerances or varying element factors. Next, we augment the SCT by a beam elimination mechanism. It evaluates the a-posteriori PDFs of the received amplitude of all candidate beams acquired via a sequential linear MMSE estimator. Then for each candidate, the current probability of outperforming the leading candidate is evaluated. This allows to eliminate unpromising beams from the remaining candidate set as soon as possible. In this way, the *Sequential Competition and Elimination Test* (SCET) significantly further reduces training time and increases efficiency w.r.t. pure competition, while enjoying the inherent adaptivity and robustness. The main proposed ideas in this chapter are presented in the following publications [MRF19, RM19, MRF20, MRF18, KMRF19].

3.1 System Model

Let us assume an environment with multipath propagation as it is sketched in Fig. 1.3, where the training sequence $s[n]$ of a certain user propagates in the form of P resolvable plane waves with different delays τ_p, Angles of Arrival (AoA) ϕ_p and corresponding path gains ρ_p to the receiver. This signal is observed with a generic antenna array of M antennas and a corresponding set of M beam patterns or codebook. Considering w.l.o.g. the simplest case of a Uniform Linear Array (ULA) and a set of orthogonal beams, the steering vectors $\mathbf{w}_m(\phi_m)$ with steering directions ϕ_m under each beam m in the azimuth plane are given by (changed to root M)

$$\mathbf{w}_m(\phi_m) = \frac{1}{\sqrt{M}}[1, e^{\frac{j2\pi d}{\lambda_0}\sin(\phi_m)}, \ldots, e^{\frac{j2\pi d(M-1)}{\lambda_0}\sin(\phi_m)}]^{\mathrm{T}}, \tag{3.1}$$

where λ_0 is the wavelength at the carrier frequency. We assume the distance between neighboring antenna elements to be $d = \lambda_0/2$ so that the spatial sampling theorem is respected. The beamformer codebook is represented by the matrix,

$$\mathbf{W} = [\mathbf{w}_1(\phi_1), \mathbf{w}_2(\phi_2), \ldots, \mathbf{w}_M(\phi_M)] \,. \tag{3.2}$$

The array propagation vector is defined in an analogous way. A path with the angle of arrival ϕ_p is described by the column vector

$$\mathbf{a}(\phi_p) = \frac{1}{\sqrt{M}}[1, e^{j\pi\sin(\phi_p)}, \ldots, e^{j\pi(M-1)\sin(\phi_p)}]^{\mathrm{T}} \,. \tag{3.3}$$

In baseband this leads to the following discrete time receive sequence after beamforming using a candidate beam m within the coherence interval as

$$b_m[n] = \sum_{p=1}^{P} \underbrace{\rho_p \mathbf{w}_m^H(\phi_m)\mathbf{a}(\phi_p)}_{A_{m,\tau_p}} s[n-\tau_p] + z_m[n] \,, \tag{3.4}$$

where n indicates the integer-valued sample index while the path gains ρ_p, the discrete delays $\tau_p \in [0, T-1]$ and AoAs ϕ_p are unknown to the receiver (the maximum delay T is to be chosen according to the environment).

A pseudo-random sequence with $s[n] \in \{\pm 1\}$, unit variance and $P\{s[n] = +1\} = P\{s[n] = -1\} = 1/2$ is assumed for training, so that $\mathbb{E}[s[n]s[n-k]] \simeq \delta[k]$ holds for its auto-correlation sequence. The additive noise $z[n] \sim \mathcal{CN}(0, \sigma^2)$ is modeled as an i.i.d. complex zero mean Gaussian r.v. with unknown variance. The combined effective channel and beamforming gain corresponding to beam m and path p can be written as the unknown complex parameter A_{m,τ_p}. Depending on the number of available RF-chains, the observations from different beams at each time instant n can be made in parallel, serially or in a combined fashion.

3.2 Statistical Channel Model

Currently, most of the proposed models for mmWave channels in 5G, such as in both 3GPP standards and IEEE 802.11ad for the 60 GHz band [MMS+10], are basically developed from the so-called double directional channel model [SMB01], which is applicable to arbitrary frequencies including the mmWave bands. Similarly, in [SR15, SR16b] a 3D statistical model for mmWave channels is introduced, which includes LoS and NLOS scenarios.

To model the mmWave channel, the key channel parameters that characterize the radio spatial propagation between the transmitter and the receiver must be characterized [LWR14]. The transmitted sequence $s[n]$ propagates in terms of plane waves to the receiver over specific paths, which can be modeled via the following three parameters:

- Path delay τ_p: propagation time over the path p.
- Angle of arrival (AoA) ϕ_p: angle between the incidence-path direction of the signal and the direction perpendicular to the antenna array.
- Path gain ρ_p: a coefficient of the channel, which can be considered being inversely proportional to the path loss generated by the corresponding path.

For simplicity of the model, we hereby treat each arriving cluster of rays in the channel model as a single propagation path, since the AoA and delay of each ray within the same cluster are almost the same. Then referring to the omnidirectional Channel Impulse Response (CIR) proposed in [SR15], we formulate the CIR at the receiver in this work as

$$h(t, \phi) = \sum_{p=1}^{P} \rho_p \delta(t - \tau_p) \delta(\phi - \phi_{AoA,p}), \tag{3.5}$$

where t is the absolute propagation time and ϕ indicates the azimuth AoA; P denotes the number of paths; ρ_p, τ_p and ϕ_p are the gains, discrete delays and azimuth AoAs of each multi-path component, respectively.

It can be assumed that in mmWave channels, each multipath component ρ_p (or coefficient of a channel tap in Eq. (3.4)) is created by the superposition of numerous micro-scattering components (e.g., due to rough surfaces) having approximately the same AoA ϕ_p and delay τ_p. It is also customary to invoke the central limit theorem and to model the superposition of these many small effects as Gaussian [PS01, Bel63]. With the additional assumption, a constant coefficient within each coherence interval for each path is modeled by Rician fading which is given by

$$\rho_p \ \sim \sqrt{\alpha_p}\left(\sqrt{\frac{\eta_p}{1+\eta_p}} + \frac{1}{\sqrt{1+\eta_p}}\epsilon_p\right), \tag{3.6}$$

where α_p stands for the overall multipath component power, $\eta_p \in [0, \infty)$ determines the power ratio between the deterministic LOS and the random scattered NLOS components, while $\epsilon_p \sim \mathcal{CN}(0,1)$ is a zero-mean unit-variance complex Gaussian random variable whose value changes in an i.i.d. fashion across the coherence intervals. Note that, $\eta_p \to \infty$ declares a pure LOS path while $\eta_p \to 0$ indicates a pure NLOS path, affected by Rayleigh fading.

On the other hand, we can also roughly quantify the propagation power α_p by the possible Path Losses (PLs), as the increasing PL will result in the descent of received power. As introduced in [LWR14], the loss after traveling distance of d_p can be calculated as

$$PL(d_p) = \beta^2(d_p)\left(\frac{4\pi d_p}{\lambda_0}\right)^2, \tag{3.7}$$

where $\beta^2(d_p)$ stands for the possible reflection loss over the path that happens in NLOS scenarios, while the latter factor in the product is the Free-Space Path Loss (FSPL) over distance d_p, which is considered as an essential path loss and can be derived from the Friis transmission formula.

3.3 Beam Selection As a Multi-Hypothesis Testing Problem

After symbolwise multiplication of the received sequence with the training sequence at different delays, the effective correlated observation sequence $y_{m,\tau}[n]$ for each beam m and delay τ becomes

$$y_{m,\tau}[n] = s[n]b_m[n+\tau] = A_{m,\tau} + z_{m,\tau}[n] \ , \tag{3.8}$$

with $z_{m,\tau}[n]$ being i.i.d. zero mean complex Gaussian noise with the same unknown variance σ^2 that belongs to $z_m[n]$ and is also assumed to be independent of m.

Through the training phase Given the size of the codebook M and the maximum assumed delay T in units of the sampling interval, the beam selection problem can be understood as an $M \times T$-ary hypothesis test where each hypothesis $\mathcal{H}_{m,\tau}$ is defined as

$$\mathcal{H}_{m,\tau} : |A_{m,\tau}| \equiv |A_{\max}| = \max_{j\in\{1,\ldots,M\},k\in\{0,\ldots,T\}} \{|A_{j,k}|\} \ , \tag{3.9}$$

with the meaning that the beam m with delay τ picks up the highest amount of signal magnitude among all possible beam and delay indices.

Defining the vector of correlated observations for beam m at all delays as $\mathbf{y}_m[n] = [\mathbf{y}_{m,0}[n], \ldots, \mathbf{y}_{m,T}[n]]$ where $\mathbf{y}_{m,\tau}[n] = [y_{m,\tau}[1], \ldots, y_{m,\tau}[n]]$, one can define the following performance measures:

Alignment Accuracy: the quality of any detector can be described in the following normalized average loss of signal magnitude taking into account a weighted selection probability associated with each hypothesis $\mathcal{H}_{m,\tau}$:

$$\bar{l} = 1 - \sum_{m=1}^{M} \sum_{\tau=0}^{T} \frac{|A_{m,\tau}|}{|A_{\max}|} \Pr\{\mathcal{H}_{m,\tau}|\mathbf{y}_1[n_1], \ldots, \mathbf{y}_M[n_M]\} \, , \tag{3.10}$$

where $\Pr\{\mathcal{H}_{m,\tau}|\mathbf{y}_1[n_1], \ldots, \mathbf{y}_M[n_M]\}$ denotes the probability of deciding for or selecting beam/delay pair (m, τ) after measuring the sequences $\mathbf{y}_1[n_1], \ldots, \mathbf{y}_M[n_M]$. The price to be paid for not selecting the strongest signal depends on the specific hypothesis that is chosen (e.g., whether a smaller or a larger fraction of the maximum signal level is missed). Therefore, a meaningful cost function for beam selection is defied as $\bar{l}$ where the aim is to maximize the average selected signal magnitude, meaning reaching higher beam alignment accuracy.

Alignment Efficiency: Efficiency of the beam selection with respect to the accuracy can be described by the total number of measurements acquired. This is given by summing the number of measurements via each individual beam as:

$$n_{\text{tot}} = \sum_{m=1}^{M} n_m, \tag{3.11}$$

where n_m stands for the sequence length observed under the beam m before the decision. A detector is more efficient if it requires less time to measure each beam on average to achieve the same accuracy in terms of $\bar{l}$.

3.4 Energy-Detection Via Fixed-length Test

A sufficient statistic to estimate the unknown amplitudes $A_{m,\tau}$ at each sequence length $n \geqslant 1$ is given by the sample mean of the correlation sequence $y_{m,\tau}[n]$, or,

$$\hat{A}_{m,\tau}[n] = \bar{y}_{m,\tau}[n] = \frac{1}{n}\sum_{i=1}^{n} y_{m,\tau}[i] \ , \tag{3.12}$$

which corresponds to the MVU estimator of $A_{m,\tau}$. A Fixed-length test [BHR+15,CRKF18] probes the channel via each beam m with a $n_m = N_{\text{fix}}$ number of observations. It decides for hypothesis $\mathcal{H}_{m,\tau}$ over all candidate beam-delay pairs if

$$|\hat{A}_{m,\tau}[N_{\text{fix}}]| = |\bar{y}_{m,\tau}[N_{\text{fix}}]| = \max_{j\in\{1,\ldots,M\},k\in\{0,\ldots,T\}} \{|\bar{y}_{j,k}[N_{\text{fix}}]|\} \ , \tag{3.13}$$

where $|.|$ denotes again the absolute value operation.

Assuming an operating point $\text{SNR} = |A_{\max}|^2/\sigma^2$ in a LOS channel (i.e., single dominant candidate), one can simply choose the test length N_{fix} based on the desired level of performance in terms of $\bar{l}$ [MRF18]. Fixed-length detectors are prone to SNR change and the achieved performance in terms of $\bar{l}$ can vary greatly. To better understand this effect, in the following we present a toy example.

The discrimination between two candidates captures the most essential features of our energy-detection problem. In the following example, we restrict the number of candidate beams to $M = 2$ and consider a single user scenario that uses correlated observations[1] $y_i[n] = s[n]r_i[n]$ for $i = 1, 2$. Considering the difference $D = A_1 - A_2$, a sufficient statistic based on the data $\{y_i[n], \ i = 1, 2\}$ equals to the MVU estimator for D, which is given by the sample mean after correlation as

$$\hat{D} = \frac{1}{N}\sum_{n=0}^{N-1} y_1[n] - y_2[n] = \bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2 \ , \tag{3.14}$$

whwere $\hat{D}$ follows a Gaussian PDF as $\hat{D} \propto \mathcal{N}(A_1 - A_2, 2\sigma^2/N)$. The performance of the test is characterized by the so-called deflection coefficient $d^2 = N(A_1 - A_2)^2/(2\sigma^2/N)$. A conventional Fixed-length test [NP33,Kayb] with arbitrary length N decides for beam $i = 1$, if $\hat{D} > 0$ or $i = 2$, if $\hat{D} < 0$. Defining $A_{\max} \equiv \max(A_1, A_2)$ this results in the

[1] Assuming a set of sequences with good auto- and cross-correlation properties so that different users can be assigned to different sequences allows a generalization to multipath and multiuser scenarios without difficulty.

average captured relative magnitude given by

$$\begin{aligned}
\bar{a} &= \mathrm{P}\{\hat{D} > 0\}\frac{A_1}{A_{\max}} + \mathrm{P}\{\hat{D} < 0\}\frac{A_2}{A_{\max}} \\
&= Q\left(\frac{A_2 - A_1}{\sqrt{2\sigma^2/N)}}\right)\frac{A_1}{A_{\max}} + Q\left(\frac{A_1 - A_2}{\sqrt{2\sigma^2/N}}\right)\frac{A_2}{A_{\max}} \ ,
\end{aligned} \tag{3.15}$$

where the Q-function returns the right tail probability of a standard Gaussian r. v. with zero mean and unit variance.

As performance criterion, we chose the normalized average loss of signal magnitude, denoted as $\bar{l}$ as explained in Eq. 3.10. Defining the ratio $r \equiv \min(A_1, A_2)/A_{\max} \in [0, 1]$ the value of $\bar{l}$ achieved with a Fixed-length test based on the correlation mean can be characterized using Eq. (3.15) as

$$\bar{l} = 1 - \bar{a} = (1 - r)Q\left(d\right) \ . \tag{3.16}$$

The fractional loss $\bar{l}$ is upper-bounded by $(1 - r)/2$, corresponding to beam selection by coin tossing (i.e., when the selection is done randomly). Note that, for $r = 1$ we cannot lose anything since both candidates are equally good. Eq. (3.16) indicates that to achieve a target performance specified by $\bar{l}_{\text{target}}$ at given r, the required deflection coefficient is

$$d^2_{\text{req}} = \left(Q^{-1}\left(\frac{\bar{l}_{\text{target}}}{1 - r}\right)\right)^2 \ . \tag{3.17}$$

This means the minimum required number of observations to achieve a certain value of $\bar{l}_{\text{target}}$ can be calculated in this case as $N_{\text{req}} = [2\sigma^2 d^2_{\text{req}}/(A_1 - A_2)^2]$.

The problem that arises in the design of the training sequence (i.e., a detector) is to choose a length N_{fix} to achieve $\bar{l}_{\text{target}}$ in a range of scenarios that occur randomly in applications where exact knowledge about $(A_1 - A_2)^2/\sigma^2$ and r is not or only roughly available. This will lead to a mismatch between the effectively achieved deflection coefficient and its required value that we describe as

$$d^2_{\text{mis}} = d^2_{\text{eff}} - d^2_{\text{req}} \ , \tag{3.18}$$

where the effective deflection coefficient $d^2_{\text{eff}} = N_{\text{fix}}(A_1 - A_2)^2/2\sigma^2$ will be based on N_{fix} used for the design. As illustrated in Fig. 3.1, the mismatch causes the achieved value of $\bar{l}$ to deviate from $\bar{l}_{\text{target}}$. The sensitivity to this mismatch between effective and required deflection coefficients increases as the target value $\bar{l}_{\text{target}}$ for the fractional loss decreases (a stricter requirement lead to stronger deviations). E.g., for $\bar{l}_{\text{target}} = 0.01$, a 3 dB deviation of d^2_{eff} from d^2_{req} leads to values of $\bar{l} \simeq 0.1$ and $\bar{l} \simeq 5 \times 10^{-4}$ respectively.

Therefore, naively fixing the test length to some value N_{fix} based on a certain assumed

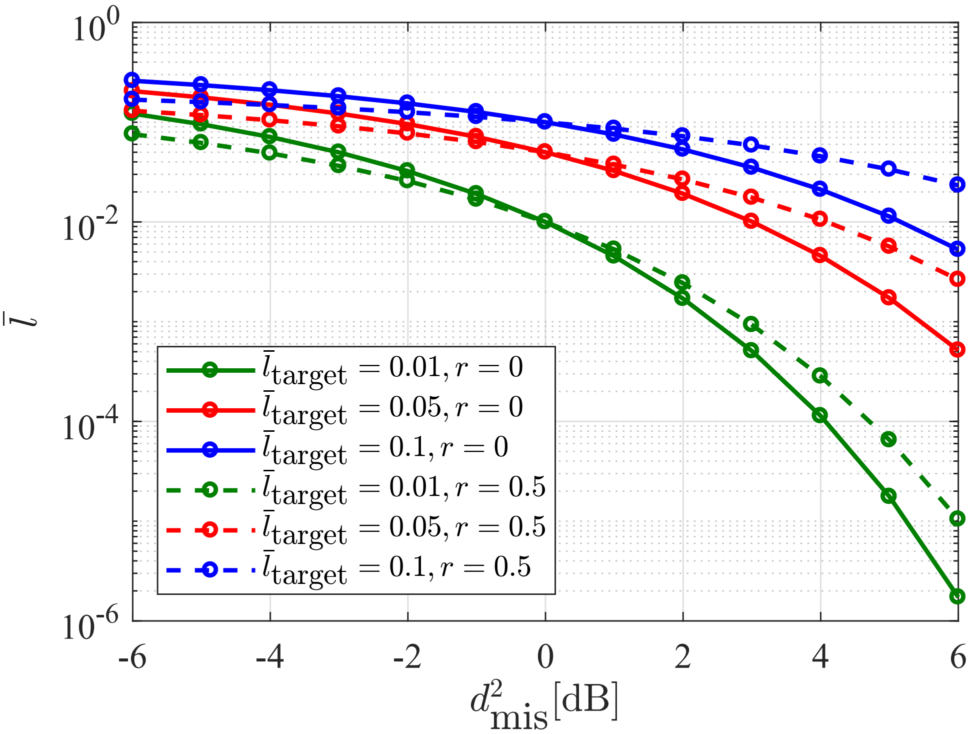


Figure 3.1: Fractional signal loss $\bar{l}$ for Fixed-length tests using 3 different target values $\bar{l}_{\text{target}}$ and ratios $r \in \{0, 1/2\}$ vs. deflection coefficient mismatch.

operating point can result in a strongly variable performance in practical scenarios. Additionally, if N_{fix} is conservatively set to a high value based on the worst still acceptable operating point, a lot of time spent will be wasted for detection of the best beam if the channel quality is better than expected.

3.5 Detection of the Signal Presence via Generalized Likelihood Ratio Test

In practice, the knowledge about the true SNR value is unavailable before beam selection. That means the underlying PDF of the strongest candidates are only partially known. Generalized Likelihood Ratio Test (GLRT) is a practical detection strategy in which the unknown parameters of the PDFs are replaced with their available estimates. The building block of our adaptive variable-length test presented in the next subsection is the GLRT to detect the presence or absence of an unknown amplitude in White Gaussian Noise (WGN) with unknown variance.

We pose the detection of a training signal of unknown strength at each given $n_m = n$ as a binary hypothesis problem: for each beam-delay candidate pair, a signal is either present (the $\mathcal{H}^1_{m,\tau}$ hypothesis) or is not present (the $\mathcal{H}^0_{m,\tau}$ hypothesis). Following Eq. (3.8), we consider a signal model for the two hypotheses of the form,

$$\begin{array}{ll} \mathcal{H}^0_{m,\tau}: & y_{m,\tau}[n] = z_{m,\tau}[n] \\ \mathcal{H}^1_{m,\tau}: & y_{m,\tau}[n] = A_{m,\tau} + z_{m,\tau}[n] \end{array} . \tag{3.19}$$

Since the variance of the noise and the signal amplitude are unknown, a proper threshold to bound the probability of deciding $\mathcal{H}^1_{m,\tau}$ when $\mathcal{H}^0_{m,\tau}$ is true, typically denoted as the probability of false alarm $P^{\mathrm{FA}}_{m,\tau}$, cannot be found using the Neyman-Pearson (NP) theorem. Instead, one can use the Maximum Likelihood (ML) estimates of the unknown parameters derived from the available observations and insert them into the likelihood functions under each hypothesis. The ML estimates of $A_{m,\tau}$ and σ^2 under $\mathcal{H}^1_{m,\tau}$ after n observations are given by

$$\hat{A}^{\mathrm{ML}}_{m,\tau} = \bar{y}_{m,\tau}[n] \tag{3.20}$$

and

$$\hat{\sigma}^2_{\mathcal{H}^1_{m,\tau}} = \frac{1}{n}(\mathbf{y}_{m,\tau}[n] - \bar{y}_{m,\tau}[n]\mathbf{1}^{\mathrm{T}})(\mathbf{y}_{m,\tau}[n] - \bar{y}_{m,\tau}[n]\mathbf{1}^{\mathrm{T}})^{\mathrm{H}}\,, \tag{3.21}$$

while under $\mathcal{H}^0_{m,\tau}$ the ML estimate of σ^2 is just

$$\begin{aligned} \hat{\sigma}^2_{\mathcal{H}^0_{m,\tau}} &= \frac{1}{n}\mathbf{y}_{m,\tau}[n]\mathbf{y}^{\mathrm{H}}_{m,\tau}[n] \\ &= \hat{\sigma}^2_{\mathcal{H}^1_{m,\tau}} + |\bar{y}_{m,\tau}[n]|^2. \end{aligned} \tag{3.22}$$

By replacing the unknown parameters with their ML estimates in the Gaussian PDFs, the GLR can be calculated as

$$L_{\mathrm{G}}(\mathbf{y}_{m,\tau}) = \frac{p(\mathbf{y}_{m,\tau}; \bar{y}_{m,\tau}[n], \hat{\sigma}^2_{\mathcal{H}^1_{m,\tau}}, \mathcal{H}^1_{m,\tau})}{p(\mathbf{y}_{m,\tau}; \hat{\sigma}^2_{\mathcal{H}^0_{m,\tau}}, \mathcal{H}^0_{m,\tau})} = \left(\frac{\hat{\sigma}^2_{\mathcal{H}^0_{m,\tau}}}{\hat{\sigma}^2_{\mathcal{H}^1_{m,\tau}}}\right)^{\frac{n}{2}} . \tag{3.23}$$

It is advantageous to work with the modified GLLR $\gamma_{m,\tau} \equiv 2\ln L_{\mathrm{G}}(\mathbf{y}_{m,\tau})$ because its statistics can be characterized in closed form. A non-trivial result that can be found in [Kayb] states that for sufficiently large n the random variable $\gamma_{m,\tau}[n]$ follows either a central or a non-central χ^2-distribution with one degree of freedom (we note that the number of degree(s) of freedom of a χ^2-variable are indicated with a subscript in the following for clarity) so that

$$\gamma_{m,\tau}[n] = n\ln\left(1 + \frac{|\bar{y}_{m,\tau}[n]|^2}{\hat{\sigma}^2_{\mathcal{H}^1_{m,\tau}}}\right) \sim \left\{ \begin{array}{ll} \chi^2_1(0), & \mathcal{H}^0_{m,\tau} \\ \chi^2_1(\lambda_{m,\tau}) & \mathcal{H}^1_{m,\tau} \end{array} \right. . \tag{3.24}$$

The ratio $\lambda_{m,\tau} = n(A^2_{m,\tau}/\sigma^2)$ is denoted in statistics as the non-centrality parameter of the $\chi^2_1(\lambda_{m,\tau})$-PDF as illustrated in Fig. 3.2.

Since now, the PDF of $\gamma_{m,\tau}[n]$ under $\mathcal{H}^0_{m,\tau}$ is completely known while being indepen-

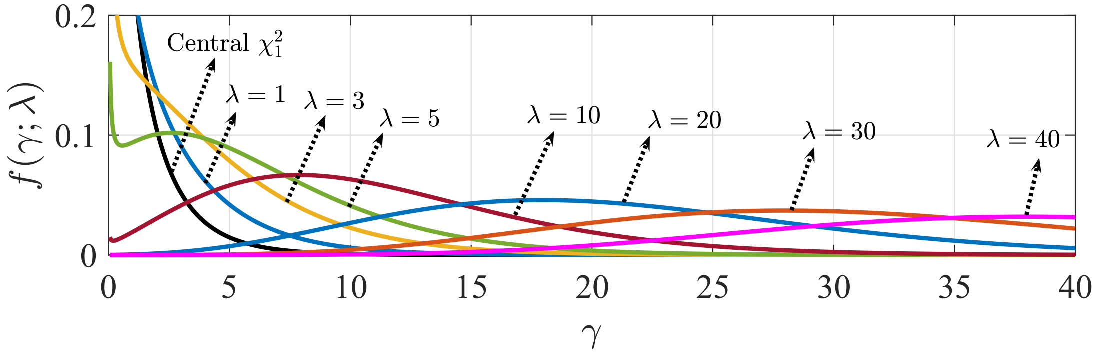


Figure 3.2: Illustration of χ_1^2 distributions with different values of the non-centrality parameter λ.

dent of n and the unknown SNR, we can ensure by using the NP design criterion that $P_{m,\tau}^{\mathrm{FA}} = \Pr\{\gamma_{m,\tau}[n] > \gamma_{\mathrm{th}}; \mathcal{H}_{m,\tau}^{0}\}$ will not surpass a predefined value by choosing a proper threshold.

Noting further that a $\chi_1^2(0)$-distributed r.v. $\gamma_{m,\tau}[n]$ is related to a standard normal r.v. $x \sim \mathcal{N}(0,1)$ by the relation $\gamma_{m,\tau}[n] = x^2$, it follows that $P_{m,\tau}^{\mathrm{FA}} = \Pr\{x > \sqrt{\gamma_{\mathrm{th}}}; \mathcal{H}_{m,\tau}^{0}\} + \Pr\{x < -\sqrt{\gamma_{\mathrm{th}}}; \mathcal{H}_{m,\tau}^{0}\}$ can be expressed as a sum of two Q-functions so that $P_{m,\tau}^{\mathrm{FA}} = 2Q(\sqrt{\gamma_{\mathrm{th}}})$. This determines γ_{th} in terms of a given constraint for $P_{m,\tau}^{\mathrm{FA}}$ as

$$\gamma_{\mathrm{th}} = \left[Q^{-1}\left(\frac{P_{m,\tau}^{\mathrm{FA}}}{2}\right)\right]^2 , \tag{3.25}$$

which leads to the following binary detection rule to be used in each binary test:

$$\gamma_{m,\tau}[n] \underset{\mathcal{H}_{m,\tau}^{0}}{\overset{\mathcal{H}_{m,\tau}^{1}}{\gtrless}} \gamma_{\mathrm{th}} \ . \tag{3.26}$$

Given the decision threshold, also the second type of error, i.e., the probability of misdetection $P_{m,\tau}^{\mathrm{MD}}[n] = 1 - \Pr\{\gamma_{m,\tau}[n] > \gamma_{\mathrm{th}}; \mathcal{H}_{m,\tau}^{1}\}$ can be obtained in closed form as

$$P_{m,\tau}^{\mathrm{MD}}[n] = Q(\sqrt{\lambda_{m,\tau}} - \sqrt{\gamma_{\mathrm{th}}}) - Q(\sqrt{\lambda_{m,\tau}} + \sqrt{\gamma_{\mathrm{th}}}). \tag{3.27}$$

It is worth noting that, the performance of the GLLRT in Eq. (3.26) will be slightly poorer than for the case when σ^2 is known. Surprisingly, for large enough n the degradation is minimal and the performance of the above detector in terms of $P_{m,\tau}^{\mathrm{MD}}[n]$ achieves the bound given by the clairvoyant NP detector that has the perfect knowledge of the parameters $A_{m,\tau}$ and σ^2. Hence, it corresponds to the Uniformly Most Powerful (UMP) test for this composite detection problem.

As shown in Fig. 3.3, the probability of misdetection $P_{m,\tau}^{\mathrm{MD}}[n]$ decreases exponentially as n increases due to the Q-function. In addition, the rate of exponential decrease depends strongly on SNR per observation $|A_{m,\tau}|^2/\sigma^2$. The sequential competition test which is

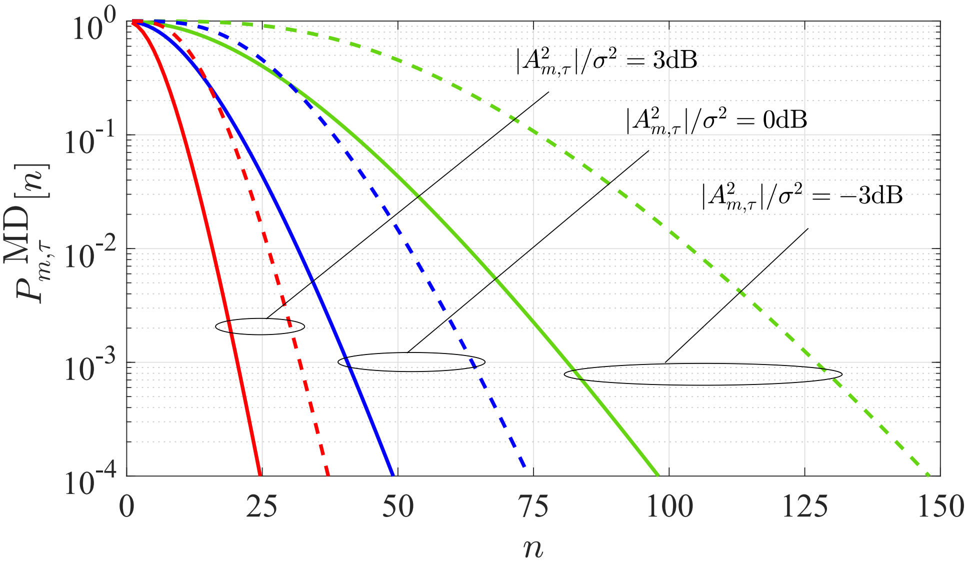


Figure 3.3:
with increasing test length n for different SNR values per observatio.]Decay of $P_{m,\tau}^{\mathrm{MD}}[n]$ with increasing test length n for different SNR values per observation, where $P_{m,\tau}^{\mathrm{FA}} = 10^{-3}$ for solid lines and $P_{m,\tau}^{\mathrm{FA}} = 10^{-6}$ for dashed lines.

elaborated next exploits this strong dependence and turns it into an advantage.

3.6 Sequential Competition Test

Returning to the initial $M \times T$-ary decision problem as treated in Sec. 3.4, where separate observed sequences were available for each beam, we now want to find optimum decision rules for a sequential test. The goal is to decide as early as possible, which among $M \times T$ unknown amplitude levels is the strongest one in a scenario where exact knowledge of the underlying PDFs except their type is not available.

We decompose the $M \times T$-ary test into $M \times T$ parallel binary tests, each formulated as in Eq. (3.19) by introducing a no-signal (null) hypothesis which acts as a virtual common reference. Allowing n to sequentially grow until a decision criterion is fulfilled, the beam-delay pair that picks up the strongest signal, will reach the given termination threshold the fastest *on average*.

As mentioned previously, the rate of exponential decrease in $P_{m,\tau}^{\mathrm{MD}}[n]$ depends strongly on $|A_{m,\tau}|^2/\sigma^2$. This is simply a consequence of the fact that the non-centrality parameter (accumulated deflection coefficient) $\lambda_{m,\tau} = n|A_{m,\tau}|^2/\sigma^2$ will grow more quickly due to enjoying the highest SNR per observation among all candidates.

Having understood that the non-central $\chi_1^2(\lambda_{m,\tau})$-PDFs move away from the invariant central $\chi_1^2(0)$-PDF with different escape velocities, the decision rule for the MT-ary

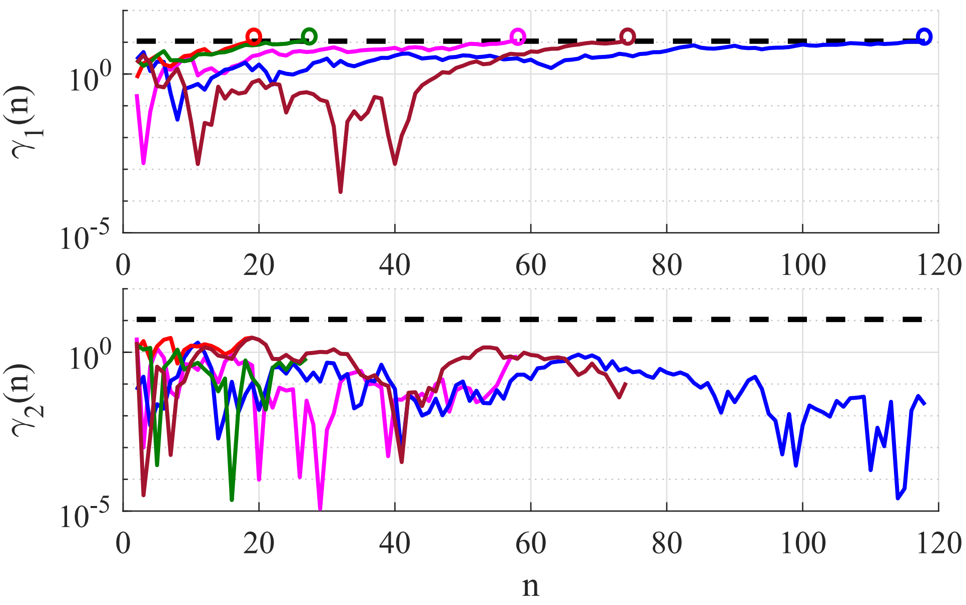


Figure 3.4: Sequential competition test visualized for $(A_1 - A_2)^2/\sigma^2 = -10$dB and $r = 0.1$. The threshold γ_{th} based on $P_{\text{FA}} = 10^{-3}$ is indicated by dashed lines, and first passages over γ_{term} are depicted via circles. Different colors represent 5 different realizations of the competition test with two candidates.

Sequential Competition Test is constructed by parallel virtual binary tests, namely,

$$\gamma_{m,\tau}[n] \underset{\text{undecided}}{\overset{\mathcal{H}_{m,\tau}}{\gtrless}} \gamma_{\text{term}} \ , \tag{3.28}$$

extended over all candidate beam/delay indices m and τ. The null hypothesis in Eq. 3.26 is now replaced with undecided. Meaning that the sequential measurements continue when all candidates are in the undecided region.

In any trial realization, the test terminates with length $n_{\text{term}} = n$ as soon as one of the sequence metrics $\gamma_{m,\tau}[n]$ surpasses the termination threshold γ_{term}, while the index of this path indicates the selected beam-delay pair. If no path metric exceeds the termination threshold, the test continues by taking the next observation $n+1$. The test length n is now a random variable with average $\bar{n} = E[n_{\text{term}}]$. Fig. 3.4 visualizes the sequential test as specified by Eq. 3.28 for $M = 2$ and a LOS channel (i.e., only two candidates).

Beam selection based on sequential competition can be understood as a specific type of first-passage-time process [Red01]. The interpretation is that we let the stochastic paths $\gamma_{m,\tau}[n]$ corresponding to different candidates compete to distinguish themselves from pure zero mean WGN by reaching the termination threshold as n increases sequentially. The one which does it fastest is declared as the 'winner' of the competition. This makes sense since the strongest candidate(s) has (have) the shortest first-passage-time on average.

The common termination threshold $\gamma_{\text{term}} = \gamma_{\text{th}}$ for all virtual binary tests is selected using Eq. (3.25). The higher we set the threshold (i.e., choosing a smaller allowed $P^{\text{FA}}_{m,\tau}$ under each virtual binary test), the longer it takes on average for the SCT to terminate

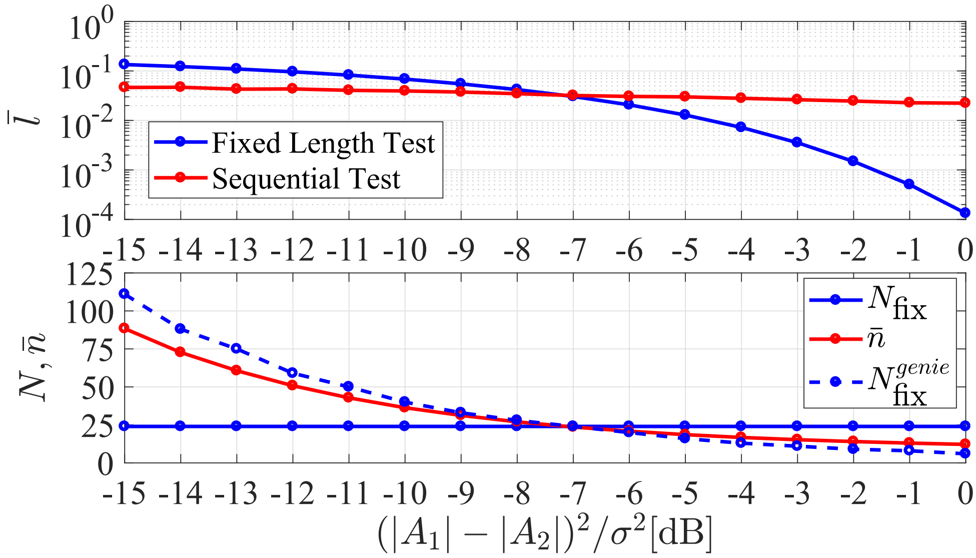


Figure 3.5: Comparing the achieved $\bar{l}$ between a Fixed-length test with length N_{fix} and the sequential test with equal γ_{th} based on $P_{\text{FA}} = 10^{-3}$.

and decide. This, however, leads to a more accurate discrimination of the strongest beam among all competitors and therefore better performance in terms of $\bar{l}$.

3.6.1 Adaptivity and Robustness

To evaluate the adaptivity and robustness of the proposed detection framework we consider the following toy example depicted in Fig. 3.5 with $M = 2$ and ratio $r = |A_1|/|A_2| = 0.5$, we observe that the sequential competition test shows an essentially invariant adaptive performance in terms of $\bar{l}$ for varying value of the differential SNR between the candidate amplitudes. This is in strong contrast to the Fixed-length test that uses length of $N_{\text{fix}} = 25$. The reason is that the average test length $\bar{n}$ (over multiple realizations) of the sequential competition test adapts itself to $(|A_1| - |A_2|)^2/\sigma^2$, as shown in the bottom plot of Fig 3.5.

In addition, as depicted in Fig. 3.6, the sequential test reacts adaptively as well to the varying ratio $r = |A_1|/|A_2|$ by changing the average number of observations. For r close to one, where the possible average loss $\bar{l}$ is negligible, but the detection of the stronger beam is more difficult, the performance of the sequential test in terms of $\bar{l}$ gets closer to the upper-bound $(1 - r)/2$ given by the random selection, i.e., coin tossing (top plot), while requiring only few observations on average (bottom plot), which makes sense.

Observe further that, the sequential competition test requires on average even less observations $\bar{n}$ in the lower SNR regime to achieve a certain value of $\bar{l}$ compared to the required number $N_{\text{fix}}^{\text{genie}}$ of a Fixed-length test tuned with genie knowledge that would achieve the same $\bar{l}$ (see Fig. 3.5, bottom). This can be understood intuitively because according to the fluctuations of the competing stochastic processes around their typical

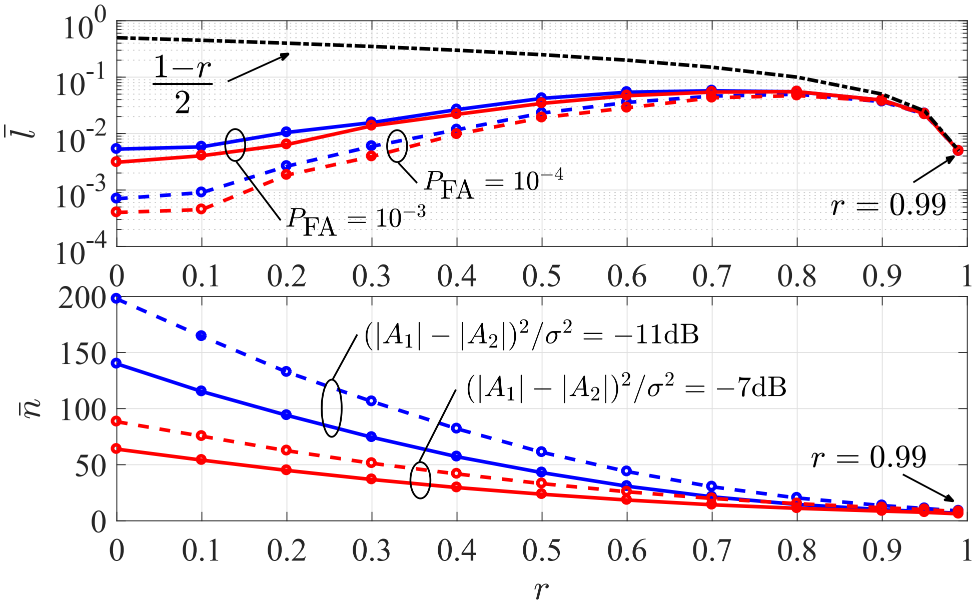


Figure 3.6: Performance of the sequential test over ratio r. The upper-bound (beam selection by coin tossing) on $\bar{l}$ is depicted as black dot dashed line.

behavior, the test exactly terminates when a reliable discrimination becomes possible, so that the test terminates earlier *on average* (which was the original motivation of *Wald* to develop his test [Wal45]). This property is of particular interest since the sequential test reduces training time at exactly those smaller values of $(|A_1| - |A_2|)^2/\sigma^2$ where many observations are needed. On the other hand, when $(|A_1| - |A_2|)^2/\sigma^2$ becomes large, the detection problem becomes easy, and we do not need many observations in the first place.

3.6.2 Training-Transmission Trade-off

In the following toy example, we numerically study the performance of the SCT in the reference channel model described in Eq. (3.4) with a single LOS path and known delay (perfect synchronization). The receiver employs a uniform linear array with 16 antenna elements using the codebook of a Butler matrix with 16 orthogonal beams. The AoA was distributed uniformly in $[-90°, 90°]$ over the channel realizations, while SNR after beamforming was defined as $|A_{\max}|^2/\sigma^2$ indicating the maximum available SNR of the best beam. The quantities $\bar{l}$ and $\bar{n}$ were estimated at each SNR point based on 10^4 simulation runs for SNR values in the interval SNR $\in [-8, 8]$ dB. For comparison, we consider the Fixed-length detector based on magnitude estimates via the sample mean as stated in Eq. (3.13) for a set of different values of the fixed test length $N_{\mathrm{fix}} \in [10, 25, 50, 75]$.

As shown in Fig. 3.7, the sequential competition test with γ_{term} based on $P^{\mathrm{FA}}_{m,\tau} = 10^{-3}$ keeps $\bar{l}$ in an interval of $[0.07, 0.14]$ or a change of at most a factor of two. This is considered to be sufficient in practice, while the average test length $\bar{n}$ is adaptively decreased as $|A_{\max}|^2/\sigma^2$ grows larger. In contrast, the Fixed-length tests show a much stronger variation

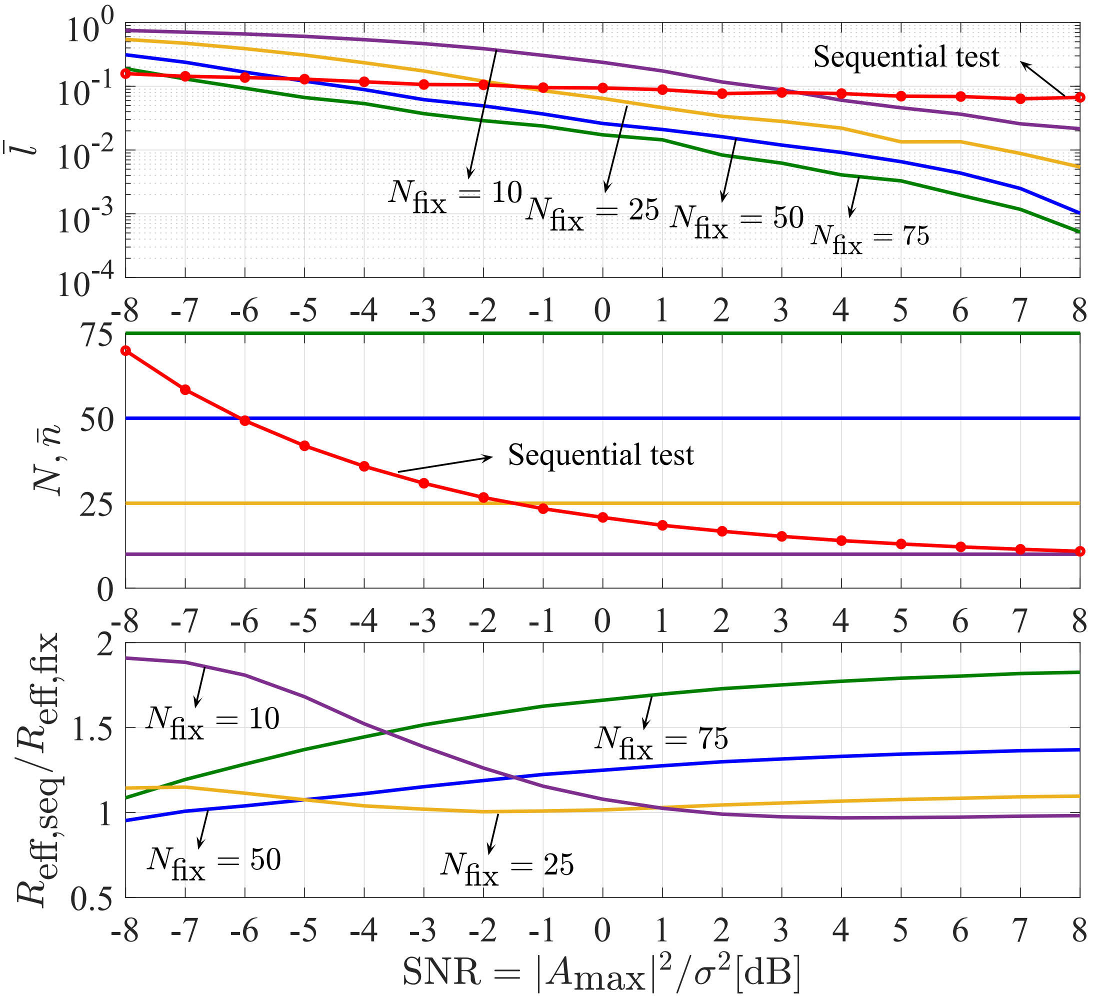


Figure 3.7: Achieved values for $\bar{l}$, $\bar{n}$ and the ratio between the effective rates using sequential competition test and Fixed-length tests with $N_{\max} = 150$ (bottom).

in performance. For instance, the Fixed-length test with $N_{\text{fix}} = 25$ results in $\bar{l}$ in an interval of $[0.005, 0.55]$ or approximately two orders of magnitude. This variation is roughly the same for the other N_{fix} values.

In case of a finite transmission or limited channel coherence time, the shorter the time we need to spend for training, the more time is left for data transmission. On the other hand, the more time we spend for training, the smaller the loss in SNR due to inaccurate decisions at the end of the training. This trade-off suggests that for $N_{\max}$ possible channel uses that can be spent for training and data transmission within some finite time interval, one can evaluate the ratio between the effective data rates achieved by the sequential and Fixed-length tests. Thus, we consider the ratio

$$\frac{R_{\text{eff,seq}}}{R_{\text{eff,fix}}} = \frac{\mathbb{E}\left[(1 - \frac{n}{N_{\max}}) \log(1 + (1 - l_{\text{seq}})^2 \frac{|A_{\max}|^2}{\sigma^2})\right]}{\mathbb{E}\left[(1 - \frac{N_{\text{fix}}}{N_{\max}}) \log(1 + (1 - l_{\text{fix}})^2 \frac{|A_{\max}|^2}{\sigma^2})\right]}, \tag{3.29}$$

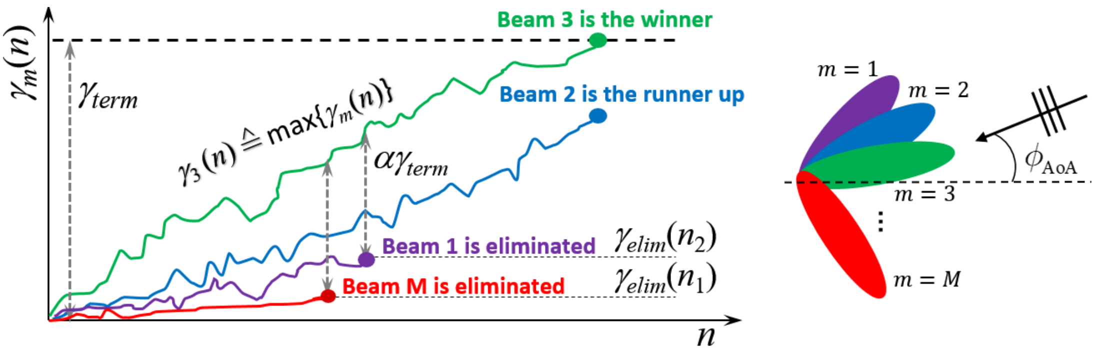


Figure 3.8: Illustration of the sequential competition and elimination test in [RM19].

where l_{seq} and l_{fix} are normalized signal magnitude losses corresponding to SCT and Fixed-length test under each channel realization, respectively. The expectation is denoted by $\mathbb{E}$ over all channel realizations.

As depicted at the bottom in Fig. 3.7 for a value of $N_{\max} = 150$ channel uses per beam (i.e., the assumed coherence block in this case amounts to $N_{\text{tot}}^{\text{coh}} = 16 \times 150 = 2400$ symbol durations), the sequential competition test fulfills the training-transmission trade-off significantly better compared to the four lengths of Fixed-length tests investigated. For the given scenario it provides an increase in effective data rate of some 10% up to a factor close to two at almost all SNR operating points for the same parametrization. This points to the adaptivity and consequently the efficiency of the proposed variable-length training strategy.

3.7 Sequential Elimination Test

As explained before, mmWave channels are sparse both in the angular and delay domains. Consequently, the majority of all $M \times T$ candidates[2] are picking up a negligible amount of signal power. This immediately suggests that the efficiency of the training procedure could be significantly increased, if much weaker beams w.r.t. to the most promising one(s) could be detected as early as possible. These could be discarded from the currently still active candidates that we denote as the 'active set'. With a reduced active set, testing continues until the final decision is achieved either by a metric crossing the termination threshold or all but one (or a few) candidate(s) are eliminated. This is the idea of a sequential elimination mechanism as proposed in [RM19], for which the question of the

[2] Actually one could add the correlation results in Eq. (3.8) along the delay dimension and consider an M-ary test instead with essentially the same numerical effort. In the algorithmic variant pursued in this work, many hypotheses are soon discarded and only the dominant path for each beam causes the decision. If the channel typically has a dominant component, both approaches are equivalent. Summing the sample means of the correlation results along the delay dimension (or all paths resolved in time) produces an estimate of the interference result observed with beam m. We note that our ideas can readily be adapted to this end by making such modifications.

optimal elimination criterion can be posed.

In general, such a variable-length test should enable us to detect candidates having a low probability of capturing the largest amount of receive energy or a high probability of *not* outperforming all other candidates. The corresponding hypotheses $\check{\mathcal{H}}_{m,\tau}$ at each time instant n are formulated as,

$$\check{\mathcal{H}}_{m,\tau} : |A_{m,\tau}| \neq \max_{(j,k)\in\mathbf{J}} \{|A_{j,k}|\} . \tag{3.30}$$

If fulfilled, we should remove the index pair (m, τ) from the set of active candidates that is initialized as $\mathbf{J} = \{(1, 1), \ldots, (M, T)\}$. Thus, appropriate probabilities associated with these hypotheses need to be derived.

In comparison to pure competition (SCT) the goal is to reduce the average number of observations as much as possible to maximize efficiency both in terms of time and energy or not squandering resources by probing unpromising beams.

3.7.1 Beam Elimination Based on GLLR Metric Differences

The original idea of beam elimination, illustrated already in Fig. 3.8, was to drop those beams that fail to keep up with the leading competitor in terms of their stochastic path metrics $\gamma_{m,\tau}[n]$ when compared to a time-dependent *elimination* threshold $\gamma_{\text{elim}}[n]$.

The latter was intuitively defined in terms of the current best metric among all competitors, from which a certain fraction of the *termination* threshold γ_{term} was subtracted:

$$\gamma_{\text{elim}}[n] = \max_{(m,\tau)\in\mathbf{J}} (\gamma_{m,\tau}[n]) - \alpha\gamma_{\text{term}} . \tag{3.31}$$

The coefficient $\alpha \in [0, 1]$ acted as a tunable parameter to optimize the cost function.

In an attempt to find the optimal elimination strategy, we developed a new probabilistic elimination criterion [MRF20]. It consists of two steps to be executed at each time instant n:

1. Based on each new observation $y_{m,\tau}[n]$ for members of the active set, update sequentially the a-posteriori PDF of each random variable $A_{m,\tau}$.

2. Determine under hypothesis c the associated probabilities of being the strongest candidate, or its complement, of not outperforming all other candidates.

As shown in [MRF20], this led to a clear improvement w.r.t. to the intuitive approach in [RM19] and will be explained next in more details.

3.7.2 Beam Elimination Based on A-Posteriori Probabilities of Winning the Competition

The goal is to find a meaningful probabilistic measure, based on which one can detect the unpromising beams at each step n during the competition. Our proposed solution evaluates the probability that the hypothesis $\mathcal{H}_{m,\tau}$ corresponds to the strongest amplitude for all candidates based on their current Bayesian estimate of amplitude levels. These probabilities can be interpreted as the estimated winning probability (i.e., Bayesian confidence) for each competitor in the competition. Once having such a measure, one can eliminate the weaker candidates during the test, with some given reliability.

In the Bayesian setup, the unknown signal amplitudes $A_{m,\tau}$ are considered as unknown random variables with some prior PDFs $f(A_{m,\tau}|\mathbf{y}_{m,\tau}[0])$. The Bayesian amplitude estimate of $A_{m,\tau}$ at each step $n \geqslant 1$ equals the mean of the a-posteriori PDF $f(A_{m,\tau}|\mathbf{y}_{m,\tau}[n])$ as $\hat{A}_{m,\tau}[n] = \mathbb{E}\{A_{m,\tau}|\mathbf{y}_{m,\tau}[n]\}$.

Assuming the amplitude prior PDFs to be zero mean complex Gaussian according to $f(A_{m,\tau}|\mathbf{y}_{m,\tau}[0]) \sim \mathcal{CN}(\hat{A}_{m,\tau}[0] = 0, \hat{\sigma}_A^2[0])$ for all beam/delay pairs (m, τ) with some assumed prior variance $\hat{\sigma}_A^2[0]$, the a-posteriori PDFs $f(A_{m,\tau}|\mathbf{y}_{m,\tau}[n])$ can be evaluated sequentially in closed form via the Sequential Linear Minimum Mean Squared Error (SLMMSE) estimator [Kaya].

The SLMMSE estimator [Kayb] treats the estimated a-posteriori PDF after the last observation, i.e. $f(A_{m,\tau}|\mathbf{y}_{m,\tau}[n-1])$, as a prior for the new observation $y_{m,\tau}[n]$. Once the measurement is made, the mean of the a-posteriori PDF $f(A_{m,\tau}|\mathbf{y}_{m,\tau}[n])$ i.e. the amplitude estimate $\hat{A}_{m,\tau}[n]$ is updated using the Kalman gain $K[n]$ and the estimate from the last step $\hat{A}_{m,\tau}[n-1]$ as

$$\hat{A}_{m,\tau}[n] = \hat{A}_{m,\tau}[n-1] + K[n] \left(y_{m,\tau}[n] - \hat{A}_{m,\tau}[n-1] \right). \tag{3.32}$$

Similarly, the variance $\hat{\sigma}_A^2[n]$ of the a-posteriori PDF $f(A_{m,\tau}|\mathbf{y}_{m,\tau}[n])$ is updated as

$$\hat{\sigma}_A^2[n] = (1 - K[n])\, \hat{\sigma}_A^2[n-1], \tag{3.33}$$

with the weighting factor $K[n]$ also known as Kalman gain[3]

$$K[n] = \frac{\hat{\sigma}_A^2[n-1]}{\hat{\sigma}_A^2[n-1] + \hat{\sigma}^2[n]}. \tag{3.34}$$

The a-posteriori PDFs $f(A_{m,\tau}|\mathbf{y}_{m,\tau}[n])$ of the unknown amplitudes $A_{m,\tau}$ are available

[3] We note that, the estimated noise variance $\hat{\sigma}^2[n]$ is employed in the denominator of $K[n]$ rather than the true value which is unknown. However, this does not hurt the performance greatly, since the noise variance can be estimated with enough accuracy in arrays with a moderately sized or large codebook.

as $\mathcal{CN}(\hat{A}_{m,\tau}[n], \hat{\sigma}_A^2[n])$ after n steps. Exploiting the independence of the observations and consequently the a-posteriori PDFs, one can express the probability of being the strongest amplitude or winning the competition for each candidate $A_{m,\tau}$ as,

$$P_{m,\tau}^{\text{win}}[n] = \int_{\mathbb{C}} f(A_{m,\tau} = x|\mathbf{y}_{m,\tau}[n]) \prod_{(j,k)\in\mathbf{J}\setminus(m,\tau)} \Pr\{|A_{j,k}| \leqslant |x|\}[n]\,\mathrm{d}x, \tag{3.35}$$

where the integral is taken over the whole complex plane $\mathbb{C}$. For each fixed value of $A_{m,\tau} = x$, the probability $\Pr\{|A_{j,k}| \leqslant |x|\}[n]$ can be evaluated for the complex Gaussian PDF $\mathcal{CN}(\hat{A}_{j,k}[n], \hat{\sigma}_A^2[n])$, employing the complementary first-order Marcum Q-function [PR06]

$$Q_1^{\text{c}}(a_{j,k}, b) = 1 - \int_b^\infty y e^{(-\frac{y^2+a_{j,k}^2}{2})} I_0(a_{j,k}y)dy\,, \tag{3.36}$$

where $a_{j,k} = |\hat{A}_{j,k}[n]|/\hat{\sigma}_A[n]$ and $b = |x|/\hat{\sigma}_A[n]$ with I_0 being the modified Bessel function of order zero.

Now that a quantitative measure is at hand, the elimination rule based on the sequential a-posteriori winning probabilities can be formulated as,

$$P_{m,\tau}^{\text{win}}[n] \underset{\text{eliminate}}{\overset{\text{keep}}{\gtrless}} P_{\text{elim}}\,, \tag{3.37}$$

where the probability P_{elim} acts as the elimination threshold. The interpretation is that, as soon as a candidate drops below a predefined minimum winning probability, it is eliminated from the competition. When all candidates corresponding to one beam are eliminated, no further observation via that beam is necessary. This will increase the efficiency both in terms of time and energy spent for beam selection.

We note that, the evaluation of the integrals that result to the winning probabilities $P_{m,\tau}^{\text{win}}[n]$, is exponentially complex and numerically cumbersome [FA14a]. However, the product within the integral in Eq. 3.35 is mostly dominated by one term. The dominating contribution from the Q-function that corresponds to the comparison with the leading competitor. The leading competitor A_{lead} is simply the beam with the largest current mean $\mu_{\text{lead}} = \hat{A}_{\text{lead}}[n]$

$$|\hat{A}_{\text{lead}}[n]| = \max_{(j,k)\in\mathbf{J}\setminus(m,\tau)} \{|\hat{A}_{j,k}[n]|\}\,. \tag{3.38}$$

This means that, there exists a simple upper bound[4] $P_{m,\tau}^{\text{win}}[n] \leqslant P_{m,\tau}^{\text{ub}}[n]$ which is expressed

[4] If there is only one close competitor while all other amplitude estimates are small, the bound will be tight. If the multiplicity of the closest competitor becomes larger than one, its tightness will be reduced accordingly.

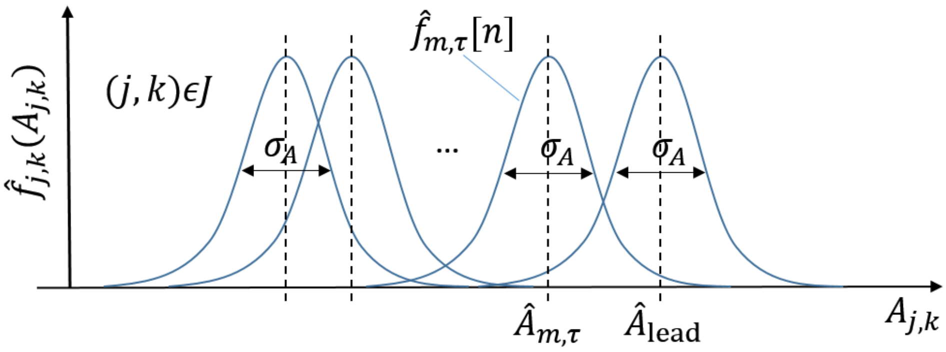


Figure 3.9: Schematic depiction of estimated a-posteriori PDFs.

as an integral of a Rician PDF times a Marcum Q-function as

$$P^{\text{ub}}_{m,\tau}[n] = \int_0^\infty \frac{r}{\hat{\sigma}^2_A[n]} e^{(-\frac{r^2+|\hat{A}_{m,\tau}[n]|^2}{2\hat{\sigma}^2_A[n]})} I_0(\frac{|\hat{A}_{m,\tau}[n]|r}{\hat{\sigma}^2_A[n]}) Q^{\text{c}}_1(\frac{|\hat{A}_{\text{lead}}[n]|}{\hat{\sigma}_A[n]}, r)\, \mathrm{d}r. \tag{3.39}$$

To get the intuition for simplifying the integral, the situation is schematically depicted (in 2D) in Fig. 3.9. The value of this upper bound is mainly dependent on three positive parameters, namely $|\hat{A}_{m,\tau}[n]|$, $|\hat{A}_{\text{lead}}[n]|$ and $\hat{\sigma}^2_A[n]$. Since the parameter $\hat{\sigma}^2_A[n]$ is common between both PDFs, the weighted distance between $|A_{m,\tau}[n]|$ and, $|\hat{A}_{\text{lead}}[n]|$ which is described by the deflection coefficient $\hat{d}^2_{m,\tau} = (|\hat{A}_{\text{lead}}[n]| - |\hat{A}_{m,\tau}|)^2/\sigma^2_A$ (a.k.a. effective SNR), characterizes the situation in reduced units.

Therefore, one can simplify the decision criterion by employing the equivalent elimination rule

$$\hat{d}^2_{m,\tau}[n] = \frac{(|\hat{A}_{\text{lead}}[n]| - |\hat{A}_{m,\tau}[n]|)^2}{\hat{\sigma}^2_A[n]} \underset{\text{keep}}{\overset{\text{eliminate}}{\gtrless}} d^2_{\text{elim}}\,, \tag{3.40}$$

where d^2_{elim} is the elimination threshold. It should be noted that, this elimination rule only makes sense if $|\hat{A}_{m,\tau}[n]| < |\hat{A}_{\text{lead}}[n]|$. This leads to the *S*equential *C*ompetition and *E*limination *T*est (SCET) stated with pseudocode in Algorithm 1.

It is worth noting that, when there remains a single (or a minimum predefined number of) survivor(s) the test will terminate by its nature and the surviving candidate will be declared as the winner. Therefore, beam elimination based on the probabilities of capturing the strongest amplitude (or rather not capturing it to decide for beam elimination) can be understood as a standalone beam selection technique. Combining it with the termination criterion, it works hand in hand with SCT.

The parameters d^2_{elim} and γ_{term} can be set to tune the algorithm and to balance probabilities between the reasons why the test finally stops, i.e., either by competition or elimination (an example of this is shown in Fig. 3.11). While the parameter $N_{\max}$ simply avoids long run length at low SNR values, the value of $N_{\min}$ is used to set a minimum number of initial observations to suppress premature elimination of strong signal amplitudes

Algorithm 1 Sequential Competition and Elimination Test (SCET)

input: $y_{m,\tau}[n], M, T, \gamma_{\text{term}}, d^2_{\text{elim}}, N_{\min}, N_{\max}$
initialization:
$\mathbf{J} \leftarrow \{(1,1), \ldots, (M,T)\}$
$n = N_{\min}$
$\bar{y}_{(m,\tau)\in\mathbf{J}}[N_{\min}] \leftarrow \sum_{i=1}^{N_{\min}} y_{m,\tau}[i]/N_{\min}$
$\hat{\sigma}^2[N_{\min}] \leftarrow \sum_{(m,\tau)\in\mathbf{J}} \sum_{i=1}^{N_{\min}} |y_{m,\tau}[i] - \bar{y}_{m,\tau}[N_{\min}]|^2/(|\mathbf{J}|N_{\min})$
$\hat{A}_{(m,\tau)\in\mathbf{J}}[0] \leftarrow 0, \hat{\sigma}^2_A[0] \leftarrow \hat{\sigma}^2[N_{\min}]$
SLMMSE $\rightarrow \hat{A}_{(m,\tau)\in\mathbf{J}}[N_{\min}], \sigma^2_A[N_{\min}]$
competition rule:
while $\max\limits_{(m,\tau)\in\mathbf{J}}(\gamma_{m,\tau}[n]) < \gamma_{\text{term}} \wedge |\mathbf{J}| \geqslant 2 \wedge n \leqslant N_{\max}$ **do**
    $n \leftarrow n + 1$
    $\bar{y}_{(m,\tau)\in\mathbf{J}}[n] \leftarrow \sum_{i=1}^{n} y_{m,\tau}[i]/n$
    $\hat{\sigma}^2[n] \leftarrow \sum_{(m,\tau)\in\mathbf{J}} \sum_{i=1}^{n} |y_{m,\tau}[i] - \bar{y}_{m,\tau}[n]|^2/(|\mathbf{J}|n)$
    GLLR:
    $\gamma_{(m,\tau)\in\mathbf{J}}[n] \leftarrow n \ln(1 + |\bar{y}_{(m,\tau)\in\mathbf{J}}[n]|^2/\hat{\sigma}^2[n])$
    SLMMSE:
    $K[n] \leftarrow \hat{\sigma}^2_A[n-1]/(\hat{\sigma}^2_A[n-1] + \hat{\sigma}^2[n])$
    $\hat{\sigma}^2_A[n] \leftarrow (1 - K[n])\hat{\sigma}^2_A[n-1]$
    $e_{(m,\tau)\in\mathbf{J}}[n] \leftarrow y_{(m,\tau)\in\mathbf{J}}[n] - \hat{A}_{(m,\tau)\in\mathbf{J}}[n-1]$
    $\hat{A}_{(m,\tau)\in\mathbf{J}}[n] \leftarrow \hat{A}_{(m,\tau)\in\mathbf{J}}[n-1] + K[n]e_{(m,\tau)\in\mathbf{J}}[n]$
    elimination rule:
    $|\hat{A}_{\text{lead}}[n]| \leftarrow \max\limits_{(j,k)\in\mathbf{J}\setminus(m,\tau)} |\hat{A}_{j,k}[n]|$
    $\hat{d}^2_{(m,\tau)\in\mathbf{J}}[n] \leftarrow (|\hat{A}_{(m,\tau)\in\mathbf{J}}[n]| - |\hat{A}_{\text{lead}}[n]|)^2/\hat{\sigma}^2_A[n]$
    if $\hat{d}^2_{(m,\tau)\in\mathbf{J}}[n] > d^2_{\text{elim}} \wedge |\hat{A}_{(m,\tau)\in\mathbf{J}}[n]| < |\hat{A}_{\text{lead}}[n]|$ **then**
        $\mathbf{J} \leftarrow \mathbf{J} - \{(m,\tau)\}$
    end if
end while
output: $\arg\max\limits_{(m,\tau)\in\mathbf{J}}(\gamma_{m,\tau}[n])$

due to the stronger fluctuations of the estimates in the early stages of the algorithm.

3.8 Numerical Evaluation

We evaluate the normalized average loss of signal magnitude $\bar{l}$ after beam selection as in Eq. (3.10) via a Monte Carlo simulation with 10^4 channel realizations. The efficiency of the beam selection with respect to the achieved accuracy is measured by the average total number of observations or average path length,

$$\bar{n}_{\text{tot}} = \mathbb{E}[\sum_{m=1}^{M} n_m^{\text{stop}}], \tag{3.41}$$

where n_m^{stop} indicates the number of observations made for the beam m before the test stops or the beam is eliminated. For the Fixed-length test, we have $\bar{n}_{\text{tot}} = M \times N_{\text{fix}}$. The quantities $\bar{l}$ and $\bar{n}_{\text{tot}}$ are evaluated for SNR values in the interval $[-9, 3]$ dB, which is the vital range in many practical applications.

We numerically studied the performance of the SCT augmented with the proposed elimination mechanism (SCET) in the reference data model described in Eq. (3.4) with a single path, with a uniform linear array with 64 antenna elements using the codebook of a Butler matrix with 64 normalized beams. Statistical performance of the proposed test is evaluated via a Monte Carlo simulation with 10^4 iterations. The AoA was distributed uniformly in $[-90°, 90°]$ over the simulation runs, while SNR after beamforming was defined as $(1^2/\sigma^2)$[dB]. Similar to $\bar{l}$ in Eq. 3.10, we evaluate the average relative effective rate $\mathbb{E}[R_{\text{eff}}/R_{\max}]$ after beam selection and average total number of observations as performance indicators, which are evaluated for SNR values in the interval $[-9, 3]$ dB.

For comparison, we consider the Fixed-length detector stated in Eq. 3.13 with $N^{\text{fix}} \in [75, 150]$, the pure SCT and the SCT augmented with elimination as proposed initially in [RM19]. Hyperparameters corresponding to different approaches has been chosen in a way that a similar range of performance in terms of average relative effective rate is achieved. As the comparison depicted in Fig. 3.10 shows, pure SCT is the most adaptive and robust in terms of performance, achieving above %99 of the maximum rate on average over the whole range of SNR values. Additionally, it is more efficient than the Fixed-length test at each N_{fix}. SCT augmented with the proposed elimination based on Sequentially Estimated a-Posteriori Probabilities (SaPPs) winning, increases the efficiency of the pure SCT by around two folds while maintaining the adaptivity and robustness of the pure SCT by achieving average effective rate of above %98 over a large range of SNR values. Besides, it outperforms significantly the previously proposed algorithm in [RM19] in both performance and efficiency while being more robust. Comparing the solid blue line (SCET) and the solid black line (Fixed-length test) at for example 0 to 3dB SNR, we observe that the number of measurements by the proposed sequential test is less than a half compared to the Fixed-length detector, while on average achieving even higher data rate. This points to the adaptivity and efficiency of the SCET.

Note that the value of the P_{elim} is heuristically set to 0.1. Choosing a large value for P_{elim} can result in higher probability of wrongful elimination of the true strongest candidate at early stages of the test and therefore hurting the performance of the pure SCT. Due to our numerical investigations, the choice of $P_{\text{elim}} \leqslant 0.1$ (roughly equal to $d^2_{\text{elim}} = 9$) results in a reasonable and robust performance in large range of SNR values, while reducing greatly the $\bar{n}_{\text{tot}}$ compared to pure SCT.

Our main goal by augmenting the SCT with a proper elimination mechanism was to increase the efficiency of the test without losing its desirable traits. Beam elimina-

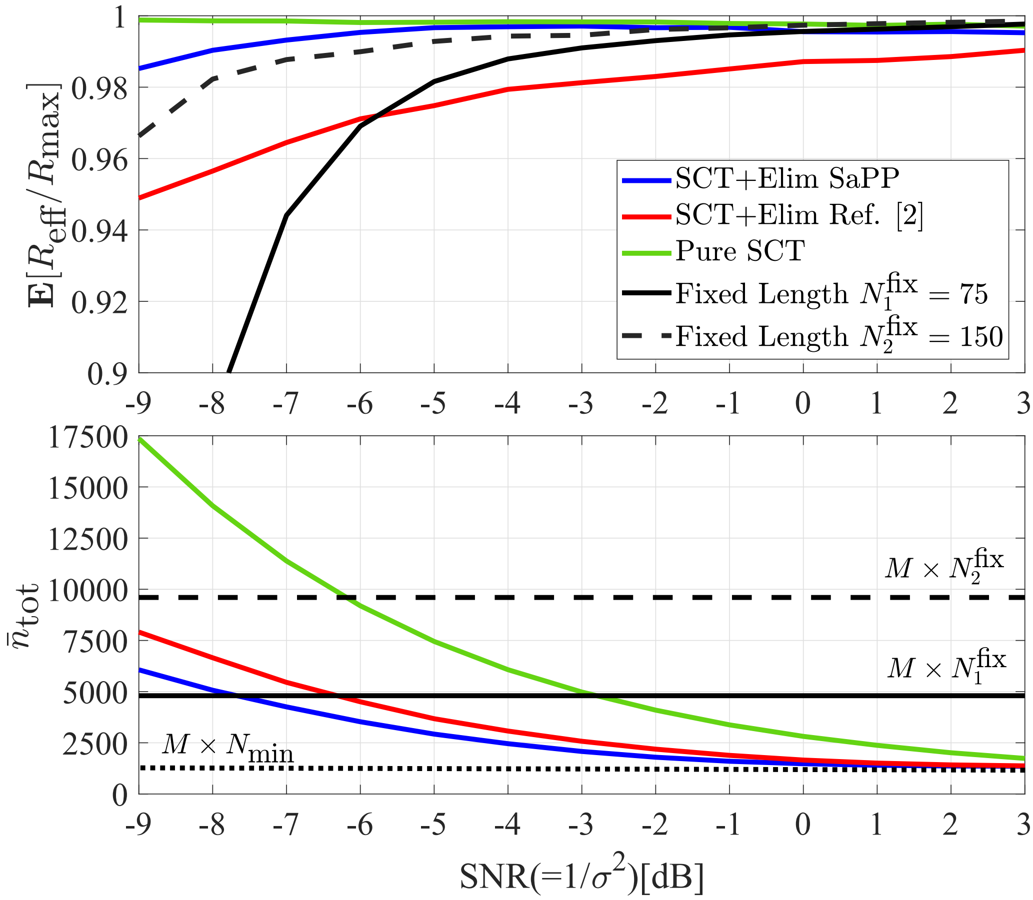


Figure 3.10: Performance Comparison using average relative achieved rate $\mathbb{E}[R_{\text{eff}}/R_{\text{max}}]$ and the corresponding average total number of observations $\bar{n}_{\text{tot}}$ for codebook of size $M = 64$ beams and $\gamma_{\text{term}} = 24 (\approx P_{\text{FA}} = 10^{-6})$. SCT+Elim SaPP uses the elimination probability of $P_{\text{elim}} = 0.1$, while SCT+Elim Ref.[2] ([RM19]) uses the fraction $\alpha = 0.5$. Initial minimum number of observations was set to $N_{\text{min}} = 20$. Limited coherence time is not considered in the rate evaluation.

tion, however, can also be considered as a standalone beam selection technique. When there remains a single (few) survivor(s) the test will terminate and the largest surviving candidate(s) will be declared as the winner. In our evaluations, the minimum number of survivors was set to 2. This means that when only two candidates survived the elimination process, the test will stop and the candidate with the larger GLLR value will be declared as the winner.

Fig. 3.11 illustrates the interplay between termination by competition and termination by elimination. As can be observed, the test stops more often by one candidate reaching the termination threshold first at lower SNR values. On the other hand, the test stops by having a single (or few) survivor(s) more often at higher SNR values. The two termination mechanisms are dominant at opposite ranges of the SNR, which demonstrates their desirable joint performance.

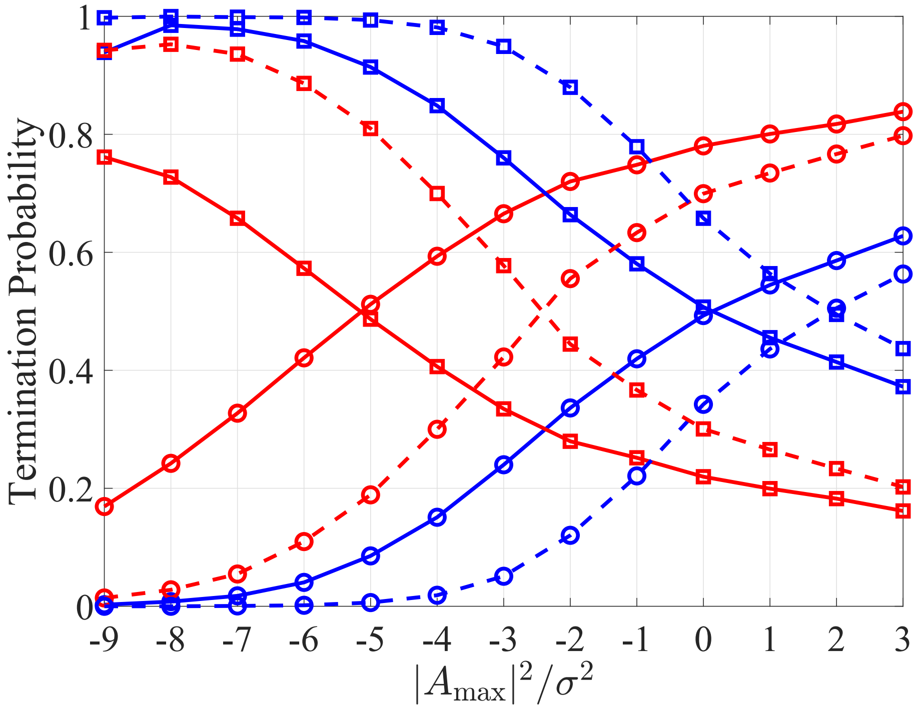


Figure 3.11: Comparing the probabilities of termination by competition (square markers) vs. termination by elimination (circle markers) over different SNR values in [dB]. Similar to Fig. 4.3, solid and dashed lines encode $P^{\mathrm{FA}}_{m,\tau} = 10^{-6}$ and $P^{\mathrm{FA}}_{m,\tau} = 10^{-3}$, while the colors blue and red represent $d^2_{\mathrm{elim}} = 9$ and $d^2_{\mathrm{elim}} = 4$, respectively.

3.9 Summary

In this chapter, we have shown that adaptivity to channel conditions during the beam selection phase, results in the rate gain and delay reduction. To achieve this, we proposed a novel MT-*ary* sequential hypothesis test based on GLR statistics to solve the composite multi-hypothesis beam selection problem with multiple sequences of observations. The proposed (pure) *Sequential Competition Test* (SCT) shows adaptivity w.r.t. the SNR operating point. Furthermore, it is more efficient compared to Fixed-length test, specially at lower SNR values where it matters the most or most time will be spent for training in a possible application. This provides the first variable-length framework for beam selection.

To improve efficiency even further, we extended the SCT to the Sequential Competition and Elimination Test (SCET) by adding a probabilistic elimination mechanism. For each beam, we estimate sequentially the a-posteriori PDF of the unknown signal amplitude using the SLMMSE estimator. Then, one can calculate the Bayesian winning probability for each candidate. This allows to discard those beams/hypotheses for which their probability drops below a predefined minimum value so that no further resources in terms of observations will be allocated to them for probing. As a result, efficiency both in terms of training time and energy is significantly increased when compared to sequential competition only. Both tests show advantageous features like adaptivity and robustness due to the strategy of using a composite hypothesis test that learns the SNR of the environment from the observations.

Therefore, these benefits can be of interest in systems with many candidate beams as in mmWave massive MIMO applications, as well as under conditions where the training time is limited due to small value of the channel coherence time.

Chapter 4

Application of Sequential Competition Test with Different Codebook Types

To this end, we learned that, the initial access step requires solving the beam alignment problem, i.e., finding the directional paths through which effective energy transfer is achievable. In this chapter, we consider the main practical codebook types which are typically used through the searching phase before data communication. These codebooks are based on the following ideas

1. Orthogonal Butler Matrix,
2. Hierarchical/Multi-Level Search,
3. Frequency Dependent Beamforming.

Furthermore, the effect of the variable- vs fixed-length detection in combination with different types of codebooks is evaluated, considering the training-transmission trade-off in channels with limited coherence interval. A fair comparison is intended throughout this study by applying the same power constraint on all scenarios. Our numerical results indicate that a variable-length test using either the frequency dependent codebook or a Butler codebook are the most promising methods achieving better efficiency, adaptivity, and robustness under varying SNR in practical scenarios. A hierarchical scheme with Fixed-length measurements shows lesser efficiency. It would only be advantageous if no power constraint would apply at lower levels. The main ideas in this chapter are presented in [KMJRG21].

4.1 Butler Codebook

A Butler matrix [But61] is a beamforming network employed to feed a phased array of antenna elements. Its purpose is to control the transmission directions corresponding to a set of M orthogonal beam patterns called Butler codebook. The orthogonal steering vectors $\mathbf{w}_m(\phi_m)$ with specific steering directions ϕ_m under each beam m in the azimuth plane are given by

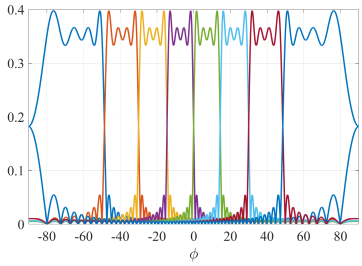
$$\mathbf{w}_m(\phi_m) = \frac{1}{\sqrt{M}}[1, e^{j\pi\sin(\phi_m)}, \ldots, e^{j\pi(M-1)\sin(\phi_m)}]^{\mathrm{T}} . \tag{4.1}$$

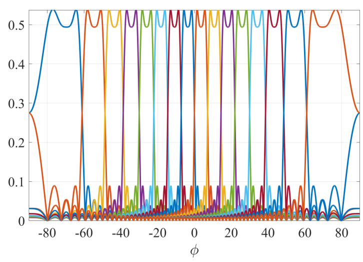
The detection performance after observing N_{fix} measurements per beam depends on the effective $\text{SNR}_{\text{eff}} = \frac{N_{\text{fix}} \times |A_{\max}|^2}{\sigma^2}$. The total time spent for beam selection in one step exhaustive search is $N_{\text{tot}}^{\text{fix}} = M \times N_{\text{fix}}$. This codebook might suffer from large number of combinations to measure if the switching from one beam to another is costly. This codebook type can particularly benefit from the elimination mechanism, since we have access to sequences observed via individual beams. This provides the chance to turn off unpromising scanning directions early. The application of the SCET with this codebook type in a single path channel has been discussed as a reference in the previous chapter.

4.2 Hierarchical Codebook

An alternative to the exhaustive search through the Butler codebook is to use the tree search approach. This is to reduce the number of tested beam combinations (less number of beam switching). This idea, introduces the design of the Hierarchical a.k.a. Multi-level (MLVL) codebooks, in which at lower levels of the search, wider beams are used and as the search proceeds to higher levels, the narrower beams are employed to find the exact direction. In this work, we use the codebook design in [NZL17b], to generate the beam patterns shown in the Fig. 4.1. In this technique, the beams in each level l are constructed via combining multiple individual beams from the Butler codebook as:

$$\begin{aligned}
\mathbf{W}^{(l)} \quad &= \quad [\mathbf{w}_1^{(l)}(\phi_1), \mathbf{w}_2^{(l)}(\phi_2), \ldots, \mathbf{w}_{N_t/N_l}^{(l)}(\phi_{N_t/N_l})] \\
&= \quad \frac{1}{N_l}[\sum_{m=1}^{N_l} \mathbf{w}_m(\phi_m)e^{jm\omega_l}, \sum_{m=N_l+1}^{2N_l} \mathbf{w}_m(\phi_m) \\
e^{jm\omega_l} \quad &, \ldots, \sum_{m=N_t-N_l+1}^{N_t} \mathbf{w}_m(\phi_m)e^{jm\omega_l}] ,
\end{aligned} \tag{4.2}$$


(a) First level $M^{(1)} = 8$.


(b) Second level $M^{(2)} = 2$.

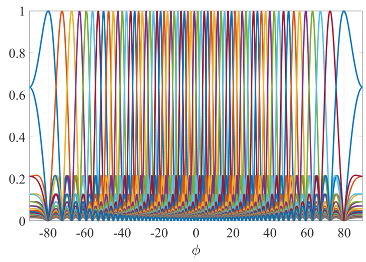

(c) Third level $M^{(3)} = 4$.

Figure 4.1: Hierarchical codebook (MLVL) design with $\omega_1 = 1.93, \omega_2 = 2.24$ and number of antennas $N_t = 64$ and $N_1 = 8, N_2 = 4, N_3 = 1$.

where N_l is the number of combined orthogonal beams to construct each beam at the level l. The optimization is done to find the parameters ω_l which satisfy

$$\min_{\omega_l \in [\frac{-\pi}{N_t}, \pi(1-\frac{1}{N_t})]} \operatorname{var}(|\mathbf{a}(\phi)\mathbf{w}^{(l)}(\phi_m)|) \ . \tag{4.3}$$

After observing $N_{\text{fix}}^{(l)}$ measurements per beam, the detection performance at level l depends on the effective $\text{SNR}_{\text{eff}}^{(l)} = \frac{N_{\text{fix}}^{(l)} \prod_{i=1}^{l} M^{(i)} |A_{\max}|^2}{N_t \times \sigma^2}$, where $M^{(l)}$ represents the number of candidate beams at level l within the beam-width of the beam selected from previous level. The total time spent for beam selection in one step exhaustive search is $N_{\text{tot}}^{\text{fix}} = \sum_{l=1}^{L} M^{(l)} \times N_{\text{fix}}^{(l)}$, L indicating the number of levels.

The SCET algorithm is simply applicable to this codebook type. It can solve the multi-hypothesis problem at each level (with different sizes) adaptively and therefore achieving a better performance compared to Fixed-length test. However, the elimination test is less effective in increasing the efficiency of the search due to smaller size of the codebook at early stages.

4.3 Frequency Dependent Codebook

A one-step beam probing procedure which does not require switching between various orthogonal beam patterns (this is an advantage which we do not quantify in the following because only somewhat arbitrary assumptions are possible about the switching cost.) has been introduced by C. Jans et al. in [JSRF20]. This technique parallelizes the task of finding the beam direction during the beam alignment phase with use of only one RF-chain. This can be achieved via exploiting the idea of a frequency dependent beamforming. In this technique, a wideband signal $s(t)$ is delayed at the M antennas relative to a reference antenna by multiples of the reciprocal bandwidth or the Nyquist sampling interval $T_s = 1/f_b$ where f_b is the bandwidth. This way, a one-to-one mapping of temporal frequencies to spatial frequencies (see Fig.4.2) is obtained. Assuming τ being k multiples of

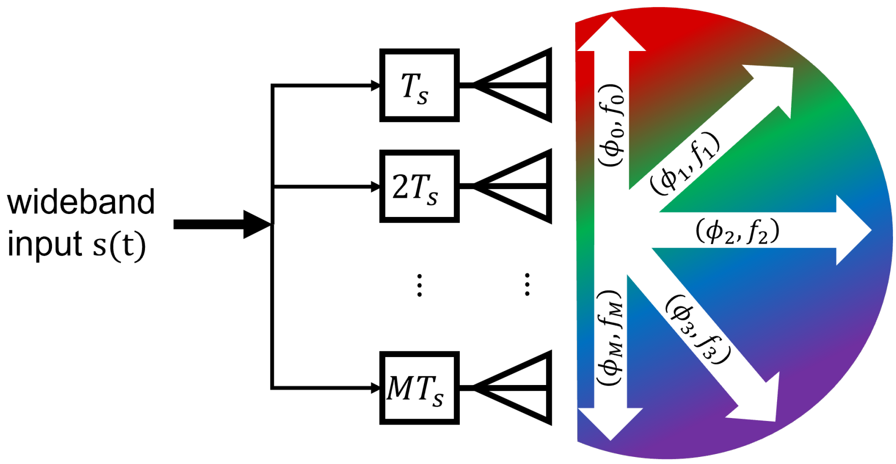


Figure 4.2: Schematic: Frequency scanning/dependent beamforming; one-to-one mapping between frequency in baseband and steering angle. (f_i, ϕ_i)

the sampling time T_s as $k = \lfloor \frac{\tau}{T_s} \rfloor$, $\tau_m^p = d/c \sin\phi$ being the common propagation delay for a ULA with d being the antenna element spacing and c being the speed of light and complex Gaussian noise $z[n] \sim \mathcal{CN}(0, \sigma_z^2)$, the received sequence is obtained within the period $0 \leqslant n < M$, where $M = \lfloor T/T_s \rfloor$ as

$$r[n] = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} s[n-m-k] e^{-j2\pi f_c (mT_s + \tau_m^p + \tau)} + z[n]. \tag{4.4}$$

$s[n]$ repeats periodically after M samples. After observing N samples, one and can express the training signal using a circulant matrix $\mathbf{S}^{M\times N}$ obtained from $\Big[s[0], \dots, s[M-1]\Big]^T$ for the transmit signal and use permutation matrix $\mathbf{\Pi}^{M\times M}$ to model the timing shift kT_s. Finally, we get the row vector for all observed receive samples

$$\mathbf{r} = \frac{1}{\sqrt{M}} [1, \dots, e^{-j\pi(M-1)\sin\phi}] \mathbf{D}^{M\times M} \mathbf{\Pi}^{M\times M} \mathbf{S}^{M\times N} + \mathbf{z} \tag{4.5}$$

with $\mathbf{D}$ being a strictly diagonal matrix with entries $d_{i,i} = e^{-j2\pi f_c \tau} e^{-j2\pi f_c i T_s}$, $i = 0, \dots, M-1$.

Observing that $\mathbf{a}^T(\phi) = \frac{1}{\sqrt{M}}[1, \dots, e^{-j\pi(M-1)\sin\phi}]$, C. Jans showed in [KMJRG21] that the coupling of array steering vector and receive samples $\mathbf{r} = [r[0], \dots, r[N-1]]$ can be obtained via

$$\mathbf{r} = [A_0, \dots, A_{M-1}] \tilde{\mathbf{D}} \mathrm{diag}(\mathbf{b}^T \mathcal{F}) \tilde{\mathbf{S}} + \mathbf{z}, \tag{4.6}$$

with unknown coupling coefficients $[A_0, \dots, A_{M-1}] = \mathbf{a}^T(\phi) \frac{1}{\sqrt{M}} \mathcal{F}$, an unknown but strictly diagonal matrix $\tilde{\mathbf{D}} = \frac{1}{M} \mathcal{F}^H \mathbf{D} \mathcal{F}$, and a known inverse Fourier transformed matrix $\tilde{\mathbf{S}} = \frac{1}{\sqrt{M}} \mathcal{F}^H \mathbf{S}$ and finally $\mathbf{b}$ being an all-zero vector having just one 1 at index $\mod(k, M)$.

4.3.1 Fixed-length Test

As our detection method is based on the power spectrum, we only need to derive an estimate on the absolute values of $[A_0, \ldots, A_{M-1}]$ and, therefore, the diagonal matrices $\tilde{\mathbf{D}}$ as well as diag($\mathbf{b}^T\mathcal{F}$), which act as phase rotations, can be neglected and the least-squares (LS) estimator can be formulated as

$$[|\hat{A}_0|, \ldots, |\hat{A}_{M-1}|] = |\mathbf{r}\tilde{\mathbf{S}}^H(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H)^{-1}|. \tag{4.7}$$

These estimates can be used to detect the strongest beamforming direction using a max detector. The effective detection SNR in this case after observing N_{fix} measurements amounts to $\text{SNR}_{\text{eff}} = \frac{N_{\text{fix}} \times |A_{\max}|^2}{N_t \times \sigma^2}$. The total time spent for beam selection in one step exhaustive search is $N^{\text{fix}}_{\text{tot}} = N_{\text{fix}}$.

4.3.2 Variable-length Test: FD-SCT

SCT can use the estimated coefficients $[|\hat{A}_0|^2, \ldots, |\hat{A}_{M-1}|^2]$ and noise variance $\hat{\sigma}_z^2$ at each step n. Both can be derived by iteratively calculating the LS solution following (4.7) and, secondly, deriving estimates on the noise samples

$$\hat{\mathbf{z}} = \mathbf{r} - \mathbf{r}\tilde{\mathbf{S}}^H(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H)^{-1}\tilde{\mathbf{S}} = \mathbf{r}(I - \tilde{\mathbf{S}}^H(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H)^{-1}\tilde{\mathbf{S}}) \tag{4.8}$$

which results to the estimate on the noise variance by $\hat{\sigma}_z^2 = \frac{1}{N}\hat{\mathbf{z}}\hat{\mathbf{z}}^H$. The candidate $\arg\max_{i=0,\ldots,M-1} |\hat{A}_i|$ which surpasses the threshold $\gamma_{\text{term}} = (\mathcal{Q}^{-1}(\frac{P^{\text{FA}}}{2}))^2$ first, meaning,

$$\max_{i=0,\ldots,M-1} \frac{n}{M} \log(1 + \frac{|\hat{A}_i|^2}{\hat{\sigma}_z^2}) > \gamma_{\text{term}}, \tag{4.9}$$

is the winner and corresponds to the optimal beamformer from the orthogonal codebook after n observations. Collecting the absolute minimum number of $n \geqslant M$ samples, which are needed, such that the LS solution in (4.7) exists. the termination of the test occurs at any given time in presence of favorable noise realizations.

4.4 Numerical Evaluation

We numerically studied the performance of different codebooks and detection frameworks in the reference data model described in Eq. (3.4). We assume a channel with $P = 3$ paths given by ($\alpha_1 = 1$,$\eta_1 = 100$), ($\alpha_2 = 0.6$,$\eta_2 = 10$) and ($\alpha_3 = 0.6$,$\eta_3 = 0$), where α_p denotes the scatterer strength, η_p indicates the strength ratio between the LOS and the NLOS propagation. Path 1 can be roughly regarded as the LOS path, while the remaining paths/scatterers represent the NLOS contributions. This is consistent with the practical mm-wave channel measurements in [SR16a] [HCGL18] [RXK+19a], where the relative power levels of the NLOS paths are around 10 dB lower than the desired LOS path. The AoAs ϕ_p were distributed uniformly in $\phi_p \in [-90°, 90°]$, while delays τ_p were randomly drawn from $\tau_p \in \{0, 1, \ldots, T_{\max} = 30\}$ over each channel realization. The receiver is equipped with a uniform linear array with 64 antenna elements. The SNR after beamforming in each realization was defined as $|A_{\max}|^2/\sigma^2$ [dB].

4.4.1 Fixed vs. variable-length Tests with a Butler Codebook

We consider the pure SCT [MRF19] without elimination and the SCT augmented with elimination (SCET) with a PN sequence of length $N_{\max} = 255$ which is observed sequentially in time. The termination threshold γ_{term} is set based on Eq. 3.25 with $P^{\text{FA}}_{m,\tau} \in \{10^{-3}, 10^{-6}\}$. The elimination threshold is set as $d^2_{\text{elim}} \in \{4, 9\}$. The hyperparameters corresponding to different approaches have been chosen in a way that a similar range of performance in terms of average relative effective rate is achieved. For comparison, we consider the Fixed-length detector based on magnitude estimates [BHR+15] stated in Eq. (3.13) with maximum length PN sequences with $N_{\text{fix}} \in \{31, 63, 127\}$.

As the comparison depicted in Fig. 4.3 shows, the pure SCT with $\{P^{\text{FA}}_{m,\tau} = 10^{-6}\}$ (solid black line) is adaptive and robust in terms of performance. The test length in SCT adaptively changes over the whole range of SNR values and hence $\bar{l}$ stays within $[0.015, 0.025]$ on average. SCT augmented with the proposed elimination based on sequentially estimated a-posteriori winning probabilities, i.e., SCET with $\{P^{\text{FA}}_{m,\tau} = 10^{-6}, d^2_{\text{elim}} = 9\}$ (solid blue line), increases the efficiency by a factor of approximately 1.5-2, while maintaining the adaptivity and robustness of the pure SCT at the cost of a mere 0.01 performance loss in terms of $\bar{l}$. Additionally, by comparing the $\bar{n}_{\text{tot}}$ for SNR values at which the loss $\bar{l}$ is the same, the SCET with $\{P^{\text{FA}}_{m,\tau} = 10^{-6}, d^2_{\text{elim}} = 9\}$ is more efficient than the Fixed-length test for all N_{fix}.

Relaxing the elimination rule, SCET with $\{P^{\text{FA}}_{m,\tau} = 10^{-6}, d^2_{\text{elim}} = 4\}$ (solid red line) results in a few percent higher loss in the lower SNR regime, while reducing further the total number of observations over the whole SNR range. Note that the value of d^2_{elim} is heuristically set, and choosing a small value for d^2_{elim} can result in higher probability of erroneous elimination of the true strongest candidate at early stages of the test. This effect can degrade the performance of SCET w.r.t. pure SCT so that it has to be chosen with care.

Relaxing the competition rule by setting the termination threshold based on $P^{\text{FA}}_{m,\tau} = 10^{-3}$, shortens the test length at the cost of increasing the loss by few percent over the whole SNR range. This is depicted by the dashed lines compared to solid lines of the same

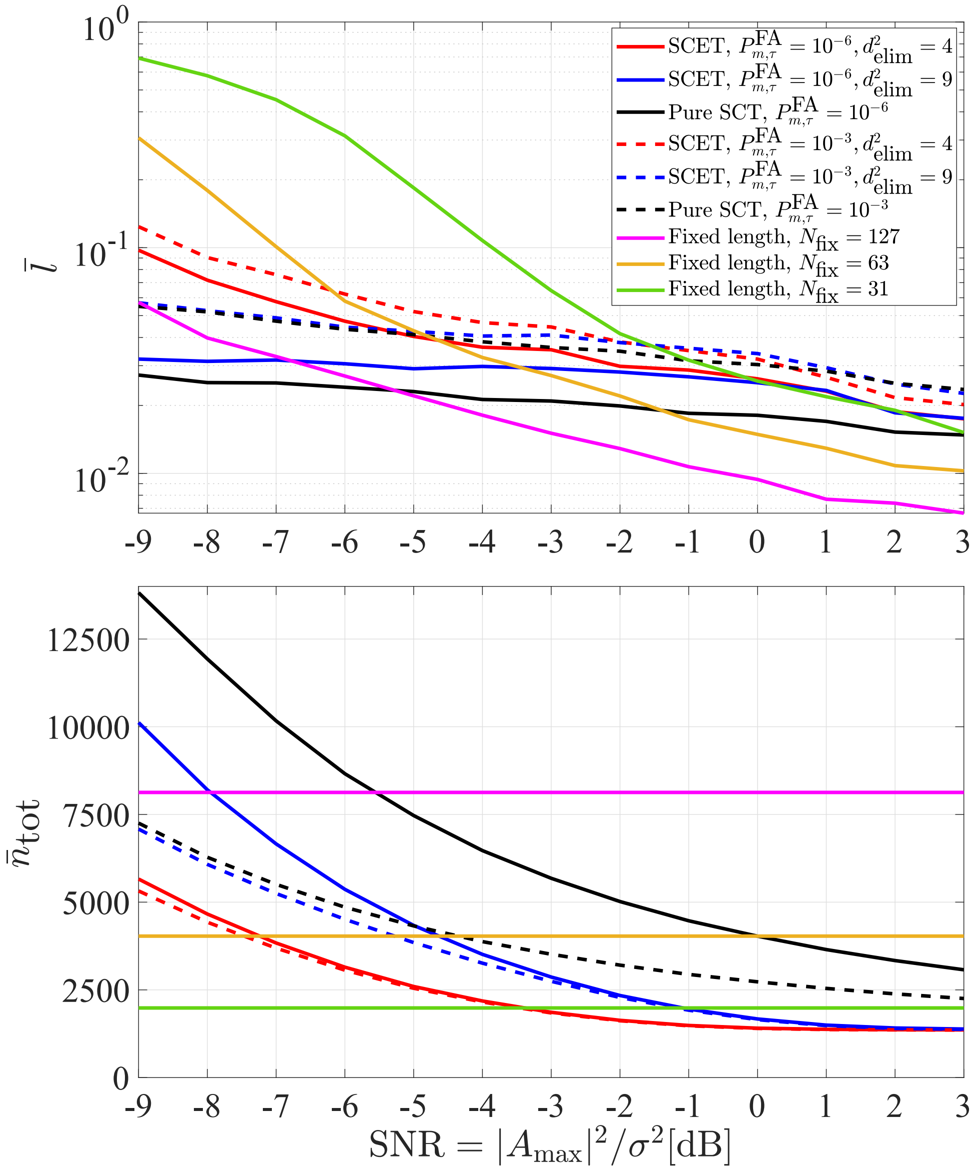
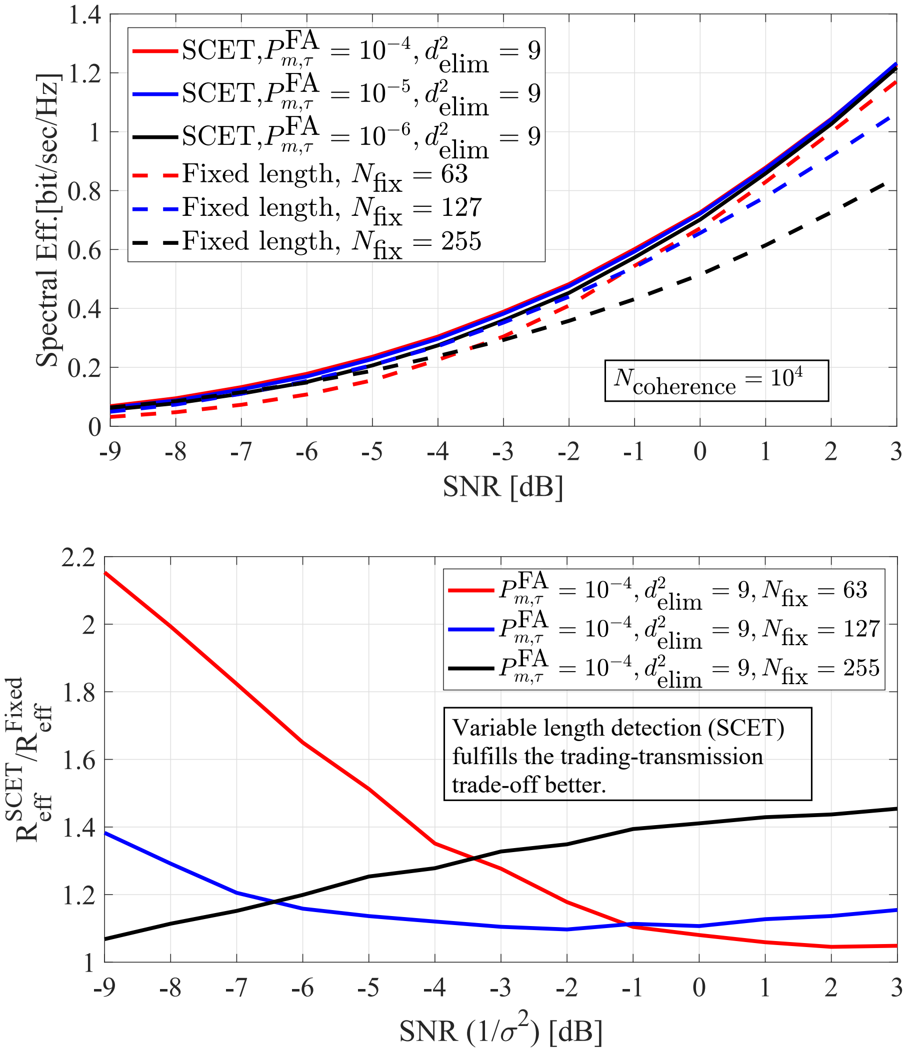


Figure 4.3: Comparing the performance of fixed and variable-length beam selection techniques while using a Butler codebook in a multi-path channel.

color.

4.4.2 Fixed vs. variable-length Tests with a Hierarchical Codebook

Next, we investigate the hierarchical (MLVL) codebook while using different detection frameworks. The codebook considered is depicted in Fig 4.1 which consists of three levels. The parameter of interest in this investigation is the achieved spectral efficiency or effective rate when coherence time is limited and equal t0 N_{coh}. As we learned previously, there exist a training-transmission trade which can be evaluated using Eq. 3.29. This trade-off suggests that for $N_{\max} = N_{\mathrm{coh}}$ possible channel uses that can be spent for training and



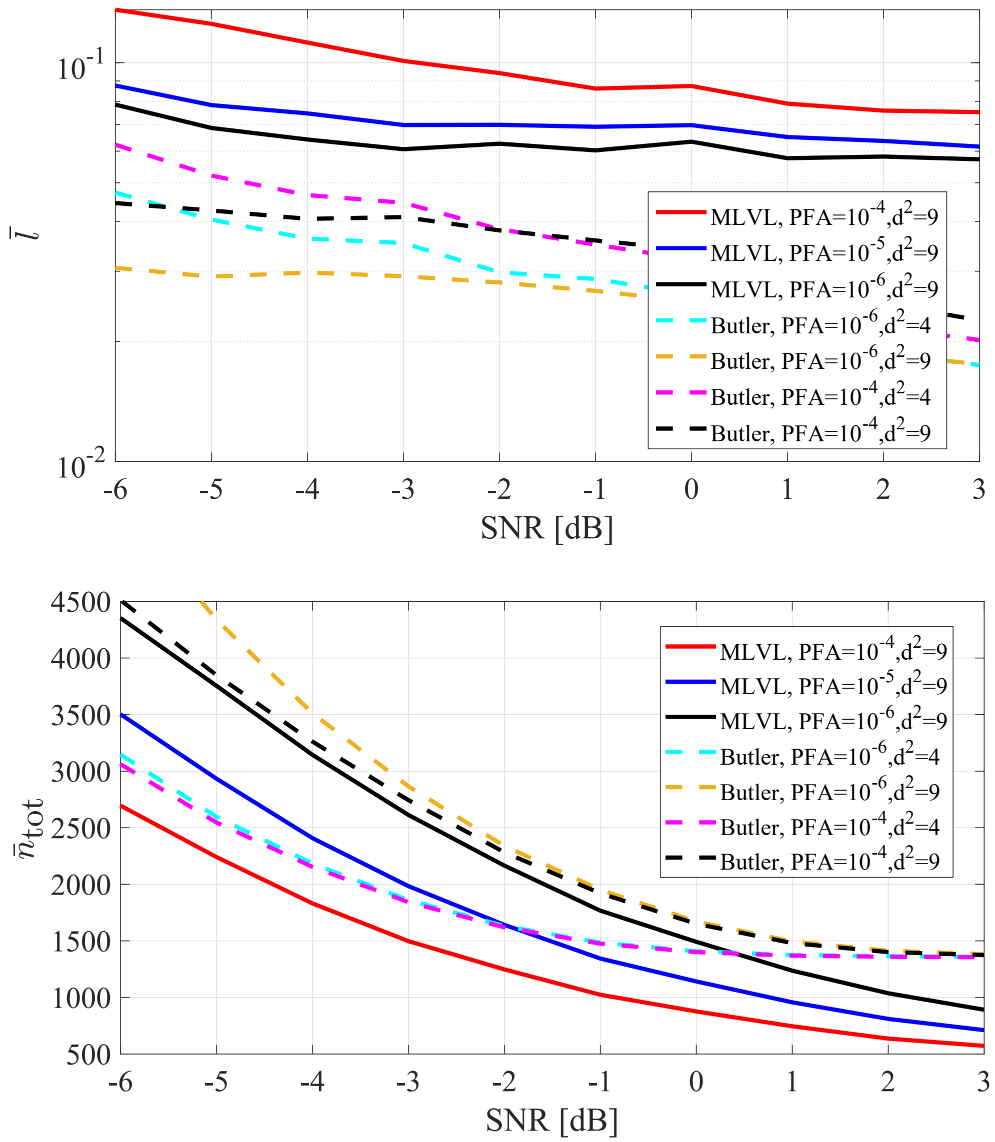
(b) Ratio of the achieved rate via variable-length detection over Fixed-length detection

Figure 4.4: Comparing the performance of fixed and variable-length detection while using a hierarchical codebook.

data transmission, one can evaluate the ratio between the effective data rates achieved by the sequential (variable-length) and Fixed-length tests. Thus, we consider the ratio

$$\frac{R_{\text{eff,seq}}}{R_{\text{eff,fix}}} = \frac{\mathbb{E}\left[(1 - \frac{N^{\text{seq}}_{\text{tot}}}{N_{\text{coh}}}) \log(1 + (1 - l_{\text{seq}})^2 \frac{|A_{\max}|^2}{\sigma^2})\right]}{\mathbb{E}\left[(1 - \frac{N^{\text{fix}}_{\text{tot}}}{N_{\text{coh}}}) \log(1 + (1 - l_{\text{fix}})^2 \frac{|A_{\max}|^2}{\sigma^2})\right]}, \tag{4.10}$$

This trade-off is evaluated in Fig. 4.4 while considering a coherence interval of size $N_{\text{coh}} = 10^4$ samples. As depicted in Fig 4.4, the variable-length detection framework using SCET fulfills the training-transmission trade-off better by achieving a ratio above one over the whole SNR range. This means adaptivity to channel conditions can increase the rate and/or reduce the delay to detection while using Hierarchical codebook as in 802.11.ad.



(b) Total number of observations

Figure 4.5: Comparison between Hierarchical (MLVL) and the Butler codebooks while using SCET.

4.4.3 Hierarchical Codebook vs. Butler Codebook using SCET

The next comparison is done between the Butler codebook of size $M = 64$ and the comparable hierarchical (MLVL) codebook while employing SCET as the detection algorithm. As visible in Fig. 4.5 the Butler codebook shows a better accuracy and efficiency with the right parametrization. For instance, for the same accuracy the SCET+Butler requires on average less number of observations in almost all SNR values. The inferior efficiency of the MLVL codebook can be associated to the lower detection SNR at earlier levels. If the cost due to switching the beams is not negligible, then MLVL codebook can show higher efficiency due to lesser number of switching required.

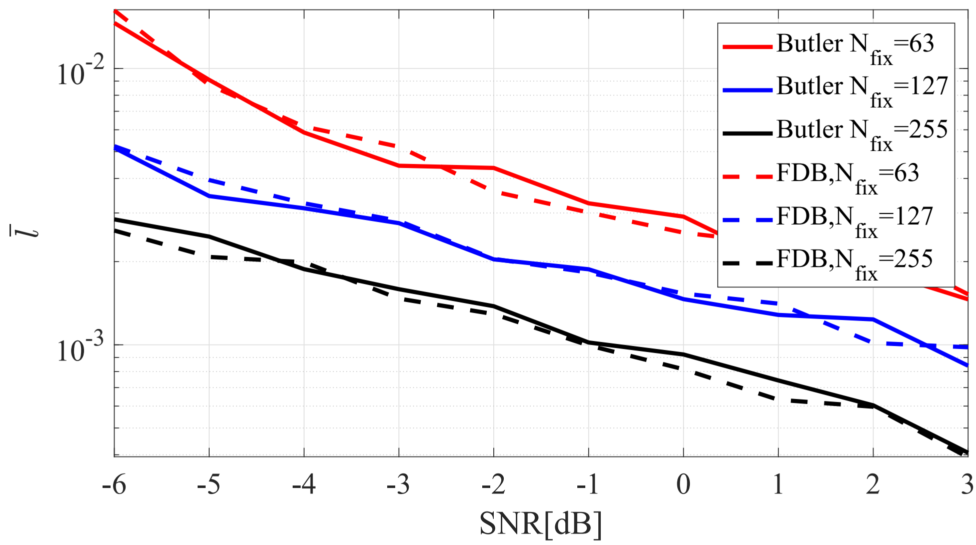


Figure 4.6: Comparison between FDB codebook and the Butler codebook while using the Fixed-length test. Simulations have been provided by C. Jans.

4.4.4 Frequency Dependent Codebook Evaluation

Finally, we investigate the FDB codebook. For the first comparison, we consider a LOS channel with one dominant path and employ a Fixed-length detector. As shown in Fig. 4.6, the FDB codebook achieves almost the same accuracy and efficiency of detection compared to Butler codebook. This means, the FDB has an advantage with one-step channel measurement without a need to measure multiple beams separately.

Next, the FDB codebook can also benefit from the variable-length detection technique called SCT. As depicted in Fig. 4.7, FDB codebook shows the same desirable adaptive performance, while outperforming the Butler codebook, reducing the acquisition time by few percents. This is due to the more elaborate data model which considers the combined observations from different beam directions, which can result in more accurate amplitude estimates.

4.5 Summary

variable-length detection via the Sequential Competition and Elimination Test (SCET) provides adaptivity to channel conditions. This adaptivity during the beam acquisition phase results in a rate gain and/or a delay reduction by better fulfilling the training transmission trade-off. Use of different codebook types can affect the detection performance: An orthogonal Butler codebook can benefit from the elimination mechanism to increase the detection efficiency. However, the performance might suffer when the settling time for the phase shifters is large or switching time of the electronic switches is long. The frequency dependent codebook removes the need for switching the beams (during the ex-

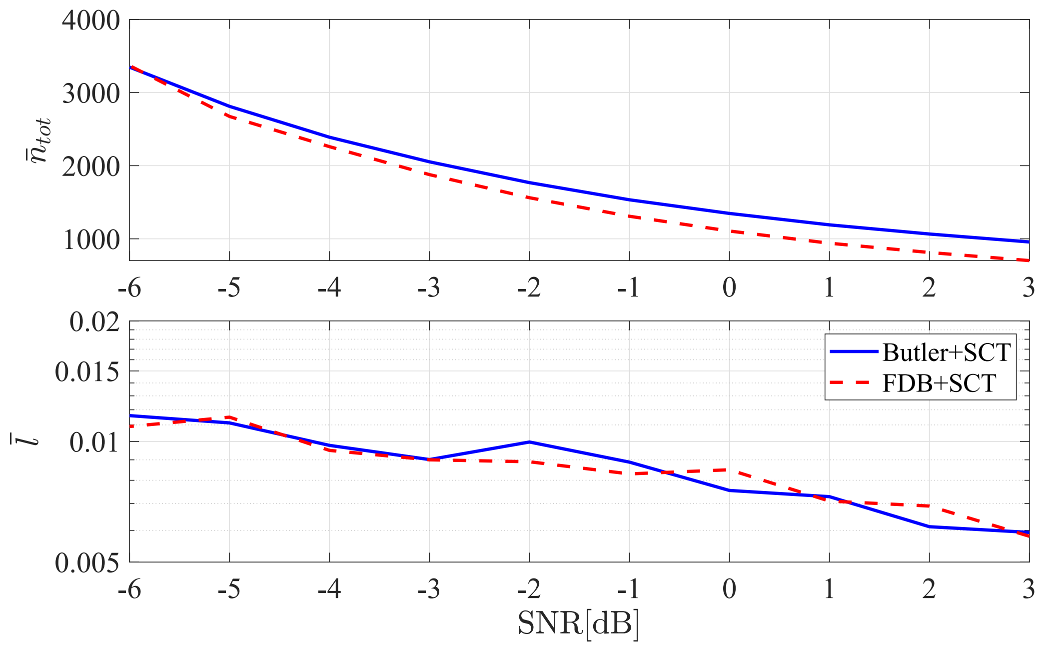


Figure 4.7: Comparison between FDB codebook and the Butler codebook while using the sequential competition test. Simulations have been provided by C. Jans.

haustive search) without any penalty. This codebook type is specially powerful when used at the base station, providing one step downlink training while employing variable-length detection based on SCT. Hierarchical codebooks show lesser detection efficiency, due to lower SNR at initial levels and the error propagation through the tree search. The use of variable-length detection based on SCET can improve the detection performance while using the hierarchical codebooks proposed in 802.11.ad.

Chapter 5

Multi-user Beam Selection and Digital Beamforming

To serve multiple users via the same time and frequency resource, the users must be separated spatially. The crucial question at the receiver side, showed schematically in Fig 5.1, is how to choose the best beam(s) for each user from a given codebook such that the overall SINR is maximized. The conventional approach [SB13] is to estimate the captured power of each user under each beam using a training sequence of *fixed* length, and then to choose the beam for each user with the highest SINR estimate. To use this method efficiently, i.e., not with a much larger/smaller number of measurements than necessary to achieve the desired performance level, knowledge of the PDFs of all estimates would be required. However, this knowledge is not or only roughly available [SHC18] to the detector due to varying operation conditions under which this problem needs to be solved. In this chapter, the multi-user sequential competition test is explained. Finally, we discuss the practical solution to the joint analog digital beamforming optimization problem in the multi-user setup to improve the sum rate performance of the MU-MIMO system. The main ideas in this chapter regarding the multi-user SCT is discussed in [KMRF19], while some ideas regarding digital beamforming are presented in [CCR+21].

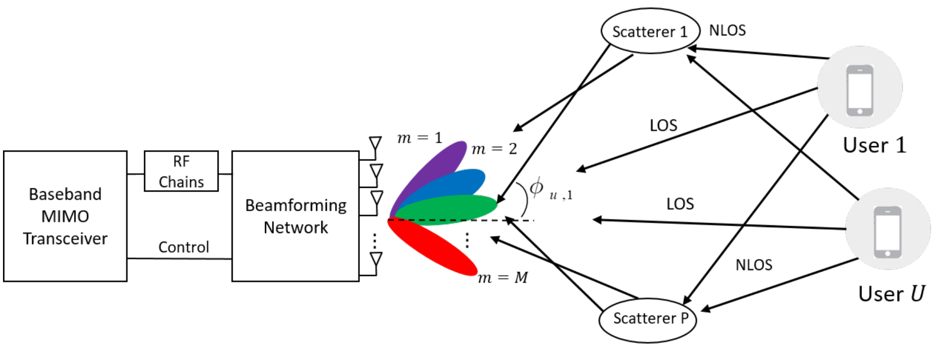


Figure 5.1: MU-MIMO transceiver using an electronically controllable beamforming network in a multi-scatterer environment.

5.1 Multi-user Beam Selection Problem

Consider the problem of allocating beams from an orthogonal codebook to different users such that they are spatially separated as much as possible. The complex valued received samples are observed separately under each candidate beam as

$$
\begin{aligned}
r_m[n] &= \sum_{u=1}^{U}\sum_{p=1}^{P_u} \underbrace{\rho_{u,p}\boldsymbol{w}(\phi_m)\boldsymbol{a}(\phi_{u,p})}_{A_{u,m,p}} s_u[n-\tau_{u,p}] \\
&+ \underbrace{\boldsymbol{w}(\phi_m)\boldsymbol{z}'[n]}_{z_m[n]} ,
\end{aligned}
$$

where $m \in \{1,\dots,M\}$, $n \in \{1,\dots,N\}$ and $u \in \{1,\dots,U\}$ indicate the beam, sample and user indices. The parameter $A_{u,m,p}$ denotes the combined effective channel and beamforming gain corresponding to user u under beam m and path p, which is treated as a deterministic unknown complex amplitude. $z_m[n]$ is a complex zero mean WGN sample with unknown variance σ^2 under beam m. A PN sequence with $s_u[n] \in \{\pm 1\}$, variance one and $P\{s_u[n] = +1\} = P\{s[n] = -1\} = 1/2$ is assumed for training of each user so that $\mathbb{E}[s_u[n]s_u[n-k]] \simeq \delta[k]$ holds for its auto-correlation sequence, while the cross-correlation between different sequences corresponding to different users is $\mathbb{E}[s_u[n]s_{u'}[n-k]] \simeq 0$ when $u \neq u'$.

Assuming a flat rank one channel[1] (i.e., a single path) for each user in the uplink, in order to allocate the best beam to each user during the training interval, the receiver should evaluate the SINR value $\beta_{u,m}$ in uplink for each user u under each beam m as

$$
\hat{\beta}_{u,m} = \frac{|\hat{A}_{u,m}|^2}{\sum_{i=1,i\neq u}^{U} |\hat{A}_{i,m}|^2 + \hat{\sigma}^2} , \tag{5.1}
$$

where the hat denotes the estimated value. Using the SINR estimates $\hat{\beta}_{u,m}$, the most promising (u,m) pairs can be decided for via exhaustive search, which involves prohibitively high complexity with $\binom{M}{U}$ searches. On the other hand, the accuracy of the Minimum Variance Unbiased Estimators [Kayb] of the parameters $A_{u,m}$ and σ^2 based on training sequences of Fixed-length N, depends on the SNR operating point at which these values are estimated. However, this information is hidden from the receiver. Therefore, fixing the test length (i.e., sequence length) can lead to drastically changing performance when the information about the SNR operating point of each user is not available. Ad-

[1] Although the channel model for mmWave systems contain features like multipath with multiple delays and angle of arrivals, we believe this simplistic channel model suffices for analyzing the optimal detection strategy.

ditionally, if N is conservatively set to a high value based on the worst still acceptable operating point, a lot of time that is spent for detection of the best beam for each user, will be wasted, if the channel quality is actually better than expected. This is particularly important when the channel coherence time is limited and wasting time for training results in loss of throughput. These problems can be potentially avoided by using an adaptive variable-length test.

We explain in the next section the GLRT for the classical linear model, and will use this result to introduce our *multi-user sequential competition test* in Section 5.3.

5.2 Generalized Likelihood Ratio Test for A Linear Model with Unknown Amplitudes and Noise Variance

Consider that the data follows the linear model $\boldsymbol{r} = \boldsymbol{SA} + \boldsymbol{w}$, where $\boldsymbol{S}$ is a known $N \times U$ $(N > U)$ matrix of rank U, $\boldsymbol{A}$ is a $U \times 1$ vector of unknown parameters, and $\boldsymbol{w}$ is a $N \times 1$ noise vector with PDF $\mathcal{N} \sim (\mathbf{0}, \sigma^2 \boldsymbol{I})$ while σ^2 is unknown.

The Generalized Likelihood Ratio (GLR) for the hypothesis testing problem

$$\begin{array}{ll} \mathcal{H}_0 & : \boldsymbol{BA} = \boldsymbol{b},\ \sigma^2 > 0 \\ \mathcal{H}_1 & : \boldsymbol{BA} \neq \boldsymbol{b},\ \sigma^2 > 0 \end{array} , \tag{5.2}$$

where $\boldsymbol{B}$ is a $K \times U$ matrix $(K \leqslant U)$ of rank K, $\boldsymbol{b}$ is a $K \times 1$ vector, can be written as

$$L_G(\boldsymbol{r}) = \frac{p\left(\boldsymbol{r}; \hat{\boldsymbol{A}}_{\mathcal{H}_1}, \hat{\sigma}^2_{\mathcal{H}_1}\right)}{p\left(\boldsymbol{r}; \hat{\boldsymbol{A}}_{\mathcal{H}_0}, \hat{\sigma}^2_{\mathcal{H}_0}\right)} = \left(\frac{\hat{\sigma}^2_{\mathcal{H}_1}}{\hat{\sigma}^2_{\mathcal{H}_0}}\right)^{N/2} , \tag{5.3}$$

where

$$\begin{aligned} \hat{\boldsymbol{A}}_{\mathcal{H}_1} &= (\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{r} \\ \hat{\sigma}^2_{\mathcal{H}_1} &= (\boldsymbol{r} - \boldsymbol{S}\hat{\boldsymbol{A}}_{\mathcal{H}_1})^{\mathrm{H}}(\boldsymbol{r} - \boldsymbol{S}\hat{\boldsymbol{A}}_{\mathcal{H}_1})/N \\ \hat{\boldsymbol{A}}_{\mathcal{H}_0} &= \hat{\boldsymbol{A}}_{\mathcal{H}_1} - (\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1}\boldsymbol{B}^{\mathrm{H}}(\boldsymbol{B}(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})\boldsymbol{B}^{\mathrm{H}})^{-1}(\boldsymbol{B}\hat{\boldsymbol{A}}_{\mathcal{H}_1} - \boldsymbol{b}) \\ \hat{\sigma}^2_{\mathcal{H}_0} &= (\boldsymbol{r} - \boldsymbol{S}\hat{\boldsymbol{A}}_{\mathcal{H}_0})^{\mathrm{H}}(\boldsymbol{r} - \boldsymbol{S}\hat{\boldsymbol{A}}_{\mathcal{H}_0})/N , \end{aligned}$$

are ML estimates of $\boldsymbol{A}$ and σ^2 under $\mathcal{H}_1$ and $\mathcal{H}_0$ respectively, while $(.)^{\mathrm{H}}$ denotes conjugate transpose operation.

It can be shown [Kayb] that the modified GLR statistic $\gamma(\boldsymbol{r})$ follows the central and non-central F distribution under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively, as

$$\gamma(\boldsymbol{r}) = \frac{N-U}{K}\left(L_G(\boldsymbol{r})^{2/N} - 1\right) \sim \begin{cases} F_{K,N-U}, & \mathcal{H}_0 \\ F'_{K,N-U}(\lambda), & \mathcal{H}_1 \end{cases} , \tag{5.4}$$

where the non-centrality parameter is

$$\lambda = \frac{(\boldsymbol{BA}-\boldsymbol{b})^{\mathrm{H}}(\boldsymbol{B}(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S})^{-1}\boldsymbol{B}^{\mathrm{H}})^{-1}(\boldsymbol{BA}-\boldsymbol{b})}{\sigma^2} .$$

This results in the following threshold criterion in terms of GLRT

$$\gamma(\boldsymbol{r}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_{\mathrm{th}} . \tag{5.5}$$

The exact detection performance in terms of probability of false alarm (P_{FA}) and probability of detection (P_{D}) is given by

$$P_{\mathrm{FA}} = Q_{F_{K,N-U}}(\gamma_{\mathrm{th}}) \tag{5.6}$$

$$P_{\mathrm{D}} = Q_{F'_{K,N-U}(\lambda)}(\gamma_{\mathrm{th}}) , \tag{5.7}$$

where $Q_{F_{K,N-U}}$ ($Q_{F'_{q,N-U}(\lambda)}$) returns the right tail probability of a central (non-central) F distribution with K degrees of freedom in the nominator and $N-U$ degrees of freedom in the denominator. Since now, the PDF of $\gamma(\boldsymbol{r})$ under $\mathcal{H}_0$ is completely known, one can ensure that P_{FA} will not surpass a predefined value by finding a proper threshold γ_{th} in Eq. 5.6.

5.3 Multi-user Beam Selection Based on Sequential Competition

For ease of exposition, consider that users are located in the same distance, while each have a single LOS path to the base station. This results in the following received sequences after beamforming under beam $m \in \{1, 2, \ldots, M\}$ at time index n as

$$r_m[n] = \sum_{u=1}^{U} \underbrace{\rho_{u,p}\boldsymbol{w}(\phi_m)\boldsymbol{a}(\phi_u)}_{A_{u,m}} s_u[n] + \underbrace{\boldsymbol{w}(\phi_m)\boldsymbol{z}'[n]}_{z_m[n]} . \tag{5.8}$$

The received sequence under beam m mentioned in Eq. (5.8) after observing n samples, follows the linear model which can be written as

$$\boldsymbol{r}_m = \boldsymbol{S}\boldsymbol{A}_m + \boldsymbol{z}_m \ , \tag{5.9}$$

where $\boldsymbol{S}$ is the known $n \times U(n > U)$ training matrix including all training sequences corresponding to different users up to sample n, $\boldsymbol{A}_m = [A_{1,m}, \ldots A_{U,m}]^{\mathrm{T}}$ is the $U \times 1$ vector of unknown complex amplitudes corresponding to different users under beam m while $\boldsymbol{z}_m$ is a $n \times 1$ vector of complex WGN noise samples with unknown variance under beam m.

Similar to our single user *sequential competition test* in [MRF19], instead of comparing the estimates of the captured power of all users under different beams with each other, we rather decompose the $M - ary$ test for each user into M parallel binary test w.r.t a virtual no signal hypothesis. Hence, the following binary test for user u under beam m can be formulated as

$$\begin{array}{ll} \mathcal{H}_{u,m,0} & : \boldsymbol{B}_u \boldsymbol{A}_m = 0, \ \sigma^2 > 0 \\ \mathcal{H}_{u,m,1} & : \boldsymbol{B}_u \boldsymbol{A}_m \neq 0, \ \sigma^2 > 0 \end{array} \ , \tag{5.10}$$

where $\boldsymbol{B}_u$ is defined as a $1 \times U$ vector with all entries equal to zero except the u^{th} entry corresponding to the user u, which is set to one. The binary test based on decision metric $\gamma_u(\boldsymbol{r}_m)$ implies that for each user u under each beam m, we check the presence of a signal corresponding to that user against its absence. Obviously, $\mathcal{H}_{u,m,0}$ is the wrong hypothesis for user u under beam m, assuming that some signal is observable but with different strength. On the other hand, $\mathcal{H}_{u,m,0}$ acts as a virtual common reference in the set of M parallel binary tests for user u. This results in total to $M \times U$ parallel binary tests.

We let all users compete to distinguish their signals from no signal hypothesis under each beam, while n can grow until a decision criterion is fulfilled. Let us denote the probability of selecting the presence of the signal corresponding to user u under beam m after n observations as $P_{\mathcal{H}_{u,m,1}}(n)$. Comparing the binary tests of users u under beams m and m', it follows from Eq. (5.7) that for $|A_{u,m}| > |A_{u,m'}|$ and a common decision threshold γ_{th}, we have $P_{\mathcal{H}_{u,m,1}}(n) > P_{\mathcal{H}_{u,m',1}}(n)$. This is simply a consequence of the fact that the accumulated deflection coefficient (equivalent to non-centrality parameter) $\lambda_{u,m}$ will grow more quickly than $\lambda_{u',m}$ as n grows. Therefore, the beam that sees the stronger signal will on average cross the threshold earlier. This fact leads to the following sequential competition test applied to stochastic paths $\gamma_u(\boldsymbol{r}_m)$ for $m = 1, \ldots, M$ and $u = 1, \ldots, U$ as,

$$\gamma_u(\boldsymbol{r}_m(n)) \underset{\text{undecided}}{\overset{\mathcal{H}_{u,m,1}}{\gtrless}} \gamma_{\mathrm{th}} \ , \tag{5.11}$$

where at each step n all stochastic paths $\gamma_u(\boldsymbol{r}_m(n))$ for $m \in \{1, \ldots, M\}$ and $u \in \{1, \ldots, U\}$ are compared to the fixed common threshold γ_{th}. As soon as one of the paths surpasses the threshold, the corresponding pair (u, m) is selected and the training stops for that user while other users continue. The same procedure repeats by taking the next observation into account until all users surpass the threshold or the maximum test length is reached.

The interpretation is that for each user, we let the beams compete to distinguish themselves from pure zero mean WGN with unknown variance and the one which does it faster is the winning beam for the corresponding user in the competition. As a result, the overall test length n for each user is now a random variable, with $\bar{n}$ as the average number of observations per user.

The multi-user sequential competition test selects the strongest beam for each user as early as possible adaptively with respect to the SNR operating point of each user. For low user density, i.e., small values of the ratio U/M which we denote as the packing ratio, the probability that the strongest beam for different users overlap, is small and therefore the selected (user, beam) pairs result in near optimal performance in terms of overall SINR. However, this result may not be optimal (specially in higher SNR regime) when the packing ratio U/M is large, since multiple users might have the same strongest beam. To combat the heavy interference in this case, the affected users may be served via time-sharing while using the same beam. The downside to time-sharing might be the fact that some RF-chains may be wasted since they will not have any contribution to the sum-rate performance. This can be addressed by interference-aware beam allocation for colliding users [AM15] [GDC+16].

5.4 Numerical Evaluation

We numerically studied the performance of the *multi-user sequential competition test* in the reference channel model described in Eq. (5.1) with a uniform linear array with 32 antenna elements using the codebook of a Butler matrix with 32 orthogonal beams with normalized magnitudes. The users have the same normalized power at the receiver before beamforming equal to 1. The AoA was distributed uniformly in $[-90^{\circ}, 90^{\circ}]$ independently for each user over the simulation runs, while SNR was defined as $1/\sigma^2$. The average SINR values were estimated at each SNR point based on 10^4 simulation runs for SNR values in the interval $[-6, 6]$ dB. For comparison, we consider the ideal beam selector based on genie knowledge on $A_{u,m}^2$ and σ^2 values. As shown in Fig. 5.2, the multi-user sequential competition test with γ_{th} based on $P_{\text{FA}} = 10^{-3}$, shows a very close performance in terms of average SINR per user compared to the ideal beam selector based on genie knowledge. This demonstrates the robustness of the sequential test under SNR variations.

This feature is achieved by adaptively changing the average test length per user $\bar{n}$ with

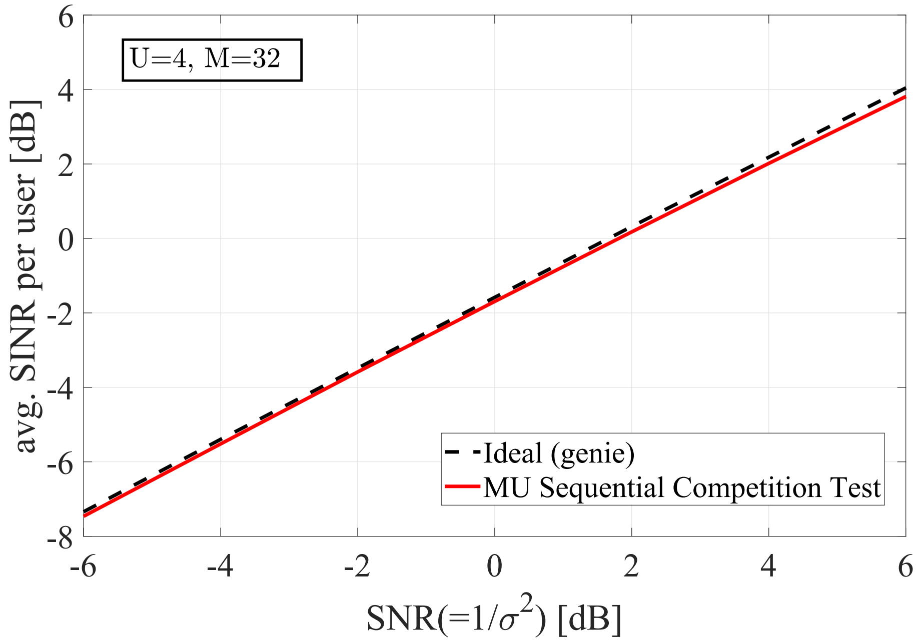


Figure 5.2: Comparison on average SINR per user after ideal beam selection based on genie knowledge versus MU sequential competition test.

respect to the SNR operating point, as illustrated in Fig. 5.3.

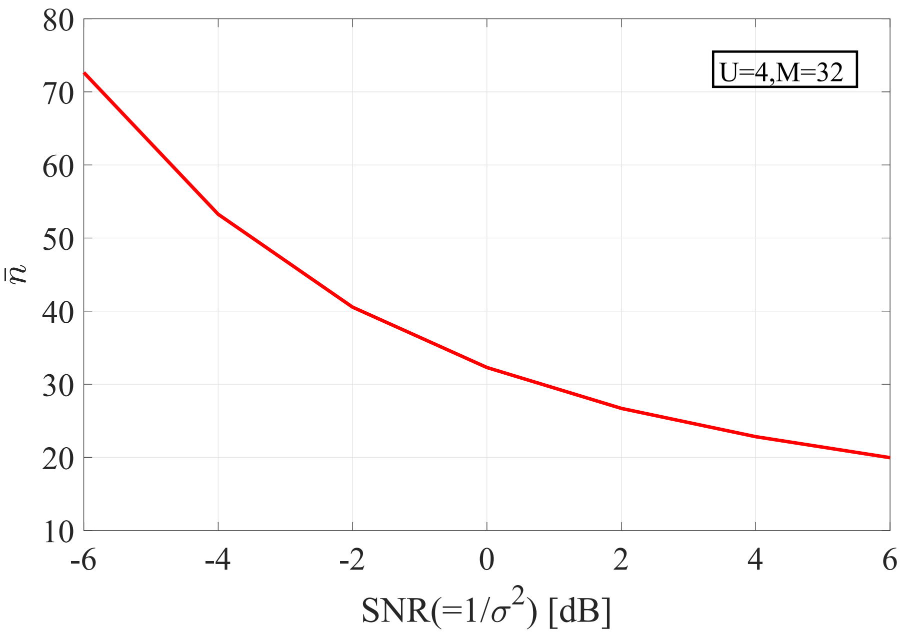


Figure 5.3: Average number of observations $\bar{n}$ per user over different SNR values resulted from MU sequential competition test.

5.5 Multi-user Digital Beamforming

Generally, the common solution for the digital beamforming is given based on the Singular Value Decomposition (SVD) of the channel matrix [AH16], which describes the coupling between the antenna arrays of the receiver and the transmitter. The overhead of the channel estimation for multiple users and large antenna arrays, make the similar solutions difficult to obtain in practice. However, we know that the channel at mmWave frequencies is sparse and therefore, the hybrid beamforming can offer a practical solution for the multi-user massive MIMO problem.

The two steps of multi-user hybrid beamforming can be viewed as follows: first, estimate the energy transfer between narrow analog beams at the BS and the users, to project a massive MIMO channel of high dimensions into a subspace of much lower dimensions, so that an 'effective' channel results based on the selected analog beamforming vectors. Only for this effective channel, the exact parameters (a matrix-valued impulse response or channel transfer function) need to be estimated. In the second step, the optimal weights for a digital precoder/combiner with the dimensions which depend on the number of RF-chains (i.e., streams/users) are determined.

Numerous works on hybrid beamforming architectures studied the optimization of RF and digital precoders/combiners during the communication phase while assuming full CSI [SY16, LM17, IE17, DXS+18, COS18], i.e., that the vectors of baseband complex channel impulse responses at each array element are known, which is not practical in mmWave systems. We note that, given candidates of analog beamforming vectors selected from the given codebooks resulting in an effective MIMO channel matrix, finding the optimal solution to the digital beamforming problem is almost equivalent to the conventional fully digital multi-user detection or precoding/combining combating the multi-user interference, apart from different power constraints [CRT01, WBKK04, WBKK03, Kuh03, KBK02].

We note that, in certain scenarios, all the users can not be spatially separated enough. In this case, if the best beam for each user is solely decided based on the estimated received power (a.k.a. magnitude maximization), the resulted effective channel matrix might not have a full-rank and a good condition number. In simpler words, the acheived performance in a MU massive MIMO system using the hybrid beamforming structure depends on the singular values of the combined matrix resulted from the multiplication of the effective channel and the digital beamforming matrices. A practical solution to this particular problem is to store a few more promising candidate beams for each user during the analog beam selection phase. Then, a subset from the set of selected candidates (this set has a reduced size compared to the codebook) will provide the best effective channel for which the successive digital beamforming yields the maximum sum rate. It is evident that the computational complexity exponentially increases depending on the size of the enlarged candidate set for each user.

We note that the performance limiting factor in multi-user applications with a single RF-chain available per user (Fig. 5.1) is the packing ratio U/M, or the ratio between the number of users U and the size of the codebook M. The Frequency of collision events (see Fig. 5.4), i.e., users with same angular signatures, depends on this packing ratio. In case of infrequent collision events: the SINR Maximization (SM) and the Magnitude Maximization (MM) are equivalent. Otherwise, the solution to MM can get near to SM using simple techniques like time-sharing, collision detection, table of best beams. In any case, Digital beamforming can be used by applying digital weights to the input of RF-chains to improve the effective channel condition, i.e., SINR values of different users and therefore achieving higher multi-user sum rates.

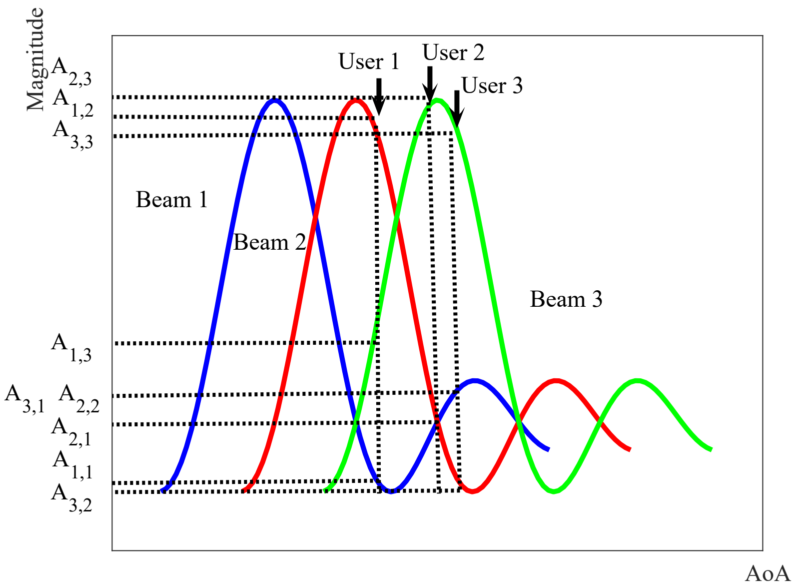


Figure 5.4: A collision event by analog beam selection process.

5.5.1 Estimation of Candidate Effective Channels

During the beam selection phase, the signal amplitudes for each user can be estimated under different beam candidates. These estimated values are used to build the $U \times U$ effective channel, which describes the coupling between the RF-chains at the BS and the users. This problem can be solved in different scenarios. Without loss of generality, we consider the uplink training in which a base station equipped with a larger array, tries to solve the multi-user beam selection and digital precoding design problems, while the users are equipped with a single antenna.

We note that, the total number of possible effective channels using the beams in the discrete codebook is M^U. When the selected beams are set to be unique and there is no repetition, this number becomes $M!/U!$.The key fact here is that the channel is sparse and only few beams capture a non-negligible amount of power for each user. Our proposed practical solution will employ the SCET at the base station separately for each user based on the received sequences from different beams as shown in Fig. 5.5. This way, a predefined number p of candidates for each user (minimum number of survivors in the elimination test) is stored in the set $J_u = \{j_{u,1}, \dots j_{u,p}\}$. The use of Gold sequences with good auto and cross correlation properties for different users, enables the efficient parallel processing at the base station and reduces the complexity.

$$\text{at BS (in parallel)} \begin{cases} \text{SCET for user } 1 \Rightarrow \text{ Set of best beams } J_1 = \{j_{1,1}, \dots, j_{1,p}\} \\ \text{SCET for user } 2 \Rightarrow \text{ Set of best beams } J_2 = \{j_{2,1}, \dots, j_{2,p}\} \\ \vdots \\ \text{SCET for user } U \Rightarrow \text{ Set of best beams } J_U = \{j_{U,1}, \dots, j_{U,p}\} \end{cases}$$

Figure 5.5: Selection of a set of promising candidate analog beams for each user at the base station.

Using the estimated complex amplitude corresponding to each user under different candidate beams as $\hat{A}_{u,J_{u'}}$ available after the analog beam selection step, one can simply build $|J_u|^U$ possible candidate effective channels described as,

$$\mathbf{H}_E = \begin{bmatrix} \hat{A}_{1,J_1} & \hat{A}_{1,J_2} & \dots & \hat{A}_{1,J_U} \\ \hat{A}_{2,J_1} & \hat{A}_{2,J_2} & \dots & \hat{A}_{2,J_U} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{A}_{U,J_1} & \hat{A}_{U,J_2} & \dots & \hat{A}_{U,J_U} \end{bmatrix}, \tag{5.12}$$

where the diagonal elements indicate the useful signal amplitudes or the energy couplings between the base station and all users. The downlink SINR for each user u considering the off-diagonal elements (interference to the user u) can be formulated as

$$SINR_u = \frac{|\hat{A}_{u,J_u}|^2}{\sum_{k \neq u}^{U} |\hat{A}_{u,J_k}|^2 + \sigma^2} . \tag{5.13}$$

We note that $|\hat{A}_{u,J_{u'}}|^2$, $u \neq u'$ denotes the resulting interference power from the beam selected on user u', present in the received signal of user u in downlink.

5.5.2 Joint Hybrid Beamforming Optimization Problem

One can now formulate the baseband downlink data transmission model including the digital beamforming matrix $\mathbf{F}_B = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_U]$ (single RF-chain per user) and the candidate effective channels $\mathbf{H}_E$ as a linear model describing the vector of received sequences $\mathbf{r}[n]$ at symbol index n as in Eq. 5.14

$$\mathbf{r}[n] = \begin{bmatrix} r_1[n] \\ r_2[n] \\ \vdots \\ r_U[n] \end{bmatrix} = \mathbf{H}_E \mathbf{F}_B \mathbf{x}[n] + \mathbf{z}[n] \ , \tag{5.14}$$

where the transmitted data vector of all users is $\mathbf{x} = [x_1[n-\tau_1], x_2[n-\tau_2], \ldots, x_U[n-\tau_U]]^T$. The delays τ_u are already detected by SCET during beam selection phase for each user. The transmit-symbols are normalized as $|x_u[n]| = 1$.

In the next step the joint analog-digital optimization problem to maximize the mutual information between the transmitted and received signals, i.e., sum rate can be formulated as

$$\max_{\mathbf{J}_1, \mathbf{J}_1, \ldots, \mathbf{J}_\mathrm{U}} \begin{Bmatrix} \max\limits_{\mathbf{F}_\mathrm{B}} \{I(\mathbf{r}, \mathbf{x})\} \\ \text{s.t.} \{\mathbf{f}_u^\mathrm{H} \mathbf{f}_u\} = 1 \end{Bmatrix} \ , \tag{5.15}$$

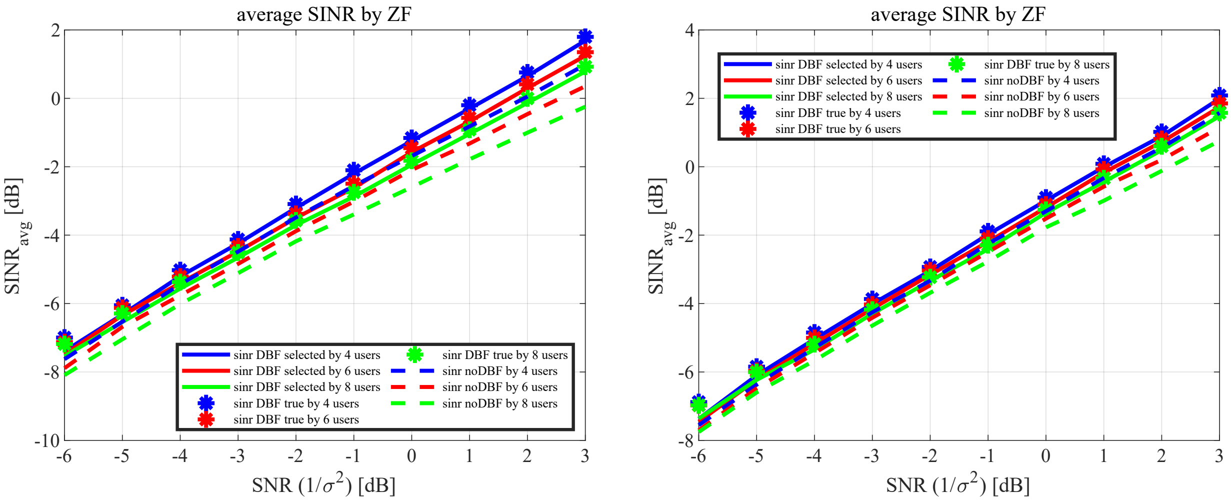
where the power constraint is applied to the digital beamforming vectors. The outer maximization problem corresponding to analog beam selection has now a reduced size since only sets of promising beams should be checked rather than the complete codebook.

One of the possible solutions to the second maximization problem known as the multi-user digital precoding problem is zero-forcing for each candidate effective channel $\mathbf{H}_E$, as

$$\mathbf{F}_B = \mathbf{H}_E^H (\mathbf{H}_E \mathbf{H}_E^H)^{-1} \ . \tag{5.16}$$

The interpretation is that we first build the most promising effective channels using the stored sets of candidate beams $\mathbf{J}_u$ for each user. This reduces the size of the optimization problem, which is initially intractable. Next, we look for the best digital beamforming matrix that maximizes the sum-rate for any given candidate $\mathbf{H}_E$. The final solution is then chosen among all candidates based on the overall achieved sum-rate. We note that, some candidate H_E matrices in particular realizations of the channel might not be full-rank, specially when the packing ratio is high. These cases are discarded while solving the optimization problem described in Eq. 5.15.

In Fig. 5.6 (single RF-chain per user) we numerically evaluated the average SINR per user for different packing ratios (i.e., $M \in \{32, 64\}$ and $U \in \{4, 6, 8\}$) in a reference channel described in Eq. (5.1) model with 3 paths per each user. Here, the ideal (visible by colored dots) case is when the beams can be selected with Genie knowledge about the coupling coefficients while using ZF. The number of candidates stored for each user for estimating the promising effective channels is set to two.



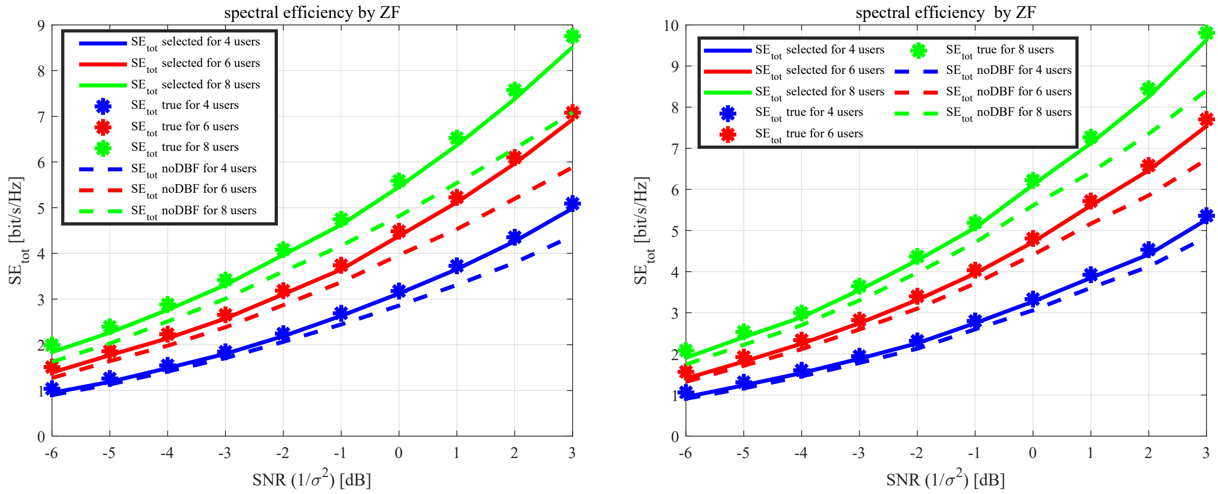
(a) $M = 32$ **(b)** $M = 64$

Figure 5.6: Achievable average SINR per user before and after ZF digital beamforming while using SCET for first step analog beam selection with $|J_u| = 2$. Comparison is done with the ideal case where beams are selected based on genie knowledge. The dashed lines show the result after beam selection, but without digital beamforming.

As expected, the benefit of digital beamforming becomes more prominent when the number of users with respect to the number of beams at the base station is larger. Furthermore, It is observable that in higher SNR regime where the effect of interference regarding the additive noise is more dominant, the application of digital beamforming is more fruitful. It is more understandable to evaluate the same experiment in terms of spectral efficiency, as depicted in Fig. 5.7. Rate gain of %10 to %15 is achievable using the simple zero forcing digital beamforming compared to simple magnitude maximization via pure analog beam selection.

5.6 Summary

In this chapter, we considered the multi-user beam selection problem in mmWave massive MIMO systems. We proposed a novel sequential hypothesis test based on GLR statistics to solve the composite multi-user beam selection problem. The proposed *multi-user sequential competition test* selects the strongest beam(s) for each user as early as possible adaptively with respect to the SNR operating point of each user. The achieved performance in terms of average overall SINR per user is close to the performance of the ideal beam selector based on genie knowledge on angle of arrivals and noise variance for different packing ratios U/M. Similar behaviors can be observed in a more complicated multipath mmWave channel, which requires additional correlation steps at different delays. In the second part, we discussed the joint multi-user hybrid beamforming optimization problem. Our practical solution with minimum complexity and performance loss is to use the SCET



(a) $M = 32$ **(b)** $M = 64$

Figure 5.7: Achievable total spectral efficiency before and after ZF digital beamforming while using SCET for first step analog beam selection with $|J_u| = 2$. Comparison is done with the ideal case where beams are selected based on genie knowledge. The dashed lines show the result after beam selection, but without digital beamforming.

algorithm in parallel at base station to save a set of promising beams for each user. These sets can be then used to build the best candidate effective channels and thereby reducing the size of the joint optimization problem. The subsequent digital beamforming problem has well-known solutions like zero-forcing, where we try to improve the effective channel matrix condition and achieving the overall maximum sum-rate among all candidate effective channels. The ratio between number of users and the size of the codebook at the base station (namely the packing ratio) determines how often some users cannot be spatially separated with different beams. Time and frequency sharing among colliding users, are simple solutions to improve the sum-rate performance.

Chapter 6

Experimental Evaluation

In this chapter, we aim to experimentally evaluate the performance of our developed variable-length beam selection framework known as Sequential Competition and Elimination Test (SCET). The experimental setup employs a hardware in the loop including the previously discussed Butler Matrix developed at TU Dresden which provides a codebook of orthogonal beams connected through an electronic switch to a single RF-chain. We have designed multiple measurement scenarios in which beams with different main-lobe directions are automatically probed at the TX and RX through different channel realizations. The acquired measurements are processed in Matlab to benchmark the beam selection accuracy and efficiency. The results are in line with the theoretical expectations and indicate the resilience of the SCET in the presence of hardware impairments and non-ideal pre-beamforming frequency and time synchronization. The presented experimental results in this chapter are published in [KMWK+22].

6.1 Experimental Test-bed

A state-of-the-art hardware setup [DBN+20] that combines a Universal Software Radio Peripheral (USRP) based platform with mmWave frontends [WLP+19a,WLP+19b,WLP+18] was developed at the Vodafone Chair for Mobile Communications Systems at TU Dresden (see Fig. 6.1). This experimental test-bed provides a software framework for design and execution of different hybrid beamforming measurement scenarios. As depicted in the Fig. 6.1, the USRP-based platform for mmWave experimentation mainly consists of the following hardware components.

- Control PC: via an Ethernet connection to the USRP-platform, this terminal PC allows users to develop and implement different hardware programming with LabVIEW to conduct various experiments.

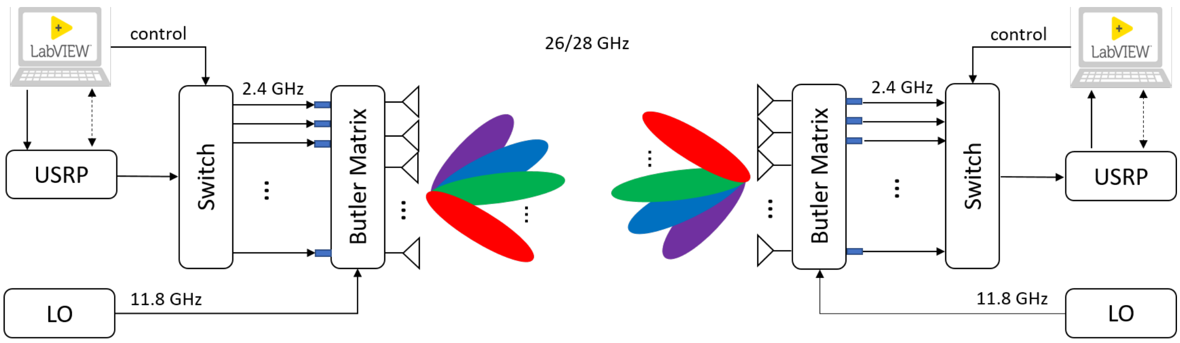


Figure 6.1: USRP baesd mmWave platform for experiments. [DBN+20].

- USRP-platform: is essentially an SDR platform. In this work, USRP- 2974 running the NI Real-Time Linux is deployed. Note that, the implemented LabVIEW programs are actually running in this NI Real- Time Linux insides the USRP.

- Switch: between the USRP and RF frontend, the 1×16 Mini-Circuits switch USB-1SP16T-83H that connects the USRP's single port (either Tx output or Rx input) to the 16 I/O-ports of the mmWave frontend.

- mmWave RF-Frontend: a multibeam mmWave antenna array with a 16×16 Butler matrix beamforming network, that has 16 I/O-ports, each one corresponding to a different beam as well as an RF path. Generally, it can operate in the frequency range from 26 GHz to 30 GHz, while in this case, the array of 16 quasi-yagi antennas is radiating at 26 GHz. Besides, it also features other components such as PAs, LNAs as well as mixers for the up/down conversion.

- Local oscillator (LO): this device produces the carrier signal at 11.8 GHz to feed the antenna array. Note that, the carrier signal can either be generated separately by two LOS for both Tx- and Rx-frontends as depicted in Fig. 6.1, or by the same one LO and then shared to both frontends via a splitter.

For any type of experiment, the USRP as the core of the transceiver can operate in either transmitter or receiver mode (Tx/Rx-mode) under the user's control. In Tx-mode the baseband signal will be shifted to an intermediate frequency (IF) of 2.4 GHz by the USRP analog frontend, while a reverse process is done in Rx-mode. Next, as each port of the mmWave frontend corresponds to an independent beam, the 1-to-16 switch can then be treated as a beam selector for the multibeam mmWave antenna array by selecting the respective port, which is controlled by the LabVIEW program via a separate TTL signal from the USRP. Eventually, in the Tx frontend, the 2.4 GHz signal will go through a serial of processes, i.e., up-conversion to 26 GHz, beamforming by the Butler matrix, and amplification by the PAs before the final radiation via the antenna array. On the Rx side, the frontend firstly amplifies the signal coming from each antenna via LNAs, then feeds them into the Butler matrix and down-converts the output back to 2.4 GHz.

A single carrier transmission has been adopted throughout this investigation. A long PN sequence of size 1024 is periodically transmitted and measured with different beams for any realized angle of arrival. This way different paths can be resolved in time by doing correlation at the receiver. For each realization, the ground truth is acquired via measuring the whole sequence for each beam combination. The received sequences can be then processed in MATLAB for any algorithmic evaluation. It also provides the freedom to add synthetic noise to the measured data to evaluate different SNR operating points.

6.2 Measurement Scenarios

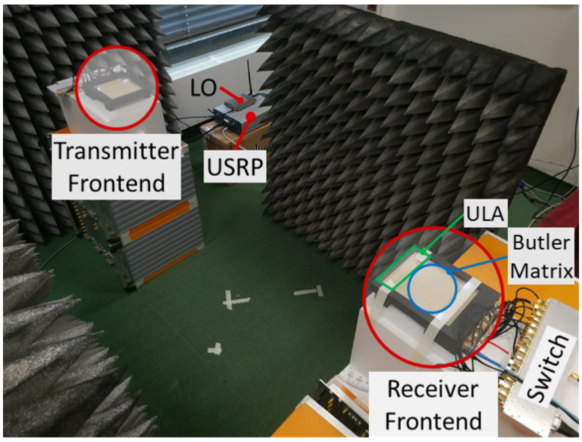
6.2.1 Line of Sight Scenario

To measure the beam, patterns generated by the mmWave frontend, i.e., a ULA with 16 quasi-yagi antennas driven by the Butler matrix as the beamforming network, we set up a LOS scenario for channel measurements. In this LOS scenario as depicted in Fig. 6.2, a stationary Tx frontend is sending the training sequence consistently with a fixed beam (e.g., beam 9 at the middle in this case) pointing directly at the receiver, while the Rx frontend is not only rotatable but also able to sweep over all $M = 16$ beams. Note that, both Tx and Rx frontends are placed at the same height. Furthermore, we also put two arrays of pyramid-shaped Radiation Absorbent Material (RAM) on both sides of the propagation path to effectively absorb and reduce incident RF radiation in NLOS directions, which leads to a pure LOS channel. TX and RX are set 1 meter apart. In summary, this whole arrangement enables measurements of all Rx-beams under different AoA realizations with good accuracy.

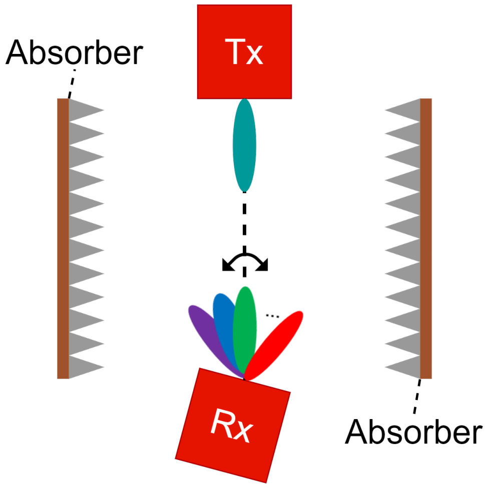
Next, we measured the beam patterns depicted in Fig. 6.3 in a 1-meter LOS scenario. The magnitudes patterns are estimated by taking the sample mean of the correlated measurements from all 16 beams under different AoAs. This is achieved through simply rotating the Rx frontend within a sector range of $[-60, 60]$ with an angular step of 2°. The element pattern of the patch antenna can be then estimated from the peaks of the measured patterns as a cosine function. It is demonstrated by the blue line in Fig. 6.3 as a MMSE fit to the peaks of the beam patterns.

6.2.2 Non Line of Sight Scenario

In the NLOS scenario, we block the direct path between TX and RX and only allow a path through a reflection by carefully positioning the arrays, the metal plate reflector and



(a) LOS Measurement setup



(b) LOS schematic

Figure 6.2: LOS scenario with stationary Tx and rotatable Rx.

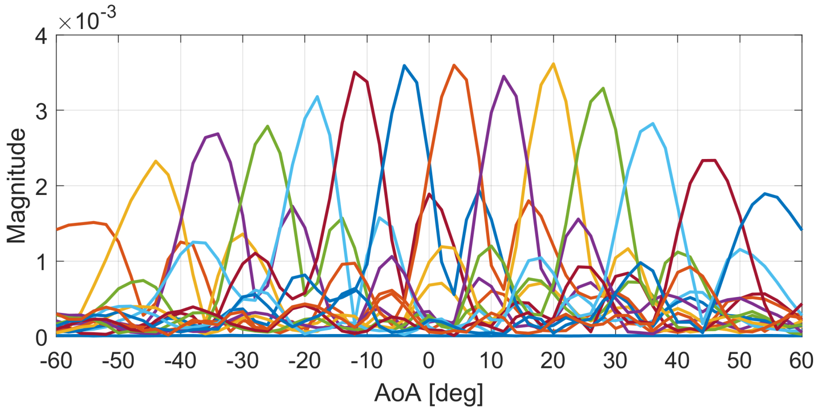
the absorbers (see Fig. 6.4). The selected beam of the transmitter will point to the metal plate with an incident angle of roughly 45 degrees. It is attempted to create a nearly perpendicular reflection. In this way, a pure NLOS scenario is built up. As for the rest of the setups for Tx and Rx frontends, we keep them unchanged: the stationary Tx with the fixed central beam and the rotatable Rx with all its beams available. The distance between the Tx/Rx frontends and the reflection point on the metal plate is approximately 0.75 meters, which makes this NLOS propagation path in total 1.5 meters.

The beam patterns measured in a path distance of 1.5 meters within the NLOS scenario are acquired similarly via estimating the received energy from each port of the Butler matrix under different AoAs. The estimated patterns are depicted in Fig. 6.5.

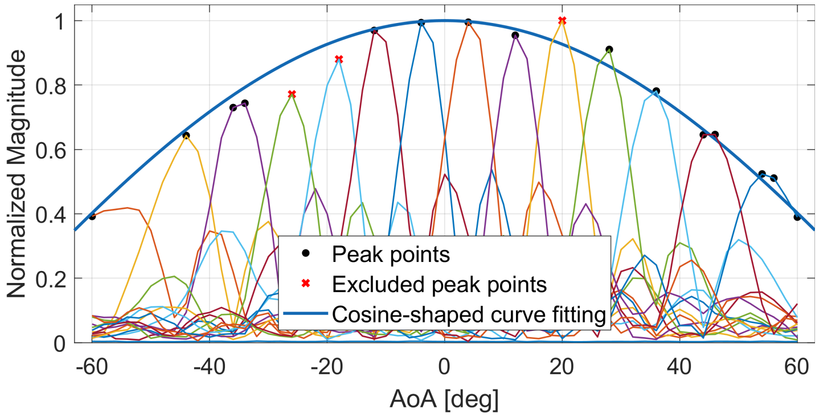
When comparing the absolute values of this magnitude pattern and compare it to the LOS scenario, we observe a certain amount of loss. This observed attenuation in peak received magnitudes is due to both the longer propagation distance and the one-time reflection. The extra free-space path loss in this case (1.5 meters instead of 1 meter) is roughly equal to 3.52 dB in signal power, while the total loss is calculated equal to 5.79 dB by comparing the maximum magnitudes from NLOS and LOS scenarios. Hence, the pure reflection loss equals 2.27 dB in power, which corresponds to a scaling factor of 0.77 in magnitude.

6.2.3 Multi-path Scenario

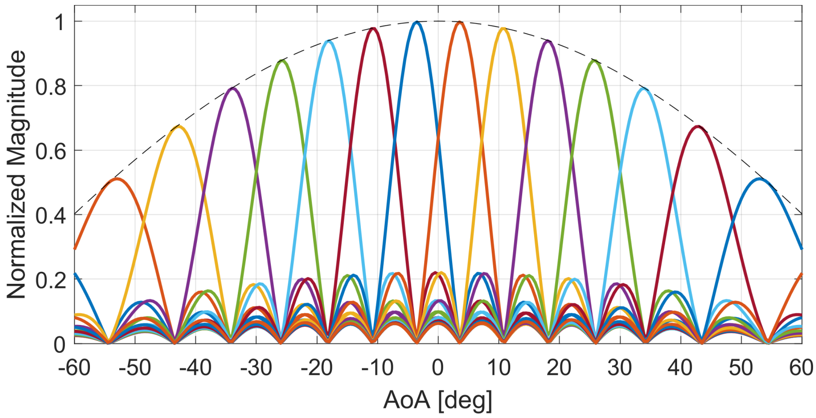
Next, we consider a multi-path scenario including a LOS and NLOS paths that are resolvable in time, as illustrated in Fig. 6.6. To generate two resolvable paths within the available space in the lab, we increase the operating sampling rate of the USRP platforms to 100 MHz, instead of the default 10 MHz in the single-path (e.g., LOS and pure NLOS)



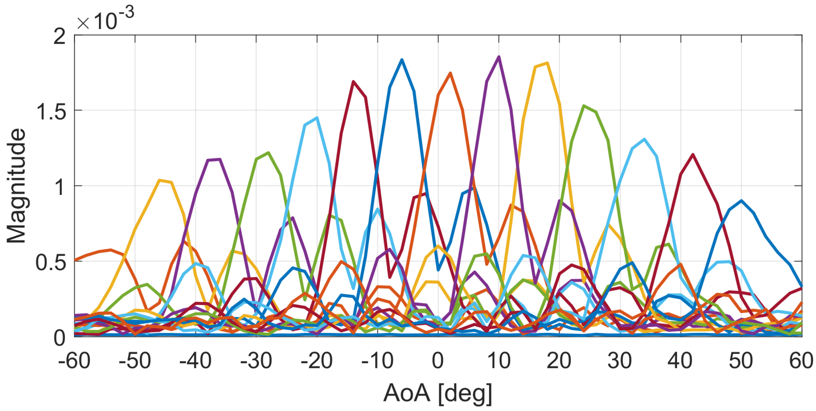
(a) Measured beam patterns in a LOS channel at 1 m distance.



(b) Estimated element pattern as a cosine function based on normalized measured patterns.

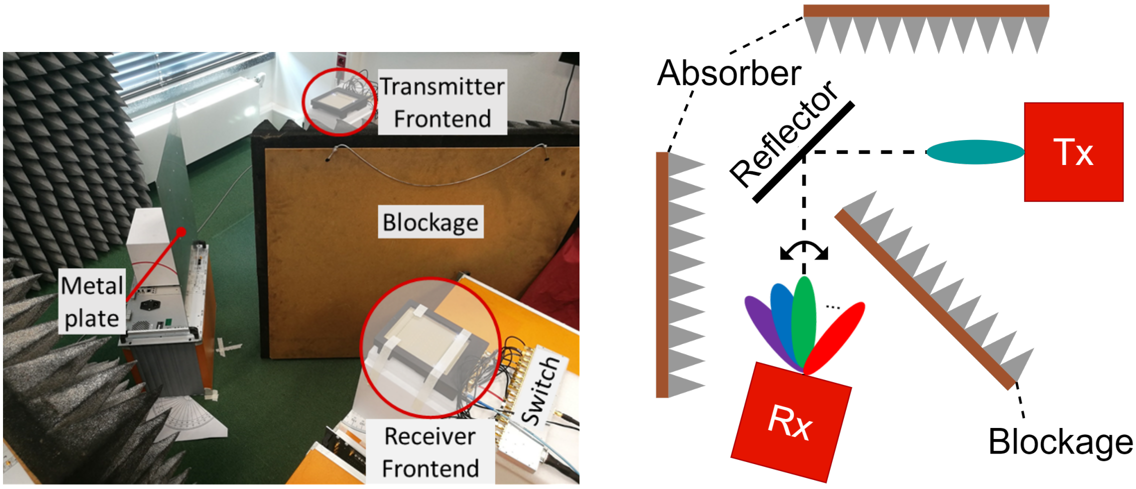


(c) Normalized Ideal beam patterns including the element factor.



(d) Measured beam patterns in a LOS channel at 1.5 m distance.

Figure 6.3: Evaluation of the beam patterns of a 16 × 16 Butler matrix and estimating the element pattern.



(a) NLOS Measurement setup

(b) NLOS schematic

Figure 6.4: NLOS scenario with stationary Tx and rotatable Rx.

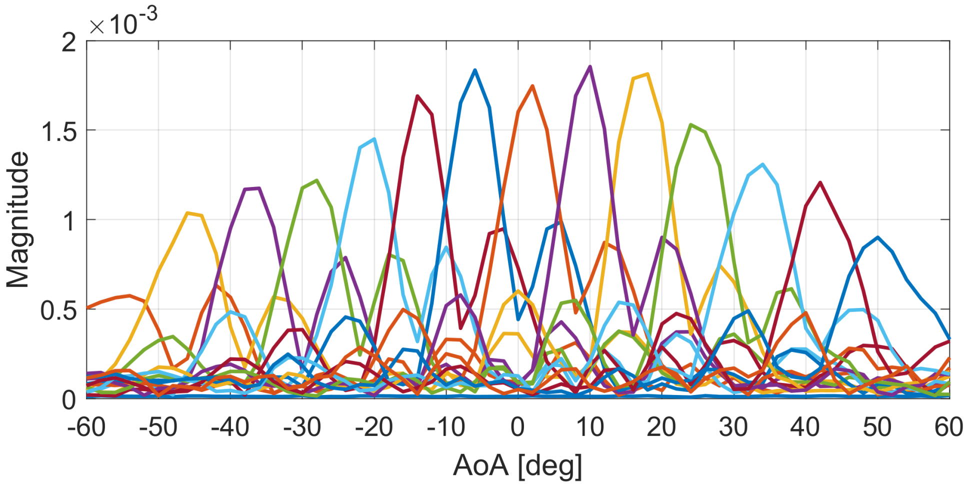
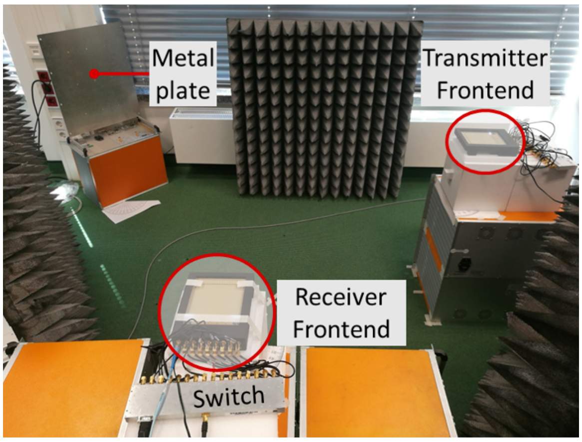
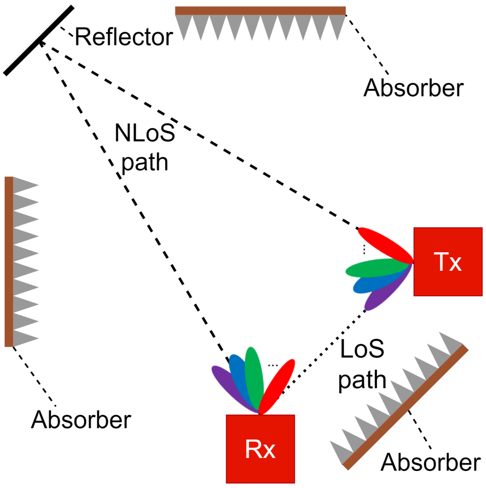


Figure 6.5: Measured beam patterns in a NLOS scenario at 1.5ms [DBN+20].

scenarios, to obtain a higher theoretical range resolution of 3 meters. The metal plate is relocated at a further distance to both frontends, so that the distance difference between LOS and NLOS paths will be larger than 3 meters. Finally, we assume that both Tx and Rx frontends are stationary with all 16 beams available in the multi-path scenario. This allows us to observe the received sequences of all 16×16 Tx- Rx-beam combinations and compute the signal power for each one of them. This way the CIR of this multi-path scenario can be described in the AoD-AoA-Delay domain, more specifically in the Tx-Rx-beam-Delay grid. Fig. 6.7 demonstrates the multi-path channel with two paths. Each path corresponds to a time index in the delay domain. According to the measurement illustrated in Fig. 6.7, the Tx-Rx-beam pair $3, 14$ captures the maximum power from the LOS path, which comes first due to the shorter distance. At the next time instant, the signal from the longer NLOS path is obtained by the beam pair $12, 5$. This matches the configuration shown in Fig. 6.6 and the measured beam patterns.



(a) Multipath measurement setup



(b) Multi-path schematic

Figure 6.6: Multi-path scenario with stationary Tx and rotatable Rx.

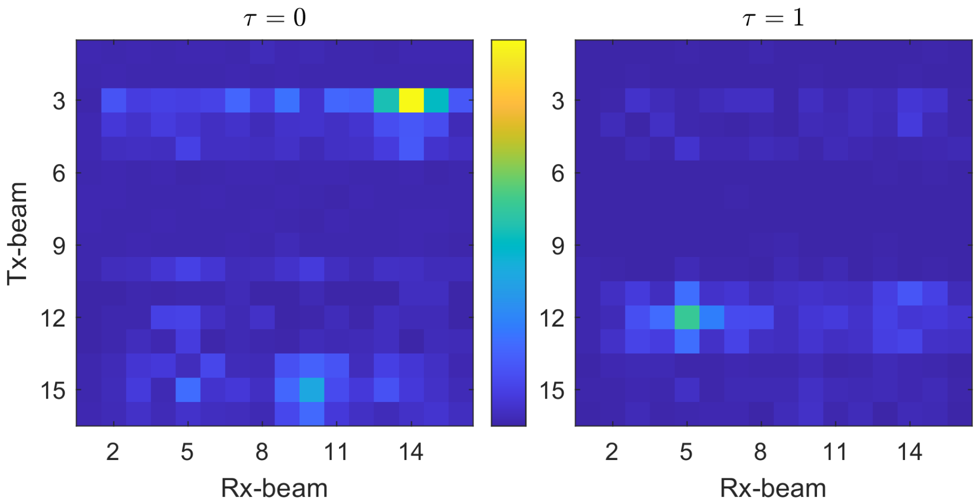
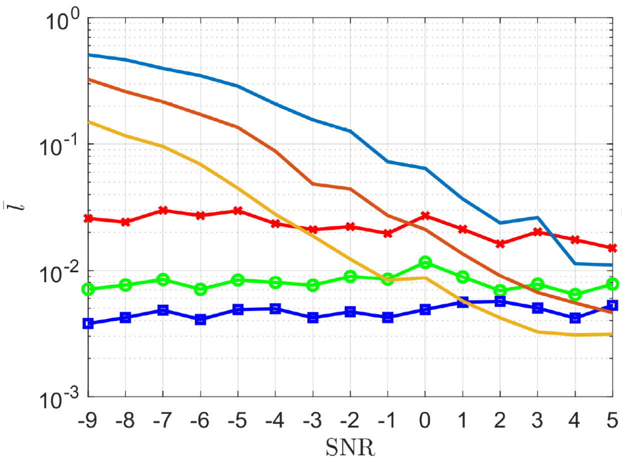


Figure 6.7: Measured sparse channel impulse response of a multi-path scenario: CIR at the delay index $\tau = 0$ (left) and CIR at the delay index $\tau = 1$ (right), correspond to the LOS path and the NLOS paths, respectively.

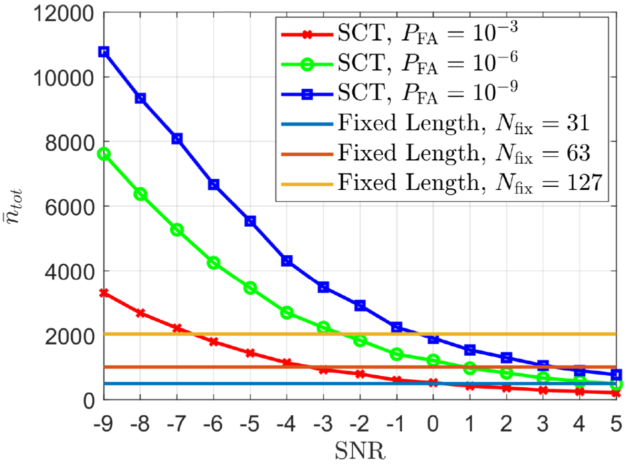
6.3 Beam Selection Performance

In contrast to the Fixed-length method, SCT has a significant advantage as a variable-length test, i.e. showing stronger adaptability to the SNR. As depicted in Fig. 6.9, the average test length for SCT is variable that can adaptively increase or decrease, as the channel condition deteriorates or improves, respectively. This leads to the normalized average loss of signal magnitude after beam selection $\bar{l}$ staying within a small interval. For instance, the fixed length detector with $N_{\text{fix}} = 63$ achieves a $\bar{l}$ in an interval of $[0.0046, 0.325]$, which is almost two orders of magnitude.

SCET also shows the same desirable adaptive and robust performance while improving the efficiency of the SCT as depicted in Fig. 6.11 by reducing the number of required

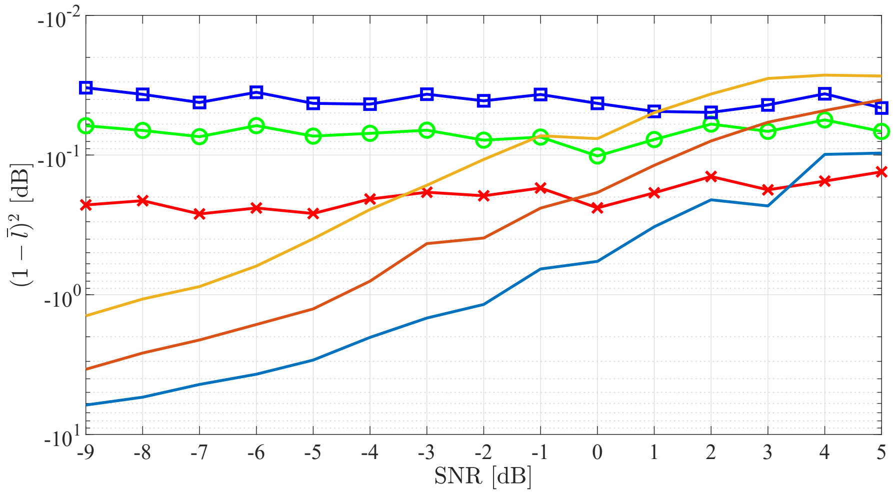


(a) Normalized average loss.

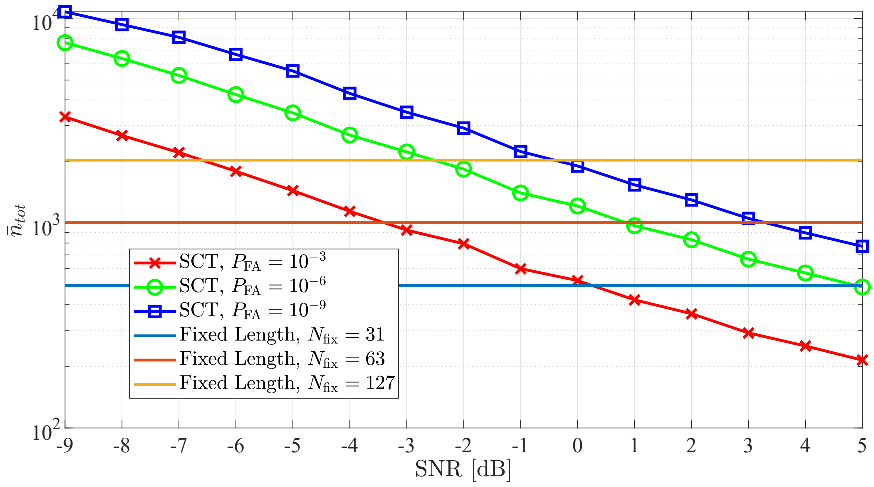


(b) Total number of samples

Figure 6.8: Beam selection accuracy and efficiency of SCT in comparison to Fixed-length detectors in a measured LOS channel.

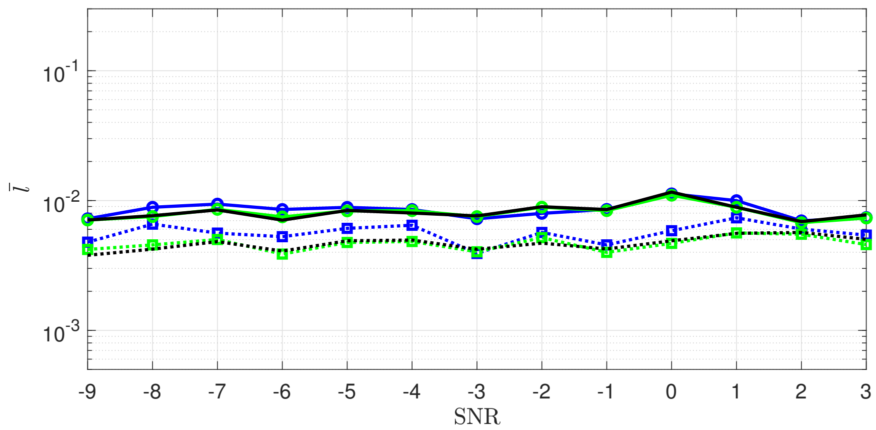


(a) Average power-loss after beam selection in dB.

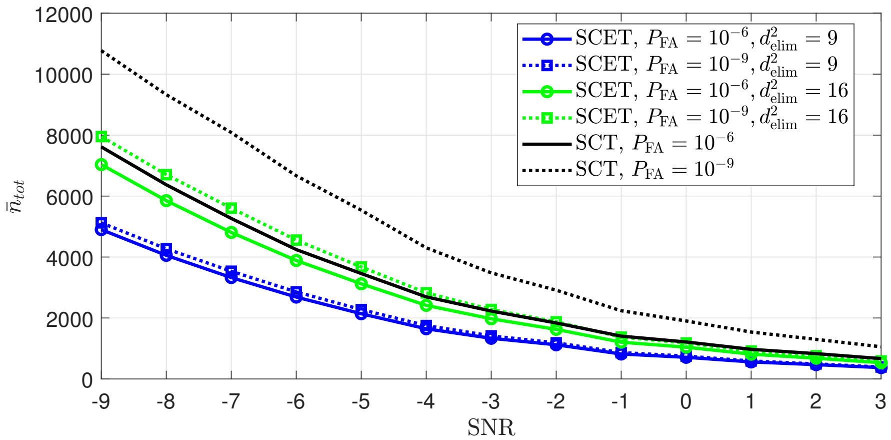


(b) Total number of samples.

Figure 6.9: Beam selection performance comparison between SCT and Fixed-length test in a measured LOS channel.



(a) Normalized average loss.



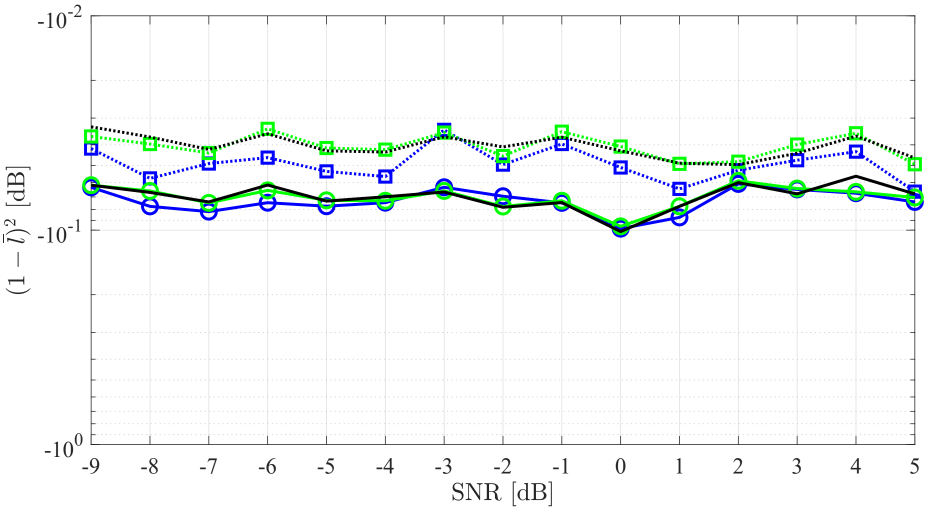
(b) Total number of samples

Figure 6.10: Beam selection accuracy and efficiency of SCET in a measured Multi-path channel.

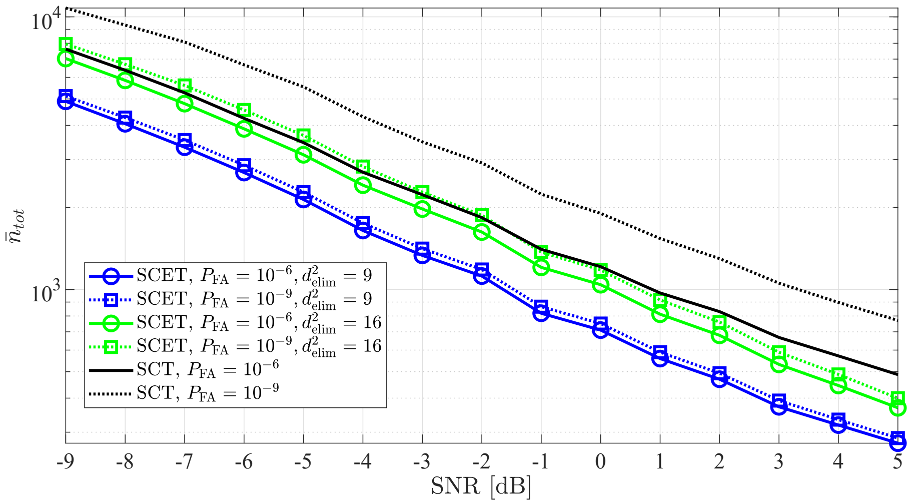
observation for reaching the same accuracy. The final evaluation has been done over the reference multi-path channel with one LOS and one NLOS path as shown in Fig. 6.6. The SCT and SCET algorithm once again prove their adaptive and robust performance. The efficiency of these test compared to the Fixed-length test with 3 different sizes is evident. Comparing the number of measurements for the same performance alignment accuracy in terms of $\bar{l}$ (for instance at -2dB) the SCET with $P_{\mathrm{FA}} = 10^{-6}, d^2_{\mathrm{elim}} = 9$ is two times more efficient than the Fixed-length test $N_{\mathrm{fix}} = 63$. Another example is the SCET with $P_{\mathrm{FA}} = 10^{-6}, d^2_{\mathrm{elim}} = 4$ at performance level of %2 loss, it outperforms the Fixed-length test with $N_{\mathrm{fix}} = 31$ by requiring two time less measurements.

6.4 Summary

In this chapter, we have employed the state-of-the-art USRP-based platform for 26/28 GHz mmWave experimentation developed at Vodafone chair to benchmark the performance of the SCET algorithm for beam selection. Through some development in LabVIEW for both Tx- and Rx-sides of the USRPs, an automated measurement procedure



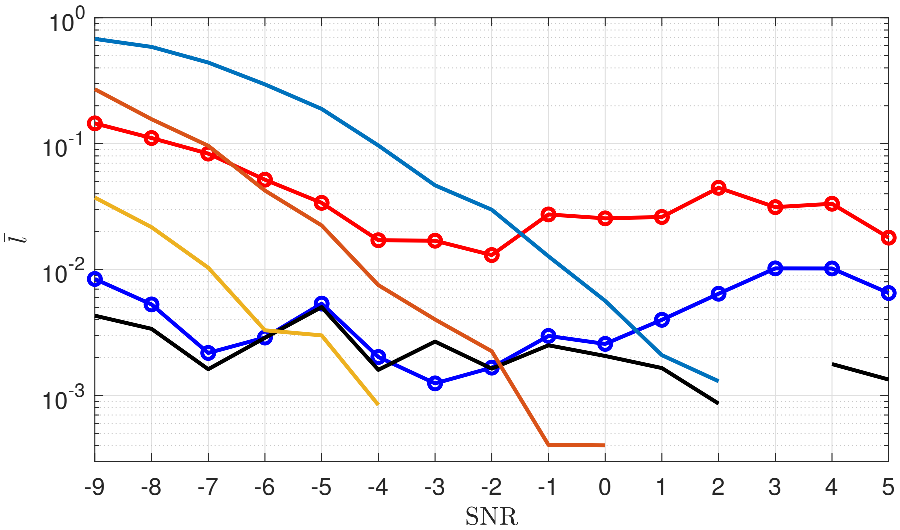
(a) Average power-loss after beam selection in dB.



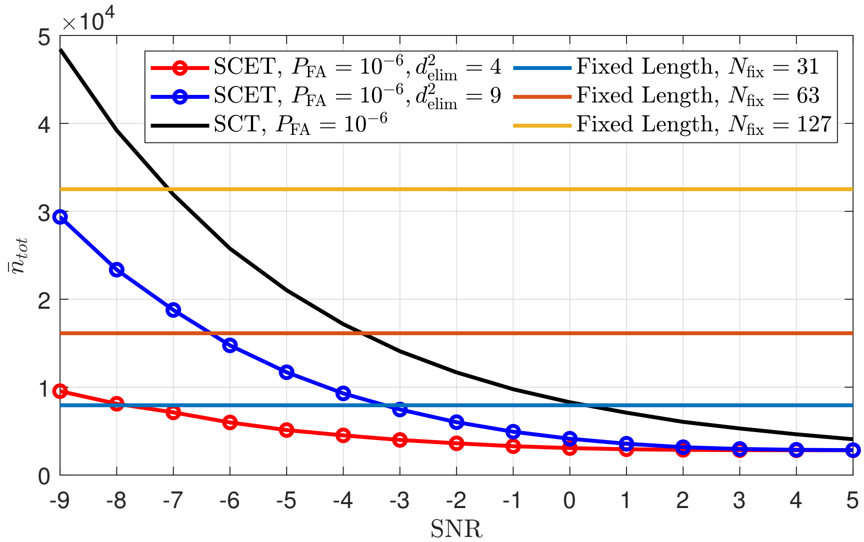
(b) Total number of samples.

Figure 6.11: SCET compared to SCT ins a LOS channel.

in each channel realizations has been attained. The beam patterns have been measured by carefully rotating the receiver antenna array and estimating the received power at each output port of the Butler matrix in angular steps of 2°. Using these measured beam patterns the element factor has been estimated which is in line with the design of the patch antenna. Next, concerning different types of channels, we designed three respective scenarios, including the LOS, pure NLOS, and multi-path. However, due to the inherent hardware limitations, practical measurements are inevitably accompanied by various nonidealities, especially the CFO. The evaluation of the measurements in MATLAB indicates the adaptivity and robustness of the variable-length detection framework SCET in presence of hardware impairments and the realistic coarse synchronization during the training phase. Moreover, the CIR was measured in a controlled two-path scenario, which shows consistency with its onsite arrangement. The corresponding evaluation in MATLAB confirms the adaptive reliable performance of SCT and the SCET algorithms for beam



(a) Normalized average loss.



(b) Total number of samples

Figure 6.12: Beam selection accuracy and efficiency of SCT and SCET in a measured Multi-path channel.

selection while achieving training efficiency of up to more than 2 times compared to the conventional Fixed-length detectors.

Chapter 7

Conclusion and Outlook

7.1 Conclusion

Throughout this thesis, we have investigated the mmWave massive MIMO systems as key enablers for the next generation of wireless communications systems 6G. We introduced two innovative pieces of hardware based on hybrid beamforming idea that has been researched and developed at Vodafone chair TU Dresnden. The analog beamforming network using a Butler matrix with a uniform linear array and a scalable D-band module with integrated beamforming capabilities. These provide practical solutions for the future requirements of the network, namely energy and cost-efficient beamforming and thereby higher data rates with larger available bandwidth.

For the performance limiting beam acquisition task during the link access phase in mmWave massive MIMO systems, we introduced a novel energy-detection framework based on variable-length channel measurements with orthogonal codebooks. The proposed beam selection technique denoted as composite *M-ary Sequential Competition Test* (SCT) solves the beam selection problem when knowledge about the SNR operating point is *not* available. It adaptively changes its test length when the SNR varies to achieve an essentially constant performance level. Moreover, to achieve the same performance in terms of captured signal power, the *sequential competition test* requires on average less observations (particularly in the lower SNR regime) in comparison to a perfectly tuned *Fixed-length* test assuming genie knowledge. Next, we augmented the SCT by a beam elimination mechanism. It evaluates the Bayesian probability of winning the competition for each candidate. Improved efficiency is attained by allowing elimination of unpromising beams from the remaining candidate set as soon as possible during the channel measurement phase. These benefits are of interest in systems with large number of candidate beams, as well as under conditions where the training time is limited due to small channel coherence time.

As mentioned, variable-length detection via the Sequential Competition and Elimination Test (SCET) provides adaptivity to channel conditions. This adaptivity during the beam acquisition phase results in a rate gain and/or a delay reduction by better fulfilling the training transmission trade-off. Use of different codebook types can affect the detection performance: An orthogonal Butler codebook can benefit from the elimination mechanism to increase the detection efficiency. However, the performance might suffer when the settling time for the phase shifters is large or switching time of the electronic switches is long. The frequency dependent codebook removes the need for switching the beams (during the exhaustive search) without any penalty. This codebook type is specially powerful when used at the base station, providing one step downlink training while employing variable-length detection based on SCT. Hierarchical codebooks show lesser detection efficiency, due to lower SNR at initial levels and the error propagation through the tree search. The use of variable-length detection based on SCET can improve the detection performance while using the hierarchical codebooks proposed in 802.11.ad.

To solve the multi-user beam selection problem, we proposed the *multi-user sequential competition test,* which selects the strongest beam(s) for each user as early as possible adaptively regarding the SNR operating point of each user. Next, we discussed the joint multi-user hybrid beamforming optimization problem. Our practical solution with minimum complexity and performance loss is to use the SCET algorithm in parallel at base station to save a set of promising beams for each user. These sets can be then used to build the best candidate effective channels and thereby reducing the size of the joint optimization problem. The subsequent digital beamforming problem has well-known solutions like zero-forcing, where we try to improve the condition of the effective channel matrix and achieving the overall maximum sum-rate among all candidate effective channels.

Finally, we experimentally evaluated the performance of our developed variable-length beam selection framework. The experimental setup combines a USRP based platform with mmWave frontends and the Butler Matrix developed at TU Dresden. The acquired measurements were processed in MATLAB to benchmark the beam selection accuracy and efficiency. The results indicate the resilience of the SCET in the presence of hardware impairments and non-ideal pre-beamforming frequency and time synchronization, while confirming the adaptive reliable performance of SCT and the SCET algorithms for beam selection achieving training efficiency of up to more than 2 times compared to the conventional Fixed-length detectors.

7.2 Outlook

As for the future work related to the proposed variable-length beam selection framework for mmWave massive MIMO systems, we outline the following interesting ideas:

Parameter optimization: The termination and elimination thresholds (γ_{term} and d^2_{elim}) used in SCET can be optimized based on considered use cases and their required accuracy/efficiency. To this end, we found interesting performance regions which can be used as bounds for the optimization problems arising for any particular application.

Frame structure: The adaptive sequential test requires a one-time ACK feedback to be exchanged between the TX and RX. A practical implementation of SCET is already possible using the available 5G-NR frame structure, where the exchange of feedback is possible via the Random Access Channel (RACH) during the training. Running SCET in a block-wise fashion with a proper hand-shaking protocol (e.g., two-step uplink-downlink training or one step downlink training) and feedback through RACH is a promising approach for a two-sided beam alignment task in different scenarios.

Geometrical priors: The beam elimination mechanism based on estimated a-posteriori probabilities of winning introduces a framework for including prior knowledge on the user positions. This can potentially be used to increase the training efficiency. The sequential nature of this algorithm makes its extension interesting for beam tracking applications as well.

Use of machine-learning: Learning methods with (deep) Artificial Neural Network (ANN) architectures can decrease the loss associated with the wrong beam selection. The ANN can learn the relationships between the beam patterns for each particular set of AoAs and thereby creating more accurate estimates of the coupling coefficients that are used in SCET. This can potentially increase the detection accuracy achieved by the SCET. The next idea involves the use of Reinforcement learning for beam selection. In this way, an agent should be trained to learn when the number of measurements is enough to draw a reliable decision, which is typically denoted as optimal stopping strategy.

Joint communication and Sensing The proposed ideas in this thesis can be of interest in joint communication and sensing scenarios which are gaining ground as a part of envisioned applications in 6G. The variable-length detection framework can be further developed towards the sensing problems where fast and adaptive detection is of greater interest.

Bibliography

[AB07] Amin M. Abbosh and Marek E. Bialkowski. Design of compact directional couplers for uwb applications. *IEEE Transactions on Microwave Theory and Techniques*, 55(2):189–194, 2007.

[AELH14] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE Journal of Selected Topics in Signal Processing*, 8(5):831–846, Oct 2014.

[AH16] Ahmed Alkhateeb and Robert W. Heath. Frequency selective hybrid precoding for limited feedback millimeter wave systems. *IEEE Transactions on Communications*, 64(5):1801–1818, 2016.

[AKM20] Berk Akgun, Marwan Krunz, and David Manzi. Impact of beamforming on delay spread in wideband millimeter-wave systems. In *2020 International Conference on Computing, Networking and Communications (ICNC)*, pages 890–896, 2020.

[ALS+14] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip. Millimeter wave channel modeling and cellular capacity evaluation. *IEEE Journal on Selected Areas in Communications*, 32(6):1164–1179, June 2014.

[AM15] Pierluigi V. Amadori and Christos Masouros. Low rf-complexity millimeter-wave beamspace-mimo systems by beam selection. *IEEE Transactions on Communications*, 63(6):2212–2223, 2015.

[BAM+18] A. A. Boulogeorgos, A. Alexiou, T. Merkle, C. Schubert, R. Elschner, A. Katsiotis, P. Stavrianos, D. Kritharidis, P. Chartsias, J. Kokkoniemi, M. Juntti, J. Lehtomaki, A. Teixeira, and F. Rodrigues. Terahertz technologies to deliver optical network quality of experience in wireless systems beyond 5g. *IEEE Communications Magazine*, 56(6):144–151, 2018.

[BDF+18] Stefano Buzzi, Carmen D'Andrea, Tommaso Foggi, Alessandro Ugolini, and Giulio Colavolpe. Single-carrier modulation versus ofdm for millimeter-wave wireless mimo. *IEEE Transactions on Communications*, 66(3):1335–1348, 2018.

[Bel63] P. Bello. Characterization of randomly time-variant linear channels. *IEEE Transactions on Communications Systems*, 11(4):360–393, 1963.

[BHM+15] C. N. Barati, S. A. Hosseini, M. Mezzavilla, P. Amiri-Eliasi, S. Rangan, T. Korakis, S. S. Panwar, and M. Zorzi. Directional initial access for millimeter wave cellular systems. In *2015 49th Asilomar Conference on Signals, Systems and Computers*, pages 307–311, 2015.

[BHM+16] C. N. Barati, S. A. Hosseini, M. Mezzavilla, T. Korakis, S. S. Panwar, S. Rangan, and M. Zorzi. Initial access in millimeter wave cellular systems. *IEEE Transactions on Wireless Communications*, 15(12):7926–7940, 2016.

[BHR+15] C. N. Barati, S. A. Hosseini, S. Rangan, P. Liu, T. Korakis, S. S. Panwar, and T. S. Rappaport. Directional cell discovery in millimeter wave cellular networks. *IEEE Transactions on Wireless Communications*, 14(12):6664–6678, 2015.

[BT18] S. Bar and J. Tabrikian. A sequential framework for composite hypothesis testing. *IEEE Transactions on Signal Processing*, 66(20):5484–5499, 2018.

[But61] Jesse Butler. Beam-forming matrix simplifies design of electronically scanned antenna. *Electron. Design*, 9:170–173, 1961.

[BV94] C. W. Baum and V. V. Veeravalli. A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, 40(6):1994–2007, 1994.

[BVBL19] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson. Massive mimo in sub-6 ghz and mmwave: Physical, practical, and use-case differences. *IEEE Wireless Communications*, 26(2):100–108, 2019.

[CCR+21] Hsiao-Lan Chiang, Kwang-Cheng Chen, Wolfgang Rave, Mostafa Khalili Marandi, and Gerhard Fettweis. Machine-learning beam tracking and weight optimization for mmwave multi-uav links. *IEEE Transactions on Wireless Communications*, 20(8):5481–5494, 2021.

[COS18] P. L. Cao, T. J. Oechtering, and M. Skoglund. Precoding design for massive mimo systems with sub-connected architecture and per-antenna power constraints. In *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, pages 1–6, 2018.

[CRKF17] H. Chiang, W. Rave, T. Kadur, and G. Fettweis. A low-complexity beamforming method by orthogonal codebooks for millimeterwave links. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3375–3379, 2017.

[CRKF18] H. Chiang, W. Rave, T. Kadur, and G. Fettweis. Hybrid beamforming based on implicit channel state information for millimeter wave links. *IEEE Journal of Selected Topics in Signal Processing*, 12(2):326–339, 2018.

[CRT01] N. Chiurtu, B. Rimoldi, and E. Telatar. On the capacity of multi-antenna gaussian channels. In *Proceedings. 2001 IEEE International Symposium on Information Theory (IEEE Cat. No.01CH37252)*, pages 53–, 2001.

[CW11] T. T. Cai and L. Wang. Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Transactions on Information Theory*, 57(7):4680–4688, 2011.

[DBN+20] Martin Danneberg, Roberto Bomfin, Ahmad Nimr, Zhongju Li, and Gerhard Fettweis. Usrp-based platform for 26/28 ghz mmwave experimentation. In *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 1–6, 2020.

[DSK+22] K. Dehkordi, O. Schwanitz, M. Khalili Marandi, T. H. Le, M. P. Kaiser, I. Ndip, and G. Caire. Sub-thz layout optimization under fabrication constraints. In *2022 IEEE European Conference on Networks and Communications (EUCNC) and 6G Summit*, 2022.

[DSM+22] Saeid K. Dehkordi, Oliver Schwanitz, Mostafa Khalili Marandi, Michael P. Kaiser, Thi Huyen Le, Ivan Ndip, and Giuseppe Caire. Sub-terahertz modular array layout optimization under fabrication constraints. In *2022 Joint European Conference on Networks and Communications 6G Summit (EuCNC/6G Summit)*, pages 31–36, 2022.

[DTV99] V. P. Draglia, A. G. Tartakovsky, and V. V. Veeravalli. Multihypothesis sequential probability ratio tests .i. asymptotic optimality. *IEEE Transactions on Information Theory*, 45(7):2448–2461, 1999.

[DXS+18] J. Du, W. Xu, H. Shen, X. Dong, and C. Zhao. Hybrid precoding architecture for massive multiuser mimo with dissipation: Sub-connected or fully connected structures? *IEEE Transactions on Wireless Communications*, 17(8):5465–5479, 2018.

[FA14a] Hatem Fayed and Amir Atiya. An evaluation of the integral of the product of the error function and the normal probability density with application to the bivariate normal integral. *Mathematics of Computation*, 83(285):235–250, 2014.

[FA14b] Gerhard Fettweis and Siavash Alamouti. 5g: Personal mobile internet beyond what cellular did to telephony. *IEEE Communications Magazine*, 52(2):140–145, 2014.

[GDC+16] Xinyu Gao, Linglong Dai, Zhijie Chen, Zhaocheng Wang, and Zhijun Zhang. Near-optimal beam selection for beamspace mmwave massive mimo systems. *IEEE Communications Letters*, 20(5):1054–1057, 2016.

[Gol05] Andrea Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[HBA15] Chong Han, A. Ozan Bicen, and Ian F. Akyildiz. Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band. *IEEE Transactions on Wireless Communications*, 14(5):2402–2412, 2015.

[HCGL18] T. Haelsig, D. Cvetkovski, E. Grass, and B. Lankl. Statistical properties and variations of los mimo channels at millimeter wave frequencies. In *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, pages 1–6, 2018.

[HM19] M. Hussain and N. Michelusi. Energy-efficient interactive beam alignment for millimeter-wave networks. *IEEE Transactions on Wireless Communications*, 18(2):838–851, 2019.

[HWF+17] Jie Huang, Cheng-Xiang Wang, Rui Feng, Jian Sun, Wensheng Zhang, and Yang Yang. Multi-frequency mmwave massive mimo channel measurements and characterization for 5g wireless communication systems. *IEEE Journal on Selected Areas in Communications*, 35(7):1591–1605, 2017.

[IE17] Shahar Stein Ioushua and Yonina C Eldar. Hybrid analog-digital beamforming for massive mimo systems. *arXiv preprint arXiv:1712.03485*, 2017.

[JSRF20] Christoph Jans, Xiaohang Song, Wolfgang Rave, and Gerhard Fettweis. Frequency-selective analog beam probing for millimeter wave communication systems. In *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2020.

[JWNC19] Furqan Jameel, Shurjeel Wyne, Syed Junaid Nawaz, and Zheng Chang. Propagation channels for mmwave vehicular communications: State-of-the-art and future research directions. *IEEE Wireless Communications*, 26(1):144–150, 2019.

[Kaya] Steven M Kay. *Fundamentals of statistical signal processing, Vol. I Estimation Theory*, volume I.

[Kayb] Steven M Kay. *Fundamentals of statistical signal processing, Vol. II Detection Theory*, volume I.

[KBK02] V Kühn, Ronald Böhnke, and Karl-Dirk Kammeyer. Multi-user detection in multicarrier-cdma systems. *e&i Elektrotechnik und Informationstechnik*, 119(11):395–402, 2002.

[KCW+17] M. Kokshoorn, H. Chen, P. Wang, Y. Li, and B. Vucetic. Millimeter wave mimo channel estimation using overlapped beam patterns and rate adaptation. *IEEE Transactions on Signal Processing*, 65(3):601–616, Feb 2017.

[KMJRG21] Mostafa Khalili Marandi, Christoph Jans, Wolfgang Rave, and Fettweis Gerhard. Evaluation of detection accuracy and efficiency of considered beam alignment strategies for mmwave massive mimo systems. In *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021.

[KMRF19] Mostafa Khalili Marandi, Wolfgang Rave, and Gerhard Fettweis. Multi user beam selection using sequential competition test. In *WSA 2019; 23rd International ITG Workshop on Smart Antennas*, pages 1–4, 2019.

[KMWK+22] Mostafa Khalili Marandi, Shizhang Wei, Behnam Khodapanah, Wolfgang Rave, and Gerhard Fettweis. Experimental evaluation of a variable length beam selection framework in a usrp based testbed with mmwave frontends and butler matrices. In *2022 IEEE International Symposium on Phased Array Systems and Technology*. IEEE, 2022.

[Kuh03] V. Kuhn. Iterative interference cancellation and channel estimation for coded ofdm-cdma. In *IEEE International Conference on Communications, 2003. ICC '03.*, volume 4, pages 2465–2469 vol.4, 2003.

[Kuh06] Volker Kuhn. *Wireless communications over MIMO channels: applications to CDMA and multiple antenna systems*. John Wiley & Sons, 2006.

[LL16] Cen Lin and Geoffrey Ye Li Li. Terahertz communications: An array-of-subarrays solution. *IEEE Communications Magazine*, 54(12):124–131, 2016.

[LM17] A. Li and C. Masouros. Hybrid analog-digital millimeter-wave mu-mimo transmission with virtual path selection. *IEEE Communications Letters*, 21(2):438–441, 2017.

[LWR14] Qian Clara Li, Geng Wu, and Theodore S Rappaport. Channel model for millimeter-wave communications based on geometry statistics. In *2014 IEEE Globecom Workshops (GC Wkshps)*, pages 427–432. IEEE, 2014.

[MMS+10] Alexander Maltsev, Roman Maslennikov, A Sevastyanov, A Lomayev, and Alexey Khoryaev. Statistical channel model for 60 ghz wlan systems in conference room environment. In *Proceedings of the Fourth European Conference on Antennas and Propagation*, pages 1–5. IEEE, 2010.

[MRF18] M. K. Marandi, W. Rave, and G. Fettweis. An adaptive sequential competition test for beam selection in massive mimo systems. In *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 104–108, 2018.

[MRF19] M. K. Marandi, W. Rave, and G. Fettweis. Beam selection based on sequential competition. *IEEE Signal Processing Letters*, 26(3):455–459, March 2019.

[MRF20] M. K. Marandi, W. Rave, and G. Fettweis. Beam elimination based on sequentially estimated a posteriori probabilities of wining. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[NP33] Jerzy Neyman and Egon Sharpe Pearson. Ix. on the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337, 1933.

[NZL17a] S. Noh, M. D. Zoltowski, and D. J. Love. Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment. *IEEE Transactions on Wireless Communications*, 16(9):5689–5701, Sep. 2017.

[NZL17b] Song Noh, Michael D. Zoltowski, and David J. Love. Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment. *IEEE Transactions on Wireless Communications*, 16(9):5689–5701, 2017.

[PDW17] J. Palacios, D. De Donno, and J. Widmer. Tracking mm-wave channel dynamics: Fast beam training strategies under mobility. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, May 2017.

[PR06] Pooi Yuen Kam and Rong Li. A new geometric view of the first-order marcum q-function and some simple tight erfc-bounds. In *2006 IEEE 63rd Vehicular Technology Conference*, volume 5, pages 2553–2557, 2006.

[PS01] John G Proakis and Masoud Salehi. *Digital communications*, volume 4. McGraw-hill New York, 2001.

[Red01] Sidney Redner. *A guide to first-passage processes*. Cambridge University Press, 2001.

[RM19] W. Rave and M. K. Marandi. The elimination game or: Beam selection based on m-ary sequential competition elimination. In *WSA 2019; 23rd International ITG Workshop on Smart Antennas*, pages 1–8, April 2019.

[RMSS15] Theodore S. Rappaport, George R. MacCartney, Mathew K. Samimi, and Shu Sun. Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design. *IEEE Transactions on Communications*, 63(9):3029–3056, 2015.

[RXK+19a] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos. Wireless communications and applications above 100 ghz: Opportunities and challenges for 6g and beyond. *IEEE Access*, 7:78729–78757, 2019.

[RXK+19b] Theodore S. Rappaport, Yunchou Xing, Ojas Kanhere, Shihao Ju, Arjuna Madanayake, Soumyajit Mandal, Ahmed Alkhateeb, and Georgios C. Trichopoulos. Wireless communications and applications above 100 ghz: Opportunities and challenges for 6g and beyond. *IEEE Access*, 7:78729–78757, 2019.

[SAAN21] Hadi Sarieddeen, Mohamed-Slim Alouini, and Tareq Y. Al-Naffouri. An overview of signal processing techniques for terahertz communications. *Proceedings of the IEEE*, 109(10):1628–1665, 2021.

[SB13] Akbar Sayeed and John Brady. Beamspace mimo for high-dimensional multiuser communication at millimeter-wave frequencies. In *2013 IEEE Global Communications Conference (GLOBECOM)*, pages 3679–3684, 2013.

[Sch58] B.M. Schiffman. A new class of broad-band microwave 90-degree phase shifters. *IRE Transactions on Microwave Theory and Techniques*, 6(2):232–237, 1958.

[SHC18] X. Song, S. Haghighatshoar, and G. Caire. A scalable and statistically robust beam alignment technique for millimeter-wave systems. *IEEE Transactions on Wireless Communications*, 17(7):4792–4805, 2018.

[SHC19a] X. Song, S. Haghighatshoar, and G. Caire. Efficient beam alignment for millimeter wave single-carrier systems with hybrid mimo transceivers. *IEEE Transactions on Wireless Communications*, 18(3):1518–1533, 2019.

[SHC19b] Xiaoshen Song, Saeid Haghighatshoar, and Giuseppe Caire. Efficient beam alignment for millimeter wave single-carrier systems with hybrid mimo transceivers. *IEEE Transactions on Wireless Communications*, 18(3):1518–1533, 2019.

[Sie02] P. H. Siegel. Terahertz technology. *IEEE Transactions on Microwave Theory and Techniques*, 50(3):910–928, 2002.

[SKC20] X. Song, T. Kühne, and G. Caire. Fully-/partially-connected hybrid beamforming architectures for mmwave mu-mimo. *IEEE Transactions on Wireless Communications*, 19(3):1754–1769, 2020.

[SMB01] Martin Steinbauer, Andreas F Molisch, and Ernst Bonek. The double-directional radio channel. *IEEE Antennas and propagation Magazine*, 43(4):51–63, 2001.

[SR15] Mathew K Samimi and Theodore S Rappaport. Statistical channel model with multi-frequency and arbitrary antenna beamwidth for millimeter-wave outdoor communications. In *2015 IEEE Globecom Workshops (GC Wkshps)*, pages 1–7. IEEE, 2015.

[SR16a] M. K. Samimi and T. S. Rappaport. 3-d millimeter-wave statistical channel model for 5g wireless system design. *IEEE Transactions on Microwave Theory and Techniques*, 64(7):2207–2225, 2016.

[SR16b] Mathew K Samimi and Theodore S Rappaport. 3-d millimeter-wave statistical channel model for 5g wireless system design. *IEEE Transactions on Microwave Theory and Techniques*, 64(7):2207–2225, 2016.

[SRB+18] Xiaohang Song, Wolfgang Rave, Nithin Babu, Sudhan Majhi, and Gerhard Fettweis. Two-level spatial multiplexing using hybrid beamforming for millimeter-wave backhaul. *IEEE Transactions on Wireless Communications*, 17(7):4830–4844, 2018.

[SY16] F. Sohrabi and W. Yu. Hybrid digital and analog beamforming design for large-scale antenna arrays. *IEEE Journal of Selected Topics in Signal Processing*, 10(3):501–513, 2016.

[TV05] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[V14] 3GPP TR 38.900 V14.3.1. Study on channel model for frequency spectrum above 6 ghz (release 14).

[VAGH17] K. Venugopal, A. Alkhateeb, N. González Prelcic, and R. W. Heath. Channel estimation for hybrid architecture-based wideband millimeter wave systems. *IEEE Journal on Selected Areas in Communications*, 35(9):1996–2009, 2017.

[Wal44] Abraham Wald. On cumulative sums of random variables. *The Annals of Mathematical Statistics*, 15(3):283–296, 1944.

[Wal45] Abraham Wald. Sequential tests of statistical hypotheses. *The annals of mathematical statistics*, 16(2):117–186, 1945.

[WBKK03] D. Wubben, R. Bohnke, V. Kuhn, and K.-D. Kammeyer. Mmse extension of v-blast based on sorted qr decomposition. In *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No.03CH37484)*, volume 1, pages 508–512 Vol.1, 2003.

[WBKK04] Dirk Wubben, Ronald Bohnke, Volker Kuhn, and K-D Kammeyer. Near-maximum-likelihood detection of mimo systems using mmse-based lattice-reduction. In *2004 IEEE International Conference on Communications (IEEE Cat. No. 04CH37577)*, volume 2, pages 798–802. IEEE, 2004.

[WHR+11] Shurjeel Wyne, Katsuyuki Haneda, Sylvain Ranvier, Fredrik Tufvesson, and Andreas F. Molisch. Beamforming effects on measured mm-wave channel characteristics. *IEEE Transactions on Wireless Communications*, 10(11):3553–3559, 2011.

[WLP+18] X. Wang, M. Laabs, D. Plettemeier, H. Chiang, M. K. Marandi, G. P. Fettweis, K. Kosaka, and Y. Matsunaga. 28 ghz multi-beam antenna array based on a compact wideband 8×8 butler matrix. In *2018 IEEE International Symposium on Antennas and Propagation USNC/URSI National Radio Science Meeting*, pages 2177–2178, 2018.

[WLP+19a] Xiaozhou Wang, Martin Laabs, Dirk Plettemeier, Keishi Kosaka, and Yasuhiko Matsunaga. 28 ghz multi-beam antenna array based on wideband high-dimension 16x16 butler matrix. In *2019 13th European Conference on Antennas and Propagation (EuCAP)*, pages 1–4, 2019.

[WLP+19b] Xiaozhou Wang, Martin Laabs, Dirk Plettermeier, Keishi Kosaka, and Yasuhiko Matsunaga. Mimo antenna array system with integrated 16x16 butler matrix and power amplifiers for 28ghz wireless communication. In *2019 12th German Microwave Conference (GeMiC)*, pages 127–130, 2019.

[XMH+17] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C. I, and A. Ghosh. Millimeter wave communications for future mobile networks. *IEEE Journal on Selected Areas in Communications*, 35(9):1909–1935, 2017.

[XR21] Yunchou Xing and Theodore S. Rappaport. Millimeter wave and terahertz urban microcell propagation measurements and models. *IEEE Communications Letters*, 25(12):3755–3759, 2021.

[ZHS+17] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang. Codebook design for beam alignment in millimeter wave communication systems. *IEEE Transactions on Communications*, 65(11):4980–4995, 2017.